A PRIMER ON NONMARKET VALUATION

THE ECONOMICS OF NON-MARKET GOODS AND RESOURCES

VOLUME 3

Series Editor: Dr. Ian J. Bateman

Dr. Ian J. Bateman is Professor of Environmental Economics at the School of Environmental Sciences, University of East Anglia (UEA) and directs the research theme Innovation in Decision Support (Tools and Methods) within the Programme on Environmental Decision Making (PEDM) at the Centre for Social and Economic Research on the Global Environment (CSERGE), UEA. The PEDM is funded by the UK Economic and Social Research Council. Professor Bateman is also a member of the Centre for the Economic and Behavioural Analysis of Risk and Decision (CEBARD) at UEA and Executive Editor of Environmental and Resource Economics, an international journal published in cooperation with the European Association of Environmental and Resource Economists (EAERE).

Aims and Scope

The volumes which comprise *The Economics of Non-Market Goods and Resources* series have been specially commissioned to bring a new perspective to the greatest economic challenge facing society in the 21^{st} Century; the successful incorporation of non-market goods within economic decision making. Only by addressing the complexity of the underlying issues raised by such a task can society hope to redirect global economies onto paths of sustainable development. To this end the series combines and contrasts perspectives from environmental, ecological and resource economics and contains a variety of volumes which will appeal to students, researchers, and decision makers at a range of expertise levels. The series will initially address two themes, the first examining the ways in which economists assess the value of non-market goods, the second looking at approaches to the sustainable use and management of such goods. These will be supplemented with further texts examining the fundamental theoretical and applied problems raised by public good decision making.

A Primer on Nonmarket Valuation

Edited by

Patricia A. Champ

U.S. Forest Service,
Rocky Mountain Research Station, Fort Collins, CO, U.S.A.

Kevin J. Boyle

Libra Professor of Environmental Economics,
Department of Resource Economics & Policy, University of Maine, U.S.A.

and

Thomas C. Brown

U.S. Forest Service,
Rocky Mountain Research Station, Fort Collins, CO, U.S.A.

SPRINGER SCIENCE+BUSINESS MEDIA, LLC

A C.I.P. Catalogue record for this book is available from the Library of Congress.

ISBN 978-1-4020-1445-1 ISBN 978-94-007-0826-6 (eBook)
DOI 10.1007/978-94-007-0826-6

Printed on acid-free paper

Contents

Contributing Authors

Vic Adamowicz, Department of Rural Economy, University of Alberta, Edmonton, Alberta T6G 2H1, CANADA

Richard C. Bishop, Dept. of Agricultural and Applied Economics, University of Wisconsin, 427 Lorch Street, Madison, WI 53706, USA

Kevin J. Boyle, Department of Resource Economics & Policy, University of Maine, 5782 Winslow Hall, Orono, ME 04469-5782, USA

Thomas C. Brown, U.S. Forest Service, Rocky Mountain Research Station, 2150 Centre Avenue, Bldg A, Fort Collins, CO 80526, USA

Patricia A. Champ, U.S. Forest Service, Rocky Mountain Research Station, 2150 Centre Avenue, Bldg A, Fort Collins, CO 80526, USA

Mark Dickie, Department of Economics, College of Business Administration, University of Central Florida, P.O. Box 161400, Orlando, FL 32816-1400, USA

Nicholas E. Flores, Department of Economics, University of Colorado,Economics 115, Campus Box 256, Boulder, CO 80309-0256, USA

Myrick Freeman, III, Department of Economics, Bowdoin College, Brunswick, ME 04011, USA

Thomas P. Holmes, U.S. Forest Service, Southern Research Station, P.O. Box 12254, Research Triangle Park, NC 27709, USA

John B. Loomis, Department of Agricultural and Resource Economics, Colorado State University, Fort Collins, CO 80523-1172, USA

Daniel W. McCollum, U.S. Forest Service, Rocky Mountain Research Station, 2150 Centre Avenue, Bldg A, Fort Collins, CO 80526, USA

George R. Parsons, College of Marine Studies, University of Delaware, Robinson Hall, Newark, DE 19716-3501, USA

George L. Peterson, U.S. Forest Service, Rocky Mountain Research Station, 2150 Centre Avenue, Bldg A, Fort Collins, CO 80526, USA

Randall Rosenberger, Department of Forest Resources, Oregon State University, 280 Peavy Hall, Corvallis, OR 97331-5703, USA

Laura Taylor, Department of Economics, School of Public Policies, Georgia State University, Atlanta, GA 30303-3083, USA

Preface

Public policy should reflect an understanding of the public's values. This is especially true with respect to the environment. Because public values about the environment are not generally expressed in the marketplace, nonmarket valuation has become an important source of information for environmental decisionmaking. While interest has grown in measuring nonmarket values, the technical nature of the nonmarket valuation literature can overwhelm newcomers. This book provides clear descriptions of the most commonly used nonmarket valuation techniques along with the steps for implementation. Individuals working for natural resource agencies, attorneys and consultants involved with natural resource damage assessments, graduate students, and others will appreciate the practical tone of this book.

Our approach for developing this primer was to invite experienced, well-respected experts to write the chapters. The next step was to get the authors together to work toward a well-integrated book. The Rocky Mountain Research Station of the U.S. Forest Service organized two author workshops. At the first meeting each contributor presented an outline of his or her chapter; as a group we critiqued the outlines and worked to define a consistent structure that would be used to write all the chapters. At the second meeting each author presented his or her draft chapter and another author in the group served as a discussant. This process facilitated development of a cohesive text and provided peer review of all the chapters.

The first section of the book, containing three chapters, provides background for the methodology sections that follow. The opening chapter explains the policy context of nonmarket valuation. The next chapter develops the economic theory that is the basis of nonmarket values. And because value estimates are only as good as the data upon which they are based, the third chapter describes the process of collecting nonmarket valuation data.

The valuation methods are presented in two sections, one section on stated preference methods and the other on revealed preference methods. Following a short introduction chapter, the section on stated preference methods includes three chapters—two covering widely used nonmarket valuation techniques (contingent valuation and attribute-based methods) and the other describing an emerging technique (the method of paired comparisons). The revealed preference section also includes an introduction chapter and three methods chapters. The methods chapters covered – the travel cost method, the hedonic method, and the defensive behavior and damage cost methods – each use "clues" from behaviors observed in markets to estimate nonmarket values. We expect readers will come away from these two sections with a thorough understanding of how to design and implement a nonmarket valuation study. The chapters also describe how to analyze the data and estimate values.

The final section of the book takes stock of the usefulness of nonmarket valuation. The first chapter in this section describes techniques for transferring values from existing studies to new situations. The next chapter discusses interesting situations in which nonmarket values were used to help make important natural resource and environmental decisions. The final chapter provides thoughts on the history, validity, and future of nonmarket valuation.

As a companion to this book, we developed a website containing datasets, nonmarket valuation surveys, and links to publications based on the data (www.fs.fed.us/nonmarketprimerdata). This website should aid readers who are curious about the format of nonmarket valuation surveys and the resulting data. The data on the website can be used to replicate statistical models reported in the linked publications.

We wish to express our gratitude for the help and support provided by the following people. This book has benefited greatly from the word processing skills of Kim Junkins. The companion website was designed and put together by Pam Froemke. Alexandar Bujak carefully read and provided comments on every chapter for us. Fred Kaiser attended our second workshop and motivated all of us with his agency perspective on why this book is important.

Finally, we dedicate this book to George Peterson. After a distinguished teaching career at Northwestern University, George became a research project leader with the Rocky Mountain Research Station in Fort Collins, Colorado. Professionally, he is known as an award-winning professor, an innovative researcher, and an outstanding project leader. Among those of us fortunate enough to work with George, he is considered a fountain of wisdom and a dear friend.

Patricia Champ, Kevin Boyle, and Thomas Brown

Chapter 1

ECONOMIC VALUATION: WHAT AND WHY

A. Myrick Freeman III
Bowdoin College

1. THE MANAGEMENT PROBLEM

Consider the choices faced by those responsible for managing releases of water at the Glen Canyon Dam on the Colorado River. The size and timing of the releases determine the value of the hydroelectric power generated. But the pattern of the releases influences the nature and value of the recreational activities on Lake Powell above the dam and on the rivers in Glen Canyon and Grand Canyon below the dam. The release patterns also affect, in complex ways, the availability of suitable habitat for and the probabilities of the survival of several endangered species. Thus, managers face choices that involve trade-offs.[1] More hydroelectric power means reduced quality and quantity of recreation opportunities and, perhaps, reductions in the survival of some endangered species. It is necessary to reconcile the sometimes conflicting desires and interests of those affected by management decisions and to make wise choices concerning these trade-offs. To do so, resource managers require information on the value of hydroelectric power and the values that people place on various attributes and types of recreation on and along the Colorado River and on changes in the survival probabilities for endangered species.[2]

We live in a world of scarcity and thus, we must make choices about how to manage the human impact on natural systems. Greater use of a particular environmental service or greater protection of a specific natural system results in less of something else. This is the trade-off that we must accept. To make the most of scarce resources, we must compare what is gained from an activity with what is sacrificed by undertaking that activity. To do this, we need a way to assess the net impacts of policy changes on human well-being to choose the best mix (in terms of the contribution to human well-being) of service flows from the environment.

Natural resource systems, such as forests and commercially exploitable fisheries, and environmental systems, such as air sheds and watersheds, are valuable assets because they yield flows of valuable services to people (Smith 1988; Freeman 1993; Daily et al. 2000). A forest included in the U.S. National Forest System is an example of a resource-environmental system that provides a wide range of services including materials, such as wood and fiber and forage for cattle and sheep, and the amenities associated with a variety of outdoor recreation activities. A forest regulates stream flow, controls erosion, absorbs atmospheric carbon dioxide, and provides habitat for a variety of species, thus, protecting biodiversity. Increasing the flow of timber can lead to a decrease in the amenity values associated with mature forests and an increase in the loss of habitat for forest dwelling species. There is also likely to be an increase in runoff and erosion and degradation of aquatic habitat that is downstream from harvested areas. Similarly, an estuary and its adjacent upland provide services such as support for commercially exploitable fisheries, space for residential, industrial, and commercial structures, and the absorption of waste products from local runoff and from upstream sources along rivers and streams. An estuary also provides amenity services associated with a variety of recreation activities including fishing, boating, and bird watching (Freeman 1995).

The natural resource-environmental complex produces four direct service flows to people. First, as in the conventional view of resource economics, the natural environment is a source of materials inputs, such as fossil fuels, wood products, minerals, water, and fish, to the economy. Second, some components of the resource-environment system provide life-support services such as a breathable atmosphere and a liveable climatic regime. Third, the resource-environment system provides a wide variety of amenity services including recreation, wildlife observation, scenic views, and services not related to direct use (passive use values). Lastly, the environment disperses, transforms, and stores the waste products of economic activity (Kneese, Ayres, and d'Arge 1970; and Freeman, Haveman, and Kneese 1973).

Many of the services provided by natural resource-environmental systems can be characterized as direct services since their benefits accrue directly to people; for example, materials flows and life-support services. Other environmental services could be better described as indirect services because they support other biological and ecological production processes that yield value to people. Examples include nutrient recycling, organic material decomposition, soil fertility generation and renewal, crop and natural vegetation pollination, and biological control of agricultural pests.[3] Indirect services might also be termed intermediate service flows.

Environmental services from a resource-environmental system include some that are joint products because increasing or decreasing the flow of one simultaneously involves increasing or decreasing the flow of the other. This means that the resource-environmental system is characterized by scarcity, opportunity costs, and the necessity for trade-offs.

Some of the service flows of resource-environmental systems are linked directly to markets and hence, are responsive to market forces. But many service flows are not properly regulated by markets because of their public goods characteristics of nonexcludability and nondepletability, externalities, and other factors. This means that a decentralized market system is unlikely to lead to the optimal pattern of service flows, that is, there will be market failure. Therefore, there is a potential for public policy in the management of resource-environmental systems and a need for information on the values of these service flows.

In its most fundamental form, the environmental management problem faced by society is to choose the mix of environmental and resource service flows that is consistent with the highest possible level of human well-being, that is, the mix with the highest aggregate value to people. Given the inability of a decentralized market system to achieve this welfare-maximizing outcome, there is a role for public intervention and resource management. Resource managers will require information on the values of the nonmarket service flows, and how they change, for all the services affected by the management choices. This need initiates a host of market and nonmarket valuation issues.

The objective of this book is to provide environmental valuation practitioners with a practical guide to the state of the art of estimating the values of nonmarket environmental services to help guide environmental and resource management policies. In this chapter, I will provide background on the role of economic valuation in policy evaluation and on the nature, definition, and measurement of economic values.

2. TYPES OF ENVIRONMENTAL SERVICES

In this section, I describe some of the ways to classify the types of environmental services that might be the subject of an economic valuation exercise. Since there are many ways of classifying a set of objects with multiple attributes, any classification system or taxonomy must be evaluated on the basis of its usefulness.

One basis for classification is the type of resource or environmental media. Environmental effects are often classified according to whether they stem from changes in air quality, water quality, land quality, and so forth. Although the current legal and administrative divisions of responsibilities for environmental management and pollution control are consistent with this basis for classification, they are becoming less relevant as cross-media effects are becoming better understood. For example, controlling emissions of nitrogen oxides from coal burning power plants might be part of a cost-effective strategy for improving water quality in estuaries because of the impact of nitrate deposition on nutrient levels in the waters. Additionally, controls on land use are a key part of the strategy for reducing nonpoint source water pollution.

A second type of classification is based on the economic channel through which human well-being is affected. Environmental and resource service flows can be classified in two ways. One manner is according to whether the flows convey their effects through the market system in the form of changes in incomes to producers and changes in the availability of and prices for marketed goods and services to consumers. The second method is through changes in the availability of goods and services not usually purchased through markets; for example, health, environmental amenities, such as visibility, and outdoor recreation opportunities. The subject of this book is the methods and techniques for measuring the values of these latter nonmarket services. However, many of the policies for managing environmental and resource systems will affect the flows of both market and nonmarket goods and services. So policy assessments should consider market and nonmarket valuation methods. See Chapter 9 in Freeman (1993) for an overview of market valuation methods.

A third way of classifying environmental and resource service flows is according to whether they impinge directly on humans or indirectly on humans either by affecting other living organisms or inanimate systems. Direct impacts on humans include morbidity and mortality effects associated with air and water pollutants, hazardous wastes, pesticide residues, and the like, and the non-health effects of pollutants such as odors, reduced visibility, and reduced visual attractiveness of outdoor settings.

Indirect impacts on humans involving biological mechanisms and other organisms include the economic productivity of both managed and natural ecosystems such as agricultural crop lands, commercial forests, and commercial fisheries. Market valuation methods are used to value these effects. There are also impacts on nonmarket direct service flows to people such as recreational uses of ecosystems for hunting, fishing, and nature observation. Additionally, there are impacts on indirect or intermediate ecosystem services such as pollination, decomposition, biological pest control, and nutrient recycling.

Indirect impacts on humans involving nonliving systems include: damages to materials and structures and increases in cleaning and repair costs of producers and firms, which would be measured by market valuation techniques; damages to materials and structures and increases in cleaning and repair costs for households, which would be measured by nonmarket valuation techniques; and impacts on weather and climate, which would be measured by either market or nonmarket valuation techniques, depending on the nature of the activity affected.

Finally, we can distinguish between services that individuals value because they make use of them in some way (use values) and those that people value independent of any kind of observable use (nonuse values or passive use values). Questions of defining use and use versus passive use values are discussed in Chapter 2.

3. POLICY EVALUATION - AN OVERVIEW

Consider the case of a public decision maker who must choose among alternative policies such as the flow management options discussed in section 1. The choices will affect large numbers of people, making some better off and others worse off to varying degrees. Presumably, the decision maker wishes to choose policies that produce the largest possible increases in social well-being or welfare. When decision makers choose from alternative policies, they should identify the favorable (beneficial) and adverse (costly) consequences associated with each alternative and choose the preferred option.

However, to determine the preferred option, the decision maker must consider three key issues. The first issue is what basis does the decision maker have for evaluating alternatives; that is, what is the criterion for determining the relative degree of preferredness of the alternatives being considered? The second issue arises when a policy affects more than one thing that matters to the decision maker. How are the relative weights assigned to the different favorable and adverse consequences so that they are commensurate and summarized in a single measure of preferredness? The third issue concerns the aggregation of favorable and adverse effects across people; that is, how to assess the net effect on the social welfare of a policy that makes some people better off and others worse off. These issues are addressed in this and the following three sections of this chapter.

There are four basic steps involved in the evaluation of environmental and resource management policies. The first is to specify the criterion for evaluation. This criterion must provide a resolution of the three issues described above; that is, the issues of identifying a basis for determining what is to be preferred, making unlike effects commensurate, and considering aggregation across individuals when some are made better off while others are made worse off. The second step is to identify and describe the alternatives to be evaluated including the option of doing nothing. Since any policy evaluation is a comparison of at least two alternatives, it is important to be clear and specific about the alternatives. The third step is to predict the effects of each of the policy alternatives on the variables of concern. Finally, the fourth step is to convert the quantified effects into commensurate measures of value, and then to aggregate them. Each of these steps is discussed in more detail.

3.1 The Criterion

The terms "benefits" and "costs" are meaningless without some social objective function or social welfare function, which defines good and bad consequences. The criterion adopted by economists for policy evaluation is the well-being of the members of society, where well-being is defined as the individuals'preferences and their willingness to pay for gains or to accept compensation for losses. The methodology of economic valuation is the set of analytical tools designed to measure the net contribution that any public policy makes to the economic well-being of each of the members of society. It describes the rules and procedures used to conduct a narrowly defined, technical economic calculation to reduce all beneficial and adverse effects that each policy alternative has on individuals to a common monetary measure. In sections 5 and 6, the questions of how economists have dealt with the issues of defining and estimating commensurate measures of value and aggregating effects across individuals is addressed.

3.2 Specifying Alternatives

Each alternative can be described as a set of activities affecting the resource-environmental system in question. The kinds of activities included in the set could be the rates of harvest or exploitation of the resource, the rates of investment or commitment of resources to managing or improving the resource, the rates of discharge of various pollutants into the environment, and so forth. The kinds of activities in each policy alternative must be described in sufficient detail to provide a basis for analysis. For example, it is insufficient to describe a policy as involving cleaner waters. Rather, the policy must specify which waters, how clean they will be, and what actions will be necessary to achieve the specified targets.

The purpose of undertaking a benefit-cost analysis is to determine whether a proposed governmental intervention is good in economic terms. A policy is considered good only if the economic outcome after p□licy implementation is preferred, in some sense, to the outcome in the absence of the policy. Thus, the benefit-cost analysis can be viewed as providing a comparison of the outcomes of two alternative scenarios: the world with the policy in place and the status quo or the world without the policy, other things being equal. In situations where there are more than one active intervention or policy under consideration, then the best policy is the one preferred to all of the other alternatives, including the alternative of doing nothing.

3.3 Quantifying the Effects

In principle a benefit-cost analysis involves a comparison of the well-being of all of the affected parties in the two different states of the world; that is, with the project and without the project. This requires having models capable of predicting the magnitudes of all of the things that will affect individuals' well-being in the two states of the world. For example, if the policy alternative is a specified temporal pattern of releases of water from a hydroelectric dam, it is necessary to model the time varying demand for electricity, since this will help to determine the value of the electricity produced by the dam. It is also necessary to model the effects of the water releases on the downstream hydrology including rates of flow, water levels and temperatures, and sediment transportation and deposition. Then, it is necessary to model how various species of plants and animals would be affected by the predicted changes in the hydrology. Finally, it is necessary to model how the various uses that people make of the downstream areas are affected by these changes. For example, the policy could result in improved opportunities for camping and white water rafting in the downstream area but a decrease in the quality of a recreational fishery. Much of the required modeling is non-economic in nature. Nevertheless, it provides the necessary foundation for economic valuation.

3.4 Commensurate Measures of Value

The basis for defining and measuring the values of benefits and costs must be consistent with the criterion selected for policy evaluation. In the next two sections of this chapter, I explain the concept of economic value, how it is derived from the underlying economic criterion for policy evaluation, and the basis for measuring economic values.

4. VALUE AS AN ECONOMIC CONCEPT

The word value has several different meanings . For example, economists and ecologists use the word in two different ways in discussions of environmental services and ecosystems. Ecologists typically use the word to mean "that which is desirable or worthy of esteem for its own sake; thing or quality having intrinsic worth" (Webster's New World Dictionary, Third College Edition). Economists use the word to mean "a fair or proper equivalent in money, commodities, etc." (Webster's New World Dictionary, Third College Edition), where "equivalent in money" represents the sum of money that would have an equivalent effect on the welfare or utilities of individuals.

These two different uses of the word correspond to a distinction made by philosophers between intrinsic value and instrumental value. According to philosophers, something has intrinsic value "if it is valuable in and for itself -- if its value is not derived from its utility, but is independent of any use or function it may have in relation to something or someone else. . . . an intrinsically valuable entity is said to be an 'end-in-itself,' not just a 'means' to another's ends" (Callicott 1989, p.131). In contrast, something has instrumental value if it is valued as a means to some other end or purpose. In this view, the value of something lies in its contribution to some other goal (Costanza and Folke 1997).

Some people have argued that nature has intrinsic value for various reasons, including because of its harmony or natural balance. But from the perspective of the new ecology, which emphasizes disturbance and change in ecosystems (for example, Botkin 1990), this justification of an intrinsic value in nature is very problematic. A conservation biologist might argue that the part of nature consisting of the variety of organisms, their interactions, and their genetic diversity has intrinsic value. But this view does not endow any particular manifestation of nature with any more or less intrinsic value than some alternative manifestation. Nature's value is preserved as long as diversity, in the broad sense, is preserved. Although the concept of intrinsic value as applied to the environment is attractive in many respects, it does not provide a basis for dealing with the kinds of environmental management questions that were identified in the first section of this chapter. In contrast, the concept of instrumental value and, in particular, the economic form of instrumental value is well suited to helping answer these questions.

To assess the instrumental value of nature, it is necessary to specify a goal and to identify the contributions that specific components of nature make toward the furtherance of that goal. Economics is the study of how societies organize themselves to provide for the sustenance and well-being of their members. Thus, in economics, the goal is increased human well-being. The economic theory of value is based on the ability of things to satisfy human needs and wants or to increase the well-being or utility of individuals. The economic value of something is a measure of its contribution to human well-being. The economic value of resource-environmental systems resides in the contributions that the variety of ecosystem functions and services make to human well-being.

5. DEFINING ECONOMIC VALUES

The economic concept of instrumental value is based on two fundamental premises of neoclassical welfare economics: that the purpose of economic activity is to increase the well-being of the individuals in the society and that individuals are the best judge of how well off they are in any given situation.

In other words, the preferences of individuals over alternative states is the basis for valuation. What determines each individual's ordering of alternative states? It is assumed that individuals act in their self interest. Specifically, it is assumed that people rank alternative states according to their own well-being in each of these states and that the well-being depends on the quantities of goods and services (represented as bundles) available to the individual in alternative economic states. Thus, if an individual prefers the set of goods and services provided in state A over the set that is provided in state B, then the individual prefers state A over state B. Using this brief description of the way individuals are assumed to order states, three important questions remain: (i) What is to be included in the sets of goods and services over which individuals are assumed to have preferences? (ii) What are the properties of the individual's preferences over alternative bundles? and (iii) Does the assumption of self interest preclude a concern for the well-being of other individuals? Each of these questions is discussed.

5.1 What is Included?

There is little controversy over including all of the goods and services that can be bought or sold in markets including consumer goods, the services of household assets such as a house or car, and consumer durables. Since time can be used in leisure activities or sold at some wage rate in the labor market, individuals must have preferences over alternative uses of time such as reading, outdoor recreation, and working at some wage rate. Governments provide a variety of services that enhance the well-being of their citizens. Finally, environmental amenities, such as clean air, clean water, scenic amenities, and so forth, also enhance individuals well-being. All these goods, services, activities, and amenities involve opportunity costs from an individual and/or social perspective. Spending more time on one activity means there is less time for other activities including earning wage income. Increasing the availability of environmental amenities means that the resources devoted to pollution control are not available to produce government services or market goods. So the question: What should be included in the analysis of economic value? is answered: include all things that people want, the provision of which has an opportunity cost.

5.2 Properties of Preferences

For our purpose, two properties of preferences are important. The first is nonsatiation, or the "more is better" property. This means that a bundle with a larger quantity of an element will be preferred to a bundle with a smaller quantity of that element, other things being equal. The second property is that preference orderings are characterized by substitutability. This means that if the quantity of one good or service provided to the individual is decreased, it is possible to increase the quantity of another good or service sufficiently to make the individual indifferent between the two bundles.

Substitutability is at the core of the economist's concept of value. This is because substitutability establishes trade-off ratios between pairs of goods that matter to people. The trade-offs that people make as they choose less of one good and substitute more of another good reveal the values that people place on these goods. If one of the goods has a monetary price, the revealed values are monetary values. The money price of a market good is a special case of a trade-off ratio because the money spent to purchase one unit of one element of the bundle is a proxy for the quantities of one or more of the other elements in the bundle that had to be reduced to make the purchase.

If individuals' preference orderings have the properties described here, they can be represented by ordinal preference functions or utility functions that assign a number to each bundle as a function of the quantities of each element of the bundle. This function is increasing in all of its arguments, and it is unique up to a monotonic transformation. This preference function is not the same thing as the cardinal utility function of the classical utilitarians. Since there is no unit of measurement for this ordinal utility, it is not possible to add or otherwise compare the utilities of different individuals.

Given the central role of the substitutability property in the definition and the measurement of economic values, it is important to consider the evidence supporting the assumption of substitutability. This assumption is the basis of most of the models of individual choice that are used to analyze and predict a wide variety of economic behavior both inside and outside of markets. These models include those of consumer demand and response to changes in prices, of savings, and of labor supply. Also included are models of a variety of individuals' behaviors related to environmental and health considerations, including participation in outdoor recreation activities, choices among jobs with varying degrees of risk of fatal accident, and choices of where to live and work when houses and urban centers offer different packages of amenities and environmental pollution. The successful development and application of these models would not be possible if substitutability was not a common feature of individuals' preferences.[4]

Another question to consider is whether the assumptions of individualism and self interest preclude altruism or a concern on the part of individuals for the well-being of others. In principle, the answer is no. One of the goods over which an individual has a preference could be the quantity of the good consumed by another individual or the money expenditure of others. In fact, it is difficult to rationalize some commonly observed behavior, for example, gifts to other individuals and charitable donations, without resorting to some form of interdependence of preference or utility functions. However, the typical practice in applied welfare economics and benefit-cost analysis has been to assume that interdependence of preferences does not exist. Furthermore, the anthropocentric focus of economic valuation does not preclude a concern for the survival and well-being of other species. Individuals can value the survival of other species because of the uses they make of them (for food and recreation, for example), and because of an altruistic or ethical concern. Such concerns are the source of passive use values (see Chapter 2).

5.3 Compensation Measures of Value

A measure of the value of an environmental change based on substitutability can be expressed either as a compensating surplus (CS) or an equivalent surplus (ES).[5] In principle, CS and ES measures can be defined in terms of any other good an individual is willing to substitute for the good being valued. In the following discussion, money is used as the numéraire in which trade-off ratios are expressed; but CS and ES could be measured in terms of any other good that mattered to the individual.[6] The CS for an environmental improvement is the maximum sum of money the individual would be willing to pay rather than do without the improvement. This sum is the amount of money that would make the individual indifferent between paying for and having the improvement and forgoing the improvement, while keeping the money to spend on other things. The CS for an improvement is also known as willingness to pay (WTP).

The ES for an improvement is the minimum sum of money the individual would require to voluntarily forgo the improvement; it is the amount that would make a person indifferent between having the improvement and forgoing the improvement, while getting the extra money. In other words, it is the amount of money that would generate an increase in utility equivalent to that realized from the improvement in the environmental amenity. The ES for an improvement is also known as willingness to accept (WTA) compensation.

Both value measures are based on the assumption of substitutability in preferences, but each adopt different reference points. The CS reference point is the absence of the improvement, while the ES reference point is the presence of the improvement. In principle, CS and ES need not be exactly equal for an equal size change in environmental quality (Q). WTP is constrained by the individual's income. But there is no upper limit on what that individual would require as compensation to forgo the improvement (WTA).[7]

These two measures can also be interpreted in terms of the implied rights and obligations associated with alternative outcomes. The CS measure presumes that the individual has no right to the improvement in Q but does have a right to the original level of Q, for proposed decreases in Q. In contrast, the ES measure presumes that the individual has a right to the new, higher level of Q and must be compensated if the new level of Q is not attained. Based on this interpretation of the two measures, some economists have argued that the choice between them is basically an ethical one; that is, a choice that depends on a value judgment concerning which underlying distribution of property rights is more equitable.[8]

5.4 Some Issues

There are three problematic aspects of basing an economic value measure on individuals' preferences and observations of their behavior, which I will briefly mention. First, individuals may have poor information about the nature of the environmental-service flows and how the flows affect their well-being, especially those indirect or intermediate services described in section 1. If individuals do not understand the contribution that an ecosystem makes to their well-being, then their observed behaviors or responses to questions will reflect that ignorance rather than the true value of the ecosystem to them. And estimates of value based on observed behavior or responses to questions will be biased downward. For example, if people do not know the role that an insect species plays in pollinating a valuable food or ornamental plant species, their valuation of the insect will not reflect its indirect contribution to their well-being.

Second, individuals' choices and responses to valuation questions are constrained by their income. If the distribution of income is deemed unfair or unjust, then the values revealed or expressed by people who are too poor or too rich lose their claim to moral standing. This issue has led to various proposals to add an equity or fairness criterion to policy evaluation. This point is discussed in section 6.3.

The third issue concerns the nature of the preferences that are relevant for public policy decisions. In the standard applied welfare economics framework, individuals are viewed as having well-defined preferences over alternative bundles of goods that contribute to well-being. Where most of the goods and services that contribute to utility are purchased in markets, the constraints of prices and limited money income reward well-informed and well-defined preferences. The theory of economic value as applied to environmental goods has expanded from this point by assuming that the domain of preferences extends to environmental goods that are unavailable through markets. The assumption is made that preferences for market and environmental goods are well-defined and guide individuals' choices.

Some critics of the use of benefit-cost analysis and economic valuation methods have questioned whether people have a single or unified set of preferences that govern private choices concerning market goods and the choices that they make over alternative public policies affecting environmental services (for example, Kelman 1981; Sagoff 1988). Their argument is that people have one set of preferences that govern their private choices (consumer preferences), and another set that governs their actions in the political arena (civic preferences). Therefore, it is a mistake to use values derived from consumer preferences to make choices about public policies toward the environment. Sen has also recently taken up this theme:

> The basic question that is raised by such a market oriented approach is whether this view of the individual as an operator in a market best captures the problems of environmental valuation. An alternative view is to see the individual as a citizen - an agent who judges the alternatives from a social perspective which includes her own well-being but also, quite possibly, many other considerations (Sen 1995, p.23).

The question of civic versus consumer preferences is an empirical one that could be settled by gathering the appropriate evidence. When decisions about environmental policy are made in the political arena, the standard methods of economic valuation could be used to infer the preferences lying behind individuals' stated intentions, or their actual behavior in the political system. Where individuals are making private and public choices about some environmental variable, it should be possible to test for the consistency of these choices. Unless such tests are conducted, and their results support the hypothesis of differences between civic and consumer preferences, the standard economic assumption should continue to be the basis for economic valuation.

6. USING ECONOMIC VALUES IN DECISION MAKING

In this section, I discuss the question of how to aggregate the measures of individual welfare changes (CS or ES) when gains and losses accrue to different individuals. The fundamental value judgment of standard welfare economics is that social ranking of alternative policies should be based on individuals' preferences over these alternatives. The key issue is what to do when individuals rank the alternatives differently. The Pareto Principle provides a minimum standard for deriving a social ranking from individuals' preferences or rankings. The principle states that if at least one person prefers policy A over policy B, and all other individuals rank policy A at least as high as policy B, then the social ranking should place A above B. For any pairwise comparison of two policies (A and B), one of four outcomes is possible:

(i) policy A will be ranked above policy B;
(ii) policy B will be ranked above policy A;
(iii) policies A and B will be ranked as equally preferred; or
(iv) the policies will be deemed Pareto noncomparable.

This last outcome would occur if at least one person preferred A, while another person preferred B. Any policy that imposes costs on some individuals will not be preferred by those who bear the costs. As long as it is possible to find one individual who would lose from a policy, that policy could not be Pareto preferred. It is difficult to imagine real world policies (from imposing pollution control requirements on firms to raising taxes to producing public goods to breaking up the market power of monopolists) that do not involve losses to some individuals either through higher prices or reduced disposable income. The potential compensation test has been proposed as a way to rank policies that impose costs on some, while benefitting others.

6.1 The Potential Compensation Test

Benefit-cost analysis is an application of the Hicks-Kaldor potential compensation test or Potential Pareto improvement (PPI) criterion used as a basis for judging which proposed policies result in welfare improvements. According to this criterion, a policy would be accepted if in principle, those who gain from the intervention could transfer income to those who would lose. Payment of the compensation makes those who are worse off with the intervention at least as well off as they would be without the intervention, while those who pay the compensation remain better off (even when paying compensation) than they would be without the intervention.[9] Thus, if the project were undertaken and the compensation actually paid, there would be some who gain and none who lose. The act of compensation would convert a PPI into a true Pareto improvement policy.

The original statement of the criterion by Kaldor (1939) did not address whether the compensation should actually be paid. The purpose of the Hicks-Kaldor potential compensation test was to separate the question of whether the policy should be undertaken from the question of whether compensation should be paid, or in other words, to separate the efficiency dimension from the equity dimension of the social choice problem. But a number of economists, most notably I. M. D. Little (1957), have criticized this effort. Little argued that if compensation were not paid, the PPI criterion should be amended so that the project would be undertaken only if the Hicks-Kaldor potential compensation test were passed and if the resulting redistribution of well-being is judged to be satisfactory. Thus, Little argued that the problem of judging equity or income distribution could not be avoided.

The potential compensation test also provides the basis for defining and measuring the benefits and costs of public policies. The costs of a policy are the losses in well-being experienced by a subset of the population. These losses could result from reductions in money income, higher prices of consumer goods bought through markets, and reduced availability of nonmarket goods such as environmental amenities. For individuals, the cost is defined and measured by the amount of money that must be paid to them to compensate for the loss; that is, the sum that enables the individual to purchase more of other goods and services to avoid a loss in well-being. Similarly, for those who gain from the policy, the benefit is defined and measured by the sum of money that could be taken away from that individual, thereby reducing the consumption of other goods and services so that the individual is no better off than in the status quo no policy situation. This is the WTP or CS measure defined earlier. Thus, the measurement of costs and benefits rests on the validity of the substitutability property of preferences, since costs cannot be compensated in the absence of substitutability.

A benefit-cost analysis of a policy is carried out by first measuring the aggregate WTP of those who gain and calling this the benefit, and then measuring the aggregate WTA of the losers and calling this the cost. If the benefits exceed the cost, that is, if the net benefits are positive, then by definition there is sufficient money that can be transferred from the beneficiaries and to the losers so that after the transaction the new position is Pareto preferred.[10] But the PPI or net benefit criterion does not require that the compensation be paid.

The potential compensation test or PPI criterion is perhaps the most controversial feature of standard welfare economics. On one hand, PPI has been criticized as being incompatible with the Pareto Principle since it allows for a ranking of projects that are Pareto noncomparable. Conversely, its application has been rationalized on the grounds that if a large enough number of PPI projects are undertaken, benefits and costs will be spread sufficiently widely that everyone will be a net gainer from the set of projects as a whole, even though some might be losers on some individual projects. Thus, benefit-cost analysis as a basis for policy choice has a shaky foundation. However, this has not deterred governments from using PPI for some kinds of policy choices and economists from advocating greater use of it in a wider range of environmental and resource policy questions. Whether this foundation can take the strain associated with its use in the emerging environmental policy issues is an important question.

The Kaldor (1939) and Hicks (1939) versions of PPI tests take different reference points as the basis for determining the required compensations. This means that there are two alternative definitions of welfare change, as pointed out in section 4.3.

The Kaldor (1939) version of the compensation test asks whether those who gain from policy implementation can transfer sufficient income to the losers from the policy without themselves becoming worse off. It takes the status quo levels of well-being as the reference point. For the gainers, the test requires determining each individual's maximum willingness to pay rather than do without the improvement brought about by the project. This is a CS measure.

For the losers, the test requires determining the minimum sum of money that, if paid when the policy is implemented, would leave the individual indifferent between the status quo and accepting the project with compensation. This is also a CS measure. If the aggregate of the compensating measures of benefits exceeds the aggregate of the compensating measures of cost, the project passes the Kaldor version of the potential compensation test. If the aggregate of compensating measures is positive and, if the compensation is paid, there are no losers and some people gain. In fact, it may be possible to overcompensate the losers so that everyone gains.

Hicks' (1939) version of the potential compensation test says that a proposed policy should be rejected if those who lose from the policy could compensate the gainers for a decision not to proceed with the project. To implement this compensation test, one would need to know how much the potential gainers would require in compensation to induce them to forgo project implementation. The amount of compensation would be the monetary equivalent of the increase in well-being that each member of this group would experience if the project were approved. Similarly, the project losers would be willing to pay an amount equivalent to their loss in well-being if the project were to proceed. Thus, the Hicks version of the compensation test requires equivalent or ES measures of welfare change. If the sum of these equivalent measures of gains and loses is greater than zero, the project passes the Hicks potential compensation test because it is impossible for those who would lose from the project to bribe the potential gainers into agreeing to forgo the project.

The principal difference between these two forms of compensation test is the position that is taken as the reference point for welfare measurement. The Hicks test measures changes in welfare with respect to the levels of well-being associated with the project. The Kaldor test measures changes in welfare with respect to levels of well-being without the project. The perspective that is adopted for benefit-cost analysis may depend on a social or collective judgment about rights and entitlements. Additionally, the judgment might vary according to the circumstances of the case at hand. For example, if the question is whether or not to build a trash disposal facility in a residential neighborhood, society might make the value judgment that residents have a right to the status quo ante regarding neighborhood amenities. Given that judgment, the question is how much must they be compensated for any loss in those amenities. Thus, compensating measures of welfare change and the Kaldor welfare test are appropriate. Alternatively, the policy question might be whether or not to impose a set of regulations to reduce air pollution levels. If society makes the value judgment that its citizens have the right to clean air, then the question is how much would they have to be compensated to forgo air quality improvement. The amount of compensation is the monetary equivalent of the increase in well-being associated with the air quality. Thus, equivalent measures of welfare change are appropriate, and the proper test is the Hicks version of the potential compensation test.

6.2 Benefit-Cost Analysis and Uncertainty

One problem in carrying out a benefit-cost analysis is that the values for some physical, technical, or economic parameters of the model may not be known with certainty. The state of the art in measurement of environmental values is not sufficiently far advanced to produce exact measures of the relevant parameter values. This leads to the questions: Must policy makers wait for further research to produce exact measures before they can attempt to measure benefits and costs? If not, how can they make the best use of the ranges of parameter values that current research has produced?

To wait for exact values is equivalent, in many cases, to never using benefit-cost analysis. The state of the art cannot be expected to advance to the point of producing exact parameter values for all kinds of environmental change. This is because of the inherent uncertainty and imprecision in measurement techniques based on sampling and statistical inference and because of the complexity and imperfect understanding of the physical, biological, and socioeconomic systems that must be modeled to produce the relevant scenarios for welfare comparisons. How are policy makers to proceed in the face of continued and inherent uncertainty?

The simplest approach is to base the calculations of benefits and costs on the expected values of the uncertain parameters and to base decisions on these expected values. But decision makers often want to know more about the magnitude of the uncertainties in the estimates of benefits and costs. They could be provided with the upper and lower bounds of the ranges of values along with the expected values. If the benefits of a policy that are calculated with the upper end of the range are less than the lower end of the range of estimated costs, the policy is unlikely to be justifiable on economic grounds. However, if the benefits of a policy that are calculated with the lower end of the range exceed the upper end of the range of costs, the economic case for the policy is quite strong.

This approach is a step in the right direction, but it can be criticized because it does not use all of the relevant information contained in the studies making up the range of values. Formally, the range reflects only the information contained in the two studies yielding the highest and lowest values. It ignores information on the quality of these two studies; and it ignores all of the information contained in the other studies that yielded values within the range.

There is a more formal approach to dealing with uncertainty that makes use of all the available information from different studies and allows for incorporation of judgments about the quality of each of the studies. This approach is based on viewing probabilities as statements about the degree of confidence held about the occurrence of some possible event. It involves assigning probabilities to all of the values produced by the available studies, where a higher probability reflects a greater degree of confidence in the results of that study.[11] Once the probabilities have been assigned, then various statistical manipulations can be performed. For example, the expected value of the parameter value (the mean of the probability distribution) can be calculated and used for benefit-cost calculations. The variance of the distribution can be used to determine confidence intervals on the value to be used, thus preserving for policy makers information on the uncertainty about values. When there are multiple uncertainties, Monte Carlo methods can be used to draw from the assumed distributions to generate a probability distribution of outcomes.

6.3 Conclusions

Faced with the problem of making judgments about the aggregate welfare consequences of policies that provide benefits to some and impose costs on others, economists have adopted the criterion of choosing the policy with the highest aggregate net benefits and have justified this criterion by indicating that the gainers could potentially compensate the losers. One problem with the aggregate net benefit criterion is that it omits any concern for the fairness of the underlying distribution of well-being or for the incidence of benefits and costs across different income levels. One way to incorporate concern for equity in the distribution of well-being, with roots going back to Bergson (1938), is to weight the measures of value or welfare change for each individual by that person's relative degree of deservingness; that is, to employ a social welfare function that attaches a higher weight to benefits going to those judged to be more deserving because of their lower level of income.[12] However, the search for a social welfare function has been unsuccessful. In practice, analysts typically use the value measures derived from the mean individual in the sample that is providing data for the valuation model in use. If value or willingness to pay is an increasing function of income, the analyst is implicitly underestimating the values of the highest income individuals and overestimating the values of the lowest income individuals. The result, in a crude qualitative sense at least, is equivalent to what would be obtained with a social welfare function.

However, it must be acknowledged that a major limitation to the economic approach to policy evaluation based on aggregate net benefits is its failure to provide a basis for including equity or fairness considerations in its evaluation of alternative policies. Recognizing this, many economists argue that although information on the benefits and costs of policies is important, it should not be decisive in decisions about policy; that is, the aggregate net benefit criterion should be a factor in policy choice, but it should not be a decision rule.[13]

7. METHODS OF VALUATION

In this book, the methods of nonmarket valuation are placed in one of two broad categories according to the nature of the data generated for modeling and estimation. Chapters 4 through 7 describe methods based on statements individuals make in response to questions about hypothetical situations such as "What would you do if . . . ? or Would you be willing to pay for . . ?", As the individuals' preferences are not observed but rather stated, these methods are referred to as "stated preference methods." The other methods, which are described in Chapters 8 through 11, rely on data that come from observations of people acting in real-world settings where people live with the consequences of their choices. The reliance of these methods on observed behavior gives rise to the term "revealed preference methods."

Before the specific nonmarket valuation techniques are presented, Chapter 2 by Flores develops the theory of nonmarket valuation more formally and explores some conceptual issues. Chapter 3 by Champ describes data collection methods for stated preference and revealed preference nonmarket valuation methods. Chapter 4 introduces the stated preference valuation methods.

Some issues and problems in the stated preference methods are specific to the particular form of the question being asked. For example, when people are asked how much they would be willing to pay for something, they might say zero, because they reject the idea of having to pay for something they consider to be rightfully theirs. Other problems are generic to all methods based on hypothetical questions. For example, problems in scenario specification, sampling, and item nonresponse. The major questions regarding all stated preference methods concern the validity and reliability of the data; that is, whether the hypothetical nature of the questions asked inevitably leads to some kind of bias or results in so much "noise" that the data are not useful for drawing inferences. These issues will be discussed in detail in Chapters 4 through 7.

The most commonly used stated preference method is the contingent valuation method covered in Chapter 5. This method describes the good or the program to be valued, then directly asks respondents to identify their maximum willingness to pay, or whether they would be willing to pay a specified amount for the good or the program. Chapter 6 describes stated preference approaches that focus on valuation of attributes of a good. Attribute-based methods obtain preferences over similar goods that differ in the levels of their common attributes. An emerging approach related to the attribute-based methods is the method of paired comparison described in Chapter 7. This method is also based on choices individuals make between goods. However, the goods may have differing attributes and the method allows for valuation of multiple goods. The commonality among the various stated preference approaches is that the data are generated based on individuals stating their preferences rather than on making observable economic decisions.

The revealed preference methods are based on the actual behaviors of individuals and reflect utility maximization subject to constraints. These methods involve a kind of detective work in which clues about the values individuals place on environmental services are pieced together from the evidence that people leave behind as they respond to prices and other economic signals.

Chapter 8 introduces the revealed preference valuation methods. The travel cost method described in Chapter 9 infers the value of a recreational experience from the costs individuals incur to travel to the recreation site. Chapter 10 describes the hedonic method, in which the values of the characteristics of a resource are inferred from observable market transactions. Chapter 11 describes some frequently used methods of nonmarket valuation based on actual expenditures to avoid or mitigate negative environmental externalities.

The concluding section of the book takes up some additional topics in nonmarket valuation. Chapter 12 by Rosenberger and Loomis outlines the principles of benefits transfer, the procedure of using existing valuation data from other studies, perhaps with adjustments for differences in circumstances, to estimate values in a particular policy setting. Chapter 13 by McCollum reviews several cases and discusses the role of nonmarket valuation data in real policy decisions. Finally, Chapter 14 by Bishop provides conclusions and a discussion of the future of nonmarket valuation.

NOTES

1 For a more complete discussion of these trade-offs, see Schmidt et al. (1998).
2 See Chapter 13 for a further discussion of this example.
3 For a complete discussion of indirect ecosystem services, see Daily (1997).
4 However, researchers have found some evidence of lexicographic preferences. For example, Common, Reid, and Blamey (1997) used a contingent valuation survey to examine people's willingness to trade income for changes in the survival probability of an endangered species.
5 These measures are defined more rigorously in Chapter 2.
6 The choice of a numéraire for the compensation measure is irrelevant in terms of its effect on how any one individual ranks alternative outcomes. But as Brekke (1997) has shown, the choice of a numéraire can affect the rankings of outcomes based on the aggregation of welfare measures across individuals.
7 For more on the difference between ES and CS, see Chapter 2.
8 See Krutilla (1967) and Mishan (1976).
9 The Hicks (1939) and Kaldor (1939) versions of the compensation tests differ in terms of which state (the pre- or post-intervention state) is to be taken as the reference point or status quo for identifying gainers and losers. The text is based on the Kaldor version of the tests.
10 If there is any consideration of actually paying the compensation, the transaction's costs, potential losses, and excess burdens of taxation associated with raising funds from the beneficiaries and transferring them to the losers must be accounted for in the calculation of net benefits.
11 For a discussion of this approach, see von Winterfeldt and Edwards (1986).
12 See Boadway and Bruce (1984), Chapter 6.
13 See Arrow et al. (1996).

REFERENCES

Arrow, K. J., M. L. Cropper, G. C. Eads, R. W. Hahn, L. B. Lave, R. G. Noll, P. R. Portney, M. Russell, R. Schmalensee, V. K. Smith, and R. N. Stavins. 1996. Is There a Role for Benefit-cost Analysis in Environmental, Health, and Safety Regulation? *Science* 27:221-222.

Bergson, A. 1938. A Reformulation of Certain Aspects of Welfare Economics. *Quarterly Journal of Economics* 52:310-334.

Boadway, R., and N. Bruce. 1984. *Welfare Economics*. Oxford, England: Basil Blackwell.

Botkin, D. 1990. *Discordant Harmonies: a New Ecology for the Twenty-first Century.* New York: Oxford University Press.

Brekke, K. A. 1997. The Numéraire Matters in Cost-benefit Analysis. *Journal of Public Economics* 64:117-123.

Callicott, J. B. 1989. *In Defense of the Land Ethic: Essays in Environmental Philosophy.* Albany, NY: State University of New York Press.

Common, M., I. Reid, and R. Blamey. 1997. Do Existence Values for Cost-benefit Analysis Exist? *Environmental and Resource Economics* 9:225-238.

Costanza, R., and C. Folke. 1997. Valuing Ecosystem Services with Efficiency, Fairness and Sustainability as Goals. In *Nature's Services: Societal Dependence on Natural Ecosystems*, Edited by G. C. Dail. Washington, DC: Island Press.

Daily, G. C. 1997. Introduction. In *Nature's Services: Societal Dependence on Natural Ecosystems*, Edited by G. C. Daily. Washington, DC: Island Press.

Daily, G. C., T. Soderqvist, S. Aniyar, K. Arrow, P. Dasgupta, P. R. Ehrlich, C. Folke, A.M. Jansson, B-O. Jansson, N. Kautsky, S. Levin, J. Lubchenco, K-G. Maler, D. Simpson, D. Starrett, D. Tilman, and B. Walker. 2000. The Value of Nature and the Nature of Value. *Science* 289:395-396.

Freeman, A. M. III. 1993. *The Measurement of Environmental and Resource Values: Theory and Methods.* Washington, DC: Resources for the Future.

Freeman, A. M. III. 1995. The Economic Valuation of Coastal Resources Supporting Recreation. In *Sustaining Coastal Resources: Roles of Economics and Environmental Sciences.* Edited by Charles S. Colgan. Portland, ME: Edmund S. Muskie Institute of Public Affairs.

Freeman, A. M. III., R. H. Havemen, and A. V. Kneese. 1973. *The Economics of Environmental Policy.* New York: John Wiley.

Hicks, J. R. 1939. The Foundation of Welfare Economics. *Economic Journal* 49(196):696-712.

Kaldor, N. 1939. Welfare Propositions of Economics and Interpersonal Comparisons of Utility. *Economic Journal* 49(195):549-552.

Kelman, S. 1970. Cost-benefit Analysis: an Ethical Critique. *Regulation* 5:33-40.

Kneese, A. V., R. U. Ayres, and R. C. d'Arge. 1970. *Economics and the Environment: A Materials Balance Approach.* Baltimore, MD: The Johns Hopkins University Press for Resources for the Future.

Krutilla, J. V. 1967. Conservation Reconsidered. *American Economic Review* 57(4):777-786.

Little, I. M. D. 1957. A Critique of Welfare Economics. 2nd ed. Oxford, England: Clarendon Press.

Mishan, E. J. 1976. The Use of Compensating and Equivalent Variations in Cost-benefit Analysis. *Economica* 43(170):185-197.

Sagoff, M. 1988. *The Economy of the Earth.* Cambridge, England: Cambridge University Press.

Schmidt, J. C., R. H. Webb, R. A. Valdez, G. R. Marzolf, and L. E. Stevens. 1998. Science and Values in Restoration in the Grand Canyon. *Bio Science* 48(9):735-747.

Sen, A. 1995. Environmental Evaluation and Social Choice: Contingent Valuation and the Market Analogy. *The Japanese Economic Review* 46(1):23-37.

Smith, V. K. 1988. Resource Evaluation at the Crossroads. *Resources* 90:2-6.

von Winterfeldt, D., and W. Edwards. 1986. *Decision Analysis and Behavioral Research.* Cambridge University Press: Cambridge.

Chapter 2

CONCEPTUAL FRAMEWORK FOR NONMARKET VALUATION

Nicholas E. Flores
University of Colorado, Boulder

1. INTRODUCTION

Serious practice of nonmarket valuation requires a working knowledge of the economic theory because it forms the basis for the explicit goals in any nonmarket valuation exercise. This chapter provides readers with the requisite theory to meaningfully apply the nonmarket valuation techniques.

To do so, I will (1) develop a model of individual choice that explicitly recognizes the public good nature of nonmarket goods. Using this model, I will (2) derive the basic welfare measures that nonmarket valuation studies attempt to measure. Moving toward a more specific framework, I will (3) examine how market behavior can be used to identify the basic welfare measures for nonmarket goods. I will also (4) provide a discussion of situations for which market demands are not sufficient to recover the basic welfare measures, cases of passive use value or visits to new recreation sites. From there, I will (5) discuss inter-temporal choice and nonmarket valuation, and finally nonmarket valuation under uncertainty.[1]

2. THEORETICAL MODEL OF NONMARKET GOODS

Nonmarket valuation is necessary and distinct from neoclassical price theory of market goods. The air quality in Boulder, the water quality of Colorado's lakes and streams, and the preservation of public lands are relevant examples of nonmarket goods. Each of these goods can change due to society's choices, but individuals may not unilaterally choose their most preferred level of air quality, water quality, or acreage of preserved public lands. In addition to being outside of the choice set of any individual, these examples have the feature that everyone experiences the same level of the good. Citizens of Boulder experience the same level of local air quality; citizens of Colorado experience the same level of water quality in the state's lakes and streams; and finally the level of preserved public lands is shared by all. Rationed, common level goods serve as our point of departure for standard neoclassical price theory in developing the theoretical framework.

The basic premise of neoclassical economic theory is that people have preferences over goods, in our case both market and nonmarket goods. Without regard to the costs, each individual is assumed to be able to order bundles of goods in terms of desirability, resulting in a complete preference ordering. The fact that each individual can preference order bundles forms the basis of choice. The most fundamental element of economic theory is the preference ordering, more simply the desires of the individual, not money. Money plays an important role since individual have a limited supply of money to buy many, but not all, of the things they want. Economic theory is silent with regard to motivation. An individual may desire improved air or water quality or the preservation of an endangered species for any reason including for personal use, bequests to future generations, or simply for the existence of the resource. As Becker (1993) offers that the reasons for enjoyment of any good can be "selfish, altruistic, loyal, spiteful, or masochistic." Economic theory provides nearly complete flexibility for accommodating competing systems of values

Preference ordering can be represented through a utility function defined over goods. For our purposes, $X = [x_1, x_2, \ldots, x_n]$ denotes a list or vector of all of the levels for the n market goods that the individual chooses. The k nonmarket goods are similarly listed as $Q = [q_1, q_2, \ldots, q_k]$. The utility function assigns a single number, $U(X, Q)$, for each bundle of goods (X, Q). For any two bundles, (X^A, Q^A) and (X^B, Q^B), the respective numbers assigned by the utility function are such that $U(X^A, Q^A) > U(X^B, Q^B)$ if and only if (X^A, Q^A) is preferred over (X^B, Q^B). The utility function is thus a complete representation of preferences.[2]

Money enters the process through scarcity and, in particular, scarcity of money to spend on obtaining the things we enjoy. For market goods, individuals choose the amount of each good to buy based on preferences, the relative prices of the market goods $P = [p_1, p_2, \ldots, p_n]$, and available income. Given this departure point, the nonmarket goods are rationed in the sense that individuals may not unilaterally choose the level of these goods.[3] The basic choice problem is how to obtain the highest possible utility level when spending income y towards the purchase of market goods is subject to a rationed level of the nonmarket goods.

$$(1) \qquad \max_{X} \; U(X, Q) \quad s.t. \quad P{\cdot}X \le y, \quad Q = Q^0$$

There are two constraints that people face in (1). First, the total expenditure on market goods cannot exceed income,[4] and second, the levels of the nonmarket goods are fixed.[5] The X that solves this problem then depends on the level of income (y), the prices of all of the market goods (P), and the level of the rationed, nonmarket goods (Q). For each market good, we have an optimal demand function that depends on these three elements, $x_i^* = x_i(P, Q, y)$; the vector of optimal demands can be written similarly, $X^* = X(P, Q, y)$ where the vector now lists the demand function for each market good. If we plug the set of optimal demands into the utility function, we obtain the indirect utility function, $U(X^*, Q) = v(P, Q, y)$. Because the demands depend on prices, the levels of the nonmarket goods, and income, the highest obtainable level of utility also depends on these elements.

As the name suggests, demand functions provide the quantity of goods demanded at a given price vector and income level. Demand functions also can be interpreted as marginal value curves since consumption of goods occurs up to the point where marginal benefits equal marginal costs. For this reason, demand has social significance.

2.1 Compensating and Equivalent Welfare Measures

Policies or projects that provide nonmarket goods often involve costs. We may assign value to these policies or projects in order to assess whether the benefits justify the costs. For example, consider a policy intended to improve the water quality of Boulder Creek, a stream that runs through my hometown of Boulder, Colorado. I care about this stream because I jog along its banks and enjoy the wildlife that it supports, including the trout that my daughters may catch when they are lucky. To pay for this clean-up, the prices of market goods might change due to an increase in sales tax and/or I might be asked to pay a lump-sum fee. Two basic measures of value, which are standard fare in welfare economics, may be used to assess the benefit of cleaning up Boulder Creek. The first is the amount of income that I would give up after the policy has been implemented that would exactly return my utility to the status quo utility level. This measure is the *compensating* welfare measure, which I will refer to as C. Letting 0 superscripts denote the initial, status quo, conditions and 1 superscripts denote the new conditions provided by the policy, define C using the indirect utility function.

(2) $$v(P^0, Q^0, y^0) = v(P^1, Q^1, y^1 - C)$$

The basic idea behind C is that if I give up C with the changes, then I am back to my original utility. C could be positive or negative depending upon how much prices increase and/or the size of any lump sum tax that I pay. If costs are less than C and the policy is implemented, then I am better off than before the policy. If costs are more than C, I am worse off.

The second basic welfare measure is the amount of additional income that I would need with the initial conditions to obtain the same utility as after the change. This is the *equivalent* welfare measure, referred to as E, and is defined as follows.

(3) $$v(P^0, Q^0, y^0 + E) = v(P^1, Q^1, y^1)$$

The two measures differ by the implied assignment of property rights. For the compensating measure, we are recognizing the initial utility level as the basis of comparison. For the equivalent measure, we are recognizing the subsequent level of utility as the basis. Whether one should consider the compensating welfare measure or the equivalent welfare measure as the appropriate measure depends on the situation. Suppose that we are considering a new policy intended to improve Boulder Creek water quality. In this case, the legal property right is the status quo; therefore, we should use the compensating welfare measure. There are, however, instances when the equivalent welfare measure is conceptually correct. Returning to the water quality example, the Clean Water Act provides minimum water quality standards. In essence, the Act assigns a property right to the standard. If water quality had declined below the standard and we are considering a project that would restore quality to this minimum standard, then the equivalent welfare measure is the appropriate measure. Both conceptual and practical matters should guide the choice between the compensating and equivalent welfare measure.[6]

2.2 Duality and the Expenditure Function

I have so far used the indirect utility function to describe the basic welfare measures used in economic policy analysis. To more easily discuss and analyze specific changes, we can equivalently use the expenditure function to develop our welfare measures. The indirect utility function represents the highest level of utility obtainable when facing prices P, nonmarket goods Q, and income y. A necessary condition for utility maximization is that we spend our income in a least-cost manner. To illustrate this, suppose that my market good purchases are facing prices P and nonmarket goods Q. I obtain a utility level of U^0. Now suppose that I am not minimizing my expenditures and U^0 could be obtained for less money. If this were true, I would not be maximizing my utility since I could spend less on U^0 and use the remaining money to buy more market goods and thus obtain a utility level higher than U^0. This reasoning is the basis of what we refer to in microeconomics as the dual problem. Instead of looking at maximizing utility subject to the budget constraint, we consider the dual problem of minimizing expenditures subject to obtaining a given level of utility. The expenditure minimization problem is stated as follows.

$$(4) \qquad \min_{X} P \cdot X \quad s.t. \quad U(X, Q) \geq U^0, \quad Q = Q^0$$

The solution to this problem is the set of compensated or Hicksian demands which are a function of prices, nonmarket goods levels, and level of utility, $X^* = X^h(P, Q, U)$. The dual relationship between the ordinary demands and the Hicksian demands is that $X(P, Q, y) = X^h(P, Q, U)$ when either

$U = v(P, Q, y)$ in the expenditure minimization problem or $y = P{\cdot}X^h(P, Q, U)$ in the utility maximization problem. As the phrase duality suggests, these relationships represent two views of the same choice process. The important conceptual feature of the compensated demands is that utility is fixed at some specified level of utility, which relates directly to our compensating and equivalent welfare measures. For the expenditure minimization problem, the expenditure function, $e(P, Q, U) = P{\cdot}X^h(P, Q, U)$, takes the place of the indirect utility function.

It is worth stressing that the expenditure function is the ticket to understanding welfare economics. Not only does the conceptual framework exactly match the utility-constant nature of welfare economics, the expenditure function itself has very convenient properties. In particular, the expenditure function approach allows us to a decompose a policy that changes multiple goods or prices into a sequence of changes which will be shown below to provide powerful insight into our welfare measures.

I have so far introduced the broad concepts of compensating and equivalent welfare measures. Hicks (1943) developed the compensating and equivalent measures distinctly for price and quantity changes and named them respectively as the price compensating/equivalent variation for changes in prices and the quantity compensating/equivalent variation for quantity changes. These two distinct measures are now typically referred to as the compensating/equivalent variation for price changes and the compensating/equivalent surplus for quantity changes. It is easy to develop these measures using the expenditure function, particularly when one understands the terms equivalent and compensating.

Before jumping directly into the variations and surpluses, I want to discuss income changes. Income changes can also occur as a result of policies, and so we deal with changes in income first. For example, regulating the actions of a polluting firm may decrease the demand for labor and result in lower incomes for workers.

2.3 The Treatment of Income Changes

Let $U^0 = v(P^0, Q^0, y^0)$ represent the status quo utility level and $U^1 = v(P^1, Q^1, y^1)$ the utility level after a generic change in income, prices, and/or nonmarket goods. Our compensating and equivalent measures introduced above are kept here, but I have added a change in income. The two measures are defined by the fundamental identities as follows.

$$(5)\qquad \begin{aligned} v(P^0, Q^0, y^0) &= v(P^1, Q^1, y^1 - C) \\ v(P^0, Q^0, y^0 + E) &= v(P^1, Q^1, y^1) \end{aligned}$$

We can also represent C and E using the expenditure function.

$$(6)\qquad \begin{aligned} C &= e(P^1, Q^1, U^1) - e(P^1, Q^1, U^0) \\ E &= e(P^0, Q^0, U^1) - e(P^0, Q^0, U^0) \end{aligned}$$

To determine how to handle income changes, I need to rewrite C and E in more workable forms. In expenditure terms, $y^0 = e(P^0, Q^0, U^0)$, $y^1 = e(P^1, Q^1, U^1)$, and $y^1 = y^0 + y^1 - y^0$. By creatively using these identities, I can rewrite C and E as follows.

$$(7)\qquad \begin{aligned} E &= e(P^0, Q^0, U^1) - e(P^1, Q^1, U^1) + (y^1 - y^0) \\ C &= e(P^0, Q^0, U^0) - e(P^1, Q^1, U^0) + (y^1 - y^0) \end{aligned}$$

The new form shows that for C, we value the changes in prices and nonmarket goods at the initial utility level and then consider the income change. For E, we value the changes in prices and nonmarket goods at the post-change utility level and then consider income change. The generalized compensated measure is subtracted from income under the subsequent conditions (equation 2), while the generalized equivalent measure is added to income under the initial conditions (equation 3). How we value the changes in prices and nonmarket goods is the next question.

2.4 Variation Welfare Measures for a Change in Price *i*

Suppose we are considering a policy that only provides a price decrease for good i, $p_i^0 > p_i^1$. Hicks referred to the compensating welfare measure for a price change as compensating variation (CV) and the equivalent welfare measure as equivalent variation (EV). Since a price decrease makes the consumer better off, both measures are positive. P_{-i} refers to the price vector left after removing p_i.

(8) $$CV = e(p_i^0, P_{-i}^0, Q^0, U^0) - e(p_i^1, P_{-i}^0, Q^0, U^0)$$

(9) $$EV = e(p_i^0, P_{-i}^0, Q^0, U^1) - e(p_i^1, P_{-i}^0, Q^0, U^1)$$

Using Roy's identity, and the fundamental theorem of calculus, compensating and equivalent variations can be expressed as the area under the Hicksian demand curve between the initial and subsequent price.[7] Here s represents p_i along the path of integration.

(10) $$\begin{aligned} CV &= e(p_i^0, P_{-i}^0, Q^0, U^0) - e(p_i^1, P_{-i}^0, Q^0, U^0) \\ &= \int_{p_i^1}^{p_i^0} x_i^h(s, P_{-i}^0, Q^0, U^0)\, ds \end{aligned}$$

(11) $$\begin{aligned} EV &= e(p_i^0, P_{-i}^0, Q^0, U^1) - e(p_i^1, P_{-i}^0, Q^0, U^1) \\ &= \int_{p_i^1}^{p_i^0} x_i^h(s, P_{-i}^0, Q^0, U^1) ds \end{aligned}$$

For the price change, compensating variation is simply the area under the Hicksian demand curve evaluated at the initial utility level and the two prices. Similarly, equivalent variation is simply the area under the Hicksian demand curve evaluated at the new utility level and the two prices. Figure 1 depicts these two measures for the price change.

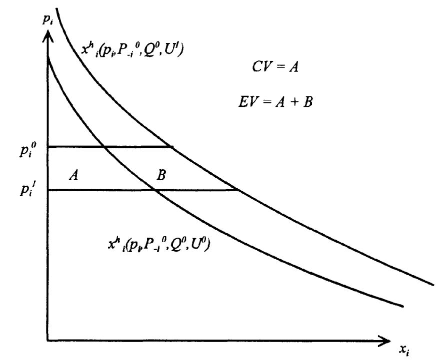

Figure 1. Compensating and Equivalent Variations for a Decrease in p_i

A few issues regarding the welfare analysis of price changes deserve mention. First, I have only presented a single price change. Multiple price changes are easily handled using a compensated framework that simply decomposes a multiple price change into a sequence of single price changes (Braeutigam and Noll 1984). I will provide an example of how to do this in the discussion of weak complementarity below. Second, the area under the ordinary (uncompensated) demand curve and between the prices is often used as a proxy for either compensating or equivalent variation. Willig (1976) has shown that in many cases this approximation is quite good, depending on the income elasticity of demand and the size of the price change. Hausman (1981) offered one approach to deriving the exact Hicksian measures from ordinary demands. Vartia (1983) offered another that uses numerical methods for deriving the exact Hicksian measures. While both methods for deriving the compensated welfare measures from ordinary demands are satisfactory, Vartia's method is very simple. Finally, we also need to consider price increases, which are conceptually the same except that the status quo price is now the lower price, $P^0 < P^1$. Both welfare measures here are negative. In the case of compensating variation, we take away a negative amount, *i.e.* give money, because the new price level makes me worse off. Similarly we would have to give up money at the old price in order to equate utility with utility at the new price, which is equivalent to saying we have a negative equivalent variation.

2.5 Welfare Measures for a Change in Nonmarket Good *j*

Now suppose we are considering an increase in nonmarket good q_j. Recall that our compensating and equivalent measures are referred to as compensating surplus (CS) and equivalent surplus (ES). The expenditure function representation of these is given as follows.

$$(12) \qquad CS = e(p^0, Q^0, U^0) - e(p^0, Q^1, U^0)$$

$$(13) \qquad \begin{aligned} CS &= e(P^0, q_j^0, Q_{-j}^0, U^0) - e(P^0, q_j^1, Q_{-j}^0, U^0) \\ &= \int_{q_j^0}^{q_j^1} p_i^v(P^0, s, Q_{-j}^0, U^0)\, ds \end{aligned}$$

$$(14) \qquad ES = e(p^0, Q^0, U^1) - e(p^0, Q^1, U^1)$$

Using the properties of the expenditure function, one can rewrite the quantity compensating and equivalent variations in an insightful form. Maler (1974) showed that the derivative of the expenditure function with respect to nonmarket good *j* is simply the negative of the inverse Hicksian demand curve for nonmarket good *j*. This derivative equals the negative of the virtual price, *i.e.* shadow value, of nonmarket good *j*. Again applying the fundamental theorem of calculus, we can rewrite the surplus measures in terms of this shadow value. Note that the properties of the integral we used to change the order of the integration limits, which then cancels out the negative factor. Similar to the notation for price changes, Q_{-j} refers to the vector of nonmarket goods left after removing q_j, and s represents q_j, along the path of integration.

$$
\begin{aligned}
ES &= e(P^0,\ q_j^0,\ Q_{-j}^0,\ U^1) - e(P^0,\ q_j^1,\ Q_{-j}^0,\ U^1) \\
&= \int_{q_j^0}^{q_j^1} p_i^v(P^0,\ s,\ Q_{-j}^0,\ U^1)\ ds
\end{aligned}
\tag{15}
$$

Figure 2 graphs the compensating and equivalent surpluses for this increase. The graph looks similar to Figure 1 except that the change is occurring in the quantity space as opposed to the price space.

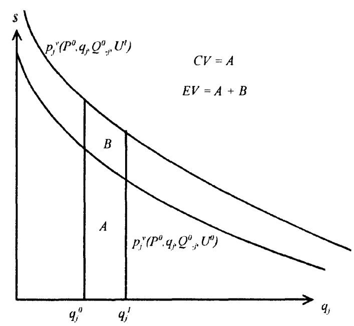

Figure 2. Compensating and Equivalent Variations for an Increase in q_i

In thinking about compensating/equivalent surpluses as opposed to the variations, it is useful to remember what is public and what is private. In the case of market goods, prices are public and the demand for the goods varies among individuals. For our nonmarket goods, the levels are public and shared by all while the marginal values vary among individuals. These rules of thumb help to differentiate between the graphic representations of compensating/equivalent variations and surpluses.

2.6 Compensating and Equivalent Variations, Willingness to Pay, and Willingness to Accept

Provided that we can agree on what constitutes the initial levels of prices and nonmarket goods, then our compensating and equivalent welfare measures are clearly defined and hopefully by now easy to understand. Two other terms, willingness to pay (WTP) and willingness to accept (WTA) compensation, are often used as substitute names for either the compensating measures or the equivalent measures. WTP is typically associated with a desirable change and WTA compensation is associated with a negative change. Consider Table 1 for a price change.

Table 1. CV, EV, WTP and WTA for a Price Change

Welfare Measure	Price Increase	Price Decrease
EV - Implied property right in the change	WTP to avoid	WTA to forgo
CV - Implied property right in the status quo	WTA to accept	WTP to obtain

Source: Freeman (1993, pg. 58)

As the table suggests, one needs to be explicit about what you are paying for when using WTP and one needs to be explicit about what you are being compensated for when using WTA. In cases where utility changes are unambiguously positive or negative, the WTP/WTA terminology works well. However, when we have combinations of desirable and undesirable changes, such as an increase in water quality accompanied by an increase in sales taxes on market goods, then WTP and WTA are less useful terms. This is true because if the policy as a whole is bad $(U^0 > U^1)$, then the compensating welfare measure is WTA and the equivalent welfare measure is WTP to avoid the policy. If the policy as a whole is good $(U^0 < U^1)$, then the compensating welfare measure is WTP to obtain the policy and the equivalent welfare measure is WTA forgoing the policy. The situation could result in mixed loses and gains leading us to measure WTA for losers and WTP for gainers the WTP/WTA terminology. For example, reintroduction of the wolf in Colorado may make conservationists better off while some ranchers are made worse off. Using equivalent or compensating welfare measures, we refer to one measure for losers and gainers.

These concepts refer to gains and losses at the individual level. As discussed in Chapter 1, there are different approaches to aggregating information from individuals to make collective choices. The Kaldor-Hicks potential compensation test is the most widely used approach to aggregating compensating and equivalent welfare measures.

3. IMPLICIT MARKETS FOR ENVIRONMENTAL GOODS

By definition, we do not explicitly purchase nonmarket goods. We do, however, purchase other goods for which demands are related to nonmarket goods. For example, my choice of where to recreate may depend on the environmental quality of the sites under consideration. Furthermore, environmental quality can influence my choice of which community to live in or which house to buy once I have decided on the community. These market links to nonmarket goods makes it possible to infer values for the demand revealed through these purchases. The specific nonmarket valuation techniques used to infer these values, called revealed preference methods, are described in Chapters 8 through 11. This section reviews some of the concepts related to inferring environmental values from market purchases.

3.1 Price Changes and Environmental Values

Suppose we are increasing nonmarket good q_1, we wish to measure the monetary value for this change, and that we determine compensating surplus to be the appropriate measure. Using the expenditure function, the only argument that changes is q_1. Q_{-1} is the vector left after removing the first element of Q.

(16) $$CS = e(P^0, q_1^0, Q_{-1}^0, U^0) - e(P^0, q_1^1, Q_{-1}^0, U^0)$$

Now I want to introduce an arbitrary price change along with this quantity change by adding and subtracting two different terms. I have not changed the size of the compensating surplus.

$$
\begin{aligned}
CS = \; & e(P^1, q_1^1, Q_{-1}^0, U^0) - e(P^0, q_1^1, Q_{-1}^0, U^0) \\
(17) \qquad & - [e(P^1, q_1^0, Q_{-1}^0, U^0) - e(P^0, q_1^0, Q_{-1}^0, U^0)] \\
& + e(P^1, q_1^0, Q_{-1}^0, U^0) - e(P^1, q_1^1, Q_{-1}^0, U^0)
\end{aligned}
$$

The second and fourth terms are the original terms in (16) and the other four are the "zero" terms. Note the arrangement of the terms. The first line is the value of the price change at the new level of q_1. The second line is the negative of the value of the price change at the initial level of q_1. The last line is the value of the change in q_1 at the new price level. If a special condition referred to as weak complementarity is satisfied, this arrangement is useful and forms the basis for the travel cost method presented in Chapter 9.

3.2 Weak Complementarity

Suppose that the compensated demand for market good one (x_1) depends upon the level of q_1 in a marginally positive way, i.e. the demand curve shifts out as q_1 increases. Further suppose that if consumption of this market good is zero, the marginal value for the change in q_1 is zero. Maler (1974) referred to this situation as weak complementarity. Now turning back to the way compensating surplus was rewritten in (17), suppose that the change in price was from the original price level to the price that chokes off demand for this weakly complementary good. This choke price is designated as $\hat{p}_1$ below.

$$
\begin{aligned}
CS = \; & e(\hat{p}_1, P_{-1}^0, q_1^1, Q_{-1}^0, U^0) - e(p_1^0, P_{-1}^0\, q_1^1, Q_{-1}^0, U^0) \\
(18) \qquad & - [e(\hat{p}_1, P_{-1}^0, q_1^0, Q_{-1}^0, U^0) - e(p_1^0, P_{-1}^0, q_1^0, Q_{-1}^0, U^0) \\
& + e(\hat{p}_1, P_{-1}^0, q_1^0, Q_{-1}^0, U^0) - e(\hat{p}_1, P_{-1}^0, q_1^1, Q_{-1}^0, U^0)
\end{aligned}
$$

By definition, demand for our weakly complementary good is zero at $\hat{p}_1$,so the last line of (18) equals zero. Now our compensating surplus is simply the change in total consumer surplus for the weakly complementary good.

$$
\begin{aligned}
CS &= e(\hat{p}_1, P^0_{-1}, q^1_1, Q^0_{-1}, U^0) - e(p^0_1, P^0_{-1}, q^1_1, Q^0_{-1}, U^0) \\
&\quad - [e(\hat{p}_1, P^0_{-1}, q^0_1, Q^0_{-1}, U^0) - e(\hat{p}_1, P^0_{-1}, q^0_1, Q^0_{-1}, U^0)] \\
&= \int_{p^0_1}^{\hat{p}_1} x^h_1(s, P^0_{-1}\ q^1_1, Q^0_{-1}, U^0)ds - \int_{p^0_1}^{\hat{p}_1} x^h_1(s, P^0_{-1}, q^0_1, Q^0_{-1}, U^0)ds
\end{aligned} \tag{19}
$$

Weak complementarity is convenient because valuing the change in the nonmarket good is possible by valuing the change in consumer surplus from the weakly complementary good. Figure 3 graphically depicts compensating surplus for this weakly complementary good.

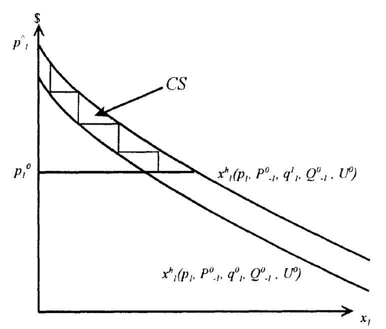

Figure 3. Weakly Complementarity of Market Good x_1 and CS for a Change in Nonmarket Good q_1

Consumption of several goods might need to be zero in order for the marginal value of q_1 to equal zero. An example is increasing the water quality at two sites along a river. The value of improving water quality might be zero if trips to both sites were zero, a joint weak complementarity condition. These concepts are similar to those I have so far presented. The difference is the way would still equal zero. However, we have multiple prices to consider. Let me give a simple example of how prices of two goods would need to be adjusted. Suppose that if demand for market good one and two are zero, then the marginal value for the change in q_1 equals zero. Compensating surplus is then given as follows. Similar to earlier notation, $P^0_{-1,-2}$ is the price vector formed by removing the first and second elements of P^0.

$$
\begin{aligned}
CS = & \int_{p_1^0}^{\hat{p}_1} x_1^h(s,\ P^0_{-1},\ q_1^1,\ Q^0_{-1},\ U^0)ds \\
& - \int_{p_1^0}^{\hat{p}_1} x_1^h(s,\ P^0_{-1},\ q_1^0,\ Q^0_{-1},\ U^0)ds \\
& + \int_{p_2^0}^{\hat{p}_2} x_2^h(\hat{p}_1,\ s,\ P^0_{-1,-2}\ q_1^1,\ Q^0_{-1},\ U^0)ds \\
& - \int_{p_2^0}^{\hat{p}_2} x_1^h(\hat{p}_1,\ s,\ P^0_{-1,-2}\ q_1^0,\ Q^0_{-1},\ U^0)ds
\end{aligned}
\tag{20}
$$

Compensating surplus is given by the change in consumer surplus resulting from the increase in q_1 for the two goods. For the first good, the change in consumer surplus is conditioned on all of the other prices being held at the original level, P^0_{-1}. For the second good, the change in consumer surplus is conditioned on the choke price of the first good, $\hat{p}_1$, and the original price for the remaining market goods, $P^0_{-1,-2}$. If we had a third good, the change in consumer surplus for the third good would be conditioned on the choke prices of goods one and two. This adjustment would be necessary for measuring changes in consumer surplus for any sequence of price changes, not just choke prices. The order of the price changes does not matter as long as we condition on the other prices correctly (Braeutigam and Noll 1984).

Before moving onto inference for marginal values, I want to briefly discuss two issues related to weak complementary goods. First, we need not rule out

Before moving onto inference for marginal values, I want to briefly discuss two issues related to weak complementary goods. First, we need not rule out a market price other than the choke price for which we obtain the condition that the marginal value of our market good is zero.[8] Any price that results in this condition allows the compensating surplus to be derived from the compensated market demands. Though we cannot rule these possibilities out, a solid case exists for this particular selection of the price change. The techniques discussed in this section will handle any such price. A second issue is the impact of incorrectly assuming weak complementarity. The last term that vanishes under weak complementarity will be positive if we incorrectly assume weak complementarity. In this case, the value inferred from the good (incorrectly) assumed to be weakly complementary will simply bound compensating surplus from below.

3.3 Household Production Framework

A slightly different approach to that presented above is the household production framework.[9] The household production framework is the basis for he defensive behavior approach to nonmarket valuation described in Chapter 11. Suppose we are interested in determining the marginal value of a single nonmarket good q_j. The household production framework posits a production relationship between a consumption goods x_p and q_j. The good produced in this process is a final product that the consumer values. Let us partition the vector X into $[X_1, x_p]$, where x_p is a good produced by the individual according to the a production process $x_p = f(I, q_j)$. I is a marketed production input, and X_1 is the vector of market goods consumed. Assuming that q_j, enters the choice problem only through production of x_p, the utility maximization problem is as follows.

$$(21) \qquad \max_{X_1, I} U(X_1, x_p) \;\; s.t. \;\; p_1 \cdot X_1 + p_I \cdot I \leq y, \;\; q_j = q_j^0, \;\; x_p = f(I, q_j)$$

(22) $$\frac{\partial U}{\partial x_p}\cdot\frac{\partial f}{\partial I} = \lambda\cdot p_I \qquad \frac{\partial U}{\partial x_p}\cdot\frac{\partial f}{\partial q} = p_{q_j}^{v}\cdot\lambda$$

From these two equations, we can solve for the marginal value of q_j.

(23) $$p_{q_j}^{v} = p_I\cdot\frac{\left(\frac{\partial f}{\partial q}\right)}{\left(\frac{\partial f}{\partial I}\right)}$$

Thus, we can derive the marginal value of q_j from the price of the input and the marginal rate of transformation between the input and q_j. The desirable property of this technique is that we do not have to model preferences. Of course we still have to model the production process. Moving away from a single input and a single produced good quickly complicates the model. For one, we need to model preferences since marginal utilities come into play. Thus we need to model the production process *and* consumer preferences thus creating an additional layer to what we have already presented above in the basic framework.[10]

3.4 The Hedonic Concept[11]

Some goods we consume can be viewed as bundles of attributes. For example, houses have distinguishing attributes such as square footage, number of bedrooms, location, and environmental attributes. Open space is a publicly owned good that is accessible to all. Being close to open space is, for some, a valuable attribute. The closer houses are to open space, the higher the price for houses that are otherwise identical. Given this price gradient, purchasers of homes can buy location relative to open space up to the point where the marginal cost of moving closer equals the marginal benefit. Hence we have an implicit market in this attribute because the home price varies according to the attribute proximity to open space. This concept underlies the hedonic nonmarket valuation technique described in Chapter 10. Other examples of attributes in the housing market are air quality, proximity to busy streets, and attribute proximity to open space. This concept underlies the hedonic nonmarket valuation technique described in Chapter 10. Other examples of attributes in the housing market are air quality, proximity to busy streets, and proximity to power lines. Environmental risk is an attribute of jobs, which are objects of choice that implicitly offer us the chance to trade off pay and on-the-job risk of injury or exposure to toxins. The important feature of the hedonic model is that an implicit market exists for attributes of goods such as proximity to open space or job risk that are not explicitly traded in markets.[12]

In the case of the home purchase, the idea is that the consumer purchases environmental quality through the house. Utility still depends on the consumption of market goods X and nonmarket goods Q, but now certain aspects of Q can be thought of as being chosen. It is important to recognize levels of rationing. For example, I do not individually purchase open space; thus, the quantity of Q is fixed. I can, however, purchase a home closer to the open space that is available. For the case of air quality, the quality gradient is fixed so far as the individual is concerned. A resident of Los Angeles cannot unilaterally affect this. The resident of Los Angeles can choose a house based on where it falls along the air quality gradient. She can care a great deal about the gradient itself in ways other than her choice of housing. For example she can live near the beach, which has relatively good air quality, and yet be subjected to really poor air quality at work downtown. Similarly I can locate near the north Boulder open space and yet care a great deal about whether the Boulder County purchases another tract of land in south Boulder. The point here is that Q can enter my utility for life at home and separately for other purposes.

The basic approach to the hedonic method is that the house is really a bundle of attributes. Since other people also care about these attributes, they are scarce and valuable. Although we pay a bundled price for the house, we are paying for the individual attributes. A way to model things on the consumer side is to partition both market goods, $X = [X_1, X_2]$ and nonmarket goods, $Q = [Q_1, Q_2]$. The second vector in both the market and nonmarket goods partition are those attributes selected through the housing purchase. The total price of the house is a function of these attributes, $p_h(X_2, Q_2)$. The maximization problem follows.

$$\text{(24)} \qquad \max_{X,\, I} \; U(X_1, X_2, Q_1, Q_2)$$

$$s.t. \;\; p_1 \cdot X_1 + p_h(X_2, Q_2) \leq y, \;\; Q_1 = Q_1^0$$

The important feature is that the consumer chooses the levels of Q_2 through the house purchase up to the point where the marginal benefit equals marginal cost. In particular, the equal marginal rate of substitution condition, called the equimarginal condition, is satisfied for any of the individual elements in Q_2 and the individual market goods in either X_1 or X_2.

$$\text{(25)} \qquad \begin{aligned} \frac{\left(\frac{\partial U}{\partial q_j}\right)}{\left(\frac{\partial U}{\partial x_i}\right)} &= \frac{\left(\frac{\partial p_h}{\partial q_j}\right)}{\left(\frac{\partial p_h}{\partial x_i}\right)} \qquad q_j \in Q_2, \;\; x_i \in X_2 \\ \frac{\left(\frac{\partial U}{\partial q_j}\right)}{\left(\frac{\partial U}{\partial x_i}\right)} &= \frac{\left(\frac{\partial p_h}{\partial q_j}\right)}{p_i} \qquad q_j \in Q_2, \;\; x_i \in X_1 \end{aligned}$$

As in the case for market goods, the combined marginal substitution relationships conceptually yield a marginal substitution curve, referred to as the bid function for the individual. Conversely, sellers typically are trying to get the most money possible for their houses. The price function, $p_h(X_2, Q_2)$, is the resulting equilibrium from the interaction of buyers and sellers. Estimating the price function using demand data provides information on the marginal values of Q_2. Additional structure will allow the estimation of the bid function which can then be used to value non-marginal changes.

3.5 When Markets Will Not Do

The concepts outlined in the earlier sections involve the use of observable market behavior to infer either the marginal value of nonmarket goods or the value for a discrete change in the nonmarket goods. All of these methods require an identifiable link between the nonmarket goods and some subset of the market goods. Furthermore, there also must be sufficient variation in the prices of the market goods and the quantities of the nonmarket goods accompanying the observed transactions to be able to statistically identify these relationships. The concepts above form the basis for revealed preference techniques: travel cost/recreational demand modeling, household production function modeling, and the hedonic price method.

Using market data to infer the value of a nonmarket good requires that values can only be inferred for individuals who used the nonmarket good, but there are cases when the demand link is unidentifiable for some individuals. No identifiable link for some people does not mean that they do not value the nonmarket good. Value for these individuals for which there is no identifiable or estimable link is referred to as non-use value or passive use value. Passive use value is the legal term used by the U.S. Federal Court of Appeals in an influential court case, *Ohio v. U.S. Department of Interior*, that gave legal standing to the concept. Drawing on earlier work from Carson, Flores, and Mitchell (1999), a brief overview of how this concept evolved follows.

In a highly influential article, Krutilla (1967) suggested that revealed preference techniques may not accurately measure societal values. The strength of his argument came through example; the paper provides no theory. Using unique resources, such as the Grand Canyon, and considering irreversible changes, Krutilla makes a number of important points.[13] First, demand for the environment has dynamic characteristics that imply value for potential use, though not current use, and that trends for future users need to be explicitly recognized in order to adequately preserve natural areas.[14] Second, some individuals may value the environment for its mere existence. Krutilla gave the example of the "spiritual descendants of John Muir, the current members of the Sierra Club, the Wilderness Society, National Wildlife Federation, Audubon Society and others to whom the loss of a species or a disfigurement of a scenic area causes acute distress and a sense of genuine relative impoverishment." Third, bequest of natural areas to future generations may be a motive for current nonusers to value preservation, particularly since given the dynamic characteristics mentioned earlier, preserving natural areas effectively provides an estate of appreciating assets. These examples obviously struck a chord with many economists. Methods and techniques were developed to formally describe the phenomena mentioned by Krutilla and to measure the associated economic value.[15]

Measuring passive use values and using them in policy analysis, particularly natural resource damage assessment, has been controversial. Much of the problem stems from the fact that passive use values, by implied definition, cannot be measured from market demand data. Economics, as a discipline, places considerable emphasis on drawing inferences regarding preference from revealed action. However, stated preference methods such as contingent valuation and the attribute-based stated preference methods discussed in Chapters 5, 6, and 7 are the only viable alternatives for measuring passive use values. These stated preference methods draw inference from hypothetical tradeoffs. From these hypothetical tradeoffs we can at least hope to learn about preferences of individuals who hold passive use values. However given the economics profession's preference for inference from actual choices, the considerable skepticism regarding passive use values is understandable. Skepticism can occur on two levels. The first level of skepticism involves the measurement of passive use values because of the need to use stated-preference techniques. The second level involves the idea of whether or not passive use values even exist. The second position is fairly extreme, and I can reject this notion based on my own preferences. The former concern is valid and plenty of discussion exists in the literature.[16]

In this section, I will discuss how passive use values have been viewed conceptually. While my discussion will focus on compensating surplus, the issues also apply to equivalent surplus. Recall the decomposition of compensating surplus into the value of a price change, the new price being the choke price, and the value of the quantity change at the higher price level. Weak complementarity called for the final term to equal zero. McConnell (1983) and Freeman (1993), define passive use value as this last term.

$$\begin{aligned} CS = {} & e(\hat{p}_1, p^0_{-1}, q^1_1, Q^0_{-1}, U^0) - e(p^0_1, p^0_{-1}\, q^1_1, Q^0_{-1}, U^0) \\ & - [e(\hat{p}_1, p^0_{-1}, q^0_1, Q^0_{-1}, U^0) - e(p^0_1, p^0_{-1}, q^0_1, Q^0_{-1}, U^0)] \\ & + PUV \end{aligned} \tag{26}$$

This definition does not have much practical appeal because we could choose any good that is not a necessary good, measure the value from the first two lines of (26) and end up with a measure of passive use value. Since we could do this for each good that is not a necessary good or any combination of goods in this category, we could derive multiple measures of passive use value.

Another conceptual definition was suggested by Hanemann (1995). Hanemann had a specific form of utility in mind, $U(X, Q) = T\,[g(X, Q), Q]$. This functional form suggests that choices of market goods will be influenced by Q and so market demand data could reveal the part of the relationship involving $g(X, Q)$, but not the part where Q enters directly.[17] Hanemann defines passive use value and use value according to the following two identities.

$$(27) \qquad T[g(X(P, Q^0, y - PUV), Q^1] = T[g(X(P, Q^0, y), Q^0]$$

$$(28) \qquad T[g(X(P, Q^1, y - PUV - UV), Q^1] = T[g(X(P, Q^0, y), Q^0]$$

The definitions implied by (26) and by (27) together with (28) decompose compensating surplus into two parts for which the sum of the parts equals the whole. Intuitively Hanemann's definition works in reverse of the decomposition in (26). Because the same preferences can be defined differently, passive use values is a somewhat tenuous theoretical concept.[18] Furthermore, neither definition is easy to implement since the first decomposition requires one to choose the market goods for which demand is choked. Using separate measurement, it is difficult to if not impossible to elicit either of these concepts to subjects in a stated preference study.

Carson et al. (1999) provided a definition based on methodological considerations. "Passive-use values are those portions of total value that are unobtainable using indirect measurement techniques which rely on observed market behavior."[19] This definition was conceived with the researcher in mind as opposed to a theoretical foundation. Revealed preference techniques can miss portions of value because of the form of preferences such as those used in the Hanemann definition. We are typically after compensating or equivalent surplus, also referred to as total value in this literature. The definition implies that passive use value is the residual based on what is available. Practically and conceptually passive use value is not unique. The important social issue is the need to incorporate the values of all of those who value the nonmarket good.

To do so requires us, at times, to turn to stated preference techniques if we believe that passive use values are likely to be decisive.

4. NONMARKET VALUES IN A DYNAMIC ENVIRONMENT

Models of inter-temporal choice abound in economics. Most models assume that utility across time periods is represented by a sum of utility functions from each of the time periods. This sum involves a time preference component that is typically assumed to be the discount factor, $\beta = 1/(1 + r)$.

$$U = \sum \beta^t u(X_t, Q_t) \tag{29}$$

The time horizon, T, can be either finite or infinite. The analog to the earlier problem is that the consumer still allocates income toward the purchase of market goods, but now total income is in present value form, $Y = \sum \beta^t y_t$ where y_t is income in period t. A simple time separable model such as this can be used to extend the earlier concepts of value developed for a single period into a dynamic framework. I am going to assume that X_t is a composite good consisting of expenditures on the market goods in period t. Thus, expenditures on market goods and levels of nonmarket goods exist in each period that drive utility. The important feature of this model is that the individual efficiently allocates income between periods. Namely that the marginal benefit of spending on market goods in each period is equated in present value terms.

$$\frac{\partial u(X_0, Q_0)}{\partial X} = \beta^t \frac{\partial u(X_t, Q_t)}{\partial X} \tag{30}$$

This condition must hold for all t under optimal income allocation. The consideration is what a marginal change in Q_t is worth in the current period. We know from the earlier static analysis that in period t, the marginal value for the change will be given by $p_t^v = \left(\partial u(X_t, Q_t)/\partial Q\right)/\left(\partial u(X_t, Q_t)/\partial X\right)$. By (30), the value today for the marginal change in the future will simply be given by $\beta^t\ p_t^v$. Thus the marginal value of Q in the dynamic model is simply the discounted value of the marginal value in the respective period.

For discrete changes, we would like the total amount of income today that the consumer is willing to give up for some change in the sequence of nonmarket goods, $\{\ Q_t\ \}$. Brackets are used because we must represent nonmarket goods' levels for T periods and T may be infinite. Assuming we are interested in a compensating measure of welfare, the logical extension from the marginal case is to use the present value discounted stream of compensating surplus in each period as our welfare measure. This generalization meets our needs provided that the allocation of income is unaffected by the sequence of nonmarket goods, $\{\ Q_t\ \}$. However, when income allocation is affected by this sequence, the proposed welfare measure, present value discounted compensating surplus in each period essentially values the changes in the sequence while imposing that the income allocation stay fixed. Thus for increases in nonmarket goods, the present value of the compensating surpluses will underestimate the desired welfare measure, and the present value of equivalent surpluses will overstate the amount. The reasoning is that for both cases, the ability to reallocate income is worth money. For the compensating measure, we would pay for this flexibility over the restricted case measured by the present value of the compensating surpluses from each period. Thus the divergence between the compensating measure with flexibility and our approximation. For equivalent surplus, the ability to reallocate income makes giving up the change in $\{\ Q_t\ \}$ not as bad. For decreases in $\{\ Q_t\ \}$, the opposite is true in both cases.

Practically speaking, the standard practice is to estimate the periodic benefits and then discount them. The choice of the discount rate is a sensitive issue that I will not address here.[20] Since we are estimating future benefits in today's dollars, the appropriate discount rate should not include an inflationary component.

4.1 Values in an Uncertain World

A great deal of uncertainty exists regarding our willingness to trade money for nonmarket goods. For example, the levels of nonmarket goods provided by a policy may be uncertain, prices of market goods that will occur once the policy is implemented may be uncertain, and the direct cost if the policy is enacted may be uncertain. Policies can affect the distributions of all of these random variables. The question then becomes one of how do we extend the welfare measures developed in the earlier section to cases of uncertainty?

To begin, I will exclusively consider uncertainty regarding the distribution of Q, assuming away time.[21] Q can accommodate things as different as the amount of open space that will actually be purchased by a bond initiative or the level of environmental risk associated with living or working in a given area. In relation to the earlier models, we now assume that individuals allocate income toward the purchase of market goods according to expected utility maximization.

$$\max_{X} E_Q[\ U(X,\ Q)\ ] \quad s.t. \quad p{\cdot}X \le y \tag{31}$$

Here, the allocation of income depends on the distribution of Q, which involves different possible levels instead of a particular level. The distribution of Q can be discrete or continuous. The maximized expected utility depends on the prices of the market goods, income, and the probability distribution of Q. Policies now act on the distribution associated with Q. Letting F denote the probability distribution of Q, maximized expected utility is then given by an indirect utility function, $v^E(P, y, F)$.[22] The concept that I focus on is option price. Option price is defined as the amount of income given that makes the individual indifferent between the status quo level of expected utility and the new expected utility under the changed distribution.

$$v^E(p,\ y\ -\ OP,\ F^1)\ =\ v^E(p,\ y,\ F^0) \tag{32}$$

In cases such as bond issues for the provision of open space, we typically pay some single, specified amount over time. The amount of open space that will actually be purchased is uncertain. In this case, option price is a very close analog to compensating surplus from before. In fact, contingent valuation surveys are generally measuring option price since some uncertainty almost always exists. Other important concepts involving environmental valuation and uncertainty are not covered here.[23]

4.2 Averting Expenditures

When facing environmental risks, individuals may independently undertake costly risk reduction. Examples include the purchase of bottled water and purchasing air bags for the car, to name a few. Since individuals spend money on providing a more favorable probability distribution of the nonmarket good, averting expenditures offer an avenue for inferring the value of collective policies that affect the distribution. The idea here is that the probability distribution can be favorably affected through individual inputs as well as collective inputs. Let E_I denote the individual's expenditure dedicated toward individual improvement of the distribution of Q, and let E_G denote the government's expenditure dedicated toward improving this distribution. Now the individual chooses both the level of market expenditures and the level of E_I subject to E_G. As in the previous section, F is the probability distribution of Q. At the optimal level of E_I, the indirect expected utility function becomes $v^E(P, y, F(E_I, E_G))$. A necessary condition for optimization is that the marginal benefit of more E_I equals the marginal utility of additional income.

$$(33) \qquad \frac{\partial v^E}{\partial F}\frac{\partial F}{\partial E_I} = \lambda$$

The marginal value of additional government expenditure dedicated toward improving the distribution of Q, denoted p_G^v, can be represented as the marginal utility of the expenditure function divided by the marginal utility of income.

$$(34) \qquad p_G^v = \frac{\partial v^E}{\partial F}\frac{\partial F}{\partial E_G}\frac{1}{\lambda}$$

From (33) and (34), we can solve for the marginal value of E_G. The way in which E_I enters the problem, the marginal value of E_G reduces to what is similar to the marginal rate of transformation for inputs.

$$p_G^v = \frac{\dfrac{\partial v^E}{\partial F}\dfrac{\partial F}{\partial E_G}}{\dfrac{\partial v^E}{\partial F}\dfrac{\partial F}{\partial E_I}} = \frac{\dfrac{\partial F}{\partial E_G}}{\dfrac{\partial F}{\partial E_I}} \tag{35}$$

In this case, we only need to understand the relative production possibilities between private and public expenditures. This technique is conceptually similar to the household production framework. As with household production framework, if expenditures made toward improving the probability distribution also affect other aspects of utility, the marginal value expression is more complicated than (35).

5. PARTING THOUGHTS

All of the models above are based on the assumptions that individuals understand their preferences and make choices so as to maximize their welfare. Even under optimal conditions, inferring values for nonmarket goods is difficult. The nonmarket valuation practitioner needs to understand these concepts before heading into the field; to do otherwise could prove costly. There has never been a shortage of critics of welfare economics, either from inside or outside the profession. The underlying concepts need to be challenged, refined if possible, and even discarded if necessary. Nonmarket valuation researchers are on the cutting edge of these conceptual issues, a necessary trend that will undoubtedly continue.

NOTES

1 These topics alone could constitute an entire book. My treatment of each must be brief. For those launching a career in this area, I recommend Freeman (1993) and Hanley, Shogren, and White (1997).

2 The utility function is ordinal in the sense that many different functions could be used to equally represent a given preference ordering. For a complete discussion of preference orderings and their representations by utility functions, see Kreps (1990) or Varian (1996).

3 As discussed below, I can choose goods that have environmental quality attributes, *e.g.* air quality and noise. These goods are rationed in the sense that I cannot unilaterally improve ambient air quality or noise level at my current house. I can move to a new location where air quality is better, but I cannot determine the level of air quality at my current location.

4 It may be the case that I have to pay for Q^0. Rather than include this payment in the budget constraint, I simply consider income to already be adjusted by this amount. Because the levels of the nonmarket goods are not individually chosen, we need not include payments for nonmarket goods in the budget constraint.

5 To clarify notation, $p{\cdot}X = p_1{\cdot}x_1 + p_2{\cdot}x_2 + \ldots + p_n{\cdot}x_n$ where p_i is the price of market good i.

6 Interest over the difference in size has received considerable attention. For price changes, Willig (1976) provides an analysis. For quantity changes, see Randall and Stoll (1980) and Hanemann (1991). Hanemann (1999) provides a comprehensive, and technical review of these issues.

7 Roy's identity states that the derivative of the expenditure function with respect to price i is simply the Hicksian demand for good i. The fundamental theorem of calculus allows us to write the difference of two differentiable functions as the integral over the derivative of that function.

8 An example is the case of weak substitutability provided in Feenberg and Mills (1980).

9 Other approaches to infer the marginal value of a nonmarket good, such as Neill (1988), Larson (1992), and Flores (1996).

10 We must model preferences and production technology if we have a produced good that uses and nonmarket good q and the consumer values q outside of the production process.

11 Hedonism - pursuit of or devotion to pleasure.

12 The classic citations in this area are Griliches (1971), Rosen (1974), and Palmquist (1984).

13 Cicchetti and Wilde (1992) have argued that Krutilla's arguments, and hence passive use value, only apply to highly unique resources. However in footnote 5, Krutilla notes "Uniqueness need not be absolute for the following arguments to hold."

14 In discussing trends, Krutilla gives the example of the evolution from a family that car-camps to a new generation of back packers, canoe cruisers, and cross-country skiers.

15 The terms option value, preservation value, stewardship value, bequest value, inherent value, intrinsic value, vicarious consumption value, and intangible value have been used to describe passive use values. Carson *et al.* (1999) note these are motivations rather than distinct values.

16 See Portney (1994), Hanemann (1994), and Diamond and Hausman (1994).

17 The special case where $g(X, Q) = g(X)$ has been referred to as the hopeless case because the ordinary demands are independent of the levels of Q, leaving no hope for recovering the value of Q from demand data.

18 Dividing passive use value into bequest value, existence value, and the like will provide similarly inconclusive results. The decompositions will not be unique.

19 Maler, Gren, and Folke (1994) similarly define use values as those values that rely on observed market behavior for inference.

20 For examples, see Fisher and Krutilla (1975), Horowitz (1996), Porter (1982), and Schelling (1997).

21 Time is an important dimension and uncertainty transcends time. However, there is not enough space to cover time and uncertainty together.

22 In accordance with standard probability theory, F consists of a sample space of outcomes and a probability law for all subsets of the sample space that satisfies the properties of a probability measure.

23 Influential papers in this area include Graham (1981), Weisbrod (1964), Schmalensee (1972), and Arrow (1974). Freeman (1993) provides a fairly comprehensive overview.

REFERENCES

Arrow, K. J. and A. C. Fisher. 1974. Environmental Preservation, Uncertainty, and Irreversibility. *Quarterly Journal of Economics* 88 (2):312-319.

Becker, G. S. 1993. Nobel Lecture: The Economic Way of Looking at Behavior. *Journal of Political Economy* 101 (3):385-409.

Braeutigam, R., and R. G. Noll. 1984. The Regulation of Surface Freight Transportation: The Welfare Effects Revisited. *Review of Economics and Statistics* 66:80-87.

Carson, R. T., N. E. Flores, and R. C. Mitchell. 1999. The Theory and Measurement of Passive-Use Value. In *Valuing Environmental Preferences, Theory and Practice of the Contingent Valuation Method in the US, EC, and Developing Countries*, edited by I. J. Bateman and K. G. Willis. Oxford: Oxford University Press.

Cicchetti, C. J., and L. L. Wilde. 1992. Uniqueness, Irreversibility, and the Theory of Nonuse Values. *American Journal of Agricultural Economics* 74:1121-1125.

Diamond, P. A., and J. A. Hausman. 1994. Contingent Valuation: Is Some Number Better than No Number? *Journal of Economic Perspectives* 8 (4):45-64.

Feenberg, D., and E. S. Mills. 1980. *Measuring the Benefits of Water Pollution Abatement.* New York: Academic Press.

Fisher, A. C., and J. V. Krutilla. 1975. Resource Conservation, Environmental Preservation, and the Rate of Discount. *Quarterly Journal of Economics* 89 (3):358-370.

Flores, N. E. 1996. Reconsidering the Use of Hicks Neutrality to Recover Total Value. *Journal of Environmental Economics and Management* 31:49-64.

Freeman, M. A. III. 1993. *The Measurement of Environmental and Resource Values, Theory and Methods.* Washington, DC: Resources for the Future.

Graham, D. A. 1981. Cost-Benefit Analysis Under Uncertainty. *American Economic Review* 71 (4):715-725.

Griliches, Z. 1971. *Price Indexes and Quality Change.* Cambridge: Harvard University Press.

Hanley, N., J. F. Shogren, and B. White, eds. 1997. *Environmental Economics: In Theory and Practice.* New York: Oxford University Press.

Hanemann, W. M. 1991. Willingness to Pay and Willingness to Accept: How Much Can They Differ? *American Economic Review* 81:635-647.

Hanemann, W. M.. 1994. Valuing the Environment Through Contingent Valuation. *Journal of Economic Perspectives* 8 (4):19-44.

Hanemann, W. M. 1995. Contingent Valuation and Economics. In *Environmental Valuation: New Perspectives*, edited by K. G. Willis and J. T. Corkindale. Wallingford: CAB International.

Hanemann, W. M. 1999. The Economic Theory of WTP and WTA. In *Valuing Environmental Preferences, Theory and Practice of the Contingent Valuation Method in the US, EU, and Developing Countries*, edited by I. J. Bateman and K. G. Willis. Oxford: Oxford University Press.

Hausman, J. A. 1981. Exact Consumer's Surplus and Deadweight Loss. *American Economic Review* 71 (4):662-676.

Hicks, J. R. 1943. The Four Consumer's Surpluses. *Review of Economic Studies* XI:31-41.

Horowitz, J. K. 1996. Environmental Policy under a Non-market Discount Rate. *Ecological Economics* 16 (1):73-78.

Kreps, D. M. 1990. *A Course in Microeconomic Theory*. Princeton: Princeton University Press.

Krutilla, J. V. 1967. Conservation Reconsidered. *American Economic Review* 57:777-786.

Larson, D. M. 1992. Further Results on Willingness to Pay for Nonmarket Goods. *Journal of Environmental Economics and Management* 23:101-122.

Maler, K. 1974. *Environmental Economics: A Theoretical Inquiry*. Baltimore: Johns Hopkins Press for Resources for the Future.

Maler, K., I. Gren, and C. Folke. 1994. Multiple Use of Environmental Resources: A Household Production Approach to Valuing Natural Capital. In *Investing in Natural Capital, The Ecological Economics Approach to Sustainability*, edited by A. Jansson, M. Hammer, C. Folke and R. Costanza. Washington, DC: Island Press.

McConnell, K. E. 1983. Existence and Bequest Value. In *Managing Air Quality and Scenic Resources at National Parks and Wilderness Areas*, edited by R. D. Rowe and L. G. Chestnut. Boulder: Westview Press.

Neill, J. R. 1988. Another Theorem on Using Market Demands to Determine Willingness to Pay for Non-traded Goods. *Journal of Environmental Economics and Management* 15:224-232.

Palmquist, R. B. 1984. Estimating the Demand for Characteristics of Housing. *Review of Economics and Statistics* 64 (3):394-404.

Porter, R. C. 1982. The New Approach to Wilderness Preservation Through Benefit-Cost Analysis. *Journal of Environmental Economics and Management* 9:59-80.

Portney, P. R. 1994. The Contingent Valuation Debate: Why Economists Should Care. *Journal of Economic Perspectives* 8 (4):3-18.

Randall, A., and J. R. Stoll. 1980. Consumer's Surplus in the Commodity Space. *American Economic Review* 71:449-457.

Rosen, S. 1974. Hedonic Prices and Implicit Markets: Product Differentiation in Perfect Competition. *Journal of Political Economy* 82 (1):34-55.

Schelling, T. C. 1997. Intergenerational Discounting. In *The Economics of Global Warming*, edited by T. Tietenberg. Cheltenham, U.K. and Lyme, N.H.: Elgar.

Schmalensee, R. 1972. Option Demand and Consumer's Surplus: Valuing Price Changes Under Uncertainty. *American Economic Review* 62 (6):813-824.

Varian, H. R. 1996. *Intermediate Microeconomics: A Modern Approach.* New York: W. W. Norton.

Vartia, Y. O. 1983. Efficient Methods of Measuring Welfare Change and Compensated Income in Terms of Ordinary Demand Functions. *Econometrica* 51 (1):79-98.

Weisbrod, B. A. 1964. Collective-Consumption Services of Individual-Consumption Goods. *Quarterly Journal of Economics* 78:471-477.

Willig, R. D. 1976. Consumer's Surplus Without Apology. *American Economic Review* 66:589-597.

Chapter 3

COLLECTING SURVEY DATA FOR NONMARKET VALUATION

Patricia A. Champ
U.S. Forest Service, Rocky Mountain Research Station

1. INTRODUCTION

The unique nature of environmental and natural resource amenities makes valuation a challenge in many respects. Prices reflect aggregate societal values for market goods but nonmarket goods lack an analogous indicator of value. Estimation of nonmarket values requires that the researcher either ask the public about their values for the good of interest or use existing data to infer the values. Either approach is sensitive to the quality of the data used for the estimation, which is why this chapter focuses on how to collect data for estimating nonmarket values. Understanding where such data come from and the procedures used to collect the data will make the descriptions of the nonmarket valuation techniques in Chapters 4-11 more meaningful.

Both primary and secondary data are used to estimate nonmarket values. The stated preference approaches (Chapters 4-7) mainly rely on primary data collected with surveys. For example, the contingent valuation method is a survey-based method that directly elicits individuals' willingness-to-pay for a particular nonmarket good. Revealed preference approaches to nonmarket valuation are more likely to use secondary data or some combination of primary and secondary data. For example, a hedonic model that values open space infers the value of an open space property based on the price of houses located in a defined geographic area near the open space property. The housing price data can be collected from an existing source such as housing sale records or property tax records.

If a researcher uses secondary data, it is especially important to understand how the data were collected and the types of errors that could be associated with the data. In Section 3, I describe some useful sources for secondary data and issues to consider when using secondary data. The rest of this chapter describes the steps involved in the design and administration of a nonmarket valuation survey. Because no standard operating procedures assure the collection of quality data, the researcher needs to make subjective decisions in the course of developing and implementing a survey. This chapter offers guidance about these decisions.

2. STUDY OBJECTIVES

The researcher must identify the objective(s) of the study prior to determining what type of data to use for the nonmarket valuation. Broad objectives might be to inform policy, assess the magnitude of damage to a natural resource, or address a research question. Within the broad study objectives, the researcher should clearly identify the specific economic question. For example, does the researcher want to measure equivalent surplus or compensating surplus? The economic question should be stated within the framework described in Chapter 2. If the study addresses a research question, the researcher should also develop testable hypotheses. The study objectives and the specific economic question will help determine the most appropriate nonmarket valuation technique. In turn, the technique will determine whether existing data can be used or whether an original data collection effort is needed. If primary data are to be collected, then decisions about the sampling techniques, the mode of administration, survey content, and type of analyses will be made in light of the specific economic question and the overall study objectives.

3. COLLECTING SECONDARY DATA

As mentioned, estimating nonmarket values using existing data is often possible. Although use of secondary data can save substantial expense and effort, secondary data should only be used if the researcher is confident about its quality. Many large national surveys such as the U. S. Decennial Census are well designed and collect data that could be used for nonmarket valuation. The U.S. Census Bureau administers many other useful demographic and economic surveys such as the National Survey of Fishing, Hunting, and Wildlife-Associated Recreation.[1] This and most of the data collected by the Census Bureau can be downloaded from its website (www.census.gov). Various types of social science data are also archived and made available from the Inter-University Consortium for Political and Social Research at the University of Michigan and the National Opinion Research Center at the University of Chicago.

A consideration when using secondary data is whether the data contain the requisite variables for estimating the desired nonmarket value. It is not just an issue of having the variables or not–the variables must be in the appropriate units. If, for example, one wants to estimate a travel cost model using some expenditures that are in annual terms and other expenditures that are reported on a per trip basis, the expenditures must be converted so that both are on a per trip basis or an annual basis.

Another consideration when using secondary data is whether the data are representative of the appropriate population. Sampling is described in Section 4. Key sampling issues that need to be considered when collecting primary data–in terms of defining the relevant study population and making sure the final data are representative of that population–should also be considered when assessing the appropriateness of secondary data.

4. DEVELOPING SURVEYS FOR PRIMARY DATA COLLECTION

Dillman (1991, p. 228) describes the general characteristics of a good survey as follows:

> A good sample survey, by whatever method, is one in which all members of a population have a known opportunity to be sampled for inclusion in the survey (noncoverage error is avoided); the people to be surveyed are sampled by random methods in sufficiently large numbers to provide a desired level of precision (sampling error is limited); questions are selected and phrased in ways that result in people

providing accurate information (measurement error is avoided); and everyone who is included in the sample responds (nonresponse error is avoided).

However, surveys rarely, if ever, avoid all these sources of errors. The realistic goal of good survey design and implementation is to minimize relevant errors and make post data collection adjustments when necessary.

The rest of this chapter focuses on developing quality surveys for collection of nonmarket valuation data. The steps involved are listed in Table 1. Although, the following sections and Table 1 are presented sequentially, decisions associated with the various steps are often made jointly. For example, if the project budget does not allow for a sufficient number of in-person surveys for accurate estimation, the researcher must consider an alternative mode of administration such as telephone or mail. The chosen mode of administration affects the choice of an appropriate sample frame, and the availability of sample frames will feed into the decision about how to administer the survey. The following sections provide options for each step in the survey process.

Table 1. Steps to Collecting Data for a Nonmarket Valuation Study

1.	Choose a mode of administration
	In-person
	Telephone
	Mail
	E-mail or web
	Central site
	Mixed modes
2.	Develop the survey materials
	Identify desired measures
	Design questions to obtain desired measures
	Implement qualitative research methods to design survey materials
3.	Administer the survey
	Conduct pilot study
	Conduct final survey
4.	Prepare data
	Develop codebook
	Data entry
	Clean data

4.1 Sample Design

The sample design specifies the population of interest, the sampling frame, and the techniques for drawing a sample from the sample frame. Throughout the survey process, the researcher should be aware of issues that could give rise to errors in the survey data. Such errors can influence the entire study. One source of error introduced in the sample design stage is coverage error, which arises when the population of interest does not correspond with the sample frame. If, for example, the population of interest is all residents of the U.S., a sample frame of registered voters will not fully represent that population, as many residents of the U.S. do not register to vote. Other errors may occur if procedures for selecting sampling units from the frame are flawed. Consider a sample frame of Colorado residents that is organized by county. If the researcher selects the sample by taking the first 5,000 sample units from the sample frame, the counties listed at the top of the sample frame will be over represented and counties listed at the end of the sample frame may be missed completely.

Another source of error, non-response error, results from discrepancies between the actual survey participants and the sample frame. Non-response error occurs when the individuals who respond to the survey are systematically different from those who do not respond. Errors are only a problem if they affect the measures of interests. If a researcher is using a travel cost model to estimate the value of recreational angling at a particular lake and uses a sample that has significantly more female anglers in it than does the actual angler population, the travel cost estimates will be biased if female anglers differ from males on the measures used to estimate the travel cost model such as distance traveled or expenditures.

4.1.1 Study Population

Nonmarket valuation is used in many contexts including benefit-cost analyses, natural resource damage assessments, and policy analyses. Each of these situations may suggest a different study population. An important determination when defining the study population is whose values are relevant. In some cases, defining the study population will be straightforward. Consider a university that is interested in measuring the benefits to students of providing Internet access funded using student fees. The appropriate study population would be students currently enrolled at the university. It would be important to include the students who have university Internet accounts as well as students who do not have such accounts since all students pay for the service. Including all students in the defined study population would also allow for measurement of the benefits to both current users and potential users of the Internet service. In this example, the benefits associated with a university-provided Internet service would largely accrue to users. Defining the relevant study population for some other use values, such as recreation opportunities in National Parks, can become more complicated when considering how to identify both current and potential users.

Identification of the relevant study population for measurement of passive use values is particularly challenging, as the benefits may extend beyond the population responsible for paying for the good. Loomis (2000, p. 312) argues "If the public good provides benefits well beyond the immediate jurisdiction where the good is located, then either federal grants-in-aid or even federal provision may be needed to improve the allocation of resources involving the public good. Comparing only the local public benefits to marginal cost of supply will result in underprovision if substantial spillover benefits to other non-payers are ignored." Likewise, this issue of whose benefits should be measured has been described by Smith (1993, p. 21) as ". . . probably more important to the value attributed to environmental amenities as assets than any changes that might arise from refining our estimates of per unit values." Empirical studies of goods with a large passive use value component (Loomis 2000; Sutherland and Walsh 1985) have verified that a only small percentage of the aggregate economic benefit is accounted for by the immediate jurisdiction where the good is located. Unfortunately, no simple rule of thumb exists for defining the study population. I recommend thoughtful consideration of who will pay for the good as well as who will benefit from its provision.

4.1.2 Sample Frame

The sample frame is the list from which sample units are chosen. Ideally the sample frame perfectly matches the study population, but that is rarely the case. Strictly speaking, generalization from the survey sample should be made only to the sample frame, not to the study population. It is recommended that a sample frame be chosen that is as close to the study population as possible. To offer another example, if the intent of a study is to provide information about all households in a specified geographic area, using the area telephone listings for some areas of the U.S. might not be a good sample frame because some households do not have telephones and many more do not list their telephone number. In such cases, a sample frame is commonly developed from multiple sources such as telephone listings augmented with voter and/or automobile registration lists. Such lists can be purchased from survey research firms. Although coverage of the study population using such lists may not be perfect, it is often very good.

Occasionally nonmarket valuation studies are conducted for well-defined study populations for which complete lists exist. For example, McCollum, Haefele, and Rosenberger (1999) conducted a mail survey to estimate the willingness-to-pay of Colorado anglers for various fishery management scenarios. The study population was Colorado resident 1997 annual fishing license holders. The sample frame was the list of Colorado residents who purchased annual fishing licenses in 1997. The sample frame used in the McCollum, Haefele, and Rosenberger study covered the entire study population.

If the population of interest is national, it is very difficult to develop a good sample frame that includes mailing addresses or telephone numbers. Therefore, many national surveys are conducted in-person using sampling techniques that do not require an initial sample frame. Another frequently used option is to conduct a phone survey using random digit dialing to identify potential respondents. Random digit dialing, a technique for contacting households without an initial list of telephone numbers, is described in section 4.2.2.

After the sample frame has been identified, budget constraints usually require the selection of only a portion of the sample units to receive the survey. The sample selection process involves two decisions: How will sample units be chosen? and How many sample units will be chosen?

4.1.3 Sampling Procedures

There are basically two types of sampling procedures: non-probability sampling and probability sampling. Examples of non-probability sampling include using students in a class or recruiting subjects in a shopping mall. With such sampling techniques, each individual in the population does not have a known non-zero probability of being chosen for the sample. Non-probability sampling is best suited for studies that do not require generalization from the survey sample to a broader population. Such samples are frequently used to investigate methodological questions such as how behaviors or responses vary under different treatments. The field of experimental economics largely relies on observations of behaviors using non-probability samples. Harrison and Lesley (1996) have even recommended the use of non-probability convenience samples to estimate a behavioral model (for estimating contingent values). This model uses sample averages of population characteristics to predict behavior for the population. This approach, however, remains controversial. I recommend the use of probability sampling if generalizations from the survey sample to a broader population are to be made.

With probability sampling, every unit in the sample frame has a known nonzero probability of being chosen for the sample. Probability sampling allows for making inferences from the sample to the broader study population (or sample frame if the frame does not adequately represent the study population). If statistical inference is the goal of a nonmarket valuation survey, probability sampling is most appropriate. Several probability sampling techniques exist, the most straightforward of which is simple random sampling. Simple random sampling requires a list of the entire study population. From the list, sample units are randomly chosen. A simple random sample is desirable because every sample unit has an equal probability of being chosen, and it is not necessary to weight the final data (unless the data need to be weighted to correct for sample non-response). If the population list is very long, simple random sampling may be cumbersome and it might be better to use a systematic sampling technique in which the first sample unit is randomly chosen and after that every n^{th} unit is chosen.

If the study population is geographically dispersed, a simple random sample or systematic sample may result in sample units that are also geographically dispersed. For in-person surveys, the geographic dispersion can be quite costly if interviewers are traveling great distances between each interview. Therefore, random sampling is best suited for telephone, mail, or e-mail surveys if a complete sample frame is available.[2]

Another commonly used probability sampling technique is stratified sampling. With stratified sampling, the entire study population is divided into non-overlapping subpopulations (strata) based on some measure that is available for the initial study population list. Within each stratum, a variety of methods can be used to select the sample units. For example, if sample units are randomly chosen from each stratum, the procedure is called stratified random sampling. Different methods for selecting the sample units can be used within the various strata.

Stratified sampling is primarily used in three situations. First, stratified sampling can be used to insure adequate sample size within a strata for separate analysis. This is important when the strata are of particular interest. For example, consider a nonmarket valuation study with the primary objective of estimating mean willingness-to-pay for improved access to a recreational area and a secondary goal of comparing the mean willingness-to-pay between urban and rural populations. In this situation, a simple random sample may not provide an adequate number of rural respondents to allow for detection of a statistical difference. The sample could be stratified into rural and urban strata with rural stratum more heavily sampled to insure an adequate number of respondents for analyses. Another situation in which stratified sampling is used is when the variance on a measure of interest is not equal among the strata. In this case, more of the sample can be drawn from the stratum with the larger variance to increase the overall sample efficiency. This situation is most common when the sample is stratified into institutional populations such as educational systems, hospitals, or corporations, and would not likely occur with samples of individuals which are most common for nonmarket valuation studies. In the rural-urban example above, there could be higher variance of willingness-to-pay for the urban population relative to the rural population. The third situation in which a stratified sample might be used is when the cost of obtaining a survey response differs by strata. An example of this would be an Internet survey combined with a phone survey for individuals who do not have an Internet account. Survey responses from the phone stratum will be more costly than those from the Internet stratum. For a fixed budget, drawing a disproportionate amount of the sample from the lower cost stratum would be optimal, in terms of sample variance. Sudman (1983) provides the details of how to stratify the sample for strata with differing costs.

A third type of probability sampling, cluster sampling, is used with in-person surveys to minimize the travel costs between interviews. The general idea is that respondents are chosen in groups or clusters. For example, a simple cluster sample may define blocks in a city as clusters. First blocks are selected, then survey participants within each of the blocks are chosen. Cluster sampling can be quite complicated, and consulting a survey statistician to design the sampling procedures is a good idea. Frankel (1983) provides a detailed explanation of cluster sampling.

Multistage area sampling is a similar technique for in-person surveys that does not require a complete sample frame. A multistage area sample involves first sampling geographic regions, then sampling areas within each region. Carson et al. (1992) implemented a multistage area sample with their contingent valuation study of the lost passive use values resulting from the Exxon Valdez oil spill. In the first stage, they sampled 61 counties or county groups from a list of all counties in the U.S. In the second stage, 330 blocks were chosen from within the 61 counties. Finally, 1,600 dwelling units were chosen from within the 330 blocks. This sample design with appropriate sample weighting allowed Carson et al. to generalize the results of their study to the population of the U.S.

4.1.4 Sample Size

Closely related to choosing the sample is deciding how many sample units to choose. Following are a couple considerations for determining the sample size. First, sample error is a function of sample size. Sample error arises because a sample does not provide complete information about the population of interest. For small study populations, a relatively large proportion of the population is needed in the sample to maintain an acceptable level of sampling error. For example, for a study population of size 1,000, approximately 200-300 observations are required for ±5% sampling error. The sensitivity of the sampling error to sample size decreases as the population increases in size. Whether the size of the study population is 10,000 or 100,000,000, a sample size of approximately 380 is needed ±5% sampling error, which is not much more than the sample size needed for the previous example with a study population of 1,000 (see Salant and Dillman 1994, p. 55 for a table of final sample sizes for various population sizes and characteristics).

A second consideration is that the power of statistical testing is related to the sample size. The power function, which measures the probability of rejecting the null hypothesis when it is false, increases as the sample size increases. One must consider that all the statistical tests that will be conducted with the data. If statistical testing will use subsamples of the survey data, the number of observations in each of the subsamples must be adequate. Mitchell and Carson (1989) provide a nice exposition of the power of contingent valuation hypothesis tests. Of course the final sample size is usually much smaller than the initial sample, so the initial sample should be selected with consideration of the expected response rate.

Ignoring project costs, bigger is better when it comes to sample size. Choosing a sample size involves deciding how much error is acceptable and determining whether the budget allows for a large enough sample given the acceptable level of error. The sample size is also related to the mode of administration. With a fixed budget, choosing a more expensive mode of administration such as in-person surveys, implies a smaller sample size than choosing a less expensive mode such as a mail survey. The next section discusses the tradeoffs in terms of cost of the various modes of administration.

4.2 Mode of Administration

Choosing the mode of administration for a nonmarket valuation survey requires consideration of several issues. No one survey administration mode is unambiguously superior to others. Tradeoffs associated with the various modes must be considered: survey administration costs, time constraints, sample coverage, sample non-response bias, and context issues. In this section, I describe the most commonly used modes of administration along with their strengths and weaknesses (Table 2). Where empirical nonmarket valuation studies have been conducted to investigate the effects of the mode of administration, I include a summary of the study results.

Table 2. Strengths (+) and Weaknesses (-) of the Various Modes of Administration

	In-person	Telephone	Mail	Web-based/e-mail	Central Facility
Control over how the survey is administered	+	+	–	+	+
Ability to use visual aids	+	–	+	+	+
Ability to use an incomplete sample frame	+	+	–	–	+/–
Interviewer Effects	–	–	+	+	+/–

4.2.1 In-person Surveys

In-person surveys are administered face-to-face with an interviewer asking the survey questions and recording respondents' answers. Such surveys can be administered in respondents' homes, in convenient locations such as shopping malls or on the street, or in places where the participants of targeted activities tend to congregate. For example, recreation use surveys often are administered at the site of the recreational activity (i.e., hiking trailhead, boat landing, or campsite). Such surveys are referred to as "intercept" surveys as the survey respondent is intercepted at the recreation site.[3]

The primary advantages of in-person surveys come from the control the interviewer has while administering the survey. The interviewer controls the question order and sometimes uses visual aids. In-person surveys also allow for control over who in the household completes the survey. Likewise, complex surveys that may need some interviewer clarification are administered more easily in-person. Another advantage of in-person surveys is that a sample frame does not need to be complete and interviewers can reach people who would not respond to a telephone, mail, or Internet survey. However, in-person interviews tend to be much more expensive than other modes of administration due to training, salary, and travel costs for interviewers.[4] Interviewers must be well trained as interviewer influences are well documented (Singer and Presser 1989). Even well-trained interviewers can contribute to measurement error such as "social desirability bias" which occurs when respondents provide interviewers with a response that they think is socially desirable. For example, respondents may overstate their contributions to charitable organizations if they perceive charitable giving as a socially desirable activity. Social desirability bias is more likely to occur with an in-person interview than with other modes of administration. Interviewing techniques have been developed to diminish measurement error associated with reports of socially desirable behaviors such as voting frequency (Belli et al. 1999).

In the past, in-person surveys were thought to produce the highest quality data of all survey administration methods. However, research over the last two decades has challenged this belief (Lyberg and Kasprzyk 1991). The circumstances specific to a study will dictate which mode of administration is most appropriate. In 1993, a panel of prominent scholars submitted a report to the National Oceanic and Atmospheric Administration on appropriate contingent valuation methodology for natural resource damage assessments. In this report

(Arrow et al. 1993), the panel recommended use of in-person contingent valuation surveys for natural resource damage assessments stating (p. 4608):

> The Panel believes it unlikely that reliable estimates of values could be elicited with mail surveys. Face-to-face interviews are usually preferable although telephone interviews have some advantages in terms of cost and centralized supervision.

This recommendation was based largely on the benefits of in-person surveys mentioned above. (See Schuman (1996) for a summary of the reasons why mail surveys are not appropriate for contingent valuation studies related to natural resource damage assessments.) This recommendation should not be generalized to all contingent valuation studies, much less all nonmarket valuation studies. The researcher must carefully consider the tradeoffs among the various modes of administration to determine the most appropriate method.

Few empirical studies have investigated the effects of the mode of administration on nonmarket valuation data. One of those, a contingent valuation study by Mannesto and Loomis (1991) compared results of an in-person survey with a mail survey. They implemented a split sample experiment with recreational boaters and anglers in the Sacramento-San Joaquin Delta of California. In one treatment, in-person interviews were conducted at public boat ramps in the delta. In the other treatment, mail-back surveys were placed on vehicles with boat trailers at the same public boat ramps. The data from the two treatments were used to assess the impact of mode of administration on overall response rate, data comparability, and item non-response. They found that the completion rate was substantially higher for the in-person surveys (97%) than for the mail-back treatment (24%).[5] Item non-response was higher for the mail-back survey. Estimated mean willingness-to-pay based on the in-person surveys conducted by experienced interviewers was $70 compared to $37 based on the interviews conducted by less experienced interviewers. The mean willingness-to-pay based on the mail-back survey ($38) was similar to that based on the data collected by the less experienced interviewers in the in-person treatment. Mannesto and Loomis conclude that one mode does not dominate the other on relevant criteria. Salant and Dillman (1994, p. 42) summarize the situations in which in-person surveys are best suited:

- surveying populations for which there is no list,
- collecting information from people who are not likely to respond willingly or accurately (or cannot be reached) by mail or telephone,

- using complex questionnaires, and
- using experienced interviewers and professional help for well-funded projects.

4.2.2 Telephone Surveys

Over 90 percent of households in the U.S. have telephones, making the telephone a plausible mode for administering nonmarket valuation surveys. Telephone surveys, like in-person surveys, require an interviewer. However, the interviewer and respondent are not face-to-face. Telephone surveys allow for use of sample frames such as telephone listings sometimes augmented with additional information such as vehicle registration lists to overcome exclusion of people who have unlisted telephone numbers. However, obtaining a representative sample is difficult because approximately one in five households moves each year and many households have multiple telephone listings or unlisted telephone numbers. One alternative to using telephone listings as a sample frame is to use random digit dialing. With random digit dialing, the researcher identifies the three-digit prefix within a particular area code, then the following four digits are randomly generated.[6] This approach avoids the need for a pre-existing list of names with telephone numbers. However, the interviewer does not know the name of the person being called.

As with an in-person survey, the telephone survey interviewer has control over the order of the survey questions and who in the household is interviewed. This control reduces missing data and error due to the respondent not properly following the skip patterns in the survey. Another advantage is that the telephone interviewers can be in the same location thus simplifying supervision relative to in-person interviewer who are geographically dispersed. Also the interviewer can directly enter survey responses into a computer making the data readily available. Further, the cost of telephone interviews is substantially less than the cost of in-person interviews.[7] The drawbacks of telephone surveys include the possibility that the use of an interviewer will result in respondents providing socially desirable answers, that any use of visual aids requires prior mailing which would be impossible if random digit dialing were used, and that communication of complex information may be difficult. Guidelines for implementing a telephone survey are provided by Groves et al. (1988).

Three nonmarket valuation studies have compared telephone and mail surveys. Loomis and King (1994) conducted a contingent valuation study about improvements of wildlife and fisheries resources in California. Some of the surveys were administered by mail and others by telephone. The telephone survey procedures called for making an initial telephone call to obtain a commitment to participate, then mailing a survey booklet to the participant to allow for the use of visual aids, and, finally, conducting a telephone interview to elicit responses to the survey. Loomis and King considered four criteria when comparing mail and telephone surveys: (1) overall response rates and item non-response on the willingness-to-pay question, (2) differences in respondent demographics, (3) differences in logistic equations and willingness-to-pay amounts, and (4) differences in willingness-to-pay amounts after corrections were made for demographic characteristics and recreational participation.

The telephone survey achieved a higher response rate (56%) than the mail response (35%). Overall, and in particular with respect to the willingness-to-pay question, item non-response was higher in the mail survey. The demographic characteristics (gender, education, age, and income) of the two treatments were significantly different, although the attitudes toward wildlife were similar. Estimated willingness-to-pay based on the mail surveys was higher than that based on the telephone surveys. Correcting for the differences in demographic characteristics between the two treatments did not significantly diminish the difference between the willingness-to-pay estimates for the two treatments. The direction of the difference in willingness-to-pay is counterintuitive: one would expect social desirability bias to lead to higher willingness-to-pay with the telephone survey respondents.

Whittaker et al. (1998) compared willingness-to-pay for increased park entrance fees, reports of recreational expenditures, and frequency of visits to a recreational site for telephone and mail surveys. Whittaker et al. posited that differences in willingness-to-pay responses between the two treatments would most likely be due to (1) the social desirability bias associated with the telephone surveys and (2) recall information, such as frequency of visits to a recreational site or expenditures on such trips, would be more accurate with the mail survey because respondents have more time to think and consult with other members of the household.[8]

The results of the study suggest evidence of social desirability bias in the telephone survey because respondents to the telephone survey expressed higher willingness-to-pay for the increased park entrance fees than did to the mail survey respondents. However, even controlling for income, Whittaker et al. did not find any statistically significant differences between the two modes regarding frequency of park visitation or expenditures.

Ethier et al. (2000) also compared mail and telephone surveys for a contingent valuation study focusing on a program sponsored by the local electric utility company that would invest in renewable energy products. Ethier et al. found that the two modes of administration resulted in similar response rates, respondent populations, and willingness-to-pay responses. However, they found evidence of social desirability bias associated with the telephone survey responses to four subjective survey questions about giving to environmental causes, rating of the service provided by the electric utility company, level of interest in renewable energy, and respondent interest in planting trees.

Based on the limited result of these three studies, it is not clear that one mode of administration is superior to the other. Again, Salant and Dillman (1994, p. 40) summarize the circumstances in which telephone surveys are most appropriate:

- members of the population are likely to have telephones,
- questions are relatively straightforward,
- experienced interviewers are available, and
- quick turnaround is important.

4.2.3 Mail Surveys

Mail is the most common mode of survey administration (Dillman 1991) for several reasons. Mail surveys require fewer resources to implement and, in turn, are less expensive than telephone or in-person surveys. The procedures for implementing a mail survey are also less complicated, making implementation without hiring a survey research firm much easier. Furthermore, interviewer effects are avoided, visual aids can be easily used, and the survey can be completed at the respondent's pace rather than the interviewer's pace.

The main drawbacks of mail surveys are the potential sources of error. Non-coverage problems arise if the sample frame does not adequately represent the population about which generalizations will be made. Errors can also arise due to non-response, which is often higher for mail surveys than other modes of administration. Also respondents can examine the survey to see if they are interested in the topic and then decide whether or not to participate. This can result in important differences between respondents and non-respondents. Completion of a mail survey usually requires that respondents be able to read the survey. Another drawback of a mail survey is the loss of control over the order in which questions are answered. Respondents in a mail survey can change their responses to earlier questions based on information presented later in the survey. It is not clear if this would be desirable or undesirable, however it is another uncontrollable factor. The time between mailing out the surveys and having useable data can also be an issue. Dillman (2000) provides a thorough description of mail survey methods. Salant and Dillman (1994) conclude mail surveys are the most appropriate mode of administration when:

- surveying people for whom a reliable address list is available,
- surveying people who are likely to respond accurately and completely in writing,
- an immediate turnaround is not required, and
- money, qualified staff, and professional help are relatively scarce.

4.2.4 E-mail and Web Surveys

Over 50% of households in the U.S. have personal connections to the Internet. While these households are not representative of the broader population, special interest populations exist for which an e-mail or Web survey could be highly successful. Use of Internet technology to administer nonmarket valuation studies has been limited to date, however, it is an emerging technology that will become more prevalent.

An Internet survey can be e-mailed to an individual or the individual can go to a website to complete the survey. If the survey is e-mailed, it can be part of the message or an attachment. The respondent completes the survey and returns it (via e-mail) to the sender. This mode of administration has many advantages. The time needed to administer the survey to the entire sample is much shorter than with mail or even telephone surveys, especially for large samples. E-mail surveys are less expensive than mail surveys because there are no postage or printing costs and much less labor is required. However, significant costs may be associated with having the survey set up for e-mail administration. Since e-mail accounts often read incoming messages and attachments differently, the survey must be adaptable to various formats. The biggest drawback of e-mail surveys is sample coverage if all individuals in the sample frame do not have e-mail accounts. Coverage error can be reduced by combining an e-mail survey with some other mode such as a mail survey. Another issue with e-mail surveys is that it is difficult to respond anonymously. Whether this is a serious issue for nonmarket valuation surveys needs to be investigated.

Some research has compared e-mail surveys to mail surveys. None of the applications have specifically dealt with nonmarket valuation; however, the results may be relevant. With respect to item nonresponse, study results have been mixed. Sproull (1986) and Bachmann et al. (1996) found higher item nonresponse in e-mail surveys relative to mail surveys while Mehta and Sivadas (1995) and Tse et al. (1995) found no difference. Bachmann et al. (1996) found longer responses to open-ended question in e-mail surveys compared to mail surveys. Schaefer and Dillman (1998) found no significant difference in response rate between an e-mail survey and a mail survey. They also found that the e-mail survey respondents provided more complete responses (lower item non-response) than mail survey respondents and the completion time was quicker with the e-mail survey. Overall, e-mail appears to be a promising mode of administration. Schaefer and Dillman (1998) have developed a standard methodology for implementing e-mail surveys.

Web surveys involve a respondent going to a designated website and completing the survey there. Surveys can be accessed using an e-mail message containing a link to the website for the respondent to click. Administering the survey at a website provides control over how the survey is administered. Questions can appear one at a time or in groups thus allowing for automatic implementation of skip patterns and control over the order in which questions are answered. Another advantage of Web surveys is that the data are available instantly. Also possible is tracking how long respondents take to complete each question and whether the respondent accesses any linked web pages for additional information during the survey. As with e-mail surveys, the survey population of a web-based survey is not likely to be representative of a general population. There can also be issues of access. To prevent the same person completing the survey multiple times, a password is often implemented.

Berrens et al. (2003) conducted a research study to compare a random digit dialing telephone survey with a web surveys using two different types of Internet samples. The Internet samples used in this study were developed by the commercial firms Harris Interactive and Knowledge Networks. Harris Interactive develops a database of willing survey participants at the time individuals select their Internet provider. Knowledge Networks recruits potential survey participants by first contacting individuals via a random digit dialing phone request to participate in Web TV research.[9] The Knowledge Networks' panel of participants is most akin to a random sample. The survey elicited attitudes toward global climate change. Berrens et al. (2003) are careful to note that they are not suggesting that the telephone survey is a perfect sample to serve as the basis for comparison. Rather this research addresses the issue of how the Internet samples compare to a high quality telephone sample of the sort commonly used in social science research. The demographic comparisons among the telephone and two Internet survey participants show similar distributions on gender and age. With respect to education levels, 41% of telephone survey respondents were college graduates. This is similar to the unweighted percent found in the Harris samples. The Knowledge Networks sample had 24% with college degrees, a result similar to the most recent current U.S. population survey (23%). Other demographic measures such as percent Hispanic, percent African-American, household mean income, percent with computers at home, percent with computers at work, and mean number of household telephone lines were found to differ among the three samples. Likewise, many of the measures of knowledge about global warming impacts differed significantly between the telephone survey sample and the two Internet samples. Interestingly, Berrens et al. found that when the Internet samples were weighted to better match the demographics of U.S. adults, the survey results were more consistent with U.S. Census data than the weighted results of the telephone survey. In the end, Berrens et al. conclude "Although many differences arose, across a variety of tests on attitudes and voting intentions, the Internet samples produced relational inferences quite similar to the telephone sample."

As the technology of conducting Web surveys is relatively new, only limited research has focused on how the format and design of the survey affect responses to the survey. Couper (2000) reviews some of the sources of errors associated with Web surveys and major types of Web surveys. Couper, Traugott, and Lamias (2001) examined three issues related to design of a Web survey: the effect of including a progress indicator on the computer screen, the effect of including multiple items on the screen versus one item at a time, and the effect of providing radio buttons for respondents to click versus having respondents type a numeric response. With respect to the first issue, Couper et al. expected that a progress indicator would increase the completion rate of the survey because, unlike a mail survey where a participant can preview the survey to assess his progress, in a Web survey the respondent may not know how long the survey is or where he is at in terms of completing the survey. Researchers worry that individuals might abandon the survey before the entire survey is completed because they do not realize they are near the end of the survey. Couper et al. did not find a statistical difference in the completion rates between the version with the progress indicator and the version without. The a priori expectation with the second investigated issue was that displaying multiple items on the screen at one time would increase the correlations among the items. However, the study results did not provide statistically significant evidence in favor of this hypothesis. The third issue investigated was whether it would take less time to complete a question with radio buttons (where the respondent uses the mouse to click on the appropriate responses) versus filling in a number using the computer keyboard. Again, the study results did not find any statistically significant difference between the two methods of responding. However, more missing data were associated with the entry box approach. More research of the type conducted by Couper et al. clearly is needed to increase our understanding of the effects of the context of Web surveys.

A few nonmarket valuation surveys have been administered using the Web. Results of these studies are largely unpublished. One study conducted by Berrens et al. (expected publication: 2004) used a Web survey to investigate the effects of information access and effort on willingness-to-pay in terms of higher energy and gasoline prices for the U.S. Senate to ratify the Kyoto Protocol to the United Nations Framework on Climate Change. The use of web-based technology facilitated investigation of the effects of information and respondent effort on willingness-to-pay measures. Berrens et al. found a positive relationship between an objective measure of the effort put forth by the respondents (time spent reading and seeking additional information on global climate change) to obtain more information and willingness-to-pay responses. However, they did not find an effect of simply being offered additional information about global climate change. The Berrens et al. study demonstrates how Web surveys can be used to investigate important context issues related to a contingent valuation survey about a topic as complex as global climate change.

4.2.5 Central Site Data Collection

A technique used in marketing research that is becoming more prevalent among nonmarket valuation researchers are central site interviews. This approach involves gathering a small group of survey respondents at a convenient site. The interview session can be conducted in many ways. Often times an oral presentation of information is made, then respondents each complete their own survey. A group discussion may follow. Alternately, the survey can be self-administered as a paper survey or on a computer. The advantages of this approach include being able to use visual aids and provide detailed explanations, being able to have an in-depth conversation that includes probes, and being able to see and interpret body language. This approach is also less costly than having interviewers travel to survey respondents' homes. Some of the drawbacks are that survey participants are recruited in manner that is not probabilistic and the research site may be an artificial environment.

Adamowicz et al. (1997) used central site interviews for their study of moose hunting in Alberta, Canada. Their survey was long and complex (they collected stated preference data, revealed preference data attribute perceptions and demographic information from the survey respondents) making phone and mail surveys infeasible. The central facility interviews were less costly than in-person interviews.

4.2.6 Mixed Mode Surveys

Many surveys are implemented using more than one mode of administration. Mixing modes of administration can serve two purposes. First, the issue of coverage error can be addressed by using different modes for different parts of the sample population. The Schaefer and Dillman (1998) study involved a study population that largely had access to the Internet (89% had e-mail addresses); however, there was an identifiable part of the population for whom e-mail would not work. Mail surveys were used to contact individuals who did not have e-mail accounts.

Likewise, in-person surveys can be used with telephone surveys to contact individuals without telephones. Another way to use mixed mode surveys is to use multiple modes to contact the same individual. For example, the initial contact or follow-up contacts can be made via a mode that differs from the mode used to administer the main survey. The motivation for mixing the modes is to increase the response rate. With respect to mixing an e-mail approach with other modes of administration, Schaefer and Dillman (1998, p. 381) suggest "The cost and speed advantages of e-mail make it ideal for a first mode of contact in surveys. Researchers can begin with an e-mail approach and use progressively more expensive methods for nonrespondents until an acceptable response level is reached."

Kramer and Eisen-Hecht (1999) investigated two mixed mode approaches for implementing a contingent valuation survey. They conducted a split sample experiment where in one treatment they administered the contingent valuation survey using a phone-mail-phone approach, and in the other treatment they used a mail-phone approach. In the phone-mail-phone treatment, individuals were initially contacted by phone using random digit dialing. If the individuals agreed to participate in the survey, a short phone survey was administered and at the same time respondents were asked if they would participate in another phone survey. If so, they were mailed a cover letter and an informational brochure "Water Quality in the Catawba River Basin." They were then called for the second interview which included the contingent valuation question. In the other mail-phone treatment, individuals were sent the cover letter and "Water Quality in the Catawba River Basin" brochure.[10] The cover letter let individuals know that they would receive a call asking them to participate in a telephone interview. Kramer and Eisen-Hecht report different response rates for the two treatments with 58% for the phone-mail-phone treatment and 36% for the mail-phone treatment.[11] The estimates of mean willingness-to-pay did not differ significantly between the two treatments. However, there was evidence of some self selection by mail-phone respondents in that there were more recreationists, more individuals who had heard about the water quality concerns, more individuals who thought protection of the Catawba River Basin was more important than other state environmental issues among the mail-phone respondents relative to the phone-mail-phone respondents. Likewise more males and individuals with higher education levels responded to the mail-phone surveys than to the phone-mail-phone surveys. In the mail-phone treatment, potential survey respondents were able to preview the materials before determining whether to participate in the survey. In the phone-mail-phone treatment, they did not know specifically what the survey was about when they decided to participate in the first interview. However they were able to preview the brochure prior to actually participating in the second interview. It is not clear how use of different sample frames for the two treatments affected the respondents population. It is possible that the mail-phone sample frame initially included more recreationists, etc. Given the phone-mail-phone approach had a much higher cost per completed survey, Kramer and Eisen-Hecht conclude that the additional cost of the phone-mail-phone approach may not be justified.

4.3 Survey Instrument

How the survey will be administered affects the design of the survey materials; a survey administered by a professional interviewer is formatted differently than a self-administered survey. However, the process of designing a survey instrument is similar for all the modes of administration. As the quality of the data analyses is only as good as the data, good survey design and implementation are essential. A well designed survey instrument will clearly communicate relevant information and present questions in an unambiguous manner. Qualitative research methods such as focus groups can facilitate decisions about what information to include in the survey. The researcher must also consider *how* to communicate relevant information.

4.3.1 Identifying Desired Measures

The first step in developing the survey instrument is to identify the measures to be elicited in the survey. The researcher will want to consider what data are required for analyses and testing of hypotheses. The researcher will usually have a good sense of the general goals of a study but may struggle with identifying the specific measures needed to address these study goals. For example, a researcher designing a contingent valuation survey to measure willingness-to-pay for improved fishing conditions may want to know about the respondents' angling experience. However, simply asking the question, Are you an experienced angler? may not provide useful information. The researcher should consider the specific dimensions of angling experience such as number of years the respondent has been fishing, how often the respondent goes fishing, and the type of gear the respondent uses when fishing. It is often useful to review studies on similar topics to facilitate identification of the measures to include in the survey.

Validity assessments will also motivate what questions to ask in the survey. One form of validity, theoretical construct validity, involves assessing the relationships between measures as suggested by economic theory. For example, testing whether income is positively related to willingness-to-pay would be a test of theoretical construct validity.

If the study is going to be used for policy purposes or is a source of information for a decision-making process, the researcher should work with the end users of the information to develop survey measures.

4.3.2 Writing Survey Questions

The survey questions should be clearly written using common vocabulary. Focus groups and other qualitative approaches can be used to investigate issues, such as appropriate language for the respondent population and baseline knowledge of the survey topic. How questions are phrased will vary with the mode of administration. In-person and telephone surveys often pose questions in the spoken language while mail and other forms of self-administered language are usually written to be grammatically correct.

When developing survey questions, it is useful to review surveys that others have done on related topics. Several good books address question design including Converse and Presser (1986), Sheatsley (1983), and Sudman and Bradburn (1982). Survey design is an iterative process with many rounds of revising and testing. Some important issues related to the design of survey questions are discussed in the following sections.

4.3.2.1 Open-versus Closed-ended Questions

Questions can be asked in either an open- or closed-ended format. Open-ended questions do not include response categories. For example, What do you like most about visiting Yellowstone National Park? suggests that the respondent is to answer in his own words. Responses to open-ended questions can be categorized subsequently, but doing so is subjective and time consuming. (See Sheatsley (1983) for detailed discussion of the advantages and disadvantages of open- and closed-ended questions.) Nonmarket valuation surveys usually consist of closed-ended questions because closed-ended questions are less burdensome on respondents and the responses can be easily used for statistical analyses. Developing response categories for closed-ended questions takes careful consideration of the range of possible answers. One technique for developing response categories is to use open-ended questions in early drafts of the survey and develop response categories based on responses to the open-ended questions.

4.3.2.2 Avoiding Confusion

One type of confusing survey question is the double-barreled question. An example is asking respondents whether they agree or disagree with the statement, I am concerned about the health effects of smoking and drinking alcohol. A respondent who is concerned about the health effects of smoking but not about drinking alcohol will have a difficult time responding to this statement.

Providing balance in a survey by eliciting responses to statements that reflect differing perspectives on a topic is important. However, changing from positive to negative statements in a series of attitude items is confusing. Notice how the statements in Figure 1 change between positive and negative:

National Parks should be easily accessible by roads National Parks should *not* include designated wilderness areas National Parks should be managed to preserve native plant species

Figure 1. Statements that switch between positive and negative

Respondents who are asked whether they agree or disagree with the above statements may not notice that the statements change between positive and negative. Preferable statements would address alternative sides of the issue stated in a consistent manner.

The response categories should be consistent with the question asked. For example, Did you go fishing in the last twelve months? should have yes and no as response categories. However, it is not uncommon to see a such a question followed up the response categories frequently, occasionally, and never. Likewise, you want to point out to the survey respondent changes in the reference time frame. For example, one set of questions in a fishing survey asks about expenditures in the last twelve months and another set of questions refers to expenditures on the most recent trip would require pointing out this change in reference to the respondent.

4.3.2.3 Eliciting Recall Data

Frequency data, such as number of times visited the doctor in the last twelve months or the number of times fishing at a particular site, are often elicited in nonmarket surveys. Frequency questions can be difficult for some respondents. Schwarz (1990, p. 99) summarizes the cognitive processes that respondents go through when responding to behavioral frequency questions:

> First, respondents need to understand what the question refers to, and which behavior they are supposed to report. Second, they have to recall or reconstruct relevant instances of this behavior from memory. Third, if the question specifies a reference period, they must determine if these instances occurred during this reference period or not. Similarly, if the

> question refers to their "usual" behavior, respondents have to determine if the recalled or reconstructed instances are reasonably representative or if they reflect a deviation from their usual behavior. Fourth, as an alternative to recalling or reconstructing instances of the behavior, respondents may rely on their general knowledge, or other salient information that may bear on their task, to infer an answer. Finally, respondents have to provide their report to the researcher.

Respondents may be better able to recall frequency of events if they are provided with cues that make the question more specific. For example, instead of asking how many times a person went fishing, asking the number of times a person went fishing for specific species may reduce the cognitive burden on the respondent. Likewise when asking about expenditures for travel to a particular recreation destination, it is helpful to ask about specific expenditures (i.e., fuel, food, equipment, etc.) rather than total trip expenditures. Given the heavy dependence of nonmarket valuation techniques, such as the travel cost method on recall data, the researcher needs to be sensitive to potential data-collection problems and implement techniques to mitigate such problems.

4.3.2.4 Combining Data Sets

If the survey data are going to be combined with existing data, then the survey data must be consistent with the existing data. For example, response categories must be the same as those used in the existing data. Likewise, definitions of terms in the new data set must be consistent with definitions in the existing data set. For example, if the existing data defines household income as the amount before taxes and deductions, the survey should use the same definition of household income.

4.3.3 Question Order and Format

After developing the survey questions, the researcher considers the order and format of the questions. Again, decisions about question order and format are closely related to the mode of administration. Sudman and Bradburn (1982) suggest starting a survey with easy, nonthreatening questions. Some nonmarket valuation surveys, such as contingent valuation surveys, require substantial description of the good or program of interest prior to asking the willingness-to-pay question. Some contingent valuation practitioners break up large sections of information by interspersing relevant questions about the provided information. There is usually a natural flow to the question order. Dillman (2000) suggests putting objectionable questions, such as income, at the end of the survey as respondents are less likely to terminate the survey if they have already spent five or ten minutes working on the survey. Asking all demographic questions at the end of the survey instrument is common practice.

One issue the researcher needs to be aware of with respect to opinion questions is that the order of the questions has been found to influence responses (Dillman 2000). This phenomenon could be of particular concern for stated preference surveys that ask the respondent to value multiple goods. While the researcher may tell the survey respondent to value each good independently, individuals may not be willing or able to do so. While such effects are difficult to mitigate, the recommendation (Dillman 2000; Mitchell and Carson 1989) is to randomize the order of questions that are susceptible to order effects.

4.4 Testing and Refining the Survey Instrument

The final version of a survey often bears little resemblance to the first draft. The iterations of revisions facilitate development of a well-designed survey instrument that will, in turn, improve the chances of an individual completing the survey and providing meaningful data. Qualitative methods, such as focus groups and one-on-one interviews, are commonly used to refine and develop the survey instrument. Since nonmarket valuation surveys often involve topics with which citizens may not be familiar, understanding the respondents' pre-survey level of knowledge is important so the survey materials can be written appropriately. Qualitative methods are also useful for developing an understanding of how respondents think and talk about the survey topic, for example, the vocabulary survey respondents use to refer to the survey topic. I also recommend review by peers and others experienced with development of nonmarket valuation survey instruments. While qualitative methods are generally considered useful for developing and refining survey materials, there is no agreement about standard procedures. The following sections briefly describe two of the more commonly used qualitative approaches for developing a survey instrument.

4.4.1 Focus Groups

Focus groups are small group (7 to 12 individuals) discussions conducted for the purpose of getting feedback to aid in the development of the survey materials. The discussion is lead by a focus group moderator using an agenda to guide the discussion. The moderator usually asks open-ended questions to facilitate discussion about the relevant topic. One challenge for the moderator is to balance adhering to the agenda and allowing the discussion to evolve. Another challenge is dealing with the diverse personalities of the focus group participants. A participant with a strong personality can dominate the focus group if the moderator is not able to manage that participant and draw out comments from the more reserved participants. Some project budgets will allow for focus groups conducted in professional focus group facilities with an experienced moderator. However, much can also be learned using less expensive focus groups that are conducted in a more informal setting such as a classroom at a university or at a local library. What matters is that the group is well planned, the participants are somewhat diverse, and the moderator is adequately trained to obtain the desired information from the participants. While focus group participants are not representative of the study population, issues that are raised by the group are often relevant for the broader population and issues are frequently raised that facilitate better design of the survey instrument. However, the researcher should carefully consider how the small group dynamic can influence what issues are raised and how important participants really think these issues are. For example, participants are sometimes susceptible to jumping on the bandwagon regarding issues that they would not have considered otherwise. Focus groups are usually audio or video taped so the moderator can concentrate on the group discussion without concern about taking notes. Afterward, the tapes are reviewed and summarized. Greenbaum (2000) provides a how to guide for conducting focus groups. Chilton and Hutchinson (1991) suggest a structured approach to analyzing the focus group data.

4.4.2 One-on-one Interviews

Another qualitative method used to develop and refine nonmarket survey materials is the one-on-one interview. One-on-one interviews allow for detailed feedback on survey materials without concern about a group effect because the interview involves only an interviewer and a survey participant. These interviews usually take one of two forms: either the survey participant is asked to complete the survey materials while verbalizing aloud all of his or her thoughts or the interviewer asks questions from a structured list and probes for more information from the respondent. The respondent answers the survey questions and is then asked about motivations for responses, assumptions made when answering the questions, or whatever issues the researcher wishes to investigate. One-on-one interviewers can be very helpful but enough interviewers should be conducted to allow for a sense of whether issues are particular to one individual or more widespread.

5. SURVEY IMPLEMENTATION

Survey implementation procedures should be developed with the goal of collecting information from the survey sample that is uniform and reliable and is obtained within time and budget constraints (Weinberg 1983). While the procedures may appear sound when developed, serious problems often are not identified until testing occurs. Researchers cannot afford to make major mistakes with implementation of the final survey. So it is important to carefully consider how the final survey will be implemented and conduct a pilot survey as a test. Ideally, all contacted individuals would complete the survey; however, only a portion of the contacted individuals will actually respond. Failure to obtain information from all individuals in the survey sample can result in response error. The severity of response error varies with each individual survey, but is a problem when individuals responding to the survey differ from those not responding on the measures of interest. Research suggests that in addition to the actual design of the survey, how the survey is implemented can affect the response rate. The specific procedures for implementing a survey are many and vary with the mode of administration. Standard procedures have been developed by Dillman (2000) for both mail and Internet surveys. Likewise Weisberg et al. (1996) outline procedures for mail, telephone, and in-person surveys.

One of the more important issues affecting response rates is the saliency of the survey topic to the sample. Hunters are much more likely to respond to a survey about hunting than the general public is to respond to a survey about renewable energy sources. Given that the survey topic is what it is, the researcher should do what he can in terms of survey design and implementation to assure a reasonable response rate. The researcher has control over the appearance of self-administered surveys, the tone and professionalism of interviewers, and the description of why an individual should complete the survey. Most survey procedures call for multiple follow-ups or contacts with individuals who do not respond to the initial contact to increase the response rate. The benefits of each additional follow-up should be carefully weighed against the additional cost.[12] A reasonable response rate is one that provides enough data for the desired analyses with an acceptable range of error and provides a group of respondents with characteristics similar to the broader population to which inferences will be made.

5.1 Pilot Survey

A pilot survey is a small scale implementation of the final survey procedures with an acknowledgment that final procedures may be modified based on the results of the pilot survey. The pilot survey serves several purposes. First, it allows for a test of the survey sample. Some samples can be problematic but the researcher may not realize it until the survey is in the field and great expense has already been incurred. For example, I was involved in the design of a mail survey in which the sample was developed using a list of registered voters. Even though the list was current during the pilot study, we found that many of the addresses were not valid and a substantial number of the surveys were undeliverable. In the end, we had to use a completely different sample for the final survey. Second, the pilot survey also provides information about the anticipated response rate, the number of follow-ups and the expected cost for the final survey. Third, the pilot study allows for a test of the survey implementation procedures. For contingent valuation studies, the pilot study also offers an opportunity to obtain information about the distribution of willingness-to-pay. If the researcher plans to use a dichotomous-choice contingent valuation question in the final survey, an open-ended question can be used in the pilot study to guide assignment of the offer amounts for the final survey.

5.2 Final Survey

After the survey has been designed and the survey implementation procedures tested, the survey is ready to go into the field. Regardless of the mode of administration, final survey implementation tends to be hectic. A well organized and realistic schedule is important. Printing of survey materials, ordering postage, hiring interviewers, setting up a database, and developing a sample frame are time consuming tasks. A timeline should be developed for each of these tasks.

The initial sample should be entered into a database so respondents and non-respondents can be properly tracked. In addition to tracking whether or not an individual responded, it is beneficial to track when they responded and when various follow-up contacts were made. This information can be used for planning future surveys and allows for the testing of differences between early and late respondents. Information about the completion of surveys should be entered into the database daily, and the database should be backed-up every time new information is entered.

Since most nonmarket valuation surveys elicit sensitive information such as annual household income, it is important to assure respondents that their responses will be confidential and their names will not be associated with their responses. The standard procedure is to assign a unique identification number to each person in the initial sample that is used to track whether or not the person has responded. Only the identification numbers, not the respondents' names or addresses, are included in the final data set. Survey interviewers have an ethical responsibility to keep all responses confidential. Surveys conducted through universities usually require informed consent, which involves informing respondents that their participation is voluntary. The specific information provided to respondents about the terms of their participation and the procedures for getting permission from a university to do a survey varies among universities. The funding organization for the survey should be knowledgeable about both the procedures for obtaining permission to conduct a survey and the procedures for obtaining consent from the survey respondents.

Organizing the completed surveys in order of the unique identification number makes data entry and verification much easier. After the data are entered, the researcher may need to go back to the surveys to check on answers that appear to be inconsistent or problematic. Finding a particular survey if not in order by identification number is very time consuming.

6. DATA SET

In the following sections, I describe the procedures for transforming completed surveys into a useable dataset. The first step is to develop a set of rules, called a codebook, for data entry. Then the data are entered and verified. Finally data analyses can be conducted.

6.1 Codebook

Coding the data refers to the process of translating the survey responses into numerical categories, which can then be entered into the computer. Coding allows the data entry person to go as quickly as possible. The format of the survey questions dictates whether and how responses are to be coded. One common format for responses to closed-ended survey questions is to have the respondent or interviewer fill in a bubble or put an x in a box. This format is not recommended because it requires an additional step of assigning a numerical response to each of the bubbles or boxes. A preferred format is to have a number printed next to every possible response to a question as shown in Figure 2. Even when responses have corresponding numbers, some judgement may be involved in coding the data where the respondent did not follow the directions. For example, when two responses are circled but only one response was supposed to be circled, the person coding the data will have to decide what number should be entered. It is important to develop consistent rules for coding the data. For example, the question in Figure 2 assigns the value 1 to a yes response and 2 to a no response. However, a numeric value also needs to be assigned if the question was not answered. A system of codes should be developed to differentiate between questions that respondents were supposed to skip and those that they were supposed to answer but chose not to. Weisberg, Krosnick and Bowen (1996) describe standard coding practices.

Coding open-ended questions is more complicated. One option is to not do any coding and have the person who does the data entry type the entire response to the open-ended questions. Using this sort of data for statistical analysis can

Have you ever visited Yellowstone National Park? ***(Circle one number)***		
	1	Yes
	2	No

Figure 2. Sample Survey Question

be difficult. Another approach is for the researcher to develop a workable number of response categories and code each open-ended response into the appropriate category.

6.2 Data Entry

Some survey modes such as telephone surveys with a computer-assisted telephone interviewing system, Web surveys or other computer-administered surveys do not require that data be entered because survey responses are directly entered when the survey is administered. However for surveys requiring separate data entry, the proliferation of database management software has made data entry much more efficient. Software can be set up to notify the data entry person when a number that is out of the relevant range has been entered. Skip patterns can be entered such that the software automatically enters the appropriate code if a question is supposed to be skipped. These innovations minimize data entry errors. However, even when using a data entry software package, the data probably should be entered twice. During the second entry of the data, the computer software signals if something is being entered that does not match the original data entry. While double entry is time consuming, the extra effort to catch data entry errors is worthwhile.

After the data have been entered, frequencies should be conducted on all variables to insure that the response values are within the appropriate range for each question and that skip patterns have been properly followed. Cross tabulations can also be used to check that the data have been correctly entered. Since a unique identification number is entered in the dataset and written on the actual survey, errors in the data can be checked using the original surveys. This step is referred to as data cleaning. As mentioned earlier, the completed surveys should be ordered by identification number to save time finding a problem survey.

If the data are going to be used to generalize to a broader population, the socioeconomic variables from the survey respondents should be compared to the those of the broader population. If the survey respondents differ from the broader population to which survey results are to be generalized, sample weights can be applied.[13] Quite a few articles describe how to weight respondent populations (see Whitehead et al. 1994; Whitehead, Groothuis, and Blomquist 1993; Dalecki, Whitehead, and Blomquist 1993; Mattsson and Li 1994, Whitehead 1991).

7. REPORTING RESULTS

Regardless of the study objectives, all studies should include proper documentation of the procedures used to collect the data and summary statistics on all measures elicited in the survey. Proper documentation will make the data more accessible. Complete documentation also will assist reviewers of the study. The best way to document data collection procedures is to do it as the study is being conducted. The entire study from the initial study plan to the final summary statistics should be documented in one location. Table 3 outlines the information about the study that should be documented.

8. SUMMARY

While this chapter largely focused on the steps to collecting original nonmarket valuation data, users of secondary data should also have a better understanding of how data they might use was collected. Most nonmarket valuation practitioners tend to focus on issues associated with the various valuation techniques and statistical issues associated with model estimation. These are important issues but irrelevant if the data are not valid. This chapter highlighted the need for careful data collection. Although there are no strict guidelines to assure collection of valid data, this chapter summarized the key issues for the researcher to consider and the steps involved in collecting data for nonmarket valuation.

Table 3. Guidelines for Documenting a Study

		Information to Document
1.	Study Plan	Date Initial budget Study proposal
2.	Survey Development	
	Focus Groups/One-on-one interviews	Dates Location Costs (amount paid to participants and other Costs such as food, facility rental, etc) Focus group moderator/interviewer Number of participants (do not include Names or addresses of participants) Hard copy of all handouts Hard copy of agenda Video and/or audio tapes Hard copy of interview or focus group Summaries
3.	Sample	Description of sample frame From which sample was developed Cost of sample Number of sample units Data included with sample Procedures for selecting sample units
4.	Final Survey	Cover letters Surveys Question and answer sheets Dates of all survey mailings/interviews/telephone calls Number of respondents to each wave of contact
5.	Data	Codebook Final sample size Frequencies including missing values for each measure in the survey Means/median for continuous measures Data file in several formats (ascii as well as software specific such as excel, spss, etc.)

ACKNOWLEDGMENTS

Thanks to the other authors in this book for their constructive comments at the two nonmarket valuation workshops. Tom Holmes' detailed review and Tom Brown's careful editing helped in the refinement of this chapter. I would also like to thank Mike Welsh, Kelly Giraud, and Denise Wickwar for reviewing early drafts.

NOTES

1 This survey, conducted every five years for the U.S. Fish and Wildlife Service, collects information on number of participants in the targeted recreational activities, number of trips and days spent on the different activities, expenditures by activity type, as well as socio-demographic characteristics. Most of the survey data and summary reports can be ordered or downloaded from the U.S. Fish and Wildlife Service's website.

2 The Sudman (1983) chapter in the Handbook of Survey Research provides a detailed discussion of sampling techniques.

3 Intercept surveys do not *always* involve an in-person survey. The intercept can be a commitment from an individual to complete a mail or phone survey at a later date.

4 In-person intercept surveys are not necessarily more expensive than other modes of administration because the interviewers may travel to only one site.

5 Standard mail survey procedures (Dillman 2000) call for multiple follow-up mailings to improve response rates. Mannesto and Loomis were not able to do follow-up mailings because they did not have the necessary addresses

6 This description is specific to telephone numbers in the U.S.. A similar approach could be used in other countries. There are also variations of random digit dialing such as "add a digit" dialing. These approaches are described in more detail in Groves et al. (1988).

7 Cost comparisons from a survey research firm suggest approximate costs per completed survey of $25-30 for a mail survey, $30-35 for a telephone survey and $50-100 for a personal interview. These rough cost estimates are subject to substantial variation depending on the nature of the survey and the study population.

8 Recall data on expenditures and visitation frequency is used extensively for travel cost models described in Chapter 10. Accuracy of recall data is an important research issue as errors will, in turn, affect the value estimates. Champ and Bishop (1996) compared recall data on deer hunting expenditures reported in ex-post surveys to expenditures reported in diaries that were filled out as expenditures were made. They found no significant differences between the expenditures reported in the surveys and those in the diaries when the surveys were mailed out soon after the hunt.

9 Generally, recruited households are provided with a Web television unit, Web access and e-mail accounts for all household members over 13 years of age. This is provided free of charge. In exchange, the participants agree to complete at most one 10-15 minute survey per week. The details of the arrangements vary.

10 Kramer and Eisen-Hecht did not describe how the name and address list was created.

11 The response rates are calculated a bit differently for the two treatments. For the phone-mail-phone they report that 73% of initial phone interviewees agreed to receive the additional materials and complete the second interview. Fifty-eight percent of these 73% that agreed to complete the second interview actually did so. Thus, 58% response rate is reported. For the mail-phone treatment, 36% of the individuals who were mailed the initial materials and later contacted, completed the phone interview.

12 A perception exists that response rates to mail and phone surveys have been declining over time. I was not able to find documentation of this trend. Curtin, Presser, and Singer (2000) have investigated the increased effort to maintain a 70 percent response rate to a survey of consumer attitudes between 1979 and 1996. This investigation implicitly operates under premise that the response rate would have significantly dropped without the increased follow-up effort.

13 Often times data are weighted using U.S. Census data.

REFERENCES

Adamowicz, W., J. Swait, P. Boxall, J. Louviere, and M. Williams. 1997. Perceptions versus Objective Measures of Environmental Quality in Combined Revealed and Stated Preference Models of Environmental Valuation. Journal of Environmental Economics and Management 32:65-84.

Arrow, K., R. Solow, P. Portney, E. Leamer, R. Radner, and H. Schuman. 1993. Report of the NOAA Panel on Contingent Valuation. Federal Register.

Bachmann, D., J. Elfrink, and G. Vazzana. 1996. Tracking the Progress of e-mail versus Snail-Mail. Marketing Research 8:31-35.

Belli, R., M. Traugott, M. Young, and K. McGonagle. 1999. Reducing Vote Overreporting in Surveys: Social Desirability, Memory Failure, and Source Monitoring. Public Opinion Quarterly 63:90-108.

Berrens, R., A. Bohara, H. Jenkins-Smith, C. Silva, and D. Weimer. 2004. Information and Effort in Contingent Valuation Surveys: Application to Global Climate Change Using National Internet Samples. Journal of Environmental Economics and Management (forthcoming).

Berrens, R., A. Bohara, H. Jenkins-Smith, C. Silva, and D. Weimer. 2003. The Advent of Internet Surveys for Political Research: A Comparison of Telephone and Internet Samples. Political Analysis 11(1):1-23

Carson, R., R. Mitchell, W. Hanemann, R. Kopp, S. Presser, and P. Ruud. 1992. A Contingent Valuation Study of Lost Passive Use Values Resulting from the Exxon Valdez Oil Spill: Report to the Attorney General of the State of Alaska.

Champ, P. A., and R. C. Bishop. 1996. Evidence on the Accuracy of Expenditures Reported in Recreational Surveys. Journal of Agricultural and Resource Economics 21(1):150-159.

Chilton, S., and W. Hutchinson. 1991. Do focus groups contribute anything to the contingent valuation process? Journal of Economic Psychology 20:465-83.

Converse, J., and S. Presser. 1986. Survey Questions: Handcrafting the Standardized Questionnaire. Edited by M.S. Lewis-Beck, Quantitative Applications in the Social Sciences. Newbury Park, CA: Sage Publications, Inc.

Couper, M. 2000. Web Surveys: A Review of Issues and Approaches. Public Opinion Quarterly 64:464-494.

Couper, M., M. Traugott, and M. Lamias. 2001. Web Survey Design and Administration. Public Opinion Quarterly Volume 65:230-253.

Curtin, R., S. Presser, E. Singer. 2000. The Effects of Response Rate Changes on the Index of Consumer Sentiment. The Public Opinion Quarterly 64 (4): 413-28.

Dalecki, M. G., J. C. Whitehead, and G. C. Blomquist. 1993. Sample Non-response Bias and Aggregate Benefits in Contingent Valuation: An Examination of Early, Late and Non-respondents. Journal of Environmental Management 38:133-143.

Dillman, D. 1991. The Design and Administration of Mail Surveys. Annual Review of Sociology 17:225-49.

Dillman, D. 2000. Mail and Internet Surveys: The Tailored Design Method. Second Edition ed. New York: John Wiley & Sons, Inc.

Ethier, R., G. Poe, W. Schulze, and J. Clark. 2000. A Comparison of Hypothetical Phone and Mail Contingent Valuation Responses for Green-Pricing Eelectricity Programs. Land Economics 76 (1):54-67.

Frankel, M. 1983. Sampling Theory. In Handbook of Survey Research, edited by P. Rossi, J. Wright, and A. Anderson. San Diego: Academic Press, Inc.

Greenbaum, T. 2000. Moderating Focus Groups. Thousand Oaks: Sage Publications, Inc.

Groves, R., P. Biemer, L. Lyberg, J. Massey, W. Nicholls, and J. Waksberg, eds. 1988. Telephone Survey Methodology. New York: John Wiley & Sons, Inc.

Harrison, G. W., and J. C. Lesley. 1996. Must Contingent Valuation Surveys Cost So Much? Journal of Environmental Economics and Management 31: 79-95.

Kramer, R., and J. Eisen-Hecht. 1999. Estimating the Economic Value of Water Quality in the Catawba River Basin of North and South Carolina. In Southern Economics Association's Sixty-Ninth Annual Conference. New Orleans.

Loomis, J. 2000. Vertically Summing Public Good Demand Curves: An Empirical Comparison of Economic Versus Political Jurisdictions. Land Economics 76 (2):312-321.

Loomis, J., and M. King. 1994. Comparison of Mail and Telephone-mail Contingent Valuation Surveys. Journal of Environmental Management 41:309-324.

Lyberg, L., and D. Kasprzyk. 1991. Data Collection Methods and Measurement Error: An Overview. In Measurement Errors in Surveys. Edited by P. P. Biemer, R. M. Groves, L. Lyberg, N. A. Mathiowetz and S. Sudman. New York: John Wiley & Sons, Inc.

Mannesto, G., and J. Loomis. 1991. Evaluation of Mail and In-person Contingent Value Surveys: Results of a Study of Recreational Boaters. Journal of Environmental Management 32:177-190.

Mattsson, L., and C. Li. 1994. Sample Nonresponse in a Maill Contingent Valuation Survey: An Empirical Test of the Effect on Value Inference. Journal of Leisure Research 26 (2):182-188.

McCollum, D. W., M. A. Haefele, R. S. Rosenberger. 1999. A Survey of 1997 Colorado Anglers and Their Willingness to Pay Increased License Fees (Project Report No. 39). Project Report for the Colorado Division of Wildlife. Fort Collins, CO: Colorado State University, Human Dimensions in Natural Resources Unit, and USDA Forest Service, Rocky Mountain Research Station. 106pp.

Mitchell, R. C. and R. T. Carson. 1989. Using Surveys to Value Public Goods: The Contingent Valuation Method. Washington, DC: Resources for the Future.

Mehta, R., and E. Sivadas. 1995. Comparing Response Rates and Response Content in Mail versus Electronic Mail Surveys. Journal of Market Research Society 37:429-39.

Salant, P., and D. Dillman. 1994. How to Conduct Your Own Survey. New York: John Wiley & Sons, Inc.

Mitchell, R. C. and R. T. Carson. 1989. Using Surveys to Value Public Goods: The Contingent Valuation Method Washington, DC: Resources for the Future.

Schaefer, D., and D. Dillman. 1998. Development of a Standard e-mail Methodology: Results of an Experiment. Public Opinion Quarterly 62:378-97.

Schaeffer, R., W. Mendanhall, and L. Ott. 1986. Elementary Survey Sampling. Boston: Duxbury.

Schuman, H. 1996. The Sensitivity of CV Outcomes to CV Survey Methods. In The Contingent Valuation of Environmental Resources. Edited by D. J. Bjornstad and J. R. Kahn. Northampton, MA: Edward Elgar.

Schwarz, N. 1990. Assessing Frequency Reports of Mundane Behaviors: Contributions of Cognitive Psychology to Questionnaire Construction. In Research Methods in Personality and Social Psychology. Edited by C. Hendrick and M. S. Clark. Newbury Park, CA: Sage.

Sheatsley, P. 1983. Questionnaire Construction and Item Writing. In Handbook of Survey Research. Edited by P. H. Rossi, J. D. Wright and A. B. Anderson. San Diego: Academic Press, Inc.

Singer, E., and S. Presser, eds. 1989. Survey Research Methods. Chicago: The University of Chicago Press.

Smith, V. 1993. Nonmarket Valuation of Environmental Resources: An Interpretive Appraisal. Land Economics 69 (1):1-26.

Sproull, L. S. 1986. Using Electronic Mail for Data Collection in Organizational Research. Academy of Management Journal 29:159-69.

Sudman, S., and N. Bradburn. 1982. Asking Questions. San Francisco: Jossey-Bass Publishers.

Sudman, S. 1983. Applied Sampling. In Handbook of Survey Research, edited by R. H. Rossi, J. D. Wright, and A. B. Anderson. San Diego: Academic Press.

Sutherland, R., and R. Walsh. 1985. Effect of Distance on the Preservation Value of Water Quality. Land Economics 61 (3):281-291.

Tse, A., K. C. Tse, C. H. Yin, C. B. Ting, K. W. Yk, K. P. Yess, and W. C. Hong. 1995. Comparing Two Methods of Sending Out Questionnaires: e-mail versus Mail. Journal of the Market Research Society 56:441-46.

Weinberg, E. 1983. Data Collection: Planning and Management. In Handbook of Survey Research. Edited by P. H. Rossi, J. D. Wright and A. B. Anderson. San Diego: Academic Press, Inc.

Weisberg, H. F., J. A. Krosnick, and B. D. Bowen. 1996. An Introduction to Survey Research, Polling, and Data Analysis. Thousand Oaks, CA: Sage Publications.

Whitehead, J. C. 1991. Environmental Interest Group Behavior and Self-Selection Bias in Contingent Valuation Mail Surveys. Growth and Change 22:10-21.

Whitehead, J. C., P. A. Groothuis, and G. C. Blomquist. 1993. Testing for Non-response and Sample Selection Bias in Contingent Valuation. Economics Letters 41:215-220.

Whitehead, J. C., P. A. Groothuis, T. J. Hoban, and W. B. Clifford. 1994. Sample Bias in Contingent Valuation: A Comparison of the Correction Methods. Leisure Sciences 16:249-258.

Whittaker, D., J. J. Vaske, M. P. Donnelly, and D. S. DeRuiter. 1998. Mail versus Telephone Surveys: Potential Biases in Expenditure and Willingness-to-Pay. Journal of Park and Recreation Administration 16 (3):15-30.

Chapter 4

INTRODUCTION TO STATED PREFERENCE METHODS

Thomas C. Brown
U.S. Forest Service, Rocky Mountain Research Station

1. INTRODUCTION

Stated preference approaches to nonmarket valuation rely on answers to carefully worded survey questions. Those answers—in the form of monetary amounts, choices, ratings, or other indications of preference—are scaled following an appropriate model of preference to yield a measure of value. Two common, and one not so common, stated preference methods are discussed in the next three chapters.

Most economists, because they tend to distrust people's willingness or ability to answer questions truthfully and carefully, have looked askance at stated preference methods of valuation. But as Manski (2000) argues, the rejection of stated preferences is both naive and limiting, naive because well-designed surveys can avoid many of the potential problems, and limiting because surveys are often the most effective way to understand people's preferences. If economics is to address important issues involving nonmarket goods, it must accept the challenge of refining stated preference methods.

Of course, most economists would prefer to rely on revealed preference methods to value nonmarket goods. Unfortunately, the complementary relation between a market and the nonmarket good that is required for use of revealed preference methods does not always exist. For example, passive use values for wilderness preservation are not clearly linked to consumption of market goods. And even when there is a relation between the nonmarket good and a market good, the assumptions of the methods may not be sufficiently met to allow the method's application. So economists have ventured into the domain of stated preferences.

Careful measurement of stated preferences probably began with L. L. Thurstone's work in the 1920s. He built upon work by psychologists who had studied judgments of phenomena such as the brightness of lights or weight of objects, for which there were observable physical correlates. Error in binary judgments of such phenomena had long been observed to increase the closer were the items on the dimension of interest (e.g., Fechner 1860). Utilizing that observed error in judgment, Thurstone developed a model of binary choice that theoretically allowed measuring judgments on an interval scale (Thurstone 1927b). He then applied the model to measurement of preference dimensions, such as the seriousness of offenses (Thurstone 1927a) and attitudes about religion (Thurstone and Chave 1929).

Economists began using surveys by at least the 1940s to understand consumer purchases,[1] but it was not until the 1960s, with Robert Davis' use of what is now called contingent valuation to value outdoor recreation opportunities (Davis 1963), that surveys were used for nonmarket valuation. Also about that time, Lancaster (1966) laid the economic foundation for methods to value attributes of goods. Owing in part to the increasing interest in nonmarket goods, much has been accomplished since the 1960s in both of these areas, as described in Chapter 5 for contingent valuation and Chapter 6 for attribute-based methods. Chapter 7 focuses on a newly emerging method, based in large part on Thurstone's early work, that uses paired comparisons to order preferences or estimate monetary values for multiple goods using a single survey.

All valuation methods rely on people's innate ability to prefer—to place one object above or below another in a given context. With stated preference methods, that valuation context is described in a survey instrument. Because surveys can describe new goods, limit the choice set, and posit hypothetical markets, they offer possibilities for valuation well beyond those available with revealed preference methods. However, the hypothetical decision making context brings with it serious concerns about the validity of the survey results, a topic to which we return following the next section on classification.

2. CLASSIFYING STATED PREFERENCE METHODS

Although "valuation" in economics commonly refers to measurement of monetary values, valuation also refers to ordering preferences among goods or attributes, whether in terms of a simple ranking or an interval scale. Contingent valuation was designed to measure monetary value, but the other stated preference methods were originally developed to order preferences, and then adapted by economists for monetary valuation (Table 1). Contingent valuation is commonly used to value a single good, although the method is sometimes used to value several closely related goods that differ in the level of a key attribute, allowing inferences about the value of the attribute. A "good," of course, may be a public program, recreation experience, habitat, environmental change, or any other object, and an attribute is a characteristic of a good. Attribute-based methods and paired comparison ask each respondent about numerous goods, thereby allowing estimation of a preference ordering. Attribute-based methods obtain responses about similar goods that differ in the levels of their common attributes; including a monetary cost or benefit as one attribute allows monetary valuation of the other attributes. When a monetary cost or benefit is not included as an attribute, the method produces a preference ordering of the attributes. The paired comparison approach to valuation obtains responses about numerous goods; including sums of money among the items to be compared allows for monetary valuation of the goods. When only goods are included, the paired comparison method allows a preference ordering of the goods.

Table 1. Characteristics of Stated Preference Nonmarket Valuation Methods

Method	Valuation objective	Number of items judged by a respondent	Objects valued
Contingent valuation	Monetary value	one (sometimes more)	Good (or attribute)
Attribute-based methods	Preference order or monetary values	several	Attributes
Paired comparison	Preference order or monetary values	several	Goods

Stated preference methods use a variety of approaches for asking valuation questions, from the straightforward request for maximum willingness to pay of open-ended contingent valuation, to indirect methods using choice, ranking, or ratings. To better understand how the methods compare, it is helpful to review the scales of measurement, which Stevens (1946) characterized as ordinal, interval, and ratio. Measurement on an ordinal scale requires no more than ordering items, as in ranking a set of items from most to least along a given dimension. No information on the distance between the items is provided. Measurement on an interval scale requires that equal differences on the scale signify equal differences on the underlying dimension, but an interval scale lacks a meaningful zero point.[2] Examples include the Fahrenheit temperature scale and the time of day. Measurement on a ratio scale adds to an interval scale a meaningful zero point, one that indicates complete absence on the dimension of interest. Examples include measuring length in inches, weight in grams, and value in money. A ratio scale allows computing ratios among items and thus supports a statement such as "item A weighs twice as much as item B."

Stated preference methods require either ordinal or ratio judgments from respondents (Table 2). Ordinal methods require that respondents be able to order items, such as goods and sums of money, along a given dimension. The simplest ordinal judgment is a binary response, either a choice between two items or an accept/reject (yes/no) choice. The dichotomous choice contingent valuation method, a binary choice approach, asks whether the respondent's willingness to pay for a good is greater than a posited cost. Paired comparison presents two items and asks respondents to choose the one they prefer. Attribute-based methods ask respondents to choose between 2 (binary choice) or more (multinomial choice) items, to rank a set of items, or to rate a set of items along a categorical scale that reveals the order of the items along a given dimension.[3] Ratio-level questions require that respondents state a numerical amount that indicates the value placed on the good. Open-ended contingent valuation, which requires a monetary response, is one obvious example.[4]

Binary choice questions are attractive because they require less of the respondent, in two ways. First, as mentioned above, they require comparison of only two items at a time, unlike ranking, rating, or multinomial choice, which require comparisons among three or more items. Second, the respondent needs to only make an ordinal judgment, unlike ratio scale questions, which require numerical estimates of value.

Table 2. Stated Preference Question Types

Method	Ordinal-level response				Ratio-level response
	Binary choice	Multi-nominal choice	Ranking	Rating	
Open-ended contingent valuation					X
Dichotomous choice contingent valuation	X				
Attribute-based methods	X	X	X	X	
Paired comparison	X				

Much of the intellectual effort invested in valuation methods has focused on increasing the information value of ordinal data (e.g., McFadden 2001). Dichotomous choice contingent valuation, some attribute-based methods for monetary valuation, and paired comparison for monetary valuation each employ binary choice scaling methods to derive ratio-level measures of value from ordinal-level survey responses. Using appropriate scaling procedures, ordinal responses obtained using preference ordering methods can be converted to an interval level of measurement. Responses obtained using a ratio question mode of course directly provide an interval-level or ratio-level measure of value. In essence, ratio-level valuation questions substitute (assumed) respondent ability to quantify values for analyst ability to infer values from ordinal responses.

3. VALIDITY

Validity, in reference to nonmarket valuation, essentially asks the question: Does the estimated value measure the theoretical construct under investigation? For example, does the estimate of mean willingness to pay derived from a contingent valuation study match the population's mean Hicksian surplus?

The surest way to assess the validity of a stated preference measure would be to compare it with a behaviorally-based measure that directly represents the construct under investigation. For example, the result of a survey of buying intentions could be compared with a measure of the same population's subsequent purchase behavior. However, assessing validity in nonmarket valuation is not so simple, because we are typically trying to predict behaviors or reveal preferences that are not directly observable in the market place.

A useful framework for considering the complex, multifaceted validity issue was offered by the American Psychological Association (1974) and summarized by, among others, Bohrnstedt (1983) and Mitchell and Carson (1989). The framework describes three approaches to assessing the validity of a measure, called criterion, construct, and content validity. Criterion validity refers to the comparison of the stated-preference measure with some other measure (the criterion) that is closer to the theoretical construct under investigation. For example, the validity of the contingent valuation referendum format for estimating Hicksian surplus for a public good might be assessed, in part, by comparing the result of a hypothetical referendum with an experimentally controlled vote on providing the same good (whereby participants actually pay if the majority votes in favor of funding the good). Here the experimentally controlled vote is the criterion that is more closely related to the theoretical construct, Hicksian surplus.

Construct validity deals with whether the measure under investigation relates to other measures as predicted by theory. This may be evaluated in two ways: by comparing the measure with other measures that theory suggests should give a similar result (e.g., comparing a contingent valuation measure of willingness to pay with a travel cost measure for the same good), or by investigating the relation between the measure and variables that theory suggests should affect the measure (e.g., examining how contingent willingness to pay varies with income or with behaviors that are thought to be related to willingness to pay).

Content validity takes a different, less objective, tack from the other two approaches. Instead of comparing the observed measure with other measures, it looks at the quality of the survey instrument used to obtain the measure, asking such questions as (1) are the items to be valued unambiguously described to respondents, (2) is the payment vehicle (e.g., in a contingent valuation study) likely to be accepted as reasonable, (3) does the sample represent the population, and (4) is the statistical model appropriate?

Two important content validity questions for stated preferences are: (1) Was the right question asked? (2) Did respondents answer the question that was asked? Asking the wrong question, meaning a question that in theory will not measure the underlying construct, is not uncommon. For example, in applications of the contingent valuation method, if a loss is being valued and willingness to accept compensation is the appropriate measure of the underlying construct, asking for willingness to pay may underestimate the desired measure. Or if a willingness to pay measure of the value of a public good is desired, asking for willingness to donate to a fund to provide the good will underestimate willingness to pay unless it can be shown that free-riding does not unduly occur with the donation payment vehicle.

Asking a question that is not answered by some respondents as the survey intended—perhaps because they misinterpreted the question or answered strategically—is also probably common. When respondents misinterpret the question or answer strategically, they in essence answer a different question. For example, if asked to choose between options for improving air quality in an attribute-based survey, some respondents might always choose the option with the greatest benefits, regardless of the posited cost, because they believe their answers may encourage policies that they favor but will not have to pay for directly. Such respondents are in essence answering the question "Which option provides the greater improvement in air quality?" Similar challenges face the paired comparison method of monetary valuation.

Stated preference valuation questions typically require respondents to estimate their willingness to pay for one or more goods. Estimating one's willingness to pay becomes more difficult the less familiar and more complex is the good. We can all estimate with considerable precision our willingness to pay for a cup of coffee or a certain music CD. We can probably do so as well for access to a well-know recreation site. But estimating our willingness to pay for goods we have never thought about in monetary terms, if at all, is more difficult. Estimating our willingness to pay for protecting a habitat or species we have never seen or thought much about is a bit like estimating willingness to pay for a thneed, the good produced by the Once'ler in Dr. Seuss' "The Lorax." This task requires learning about the good and then searching in memory for relevant information, such as information about related goods that we have some experience purchasing. Given the hypothetical nature of a survey, that mental search will probably not be as thorough or context-free as the researcher would like (Sudman, Bradburn, and Schwarz 1996). Furthermore, even if highly motivated, some respondents may simply not be able to achieve more than a vague notion of their willingness to pay for an unfamiliar good (Gregory et al. 1995). Some error in the response is therefore to be expected. If this error is random, it will not threaten the validity of the welfare estimate as long as proper statistical analysis procedures are used. However, if the error is overly influenced by contextual factors, and therefore nonrandom, the welfare estimate is likely to be biased.

Good survey design helps avoid context effects, and such effects are less likely in any case when valuing familiar nonmarket goods, such as the recreation site mentioned above. The difficulty of avoiding contextual influences increases as the good becomes less familiar and more complex, because respondents become more open to extraneous influences when they have less past experience to draw upon or have difficulty in conceptualizing the good to be valued. Monetary valuation of complex, unfamiliar goods—whether using contingent valuation, attribute-based methods, or paired comparisons—is a significant challenge, as the reader will learn in the following three chapters.

Validity is obviously not an either/or matter. All survey-based measures face validity issues, and choosing the best measure for a given application may require comparing degrees of validity. Examining the validity of a stated preference measure requires good judgment—judgment that is enhanced by a thorough understanding of the theoretical and methodological issues underlying the valuation method.

Because stated preference methods rely on answers to questions about hypothetical situations, the methods are rightly subjected to careful scrutiny. However, that scrutiny should not be interpreted as an indication that stated preference estimates are necessarily less valid than revealed preference estimates. As is seen in later chapters, revealed preference methods rely on many assumptions and thus also face serious validity issues.

4. STEPS OF MEASUREMENT

Regardless of the method used, the process of estimating nonmarket values via stated preferences has four major steps (Table 3). During the first step, critical decisions are made about the valuation objective of the study and how that objective will be met. The objective of the study should of course respond to the policy question of interest. For example, if the question is whether an area of national forest should be designated as wilderness, the objective might be to estimate the economic benefits of designation. Next the population of interest is specified. Continuing with the example, the relevant population for

Table 3. Process of Measuring Values Using Stated Preference Methods

1. Map out the valuation approach by specifying the following:
   a. measurement objective to respond to the policy question
   b. population to be sampled
   c. theoretical construct that satisfies the measurement objective
   d. valuation method that suits the theoretical construct
   e. response mode of the valuation question(s)
   f. measure of value
   g. statistical model used to scale the valuation responses
2. Design the survey instrument and sampling plan by specifying the following:
   a. item (goods or attributes) to be valued
   b. monetary amounts to be used (necessary for some methods, such as dichotomous choice contingent valuation)
   c. independent variables (if any) to be measured
   d. method of administration (e.g., mail, phone)
   e. other details of the instrument, such as background information about the good, information about substitutes, order of the questions, and use of graphics.
   f. the sample
   g. sampling details, such as method of contacting respondents, method of encouraging response, and schedule of activities.
3. Administer the survey
4. Clean and analyze the data

deciding about a national forest wilderness area is the entire U.S. public. Then the theoretical construct is specified (e.g., a compensating surplus measure), the valuation method is chosen (e.g., contingent valuation), a response mode is selected (e.g., dichotomous choice), the specific measure of value is specified (e.g., mean willingness to pay per year in increased income tax for the next ten years), and a statistical model is chosen (e.g., logit model).

During step two, the survey instrument and sampling plan are designed. Design of the survey instrument involves specifying the goods or attributes to be valued (e.g., a careful description of the proposed wilderness area) and any independent variables to be collected in the survey (e.g., income, education), the method of administration (e.g., mail survey), and other details such as descriptions of substitutes the presence of which may affect value for the good of interest, the order of the questions, and use of graphics. Qualitative research methods (e.g., focus groups) may be used for this step to help decide what information should be included in the survey materials and how that information should be conveyed. For some methods, such as attribute-based methods, this stage would include developing an experimental design. The sample frame would be defined to represent the study population as closely as study resources allow, and a plan for contacting the sample and encouraging response would be made.

During step three the survey is administered. During step four the data are cleaned and analyzed, which includes scaling the responses to estimate the magnitudes on the dimension of interest (e.g., binary choices are converted to an estimate of mean willingness to pay). Analysis may also include estimating a model relating dependent to independent variables.

Step 1 relies most importantly on a thorough understanding of the theory of nonmarket valuation. Step 2 relies critically on an understanding of the nuances of survey design, both because survey design leaves the researcher with so much latitude, and because almost any question will receive an answer—though not necessarily an answer to the question the researcher sought to ask. Step 3 relies on diligence in carrying out the plan. Step 4 relies on good training in econometrics. Each of these four steps is replete with judgment calls, as the next three chapters explain in more detail.

NOTES

1 As Juster (1966) explains, the Federal Reserve Board began in 1945 sponsoring surveys to, among other things, understand consumer attitudes and predict their purchases. Unlike nonmarket valuation, purchase intentions could easily be compared with subsequent actual purchases, allowing an evaluation of response validity.

2 When economists refer to a "cardinal" measure they mean either an interval or ratio scale measure; for example, when a cardinal measure of utility was thought possible an interval scale measure was implied (von Neumann and Morgenstern 1944), and when price or income is measured a ratio scale measure is implied.

3 Respondents could use a rating scale as an interval scale (such that, for example, a rating difference of 4 indicates twice as much difference on the underlying dimension as does a rating difference of 2), but interval properties are not necessarily assumed for rating data.

4 A approach using ratio-level questions to order preferences, not included in Table 2, is magnitude estimation. Respondents may be asked to allocate a given total (say, 100 points) among a set of alternatives, or asked to assign a number (any number) to one item and then to assign numbers to remaining items that reflect the ratio of their values to the value of the first item (Stevens 1966).

REFERENCES

American Psychological Association. 1974. *Standards of Educational and Psychological Tests.* Washington, D. C.: American Psychological Association.

Bohrnstedt, G. W. 1983. Measurement. Pages 69-121 in *Handbook of Survey Research*, edited by P. H. Rossi, J. D. Wright and A. B. Anderson. New York: Academic Press.

Davis, R. K. 1963. Recreation Planning as an Economic Problem. *Natural Resources Journal* 3(2):239-249.

Fechner, G. T. 1860. *Elemente der Psychophysik.* Leipzip: Breitkopf and Hartel.

Juster, F. T. 1966. Consumer Buying Intentions and Purchase Probability: An Experiment in Survey Design. *Journal of the American Statistical Association* 61:658-696.

Lancaster, K. J. 1966. A New Approach to Consumer Theory. *Journal of Political Economy* 74:132-147.

Gregory, R., S. Lichtenstein, T. Brown, G. Peterson, and P. Slovic. 1995. How Precise Are Monetary Representations of Environmental Improvements? *Land Economics* 71(4):462-473.

Manski, C. F. 2000. Economic Analysis of Social Interactions. *Journal of Economic Perspectives* 14(3):115-136.

McFadden, D. 2001. Economic Choices. *American Economic Review* 91(3):351-378.

Mitchell, R. C., and R. T. Carson. 1989. *Using Surveys to Value Public Goods: The Contingent Valuation Method.* Washington, D. C.: Resources for the Future.

Stevens, S. S. 1946. On the Theory of Scales and Measurement. *Science* 103:667-680.

Stevens, S. S. 1966. A Metric for the Social Consensus. *Science* 151(Februray 4):530-541.

Sudman, S., N. M. Bradburn, and N. Schwarz. 1996. *Thinking About Answers.* San Francisco: Jossey-Bass Publishers.

Thurstone, L. L. 1927a. The Method of Paired Comparisons for Social Values. *Journal of Abnormal Social Psychology* 21:384-400.

Thurstone, L. L. 1927b. Psychophysical Analysis. *American Journal of Psychology* 38:368-389.

Thurstone, L. L., and E. J. Chave. 1929. *The Measurement of Attitude*. Chicago, IL: University of Chicago Press.

von Neumann, S, and O. Morgenstern. 1944. *The Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press.

Chapter 5

CONTINGENT VALUATION IN PRACTICE

Kevin J. Boyle
University of Maine

1. INTRODUCTION

Contingent valuation is a survey-based methodology for eliciting values people place on goods, services, and amenities. The first contingent valuation study was conducted by Davis (1963) to estimate the value of big game hunting in Maine. A decade later, Hammack and Brown (1974) applied contingent valuation to valuing waterfowl hunting. Simultaneously, an application to valuing visibility in the Four Corners region of the Southwest represented a turning point after which contingent valuation gained recognition as a methodology for estimating Hicksian surplus for public goods (Randall, Ives, and Eastman 1974). Contingent valuation filled a substantial void by providing a way to estimate values when markets do not exist and revealed preference methods are not applicable.

Results from early applications of contingent valuation met with skepticism and criticism. One of the more notorious comments was expressed by Scott (1965) who referred to contingent valuation as a "short cut" and concluded: "ask a hypothetical question and you get a hypothetical answer" (p. 37). Some of this criticism was deflected by Bishop and Heberlein's (1979) landmark validity study in which they compared welfare estimates for goose hunting from actual cash transactions, contingent valuation, and travel cost. This study showed that the contingent valuation estimate of willingness-to-pay (WTP) was of similar magnitude to estimates of WTP provided by a travel cost model and the cash transactions. These results suggested that contingent valuation met the conditions of convergent validity (comparable to travel cost estimates) and provided a conservative estimate from the perspective of criterion validity (less than the cash transaction estimate) (Carmines and Zeller 1979).

A workshop sponsored by the U. S. Environmental Protection Agency was the first attempt to synthesize what was known about contingent valuation (Cummings, Brookshire, and Schulze 1986). The notable outcome of this "state of-the-art assessment" was a set of reference operating conditions for the conduct of a credible contingent valuation study. The conditions require respondent familiarity and choice experience with the commodity, little uncertainty in the valuation exercise, and the use of willingness-to-pay. While the conditions provided some guidance for the types of applications where contingent valuation could be credibly applied, Freeman (1986) noted that the restrictive nature of the conditions essentially implied that contingent valuation ". . . is likely to work best for those kinds of problems where we need it the least. . . ." (p. 160). That is, the conditions imply that contingent valuation works well only where travel cost or other revealed-preference methods are readily applied. The assessment in the Cummings, Brookshire, and Schulze (1986) book set off a multitude of research projects to evaluate the validity of contingent valuation and to probe the limits of the types of applications where contingent valuation could provide credible welfare estimates.

Perhaps the most substantive contribution to the contingent valuation literature is a book by Mitchell and Carson (1989), which presented the first attempt to develop detailed recommendations for designing a contingent valuation study. The book provided the broad overview that novices require when conducting their first contingent valuation study, as well as prescriptive recommendations that set off a whole new wave of validity research. Mitchell and Carson fundamentally shifted the research focus to one that considered the details of study design. Validity, rather than being a global, all-or-nothing criterion, was now viewed as a function of specific aspects of study design. The design of specific study elements may enhance or reduce validity by reducing or increasing bias in point estimates of central tendency, and such design may increase or decrease efficiency by affecting the dispersion around the estimates of central tendency.

Through the first 25 or more years of the use of contingent valuation, critiques of this methodology seemed to ebb and flow, without a specific focal point of attack. This all changed when contingent valuation estimates began to be used in legal cases as the basis of damage payments by parties responsible for large-scale pollution under the Comprehensive Environmental Response, Compensation and Liability Act (CERCLA) of 1980 (Ward and Duffield 1992). The controversy became particularly heated after the settlement of the Natural Resources Damage claim for the Exxon Valdez oil spill. Exxon supported the publication of a book that critiqued the fundamental premises of contingent valuation (Hausman 1993), and the National Oceanic and Atmospheric Administration (NOAA) responded with a blue ribbon panel to evaluate the credibility of using contingent valuation to estimate nonuse values (NOAA 1993).

The NOAA panel provided specific recommendations on how a contingent valuation study should be designed and conducted to develop "reliable" estimates of non-use values. Unlike the assessments by Cummings, Brookshire and Schulze (1986) and Mitchell and Carson (1989), the NOAA panel's recommendations were not clearly grounded in the existing contingent valuation literature and can best be interpreted as informed opinions. The panel's recommendations did set off another wave of research designed to investigate the credibility of contingent valuation, particularly in the context of estimating non-use values.[1]

While the books by Cummings, Brookshire, and Schulze (1986), Mitchell and Carson (1989), and the NOAA (1993) panel report have been landmark contributions, they have not dispelled the critiques of contingent valuation. In fact, the critiques of contingent valuation, which have become more direct and focused over time, have made contingent valuation practitioners much more cautious about using estimates of non-use values when evaluating public policies and developing damage claims for court cases. However, the use of contingent valuation for policy analyses and damage claims has not stopped. The critiques have helped to focus the research agenda in a manner that has led to more credible welfare estimates.

The research agenda, like that of any area of study, has been motivated by interests of individual researchers, funding availability, and various critiques. Thus, research investigating the validity of contingent valuation has not followed a systematic path. The lack of a systematic research agenda, combined with the large variability in procedures used to implement contingent valuation studies, has led to confusion among novices and those who use the welfare estimates in decisionmaking regarding what constitutes a credible study design. Even those who are thoroughly familiar with contingent valuation are not always clear when designing a study what will be considered a credible design by the time their value estimates are published.

Mitchell and Carson's book is still a very useful reference on the basic issues to be addressed in a contingent valuation study, but much research has been conducted since the book was published in 1989. Moreover, anecdotal evidence suggests that there is a demand by students, policy makers, and others for a concise overview of contingent valuation. This chapter is intended to fill these voids. The concise nature of this chapter precludes coverage of all design features in substantive detail, and I will focus on the literature subsequent to 1990. There are contributions prior to this date that still have a major impact today; these earlier contributions are generally summarized in Mitchell and Carson. Furthermore, given the large number of contingent valuation studies being conducted today and the wide range of journals in which they are being published, it is impossible to cover all contributions to the literature since 1990. Readers are encouraged to undertake electronic literature searches to identify additional contributions specific to your particular topic of interest.

Valuation of groundwater will be used as an example to help explain some of the steps involved in designing and conducting a contingent valuation study. However, contingent valuation can be applied to a wide array of policy issues ranging from the early hunting applications cited above to more recent health-care applications (Alberini et al. 1997; Alberini, and Krupnick 2000; Dickie and Gerking 1996). Methodological citations will be drawn from all types of resource applications in the peer-reviewed literature, not just groundwater valuation.

It is worth noting that the number of journals where contingent valuation studies have been published is expanding, both in terms of economics journals and journals sponsored by other academic disciplines (Hanemann 1984; Randall 1998; and Smith 1993). This suggests a wider academic and professional acceptance of contingent valuation. In addition, economists outside the area of environmental economics have joined the discourse (e.g., Hausmann 1993) as well as people from other academic disciplines (e.g., Ajzen, Rosenthal and Brown 2000; Fischoff and Furby 1988; Sckade and Payne 1994). The simple fact that people of this caliber take the time to consider contingent valuation, even if the consideration is often critical, is a testament to the quality and progress of the design of contingent valuation studies.

2. STEPS IN CONDUCTING A CONTINGENT VALUATION STUDY

Most of the action in designing a contingent valuation study occurs in the development of the survey instrument and during data analyses (Table 1). It is at these two stages of a study that investigator choices can substantially affect welfare estimates. Thus, careful design of a contingent valuation survey and careful analysis of the resultant data are crucial to the estimation of credible welfare estimates.

2.1 Identifying the Change in Quantity or Quality to be Valued

This step involves developing a theoretical model of the value(s) to be estimated, which is based on the difference between the baseline utility with the current environmental condition and the utility with the new environmental condition. Following the notation protocol established in Chapter 2 (see equation 2), Hicksian surplus for a program to protect groundwater from contamination an be defined as

$$(1) \qquad v(p^0, Q^0, y-c) = v(p^1, Q^1, y)$$

where $v(\cdot)$ is indirect utility function, p is the price of drinking water obtained from groundwater, Q is groundwater quality, 0 denotes current water quality, 1 denotes quality if the protection program is not implemented, y is income, and c is Hicksian compensating variation (WTP).[2,3] The value of interest is that of protecting groundwater from contamination ($Q^0>Q^1$), i.e., maintaining the status quo. The purpose here is not to delve into details of theoretical definitions of changes in welfare, which is the domain of Chapter 2, but to discuss aspects of the value definition that must be considered in the design of any contingent-valuation study.

The theoretical definition of value is fundamental to three key components of any contingent valuation study. First, the definition plays a central role in developing wording in the survey to describe conditions with and without the policy to be valued (Step 5.1). Second, the definition frames the statistical

Table 1. Steps in Conducting a Contingent Valuation Study

1. Identify the change(s) in quantity or quality to be valued.
2. Identify whose values are to be estimated.
3. Select a data collection mode.
4. Choose a sample size.
5. Design the information component of the survey instrument.
   - 5.1 Describe the item to be valued.
   - 5.2 Explain the method of provision.
   - 5.3 Select a payment vehicle.
   - 5.4 Select a decision rule.
   - 5.5 Select a time frame of payment.
6. Design the contingent-valuation question.
   - 6.1 Select a response format.
   - 6.2 Allow for people to respond with values of $0.
   - 6.3 Develop questions to screen for protest and other types of misleading responses.
7. Develop auxiliary questions for inclusion in the survey instrument.
   - 7.1 Develop questions that provide covariates for statistical analyses of valuation responses.
   - 7.2 Develop questions that help to assess the validity of valuation responses.
8. Pretest and implement the survey.
9. Develop data analysis procedures and conduct statistical analyses.
10. Report value estimates.

analysis of the contingent valuation responses (step 9). Third, having a theoretical definition allows for clear interpretations of value estimates in specific policy contexts.[4] Policy is used in a very general sense to refer to any condition that may change utility and for which welfare estimates are required.

This broad definition includes new laws and regulations as well as fines or legal damage assessments.

The difficult part of step 1 is identifying the physical change and how it affects utility. Economists depend on physical and biological information to establish status quo conditions (Q^0) and predictions of conditions without the policy (Q^1). This information may not be known with certainty, and applied valuation further requires information about this uncertainty. For example, if a policy were designed to reduce the probability that groundwater will become contaminated, the definition of value for the groundwater protection program would be as follows:

$$(2) \qquad \Pi_1 v(p^0, Q^0, y - op) + (1 - \Pi_1) v(p^1, Q^1, y - op) = \Pi_0 v(p^0, Q^0, y) + (1 - \Pi_0) v(p^1, Q^1, y)$$

where (1-Π) is the probability of contamination, $\Pi_0 < \Pi_1$, and op is WTP (option price as defined in equation 32, Chapter 2) to reduce the probability of contamination (Bishop 1982). Here, Π_0 is the probability that the current water quality Q^0 will not be degraded, and Π_1 is the policy-enhanced probability that Q^0 will not be degraded. Thus, op is maximum WTP to increase the probability that the current water quality will be maintained. Applying contingent valuation to conditions of uncertainty, especially for groundwater, is a common practice.

The fundamental information transmitted in the survey are descriptions of the changes in resource conditions due to the policy being valued. For example, a policy to ration use of groundwater as a source of drinking water requires a survey description detailing current conditions in the aquifer and what would happen if current extraction rates continued. The policy condition would describe the reduction in groundwater extraction and the consequent effect on the availability of drinking water over time.

Physical descriptions of changes in resource conditions frequently are not available. In this case, contingent valuation questions often are framed to value the policy change. With vague or nonexistent information on the resource change, survey respondents are left to their own assumptions regarding what the policy change will accomplish. This is a problem because different respondents will use different assumptions, i.e., they are valuing different resource changes. In addition, it is usually the change in the resource, not the policy itself, that directly affects the services people enjoy; is the change in resource services that should conceivably enter as arguments in the indirect utility function. Without that information, survey respondents are left to make two assumptions: (1) how the policy change affects resource conditions, and (2) how the change in the resource affects the services they receive. This issue really has not been directly and extensively addressed in the contingent valuation literature and deserves more consideration.

It may also be the case that a contingent valuation study is being conducted in advance of an anticipated policy, and the details of the policy and the consequent effects of the policy are not known. Contingent valuation studies are often designed to estimate an array of values that represent plausible conditions that might occur when the physical information becomes known or the actual policy is finalized. The practitioner is often left to proceed with the design of a contingent valuation study using a "best guess" of the actual policy that will be implemented and of the ultimate effects with and without the policy. Here is where a careful theoretical definition of value is crucial; it allows an *ex post* interpretation of whether the value estimate(s) are relevant for the realized change in a resource.

In addressing this step it is important to recognize that there are two types of contingent valuation studies. Many studies in the peer-reviewed literature focus on methodological contributions and are not designed to address a specific policy issue (e.g., Alberini 1995a), while other studies have been designed to address a specific policy issue (Carson et al. 1998). This difference is crucial because methodological studies can use whatever definition of value the investigator finds convenient. In a policy study, however, the value definition must be linked to the specific change in utility that will occur if the proposed policy is implemented. Methodological studies still require theoretical definitions of estimated values to guide data analyses. Theoretical definitions of value allow estimates from methodological studies to be useful for benefit transfers.

The discussion in this section has focused on Hicksian measures of value, which is the theoretical concept typically measured in contingent valuations studies. Step 5.1 involves an explanation of the theoretical definition of value presented within the survey instrument and the policy change that respondents are asked to value.

2.2 Identify Whose Values Are to Be Estimated

Once the policy change has been specified, the affected population can be identified. Step 2 involves translating the policy change into effects on people, and identifying who will be affected. This information is important in selecting a sample frame so that the contingent valuation survey will be administered to a representative sample of people. Contingent valuation studies result in point estimates of values on a per-capita or per-household basis. It is necessary to know the size of the affected population to expand the per-unit values to a population value for policy purposes. The affected population may also indicate a mode of data collection that is most desirable or modes of data collection that are clearly unacceptable for the particular application.

In the example where groundwater usage is to be rationed, the affected population constitutes those individuals who use the aquifer as a source of drinking water. If everyone in a community obtains water for household usage from a community supply and that supply is the only extraction from the aquifer, then it is fairly clear that the affected population is the community.

A more complicated issue is the unit of measurement for values. Some contingent valuation surveys elicit values for individuals (e.g., Bateman et al. 1995) and others elicit household values (e.g., Poe and Bishop 2001). It is important that the framing of contingent valuation questions makes clear whether a household valuation or an individual valuation is being requested. In studies asking for household values, respondents generally understand that a household value is being asked for, but this is not always the case in studies asking for individual values. For example, in the Bateman study respondents were asked for their WTP to "prevent flooding and preserve the Broads" (p. 164). It is unclear whether people answered for themselves, for their households, or whether some answered for themselves and others answered for their households. One might argue that applications to recreation activities, such as Bishop and Heberlein's (1979) study of goose hunting, may result in a clear interpretation that the contingent valuation question is asking for an individual value. However, even here the interpretation of valuation responses may be clouded if hunting is a group decision among a few friends.

Mitchell and Carson (1989) assert that ". . . payments for most pure public goods are made at the household level. [...] When this is the case, the appropriate sampling procedure is to allow any adult who claims to be a household head to be a spokesperson for the household—the current practice of the U.S. Census Bureau" (p. 265-266). (Mitchell and Carson cite Becker's (1981) *Treatise on the Family* in support of this assertion.) However, when the household decision maker is the response unit in a contingent valuation study, it is important to identify who the appropriate decision maker is in the household. Do households pool their income and make group decisions? Do individuals make their own decisions? Do some households make group decisions and others make individual decisions? Or, are some decisions within the household group decisions and others individual decisions? Answers to these questions have important implications for the credibility of the welfare estimates from contingent valuation studies. Researchers make decisions on individual or household units of measurement based on the specific policy change being valued, pretesting, the availability of a sampling frame, and perhaps personal intuition.

When selecting a sample frame, each unit in the sample should have a known probability of selection from a specified population of known size. The choice of a unit of measurement can be refined in focus groups or one-on-one interviews and should be clearly stated within the valuation question.

2.3 Select a Data Collection Mode

A contingent valuation study requires the collection of primary data (step 3). The various modes of data collection are discussed in detail in Chapter 3 of this volume, so this section focuses on insights for contingent valuation applications.

The most common way to implement contingent valuation surveys is by mail (Schneemann 1997), but Mitchell and Carson (1989) and the NOAA panel (NOAA 1993) both advocated the use of personal interviews. Telephone surveys are also an option (Schuman 1996). Each method has its relative strengths and weaknesses.

The primary reason that mail surveys are used is that they are the cheapest way to collect data. A call to a survey-research firm that has administered a number of contingent valuation studies suggests that the costs of a completed survey ranges from \$25–30 for a mail survey to, \$30–35 for a telephone survey, to \$50–100 for personal interviews. Thus, the cost advantage of a mail survey over a telephone survey is slight, while the cost advantage of mail and telephone surveys over personal interviews is substantial. With a limited budget a less expensive mode enables a larger sample size.

Another factor that affects the choice of a survey mode is the expected survey response rate. If a specific segment of the sample chooses not to respond to a survey or to not answer the contingent valuation question, then a study can result in biased welfare estimates (Edwards and Anderson 1987). Even with careful design, Schneemann (1997) has shown that response rates to contingent valuation surveys are affected by factors that are not under the control of the researcher. For example, policy issues that involve specific user groups (e.g., anglers or hunters) are likely to result in higher response rates than general population surveys. Schneemann finds that survey response rates are related to the resource the study is designed to value, the affected population, and components of the contingent valuation question itself. Thus, even with a good design process, contingent valuation practitioners need to recognize that features of the study application, as well as the design of contingent valuation exercise, affect survey response rates.

Provision of information on the item being valued is the fundamental component of a contingent valuation survey. Personal interviews have the highest ability because visual information is provided and an interviewer is available to explain the information and answer questions.[5] A mail survey is more limited because no interviewer is present to explain the visual information. Ability to provide information in a telephone survey is much more limited because no visual information is available. Mixed mode surveys using a telephone interview after respondents have received written and visual information in the mail, similar to the content of a mail survey, is one way to overcome the informational deficiencies of telephone interviews (Hanemann, Loomis, and Kanninen 1991). However, the repeated contacts of this mixed mode may reduce response rates.

Other methods of implementing contingent valuation surveys include mall intercepts (Boyle et al. 1994) and convenience samples of students or other groups who are brought to a central location to participate in the study (Cummings and Taylor 1998). These alternative survey settings are clearly appropriate for doing methodological tests. Davis and Holt (1993) note that "behavior of (real) decision makers has typically not differed from that exhibited by more standard . . . student subject pools." (See also Smith, Suchanek, and Williams 1988.) However, Davis and Holt were referring to induced value studies, and questions remain regarding whether contingent valuation estimates derived from samples of convenience can be used to develop population inferences of gains and losses in economic welfare.

As noted in Chapter 3, the use of the Internet and web-based surveys provide new opportunities for conducting contingent valuation surveys. Many of the same concerns discussed for mail surveys apply to these modes, and the limited sample issues of mall intercepts and central location settings also apply. Several studies are underway that use these newer modes.

There have been some comparisons of contingent valuation studies conducted with different modes (Ethier et al. 2000; Loomis and King 1994; Mannesto and Loomis 1991; Whitaker et al. 1998). Such comparisons carry five questions: (1) is it possible to develop identical sample frames, (2) are the response rates the same across modes, (3) do the same types of people respond to each mode, (4) do the different modes result in differing levels of item nonresponse to the contingent valuation question, and (5) are welfare estimates affected by the survey mode? When differences in welfare estimates are observed among survey modes it is difficult to assess whether the differences are due to variation in sample frames or to the basic distinctions in the survey modes. The comparison studies cited above have not identified clear, systematic distinctions among survey modes. The general finding, however, is that survey modes do affect estimates of value estimates, although the source of the differences is not clearly attributable to one of the five questions posed above.

Practitioners need to consider the advantages of each mode when implementing a contingent valuation survey. A mail survey does not always dominate just because of cost advantages and personal interviews do not always dominate just because of informational advantages. Moreover, investigator choices can affect response rates for each of these survey modes.

2.4 Choose a Sample Size

As noted in Chapter 3, selection of a sample size (step 4) is a matter of choosing an acceptable level of precision within a given budget. Mitchell and Carson (1989) note that contingent valuation studies "require large sample sizes because of the large variance in the (WTP) responses" (p. 224). That is, the standard error of mean WTP is:

$$(3) \quad se_{\overline{\mathrm{wtp}}} = \frac{\sigma}{\sqrt{n}}$$

where σ is the standard deviation and n is the sample size. Thus, for a given variance, the absolute value of the standard error can be reduced by increasing the sample size. This means that the same sample size does not fit all contingent valuation applications, and it is not appropriate to select samples as a percentage of the affected population. The larger is σ, the larger the required sample size to attain the desired level of precision. Additionally, for a fixed σ, a sample size will represent a larger percentage of the population for a small population than it will for a large population (Salant and Dillman 1994).

For policy applications where there have been a lot of contingent valuation studies conducted (e.g., recreational fishing), there may be sufficient information to develop a reasonable estimate of σ for a new application. Where existing studies are not available, an estimate of σ can be obtained through a field test of the survey instrument. In practice, most studies choose the largest sample size possible given the available budget.

Selecting a sample size also involves consideration of the expected response rate; the expected percentage of invalid mailing addresses, phone numbers, etc.; and the expected item nonresponse to the contingent valuation question and other co-variates that will be used to analyze contingent-valuation responses.[6] Other considerations include whether the sample will be stratified into subsamples for analysis and data reporting. Another consideration is whether every person or household contacted is eligible to participate in the survey.[7]

Concerns about sample sizes are important for two reasons. First, the precision of estimated values affects their usefulness in policy analyses. An estimate with large error bounds can leave questions about whether benefits really exceed costs or what should be a specific damage payment in an Natural Resources Damage Assessment case. Second, statistical precision affects the ability to detect differences among value estimates in methodological studies designed to investigate the validity and reliability of contingent valuation estimates.

2.5 Design the Information Component of the Survey Instrument

Step 5 focuses on the information provided to respondents in the survey instrument. This includes telling respondents what it is they are being asked to value, how it would be provided, and how they would pay for it. While this is a crucial component in the design of the contingent valuation survey, most of the literature has focused on the effects on welfare analyses of different response formats and analysis procedures. That is, while most practitioners believe that the design of the information component is a crucial component of any contingent valuation study, the implicit assumption of many methodological analyses in the peer-reviewed literature is that respondents understand the information and that the information presentation does not affect the outcome of statistical tests. This implicit assumption may be invalid, and careful consideration must be given to all the information provided to respondents in a contingent valuation survey instrument.

2.5.1 Describe the Item to Be Valued

This information presents the change in the quantity, quality, or probability to be valued that was identified in step 1. The information is presented in written or verbal form and is accompanied by graphs, pictures, and other visual stimuli to facilitate respondent understanding. The information scenario is not a marketing or sales pitch but a neutral and fair description of the change to be valued, and it may include statements about why some people would desire the change and others would not.

While the description of the item to be valued is the fundamental component in the design of any contingent valuation study, it seems that the information is rarely complete in terms of the baseline condition(s) and the new condition(s) that will result as a consequence of the policy change. While either the current condition or the new condition is defined, the other condition often is missing. This problem arises most frequently in the estimation of values for marginal changes in resources, and to a lesser extent when all-or-nothing values are elicited. If respondents must infer the change being valued, it is likely that different respondents will use different subjective perceptions. This reduces the credibility of value estimates for the original policy application and for transfers to new policy applications.

Some have concluded that a credible contingent valuation study requires that survey respondents be provided extensive, detailed information. The literature does not support such a conclusion, but the literature does indicate that specific types of information should be provided to respondents. While there have been only a handful of studies that have investigated the effects of information on contingent valuation estimates, these studies collectively tell an important story.

Samples, Dixon, and Gowen (1986) considered three different levels of information in the valuation of wildlife. The baseline valuation scenario presented the species simply as A, B and C, with no specific information about each. Despite this lack of information, estimates of central tendency were obtained for each species; all three estimates were significantly different from zero, but were not significantly different from each other. When respondents were told each species' identify (rabbit, monkey, and rat) and endangered status, the welfare estimates for each species changed significantly in a manner consistent with what would be expected from economic theory; the estimates also were significantly different from each other. This study confirms the obvious: you must tell people what it is they are being asked to value in order to establish content validity (Carmines and Zeller 1979). Absence of the specific information about the species resulted in biased estimates of WTP.

Boyle (1989), in a study of brown trout fishing, provided different types of information to three independent samples of respondents. The baseline information told respondents that values were being estimated for trout fishing in southern Wisconsin streams, and a map was provided to show respondents where the streams are located. The second group was given the baseline information and was told about the stocking of brown trout in the affected area as well as the composition of catches among brown trout, rainbow trout, and brook trout. The third group was further told about the cost of stocking brown trout in the study area. While estimates of central tendency decreased as respondents received additional information, the reductions were not statistically significant. The standard errors of the welfare estimates decreased significantly with additional information. This reduction in dispersion suggests information also affects the efficiency of value estimates.

Bergstrom, Stoll, and Randall (1990) investigated the effect on value estimates of providing information about the services provided by wetlands. From a theoretical perspective, these authors did not posit whether service information should increase or decrease welfare estimates, only that there should be an effect. However, value estimates were affected by different types of service information, which suggests that in the absence of this information respondents did not fully consider, or were not aware, of the services provided by the resources being valued. This outcome suggests that specific information on services may be important in applications where respondents are not fully aware of how they currently benefit from a resource or may benefit from the resource in the future.

Poe and Bishop (1999) demonstrated that specific information on well-water contamination was required in a study of groundwater protection. The specific information told respondents the level of contamination in their own well in addition to general information about groundwater contamination in their area. This specific information helped respondents to better understand (a) the change in quality they were being asked to value and (b) how the groundwater protection program would affect their specific service flow.

These studies clearly indicate that specific information about the item being valued is required in order to elicit credible responses to contingent valuation questions. Ajzen, Brown, and Rosenthal (1996) found that information effects on value estimates are related to the personal relevance of the good being valued. The specific information being called for enhances the personal relevance of the policy change to survey respondents.

The specific information includes the item and change being valued, services provided, and the effects on services to individual respondents and their households. This does not require an extensive, detailed valuation scenario, but rather a scenario that is clear so that respondents understand the resource change they are being asked to value. This appears to be what the NOAA panel (NOAA 1993) was calling for when they called for an "accurate description of the . . . policy" (p. 4608) as one guideline for reliable and useful contingent valuation estimates. There are no "cookie cutter" or "one-size-fits-all" rules for framing specific and accurate information scenarios, but testing draft information scenarios in focus groups can be quite helpful. Focus group testing allows the investigator to learn if respondents are using the information, understand and believe the information, and are basing valuation responses on the actual change being valued.

The NOAA panel also called for giving respondents specific information about substitutes and reminders that they can spend their money on other items (budget constraint). This is logical from the perspective of economic theory. Both the availability of substitutes (Freeman 1993, Chapter 3; Hoehn and Loomis 1993; Flores, Chapter 2, this book) and income (Flores and Carson 1997) fundamentally affect the magnitude of Hicksian surplus. The presumption of this recommendation is that people will not think about substitutes or their budget constraints unless they are prompted to do so.

These design recommendations are intuitively straight forward, but it is difficult to test their effectiveness. What may be considered as a substitute by one respondent is not a substitute for another respondent. In addition, a variety of factors could be causing the lack of a statistically detectable effect of substitute reminders: respondents are already considering substitutes, the wrong substitutes are mentioned in the survey, or the reminder is not having the desired effect. The first and third interpretations of an insignificant effect also apply to any investigation of budget constraint reminders. However, split-sample studies, where one sample is reminded of substitutes and their budget constraints and another sample is not, reveal that information on substitutes, complements, and budget constraints affect contingent valuation estimates of central tendency and dispersion (Kotchen and Reiling 1999; Whitehead and Blomquist 1995). Given the role that substitutes, complements, and income play in theoretical definitions of economic values, theoretical validity suggests that respondents should be prompted to consider likely substitutes and complements, and they should be reminded that they could spend their money otherwise.

The NOAA panel (NOAA 1993) also recommended the use of pictures in the study design. While some studies have used pictures in valuation scenarios, there do not appear to be any studies that have evaluated whether the use of pictures affects value estimates. In three recent studies I conducted, while pictures seemed like a good idea in the beginning, pretesting suggested that pictures were problematic. In two of these instances pictures reduced the credibility of the contingent valuation scenario, and in the third instance respondents kept using the pictures to seek unintended clues. For example, in one of these studies respondents indicated that presenting dusky, brown birds was an attempt to make them dislike the birds, and we lost credibility with focus-group participants. In a current study of mine investigating values for farmland preservation, respondents continually tried to infer whether the scenes represented a family farm and did not focus on other attributes of the land that framed the change to be valued. This problem was overcome with verbal descriptions that explicitly stated whether a family farm was considered, and the pictures were omitted. Thus, while pictures and other graphics can be helpful in conveying information in a contingent-valuation survey, they can also generate unwanted effects.

Other types of visual stimuli can also be effective. A number of studies have used maps (Boyle 1989; Roe, Boyle, and Teisl 1996) to present physically affected areas and graphs (Ahearn, Boyle, and Hellerstein 2003) to portray the magnitude of the change. The graphs were particularly helpful in a study where we were estimating a value that would change the populations of sixteen species of grassland birds. Focus groups revealed that some people like the changes presented with numbers and others liked the changes presented with graphs. Because of this finding, both graphs and actual numbers were used to portray the changes for the final survey instrument.

If the benchmark by which contingent valuation estimates are judged is how people would have made the decisions in a revealed preference setting, then this benchmarking also suggests providing specific information. In market analyses it is assumed that people are fully informed, or at least informed to the level they choose. Collectively, the lesson is that respondents to a contingent valuation survey need to be presented with information that clearly explains the policy change in a context that is specific to them. Refinement of this information occurs in focus groups, one-on-one interviews and, if necessary, in small-scale field pretests.

2.5.2 Explain the Method of Provision

The method of provision is the mechanism by which a policy will be implemented. In a general sense the method of provision is the "production process" for a policy change. Suppose the policy were protection of well water from contamination. One method that has been used to provide such protection is to establish protection zones around well heads, which preclude any activities that might contaminate the groundwater.

This issue has been largely ignored in the contingent valuation literature. The implicit perception of practitioners appears to be that the particular method of provision does not matter as long as it is credible to respondents. Credibility has generally been investigated, if at all, in focus groups and one-on-one pretests in the survey design phase. Some studies have not specified the method of provision to survey participants.

Choosing the method of provision is complicated because the chosen method may provide unintended clues to respondents that affect their responses to contingent valuation questions. Consider public concern over chemical residues in fruits and vegetables, genetically modified foods, "sweat-shop" production of clothing, dolphin-free tuna, etc. These attributes affect purchase decisions for market goods (Foster and Just 1989; Teisl, Roe, and Hicks 2002) and there is no reason why there should not be similar types of provision-method effects in responses to contingent valuation questions. Pesticide residues and genetic modifications carry information on the quality of the food products. While sweat-shop production may not affect the quality of the clothes in terms of their use by the purchaser, this production process may represent an undesirable externality.

In some policy studies, the method of provision has already been specified by the policy and is the appropriate first consideration for use in the contingent valuation scenario. Yet, pretesting may reveal that the actual method of provision engenders protest responses that obscure respondents true values for empir-ical estimation. In many *ex ante* analyses, the method of provision has not been decided upon and the analyst is left to make a choice when designing the contingent valuation scenario. While pretesting a method of provision in focus groups and one-on-one interviews is helpful and useful, this is a design issue that requires explicit research. If the choice of a method of provision affects welfare estimates and the most desirable method is chosen after pretesting, then welfare estimates might be overstated if a less desirable method is ultimately implemented as part of the actual policy. Careful pretesting in focus groups can identify whether the choice of a provision mechanism is likely to substantially affect value estimates.

2.5.3 Select a Payment Vehicle

This is a design area where the tradeoff between credibility and unintended clues has been clearly noted in the literature. Mitchell and Carson (1989) argue that the choice of a payment vehicle requires balancing realism against payment-vehicle rejection. That is, as realism increases, the likelihood that the payment vehicle will engender responses that protest the vehicle may also increase. For example, water-use fees are very realistic vehicles, but someone who values protecting potable groundwater may still give a valuation response of $0 to protest an increase in water rates. Income tax vehicles can run into problems due to resistance to higher taxes. On the other hand, failure to provide a realistic payment vehicle can also lead to protest responses. A local-option sales tax would not be realistic in an area that does not use this funding mechanism. Thus, respondents may reject the valuation exercises, even if they value the change, because the payment mechanism is not believable. The realism of some vehicles may lead people to give what they feel is a reasonable response, not their Hicksian surplus. For example, where there are only nominal entry fees (e.g., entrance fees at national parks), a fee vehicle may engender responses of a reasonable increase in the current nominal price rather than statements of Hicksian surplus (which could be very large for national park visits).

A variety of payment vehicles have been used in studies and a sampling of those in the recent literature is presented in Table 2. While some older studies demonstrate that payment vehicles do influence welfare estimates (Rowe, d'Arge, and Brookshire 1980; Greenley, Walsh, and Young 1981), the line of inquiry has not been prominent in recent years.[8] Practitioners appear to take for granted that payment vehicles affect welfare estimates, and any testing is undertaken in pretests, if at all.

This research indicates that the choice of a payment vehicle is another area that warrants further research. We know that payment vehicles can

Table 2. Payment Vehicles Used in Recent Studies

Payment Vehicle	Study Citation
Income taxes	Loomis and duVair (1993)
General increase in prices and taxes	Boyle et al. (1994)
Admission fee	Lunander (1998)
Utility bill	Powell, Allee, and McClintook (1994)
Recreation trip cost	Duffield, Neher, and Brown (1992)
Donations	Champ et al. (1997)

significantly influence welfare estimates. The research issue is to try to identify payment vehicles that have relatively small impacts on welfare estimates and to consider what magnitude of impact is acceptable for policy analyses. The design of the payment vehicle is another feature of the survey that can be refined through pretesting that uses focus groups. If a funding mechanism has been identified for a policy change, then this is the appropriate starting point for the design of the payment vehicle. However, if the actual payment method for the project is rejected in pretesting, alternative vehicles need to be considered.

2.5.4 Select a Decision Rule

The decision rule is the mechanism by which the results of the contingent valuation study, individual valuation responses, or summary statistics on valuation responses are used to inform the decision as to whether the item will be provided. Many studies neglect to specify the decision rule or leave respondents with a vague inference. Such an inference may be that the policy will be implemented if at least 50% of respondents answer yes to a dichotomous-choice question. The NOAA panel (NOAA 1993) recommended the framing of the contingent valuation question as a referendum with this implied decision rule. However, there is evidence that people may vote as "good citizens" rather than revealing their individual WTP when answering a question with a referendum-voting decision rule (Blamey, Bennett, and Morrison 1999). When voting as good citizens, respondents respond in a way that they think is in the best interest of the affected group, and they may overstate or understate their individual value. While voting as a good citizen may be perfectly legitimate in some policy contexts, it is not a valid response when the objective is to measure individual WTP and when the aggregation of individual WTP responses is used as input to help decide what action is in the public interest.

The choice of a decision rule is closely linked to the payment vehicle. A referendum is clearly applicable when the issue relates to the provision of a public good, such as groundwater protection, and the payment vehicle is an increase in taxes. However, a referendum is not applicable when dealing with use values, such as recreational fishing, and the payment vehicle is an increase in individual trip costs. In this later example an increase in travel costs, perhaps the cost of gasoline, may be an appropriate payment vehicle. In contrast to the groundwater example where a group decision rule is used, the decision rule in this case may simply be whether each respondent will take a fishing trip or not at the higher travel costs.

The decision rule is a key element for evaluating whether a contingent valuation survey is incentive compatible in the elicitation of valuation responses (Carson, Groves, and Machina 2000). An incentive compatible contingent valuation survey design is one that results in respondents providing truthful and accurate responses to the contingent valuation question(s).

The theoretical work on incentive compatibility is rather sterile in that it assumes that survey respondents act like rational economic agents. There is evidence in the literature to suggest that this naive but important assumption is implausible. For example, Carson, Groves, and Machina (2000) advocated the use of dichotomous choice questions framed as referendum votes, while it will be shown later in this chapter that this question format is susceptible to anchoring and yea saying bias. On the other hand, recent empirical research has shown that incentive compatible decision rules can reduce the difference between real and hypothetical value estimates (Boyle, Morrison, and Taylor 2002).

Another concern is that discussions of incentive compatibility focus on whether a contingent valuation question is or is not incentive compatible. All of the design elements discussed in this section (2.5) can affect incentive compatibility. While one component of a contingent valuation question can induce truthful (incentive compatible) responses, another component may not be incentive compatible. Therefore, incentive compatibility must be considered in terms of the multiple elements of a contingent valuation question.

Thus, incentive compatibility is one element, but not the deciding element in the design of a contingent valuation survey. The decision rule is a component of the design that is likely to evolve over the next few years as ongoing theoretical and empirical research progresses.

2.5.5 Select a Time Frame of Payment

This step tells respondents how many payments they will be required to make and how frequently the payments are required for the policy. For example, a valuation scenario might be posed where a new filtration system might be installed to remove contaminants from a public water supply. Values could be elicited as a one-time payment now or as annual payments over the lifetime of the system, say 20 years. Examples of studies using different payment vehicles are presented in Table 3.

Stevens, DeCoteau, and Willis (1997) have shown that repeated payments compared with a lump-sum payment yield statistically different estimates of WTP, and the implied discount rate is very high. Thus, the repeated-payment estimate is not substantially larger than the lump-sum-payment estimate. Their results suggest that the choice of a payment vehicle must proceed with caution.

It is important to recognize that there is often a disconnect between the time frame of payment in a contingent valuation question and the time frame over which survey respondents will enjoy the benefits of the policy change. Typical policy changes result in benefits that accrue over a number of years, and the time frame of payment, which is supposed to provide a picture of these benefits, is often not the same. Thus, survey respondents are asked to undertake personal discounting to answer valuation questions. The Stevens' research suggest that survey respondents may not do this well. This is a topic that has received scant attention in the literature.

Table 3. Payment Time Frames Used in Selected Studies

Payment Timing	Study Citation
One-time	Champ et al. (1997)
Each time participate	Stevens, DeCoteau, and Willis (1997)
Forever	Hanemann, Loomis, and Kanninen (1991)
Annual payments for x years	Shabman and Stephenson (1996)

2.5.6 Summary

The research on the provision of information suggests that this information should be as specific as possible regarding the policy change to be valued. While Mitchell and Carson (1989) argue that selection of a payment vehicle requires balancing realism/plausibility with payment vehicle rejection/failure, this need for balance applies to all components of the information package. Therefore, careful pretesting of the information scenario is required before the survey is put into the field. Even seemingly innocuous statements and terms can cause problems. For example, in a recent study of preserving agricultural lands we were quite cavalier in the use of the term "open space" in the first focus groups. We quickly found that open space has entirely different meanings to people involved in land-use policy and to the general public. Focus group participants told us that open space conveyed a sense of "outer space," "large foyers," etc., not undeveloped land. Pretesting in focus groups and/or one-on-one interviews is the best way to avoid pitfalls that can bias welfare estimates because of incorrect interpretation of information by respondents, the provision of unintended clues to respondents, and information rejection by respondents. This pretesting must be carefully conducted and is not a substitute for more research to understand the effects of each element of the information in a contingent valuation scenario.

There is no set of standards or guidelines for the design and presentation of information within contingent valuation surveys. This means that careful design must also be accompanied by conceptual and methodological research to refine what information should be presented to respondents and how it should be presented.

2.6 Design the Contingent Valuation Question

This section focuses on how the selection of a contingent valuation response format affects welfare estimates.

2.6.1 Select a Response Format

The key characteristic that differentiates the various types of contingent valuation questions is the response format; the information scenario components described above are portable from one question format to another with only slight modifications. The three primary formats are open ended, payment card, and dichotomous choice. Variants of the dichotomous-choice format are used most common.

Early contingent valuation studies used either an open-ended question (Hammack and Brown 1974) or an iterative bidding questions (Randall, Ives, and Eastman 1974). An open-ended question asks respondents "how much they would pay" for the specified change in a resource. An iterative-bidding question starts by asking respondents, "would you pay $B" for a specified policy. If respondents answer yes, then the bid is increased until they say no, and decreased until they answered yes if the initial response was no. The magnitudes of starting bids, magnitudes of bid iterations and number of iterations varied from study to study. While the open-ended format has persisted, the iterative-bidding format is no longer used because of an anchoring effect where the final bid at the end of the iterations was found to be significantly correlated with the starting bid—i.e., the higher the starting bid the higher the final bid to which people would answer yes (Boyle, Bishop, and Welsh 1985; Thayer 1981).

Open-ended questions are still used in some studies. The wording of an open-ended question used by Welsh and Poe (1998) is:

> If passage of the proposal would cost you some amount of money **every year** for the foreseeable future, what is the highest amount that you would pay annually and still vote for the program? (WRITE IN THE HIGHEST DOLLAR AMOUNT AT WHICH YOU WOULD STILL VOTE FOR THE PROGRAM) (p. 183)

Respondents are provided with a blank line where they can write in the maximum they would pay.

Dichotomous-choice questions, introduced by Bishop and Heberlein (1979), ask respondents, "would you pay $B" for the specified change in the resource, which is simply the first round in an iterative-bidding question. The bid amount is varied over different respondents. The starting-point problems with iterative-bidding questions subsequently led practitioners to quickly adopt the simpler dichotomous-response question, and the single-shot question is easier to administer than the iterative framework. Some practitioners also posed the heuristic argument that mimic the take-it-or-leave-it nature of many market purchases. Such a heuristic argument can not be made for open-ended and payment-card questions.

The dichotomous-choice question used by Welsh and Poe (1998) was worded as follows:

> Would you vote for this proposal if the proposal would cost you $______ **every year** for the foreseeable future? (CIRCLE ONE NUMBER) (p.183)

This is a dichotomous-choice question framed as a referendum, and the respondents can answer yes or no. The bid amounts are entered on the blank line.[9] Recent studies have used double-bound questions that include a second round of bids in which respondents are also asked to indicate if they would pay, a higher bid if yes was the response to the initial bid and a lower bid if no was the response to the initial bid (Hanemann, Loomis, and Kanninen 1991).

The NOAA panel (1993) recommended the use of a referendum question, and some practitioners have begun to use the terms referendum question and dichotomous-choice question as synonyms. This is incorrect and inappropriate. The referendum is the decision rule. A dichotomous-choice question can be framed as a referendum in examples such as the groundwater protection program noted in the text. A dichotomous-choice question can also be framed as agreeing to pay or not pay an entrance fee to a national park, which is not a referendum.

In the early 1980s, Mitchell and Carson (1981) introduced an anchored payment card (see also Mitchell and Carson 1993). This was a card with k bid amounts and it showed respondents how much they pay for selected public services, which in essence is very general information on substitutes.

Respondents are asked to "circle the dollar amount that is the most they would pay" for the change in the resource. Current applications of payment cards have proceeded without anchors.

Welsh and Poe (1998) also used an unanchored, payment-card question. The wording of this question was:

> If the passage of the proposal would cost you these amounts **every year** for the foreseeable future, what is the highest amount you would pay and still vote for the program? (CIRCLE THE HIGHEST AMOUNT THAT YOU WOULD STILL VOTE FOR THE PROGRAM) (p. 183)

The associated payment card was:

10¢	50¢	$1	$5	$10	$20
$30	$40	$50	$75	$100	$150
$200	MORE THAN $200				

More recently, some investigators have considered the use of a multiple-bounded question that is a hybrid of a dichotomous-choice and payment-card question; respondents are asked to indicate if they will pay each of the amounts that would occur on a payment card. The wording of the Welsh and Poe (1998) question is:

> Would you vote for this proposal if passage of the proposal could cost you these amounts **every year** for the foreseeable future? (CIRCLE ONE LETTER FOR EACH DOLLAR AMOUNT TO SHOW HOW YOU WOULD VOTE) (p. 172)

While researchers have experimented with different multiple-bounded response formats (Alberini, Boyle, and Welsh 2003), the simplest form is the following:

	YES	NO
10¢	A	B
5¢	A	B
1$	A	B
⋮	⋮	⋮
$200	A	B

This multiple-bounded question is a repeated dichotomous choice where a response is required for every bid amount on the payment card.

As you can see from the questions used in the Welsh and Poe study, the framing of the contingent-valuation question is really quite simple. While particulars of the choice of wording and sentence structure vary from study to study, this simplicity of framing is quite standard. The difficult design issue, as noted above, involves the development of the information scenario that immediately precedes the contingent valuation question. The issue here is not so much how these questions are asked, but whether the bids and response formats affect responses in ways that introduce experimentally induced biases into welfare estimates. It is easy to see why much of the methodological research has focused on the response formats—they are the penultimate component of the contingent valuation exercise and their structured nature facilitates the development of well-defined hypotheses to test statistically.

2.6.1.1 Relative Strengths and Weaknesses of Response Formats

While dichotomous-choice questions are most commonly used, each of the three main response formats has strengths and weaknesses (Table 4). Conceptual arguments by Carson, Groves, and Machina (2000) and Hoehn and Randall (1987) suggest that the "take-it-or-leave-it" nature of dichotomous-choice questions, when framed as a referendum vote, have desirable properties for incentive compatible (truthful) revelation of preferences. There is a single bid amount to which respondents respond, and respondents are not able to pick very high or very low dollar amounts to purposely misstate their values. This

Table 4. Comparison of Contingent Valuation Response Formats

Characteristics	Open Ended	Payment Card	Dichotomous Choice
Theoretically incentive compatible	No	No	Has some desirable properties
Bid design required	No	Yes	Yes
Responses/statistical efficiency	Continuous	Interval	Interval
Potential problems	Zero bids, fair share responses	Anchoring	Anchoring, yea saying, voting as good citizen

is not the case for open-ended and payment-card questions where respondents can influence the outcome of a study by the value they state or dollar amount they pick. For example, if respondents want to see a policy change occur they can state an open-ended value or pick a payment card amount that exceeds their WTP. The opportunity for such misstatements of value are not consistent with incentive comparability.

Cummings and Taylor (1998) argue that dichotomous-choice questions must be accompanied by the realism that the referendum vote will be binding, i.e., the policy change *will* be implemented if more than 50% of respondents vote yes. That is, a hypothetical vote yields more yes responses than would occur in a real vote and consequently results in overestimates of WTP.[10] Thus, incentive compatibility is not merely a function of the response format, but it includes the decision rule and other elements of the scenario that precedes the contingent valuation question.

Both payment-card and dichotomous-choice questions require selection of bids. There have been a number of contributions to the literature regarding the selection of bid amounts for dichotomous-choice questions. Alberini (1995a and b) and Kanninen (1993a and b 1995) have shown that an optimal design has a small number of bids (5 to 8), and the bid amounts should be clustered near median WTP and not placed in the tails of the distribution. These conceptual results were confirmed in an empirical investigation of bid designs by Boyle et al. (1998).

Cooper (1993) has developed computerized programs to select bid designs that require data on estimates of central tendency and dispersion from a pretest or prior study, and can use data from an open-ended question or a dichotomous-choice question.[11] While Cooper's approach is not the same as Alberini's (1995a) optimal design, the bids generated by Cooper's programs generally follow Alberini's qualitative recommendation that there be a small number of bids clustered around the mean.

Optimal bid designs are only as good as the available information on the central tendency and dispersion of the value to be estimated. When the investigator is unsure of the appropriate bid design, it is reasonable to consider a field pretest with at least 50 to 100 observations. This is more effective than the smaller sample sizes that occur in a focus group (n=8–12).

Responses to open-ended questions result in a continuous distribution of responses on the interval $[0, +\infty)$, while responses to dichotomous-choice questions indicate only whether each respondent's values lie below ($-\infty$, \$B) or above [\$B, $+\infty$) the bid threshold. Payment-card responses reveal whether the respondents' values reside within a k+1 interval where k is the number of bid amounts on the payment card. Assuming truthful revelation of responses, a person with a value of \$15 would respond in the following manner to each of the three basic contingent valuation response formats:

- The response to an open-ended question would be "\$15."
- The response to a dichotomous-choice question with a bid amount of \$10 would be "yes."
- The response to a payment-card question with bids of \$1, \$10, \$20, and \$30 would be "\$10."

Thus, for the dichotomous-choice question the empirical investigator knows only that the respondent's value lies in the interval [\$10, $+\infty$) and for the payment-card question the investigation knows that respondent's value resides in a narrower interval [\$10, \$20]. Thus, in terms of estimating central tendency, open-ended questions provide the most efficient estimates while dichotomous-choice questions provide the least efficient estimates ($se_{oe} < se_{pc} < se_{dc}$).[12] This relationship also assumes that the response formats do not have unintended experimental effects on responses that could affect the dispersion of the welfare estimates.

In addition to not being incentive compatible, open-ended questions are believed to yield an unusually high percentage of responses of \$0, which may be due, at least in part, to the lack of incentive compatibility (Carson, Graves, and Machina 2000; Hoehn and Randall 1987). It is also argued that people have difficulty coming up with a specific dollar amount for policies with which they are not familiar. It is not clear whether respondents can not respond with specific dollar amounts, or that open-ended questions do not provide enough structure to motivate respondents to put in the effort to answer with a specific dollar amount. A manifestation of this issue is that the empirical distribution of responses to open-ended questions are not smooth but tend to have spikes at \$5 increments, with a smaller number of responses at the dollar amounts within the \$5 intervals. This rounding to the nearest \$5 further attests to the difficulty respondents have giving a precise dollar value.

On a positive note, Green et al. (1998) argue for the use of open-ended questions over dichotomous-choice questions to avoid anchoring on bid amounts and to avoid the inefficiency of knowing only if respondents' values lie above or below the threshold established by the bid amounts to which they are asked to respond. However, knowing the interval where a respondents value lies, using payment cards and dichotomous choice, may be the best we can do.

While dichotomous-choice questions gained popularity to avoid the anchoring in iterative-bidding questions, dichotomous-choice questions are not free from anchoring problems (Boyle, Johnson, and McCollum 1997; Boyle et al. 1998; Green et al. 1998). That is, respondents have a propensity to say they would pay high bid amounts that are likely to exceed their true values. The converse occurs with low bid amounts. The problem seems to be most problematic with high bids, which serves to inflate value estimates (Boyle et al. 1998). Prices and quality are often perceived as being correlated in market goods, and this market intuition may lead respondents to interpret single bids as implicit signals of quality that lead to anchoring (Gabor and Granger 1966; Shapiro 1968).

Another problem has been termed "yea saying," which is the propensity of some respondents to answer yes to any bid amount presented to them (Berrens, Bohara, and Kerkvliet 1997; Blamey, Bennett, and Morrison 1999; Holmes and Kramer 1995). Here it seems that bid amounts are not acting as a quality or price cue. The manifestation of this problem has been the so-called "fat-tails" problem, with as much as 30% of a sample answering yes to any bid amount (Desvousges et al. 1993). When the inverse of the empirical cumulative distribution function (cdf) asymptotically approaches 0.30, rather than 0.00, the result is an extremely large estimate of central tendency with a large standard error. This effect further reduces the efficiency of dichotomous-choice data. Herriges and Shogren (1996) have shown that a double-bounded question only exacerbates anchoring, because responses to the second bid are influenced by the magnitude of the initial bid.

Payment cards appear to remove the proclivity to anchor and yea say, because there is not a single bid for respondents to focus on. It has been shown, however, that cutting bid amounts from the lower or upper end of the bid distribution affects welfare estimates (Rowe, Schulze, and Breffle 1996; Roach, Boyle, and Welsh 2002). Some, but not all, people who hold values below the lowest threshold of the truncated bid distribution tend to answer yes to the lowest bid amount available. The reverse logic applies if the upper end of the bid distribution is truncated. This empirical result is contrary to Alberini's (1995a) recommendation to avoid putting bids in the tails of the bid distribution for a dichotomous-choice question. Thus, when respondents see all bid amounts, not just one, it is important *not* to cluster bid amounts near the median.

Payment cards appear to avoid anchoring because there is no one bid amount to anchor on; respondents see all k bids and must circle just one of these bids. An adequate bid design might include all bids that would be used in the design of bids for a dichotomous choice question, as in Alberini (1995a), and might include low and high bid amounts that would be excluded from the dichotomous-choice bid design. As noted above, payment-card questions provide more efficient statistical information by narrowing the interval where the respondents' values reside.

Multiple-bounded questions appear to have these same properties as payment cards. Alberini, Boyle, and Welsh (2003) have shown that responses, and consequently estimated values, are affected by whether bids are presented in ascending or descending order.[13] In their study, the ascending bid distribution, which is the intuitively logical sequence, resulted in the smaller value estimate. This issue has not been investigated in the context of a pure payment card. All payment card studies have presented bids in ascending order.

It is not the intent of this section to advocate the use of payment cards and multiple-bounded response formats, but one could reasonably come to such a conclusion from the arguments presented above. While these two response formats may avoid the pitfalls associated with one-shot, dichotomous-choice questions, they are not without their own set of issues, including lack of incentive compatibility. Incentive compatibility is one consideration in choosing a contingent-valuation response format, but it is not a deciding factor. The literature does not support the choice of a single-response format (dichotomous choice) and it does not exclude the use of payment-card and multiple-bounded questions. Given the empirical issues associated with dichotomous-choice questions, it is safe to say that more research is warranted in the use of payment-card and multiple-bounded response formats.

2.6.1.2 Comparisons of Response Formats

A number of studies have compared open-ended, payment-card, and dichotomous-choice response data (e.g., Boyle et al. 1996; Kealy and Turner 1993; Ready, Buzby, and Hu 1996).[14] While most comparison studies indicate that estimates of central tendency from dichotomous-choice data exceed those from payment-card and open-ended data, this is not always the case. Boyle et al. (1996) found no significant difference between dichotomous-choice and open-ended responses for three of four comparisons. More recently, Halvorsen and Sœlensmihde (1998) demonstrated that differences are due to heteroskedasticity. Boyle et al. (1996) also allowed for differing estimates of dispersion between the dichotomous-choice and open-ended data.

The most convincing comparison study was conducted by Welsh and Poe (1998) using open-ended, payment-card, single-bounded and multiple-bounded response data. Welsh and Poe used a multiple-bounded question with polychotomous responses of "definitely" and "probably" yes, "unsure," and "definitely" and "probably" no. They found that when the definitely and probably yes responses were coded as yes, and no otherwise, the multiple-bounded cdf was indistinguishable from those for the open-ended and payment-card data. When definitely and probably yes and unsure responses were coded as yes, and no otherwise, the multiple-bounded cdf replicated that of the traditional dichotomous-choice question with a single bid amount and yes/no response options. This study supports the notion of convergent validity of open-ended and payment-card, but it clearly shows that convergent validity does not hold for these two question formats when each is compared to a dichotomous-choice question. In contrast, Bohara et al. (1998) show that open-ended questions are susceptible to eliciting "fair share" responses such that open-ended value estimates are less than those obtained from a payment-card question.

Why do dichotomous-choice questions yield welfare estimates that are greater than those of open-ended and payment-card questions? The exact reason for this difference is not known, but the high proportion of $0 responses to open-ended questions and anchoring on the bid amounts in dichotomous-choice questions both are likely to contribute to this disparity. The Welsh and Poe (1998) comparison presents another reason: respondents who are unsure will answer yes to a traditional dichotomous-choice question. This shifts the estimated cdf to the right and increases welfare estimates. The realism that Cummings and Taylor (1998) argue for a binding referendum may serve to reduce the disparity.

2.6.1.3 Summing Up on Response Formats

Having reviewed this literature, it is not absolutely clear that dichotomous-choice questions clearly represent the best approach. Dichotomous-choice questions, framed as a referendum vote, are the safe approach given the support this question format has received from the NOAA panel (NOAA 1993). (As noted above, the referendum framing of a dichotomous-choice question is not always practical, e.g., in the context of estimating recreation use values). This safe approach, regardless of whether referendum framing is used or not, is less likely to be challenged when welfare estimates are used in policy analyses and when the study results are submitted to a peer-reviewed journal for publication. There are also reasons to consider using unanchored payment-card questions and multiple-bounded questions; responses do not appear to suffer from anchoring or yea saying.

Both payment-card and dichotomous-choice questions require careful consideration in the selection of bid amounts. The advice of Alberini (1995a) to have a small number of bids clustered around the median is well worth heeding from both the perspectives of statistical efficiency and reduction of anchoring effects. Payment cards, on the other hand, while perhaps having the same number of bid amounts as a comparable dichotomous-choice question, should not cluster bid amounts near the median.

Despite the increased efficiency of double-bounded, dichotomous-choice questions, the anchoring that is introduced by the second bid amount reduces the usefulness of this question format. Multiple-bounded questions, however, may reduce anchoring but are not ready for policy analyses because of their experimental status.

2.6.2 Allowing for People to Respond with Values of $0

Some people include the issue of zero bidders under the general heading of protest responses, but there are two issues here. The first relates to people who give a response of $0 because they reject some component of the contingent valuation exercise; these are protest responses that will be dealt with in the next subsection. What is considered here are those people who truly hold values of $0 for the item being valued. It is quite possible that any policy may not be utility increasing for some segment of the sampled population, and respondents need a way to indicate such a lack of value.

With an open-ended question a respondent can simply enter a response of $0, and a payment card can include a value of $0 for respondents to circle. It is necessary to distinguish between people who have values of $0 and those who provide this answer as a protest. This can best be approached through follow-up questions to identify those who truly hold values of $0 for the policy.

The more problematic case is a dichotomous-choice question where respondents can answer no to the bid but do not have an opportunity to express a value of \$0. In these cases we know only whether respondents' values lie within the interval (-∞, \$B) or [0, \$B), depending on the modeling assumptions used to analyze the response data. We do not know if there is a spike in the probability distribution at \$0, and it is necessary to have a separate question to identify respondents whose value is \$0.[15] This \$0-value screen question has been implemented by posing it before the contingent valuation question and then administering the valuation question only to those who answer yes to this screen. Alternatively, this question could probe respondents after they have answered the contingent-valuation question by asking respondents who answer no to the bid if they would "pay anything" for the resource change.

For example, Ahearn, Boyle, and Hellerstein (2003) used the following question that preceded the contingent valuation question: "Would you vote for the proposal if passage of the proposal would increase your households' 1998 income tax?" Respondents who answered no were not asked the contingent valuation question that used the income tax as a payment vehicle. Procedures for analyzing contingent valuation data using screening questions to identifying people with values of \$0 will be discussed below.

A related issue is that policies might actually give some people disutility, which would imply that their values would be strictly negative. It appears that most studies treat people with negative values as \$0s. This treatment would tend to overstate estimates of central tendency. The issue of negative values has been investigated by Berrens et al. (1998) and Bohara, Kerkvliet, and Berrens (2003).

2.6.3 Protest and Other Types of Misleading Responses

There are at least three types of potential response categories under the heading of protests, all based on a presumption that these are respondents who do not report their true values. The first category includes people who protest some component of the contingent valuation exercise. These respondents may answer \$0, which biases the estimate of central tendency downward, or they may choose not to complete the survey, leaving the effect on central tendency dependent on how the analyst treats these respondents in the analysis of the contingent valuation data. The second category comprises people who do not understand what they are being asked to do in the survey and who answer the contingent valuation question anyway. The effect of this misunderstanding may not introduce a bias into estimates of central tendency, but it most likely will increase noise in the data that will increase the standard error of the mean. The third category is people who behave strategically in an attempt to influence survey results and ultimately the policy decision. If everyone behaving strategically acts in a similar manner, the effect will be to introduce a bias into the estimate of central tendency. However, some people may have incentives to understate values and others may have incentives to overstate the values, leaving the overall effect on estimates of central tendency indeterminate. In the second category the lack of a truthful response may be inadvertent, while this is not the case for people in the first and third categories.

People have used a variety of techniques to identify these types of responses. Some have included questions in the survey to probe respondents' understanding and motivations when answering the contingent valuation question (Ajzen, Brown, and Rosenthal 1996; Berrens et al. 1998; Blamey, Bennett, and Morrison 1999; Stevens, More, and Glass 1994). Others have trimmed the upper values if they are greater than a certain percentage (e.g., 10%) of a respondent's income (Mitchell and Carson 1989, p. 226–227). Others have used statistical routines as described in Belsey, Kuh, and Welsch (1980) to identify responses that have undue influence on estimation results (Desvousges, Smith, and Fisher 1987).

While most practitioners acknowledge that there are misleading responses in contingent valuation data, there is no established procedure with a sound conceptual basis for excluding responses. The reasons for this are varied. The problem with all of these approaches is that they are all based on *ad hoc* heuristics. Indirect questioning to identify respondents who are protesting the valuation exercise can be quite problematic. What if a person gives one response that suggests a protest and provides another response that indicates they are not protesting? Which response is correct or more meaningful? A variety of questions throughout the survey can be used to probe respondents' understanding of key parts of the valuation task. We are still left with some tough issues. What constitutes a sufficient lack of understanding such that a respondent's valuation response is excluded from statistical analyses? For example, it would not be appropriate to exclude respondents with low levels of education because they still have preferences and hold values. They may not be able to understand the valuation scenario as well as other respondents, but they make market decisions with a similar type of knowledge on a daily basis.

Questioning people after they answered the valuation question is nonsensical to a certain degree; people who are behaving strategically would be unlikely to tell you that they are doing this. People who do not understand the valuation question may also not understand the follow-up question.

Trimming the tails of the distribution by deleting outliers is also problematic, because dropping people who have high values may disenfranchise those who have the most to lose. For example, some people give up income to live in areas that are near desirable resources; for these people, welfare loses could be quite large relative to income.

Another question deals with how much of an effect misleading responses actually have on estimates of central tendency. The famous Marwell and Ames (1981) study found that "economists free ride," but others do not. This suggests that strategic behavior might be relegated to a small segment of any sample. In addition, those who behave strategically may not do so in a way that actually influences sample statistics. The first contingent valuation study that I ever conducted was undertaken with personal interviews of people while they were recreating on-site. An environmental group came through one of the survey locations; they talked among each other and encouraged each other to behave strategically by giving high value responses to influence a desirable environmental outcome. We marked all of the surveys so that we could identify these individuals in data analyses. Despite behaving strategically, none of their responses were statistical outliers and most responses were quite close to the sample mean; their strategic behavior was not effective at manipulating sample statistics. Thus, while strategic behavior may occur, it is possible that it is not sufficiently pervasive and of a magnitude to affect on welfare estimates.

At a minimum, studies should include questions to differentiate true \$0 responses from protest \$0s, and they should use statistical routines to search for data outliers. This latter suggestion requires searching through all data. Questions to detect misleading \$0 responses usually follow the contingent valuation question and typically focus on the method of provision, payment vehicle, decision rule, time frame of payment, and whether people hold a value for the resource. A cursory review of the literature indicates that investigators typically probe only the people who answer no to the bid amount in dichotomous-choice studies. This approach ignores the effects of strategic behavior and confusion on yes responses. Misleading responses and data outliers are generally excluded from welfare estimation, but it is desirable to conduct analyses with and without these data to investigate the influence on welfare estimates and the outcomes of the statistical test.

While the issue of misleading responses to contingent valuation questions deserves consideration, it is a tough conceptual and empirical issue that has not been addressed well in any study in the literature. Even if the conceptual and empirical issues are addressed, questions remain regarding how extensive the problem is and whether there is a systematic directional effect on estimates of central tendency and dispersion.

2.7 Auxiliary Questions

Step 7 calls for development of auxiliary questions, which are questions that are designed to collect data to be used in the analyses of responses to the contingent valuation questions. The most obvious candidates are income and other variables that the theoretical definition of the value suggests affect values. For payment-card, dichotomous-choice, and multiple-bounded questions, data on these variables are often used in the estimation of econometric models that are used to derive welfare estimates. That is, these formats reveal only the intervals where each respondents' value resides, and it is necessary to use econometric models to develop value estimates. For open-ended questions, value responses are generally averaged and covariates are not needed to estimate a WTP equation.

Secondary data may also be incorporated on baseline levels of environmental quality and quality with the policy. For example, in a study of lost value from contamination of an aquifer, Poe and Bishop (1999) argue that well-specific water quality data is needed in the valuation question. This is also relevant for the statistical analyses of valuation responses. GIS data may also be useful to develop variables that describe proximity of respondents' households to wells with known levels of contamination or proximity to the source of the contamination.

As noted in the preceding section, questions that probe for protest responses should be included in the survey. Questions on respondent understanding are relevant for evaluating the validity of the study even if they are not included in the analysis of valuation responses. These questions should focus on respondents' understanding of the resource being valued and the specific change in the resource they are being asked to consider. One area of particular consideration is whether valuation responses are based on the specific information presented in survey or on the respondents' own subjective perceptions of the resource and its change (Adamowicz et al. 1997). With a groundwater valuation study, for example, are valuation responses based on the actual risks of drinking contaminated water or on survey respondents' subjective perceptions of risk? If responses are based on subjective perceptions of risk, additional questions may be needed in the survey to help link values based on subjective risk perceptions to actual risks for policy analyses.

In designing a survey it is important to consider whether existing surveys (e.g., the U.S. Census) have similar questions. Using the same framing of questions as existing surveys allows for direct comparisons of the data for assessing whether sample selection has occurred, merging data sets for richer statistical analyses, and filling in missing data due to item nonresponse. Cameron, Shaw and Ragland (1999) used such an approach to address nonresponse bias in a mail survey.

2.8 Pretest and Implement the Survey

The intent of this chapter is not to delve into the details of survey administration, which is the topic of Chapter 3. Having covered the basic elements of designing a contingent valuation study, we turn to step 8, in which the survey is pretested through one-on-one interviews, focus groups, or a field trial. This pretesting allows the investigator to ensure that survey questions are understandable to respondents and are actually eliciting the information they are designed to elicit. Following the pretesting, the survey should be implemented using best practices for survey administration.

2.9 Data Analyses

In step 9, responses to open-ended questions are typically analyzed by computing the arithmetic mean:

$$(4) \qquad \bar{x} = \sum_{i}^{n} x_i / n$$

where x_i is the open-ended response for the ith respondent and n is the number of observations. The responses to an open-ended question (x_i) are individual statements of value, c or op in equations (1) and (2), respectively. Sometimes, as noted above, the data are screened for protest responses and data outliers before calculating the mean. If open-ended responses are analyzed as a function of variables that explain WTP, then a theoretical specification would be based on a definition of the value, as in equations (1) and (2), which would be solved for c or op as the dependent variable (Boyle, Johnson, and McCollum 1997). However, econometric analyses of open-ended data typically use *ad hoc* specifications that include logically intuitive variables (Stevens, More, and Glass 1994). These equations should be estimated using a tobit model because negative values are not allowed and there is a probability spike at $0.

Analysis of payment-card data proceeds by modeling the interval where respondents have revealed that their values reside. These intervals are bounded by the bid amounts each respondent circled and the next highest amount on the payment card. Following Cameron and Huppert (1989), respondents' true values (c_i) reside in the interval ($\$B_{li}$, $\$B_{ui}$), where "l" denotes the lower bid that respondent i circled and "u" denotes the next highest bid amount on the payment card. The true values can be expressed as a function of the arguments of equation (1) such that:

$$(5) \qquad log\, c_i = z_i' b + u_i$$

where u_i is the random econometric error that is assumed to be distributed normally with mean 0 and stand deviation σ, z_i is a vector of variables that are arguments from equation (1) that explain variation in responses to the valuation question, and β is a vector of coefficients to be estimated. The function $z'\beta$ is specified as the solution to equations such as (1) or (2) that define the value being estimated. The explanatory variables are those that the investigator believes should theoretically be included in the indirect utility function and affect the magnitudes of respondents' WTPs. Thus, since c_i is not actually observed and the empirical investigator does not know all variables that influence WTP, the probability that c_i falls into the interval on the payment card chosen by individual i is:

$$(6) \qquad Pr(c_i \subseteq (\$B_{li}, \$B_{ui})) = $$
$$Pr((log\$B_{li} - z_i'\beta)/\sigma < t_i < (log\$B_{ui} - z_i'\beta/\sigma))$$

and t_i is a standard normal variable. The log likelihood function for a sample size n is:

$$\text{(7)} \qquad logL(\beta,\sigma \mid \$B_{li},\$B_{ui},z_i) = \sum_{i=1}^{n} log[\phi(\$B_{ui}) - \phi(\$B_{li})].$$

where $\varphi(\cdot)$ is a normal cumulative distribution function. Using estimates of the coefficient parameter $(\hat{\beta})$, estimates of willingness to pay can be derived, where:

$$\text{(8a)} \qquad E(log\ c) = z'\hat{\beta}$$

or

$$\text{(8b)} \qquad E(c) = exp(z'\hat{\beta})exp(\hat{\sigma}^2/2)$$

The point estimates of WTP that result from equation (8) are functions of chosen levels of the z vector. Suppose a question elicited values for differing concentrations of ground water contamination; one of the arguments in z would be a variable that represents the change in the contaminant concentration. For policy purposes, the mean for a reduction in the concentration would be defined as:

$$\text{(8c)} \qquad E(c_{\Delta z_2}) = exp((z_1, z_{2SQ} - z_{2R}, \ldots, z_K)'\hat{\beta})exp(\hat{\sigma}^2 / 2)$$

where z_2 is the variable that represents the concentration of the pollutant, SQ is the status quo (current) concentration, and R denotes the reduced concentration of the pollutant as a consequence of the policy. The other arguments of z could be set at levels that represent no change, which may be their sample means for respondent-specific characteristics and the current conditions for other policy-specific characteristics that are not changing.

Again, following Cameron and Huppert (1989) and continuing with the notation established for the analysis of payment-card data, analysis of dichotomous-choice data can proceed as follows:

$$\begin{aligned} \text{(9)} \qquad Pr(YES_i) &= Pr(log\ c_i > log\ \$B_i) \\ &= Pr(u_i / \sigma > (log\ \$B_i - z'\beta) / \sigma) \\ &= 1 - \phi((log\$B_i - z_i'\beta) / \sigma) \end{aligned}$$

The log-likelihood function is:

$$(10) \qquad log\,L = \sum_{i=1}^{n} \{I_i \, log\,[1 - \phi((log\$B_i - z_i^{'}\beta)/\sigma)] + (1 - I_i)\, log\,[\phi((log\,\$B_i - z_i^{'}\beta/\sigma)]$$

where I_i equals one if a person answered yes and zero if they answered no. The computation of mean and median values proceeds as described above for payment-card data.

Hanemann (1984) proposed an alternative random-utility approach to analyzing dichotomous-choice data, but the Cameron approach is most commonly used in the literature. McConnell (1990) has shown the Hanemann and Cameron approaches to be the duals of each other.

In addition to the parametric approach described above, some investigators have considered semi-parametric and non-parametric approaches to analyzing dichotomous-choice data (Creel and Loomis 1997; Haab and McConnell 1998; Kriström 1990; Li 1996). While these approaches have not gained widespread acceptance in the literature, they deserve further consideration in the future.

The issue of a spike in the probability distribution at \$0 for people who do not value a good has received very little attention in the literature (Kriström 1997; Bohara, Kerkvliet, and Berrens 2003), but this issue needs to be considered in the design of payment-card and dichotomous-choice questions and analyses of data from these questions.

A number of approaches have been used to develop confidence intervals for welfare estimates derived from payment-card and dichotomous-choice data (Cooper 1994). While the approach used is not always clearly stated in journal articles, a quick review of recent articles suggests that the most commonly used method is the Krinsky-Robb approach that Park, Loomis, and Creel (1991) introduced to the nonmarket valuation literature. It appears that the Krinsky-Robb approach has been accepted for its computation ease (Kling 1991). However, Cooper (1994) found, in the context of data from a dichotomous-choice question, that the relative performance of different methods depends on sample size, the assumed distribution of the random variable, and the functional specification of $z_i^{'}\beta$. Cooper found that all of the methods are more likely to underpredict the standard deviation as sample sizes increase, and that bootstrap and jackknife procedures were equal to or better than the Krinsky-Robb approach. Thus, given the movement to larger samples for contingent valuation studies, Krinsky-Robb may not actually be the best choice.

2.10 Reporting Study Results

Reporting of study results, step 10, requires an understanding that the findings serve two purposes—the current policy analysis, and transfers to new applications where primary data are not available. Even methodological studies, while not designed for policy analyses, may still be used in benefit transfer. This requires clear reporting of information on crucial steps, including details on how each of the steps in Table 1 were addressed, such as:

- The study application.
- The theoretical definition of the value collected.
- The sample frame.
- Survey mode and response rates reported according to the American Association of Public Opinion Research guidelines (www.aapor.org/ethics/best.html).
- The verbatim commodity description and valuation scenario from the survey.
- The contingent valuation format used, including the question wording.
- Respondents' demographic characteristics and use of the resource.
- Methods of data analysis, including treatment of $0 values and protest responses, and equation estimates.
- Estimates of central tendency and dispersion, and the methods used to calculate these sample statistics.

This information allows one to evaluate the content validity of value estimates for the current policy application and to evaluate the transferability of value estimates to new policy applications.

3. VALIDITY AND RELIABILITY

The NOAA panel (NOAA 1993) concluded that contingent valuation provides "useful information" within the context of estimating non-use values and is "reliable by the standards that seem implicit in similar contexts" (p. 4610). What the NOAA panel refers to as reliability is what has been commonly referred to as validity in the contingent valuation literature. Tests of validity ask whether a contingent valuation study accurately measures the value it is designed to estimate. Three types of validity are commonly investigated: criterion, content, and convergent (Carmines and Zeller 1979). Criterion validity compares contingent valuation estimates to a measurement that is external to the contingent valuation study. The cash transactions in the Bishop and Heberlein (1979) study provided a criterion upon which the parallel contingent valuation estimate is validated. Content validity asks whether the elements in the design of the contingent valuation survey and data analyses are consistent with economic theory, established practice, and the valuation objective. Convergent validity investigates the consistency of contingent valuation estimates with estimates provided by another nonmarket valuation method. This is what was done when Bishop and Heberlein (1979) compared the contingent valuation estimate with those derived from a travel cost model and when Welsh and Poe (1998) compared different contingent valuation response formats. Reliability involves the extent to which a contingent valuation survey will yield the same estimates in repeated trials. The consideration of validity and reliability, as well as the efficiency of point estimates of value, collectively constitute what is referred to as a credible contingent valuation study.

Many of the critics of contingent valuation appear to hold contingent valuation up to a criterion of perfection, but this is not realistic because perfection does not exist even in actual market decisions (Randall and Hoehn 1996; Shapiro 1968; Yadov 1994). The key is to consider where contingent valuation has been shown to work reasonably well and where there may be problems.

As noted in the Introduction to this chapter, the seminal validity study by Bishop and Heberlein (1979) demonstrated that contingent valuation can provide plausible estimates of use values from the perspectives of both criterion and convergent validity; subsequent research has not reversed this conclusion. Dickie, Fisher and Gerking (1987) found a similar result when contingent valuation was applied to a market good, strawberries. In fact, many of the studies cited in this chapter, while identifying methodological issues that need to be addressed in the design process, have served to further support the credibility of contingent valuation estimates of use values, primarily through tests of convergent validity.

Criterion validity studies that compare contingent valuation estimates of non-use values with cash transactions generally find that contingent valuation estimates exceed those obtained from the cash transaction (Brown et al. 1996;

Champ et al. 1997; Duffield and Paterson 1991; Kealy, Dovidio, and Rockel 1988). Two issues arise here. First, many of these studies use a donation as a payment vehicle; real transactions are very difficult to implement using payment vehicles such as the alternatives listed in Table 3. Andreoni (1989) has argued that donations are likely to understate Hicksian surplus, which is the conceptual objective in a contingent valuation study. One may question whether it is advisable to use a payment vehicle that is not incentive compatible in tests of criterion validity. Second, Cummings and Taylor (1998), as noted earlier, argue that the valuation choice must be binding, and Kealy, Montgomery, and Dovidio (1990) argue that the payment obligation must be explicit, which are related. This suggests that innovations in the payment-vehicle and decision mechanisms may help to reduce overestimation by contingent valuation in the context of non-use and use values.

Another issue is the choice of a contingent valuation question. Brown et al. (1996) found that the difference between contingent valuation and cash estimates of central tendency was smaller when an open-ended question was used than when a dichotomous-choice question was used, a ratio of contingent valuation to actual transactions of 4.11 compared with 6.45. One reason for this difference in ratios may be that the anchoring and yea saying that arise in dichotomous-choice questions are more prevalent in contingent valuation than in the cash component of the experiment. Thus, investigating the use of unanchored, payment-card, and multiple-bounded questions may also help to reduce the disparity between contingent valuation estimates and those obtained from actual cash transactions.

The ability to detect scope appears to be purely an issue for the estimation of non-use values. Scope is present when contingent valuation estimates are responsive to the magnitude of a policy change; a policy that provides more should have a higher value than a policy that provides less of the same item. In a study of use values for white-water boating, Boyle, Welsh, and Bishop (1993) found that value estimates for white-water boating were significantly and systematically affected by river flows. In a study of non-use values, however, Boyle et al. (1994) found that values for protecting different numbers of migratory bird deaths were not responsive to the number of bird deaths prevented. In the white-water boating study, values increased with higher flows to a point and then deceased, in part because of safety concerns. For the migratory bird study, value estimates were not significantly different for preventing the deaths of 2,000, 20,000 or 200,000 birds. Carson (1997) provides a list of contingent valuation studies that have detected scope effects, but a review of this list reveals that many of the studies estimated use values or values for use and non-use combined. The fundamental question is, can contingent valuation detect scope in applications to pure non-use values? Until scope is clearly demonstrated in the estimation of non-use values, it is advisable to follow the NOAA panel's (NOAA 1993) recommendation to conduct a test for scope as part of the study design.

An area where the validity and credibility of contingent valuation is questioned is the many subjective decisions made by investigators in the design of contingent valuation surveys. This is unavoidable in survey research, and critiquing contingent valuation on this point *vis-a-vis* market transactions is naive. All data come from a survey of some sort, even market data, and there are always design features that are left to the investigator. Undesirable design effects in a survey can be minimized by following the steps listed in this chapter, pretesting carefully, and seeking peer review of survey instrument.

The consensus in the literature appears to support a conclusion that contingent valuation estimates are reliable (Carson et al. 1997; Kealy, Dovidio, and Rockel 1988; Kealy, Montgomery, and Dovidio 1990; Loomis 1989; Loomis 1990; Reiling et al. 1990; Stevens, More, and Glass 1994; Teisl et al. 1995). Thus, reliability of contingent valuation estimates is not an issue of concern.

Finally, as noted above, variance can be reduced by the information component of a contingent valuation survey by the choice of a payment-card question over a dichotomous-choice question and by bid designs; current econometric techniques provide more efficient estimates of central tendency than those used 10 or 20 years ago. Thus, features of the design of a contingent valuation survey and methods of analyzing the resultant data can reduce dispersion around estimates of central tendency. Improvements in these aspects of contingent valuation studies over the past 10 years appear to have increased the efficiency of welfare estimates.

The preponderance of evidence indicates that contingent valuation can provide estimates of Hicksian surplus that inform policy analyses and litigation, but... The "but" is that contingent valuation appears to overestimate values. Further validity research needs to focus on design features that would reduce the overestimation bias.

4. FRONTIER ISSUES

While other individuals may develop a different list of frontier research issues, I consider five issues here. The first issue is that additional research is required to identify the types of information respondents need in order to provide credible answers to contingent valuation questions. This research should also consider how the information should be presented to be understandable to respondents and to avoid providing unintended clues that bias welfare estimates. As noted earlier, most of the validity and reliability research has focused on the various response formats to the contingent valuation question, not on the information scenario that precedes the questions. While the response formats were the logical first area of inquiry, the marginal payoff for research to refine how to describe the item being valued may be relatively higher now than for research on contingent valuation response formats.

The second issue is the choice of a response format. While the use of dichotomous-choice questions should not be dismissed, there are reasons to consider using unanchored, payment-card questions or multiple-bounded questions. In fact, including more than one question format in a study to conduct tests of convergent validity is an effective way to assess the overall credibility of the study. Alternative question formats where respondents see all k bids are appealing because this format appears to reduce anchoring on a single bid amount and yea saying. Presenting k bids makes it more difficult for respondents to draw inferences about an implicit correlation between price and quality. In addition, payment cards and multiple bids narrow the intervals where respondents' values are known to reside, which increases the estimate of the welfare estimate.

The third issue is allowing respondents to express uncertainty in their responses to contingent valuation questions (Alberini, Boyle, and Welsh 2003; Berrens et al. 2002; Champ et al. 1997; Welsh and Poe 1998). The conceptual and methodological implications of allowing respondents to express uncertain responses have not been fully explored, but the use of uncertain response options have allowed investigators to better understand differences between contingent valuation and actual transaction estimates, bid designs, and contingent valuation response formats. While the use of uncertain responses may not be ready for prime-time policy analyses, these responses do appear to be useful in understanding how people answer contingent valuation questions.

The fourth issue relates to the inclusion of multiple valuation questions in a survey instrument. While the number of questions included in a contingent valuation survey varies from study to study, the recent introduction of conjoint analysis into the nonmarket valuation literature (see Chapter 7) has led researchers to consider asking respondents multiple valuation questions in a single survey instrument. The history of conjoint analysis has been to ask respondents multiple trade-off questions, sometimes to the extent that indifference curves can be estimated. This design aspect appears to be spreading to contingent valuation studies. While Boyle, Welsh, and Bishop (1993) did not find a sequence effect when respondents were asked to answer multiple contingent-valuation questions, recent conjoint research indicates that asking multiple valuation questions in a single survey instrument does create a sequencing effect (Holmes and Boyle 2002). Conjoint practitioners argue that learning occurs in responses to early questions and it is only responses to latter questions in a sequence that are reliable. This may be true, but if responses to latter questions are anchored on information in, and responses to, prior questions one must question the credibility of these subsequent responses.

The fifth area is combing contingent valuation data with revealed preference data, typically in a random utility model (Adamowicz et al. 1997; Cameron 1992; Kling 1997; McConnell, Weninger, and Strand 1999). This allows for the consistency of revealed preference data to be imposed on the contingent valuation data in the estimation, and contingent valuation can fill in gaps where revealed preferences do not exist. For example, in the case of long-term contamination there may not be use behavior for an improved level of water quality, and contingent valuation data could help fill the void. Some of these analyses use data from contingent behavior questions, not contingent valuation questions. (Contingent behavior questions ask people how they would behave under certain conditions.) Most of the design issues discussed in this chapter apply to surveys designed to collect contingent behavior data.[16]

These are just a sampling of frontier issues that are useful for students to consider as thesis topics and for practitioners to watch the peer-reviewed literature to stay abreast of recent developments.

5. CONCLUSIONS

A contingent valuation study requires very careful design and data analysis. A casual contingent valuation study can be truly horrendous, while a well-designed study can provide important insights to guide public policy. My experience is that nearly all contingent valuation studies fall short on at least one of the steps outlined above. This means that even people who have extensive experience conducting these studies still have room for improvement. While these shortcomings and the "art" component of survey design leave ample opportunities for critics to question the credibility of contingent valuation estimates, the criticism have led to better designed studies and more focused validity research. The outcome has been improved contingent valuation studies that are in high demand for policy analyses and litigation.

Even with additional research and careful pretesting, there will remain art in the design of any contingent valuation study. This subjective component will continue to be an Achilles' heel for contingent valuation. However, the art component can be enhanced with experience. More importantly, the influence that the researcher has over the design and ultimately the outcome of any contingent valuation study is **not** any different from any other line of research or empirical analysis. Simply put, contingent valuation like any other empirical method requires a high degree of skill and considerable art that must be combined with careful pretesting and validity checks.

ACKNOWLEDGMENTS

This chapter could not have been completed without the assistance of a number of people. Kim Junkins deciphered and edited the hieroglyphics in early drafts and massaged the text into the required template for publication—I'm amazed that she is still willing to work with me. Robert Paterson and Janice Huntley did extensive background research on the contingent valuation literature to help me select and refine the references cited in the text. Laura Taylor provide two rounds of constructive comments. John Loomis and Bob Berrens provided helpful comments on previous drafts. All errors and omissions are solely my responsibility. This chapter also benefits

from comments by other authors in this book at the two USDA Forest Service-sponsored workshops in Estes Park, Colorado.

NOTES

1 The use of contingent valuation to estimate use values, which was the focus of the Bishop and Heberlein (1979) study, was not called into question. Some people have subsequently maintained that use-value studies must also meet the conditions set out by the panel, and some researchers have investigated the panel's recommendations in the context of estimating use values.

2 All other arguments in the indirect utility function are assumed to be constant between the status quo (Q^0) and the condition with diminished groundwater quality (Q^1) and are suppressed for notational convenience. It is assumed here that some activity is undertaken to remediate the contamination such that the price of providing potable water increases ($p^0<p^1$), but the activity is not fully effective so there is a residual value to be estimated.

3 I will use willingness to pay in the presentation of this chapter, while fully recognizing that many policy applications require estimates of willingness to accept (Chapter 2, equation 3).

4 Having a clear theoretical definition helps to determine if estimated values are appropriate for the policy question at hand, as well as for other uses of the estimate. This is particularly important because many policy analyses involve benefit transfers (Chapter 12) and it is imperative that the value transferred from an original study be consistent with the policy question at the transfer site.

5 Personal interviews can also result in interviewer bias in responses to the contingent valuation question. While Boyle and Bishop (1988) found interviewer effects with graduate students doing the interviewing, this effect may not persist if professional interviewers are used.

6 Consider a case where the desired sample size for statistical analyses is 500. Further consider that a mail survey will be used with a sample list that has 10% bad addresses, an expected response rate of 60% and 10% item nonresponse to a valuation question. The initial sample size would need to be at least 1,029 [500 x (1/0.9) x (1/0.6) x (1/0.9)].

7 For example, if a survey dealt with techniques to avoid contaminated drinking water for households on private wells, but the available sample frame included people on public water supplies and those who don't engage in any avoidance activities, then the initial sample size from the previous note would increase to 2,058 if half of the households were expected to be on private wells and engage in avoidance activities.

8 In a recent conjoint study, Johnston, Swallow, and Weaver (1999) demonstrate that the choice of a payment mechanism, particularly one that guarantees funding for a policy, can influence welfare estimates.

9 Welsh and Poe used nine bid amounts that ranged from \$1 to \$200.

10 The reader should also see the debate on this point by Cummings et al. (1997) and Habb, Huang, and Whitehead (1999).

11 The programs can be obtained by emailing Cooper (jcooper@ers.usda.gov).

12 These denote the standard error (se) of the mean and the open ended (oe) payment card (pc) and dichotomous choice (dc) response formats.

13 Hellerstein and Boyle (2002) found, in a split-sample experiment, that less than 5% of respondents answer yes to the highest bid amount in a multiple-bounded question, while over 30% of respondents answer yes to this same bid amount in a single-bounded question. Thus, the multiple-bounded question also appears to have advantages over single-bounded and double-bounded questions in terms of statistical efficiency, and in reducing anchoring and yea saying.

14 See Table 1 in Huang and Smith (1998) for a summary of the comparison studies and the outcomes of the comparisons.

15 A similar question is required for a payment card if a bid amount of $0 is not included.

16 For example, one exception would be that a payment vehicle is not necessary.

REFERENCES

Adamowicz, W., J. Swait, P. Boxall, J. Louviere, and M. Williams. 1997. Perceptions versus Objective Measures of Environmental Quality in Combined Revealed and Stated Preference Models of Environmental Valuation. *Journal of Environmental Economics and Management* 32 (1):65-84.

Ahearn, M., D. Hellerstein, and K. J. Boyle. 2003. Designing a Contingent-Valuation Study to Estimate the Benefits of the Conservation Reserve Program on Grassland Bird Populations. In *The Contingent-Valuation Handbook*. Edited by J. Kahn, D. Bjornstad, and A. Alberini. Cheltenham, UK: Edward Elger (forthcoming).

Ajzen, I., T. C. Brown, and L. H. Rosenthal. 1996. Information Bias in Contingent Valuation: Effects of Personal Relevance, Quality of Information, and Motivational Orientation. *Journal of Environmental Economics and Management* 30(1):43-57.

Ajzen, I., L. H. Rosenthal, and T. C. Brown. 2000. Effects of Perceived Fairness on Willingness to Pay. *Journal of Applied Social Psychology* 30(12):2439-2450.

Alberini, A. 1995a. Optimal Designs for Discrete Choice Contingent Valuation Surveys: Single-Bound, Double-Bound, and Bivariate Models. *Journal of Environmental Economics and Management* 28(3):287-306.

Alberini, A. 1995b. Willingness-to-Pay Models of Discrete Choice Contingent Valuation Survey Data. *Land Economics* 71(1):83-95.

Alberini, A., K. J. Boyle, and M. P. Welsh. 2003. Analysis of Contingent Valuation Data with Multiple Bids and Response Options Allowing Respondents to Express Uncertainty. *Journal of Environmental Economics and Management* (forthcoming).

Alberini, A., M. Cropper, T. Fu, A. Krupnick, J. Liu, D. Shaw, and W. Harrington. 1997. Valuing Health Effects of Air Pollution in Developing Countries: The Case of Taiwan. *Journal of Environmental Economics and Management* 34(2):107-126.

Alberini, A., and A. Krupnick. 2000. Cost-of-Illness and Willingness-to-Pay Estimates of Improved Air Quality: Evidence from Taiwan. *Land Economics* 76(1):37-53.

Andreoni, J. 1989. Giving with Impure Altruism: Applications to Charith and Richardian Equivalence. *Journal of Political Economy* 97(6):1447-1458.

Bateman, I. J., I. H. Langford, R. K. Turner, K. G. Willis, and G. D. Garrod. 1995. Elicitation and Truncation Effects in Contingent Valuation Studies. *Ecological Economics* 12(2):161-180.

Becker, G. S. 1981. *Treatise on the Family.* Cambridge, MA: Howard University Press.

Belsey, D.A., E. Kuh, and R.E. Welsch. 1980. "Regression Diagnostics." New York: Wiley.

Bergstrom, J. C., J. R. Stoll, and A. Randall. 1990. The Impact of Information on Environmental Commodity Valuation Decisions. *American Journal of Agricultural Economics* 72(3):614-621.

Berrens, R. P., H. Jenkins-Smith, A. K. Bohara, and C. L. Silva. 2002. Further Investigation of Voluntary Contribution Contingent Valuation: Fair Share, Time of Contribution and Respondent Uncertainty. *Journal of Environmental Economics and Management* 44(1):144-168.

Berrens, R P., A. K. Bohara, H. Jenkins-Smith, C. L. Silva, P. Ganderton, and D. Brookshire. 1998. Joint Investigation of Public Support and Public Values: Case of Instream Flows in New Mexico. *Ecological Economics* 27(2):189-203.

Berrens, R., A. Bohara, and J. Kerkvliet. 1997. A Randomized Response Approach to Dichotomous Choice Contingent Valuation. *American Journal of Agricultural Economics* 79(1):252-266.

Berrens, R. P., D. Brookshire, P. Ganderton, and M. McKee. 1998. Exploring Nonmarket Values for the Social Impacts of Environmental Policy Change. *Resource and Energy Economics* 20(2):117-137.

Bishop, R. C. 1982. Option Value: An Exposition and Extension. *Land Economics* 58(1):1-15.

Bishop, R. C., and T. A. Heberlein. 1979. Measuring Values of Extra-Market Goods: Are Indirect Measures Biased? *American Journal of Agricultural Economics* 61(5):926-930.

Blamey, R. K., J. W. Bennett, and M. D. Morrison. 1999. Yea-Saying in Contingent Valuation Surveys. *Land Economics* 75(1):126-141.

Bohara, A. K., J. Kerkvliet, and R. Berrens. 2003. Addressing Negative Willingness to Pay in Dichotomous Choice Contingent Valuation: A Monte Carlo Simulation. *Environmental and Resource Economics* (forthcoming).

Bohara, A. M. McKee, R. P. Berrens, H. Jenkins-Smith, C. L. Silva, and D. S. Brookshire. 1998. Effect of Total Cost and Group-Size Information on Willingness to Pay Responses: Open Ended vs. Dichotomous Choice. *Journal of Environmental Economics and Management* 35(2):142-163.

Boyle, K. J. 1989. Commodity Specification and the Framing of Contingent-Valuation Questions. *Land Economics* 65(1):57-63.

Boyle, K. J., and R. C. Bishop. 1988. Welfare Measurements Using Contingent Valuation: A Comparison of Techniques. *American Journal of Agricultural Economics* 70(1):20-28.

Boyle, K. J., R. C. Bishop, and M. P. Welsh. 1985. Starting Point Bias in Contingent Valuation Bidding Games. *Land Economics* 61(2):188-196.

Boyle, K. J., W. H. Desvousges, F. R. Johnson, R. W. Dunford, and S. P. Hudson. 1994. An Investigation of Part-Whole Biases in Contingent-Valuation Studies. *Journal of Environmental Economics and Management* 27(1):64-83.

Boyle, K. J., F. R. Johnson, and D. W. McCollum. 1997. Anchoring and Adjustment in Single-Bounded, Contingent-Valuation Questions. *American Journal of Agricultural Economics* 79(5):1495-1500.

Boyle, K. J., F. R. Johnson, D. W. McCollum, W. H. Desvousges, R. W. Dunford, and S. P. Hudson. 1996. Valuing Public Goods: Discrete versus Continuous Contingent-Valuation Responses. *Land Economics* 72(3):381-396.

Boyle, K. J., H. F. MacDonald, H. Cheng, and D. W. McCollum. 1998. Bid Design and Yea Saying in Single-Bounded, Dichotomous-Choice Questions. *Land Economics* 74(1):49-64.

Boyle, K. J., M. M. Morrison, and L. O. Taylor. 2002. Provision Rules and the Incentive Capability of Choice Surveys. Unpublished Manuscript, Georgia State University.

Boyle, K. J., P. P. Welsh, and R. C. Bishop. 1993. The Role of Question Order and Respondent Experience in Contingent-Valuation Studies. *Journal of Environmental Economics and Management* 25(1, Part 2):580-599.

Brown, T. C., P. A. Champ, R. C. Bishop, and D. W. McCollum. 1996. Which Response Format Reveals the Truth About Donations to a Public Good. *Land Economics* 72(2):152-166.

Cameron, T. A. 1992. Combining Contingent Valuation and Travel Cost Data for the Valuation of Nonmarket Goods. *Land Economics* 68(3):302-317.

Cameron, T. A., and D. D. Huppert. 1989. OLS Versus ML Estimation of Non-market Resource Values with Payment Card Interval Data. *Journal of Environmental Economics and Management* 17(3):230-246.

Cameron, T. A., W. D. Shaw, and S. R. Ragland. 1999. Nonresponse Bias in Maine Survey Data: Salience vs. Endogenous Survey Complexity. In *Valuing Recreation and the Environment: Revealed Preference Methods in Theory and Practice*. Edited by J. A. Herriges, and K. L. Kling. Cheltenham, UK: Edward Elgar Publishing.

Carmines, E. G., and R. A. Zeller. 1979. *Reliability and Validity Assessment*. Beverly Hills, CA: Sage.

Carson, R. T. 1997. Contingent Valuation Surveys and Test for Insensitivity to Scope. In *Determining the Value of Non-marketed Goods: Economics, Psychological, and Policy Relevant Aspects of Contingent Valuation Methods*. Edited by R. J. Kopp, W. W. Pommerehne, and N. Schwarz (eds.). Boston, MA: Kluwer Academic Publishers.

Carson, R. T., T. Groves, and M. Machina. 2000. Incentive and Informational Properties of Preference Questions. Unpublished paper, Department of Economics, University of California, San Diego.

Carson, R. T., W. M. Hanemann, R. Kopp, J. A. Krosnick, R. C. Mitchell, S. Presser, P. A. Ruud, and V. K. Smith. 1998. Referendum Design and Contingent Valuation: The NOAA Panel's No-Vote Recommendation. *Review of Economics and Statistics* 80(3):484-487.

Carson, R. T., W. M. Hanemann, R. Kopp, J. A. Krosnick, R. C. Mitchell, S. Presser, P. A. Ruud, V. K. Smith, M. Conaway, and K. Martin. 1997. Temporal Reliability of Estimates from Contingent Valuation. *Land Economics* 73(2):151-163.

Champ, P. A., R. C. Bishop, T. C. Brown, and D. W. McCollum. 1997. Using Donation Mechanisms to Value Nonuse Benefits from Public Goods. *Journal of Environmental Economics and Management* 33(2):151-162.

Cooper, J. C. 1993. Optimal Bid Selection for Dichotomous Choice Contingent Valuation Surveys. *Journal of Environmental Economics and Management* 24(1):25-40.

Cooper, J. C. 1994. A Comparison of Approaches to Calculating Confidence Intervals for Benefit Measures from Dichotomous-Choice Contingent Valuation Surveys. *Land Economics* 70(1):111-122.

Creel, M., and J. Loomis. 1997. Semi-nonparametric Distribution Free Dichotomous Choice CV. *Journal of Environmental Economics and Management* 32(3):341-358

Cummings, R. G., D. S. Brookshire, and W. D. Schulze (eds.) 1986. *Valuing Environmental Goods: An Assessment of the Contingent Valuation Method*, Totowa, NJ: Rowman & Allanheld.

Cummings, R. G., G. Harrison, S. Elliot, and J. Murphy. 1997. Are Hypothetical Referenda Incentive Compatible? *Journal of Political Economy* 105(3):609-621.

Cummings, R. G., and L. O. Taylor. 1998. Does Realism Matter in Contingent Valuation Surveys? *Land Economics* 74(2):203-215.

Davis, R. K. 1963. Recreation Planning as an Economic Problem. *Natural Resources Journal* 3(3):239-249.

Davis, D. D., and C. A. Holt. 1993. *Experimental Economics*. Princeton, NJ: Princeton University Press.

Desvousges, W. H., F. R. Johnson, R. W. Dunford, S. P. Hudson, K. N. Wilson, and K. J. Boyle. 1993. Measuring Natural Resource Damages with Contingent Valuation: Tests of Validity and Reliability. In *Contributions to Economic Analysis*. Edited by J. A. Hausman. Amsterdam, The Netherlands: Elsevier Science Publishers B.B.

Desvousges, W. H., V. K. Smith, and A. Fisher. 1987. Option Price Estimates for Water Quality Improvements: A Contingent Valuation Study for the Monongahela River. *Journal of Environmental Economics and Management* 14:248-267.

Dickie, M., A. Fisher, and S. Gerking. 1987. Market Transactions and Hypothetical Demand Data: A Comparative Study. *Journal of the American Statistical Association* 82(3):69-75.

Dickie, M., and S. Gerking. 1996. Formation of Risk Beliefs, Joint Production and Willingness to Pay to Avoid Skin Cancer. *The Review of Economics and Statistics* 78(3):451-464.

Duffield, J., C. Neher, and T. Brown. 1992. Recreation Benefits of Instream Flow: Application to Montana's Big Hole and Bitterroot Rivers. *Water Resources Research* 28(9):2169-2181.

Duffield, J. W., and D. A. Paterson. 1991. Field Testing Existence Values: An Instream Flow Trust Fund for Montana Rivers. Paper presented at the annual meetings of the American Economics Association.

Edwards, S. F., and G. D. Anderson. 1987. Overlooked Biases in Contingent Valuation Surveys: Some Considerations. *Land Economics* 63(2):168-178.

Ethier, R. G., G. L. Poe, W. D. Schulze, and J. Clark. 2000. A Comparison of Hypothetical Phone and Mail Contingent Value Responses for Green-Pricing Electricity Programs. *Land Economics* 76(1):54-67.

Fischoff, B., and L. Furby. 1988. Measuring Values: A Conceptual Framework for Interpreting Transactions with Special Reference to Contingent Valuation of Visibility. *Journal of Risk and Uncertainty* 1:147-184.

Flores, N. E., and R. T. Carson. 1997. The Relationship between the Income Elasticities of Demand and Willingness to Pay. *Journal of Environmental Economics and Management* 33(3):287-295.

Foster, W., and R. E. Just. 1989. Measuring Welfare Effects of Product Contamination with Consumer Uncertainty. *Journal of Environmental Economics and Management* 17(3):266-283.

Freeman, A. M. III. 1986. On Assessing the State of the Arts of the Contingent Valuation Method of Valuing Environmental Change. In *Valuing Environmental Goods: An Assessment of the Contingent Valuation Method.* Edited by R. G. Cummings, D. S. Brookshire, and W. D. Schulze. Totowa, NJ: Rowman and Allenheld Publishers.

Freeman, A. M. III. 1993. *The Measurement of Environmental and Resource Values: Theory and Methods.* Washington, DC: Resources for the Future.

Gabor, A., and C. W. J. Granger. 1966. Price as an Indicator of Quality. *Economica* 33:43-70.

Green D., K. Jacowitz, D. Kahneman, and D. McFadden. 1998. Referendum Contingent Valuation, Anchoring and Willingness to Pay for Public Goods. *Resource and Energy Economics* 20(2):85-116.

Greenley, D. A., R. G. Walsh, and R. A. Young. 1981. Option Value: Empirical Evidence from a Case Study of Recreation and Water Quality. *Quarterly Journal of Economics* 96(4):657-672.

Habb, T. C., J. C. Huang, and J. C. Whitehead. 1999. Are Hypothetical Referenda Incentive Compatible? A Comment. *Journal of Political Economy* 107(1):186-196

Haab, T. C., and K. E. McConnell. 1998. Referendum Models and Economic Values: Theoretical, Intuitive, and Practical Bounds on Willingness to Pay. *Land Economics* 74(2):216-229.

Halvorsen, B., and K. Sœlensminde. 1998. Differences Between Willingness-to-Pay Estimates from Open-Ended and Discrete Choice Contingent Valuation Methods: The Effects of Heteroskedasticity. *Land Economics* 74(2):262-282.

Hammack, J., and G. M. Brown, Jr. 1974. *Waterfowl and Wetlands: Towards Bioeconomic Analysis*. Baltimore, MD: Johns Hopkins University Press.

Hanemann, W. M. 1984. Welfare Evaluations in Contingent Valuation Experiments with Discrete Responses. *American Journal of Agricultural Economics* 66(3):332-341.

Hanemann, W. M. 1994. Valuing the Environment Through Contingent Valuation. *The Journal of Economic Perspectives* 8(4):19-44.

Hanemann, W. M., J. Loomis, and B. Kanninen. 1991. Statistical Efficiency of Double-Bounded Dichotomous Choice Contingent Valuation. *American Journal of Agricultural Economics* 73(1):1255-1263.

Hausman, J. A. (ed.) 1993. *Contingent Valuation: A Critical Assessment*. Amsterdam: Elsevier Science Publishers BV.

Hellerstein, D., and K. Boyle. 2002. Single Bounded versus Multiple Bounded Estimators: Some Evidence on Efficiency Gains. Unpublished paper Economic Research Service, U.S. Department of Agriculture, Washington, DC.

Herriges, J. A., and J. F. Shogren. 1996. Starting Point Bias in Dichotomous Choice Valuation with Follow-up Questioning. *Journal of Environmental Economics and Management* 30:112-131.

Hoehn, J. P., and J. B. Loomis. 1993. Substitution Effects in the Valuation of Multiple Environmental Programs. *Journal of Environmental Economics and Management* 25(1, Part 1):5-75.

Hoehn, J. P., and A. Randall. 1987. A Satisfactory Benefit-Cost Indicator from Contingent Valuation. *Journal of Environmental Economics and Management* 14:226-247.

Holmes, T. P., and R. A. Kramer. 1995. An Independent Sample Test of Yea-Saying and Starting Point Bias in Dichotomous Choice Contingent Valuation. *Journal of Environmental Economics and Management* 29(1):121-132.

Holmes, T. P., and K. J. Boyle. 2002. Dynamic Bid Anchoring and Sequential Policy Interactions: Evidence from an Attribute-Based Dichotomous Choice Experiment. Unpublished paper, Southern Research Station, USDA Forest Service, Research Triangle Park, NC.

Huang, J., and V. K. Smith. 1998. "Monte Carlo Benchmarks for Discrete Response Valuation Methods." *Land Economics* 74(2):186-202.

Johnston, R. J., S. K. Swallow, and T. F. Weaver. 1999. Estimating Willingness to Pay and Resource Tradeoffs with Different Payment Mechanisms: An Evaluation of a Funding Guarantee for Watershed Management. *Journal of Environmental Economics and Management* 38(1):97-120.

Kanninen, B. J. 1993a. Design of Sequential Experiments for Contingent Valuation Studies. *Journal of Environmental Economics and Management* 25(1):S1-S11.

Kanninen, B. J. 1993b. Optimal Experimental Design for Double-Bounded Dichotomous Choice Contingent Valuation. *Land Economics* 69(2):138-146.

Kanninen, B. J. 1995. Bias in Discrete Response Contingent Valuation. *Journal of Environmental Economics and Management* 28(1):114-125.

Kealy, M. J., J. F. Dovidio, and M. L. Rockel. 1988. Accuracy in Valuation is a Matter of Degree. *Land Economics* 64(2):158-171.

Kealy, M. J., M. Montgomery, and J. F. Dovidio. 1990. Reliability and Predictive Validity of Contingent Values: Does the Nature of the Good Matter? *Journal of Environmental Economics and Management* 19(3):244-263.

Kealy, M. J., and R. W. Turner. 1993. A Test of the Equality of Closed-Ended and Open-Ended Contingent Valuation. *American Journal of Agricultural Economics* 75(2):321-331.

Kling, C. 1997. The Gains from Combining Travel Cost and Contingent Valuation Data to Value Nonmarket Goods. *Land Economics* 73(3):428-439.

Kling, C. L. 1991. Estimating the Precision of Welfare Measures. *Journal of Environmental Economics and Management* 21(3):244-259.

Kotchen, M. J., and S. D. Reiling. 1999. Do Reminders of Substitutes and Budget Constrains Influence Contingent Valuation Estimates? Another Comment. *Land Economics* 75(3):478-482.

Kriström, B. 1990. A Non-parametric Approach to the Estimation of Welfare Measures in Discrete Response Valuation Studies. *Land Economics* 66(2):135-139.

Kriström, B. 1997. Spike Models in Contingent Valuation. *American Journal of Agricultural Economics* 79(3):1013-1023.

Li, C. Z. 1996. Semiparametric Estimation of the Binary Choice Model for Contingent Valuation. *Land Economics* 72(4):462-473.

Li, C. Z., and L. Mattsson. 1995. Discrete Choice Under Preference Uncertainty: An Improved Structural Model for Contingent Valuation. *Journal of Environmental Economics and Management* 28(2):256-269.

Loomis, J. B. 1989. Test-Retest Reliability of the Contingent Valuation Method: A Comparison of General Population and Visitor Responses. *American Journal of Agricultural Economics* 71(1):76-84.

Loomis, J. B. 1990. Comparative Reliability of the Dichotomous Choice and Open-Ended Contingent Valuation Techniques. *Journal of Environmental Economics and Management* 18(1):78-85.

Loomis, J. B., and P. H. duVair. 1993. Evaluating the Effect of Alternative Risk Communication Devices on Willingness to Pay: Results from a Dichotomous Choice Contingent Valuation Experiment. *Land Economics* 69(3):287-298.

Loomis, J., and M. King. 1994. Comparison of Mail and Telephone-Mail Contingent Valuation Surveys. *Journal of Environmental Management* 41(4):309-324.

Lunander, A. 1998. Inducing Incentives to Understate and to Overstate Willingness to Pay within the Open-Ended and Dichotomous-Choice Elicitation Formats: An Experimental Study. *Journal of Environmental Economics and Management* 35(1):88-102.

Mannesto, G., and J. B. Loomis. 1991. Evaluation of Mail and In-Person Contingent Value Surveys: Results of a Study of Recreational Boaters. *Journal of Environmental Management* 32:177-190.

Marwell, G., and R. E. Ames. 1981. Economists Free Ride, Does Anyone Else? Experiments on the Provision of Public Goods. *Journal of Public Economics* 15(3):295-310.

McConnell, K. E. 1990. Models for Referendum Data: The Structure of Discrete Choice Models for Contingent Valuation. *Journal of Environmental Economics and Management* 18(1):19-34.

McConnell, R. E., Q. Weninger, and I. Strand. 1999. Joint Estimation of Contingent Valuation and Truncated Recreational Demands. In *Valuing Recreation and the Environment: Revealed Preference Methods in Theory and Practice*. Edited by J. A. Herriges and C. L. Kling. Cheltenham, UK: Edward Elger.

Mitchell, R. C., and R. T. Carson. 1981. An Experiment in Determining Willingness to Pay for National Water Quality Improvements. Unpublished report, Resources for the Future, Washington, DC.

Mitchell, R. C., and R. T. Carson. 1985. Comment on Option Value: Empirical Evidence from a Case Study of Recreation and Water Quality. *Quarterly Journal of Economics* 100:291-294.

Mitchell, R. C., and R. T. Carson. 1989. *Using Surveys to Value Public Goods: the Contingent Valuation Method*. Washington, DC: Resources for the Future.

Mitchell, R. C., and R. T. Carson. 1993. The Value of Clean Water: The Public's Willingness-to-Pay for Boatable, Fishable, and Swimmable Quality Water. *Water Resources Research* 29(7):2445-2454.

National Oceanic and Atmospheric Administration (NOAA). 1993. Natural Resource Damage Assessments under the Oil Pollution Act of 1990. *Federal Register* 58:4601-4614.

Park, T., J. Loomis, and M. Creel. 1991. Confidence Intervals for Evaluating Benefit Estimates from Dichotomous Choice Contingent Valuation Studies. *Land Economics* 67(1):64-73.

Poe, G. L., and R. C. Bishop. 1999. Valuing the Incremental Benefits of Groundwater Protection When Exposure Levels are Known. *Environmental and Resource Economics* 13(3):347-373.

Poe, G. L., and R. C. Bishop. 2001. Information and the Valuation of Nitrates in Ground Water, Portage County, Wisconsin. In *The Economic Value of Water Quality*. Edited by Bergstrom, J. C., K. J. Boyle, and G. L. Poe. Cheltenham, UK: Edward Elgar Publishing.

Powell, J. R., D. S. Allee, and C. McClintook. 1994. Groundwater Protection Benefits and Local Government Planning. Impact of Contingent Valuation Information. *American Journal of Agricultural Economics* 76(5):1068-1075.

Randall, A. 1998. Beyond the Crucial Experiment: Mapping the Performance Characteristics of Contingent Valuation. *Resource and Energy Economics* 20(2):197-206.

Randall, A., and J. P. Hoehn. 1996. Embedding in Market Demand Systems. *Journal of Environmental Economics and Management* 30(3):369-380.

Randall, A., B. Ives, and C. Eastman. 1974. Bidding Games for Evaluation of Aesthetic Environmental Improvements. *Journal of Environmental Economics and Management* 1:132-149.

Ready, R. C., J. C. Buzby, and D. Hu. 1996. Differences Between Continuous and Discrete Estimates. *Land Economics* 72(3):397-411.

Ready, R. C., J. C. Whitehead, and G. C. Bloomquist. 1995. Contingent Valuation When Respondents Are Ambivalent. *Journal of Environmental Economics and Management* 29(2):178-185.

Reiling, S. D., K. J. Boyle, M. L. Phillips, and M. W. Anderson. 1990. Temporal Reliability of Contingent Values. *Land Economics* 66(2):128-134.

Roach, B., K. J. Boyle, and M. P. Welsh. 2002. Testing Bid Design Effects in Multiple Bounded Contingent Valuation. *Land Economics* Vol 78(1):121-131.

Roe, B., K. J. Boyle, and M. F. Teisl. 1996. Deriving Estimates of Compensation Variation from Conjoint Data. *Journal of Environmental Economics and Management* 31(2):145-159.

Rowe, R. D., R. C. d'Arge, and D. S. Brookshire. 1980. An Experiment on the Economic Value of Visibility. *Journal of Environmental Economics and Management* 7:1-19.

Rowe, R. D., W. D. Schulze, and W. S. Breffle. 1996. A Test for Payment Card Biases. *Journal of Environmental Economics and Management* 31(2):178-185.

Salant, P., and D. A. Dillman. 1994. *How to Conduct Your Own Survey.* New York, NY: John Wiley & Sons, Inc.

Samples, K. C., J. A. Dixon, and M. M. Gowen. 1986. Information Disclosure and Endangered Species Valuation. *Land Economics* 62(3):306-312.

Schkade, D. A., and J. W. Payne. 1994. How People Respond to Contingent Valuation Questions: A Verbal Protocol Analysis. *Journal of Environmental Economics and Management* 26(1):88-100.

Schneemann, M. 1997. A Meta-Analysis of Response Rates to Contingent Valuation Surveys Conducted by Mail. Unpublished M.S. Thesis, University of Maine.

Schuman, H. 1996. The Sensitivity of CV Outcomes to CV Survey Methods. In *The Contingent Valuation of Environmental Resources: Methodological Issues and Research Needs.* Edited by D. Bjornstad and J. Kahn. Cheltenham, UK: Edward Elgan Publishing.

Scott, A. 1965. The Valuation of Game Resources: Some Theoretical Aspects. *Canadian Fisheries Report,* iv, Department of Fisheries of Canada, Ottawa, Ontario.

Shabman, L., and K. Stephenson. 1996. Searching for the Correct Benefit Estimate: Empirical Evidence for an Alternative Perspective. *Land Economics* 72(4):433-449.

Shapiro, B. P. 1968. The Psychology of Pricing. *Harvard Business Review* 46(7):14-25

Smith, V. K. 1993. Nonmarket Valuation of Environmental Resources: An Interpretive Appraisal. *Land Economics* 69(1):1-26.

Smith, V. L., G. L. Suchanek, and A. W. Williams. 1988. Bubbles, Crashes and Endogenous Expectations in Experimental Spot Asset Markets. *Econometrica* 56:1119-1151.

Stevens, T. H., N. E. DeCoteau, and C. E. Willis. 1997. Sensitivity of Contingent Valuation to Alternative Payment Schedules. *Land Economics* 73(1):140-148.

Stevens, T. H., T. A. More, and R. J. Glass. 1994. Interpretation and Temporal Stability of CV Bids for Wildlife Existence: A Panel Study. *Land Economics* 70(3):355-363.

Teisl, M. F., K. J. Boyle, D. W. McCollum, and S. D. Reiling. 1995. Test-Retest Reliability of Contingent Valuation with Independent Sample Pretest and Post-Test Control Groups. *American Journal of Agricultural Economics* 77(3):613-619.

Teisl, M. F., B. Roe and R. D. Hicks. 2002. Can Eco-labels Tuna Market? *Journal of Environmental Economics and Management* 43(3):339-359.

Thayer, M. 1981. Contingent Valuation Techniques for Assessing Environmental Impacts: Further Evidence. *Journal of Environmental Economics and Management* 8:27-44.

Ward, K. L., and J. W. Duffield. 1992. *Natural Resource Damages: Law and Economics*. New York: John Wiley & Sons, Inc.

Welsh, M. P., and G. L. Poe. 1998. Elicitation Effects in Contingent Valuation: Comparisons to a Multiple Bounded Discrete Choice Approach. *Journal of Environmental Economics and Management* 36(2):170-185.

Whitehead, J., and G. Bloomquist. 1995. Do Reminders of Substitutes and Budget Constraints Influence Contingent Valuation Estimates? Comment. *Land Economics* 71(4):541-543.

Whitaker, D., J. Vase, M. P. Donnell, and D. S. Dernier. 1998. Mail versus Telephone Surveys: Potential Biases in Expenditures and Willingness-to-Pay Data. *Journal of Park and Recreation Administration* 15(3):15-30.

Yadov, M. S. 1994. How Buyers Evaluate Product Bundles: A Model of Anchoring and Adjustments. *Journal of Consumer Research* 21:342-353.

Chapter 6

ATTRIBUTE-BASED METHODS

Thomas P. Holmes and Wiktor L. Adamowicz
U.S. Forest Service, Southern Research Station, and University of Alberta

1. INTRODUCTION

Stated preference methods of environmental valuation have been used by economists for decades where behavioral data have limitations. The contingent valuation method (Chapter 5) is the oldest stated preference approach, and hundreds of contingent valuation studies have been conducted. More recently, and especially over the last decade, a new class of stated preference methods has been developed, which we generically refer to as attribute-based methods (ABMs). As with contingent valuation, numerous ABM variants exist, employing, for example, different constructs for eliciting preferences. In this chapter, we describe the various ABMs currently used, explain how to construct an attribute-based experiment, and recommend methods for environmental valuation.

The objective of an ABM stated preference study is to estimate economic values for a technically divisible set of attributes of an environmental good. Responses to survey questions regarding versions of an environmental good that vary in levels of its attributes can provide resource managers and policy makers with detailed information about public preferences for multiple states of the environment. The inclusion of price as an attribute permits a multi-dimensional valuation surface to be estimated for use in benefit-cost analysis. The focus on economic welfare and willingness to pay (WTP) distinguishes the environmental economists' use of ABMs from other applications of conjoint analysis.

ABMs can offer several advantages relative to other valuation methods:

- The experimental stimuli are under the control of the researcher, as opposed to the lack of control generally afforded by observing the real market place. This includes the introduction of new attributes and attributes associated with passive uses that cannot be observed in the marketplace.
- The use of statistical design theory yields greater statistical efficiency and eliminates collinearity between explanatory variables.
- A multi-dimensional response surface is modeled that provides a richer description of preferences than can be obtained by the valuation of single "with versus without" scenarios. This richness enhances the application of ABMs to managerial decision making.
- Salient attributes of the valuation problem are clearly circumscribed. Attributes are traded off in the process of value elicitation so that a reduction in one attribute may be compensated by an increase in another attribute.

Modern applications of ABMs are based on theoretical and empirical foundations spanning several decades. To convey a sense of the richness of ABMs as developed in a variety of academic disciplines, this chapter provides an overview of the conceptual foundations that support contemporary applications of ABMs. After providing a historical perspective, we describe the basic steps for conducting an attribute-based experiment. Then we expand upon a set of selected topics in experimental design that are important to understand when developing an attribute-based experiment. Next, we review the three most popular response formats for conducting ABMs: ratings, rankings, and choice. An application of a choice experiment to a forestry issue is presented to illustrate the implementation and interpretation of a choice-based model. We then provide descriptions of models that relax the standard assumptions, which are the subject of much current research. We end with an overview of the future directions of ABM research.

2. AN INTERPRETIVE HISTORY

The origins of currently popular ABMs are found in various social science disciplines. This creative merging of disciplines has generated some confusion in terminology and classification. By presenting an interpretive overview of the literature, we hope to clarify the main concepts needed to apply ABMs to nonmarket valuation, and to distinguish between non-market valuation and other applications of ABMs.

Within economics, the conceptual foundation for ABMs finds its source in the "hedonic" method that views the demand for goods as derived from the demand for attributes. This approach can be traced to Court (1939) who used hedonic regressions to study the demand for automobiles, and Griliches (1961) who used hedonic regressions in the construction of hedonic price indices. The hedonic model was put on a firm theoretical foundation by Lancaster (1966) using household production theory. Although theories of information processing in the judgment and decision making literature in psychology (Hammond 1955; Anderson 1970) have also included discussions of how consumers evaluate characteristics of items and use these evaluations in choosing between items, Lancaster's theory of consumer demand provides the basic conceptual structure that underlies economic applications of ABMs.

At the same time that Lancaster was writing about consumer demand being driven by commodity attributes, a new measurement technique in mathematical psychology was articulated for decomposing overall judgments regarding a set of complex alternatives into the sum of weights on attributes of the alternatives (Luce and Tukey 1964). This method, known as "conjoint measurement", was rapidly embraced by marketing researchers who recognized the value of information about the relative importance of commodity attributes in the design of new products (Green and Rao 1971; Green and Wind 1975). This new marketing research method became generally known as "conjoint analysis".[1]

Many commercial applications for conjoint analysis were soon found, particularly the prediction of market share for new products (Cattin and Wittink 1982). The typical procedure would ask respondents to rate the attractiveness of a set of products and then model the preferences of each respondent (see Section 9).[2] Predicted utilities for competing products would then be computed for each individual and entered into a choice simulator to estimate the market share, computed over the sample, for each competing product (e.g., see Green et al. 1981).[3] This approach emphasized the importance of capturing individual-level preference heterogeneity as a key element in predicting market share.

Despite these advances, two primary concerns arose regarding the typical conjoint procedure. First, it was not clear that the information contained in rating data was the same as the information contained in choice data. Second, implementation of choice simulators was cumbersome and often confusing to managers who used the predictions of market share models.

A simpler, more direct approach to predicting choices in the market place was provided by discrete choice theory, particularly as formulated for economic analysis by McFadden (1974). The conceptual foundation for McFadden's analysis of economic choice lay in Thurstone's (1927) idea of random utility (discussed in greater detail in Section 6). By positing that individuals make choices that maximize their utility, and that utility is "subject to the vagaries of whim and perception", McFadden (1986, p. 278) was able to place choice theory on a strong economic foundation that included a richness of behavior not found in standard Hicks-Samuelson theory.[4] In addition, starting with Luce's choice axiom (1959), as linked to the random utility model by Marschak (1960), McFadden developed an econometric model that combined hedonic analysis of alternatives and random utility maximization.[5] This model is known as the multinomial logit (conditional logit) model.

A further advance identified by McFadden and others is the linkage between random utility models and welfare economics. The utility function in random utility models is actually a conditional indirect utility function (conditional on the choice of the alternative). Thus, including price, or more formally income minus price, as an attribute in the conditional indirect utility function allows one to assess economic welfare measures (e.g., compensating variation; see Small and Rosen, 1981). This differentiates random utility applications of ABMs in economics from other non-economic applications since economists are often interested in welfare measures and are always cognizant of the need to be consistent with theory.

The conceptual richness of random utility theory, and the practical advantages of the multinomial logit (MNL) model, were embraced by marketing researchers who promoted the use of MNL to analyze aggregate marketing data (Louviere and Woodworth 1983; Louviere and Hensher 1983; Louviere 1988a). The random utility model also found wide application in modeling transportation demand (a comprehensive treatment is provided in Ben-Akiva and Lerman 1985).[6] Initial work using the MNL model was based on the analysis of aggregate data but recent methodological developments have focused on understanding sources of individual preference heterogeneity in random utility models (see Section 12), reminiscent of the focus on individual-level modeling used in early applications of conjoint analysis.

In addition to rating and choice response formats, another variant of ABMs developed in marketing and transportation research was to ask respondents to rank bundles of attributes from most preferred to least preferred. Ranking data have the advantage of not requiring the assumption of cardinal utility that was typically relied on to analyze rating data. A popular interpretation of ranking data is based on a random utility model of choice behavior in which respondents make a sequence of choices, and the number of alternatives in the choice set decreases as ranking depth increases (this model is described in greater detail in Section 8). Thus, ranking data could be analyzed using a special form of the MNL model (Beggs, Cardell and Hausman 1981; Chapman and Staelin 1982).

The ability to decompose values of environmental programs into implicit values associated with particular attributes of those programs has made ABMs attractive to environmental economists. Although the three major response formats (rating, ranking and choice) have all been used by economists, the first application of ABMs to environmental valuation that we are aware of was Rae's (1983) work using rankings to value visibility impairments at Mesa Verde and Great Smoky Mountains National Parks. However, only a weak empirical association between rankings and visibility was observed. Stronger empirical support for ranking models was later provided by Smith and Desvousges (1986) who evaluated water quality in the Monongahela River, and Lareau and Rae (1989) who evaluated WTP for diesel odor reductions. After a hiatus of nearly a decade, a number of recent studies have been conducted using the ranking model for non-market valuation of environmental amenities (Garrod and Willis 1996 and 1998; Foster and Mourato 2000; Layton 2000; Morrison and Boyle 2001).

ABMs using rating data to value environmental quality began growing in popularity during the early 1990's. Mackenzie (1993) showed how rating data could be converted to rank and choice data. Gan and Luzar (1993) used ratings to model waterfowl hunting site decisions. Roe, Boyle, and Teisl (1996) showed how compensating variation can be estimated from rating data.

During the same period that rating models for environmental valuation were being developed, a number of studies were reported using random utility models of choice. Adamowicz, Louviere, and Williams (1994) recognized that random utility theory provides a common conceptual foundation for a class of stated preference and revealed preference models and demonstrated how revealed and stated preference data can be combined.[7] At present, choice-based ABMs are receiving the most attention.

3. STEPS IN CONDUCTING AN ATTRIBUTE-BASED EXPERIMENT

Implementation of an attribute-based experiment should follow the seven steps outlined in Table 1 (Adamowicz, Louviere, and Swait 1998; Louviere, Hensher, and Swait 2000). Each step is briefly described below.

Table 1. Steps in an Attributed-Based Experiment

1.	Characterize the decision problem
2.	Identify and describe the attributes
3.	Develop an experimental design
4.	Develop the questionnaire
5.	Collect data
6.	Estimate model
7.	Interpret results for policy analysis or decision support

The initial step is to clearly identify the economic and environmental problem. This requires thinking about two key issues: (1) the geographic and temporal scope of the change in environmental quality, and (2) the types of values that are associated with changes in environmental quality. Regarding the first key issue, several questions should be considered: Are changes in environmental quality limited to a single site or will they impact multiple sites? Are there any possible spill-overs between changes at one site and changes at other sites? Will changes be implemented instantaneously or will they take time to be fully realized?

The second key issue focuses attention on the types of values that are affected by changes in environmental quality. This requires consideration of the following questions: Who will benefit from changes in environmental quality? Will passive uses be affected? And, if the changes in environmental quality affect use value, what is the behavior that best captures this value?

Consider, for example, valuation of the benefits from improving a specific beach recreation site. The relevant values would be associated with changes in various beach attributes (such as water clarity, showers, picnic areas, and so forth). The relevant behavior to model is beach choice (from a set of beaches) and the frequency of trips. And, if changes in environmental quality also impact people who do not use the beach, passive use values need to be considered as well.

Once the decision problem is specified, the relevant attributes are identified and characterized (step 2). Continuing with the beach choice example, the researcher must identify the most important attributes of beaches that influence decisions regarding which site(s) to visit. Focus groups, or structured conversations with people who are broadly representative of the population that will be sampled, are used to identify the important attributes. For example, we might ask members of a focus group "How would you describe an excellent beach, or a poor beach?" or "What things do you consider when choosing a beach to visit?" At this stage, it is also necessary to decide how many attributes to include in the experiment as well as the particular levels that each attribute can take. How people respond to highly complex survey questions is unknown (e.g., see Mazzotta and Opaluch 1995; Swait and Adamowicz 2001a and 2001b), so it is good to keep the set of attributes as simple as possible.

Steps 1 and 2 are critically important to the successful application of ABMs but these steps are often not given the due consideration that they require. If the researcher either inappropriately frames the choice problem or omits important attributes, the entire experiment is jeopardized. We encourage practitioners to spend significant time and effort in scoping the problem, using focus groups and pre-tests, and making sure that the choice context and scenario descriptions are well developed.

After attributes and levels have been determined, in step 3 experimental design procedures are used to construct the alternatives that will be presented to the respondents. As mentioned above, the objective of an ABM stated preference study is to identify WTP for the attributes of an environmental good. WTP values are constructed from econometric estimates of the preference or taste parameters (coefficients of a utility model). The scenarios presented to respondents must provide sufficient variation over the attributes to allow the researcher to identify the taste parameters. In most cases, presenting all combinations of attributes and levels will be impossible. Thus, experimental design procedures must be used to identify subsets of the possible combinations of attributes and levels that will "best" identify the attribute preferences. Economists tend not to receive formal training in experimental design because they seldom construct controlled experiments. Therefore, we present Section 4 as a primer to this important topic.

In step 4 the questionnaire is developed. All ABMs involve surveys of some sort. As with other stated preference methods, various modes of administration are available:

- mail-out, mail-back surveys
- telephone recruitment, mail-out, mail-back surveys
- telephone recruitment, mail-out, telephone surveys
- computer-assisted surveys at centralized facilities
- intercept surveys, which may be paper-and-pencil or computer assisted
- internet-based surveys.

To date, the performance characteristics associated with various administration modes (in terms of overall response rate and item non-response) are not known. Thus, selection of the mode of administration is usually based on pragmatic concerns such as geographic specificity of the target population and budget limitations.

Various methods can be used to communicate information about the attributes of the valuation problem. In addition to verbal descriptions, graphic displays such as maps, photographs, and line drawings should be considered. As in any survey-based research, pre-testing of the questionnaire is absolutely necessary to assure that respondents clearly understand the information being communicated (see Chapter 3 for more detail on survey methods).

In step 5, the data are collected using the best survey practices (e.g., Dillman 1978). Chapter 5 outlines a number of issues in data collection for contingent valuation studies that apply as well to the implementation of ABMs.

In step 6, the taste parameters in the utility model are estimated econometrically. The choice of econometric method depends on the response format (choice, ranking, or rating) and on a variety of econometric considerations, as discussed in Sections 7 through 9.

Finally, the results are interpreted for policy analysis and decision support. ABM applications are targeted to generating welfare measures, predictions of behavior, or both. These models are used to simulate outcomes that can be used in policy analysis or as components of decision-support tools. Estimation of welfare measures is described in Section 10.

4. EXPERIMENTAL DESIGN

A strength of attribute-based experiments is that they allow the researcher to manipulate the set of explanatory variables associated with the attributes of the environmental valuation problem. However, without a proper understanding of experimental design, this asset can become a liability. The design determines both the types of effects that can be identified in the data and the interpretation of those effects. Without a proper design, an improperly specified model with biased parameter estimates and collinear variables may result.

Designed experiments are widely used in biological, physical, and behavioral sciences but are not as familiar to economists, who have historically favored the analysis of secondary data generated by social processes. A designed experiment involves the manipulation of independent variables, called *factors*, over pre-specified factor levels. Factors that represent features or characteristics of a consumer good or service are typically referred to as *attributes*.

4.1 Factorial Designs

A factorial design combines every level of each attribute with every level of all other attributes (e.g., Cochrane and Cox 1957; Snedecor and Cochrane 1974; Winer 1971). Each combination of attribute levels is called an *alternative*, *profile*, or *treatment combination*. We use these terms interchangeably (although profile is more commonly used in conjoint analysis, since combinations of attributes are often examined one-at-a-time and, thus, are not truly "alternatives"). As you can anticipate, a problem of the full factorial design is that a large number of alternatives are generated as the numbers of attributes and levels are increased.

To set the stage, consider a state parks and recreation agency that is evaluating various designs for a new campground. Suppose that agency managers need to decide whether or not to build picnic shelters, playgrounds, and showers at the new campground. Each of the three "facility" attributes takes two levels ("build" or "do not build"). Thus, there are 2^3 possible combinations of facilities. This is referred to as an L^n design, where L refer to the number of levels and n refers to the number of attributes. In this case, the full factorial design includes 8 possible combinations of attributes.

The primary advantage of a factorial design is that all "main" and "interaction" effects are independent (orthogonal) and can be identified. A "main effect" is the difference between the average (mean) response to each attribute level and the overall average (or "grand mean"). In multiple regression, main effects are represented by parameter estimates for the attribute levels and the grand mean is the intercept. An "interaction effect" occurs if the response to the level of one attribute is affected by the level of another attribute. Interaction effects are represented by the parameter estimates for the interaction (cross-product) of two (or more) variables in a multiple regression model.

Interaction effects are important in economics because they represent the concepts of *complementarity* and *substitutability*. In the example above, the average consumer may respond more favorably to a new campground with picnic shelters if playgrounds are also included in the campground description. If so, picnic shelters and playgrounds are complements. A less than full factorial design may fail to detect the interaction between picnic shelters and playgrounds and could possibly confound the interaction with one of the main effects. The reasoning behind this result follows.

4.2 Fractional Factorial Designs

Fractional designs reduce the number of profiles or alternatives included in a design to reduce the cognitive burden faced by respondents. However, information is potentially lost when fractional designs are used. To understand why fractional designs typically omit information on interactions among attributes, and the potential impact of fractional designs on parameter estimates, an understanding of *aliasing* (or *confounding*) is required.

The alias of an included effect consists of the correlated omitted effects in a fractional factorial design. Attribute codes that are completely uncorrelated are useful for identifying correlated effects. Orthogonal polynomial codes are used for this purpose. Two-level variables are represented by -1 and $+1$ rather than 0 and $+1$ used for dummy variables (for details see Louviere 1988a; Louviere, Hensher, and Swait 2000).

The concept of aliasing can be illustrated with a one-half fraction of a 2^3 design. Returning to our campground example, let A1 represent "picnic shelters", A2 represent "showers", and A3 represent "playgrounds". Table 2 shows the main and interaction effects for the full-factorial and two ½ fractions of the full design. The main effects in the full-factorial are specified using all possible combinations of attributes. Interaction effects are defined by multiplying columns (cross-products) of the orthogonal polynomial codes for each attribute.[8] Now, note that in the first one-half fraction of the full factorial (combinations 1 through 4), the vector of 2-way interactions A1A2 [+1, - 1, - 1, +1] is exactly the same as the vector of main effects for A3. Thus, A1A2 is perfectly collinear (confounded) with A3 (A3 is an alias for the A1A2 interaction). If only the first four attribute combinations in Table 2 were used for a 2^3 factorial, and if regression analysis showed that the parameter estimate on A3 was significantly different from zero, we could not be certain whether playgrounds were significant, the combination of picnic shelters and showers was significant, or both. The parameter estimate on A3 is unbiased only if the A1A2 interaction equals zero.

Table 2. Orthogonal codes showing two ½ fractions of a 2^3 factorial design

	Main effects			*2-way interactions*			*3-way interactions*
Profile	Picnic shelter A1	Showers A2	Playground A3	A1 A2	A1 A3	A2 A3	A1 A2 A3
		First ½ fraction					
1	-1	-1	+1	+1	-1	-1	+1
2	-1	+1	-1	-1	+1	-1	+1
3	+1	-1	-1	-1	-1	+1	+1
4	+1	+1	+1	+1	+1	+1	+1
		Second ½ fraction					
5	-1	-1	-1	+1	+1	+1	-1
6	-1	+1	+1	-1	-1	+1	-1
7	+1	-1	+1	-1	+1	-1	-1
8	+1	+1	-1	+1	-1	-1	-1

We also note that the 3-way interactions in the first ½ fraction in Table 2 always take the value "+1". Thus, the intercept in a regression model is perfectly collinear with the three-way interaction A1A2A3.[9]

From a practical perspective, it is generally not known *a priori* which attributes are complements or substitutes. To shed some light on the issue,first-order (2-way) interactions can be evaluated during the focus group stage of survey development. If focus group participants indicate that the attractiveness of a particular attribute depends in part on the level of other attributes, then "main effects plus selected interaction" designs can be constructed (Carmone and Green 1981). Higher order interaction effects typically have little explanatory power and probably can be ignored.

If focus groups and pre-tests reveal that interactions can be safely omitted, design catalogues are available for orthogonal fractional factorial main effects plans (e.g., Adelman 1962; Hahn and Shapiro 1966; McLean and Anderson 1984). However, if complements and substitutes are important elements of preferences for the environmental good(s) under consideration, specialized software that supports "main effects plus interactions" plans can be used.

4.3 Randomized Designs

In addition to factorial and fractional factorial designs, other strategies are available for designed experiments.[10] In principle, random sampling of attribute levels from the full factorial design space will maintain orthogonality of the design. Of course, this result is valid only for large samples. For small samples, random sampling may induce unwanted correlation among attributes. For example, consider a design for 5 attributes each with 4 levels. The full factorial for a 4^5 design yields 1024 possible profiles. Consider constructing a 1/32 fraction design by random sampling of the design space until 32 profiles are selected. Because 32 random profiles represent a small fraction of the design space, the randomly generated profiles could be clustered by the "luck of the draw". If all respondents are shown profiles from the same sample of profiles, sampling a small proportion of the design space may result in a set of correlated attributes, which reduces the efficiency of the design.

The ability of computers to randomly sample and store large amounts of data offers a second random sampling technique, the completely randomized design, wherein a randomly sampled profile is generated for each respondent in the sample. If, for example, a researcher anticipates that 1,000 people will respond to an ABM questionnaire, random assignment of attribute levels to profiles for each respondent would nearly span the entire design space in a 4^5 full factorial. Of course, it is not guaranteed that every randomly generated profile will be unique. However, if each respondent is presented with 2 or more profiles, as is usual practice in an attribute-based experiment, then it is likely that the entire design space will be sampled by randomly generated profiles.

After randomly generating profiles, it is a good idea to evaluate the experimental design by examining the correlation matrix of main effects to assure that the design is orthogonal. In addition, the correlation matrix of main effects and 2-way interaction effects should be examined for evidence of confounding.

4.4 Correlated Attributes

Attributes encountered in environmental valuation problems may be highly correlated by natural processes and, thus, they are not intrinsically separable. If two correlated attributes were treated as independent in a valuation experiment, respondents might become confused, reject the scenario, and fail to answer the question. Although some empirical studies indicate that treating correlated attributes as independent factors does not cause serious problems (Huber and McCann 1982; Moore and Holbrook 1990), it is safest to use only feasible combinations of attributes. In general, the problem of correlated attributes is best solved by selecting attributes that represent separable dimensions of the valuation problem.

4.5 Designs for Choice Experiments

When the rating response format is used in an attribute-based experiment, the efficiency of an experimental design is maintained by constructing profiles that are independent (uncorrelated) over the iterations (sequence of rating tasks) of the experiment. However, the design of a choice experiment is complicated by requiring respondents to compare two or more alternatives simultaneously. Maximum design efficiency requires selection of attribute levels that are independent of one another both within and between alternatives. This results in a L^{mn} factorial design, where *m* refers to the number of designed (non-status quo) alternatives in each choice set presented to respondents.

Let's revisit the campground design problem where the full factorial design is represented by 2^3 (L^n) possible combinations of attributes. If a rating scale response format were used, then the full factorial for this problem would be represented by 8 profiles (as in Table 2). It is possible that people could meaningfully respond to all profiles in the factorial design, and you could test hypotheses about all main and interaction effects. However, if a two-alternative choice response format were used, the full factorial would include 2^3 x 2^3 (L^n x $L^n = L^{2n}$, where $m = 2$) combinations of attribute levels and choice alternatives, or 64 (8 x 8) possible pairs of profiles (choice sets). Choice formats with more alternatives would clearly require even larger designs. Although there is no definitive number of choices that people can respond to without being fatigued, most researchers use no more than 8 or sometimes 16 choice sets. A design with 64 choice sets would be too large a design for people to respond to. A main effects design could be selected from this collective factorial if one assumed that there were no interaction effects. An example choice set is shown in Table 3.

Table 3. A campground choice set taken from a 2^6 (2^3 * 2^3) factorial

Attribute	Alternative	
	Camp site A	Camp site B
Showers	No	Yes
Playgrounds	Yes	No
Picnic shelters	No	Yes
I would choose: please check one box	□	□

In actuality, the choice experiment presented above would not be very useful to economists because no price variable (distance) is included and because choosing neither of the alternatives (opting out) is not allowed. Let us expand the example to include a 4th attribute, distance, that also has two levels (Table 4). Now the campground problem contains 4 two-level attributes in each alternative and the overall problem can be represented by a 2^4 * 2^4 or 2^8 main effects plan.

What is the smallest main effects plan (design with no interaction effects) that could be selected for this campground choice problem? This is determined by first evaluating the number of degrees of freedom needed to estimate the entire set of main effects (Louviere, Hensher, and Swait 2000). In our example with L=2 and n=4, there are 8 main effects (L x n), and each main effect has 1 (or L - 1) degree(s) of freedom. There are 8 main effects because each level of each attribute constitutes a main effect. Thus, there are a total of (L x n) x (L - 1) = 8 degrees of freedom plus one degree of freedom for the equation intercept. Next, the number of orthogonal choice sets in the fractional factorial must exceed the number of degrees of freedom. An orthogonal main effects $2^{(8-4)}$ fraction of the 2^8 factorial satisfies this requirement. Thus, the smallest orthogonal main effects plan for this example requires 16 choice sets.[11] The number of choice sets offered defines the number of iterations or replications of the choice experiment.

Table 4. A campground choice set taken from a 2^8 factorial

Attribute	Alternative		
	Camp site A	Camp site B	Stay at home: I would not choose either camp site A or B and would stay at home instead
Distance	50 miles	100 miles	
Showers	No	Yes	
Playgrounds	Yes	No	
Picnic shelters	No	Yes	
I would choose: please check one box	□	□	□

Suppose that in the campground choice problem it was decided that 16 choice sets were too many for people to reasonably consider. In this case, choice sets could be assigned to "blocks" (independent subsets of the overall design) and each respondent assigned randomly to a particular block. Two methods can be used for blocking. The first method is to list the choice sets in random order and then subdivide the list to obtain blocks of "reasonable" size. For example, 16 choice sets may be reordered and separated into 4 blocks of 4 choice sets each. With the second method, blocks are considered another attribute in the experimental design, where the number of levels is represented by the number of desired blocks. Including blocks as attributes in an orthogonal design assures that every level of all attributes will be present in every block (Adamowicz, Louviere, and Swait 1998).

When considering the number of choice alternatives to present in a choice set, one attractive option is the binary choice experiment. This is simply the ABM version of the binary (or dichotomous) choice model used in contingent valuation. A binary choice experiment can be posed as a referendum ("Would you vote for a given profile?") given a certain specification of environmental attributes. The binary choice experiment reduces the full factorial design in a choice experiment from L^{mn} to L^n because in this case m = 1. Econometric models are widely available for analyzing binary choice experiments that are not generally available for multinomial choice experiments.[12]

Another issue to consider when designing choice alternatives is whether choice alternatives should be generic or "branded". A branded alternative in the camping example would include camp site names as labels for the alternatives (such as "Jasper National Park"), and ask respondents to consider choices from the labeled alternatives, with the attributes and levels as specified. However, it is important to recognize that the brand name might be highly collinear with attributes omitted from the choice problem. If brand name is collinear with omitted attributes and is not included in the model specification, then parameter estimates are affected by omitted variable bias. Fortunately, this can be simply handled by including alternative-specific constants in the econometric specification to account for the utility associated with the alternative that is independent of the attributes (see Section 5).

In the design of a choice experiment, a common recommendation (e.g., Louviere, Hensher, and Swait, 2000) is to mimic an actual market situation by including a constant opt-out option (e.g., "I would not choose any of the available alternatives"). Continuing with the campground selection problem shown in Table 4, a typical question would ask the respondent to choose among camp site A, camp site B, or the option to stay at home. In this case, adding the choice to not go camping would allow for the possibility that when individuals are presented with camping alternatives that are not satisfactory to them, they will respond by choosing not to go camping. Without the stay-at-home option, respondents are required to choose a camping alternative, even in cases where they would never choose to go camping under the specified conditions. In practice, the opt-out alternative is not modeled in a sophisticated fashion (for a discussion of inclusion of the opt-out alternative see Louviere, Hensher, and Swait, 2000, and the discussion of alternative-specific constants in the section immediately following). If an "opt-out" alternative is not presented, the choice provides information on preferences, conditional on choosing one of the alternatives, but it does not provide information on whether the individual would choose one of the alternatives or not. We believe that choice scenarios should include opt-out options because in most real world choice situations, individuals are not in a situation of "forced choice" and they have the option to choose not to choose. A more important issue for debate may be the form of the opt-out option and the econometric modeling of this option.[13]

5. ATTRIBUTE CODING SCHEMES

Coding quantitative attributes, such as travel distance, is straightforward because the attribute level is a quantity. However, qualitative attributes pose a problem. Of course, dummy variables can be defined for $L-1$ qualitative attribute levels in the usual manner and, for ease of interpretation, the status quo level can be designated as the "omitted" level so that parameter estimates on included levels represent changes from the status quo. The problem, however, is that when dummy variables are used to code attribute levels, the attribute level associated with the omitted category is perfectly collinear with the intercept in a regression model. Thus, no information is recovered about preferences regarding the omitted level.

This limitation can be overcome by using effects codes. Because effects codes are uncorrelated with the intercept, the values of omitted levels for each attribute can be estimated (Louviere, Hensher, and Swait 2000).

Effects codes are created as follows. Begin by creating an effects-coded variable, EC_1, for the first attribute using 3 criteria:

1. If the profile contains the first level of the attribute, set $EC_1 = 1$.
2. If the profile contains the L^{th} level of the attribute, set $EC_1 = -1$.
3. If neither step 1 or 2 apply, set $EC_1 = 0$.

If an attribute has two levels, we only need to create one effects-coded variable using the preceding 3 criteria for that attribute. However, if an attribute has three levels, we continue the coding process by creating a second effects coded variable, EC_2, for that attribute using 3 additional criteria:

4. If the profile contains the second level of the attribute, set $EC_2 = 1$.
5. If the profile contains the L^{th} level of the attribute, set $EC_2 = -1$.
6. If neither step 4 or 5 apply, set $EC_2 = 0$.

If an attribute has more than three levels, we continue creating effects codes in this manner until $L-1$ effects codes are created for each L-level attribute.

Using this coding scheme, the parameter value for the omitted attribute level can be simply computed. For example, the value of the parameter for the L^{th}-level of an attribute is the sum $b_1(-1) + b_2(-1) + \ldots + b_{L-1}(-1)$, where b_n is the parameter estimate on the n^{th} level ($n \neq L$) of an effects coded variable. An example of effects coding is presented in Section 11 below.

If an attribute-based experiment contains branded alternatives or an opt- out option (e.g., stay at home option in a recreation choice experiment), it is necessary to use dummy variables known as *alternative-specific constants* (ASCs). As previously suggested, people might respond in some degree to a brand name independent of the attribute levels. ASCs identify the utility of branded alternatives not accounted for by the attributes of those alternatives. It is also essential to create an ASC for the opt out option to capture the utility associated with that option. Since the opt-out alternative usually has no attributes, an ASC is necessary to model this alternative's utility. If there are K alternatives in the choice set, then (K - 1) ASCs are included in the econometric specification.

6. RANDOM UTILITY

Models used to implement an attribute-based experiment for environmental valuation should be based on an explicit utility theory. Much of the recent work in environmental valuation is based on random utility maximization (RUM).[14] The RUM model assumes that utility is the sum of systematic (v) and random components:

$$(1) \qquad U_j = v(x_j, p_j; \beta) + \varepsilon_j$$

where U_j is the true but unobservable indirect utility associated with profile *j*, x_j is a vector of attributes associated with profile *j*, p_j is the cost of profile *j*, β is a vector of preference parameters, and ε_j is a random error term with zero mean.[15] Choice behavior is assumed to be deterministic (without error) from the perspective of the individual, but stochastic from the perspective of the researcher because the researcher does not observe everything about the individual. Thus the error term in the random utility expression reflects researcher uncertainty about the choice. It is usually assumed that utility is linear-in-parameters:

$$(2) \qquad U_j = \sum_{k=1}^{I} \beta_k x_{jk} + \beta_p p_j + \varepsilon_j$$

where β_k is the preference parameter associated with attribute *k*, x_{jk} is attribute *k* in profile *j*, and β_p is the parameter on profile cost. However, if interactions are included in the experimental design, a utility function that includes interactions (quadratic terms) can be specified as:

$$(3) \qquad U_j = \sum_{k=1}^{1} \beta_k x_{jk} + \beta_p p_j + \sum_{m=1}^{1} \sum_{k=1}^{1} \beta_{km} x_{jk} x_{jm} + \varepsilon_j$$

where β_{km} is a vector of preference parameters for interactions between attributes *k* and *m* in profile *j*, and x_{jk} and x_{jm} are attributes *k* and *m* in profile *j*. Equation (3) includes all possible substitute/complementary relations between attributes. In practice, only a subset of all possible attribute interactions would likely be specified in an attribute-based model.

By differentiating equation (2), it is seen that parameter estimates (β's) in an additively separable linear utility model represent marginal utilities: $\beta_k = \partial U/\partial x_k$. The parameter estimate on profile cost, β_p, has a special interpretation. Because an increase in profile price decreases income, β_p registers the change in utility associated with a marginal decrease in income. Thus, the negative of the parameter estimate on profile cost, $-\beta_p$, is interpreted as the marginal utility of money.

The marginal rates of substitution between any two attributes *k* and *m* is easily computed as the ratio of two parameter estimates ($MRS_{km} = \beta_k / \beta_m$). The marginal value (implicit price) of attribute *k* is computed as the ratio $\beta_k/\beta_p = (\partial U/\partial x_k)/(\partial U/\partial p_j)$. Differentiation of equation (3) shows that the marginal utility of attribute x_k in a quadratic utility function depends on the level of x_m: $\partial U/\partial x_k = \beta_k + \beta_{km} x_m$.

7. CHOICE

RUM provides the theoretical foundation for a class of empirical models based on consumer choices between competing alternatives. The choice problem asks respondents to choose the most preferred alternative from a choice set. This response format mimics actual market behavior, such as choosing a brand of cereal from among brands with different attributes. The choice format focuses the consumer's attention on the tradeoffs between attributes that are implicit in making a choice. Model estimates are based on utility differences across the alternatives contained in choice sets.

The stochastic term in the random utility function shown in equation (1) allows probabilistic statements to be made about choice behavior. The probability that a consumer will choose alternative *i* from a choice set containing competing alternatives can be expressed as:

$$(4) \qquad P(i|C) = P(U_i > U_j) = P(v_i + \varepsilon_i > v_j + \varepsilon_j), \forall\, j \in C$$

where C contains all of the alternatives in the choice set. Different probabilistic choice models can be derived depending on the specific assumptions that are made about the distribution of the random error term. If errors are assumed to be distributed according to a bivariate normal distribution, a binary probit model can be specified (Thurstone 1927) which can be generalized to the multivariate case via a multinomial probit model.[16] A type 1 extreme value (Gumbel) distribution yields the conditional or multinomial logit (MNL) model (McFadden 1974). A generalized extreme value distribution gives rise to the nested MNL model (McFadden 1981).

The standard assumption in using RUM has been that errors are independently and identically distributed (IID) following a type 1 extreme value distribution. However, the associated MNL model imposes the restrictions that: (1) preference structure is homogeneous over respondents (this assumption is relaxed in Section 12.1 below), (2) choices conform to the Independence from Irrelevant Alternatives (IIA) assumption[17] (this assumption is relaxed in Section 12.2 below), and (3) all errors have the same scale parameter.[18]

Equation (4) can be rearranged to show that, in RUM, choices are made based on utility differences across alternatives:

$$(5) \qquad P(i|C) = P(v_i - v_j > \varepsilon_j - \varepsilon_i), \forall\, j \in C.$$

Thus, any variable that remains the same across profiles, such as respondent income, drops out of the model. If errors are distributed as type 1 extreme value, the MNL model applies and the choice probability is written as:

$$(6) \qquad P(i|C) = \frac{\exp(\mu v_i)}{\sum_{j \in C} \exp(\mu v_j)}$$

where μ is the scale parameter.[19] Given an additively separable specification of utility, and assuming that $\mu = 1$, the probability of choosing profile *i* from the set C is written as:

$$(7) \qquad P(i|C) = \frac{\exp(\sum_{k=1}^{1} \beta_k x_{ik} + \beta_p p_i)}{\sum_{j \in C} \exp(\beta_k x_{jk} + \beta_p p_j)}.$$

If we let N represent the sample size and define

$$y_{in} = \begin{cases} 1 & \text{if respondent n chose profile i} \\ 0 & \text{otherwise} \end{cases}$$

then the likelihood function for the MNL model is:

$$(8) \qquad L = \prod_{n=1}^{N} \prod_{i \in C} P_n(i)^{y_{in}}.$$

Substituting equation (7) into equation (8) and taking the natural logarithm, the MNL model is estimated by finding the values of the β's that maximize the log-likelihood function:

$$(9) \qquad \ln L = \sum_{n=1}^{N} \sum_{i \in C} y_{in} (\sum_{k=1}^{1} \beta_k x_{ikn} + \beta_p p_{in} - \ln \sum_{j \in C} (\sum_{k=1}^{1} \beta_k x_{jkn} + \beta_p p_{jn}).$$

Choice based ABMs have been found to be useful for modeling use values (Adamowicz et al. 1997) and they were found to be useful in measuring passive use values as well (Adamowicz et al. 1998). Random utility models of choice have been used in a number of other studies including recreational site choice (Boxall et al. 1996) and policy/program evaluation (Viscusi et al. 1991; Opaluch et al. 1993; Hanley et al. 1998; Hanley, Wright, and Adamowicz 1998).

8. RANKING

Contingent ranking questions ask respondents to rank a set of profiles from most preferred to least preferred. This question format results in a series of responses from 1 to J for a set of J profiles. Ranking of responses ostensibly provides more information than a single choice because, in addition to the most preferred choice from a choice set, rankings provide information on preferences for all of the profiles included in the set. The standard interpretation of a ranking task views the ranking problem as a series of choices, and indeed a series of carefully constructed individual choice questions could provide the same information as a single ranking task.[20]

Analysis of ranking responses is typically conducted using random utility theory. Consider the problem of ranking profiles from the set $\{j, k, l, \ldots, J\}$. Marschak (1960) showed that a ranking problem can be modeled as a sequence of choices that, in turn, can be considered the product of independent probabilities:

$$\begin{aligned}&\Pr[j \textit{ ranked 1st, } k \textit{ ranked 2nd, } .., J \textit{ ranked last}] = \\ &P(j|j, k, l, \ldots, J) \bullet P(k|k, l, \ldots, J) \bullet \ldots \bullet P(J-1|J-1, J)\,.\end{aligned} \tag{10}$$

Equation (10) is based on the assumption that the respondent chooses the most preferred profile from the entire choice set, then the second ranked profile is chosen from the remaining choice set, and so forth. If it is assumed that the J - 1 choices in a set of J profiles are independent, and if the additively separable linear utility model adequately represents the data, then the rank-ordered logit model (Beggs, Cardell, and Hausman 1981) describing the probability of a given ranking is written as a function of the probability of the utility of alternative j being greater than that of alternative k, the utility of k being greater than that of l, and so on:

$$P(U_j > U_k > \ldots > U_J) = \prod_{j=1}^{J-1} \left(\frac{\exp[\mu(\sum_{k=1}^{l} \beta_k x_{jk} + \beta_p p_j)]}{\sum_{i=j}^{J} \exp[\mu(\sum_{k=1}^{l} \beta_k x_{ik} + \beta_p p_i)]} \right). \tag{11}$$

From a statistical perspective, the additional information provided by rankings should lead to smaller standard errors for parameter estimates or, equivalently, smaller sample sizes for a given level of precision (Hausman and Ruud 1987). However, practical experience has shown that this is not always the case. Rankings are cognitively more demanding than a single choice, and respondents may become fatigued or confused as they proceed through the sequence of choices required in ranking. Consequently, parameter estimates may lack stability and "noise" may increase for lower ranks (Chapman and Staelin 1982; Ben-Akiva, Morikawa, and Shiroishi 1992; Holmes and Boyle 2002).

9. RATING

Rating scale questions require individuals to make judgments about the magnitude of utility associated with profiles presented in an attribute-based experiment. It is implicitly assumed that judgments directly transform utility to the rating scale. Rating models can be simply estimated by regressing the vector of rating responses on the attribute levels included in each profile. Errors in rating models are often treated as additive nuisance parameters rather than having a structural interpretation as in RUM models.

Rating data are most often assumed to contain information on ordinal, not cardinal, preferences. An ordinal interpretation of rating data only requires, for example, that a response of *4* on a rating scale represents a higher intensity of preference than a *3*, but does not necessarily represent the same cardinal difference as a score of *2* relative to a score of *1*. Taking this view of rating data, it is appropriate to use an ordered probit or ordered logit model[21] although many analysts employ ordinary least squares procedures that can be implemented easily with rating data.

The use of ratings is appealing because of the simplicity of the econometric analysis and the ease with which respondents can answer rating questions. However, problems arise in using such an approach. First, ratings must be adjusted so that a common metric is used across individuals (Torgerson 1958; Mackenzie 1993; Roe, Boyle, and Teisl 1996; Holmes et al. 1998). Second, a status quo or base situation (current choice) must be evaluated using the rating to judge whether an individual would rate a new alternative higher than the status quo or base situation (which would imply choice of the alternative over the current situation) (Roe, Boyle, and Teisl 1996). Respondents may suggest that alternatives have equal ratings (ties), which presents problems when one is attempting to estimate ordinal econometric models and predict demand behavior. Decisions to include or exclude ties can effect parameter estimates (Boyle et al. 2001). Most of these challenges can be addressed using econometric procedures or by restructuring the data.

However, despite potential econometric "fixes" to rating data, we do not recommend their use for environmental valuation. Choice or ranking methods provide information on choice directly and do not require such econometric and data restructuring steps. Rating scales do not have a natural analogue in actual markets. Economic theory, in its most basic form, involves the preference of one object over another. Thus, choice methods correspond most directly with such a theory and form the most direct method of eliciting preference information. While ratings data may be used to develop welfare measures, choices or rankings are more direct.[22]

10. POLICY ANALYSIS

The goal of many ABM nonmarket valuation studies is to estimate welfare impacts so they can be used in policy analysis. Welfare measures for the random utility model underlying stated choice methods are relatively well founded and presented in the literature (Small and Rosen 1981; Hanemann 1999; Morey 1999). Since utility is random, the evaluation of welfare measures involves examination of the systematic components of utility as well as the stochastic elements. ABMs provide quantitative measures of tradeoffs between attributes (including price). Thus, they can be used to examine, after an attribute change, how much money would be required to make a person as well off as they were before the change. The fact that ABMs provide estimates of the indirect utility function allows one to calculate these welfare measures for gains, losses, or any combination of change in attributes (assuming that the specification is accurate.)

As defined in equation (1), utility is characterized by systematic (v) and stochastic (ε) components. The maximal elements of the utilities over the set of alternatives is defined as max (U_j) = max ($v_j + \varepsilon_j$) $\forall j$. Following Morey (1999), we can express the expected value of the maximum as:

$$(12) \qquad E(U) = \int_{\varepsilon_1=-\infty}^{\infty} \dots \int_{\varepsilon_J=-\infty}^{\infty} \max(v_1 + \varepsilon_1, \dots, v_J + \varepsilon_J) f(\varepsilon_1, \dots, \varepsilon_J) d\varepsilon_1 \dots d\varepsilon_J$$

where equation (12) integrates utilities over all stochastic terms (densities defined by f(.) associated with each alternative).[23] If a type 1 extreme value distribution is assumed for the stochastic elements, the expected value of the maximum can be specified as

$$(13) \qquad E(U) = \ln\left(\sum_{j=1}^{J} \exp(V_j)\right) + D$$

where expression (13) is the "log sum" plus a term known as Euler's constant (D).[24] This expression forms the basis for welfare measurement in the multi-alternative case.

In a simple situation where the marginal utility of money is constant (and expressed as λ_Y), an expression for compensating variation can be formulated as follows. Compensating variation (CV) is the amount of money that must be given to or taken away from a person to make him or her as well off after a change as they were before a change. Thus, let "before the change" be expressed as the expected value of the maximum utility in the base case: $E(U)^0 = E(U(Y^0, P^0, X^0))$, where Y is income, P is price, X is the set of attributes, and E is the expectation operator. Let "after the change" be represented by $E(U)^1 = E(U(Y^0, P^1, X^1))$, where for generality either price, or attributes, or both can change. Compensating variation is calculated by solving the expression $E(U(Y^0, P^0, X^0)) = E(U(Y^0 - CV, P^1, X^1))$ for the value CV. Using the expression for the expected value of the maximum, and assuming zero income effects, CV becomes

$$(14) \qquad CV = (1/\lambda_Y)\left[\ln\left(\sum_{j=1}^{J} \exp(V_j^1)\right) - \ln\left(\sum_{j=1}^{J} \exp(V_j^0)\right)\right]$$

which is simply the difference in the two expected values of maximum utility (change in utility), divided by the marginal utility of money.[25] The marginal utility of money parameter, in this simple case, is just the parameter on the price variable, with the sign change to reflect increasing utility with income increases. Note that the welfare measure described in equation (14) is for single choice occasions (e.g., one camping trip) or is per choice occasion.[26] That is, the random utility model is implicitly specified for a given time period (such as a week or day) and the welfare measure applies to this time period. A model of camping destination choice, for example, may be applied to the choice of site each week, where in many weeks the choice will be to "stay at home" or not go camping.

The expression in equation (14) is relevant to cases with multiple alternatives as in the case of recreation sites, alternative products, and so forth. However, choice experiments are also used to compare "states of the world", or a base case described by a single alternative against an altered case described by a single alternative. For example, two new states of the world described by attributes could be presented along with the current situation. These new states of the world could involve improved attribute levels and a positive payment amount, reduced attribute levels and a negative payment amount (refund), or some combination of these conditions. Expression (14) then reduces to

$$(15) \qquad CV = (1/\lambda_Y)[V^1 - V^0]$$

where V^1 and V^0 are the expressions of utility for the base and altered cases. Finally, if V^1 and V^0 are linear in attributes, and the goal is to evaluate a change in a single attribute, equation (15) reduces to the ratio of the attribute coefficient and the marginal utility of money.[27] The resulting values are known as "implicit prices" or marginal willingness to pay.

Note that in most simple ABMs, income is not included in the utility function (since income drops out of the utility difference expression). This means that income effects, to the extent that they exist, are ignored. In this case, the utility function specified can be used to measure compensating or equivalent variation, and they will be identical. More complex forms of random utility models do include income effects.

Readers will notice that we have not discussed willingness to accept (WTA) nor have we discussed the difference between willingness to pay (WTP) and WTA as this relates to ABMs. That is because ABMs result in specification of indirect utility functions and the specification of the indirect utility function will dictate the difference between WTA and WTP, if any. If a simple linear utility function is specified, income effects are assumed to be zero. Furthermore, these simple utility functions seldom contain any reference point measures or endowment effects. Therefore, in this case it is assumed that there is no difference between WTP and WTA. Some researchers have examined indirect utility functions with income effects and with reference points (e.g. Adamowicz et al. 1998) and we expect this to become more common practice in the future.

11. APPLICATION

The foregoing concepts lay out the basic methods used in designing, analyzing, and interpreting an attribute-based experiment. To clarify the concepts and fill in some of the details that we have omitted so far, we present an empirical example based on data collected in a mail survey regarding Maine residents' preferences for alternative timber harvesting practices. For purposes of this chapter, the following example is modified from data descriptions and analyses presented elsewhere (Boyle et al. 2001; Holmes and Boyle 2002).[28]

Timber harvesting practices in Maine have received a great deal of public attention. In 1989, the Maine legislature passed a Forest Practices Act that provides rules regarding timber harvesting standards. However, public concern about some provisions of the Act, particularly regarding clearcutting (removing all trees from a harvest area), led to a number of initiatives to modify the Act. Although none of the initiatives have succeeded to date, it is clear that many among the voting public are dissatisfied with status quo forest practices and are seeking alternatives that reduce timber harvesting impacts on the goods and services provided by Maine forests.

After discussions with forestry experts, stakeholders, and focus groups, a policy proposal and a set of timber harvesting attributes were selected for the experiment. The policy proposal was for the state of Maine to purchase a large tract of forest land from the timber industry and to manage a set of forestry attributes on the tract. Table 5 presents a set of forestry attributes and the attribute levels used for our example. In addition, thirteen different tax prices, ranging from $1 to $1,600 for a one-time tax payment, were included in the experimental design. Alternatives were created by randomly selecting attribute levels for each individual in the sample. The data here consist of N = 156 observations in which the choice set included 4 alternative management plans plus the option to select the status quo (no public purchase of private forest land). An example choice question is shown in Table 6.

Table 5. Forestry attributes and levels for Maine timber harvesting example

Attribute	Level
Live trees left after harvest	No trees (clearcut)
	153 trees/acre (heavy selection harvest)
	459 trees/acre (light selection harvest)
Dead trees left after harvest	Remove all
	5 trees/acre
	10 trees/acre
Percent of forest set aside from harvesting	20% set aside from harvesting
	50% set aside from harvesting
	80% set aside from harvesting

For each L-level, non-price attribute, L - 1 variables were constructed to specify the qualitative timber harvesting attributes. Table 7 presents the effects codes associated with each attribute level. Using effects codes, a base level is chosen. If dummy variables were used, this would be the level assigned zeros throughout the data set. With 3 levels only 2 unique parameters can be estimated. When using effects codes, the two unique parameters are summed minus one (the omitted level would be coded as 0 if we used dummy codes)

Table 6. A timber harvesting plan choice set

Attribute	Alternative				
	Plan A	Plan B	Plan C	Plan D	I would not vote for any of the plans
Live trees	No trees	459	No trees	153	
Dead trees	Remove all	Remove all	5	10	
Percent set aside	80%	20%	50%	20%	
Tax	$40	$200	$10	$80	
I would vote for: (please choose one box)	☐	☐	☐	☐	☐

and multiplied by −1 to create the parameter value for the base level. When using dummy variables, the parameter value for the base level is assumed to be zero. Note how the base (omitted) levels for the attributes are coded using −1 in the effects codes. In addition, an ASC was included in the specification to estimate the change in utility associated with choosing the status quo alternative.

The results from maximum likelihood estimation of a MNL model are shown in Table 8. All of the preference weight parameters of the indirect utility function have t-statistics greater than 1.64 (90 percent confidence level) except "selection harvest-light" and "5 dead trees/acre". Preference weights for the base (omitted) attribute levels were computed as the sum of −1 times the preference weights on the included levels for each attribute. Marginal WTP values (the WTP for a marginal change in the attribute) were then computed by dividing the preference weights by the marginal utility of money (−1 times the preference weight for the tax attribute). As can be seen, clearcutting, leaving no dead trees after harvest and setting aside 80 percent of the forest from

Table 7. Effects codes for the timber harvesting attributes

Attribute	Effects code 1	Effects code 2
Live trees left after harvest		
Clear cut	1	0
Selection harvest - heavy (base level)	−1	−1
Selection harvest - light	0	1
Dead trees left after harvest		
Remove all (base level)	−1	−1
5 dead trees/acre	1	0
10 dead trees/acre	0	1
Percent forest set aside from harvest		
20% set aside (base level)	−1	−1
50% set aside	1	0
80% set aside	0	1

harvest have relatively large negative impacts on indirect utility. Conversely, a heavy selection harvest, leaving 10 dead trees per acre and setting aside 50 percent of the forest from harvest have relatively large positive impacts on indirect utility.

An interesting interpretation can be made for the preference weight on the ASC. Recall that, as defined here, the ASC represents the utility of choosing the status quo alternative, everything else held constant. The negative sign indicates that choosing the status quo decreases indirect utility (choosing alternatives to the status quo increase indirect utility). This result is consistent with the degree of political activity in Maine seeking alternatives to current timber harvesting practices. The respondents would prefer to see a change from the status quo even if all attributes were held constant. This indicates a significant desire to have some change in the policy environment. If a positive sign on the ASC were found, it would indicate a positive preference for the status quo (everything else held constant) and would be consistent with the more common status quo "bias" found in the literature, in which individuals attach some positive utility to the status quo situation.

The pseudo R^2 for the overall model, computed as 1 minus the ratio of log-likelihood at convergence and log-likelihood at zero, is 0.14. The ASC accounts for 0.03 of the pseudo R^2 value. The attributes included in the example clearly had a dominant role in explaining choice among the timber harvesting alternatives.

12. RELAXING THE ASSUMPTIONS OF THE CONDITIONAL LOGIT MODEL

Up to this point, two assumptions have been made to simplify the econometric analysis of the conditional logit model. First, we assumed that everyone in the population has the same preference structure. This assumption restricts the β's to be the same for all members of the population. Second, we assumed that the ratio of probabilities between any two alternatives was unaffected by other alternatives in the choice set. This property (IIA, section 7) results in limited substitution possibilities.

Table 8. Parameter estimates for the timber harvesting choice experiment example

Attribute	Preference weight	t-statistic	Marginal WTP
ASC (for status quo alternative)	-1.15	-4.36	—
Clearcut	-0.42	-2.79	-221.05
Selection harvest - light	0.08	0.57	42.11
Selection harvest - heavy (base level)	0.34	—	178.95
5 dead trees/acre	0.09	0.66	47.37
10 dead trees/acre	0.32	2.42	173.68
No dead trees (base level)	-0.41	—	-215.79
Set aside 80 percent	-0.29	-1.88	-152.63
Set aside 50 percent	0.34	2.50	178.95
Set aside 20 percent (base level)	-0.05	—	-26.32
Tax	-0.0019	-4.361	—
Log-likelihood at zero	-251.07		
Log-likelihood at constants	-243.46		
Log-likelihood at convergence	-215.39		
Likelihood ratio (pseudo-R^2)	0.14		

12.1 Relaxing the Assumption of Common Preferences: Heterogeneity

The basic conditional logit model described in equation (6) implicitly assumes that preferences are identical for all respondents (the parameters in the conditional indirect utility function are constant). This simplifying assumption can be altered by three modifications: (1) including interaction effects, (2) estimating a latent class/finite mixture model, and (3) using a random parameter/mixed logit approach.

12.1.1 Interaction Effects

Individual (respondent) specific variables (age, wealth, etc.) cannot be examined directly in a conditional logit model because these variables do not vary across alternatives. Thus, individual specific variables drop out of the utility difference. However, individual specific variables can interact with alternative-specific attributes to provide some identification of attribute parameter differences in response to changes in individual factors. For example, interacting age with the price attribute would generate information on the marginal utility of money (price) as a function of age. This is a simple approach that provides insight into heterogeneity of consumers, but it assumes the researcher already knows the elements that lead to heterogeneity (those items included as interaction effects) and results in many parameters and potential collinearity problems.

12.1.2 Latent Class/Finite Mixture Approach

A better, although somewhat more complicated, approach is to use a latent class/finite mixture model. Suppose S segments exist in the population, each with different preference structures, and that individual n belongs to segment s (s = 1,...,S). The conditional indirect utility function presented above can now be expressed as $U_{in|s} = \beta_s X_{in} + \varepsilon_{in|s}$. The preference parameters (β) vary by segment. The probability of choosing alternative i depends on the segment that one belongs to and can be expressed as:

$$(16) \qquad P_{n|s}(i) = \frac{\exp(\beta_s X_i)}{\sum_{k \in C} \exp(\beta_s X_k)}$$

where β_s are segment-specific utility parameters (and scale is fixed at 1).

Now let there be a process describing the probability of being in a particular segment, as a function of demographic (and other) information. Following Boxall and Adamowicz (1999), Swait (1994), and Gupta and Chintagunta (1993), that process can be specified as a separate logit model to identify segment membership as:

$$(17) \qquad P_{ns} = \frac{\exp(\lambda_s Z_n)}{\sum_{s=1}^{S} \exp(\lambda_s Z_n)}.$$

where Z is a set of individual characteristics and λ is a vector of parameters.

Let $P_{ns}(i)$ be the joint probability that individual n belongs to segment s and chooses alternative i. This is also the product of the probabilities defined in equations (16) and (17): $P_{ns}(i) = P_{ns} P_{n|s}(i)$. The probability that individual n chooses i becomes the key component in the finite mixture or latent class approach:

$$(18) \qquad P_n(i) = \sum_{s=1}^{S} \pi_{ns} \, \pi_{n|s}(i).$$

Note that this approach provides information on factors that affect or result in preference differences. That is, the parameters in the segment membership function indicate how the probability of being in a specific segment is affected by age, wealth, or other elements included in the segment membership function. Further detail on this approach to heterogeneity can be found in Swait (1994), Boxall and Adamowicz (1999) or Shonkwiler and Shaw (1997).

12.1.3 Random Parameter/Mixed Logit Approach

Another approach to identifying preference heterogeneity is based on the assumption that parameters are randomly distributed in the population. Then, the heterogeneity in the sample can be captured by estimating the mean and variance of the random parameter distribution. This approach is referred to as random parameter logit (RPL) or mixed logit (Train 1999) modeling.

Let the conditional indirect utility function be as specified in equation (1). Assume that the parameters (β) are not fixed coefficients but rather are random coefficients that follow a predetermined distributional form. The probability expression from the conditional logit model

$$(19) \qquad P(j) = \frac{\exp(X_j \beta)}{\sum_{k \varepsilon C} \exp(X_k \beta)}$$

is modified to reflect the fact that β has a distribution. Following Train (1999) the overall probability is expressed as the conditional probability (conditional on β) integrated over values of β, or:

$$(20) \qquad P(j) = \int \pi_j(\beta)\; g(\beta)\; d(\beta).$$

Given a choice of a specific distribution for β (or assumptions on $g(\beta)$) such as normal or log-normal, estimation of the choice probabilities proceeds with providing estimates of the mean and variance of those parameters assumed to be random. Note that if $g(\beta)$ is constant or degenerate, this model reduces to the standard conditional logit model. Also, note the similarity of equation (20) to equation (18). Both are essentially weighted conditional logit models. Equation (18) reflects a finite weighting or mixture, whereas equation (20) is a continuous mixture. See Train (1999), Revelt and Train (1998), or Layton (2000) for details.

12.2 Relaxing the IIA Assumption

The simple conditional logit model produces probabilities of the form expressed in equation (19). However, the ratio of probabilities for any two alternatives (i and j) results in:

$$(21) \qquad \frac{P(i)}{P(j)} = \frac{\exp(V_i)}{\exp(V_j)}$$

Thus the ratio of probabilities between i and j is unaffected by any other alternative in the choice set, and the conditional logit model depends on the IIA property. This property results in elasticities that are limited in flexibility and generally produces substitution patterns that are simplistic (the elasticities of the probability of choosing alternative j with respect to a change in an attribute in alternatives other than j are all equal). In the simple camping choice experiment, for example, the two camping alternatives would likely be more similar or there would be unobserved correlation between these alternatives, relative to the opt-out alternative. However, in the conditional logit formulation, there is no correlation between the unobserved effects (errors) of the alternatives. A further implication of choosing the conditional logit model is that the cross elasticities (the percent change in probability of choosing i for a percent change in an attribute level in any alternative j) are identical. This is a highly restrictive form of preference.

12.2.1 Nested Logit

An approach to address these issues is to estimate a nested logit model (McFadden 1981; Ben-Akiva and Lerman 1985; Louviere, Hensher, and Swait 2000). Suppose we consider the camping example above, but assume that camping alternatives (A or B) are similar relative to the alternative of not going camping (C). The choice of alternatives A, B, or C could be specified as the probability of choosing an alternative, conditional on the probability of going camping (A or B) versus (C). Utility would be decomposed into utility associated with camping versus not camping, and utility arising from camping sites (A or B) conditional on going camping. In terms of probability expressions, this is reflected as follows. Let j index alternative sites and m index activities (going camping or not). The utility of choosing site j in activity m (camping) can be expressed as:

$$(22) \qquad U_{jm} = U_{j|m} + U_m = V_{j|m} + V_m + e_{j|m} + e_m$$

The two error terms ($e_{j|m}$ and e_m) reflect the unobserved variation in alternatives j (conditional on m) and m. Assuming independence between the two error terms, one can show that the joint probability of choosing alternative jm is

$$(23) \qquad P(jm) = \frac{\exp a_m(V_m + V_{m*})}{\sum_{m'=1}^{M} [\exp a_m(V_{m'} + V_{m'*})]} \cdot \frac{\exp(V_{j|m})}{\sum_{j'=1}^{J} [\exp(V_{j'|m})]}$$

where V_{m*} is $(1/a_m) \log \sum \exp (V_{j|m})$ or the "inclusive value" or "log-sum" and a_m is the parameter on the inclusive value. The inclusive value term captures the utilities (the expected value of the maximum utility) of the camping alternatives within the utility associated with the activity camping. If $a_m = 1$, then the expression collapses to the simple logit expression. An inclusive value parameter of 1 corresponds to equal correlation between the alternatives and an inclusive value parameter between zero and 1 indicates the degree of correlation (or similarity) between alternatives within a particular activity.

Expression (22) can also be considered to be the product of probabilities. The probability of choosing alternative j and activity m can be expressed as the product of the probability of choosing alternative j, conditional on choosing activity m, times the probability of choosing activity m. In other words, the probability of choosing camping alternative j is the product of the probability of choosing camping (versus not camping) times the probability of choosing alternative j conditional on choosing camping, or:

$$(24) \qquad P(j,m) = P(j|m)\ P(m).$$

The nested logit model (nesting the decision of where to go camping within the decision to go camping or not) does not have the IIA property and relaxes the assumption of identical substitution elasticities. However, a more interesting interpretation of the nested logit model, in terms of error variance components, is provided below through the description of the mixed multinomial logit model.

12.2.2 Mixed Multinomial Logit Models: Error Components

Random parameter models were described in Section 12.1.3 as one outcome of a mixed logit structure. An alternative interpretation of mixed logit can be used to construct nested logit models, as well as a variety of other models that involve correlation between the unobserved elements of the alternatives. Following Train (1998, 1999), let the conditional indirect utility of alternative j be expressed as

$$(25) \qquad V_j = \beta X_j + \mu Z_j + \varepsilon_j$$

where ε_j is an IID extreme value error term (extreme value is chosen to be consistent with the logit framework), and μZ_j represents an additional stochastic component of the utility. Let μ be a mean zero term. The inclusion of μZ_j in the stochastic component of the utility function allows alternative-specific elements to enter the stochastic portion of utility, and thus allows for the examination of various correlations of unobserved effects. As Train (1999) illustrates, defining Z_j as a dummy variable for a subset of the overall set of alternatives (e.g., camping alternatives) provides an estimate of the error correlation among this subset of alternatives, and the variance on μ becomes an estimate of the correlation or the inclusive value parameter. Note that if μ is zero and non-random, the conditional logit model results.

Estimation of such a model relies on the relationship between the mixed logit/random parameters model specified above and the error components model. If $X_j = Z_j$, the parameter β can be interpreted as the mean while μ can be interpreted as the variance. In an error components interpretation, one is most interested in the correlations between alternatives (as in nested logit models) as captured by the stochastic terms (Brownstone and Train 1996; Revelt and Train 1996). Nevertheless, the estimation of these models follows the approach presented in equations (19) and (20) above.

13. FUTURE DIRECTIONS

These are still early days in the application of ABMs to environmental valuation. Researchers continue to evaluate the effectiveness of these methods. Efforts to improve design and analysis of data generated by ABMs are on-going. The current literature can be divided into these components: evaluating and testing ABM performance, improving econometric analysis of ABM data, and improving ABM designs.

13.1 Evaluation and Testing of ABM Performance

Many writers have speculated that ABMs may outperform contingent valuation with respect to strategic behavior, hypothetical bias, or a variety of other challenging issues associated with stated preference methods. However, very few tests of ABM performance have been conducted. Recent results from Carlsson and Martinsson (2001) suggest that ABMs perform very well relative to market or experimental market choices. In addition, studies like that of Haener, Boxall, and Adamowicz (2000) show that ABMs do a good job in predicting "out of sample" (data not included in the sample used for estimation) choices. Nevertheless, additional research is required to evaluate ABM performance and subject ABMs to the same level of scrutiny as contingent valuation methods have received in the past.

13.2 Econometric Analysis

Attribute-based methods are often administered such that individuals respond to several ranking, rating, or choice tasks. Presenting respondents with as many as 16 such tasks is not unusual. In simple econometric analysis, these tasks are assumed to be independent. However, some empirical and much anecdotal evidence suggests that these responses are not independent. Respondents may learn about their preferences, or they may become fatigued during the survey. In general, the responses may be serially correlated, or at least should be treated as arising from panel data. Mixed multinomial logit models offer econometric methods to address correlations between choice sets and panel data considerations within discrete choice/random utility data (e.g., Train 1999; Revelt and Train 1998, McFadden and Train 2000). However, in addition to simple correlation between alternatives, issues of fatigue and learning may be better represented as systematic preference changes in response to sequences of questions. Swait and Adamowicz (2001a) provide one approach to such an issue by examining preference variation with a finite mixture model operating on question order and task complexity. Certainly, other approaches also could be explored to assess the implications of question order, serial correlation, and stated preference question response.

In addition to serial correlation, research on combining data types, or data fusion, is on-going. If revealed preference responses suffer from collinearity, or from limited data range, ABMs can facilitate the estimation of parameters that are difficult or impossible to measure using revealed preference data alone. Evidence suggests that joint revealed and stated preference models outperform revealed preference methods within samples (Adamowicz, Louviere, and Williams 1994) as well as in out-of-sample prediction tests (Haener, Adamowicz and Boxall 2000). However, many unanswered questions in data fusion remain including the following three: What weight should be placed on each data type? Are there more efficient ways to combine data? Can combining ABM data with small samples of revealed preference data provide better benefits transfers than transfers of revealed preference data from other regions?

13.3 Design Issues

Psychologists and researchers in human judgment and decision making have long focused on the effect of changes in decision context on response and implied behavior. Similar issues arise in ABM surveys. Do changes in context affect responses? Are these effects systematic and could they be examined econometrically? Currently rules-of-thumb are used to determine the number of attributes, alternatives, and questions, and orthogonal designs are heavily relied upon to generate the correlation structure between alternatives. However, these rules-of-thumb have not been rigorously examined. In addition to complexity, other context effects arise, such as the respondent's reference group (family, peer group, etc.) and the degree to which these elements affect preferences. Although economists have historically focused on individual responses, there is increasing interest in examining demand and preference as arising from groups such as households (Smith and van Houtven 1998 among others) or as being affected by reference groups (Manski 2000 or Brock and Durlauf 1995 among others).

14. CONCLUSIONS

ABMs have emerged from a creative linkage of research across disciplines including marketing, psychology, transportation and economics. Through this process, the hedonic framework articulated by Lancaster more than 3 decades ago has been refined by developments in random utility theory, econometrics and experimental design into a set of powerful tools that provide economists with new methods for environmental valuation. If carefully designed and administered, ABMs can provide defensible estimates of environmental value for behavioral analysis (such as recreational choice) or passive use valuation. However, without careful attention to framing the decision context, applying an appropriate experimental design, developing a focused survey instrument and implementing robust empirical procedures, ABM applications will not provide the desired information.

These are still early and exciting days in the application of ABMs to environmental valuation. As stated preference methods, ABMs are closely related to contingent valuation methods and face similar issues relating to the validity of responses. Assessment of the validity and consistency of ABM responses will undoubtedly be an important avenue of future research. However, research to date combining stated and revealed preference data indicates that ABMs, when properly applied, can provide information on preferences that is consistent with actual behavior. We anticipate that future research on ABMs will not only provide a deeper understanding of environmental preferences but will also enhance other applications of stated preference methods.

ACKNOWLEDGMENTS

The development of this book has been a wonderful collegial experience. We would like to thank Patty Champ, Tom Brown and Kevin Boyle for assembling this group of authors and for inviting us to be part of this team. We are indebted to George Peterson for his leadership over the years in promoting the scientific assessment of nonmarket values and hope that this book makes a positive contribution to his intellectual legacy. We gratefully acknowledge the authors of other chapters in this text for their comments during our workshop sessions. In particular, many thanks are due to Mark Dickie for his insightful review of our chapter during its development and to Tom Brown whose editing of our chapter substantially improved its quality. Vic Adamowicz would like to acknowledge that this chapter was completed while he was a Gilbert White Visiting Fellow at Resources for the Future, Washington DC.

NOTES

1 The term "conjoint" arose from early attempts to examine a set of attributes at the same time or "consider jointly". Traditional conjoint analysis involved ratings. Some authors refer to the methods we describe in this chapter as being examples of conjoint analysis. Others refer to choice methods as "choice-based conjoint". We prefer the term "attribute-based method" because it more explicitly highlights the focus on examination of a bundle of attributes associated with a good, and because it leaves open the choice of elicitation approach, whether it be **choices** between bundles, **ranking** of bundles, or **ratings** of individual bundles.

2 Rating scale approaches, or traditional conjoint analysis, are based on Torgerson's (1958) Law of Comparative Judgment. This approach presents individuals with profiles (alternatives) or bundles of attributes and asks them to provide a rating of each profile (e.g., 1 to 10, where 10 is very good, and 1 is very poor). The development of rating-based conjoint is discussed in Green and Srinivasan (1978) and Louviere (1988b). Axiomatic theories of conjoint analysis have also been developed (Krantz et al 1969, Barron 1977) that deal with the relationship between the ordinal numerical scores provided in conjoint rating tasks and various forms of preferences or utility. One of the earliest empirical studies using rating scales to measure preference parameters concerned preferences for the visual appearance of residential neighborhoods (Peterson 1967).

3 Individual prediction models typically assumed that the best or first choice alternative would be the product that received the highest predicted utility. Summation of first choices across the sample provided estimates of market share. An alternative approach introduced by Louviere and Woodworth (1983) used the predicted utilities in a modified Luce model to predict the probability that an individual would choose competing products. The summation of predicted probabilities over the sample then generated choice frequencies, which were used in a weighted least squares regression of the multinomial logistic model.

4 McFadden (1986) goes on to state that "...it is unnecessary to provide accurate behavioral models individual-by-individual to obtain good market forecasts. It is sufficient to determine the distribution of behavior in the population" (p. 278). The multinomial logit model is based on the assumption that behavior in the population follows an extreme-value type 1 distribution.

5 See also subsequent work by Manski (1977) and Yellot (1977).

6 In addition, Hensher (1994) provides an overview of RUM models in the transportation literature while Louviere (1994) reviews applications in marketing (see also the case study chapters in Louviere, Hensher, and Swait, 2000).

7 See also Swait and Louviere (1993) in the marketing research literature.

8 In Table 2, note that the elements of each column vector sum to zero and that, within sets of main effects and 2-way interactions, the inner product of two column vectors equals zero. This second property defines orthogonality or statistical independence.

9 The alias of any factorial effect can be determined using what the experimental design literature refers to as a defining contrast. In Table 2, note that the 3-way interaction A1A2A3 contains a vector of +1's for the first ½ fraction (and a vector of – 1's for the second ½ fraction). A1A2A3 is the "defining contrast" because it was used to split the factorial into two fractions. In a 2^n design, the alias of any factorial effect is found by multiplying the effect by the defining contrast (Cochran and Cox 1957). So, for example, the alias of A3 is the generalized interaction $A1A2A3^2$. Squared terms are canceled in interpreting generalized interactions, so $A1A2A3^2$ is read as A1A2 which, as shown previously, is the alias for A3.

10 For a good discussion of various design strategies, see Louviere, Hensher, and Swait (2000).

11 The smallest orthogonal main effects designs for various combinations of attributes, levels and choice options is presented in Louviere, Hensher, and Swait (2000; Table 5.3, p.121).

12 For example, methods for analyzing panel data when the responses are binary are well known (e.g., Hsiao 1986).

13 Choice experiments often include "opt-out" (none of the offered alternatives) or "status quo" options as fixed alternatives. These are referred to as fixed alternatives because their attribute levels or descriptions do not vary over the set of choices presented to the respondent. Coding of the attribute levels for the status quo option may cause confusion. If attribute levels for the status quo option are known, then they are coded in the usual manner. This would be the case, for example, if the status quo were one of the options in a policy choice set. For another example, consider the options for a recreation choice experiment. If the options included two hypothetical alternatives and an alternative for a currently existing substitute, then the fixed alternative might ask people to provide descriptions of the substitute option (for an evaluation of this approach, see Banzhaf, Johnson, and Matthews 2001). Alternatively, it might be the case that a "generic" opt-out option is included such as "I would choose neither of the hypothetical alternatives" in a recreation choice experiment. In this case, because nothing is known about the attributes of this option, attribute coding is typically handled using zeros for attribute levels of the opt-out option. Because the MNL model is based on utility differences, this approach normalizes utility relative to the opt-out option.

14 For an application of random utility theory to travel cost models, see Chapter 9.

15 Indirect utility in equation (1) modifies the indirect utility function described in Chapter 2 by (1) considering non-market goods to be described by a vector of attributes, and (2) the addition of a random error term. Further, indirect utility described in Chapter 2 is global, in that it encompasses all goods and services in an individual's consumption bundle. The indirect utility function described here is an additively separable sub-utility function that is specific to the particular non-market good under consideration.

16 The multinomial probit model relaxes the assumption of independence of irrelevant alternatives and allows a more general pattern of covariance among alternatives. Empirical difficulties formerly associated with estimating this model have recently been addressed and applications to ABMs are anticipated.

17 The IIA axiom states that the ratio of the probabilities of choosing any two alternatives is independent of the attributes or the availability of other alternatives (e.g. Ben-Akiva and Lerman 1985).

18 Given a single data set, it is assumed that the scale parameter μ equals unity. If $n > 1$ data sets are available, it is possible to estimate the value of μ for $n - 1$ of the data sets (Swait and Louviere 1993).

19 In the MNL model, scale is inversely proportional to the error variance: $\sigma^2 = \pi^2 / 6\mu^2$ where π is the mathematical value 2.1415... and μ is the scale.

20 Ranking data have also been analyzed using an ordered probit specification (e.g., see Boyle et al. 2001).

21 This method is appropriate if a limited number of values are included on the rating scale.

22 There is some literature that suggests that the structure of preferences varies when elicited by different response formats (Huber 1997; Boyle et al. 2001).

23 See Morey (1999) for additional detail on the derivation of compensating variation in logit models.

24 D = 0.57722.

25 See Morey (1999) and Choi and Moon (1997) for details regarding estimation of welfare effects in more complex cases (nested logit models, etc.). Morey (1999) also provides a very good discussion of estimating confidence intervals for welfare measures.

26 Details on integrating welfare measures from choice occasions with changing frequencies of choice occasions (number of trips) can be found in Morey and Waldman (1998), Hausman, Leonard, and McFadden (1995).

27 This still assumes zero income effects.

28 In the original experiment, survey respondents were asked to consider seven timber harvesting attributes. Here we reduce the number of attributes to three. Also, in the original experiment, the "opt-out" option was presented in a sequential manner. Here we treat the opt-out option as being presented simultaneously with the other alternatives.

REFERENCES

Adamowicz, W., P. Boxall, M. Williams, and J. Louviere. 1998. Stated Preference Approaches for Measuring Passive Use Values: Choice Experiments and Contingent Valuation. *American Journal of Agricultural Economics* 80:64-75.

Adamowicz, W., J. Louviere, and J. Swait. 1998. *Introduction to Attribute-Based Stated Choice Methods*. Final Report to Resource Valuation Branch, Damage Assessment Center, NOAA, U.S. Department of Commerce.

Adamowicz, W., J. Swait, P. Boxall, J. Louviere, and M. Williams. 1997. Perceptions versus Objective Measures of Environmental Quality in Combined Revealed and Stated Preference Models of Environmental Valuation. *Journal of Environmental Economics and Management* 32:65-84.

Adamowicz, W. L., J. Louviere, and M. Williams. 1994. Combining Stated and Revealed Preference Methods for Valuing Environmental Amenities. *Journal of Environmental Economics and Management* 26:271-292.

Adelman, S. 1962. Orthogonal Main-Effect Plans for Asymmetrical Factorial Experiments. *Technometrics* 4:21-46.

Anderson, N. H. 1970. Functional Measurement and Psychophysical Judgement. *Psychological Review* 77:153-170.

Banzhaf, M. R., F. R. Johnson, and K. E. Matthews. 2001. Opt-out Alternatives and Anglers' Stated Preferences. In *The Choice Modelling Approach to Environmental Valuation*. Edited by J. Bennett, and R. Blamey. Northampton, MA: Edward Elgar.

Barron, F. H. 1977. Axiomatic Conjoint Measurement. *Decision Sciences* 8:548-559.

Beggs, S., S. Cardell, and J. Hausman. 1981. Assessing the Potential Demand for Electric Cars. *Journal of Econometrics* 16:1-19.

Ben-Akiva, M., and S. R. Lerman. 1985. *Discrete Choice Analysis: Theory and Application To Travel Demand*. Cambridge, MA: MIT Press.

Ben-Akiva, M., T. Morikawa, and F. Shiroishi. 1992. Analysis of the Reliability of Preference Ranking Data. *Journal of Business Research* 24:149-164.

Boxall, P., W. Adamowicz, M. Williams, J. Swait, and J. Louviere. 1996. A Comparison of Stated Preference Approaches to the Measurement of Environmental Values. *Ecological Economics* 18:243-253.

Boxall, P. C., and W. L. Adamowicz. 1999. Understanding Heterogeneous Preferences in Random Utility Models: The Use of Latent Class Analysis. Staff Paper 99-02, Department of Rural Economy, University of Alberta, Edmonton, Alberta.

Boyle, K. J., T. P. Holmes, M. F. Teisl, and B. Roe. 2001. A Comparison of Conjoint Analysis Response Formats. *American Journal of Agricultural Economics* 83:441-454.

Brock, W. A., and S. N. Durlauf. 1995. Discrete Choice with Social Interactions I: Theory. NBER Working Paper Series, 5291. October. Cambridge, MA.

Brownstone, D., and K. Train. 1996. Forecasting New Product Penetration with Flexible Substitution Patterns. Paper presented at American Economic Association meetings, New Orleans.

Carlsson, F., and P. Martinsson. 2001. Do Hypothetical and Actual Marginal Willingness to Pay Differ in Choice Experiments? Application to Valuation of the Environment. *Journal of Environmental Economics and Management* 41:179-192.

Carmone, F. J., and P. E. Green. 1981. Model Misspecification in Multiattribute Parameter Estimation. *Journal of Marketing Research* 18:87-93.

Cattin, P., and D. R. Wittink. 1982. Commercial Use of Conjoint Analysis: A Survey. *Journal of Marketing* 46:44-53.

Chapman, R. G., and R. Staelin. 1982. Exploiting Rank-Ordered Choice Set Data within the Stochastic Utility Model. *Journal of Marketing Research* 19:288-301.

Choi, K-H., and C-G Moon. 1997. Generalized Extreme Value Model and Additively Separable Generator Function. *Journal of Econometrics* 76:129-140.

Cochran, W. G., and G. M. Cox. 1957. *Experimental Designs*. New York: John Wiley & Sons.

Court, A. T. 1939. Hedonic Price Indexes with Automotive Examples. In *The Dynamics of Automobile Demand*. New York: General Motors.

Dillman, D. A. 1978. *Mail and Telephone Surveys - the Total Design Method*. New York: Wiley.

Foster, V., and S. Mourato. 2000. Valuing the Multiple Impacts of Pesticide Use in the UK: A Contingent Ranking Approach. *Journal of Agricultural Economics* 51:1-21.

Gan, C., and E. J. Luzar. 1993. A Conjoint Analysis of Waterfowl Hunting in Louisiana. *Journal of Agricultural and Applied Economics* 25:36-45.

Garrod, G. D., and K. G. Willis. 1996. The Non-Use Benefits of Enhancing Forest Biodiversity: A Contingent Rank Study. *Ecological Economics* 21:45-61.

Garrod, G. D., and K. G. Willis. 1998. Using Contingent Ranking to Estimate the Loss of Amenity Value for Inland Waterways from Public Utility Service Structures. *Environmental and Resource Economics* 12: 241-247.

Green, P. E., and V. R. Rao. 1971. Conjoint Measurement for Quantifying Judgmental Data. *Journal of Marketing Research* 8:355-363.

Green, P. E., and V. Srinivasan. 1978. Conjoint Analysis in Consumer Research: Issues and Outlook. *Journal of Consumer Research* 5:103-123.

Green, P. E., and Y. Wind. 1975. New Way to Measure Consumers' Judgments. *Harvard Business Review* 53:107-117.

Green, P. E., J. D. Carroll, and S. M. Goldberg. 1981. A General Approach to Product Design Optimization via Conjoint Analysis. *Journal of Marketing* 45:17-37.

Griliches, Z. 1961. Hedonic Price Indexes for Automobiles: An Econometric Analysis of Quality Change. In *Price Indexes and Quality Change: Studies in New Methods of Measurement.* Edited by Z. Griliches. Cambridge, MA: Harvard University Press.

Gupta, S., and P. K. Chintagunta. 1994. On Using Demographic Variables to Determine Segment Membership in Logit Mixture Models. *Journal of Marketing Research* 31:128-136.

Haener, M., P. C. Boxall, and W. L. Adamowicz. 2000. Modeling Recreation Site Choice: Do Hypothetical Choices Reflect Actual Behavior? *American Journal of Agricultural Economics.* 83(3): 629-642.

Hahn, G. J., and S. S. Shapiro. 1966. *A Catalog and Computer Program for the Design and Analysis of Orthogonal Symmetric and Asymmetric Fractional Factorial Experiments.* Technical Report 66-C 165. Schnectady, NY: General Electric Research and Development Center.

Hammond, K. R. 1955. Probabilistic Functioning and the Clinical Method. *Psychological Review* 62:255-262.

Hanemann, W. M. 1999. Welfare Analysis with Discrete Choice Models. In *Valuing Recreation and the Environment: Revealed Preference Methods in Theory and Practice.* Edited by C. Kling, and J. Herriges. Northampton MA: Edward Elgar.

Hanley, N., D. Macmillan, R. Wright, C. Bullock, I. Simpson, D. Parrsisson, and B. Crabtree. 1998. Contingent Valuation versus Choice Experiments: Estimating the Benefits of Environmentally Sensitive Areas in Scotland. *Journal of Agricultural Economics* 49:1-15.

Hanley, N., R. Wright, and W. Adamowicz. 1998. Using Choice Experiments to Value the Environment. *Environmental and Resource Economics* 11:413-428.

Hausman, J. A., G. K. Leonard, and D. McFadden. 1995. A Utility-Consistent Combined Discrete Choice and Count Data Model Assessing Recreational Use Losses Due to Natural Resource Damage. *Journal of Public Economics* 56:1-30.

Hausman, J. A., and P. A. Ruud. 1987. Specifying and Testing Econometric Models for Rank-Ordered Data. *Journal of Econometrics* 34:83-104.

Hensher, D. A. 1994. Stated Preference Analysis of Travel Choices: The State of Practice. *Transportation* 21:107-133.

Holmes, T. P., K. Alger, C. Zinkhan, and E. Mercer. 1998. The Effect of Response Time on Conjoint Analysis Estimates of Rainforest Protection Values. *Journal of Forest Economics* 4:7-28.

Holmes, T. P., and K. J. Boyle. 2002. Using Stated Preference Methods to Estimate the Value of Forest Attributes. In *Forests in a Market Economy*. Edited by K. A. Lee, and E. O. Sills Dordrecht, Netherlands: Kluwer Academic Publishers. Forthcoming.

Hsiao, C. 1986. *Analysis of Panel Data*. New York: Cambridge University Press.

Huber, J., and J. McCann. 1982. The Impact of Inferential Beliefs on Product Evaluations. *Journal of Marketing Research* 19:324-333.

Huber, J. 1997. What Have We Learned from 20 Years of Conjoint Research: When to Use Self-Explicated, Graded Pairs, Full Profiles, or Choice Experiments. Paper presented at Sawtooth Software Conference, August. Accessed at www.sawtoothsoftware.com/download/techpap/whatlrnd.pdf

Johnson, F. R., and W. Desvousges. 1997. Estimating Stated Preferences with Rated-Pair Data: Environmental, Health, and Employment Effects of Energy Programs. *Journal of Environmental Economics and Management* 34:79-99.

Krantz, D. H., R. D. Luce, P. Suppes, and A. Tversky. 1969. *Foundations of Measurement*. New York: Academic Press.

Lancaster, K. 1966. A New Approach to Consumer Theory. *Journal of Political Economy* 74:132-157.

Lareau, T. J., and D. A. Rae. 1989. Valuing WTP for Diesel Odor Reductions: An Application of Contingent Ranking Technique. *Southern Economic Journal* 55:728-742.

Layton, D. F. 2000. Random Coefficient Models for Stated Preference Surveys. *Journal of Environmental Economics and Management* 40:21-36.

Louviere, J., D. Hensher, and J. Swait. 2000. *Stated Choice Methods - Analysis and Application*. Cambridge: Cambridge University Press.

Louviere, J. J. 1988a. *Analyzing Individual Decision Making: Metric Conjoint Analysis*. Sage University Series on Quantitative Applications in the Social Sciences, Series No. 67. Newbury Park, Ca: Sage Publications, Inc.

Louviere, J. J. 1988b. Conjoint Analysis Modeling of Stated Preferences: A Review of Theory, Methods, Recent Developments and External Validity. *Journal of Transport Economics and Policy* 10:93-119.

Louviere, J. J. 1994. Conjoint Analysis. In *Advances in Marketing Research*. Edited by R. Bagozzi. Cambridge, MA: Blackwell Publishers.

Louviere, J. L., and D. A. Hensher. 1983. Using Discrete Choice Models with Experimental Design Data to Forecast Consumer Demand for a Unique Cultural Event. *Journal of Consumer Research* 10:348-361.

Louviere, J. L., and G. Woodworth. 1983. Design and Analysis of Simulated Consumer Choice or Allocation Experiments: An Approach Based on Aggregate Data. *Journal of Marketing Research* 20:350-367.

Luce, R. D. 1959. *Individual Choice Behavior*. New York: John Wiley & Sons.

Luce, R. D., and J. W. Tukey. 1964. Simultaneous Conjoint Measurement: A New Type of Fundamental Measurement. *Journal of Mathematical Psychology* 1:1-27.

Mackenzie, J. 1993. A Comparison of Contingent Preference Models. *American Journal of Agricultural Economics* 75:593-603.

Manski, C. 1977. The Structure of Random Utility Models. *Theory and Decision* 8:229-254.

Manski, C. F. 2000 Economic Analysis of Social Interactions. *Journal of Economic Perspectives*. 14:115-136

Marschak, J. 1960. Binary Choice Constraints on Random Utility Indicators. In *Stanford Symposium on Mathematical Methods in the Social Sciences*. Edited by K. Arrow. Stanford, CA: Stanford University Press.

Mazzotta, M. J., and J. J. Opaluch. 1995. Decision Making When Choices Are Complex: A Test of Heiner's Hypothesis. *Land Economics* 71:500-515.

McFadden, D. 1974. Conditional Logit Analysis of Qualitative Choice Behavior. In *Frontiers in Econometrics*. Edited by P. Zarembka. New York: Academic Press.

McFadden, D. 1981. Econometric Models of Probabilistic Choice. In *Structural Analysis of Discrete Data with Econometric Applications*. Edited by C. Manski, and D. McFadden. Cambridge, MA: MIT Press.

McFadden, D. 1986. The Choice Theory Approach to Market Research. *Marketing Science* 5:275-297.

McFadden, D., and K. Train. 2000. Mixed MNL Models for Discrete Response. *Journal of Applied Econometrics*. 15:447-470.

McLean, R., and V. Anderson. 1984. *Applied Factorial and Fractional Designs*. New York: Marcel Dekker.

Moore, W. L., and M. B. Holbrook. 1990. Conjoint Analysis on Objects with Environmentally Correlated Attributes: The Questionable Importance of Representative Design. *Journal of Consumer Research* 16: 490-497.

Morey, E. R. 1999. Two RUMs uncloaked: nested logit models of site choice and nested logit models of participation and site choice. In *Valuing Recreation and the Environment: Revealed Preference Methods in Theory and Practice*. Edited by J. A. Herriges, and C. L. Kling. Northampton MA: Edward Elgar.

Morey, E. R., and D. M. Waldman. 1998. Measurement Error in Recreation Demand Models: The Joint Estimation of Participation, Site Choice and Site Characteristics. *Journal of Environmental Economics and Management* 35:262-276.

Morrison, M., and K. Boyle. 2001. Comparative Reliability of Rank and Choice Data in Stated Preference Models. Paper presented at the European Association of Environmental and Resource Economics 2001 Conference, Southampton, UK, June .

Opaluch, J. J., S. Swallow, T. Weaver, C. Wessels, and D. Wichlens. 1993. Evaluating Impacts from Noxious Waste Facilities: Including Public Preferences in Current Siting Mechanisms. *Journal of Environmental Economics and Management* 24:41-59.

Peterson G. L. 1967. A Model of Preference: Quantitative Analysis of the Perception of the Visual Appearance of Residential Neighborhoods. *Journal of Regional Science* 7:19-31.

Rae, D. A. 1983. The Value to Visitors of Improving Visibility at Mesa Verde and Great Smoky National Parks. In *Managing Air Quality and Scenic Resources at National Parks and Wilderness Areas*. Edited by R. D. Rowe, and L. G. Chestnut. Boulder, CO: Westview Press.

Revelt, D., and K. Train. 1998. Mixed Logit with Repeated Choices: Households' Choice of Appliance Efficiency Level. *Review of Economics and Statistics* 53:647-657.

Roe, B., K. J. Boyle, and M. F. Teisl. 1996. Using Conjoint Analysis to Derive Estimates of Compensating Variation. *Journal of Environmental Economics and Management* 31:145-159.

Shonkwiler, J. S., and W. D. Shaw. 1997. Shaken, not Stirred: A Finite Mixture Approach to Analyzing Income Effects in Random Utility Models. Paper Presented at the 1997 Annual Meeting of the American Agricultural Economics Association. August 2-4, Toronto, Ontario.

Small K. A., and H. S. Rosen. 1981. Applied Welfare Economics with Discrete Choice Models. *Econometrica* 49:105-130.

Smith, V. K., and W. H. Desvousges. 1986. *Measuring Water Quality Benefits*. Boston: Kluwer-Nijhoff Publishing.

Smith, V. K., and G. van Houtven. 1998. Non-market Valuation and the Household. Resources for the Future Discussion Paper 98-31. Washington, D.C.

Snedecor, G. W., and W. G. Cochran. 1974. *Statistical Methods*. Ames: Iowa University Press.

Swait, J. 1994. A Structural Equation Model of Latent Segmentation and Product Choice for Cross-sectional, Revealed Preference Choice Data. *Journal of Retailing and Consumer Services* 1:77-89.

Swait, J., and W. Adamowicz. 2001a. The Influence of Task Complexity on Consumer Choice: A Latent Class Model of Decision Strategy Switching. *Journal of Consumer Research* 28:135-148.

Swait, J., and W. Adamowicz. 2001b. Choice Environment, Market Complexity and Consumer Behavior: A Theoretical and Empirical Approach for Incorporating Decision Complexity into Models of Consumer Choice. *Organizational Behavior and Human Decision Processes*. 86(2): 141-167.

Swait, J., and J. Louviere. 1993. The Role of the Scale Parameter in the Estimation and Use of Multinomial Logit Models. *Journal of Marketing Research* 30:305-314.

Thurstone, L. L. 1927. A Law of Comparative Judgment. *Psychology Review* 34:273-286.

Torgerson, W. S. 1958. *Theory and Method of Scaling*. New York: John Wiley & Sons.

Train, K. E. 1998. Recreation Demand Models with Taste Differences over People. *Land Economics* 74(2):230-239.

Train, K. E. 1999. Mixed Logit Models for Recreation Demand. In *Valuing Recreation and the Environment: Revealed Preference Methods in Theory and Practice*. Edited by J. A. Herriges, and C. L. Kling. Northampton MA: Edward Elgar.

Viscusi, W. K., W. A. Magat, and J. Huber. 1991. Pricing Environmental Health Risks: Survey Assessments of Risk-Risk and Risk-Dollar Tradeoffs for Chronic Bronchitis. *Journal of Environmental Economics and Management* 21:32-51.

Winer, B. J. 1971. *Statistical Principles in Experimental Design.* New York: McGraw-Hill.

Yellot, J. I. 1977. The Relationship between Luce's Choice Axiom, Thurstone's Theory of Comparative Judgment and the Double Exponential Distribution. *Journal of Mathematical Psychology* 15:109-144.

Chapter 7

MULTIPLE GOOD VALUATION

With Focus on the Method of Paired Comparisons

Thomas C. Brown and George L. Peterson
U.S. Forest Service, Rocky Mountain Research Station

1. INTRODUCTION

Assume for the moment that you are the supervisor of the Roosevelt National Forest (it's located near our town of Fort Collins, Colorado). You have been asked by the Chief of the Forest Service how you would spend a specified increase in your budget. You could spend the increase on such projects as improved campgrounds, better roads for reaching the backcountry, reduction in forest fuels to lower the risk of wildfire, and watershed management to lower erosion and thereby improve fish habitat. You know your chances of getting an increase depend on how well you support your proposal. You can design options that use up the budget increase, but you don't have good information about the benefits of, or even the public preferences for, the options. You know what the vocal interest groups want, but you would like your proposal to have some quantitative justification that reflects the values of the wider citizenry who care about the National Forest. You examined the economic analyses that had been done about forest resources in the area and found no studies for most of the major resources you manage. You don't have time for separate valuation studies of each of the options. However, you could commission a single study directly comparing the values people place on the options—a multiple good valuation study. Furthermore, because the options cost the same amount, a preference ordering of the options is all you need to choose the best one. That option might then become the focus of an economic valuation study, which would allow a benefit-cost comparison.[1]

At the very least, a multiple good valuation study provides a reliable ranking of the goods—that is, it provides an ordinal scale measure of the values. A next step would be to provide an interval scale measure (defined in Chapter 4) of the values of the goods. Both ordinal and interval scales are preference orderings, but the latter has useful properties not found in the former, as seen later in this chapter. A further step would be to obtain a set of values that each have meaning in an absolute (i.e., ratio scale) sense, as would a set of economic values. In this chapter "value" refers to assigned value in general, which includes but is not restricted to monetary value (Brown 1984).

Stated preference methods for ordering preferences—such as ranking, rating, binary choice, and multiple choice—rely on asking each respondent to consider numerous items. Because the items are compared by each respondent, the resulting values are by definition comparable—a condition that is not necessarily assured for values that are each obtained from individual studies, where uncontrolled differences between studies may affect results.

Given the limited space available here, we will focus on only one multiple good valuation method, paired comparisons, which is a form of binary choice. Paired comparisons provide a rich set of data that allows estimation of interval or, in certain applications, ratio scale values as well as estimates of individual respondent reliability. Paired comparisons require more questions of respondents to evaluate a given set of goods than do, say, ratings, but the measures of respondent reliability that paired comparisons provide may justify the extra effort.

With paired comparisons, items in a choice set are presented in pairs and respondents are asked to choose the item in each pair that is superior on a specified dimension. Early use of paired comparisons, such as by Fechner (1860), focused on perception of physical dimensions (e.g., weight, length). The method was later extended by psychologists, marketing researchers, and others to measure preference dimensions such as product attractiveness (Ferber 1974), landscape preference (Buhyoff and Leuschner 1978), and children's choice of playground equipment (Peterson et al. 1973). In the national forest example above, respondents might be asked to select from each pair the project that they would prefer be funded. Most recently, paired comparisons have been used in economic valuation of attributes (Chapter 6) and goods (section 6 of this chapter).

We will present two applications of paired comparisons to value multiple goods. The first application achieves a preference ordering among a set of resource losses, an ordering that could support a schedule of costs to be imposed if the resources were to be damaged. The second application estimates economic values of a mix of public and private goods. Before we present these applications, we explain the theoretical model upon which the analysis of paired comparisons is based, provide some basics about the method, and summarize the steps to follow in implementing a paired comparison study.

In describing the mix of items presented to respondents when using the paired comparison method, we use "goods" generally to indicate public or private goods, resource conditions, or resource losses. "Items" may include such goods and any other stimuli included in the mix, such as monetary amounts.

2. A CHOICE MODEL FOR PAIRED COMPARISONS

Psychologists in the 19th century found that comparative judgments of physical stimuli (such as weight of objects, or loudness of noises) became less accurate the more alike were the stimuli. For example, the closer the weights of paired objects, the greater was the proportion of subjects who misjudged which object of a pair was heaviest, and the greater was the likelihood that a given subject would misjudge the relative weights some of the time in repeated trials (Guilford 1954). Differences in consistency of judgment were later also observed with qualitative judgments, such as of the excellence of hand writing samples or the seriousness of offenses (Thurstone 1927b). Subjects had more difficulty consistently judging some pairs than others. It was assumed that, as with physical stimuli, increasing inconsistency was associated with increasing similarity of the items on the dimension of interest.

In an effort to explain inconsistent paired comparison responses, Thurstone (1927a) proposed a model characterizing judgment as a stochastic or random process, wherein a stimulus, or item, falls along a "discriminal dispersion" around the modal value for the item on the "psychological continuum." The dispersion was attributed to random errors in judgment. This model was soon applied to preference as well as judgment, with the dispersion attributed to random fluctuations in preference.

The characterization of preference as a stochastic process was formalized as a direct random utility function (U) consisting of systematic (deterministic) and random (error) components. For example, the utility (U) of item i to respondent n can be represented by the following relation:

$$(1) \qquad U_{in} = E(U)_{in} + \varepsilon_{in}$$

U is a momentary relative magnitude internal to the person.[2] The systematic component, E(U), represented, for Thurstone, the expected value of U; the error component, ε, represented momentary variability about the expected value due to unobservable influences within a given respondent.[3]

More recent research has indicated that preference and judgment are sensitive to minor changes in the decision context. Contextual influences include phrasing of the questions and the characteristics of the survey setting (e.g., time of day or week, personality of the interviewer) (summarized by Payne, Bettman, and Johnson 1992; Slovic 1995). Some contextual influences are inherent to multiple-item stated preference methods. For example, in judging a set of multi-attribute items, the utility of an item may depend in part on characteristics of the other items in the choice set (Tversky 1996; Brown et. al. 2002) or on the order in which the items appear. One explanation of this behavior is that people simplify their decisions by more heavily weighting certain attributes in certain contexts (Tversky, Sattath, and Slovic 1988). In paired comparisons, for example, the weights for a given comparison may depend on which attributes are shared by the items of the pair; as the items change from one comparison to the next, the weights may change slightly, perhaps as respondents focus on the most obvious differences between the items, causing dispersion in preference or judgment for a given item across the several times it appeared in the pairs presented to the respondent.[4] Such minor contextual influences are difficult to detect and are typically left within the error term ε of equation 1.

Whereas attempts to model responses in psychology have often focused on understanding intra-personal variation, the use of the random utility model in economics has focused on explaining inter-personal variation. In modern demand modeling, the systematic component represents measured attributes of items (e.g., size, color, availability) and measured characteristics of people (e.g., income, age, education) (Ben-Akiva and Lerman 1985). Error is attributed to the analyst's incomplete ability or effort to measure and model.[5] A case for modeling is typically an individual or household—not a given respondent at a specific moment during a survey—and the variability of interest is across people.

Both interpretations of the random utility model—one from psychological scaling and decision making and the other from economic demand modeling—recognize the influence of unmeasured variables; both interpretations model that influence as a random component. However, the two interpretations have historically served different purposes, in keeping with the different objectives of the analyses. The psychological approach has served as a model of respondent behavior, allowing an interpretation of inconsistent responses and forming the theoretical basis for scaling the responses to order preferences for the items of interest. The economic demand interpretation has provided the theoretical structure for modeling preferences as a function of explanatory variables by utility maximization.[6] It is the psychological interpretation that provides the theoretical structure of methods for multiple good valuation, to which we return.

In the absence of strategic behavior, items of higher $E(U)$ will tend to be chosen above other items in a paired comparison exercise. However, because of the randomness inherent in preference or judgment, responses may not always match the order of the expected values of items. Consider the utility of two items i and j:

$$\text{(2)} \qquad \begin{aligned} U_{in} &= E(U)_{in} + \varepsilon_{in} \\ U_{jn} &= E(U)_{jn} + \varepsilon_{jn} \end{aligned}$$

If the error distributions of the items (ε_{in} and ε_{jn}) overlap, the order of the utilities of the items at a given moment (U_{in} and U_{jn}), and therefore the response, may be inconsistent with their respective expected values ($E(U)_{in}$ and $E(U)_{jn}$). The probability (P) that item i will be considered to be of greater utility than item j is:

$$\text{(3)} \qquad P(U_{in} > U_{jn}) = P(E(U)_{in} + \varepsilon_{in} > E(U)_{jn} + \varepsilon_{jn})$$

This probability increases as the difference between $E(U)_{in}$ and $E(U)_{jn}$ becomes larger, and the distributions ε_{in} and ε_{jn} become narrower.

Figure 1 depicts the utility of three items along the psychological continuum, assuming the errors are normally distributed. Given the preferences of Figure 1, responses will nearly always indicate that item k is preferred to

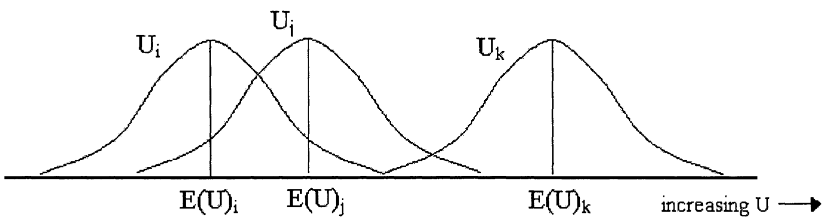

Figure 1. Utility of Three Items Along the Psychological Continuum

item *i*. However, as indicated by the over-lapping error distributions of items *i* and *j*, item *i* will sometimes be preferred to item *j*. Similarly, item *j* will sometimes, although relatively rarely, be preferred to item *k*. A preference for item *i* over item *j*, or of *j* over *k*—although perfectly reasonable given the utility function of Figure 1—would be inconsistent with the expected values of the preference distributions. A successful valuation method will estimate the expected values despite the variability of the response process.

Aside from these item-by-item concerns, multiple item valuation methods are also susceptible to two types of systematic changes that may occur over the course of a series of preference responses: the expected values of the items may change, and the error distributions may change. A change in the expected values may occur, for example, because the order in which the pairs of items are presented affects preference. Such an order effect, to the extent it occurs, cannot be avoided at the individual respondent level, but it can be neutralized across the sample by randomizing for each respondent the order in which the pairs appear.

Regardless of the order in which pairs of items are presented, systematic changes in the error distributions may occur over the course of a series of responses. Two such changes seem plausible. First, fatigue may cause responses to become erratic, leading to increasing inconsistency in the data. Second, the processing of multiple valuations may lead respondents to become more certain about their preferences, leading to decreasing εs and thus decreasing inconsistency with sequence.

3. PAIRED COMPARISON BASICS

With paired comparisons, respondents choose the item in each pair that has the greater magnitude on the given dimension. For valuation applications, the items are either all gains or all losses. The simplest approach, which we will use here, is to present all possible pairs of the items. With t items, there are $t(t-1)/2$ pairs in total. Each pair results in a binary choice that is assumed to be independent of all other choices. The choices allow calculation of a set of scale values indicating the position of the items along the specified dimension.

When presented with a pair of items, respondents typically are not offered an indifference option. This practice is supported by the theory of stochastic preference, wherein the probability of true indifference at any one moment is assumed to be very small. The practice also has the practical benefit of maximizing the amount of information potentially learned about the respondent's preferences. Although the lack of an indifference option may sometimes force respondents to make what seems like unwarranted distinctions, this is not worrisome because, across many comparisons, indifference between two items will be revealed in the data as an equality of preference scores.

3.1 Preference Scores

The full set of choices yields a *preference score* for each item, which is the number of times the respondent prefers an item to other items in the set. Preference scores are easily calculated by creating a t by t matrix and entering a 1 in each cell where the column item was preferred to the row item, and a 0 otherwise. Column sums give the preference scores. For example, Figure 2 contains a hypothetical matrix for a five-item choice set.[7] Preference scores, at the bottom, indicate that, for example, item j, with a preference score of 2, was chosen two of the times it appeared.

A respondent's vector of preference scores describes the individual's preference order among the items in the choice set, with larger integers indicating more preferred items. In the case of a five-item choice set, an individual preference profile with no circular triads (defined in section 3.2) contains all five integers from 0 through 4 (Figure 2).[8] For a given respondent,

	i	j	k	l	m
i	--	1	1	1	1
j	0	--	0	1	1
k	0	1	--	1	1
l	0	0	0	--	1
m	0	0	0	0	--
total	0	2	1	3	4

Figure 2. Five-item Response Matrix for One Respondent, without Intransitivity

the difference between the preference scores of items in a pair is that pair's *preference score difference*. This integer can range from 0 to 4 for a five-item choice set.

A response that is inconsistent with the expected values of a respondent's utility function will not necessarily be detected as inconsistent in the paired comparison data. For example, the responses of Figure 2 for items i, j, and k could have been produced from the utility function depicted in Figure 1. However, because of the theoretically inconsistent choice for the pairing of items j and k, the preference scores of Figure 2 misidentify the preference order of the Figure 1 utility function. The perfectly correct preference order for a single respondent could only be obtained with certainty if the respondent could be independently re-sampled several times and the several data sets were combined.

3.2 Circular Triads and Reliability

If an individual respondent is sampled only once, inconsistency is detectable only if it causes a lack of internal consistency in the data, which appears as one or more *circular triads*. A circular triad is an intransitive preference order among three items. For example, Figure 3 shows a five-item data set where one inconsistent choice causes the following circular triad: $l > k > j > l$. Such an inconsistent choice can occur because of the stochastic nature of preference, because of a context effect, or if the respondent makes a mistake in recording the choice. Circular triads cause some integers to appear more than once in the vector of preference scores (as a preference score of 2 does in Figure 3), while others disappear. In general, the greater the preference score difference of the items involved in the inconsistent response, the greater the number of circular triads that result.[9]

	i	j	k	l	m
i	--	1	1	1	1
j	0	--	1	0	1
k	0	0	--	1	1
l	0	1	0	--	1
m	0	0	0	0	--
total	0	2	2	2	4

Figure 3. Five-item Response Matrix for One Respondent, with Intransitivity

Because respondents enter choices for all pairs of items, reliability can be assessed individually for each respondent. A primary measure of reliability for a set of paired comparisons is the *coefficient of consistency*, which relates the number of circular triads in the respondent's choices to the maximum possible number. The maximum possible number of circular triads, m, is determined by the number of items in the choice set, t; m is equal to $(t/24)(t^2 - 1)$ when t is an odd number and $(t/24)(t^2 - 4)$ when t is even. The number of circular triads in each individual's responses can be calculated directly from the preference scores. Letting a_i equal the preference score of item i (i.e., the number of items in the choice set dominated by the ith item) and b equal the average preference score (i.e., $(t - 1)/2$), the number of circular triads for an individual respondent, c, is (David 1988):

$$(4) \qquad c = \frac{t}{24}(t^2 - 1) - \frac{1}{2}\sum (a_i - b)^2$$

The coefficient of consistency is then *1 - (c/m)* (Kendall and Smith 1940). The coefficient varies from one, indicating that there are no circular triads in a person's choices, to zero, indicating the maximum possible number of circular triads. Brown et al. (2001) found, based on paired comparisons from 1,230 respondents, that the median coefficient of consistency was 0.93, and that 95 percent of respondents had a coefficient of consistency of at least 0.77 (Figure 4). Over one-half of the range in coefficient of consistency was contributed by the remaining 5 percent of respondents. Clearly, a small minority of respondents had highly variable preferences or, more likely, made little or no effort to answer carefully.

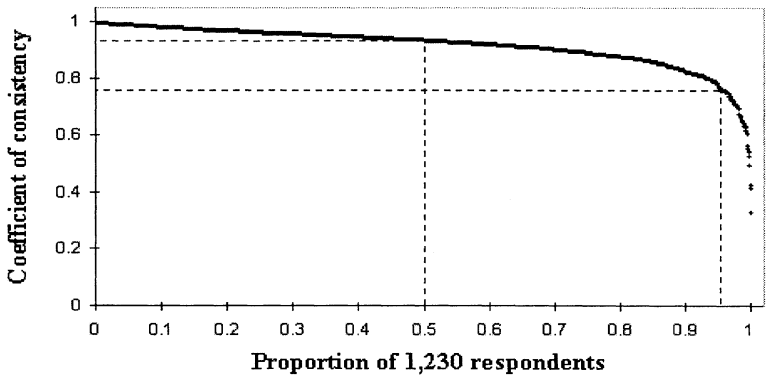

Figure 4. Distribution of Coefficient of Consistency

A short-term test-retest measure of reliability is made possible by repeating a random selection of pairs at the end of a paired comparison session. Some switching of choice over multiple presentations of the same pair is expected because of the stochastic nature of preference, whereby the utilities (*U*) of items change slightly from one moment to the next, as in equation 1.[10] Switching is more likely the closer are the *E(U)*s of the items of the pair. The closer are the *E(U)*s of the items, the smaller is the preference score difference, all else equal. Figure 5 shows the relation for the large set of paired comparisons summarized by Brown et al. (2001). For their data, the likelihood of switching drops from about 0.4 at a preference score difference of 0 to near 0 at a preference score difference of 12 or more.[11] Overall, 12 percent of the pairs were switched on retest.

3.3 Scale Values—Estimating the *E(U)*s

The response matrices of all respondents in the sample can be summed to provide a frequency matrix for the sample. For example, Figure 6 shows a hypothetical frequency matrix for a sample of ten respondents who each judged all possible pairs of five items. Column sums (next to last line) give the

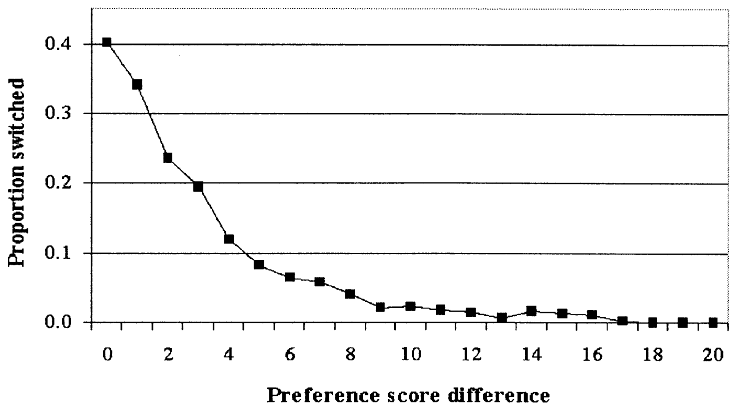

Figure 5. Likelihood of Choice Switching on Retrial

aggregate preference scores for the sample, specifying the number of times each item was chosen across all paired comparisons made by the respondents, and indicting the ordinal position of the items. It is also common to convert aggregate preference scores into percentages expressing the number of times that an item was chosen over the number of times it could have been chosen (Dunn-Rankin 1983). In the example of Figure 6, ten respondents judged each item four times, giving a maximum possible frequency of 40. The scale values (SV) vary from 15 percent to 90 percent of the maximum (last line of Figure 6). Such scale values are a linear transformation of the aggregate preference scores; thus, a zero on this scale indicates only that the item was never chosen, not that the item is of zero utility.

The aggregate preference scores, and thus the scale values, not only indicate the ordinal position of the items; they also approximate an interval scale measure of preference, revealing the sizes of the intervals between items. The approximation, given a sufficient sample size, tends to be very accurate except near the ends of the scale (i.e., except close to the most and least preferred items). The accuracy of the interval information is less near the ends of the scale because the paired comparison choice data are less rich there (i.e., the lack of items beyond those of minimum and maximum *E(U)* limits the number of choices that could help map out interval sizes in the regions of those items).[12]

	i	j	k	l	m
i	--	7	8	9	10
j	3	--	9	7	9
k	2	1	--	8	9
l	1	3	2	--	8
m	0	1	1	2	--
total	6	12	20	26	36
SV	15	30	50	65	90

Figure 6. Example of Frequency Matrix, with Aggregate Preference Scores and Scale Values

Thurstone (1927c) proposed a more involved scaling procedure based on what he called the "law of comparative judgment" that in theory corrects for the problem near the ends of the scale.[13] His approach models $(E(U)_{in} - E(U)_{jn})$ as a function of the probability of $(U_{in} > U_{jn})$. His scaling model relies on the assumption that U is normally distributed about $E(U)$, along with other assumptions regarding the standard deviations of and correlations between the disturbance terms. The most easily applied version of his approach, which assumes independent and identically distributed errors, leads to the probit model (Ben-Akiva and Lerman 1985).

McFadden (1974) proposed another approach to analyzing binary response data. If Thurstone's assumption about the error term is replaced by one of independent double-exponential random disturbances, we have the basis of the logit model, as the difference distribution of two independent double-exponential random variables is the logistic distribution. It is important to note, however, that Thurstone's and McFadden's methods each require data sets that comply with some rather demanding assumptions, and that contain enough cases to allow capture of the full extent of the variability of preference within the population. Without a rich data set, the improvements over aggregate preference scores promised by their methods are not likely to be achieved—a conclusion that would be of more concern if it were not for the fact that scale values computed using their approaches typically correlate very highly with the aggregate preference scores. We do not describe how to apply Thurstone's or McFadden's methods here. The logit model is described in chapter 5 for dichotomous choice contingent valuation and in many other books and articles. Thurstone's method is described by Guilford (1954) and Torgerson (1958).

These concerns about the degree to which a true interval scale measure is achieved become moot if well-chosen anchors of known value, such as monetary amounts, are included among the items to be judged. If anchors of known value are included, ratio scale values of the goods can be approximated from the individual vectors of preference scores (explained in section 6 of this chapter). These approximations can then be averaged to form sample estimates of the values of the goods.

When the known anchors are monetary amounts, each respondent makes multiple binary choices between goods and sums of money. Such paired comparison responses can be analyzed using discrete choice methods, such as binary logit analysis.

4. STEPS IN A PAIRED COMPARISON VALUATION STUDY

The first step in using the method of paired comparisons is to decide on the goal of the study (Table 1). As mentioned above, the method of paired comparisons can be used to order preferences among a set of goods or to estimate monetary values of goods. In keeping with the goal of the study and the types of goods involved, the dimension that respondents are instructed to use in making their choices is also specified. There are many options. For example, if the goal is a preference order among certain public goods, respondents may be asked to choose the good of each pair that they prefer or that they think is the most important for society. If the goal is economic valuation, the analyst must decide whether a payment or compensation measure is desired before the response dimension is specified, as described in the economic valuation application at section 6.1.

Second, the items to be valued are specified. If only a preference order is desired, monetary amounts do not need to be included among the items to be presented to respondents. In this case, specification involves deciding precisely what goods will be valued and how they will be described to respondents. The goods must be carefully defined in terms that respondents will understand, a process that may require focus groups and pre-testing involving persons selected from the likely respondent population. The decision about how much

Table 1. Steps in Using Paired Comparisons to Value Multiple Goods

1.	Decide what is to be measured (a preference order, WTPc*, or WTAc) and what dimension is to be used by respondents in making their choices.
2.	Specify the items to be presented (the goods and, if necessary, the monetary amounts).
3.	Select the respondent population and sample frame.
4.	Choose the method of administration and design the instrument for presenting the pairs of items.
5.	Apply the instrument with the sample.
6.	Analyze the data, including assessing reliability and scaling the choices.
7.	Interpret the results for the application at hand.

*WTP = willingness to pay; WTA = willingness to accept

information to present about the goods is not unlike the similar decision faced in contingent valuation: the analyst wants to present a rich description of the good but not so much information that respondents become bored or confused. However, with multiple good valuation the decision is faced multiple times, once for each of the goods that must be described. The larger the number of goods, the shorter must be the descriptions, all else being equal, to avoid over-burdening the respondent.

We have used two methods of presenting descriptions of the goods when they are complex, which is commonly the case with nonmarket goods. The first is to describe all of the goods before respondents are presented with the paired comparisons. Then for the comparisons, each good is indicated by an abbreviated description or just a title, with the full descriptions still available at all times on one or more separate sheets of paper. With the other approach the goods are not described before the paired comparisons begin, but for each comparison the full description is presented and an abbreviated description is presented just below it. Once respondents become familiar with the goods—which they encounter many times in the course of making their choices—they tend to rely on the short descriptions, lessening the respondents' burden.

The set of goods need not include only those of primary interest. It may be useful to also include goods selected to help respondents choose between the goods of primary interest. Two approaches are useful. First, close substitutes for the goods of primary interest may be included, to help bring to respondents' attention the characteristics of the goods of primary interest. Second, familiar, commonly available market goods may be included. These may serve two purposes: to help familiarize respondents with the paired comparison task by having them make choices between goods with which they are more familiar, and to provide the researcher with data to test whether respondents are making sensible choices.[14] Additional private goods were included in the item mix in the monetary valuation application presented in section 6.

When the paired comparison approach is being used to estimate economic values, monetary amounts (called "bid levels" in the contingent valuation literature) must be included in the mix of items to be compared. Because the paired comparison approach to economic valuation is still being developed, less thought has gone into the details of bid level selection than in the case of dichotomous choice contingent valuation. However, two basic considerations can be offered. First, the bid levels should span the values of the goods for the bulk of the respondent population so that respondents' values for the goods are bracketed by their values for sums of money, allowing the values of the goods to be estimated. Second, increasing the number of bid levels within the specified range allows the values of the goods to be more precisely estimated, but the bid levels must not be so numerous that the pairs of items cannot all be judged within a reasonable amount of time for the respondent population.

The number of items (goods, or goods plus bid levels) to be included in the mix should be large enough to help maintain independence among the choices (i.e., to make it difficult for respondents to remember prior choices or make their task easier by memorizing a ranking of the items), but not so large that respondents become fatigued or lose interest and answer without care. Research has not been done on what number of items is optimal. From our experience we suggest that at least ten items be used. As for the maximum number, when the pairs are presented and responses are recorded on the computer, we have found that respondents can judge at least 200 pairs without loss of reliability. With a 200-pair maximum, if all items are goods, this allows for a maximum of twenty goods ($20(20-1)/2=190$). If, say, ten of the items are bid levels, the 200-pair limit allows at most twelve goods, not ten, because the bid-level-by-bid-level pairs are not presented (since choices between amounts of money are obvious).

In the third step, the respondent population and sample frame are selected. Considerations in making these selections are no different in concept from those with the other stated preference methods. See Chapter 3 of this book for details.

Fourth, the instrument for presenting the pairs and recording the choices is designed. Perhaps the simplest approach is to gather respondents in a room and present the pairs to all of them at the same time, using whatever visual aids are appropriate. However, this approach does not allow the order of presentation to be randomized for each respondent, which can be done by presenting the pairs on paper. Presenting each pair on a separate half sheet of paper helps maintain independence among the choices; however, one cannot be assured that respondents will abide by the request to not look back at prior choices, which would compromise the attempt to maintain independence among choices. Independence is more likely to be maintained if respondents are contacted in person, rather than by mail, because the interviewer is present to observe if respondents ignore the request to not look back at prior choices. Perhaps the most satisfying approach is to use the computer, either via the Internet or by bringing computers (e.g., laptops) to the respondents or bringing respondents to the computers.

Use of a computer program for presenting the pairs of items and recording respondents' choices has several advantages. First, the order of the pairs can be randomized easily for each respondent. Second, independence among an individual's responses is more likely maintained, because respondents cannot go back to see how they responded to prior pairs. Third, if some pairs are to be repeated to check for reliability, the retest pairs can easily and quickly be selected based on the initial responses. Fourth, the computer can keep track of the time required to enter each response; these data can later be used to see if, for example, respondents take more time for early versus later pairs, or more time for difficult (i.e., small preference score difference) versus more easy pairs (Peterson and Brown 1998).

The fifth step is application of the instrument, ideally over a short enough time period that events or news stories do not change respondents' knowledge and experience over the course of data collection. To help maintain independence among the choices, respondents are instructed at the start that each choice is a change from their situation when they began the experiment (i.e., that each choice is made as if it were the first and only choice).

Sixth, the data are analyzed. Methods of assessing reliability and scaling the choices were discussed section 3 and are demonstrated in the two applications that follow in sections 5 and 6.

Finally, the results are interpreted for the policy maker. If only a preference ordering was produced, the interpretation must include the warning that economic benefits have not been estimated and that such benefits may not exceed the costs of providing the goods that were assessed. It may also involve the mapping of monetary values onto the preference ordering, as envisioned with the damage schedule approach described in the next section. If monetary values were produced, the interpretation must include an explanation of the nature of the values that were estimated, which are not necessarily identical to those estimated with more traditional nonmarket valuation methods. The differences are explained in the monetary valuation application in section 6. Also, because independence among choices is assumed, the monetary values are not additive.

5. APPLICATION 1: ORDERING LOSSES AND THE DAMAGE SCHEDULE

It has long been maintained that comparative judgments are easier for people than are absolute judgments. As Nunnally (1976) states, "People simply are not accustomed to making absolute judgments in daily life, since most judgments are inherently comparative...people are notoriously inaccurate when judging the absolute magnitudes of stimuli...and notoriously accurate when making comparative judgments" (p. 40). Nunnally's statement focused largely on judgments of physical phenomena such as the length of lines or the brightness of lights, but his notion can perhaps be applied as well to judgments of monetary value. If so, in contingent valuation we would prefer to use a binary choice question to an open-ended question, because the latter would require an absolute judgment of willingness to pay (WTP). However, even a dichotomous choice contingent valuation question demands some measure of quantification, because to know whether one would pay a specified bid amount, say \$30, for an item, one's maximum WTP must be compared to the \$30. It may simply be more difficult for people to know if they would pay \$30 for an item than to decide which good in a pair they would pay more for.

Binary valuation questions involving a distinct monetary amount cannot avoid requiring the respondent to quantify their values to some degree. This difficulty of binary choice WTP questions perhaps contributes to the sensitivity of such questions to anchors provided by the bid amounts (Boyle, Johnson, and McCollum 1997; Green et al. 1998). It may also account for the recent findings of Breffle and Rowe (2002) that the randomness in paired comparisons of resource conditions was less than that for two other kinds of binary choices involving monetary amounts, referendum contingent valuation and an attribute-based approach. Thus, simplifying the judgments required in binary choices by removing the need for the respondent to quantify his or her values in monetary terms has potential benefits. Such simplification also incurs a cost, in that the researcher receives only a preference ordering of the goods, not quantified values. However, in some cases, such as the situation described in the introduction, a preference ordering is sufficient.

A preference ordering may also be sufficient is when it is used as input for a damage schedule. A damage schedule, as described by Rutherford, Knetsch, and Brown (1998), is a predetermined set of sanctions against resource losses. It relies on a community-based preference ordering with respect to deteriorations in environmental conditions. Only after citizens' judgments about the relative importance of a series of resource losses are obtained, and then scaled to provide an interval scale measure of importance of loss, are monetary damage payments and other sanctions mapped onto the loss scale. The damage schedule is described at the end of this section. First, we summarize how a preference ordering was obtained for coastal resources in Thailand.

5.1 Importance Judgments for Natural Resources of Phangnga Bay, Thailand

The viability of a damage schedule based on citizens' judgments of importance of loss depends critically on the reliability of those judgments. One measure of reliability is the level of agreement among different groups of respondents. Chuenpagdee, Knetsch, and Brown (2001a) tested this agreement for importance judgments from paired comparisons of individuals familiar with the natural resources of Phangnga Bay, a coastal area of southern Thailand.[15]

Phangnga Bay, like other Thai coastal regions, is rich in natural resources. Rivers flowing into the bay supply nutrients and minerals, making it an important spawning ground, nursery area, and habitat for many commercially important species including marine shrimps, lobsters, crabs, clams, Indian mackerel, and pomfret. Several species of molluscs and crustaceans inhabit the remaining old-growth stands of mangroves. During the past decade, coastal aquaculture—involving black tiger prawns, cockles, oysters, and cage culture of snapper and groupers—has joined traditional fishing and gathering as an important economic activity. Furthermore, the coast is rapidly being developed for housing, tourism, and a variety of other industries. All this growth has enhanced conflicts among resource users and increased the probability of resource losses.

The eight resource losses used in Chuenpagdee, Knetsch, and Brown's test of the paired comparison approach focused on four ecosystems at risk in Phangnga Bay: sandy beaches, mangrove forests, coral reefs, and seagrass beds.[16] The eight losses were developed from personal visits to the area, interviews of resource users and other residents, discussions with local resource managers, and the results of an extensive pre-test of the survey. The losses (listed in the left-hand column of Table 2) included two levels of damage for each of the four ecosystems.

Respondents were given information about the nature and productivity of the resource, the extent of the human-caused damage at issue, the expected changes in the level of productivity due to the losses, and the length of the likely recovery time for the resource losses where recovery was possible. For example, in the case of coral reefs, the reefs were shown on a map of the bay; the importance of the reefs to marine organisms, recreation, and natural beauty were outlined; and the possible sources of damage (pollution, sedimentation, boat anchoring, discarded fishing nets, and tourist activities) were listed. "Partial" damage was then specified as a 50 percent reduction in resource productivity of the reefs, with a recovery period of six to ten years, whereas "serious" damage was specified as productivity being reduced to almost nothing and requiring twelve to fifteen years to recover.

The respondent population, depending on the objectives of the damage schedule, may consist of persons very knowledgeable about the resources, or the general public. Two sets of respondents were sampled, experts and resource users. A comparison of the preferences of these two populations is perhaps most instructive, because it shows whether, and to what extent, they differ. The

Table 2. Scale Values[1] of Resource Losses in Phangnga Bay, with Two Measures of Consistency

Resource Loss	Experts (51)[2]	All Resource Users (170)	Fishers (45)	Shrimp Farmers (40)	Tourism (39)	Others (46)
Clear-cutting mangroves	85	83	84	81	80	84
Severe damage to coral reefs	83	76	73	76	79	76
Severe damage to mangroves	62	67	72	67	64	65
Partial damage to coral reefs	59	53	51	53	56	51
Severe damage to seagrass beds	51	42	42	41	45	41
Severe damage to sandy beaches	31	43	41	44	41	47
Partial damage to seagrass beds	24	19	18	20	23	16
Partial damage to sandy beaches	6	17	19	18	12	19
Average number of circular triads	1.5	3.7	3.5	3.9	3.5	3.9
Average coefficient of consistency	0.93	0.81	0.82	0.81	0.82	0.81

[1]Scale values are presented as the percent of the total number of times the loss could have been considered more important.
[2]Sample size in parentheses.

former included researchers, academics, administrators, and other government officials with experience and knowledge of the area and the resources at issue. The resource users were people living in the area and dependent on the resources. Four groups of resource users were sampled: (1) fishers, (2) shrimp farmers, (3) people in tourism-related businesses, and (4) others living in the area whose dependence on coastal resources was less specific. Convenience samples of these four groups of resource users in the Phangnga Bay area were selected. Sample sizes for the five groups (four resource user groups plus the experts) are listed in Table 2, as is the sample size for the entire set of resource users. Altogether 221 people were surveyed.

Use of a computer to present the pairs of losses was not practical with some of these samples, in part because of respondents' lack of familiarity with computer technology. After being given detailed descriptions of the eight losses, plus a map and a reference table summarizing the magnitudes and recovery times of the losses, each participant was given a set of paired losses, with the losses of each pair presented side-by-side on a separate half sheet of paper. The pairs were arranged in random order and the losses in each pair were randomly placed (left versus right) to randomize order effects. For each paired comparison, participants were asked to choose "the more important loss, not only to yourselves, but also to the environment, to the economic and social values of the community, and to the future of the area."

Of the 28 possible pairs of the eight losses, three were not included in the questionnaires because they compared a more severe loss to a less severe loss of the same resource; the assumed answers to these three were included in the analysis. All participants answered all 25 of the remaining paired comparison questions.

5.1.1 Reliability

The 221 respondents averaged 3.2 circular triads of a possible maximum of 20, yielding an average coefficient of consistency of 0.84. Thirty-nine percent of the respondents had no circular triads, 68 percent had 3 or less, but 9 percent had more than 10, indicating that a minority of respondents was responsible for much of the range in coefficient of consistency, as in Figure 4. As reported in the bottom two rows of Table 2, the experts were more consistent in their choices than were the user groups. The experts averaged only 1.5 circular triads per respondent, whereas the user groups averaged roughly 3.7; corresponding mean coefficients of consistency were 0.93 and 0.81, respectively.

5.1.2 Scale Values

The scale values for the eight losses, presented as the percent of the total number of times the loss could have been considered more important, are listed in Table 2 for each sample and for all resource users together.

The close correspondence of the scale values across the different samples is apparent in Table 2. Not only did resource users generally agree with the experts, but the scale values among the groups of users did not vary widely. All samples, for example, considered clear-cutting of mangrove forests to be the most important loss, followed by severe damage to coral reefs. The overall level of agreement is indicated by the correlation coefficients comparing the scale values across the five samples, which range from 0.934 to 0.999, with a median of 0.976. The general agreement about the rankings of resource losses suggests that the groups did not act strategically in favoring resources of particular interest to them.

The high level of agreement among groups supports the combination of responses from the various groups to form a single importance scale for the 221 respondents. The aggregate preference scores, expressed as before in terms of percent of times, in relation to the maximum possible number of times, the loss was considered most important (as in Table 2), are in parentheses in Figure 7.

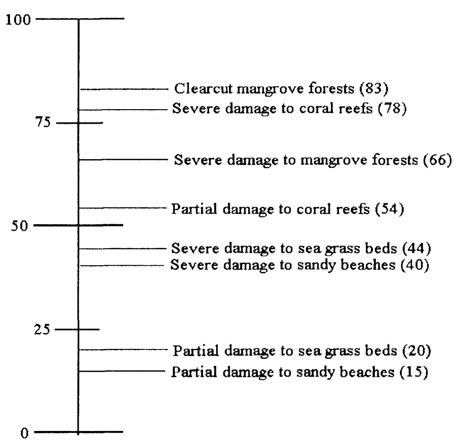

Figure 7. Importance Scale of Selected Losses in Phangnga Bay

5.2 The Damage Schedule

Once scale values are computed, they may be used to support a damage schedule. As mentioned earlier, a damage schedule is a set of sanctions, perhaps including monetary payments, that are mapped onto a preference ordering such as that of Figure 7. If monetary values are to be mapped onto the scale, the degree of measurement represented by the importance scale must be determined. If it is assumed to be only ordinal, a separate monetary amount must be specified for each loss. However, if interval properties are assumed, the mapping requires that only two points along the scale be specified in monetary terms; all other points along the scale follow from these two in a linear fashion.

The specification of sanctions is more a political than a technical task. Sanctions would be specified by the elected or appointed officials with statutory authority for protecting the public's resources, perhaps in consultation with economists and others knowledgeable about the full range of evidence on the economic value of the resources. Such a damage schedule is not intended to provide accurate monetary measures of value. However, it is based on a carefully derived community-based ordering of the importance of alternative resource losses. To the extent that the ordering provides a consistent set of comparable judgments of the importance of alternative resource losses, it may provide a well-founded basis for damage payments (Rutherford, Knetsch, and Brown 1998).

Damage schedules in general have several advantages, compared to approaches that require post-incident valuation. Two are mentioned here. First, because damage schedules specify remedies in advance rather than after an event (such as an oil spill or degradation of wildlife habitat) has occurred, they can provide more effective deterrence incentives, because those responsible for potential losses would be more fully aware of the cost to them of their actions, allowing them to undertake appropriate levels of precaution. Second, enforcement of sanctions would likely be easier and acceptance greater, because once liability is established in any particular case, the consequence is predetermined from the schedule rather than subject to the uncertainties of contentious adjudication.

Schedules, or their equivalent, are accepted in other areas in which specific assessments of the value of losses are difficult or expensive. An example is the schedule of awards for injuries used in many workers' compensation schemes.

Although the specified sums are not taken to actually reflect the value of such losses to individuals, they do reflect relative values and are therefore widely accepted and achieve many of the goals of sanctions such as efficiency enhancement. Damage schedules, or replacement tables, have also been used for environmental losses, especially for small oil spills. However, nearly all instances of such use have based sanctions on notions of replacement costs or on fairly arbitrary legislative directives rather than on an empirical assessment of community preferences regarding the importance of different losses.

A schedule for dealing with unauthorized loss is not the only use of the importance scale. It could also be used to guide further development and use of an area's resources. For example, absolute prohibitions or more onerous sanctions might be adopted to severely restrict uses that could cause losses judged to be of the highest importance, such as clear-cutting of mangrove forests and severe damage to coral reefs. Uses that could cause somewhat less serious losses, such as partial damage to seagrass beds and sandy beaches, might be subjected to somewhat less stringent restrictions or payments to discourage the loss but to allow compromise and accommodation in cases of extremely valuable alternative uses. Uses that could cause losses considered to be increasingly less serious might be made subject to successively more lenient restrictions and smaller payments. And in the cases of losses judged to be trivial, an absence or near absence of sanctions could reflect this valuation.

6. APPLICATION 2: ECONOMIC VALUATION

Economic valuation of multiple goods using the method of paired comparisons is made possible by including sums of money among the items in the choice set. Choices between a good and a monetary amount indicate which is preferred. A series of choices involving a range of carefully chosen monetary amounts allows us to estimate the break point—where the respondent switches from the good to the money. The inclusion of several different goods in the choice set potentially allows the values of each of those goods to be approximated for each respondent.

This approach has several attractive features related to the inclusion of multiple goods in the choice set. First, it produces individual respondent vectors of preference scores involving several goods, allowing multiple tests of individual reliability. Second, requiring respondents to compare several goods may increase the likelihood that respondents will think more carefully about the characteristics and relative worth of the goods. Third, including several public goods in the choice set may reduce the tendency of respondents to exaggerate the importance of a single good to which attention has been drawn. In addition, the method uses a range of monetary amounts, perhaps lessening the tendency to anchor on a given amount. However, the method faces limitations as well, related to the quantification of value implied by the good versus money choices (as described in the introduction to section 5) and to the need to describe numerous goods to each respondent. Because several goods must be described, the tendency is to shorten the descriptions, compared with contingent valuation (where typically only one good is described), so as to not over-burden the respondent. The challenge of any stated preference method, to maintain content validity by adequately specifying the item(s) of interest, is magnified when the items are numerous.

6.1 Theory

To understand what is being measured when monetary amounts are included in the choice set of a paired comparison survey, consider Figure 8, which shows three indifference curves. The horizontal axis measures quantity of the good of interest (i.e., one of the goods included in the choice set), which we will call good X, and the vertical axis measures money and all other goods. Assume in all cases that the individual is at point A along U_2, that point B falls on a higher indifference curve U_3, that point D falls on a lower indifference curve U_1, that moving from A to B represents a zero price change increase in X, that moving from A to D represents a zero price change decrease in X, and that AB=AD (see Freeman (1993) for background on the economics of quantity changes). Because the items included in the paired comparisons can be losses or gains, both WTP and WTA measures are obtainable, as will be seen shortly.

Unlike the standard measures of WTP and WTA, which are computed by holding utility constant (thus providing compensating surplus measures), the paired comparison measures involve moving to a new utility level (providing equivalent surplus measures). The two familiar economic measures for a zero price quantity change are represented in Figure 8 as BC, the maximum WTP to obtain AB of good X, and DE, the minimum WTA to give up AD of good X.

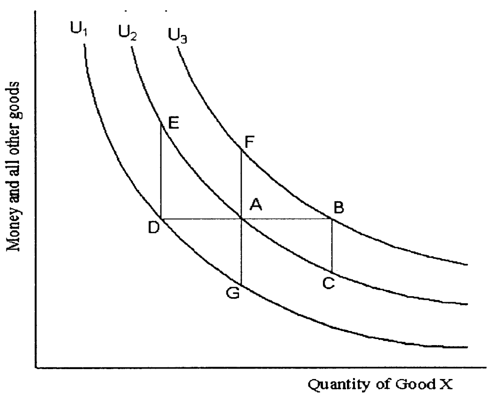

Figure 8. Measures of Economic Value

Paired comparisons offer two additional measures—a WTP measure obtained by offering the respondent a choice between losses, and a WTA measure obtained by offering the respondent a choice between gains. If offered a choice between losing AD of good X or losing various sums of money, the respondent will choose to retain the quantity of good X at small sums of money but will switch to retaining the money once the amount passes the maximum WTP to avoid losing AD of good X, which is equal to AG. If offered a choice between gaining (at no cost) AB of good X or gaining various sums of money, the respondent will choose the quantity of good X at small sums of money but will switch to taking the money once the amount passes the minimum WTA to forego AB of good X, which is equal to AF.

Equality of the two measures of WTP, or the two measures of WTA, is unlikely. For example, for WTA, DE (the standard measure) is not necessarily equal to AF (the paired comparison measure). (Of course, the standard measure of WTA assuming the individual is at point B, which is equal to AF, would be identical to the paired comparison measure of WTA assuming the individual is at point A, which is again AF.) Given the individual is at point A, the standard measure DE will equal the paired comparison measure AF only if the marginal rate of substitution (MRS) of good X for money and other goods is constant for all levels of good X at a given level of money and other goods (i.e., if the indifference curves are horizontally parallel).

In addition to effects of changing MRS, the standard measure may be affected by loss aversion (Knetsch 1989), an effect not depicted in Figure 8.

The potential effect of loss aversion is seen by noting the differences in questions put to the respondent. With WTP, the standard question asks WTP to *obtain* the good, whereas the paired comparison question essentially (though not literally) asks WTP to *keep* (to not lose) the good. If, as the loss aversion notion suggests, WTP to avoid losing a good is greater than WTP to obtain the good, the paired comparison measure will, all else being equal, exceed the standard measure. And with WTA, the standard question asks WTA to *give up* (to lose) the good, whereas the paired comparison question essen-tially (though not literally) asks WTA to *forego* (to not gain) the good. If, as loss aversion suggests, WTA to give up a good is greater than WTA to forego a good, the paired comparison measure will, all else being equal, fall below the standard measure as enhanced by loss aversion. Loss aversion causes a kink in the indifference curve at the current reference point, which causes a divergence from the expectation of standard economic theory depicted in Figure 8.

Both a diminishing MRS (of good X for money and other goods as amount of X increases at a given level of money and other goods) and loss aversion cause: (1) the paired comparison measure of WTP to exceed the standard measure, and (2) the standard measure of WTA to exceed the paired comparison measure. Because of the potential differences between the standard measures and the measures available with use of paired comparisons, we designate the paired comparison measures WTPc and WTAc, where the "c" signifies the *chooser* reference point. Thus, in the presence of either diminishing MRS or loss aversion, we would expect the magnitudes to order as follows: WTP ≤ WTPc ≤ WTAc ≤ WTA.

The sums of money included in the choice set must be chosen with care to bound the values of the goods of interest. Pre-testing may be necessary to select the monetary amounts. If the amounts are properly chosen, the goods of interest are bounded for each respondent by monetary amounts along the vector of preference scores that specify upper and lower bounds on maximum WTPc or minimum WTAc for each good.[17] For WTPc, if respondents were asked to choose the item they preferred to keep, the upper and lower bounds are the proximate sums of money of higher and lower preference score, respectively. For WTAc, if respondents were asked to choose the item they preferred to gain, the upper and lower bounds are again the proximate sums of money of higher and lower preference score, respectively.[18]

These upper or lower bounds for a good, or an interpolated value between these bounds, can be aggregated across respondents to estimate mean or median values for the good. Empirical bid curves can also be estimated for a good based simply on the proportion of respondents rejecting the good (i.e., accepting the money) at each bid level. As mentioned in section 3.3, point estimates of the value of a good may also be approximated from the aggregate preference scores, estimated using binary discrete choice methods common in dichotomous choice contingent valuation (based on respondents' choices between a good and the various sums of money), or estimated using scaling based on Thurstone's (1927c) "law of comparative judgment."

When respondents are presented with a series of paired choices, either gains or losses, they are assumed, as with contingent valuation, to take their current situation as the status quo. Respondents are assumed for each choice to be at point A in Figure 8, such that each choice is a mutually exclusive change from their current condition. Instructions to the respondent should make this clear.

People often consider choices between items and thus are familiar with the chooser reference point. Such situations may involve private items (e.g., a choice between chocolate bars) or public items (e.g., a choice between candidates, or between rapid transit options). However, economic decisions about public goods are not commonly made from this reference point. That is, citizens are not typically asked to choose between receiving a public good and receiving a sum of money, or between giving up a public good and giving up a sum of money, so such choices tend to appear distinctly hypothetical. For example, it is not common to be asked: "Which would be the greater gain to the citizens of this city, the previously described improvement in air quality or each receiving $100?"

Because of the hypothetical character of the chooser reference point when valuing pubic goods, and because numerous public goods are included in the choice set, respondents to a paired comparisons survey may not be inclined to believe that their choices will directly affect policy. Thus, use of the method of paired comparisons to value public goods may interfere with establishing the kind of "incentive compatibility" for which some contingent valuation practitioners strive (Carson, Groves, and Machina 1999).

6.2 Valuing Public Goods in Fort Collins, Colorado

Peterson and Brown (1998) made a first attempt to use paired comparisons to estimate monetary values of multiple public goods. Their study evaluated the reliability and transitivity of respondent choices; the study was not intended to provide values for benefit-cost analysis, but it does demonstrate the method.

The choice set in this experiment consists of six locally relevant public goods, four private goods, and eleven sums of money. Each respondent made 155 choices, 45 between goods and 110 between goods and sums of money. They did not chose between sums of money. Three hundred thirty students at Colorado State University, located in Fort Collins, participated in the study. Three were dropped because of incomplete data, leaving a total of 327 respondents.

The four private goods were familiar market goods with suggested retail prices: a restaurant meal not to exceed $15, a nontransferable $200 certificate for purchase of clothing, two tickets to a cultural or sporting event worth $75, and a nontransferable $500 certificate for purchase of airline tickets. The private goods were included to encourage respondents to consider a wide range of goods and trade-offs, to avoid inducing value by focusing too much attention on any one good, and to examine WTAc for familiar private goods with suggested prices.

The six public goods were of mixed types. Two of the goods—a 2,000 acre wildlife refuge in the mountains west of Fort Collins purchased by the university (Wildlife Preserve) and an improvement in the air and water of Fort Collins (Clean Arrangement)—were pure public environmental goods (i.e., environmental goods that are nonrival and nonexcludable in consumption). The remaining four public goods—a no-cost on-campus weekend festival of music and other events (Spring Festival), a no-fee service of videotapes of all course lectures (Video Service), parking garages to eliminate parking problems on campus (Parking Capacity), and expansion of the eating area in the student center (Eating Area)—were excludable by nature but stated as nonexcludable by policy. Wildlife Preserve and Clean Arrangement benefit all people in the broader community, whereas the other goods benefit only the students. Respondents had a table describing each of the goods in front of them during the experiment and were free to refer to it at any time.

The eleven sums of money were $1, $25, $50, $75, and $100 to $700 in intervals of $100. The public and private goods used in the experiment were derived from pilot studies in order to have good variation and distribution across the dollar magnitudes. Respondents were asked to choose one or the other item under the assumption that either would be provided at no cost to the respondent.

The experiment was administered on laptop computers that presented the items on the monitor in random order for each respondent. The goods had short names which appeared side-by-side on the monitor, with their position (right versus left) also randomized. The respondent entered a choice by pressing the right or left arrow key and could correct mistakes by pressing "backspace."

6.2.1 Reliability

Across the 327 respondents, the coefficient of consistency ranged from 1 to 0.51, with a mean of 0.92. As suggested by the larger median, 0.94, a few respondents were particularly inconsistent; indeed, only 10 respondents (3 percent) had a coefficient of consistency below 0.75, the midpoint of the range.[19]

6.2.2 Scale Values

Application of various scaling options—averaging interpolated values from respondents' individual vectors of preference scores, plotting empirical bid curves, interpolation based on the aggregate preference scores, discrete choice analysis,[20] and Thurstone scaling—yielded similar results. We present results for the former two, more straightforward approaches.

Table 3 displays a vector of preference scores for a typical respondent. As described above, each respondent's monetary values can be obtained for each good from the preference profile by using the lower bound, upper bound, or an interpolation between these bounds. The conservative approach is to use the lower bounds. In this case, for the respondent in Table 3, Parking Capacity would be assigned a value of $100 and Video Tape Service and Wildlife Refuge would each be assigned a value of $400. However, when the preference score of a good coincides with that of a monetary amount, the good can be assigned that amount; thus, Clothing would be assigned a value of $200. Using the linear interpolation option, Parking Capacity would be assigned a value of $150 and the Video Tape Service and Wildlife Refuge would each be assigned a value of $433. Clothing would again be assigned a value of $200.

Table 4 shows mean and median estimates for the ten goods calculated from interpolated estimates of WTAc from each respondent's preference scores. The median is the value that is acceptable to at least 50 percent of the sample

Table 3. Preference Scores of One Respondent

Item	Preference Score
$700	20
$600	18
Air travel worth $500	17
Clean arrangement	17
$500	16
Video tape service	14
Wildlife preserve	14
$400	13
$300	12
Clothing worth $200	10
$200	10
Parking capacity	9
$100	8
$75	7
$50	6
Tickets worth $75	5
A meal worth $15	5
$25	4
Eating area capacity	3
Spring festival	1
$1	0

and therefore identifies the value that a majority will accept. The mean is an estimate of the expected value of the response. The means and medians differ substantially in some cases because the distributions are highly skewed. The reader must not generalize the values reported here because they are merely illustrative and do not necessarily represent any population beyond the sample observed. To generalize such values beyond the sample requires rigorous sample design and more rigorous examination of the estimates.[21]

Table 4. Monetary Values (WTAc) Estimated by Interpolation of Individual Vectors of Preference Scores

	Median	Mean
Wildlife preserve	$400	$390
Air travel worth $500	400	382
Clean air arrangement	220	328
Clothing worth $200	150	179
Video tape service	88	190
Tickets worth $75	88	144
Parking capacity	81	165
Spring festival	63	151
Eating area capacity	36	94
A meal worth $15	17	44

Figure 9 shows empirical bid curves for Wildlife Preserve and Clothing, and another way to estimate medians. Each dot along the curves indicates the proportion of respondents who chose the good over the respective monetary amount. The straight lines connecting the dots allow estimates of the medians, which are the monetary amounts that would be rejected by 50 percent of the respondents. Using this approach, the median WTAc is $382 for Wildlife Preserve and $152 for Clothing.

Inclusion of familiar private goods with known prices in the choice set offers some information about the validity of the choices. Air Travel and Clothes, for example, have listed values of $500 and $200, respectively; their medians from Table 4 are $400 and $150, respectively. Their mean values are also lower than the suggested retail prices. The other two private goods, Tickets and Meal, have medians and means higher than their stated retail prices, of $75 and $15, respectively. The medians of the lower bounds for these two goods ($75 and $1, respectively) are equal to or lower than the retail prices.

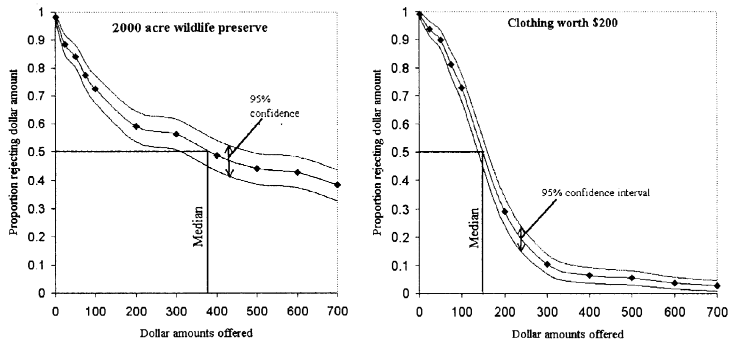

Figure 9. Empirical Bid Curves for Two Goods

Excess values for some private goods occurred for two reasons. First, some respondents sometimes actually chose the good over a dollar amount of greater magnitude than the market value of the good. Removing these respondents from the sample lowers all the median and mean values. For example, the median and mean interpolated values for the $15 Meal drop to $15 and $17, respectively, and the values of Wildlife Preserve drop to $200 and $312, respectively. Second, the good versus good choices also affect the preference scores from which the values reported in Table 4 were computed. Some respondents tended to value some private goods more highly when comparing them with public goods than when comparing them with dollar amounts, which elevates the standing of these goods in the frequency matrix relative to the dollar amounts.

A minority of the respondents appears to have placed considerably more value on the public goods than the other respondents. For example, 33 percent chose Wildlife Preserve over $700. We cannot know for sure whether such responses are valid, but the presence of such high values suggests that paired comparisons may be subject to hypothetical bias similar to that affecting dichotomous choice contingent valuation. That is, some respondents may be treating the choice as an opportunity to express attitudes towards certain kinds of goods rather than as an actual choice between a good and money.

7. FINAL COMMENTS

Multiple good valuation can provide a reliable ordering among goods, as well as a set of values that place the goods along an interval-level scale. An ordering of goods can support, for example, a ranking of environmental conditions along some dimension of interest or a ranking of desirability of equally costly public projects. An interval scale of value can support additional policy-relevant situations, such as development of a damage schedule or the grouping of goods of similar value.

For some decisions, a preference ordering will not be adequate or appropriate, and the analyst must estimate the economic value of one or more goods. Several options are available, as described elsewhere in this book. In addition to those more thoroughly tested methods, the inclusion of monetary amounts among the items to be judged in a multiple good valuation study allows an estimation of monetary values for a set of goods. Multiple good economic valuation has the advantage of encouraging respondents to compare goods in terms of their characteristics and desirability, but the validity of this procedure has not been thoroughly tested.

NOTES

1 Only by estimating the economic benefits of the preferred option and comparing them with the cost can one know whether implementing it will improve aggregate social welfare. However, benefit-cost analyses are expensive and, therefore, are not always realistic. Indeed, they may be counter productive if the cost of the analysis is substantial compared to the cost of the project. For relatively small-scale decisions, such as allocating a modest increment in funding at a national forest, something short of an economic benefit analysis may be appropriate.

2 Strictly speaking, utility is an ordinal concept and has no cardinal expected value; all that matters is that the relative positions of the $E(U)$s are maintained, indicating the order of the items. However, it is convenient to describe U as made up of a central tendency and a disturbance, which requires an assumption of an interval scale measure of utility.

3 Thurstone thought of ε as normally distributed about mean V_n. Other distributional forms for ε are feasible and commonly assumed in modern discrete choice analysis (Amemiya 1981).

4 Inconsistent responses due only to momentary fluctuations in preference are not systematically repeatable, but inconsistent responses resulting from subtle context changes are potentially repeatable. At the individual respondent level, and assuming the respondent completes the exercise only once, inconsistency due to context is indistinguishable from completely random momentary fluctuations in preference. Context dependent inconsistency is detectable only by having a respondent repeat the exercise enough times, or by combining responses across enough respondents of homogenous preferences, to accurately measure the subject's or group's choices for each pair of items. As Tversky (1969) argues, detecting systematic intransitivity in paired comparisons requires a carefully designed study. Detection is complicated because intransitive responses, like transitive responses, are subject to momentary disturbances, as in equation 1.

5 Strictly speaking, variations in U that depend on measurement effort—or, for that matter, context—are not truly random. Rather, they have observable causes and are repeatable. However, from the standpoint of the analyst who fails to notice errors or define and measure the variables causing the variation, the convention is to assume randomness.

6 Economic demand modeling, of course, requires knowing the functional form of the relation of preference to the explanatory variables, and measuring those variables.

7 A five-item choice set is used for illustrative purposes only. In practice, ranking would be the more efficient method of ordering preferences for such a small set of items. However, when the choice set contains more than about ten items, ranking becomes less effective and other methods, such as rating and paired comparisons, must be considered.

8 A response matrix with no intransitive responses can be reordered so that the upper right triangle contains all 1s and the bottom left triangle contains all 0s. In Figure 2, the order of the items would be: $i < k < j < l < m$.

9 For example, if all choices reflected the preference order $m > l > k > j > i$ except for the choice between i and k, one circular triad results and the vector of preference scores contains one three-way tie among items i, j, and k. If the only exception is between i and l, two circular triads result and the vector of preference scores contains two two-way ties (one between i and j and the other between k and l). And if the only exception is between i and m, three circular triads result and the set of preference scores contains two two-way ties (one between items i and j and the other between items l and m).

10 Switching is also expected in the case where the respondent made a mistake (e.g., pushed the wrong key) in recording the original choice.

11 A preference score difference of 0, when it occurs, reflects the full set of choices made by the respondent. It does not necessarily mean that the respondent is indifferent between the two items. When the respondent is not indifferent, the likelihood of the respondent making different choices on different occasions, and thus of switching, is something less than 0.5.

12 Space does not allow us to fully develop this point here, but we have verified it via extensive testing using a model that simulates choices assuming the random utility function of equation 1. See also Torgerson (1958).

13 Thurstone's approach was only the first attempt to improve upon the use of aggregate preference scores to summarize paired comparisons. See David (1988) for other methods.

14 Such an observation is possible when monetary amounts are included among the items to be compared. A nonsensical choice would be one indicating a value for the good well above its market price. Although respondents may have a maximum willingness to pay in excess of the market price, they would not choose the good much above a sum of money equal to its market price because they could always choose the money and then buy the good.

15 A similar application for another Thai coastal area is found in Chuenpagdee, Knetsch, and Brown (2001b). Resource damaging activities as well as resource losses were assessed in this study.

16 A set of eight items is slightly smaller than we would now recommend, as mentioned in section 4.

17 An exception occurs if a respondent would choose any amount of money, no matter how small, over the good (precluding a monetary amount with a lower preference score), or would choose the good over any amount of money (precluding a monetary amount of higher preference score). In the former case, the lower bound is presumed to be $0 unless the good is in fact a bad.

18 Alternatively, for WTPc, if respondents were asked to choose the item they preferred to give up, the upper and lower bounds are the proximate sums of money of lower and higher preference score, respectively. And for WTAc, if respondents were asked to choose the item they preferred to forego, the upper and lower bounds are the proximate sums of money of lower and higher preference score, respectively.

19 As explained in Peterson and Brown (1998), after presenting the 155 paired comparisons, the computer program administering the experiment selected those pairs for which the respondent's choices were not consistent with his or her dominant preference order established from all 155 choices. The computer also randomly selected ten consistent pairs. These two sets of selected pairs were randomly ordered and presented again to the respondent, with no break in the presentation, so that the respondent was unlikely to notice when the 155 original pairs ended and the repeats began. For the originally inconsistent pairs, respondents switched many of their earlier choices, suggesting that respondents' earlier choices were mistakes or that their preferences became more defined in the course of considering the various pairs. When the original choices for the originally inconsistent pairs are replaced by the new choices, the number of circular triads drops dramatically, causing both the mean and median coefficient of consistency to rise to 0.99, compared with 0.92 and 0.94, respectively.

20 Discrete choice analysis relies on the assumption of independent and identically distributed error terms. This assumption was not met in the Peterson and Brown experiment, especially for the choices involving dollar amounts and private goods. As indicated in Figure 9, later in this section, the variances of the disturbance distributions for private goods were relatively narrow compared with those of the public goods.

21 Updating the choices for the originally inconsistent pairs, as described in footnote 19, had only a small effect on the estimated monetary values. For example, using the method of interpolation of individual vectors of preference scores, updating lowered the mean values an average of $6 per good, compared with those for the original choices reported in Table 4.

REFERENCES

Ben-Akiva, M., and S. R. Lerman. 1985. *Discrete Choice Analysis: Theory and Application to Travel Demand.* Cambridge, MA: Massachusetts Institute of Technology. 390 p.

Boyle, K. J., F. R. Johnson, and D. W. McCollum. 1997. Anchoring and Adjustment in Single-bounded, Contingent-valuation Questions. *American Journal of Agricultural Economics* 79(5):1495-1500.

Breffle, W. S., and R. D. Rowe. 2002. Comparing Choice Question Formats for Evaluating Natural Resource Tradeoffs. *Land Economics* 78(2):298-314.

Brown, T. C. 1984. The Concept of Value in Resource Allocation. *Land Economics* 60(3):231-246.

Brown, T. C., G. L. Peterson, A. Clarke, and A. Birjulin. 2001. Reliability of Individual Valuations of Public and Private Goods. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station.

Brown, T. C., D. Nannini, R. B. Gorter, P. A. Bell, and G. L. Peterson. 2002. Judged Seriousness of Environmental Losses: Reliability and Cause of Loss. *Ecological Economics* 42(3):479-491.

Buhyoff, G. J., and W. A. Leuschner. 1978. Estimating Psychological Disutility from Damaged Forest Stands. *Forest Science* 24 (3):424-432.

Carson, R. T., T. Groves, and M. J. Machina. 1999. Incentive and Informational Properties of Preference Questions. Paper presented at the Meeting of the European Association of Resource and Environmental Economists, Oslo, Norway: June, 1999.

Chuenpagdee, R., J. L. Knetsch, and T. C. Brown. 2001a. Environmental Damage Schedules: Community Judgments of Importance and Assessments of Losses. *Land Economics* 77(1):1-11.

Chuenpagdee, R., J. L. Knetsch, and T. C. Brown. 2001b. Coastal Management Using Public Judgments, Importance Scales, and Predetermined Schedule. *Coastal Management* 29:253-270.

David, H. A. 1988. *The Method of Paired Comparisons.* Second ed. Vol. 41, *Griffin's Statistical Monographs and Courses.* New York, NY: Oxford University Press. 188 p.

Dunn-Rankin, P. 1983. *Scaling Methods.* Hillsdale, NJ: Lawrence Erlbaum Associates. 429 p.

Fechner, G. T. 1860. *Elemente der Psychophysik.* Leipzip: Breitkopf and Hartel.

Ferber, R. 1974. *Handbook of Marketing Research.* New York: McGraw-Hill. 1,358 p.

Freeman, III, A. M. 1993. *The Measurement of Environmental and Resource Values: Theory and Methods.* Washington, D.C.: Resources for the Future. 516 p.

Green, D., K. E. Jacowitz, D. Kahneman, and D. McFadden. 1998. Referendum Contingent Valuation, Anchoring, and Willingness to Pay for Public Goods. *Resource and Energy Economics* 20:85-116.

Guilford, J. P. 1954. *Psychometric Methods.* Second ed. New York: McGraw-Hill. 597 p.

Kendall, M. G., and B. B. Smith. 1940. On the Method of Paired Comparisons. *Biometrika* 31:324-345.

Knetsch, J. L. 1989. The Endowment Effect and Evidence of Nonreversible Indifference Curves. *American Economic Review* 79(5):1277-1284.

McFadden, D. 1974. Conditional Logit Analysis of Qualitative Choice Behavior. Pages 105–142 in *Frontiers in Econometrics*, edited by P. Zarembka. New York: Academic Press. 252 p.

Nunnally, J. C. 1976. *Psychometric Theory.* New York: McGraw Hill. 640 p.

Payne, J. W., J. R. Bettman, and E. J. Johnson. 1992. Behavioral Decision Research: a Constructive Processing Perspective. *Annual Review of Psychology* 43:87-131.

Peterson, G. L., R. L. Bishop, and R. M. Michaels. 1973. Children's choice of playground equipment: Developing methodology for integrating preferences into environmental engineering. *Journal of Applied Psychology* 58:233-238.

Peterson, G. L., and T. C. Brown. 1998. Economic Valuation by the Method of Paired Comparison, with Emphasis on Evaluation of the Transitivity Axiom. *Land Economics* 74(2):240-261.

Rutherford, M. B., J. L. Knetsch, and T. C. Brown. 1998. Assessing Environmental Losses: Judgments of Importance and Damage Schedules. *Harvard Environmental Law Review* 22:51-101.

Slovic, P. 1995. The Construction of Preference. *American Psychologist* 50(5):364-371.

Thurstone, L. L. 1927a. A Law of Comparative Judgment. *Psychology Review* 34:273-286.

Thurstone, L. L. 1927b. The Method of Paired Comparisons for Social Values. *Journal of Abnormal Social Psychology* 21:384-400.

Thurstone, L. L. 1927c. Psychophysical Analysis. *American Journal of Psychology* 38:368-389.

Torgerson, W. S. 1958. *Theory and Methods of Scaling.* New York, NY: John Wiley & Sons. 460 p.

Tversky, A. 1996. Contrasting Rational and Psychological Principles in Choice. In *Wise Choices: Decisions, Games, and Negotiations*, edited by R. J. Zeckhauser, R. L. Keeney and J. K. Sebenius. Boston, MA: Harvard Business School Press.

Tversky, A., S. Sattath, and P. Slovic. 1988. Contingent Weighting in Judgment and Choice. *Psychological Review* 95(3):371-384.

Chapter 8

INTRODUCTION TO REVEALED PREFERENCE METHODS

Kevin J. Boyle
University of Maine

Revealed preference methods draw statistical inferences on values from actual choices people make within markets. Estimation of the values people place on environmental amenities and disamenities proceeds by specifying a theoretical framework and conducting data analyses from purchase decisions (prices paid and quantities purchased) according to this conceptual framework. Four commonly used revealed preference methods are discussed in the next three chapters (Chapter 9, travel cost; Chapter 10, hedonics; Chapter 11, defensive behavior and damage costs).

Travel cost models are typically used to estimate use values for recreation activities and changes in these use values associated with changes in environmental quality. Hedonic models, generally property value models, are used to infer the premium that households pay to purchase a property near an environmental amenity or away from an environmental disamenity. Models of defensive behavior focus on expenditures that people make to reduce exposure to disamenities or to offset adverse effects of exposure. Damage cost methods measure the resource costs caused by environmental contamination. Defensive behavior and damage cost methods are typically applied to value health effects of pollution, in which case the damage cost method is known as the cost of illness.

The key difference between these revealed preference methods and the stated preference methods previously discussed is the types of data used to estimate values. Stated preference methods rely on data from carefully worded survey questions asking people what choice they would make for alternative levels of an environmental amenity (intended behavior). Revealed preference methods rely on data that record people's actual choices (revealed behavior).

From a conceptual perspective, stated preference methods can provide estimates of Hicksian surplus, whereas revealed preference methods typically provide estimates of Marshalian surplus (Freeman 1993). The exception is when utility-theoretic models are estimated in order to derive exact (Hicksian) measures of welfare. Even with utility theoretic formulations, no one has developed an established means of using revealed preference methods to estimate nonuse values. Estimation of nonuse values remains solely in the domain of stated preference methods.

1. CLASSIFYING METHODS

Recognizing that any categorization of methods is to some degree arbitrary and is not acceptable to everyone, the classification proposed here is designed to highlight key features and differences in the revealed preference methods to be discussed. Although each of the methods can be used to estimate values for environmental amenities, the conceptual frameworks, data, and applications differ substantially.

Travel cost, hedonics, and defensive behavior share a common feature: values are inferred from individual or household choices (Table 1). Travel cost models are based on decisions to visit recreation sites that differ in travel cost and quality. Hedonic models are based on decisions to purchase a house from among choices that have different levels of attributes, including environmental amenities and disamenities. Defensive behavior models are based on expenditures that households make to avoid exposure to an environmental disamenity. Individual or household decisions are not the basis for damage cost calculations. The cost of illness, for example, is simply a summation of the direct (e.g., doctors visits and medicine) and indirect (e.g., lost time of work) costs of treating an environmentally-induced illness. These costs result from a combination of individual and societal decisions.

Both travel cost and defensive behavior arise from a household production framework whereby people combine market goods that they purchase and their own time to produce a desired outcome. In a travel cost model, people combine

Table 1. Revealed Preference Valuation Methods.

Method	Revealed Behavior	Conceptual Framework	Types of Application
Travel Cost	Participation in recreation activity and site chosen	Household production, weak complementarity	Recreation demand
Hedonics	Property purchased or choice of employment	Demand for differentiated goods	Property value and wage models
Defensive Behavior	Expenditures to avoid illness or death	Household production, perfect substitutes	Morbidity/mortality
Cost of Illness	Expenditures to treat illness	Treatment costs	Morbidity

the travel costs and other marginal costs of participating in a recreation activity with their travel time to produce a recreation experience, a fishing trip for example. The travel and time costs comprise the implicit price of participation. It is presumed that people will pay more for a higher quality recreation experience. With defensive behavior, people also combine purchased inputs with time and the outcomes are improved health or well being from reductions in exposure to a disamenity. The purchased inputs and time comprise the implicit price of improved health.

Despite these similarities, the theoretical foundations of travel cost models and defensive behavior models differ. Travel cost is based on the premise of weak complementarity (Maler 1974). Weak complementarity restricts the consumption of an environmental amenity to zero when consumption of a related market good is zero. With a travel-cost model, the quantity (x) of recreation and the quality (q) of recreation directly enter the utility function, which is specified as U(Y,x,q), where Y is a vector of all other goods and services consumed. Assuming the price of visits (P_x) represents the full marginal cost of participation (purchased inputs and the opportunity cost of travel time), then the budget constraint is specified as $I = P_Y Y + P_x x$. This specification allows the estimation of a demand or utility function from which surplus estimates can be derived.

With defensive behavior, neither the purchased inputs nor environmental quality are assumed to directly enter the utility function. Rather, x and q are assumed to be substitutes in the production of a good (z) that enters the utility function as U(Y,z) and z = f(x,q) (Smith 1992). Consider the case of contaminated drinking water where z would be the health effects from consumption. If q is the level of contamination in the tap water, and x is the quantity of bottled water purchased to avoid exposure to q, then an improvement in q such that it is no longer necessary to purchase bottled water results in P_x being a lower bound estimate of the marginal value of improving water quality. This is a lower bound estimate because people may not be able to fully avert or there may be inconvenience in defensive actions that is not captured in the monetary cost.

Hedonic price models assume that a property is a heterogeneous good whose component attributes yield utility (Rosen 1974). The attributes can be represented by the vector $Z = (z_1, z_2, z_3, \ldots, z_n)$ such that $U(Y,Z) = U(Y, z_1, z_2, z_3, \ldots, z_n)$. Thus, the price paid for a property is a function of the attributes where $P_Z = f(z_1, z_2, z_3, \ldots, z_n)$, which is commonly referred to as the hedonic price function. One or more of the attributes of the property may be an environmental amenity (z_e) with the implicit price $\partial P_Z/\partial z_e = p_e$. While hedonic models are generally estimated for developed property with structures, the value of the environmental amenity is actually embedded in the price of the land. Under specific conditions described in Chapter 10, estimated implicit prices can be used to estimate the demand for environmental amenities, which allows welfare estimation for non-marginal changes in quality.

Cost of illness does not include any estimate of consumer surplus or marginal prices. The cost of illness method simply attempts to measure the full cost of illness, including treatment costs. Treatment costs are based on individual and societal decisions about the level of care to be provided. The presumption is that if a certain level of care is provided, then it must be worth at least that much to society. No inferences are made based on individual or household decisions. From an economic perspective, cost of illness is a less desirable benefit measure than the other three methods that are firmly grounded in consumer choice and welfare theory. Cost of illness is discussed in Chapter 11 because it is the method most commonly used to estimate the benefits of improvements in health.

2. APPLICATIONS

There are complementary and substitute applications of each of the revealed preference methods. Consider the case where a river is polluted with toxic materials. Both travel cost and hedonic models may be appropriate, but for different affected groups. Travel cost could be used to estimate the benefits of cleaning up the river for people who travel to the river to recreate, while a hedonic model could be used to estimate the benefits to people who own property along the river. In this type of application, two revealed preference methods are complementary in the estimation of the benefits of cleaning up the river; the benefit estimates are additive in the calculation of the total benefits of cleaning up the river. This result holds because the methods have been used to estimate values for different affected groups. If applications are not mutually exclusive, double counting of benefits will occur (McConnell, 1990).

Now consider the case of adverse health effects caused by air pollution. In this case, a hedonic model, defensive behavior, and cost of illness are all approaches to estimating the same benefits of cleaning up air pollution. Here a choice must be made regarding which approach will provide the best estimates of benefits for policy analyses.

The key to using the various revealed preference methods is to identify first the change to be valued and then the affected groups. It is then possible to determine whether various approaches are complementary or substitutes, and if more than one approach should be used. Although situations like the river example clearly call for the use of multiple methods applied to different affected groups, other situations might dictate the use of multiple methods even when they are substitutes. Multiple revealed preference methods might be employed to investigate convergent validity of the resultant welfare estimates (Carmines and Zeller 1979) or perhaps to bound the welfare estimate. For example, defensive behavior is generally accepted as a lower-bound estimate, but welfare estimates from a hedonic model can also include the values of additional undesirable features that are perfectly correlated with the environmental disamenity of interest, which would provide an upper bound. In such a case, the defensive and hedonic estimates may be assumed to bound the true marginal value of improving quality.

3. IMPLEMENTATION STEPS

The basic steps in a revealed preference study are the same as for a stated preference study (steps 1, 2, and 3; Table 2).

During step 1, the valuation objective is identified so that policy-relevant welfare estimates will be obtained. For example, if a policy involves a program to protect drinking water from contaminants, then it is necessary to identify the precise change in water quality to be valued. Next (step 2), it is necessary to identify whose values will be estimated. If the program is designed to protect people on private wells, then people who obtain their water from private wells that are affected by the contaminant make up the relevant population. Using information from steps 1 and 2, step 3 involves developing a theoretical definition of the value that households place on reducing contaminants in their drinking water. This is done using the basic welfare precepts discussed in Chapter 2. Any concerns about the values for specific subpopulations (e.g., children and women of child-bearing age) would need to be considered in the formulation of the theoretical model.

Step 4 is the choice of a revealed preference method. Continuing with the drinking water example, both hedonic and defensive-behavior models are logical choices and might be employed to bound the theoretically correct value. The defensive behavior model may be a more desirable approach if there were interest in subpopulations, because information about the household composition of the sellers and buyers of properties in the study for a hedonic model may be unknown.

Hedonic models and cost-of-illness calculations are typically based on secondary data (steps 5 and 6). Data sources must be identified and data evaluated once obtained. For example, property sales data may not be comprised solely of arms-length transactions. Some intra-family sales for very modest costs of $1 or $100 may exist. These types of sales would be excluded from the data used to estimate the hedonic price function. The credibility of value estimates always relies on careful attention to data quality.

Multiple sources of secondary data may be required. Sales prices and housing characteristics may be obtained from one source, environmental quality

Table 2. Process of Estimating Values Using Revealed-Preference Methods

1.	Identify change(s) in quantity or quality to be valued.
2.	Identify population whose values are to be estimated.
3.	Develop theoretical definition(s) of value(s) to be estimated.
4.	Select revealed preference valuation method(s).
5.	Identify appropriate sources of secondary data.
6.	Obtain secondary data and check the coding of the data.
7.	Determine if any primary data are needed.
8.	If primary data are needed, design survey instrument and collect the data.
9.	Estimate model(s).
10.	Derive welfare estimate(s) from estimated model(s).

data from another source, demographic characteristics of neighborhoods from census data, and spatial distribution of amenities from GIS data. Merging multiple data sources can require considerable time and attention to detail.

Travel cost and defensive-behavior models typically require the collection of primary data (steps 7 and 8). The data collection process requires the selection of a survey mode, sampling frame, and survey design (described in Chapter 3). Although careful survey design is required so that appropriate survey data are collected, the level of detail is less than would be required in the design of a survey to elicit stated preferences. A common problem with these revealed preference data is recall bias that arises when people are asked to recall recreation trips or defensive expenditures over a long time period (Chu et al. 1992). Travel cost and defensive behavior applications also generally require secondary data on environmental quality that are used in conjunction with primary data on individual choices to estimate models and derive value estimates.

Finally given the theoretical definition of value and the data collected, econometric models are estimated from which welfare estimates are derived. Here, the key issue is the selection of the model to be estimated. This is particularly true for travel cost and hedonics, because different functional specifications of recreational demand or the hedonic price function dramatically affect welfare estimates (Smith and Huang 1995; Smith and Kaoru 1990)

4. LIMITATIONS OF REVEALED PREFERENCE METHODS

A key limitation of revealed preference methods is the inability to estimate nonuse values. It is not necessary to exhibit any type of use behavior to hold nonuse values, which implies that revealed preference estimates of nonuse values are not feasible. This conclusion has been challenged by Larson (1992), but his challenge has not received any substantial credence in the literature.

Another limitation of revealed preference methods is the inability to estimate values for levels of quality that have not been experienced. Revealed behavior for new levels of environmental quality may not exist in many policy applications. For example, in terms of long-term contamination, people may not have had the opportunity to exhibit choice behavior for the new condition that arises after remediation. Likewise, people may not have experienced the degraded quality associated with a sudden and substantial environmental accident. Or in the case of mercury deposition, all lakes may have fish with elevated levels of mercury in their tissue so that anglers have no lakes without contaminated fish to choose. In these instances, models have been estimated using pooled stated preference and revealed preference data (Adamowicz, Louviere, and Williams 1994; Cameron 1992). This approach involves surveying people about how they would behave with the new level of environmental quality; their survey responses are pooled with their actual choices for existing qualities to estimate a joint model. Thus, while lack of actual choices is a limitation for standard revealed preference methods, estimation of models that combine behavioral data with stated preference data is an important frontier issue in the estimation of nonmarket values.

Revealed preference methods face other limitations. For example, random utility models are typically estimated using single-day trips to a site, and multiple-day visits are not modeled. It is also the case that the measurement of travel costs and accounting for multiple-destination trips for travel cost models remain unresolved issues. Alternatively, people may not have full information when they make purchase decisions. For instance, the lack of knowledge of groundwater contamination would severely compromise the use of a hedonic model.

Revealed preference methods are generally accepted by economists because they are based on actual decisions people have made. However, a quality study requires careful attention to the detailed design steps presented in each to the following three chapters. Even a good design requires a cognizance of the limitations of each method and the assumptions made. These considerations directly affect the usefulness of value estimates in specific policy applications.

REFERENCES

Adamowicz, W. L., J. Louviere, and M. Williams. 1994. Combining Revealed and Stated Preference Methods for Valuing Environmental Amenities. *Journal of Environmental Economics and Management* 26:271-292

Cameron, T. A. 1992. Combining Contingent Valuation and Travel Cost Data for the Valuation of Nonmarket Goods. *Land Economics* 68:302-317.

Carmines, E. G., and R. A. Zeller. 1979. *Reliability and Validity Assessment.* Beverly Hills, CA: Sage Publications.

Chu, A., D. Eisenhower, M. Hay, D. Morganstein, J. Neter and J. Waksberg. 1992. Measuring Recall Error in Self-Reported Fishing and Hunting Activities. *Journal of Official Statistics* 5:13-39.

Freeman, A. M. III. 1993. *The Measurement of Environmental and Resource Values: Theory and Methods.* Washington, DC: Resources for the Future.

Larson, D. 1992. Further Results on Willingness to Pay for Nonmarket Goods. *Journal of Environmental Economics and Management* 23:101-122.

Rosen, S. 1974. Hedonic Prices and Implicit Markets: Product Differentiation in Perfect Competition. *Journal of Political Economy* 82:34-55.

Smith, V. K. 1992. Household Production Functions and Environmental Benefit Estimation. In *Measuring the Demand for Environmental Quality.* Edited by J. B. Braden and C. D. Kolstad. Amsterdam: North-Holland.

Smith, V. K., and J. Huang. 1995. Can Markets Value Air Quality? A Meta-Analysis of Hedonic Property Value Models. *Journal of Political Economy* 103:209-227.

Smith, V. K., and Y. Kaoru. 1990. Signals or Noise? Explaining the Variation in Recreation Benefits Estimates. *American Journal of Agricultural Economics* 72:419-433.

Chapter 9

THE TRAVEL COST MODEL

George R. Parsons
University of Delaware

1. INTRODUCTION

The travel cost model is used to value recreational uses of the environment. For example, it may be used to value the recreation loss associated with a beach closure due to an oil spill or to value the recreation gain associated with improved water quality on a river. The model is commonly applied in benefit-cost analyses and in natural resource damage assessments where recreation values play a role. Since the model is based on observed behavior, it is used to estimate use values only.

The travel cost model is a demand-based model for use of a recreation site or sites. A site might be a river for fishing, a trail for hiking, a park for wildlife viewing, a beach for swimming, or some other area where outdoor recreation takes place. It is useful to separate travel cost models by those that estimate demand for a single site and those that estimate demand for many sites.

Single-site models work like conventional downward sloping demand functions. The "quantity demanded" for a person is the number of trips taken to a recreation site in a season and the "price" is the trip cost of reaching the site. Variation in price is generated by observing people living at different distances from the site. Price is low for people near the site and high for those living farther away. The demand function slopes downward if trips decline with distance to the site.

Single-site models are useful when the goal is to estimate the total use or "access value" of a site. The elimination of a site is the usual application. The lost value is the total consumer surplus under the single-site demand function—the difference between a person's total willingness to pay for trips and the actual trip cost incurred over a season. Some valuation examples include:

- A beach closure due to an oil spill
- A fish consumption advisory that closes a lake for fishing, or
- A development that eliminates a natural area for wildlife viewing.

It is also possible to use a single-site model to estimate the value associated with a change in the cost of access to a site. For example, an increase in an entry fee or the opening of a new entrance could be evaluated using a single-site model. More difficult, and really calling for the transfer of a single-site model from one site to another, is valuing the addition of a new site such as a reservoir created by a dam.

There are some variations of the single-site model that can be used for valuing changes in site characteristics such as improved water quality on a lake or an increase in the number of hiking trails in a wilderness area. However, this is not the strength of the single-site model. When the goal is to value changes in site characteristics at one or more sites or to value the access to more than one site simultaneously, a multiple-site model is preferred.

The random utility maximization (RUM) model is the most widely used multiple-site model. A travel cost RUM model considers an individual's discrete choice of one recreation site from a set of many possible sites on a single choice occasion in a season. The choice of site is assumed to depend on the characteristics of the sites. For example, a person making a fishing trip may consider trip cost, catch rate of fish, and site amenities. The choice of site implicitly reveals how a person trades off one site characteristic for another. Since trip cost is always included as one of the characteristics, the model implicitly captures trade-offs between money and the other characteristics.

Some examples of valuing changes in site characteristics (at one or more sites) using the RUM model include:

- improvement in water quality on lakes and rivers,
- increase in the catch rate of fish on lakes and rivers,
- improvement in the conditions of access to several local parks,
- increase in moose populations in several hunting areas,
- increase in the number of mountain biking trails in a state park.

The RUM model may also be used to value access to one or more sites simultaneously. For instance, it may be used to value the loss of several beaches closed because of swimming advisories or to value several ski slopes opened in a development project.

This chapter is organized into two sections. The first covers the single-site model; the second covers the RUM model.[1] Both are organized around a table listing the steps required to estimate a basic version of each model. The emphasis is placed on how one applies a modern version of the model, not on the frontiers of modeling. For more on the frontiers and recent developments see Herriges and Kling (1999) or Phaneuf and Smith (2002). For a good historical account of travel cost models, which were first applied more than 50 years ago, see Ward and Beal (2000).

2. THE SINGLE-SITE MODEL

This section is divided into four parts: basic model, steps in estimation, an example, and variations. I will focus on using the single-site model to value site access since that is the strength and most common application of the model.

2.1 Basic Model

The single-site model is a demand model for trips to a recreation site by a person over a season. The quantity demanded is the number of trips a person takes to the site. The price is the trip cost of reaching the site, which includes a person's travel expenses and time cost necessary to make the trip possible. In its simplest form the single-site model is

$$(1) \quad r = f(tc_r)$$

where r is the number of trips taken by a person in a season to the site and tc_r is the trip cost of reaching the site. Like any demand function, one expects a negative relationship between quantity demanded (trips r) and price (trip cost tc_r). People living closer to the site face a lower cost of reaching the site and, all else constant, probably take more trips.

Trip costs alone will not explain an individual's demand for recreation trips. Demand will also depend on factors such as income, age, experience in the recreation activities available at the site, and proximity to other recreation sites. Equation (1) can be respecified to include these factors

$$(2) \quad r = f(tc_r, tc_s, y, z)$$

where tc_s is a vector of trip costs to other recreation sites, y is income, and z is a vector of demographic variables believed to influence the number of trips.

By incorporating trip cost to other sites, the model now accounts for substitutes. The tc_ss are the prices of trips to substitute sites. If a person lives near a substitute site, the number of trips (r) is likely to decline as the person substitutes trips to that site and away from the site of interest in the analysis. A positive coefficient on tc_s would pick up this substitution effect. The other shifters work much as one would expect. A positive coefficient on income, for example, implies that the number of trips taken increases with income, and so forth.

Figure 1 shows a linear version of equation (2) corresponding to

$$(3) \quad r = \beta_{tc_r} tc_r + \beta_{tc_s} tc_s + \beta_y y + \beta_z z$$

where the β_i's are coefficients to be estimated. If a person faces a trip cost of tc_r^0 in this model, he or she takes r^0 trips. The area A is his or her total consumer surplus for trips to the site during the season—the difference between total willingness to pay for trips (area A+B) and total trip cost (area B). This is also called the individual's access value for the site. If the site were closed for a season, the individual would lose access to the site and consequently the

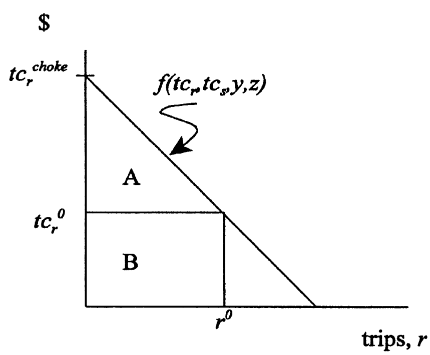

Figure 1. Access Value in a Linear Single-site Model

area A. A more general expression for consumer surplus or access value, which applies to any functional form, is the area under the demand curve between an individual's current price and the choke price. The choke price is the price at which trips fall to zero in the model. Mathematically the consumer surplus is

$$(4) \quad \Delta w = \int_{tc_r^0}^{tc_r^{choke}} f(tc_r, tc_s, y, z)dtc_r$$

where tc_r^{choke} is the choke price and $tc_r^{\,0}$ is the individual's trip cost.

In application, one seeks to estimate an equation like (3) using visitation data to a site. Table 1 shows the form of a typical single-site data set— trip count, trip cost, and demographic data for a sample of individuals. The data are gathered by survey. Once assembled, equation (3) is estimated by regressing trips (r) on the relevant explanatory variables in the data set (tc_r, tc_s, y, z_1, z_2). After the model is estimated, the parameters are used to compute access value (area A) for each individual in the sample. Means are computed and an estimate is extrapolated to the population. (The model is usually not linear but the principle is the same regardless of functional form.) The next section describes these steps in detail. For a derivation of the basic model from utility theory or from a household production function, see Freeman (1993) or Bockstael (1995).

Table 1. Typical Form of a Data Set for Estimating a Single-site Model

Observation Number	Number of Trips to the Site for the Season	Trip Cost to Site *$*	Trip Cost to Substitute Site *$*	Annual Household Income *$000*	Years Engaged in this Form of Recreation	Number of Children
	(r)	(tc_r)	(tc_s)	(y)	(z_1)	(z_2)
1	7	45	200	45	17	0
2	1	150	20	100	2	2
3	21	12	65	77	22	1
.	.	.	.	.	.	.
.	.	.	.	.	.	.
N	3	98	25	55	4	1

2.2 Steps in Estimation

The steps are shown in Table 2. I will discuss each of these in a separate subsection. Assume the model is being estimated for the purpose of valuing site access, that the analyst is using day-trip data only, and that the data are individual-based. This is the most common application of the model. Valuing quality changes at a site, as already noted, is not the strength of the single-site model, and incorporating overnight trips, while possible, is less common and presents special complications.

Following this section is an actual single-site application by Brent Sohngen (2000). Each step of the analysis for a model of beach use for two Lake Erie beaches is described.

Table 2. Steps in Estimating a Single-site Model

Step 1	Define the Site to be Valued
Step 2	Define the Recreation Uses and the Season
Step 3	Develop a Sampling Strategy
Step 4	Specify the Model
Step 5	Decide on the Treatment of Multiple Purpose Trips
Step 6	Design and Implement the Survey
Step 7	Measure Trip Cost
Step 8	Estimate the Model
Step 9	Calculate Access Value

2.2.1 Define the Site to be Valued

The analysis begins with a definition of the site to be valued. The site might be a park, lake, beach, wilderness area, river segment, or some other area used for outdoor recreation. The boundaries are often easy to delineate, such as in the case of a park, lake, or wildlife reserve. In other cases the delineation is not so clear. A river for fishing or white-water rafting, for example, calls for the analyst to define the length of the segment to be analyzed. A site for hunting or ocean fishing usually requires researcher-defined boundaries as well. In these cases, one seeks to define a site broadly enough to include most, if not all, of the area affected by the policy being analyzed.

Government agencies, park services, and tourist bureaus often have literature and maps, which help in defining a site. Sometimes government agencies charged with managing natural resource use have their own definitions for sites, such as wildlife management units. These site definitions often work for a travel cost model.

2.2.2 Define the Recreation Uses and the Season

In some cases the site will have a single or dominant recreation use such as fishing or beach visitation. In other cases there will be multiple uses such as fishing, swimming, boating, and viewing on a lake. Ideally one would include all recreation types and estimate separate demand functions for each. Sometimes, policies call for focusing on a single recreation type. Again, government agencies, park services, and tourist bureaus often have information and data to help identify major uses.

If recreation types are similar enough, they may be aggregated. For example, if there are many types of boating, one might treat these as a single group. The more similar the recreation types, the less problematic the aggregation; aggregating sail-boating and motor-boating is less problematic than aggregating motor-boating and swimming. Most studies have some aggregation. Beach use— which can include sunbathing, swimming, surfing, jogging, and even surf fishing—is often treated as a single recreation type. Aggregation simplifies data collection and analysis. It means less information from each respondent, fewer observations, and less modeling. Again, one must be careful not to bundle recreation types that are too dissimilar.

Since individuals are sometimes observed engaging in more that one type of recreation on a single visit, a common practice is to identify the primary purpose of the recreation trip and classify the use accordingly. For example, one might ask respondents in a survey to report the number of trips taken "*primarily* for the purpose of fishing," and then "*primarily* for the purpose of boating," and, so on.

Along with defining the uses of the site one must also define the season for each use. Hunting and fishing may have a season defined by law. Skiing, rock climbing, and beach use will have a season defined by periods of favorable weather. Others uses, such a viewing and hiking, may be year-round.

2.2.3 Develop a Sampling Strategy

Next, a strategy is developed to sample the users and potential users of the site. There are essentially two approaches: on-site and off-site sampling.

In on-site sampling, recreationists are intercepted at the site and asked to complete an oral or written survey. The survey may be completed on the spot or handed to people to mail to a specified address at a later time. It is even conceivable to recruit respondents on-site and later mail a survey to their home. In this case, the recreationist reports only an address at the site.

On-site samples have the advantage of hitting the target population directly. Every person interviewed has visited the site. Compare this to a random survey of the general population where the percent of people who have visited the site in the current season is likely to be quite small for most types of outdoor recreation. In this case, the number of contacts or interviews required to get a reasonable sample of users can be rather large. For this reason, on-site samples are popular for single-site studies.

There are a number of issues to be aware of when using on-site samples. First, people who do not visit the site are missed. This implies a sample with no observations taking zero trips. This compromises the accuracy of the estimated intercept, the "choke price", for the demand function. Think of a scatter of data points being used to estimate the demand function in Figure 1. With on-site sampling the data are truncated at one trip. We have no scatter of data points at zero trips and are forced to estimate the intercept using trip data over individuals having taken one or more trips. This is extrapolating outside the range of the observed data.

Second, on-site samples can be difficult to conduct in such a way that a random sample of users is obtained. Think about randomly drawing a user on a beach. When and where do you sample? A strategy must be devised. For example, randomly drawing several week days and weekends during the season for interviewing and then interviewing every tenth person is a strategy that attempts to approximate a random sample. Clear entry points, such as gates, help in on-site random sampling. Consideration must also be given to how one conducts a survey on-site. Interrupting someone as they sleep on the beach or as they put a boat in the water is hardly advisable. Catching respondents at an opportune time with minimum disruption will help response rates and extend common courtesy. Another consideration is whether or not one samples people as they arrive or depart. The latter has the advantage that respondents know more about the actual recreation experience—catch rate of fish, time spent on the site, activities, and so forth.

Third, in estimation one must correct for selection bias inherent in on-site samples. The error term implicit in equation (3) in an empirical analysis will be truncated such that no observations less than one are observable. This will cause the estimated demand function to be too steep, giving biased parameter and welfare estimates unless corrected in estimation. See Creel and Loomis (1990). On-site sampling will also over-sample more frequent users. This is called endogenous stratification. A person who visits a site ten times during the season is ten times more likely to be sampled than someone visiting the site only once. Such stratification introduces more bias into the estimation of demand coefficients that must also be corrected. See Shaw (1988) or Haab and McConnell (2002, p. 174-181).

The alternative sampling strategy is off-site sampling. In off-site sampling, individuals from the general population are contacted, usually by mail or phone, and asked to complete a survey. Unlike an on-site sample, an off-site sample will include people who take trips (participants) and people who do not take trips (nonparticipants). This gives the information needed to estimate the intercept, avoids the selection biases, and is simpler to design for random response. It also is possible to model the individual's decision of whether or not to visit the site (see Step 8). Unfortunately, as noted above, off-site samples can be costly to assemble because the participation rates from the general population for a given site tend to be low.

Off-site sampling also raises the issue of determining the extent of the market. Is the full day-trip population located within twenty minutes of the site, three hours of the site, or farther? A local fishing pond may have a small geographic market. A popular ocean beach may have a large geographic market. In principle one would want to sample the entire market randomly.

In a day-trip model of recreation use, a safe bet for the extent of the market is a maximum day's drive to reach the site—perhaps three to four hours. However, the farther one gets away from the site, the lower the rate of participation and the higher the cost of assembling a sample with participants. Prior knowledge of where users come from may help establish the extent of the market. For example, parks may keep records of where their visitors come from. Stratified sampling is also possible whereby residents living closer to the site are sampled more heavily than those living farther away. In this case one must correct for the sampling in estimation and welfare computation.

Using an off-site targeted population is a common and effective way to circumvent the participation rate problem. For many types of recreation, users must obtain a license (such as to fish, boat, or hunt) or pay a registration fee. Participation rates among the set of license holders is significantly higher than the general population. Hence, if one can sample from the set of license holders, the number of people one must contact will be considerably lower than sampling over the general population. Agencies that issue fishing licenses or boating permits often will give an analyst access to addresses of license or permit holders. If so, one can randomly sample from this set.

For a good example of studies using on-site surveys see Sohngen (2000) or Siderelis and Moore (1995).

2.2.4 Specify the Model

Before data collection begins the right-hand side variables for equation (2) must be identified to establish the information the analyst needs to gather in the survey.

Every model includes an individual's trip cost to the site (tc_r). Step 7 addresses the measurement of trip cost at length. Most models also include a measure of trip cost to substitute sites (tc_s); there are usually three or fewer included. The analyst looks for sites frequently visited by the same population of users, sites similar in character to the site of interest in the analysis, and sites nearby. From this mix, a set is formed. Occasionally, proxies for substitute sites are used, such as the number of lakes or acreage of water within a certain distance of one's home.

As shown in equation (2) most models also include a measure of income (y) and a set of demand shifters (z) or demographic variables. The shifters are factors other than trip cost believed to influence the number of trips taken over a season. Some common shifters include:

- Family size
- Age
- Gender
- Urban/rural residence
- Occupation
- Level of education
- Club membership
- Equipment ownership
- Attitudinal information
- Experience in activity.

Family size or family composition may matter for many types of recreation such as beach use or hiking; families with young children are more likely to use a beach, for example. Club membership might include a fishing or hunting club, environmental group, or some other such association. One might use subscriptions to specialized recreation magazines as an explanatory variable as well. These variables pick up unobserved measures of intensity of interest in the recreation activity.

Urban/rural, occupation, and education are sometimes used as shifters as well. Occupation and education are usually categorical variables such as unemployed (yes/no), student (yes/no), retired (yes/no), and so forth. Education might be high school completed (yes/no) and college completed (yes/no). Attitudinal information is response data based on questions like "Would you consider yourself an advocate for the environment?" Experience includes elements such as number of years a person has been rock climbing or a self evaluation of level of expertise. This might come from the answer to a question like "How would you rate your level of rock climbing expertise? Novice, intermediate, or expert?"

Although the complete list of possible shifters is long, most analysts are parsimonious in their selection. Studies typically include anywhere from one to five variables. Sometimes analysts will report two or three model specifications, such as a short model that includes only trip cost and income and a longer model that adds a set of shifters.

2.2.5 Decide on the Treatment of Multiple Purpose Trips

Recreation trips may be single or multiple purpose. On a single-purpose trip, an individual engages only in recreation at the site. On a multiple-purpose trip a person does more—for example, they may visit family or friends along the way, take side trips for business, go shopping, go site seeing, visit other recreation sites, and so forth.

Single-purpose trips fit the travel cost model well. An individual leaves home, travels to a recreation site, engages in some type of recreation, and returns home. All travel expenses can, more or less, be attributed to creating the recreation experience.

Multiple-purpose trips are more complicated. Trip expenses no longer purchase a recreation experience alone. Instead, they purchase a package of experiences. The logic of treating trip cost as the price of a recreation trip or recreation experience becomes tenuous. Attempts to apportion trip cost across the different purposes have met with little success. There is no logical way to identify the marginal cost of the recreation portion of the trip unless some restrictions are placed on the model (Parsons and Wilson 1997). How then does one handle multiple-purpose trips in the analysis?

With day-trip data the most common approach is to assume all trips are single-purpose. If a case can be made that people engage in few other purposes along the way and that these other purposes are mostly incidental, then all trips may be treated as single-purpose with little objection. With day-trip data this is often a reasonable assumption. With overnight data it is not. Indeed, this is one reason many analysts confine their attention to day trips.

Another approach is to drop multiple-purpose trips from the analysis. In this case, the analyst ask respondents to report multiple-and single-purpose trips separately or to report only single-purpose trips. The model is then confined to single-purpose trips.

A last approach is to amend to the basic model to accommodate multiple-purpose trips. Mendelsohn et al. (1992) and Parsons and Wilson (1997) are recent examples. Mendelsohn's framework broadens the definition of a site to include multiple purposes and proceeds with the same logic as the basic travel cost model. Their application is to a multiple-site model where the other purposes are recreation at other sites, but the reasoning would apply to a single-site model. Parsons and Wilson (1997) present an empirical model, with theoretical support, that introduces a simple demand shifter for multiple-purpose trips in a single-site model in which all trips are analyzed together. They also offer an argument for why access values for a site may embody the consumer surplus associated with the other purposes of the trip.

2.2.6 Design and Implement the Survey

The next step is to design and implement the survey. It is useful to start with a survey from a past study as a guide. The survey is usually divided into four parts:

- Introductory material,
- Trip count questions,
- Last trip questions,
- Demographic/household characteristic questions.

The introductory material introduces the interviewer, identifies his or her affiliation, explains the purpose of the study, and provides assurances that keep respondents interested (e.g., the survey is short and the responses are confidential). There are usually opening questions as well that are easy to answer, familiarize the respondent with the material, set the stage for the trip count questions, and help define the site and the season.

The next three parts of the survey are used to form a data set like that shown in Table 1, earlier in this chapter. The trip count questions ask the respondent to report the number of trips taken to the site over a designated time period. These questions may be divided by recreation type (number of fishing trips, number of boating trips, and so forth), by day and overnight trips, and/or by multiple-and single-purpose trips.

The last trip questions pertain to the most recent trip taken by the respondent and typically include information such as time spent on-site, number of people sharing travel expenses, other expenses incurred, and information on the trip experience such as number of fish caught. These data are used to construct trip cost and sometimes to create other explanatory variables in the demand model. These are gathered for the last trip only because gathering them for each trip over the season can lengthen a survey considerably and are difficult for respondents to recall for every trip.

The demographic/household characteristic questions correspond to the vector z in equation (3). These questions close out the survey and include the respondent's income and the location of the respondent's hometown, which are required to estimate trip cost. In principle, one needs to know how trip cost will be calculated in this step to know what data to gather. For this reason steps 6 and 7 overlap somewhat.

It is essential to use good survey research methodology in developing the survey. Champ's Chapter 3 in this volume is a good reference on survey research methods, which cover issues pertaining to effective question writing, keeping response rates high, sampling, and more. While there is no need or room to repeat the principles of good survey research here, there are a couple of issues of special concern in travel cost surveys that deserve some attention: trip recall and trip categorization.

When people are asked to report their number of trips, the analyst assumes they will remember how many were taken. Since the model calls for a count of trips over a season, respondents are often asked to recall the number of trips taken over many months or even a year. This raises obvious questions. How accurate is an individual's recall of past trips? Will approximations be valid?

There is no evidence I am aware of from controlled experiments as to how serious recall error may be, but it is hard to deny the potential. One approach to improve recall is to survey at several intervals over the season asking respondents to report trips taken only over the preceding month or so. While sensible, this approach can raise survey costs considerably and lower responses rates through respondent attrition over the season.

Off-site surveys are usually conducted immediately following a season while trip recall is good. On-site surveys are done within a season, which is also likely to lessen recall problems. However, with on-site surveys the seasonal data are truncated because the respondent can report only the number of trips taken to date; the balance of the season is unknown. Two approaches are used to fill in the data for the end of the season: ask respondents to estimate the number of trips they plan to take or predict the number trips based on reported trip behavior. The later assumes an individual's rate of trips is constant throughout the season. Note that on-site samples that gather addresses only for an end-of-the-season mail survey have a complete trip count but probably have a greater recall problem.

As mentioned in step 2, there is often more than one type of recreation at the site. If so, the survey may be designed for multiple recreation types. The most common survey strategy is to proceed use-by-use in the questioning. For example, in valuing recreation at a local lake one might begin by asking "How many day trips did you take to the site for the primary purpose of fishing?" Then, following the fishing questions, there would be a similar block of questions about swimming, and then boating, and so on.

There are short-cuts that reduce the length of the survey. One might simply ask people to report "How many trips did you take to the site for purposes of recreation?" And then, "What is your primary recreation use of the lake?" This avoids questioning about each recreation use. People are then classified by their primary use. There is a model of lake visitation for people who use the site primarily for fishing, one for people who use the site primarily for boating, and so on. There is some mixing of trips within models in this case; however, if people tend to have a single dominant use for a site, this strategy can be effective.

It is also useful to isolate side trips and trips originating from somewhere other than the resident's hometown. For example, if a person is on a business trip and takes a trip to the beach or goes on a whale-watching tour while on the trip, the beach or whale-watching trip is a side trip. The trip cost from the individual's home to the site would clearly overstate the real marginal cost of the trip. It is easiest to isolate and delete side trips from the analysis by clearly defining a day trip as being to and from one's primary residence when the trip count is being made. See Parsons and Wilson (1997) for an approach that adjusts trip cost and then incorporates side trips into an analysis.

A similar issue arises when an individual owns a cabin or cottage near the recreation site of interest. Suppose the person spends most of the summer at the cottage and makes frequent day trips to the site from that second residence. How should these trips be handled in the analysis—as one three month-long overnight trip from their permanent residence, or, as many day trips to the site from their cottage? While the former choice should, in principle, embody all the individual's consumer surplus, it is simply not amenable to the single-site model. One might consider two strategies: either drop the observation from the sample and lose a heavy recreation user, or include the observation and count the trips as day trips from the cottage. The latter approach understates surplus but avoids deleting important observations. To identify these persons it is necessary to include a question that asks whether or not the individual owns property near the recreation site and makes trips to the site from there.

2.2.7 Measure Trip Cost

Once the raw data are assembled and organized, the trip costs to the site and any substitutes are computed. Trip cost is the sum of the expenses required to make a trip possible. Typical costs for a day trip include:

- Travel cost,
- Access fees,
- Equipment cost,
- Time cost.

Travel cost includes all transit expenses. In a model of day trips where most or all of the trips are made by car, travel cost is measured using the U.S. Department of Transportation's or the American Automobile Association's estimate of the average cost of operating a vehicle per mile. Current studies are using about thirty-five cents per mile. These costs include fuel and upkeep. The round-trip distance to the sites is usually calculated using a software package such PC*Miler. The per-mile cost is multiplied by the round-trip distance to arrive at trip cost. Tolls, if any, are added.

Since travel costs may be shared by several people, efforts are sometimes made to apportion the costs. For example, one might ask respondents to report the number of people sharing travel cost on a last trip and divide the cost equally. Or, one might ask directly for an individual's share of the cost. In either case, this is an example of a "last trip" question mentioned in Step 6. Since it difficult to pose these questions for each trip and since there is no logical way to handle separate trips within a season differently in a single-site model, the analyst typically relies on last trip data.

If the site or any of its substitutes have an access fee, that fee is included in the trip cost. Sometimes sites will have annual or weekly passes, senior discounts, or free admission for children. Making adjustments for seniors and children is easy, but accounting for discounts is more difficult and usually ignored. Typically the daily fee is used.

Equipment costs vary by type of recreation. For fishing one needs bait, tackle, a rod, and sometimes use of a boat. For beach use there may be chairs, umbrellas, surf boards, and so on. For bird watching, there are binoculars and film. For an item like bait the cost is simply the market price of bait for a day of fishing. For durable goods an imputed rent is needed. If one rents or charters a boat for fishing, the cost is simply the fee; if one owns a boat, the rent or cost of its service flow needs to be imputed. One approach for imputing such costs is to use the rental fee for comparable services. This is, no doubt, an overstatement of the cost, but is usually easy to obtain. Equipment cost is often excluded fromthe trip cost estimate. It isdifficult to estimate and, is generally a negligible portion of trip cost when the full cost is amortized over the life of the equipment. If incorporated, equipment cost is usually obtained using a last trip question.

An alternative strategy for estimating travel expense, access fees, and equipment cost is simply to ask individuals to report these expenses on the last trip to the site. The advantages of this approach are that it uses perceived cost information and the researcher need not construct the cost estimates. Since individuals base trip decisions in part on perceptions of cost, which may diverge from actual costs, the respondent-reported estimate is compelling. However, objective estimates based on researcher computation as described above is most common. The data are cleaner in the sense that they are uniform across individuals and there are no missing or otherwise peculiar numbers.

The most difficult issue in computing trip cost, and certainly the one that has received the most attention in the literature, is estimating the time cost of the trip. Time lost traveling to and from the site and time spent on the site constitute time that could have been devoted to other endeavors. The value of those lost opportunities is the time cost of the trip. Time cost often accounts for a sizable portion of the total trip cost and deserves careful attention.

In most applications the estimate of time cost is related to a person's wage in some way. This relationship has a theoretical basis so long as the individual has a flexible working arrangement and can substitute work time for leisure time at the margin. Under such conditions, in theory, an individual increases the number of hours worked until the wage at the margin is equal to the value of an hour in leisure. Multiplying the hourly wage times travel and on-site time, in this case, results in a fair estimate of time cost. Unfortunately, this simple model breaks down for many individuals. The simple leisure/work trade off does not apply to individuals working a fixed forty hour-a-week job for a salary. These people do not have the flexibility to shift time in and out of work in exchange for leisure. The tradeoff is also implausible for retired folks, homemakers, students, and unemployed persons.

Despite the difficulty of extrapolating the simple flexible leisure/work model to many individuals in a recreation data set, the most commonly used approach to value time is still wage-based. For people with fixed work schedules most studies impute an hourly wage using annual income. Reported annual income is divided by the number of hours worked in a year—a number in the range 2,000 to 2,080. Another approach is to impute an individual's wage using a simple wage regression over the subset of individuals in the sample earning an hourly wage (Smith, Desvousges, and McGivney 1983). In this case, wage is regressed on income and a vector of individual characteristics such as age, gender, and education. The fitted regression is then simulated over non-wage earners to impute a wage. For a recent application see McConnell and Strand (1994, p.100).

In wage-based applications, it is also common to see some fraction of the imputed wage used to value time, anywhere from one-third of the wage to the full wage, as the value of time. According to Feather and Shaw (1999), this practice stems from early transportation literature wherein analysts had imputed the time cost in empirical travel studies in this range. The recreation literature has more or less accepted one-third as the lower bound and the full wage as the upper bound, but neither is on firm footing. For example, Feather and Shaw (1999) show that for those on a fixed work week it is possible for the value of time to be greater than the wage. Finally, there are approaches for inferring values of time from market data in the recreation context. See McConnell and Strand (1981); Bockstael, McConnell, and Strand (1988); and Feather and Shaw (1999).

Time traveling to the site as well as time spent on-site should be included in any calculation of time cost. While the time of getting to and from the site is more or less fixed, time at the site is chosen by each individual and may vary across the sample. Nevertheless, on-site time is typically assumed to be constant across individuals and valued the same as travel time. Sometimes analysts use the sample average length of stay on the last trip as an estimate of the fixed on-site time. Others allow the on-site time to vary across the sample using last trip data as each person's on-site time estimate for each site.

It should be evident that measuring trip cost calls for considerable researcher judgement. Moreover, many of the trip cost components are endogenous or chosen by individuals. For example, the purchase of equipment, number of people with whom a person fishes, mode and route of travel, time on-site, choice of residence (starting point for trip), and so forth are all chosen. Yet the model assumes that individuals take price as given. This creates the potential for biased parameter estimates. There are a few attempts in the literature to deal with this endogeneity. McConnell (1992) suggests an approach that allows on-site time to be endogenous, and Parsons (1991) suggests an approach for purging trip cost of endogeneity that is due to choice of residence. See Randall (1994) for a criticism of the trip cost model fueled chiefly by the issue of trip cost endogeneity.

2.2.8 Estimate the Model

The next step is to estimate the model specified in Step 4. In most modern single-site applications, the model is estimated as a count data model. The dependent variable (number of trips) is a nonnegative integer, and the frequency of zero and small numbers of trips typically make up a sizable fraction of the data set. Count models are well suited for these data. See Hellerstein (1999), Creel and Loomis (1990), and Greene (1997, p. 931-946) for details.

The basic count data travel cost model is a Poisson regression. The number of trips taken by a person to a site in a given season is assumed to be generated by a Poisson process. The probability of observing an individual take r trips in a season is

(5) $\Pr(r) = \frac{\exp(-\lambda) \cdot \lambda^r}{r!}$.

The parameter λ is the expected number of trips and is assumed to be a function of the variables specified in the demand model. To ensure nonnegative probabilities, λ usually takes a log-linear for

(6) $\ln(\lambda) = \beta_{tc_r} tc_r + \beta_{tc_s} tc_s + \beta_y y + \beta_z z$

Substituting equation (6) into (5) gives an expression for the probability of observing an individual take r trips as a function of trip cost, income, and individual characteristics. Equation (6) is the Poisson form of the recreation demand specified in equation (2).

The parameters in equation (6) are estimated by maximum likelihood. For each person in the sample, the analyst knows r, tc_r, tc_s, y, and z. Using these data and equations (5) and (6), the probability of observing the number of trips actually taken is constructed for each person in the sample. The likelihood of observing the actual pattern of visits is the product of these probabilities

(7) $L = \prod_{n=1}^{N} \frac{\exp(-\lambda_n) \cdot \lambda_n^{r_n}}{r_n!}$

An individual is denoted by $n = 1, \ldots, N$, so r_n is the number of trips taken by person n. In estimation, the parameters β, on which λ depends according to equation (6), are chosen to maximum L. Many software packages are available for estimating by maximum likelihood. GAUSS and LIMDEP are popular for Poisson forms.

Consumer surplus, or access value, for each person in the sample (area A in Figure 1) has an explicit form in the Poisson model. For individual n the surplus is

(8) $S_n = \lambda_n / -\beta_{tc_r}$

where λ_n is the expected number of trips from equation (6). Once the parameters of the model are estimated, equation (8) is used to calculate the surplus value for each individual in the sample and then aggregated over the population of users to arrive at a total access value. This is discussed in detail in step 9.

The actual form of the probability used in estimation varies somewhat from equation (5) depending on whether one uses on-site or off-site sampling. Recall from step 3 that on-site random samples are truncated at one trip and oversample more frequent users. Either of these complications will bias parameter estimates unless corrected statistically.

The corrected probably for an on-site sample is a slight variation on the basic Poisson probability in equation (5). It takes the form

$$(9)\quad pr(r_n \mid r_n > 0) = \frac{\exp(-\lambda_n)\lambda_n^{r_n-1}}{(r_n-1)!}$$

This corrects for truncation and endogenous stratification and differs from the basic Poisson regression only by r_n-1 replacing r_n. See Haab and McConnell (2002, p. 174-81) for the derivation. With on-site sampling, then, equation (9), instead of (5), enters the likelihood function for each individual. Consumer surplus is still measured as shown in equation (8). See Shaw (1988) and Greene (1997, p. 936-937) for more on truncated regressions in the Poisson model.

Off-site random samples avoid the problem of truncation and endogenous stratification and have the advantage of including non-participants. This allows the researcher to model the decision to participate in recreation at the site or not. For example, in the simple Poisson model of equations (5) and (6) the probability of not participating (r = 0) can be modeled along with the decision of how many trips to take (r = 1,2,3,...). The same Poisson process is assumed to generate the outcome for any count of trips, including zero. For individuals taking zero trips, $\Pr(r_n=0) = \exp(-\lambda_n)$ enters the likelihood function. Again, modeling participation is important because it helps pin down the choke price or intercept on the demand function by using observed choices to estimate $\Pr(r_n=0)$.

A more complicated version of the model is the hurdled Poisson which assumes that the decision of whether or not to take a trip and the decision of how many trips to take are generated from different Poisson processes. The probability of taking zero trips is assumed to be $\exp(-\theta_n)$, where θ_n is some function of individual characteristics that govern whether or not a person takes trips at all. This may include some of the same variables used in the trip frequency portion of the model as well as some new variables. The individual's probability of taking one or more trips, then, is just (1-exp(-θ_n)). The model becomes

$$\begin{aligned} (10) \quad & \Pr(r_n = 0) = \exp(-\theta_n) \\ & \Pr(r_n \mid r_n > 0) = (1 - \exp(-\theta_n)) \frac{\exp(-\lambda_n)\lambda_n^{r_n}}{r_n!(1 - \exp(-\lambda_n^{r_n}))} \end{aligned}$$

The term 1-exp(-λ_n) scales all the non-zero probabilities so that they sum to one. The likelihood function is then constructed using the probabilities in equation (10). See Greene (1997, p. 943) for more.

In some cases the hurdled model is estimated with a double hurdle. For example, the population of non-participants in a given season may be divided by those who never take trips to the site of interest and those who do generally but for one reason or another have not in this particular season. For example, in a study of fishing on a river, there may be two types of nonparticipants: those who never fish and those who fish but not at this site in this season. The analyst may wish to treat these groups differently—the latter group may suffer from a site loss while that former will not (at least in terms of use values). See Shonkweiler and Shaw (1996) and Haab and McConnell (1996) for more on the double hurdled model.

Finally, it is common to see a variation of the Poisson Model known as the Negative Binomial Model to estimate travel cost models. In the Poisson Model the mean and variance of r_n are constrained to be equal. To the extent that this constraint is unreasonable, the model is in error. The Negative Binomial Model is an approach for relaxing this constraint. The basic reasoning and structure above still applies, but the estimation is somewhat more complex. See Greene (1997, pp. 939-940) and Haab and McConnell (2002).

2.2.9 Calculate Access Value

In the final step, access value for the site is computed using the estimated model. Access value may be reported as

- A mean seasonal value per person,
- A total seasonal value for the population,
- A per-trip value per person, and/or
- A discounted present value of the site.

The seasonal per-person estimate for any observation in the sample is the area A in Figure 1. In the Poisson Model this is S_n in equation (8). The estimated seasonal access value for the n^{th} individual in the sample then is

$$(11)\quad \hat{S}_n = \frac{\hat{\lambda}_n}{-\hat{\beta}_{tc_r}} = \frac{\exp(\hat{\beta}_{tc_r} tc_r + \hat{\beta}_{tc_s} tc_s + \hat{\beta}_y y + \hat{\beta}_z z)}{-\hat{\beta}_{tc_r}}$$

where ^ denotes an estimated value using the results of the Poisson regression, and the subscript n on an explanatory variable denotes the value of that variable for individual n.

If one has estimated a Poisson Model using a randomly drawn <u>off-site sample</u>, the sample mean access value is

$$(12)\quad \bar{S}_{off} = \frac{\sum_{n=1}^{N} \hat{S}_n}{N}$$

where $\bar{S}_{off}$ is an unbiased estimate of the mean access value over the population of participants and non-participants in the market area sampled. N is the sample size. A reasonable estimate of aggregate seasonal access value is

$$(13)\quad AS = \bar{S}_{off} \cdot POP_{off}$$

where POP_{off} is the total number of people in the relevant geographic market. POP_{off} is typically from census data. For example, if one samples the entire population over 16 years old within a day's drive, that population is the relevant value for POP_{off}. If one is using a targeted population, such as all individuals holding a fishing license, POP_{off} is the total number of people holding a license in the relevant time period.

If one is estimating a model with an <u>on-site sample</u>, the sample mean access value is a biased estimate of the population mean because it oversamples more frequent visitors to the site. A "corrected" sample mean is

$$(14)\quad \bar{S}^{c}_{on} = \frac{1}{N^*}\sum_{n=1}^{N}\frac{\hat{S}_n}{r_n}\text{, where } N^* = \sum_{j=1}^{R}\frac{n_j}{j}.$$

where dividing the individual surpluses by r_n and the summation by N^* are corrective weights. The term n_j is the number of persons in the sample taking j trips, and R is the largest number of trips taken by a person in the sample. A reasonable estimate of aggregate seasonal surplus in this case is

$$(15)\quad AS = \bar{S}^{c}_{on} \cdot POP_{on}$$

where POP_{on} is the total number of participants at the site over the season. POP_{on} may be gathered from an independent data source on participation rates at the site or may be estimated from survey data.

In principle, an off-site study using equation (13) and an on-site study using equation (15) estimate the same aggregate value *AS*. $\bar{S}^{c}_{on}$ is larger than $\bar{S}_{off}$ because it excludes nonparticipants. But, POP_{on} is smaller than POP_{off} for the same reason. One way of thinking about $\bar{S}^{c}_{on} \cdot POP_{on}$ is that it implicitly assumes that anyone who has not taken a trip has zero surplus and is excluded in the calculation of *AS*. $\bar{S}_{off} \cdot POP_{off}$, on the other hand, includes non-participants, and each non-participant will have a $\hat{S}_n > 0$ since everyone has some positive probability of taking one or more trips in a Poisson model. The non-participants contribute low surplus values to the overall mean, giving the off-site model its lower seasonal mean.

An alternative method of estimating aggregate surplus is to compute an average per-trip-per-person value and then multiply this by an estimate of the total number of trips taken to the site. Since average per-trip values are the per-person seasonal value of the site divided by the number of trips taken, the average per-trip value in a Poisson Model is

$$(16)\quad \hat{t} = \frac{(\hat{\lambda}_n / -\hat{\beta}_{tc_r})}{\hat{\lambda}_n} = \frac{1}{-\hat{\beta}_{tc_r}}$$

This applies to on-site and off-site models alike. To arrive at an aggregate value for the site, multiply the average per-trip value by the number of trips taken to the site during the relevant season. This gives

$$(17)\quad AS = \hat{t} \cdot TRIPS$$

where *TRIPS* is the total number of day trips to the site over the relevant season. Many parks and major recreation sites collect data or at least estimate *TRIPS*, which makes this a popular method of measuring aggregate surplus.

Finally, it is common to see the discounted present value of a site computed using the seasonal aggregate estimate. The easiest approach is to assume no change in the use of the site over time, no change in the character of the site, and a constant rate of discount. Then, using the conventional formula for the value of a perpetuity, the discounted present value of the site is

$$(18) \quad \mathrm{PV} = AS / i.$$

where i is the real rate of discount, usually a number in the range of .01 to .05.

In reporting site values,—be it S, AS, t, or PV,—it is important to be clear what is and is not included in the value. For example, you may be reporting a value of day trips for rocking climbing in a park. If so, it should be noted that this excludes overnight and side trips, other types of recreation taking place in the park, and non-use values.

2.3 A Single-Site Application

Sohngen's (2000) model of beach recreation on Lake Erie in 1997 is good example of a modern application of the single-site model. He estimated two models— one for Maumee Bay State Park (in western Ohio) and the other for Headlands State Park (in eastern Ohio). Maumee Bay offers opportunities for recreation beyond beach use including golfing, camping, and so forth. Headlands is more natural.

Shongen defines the sites as the beaches located withing these two parks (step 1). The recreation users are people visiting a beach within either park for recreation or pleasure (step 2). All forms of beach recreation are aggregated into a single type, and the data were gathered on-site (step 3).

Both models were specified with own price (tc_r), income (y), substitute prices (tc_{s1} and tc_{s2}), and a variety of explanatory variables (z) (step 4). The Maumee Bay model used one substitute site; the Headlands model used two. In both cases the substitutes were nearby beaches similar in character to the beaches under study. The Headlands substitutes were located on either side of Headlands Park. Income was measured as annual household income divided by 10,000. The demand shifters were the scaled responses to five attitudinal questions. According to Sohngen

> These questions asked individuals to rank (from 1 to 5) how important certain issues were in the choice of making a visit. The issues included water quality, maintenance, cleanliness, congestion, and facilities. Higher rankings indicate the issue is more important to the individual visitor.

Finally, the model included a dummy variable for whether or not the primary purpose of the last trip was beach use.

Multiple-purpose trips were included in the demand model along with single purpose trips (step 5). Since all trips were single-day visits by people living within 150 miles of the site, other purposes are likely to be incidental side trips of little consequence. As noted above, they included a demand shifter in the model for trips with the sole purpose of visiting the beach. *Sole = 1* if the sole purpose was to visit the beach and *Sole = 0* otherwise.

The survey was conducted on-site. Random beach users were handed a survey and asked to return it by mail (step 6). Respondents reported their number of day as well as overnight trips to the beach over the entire season. Over 90% of the trips to Headlands were day trips; over 66% to Maumee Bay were day trips. Only day trips were considered in the analysis. The researchers also gathered data on other activities while on a typical beach trip and the attitudinal data on factors that affected a decision to visit the beach. The latter were gathered to construct the demand shifters specified in step 4. The response rate was 52% for Headlands and 62% for Maumee Bay.

Trip cost was measured as the sum of travel expenses and time cost (step 7). Distances to the site were measured as the linear distance from the center of an individual's home zip code to the beach using latitude and longitude coordinates. That distance was doubled (round trip) and then multiplied by 33 cents per mile to get total transit cost. The average distance traveled to Headlands was 26 miles and to Maumee Bay was 35 miles. Time cost was measured as an imputed wage times travel time. Travel time was calculated using the estimated round trip distance and assuming people traveled at 40 miles per hour. The imputed wage was 30% of annual household income divided by 2,040. On-site costs were ignored.

Four different functional forms were estimated (step 8), two continuous (linear and log-linear) and two count (Poisson and Negative Binomial). Truncation at zero trips due to on-site sampling was accounted for in each model; endogenous stratification (oversampling of more frequent users in on site samples) was not. For this reason, their probabilities corresponded to Greene's truncated model (1997, p. 937) instead of to my equation (9). The parameter estimates for the Poisson models for Maumee Bay and Headlands are shown in Table 3.

The coefficient on own trip cost (tc_r) is negative and significant in both regressions, so the demand function is downward sloping. The coefficient on income is positive in both regressions but significant only in the Headlands regression. The effect of substitute sites on trips works as expected in the Maumee Bay model: the higher the cost of reaching a substitute all else constant, the more trips taken to the site. The results are a bit weaker in the Headlands model, where one of the substitute prices has a negative but insignificant coefficient and the other is positive and significant.

Table 3. Single-site Poisson Models for Maumee Bay and Headlands

	Parameter Estimates	
Variables	*Maumee Bay*	*Headlands*
tc_r	-.040***	-.026***
Income	.018	.040***
Sole	-.016	.292***
tc_{s1}	.004***	.005
tc_{s2}	--	-.004
Water Quality	-.053	-.139***
Maintenance	-.270***	.033
Cleanliness	.176**	.028
Congestion	-.065*	-.066***
Facilities	.098**	-.004
Constant	2.648***	2.433***
R^2	.38	.29
Sample Size	230	345

Notes: Three asterisks denotes a coefficient that is statistically significant at the 99% level of confidence; two denotes significance at 95%; and one significance at 90%. *Water quality, maintenance, cleanliness, congestion,* and *facilities* are individual, not site characteristics, based on respondents' ranking of the importance of each beach characteristic in making a trip decision. The ranking ran from 1 to 5, where 1 = strongly disagree it is important and 5 = strongly agree it is important. Source: Sohngen (2000).

The dummy variable for whether or not the primary purpose of the trip is for beach use works in opposite directions in the two models. The attitudinal variables also produced mixed results. Recall that these variables are characteristics of the respondents, not of the sites. One would expect the coefficients on these variables, however, to be correlated with the characteristics of the site. For example, if a site is congested, one would expect a negative and significant coefficient on congestion. This would indicate that people who consider congestion important in making a trip decision are less likely to make a trip to a congested site.

The results are then used to estimate access values (step 9). A per-trip, an annual aggregate, and a discounted present value estimate were computed for each park. The per-person-per-trip values using the model above are \$25 (= 1/.04) for Maumee Bay and \$38 (= 1/.026) for Headlands. The range of per trip values from all models (not reported here) was \$14 to \$33 for Maumee Bay with a midpoint of \$23.50, and \$11 to \$39 for Headlands with a midpoint of \$25.

The Ohio Department of Natural Resources reported the total number of trips taken to the beaches during the 1997 season at 224,000 for Maumee Bay and 238,000 for Headlands. The total access value of the beaches for day trips was estimated by multiplying total trips to each beach by its per-trip value. This is my equation (17). Sohongen reported these using the midpoint of the per-trip values across the models, so I report the same. These were \$5.6 million (= 238,000 · \$23.50) for Maumee Bay and \$5.6 million (= 224,000 · \$25) for Headlands. Last, assuming a real discount rate of 3% and no future change in the use of the beaches, the total discounted present value of each beach (equation (18)) is \$187 million (= \$5.6/.03). This accounts for day-trip beach use only, excluding all other uses as well as non-use value.

2.4 Variations

While the single-site travel cost model using individual data to estimate the access value of a site is the most defensible and widely used application, there are some variations on this theme. I will briefly mention a few of these and direct the reader to the latest literature.

First, there are single-site zonal travel cost models. These are estimated using aggregate visitation rate data and average trip costs from predefined zones near a recreation site. The zonal model has fallen out of favor because of its lack of consistency with basic theory. Nevertheless, when data are limited, the zonal model can provide a useful approximation. See Hackett (2000) for a recent application and Loomis and Walsh (1997) for more detail.

Second, there have been some efforts at valuing quality changes with single-site models. These all seek to estimate a shift in a single-site demand function due a change in quality, and to then use the area between the functions as an estimate of the value of the quality change. There are two versions of these models: pooled and the varying parameter. These can be estimated with cross section data on many sites (Smith and Desvousges 1985, Loomis 1988), using time series data on a single site (Brown et al. 1983), or using data that combine actual trips to a site with hypothetical visits to a site (Layman, Boyce, and Griddle 1996, McConnell 1986).

Third, there are systems of single-site demand equations. These are stacked single-site models for a group of substitute sites. See Bockstael, McConnell, and Strand (1991, pp.254-256) for a discussion and Burt and Brewer (1971) for the first application to a system of demands. Morey (1981) estimated a system of share equations and Ozuna and Gomez (1994) and Shonkwiler (1999) have estimated systems of count data models. These models account for substitutes and allow one to estimate access value for more than one site simultaneously. However, as Bockstael, McConnell, and Strand (1991) explain, it is impossible to value quality changes without placing rather stringent restrictions on the model. Furthermore, it is difficult to estimate such models when the number of sites rise above more than a half dozen or so. The random utility model (RUM), on the other hand, can handle hundreds or even thousands of substitute sites and works well in valuing quality changes.

3. THE RANDOM UTILITY MODEL

This section is divided into four parts as was the previous section: basic model, steps in estimation, an application, and variations. I will discuss the use of the RUM model for valuing site access and changes in site quality. As a general rule the RUM model is preferred to a single-site model. It does a better job of capturing site substitutes and valuing quality changes.

3.1 Basic Model

The RUM model considers a person's choice of a site for a recreation trip. Instead of a "quantity demanded" as in the single-site model, there is a site chosen. In choosing a site a person is assumed to consider its "price" and its characteristics. The "price" is the trip cost. The characteristics are features of the site that matter to people, such as ease of access and environmental quality. While the time frame for a single-site model is a season, the time frame for the RUM model is a choice occasion. When analyzing day trips a choice occasion is simply a day.

The theory works as follows. On a given choice occasion, a person considers visiting one of C sites denoted as $i = 1,2,....,C$. Each site is assumed to give the person a *site utility* v_i. The utilities are assumed to be a function of trip cost and site characteristics. The utility for site i assuming a linear form is

$$(19)\quad v_i = \beta_{tc} tc_i + \beta_q q_i + e_i$$

where tc_i is the trip cost of reaching site i, q_i is a vector of site characteristics, e_i is a random error term, and the βs are parameters. One expects site utility to decline with trip cost ($\beta_{tc} < 0$), to increase with desirable characteristics such a easy access and good environmental quality, and to decrease with undesirable characteristics. The random error term accounts for unobserved factors.

In theory, a person chooses the site with highest utility. Site k, for example, is chosen if

$$(20)\quad \beta_{tc} tc_k + \beta_q q_k + e_k \geq \beta_{tc} tc_i + \beta_q q_i + e_i \text{ for all } i \in C$$

A useful way of expressing this result is in terms of *trip utility*. An individual's trip utility is

$$(21)\quad u = \max(v_1, v_2, ..., v_C)$$

Trip utility is the maximum attainable site utility on a given choice occasion assuming a person visits a site. If site k gives the highest utility, the person visits site k and attains trip utility $u = v_k$.

Since people may choose not to take a trip on a given choice occasion, it is common to see a no-trip utility included in the choice set. Let a person's no-trip utility be

$$(22)\quad v_0 = \alpha_0 + e_0$$

No-trip utility is the highest utility a person can attain in any activity other than visiting one of the C sites. A person now chooses from $C + 1$ alternatives—C sites and no-trip. Now, it is useful to define a *choice occasion utility* as

$$(23)\quad u^* = \max\{v_0, v_1, \ldots, v_C\}$$

Choice occasion utility is the maximum attainable utility on a given choice occasion. If no-trip gives the highest utility, choice occasion utility is $u^* = v_0$. If visiting site k gives the highest utility, choice occasion utility is $u^* = v_k$. And, so forth.

Individual characteristics similar to the vector z in the single-site model (age, family size, years of experience in a recreation activity, and so on) may also enter the model and do so in one of two ways. First, they may be used to capture differences in participation in recreation across the sample. Some people like to fish, and some do not; some people like to go to the beach, and some do not; and so forth. To capture differences in participation, characteristics are entered as shifters in the no-trip utility function

$$(24)\quad v_0 = \alpha_0 + \alpha_1 z + e_i$$

where z is a vector of characteristics believed to influence a person's propensity for recreation. In this way, no-trip utility varies across the population. Consider a model of recreational fishing. If gender is a dummy variable in the vector z ($z_g = 1$ if female) and men like to fish more than women, a positive coefficient on z_g allows women to have a higher no-trip utility than men and captures this effect. Or consider rock climbing: if interest in climbing decreases with age, a positive coefficient on age in the vector z would capture this effect.

Individual characteristics may also be used to capture differences in preferences across the population for different sites. Some people like natural beaches, while others like more developed beaches. Some people trailer a boat and prefer a site with a boat ramp, while others look to rent a boat or moor a boat at some preselected site. To capture differences in preferences for different sites, individual characteristics are interacted with the relevant site characteristics.

Assume there are m site characteristics and let $q_i = (q_{1i}, q_{2i}, \ldots q_{mi})$. Suppose in an analysis of beach use that there are people for whom surf fishing at a beach is an important part of their recreation experience, and others for whom it is not. Suppose I have data on whether or not a person owns a surf fishing license. I might specify site utility as

$$(25)\quad v_i = \beta_{tc}tc_i + \beta_{q_1}(q_{1i} \cdot z_1) + \beta_{q_2}q_{2i} + \ldots + \beta_{q_m}q_{mi} + e_i$$

where q_{1i} is a measure of the quality of surf fishing and $z_1 = 1$ if the person owns a surf fishing license and $= 0$ if not. In this way $\beta_{q1}q_{1i}$ affects site utility only for individuals who own a surf fishing license. Assuming $\beta_{q1} > 0$, sites with good surf fishing increase utility for the fishing population but not for the general population. It is possible to concoct numerous interactions of this sort.

Now, let's turn to how the RUM theory can be used to value site access at one of the C sites. Suppose site 1 is closed because of an oil spill. Using RUM theory, I can express a person's choice occasion utility with and without a spill. The utility without a spill, the baseline, is

$$(26)\quad u^*(baseline) = \max\{v_0, v_1, v_2 \ldots, v_C\}$$

The utility with a spill is

$$(27)\quad u^*(spill) = \max\{v_0, v_2, v_3 \ldots, v_C\}$$

The utility $u^*(spill)$ excludes site 1 since it is closed. The decline in utility or welfare loss due to the spill is $u^*(spill)$ - $u^*(baseline)$. This is the difference in the utility a person can attain when the site is and is not available. The change in welfare is

$$(28)\quad \Delta u^* = \max\{v_0, v_2 \ldots, v_C\} - \max\{v_0, v_1, v_2 \ldots, v_C\}$$

Using equation (28), suppose a person visits site 1 if there is no closure. It must be true for that person that $v_1 > v_i$ for all i. If site 1 were to close, the person would choose the alternative with the second highest utility. Suppose that is site k. If so, the person visits k instead of 1 and utility declines from v_1 to v_k. By the same reasoning, if no-trip has the next highest utility, the person no longer takes a trip and utility declines from v_1 to v_0. The closer the second highest utility is to site 1 utility, the smaller the decline in welfare due to the spill.

If a person visits a site other than site 1 if there is no closure, then there is no loss in utility. For example, if a person visits site k when site 1 is open, then *u*(spill)* = *u*(baseline)* = v_k. The same alternative that maximized utility without closure maximizes it with the closure. Note that there is no accounting here for congestion at site k that may follow from the closure of site 1.

To convert the decline in utility in equation (28) into monetary terms, divide by an individual's marginal utility of income. In the RUM model the negative of the coefficient on trip cost, $-\beta_{tc}$, is a measure of the marginal utility of income. It tells us how much an individual's site utility would increase if trip cost were to decline or, what is the same, if income were to rise for that trip.[2] The welfare loss in monetary terms then is

$$(29)\quad \Delta w = \left[\max\{v_0, v_2 \ldots, v_C\} - \max\{v_0, v_1, v_2 \ldots, v_C\}\right] / -\beta_{tc}$$

Δw is a *per choice occasion* value. In a day trip model it gives the welfare loss for a day due to the oil spill. To convert it to a seasonal value comparable to the single-site model, multiply by the number of choice occasions in the season. The seasonal value for the loss of site 1 is

$$(30)\quad \Delta W = T \cdot \Delta w$$

where T is the number of choice occasions in a season.

The value of access to more than one site is computed by dropping many sites from equation (27). For example, if sites 1 through 5 are closed because of the oil spill, sites 1 through 5 are dropped in equation (27). It also possible to value new sites by adding them to the choice set. In this case, the trip cost and site characteristics for the new site or sites must be specified as well.

The model is used in a similar way to value changes in site characteristics. Instead of dropping a site from the choice set, a site characteristic is altered which, in turn, changes site utility. For example, if water quality is a site characteristic in the vector q_i, a decline in water quality at site 1 is captured by writing q_1 as $q_1^{\#}$, where the element in q_i measuring water quality is smaller in $q_1^{\#}$ than in q_1. The utility at site 1 then declines from $v_1 = \beta_{tc}tc_1 + \beta_q q_1 + e_1$ to $v_1^{\#} = \beta_{tc}tc_1 + \beta_q q_1^{\#} + e_1$.

In this case an individual's choice occasion utility without the decline in water quality is *u*(baseline)* in equation (26). The choice occasion utility with the decline in water quality is

$$(31)\quad u^*(dirty) = \max\{v_0, v_1^{\#}, v_2 \ldots, v_C\}$$

where $v_1^{\#}$ is the utility at site 1 with the now poorer water quality. Following the same reasoning as a site closure, the welfare loss is

$$(32)\quad \Delta w = \left[\max\{v_0, v_1^{\#}, v_2 ..., v_C\} - \max\{v_0, v_1, v_2 ..., v_C\}\right] / -\beta_{tc}$$

Equation (32), like equation (29), captures an array of possible behavioral responses. If an individual does not take a trip or visits a site other than site 1 without the decline in water quality, there is no change in utility. The same alternative maximizes utility with and without the change in water quality, so choice occasion utility is unchanged. If an individual visits site 1 without the change, he or she may either continue to visit site 1, visit another site, or take no trip with the change. In each case, the choice occasion utility declines.

If the individual continues to goes to site 1, utility drops from v_1 to $v_1^{\#}$. The individual maximizes utility by making the same trip but the recreation experience is diminished because of the decline in the water quality. If the individual goes to another site (say, k) or no longer takes a trip, then utility declines from v_1 to v_k or from v_1 to v_0. All of these manifestations are captured in equation (32). Again, this gives a *per choice occasion* value. A seasonal value is computed as before using equation (30). And again, as with site access value, quality changes can be evaluated at many sites simultaneously by altering the relevant attribute for each affected site in equation (31).

So far, I have treated the RUM model as deterministic—assuming all parameters and the error terms are known. In application, however, the parameters are estimated and the error terms are unknown. The error terms are assumed to come from some known random distribution. The choice occasion utility in equation (23) then is also random. This is easy to see by substituting the random site and no-trip utilities into equation (26) giving

$$(33)\quad u^*(baseline) = \max\{\alpha_0 + e_0, \beta_{tc} tc_1 + \beta_q q_1 + e_1, ..., \beta_{tc} tc_C + \beta_q q_C + e_C\}$$

If the error terms in equation (33) are random, choice occasion utility u^* is random. For this reason, in application, one uses the expected value of choice occasion utility, instead of a deterministic value. Expected choice occasion utility, the applied counterpart to equation (26), is

$$(34)\quad u^*(baseline) = E(\max\{\alpha_0 + e_0, \beta_{tc} tc_1 + \beta_q q_1 + e_1, ..., \beta_{tc} tc_C + \beta_q q_C + e_C\})$$

Equation (29) for access value and (32) for a quality change now take the forms

$$(35)\quad \Delta w = \left[E(\max\{v_0, v_2, ... v_C\}) - E(\max\{v_0, v_1, v_2, ... v_C\})\right] / -\beta_{tc}$$

$$(36)\quad \Delta w = \left[E(\max\{v_0, v_1^{\#}, v_2, ... v_C\}) - E(\max\{v_0, v_1, v_2, ... v_C\})\right] / -\beta_{tc}$$

In estimation, the form of the distribution for the error terms determines the form of the expected value of the choice occasion utility. Each of the behavioral responses to a site closure and a quality change as described above still apply. However, now, each response occurs with some probability. Annual values, once again, are calculated by multiplying the per choice occasion values in equations (35) and (36) by the number of choice occasions in the season.

In estimation then, one seeks to estimate the parameters of the site and no-trip utilities in equations (19) and (24) using visitation data. Table 4 shows the form of a typical data set—a count of trips to each site in the choice set, trip cost to each site, detailed site characteristics for each site, and demographic data across a sample of individuals. For simplicity the table pertains a simple model with three sites and one site characteristic (water quality). In most applications the number of sites and site characteristics is much larger. Once the data are assembled, the parameters of site and no-trip utility are estimated using some form of a discrete choice multinomial logit model (more on this in the following sections). The estimated parameters are then used to value site access or quality changes at specified sites for each individual in the sample using equation (35) and (36). Seasonal measures and means across the sample are computed. Finally, the estimates are used to compute aggregate values across the population. The next section covers these steps in detail.

3.2 Steps in Estimation

The steps in estimating the basic RUM model are listed in Table 5 and are similar to the steps for a single-site model. To avoid repetition I will focus on aspects of the steps in the RUM model that are different from the single-site model. It will be useful to refer the single-site counterpart for most steps.

Table 4. Typical Data Set for Estimating a RUM Model (Three-site Choice Set)

ID	Number of Trips to Site			Trip Cost to Site $			Water Quality Index at Site *1-10*			Income *$000*	Age *yrs*
	Site 1 (r_1)	Site 2 (r_2)	Site 3 (r_3)	Site 1 (tc_1)	Site 2 (tc_2)	Site 3 (tc_3)	Site 1 (q_1)	Site 2 (q_2)	Site 3 (q_3)	(y)	(d_1)
1	2	0	17	45	158	15	10	2	1	67	43
2	1	0	3	111	201	35	8	7	5	22	28
3	0	3	0	29	33	345	2	8	9	109	39
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
N	12	0	0	12	66	123	5	2	1	78	51

Throughout this section assume that a RUM model is being estimated for the purpose of valuing site access and quality change at several sites, that the analyst is using day trip data only, and that the data are individual-based.

Following this section is a brief discussion of a RUM application of my own, done with Matt Massey. We estimated a model of Mid-Atlantic beach use by Delaware residents to value beach closures and loss due to erosion. Each step of the analysis is described.

Some other good applications to read if you are just beginning to learn about the use of travel cost RUM models are Montgomery and Needelman (1997) and Murray, Sohngen, and Pendelton (2001). I also recommend two publications that go into considerable depth in laying out the development of their model, including basic theory, survey design, data collection, computer programs, and so forth. These are McConnell and Strand (1994) and Hoehn et al. (1996).

3.2.1 Identify the Impacts to be Valued

The analysis begins by identifying the impacts to be valued. These will take the form of site closures, openings, or quality changes at one or more sites. Some examples of impacts are the closure of many lakes and rivers due to a fish consumption advisory, the opening of a new park or skiing area, the expansion

Table 5: Steps in Estimating a RUM Model

Step 1	Identify the Impacts to be Valued
Step 2	Define the Population of Users to be Analyzed
Step 3	Define the Choice Set
Step 4	Develop a Sampling Strategy
Step 5	Specify the Model
Step 6	Gather Site Characteristic Data
Step 7	Decide on the Treatment of Multiple Purpose Trips
Step 8	Design and Implement the Survey
Step 9	Measure Trip Cost
Step 10	Estimate Model
Step 11	Calculate Access and/or Quality Change Values

of several hunting areas, a change in water quality across several urban beaches, and so forth. Identifying the impacts to be evaluated at the outset defines the research problem and gives direction for the succeeding steps.

The impacts may be hypothetical. For example, a site is currently open and you want to analyze the loss associated with its possible closure, or a site presently has no hiking trials and you want to consider the gain associated with adding trails. The impacts may also be actual events. For example, several lakes are closed because of contamination and you want to analyze the welfare loss.

With a site closure or opening one identifies the areas and, in turn, the sites that are affected. With a quality change one is identifying the affected sites and the changes that will occur there. In this case, it is useful to begin thinking early in the analysis about how that quality will be measured for each site. For example, will objective measures of quality be used, such as levels of dissolved oxygen for water quality? Or will perceived measures of quality be used, such as a rating based on the reporting from respondents in a survey ? Also, are there published measures or will you need to construct new measures as part of the analysis? How will these quality changes map into the policy being evaluated? It is also necessary to establish early on that there is sufficient variation in quality across sites to measure its effect on site choice. If quality is more or less uniform across the sites for the specific characteristic of interest, it is impossible to measure the effect of quality on site choice. See Parsons and Kealy (1992) for an example using objective measures and Adamowicz et al. (1997) for an analysis using objective and perceived measures.

In some cases the impact may be analyzed as either a change in access or a change in quality. For example, a fish consumption advisory might be analyzed as a quality change instead of a closure, especially if people continue to use the site. In this case one includes a dummy variable for fish consumption advisories in the characteristic vector q. For this approach to work, some sites in the choice set must have advisories currently.

3.2.2 Define the Population of Users to be Analyzed

The next step is to define the population of users for whom values will be estimated. In principle, this includes all users and potential users of the affected sites—persons who use the sites without the changes and who might use the site with the changes. A geographic area encompassing all users is the market for the changes under consideration.

One way to capture the market in a day-trip analysis is to define the population as all individuals residing within a day's drive of the affected sites. In practice one approximates this market using boundaries that are convenient for sampling. The most common market definition is the set of residents in one or more states within a day's drive of the sites. In some cases analysts will focus on a particular state. This may be called for by the decision makers funding the study or simply due to limited resources. As with the single-site model, one also needs to consider what types of recreation to include and whether or not to aggregate recreation uses.

Here are some examples of day trip market definitions: Parsons and Kealy (1992) define the market for water-based recreation on Wisconsin lakes as all Wisconsin residents. Adamowicz et al. (1997) define the market for hunting in Alberta as Alberta residents holding provincial moose licenses. Bockstael, Hanemann, and Kling (1987) define the market for Boston beaches as the Boston metropolitan area. See step 4 for a discussion of defining the market before sampling begins.

3.2.3 Define the Choice Set

The next step is to define the choice set C in equation (21). This includes defining and determining which sites belong. In principle, one wants to include all the sites with impacts identified in step 1 plus all other sites that may serve as substitutes for these sites for the population of users defined in step 2. In practice, one always approximates this set. Again, political boundaries play a role in constructing the choice set and defining the sites.

It is easiest to see how analysts go about defining sites and determining what to include in the choice set by giving some examples. In Parsons and Kealy (1992), we analyze lake recreation in Wisconsin. Individual lakes over 100 acres large are defined as the sites; there are over 1,000 such lakes. If a lake is within 150 miles of a person's home, it is included in his or her choice set. Few people traveled more than 150 miles to make a day trip.

Andrews (1996) analyzes trout fishing in eastern Pennsylvania. He defines sites as the management units used by the state fish and wildlife service, which are stream segments and lakes known to have trout. Each is different enough in character to be managed separately and data are organized by these units by the state. There are over 2,000 sites. The choice set includes any site within 185 miles of a person's home. The longest day trip in the data set is 183 miles.

Parsons, Jakus, and Tomasi (1999) study lake recreation at reservoirs operated by the Tennessee Valley Authority in the southeastern United States. Each person's choice set is defined as fourteen specific reservoirs in the TVA system. Shaw and Jakus (1996) study rock climbing in northeastern United States. Their sites are defined as the four major climbing areas in the region. Each person has all four sites in his or her choice set. Morey, Watson, and Rowe (1993) study salmon fishing in Maine and Canada. Nine rivers in the region known for salmon fishing are included each person's choice set.

McConnell and Strand (1994) study marine recreational fishing on the east coast of the United States and define coastal counties as the sites. Again, if the site is within 150 miles of the person's home it is included in the choice set. Hausman, Leonard, and McFadden (1995) study recreational fishing in Alaska and define sites as one of seventeen large fishing regions in the state. These regions serve as each person's choice set.

As you see, choice sets vary in size from three or four sites to more than a thousand sites. Site definitions can vary from highly aggregated regions or counties to rather narrowly defined units such as beaches or small segments of a river. In choosing the number of sites and the degree of aggregation of sites, it is best to err on the side of many sites and narrow site definitions. Modern econometric software packages can handle a large number of sites. The feasibility of randomly drawing sites in estimation to approximate larger choice sets is also an option (Parsons and Kealy 1992).

If you must aggregate sites into regions or counties, the general rule is to group similar sites together. The sites should be similar in all characteristics including trip cost. The less similar the sites, the more bias one is likely to encounter (Parsons and Needelman 1992). For approaches that mix aggregated and non-aggregated sites together, see Lupi and Feather (1998) and Parsons, Plantinga, and Boyle (2000).

Finally, there has been some concern about using researcher- vs. individual-defined choice sets in RUM models. The definitions described above are all researcher defined. Peters, Adamowicz, and Boxall (1995) and Hicks and Strand (2000) have argued that people cannot perceive this many sites in making a choice and suggest an approach using choice sets determined by people in the survey. Individuals identify sites they consider in site choice, and this makes up the choice set. For a counter argument to this approach see Parsons, Massey, and Tomasi (1999). Researcher-defined choice sets still dominate the RUM literature.

3.2.4 Develop a Sampling Strategy

Next, a sampling strategy is developed. Virtually all published studies use some form of off-site random sampling. The users identified in step 2 are contacted by phone or mail and asked to report trips taken to the sites defined in the choice set in step 3. See Champ's Chapter 3 in this volume or Dilman (1999) for random sampling with surveys. In some cases a targeted population of users, such as people with boating licenses, will be contacted randomly. In most applications participants and non-participants will be sampled. This enables the analyst to model the decision of whether or not to participate in the recreation activity.

Like the single-site model, sampling for a RUM model runs into the issue of low participation rates from the general population. The fraction of people participating in most forms of outdoor recreation is low enough that random samples from the general population yield a small sample of actual users. The conventional way of dealing with this problem is to stratify the sample—sampling counties or communities nearer the sites more heavily. Participation rates are higher in the nearby communities, but in some cases stratification is the only way to get nearby users in the sample if local communities are relatively sparsely populated. Stratification not only increases the number of participants but also allows for a more even geographic distribution across the sample of users, which can sharpen parameter estimates by ensuring variation of trip cost across the set of sites. As with the single-site model, a target population of license holders circumvents the problem of low participation rates.

Unlike the single-site model, on-site sampling is usually not an option. On-site sampling, or what is known as "choice-based sampling," is covered extensively in the RUM literature. To estimate parameters without bias, site weights are needed. The weights are based on the actual distribution of day trips by the population of users being studied and are used to adjust the count of trips to the different sites to reflect their true occurrence. I know of no RUM applications of the travel cost model using a choice-based sample. See Train (1986, pp. 48-49), Ben-Akiva and Lerman (1985), or Laitila (1999) for more on working with on-site samples.

3.2.5 Specify the Model

Before data collection begins one specifies the elements that will be included in the model in equations (19) and (24).

Site utility in equation (19) includes trip cost (tc_i) and a vector of site characteristics (q_i) for to each site. In some cases it will include interactions with individual characteristics. Trip cost is the same as in the single-site model, the sum of travel and time cost. The vector q_i includes characteristics that matter to people in making a site choice. This will vary depending on the type of recreation being studied. Fishing might include catch rate of fish, rock climbing might include difficulty of the climb, boating might include presence of boat ramps, and so forth. It is difficult to make a generic list, but here are some common attributes one sees in a RUM model:

- Amenities
- Size
- Access
- Environmental quality
- Park (yes/no)
- Opportunities on-site
- Sport success
- Remote location (yes/no)
- Character of nearby area
- Special features.

Amenities are measures such as acreage of tree coverage. Size might be total acres in the site, length of beach, or number of trails. Access might be a variable indicating the presence of a boat ramp or one indicating that the site is accessible only by 4-wheel-drive vehicles. Environmental quality includes such measures as water quality or the presence of a fish consumption advisory. A variable designating whether or not it is a park is a common site characteristic. By opportunities on-site, I mean items such as camping sites or a museum. Success is quantified as catch rate of fishing, bag rate for hunting, sightings for birding watchers, and so forth. The character of the nearby area might be an indication of whether a beach is developed or natural. Special features are unusual characteristics, such as a dam on a river, that may make a difference in site choice. Finally, if a site or group of sites is distinguished from others sites in the choice set in some more general way, it is common to see an alternative specific constant assigned to that site. An alternative specific constant is a dummy variable in the site utility for one or more sites.

The no-trip utility in equation (24) is a vector of individual characteristics that govern how often and whether or not a person makes a trip. These characteristics play the same role as the vector z in the single-site model. Some common characteristics are age, occupation, and urban/rural residence—essentially the same list shown in single-site step 4.

In some cases individual characteristics enter the site choice model as interaction terms. As explained in earlier, this happens when one believes that a certain characteristic affects people differently. For example, the presence a boat ramp may matter to people who trailer a boat but may not matter to people who keep a boat at a marina. The difficulty of a rock climb may matter to expert climbers but not to novice climbers, or at least the degree of its importance may be different. An interaction is entered into the model as shown in equation (25).

Specification always depends on available data. For this reason steps 5 and 6, model specification and collection of site characteristic data, work in parallel. The analyst has a certain specification in mind but amends it to accommodate available data. For some good examples of differences in specification by recreation types, see Sidderelis (1995) for boating, Shaw and Jakus (1996) for rock climbing, Karou (1995) for fishing, Morey (1981) for skiing, Parsons and Massey (2003) for beach recreation, and Adamowicz (1997) for hunting.

3.2.6 Gather Site Characteristic Data

The next step is to gather site characteristic data. As noted, step 5 works in tandem with step 6. The typical sources for characteristic data are the state and federal agencies responsible for managing the resource. Environmental protection, natural resource management, and fish and wildlife agencies are often good sources. These agencies may have data on indicators of environmental quality and physical measures such as size and elevation. Furthermore, the agencies often have ready made site definitions. A fish and game agency, for example, may have management units for which data are gathered for their own purposes. In most cases these agencies are a good starting point for data collection.

Other sources of data are tourist bureaus, clubs and associations, universities, scientists, and newspapers. Some cases need fieldwork where the analyst constructs the primary data independently or interviews knowledgeable people. Even in cases where variables are not constructed by field observation, such visits can confirm or amend existing site and variable definitions.

In some circumstances the site characteristic data are gathered in the survey. For example, individuals may be asked to rate the quality of hunting at each site visited in the current season. Using these responses the analyst constructs a hunting quality index for each site. One could do the same with catch rate of fish, view amenities, and so on.

There are several problems with this approach. First, the data for each site are usually confined to those who have visited the site. This is likely to bias quality measures upward. People who visit a site are those who find the site desirable. Second, popular sites have more data than unpopular sites. Many sites may have a single visitor or no visitors, which causes an asymmetry in the quality of variable measurement across sites. Third, perceived variables are more difficult to map into actual policy changes.

An alternative method for working with survey data is to perform an auxiliary regression using the reported measure for the characteristic (such as number of fish caught) as a dependent variable and observable site characteristics (such as lake size, depth, elevation, and presence of regulations) as explanatory variables. The unit of observation is a site and the number of observations is the number of respondents times the number of different sites visited by each. The fitted regression is then used to predict catch at each site. For an example see McConnell, Strand, and Blake-Hedges (1995).

3.2.7 Decide on the Treatment of Multiple Purpose Trips

In a day-trip RUM model, multiple-purpose trips are handled in much the same way as in the single-site model. Either all trips are assumed to be single-purpose or multiple-purpose trips are identified in the survey and dropped from the analysis. Again, this is a larger issue when overnight trips are being analyzed.

Another way of dealing with multiple-purpose trips is to define site characteristics in such a way that they account for the potential of other activities while visiting the site. For example, a beach characteristic in a RUM model might be a dummy variable for the presence of shopping nearby. This method at least recognizes that a recreation experience may be broader than activity at the site alone.

It also possible to imagine a variant of Mendelsohn et al.'s (1992) approach for the RUM model. Alternatives would be defined as portfolios of sites. Each site would be represented by a dummy variable in the *portfolio utility*, and trip costs would be measured as the cost of visiting the entire portfolio. People would choose portfolios. Welfare losses could be computed for losses of one or more sites from the portfolios.

3.2.8 Design and Implement Survey

Next is the design and implementation of the survey. It is organized in the same four parts as a single-site survey:

- Introductory material,
- Trip count questions,
- Last trip questions,
- Demographic/Household Characteristic questions.

The introductory material, last trip questions, and demographic questions are essentially the same as in a single-site survey. However, unlike the single-site survey, in the RUM survey one gathers data on trips to every site in the choice set. The responses are used to form a data set like that shown earlier in Table 3. RUM surveys face the same recall and trips categorization issues faced with single-site models. (See single-site step 6 for a discussion.) Recall is perhaps a larger problem in a RUM survey, because we are concerned about many sites.

There are essentially three approaches for counting trips to the sites. One provides the respondent a pre-established list of sites. The respondent reviews the list and records the trips over the relevant time period. This approach is the easiest for data handling; the sites are defined and each respondent's count fits neatly into the definition. For sites off the list, there is usually a catch-all category such as "some other site." This approach also helps recall. The site names serve as a reminder to people, and recalling exact site names is not necessary.

The second approach is open-ended. People are asked to list all the sites they have visited over the relevant time period along with a count of trips to each. Once gathered the data must be reworked such that site definitions are consistent across respondents. This can be time consuming. Sites often have more than one name, people may use nearby town names, and sites with similar names can lead to confusion. When the number of potential sites runs into the hundreds or thousands, this approach may be hard to avoid. Even in these cases an insert, separate from the survey, listing (and numbering) the sites is worth considering.

One useful innovation to use in conjunction with the open-ended approach is a map-and-sticker. Respondents are asked to place a sticker on a map for each site visited. The sticker is numbered to correspond to a table in which the respondent records the number of trips taken to the site. This eases the final assembly of the data and avoids any confusion about the name and location of the site. It also has the advantage of aiding in finer site definitions depending on individuals' use of sites. For example if there is heavy visitation to a site that seems divided by trips to the northern and southern portion of the site, one might opt for dividing the site along these lines.

Mail surveys or mail–phone surveys favor the approach in which a list of sites is provided to the respondents. Phone surveys call for an open-ended approach unless the number of sites is fewer than a dozen or so. However, one can approximate the preestablished list approach by being interactive. The caller has a list of sites, and during the survey respondents are asked to count trips region by region. For example, "Did you visit any beaches on the Outer Banks? If so, which beaches?" When the respondent gives a beach, the caller verifies that the beach fits the list and proceeds. If the respondent has trouble recalling the name, the caller can ask about the general area and provides some clues till the correct beach is identified. If the respondent gives an unknown name, it is even possible to find the closest beach. This can be an extremely valuable way to construct the trip data file and can eliminate the post-survey site definitions.

A third approach is to work with last trip data alone. In this case one gathers information only on the last site visited—what site was it, how many trips were taken to that site over the season, and how many were taken to all sites in total? A basic RUM model can be estimated with these data. The survey is less complex, and recall is a smaller issue. This comes at the expense of far fewer trips and hence less data for the final analysis.

3.2.9 Measure Trip Cost

This step is essentially the same a single-site step 7 except that trip cost is measured to every site in a person's choice set, not just the visited sites. This invariably calls for a software package (such as PC*Miler) to compute travel times and distances between respondents' hometowns and the sites. Respondents' hometowns and the site names must defined so they map into the chosen software package. Small communities are sometimes excluded. Spellings must be exact. Some packages require state abbreviations as part of the title. Most software packages will compute time and distances between zip codes as well as an approximation.

For more discussion of the issues surrounding the measurement of trip cost, such as the value of time and the use of individual versus household measures, see single-site step 7.

3.2.10 Estimate Model

Next, the analyst seeks to estimate the parameters α and β in equations (19) and (24) specified in step 5. This is done in a probabilistic framework, in which an expression for the probability of visiting a site is formed. The probability that a person visits site k is

$$(37)\quad pr(\beta_{tc}tc_k + \beta_q q_k + e_k \geq \beta_{tc}tc_i + \beta_q q_i + e_i \text{ for all } i \in C \\ \text{and } \geq \alpha_0 + \alpha_1 z + e_0)$$

Under different assumptions about the distribution of the error terms e_i, different forms for equation (37) are derived. The simplest is the multinomial logit (ML) for which the probability that an individual visits site k is

$$(38)\quad pr(k) = \frac{\exp(\beta_{tc}tc_k + \beta_q q_k)}{\exp(\alpha_0 + \alpha_1 z) + \sum_{i=1}^{C} \exp(\beta_{tc}tc_i + \beta_q q_i)}$$

Notice that the probability of visiting site k depends the characteristics of site k (in the numerator and denominator of equation (38)) and on the characteristics of all other sites (in the denominator). Also notice that each person has exp (α_0 $_+$ α_1z) a probability for each site and no trip. The no-trip probability has in its numerator, and its denominator is the same as the denominator in equation (38). This ML probability assumes that error terms in equation (19) and (24) are independent and identically distributed with Weibull distribution. See Greene (1997, p. 913).

The parameters are estimated by maximum likelihood using the probabilities in equation (38). If one has data on N persons each visiting one of the S sites or taking no trip for a given choice occasion, the likelihood of observing that pattern visits in the data using the ML probability model is

$$(39)\quad L = \prod_{n=1}^{N} \prod_{i=0}^{C} pr(i)^{r_{in}}$$

where r_{in}= 1 if individual n visited site i and = 0 otherwise. The probability *pr(i)* is the logit form from equation (38). The parameters are estimated by choosing values of α and β to maximize L. These are the maximum likelihood estimates for the model. This is the form of the model that is most likely to generate the patterns of visits actually observed in the data.

In circumstances where individuals are observed taking multiple trips to sites over the season, the same likelihood function is used, with r_{in} now equal to the number of trips taken to site i by individual n. Many software packages are available for estimating logit models; LIMDEP, GAUSS, and TSP are common.

This ML model is criticized for a restriction known as the independence of irrelevant alternatives (iia). This restriction implies that the relative odds of choosing betweenany two alternatives is independent of changes that may occur in other alternatives in the choice set. For example, the iia property implies the following: if an improvement in a characteristic at site k causes a 10% increase in the probability of visiting that site, then the percent change in the probability of visiting each of the remaining sites in the choice set must decrease by 10% (a proportional reduction in all other probabilities). If some sites are better substitutes for the site experiencing the improvement, this result is unrealistic. One expects good substitutes to have a larger percentage reduction in probability than poorer substitutes.

There are essentially two methods for relaxing iia: the nested logit (NL) and the mixed logit (MX). Both introduce correlation among the site and no-trip utility error terms, which allows for more general patterns of substitution in the model.

NL and MX models may be estimated using GAUSS or LIMDEP. The NL model nests alternatives in groups believed to be close substitutes. For example, in an application of beach use, where sites include ocean and bay beaches, nesting the ocean and bay beaches into separate groups may make sense. This has the effect of allowing beaches within each nest to be better substitutes for one another. The iia assumption still holds for sites within a nest, but it is relaxed for sites from different nests. In this way the model allows for a richer pattern of substitution. The error terms within each nest are viewed as having shared unobserved characteristics, which leads to the correlation of the error terms within a nest (e.g., all ocean beaches may have large waves and similar view amenities). A common application of the NL model is to nest sites and no-trip in separate nests. The reasoning here is that sites are more likely to serve as a better substitutes for one another than no-trip. Sites are likely to share unobserved characteristics. See Morey (1999) for more on NL models in recreation demand; Greene (1997, p.121) for an introductory discussion; and Morey, Rowe and Watson (1993) or Parsons and Hauber (1998) for applications.

The second method for relaxing the iia assumption is a mixed (MX) or random parameters logit. MX models are estimated using simulated probability techniques. Like the nested model, the mixed model is a generalization of a basic multinomial logit model that allows the parameters α and β to be random. The variation in each parameter is interpreted as a component of the error term in the site utilities, which leads to correlation among the utilities and a more general pattern of substitution. See Train (1999) for an example.

3.2.11 Calculate Access and Quality Change Values

In the final step, access or quality change values are computed. Value may be reported as:

- A mean per choice occasion value per person,
- A mean seasonal value per person,
- A total seasonal value for the population,
- A per trip value per person, and/or
- A total discounted present value of the site.

The parameters estimated in the previous step are used to calculate welfare changes using equations (35) and (36). The form of the expected maximum utility depends on the assumed distribution for the error terms in the model. With the Weibull distribution in the ML model, the expect maximum utility of a choice occasion is

$$(40)\quad eu^* = \ln\left\{\exp(\hat{\alpha}_0 + \hat{\alpha}_1 z) + \sum_{i=1}^{C}\exp(\hat{\beta}_{tc}tc_i + \hat{\beta}_q q_i)\right\}$$

This is the log of the denominator in equation (38). It is a preference-weighted utility index for a choice occasion in the sense that all alternatives are included and the higher the alternative's utility the larger its roll in the expression. Again, its form follows directly from the assumed distribution for the error terms. With equation (40), one can value per choice occasion site access and quality changes for one or more sites using equations (35) and (36).

Per choice occasion site loss for person n is

$$(41)\quad \hat{\mathrm{S}}_n = \frac{\ln\left\{\exp(\hat{\alpha}_0 + \hat{\alpha}_1 z) + \sum_{i=6}^{C}\exp(\hat{\beta}_{tc}tc_i + \hat{\beta}_q q_i)\right\} - eu^*}{-\hat{\beta}_{tc}}$$

where the first five sites are lost. A ^ denotes an estimated value using the estimation results and the subscript n on an explanatory variable that denotes the value of that variable for individual n.

For a quality change the per choice occasion value is

$$(42)\quad \hat{\mathrm{S}}_n = \frac{\ln\left\{\exp(\hat{\alpha}_0 + \hat{\alpha}_1 z) + \sum_{i=6}^{C}\exp(\hat{\beta}_{tc}tc_i + \hat{\beta}_q q_i^*)\right\} - eu^*}{-\hat{\beta}_{tc}}$$

where q^*_i is a vector indicating a quality change at some or all of the C sites. Again, as the distribution of the error term changes, the form of the expected maximum utilities in equation (42) changes. If the sample is randomly drawn, a simple mean of $\hat{s}_n$ is presented for the per choice occasion value. If the sample is stratified, the sample mean is adjusted to represent the population.

The seasonal value for each individual is the total number of choice occasions times the person's per choice occasion value. So, the mean seasonal per person value is

$$(43) \quad \bar{S} = T \cdot \bar{s}$$

where $\bar{s}$ is the sample mean per choice occasion value (adjusted for stratification if necessary) and T is the total number of choice occasions in the seasons. In a day trip model, T is the number of days in the season. The aggregate seasonal value over the population is

$$(44) \quad AS = \bar{S} \cdot POP$$

where *POP* is population of users and potential users. This might be all residents within driving distance of the site, all people owning a fishing license, or whatever defined the population from which the sample was drawn. *AS* is occasionally converted to a discounted present value assuming a constant following of recreation services from the site or sites. As shown with the single-site model in section 2.2.9, this is

$$(45) \quad PV = AS / i$$

where i is rate of discount. Equations (43) through (45) work for site access and quality changes alike.

Lastly, it is not unusual to see per-trip access or quality change values presented is a RUM analysis. A per-trip per person value is

$$(46) \quad \hat{t} = AS / TRIPS$$

where *TRIPS* is the total number of day trips by the relevant population. One can use an external estimate of the total number of day trips taken to the site over the relevant season or estimate the number of trips using the RUM model. In either case the per-trip value applies to the same scenario considered for site access or quality change above.

An alternative approach for estimating per-trip values arises when one has estimated a model that excludes no-trip form the choice set. In this case, one estimates per trip, instead of per choice occasion values, in the basic RUM model, and per-trip values flow naturally from the results. See Parsons and Massey (2003) for an example. Although convenient, these results come from a restricted model that disallows no-trip as an alternative.

3.3 A RUM Application

Matt Massey and I have estimated several RUM models of beach use in the Mid-Atlantic region of the United States using a 1997 beach use survey. See Parsons, Massey, and Tomasi (1999) or Parsons and Massey (2003). Here, I present another version of that model following the steps outlined in the previous section. The Mid-Atlantic beaches in our analysis include all of New Jersey, Delaware, and Maryland's ocean beaches—62 beaches.

We identified two impacts to analyze in our study (step 1). The first was the potential closure of beaches in the state of Delaware due to oil spills, water pollution, or other environmental episode. These were analyzed as lost site access. The second was beach erosion or narrowing of beaches in the state; this was analyzed as a quality change. Our quality measure was beach width. Beach width data was provided by the states, and there was sufficient variation across the beaches in this characteristic.

We defined our market as all residents in the state of Delaware (step 2). We recognized that there were a large number of out-of-state users. However, budget limitations and our key interest in exploring methodological issues in the model resulted in our narrow market definition. We also treated all uses of the beach as a single recreation type. We aggregated sunbathing, swimming, surf fishing, and so on.

We defined our choice set as all ocean beaches within a day's drive of Delaware residents (step 3). This included sixty-two beaches in four states. Beaches were defined using the political boundaries of beach communities, which is consistent with how people identify beaches in this area. For example, Ocean City, Maryland, Rehoboth, Delaware and Cape May, New Jersey were all separate beaches in our analysis. Every person had all sixty-two sites in their choice set.

Delaware is a small state with three counties. The most populated county is in the north. The ocean beaches are in the southern most county. We randomly sampled an equal number of residents over the age of sixteen from each of the three counties (step 4). We stratified in this way to avoid a population dominated by residents from the northern most county. We were not too concerned about low participation rates, because historical data convinced us that half or more of the population used the beaches in a typical year.

We thought a number of factors would influence day trips to the beaches in this region (step 5). Among these were:

- Trip cost
- Natural vs. developed beach
- Private or limited access
- Presence of boardwalk
- Availability of parking
- Availability of bathhouses and other facilities
- Beach width

These, more or less, were our targets as we set out to gather the data in the next step. For individual characteristic data we thought occupation, education, family composition, and flexibility in work would be important.

The site characteristic data were gathered from various sources: state departments of natural resources (including interviews with experts in those agencies), field trips, interviews with a scientist working on a data on the physical characteristics of the New Jersey beaches, tourist guides, maps, newspapers, and web sites (step 6).

We analyzed day trips only and assumed that all trips were single-purpose or at least that any side trips were incidental and easily ignored without introducing error (step 7).

We used a random mail survey of 1,000 Delaware residents in fall 1997 (step 8). An initial mailing of the survey was followed by a reminder postcard one week later and a second mailing of the survey after three weeks. Our response rate was 55%. Individuals were asked to report day, short overnight, long overnight, extended stay, and side trips separately. The survey was eight pages long. Respondents were asked to complete a two-page table of trips to sixty-two beaches. A map insert was provided to help identify beaches.

Trip cost was measured as the sum of travel expense, time expense, and beach fees (step 9). Travel cost was thirty-five cents times round trip distance plus tolls and parking fees. Many trips to New Jersey beaches are via toll roads, and on some routes a ferry is used to cross the mouth of the Delaware Bay. After the shortest route was determined using PC*Miler, the toll routes were identified and their cost computed. Many New Jersey beaches have fees. We used a per-day fee that was published in a local newspaper by beach for the 1997 season. Time costs were estimated as a wage proxy times round-trip travel time. The wage proxy was annual household income divided by 2,080. PC*Miler was used to compute round travel times.

We estimated a three-level nested logit model (step 10). At the first level a person decides to visit a beach or take no-trip. At the second level, and given that a person visits a beach, he or she decides to go to a beach in New Jersey or Delaware/Maryland. Finally, the person selects a site within the chosen region. The results are in Table 6. The site characteristics are variables appearing in site utility. The individual characteristics and inclusive value coefficients are variables appearing in no-trip utility.

As shown, the site characteristics that increase a site's day trip utility are boardwalk, amusements, good surfing, having a park, and good parking. The characteristics that decrease a site's day trip utility are private, being too wide or too narrow, and having high rises nearby. None of these are surprising. Length of the beach and park (state or federal) were insignificant, and facilities had the "wrong" sign. The alternative specific constant for Atlantic City and for New Jersey were positive and significant.

The inclusive value coefficients are parameters that characterize the degree of substitutability of among alternatives within a nest. For consistency with utility maximization, these coefficients should fall between 0 and 1. The closer the coefficient is to 0, the greater the degree of substitutability. A coefficient equal 1 is the same as not nesting. See Morey (1999) for an excellent discussion of NL models and interpreting inclusive values.

Table 6: RUM Model for Mid-Atlantic Beaches

Site Characteristics:

Variable	*Definition*	*Parameter Estimate (t-stat)*
tc	*Travel plus time cost*	*-.04 (63.9)*
Length	*Log of length of beach in miles*	*.13 (0.3)*
Boardwalk	*Boardwalk present = 1*	*.41 (6.3)*
Amusement	*Amusements nearby = 1*	*.48 (15.4)*
Private	*Private or limited access beach = 1*	*-.17 (6.3)*
Park	*State or federal park = 1*	*.04 (0.6)*
Wide	*Wide beach (= 1 if > 200 feet)*	*-.33 (12.8)*
Narrow	*Narrow beach (= 1 if < 75 feet)*	*-.20 (5.4)*
AC	*Atlantic City = 1*	*.42 (7.2)*

Variable	*Definition*	*Parameter Estimate (t-stat)*
Surf	*Good surfing = 1*	*.40 (15.5)*
HighRise	*High rises present on beach = 1*	*-.30 (9.4)*
ParkWithin	*Park located within the beach = 1*	*.25 (5.1)*
Facilities	*Bathhouse, restroom facilities present = 1*	*-.05 (1.1)*
Parking	*Parking available at beach = 1*	*.13 (1.8)*
New Jersey	*New Jersey beach = 1*	*.51 (33.9)*
Inclusive Value Coefficients:		
IV(NJ)	*Inclusive Value on New Jersey Beaches*	*.49 (36.9)*
IV(DE)	*Inclusive Value on Delaware/Maryland Beaches*	*.99 (38.7)*
IV (Beaches)	*Inclusive Value on All Beaches*	*2.06 (11.0)*
Individual Characteristics:		
Constant		*.25 (5.3)*
ln(age)	*Log of age*	*.20 (7.0)*
Kidsu10	*Number of kids under 10 in household*	*-.26 (9.4)*
Flextime	*Flexible time available in work schedule = 1*	*-.14 (3.4)*
Cottage (DE)	*Own beach property in Delaware = 1*	*-1.3 (25.5)*
Cottage (NJ)	*Own beach property in New Jersey =1*	*-.80 (16.4)*
Retired	*Retired = 1*	*.53 (10.5)*
Student	*Student = 1*	*-.90 (19.5)*
Parttime	*Work part time =1*	*-.56 (13.3)*
Workhome	*Work at home = 1*	*.94 (12.0)*
Volunteer	*Work as a volunteer = 1*	*-.16 (2.6)*
Sample Size	*565*	
Mean Log- Like	*-94.05*	

The individual characteristics that decrease no-trip utility and hence increase the probability of taking a beach trip are number of kids between 10 and 16 years old in the household, flexible time, owning a beach cottage in Delaware or New Jersey, student, working part time, or volunteer. No-trip utility increases with number of kids under 10, age, retired, and working at home.

For site access values, we considered the closure of each of the sixty-two beaches separately and the closure of groups of beaches (step 11). For the groups of beaches we considered the six northernmost Delaware beaches and eight southernmost beaches. For the beach erosion scenarios we considered a narrowing of all developed beaches in Delaware to less than 75 feet wide. We report seasonal per-person, seasonal aggregate (for Delaware residents), and discounted present values (for Delaware residents) in each case.

The mean seasonal per-person value, equation (41), for the loss of a single beach ranged from about \$5 for the northernmost beaches in New Jersey to about \$135 for the most popular beaches in Delaware and Maryland. These were Rehoboth and Cape Henlopen in Delaware, and Ocean City in Maryland. Recall that our sample considers only Delaware residents, and the scenario assumes that all other beaches remain open. These relative sizes make sense. The northern New Jersey beaches are far away and have many nearby substitutes. The more highly valued beaches are close to population centers and have many of the desirable characteristics noted in the results above.

These values translate into mean seasonal aggregate losses, using equation (43), that range from \$2.9 million for the least valued northern New Jersey beaches to \$77.8 million for the highest valued Delaware beaches. The relevant population is all residents of Delaware over the age of 16 in 1997. It is important to keep in mind that this value excludes overnight trips, people from other states, and non-use values.

We are particularly interested the loss of groups of beaches. For example, if an oil spill should occur, it is likely that more than a single beach would be lost. The northern beaches in Delaware are particularly vulnerable. Losing these simultaneously gives a mean seasonal per person loss of \$698, which translates into an aggregate seasonal loss of \$402 million. Losing the southern beaches gives a mean seasonal per person loss of \$554, an aggregate loss of \$319 million. Again, overnight trip, out-of-state residents, and non-use value are excluded.

Loss of beach width due to erosion is a major issue on beaches in the Mid-Atlantic. So we considered a scenario in which all developed beaches with a width of 75 feet or more are narrowed to less than 75 feet. If a beach is not developed it not as likely to erode and so was excluded. To calculate an individual's expected maximum utility with erosion, all developed beaches have site utilities computed with *wide = 0* and *narrow = 1*. The mean seasonal per-person loss for a narrowing of Delaware's beaches is about $76. The corresponding aggregate loss is $44 million.

3.4 Variations

There are a number of variations on the basic RUM model and an active research agenda advancing the technique (Herriges and Kling 1999). I will mention a couple of variations on the model I consider particularly important.

First, there is the application of Kuhn-Tucker models to recreation demand. These models have features of both RUM models and demand systems. They can be used to value quality changes and access, and it has been shown recently that these models may be used in settings with a large number of sites (Phaneuf, Kling, and Herriges 2000, Phaneuf and Smith 2002, and von Haefen, Phaneuf, and Parsons 2003). Kuhn-Tucker models are noted for their theoretical utility link between participation and site choice.

Second, there is increasing use of revealed and stated preference data in combination in the context of RUM models. This enables an analyst to consider behavior beyond the range of observable data and hence a much wider range of policy options see Adamowicz et al. (1994).

Third, a time dimension is being introduced into RUM models, which allows trips to have interdependence over time and allows for time periods (e.g., weekend vs. weekday) to be treated differently, see Adamowicz (1994).

Fourth, the mixed logit (MX) mentioned in section 3.2.10 has been widely adopted as a way to introduce complex patterns of substitution and unobserved heterogeneity into the model. This model is almost certain to see wider use (Train 1999).

Fifth, RUM models are often extended to include more than site choice. They may incorporate choice target species for fishing, choice of boat or shore in fishing, and even type of recreation. See Parsons and Hauber (1998) or McConnell and Strand (1994).

Finally, the basic model presented here may be extended to overnight trips, but care must be taken to account for multiple-purpose trips and to calculate lodging cost accurately. Neither is easy. See Shaw and Ozog (1999) and Hoehn et al. (1996).

4. CONCLUSION

The traditional single-site model and the contemporary RUM model are the two most widely used travel cost models in recreation demand. The RUM model is the modern workhorse. It is able to account to a broad array of substitutes, provides the possibility to value changes in site access as well changes in quality at one or more sites and it tells a convincing and defensible economic story. Furthermore, software is readily available for estimation. Most advances in travel cost modeling are taking place in the context of RUM models. The single-site model requires less data and is easier to apply. In circumstances where one is interested in access value at only one site and the number of substitute sites is not large, the single-site model is often used and is defensible.

ACKNOWLEDGMENTS

Several people reviewed all or parts of early drafts. For this I thank Vic Adamowicz, Chris Leggett, Joe Herriges, Tim Haab, Matt Massey, Ken Train, Dan Hellerstein, John Loomis, and Rick Freeman. I also thank Kim for her careful and quick proofreading of my messy manuscripts. Kevin, Patty, and Tom were superb editors. Thanks to them for the fun in the Rockies!

NOTES

1. This chapter overlooks other multiple-site approaches such as conventional demand systems and the hedonic travel cost model. Neither have been as popular as the RUM model. See Ward and Beal (2000) for a discussion of conventional demand systems and Brown and Mendelsohn (1984) for the hedonic travel cost. See Bockstael, McConnell, and Strand (1991) for a critique of the hedonic travel cost model.
2. Sometimes site utilities are written as $v_i = \beta_{tc}(y\text{-}tc_i) + \beta_q q_i$, and no-trip utility is written as $v_{0i} = \beta_{tc} y + \alpha$ where y is a measure of per-trip income. In this way trip cost is seen quite explicitly as reducing income available for other uses and hence lowering welfare. The coefficient β_{tc} (now positive) is also readily interpreted as a marginal utility of income. In estimation y is constant across sites and no-trip utility, provides no explanatory power for choice of alternative, and hence drops out in estimation.

REFERENCES

Adamowicz, W. L. 1994. Habit Formation and Variety Seeking in a Discrete Choice Model of Recreation Demand. *Journal of Agricultural and Resource Economics* 19(1):19-31.

Adamowicz, W., J. Louviere, and M. Williams. 1994. Combining Revealed and Stated Preference Methods for Valuing Environmental Amenities. *Journal of Environmental Economics and Management* 26:271-92.

Adamowicz, W. L., J. Swait, P. Boxall, and M. Williams. 1997. Perception Versus Objective Measures of Environmental Quality in Combined Revealed and Stated Preference Models of Environmental Evaluation. *Journal of Environmental Economics and Management* 32: 65-84.

Andrews, T. 1996. A Discrete Choice Model of Recreational Trout Angler Benefits in Pennsylvania. Unpublished manuscript, Department of Economics, West Chester University.

Ben-Akiva, M., and S. Lerman. 1985. *Discrete Choice Analysis.* Cambridge, MA: MIT Press.

Bockstael, N. E. 1995. Travel Cost Models. *Handbook of Environmental Economics.* Cambridge, MA: Blackwell Publishers.

Bockstael, N. E., and I. E. Strand. 1987. The Effect of Common Sources of Regression Error on Benefit Estimates. *Land Economics* 63(1):11-20.

Bockstael, N., W. M. Hanemann, and C. L. Kling. 1987. Estimating the Value of Water Quality Improvements in a Recreational Demand Framework. *Water Resources Research* 23(5):951-960.

Bockstael, N. E., W. M. Hanemann, and I. E. Strand, Jr. 1984. Measuring the Benefits of Water Quality Improvements Using Recreation Demand Models. Report to the U.S. Environmental Protection Agency. College Park, MD: University of Maryland.

Bockstael, N. E., K. E. McConnell, and I. E. Strand. 1988. Benefits from Improvements in Chesapeake Bay Water Quality. Report to the U.S. Environmental Protection Agency. College Park, MD: University of Maryland.

Bockstael, N. E., K.E. McConnell, and I. E. Strand. 1989. A Random Utility Model for Sportfishing: Some Preliminary Results for Florida. *Marine Resource Economics* 6:245-260.

Bockstael, N. E., K. E. McConnell, and I. E. Strand, Jr. 1991. Recreation: Measuring the Demand for Environmental Quality. New York: North Holland.

Bockstael, N. E., I. E. Strand, and W. M. Hanemann. 1987. Time and the Recreational Demand Model. *American Journal of Agricultural Economics* :293-302.

Brown, G. M., R. Congar, and E. A. Wilman. 1983. Recreation: Tourists and Residents. In *Assessing the Social Costs of Oil Spills: The Amoco Cadiz Case Study.* U.S. National Ocean Service, National Oceanic and Atmospheric Administration Report (NTIS PB83-100536). Washington, DC.

Brown, G., and R. Mendelsohn. 1984. The Hedonic Travel Cost Method. *Review of Economics and Statistics* 66:427-433.

Burt, O., and D. Brewer. 1971. Estimation of net Social Benefits from Outdoor Recreation. *Econometrica* 39(5):813-828.

Chen, H. Z., and S. R. Cosslett. 1998. Environmental Quality Preference and Benefit Estimation in Multinomial Probit Models: A Simulation Approach. *American Journal of Agricultural Economics* 80(3):512-520.

Cicchetti, C. J., A. C. Fisher, and V. K. Smith. 1976. An Econometric Evaluation of a Generalized Consumer Surplus Measure: The Mineral King Controversy. *Econometrica* 44(6):1259-1276.

Creel, M. D., and J. B. Loomis. 1990. Theoretical and Empirical Advantages of Truncated Count Data Estimators for Analysis of Deer Hunting in California. *American Journal of Agricultural Economics* 72:434-441.

Dilman, D. A. 2000. *Mail and Internet Surveys: The Tailored Design Method* (2nd edition). New York: Wiley.

Feather, P., and W. D. Shaw. 1999. Estimating the Cost of Leisure Time for Recreation Demand Models. *Journal of Environmental Economics and Management* 38:49-65.

Freeman, A. M. III. 1993. The Measurement of Environmental and Resource Values: Theory and Methods. Washington, DC: Resources for the Future.

Greene, W. H. 1997. *Econometric Analysis* (3rd Ed.). Upper Saddle River, NJ: Prentice Hall.

Haab, T. C., and K. E. McConnell. 1996. Count Data Models and the Problem of Zeros in Recreation Demand Analysis. *American Journal of Agricultural Economics* 78:89-102.

Haab, T. C., and K. E. McConnell. 2002. *Valuing Environmental and Natural Resources.* Cheltenham, UK: Edward Elgar.

Hackett, S. C. 2000. The Recreational Economic Value of the Eastern Trinity Alps Wilderness. School of Business and Economics, Humbolt State University.

Hanemann, W. M. 1999. Welfare Analysis with Discrete Choice Models. In *Valuing Recreation and the Environment: Revealed Preference Methods in Theory and Practice*. Edited by Herriges, J. A., and C. L. Kling. Cheltenham, UK: Edward Elgar.

Hauber, A. H. IV, and G. R. Parsons. 2000. The Effect of nesting Structure on Welfare Estimation in Random Utility Models: An Application to the Demand for Recreational Fishing. *American Journal of Agricultural Economics* 82:501-514.

Hausman, J. A., G. K. Leonard, and D. McFadden. 1995. A Utility-Consistent, Combined Discrete Choice and Count Data Model: Assessing Recreational Use Losses Due to Natural Resource Damage. *Journal of Public Economics* 56:1-30.

Hellerstein, D. 1999. Can We Count on Count Models? In *Valuing Recreation and the Environment: Revealed Preference Methods in Theory and Practice*. Edited by Herriges, J. A., and C. L. Kling. Cheltenham, UK: Edward Elgar.

Hellerstein, D., and R. Mendelsohn. 1993. A Theoretical Foundation for Count Data Models. *American Journal of Agricultural Economics* 75:604-11.

Herriges, J. A., and C. L. Kling, eds. 1999. *Valuing Recreation and the Environment: Revealed Preference Methods in Theory and Practice.* Aldershot, UK: Edward Elgar.

Herriges, J. A., C. L. Kling, and D. J. Phaneuf. 1999. Corner Solution Models of Recreation Demand: A Comparison of Competing Frameworks. In *Valuing Recreation and the Environment: Revealed Preference Methods in Theory and Practice.* Edited by Herriges, J. A., and C. L. Kling. Cheltenham, UK: Edward Elgar.

Hicks, R. L. and I. Strand. 2000. The Extent of Information: Its Relevance for Random Utility Models. *Land Economics* 76(3):374-385.

Hoehn, J. P., T. Tomasi, F. Lupi, and H. Z. Chen. 1996. An Economic Model for Valuing Recreational Angling Resources in Michigan. Michigan State University, Report to the Michigan Department of Environmental Quality.

Karou, Y. S1995. Measuring Marine Recreation Benefits of Water Quality Improvements by the Nested Random Utility Model. *Resource and Energy Economics* 17:119-136.

Kolstad, C. D. 2000. *Environmental Economics.* New York: Oxford University Press.

Laitila, T. 1999. Estimation of Combined Site-Choice and Trip-Frequency Models of Recreational Demand Using Choice-Based and On site Samples. *Economic Letters* 64:17-23.

Layman, R. C., J. R. Boyce, and K. R. Criddle. 1996. Economic Valuation of the Chinook Salmon Sport Fishery of the Gulkana River, Alaska, under Current and Alternate Management Plans. *Land Economics* 72(1):113-128.

Loomis, J. B., and R. G. Walsh. 1997. *Recreation Economic Decisions: Comparing Benefits and Costs.* State College, PA: Venture.

Lupi, F., and P. M. Feather. 1998. Using Partial Aggregation to Reduce Bias in Random Utility Travel Cost Models. *Water Resources Research* 34(12):3595-3603.

McConnell, K. E. 1986. The Damages to Recreational Activities from PCB's in New Bedford Harbor. University of Maryland, Department of Agricultural and Resource Economics. Report prepared for the Ocean Assessment Division, National Oceanic and Atmospheric Administration.

McConnell, K. E. 1992. On site Time in the Demand for Recreation. *American Journal of Agricultural Economics* 74:918-925.

McConnell, K E., I. Strand, and L. Blake-Hedges. 1995. Random Utility Models of Recreational Fishing: Catching Fish Using a Poisson Process. *Marine Resource Economics* 10:247-261.

McConnell, K. E., and I. E. Strand. 1981. Measuring the Cost of Time in Recreation Demand Analysis. *American Journal of Agricultural Economics* 63:153-156.

McConnell, K. E. and I. E. Strand. 1994. The Economic Value of Mid and South Atlantic Sportfishing. University of Maryland, Report to the U.S. EPA and NOAA. Department of Agricultural and Resource Economics, University of Maryland.

McFadden, D. 2001. Economic Choices. *American Economic Review* 91(3):351-378.

Mendelsohn, R., J. Hof, G. Peterson, and R. Johnson. 1992. Measuring Recreation Values with Multiple Destination Trips. *American Journal of Agricultural Economics* 74:926-933.

Montgomery, M. and M. Needelman. 1997. The Welfare Effects of Toxic Contamination in Freshwater Fish. *Land Economics* 73:211-223.

Morey, E. 1981. The Demand for Site-Specific Recreational Activities: A Characteristics Approach. *Journal of Environmental Economics and Management* 8:345-371.

Morey, E. 1999. Two Rums Uncloaked: Nested-Logit Models of Site Choice and Nested-Logit Models of Participation and Site Choice. Chapter 3 in *Valuing Recreation and the Environment: Revealed Preference Methods in Theory and Practice*. Edited by Herriges, J. A., and C. L. Kling. Cheltenham, UK: Edward Elgar.

Morey, E., R. D. Rowe, and M. Watson. 1993. A Repeated Nested-Logit Model of Atlantic Salmon Fishing. *American Journal of Agricultural Economics* 75:578-592.

Murray, C., B. Sohngen, and L. Pendelton. 2001. Valuing Water Quality Advisories and Beach Amenities in the Great Lakes. *Water Resources Research* 37(10):2583-2590.

Ozuna, T., and I. A. Gomez 1994. Estimating a System of Recreation Demand Function Using a Seemingly Unrelated Poisson Regression Approach. *Review of Economics and Statistics* 76:356-360.

Parsons, G. R. 1991. A Note on Choice of Residential Location in Travel Cost Demand Models. *Land Economics* 67(3):360-364.

Parsons, G. R., and M J. Kealy. 1992. Randomly Drawn Opportunity Sets in a Random Utility Model of Lake Recreation. *Land Economics* 68(1):93-106.

Parsons, G. R., and M. Needelman. 1992. Site Aggregation in a Random Utility Model of Recreation. *Land Economics* 68(4):418-433.

Parsons, G. R., and A. J. Wilson. 1997. Incidental and Joint Consumption in Recreation Demand. *Agricultural and Resource Economics Review* 26:1-6.

Parsons, G. R., and A. B. Hauber. 1998. Spatial Boundaries and Choice Set Definition in a Random Utility Model of Recreation Demand. *Land Economics* 74(1):32-48.

Parsons, G. R., P. M. Jakus, and T. Tomasi. 1999. A Comparison of Welfare Estimates from Four Models for Linking Seasonal Recreational Trips to Multinomial Logit Models of Site Choice. *Journal of Environmental Economics and Management* 38:143-157.

Parsons, G. R., D. M. Massey, and T. Tomasi. 1999. Familiar and Favorite Sites in a Random Utility Model of Beach Recreation. *Marine Resource Economics* 14:299-315.

Parsons, G. R., and D. M. Massey. 2003. A RUM Model of Beach Recreation. In *The New Economics of Outdoor Recreation.* Edited by Hanely, N., W. D. Shaw, R. E. Wright, and J. C. Coleman. Cheltenham, UK: Edward Elgar.

Parsons, G. R., A. J. Plantinga, and K. J. Boyle. 2000. Narrow Choice Sets in a Random Utility Model of Recreation Demand. *Land Economics* 76(1):86-99.

Peters, T., W. L. Adamowicz, and P. C. Boxall. 1995. Influence of Choice Set Considerations in Modeling the Benefits from Improved Water Quality. *Water Resources Research* 31(7):1781-1787.

Phaneuf, D. J., C. L. Kling, and J. Herriges. 2000. Estimation and Welfare Calculation in a Generalized Corner Solution Model With an Application to Recreation Demand. *Review of Economics and Statistics* 82:83-92.

Phaneuf, D. J. and V. K. Smith. 2002. Recreation Demand Models. Unpublished manuscript, North Carolina State University.

Randall, A. 1994. A Difficulty with the Travel Cost Method. *Land Economics* 70(1):88-96.

Shaw, D. 1988. 'On site Samples' Regression Problems of Nonnegative Integers, Truncation, and Enodgenous Stratification. *Journal of Econometrics* 37:211-223.

Shaw, W. D., and P. Jakus. 1996. Travel Cost Models of the Demand for Rock Climbing. *Agricultural and Resource Economics Review* 25(2):133-142.

Shaw, W. D., and M. Ozog. 1999. Modeling Overnight Recreation Trip Choice: Application of a Repeated Multinomial Logit Model. *Environmental and Resource Economics* 13(4):397-414.

Shonkwiler, J. S. 1999. Recreation Demand Systems for Multiple Site Count Data Travel Cost Models. Chapter 9 in *Valuing Recreation and the Environment: Revealed Preference Methods in Theory and Practice*. Edited by Herriges, J. A., and C. L. Kling. Cheltenham, UK: Edward Elgar.

Shonkwiler, J. S. and W. D. Shaw. 1996. Hurdle Count-Data Models in recreation Demand Analysis. *Journal of Agricultural and Resource Economics* 21(2):210-219.

Siderelis, C., G. Brothers, and P. Rea. 1995. A Boating Choice Model for the Valuation of Lake Access. *Journal of Leisure Research* 27(3):264-282.

Siderelis, C., and R. Moore. 1995. Outdoor Recreation Net Benefits of Rail-Trails. *Journal of Leisure Research* 27 (4):344-359.

Smith, V. K., W. H. Desvousges, and M. P. McGivney. 1983. The Opportunity Cost of Travel Time in Recreation Demand Models. *Land Economics* 59(3):259-277.

Smith, V. K., and W. H. Desvousges. 1985. The Generalized Travel Cost Model and Water Quality Benefits: A Reconsideration. *Southern Economics Journal* 52(2):371-381.

Smith, V. K., W. H. Desvousges, and M. P. McGivney. 1983. Estimating Water Quality Benefits: An Econometric Analysis. *Southern Economic Journal* 50(2):422-437.

Sohngen, B. 2000. The Value of Day Trips to Lake Erie Beaches. Unpublished report, Dept. of Agricultural, Environmental, and Development Economics, Ohio State University.

Train, K. E. 1999. Mixed Logit Models for Recreation Demand. Chapter 4 in *Valuing Recreation and the Environment: Revealed Preference Methods in Theory and Practice.* Edited by Herriges, J. A., and C. L. Kling. Cheltenham, UK: Edward Elgar.

Ward, F. A., and D. Beal. 2000. Valuing Nature with Travel Cost Models: A Manuel. Cheltenham, UK: Edward Elgar.

Chapter 10

THE HEDONIC METHOD

Laura O. Taylor
Georgia State University

1. INTRODUCTION

Heterogeneous or differentiated goods are products whose characteristics vary in such a way that there are distinct product varieties even though the commodity is sold in one market (e.g., cars, computers, houses). The variation in product variety gives rise to variations in product prices within each market. The hedonic method for non-market valuation relies on market transactions for these differentiated goods to determine the value of key underlying characteristics. For instance, by observing the price differential between two product varieties that vary only by one characteristic (e.g., two identical cars, but with one having more horsepower than the other), we indirectly observe the monetary trade-offs individuals are willing to make with respect to the changes in this characteristic. As such, the hedonic method is an "indirect" valuation method in which we do not observe the value consumers have for the characteristic directly, but infer it from observable market transactions.

The most common application of hedonic theory to environmental valuation involves housing markets. Analyzing the choices consumers make over housing is particularly well suited to many valuation exercises. The choices of housing location, and therefore neighborhood amenities, are observable. Often location choice is directly linked to an environmental amenity of interest. For example, housing locations can offer different scenic vistas (Patterson and Boyle, forthcoming). As such, the choice of a house and its associated price, implies an implicit choice over the environmental amenities linked to the house and their implicit prices.

Imagine the following hypothetical scenario in which there are two identical lakes, each with 100 identical homes surrounding them. All homes are lake-front, and all the characteristics of the homes themselves, the land, and the neighborhoods are identical across the properties. At the current equilibrium price of $200,000 per house, all 200 homes on either lake are equally preferred. Now, lets imagine that water clarity at one lake, Lake A for example, is improved. We assume that the improved water clarity is preferred by all households. Now if any home on Lake A were offered at the original equilibrium price of $200,000, consumers would uniformly prefer this house to any house on Lake B. In other words, at the current prices, there would be excess demand for the houses located on Lake A, and as such, the price of these houses must rise to bring the market into equilibrium. The price differential that results from the change in water clarity at Lake A is the implicit price consumers are willing to pay for that incremental increase in water clarity. This willingness to pay for water clarity is indirectly revealed to us through the market prices of the homes. For instance, if in the new equilibrium, houses on Lake A sell for $210,000, while houses at Lake B sell for $200,000, the "implicit price" associated with the increased water clarity is $10,000.

Of course, housing markets aren't so simple: housing choice depends on many characteristics such as structure of the house, amenities of the land, neighborhood and location. Yet, the fundamental intuition behind the hedonic method extends easily. By observing the choices consumers make over heterogenous commodities with varying prices, we can estimate the implicit price of one of the component characteristics of the commodity. These implicit prices or hedonic prices, under certain conditions, are equal to WTP or allow us to recover WTP.

Hedonic analyses have been reported as early as Fred Waugh's (1928) analysis of quality factors influencing asparagus pricing, and have been applied to markets as varied as automobiles, computers, VCRs, and appliances, and agricultural commodities.[1] Beach and Carlson (1993) estimated the factors affecting herbicide choice by farmers. Nimon and Beghin (1999) applied hedonic pricing to the apparel market to determine if consumers place a premium on clothes made from organically grown cotton.

Hedonic price functions have also been used to determine factors influencing the pricing of prescription drugs (Danzon and Chao 2000;

Cockburn and Anis 1998) and to evaluate the treatment of clinical treatment of illnesses such as depression (Berndt, Busch, and Frank 1998). Cultural applications of the hedonic method have been used to analyze market sales for Picasso paintings (Czujack 1997), Impressionist paintings (Chanel, Gerard-Varet, and Ginsburgh 1996), classical music (Harchaoui and Hamdad 2000), and have even been extended to examining the factors that influence dowries in South Asia (Rao 1993).

This chapter focuses primarily on housing markets and how they may be used to value environmental amenities. The application of the hedonic method to labor markets is also briefly reviewed in section 6. Other reviews of the hedonic method can be found in Palmquist (1991, 2000) and Freeman (1993).

Hedonic analysis of markets for differentiated goods consists of two related steps often referred to as a first-stage and second-stage analysis. In a first-stage analysis, the hedonic price function is estimated using information about the prices of a differentiated commodity and the characteristics of the commodity. This analysis allows researchers to recover the implicit prices of characteristics and reveals information on the underlying preferences for these characteristics as discussed in section 4. Once the first-stage analysis is completed, researchers may then use the implicit prices obtained in the first stage to estimate the demand functions for the *characteristics* of the commodity. This step is referred to as a second-stage analysis and is discussed in section 5. First-stage analyses are the most common application of the hedonic method because the data requirements are minimal and the needed economic insights often only require marginal price information. Second-stage analyses require significantly more data that are often difficult to obtain, and the modeling is more complex.

For the purposes of non-market valuation, Rosen's (1974) seminal article was important because it developed a utility theoretic framework for understanding the market process generating a hedonic equilibrium, and thus established the connections between consumers' preferences for characteristics of heterogeneous goods and the equilibrium price function. Next we review this theory to provide the basis for our later discussion of implementing the method and the econometric issues related to a first- or second-stage analysis.

2. THE HEDONIC PRICE FUNCTION: THEORY

An example of hedonic price function for housing is presented to facilitate discussions throughout this chapter. This example has common elements with almost every hedonic price function estimated with housing markets for the purposes of non-market valuation. The example is based on a study by Boyle, Poor, and Taylor (1999), but I will present some of their results in this chapter that were not directly reported in the journal article.

Boyle, Poor, and Taylor estimated hedonic price functions for lakefront properties in Maine. Sales prices of properties, mostly summer cottages, were analyzed as a function of the characteristics of the structure on the property and important characteristics of the land such as its location relative to the nearest town and the water quality of the lake on which the property is located. Maine lakes are known for their high water quality, however this quality is being compromised in many Maine lakes by eutrophication resulting from nonpoint pollution. The physical manifestation of eutrophication is reduced water clarity. Thus, water clarity is the measure of lake water quality that is used by the authors in the hedonic price function.

Boyle, Poor, and Taylor (1999) estimate a hedonic price function for each of four geographically distinct markets. The estimated hedonic price function for one market is:

$$\begin{aligned} P = {} & 25{,}899 + 6{,}790 \cdot \ln(SQFT) + 83 \cdot FRONT - 3{,}919 \cdot DNSTY + 1{,}516 \cdot DIST + \\ & 11{,}572 \cdot HEAT + 23{,}465 \cdot BATH - 17{,}579 \cdot LKWATER + 2.057 \cdot WQ, \end{aligned} \tag{1}$$

where the dependent variable is the sales price of a property, and the independent variables are, respectively: square feet of the structure on the property (mean = 750 square feet); the length of the property's frontage on the lake (mean = 143 feet); the number of lots per 1000 feet of frontage adjacent to either side of the property (mean = 8.9 lots); the distance between the property and the nearest town (mean = 9.4 miles); dummy variables included that designate whether or not the structure has central heating, a full-bath, or if the property uses lake-water as its primary source of water. The last independent variable in equation (1) is a measure of water quality (WQ) which is equal to LKAREA*ln(WC), where LKAREA is area of the lake on which the property is located (mean = 4,756 acres), and WC is the depth of water clarity of lake (mean = 3.9 meters).

This application of the hedonic method to value lake water clarity in Maine will be used to illustrate conceptual issues discussed throughout this chapter. Next, the theoretical underpinnings of the hedonic method are presented which allows us to precisely interpret the hedonic price function coefficient estimates. Section 3 summarizes the steps involved in implementing the theory and these steps are then discussed in detail in the following sections.

2.1 The Hedonic Equilibrium

The basic utility maximization problem including the consumer's choice over a differentiated product is presented in chapter 2, section 3.4. In this chapter, we will discuss this theory in more depth using the following notation and assumptions. Let Z represent the differentiated commodity with characteristics $\underline{z} = z_1, z_2, z_3, \ldots z_n$. The differentiated commodity is assumed to be sold in a perfectly competitive market and the interactions of the many producers and consumers together determine an equilibrium price schedule for the differentiated commodity, $P(\underline{z})$.[2] The equilibrium price for any one model of the differentiated good is a function of the characteristics of that particular model. As such, the consumer can determine the price he/she pays for the good by choosing which model to purchase. However, it is important to note that the consumer takes the entire price schedule $P(\underline{z})$ as exogenous.

Consumer utility is defined over two goods: Z, the differentiated good, and x, a composite commodity representing all other goods (i.e., income left over after purchasing Z). Consumer j, with demographic characteristics α^j has utility defined as:

(2) $$U^j(x, z_1, z_2, \ldots, z_n; \alpha^j).$$

If we assume that the consumer purchases only one unit of the differentiated commodity, a reasonable assumption for the housing market, the budget constraint is $y^j = x + P(\underline{z})$. The consumer seeks to maximize utility by choosing the model of the differentiated product, $\underline{z}$, and the amount of x to purchase, subject to this budget constraint. The consumer will choose $\underline{z}$ and x such that the following is satisfied for each z_i:

(3) $$\frac{\partial P}{\partial z_i} = \frac{\partial U/\partial z_i}{\partial U/\partial x},$$

which is equivalent to equation (32) in chapter 2, and indicates that the marginal rate of substitution between any characteristic, z_i, and the composite numeraire commodity, x, is equal to the rate at which the consumer can trade z_i for x in the market (i.e., the ratio of the implicit price for z_i and the price of the numeraire =1)).

For our purposes, it is convenient to describe the optimal bid a consumer will make for any specific product variety. The bid function, θ, describes the relationship between the dollar bid consumer j will make for Z as one or more of its component characteristics are changed while utility and income remain constant. Equation (2) can be used to define the bid function formally by recognizing that income less the bid a consumer makes for Z is the amount of money left over to spend on the numeraire, x. Thus, the relationship:

(4) $$U^j(y_0 - \theta, \underline{z}, \alpha^j) \equiv U_0$$

indicates how a consumer's optimal bid must vary in response to changes in $\underline{z}$ if utility and income are held constant. Solving equation (4) for θ indicates that $\theta^j = \theta(\underline{z}, y_0, U_0^j, \alpha^j)$, where y is exogenous income and U_0^j is a fixed level of utility. Maximizing utility in (4) yields the result that the the marginal bid a consumer is willing to make for z_i ($=\partial\theta/\partial z_i$, which is denoted as θ_{zi}), will equal the marginal rate of substitution between any characteristic, z_i, and x. Relating this result to that in equation (3) indicates that the necessary conditions for utility maximization are that the marginal bid a consumer places for a characteristic must equal the marginal price of that characteristic ($=\partial P(\underline{z})/\partial z_i$, which is denoted as P_{zi}).

For the supply-side of the market, we can describe a firm with characteristics δ^k as seeking to maximize profits, $\Pi = H^*P(\underline{z}) - C(H, \underline{z}, \delta^k)$, where H is the number of units of Z that the firm produces and C(·) is a well-behaved cost function. Again, the firm faces the exogenous, equilibrium price schedule $P(\underline{z})$ when determining its choices. Although firms can affect the prices they receive for their products by varying the characteristics of the product, no single firm can affect the price schedule. In this formulation, we assume the firm produces only one model of Z. Thus, the firm chooses what type to produce, Z^k, and then chooses how many of that type to produce.[3] Similar to the consumer's problem, we may describe the behavior of a firm in this differentiated goods market by an offer function: $\varphi^k = \varphi(\underline{z}; H, \Pi_0, \delta^k)$, which describes the amount of money a firm is willing to accept for any particular variety of Z, holding constant the number of units of that variety the firm produces, H, and its level of profit, Π_0. The offer function is defined by: $\Pi_0 = H^*\varphi^k - C(H, \underline{z}, \delta^k)$, and at the optimum, it will be the case that the marginal price a firm is willing to accept for z_i, φ_{zi}, will equal the marginal cost of producing that characteristic per unit of the differentiated good, C_{zi}/H.

The bid and offer functions, and their properties may be easily described using Figure 1, which originally appears in Rosen (1974). Figure 1 illustrates the equilibrium price schedule, $P(\underline{z})$ as it varies with changes in z_1, holding the level of all other characteristics constant. $P(\underline{z})$ is drawn such that the total price paid for z_1 increases at a decreasing rate. We might expect this relationship in many cases.

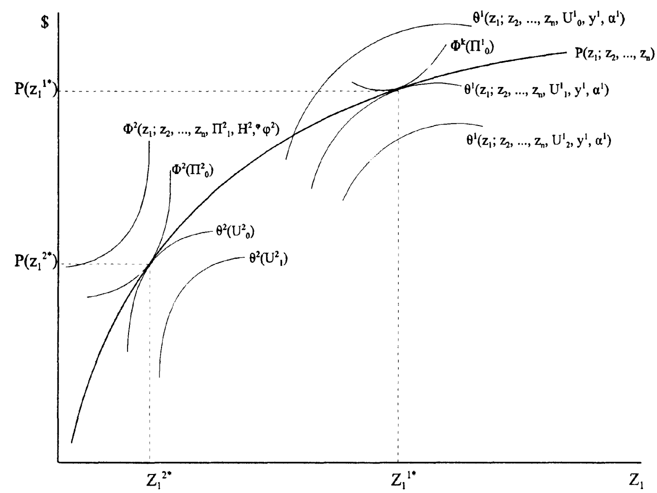

Figure 1. The Hedonic Price Function

For instance, let's say z_1 represents the square feet of living space in a house. We might expect a smaller price differential between a 5,000 and a 5,300 square foot house as compared to the price differential between a 1,000 and a 1,300 square foot house. However, the equilibrium price schedule $P(\underline{z})$ may theoretically take any form, a point discussed in more detail later.

Also depicted in Figure 1 are the bid functions for two consumers, θ^1 and θ^2. Along any bid function contour, only the level of z_1 changes; the level of all other characteristics, income, and utility are constant. Bid functions are concave in $\underline{z}$ (i.e., optimal bids increase at a decreasing rate in $\underline{z}$) and higher levels of utility for a consumer are represented by bid function contours closer to the horizontal axis. Intuitively, a lower bid for the same level of z_1, implies a higher level of utility because more money is left over to spend on x.[4] The optimal choice of z_1 is where the consumer reaches the lowest possible bid-function while still being able to participate in the market. For consumer 1, this occurs at a quantity z_1^{1*} and a total bid price of $P(z_1^{1*})$, which is the point at which the bid function is tangent to the equilibrium price schedule in Figure 1. For consumer 2, this optimal choice occurs at z_1^{2*} and the consumer pays a total price of $P(z_1^{2*})$. At each consumer's optimal choice of z_1, the marginal bid $(\partial\theta/\partial z_1)$ equals the marginal price $(\partial P(\underline{z})/\partial z_1.)$.

Offer functions for two firms, φ^1 and φ^2, are also depicted in Figure 1. Offer functions are convex in $\underline{z}$ as optimal offers increase at an increasing rate, holding profits and all else constant. Offer functions further away from the horizontal axis represent higher levels of total profit. The firm's optimal choice of what level of z_1 to produce in each of its H units of the differentiated product is where the firm reaches the highest possible offer function while still being able to participate in the market. For firm 2, this occurs at quantity z_1^{2*} and a total offer price of $P(z_1^{2*})$.

Figure 1 illustrates that the hedonic price function is simply an envelope of the equilibrium interactions between all buyers and sellers of a differentiated good. As such, it is possible for the hedonic price function to take any shape. The key point is that with relatively small data requirements, such as information on product types and their sales prices, we can recover the marginal implicit prices for any component characteristic of Z. The marginal price is equal to the marginal WTP for that characteristic by consumers, $\partial\theta/\partial z_i$. In the Maine lakes example, equation (1) indicates that consumers' value for a one-unit (one foot) increase in a property's feet of lake frontage is $\partial\theta/\partial FRONT = \partial P/\partial FRONT =$ \$83. This is the fundamental insight underlying the hedonic method. We seek to estimate the parameters of the hedonic price function so we may recover information about the marginal value consumers place on characteristics.

3. THE HEDONIC PRICE FUNCTION: ESTIMATION

The data needs for estimating the marginal implicit prices through the hedonic price function are fairly simple. In this section, we discuss the details of estimating the hedonic price function and some important considerations Table 1 summarizes the steps and considerations for estimating the hedonic price function.

Table 1. Summary of Steps for Estimating the Hedonic Price Function

Define Value to be Estimated
- Marginal willingness to pay as revealed by marginal implicit prices.

Collect Data on Property Value (Dependent Variable)
- Sales Price: preferred measure of value, may need to consider selection bias.
- Tax Assessment or Homeowner Survey of Value: measurement error may be a significant concern.
- Rental or Lease Prices: appropriate for some applications, timing issues can be of concern, care should be taken when interpreting the implicit prices.

Choose Functional Form for the Hedonic Price Function
- Linear usually not appropriate. Non-linear functions imply non-constant marginal prices.
- Semi-log functional form often used, care must be taken when interpreting the coefficient estimates for the dummy variables.
- Researcher judgement must be applied, and expectations about relationships between certain characteristics and sales price will guide choice of functional form.

Address Spatial Dependence and Correlation
- Spatial relationships among properties similar to autocorrelation in time-series is likely. Tests and corrections for spatial dependence are available pre-packaged statistical packages and should be implemented when possible. These packages require properties to be spatially-located (i.e., assigned latitude/longitude markers).

Compute Welfare Measures (Section 4)
- For localized changes in amenities, the change in sales price resulting from the change in the amenity is the measure of net-benefits if there are no transactions costs associated with moving between properties. If there are transactions costs, the change in price net of transactions costs measures net-benefits (or is an upper-bound on net benefits).
- For non-localized changes in amenities, a second-stage demand analysis is most appropriate for computing a bound on net-benefits.

3.1 Dependent Variables

We seek to estimate the equilibrium price schedule, P(z), which is the function relating transaction prices (sales price) to the characteristics of the product. Thus, the appropriate dependent variable is sales price, which is the observed outcome of a market assumed to be in equilibrium. As with any durable asset, sales price of a property represents the discounted present value (PV) of all future rents (R) from the property:

$$PV = \sum_{t=1}^{T} \frac{R_t}{(1+r)^t} \tag{5}$$

where r is the discount rate, and T is the expected life of the property. The implicit price for a characteristic of a house can then be thought of as representing the discounted presented value of the stream of benefits expected from that characteristic. Thus, expected changes in future benefits of the house would be incorporated into current sales prices.

Sales prices may be obtained from several sources. First, local government offices such as county courthouses record each deed transfer and its transfer price. This information is publicly available; however, these records are almost never recorded electronically by these government offices. The data used by Boyle, Poor, and Taylor (1999) were collected at town offices using researchers' laptop computers. Collection of sales price data in this manner can be quite time consuming, if not prohibitive when large numbers of observations are needed. For larger metropolitan areas, private firms have begun recording local sales data for re-sale in electronic formats. A recent internet search revealed many private vendors of property data. While prices vary, one company's price for academic purchasers was $1,000 per county in the Atlanta, Georgia metropolitan area for a data file containing the most recent sales price of every property in the county (well over 250,000 observations in one county alone).

When sales prices are too difficult to obtain, other estimates of housing value may be used such as homeowner or tax assessor estimates of value. Tax assessors values are often more readily available. Often, county tax assessor offices maintain electronic databases of every property's assessed value within the county. These records are public information and generally are available upon request.

Surveys of homeowners may also be used to estimate property values. For instance, a researcher may conduct a mail survey and ask owners the price they paid for their house and the purchase year. National surveys of homeowners are available which ask homeowners to estimate the value of their homes. Example databases include the American Housing Survey collected by the U.S. Census Bureau (www.census.gov/hhes/www/ahs.html) and the Panel Study of Income Dynamics, which is collected annually by the University of Michigan (www.isr.umich.edu/src/psid/pdf.html).

Tax assessor and homeowner estimates of value may not correspond exactly with actual sales prices, but it is hoped they are reasonably good approximations. As approximations, the potential for measurement error must be considered as discussed in Section 3.1.1. Also, in the case of tax-assessed prices, it is of some concern that the implicit price estimated in a hedonic price function may not reflect the owners' (or market) value of these characteristics, but rather the values of a handful of tax assessors who work for a county.

Rental prices of properties have also been used as the dependent variable in some hedonic analyses of housing markets (Pollard, 1980). Taylor and Smith (2000), Osborne (1995), and Wilman (1984) used rental prices of vacation properties to estimate values for proximity to the shoreline. In this case, implicit prices are actually implicit rents, representing the additional value to rent from an additional unit of a particular characteristic. As such, it is important to note that while future changes in amenities may be capitalized into sales prices, they are not expected to be capitalized into current rents (as equation 5 makes clear). Although this does not diminish the usefulness of rental prices for hedonic analysis, the researcher has to be clear in the interpretation of the hedonic price function that implicit prices represent current-period values, rather than changes in asset values.

3.1.1 Measurement Error in the Dependent Variable

In terms of estimating the hedonic price function, measurement error in the dependent variable is not a problem if the measurement error is uncorrelated with the set of independent variables. In this case, the estimated implicit prices would remain unbiased, although efficiency will be reduced. The empirical question is whether tax assessor or homeowner estimates of value (sale prices) are uncorrelated with the characteristics of the properties. Ihlanfeldt and Martinez-Vazquez (1986) investigate this question and compare tax assessor and homeowner estimates of property values to actual sales prices. Their results indicate that the errors in both tax assessor and homeowner data were related to a number of structural characteristics, thus causing potentially biased estimates. In the sample used by Ihlanfeldt and Martinez-Vazquez, a smaller number of variables was affected by tax assessor estimates (5 of 22 variables) as compared to homeowner estimates (10 of 22 variables).

Kiel and Zabel (1999) provided a comparison of owner assessments with actual sales prices for 3 metropolitan areas over approximately an 11-year period. Their results indicate that although owners generally overestimate the value of their homes by approximately 5%, the difference in owner valuation and actual sales price is not related to any housing or owner characteristic other than the tenure in the home (those living longer periods of time in their house at the time of the survey provided more accurate values). These results held regardless of housing cycles, and they suggest a considerable promise for the use of data from owner surveys, such as the American Housing Survey and the Panel Study of Income and Dynamics, for estimating unbiased implicit prices for housing characteristics.

It seems likely that the appropriateness of survey or assessor data will be case specific, and one can only be assured of the appropriateness of these measures of housing value if sales prices are available for comparison. Unfortunately, this prescription is not particularly helpful for the researcher who does not have actual sales prices available.

Also, sales prices themselves are not without shortcomings. While sales prices may be recorded without error by county transfer deed records, not all sales are "arms-length" transactions. In some cases, properties may be transferred to family members or from one business entity to another at prices much less than market value. These observations would not be appropriate for inclusion in the data set as the hedonic price function is an equilibrium function arising from competitive market bidding for each property. Unfortunately, recognizing transactions that are not arms-length is not always easy. An intuitive approach is to omit any observations on prices that are implausibly low. However, the researcher must be carefully define what is "too low" and should have a good understanding of the market.

3.1.2 Sample Selection Bias

A second issue with sales prices, which the use of homeowner or tax assessor values may overcome, is the potential for sample selection bias. Here, the concern is that properties with specific unobservable characteristics may be less likely to be placed on the market for sale, or less likely to sell once placed on the market. Thus, sales prices for these properties are never observed, so they are systematically not included in the hedonic data. In this case, Ordinary Least Squares (OLS) estimation produces biased coefficient estimates.

One way to test for sample selection bias in sales price data is to compare hedonic implicit price estimates obtained from a database of sales prices to those obtained from a database of all property values. Ihlanfeldt and Martinez-Vazquez (1986) took this approach and found no evidence of selection bias in a large, general sample of Atlanta, Georgia homeowners.

Sample selection bias may be particularly important in studies focusing on housing that are proximate to amenities or dis-amenities. If the selection criteria for what properties are included in the analysis (e.g., homes that are located within a certain distance of a hazardous waste site) are related to unobserved characteristics that determine if we observe a sales price, then OLS estimation of the implicit prices based on this truncated sample will be biased. To my knowledge, an explicit consideration of this issue has not been undertaken. Huang and Palmquist (2001) developed a related model in which they explicitly consider the simultaneity between the duration a property is on the market and its sales price. They argued that housing located near a disamenity may suffer two impacts: reduced sales prices and reduced probability of sales. They found that location near a noisy highway does not adversely affect duration on the market, but does adversely affect sales prices. However, they do not test for selection effects among the mix of properties offered for sale near the highways as compared to those further away, clearly an area where further research is warranted.

3.1.3 Aggregation Level

Ideally, one would use observations on individual house transactions to estimate a hedonic price function. While most hedonic studies do use data on individual housing units, some studies still use census-level data (e.g., median house value in a census tract or block, and associated median housing characteristics). If available, information on individual units is preferable, especially when a study is focusing on externalities that are likely to be localized. For instance, the effects of an environmentally contaminated site on nearby property values may be limited to a small geographic area. Aggregate housing data might not be able to distinguish these price effects if the geographic extent of the negative impacts is small relative to the geographic area that is aggregated.

The potential shortcomings of aggregated data are highlighted by Schultz and King (2001), who compared the results of hedonic regressions using census block-level and tract-level information to value open space in Tucson, Arizona. Although the authors did not find significant variations in implicit price estimates across these two data, they did find important variations between their results, and those of King, White, and Shaw (1991) who estimated a similar hedonic using individual house-sale information for the same market. As compared to the implicit prices estimated with house-level sales data, the implicit prices estimated using Census data understated the value of regional or district parks and overstated the value of neighborhood parks, golf courses, and some types of natural habitats (up to three-fold), and had the opposite sign for some types of wildlife habitat.

3.2 Independent Variables

As discussed in section 2.1, the hedonic price function is an envelope function relating sales price of a differentiated good to its characteristics. As such, the independent variables in a hedonic price regression are only those product characteristics that affect price. Characteristics of the consumers and sellers of the product *do not* belong in the hedonic price regression. Not all product characteristics belong in the regression either. Characteristics that vary across product types, but do not affect price are not included as a regressor (e.g., the color of a product type often does not affect its price).

There is no simple way to determine exactly what characteristics belong in a particular hedonic regression. Researchers must use their knowledge of the market to determine what characteristics are relevant for determining price in their market. In general, most property value studies include three types of characteristics: (1) the house and the lot, (2) features of the neighborhood such as the quality of the school district, the level of crime, and the environmental health and (3) the property's location such as its proximity to a recreation area or an employment center. How a researcher might assemble these three categories of information follows. Also discussed in more detail regarding the choice of characteristics and how to formally model the hedonic price function.

Housing and lot characteristics are typically gathered from tax assessor records or homeowner surveys. The retail vendors of property data typically begin with the tax assessor database and then append the most recent sales price of each property. For instance, the vendor mentioned in section 3.1 (charging $1,000 per county for the most recent sales price of every property on the tax-assessor's roles) also included in the database the address and all property characteristics as recorded by the county tax assessors. The Panel Study on Income Dynamics survey of homeowners (see section 3.1) includes information on housing characteristics. Housing characteristics are also available at the census tract level from the U.S. Census Bureau (each observation is an average of the individual houses within a census tract). As micro-level data has become so widely available, census-level data is seldom applied in the recent studies (see section 3.3 for more on this point).

As with all secondary data, the researcher should carefully examine the data for apparent errors and outliers. For example, a house with 1 bedroom and 6 baths is likely a data error. This information should be verified by the researcher, and if it cannot be verified, the observation can be dropped from analysis. However, considerable controversy exists about data trimming, and observations should only be dropped judiciously.

Lastly, variables may have missing observations for a significant portion of the data. There are two ways to handle missing observations. The first is to allow these observations to be dropped from your statistical analysis. Dropping these observations is a problem only if there is a systematic process determining which observations have missing data, thus introducing selection bias. Alternatively, it is possible to replace missing values with the mean value for this variable in the sample. This approach is suggested in Maddalla (1988), but it is not common in the hedonic housing literature.

Neighborhood and location-related characteristics are usually sparse in tax assessor files, and so this information is generally appended to each property's record by the researcher. A common source of neighborhood information is the U.S. Census Bureau, which publishes information such as racial and income composition of neighborhoods by census tracts (www.census.gov). Census data may be matched to property data using either zip codes (a cross-walk between zip codes and census tracts may be found at the Census website) or with a base-map of census tract boundaries (also available at the Census website). Of course, a researcher may have other sources of neighborhood information, such as school district test scores or environmental quality of a neighborhood (e.g., ambient air quality conditions), that is available from state and local governmental or non-governmental sources.

Location-related characteristics are generally created by the researcher and are specific to each application. For instance, the only location-related characteristic in the Boyle, Poor, and Taylor (1999) study was the distance of the property to the nearest town. Ihlanfeldt and Taylor (2001) conducted a study of commercial and industrial property sales in Atlanta, Georgia and location-oriented variables included the distance from each property to the central business district, the nearest highway exit, the airport, and the nearest light rail station. Also included was the distance from each property to the nearest site with known environmental contamination. Each variable was created using a geographic information systems software package (see section 3.5).

In addition to specifying which characteristics belong in a hedonic price function, the researcher must consider how the relevant characteristics should be measured and how they should enter the hedonic price function. Major considerations for choosing regressors and developing the hedonic price function specification for estimation are outlined below.

3.2.1 Selection of Independent Variables

What characteristics are relevant to include in a hedonic price function? There is no easy answer to this question. Basic components are common to all studies: size of the lot, size of the structure, neighborhood descriptors, and location descriptors. The variables used to describe each feature varies, but typically at least one element of each category is included in a hedonic price function for housing. Lot size is typically measured in acres, and is commonly available. The structure size is typically measured by square footage of the house or the number of bedrooms and baths. Quality indicators can include the age of the structure and amenities, such as the presence of a fireplace. Quality indicators are often sparse. Neighborhood characteristics often include the quality of the school district and census information about the neighborhood such as racial composition and median household income. Neighborhood characteristics may also include an environmental amenity such as ambient air quality. Lastly, locational variables such as the proximity to downtown (important for urban property markets), proximity to amenities and disamenities such as open space, shopping centers, mass transit station, and landfills are often included.

Which variables within a category should a researcher include? Should as many variables as possible be included within each category? While a "kitchen sink" approach might be tempting, it is not necessarily advisable. Adding irrelevant variables to a regression, if they are systematically related to other independent variables, can lead to increased standard errors and Type II errors in hypothesis testing (failing to reject the null hypothesis of no significant influence of a characteristic on price when it indeed *does* influence price). On the other hand, using too few variables can lead to dropping relevant variables, which would bias the coefficient estimates. Previous research by Atkinson and Crocker (1987) found that including large numbers of housing characteristics in hedonic price functions may result in increased unreliability of parameter estimates. However, Ihlanfeldt and Taylor (2001) found that increasing the number of explanatory variables introduces little bias with respect to the coefficient estimates of the environmental variables of interest.

Standard diagnostic methods for evaluating multi-collinearity exist (e.g., Belsley, Kuh, and Welsch 1980), and remedies that can be applied when variables in the hedonic price specification are suspect (e.g., principle component methods (Greene 1994)). Unfortunately, no simple analytic answer exists, and the researcher must thoughtfully develop a modeling approach, review related hedonic studies, and test the robustness of results to assumptions regarding which variables to include whenever possible.

3.2.2 Measurement Error in the Independent Variables

In addition to choosing the variables that determine price, the researcher has to be concerned with how the variables are measured. Ideally, one would have each characteristic measured in a manner consistent with the perceptions or understanding of the characteristic by the market participants. Again, this is because the hedonic price function is capturing the equilibrium outcomes of the real-world interactions between buyers and sellers of a commodity. Thus, the manner in which each characteristic is measured and used by the researcher should be consistent with the measures perceived by the market participants. For many structural or lot characteristics, this is a relatively straightforward task. For instance, the number of bedrooms as measured in a current tax assessor database is likely to be equal to the homeowner's assessment of the number of bedrooms in the house at the time of purchase. However, for other characteristics important to sales prices, such as ambient environmental conditions (e.g., air quality at the site), it may be difficult to quantify this characteristic in a manner that accurately reflects buyers' and sellers' perceptions of the characteristic. For instance, in the Boyle, Poor, and Taylor (1999) study of Maine freshwater lakes, the measure of lake water clarity was provided by the Maine Department of Environmental Protection (DEP), and the clarity was measured monthly at a specific site on each lake during the summer months. Prospective purchasers may not have visited the lake during those months (and water clarity varies throughout the year on Maine lakes), or the properties may not have been located near the site where the DEP took measurements. As such, the objective or scientific measures of clarity collected by the DEP may not have coincided with water clarity as perceived by prospective property purchasers (see Poor et. al. 2001).

Although scientific measures or the proxy variables used by researchers are hoped to be reasonably well correlated with the "true" factors that are capitalized into a property's value, the use of proxy variables can nonetheless introduce the potential for errors-in-variables problems leading to biased estimates of *all* coefficients. If only one variable is measured with error, it is the case that improving the measurement of that variable will reduce the bias in all coefficients. However, when more than one variable is measured with error, reducing measurement error in one variable will not necessarily reduce the bias in the other regression coefficients (Garber and Klepper 1980).

Several studies have focused on the impacts of measurement error on hedonic price estimates (e.g., Atkinson and Crocker 1987, Graves et al. 1988). Graves et al. (1988) assumed that the proxy variables used for neighborhood characteristics were measured with error and tested to determine the extent to which these measurement errors may bias the coefficients of two variables measuring air quality: an index of visibility and a measure of the total suspended particulates (TSP). Their results indicated that measurement error in the neighborhood characteristics affected the index of visibility, changing both sign and significance of the variable, but did not affect measures of TSP. Boyle and Taylor (2001) also found that measurement error in structural and property characteristics of lakefront properties in Maine did not affect hedonic price estimates of the impacts of lake water clarity on sales price.

Results from previous research are more mixed when the measurement error is in the environmental variable of interest. Graves et al. (1988), find that even small measurement error in the two air quality measures resulted in unstable coefficient estimates such that even qualitative conclusions were "dubious," indicating "the need for additional efforts to ensure that the environmental quality variables are carefully measured" (p. 232). More recently, Poor et al. (2001) used the Maine lake data described earlier to compare scientific readings of water clarity with subjective measures of clarity gathered from a survey of property owners. Both the objective and subjective measures of quality were measured as linear feet of water clarity, and so the objective and subjective measurements were directly comparable. Their data suggested that property owners tended to *under*-report water clarity as compared to objective measures, and non-nested J-tests indicated that the objective measures were better predictors of sales prices than were the subjective measures. Although this finding is encouraging for the use of objective measures of environmental quality in hedonic regressions, clearly additional research is needed to determine whether or not this result extends to situations in which the objective measures of quality may be very different than consumer perceptions of the amenities.

Again, no definitive answer can be offered with respect to building the data needed for the hedonic model. Common sense, research and knowledge about the study market, and an understanding of the consequences of the choices (assumptions) made are key to conducting a thoughtful and careful hedonic price study.

3.3 Sample Frame

When choosing a sample frame for a hedonic analysis, one must consider the geographic coverage of the data selected as well as the time period. First, consider the geographic coverage of the data. If the study focus is an environmental good, then the data has to have geographic coverage sufficient to ensure variation in the environmental amenity or disamenity across properties. Depending on the amenity, variation may be in the form of proximity of each property to the amenity (e.g., proximity to a park), or the variation may be the ambient level of the amenity (e.g., the level of air quality at the site). While variation of the first type is typically easy to ensure, ambient environmental quality can sometimes be more difficult to ensure in a sample frame. Geographic dispersion of properties sufficient to ensure variation in an ambient environmental variable may result in a sample frame that is comprised of properties from multiple markets. If, in order to get sufficient variation in a environmental variable, the geographic dispersion of properties in a sample is increased so that properties are now drawn from different markets, estimating one hedonic price function for the entire sample is inappropriate. This is because a hedonic price function is an equilibrium function describing a specific market. As such, all properties used in a hedonic regression must be part of the same housing market. The question that arises is how to determine what set of properties comprise a single market. In other words, we wish to know the extent of the market.

Markets are truly separate if participants in one market do not consider houses in the other market when making purchase decisions. One common assumption is that each urban area represents a separate housing market. If the focus of the study is proximity to an amenity or disamenity, each urban area typically has sufficient variation for estimation. Although considering each urban area a separate market is likely a reasonable assumption, it may also be the case that additional market segments exist within a city.[5] For instance, markets may be segmented by time, product variety, and perhaps social barriers such as income and race.

The evidence for clear market segmentation across any barrier is mixed across applications, including whether or not markets are segmented across urban areas. Palmquist (1991) reviewed early studies with evidence for market segments across urban areas (Butler 1980) and within urban areas because of geography (Straszheim 1974), and income or accessibility (Schnare and Struyk 1976). Michaels and Smith (1990) asked realtors in the Boston area to specify what they thought were different market segments and then used this as a basis for segmentation. A similar reasoning was employed by Boyle, Poor and Taylor (1999) in the Maine lakes example (see section 2). They segmented markets according to areas served by different multiple-listing services that were also geographically separate.

It is difficult to test conclusively for market segmentation, but there are commonly employed methods that are used as supporting (or refuting) evidence for the researcher's priors about segmentation. The most common approach is to apply F-tests to determine if the hedonic price functions differ across segments. Taylor and Smith (2000) used a variant of the F-test, the Tiao-Goldberger F-test, which allowed them to test easily for differences in the hedonic price functions by individual characteristics — a particularly useful extension to determine if there are key characteristics that are priced differently across markets. The problem with F-tests, however, is that results indicating that market segmentation exists may be due to mis-specification of the hedonic price function, and not actual segmentation. F-tests are also likely to reject aggregation with large samples (Ohta and Griliches 1979). Once again, we see that there are no definitive answers. However, researcher judgement along with supporting statistical tests can be used as a guide for determining the market extent in a particular study.

In addition to needing sufficient geographic variation in the data, it may be important to have appropriate time-variations in the data. Often, a researcher may be interested in how the valuation of an amenity changes over time. Or the researcher is interested in how prices changed in response to the creation of an amenity or disamenity (such as the siting of a landfill near a neighborhood). In some instances, the researcher may simply need observations from multiple years in order to increase sample sizes if sales in a market are sparse. In each case, several issues must be considered when drawing a sample frame from multiple years.

First, equation (5) makes clear the importance of considering the nature of the cross section of observations used in the analysis with respect to timing of changes in amenity levels. Future changes in amenity levels will be capitalized into current housing prices. As the expectations about the quality of the amenity or the impact of a disamenity change, so too will the housing prices. Thus, the same amenity or disamenity may have differential impacts on sales prices over time. A researcher may estimate different implicit prices depending on the sample frame chosen. Kiel and McClain (1995) underscore this point with their research on property sales prices around an incineration plant in North Andover, Massachusetts. Their results indicated that negative expectations regarding the impacts of the incinerator were capitalized into nearby housing prices prior to operation of the incinerator. However, the negative effects of the incinerator were greatest within the first three years of operation, attenuating thereafter. Being located one mile closer to the incinerator decreased property sales prices by $2,283 during the construction phase; $8,100 during the first three years of operation; and $6,607 in the following three years of operation.[6]

The study by Keil and McClain (1995) highlights how the timing of events relative to a sample of property sales can be quite important for properly interpreting the long-run impacts of environmental amenities or disamenities on property values (see also Kohlhase 1991). This is especially important for amenities and disamenities that have announcement effects associated with them, such as the discovery of hazardous waste sites, or for those amenities and disamenities that may be finite in duration (i.e., the disamenity may be removed at some point in the future). Timing effects are likely to be less important for amenities/disamenities that vary spatially, but not over time, such as proximity to urban parks.

When multiple years of data are needed, either for variation in the amenity of interest or to increase sample sizes, one must be careful to consider whether or not the hedonic price function is stable over the time period considered. In the simplest case, one only need adjust for general inflation in a market over time. Prices may be deflated prior to estimation by an appropriate, preferably local price index. If an appropriate price index is not available, a series of dummy variables may be included in the hedonic regression to control for the year in which each property was sold. Either of these methods is reasonable to use when general inflationary trends exist in the market. They essentially control for changes in the intercept of the hedonic price function. In this approach, all implicit prices are assumed to change by the same factor, and thus adjusting the intercept of the hedonic price function is all that is needed.

The more complicated case is when changes in supply and demand conditions occur over time. If structural changes in the market have occurred, implicit prices of characteristics will change, and prices may change by different amounts for each characteristic. In this case, simple price deflators are not sufficient. While F-tests or Tiao-Goldberger F-tests may be applied to test for the stability of the hedonic function and the marginal prices of individual characteristics over time, the same weaknesses in the tests applies as stated above. Ohta & Griliches (1975) suggest a less stringent rule in which the standard errors for the constrained and unconstrained regressions are compared, and aggregation over time is rejected if errors increase by more than 10% (Palmquist 2000). A similar rationale may be applied to Tiao-Goldberger F-tests.

3.4 Functional Form of the Hedonic Price Function

In addition to specifying which variables belong in the hedonic regression, researchers must determine how the characteristics influence price. In other words, the researcher must decide on the functional form of the hedonic price function that is estimated. Little theoretical guidance exists for the choice of functional form for the hedonic price function because the price schedule is determined in the marketplace by the interactions between many different buyers and sellers. If a product can be costlessly repackaged, the total price of the product would simply be the sum of the implicit prices of its component characteristics.[7] In this case, the appropriate specification for the hedonic price function would be a simple linear function, as illustrated for the case of housing:

$$(6) \qquad P = \alpha_0 + \sum_{i=1}^{h} \beta_i H_i + \sum_{j=1}^{n} \beta_j N_j + \sum_{k=1}^{L} \beta_k L_k + \varepsilon ,$$

where P is the sales price of a house; H represents structural and property characteristics of the house, such as number of bedrooms and lotsize; N represents neighborhood characteristics, such as median income and quality of schools, and could include amenities such as ambient air quality; and L represents locational characteristics, such as proximity to the central business district, and could include proximity to environmental amenities and disamenities. In the case of the linear functional form, the implicit price for any specific characteristic, z_i, is simply equal to the estimated coefficient for that variable (i.e., $\partial P(\underline{z})/\partial z_i = \beta_i$). With a linear hedonic price function, the incremental price of the house induced by an incremental increase in a characteristic is constant across levels of the characteristic. This implies, for example, that the incremental value of lakeshore frontage would be the same for lakefront houses with 50 feet of frontage as for houses with 150 feet of frontage.

Of course, marginal prices are not likely to be constant for all characteristics, and so alternative functional forms are often used where some or all of the variables are transformed. For instance, most researchers assume that a home's price is related non-linearly to its square footage. In particular, it is generally assumed that the total price of a house increases at a decreasing rate with square footage (see section 2.1, figure 2.1 and its associated discussion). To capture this empirically, price may be estimated as a function of the natural log of square footage or price may be estimated as a function of square feet and (square feet)^2.

A hedonic price function that allows for sales price to be affected non-linearly by the characteristics of the property may be specified in many possible ways. Some common functional forms are given in Table 2. Also shown for each functional form of the price function is the specific form for computing the marginal implicit price, $\partial P(\underline{z})/\partial z_i$. As can be seen from Table 2, these formulas can be complicated, and involve not only the estimated coefficients from the hedonic price function but also the levels of the characteristics themselves.

Two possible approaches may be used for computing implicit prices when the formula involves the level of a variable. Consider the semi-log functional form in which $\partial P(\underline{z})/\partial z_i = \beta_i P$. In this case, the implicit price must be evaluated at some level of sales price. Most commonly, the mean or median housing price of the sample over which the coefficients were estimated is chosen. Alternatively, the implicit price may be computed for each house in the sample and then the mean of these prices may be used.No one way is correct, and it is researcher judgement guided by the application at hand that determines the appropriate level of price (or any variable) at which one evaluates the implicit price.

Product characteristics are often described by categorical variables. For instance, rather than recording the number of fireplaces in a home, a variable

Table 2. Functional Forms for the Hedonic Price Function

Name	Equation	Implicit Prices
Linear:	$P = \alpha_0 + \sum \beta_i z_i$	$\partial P/\partial z_i = \beta_i$
Semi-Log:	$\ln P = \alpha_0 + \sum \beta_i z_i$	$\partial P/\partial z_i = \beta_i \cdot P$
Double-Log:	$\ln P = \alpha_0 + \sum \beta_i \ln z_i$	$\partial P/\partial z_i = \beta_i \cdot P/z_i$
Quadratic:	$P = \alpha + \sum_{i=1}^{N} \beta_i z_i + \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \delta_{ij} z_i z_j$	$\partial P/\partial z_i = \beta_i + \frac{1}{2}\sum_{j \neq i} \delta_{ij} z_j + \delta_{ii} z_i$
Quadratic Box-Cox:[a]	$P^{(\theta)} = \alpha + \sum_{i=1}^{N} \beta_i z_i^{(\lambda)} + \frac{1}{2}\sum_{i,j=1}^{N} \delta_{ij} z_i^{(\lambda)} z_j^{(\lambda)}$	
		$\partial P/\partial z_i = \left(\beta_i z_i^{\lambda-1} + \sum_{j=1}^{N} \delta_{ij} z_i^{\lambda-1} z_j^{(\lambda)}\right) P^{1-\theta}$

[a] The transformation function is $P^{(\theta)} = (P^\theta - 1)/\theta$ for $\theta \neq 0$, and $P^{(0)} = \ln(P)$ for $\theta = 0$. The same transformation applies to lambda. The marginal effect is computed for $\theta \neq 0$. The marginal effect when $\theta=0$ and $\lambda=0$ is the same as for the double-log. The quadratic Box-Cox is equivalent to the linear Box-Cox, another common functional form when $\delta_{ij} = 0$ for all i and j. The marginal price of the linear Box-Cox is simply $\beta_i z_i^{\lambda-1} P^{1-\theta}$.

may be created to equal one if there are any fireplaces in the home and equal to zero if there are none. If the dependent variable is log of sales price, then the interpretation of the coefficient estimate is the approximate percentage change in price when the characteristic in question is present. It is an approximation because the coefficients estimated for the dummy variables are transformations of the percentage effect (Halvorsen and Palmquist 1980 and Kennedy 1982). For a coefficient estimate, b, the percentage effect, g, is given by $100 \cdot g = 100(e^b - 1)$. When b is small, the error in interpretation (without adjustment) is small. For instance, a coefficient estimate of 0.10 implies that the percentage change in price is actually 10.5%.

To examine the potential impact of functional form choices, Cropper, Deck, and McConnell (1988) conducted simulations in which the true marginal bid functions of households and equilibrium marginal prices were known and could be compared to the estimated implicit prices from hedonic regressions. Comparisons were made for each of the functional forms are listed in table 2. Results indicated that when all attributes of housing are observed without error, the more complicated functional forms (quadratic, linear Box-Cox, and quadratic Box-Cox) can be used to most accurately estimate implicit prices. However, when some variables are not observed or are replaced by proxy variables, the quadratic and quadratic Box-Cox produced biased estimates of the marginal prices. Instead, it was the simpler forms of the linear, semi-log, double-log and linear Box-Cox that performed best. The linear Box-Cox provides accurate marginal price estimates when all attributes are measured correctly and also performs well in the presence of mis-specification of the hedonic function, so it is the preferred functional form of Cropper, Deck, and McConnell (1988).

Recent improvements allow the transformation parameter in a Box-Cox specification to vary across independent variables as done in Israngkura (1994) and Cheshire and Sheppard (1995). Palmquist and Israngkura (1999) also allow a flexible Box-Cox specification in a study focusing on air quality in 13 major cities. They use an iterative procedure in which they first estimated the Box-Cox parameter for the non-environmental characteristics of housing. This parameter is then imposed, and the Box-Cox parameter for air-quality is estimated.

What can serve as a guide for choice in functional form? First, one should be careful not to rely on a linear hedonic (with no transformed variables) unless there are compelling theoretical reasons to do so. In general, non-linear relationships between size attributes and some quality attributes are expected.

For this reason, there is a preponderance of empirical studies that rely on the semi-log functional form. The semi-log allows for incremental changes in characteristics to have a constant effect on the percentage change in price and a non-linear relationship on the price-level.

Second, researchers should pay particular attention to the relationship between price and the key characteristics of the study. For example, studies that have investigated the relationship between the sales prices of single-family residences and distance to a contaminated site have assumed that price increases with distance at a constant rate (i.e., a linear relationship; see Michaels and Smith 1990 and Kiel 1995) or that prices increase with distance at a declining rate (i.e., price increases, peaks, and then decreases with additional distance; see Kohlhase 1991). Neither of these functional forms is consistent with the expectation that price increases with greater distance from a contaminated site at a declining rate, with the effect of distance disappearing completely beyond some distance. Ihlanfeldt and Taylor (2001) instead implement a reciprocal transformation in which sales prices of commercial and industrial properties are modeled as a function of the inverse of distance to a site with known environmental contamination. This specification is more attractive *a-priori* than what has been used in the past because it is consistent with our expectations regarding how sales prices will be impacted by proximity to a contaminated site: sales prices for properties near contaminated sites are negatively impacted, and prices asymptotically increase to some baseline level as distance from the site increases.

3.5 Spatial Considerations

Specific spatial relationships, such as proximity of a property to a specific amenity such as a park or a shoreline, have been directly addressed in hedonic models for over 30 years. However, more general intrinsic spatial relationships among properties have only been taken into account more recently. Spatial effects that require specific econometric tools (often referred to as spatial econometrics) are spatial dependence and spatial heterogeneity. The following is a necessarily brief introduction to the topic, and the reader is directed to a series of reference books and articles by Luc Anselin providing a comprehensive overview of spatial econometric techniques (Anselin 1988, Anselin and Florax 1995, Anselin et al. 1996, and Anselin and Bera 1998).

Spatial dependence or spatial autocorrelation implies a lack of independence across observations in cross-sectional, spatially organized data. With spatial dependence, a functional relationship exists between what happens at one location and what happens at another location. Spatial autocorrelation is similar in concept to autocorrelation in time-series data; however, spatial correlation is multi-directional while correlation in time-series data is uni-directional. In time series data, only past events are assumed to affect current events and a single direction of dependence exists. But in spatial data, a property is likely to be related to other properties in a multitude of directions. For a variable, say sales price, associated with property i, p_i, there is a general spatial process that can be written as $p_i = f(p_1, p_2, \ldots p_N)$, where N is the number of properties in total. This relationship is modeled parametrically so that the process may be estimated and tested empirically.

Spatial heterogeneity is a related concept and refers to spatial correlation of the error terms of a regression model. Here, the unobservable components of one dependent variable are related to those of a neighboring dependent variable. Again, as with spatial dependence, the spatial relationship between errors must be specified.

Following Anselin (1988), a general model of housing price determination including spatial effects is:

$$\begin{aligned} P &= \rho W_1 P + Z\mathrm{B} + \varepsilon \\ \varepsilon &= \lambda W_2 \varepsilon + \mu \\ \mu &\sim N(0, \sigma^2 I) \end{aligned} \tag{7}$$

where Z is a NxK matrix of property characteristics, P is sales price, B is a Kx1 vector of coefficients, W_1 and W_2 are NxN spatial weight matrices, ε is a Nx1 spatial autoregressive error, μ is a Nx1 random error term with variance σ^2, and ρ and λ are coefficients on the spatially lagged variables, P and ε. The equations in (7) may be rewritten as:

$$P = \rho W_1 P + Z\mathrm{B} + (I - \lambda W_2)^{-1}\mu, \tag{8}$$

and if ρ and λ are estimated to be equal to zero, equation (8) reduces to a simple linear regression model.

In (8), ρ, λ, and B are estimated and W_1 and W_2 are arbitrarily chosen by the researcher. Specification of the spatial weights matrices are one of the more controversial aspects of spatial econometrics. The spatial weight matrix defines the sense in which properties are believed to be neighbors and determines the importance of any one observation of the variable of interest (sales price in our example) for another observation. The spatial weights matrix is akin to a lag operator in time series, except for spatial lags are more complicated as they are multi-directional. Several possible structures for this matrix might be specified. One could use distance between observations as the weights. In this approach, the importance of each property on the current property decays as distance increases. Often, the rows of the weight matrix are standardized to sum to one, which eases the statistical computation of ρ and λ. Another approach imposes a lattice structure on the relationship between properties. Here, elements of the spatial weight matrix, w_{ij} equal one if a property shares a border with the observation of interest, and equals zero otherwise. Other lattice structures may be imposed. For instance, weights may equal one if properties are contained within some predefined radius of the property of interest. In equation (8), the matrices W_1 and W_2 are allowed to differ, implying differences in the autoregressive processes of the dependent variable and the error term.

Lagrange multiplier tests, Moran's I statistics (Anselin 1988), and Kelejian and Robinson tests (Kelejian and Robinson 1992) may be used to test for the presence of either spatial dependence or spatial heterogeneity or both. Both tests are pre-packaged and easily implemented using the software package SpaceStat Software for Spatial Data Analysis (Allen, Texas; SpaceStat http://www.spacestat.com 1998). Given evidence of spatial dependence or spatial heterogeneity, spatial weights matrices must be specified and estimation may be done using maximum likelihood or generalized moments estimators. Estimation of models with spatial dependence, but not spatial heterogeneity (i.e., λ=0 in equation 8) are computationally straightforward; however, spatial heterogeneity estimation of λ via maximum likelihood becomes problematic with large samples sizes.[8] Indeed, maximum likelihood estimates with sample sizes over 400 observations may not produce reliable estimates of λ (Kelejian and Prucha 1999). An alternative, suggested by Kelejian and Prucha (1999), and implemented in a hedonic property value model by Bell and Bockstael (2000) is a generalized moments estimator that reduces the computational burden and allows more flexibility in the specification of the weight matrix.

Recent hedonic applications that explicitly control for spatial effects as discussed above are Gwande and Jenkins-Smith (2001), Leggett and Bockstael (2000), Geoghegan, Wainger, and Bockstael (1997), and Bockstael (1996). Gwande and Jenkins-Smith estimate a hedonic property value model to examine the impacts of nuclear waste transport on property values near the transportation route. Using several test statistics, Gwande and Jenkins-Smith (2001) found significant evidence of spatial correlation in the data. Because they had a large data set (over 9,000 observations), they estimated a model with spatially lagged dependent variables and assumed the proper specification of the spatial weights matrix (W_1) results in no residual spatial heterogeneity in the error terms. Leggett and Bockstael (2000) on the other hand, allowed for spatial heterogeneity in the error terms but no spatial dependence (i.e., $\rho=0$ in equation (8)). Each element of their spatial weights matrix, w_{ij}, is defined as the inverse distance between observations i and j, and following Bell and Bockstael (2000), this measure is set equal to zero if distance is greater than one mile. Their application is to water quality effects on property values around Chesapeake Bay. Results indicated that fecal coliform counts negatively impact property values, and these results were *not* sensitive to whether or not they corrected for spatial heterogeneity. However, some descriptive variables of the properties, such as density of development that were significant when spatial heterogeneity was not taken into account, are not significant in the spatially-corrected model.

To implement the spatial models described, the databases must be spatially referenced or geocoded. In geocoding, typically latitude and longitude markers are assigned to the center-point of each property. Geographic information systems software may be used to compute spatial variables, such as the linear distance between properties, quite easily. The accuracy of the spatial relationships depends critically on the accuracy of the geocoding. The most common method of geocoding data is to use address matching which is available in software packages such as ArcView and Maptitude (www.caliper.com/maptovu.htm). These programs use base maps, such as the TIGER files (www.census.gov/geo/www/tiger/index.html), to assign latitude and longitude markers to each property. Address matching is not exact because the location of a property is only approximate along the street segment in which the property is located. For highly localized externalities, address matching may not place properties accurately enough. Fortunately, many urban areas have developed, or are developing, base maps of their tax parcels that are accurate to within a few feet by combining geo-referenced aerial photos with digitized tax maps.

As spatially referenced property value data sets are becoming increasingly available, there is likely to be a continued expansion of spatial econometric techniques applied to hedonic property value studies. Indeed, tests and correction for spatial correlation are likely to become as commonplace as tests and corrections for heteroskedasticity in cross-sectional data.

4. WELFARE MEASUREMENT WITH THE HEDONIC PRICE FUNCTION

Our model indicates that the implicit price of amenity i ($p_{zi} = \partial P(\underline{z})/\partial z_i$) is equal to the consumer's marginal WTP for the amenity ($\theta_{zi} = \partial\theta(\underline{z}, y, U)/\partial z_i$). Implicit prices are the most commonly reported result from hedonic studies. If one is interested in whether or not a current stock of an amenity is capitalized into the market for housing, estimates of the marginal implicit prices (sign, magnitude, and statistical significance) are the appropriate measures to report. In the context of expenditures on other aspects of housing, these relative prices can provide interesting insights about the importance of various housing amenities.

In instances where we want to know the value consumers might place on a *change* in an environmental amenity, the relationship between implicit prices and appropriate measure of WTP for the change depends on the situation. To examine this issue, we will consider the owners of properties and the renters of these properties separately. Even though the owner and renter can be the same person (implicitly renting from oneself), this discussion considers these two sides of the market separately. Two types of changes are considered: a change in a localized amenity and a change in a non-localized amenity. Examples of localized externalities that have been studied in the hedonic context are highway noise (Palmquist 1992a), a hazardous waste site (Kohlhase 1991), an incinerator (Kiel and McClain 1995), and local parks (Shultz and King 2001). In these cases, a change in the amenity affects a relatively small number of properties, and so the equilibrium hedonic price function for the entire market is unaffected. When changes in non-localized amenities such as the air quality of a city occur (Zabel and Kiel 2000, Sieg et al. 2000), the amenity change affects a large enough number of houses to reflect a marketwide shift in supply, and thus we would expect a change in the market-clearing equilibrium hedonic price function.

For a localized change in an amenity, first consider the effects on renters of the property if no transactions costs are associated with moving. When a decrease of an amenity occurs, the renter is no longer at their optimal bundle of housing given that they face the same hedonic price schedule as before the change at their home. If renters can move costlessly to a house with the original bundle of characteristics, there will be no change in welfare for the renter. However, the owners of the property realize a capital loss on the property because of the decrease in the amenity level associated with the property. The owner would be WTP an amount of money up to the value loss of the property value to avoid the amenity change. If, in addition to being a localized change, the amenity change is marginal (i.e., a one-unit change), then WTP is simply the implicit price, $\partial P(\underline{z})/\partial z_i$, for each property owner. The total WTP is the sum of the implicit prices across property owners that receive a change in the amenity.

If the amenity change is localized and non-marginal in the magnitude of change, owners would be willing to pay an amount equal to $P^1 - P^0$, where P represents the sales price of the property with the initial level of the amenity (P^0) and the new level of the amenity (P^1). The total WTP is the sum of the price changes for houses that receive the amenity change. If the analysis is *ex-ante*, as is typical for policy analyses, the price change is forecast using the estimated hedonic price function.[9] Thus, for a change in an amenity, say characteristic z_1, the total change in welfare or net benefit (NB) is computed in the following way:

$$(9) \qquad NB = \sum_{k=1}^{N} P_k(z_{1k}^1, z_{2k}^0, \ldots, z_{nk}^0) - P_k(z_{1k}^0, z_{2k}^0, \ldots, z_{nk}^0).$$

where $P_k(z_{1k}, z_{2k}, \ldots, z_{nk})$ is the *ex-ante* hedonic price function evaluated at the characteristics associated with k^{th} property in either the initial or new state.

For example, using the estimated hedonic price function in equation (1), a property with characteristics equal to the sample means (see discussion following equation (1) for those means) is predicted to sell for \$110,466 (the mean sales price in the data is \$97,761). If the water clarity at the lake where the property is located (a small part of the overall market) were increased from 3.9 to 6.3 meters (the highest observation for this data) this property would be predicted to sell for \$115,133. The NB for this house is \$4,667. The total NB would be the sum of the net benefits estimated for each house located on this particular lake.

If we relax the assumption of costless moving on the part of renters, then the computation above represents an upper bound on the welfare change associated with the amenity change (i.e., overstate gains and understate losses). From the price change given in equation (9), one would need to subtract the transactions costs (TC) associated with each renter having to move to their new optimal location to compute the total change in welfare. For each property, net benefits are given by P^1 - P^0 - TC. Palmquist (1992b) summarizes the manner in which transactions and moving costs may be quantified and incorporated empirically into a hedonic analysis.

The above analysis assumes consumer utility is unchanged as a result of the change in the amenity (Δz_1) since renters move to a new house with the exact same bundle of characteristics as before the amenity change. This follows from our assumption that the change in the amenity does not affect the overall hedonic price function in the market, but only affects the price of houses experiencing Δz_1. Thus, in optimizing utility, a house identical to the original dwelling before the amenity change will be chosen. If identical housing is not available, and utility changes as a result of the move, then the correct welfare measure would require quantifying the difference in utility arising from the consumer no longer being at the optimal consumption bundle (see Bartik 1988 for a complete description of this problem). While this computation is not possible using information from only the hedonic price function, the change in rents less transactions costs would provide a upper bound on the resulting net benefits from an amenity change. The degree to which this bound is close to the true net benefits will depend on the degree of substitutability between the housing affected by the amenity change and the amenity level at the housing renters move to.

Table 3 summarizes the benefit estimates possible with the first-stage hedonic price function when the change in the amenity is localized. Note that the amenity change that occurs locally may be marginal or non-marginal. The table highlights that simply computing the change in property values at locations will overstate the benefits of an amenity improvement and will understate the losses of an amenity decline when transactions costs are associated with moving (Bartik 1986).

The last case to consider for a localized externality is when transactions costs (TC) are great enough to prohibit moving. For a homeowner, closing costs, changes in interest rates on mortgages, and a tight housing market may make transactions costs prohibitively high, especially in the short run when costs associated with searching for suitable new housing would be high. For owner-occupied housing, the short run may be as much as six months to a year or possibly longer. For renters, it is more likely that the short run would be a few months.

Table 3. Summary of Welfare Computations Appropriate with 1st Stage

Computation	Appropriate Situation
$\partial P/\partial z_i$	Net benefits of an ↑/↓ in the amenity by one-unit in a localized context with zero transactions costs associated with moving.
$P^1 - P^0$	Net benefits of an ↑/↓ in the amenity in a localized context with zero transactions costs associated with moving.
$P^1 - P^0 - TC$	Net benefits of an ↑/↓ in the amenity in a localized context when transactions costs are positive, but do not prohibit moving. Household can relocate to an identical house as before the amenity change.
$P^1 - P^0 - TC$	*Upper-bound* on net benefits of an ↑/↓ in an amenity in a localized context when (1) moving is possible, but households cannot relocate to an identical house as before the amenity change or (2) when it is *not* possible to move because transactions costs are prohibitively high.

When transactions costs prohibit moving, the net benefits of an amenity change cannot be exactly measured, but an upper bound can again be computed as the price differential for the property minus the transactions costs associated with moving. Consider a decline in an amenity. The landlord's capital loss as a result of the amenity change is exactly offset by the tenant's capital gain associated with paying a lower rent (recall, the tenant does not move). Thus, the price change is a transfer between owner and tenant, and no net benefit is associated with the change in rent. However, the reduction in rent only partially offsets the consumer's utility loss. Transactions costs provide an upper bound on the remaining utility loss (after the change in rent is taken into account). Thus, the difference between the change in rent and transactions costs associated with moving will provide an upper bound on the total welfare change associated with the change in the amenity for a consumer (Table 2). Palmquist (1992b) presents a graphical exposition of this point.

In the case of *non-localized* amenities an exact measure of welfare change cannot be computed with only information from the hedonic price function. This is because a non-localized change implies a shift in the overall supply, and the hedonic equilibrium itself will change. The new hedonic price function cannot be forecast using information from only the current price function. However, as Bartik (1988) demonstrates, the change in hedonic rent at sites which are improved, as predicted by the *ex-ante* hedonic price function (call this measure $\triangle$HP), is likely to be an upper bound for the net benefit of the amenity change. This bound is computed as given in equation (9) and is, therefore, quite easy to implement. However, $\triangle$HP is only an upper bound under some conditions on the profit changes landlords may experience when allowed to make equilibrium adjustments to housing characteristics in response to the amenity change (Bartik 1988, 179). These profit changes are likely to be in a range such that $\triangle$HP will be an upper bound only if no transactions costs are associated with moving. If there are significant transactions costs, this measure is less likely to be an upper bound.

Unfortunately, one cannot know *a-priori* if $\triangle$HP is an upper bound, and empirical tests have not determined how likely it is to be an upper bound. In the next section, we discuss methods for estimating the *demands* for amenities by households and associated WTP measures. While more complicated and data-intensive, there are several reasons why estimating the demand for a characteristic is desirable. In the case of localized externalities, if transactions costs cannot be measured and are expected to be significant, then household WTP must be computed. This is easily computed using estimates of the demand parameters. In the case of non-localized amenity changes where the hedonic price function is expected to shift, household WTP provides a definite lower bound, as compared to the ambiguous bound provided by $\triangle$HP. Lastly, estimating the demand for an amenity allows us to explore the relationships between WTP and the characteristics of a household such as income, race, gender, and family composition.

5. SECOND-STAGE DEMAND ANALYSIS

To estimate uncompensated demands for the *characteristics* of the differentiated good, information on the quantities of characteristics purchased and the marginal implicit prices of those characteristics from a hedonic analysis are combined with information on the socio-economic characteristics of the purchasers. Estimating the demand for characteristics of a differentiated product using hedonic prices is often referred to as a second-stage analysis (estimating the hedonic price function was described as the first stage in the previous section). Rosen's (1974) seminal article provided the framework for this method, however, some confusion exists in the literature regarding the nature of, and proper estimation of, these characteristic demand functions. This section provides an overview of second stage methods. See Palmquist (2000) for a more technical discussion of the method.

For welfare measurement, we are typically interested in compensated demands, which indicate the WTP for z_i holding the levels of all other goods/characteristics and utility constant. Recall that the bid function for each individual, θ^j, is a function describing the maximum WTP for an specific bundle of characteristics, holding utility and income constant. Because bid functions change directly with income (i.e., $\partial\theta/\partial y = 1$), the marginal bid function for characteristic z_i, θ_{zi}, only depends on $\underline{z}$, utility, and exogenous consumer characteristics, α^j (Palmquist 2000). As such, the marginal bid function is equivalent to an inverse compensated demand function for characteristic z_i. It describes the change in WTP for z_i as the level of z_i changes, holding utility and all other characteristics and goods constant. Figure 2 illustrates the marginal bid function for two individuals, θ^1_{zi} and θ^2_{zi}. For the moment, consider only the marginal bid function for consumer 1 as depicted by the heavy dark line passing through points AB and the associated implicit price function P^A_{zi}. We can see

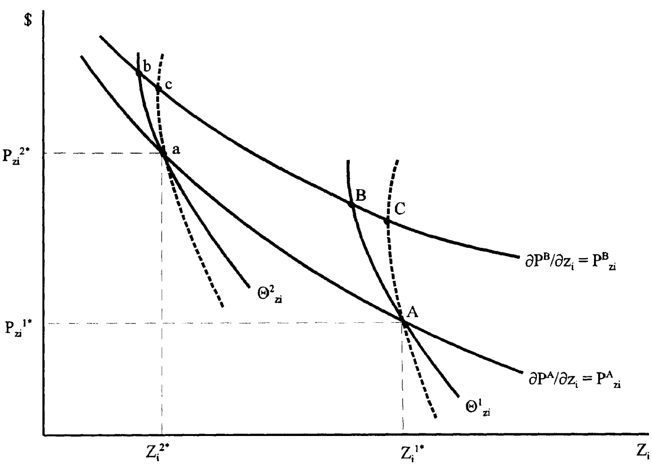

Figure 2. Demand Identification with the Hedonic Price Function

that the marginal bid function is equal to the implicit price function at the optimal level of consumption for consumer 1, z_i^{1*} (point A). This point corresponds in Figure 1 to the optimal consumption level where the bid function is tangent to the hedonic price function.

Considerable confusion exists in the literature regarding the interpretation of the marginal bid function as related to the inverse uncompensated and compensated demand functions. It has been suggested that the marginal bid function is actually akin to a marginal rate of substitution function given by $(\partial U/\partial z_i)/(\partial U/\partial x)$, which is incorrect. It is true that, at the *optimal level of consumption*, the marginal rate of substitution is equal to the marginal bid and the marginal price $(\partial U/\partial z_i)/(\partial U/\partial x) = \partial P/\partial z_i = \partial\theta/\partial z_i$ as discussed in section 2.1. In other words, the two functions cross at the optimal consumption level, but they are not equivalent functions. The marginal rate of substitution function may be combined with either utility or the budget constraint to derive the compensated or uncompensated inverse demand function, respectively (Palmquist 2000). This parallels the derivation of these two functions for any homogenous good.[10]

Theoretically correct measures of welfare change for a characteristic of a differentiated good are no different than as described for a quantity change in an amenity in Chapter 2. The literature typically makes the distinction between measures of welfare change when consumers face a price change (compensating and equivalent variation) and when consumers face an exogenous quantity change (compensating and equivalent surplus). These are analogous measures (see Chapter 2). In the case of an exogenous change for consumers who cannot move, compensating and equivalent surplus would be the appropriate measure. In the case of an exogenous change in z_i in which moving is possible, compensating and equivalent variation would be more appropriate. However, these measures will only measure the benefits to homeowners. If the change in the amenity affects consumers in other ways (say, increases recreation benefits to consumers who do not own nearby property, but visit the park after it is improved), these benefits will not be captured using the measures discussed in this chapter.

Here, we focus on compensating and equivalent surplus, as one typically assumes households cannot move in response to a change and compute WTP of the households as a lower bound to the true welfare change (see section 4). Measures of compensating or equivalent surplus are simply computed by integrating under the compensated inverse demand or the marginal bid function:

(10) $$W(\Delta z_i) = \int_{z_{i0}}^{z_{i1}} \frac{\partial \theta(z_i; \underline{z}, U^j)}{\partial z_i} dz_i ,$$

where $W(\Delta z_i)$ is compensating or equivalent surplus for a change in z_i depending on whether U^j is equal to the initial level of utility or the level of utility after the amenity change, respectively. This measure of welfare change is appropriate for the situation in which consumers receive an exogenous quantity change in an amenity, and consumers are compensated in terms of the numeraire (holding $\underline{z}$ constant). Palmquist (2000) discusses welfare measures when other types of compensation are used to return consumers to their original (or new) level of utility.

Of course, estimating the marginal bid function directly is not possible, because the marginal bid depends on unobserved utility. However, two alternatives allow us to recover the information necessary to compute welfare measures. The first is to estimate an uncompensated demand and use duality to recover estimates of welfare change. The second is to estimate the equilibrium condition given in equation (3) and use the estimated parameters of the marginal rate of substitution function to recover estimates of welfare change. Each method is described next.

5.1 Uncompensated Demands for Characteristics

Uncompensated demand functions can be derived analytically in a manner analogous to a homogenous goods market if the hedonic price function is linear so that the marginal prices are constant. However, we often expect the hedonic price function to be non-linear and the marginal prices to be non-constant. In this instance, the budget constraint is no longer an exogenous constraint; the marginal price of any characteristic is simultaneously determined by the choice of the amount of that characteristic to purchase. Traditional optimization methods are inappropriate in this case.

Uncompensated demands with non-constant marginal prices can be derived, however, if we linearize the budget constraint around the optimal consumption bundle of the consumer. Palmquist (1988), demonstrates that utility maximization with this constraint set yields the same choices as would be observed with the original, non-linear budget constraint. The linear budget constraint is of the form:

$$(11) \qquad Y_a^j = x + \sum_{i=1}^{n} p_i^* z_i ,$$

where Y_a^j is consumer j's income adjusted in the following manner:

$$(12) \qquad Y_a^j \equiv y - P(Z^*) + \sum_{i=1}^{n} p_i^* z_i^* .$$

The linearized budget constraint in (11) is exogenous as the implicit prices, p_i^*, faced by consumer j are held constant at the level associated with js this consumer's actual purchase, z_i^*. Income must be adjusted as shown in equation (12) because a linear budget constraint will imply a larger consumption set than is affordable for the consumer with income y facing a non-linear budget constraint in which prices for z_i are decreasing as amounts of z_i are increased.

The linearized budget constraint in (11) may be used in conjunction with the first-order condition for utility maximization (given in equation (3) above) to solve for the inverse uncompensated demand in a manner analogous to that for homogeneous goods:

$$(13) \qquad p_1^j = f(z_1, z_2, ...z_n, x, Y_a, \alpha^j) .$$

This demand function can be estimated using quantities of characteristics chosen, the implicit prices of those characteristics, adjusted income, and other socio-economic characteristics assumed to affect demand.

Consumer surplus estimates for a change in z_i may be computed by integrating vertically under the inverse demand curve for z_i between the initial and new level of z_i (Parsons 1986). For example, suppose the following simple linear demand for z_1 is estimated (note, this is not an inverse demand function):

$$(14) \qquad z_1 = \alpha + \beta_1 p_1 + \beta_2 p_2 + \delta y ,$$

where p_2 is the implicit price of a substitute/complement for z_1 and y is adjusted income. The integral of the inverse demand between the initial level z_{10} and new level z_{11} is:

$$(15) \qquad \int_{z_{10}}^{z_{11}} \frac{1}{\beta_1}(z_1 - k)\, dz_1 = \left[\frac{1}{\beta_1} z_1^2 - \frac{k}{\beta_1} z_1\right] \Bigg|_{z_{10}}^{z_{11}}$$

where k is a constant equal to $\alpha + \beta_2 z_2 + \delta y$ evaluated at some appropriate level, such as the mean observed in the data.

The demand function estimated should include relevant substitute and complement characteristics to the characteristic of interest. Choosing the appropriate set of substitutes and complements may be somewhat *ad hoc* because no theoretical guidance defines exactly what is a substitute or complement for any one characteristic. A reasonable approach is to assume that the major characteristics of a property (such as house and lot size and a few key neighborhood characteristics) are likely substitutes and complements to include. Boyle, Poor, and Taylor (1999) estimated the following uncompensated demand for water clarity (Q_{wc}):

$$(16) \qquad Q_{WC} = f(P_{WC}, P_{SQFT}, P_{FRONT}, Y_{adj}, \Gamma),$$

and computed consumer surplus for changes in water clarity as described in equations (14) and (15). In equation (16), Y is income as adjusted in equation (12) and P_{WC}, P_{SQFT}, P_{FRONT} are the marginal implicit prices of water clarity, square feet of living space, and lake frontage of the property, respectively. They chose two major characteristics of the property: size of the house and the amount of lake frontage. Results indicated that both characteristics were substitutes for water clarity as the price coefficients for house size and lake-frontage were positive and significant.

In addition to income, the researcher must determine the other relevant factors expected to shift demand, represented by the vector Γ in equation (16). As with any demand function, typical factors might include the age and education of the purchaser. For property markets, factors such as the number of children under 18 years old would also be expected to be important demand shifters. Other factors specific to each application might be included. For example, in the Maine lakes example, Boyle, Poor, and Taylor (1999) included a dummy variable indicating whether or not the owner had friends who had bought property at the lake prior to their own purchase of property.

After the set of substitute and complements and demand shifters are chosen, the researcher must still specify the function to be estimated. A common functional form for demand is the semi-log functional form. Boyle, Poor, and Taylor (1999) estimate a linear, semi-log, and Cobb-Douglass specification for equation (16), preferring the non-linear specifications. Based on the semi-log specification, the estimated own-price elasticity for water clarity was -0.158, and the estimated surplus for an improvement in lake water clarity from 3.78

to 5.15 meters was \$3,765. The estimated surplus for the same change in quality was \$12,870 using linear specifications and \$3,677 using Cobb-Douglas specifications. As might be expected, the results are sensitive to the choice of functional form. A careful researcher investigates how sensitive their results are to choice of functional form (and specification of variables included). If they have a preferred equation for reporting, reasons for this preference should be explained.

Consumer surplus measures can be computed using estimates of the parameters of the uncompensated demand. However, consumer surplus is only an approximation of the theoretically correct measures of welfare change (see Chapter 2 and Freeman 1993). Two approaches may be used to recover compensating or equivalent variation when uncompensated demands are estimated. In the first, utility is specified and a system of uncompensated demands are derived analytically and estimated. Given estimates of the demand parameters, duality may be used to analytically recover estimates of compensating or equivalent variation by solving for indirect utility or expenditure functions. This approach is taken by Palmquist and Israngkura (1999) and Parsons (1986).

Alternatively, one can estimate a single demand for a characteristic of interest and use differential equation methods to recover an associated utility or indirect utility function. Hausman (1981) and Vartia (1983) demonstrate how this may be accomplished when prices are changing. Palmquist (2000) demonstrates this method for quantity changes to solve for the bid function.

Consider the inverse uncompensated demand in equation (13). Recognizing that the marginal implicit price is equal to the marginal bid function at the optimal level of consumption, we can substitute $\partial\theta/\partial z_i$ for left hand side of equation (13). Also, since our numeraire x is identically equal to Y - θ, we can substitute for x in the RHS, and equation (13) becomes a differential equation in θ. Solving this differential equation for the bid function allows us to compute welfare measures as follows:

$$(17) \qquad W(\Delta z_i) = \int_{z_i^0}^{z_i^1} \frac{\partial\theta(z_i, \underline{z}, U^j)}{\partial z_i} dz_i \equiv \int_{z_i^0}^{z_i^1} MRS(z_i, \underline{z}, x(z_i, \underline{z}, U^j)) .$$

which is equal to compensating surplus if utility is held constant at the original level of utility, and equal to equivalent surplus if utility is held constant at the level of utility after the change (note, income is held constant).

5.2 MRS Function Approach

An alternative method for recovering the information necessary to compute compensating and equivalent variation is to directly specify the functional form for household utility and derive the functions to be estimated (see Quigley 1982). In other words, the exact specification of the regression is determined by the analytical solution to the utility maximization problem. What function is estimated depends on the goal of the research. For estimating welfare changes, Cropper et al. (1993) and Chattopadhyay (1999) estimate the equilibrium condition given in equation (3) in which marginal prices (computed from the first stage hedonic regression) are equal to the marginal rate of substitution (MRS) function. While this function may not be used directly for computing welfare changes, it is useful for estimating demand parameters that can then be used to recover estimates of welfare.

Chattopadhyay (1999) provides a recent example of this approach for estimating the welfare change associated with changes in air quality in Chicago. He tried two alternative forms for utility: the translog and Diewert utility function. The translog utility he specifies is of the form:

$$(18) \qquad U(Z,x) = \ln x + \sum_{i=1}^{k} (\alpha_i + \delta_i C + \tau_i R) \ln z_i + \frac{1}{2} \sum_{i,j=1}^{k} \gamma_{ij} \ln z_i \ln z_j ,$$

where x is the numeraire, Z is the vector of z_i housing characteristics, C and R are demographic characteristics of the owners (C, number of dependent children; R, race of owner), and k is the number of housing characteristics (Chattopadhyay 1999, 24). If a linear Box-Cox functional form for the hedonic price function is assumed, then the marginal rate of substitution function to be estimated is given by[11]:

$$(19) \qquad \beta_i P^{1-\theta} z_i^{\lambda-1} = \frac{x}{z_i} \left[(\alpha_i + \delta_i C + \tau_i R) + \sum_{j=1}^{k} \gamma_{ij} \ln z_j \right],$$

where the regressand is the implicit price of z_i as derived from a linear Box-Cox hedonic price equation (see table 2, footnote a), and the right-hand side is equal to $(\partial U/\partial z_i)/(\partial U/\partial x)$.[12] The estimated hedonic equation is used to compute the dependent variable for each observation, $\beta_i P^{1-\theta} z_i^{\lambda-1}$. This information is combined with information on the chosen levels of characteristics, the numeraire, and homeowner demographics to estimate equation (19) and recover the parameters of the utility function α, δ, τ, and γ. The level of the numeraire, x is equal to income less the price of housing. Typically, one would compute this with annual income and annualized housing expenditures.

Recall that $\theta_{zi}(z_i, \underline{z}, U) = MRS_{zi,x}(z_i, \underline{z}, x)$ at the optimal level of consumption, where θ_{zi} is the marginal bid function for characteristic z_i, and $MRS_{zi,x}$ is the marginal rate of substitution between z_i and the numeraire x. It is also true that $\theta_{zi}(z_i, \underline{z}, U)$ is identically equal to $MRS_{zi,x}(z_i, \underline{z}, x(z_i, \underline{z}, U))$, where $x(z_i, \underline{z}, U)$ is derived by solving the utility function for x. In other words, if we substitute utility into the marginal rate of substitution function, we recover the compensated demand curve. If a MRS function such as that given in (19) is estimated, welfare measures are given by[13]:

$$W(\Delta z_i) = \int_{z_i^0}^{z_i^1} \frac{\partial\theta(z_i, \underline{z}, U^j)}{\partial z_i} dz_i \equiv \int_{z_i^0}^{z_i^1} MRS(z_i, \underline{z}, x(z_i, \underline{z}, U^j))\, dz_i . \tag{20}$$

In equation (20), utility at either the original level of consumption or the new level of consumption may be computed directly from the utility function since its parameters have been estimated.

5.3 Identification of Demand

Recall, each consumer will choose an optimal level of each characteristic by finding the level of consumption at which the marginal bid for z_i equals the implicit price of z_i. By estimating the parameters of P(z), and computing $\partial P(\underline{z})/\partial z$, we observe one point on each consumer's uncompensated demand, as well as their compensated demand function or marginal bid function. (The term, marginal bid function, rather than inverse compensated demand function, is used for consistency with past discussions.) However, no other information on the shape of the bid function is observed. Consider Figure 2 again, which shows the marginal bid functions for two consumers and the marginal price function P^A_{zi} (ignore for the moment the second marginal price function, P^B_{zi}). For reasons discussed in section 3.4, the hedonic price function is likely to be non-linear, and so the marginal price function is likely to be non-constant. In Figure 2, the marginal price of z_i is shown to decrease as more of z_i is consumed.

Figure 2 indicates that each marginal price reveals only one point on each consumer's marginal bid function, and so we cannot make inferences about the shape of the marginal bid function using only information on marginal prices. By observing P_{zi}, we know that one point on the marginal bid function of consumer 1 is $\{P_{zi}^{1*}, z_i^{1*}\}$; however, we cannot determine if the demand function is shaped as shown by the solid line, or the dashed line. Indeed, an infinite number of possible functions could pass through the point $\{P_{zi}^{1*}, z_i^{1*}\}$. This is the standard identification problem of any demand analysis: additional information besides the market-clearing prices for each property in a market is needed to identify demand.

One approach to identifying demand is to use information from separate markets, such as different cities that are geographically distinct. In this approach, one assumes that individuals in one market do not consider housing in the other markets as substitutes when making their purchase decisions. Thus, the hedonic price function in each market is distinct and arises from interactions of only those buyers and sellers located in that market. Individuals with any given vector of socio-economic characteristics are assumed to have preferences over attributes that are identical across markets. However, because of differences in supply, or perhaps the distribution of socio-economic characteristics among individuals within a market, the equilibrium hedonic price functions will vary across markets, and so similar people will be observed making different housing choices across the different markets. If this is the case, estimating separate hedonic price functions in each market will identify demand. This point is illustrated in Figure 2. In figure 2, P^B_{zi} represents the marginal price function from a separate market. Given this additional information, we can now determine if it is point B or C that is the optimal choice for that type of consumer, and whether the marginal bid function is represented by the dashed or solid line.

Thus, to estimate an inverse uncompensated demand (such as in equation (13)) or a system of demands, hedonic price functions are estimated for multiple markets. The marginal prices from each market are pooled with information on property owners demographics and characteristics of their properties to estimate demand. For example, to estimate the demand given in (16), Boyle, Poor, and Taylor (1999) specified a hedonic price function like that given in (1) for four markets in Maine. Each market was distinguished by distances between each other, by unique regional characteristics, and by different real estate multiple listing service regions. The estimated marginal implicit prices, along with corresponding quantities purchased, were pooled with demographic information from a survey of property owners.

The use of multiple markets to identify demand for housing attributes has also been used by Palmquist (1984), Parsons (1986), Bartik (1987), Cheshire and Sheppard (1998), Palmquist and Israngkura (1999), and Zabel and Kiel (2000), with the latter two studies focusing on environmental amenities. The number of markets used to identify demand has varied from two (Cheshire and Sheppard 1998; Bartik 1987) to seven (Palmquist 1984; Parsons 1986), to thirteen (Palmquist and Israngkura 1999). There is no established correct minimum number of individual markets required. Of course, what must be established is that the hedonic price functions do vary across markets, and thus researchers must provide for statistically significant differences in the implicit prices across markets. The issues associated with determining appropriate assumptions for market segmentation are the same here as discussed in section 3.3.

For estimation purposes, some key characteristics are typically assumed to be separable in consumption so that the dimensionality of the problem is reduced. For example, Palmquist (1984) estimates the demand for four housing characteristics, and includes the price of five substitute/complement variables in each demand equation. Cheshire and Sheppard (1998) estimate demand equations for twelve housing characteristics and include eleven possible substitutes and/or complements in each demand equation.[14] Systems methods have been estimated by Palmquist and Israngkura (1999), Parsons (1986), and Stewart and Jones (1998), although the latter study was in a non-housing framework. Regardless of approach, when making weak separability assumptions about utility to reduce the dimensionality of the demand system, one has to recognize that the welfare measures estimated are a lower bound on the true measures (Hanemann and Morey 1998). This is, of course, in addition to the welfare measure being a lower bound to total net benefits in the housing market if one is assuming (say) no moving among residents in response to the amenity change.

If the alternative approach of estimating an MRS function is taken, data from only a single market is needed because the exact specification of the regression is determined by the analytical solution to the utility maximization problem. Here, the researcher simply specifies a different form for the hedonic price function and the MRS function to achieve identification. The researcher must still make sure that the functional forms chosen imply the rank conditions for identification are met (see Quigley 1982 and Chattopadhyay 1999, Appendix 1).

The utility restrictions that must be imposed when estimating the MRS function are not testable, and so this approach is often considered less desirable particularly when data from multiple markets are available. Alternatively, the multiple market approach requires consideration of the appropriateness of the market segments.

5.4 Endogeneity of Prices and Income in Demands

In addition to identifying demand, an important econometric issue that must be taken into account is the possible endogeneity of implicit prices and income. Recall, that for any functional form of the hedonic price function other than linear as given in equation (6), implicit prices may be non-constant. If prices are non-constant, then they are endogenous because consumers can choose the marginal price they pay per unit of the characteristic by choosing the level of the characteristic they consume. Also, in the case of non-constant implicit prices, we linearized the budget constraint to analytically derive the demand function, which involved adjusting income by the implicit prices (equations (11) and (12)). When including this adjusted income in the demand specification, it too will be endogenous because the adjusted income relies on the non-constant implicit prices.

For instance, the hedonic price function estimated by Boyle, Poor, and Taylor (1999) in equation (1) indicates that the marginal price for water clarity and square feet of living space are non-constant. Thus, by choosing the quantity of living area, for instance, the purchasers are also choosing the marginal price per foot of living area they must pay. The marginal price of living area and the choice of square feet of living area are both endogenous.

Endogeneity is typically handled by instrumental variable techniques. First, each price that is endogenous and adjusted income is regressed on a set of exogenous explanatory variables. These exogenous variables should be related to the endogenous variable of interest, but exogenous to the system. The instruments must be (1) correlated with the regressors, (2) uncorrelated with the error term, and (3) of full-rank (add new information). The resulting parameter estimates are used to predict values for the endogenous prices and income. The predicted prices and predicted income are then included in the demand equation, and they are expected to have less correlation with the error term than the original variables (see Greene 1994 for a standard discussion of two-stage least squares).

Choosing proper instruments for the endogenous variables can be a difficult task. For the demand given in equation (16), Boyle, Poor, and Taylor (1999) developed nine instruments for the marginal prices that described factors such as number of lakes in a market area, distance of the market to Boston, a time trend, and local economic and demographic conditions (e.g., an index of real estate activity in the area and the percentage of the current population that is employed). For adjusted income, the instruments were socio-economic characteristics of the property purchasers (as reported in a survey of the owners) and included age and gender of the owner, number of children under 18 years old, educational level, retirement status, the number of people in the household. Also included were the squared terms for age, the number of children under 18 years old, and the number of people in the household.

Palmquist (1984) estimated the demand for four housing characteristics and uses exogenous socio-economic characteristics (age, marital status, number of dependents, race, and their square where appropriate) and used a set of dummy variables for each of his seven urban areas to instrument for the non-constant implicit prices and adjusted income. In his single market, utility restriction approach, Chattopadhay (1999) uses household attributes such as income, income squared, number of dependent children and its square, race, and marital status as instruments for the levels of housing characteristics.

Cheshire and Sheppard (1998) used an innovative approach with "spatially lagged" variables as their instruments. The authors developed a spatial relationship between each house in their data set, and each next "closest" house.[15] For each observation, the characteristic prices and income levels associated with the two houses closest to the house in question were used as instruments. The authors argue that these variables clearly meet requirements (1) and (3) above for instrumental variables, and are likely to meet requirement (2) as they argue there is enough variability in neighborhood housing characteristics (and individual characteristics within a neighborhood). Palmquist (2000) makes the point that because this approach is not unidirectional (i.e., the spatial lags are not unidirectional in the same way time lags are), these instruments are invalid. To alleviate this problem, one might combine time and the spatial lags to create an instrument for the implicit price that is the closest house sold prior to the current house.

A common issue is that instruments do not explain the variation in the endogenous variables very well. In finite samples, weak instruments result in biased estimates of the demand parameters just as in the case of using simple ordinary least squares (Bound, Jaeger, and Baker 1995). Thus, with weak instruments, one may be accepting higher variance of the parameter estimates without reducing the bias. Also if instruments are weak and the loss of efficiency is large, common tests for endogeneity (Hausman 1978) are likely to have weak power (Nakamura and Nakamura 1998). Lastly, a "loss of relevance" can occur when the instruments chosen for the endogenous regressors do not adequately capture the type of variation in the regressor for the application at hand. For these reasons, Nakamura and Nakamura (1998) reject the "always instrument" policy, especially for variables that are not of primary interest to the application or when it is not clear that the potential instruments are truly exogenous. While there is no clear guidance for what are "weak" instruments, R^2 values for the auxiliary equations of less than 0.2 caused concern for Nakamura and Nakamura (1998) in earlier research on female labor supply.

5.5 Summary of Demand Estimation

Table 4 summarizes the steps involved in estimating a demand function for a characteristics of a differentiated good and highlights a few of the main issues to be considered in each step. These steps apply regardless of whether utility restrictions in a single market are used to identify demand or multiple markets are used to identify demand. For a single market, steps 3 and 4 must still be undertaken if the hedonic price function in the first stage (step 1) is non-linear. Also, in the single market approach, the functional form for the equation to be estimated in step 6 is derived from an underlying utility specification and is usually the marginal rate of substitution function. Exact welfare measures are then easily obtained (step 7) as described in section 5.2.

6. LABOR MARKETS AND OTHER HEDONIC APPLICATIONS

Hedonic theory has been applied to differentiated goods markets as varied as labor markets, child care services, agricultural products, and computers. Most often, these applications are first-stage analyses estimating only the hedonic price function, and not underlying demands for characteristics of the

Table 4. Steps for Conducting a Hedonic Demand Analysis when Multiple Markets are Used to Identify Demand

1. Estimate hedonic price function for each market.
   - Choose functional form (see section 3.4)
   - Test for market segmentation (see section 3.3)

2. Obtain marginal prices for characteristics of interest (see Table 2) for each observation in each market, $\mathbf{p}_i$, including price of the characteristic of interest and all relevant substitutes and complements.
   - Test for price differences across markets

3. Instrument non-constant (endogenous) prices to obtain predicted prices: **p-hat**$_i$. This is accomplished by pooling data across markets and estimating p_i as a function of exogenous variables that are related to p_i.
   - Choose of instrumental variables (see section 5.4)
   - Choose of functional form

4. Adjust income if hedonic price function is non-linear by computing the following for each observation:

   Y_{adj} = Y - Price of House + Sum over characteristics of: marginal implicit price of characteristic multiplied by the chosen characteristic level (see equation 12)

5. Instrument adjusted income to obtain predicted adjusted income: **Y-hat**$_{adj}$. This is accomplished by pooling income data across markets and estimating Y_{adj} as a function of exogenous variables that are related to income.
   - Choose instrumental variables (see section 5.4)
   - Choose functional form

6. Estimate demand: $\mathbf{z}_i = z(\mathbf{p}_i, \mathbf{p}_{j\neq i}, \mathbf{Y\text{-}hat}_{adj}, \boldsymbol{\tau})$, where τ are a set of demand shifters. Data from all markets are pooled.
   - Choose functional form
   - Choose relevant demand shifters

7. Compute welfare measures (see sections 5.1 and 5.2):
   a) Integrate under demand for consumer surplus changes. Levels of all other variables (besides own-price and quantity) must be fixed at some level, often the mean of the observed data.
   b) Use duality to obtain exact welfare changes.

good. This section provides an overview of hedonic applications to markets other than housing with particular attention to labor markets.

6.1 Labor Market Applications

The hedonic approach to understanding equilibrium pricing in markets for differentiated goods can easily be applied to the labor market. The hedonic wage equation is the envelope of individual firm's offer functions and worker's indifference curves. Reconsider Figure 1, and let z_1 represent the risk of on-the-job injury. In this case, Φ now represents the wage tradeoffs workers are willing to make for additional units of risk of on-the-job injury, holding utility constant. Higher contour levels of Φ would represent higher levels of utility. Similarly, θ are now iso-profit curves of the firm, representing the tradeoffs the firm can make between higher wages and expenditures to reduce the risk of on-the-job injury, holding profit constant. Lower contour levels of θ represent higher levels of profits (see Rosen 1979 for a more detailed treatment).

Using a wage hedonic to estimate the tradeoffs workers are willing to make between wages and the risk of death on the job is an area of nonmarket valuation that has stimulated much research over the past 30 years. Estimates of these dollar tradeoffs are used to compute the value of a statistical life; not on the value of a particular individual (an "identified life"), but the value society places on a statistical reduction in the probability of one death among them (see Viscusi 1992, 1993). For instance, consider the following simple wage hedonic:

$$wage_k = \alpha + \beta_r risk_k + \sum_{n=1}^{N} \lambda_n X_{kn} + \sum_{m=1}^{M} \gamma_m D_{km} + \varepsilon_k, \tag{21}$$

in which the wage of the k^{th} worker is estimated to be a function of the risk of death on the job ($risk_k$), N variables describing human capital and demographic characteristics of the worker (X_{kn}) such as age and education, and M job characteristics (D_{km}) other than the risk of death such as whether or not supervisory activities are associated with the job.[16] In a linear specification such as (21), the implicit wage for risk, $\partial w/\partial r = \beta_r$, is the additional wages a worker would require to assume an additional increment of risk of death on the job. By normalizing over risk, the compensating wage differential for risk is converted to the value of a statistical life (VSL). For instance, suppose risk is measured in units of deaths per 10,000 workers, wages are hourly earnings, and a simple linear compensating wage equation is estimated as illustrated in (21). To compute the value of a statistical life, $\hat{\beta}_r$ is multiplied by 2000 hours/year to arrive at the annual compensation required for an additional unit of risk, and then by 10,000, the number of workers over which the risk is spread. In this example, an estimate of $\hat{\beta}_r$ equal to 0.35 would imply a value of statistical life estimate of $7 million.

In a policy context, the value of statistical life is particularly important for evaluating policies that reduce human mortality. Evaluating these policies in a benefit-cost framework requires an estimate of the value that society places on a life saved as a result of the policy. The appropriate measure of such a value is the value of statistical life (VSL). A recent example of the application of this methodology to benefit/cost analyses is the U.S. Environmental Protection Agency's (EPA) retrospective (1970-1990) and prospective (1990-2010) analyses of the Clean Air Act (Environmental Protection Agency 1997 and 1999). In these analyses, the EPA evaluated the benefit and costs of emissions controls imposed by the Clean Air Act and its associated regulations for the period 1970 to 2010. To monetize the benefits associated with mortality reductions, the single best estimates of the value of a statistical life reported in 26 different studies (21 of which were labor market studies) were used to compute a mean value of $4.8 million (1990 dollars, Environmental Protection Agency, 1997, 44). This value was then aggregated over the number of lives saved to compute the monetized benefit of reduced mortality in both the retrospective and prospective studies. For example, in the retrospective analysis, the mean benefits of reduced mortality were estimated to be $17.97 trillion (1990 dollars; Environmental Protection Agency 1997, 52-53), just over 80% of the total benefits estimated to be associated with the Clean Air Act from 1970 to 1990.

While the VSL is an appropriate value to use in benefit/cost analyses, there is much controversy surrounding the empirical estimates of this value. First, the large variation in the VSL estimates that make the choice of which value is best for any specific policy application difficult. In a recent quantitative review of this literature, Mrozek and Taylor (2002) conducted a meta-analysis drawing on VSL estimates from over 30 studies using labor markets to estimate the VSL. The VSL estimates from these studies ranged from under $100,000 to over $20 million (1998 dollars). Based on their review, Mrozek and Taylor (2002) conclude that the appropriate estimate of the VSL that can be drawn from this past literature, when correcting original studies for design weaknesses, is approximately $2 million (1998 dollars) — a decrease of 50%

or more over the value suggested by previous reviewers of the literature (see for example, Viscusi 1993).

A second, perhaps more important, concern is whether or not a VSL estimate derived from studies of fatal workplace risks is an appropriate value for reducing the risks of death from illnesses, where there may be prolonged reductions in the quality of life or more dread and/or pain involved. As of yet, there is no systematic evidence as to how these latent risks are treated by individuals relative to immediate risks in an economic framework. An important extension to the VSL literature is to estimate the wage tradeoffs workers make for on the job exposure to latent risks of death (such as exposure to cancer-causing agents). However, this may be difficult to implement due to a lack of data and measurement difficulties. A related concern is whether or not the VSL derived from labor market studies, which are based on the working-age population of healthy adults, are appropriate to extend to other populations at risk such as the elderly or children. Again, empirical research is needed to determine the systematic differences, if any, between the different populations' responses to risks in an economic context.

An extension of the wage hedonic is to recognize that not only will characteristics of the individual and the job affect wages, but also the characteristics of the location of the job. The recognition that amenities are capitalized into wages as well as housing prices is the basis of "quality of life indexes" (QOLIs) that have been constructed to rank amenities across urban areas in the U.S. The framework for estimating QOLIs is as follows. Consider two identical jobs, one located in a desirable community in terms of school and environmental quality, the other located in a less desirable community. If the communities are located a distance from each other sufficient to make commuting between them unlikely, then the firm located in the less desirable community will have to offer higher wages, *ceteris paribus*, to attract workers to the firm. The firm location and the amenities associated with that location affect wages and should be included in the compensating wage equation (Rosen 1979; Smith 1983). The important assumption here is that labor markets are large enough so that amenities vary across communities within a market (usually, labor markets are assumed to be national in scope), but the communities are distant enough from each other to prevent commuting between them for work purposes. A single wage hedonic applies to the entire market, and workers face different implicit prices (or wage differentials) for different levels of amenities in various communities.

To compare quantities of amenities across locations, the quantities must be aggregated using consistent units of measurement. A consistent index of the level of amenities at a specific location is one where each amenity is weighted by its full implicit price and then aggregated (Rosen 1979 and Roback 1982). The full implicit price (FP) is equal to the sum of the housing expenditure differential associated with the amenity and the negative of the wage differential associated with the amenity or:

$$FP_k = h_k(\partial P_k / \partial a_k) - \partial w_k / \partial a_k, \tag{22}$$

where $\partial P_k/\partial a_k$ is the implicit price for amenity k in the housing market, h_k is the quantity of k associated with a property, and $\partial w_k/\partial a_k$ is the wage differential associated with amenity k. The full implicit price of locational amenities may then be combined with the levels of amenities at each location to compute the aggregate household expenditure on amenities, which is equal to total household WTP for marginal differences in these amenities (Blomquist, Berger, and Hoehn 1988).

Blomquist, Berger, and Hoehn (1988) computed a quality of life index (QOLI) for 253 urban areas across the U.S. Their index aggregated expenditures (based on full implicit prices) on 15 environmental and social amenities such as air quality, rainfall, violent crime, and teacher-pupil ratios. Based on the average compensation for these amenities, they ranked 253 urban counties, and found Pueblo, Colorado to be ranked highest (with a QOLI = \$3,289 in 1980 dollars), and found St. Louis, Missouri to be ranked lowest (with a QOLI = \$-1,857 in 1980 dollars). Because the unit of analysis is each county, Blomquist, Berger and Hoehn were also able to identify large QOLI differentials within metropolitan areas. For instance, two counties in Milwaukee, Wisconsin had different QOLIs, differing by over \$2,000. One county ranked 68th, and the other ranked 252 (out of 253 total). More recent applications and extensions of this literature include using a revealed preference approach based on hedonic prices to rank quality of life in cities (Kahn 1995) and to examine how firms respond to quality of life measures in their location decisions (Blomquist and Maury 1999).

6.2 Non-Housing Hedonic Demand Applications

The first stage hedonic method has been applied to markets as varied as consumer durables (e.g., automobiles and computers), agricultural commodities, labor markets, and cultural commodities. There are fewer examples of hedonic demand (second stage) analyses outside of housing markets. Hedonic demand applications are sparse likely due to the additional data required: information on the individuals (or firms) who purchased the products. Of course, this is true for estimating demand in the market for a homogeneous good. But in a market for a homogeneous good, one may often rely on aggregate observations such as sales per month, monthly average price, and aggregate demand information such as mean income across households. In a market for differentiated goods, this type of aggregation is undesirable because the individual variation in product variety drives the model. Individual-level sales information is needed and thus, demographic information on individual purchasers is needed.

A non-housing example of a single market approach to estimating demand is Stewart and Jones (1998) who estimated a system of factor demands for the characteristics of major league baseball players. They identified demand by assuming a translog or generalized Leontief cost function. Their estimated system of demands are largely consistent with predictions of economic theory (i.e., demands are homogeneous of degree zero and cross-substitution effects are symmetric), and are not sensitive to the form assumed for the cost function.

Hagy (1998) and Blau and Hagy (1998) used a multiple market approach in a hedonic analysis estimating the demand for attributes of child care. Hagy (1998) focused on staff-to-child ratio, and matches data on households and their choices over the quantity and quality of child care with the estimated implicit price of staff-to-child ratio in their area. She used 55 geographically distinct markets to identify demand. Her results indicate that demand is not sensitive to economic factors such as mother's wage rate and the implicit price of staff-to-child ratio; rather, demand is sensitive to factors such as the age of the youngest child and the availability of a relative near the household. These results are important for public policy because they suggest that government subsidization of child care, which effectively increases income or decreases the implicit prices, is unlikely to have a significant effect on the quality of care purchased.

Although second-stage analyses may be more difficult and data intensive, the demand analysis can provide much richer insights than just a first stage analysis alone as well illustrated by Hagy (1998).

7. SUMMARY

This chapter focused on the possibilities for non-market valuation using hedonic methods, primarily with property markets. Estimation of the hedonic price function alone (a first-stage analysis), is by far the most common hedonic application. Indeed, a search using a popular academic search engine using the keyword "hedonic" revealed over 850 published works in the economics literature. The popularity of the hedonic method is not surprising given the relatively minimal data requirements and straightforward empirical implementation. In addition, the insights we may draw from a first-stage analysis are quite powerful: the of a change in an environmental amenity that accrue to a population of residents in an area.

But, we also saw that many details require careful consideration when estimating the hedonic price function. Choice of the independent variables, their measurement, the functional form of the price function, and sample frame all require thoughtful treatment. Not all issues will require in-depth treatment. The nature of the question being asked and the nature of the data available for estimation will focus the attention of the researcher on the key issues that must be considered in detail.

Even very many first-stage hedonic analyses have been conducted, this is still an exciting area of research. We are just beginning to incorporate the spatial nature of housing data fully into hedonic models. Spatial aspects of markets have always been included in empirical applications in rudimentary manners (such as computing the linear distance of a property to the center of a city), the advent of inexpensive computing and widely available geographic information systems software have made possible the consideration of more complex spatial relationships. In addition, spatial econometric methods are increasingly applied in hedonic studies and are likely to be standard in hedonic analysis in the near future.

Beyond just the first-stage estimation of the hedonic price function, much additional research can be conducted on estimating welfare changes within the hedonic framework. Although welfare analysis in this context has been actively discussed since Rosen's (1979) seminal article, in the twenty-odd years since then, confusion remains as to how, and under what conditions, the hedonic method may be used to precisely estimate household welfare changes. Relatively few studies have attempted to address this issue. Second-stage analyses that estimate the demand for amenities not directly traded in markets are particularly useful because they provide rich information on the links between demographics and preferences for amenities. Understanding these links is essential for targeting effective policy. Future studies can easily combine property value databases with surveys of homeowners to estimate such models.

ACKNOWLEDGMENTS

The author would like to thank Kevin Boyle, Tom Brown, Patty Champ, George Parsons, and the participants of the non-market valuation workshops sponsored by the Rocky Mountain Research Station, April and October 2000, for helpful comments. Thanks also go to Peter Grigelis for his research assistance. Special thanks is extended to Ray Palmquist for his comments on this chapter as well as numerous conversations that helped to clarify my understanding and exposition of several topics.

NOTES

1 Examples of hedonic applications are Berry, Kortum, and Pakes (1996) and Dreyfus and Viscusi (1995) for automobiles; Stavins (1997), Berndt, Griliches, and Rappaport (1995) for computers; Liegey and Shepler (1999) for VCRs; and Greening, Sanstad, and McMahon (1997) for appliances. Agricultural commodities to which the hedonic method has been applied are milk, shrimp, wine, and cigars (see, respectively, Gillmeister, Yonders, and Dunn 1996, Salayo, Voon, Selvanathan 1999, Combris, Sebastien, and Visser 1997, and Stover 1996). Examples of applications to crops such as wheat and cotton are Espinosa and Goodwin (1991) and Bowman and Ethridge (1992), respectively.

2 For example, the lakefront cottages used by Boyle, Poor, and Taylor (1999) were assumed to have eight primary characteristics that determined price (as shown in equation 1), which could be labeled with variable names z_1 to z_8. In this case, the equilibrium price schedule for lakefront properties is generally represented by $P(z_1, z_2, z_3, z_4, z_5, z_6, z_7, z_8)$ or $P(\underline{z})$ or $P(Z)$. Recall that Z represents houses (so Z^1, Z^2, etc. represent house 1 and house 2, respectively, each with their own unique levels of the characteristics z_1 to z_8). Each of these representations indicates that the price of a house is a function of the characteristics of the house.

3 It may also be assumed that firms produce multiple types of the good, and that the cost function is separable in each type of product. The maximization problem for the firm choosing an entire product line in this case is equivalent to choosing each product type separately as described here.

4 Another property of the bid function is that optimal bids increase proportionally with income $\partial\theta/\partial y = 1$. If income increases by $1, the consumer's optimal bid must increase by the same amount to hold utility constant.

5 Linneman (1980) suggested that housing markets may be national, and compares city-specific hedonic price functions for housing with a hedonic price function estimated on a national sample of housing. His hypothesis of a national market is neither fully supported nor rejected.

6 Other studies focus on the impacts of similar disamenities on property values, but do not consider timing effects are Keil (1995) and Dale, et al. (1999).

7 The common example here is to think of a full grocery cart as "the commodity" which has component characteristics that are the grocery items contained within. In this case, where it is costless to repackage the component characteristics of the grocery cart, so the total price for the product is simply the sum of the prices for each of its component characteristics.

8 Maximum likelihood estimates involve calculating the N eigenvalues of the spatial weight matrix, and the precision of the numerical procedures used to calculate these eigenvalues decreases as the number of observations increases (Kelejian and Prucha, 1999).

9 In the case of a localized amenity, we are able to forecast the change in price prior to the amenity change actually occurring because the hedonic equilibrium does not change. One still must be careful that the magnitude in the change of the amenity is not outside the range observed in the total sample. If so, the prediction is out-of-sample and becomes less reliable the further out-of-sample the change.

10 This derivation requires, for non-constant marginal prices, that the budget constraint be linearized as discussed in the next section.

11 Chattopadhyay notes in his discussion that this function (equation [1b] in his article) is equal to the inverse demand for z_i (Chattopadhyay, 1999, 23). Although the inverse demand and marginal rate of substitution function are equal at one point (the optimal consumption bundle), they are not equivalent functions.

12 This function does not duplicate Chattopadhyay (1999), equation (5b), 25 because of a minor typographical error in the paper.

13 An equivalent to equation (20) is reported in Cropper et al. (1993, p.226, equation (3)). However, Cropper et al. do not make explicit the substitution that must be made in the marginal rate of substitution function for it to be equivalent to the marginal bid function.

14 The choice of substitutes and complements to include is often somewhat ad-hoc and judgements must be applied to determine the important factors to include.

15 Closeness is defined by both geographic proximity and similarity of the structures.

16 The compensating wage equation describes the equilibrium wage as a function of the factors affecting the wage negotiation. This negotiation will be in regard to factors such as job duties and job working conditions, as well as characteristics of the workers themselves that affect expected productivity.

REFERENCES

Anselin, L. 1988. *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer Academic Press.

Anselin, L., and A. K. Bera. 1998. Spatial dependence in linear regression models with an introduction to spatial econometrics. In *Handbook of Applied Economic Statistics*. Edited by A. Ullah and D. E. Giles New York: Marcel Dekker:237-89.

Anselin, L., A. K. Bera, R. Florax, and M. J. Yoon. 1996. Simple diagnostic tests for spatial dependence. *Regional Science and Urban Economics* 26:77-104.

Anselin, L., and R. J. Florax. 1995. *New Directions in Spatial Econometrics* New York:Springer Verlag.

Atkinson, S. E., and T. D. Crocker. 1987. A Baysian approach to assessing the robustness of hedonic property value studies. *Journal of Applied Econometrics* 2:27-45.

Bartik, T. J. 1986. Neighborhood revitalization's effects on tenants and the benefit-cost analysis of government neighborhood programs. *Journal of Urban Economics* 19:234-248.

Bartik, T. J. 1987. The estimation of demand parameters in hedonic price models. *Journal of Political Economy* 95:81-88.

Bartik, T. J. 1988. Measuring the benefits of amenity improvements in hedonic price models. *Land Economics* 64:172-83.

Beach, E. D., and G. A. Carlson. 1993. A hedonic analysis of herbicides: do user safety and water quality matter? *American Journal of Agricultural Economics* 75(3):612-23.

Bell, K. P., and N. E. Bockstael. 2000. Applying the generalized-moments estimation approach to spatial problems involving microlevel data. *Review of Economics and Statistics* 82:72-82.

Belsley, D., E. Kuh, and R. Welsch. 1980. *Regression Diagnostics*. New York: Wiley.

Berndt, E. R., S. H. Busch, and R. G. Frank. 1998. *Price indexes for acute phase treatment of depression*. National Bureau of Economic Research Working Paper: 6799.

Berndt, E. R., Z. Griliches, and N. J. Rappaport. 1995. Econometric estimates of price indexes for personal computers in the 1990's. *Journal of Econometrics* 68(1):243-68.

Berry, S., S. Kortum, and A. Pakes. 1996. *Environmental change and hedonic cost functions for automobiles*. National Bureau of Economic Research Working Paper: 5746.

Blau, D. M., and A. P. Hagy. 1998. The demand for quality in child care. *Journal of Political Economy* 106(1):104-46.

Blomquist, G. C., M. C. Berger, and J. P. Hoehn. 1988. New estimates of quality of life in urban areas. *American Economic Review* 78(1):89-107.

Blomquist, G. C., and D. Maury. 1999. Evaluating the Influence of Amenities on the Location of Manufacturing Establishments in Urban Areas. *Urban Studies* 36(October): 1859-1873.

Bockstael, N. E. 1996. Modeling economics and ecology: the importance of a spatial perspective. *American Journal of Agricultural Economics* 78:1168-1180.

Bound, J., D. Jaeger, and R. Baker. 1995. Problems with instrumental variables estimation when the correlation between instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association* 90:443-450.

Bowman, K. R., and D. E. Ethridge. 1992. Characteristics supplies and demands in a hedonic framework: U.S. market for cotton fiber attributes. *American Journal of Agricultural Economics* 74(4):991-1002.

Boyle, K. J., J. Poor, and L. O. Taylor. 1999. Estimating the demand for protecting freshwater lakes from eutrophication. *American Journal of Agricultural Economics* 81(5):1118-22.

Boyle, K. J., and L. O. Taylor. 2001. Does the measurement of property and structural characteristics affect estimated implicit prices for environmental amenities in a hedonic model? *Journal of Real Estate Finance and Economics*. Forthcoming.

Butler, R. V. 1980. Cross-sectional variation in the hedonic relationship for urban housing markets. *Journal of Regional Science* 20:439-453.

Chanel, O., L. Gerard-Varet, and V. Ginsburgh. 1996. The relevance of hedonic price indices: the case of paintings. *Journal of Cultural Economics* 20(1):1-24.

Chattopadhyay, S. 1999. Estimating the demand for air quality: new evidence based on the Chicago housing markets. *Land Economics* 75(1):22-38.

Cheshire, P., and S. Sheppard. 1995. On the price of land and the value of amenities. *Economica* 66:247-267.

Cheshire, P., and S. Sheppard. 1998. Estimating the demand for housing, land, and neighbourhood characteristics. *Oxford Bulletin of Economics and Statistics* 60(3):357-382.

Cockburn, I. M., and A. H. Anis. 1998. *Hedonic analysis of arthritis drugs*. National Bureau of Economic Research Working Paper: 6574.

Combris, P., L. Sebastien, and M. Visser. 1997. Estimation for a hedonic price equation for Bordeaux wine: does quality matter? *Economic Journal* 107(441):390-402.

Cropper, M. L., L. Deck, and K. E. McConnell. 1988. On the choice of functional form for hedonic price functions. *Review of Economics and Statistics* 70(4):668-675.

Cropper, M. L., L. Deck, N. Kishor, and K. E. McConnell. 1993. Valuing product attributes using single market data: a comparison of hedonic and discrete choice approaches. *Review of Economics and Statistics* 75(2):225-32.

Czujack, C. 1997. Picasso paintings at auction, 1963-1994. *Journal of Cultural Economics* 21(3):229-47.

Dale, L., J. Murdoch, M. Thayer, and P. Waddell. 1999. Do property values rebound from environmental stigmas? Evidence from Dallas. *Land Economics* 75:311-326.

Danzon, P.M., and L W Chao. 2000. Cross-national price differences for pharmaceuticals: how large and why? *Journal of Health Economics* 19(2):159-95.

Dreyfus, M. K., and W. K. Viscusi. 1995. Rates of time preference and consumer valuations of automobile safety and fuel efficiency. *Journal of Law and Economics* 38(April):79-105.

Espinosa, J. A., and Goodwin, B. K. 1991. Hedonic price estimation for Kansas wheat characteristics. *Western Journal of Agricultural Economics* 16(1):72-85.

Freeman, A. Myrick III. 1993. *The Measurement of Environmental and Resource Values*, (Resources for the Future, Washington, DC).

Garber, S., and S. Klepper. 1980. Extending the classical normal errors-in-variables model. *Econometrica* 48(6):1541-46.

Gillmeister, W. J., R. D. Yonkers, and J. W. Dunn. 1996. Hedonic pricing of milk components at the farm level. *Review of Agricultural Economics* 18(2):181-92.

Geohegan, J., L. A. Wainger, and N. E. Bockstael. 1997. Spatial landscape indices in a hedonic framework: an ecological economics analysis using GIS. *Ecological Economics* 23:251-64.

Granger, M. D., and G. C. Blomquist. 1999. Evaluating the influence of amenities on the location of manufacturing establishments in urban areas. 36(11):1859-73.

Graves, P., J. C. Murdoch, M. A. Thayer, and D. Waldman. 1988. The Robustness of Hedonic Price Estimation: Urban Air Quality. *Land Economics* 64(3): 220-33.

Greene, W. H. 1994. *Econometric Analysis*. Upper Saddle River, NJ: Prentice Hall.

Greening, L. A., A. H. Sanstad, and J. E. McMahon. 1997. Effects of appliance standards on product price and attributes: an hedonic pricing model. *Journal of Regulatory Economics* 11(2):181-94.

Gwande, K., and H. Jenkins-Smith. 2001. Nuclear waste transport and residential property values: estimating the effects of perceived risk. *Journal of Environmental Economics and Management*. Forthcoming.

Hagy, A. P. 1998. The demand for child care quality. *Journal of Human Resources* 33(3):683-710.

Halvorsen, R. and R. Palmquist. 1980. The interpretation of dummy variables in semilogarithmic equations. *American Economic Review* 70:474-75.

Hanemann, M., and E. Morey. 1998. Separability, partial demand systems, and consumer's surplus measures. *Journal of Environmental Economics and Management* 22:241-58.

Harchaoui, T. M., and M. Hamdad. 2000. The prices of classical recorded music: a hedonic approach. *International Journal of Industrial Organization* 18(3):497-514.

Hausman, J. A. 1978. Specification Tests in Econometrics. *Econometrica* 46:1251-71.

Hausman, J. A. 1981. Exact consumer's surplus and deadweight loss. *American Economic Review* 71:662-676.

Huang, J., and R. B. Palmquist. 2001. Environmental conditions, reservation prices, and time on the market for housing. *Journal of Real Estate Finance and Economics*. Forthcoming.

Ihlanfeldt, K. R., and J. Martinez-Vazquez. 1986. Alternative value estimates of owner-occupied housing: evidence on sample selection bias and systematic errors. *Journal of Urban Economics* 20:356-69.

Ihlanfeldt, K. R., and L. O. Taylor. 2001. Externality Effects of Small-Scale Hazardous Waste Sites: Evidence from Urban Commercial Property Markets. Environmental Policy Program Working Paper #2001-002, Georgia State University, Atlanta, GA.

Israngkura, A. 1994. Environmental benefit measures: a comparison between hedonic and discrete choice models, Ph.D. dissertation, North Carolina State University, Raleigh, NC.

Kahn, M. E. 1995. A revealed preference approach to ranking city quality-of-life. *Journal of Urban Economics* 38(2): 221-35.

Kelejian, H., and I. Prucha. 1999. A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review* 40(2):509-33.

Kelejian, H., and D. Robinson. 1992. Spatial autocorrelation: a new computationally simple test with an application to per capita county police expenditures. *Regional Science and Urban Economics* 22:317-31.

Kennedy, P. E. 1982. Estimation with correctly interpreted dummy variables in semilogarithmic equations. *American Economic Review* 72:801.

Kiel, K. A. 1995. Measuring the impact of the discovery and cleaning of identified hazardous waste sites on housing values. *Land Economics* 71(4):428-35.

Kiel, K. A., and K. T. McClain. 1995. Housing prices during siting decision stages: the case of an incinerator from rumor through operation. *Journal of Environmental Economics and Management* 28:241-255.

Kiel, K. A., and J. E. Zabel. 1999. The accuracy of owner-provided house values: the 1978-1991 American Housing Survey. *Real Estate Economics* 27:263-398.

King, X., X. White, and D. Shaw. 1991.Influence of urban wildlife habitats on the value of residential properties. In *Wildlife Conservation in Metropolitan Environments: A National Symposium on Urban Wildlife*. Edited by Adams and Leedy NIUW Symposium Series 2, National Institute for Urban Wildlife, Columbia, MD.

Kohlhase, J. E. 1991. The impact of toxic waste sites on housing values. *Journal of Urban Economics* 30:1-26.

Leggett, C. G., and N. E. Bockstael. 2000. Evidence of the effects of water quality on residential land prices. *Journal of Environmental Economics and Management* 39(2):121-144.

Liegy, P. R., and N. Shepler. 1999. Adjusting VCR prices for quality change: a study using hedonic methods. *Monthly Labor Review* 122(9):22-37.

Linneman, P. 1980. Some empirical results on the nature of the hedonic price function for the urban housing market. *Journal of Urban Economics* 8:47-68.

Maddalla, G. S. 1988. *Introduction to Econometrics*. New York, NY: Macmillan.

Michaels, R. G., and V. K. Smith. 1990. Market segmentation and valuing amenities with hedonic models: the case of hazardous waste sites. *Journal of Urban Economics* 28:223-42.

Mrozek, J. R., and L. O. Taylor. 2002. What determines the value of life? A meta analysis. *Journal of Policy Analysis and Management* 21(2):253-270.

Nakamura, A., and M. Nakamura. 1998. Model specification and endogeneity. *Journal of Econometrics* 83:213-37.

Nimon, W., and J. Beghin. 1999. Are eco-labels valuable? Evidence from the apparel industry. *American Journal of Agricultural Economics* 81:801-11.

Ohta, M., and Z. Griliches. 1975. Automobile prices revisited: extensions of the hedonic hypothesis. *Household Production and Consumption.* Edited by N. Terleckyj. New York: National Bureau of Economic Research.

Osborne, L. 1995. *Market Structure, Hedonic Models, and the Valuation of Environmental Amenities*, Ph.D. Dissertation, North Carolina State University, Raleigh, NC.

Palmquist, R. B. 1984. Estimating the demand for the characteristics of housing. *The Review of Economics and Statistics* 66(3):394-404.

Palmquist, R. B. 1988. Welfare measurement for environmental improvements using the hedonic model: the case of nonparametric marginal prices. *Journal of Environmental Economics and Management* 15:297-312.

Palmquist, R. B. 1991. Hedonic Methods. In *Measuring the Demand for Environmental Quality.* Edited by J. B. Braden and C. D. Kolstad. North Holland, Amsterdam, 77-120.

Palmquist, R. B. 1992a. Valuing Localized Externalities. *Journal of Urban Economics* 31:59-68.

Palmquist, R. B. 1992b. A note on transactions costs, moving costs, and benefit measurement. *Journal of Urban Economics* 32:40-44.

Palmquist, R. B. 2000. Property value models. In *Handbook of Environmental Economics.* Edited by K. G. Mäler and Vincent. Department of Economics, North Carolina State University, Elsevier North Holland. Forthcoming.

Palmquist, R. B., and A. Israngkura. 1999. Valuing air quality with hedonic and discrete choice models. *American Journal of Agricultural Economics* 81(5):1128-1133.

Parsons, G. R. 1986. An almost ideal demand system for housing attributes. *Southern Economic Journal* 53(2):347-363.

Patterson, R. W., and K. J. Boyle. Out of sight, out of mind: Using GIS to incorporate topography in hedonic property value models. *Land Economics.* Forthcoming.

Pollard, R. 1980. Topographic Amenities, Building Height, and the Supply of Urban Housing. *Regional Science and Urban Economics* 10:180-99.

Poor, J., K. Boyle, L. Taylor, and R. Bouchard. 2001. Objective versus Subjective Measure of Environmental Quality in Hedonic Property-Value Models. *Land Economics* 77(4):482-493.

Quigley, J. M. 1982. Nonlinear budget constraints and consumer demand: an application to public programs for residential housing. *Journal of Urban Economics* 12:177-201.

Rao, V. 1993. The rising price of husbands: a hedonic analysis of dowry increases in rural India. *Journal of Political Economy* 101(4):666-77.

Roback, J. 1982. Wages, rents, and the quality of life. *Journal of Political Economy* 90:12257-78.

Rosen, S. 1974. Hedonic prices and implicit markets: product differentiation in pure competition. *Journal of Political Economy* 82(January/February):34-55.

Rosen, S. 1979. Wage-based indexes of urban quality of life. In *Current Issues in Urban Economics*, Edited by P. Mieszkowski and M. Straszheim. Baltimore:Johns Hopkins University Press.

Salayo, N. D., T. Voon, and S. Selvanathan. 1999. Implicit prices of prawn and shrimp attributes in the Philippine domestic market. *Marine Resource Economics* 14(1):65-78.

Schnare, A., and R. Struyk. 1976. Segmentation in urban housing markets. *Journal of Urban Economics* 3:146-166.

Schultz, S. D., and D. A. King. 2001. The use of census data for hedonic price estimates of open space amenities and land use. *Journal of Real Estate Finance and Economics* 22(2/3):239-252.

Sieg, H., V. K. Smith, H. S. Banzhaf, and R. Walsh. 2000. *Estimating the general equilibrium benefits of large policy changes: the clean air act revisited.* NBER Working Paper 7744, Cambridge: nber.org/papers/w7744

Smith, V. K. 1983. The role of site and job characteristics in hedonic wage models. *Journal of Urban Economics* 12:296-321.

Stewart, K. G., and J. C. H. Jones. 1998. Hedonics and demand analysis: the implicit demand for player attributes. *Economic Inquiry* 36(2):192-202.

Stavins, J. 1997. Estimating demand elasticities in a differentiated product industry: the personal computer market. *Journal of Economics and Business* 49(4):347-67.

Stover, M. E. 1996. A hedonic price index for premium cigars. *Journal of Economics* 22(2):63-73.

Strazheim, M. R. 1974. Hedonic estimation of housing market prices: a further comment. *Review of Economics and Statistics* 55:1-8.

Taylor, L. O., and V. Kerry Smith. 2000. Environmental amenities as a source of market power. *Land Economics* 76(4):550-568.

United States Environmental Protection Agency. 1997. *The Benefits and Costs of the Clean Air Act: 1970 to 1990.* Prepared by the Office of Administration and Resources Management, Office of Policy, Planning, and Evaluation for the U.S. Congress, October, 1997.

United States Environmental Protection Agency. 1999. *The Benefits and Costs of the Clean Air Act: 1990 to 2010.* Prepared by the Office of Air and Radiation, Office of Policy for the U.S. Congress, November, 1999.

Vartia, Y. 1983. Efficient Methods of Measuring Welfare Change and Compensated Income in Terms of Ordinary Demand Functions. *Econometrica* 51(1):79-98.

Viscusi, W. K. 1992. *Fatal tradeoffs: Public and private responsibilities for risk.* New York: Oxford University Press.

Viscusi, W. K. 1993. The Value of Risks to Life and Health. *Journal of Economic Literature* 31(4): 1912-46.

Waugh, F. 1928. Quality factors influencing vegetable prices. *Journal of Farm Economics* 10:185-196.

Wilman, E. 1984. *The External Costs of Coastal Beach Pollution: A Hedonic Approach.* Washington, DC: Resources for the Future.

Zabel, J. E., and K. A. Kiel. 2000. Estimating the demand for air quality in four U.S. cities. *Land Economics* 76(2): 174-194.

Chapter 11

DEFENSIVE BEHAVIOR AND DAMAGE COST METHODS

Mark Dickie
University of Central Florida

1. INTRODUCTION

Suppose you wake up one morning and stumble into the kitchen to make coffee. You flip on the TV. While filling the coffee pot with tap water, you learn that your water supply is contaminated with *Giardia lamblia.* The newscaster reports that consuming contaminated water can lead to giardiasis, a common diarrheal illness that typically lasts a few days but can drag on for weeks. Infected persons may have to miss work or school or, in rare cases, may require hospitalization. Local authorities recommend boiling tap water before consuming it or finding alternate sources like bottled water. What do you do?

Maybe you decide to skip the coffee until you can buy some bottled water, and that you will boil any tap water used for drinking or cooking. You know that taking these precautions will cost some money and take some time but you figure the inconvenience is better than getting sick. And as it turns out, you never contract giardiasis. If this sounds like a choice that you might make, then this chapter describes a valuation method that might work for you: the defensive behavior method.

But maybe you don't react that way at all. Maybe water contamination is too much to think about in the morning, so you flip the TV to a cartoon channel and enjoy a few re-runs with your coffee. Soon you've forgotten about the health advisory, and you go about your usual daily business. Later you find yourself sick in bed and you don't understand how this happened or what you might have done to prevent it. You've lost income from missing work, and you had to pay for a doctor's visit and medication. But you don't feel any pain or emotional distress from the illness itself. If that sounds like a choice you might make, then this chapter describes a different valuation method for you: the damage cost method.

Defensive behavior and damage cost methods are applied to giardia contamination as follows (Harrington, Krupnick, and Spofford 1989). Suppose that boiling water and buying bottled water costs you $500 in money and time costs during a three-month incident of contamination. But taking these defensive behaviors is so effective that you avoid the case of giardiasis you would have otherwise experienced. According to the defensive behavior method, your actions reveal that avoiding the illness is worth at least $500 to you. On the other hand, had you not taken the defensive behaviors, you would have revealed that avoiding the illness is worth less than $500. Now suppose that if you do get giardiasis, you incur $250 in medical costs and lost earnings. According to the damage cost method, avoiding the illness is worth $250.

Defensive behavior generally refers to actions that people take to reduce environmental damages. Defensive behavior includes actions that reduce exposure to pollution, as well as actions that mitigate adverse effects of exposure. The defensive behavior method simply assumes that a rational person will take defensive behaviors as long as the value of the damage avoided exceeds the cost of the defensive action. Thus, choices of defensive behaviors reveal something about the value of avoiding environmental damage.

Damage costs, on the other hand, refer to the real resource costs associated with pollution, including both direct and indirect costs. Direct costs are expenditures to treat illness or to repair, replace or maintain damaged materials; indirect costs reflect the opportunity costs of reduced productivity or output foregone because of environmental contamination. The damage cost method simply uses the reduction in real resource costs to measure the benefits of reduced pollution.

There are two main differences between the defensive behavior and damage cost methods. First, the defensive behavior method emphasizes how human behavior responds to changes in the environment and the impact of behavior on the outcomes experienced. The damage cost method, however, implicitly assumes that there is no behavioral response to environmental changes or that behavioral responses are not effective. Second, the defensive behavior method is designed to estimate or at least to bound a theoretically consistent measure of economic value like willingness-to-pay (WTP), while the damage cost method is not.

If damage costs do not measure WTP, why would you ever use them? Damage costs are almost surely less than WTP, and are often easier to estimate and easier to explain to non-economists. If the damage costs avoided by a regulation exceed the costs of implementing the regulation, you can be fairly confident that using WTP to measure benefits would only strengthen the conclusion that the regulation passes a benefit-cost test. If quickly determining a lower-bound estimate of benefits is more desirable than waiting for a better estimate, then using a damage cost approach may be warranted.

Despite their differences, the defensive behavior and damage cost methods share several common features. From a conceptual standpoint, the two methods can be developed and compared in the context of a simple household production model. More practically, the two methods are used to value the same kinds of endpoints: materials and, most frequently, health. Both methods are used to value changes in outcomes of pollution -- such as length of illness or soiling of household property -- as well as to value the underlying changes in pollution that cause changes in outcomes.

This chapter introduces the applications of the defensive behavior and damage cost methods of nonmarket valuation. Because these methods are mainly applied to health valuation (where damage costs are known as the costs of illness), the chapter reflects this emphasis. Section 2 of this chapter presents a simple household production model that includes both defensive behavior and damage costs. Section 3 describes how to conduct a defensive behavior study. Section 4 describes how to conduct a damage cost study, and conclusions follow in Section 5.

Now, if few people actually take actions to defend against damages from pollution, you might as well skip to Section 4. So let's take that question first: Does anyone defend? Prior research indicates that the estimated proportion of people who engage in defensive behavior varies widely with the duration of exposure and the cost of defensive action.

Smith, Desvousges, and Payne (1995) report that 15 percent of their sample took action to reduce home radon concentrations, when the actions were as costly as the purchase of a durable appliance. Akerman, Johnson, and Bergman (1991) report a radon mitigation rate that is three times larger, but their sample consists entirely of individuals who presumably were more concerned about radon than the average person. Berger et al. (1987) also examined purchases of durable goods and found that 11 to 15 percent of their sample bought air conditioners or air purifiers partly for health reasons.

At the other extreme, Harrington, Krupnick, and Spofford (1989) report that 98 percent of their sample engaged in defensive behavior during a giardiasis outbreak, compared to 76 percent of respondents in Abdalla's (1990) study of perchloroethylene contamination. Public notification occurred in both these incidents of water contamination, and relatively few respondents defended themselves by purchasing durable goods such as home filtration systems. Other studies generally report rates of defensive action higher than those involving durable goods purchases but lower than those following water contamination incidents. For example, Bresnahan, Dickie, and Gerking (1997) report that 65 percent of their sample changed behavior in some way on days of poor air quality.

In summary, the propensity to take defensive action appears the smallest when the purchase of a durable good is involved, and largest during temporary, but extended, periods of water contamination, particularly with public notification of health dangers. Although the extent of defensive action varies widely, significant fractions of people appear to take some action to avoid environmental contaminants. Applying defensive behavior to the estimation of values for non-market goods, however, has proven difficult and is the subject of the next two sections.

2. A MODEL OF DEFENSIVE BEHAVIOR AND DAMAGE COST

This section has three purposes. The first is to derive illustrative versions of each of the four equations used for non-market valuation in the defensive behavior framework. The four equations are expressions for: (1) the marginal value of reduced pollution (section 2.2); (2) the marginal value of reducing an adverse outcome of pollution (section 2.3); (3) the compensating variation for a non-marginal change in pollution (section 2.4); and (4) the change in defensive expenditures that bounds the compensating or equivalent variation for non-marginal change (section 2.5). These valuation equations have been developed and applied in separate papers, and this section clarifies the connections among them and provides a few additional equations to tie all four together. The second purpose of this section is to compare WTP in the presence of defensive behavior to damage costs (section 2.6). Two results are derived. (1) For a reduction in pollution, marginal WTP does not necessarily exceed marginal damage cost. But if the terms involved in the comparison take the signs suggested by empirical evidence, WTP will exceed damage cost. (2) For a reduction in an outcome affected by pollution, marginal WTP unambiguously exceeds marginal damage cost. The first of these conclusions is well known; the second is new to my knowledge. The third purpose is to discuss extensions and limitations of the defensive behavior model (section 2.7).

The defensive behavior and damage cost methods can be illustrated and compared using a simple household production model. Variations of this basic model have been used for many years in health economics (Grossman 1972) and environmental economics (Cropper 1981). The model developed by Harrington and Portney (1987) clearly illustrates how defensive behavior, damage costs, and welfare are related. A slightly modified version of the Harrington-Portney model is presented below. The model focuses on health damages of pollution but is easily modified to treat other damages (Bartik 1988; Courant and Porter 1981; Jakus 1994).

2.1 Theoretical Framework

In the Harrington-Portney model, an individual's utility is given by

$$U = U(X, L, S), \tag{1}$$

where X denotes consumption expenditures, L denotes leisure time, and S denotes time spent sick. The marginal utility of consumption and leisure are positive, but the marginal utility of sick time is negative. The distinguishing feature of a defensive behavior model is that an adverse outcome like illness does not just happen, but is influenced by behavior. Sick time is "produced" according to the health production function

$$S = S(E, G, Z), \tag{2}$$

where E represents exposure to pollution, G represents a mitigating activity, and Z represents a set of exogenous factors that affect length of illness, such as chronic health status and age. Exposure increases the length of illness while mitigation reduces length of illness. Exposure to pollution also is influenced by behavior, according to the production function

(3) $E = E(A, \alpha),$

where α represents the ambient level of pollution, and A represents an averting behavior. Ambient pollution increases exposure while averting behavior reduces exposure.

The model includes two types of defensive behaviors: averting behavior (A) that reduces exposure to pollution, and mitigating behavior (G) that reduces adverse effects of exposure. To illustrate the distinction, people with asthma may reduce exposure to ambient air pollution by spending less time outdoors when pollution concentrations are high (A), and/or they may use an inhaled bronchodilator (G) to reduce asthma symptoms. Averting behavior involves changing activities or buying goods to reduce exposure, while mitigating behavior often involves using medical care or medication.

But in this model, the distinction between averting and mitigating behavior is only expository. If you substitute equation (3) into equation (2), the health production function becomes

(4) $S = S\left[G, E(A, \alpha), Z\right] = \hat{S}(G, A, \alpha, Z),$

This equation shows that G and A are just two different ways of reducing illness for a given level of ambient pollution.

The budget constraint is

(5) $I + wT_w = X + p_g G + p_a A + M(S),$

where I represents non-labor income, T_w denotes time spent working at wage rate, w; p_g and p_a respectively denote unit prices of G and A; and $M(S)$ denotes remedial expenses, typically interpreted as medical expenses, as a function of sick time. The interpretation of this function is that increasing illness forces the individual to incur additional medical costs.

The remedial expenditure function is a useful modeling device but fits somewhat awkwardly in the model. The way the model is specified, the individual does not directly choose the amount remedial expenditure: once a given amount of illness occurs, medical costs are determined by the function $M(S)$. The individual indirectly influences medical costs by choosing G and A, but once S is determined, there is no further choice over medical costs. In fact, people who are sick do exercise some control over their medical expenses and assuming they do not is inconsistent with the model's emphasis on individual decision-making to control environmental damages. Nevertheless, I include the remedial expenditure function in the illustrative model because several authors (including Harrington and Portney) employ this device to represent a monetary cost of pollution damage, and because the function simplifies the comparison of WTP and damage costs.

One more constraint is needed to complete the model. Total time available (T) is allocated to work, leisure, and illness: $T = T_w + L + S$. If you solve this time constraint for T_w and substitute it into the budget constraint, you get the full income budget constraint

$$(6) \qquad I + wT = X + wL + p_g G + p_a A + M(S) + wS.$$

This equation shows how the value of the individual's resource endowment ($I + wT$) is spent on consumption of goods and leisure ($X + wL$), defensive expenditures ($p_g G + p_a A$), and damage costs ($M(S) + wS$). Damage costs in the health context are the individual's costs of illness. These consist of direct costs of remedial medical care and indirect costs of time lost due to illness:

$$C(S) = M(S) + wS.$$

If you want to estimate marginal WTP or to use a consumer surplus approximation to the compensating variation for a change in pollution, the most direct way is to solve the utility maximization problem after substituting equation (4) into the utility function (1) and the full-income budget constraint (6). (See Cropper (1981) or Gerking and Stanley (1986) for examples with empirical implementation.) I'm going to take a more circuitous route to bring in the defensive expenditure function used to bound the value of a non-marginal change. This involves solving the utility maximization problem in two stages (Smith 1991). In the first stage, the individual chooses defensive behaviors to minimize defensive expenditures ($p_g G + p_a A$), subject to producing a given amount of sick time according to equation (4). In the second stage, the individual chooses the amount of sick time along with amounts of consumption and leisure to maximize utility.

The first-order conditions for the cost-minimization problem are

$$\begin{aligned} &P_g - \pi(\partial S / \partial G) = 0 \\ (7) \qquad &P_a - \pi(\partial S / \partial E)(\partial E / \partial A) = 0 \\ &S^0 - S\left[G, E(A, a), Z\right] = 0 \end{aligned}$$

In these equations, S^0 denotes the given amount of sick time, and π denotes the Lagrangean multiplier. Assuming second-order sufficient conditions hold, you can solve the first-order conditions to obtain the values of G, A, and π that minimize the defensive expenditures necessary to produce sick time S^0, as functions of the exogenous variables (p_g, p_a, S^0, α, and Z). These functions are

$$\begin{aligned} &G^0 = G(P_g, P_a, S^0, \alpha, Z) \\ (8) \qquad &A^0 = A(p_g, p_a, S^0, \alpha, Z) \\ &\pi^0 = \pi(p_g, p_a, S^0, \alpha, Z) \end{aligned}$$

With these functions you can define the *defensive expenditure function* (Bartik 1988) that gives the minimum cost of producing the given amount of sick time S^0 when prices of defensive behaviors are p_g and p_a, and ambient pollution is α:

$$(9) \qquad D(p_g, p_a, S^0, \alpha, Z) = p_g G^0 + p_a A^0.$$

The minimum defensive expenditure is a function of prices, sick time, and ambient pollution because G^0 and A^0 are functions of these variables. The properties of this function are useful in interpreting the defensive behavior model.

According to the envelope theorem, the marginal cost of sick time equals the Lagrangean multiplier, π. Using the first-order conditions (7), this equals

$$(10) \qquad \partial D / \partial S = \pi = \frac{p_g}{(\partial S / \partial G)} = \frac{p_a}{(\partial S / \partial E)(\partial E / \partial A)} < 0.$$

The first equality in (10) follows from the envelope theorem while the second two are implied by the first-order conditions and reflect the general result from production theory that marginal cost equals the ratio of an input price to its marginal product. The inequality indicates that defensive expenditures must rise if sick time is to be reduced.

The marginal effect of ambient pollution on defensive expenditures also can be obtained using the envelope theorem:

$$(11)\qquad \partial D/\partial\alpha = -(\partial D/\partial S)(\partial S/\partial E)(\partial E/\partial A)$$
$$= -p_g \frac{(\partial S/\partial E)(\partial E/\partial\alpha)}{(\partial S/\partial G)} = -p_a \frac{(\partial E/\partial a)}{(\partial E/\partial A)} > 0.$$

Expression (11) says that the marginal effect of ambient pollution on defensive expenditure equals the price of a defensive behavior times the marginal rate of technical substitution between the defensive behavior and ambient pollution. Of course, the marginal effect of pollution on defensive expenditure is positive because higher levels of pollution require greater defensive expenditures to hold the amount of sick time constant.

To solve the second stage of the utility maximization problem, substitute the defensive expenditure function into the full income budget constraint to rewrite equation (6) as

$$I + wT = X + wL + D(p_g, p_a, S^0, \alpha) + M(S) + wS.$$

The first-order condition involving sick time is

$$(12)\qquad (\partial U/\partial S) - \lambda[(\partial D/\partial S) + (dM/dS) + w] = 0,$$

where λ is the Lagrangean multiplier and is interpreted as the marginal utility of full income. Now re-arrange this to read

$$(13)\qquad -(\partial U/\partial S)/\lambda + (dM/dS) + w = -(\partial D/\partial S).$$

The left-hand side of this equation gives the marginal benefit of reducing sick time as the sum of three components: the consumption value of health (the monetized marginal utility($-(\partial U/\partial S)/\lambda$), the reduction in medical expenditures resulting from reduced sick time (dM/dS), and the marginal value of the increased time made available for leisure and work (w). Utility maximization requires that the marginal benefit of reduced sick time equals the marginal cost of home production of the reduction in sick time given by $-\partial D/\partial S$.

2.2 Valuing Marginal Changes in Pollution

Let $V(I, w, p_g, p_a, \alpha)$ denote the indirect utility function (see Chapter 2). According to the envelope theorem, the marginal WTP for reduced pollution is

$$(14) \qquad -(\partial V / \partial \alpha) / \lambda = \partial D / \partial \alpha.$$

At the margin, the WTP for reduced pollution equals the savings in defensive expenditure that just offsets by the change in pollution, holding sick time constant.

This equation is rewritten two ways to better understand how WTP is measured in a defensive behavior model. Using equations (11) and (12), marginal WTP equals

$$(15) \qquad \begin{aligned} &-(\partial V / \partial \alpha) / \lambda \\ &= [-(\partial U / \partial S) / \lambda + dM / dS + w](\partial S / \partial E)(\partial E / \partial \alpha). \end{aligned}$$

Thus, the value of reduced pollution equals the value of reduced sick time, weighted by the effect of pollution on sick time. This form of the WTP expression is not convenient for estimation, however, because of the presence of unobservable marginal utility terms. To obtain a more empirically tractable form of the WTP expression, use equation (11) again to get

$$(16) \qquad \begin{aligned} &-(\partial V / \partial \alpha) / \lambda \\ &= -p_g \frac{(\partial S / \partial E)(\partial E / \partial \alpha)}{\partial S / \partial G} = -p_a \frac{\partial E / \partial \alpha}{\partial E / \partial A}. \end{aligned}$$

This is the same result as equation (23) in Chapter 2, where WTP is given by the marginal rate of technical substitution between defensive behavior and pollution. Equation (16) is convenient for estimation because marginal utility terms have been eliminated; therefore, empirical researchers often estimate an equation like (16) to compute WTP.

Apart from being convenient for estimation purposes, equation (16) illustrates two important points about the defensive behavior model. First, the individual is not willing to pay much to reduce pollution when defensive behavior is cheap (small p_g or p_a) and effective (large $\partial S/\partial G$ or $\partial E/\partial A$), or when pollution has little impact on health (small $\partial S/\partial E$ or $\partial E/\partial \alpha$). Conversely, the individual will pay a lot to reduce pollution if defensive behavior is expensive and ineffective, or if pollution has a large impact on health. Thus, WTP is driven by the individual's opportunities to substitute defensive behavior for reduced pollution. Second, if two defensive behaviors exists, WTP must be the same no matter which defensive behavior (G or A in this model) is used to compute the marginal rate of technical substitution. If this condition does not hold, the individual is not behaving as predicted by the model, and the whole approach is called into question.

2.3 Valuing Marginal Changes in Outcomes

Estimating the marginal WTP for reduced pollution along the lines of equation (14) does not fit very well with the way benefit-cost analyses are conducted for pollution regulations. Economists usually are not asked to value a change in pollution directly, but instead to value a given change in health status or a change in health risk. The health change to be valued and its link to pollution are estimated from research in epidemiology or toxicology. For example, health scientists estimate how pollution affects days of symptoms or risk of death. Economists estimate a value for reducing days of symptoms or risk of death. Policy analysts put the two estimates together to estimate the value of reduced pollution. This division of labor takes advantage of the specialized expertise of health scientists and economists, but at a cost. Health scientists commonly ignore defensive behavior and thus may produce biased estimators of the health effects of pollution.

In any event, the task now is to estimate the marginal value of a given change in illness dS. This is easy because the first-order condition for sick-time in the utility-maximization problem says that the individual chooses sick time to equate the marginal benefit of reducing illness to the marginal cost of reducing illness. Thus using (11) and (12), the marginal value of a given change in illness dS is

$$(17) \quad [-(\partial U / \partial S) / \lambda + (dM / dS) + w]dS = -(\partial D / \partial S)dS$$
$$= -[p_g / (\partial S / \partial G)]dS = -[P_a / [(\partial S / \partial E)(\partial E / \partial A)]]ds.$$

Equation (17) gives the value of a change in outcome, while equation (16) gives the value of the underlying change in the environment that causes a change in outcome.

2.4 Valuing Non-Marginal Changes in Pollution

Estimating the value of non-marginal changes in environmental quality in the defensive behavior framework is more difficult than estimating marginal WTP. Bockstael and McConnell (1983) developed an approach for using the area behind the Hicksian (compensated) demand curve for a defensive input to measure the compensating variation for a change in environmental quality.

The Bockstael-McConnell model is a general household production model where the outputs would not necessarily involve health, and the inputs would not necessarily be defensive behaviors. Thus an individual might use fishing gear, time, and water quality in a public lake as inputs to "produce" fishing trips. In their model α is a good rather than a bad (for example, water quality rather than water pollution). Also, they allow α to enter the utility function directly (as opposed to being only an input into a household production function), and their model allows for multiple home-produced goods.

The Bockstael-McConnell approach relies on weak complementarity (Chapter 2). Specifically, the public good α must be weakly complementary to a private input (i.e., a defensive behavior in the present context) in the household production functions, so that the individual is indifferent to changes in α when the private input is not purchased. Bockstael and McConnell then show that two conditions are sufficient for weak complementarity to hold.

1. The public good α is weakly complementary to a subset of the home-produced outputs, so that $\partial U/\partial \alpha = 0$ when all of the outputs in the subset are zero. If α does not enter the utility function directly, as in the illustrative model used here, this condition holds trivially.
2. The private input is essential to the production of all of the home produced outputs in the subset. The essential (or necessary) input condition means that the output cannot be produced if the input is not used. Under these conditions, the change in area behind the compensated demand curve for the private input when environmental quality changes measures the compensating variation for the change in environmental quality.

The compensating variation may be approximated using the Marshallian (ordinary) demand curve for a defensive input. Suppose α is environmental quality and G is a necessary input: it is impossible to reduce sick time if G is not used. We need the utility-maximizing demand function for G. Equation (7) gives the defensive-expenditure-minimizing demand for G, at an arbitrary value of sick time, S^0. To convert this to the utility-maximizing demand, replace the arbitrary S^0 with the utility-maximizing choice of sick time, say S^*. If the second-order conditions for the utility-maximization problem hold, S^* is a function of the exogenous variables I, w, p_g, p_a, α, and Z and may be written as

$$(18) \qquad S^* = S^*(I, w, p_g, p_a, \alpha, Z).$$

Substituting this function into equation (8) yields the utility-maximizing demand for G:

$$(19) \qquad \begin{aligned} G^* &= G[p_g, p_a, S^*(I, w, p_g, p_a, \alpha, Z), \alpha, Z] \\ &= G^*(I, w, p_g, p_a, \alpha, Z). \end{aligned}$$

In general, the utility-maximizing demand for a defensive input is a function of all of the exogenous variables in the model. Of course, a similar function could be defined for the utility-maximizing choice of A, A^*. Now let p_g^0 be the current price of G and p_g^c be the choke price (the price that drives the quantity demanded to zero). Then the compensating variation for a change in α from α^0 to α^1 is approximated by the change in the Marshallian area:

$$(20) \qquad \int_{p_g^0}^{p_g^c} G(I, w, p_g, p_a, \alpha^1, Z)dp_g - \int_{p_g^0}^{p_g^c} G(I, w, p_g, p_a, \alpha^0, Z)dp_g$$

An important advantage of the Bockstael-McConnell approach is the ability to estimate welfare changes using the demand for only one (necessary) input, as opposed to estimating all the household production functions. This approach reduces data requirements and simplifies estimation procedures. As discussed more fully in section 3, the demand for an input is a reduced-form equation, while the household-production functions are structural forms that should be estimated by simultaneous equations methods. The simplification comes at the expense of assuming the input is necessary, an assumption that may be difficult to justify in many circumstances (Smith 1991). Also, the upper limit of the integration required to find the area behind the input demand curve is the choke price, a point on the demand curve where confidence intervals may be relatively wide (Just, Hueth, and Schmitz 1982).

Applications of the Bockstael and McConnell approach in the defensive behavior framework include Dickie and Gerking (1991b) and Agee and Crocker (1996). Dickie and Gerking estimate benefits of ozone control assuming that medical care is a necessary input in health production, while Agee and Crocker (1996) estimate benefits of reduced lead burdens in children's bodies using chelation as a necessary input.

2.5 Bounding Values of Non-Marginal Changes in Pollution

An alternative approach to nonmarginal welfare measurement in the defensive behavior framework is to use the change in defensive expenditures as a bound on the compensating or equivalent variation. Bartik (1988) showed that the compensating variation for a pollution reduction is bounded from below by the savings in defensive expenditures necessary to maintain the initial level of the home-produced output. Using the defensive expenditure function defined in equation (9) and considering a reduction in pollution from α^0 to α^1, the lower bound on the compensating variation is

$$(21) \qquad D(p_g, p_a, S^0, \alpha^0, Z) - D(p_g, p_a, S^0, \alpha^1, Z).$$

Notice that the change in defensive expenditure is computed holding the amount of sick time constant. Bartik's bounding result does not necessarily apply to the *observed* change in defensive expenditures, because the individual presumably will adjust sick time in response to the change in pollution according to equation (18) (Courant and Porter 1981). But if the individual consumes less sick time as pollution falls, the observed change in defensive expenditures is a lower bound on compensating variation, albeit a less accurate bound than the change in expenditures holding output constant. As a practical matter, observed defensive expenditures are much easier to calculate because they do not require estimation of the defensive expenditure function.

Several researchers have applied Bartik's analysis to estimate a bound on WTP. Most applications use observed defensive expenditures and involve avoidance of contaminated water supplies by some combination of purchasing bottled water, boiling or treating water at home, hauling water from an alternate source, or installing water filtration systems (Abdalla 1990, Abdalla, Roach, and Epp 1992, Harrington, Krupnick, and Spofford 1989, Laughland et al. 1996). Also, Murdoch and Thayer (1990) estimated the change in expenditures on sun screen lotion that would hold incidence of non-melanoma skin cancer constant despite a change in ultraviolet radiation exposure due to ozone depletion.

2.6 Comparing WTP and Damage Costs

The Harrington-Portney model illustrates the difference between WTP and damage costs. As indicated previously, damage costs in the health context are defined as the cost of illness. The individual's total cost of illness is $C(S) = M(S) + wS$, and the marginal impact of a given change in illness on cost is

(22) $$dC = (dM / dS + w)dS.$$

Now compare marginal WTP in equation (15) to the marginal damage cost in equation (22). For a given reduction in illness of $dS < 0$, WTP exceeds the absolute value of the reduction in illness costs by the amount $(\partial U / \partial S) / \lambda > 0$.

Thus, WTP *for a given reduction in illness* unambiguously exceeds the cost of illness, because the cost of illness does not account for the utility value of health or for pain and suffering. More generally, damage costs fall short of WTP because they focus on resource costs associated with changes in stocks or flows of materials or health while ignoring the utility value of the change in service flows.

The divergence between WTP and the cost of illness for *a change in pollution* is less clear-cut. In addition to ignoring the utility value of health, the cost-of-illness method relies on an inappropriate estimate of the effect of pollution on health.

To see this, consider how pollution affects health. The whole point of the defensive behavior model is that the individual adjusts behavior in response to changes in pollution. These behavioral adjustments then affect health. Thus the overall effect of pollution on health includes a direct effect that would occur if behavior did not change

(23) $$(\partial S / \partial E)(\partial E / \partial \alpha),$$

plus an indirect effect operating through the change in behavior

$$(24) \quad (\partial S / \partial G)(\partial G^* / \partial \alpha) + (\partial S / \partial E)(\partial E / \partial A)(\partial A^{**} / \partial \alpha).$$

where G^* and A^* are the utility-maximizing choices of mitigating and averting behavior. The change in sick time that actually occurs when pollution changes is the sum of the direct and indirect effects:

$$(25) \quad dS / d\alpha = (\partial S / \partial E)(\partial E / \partial \alpha) + (\partial S / \partial G)(\partial G^* / \partial \alpha) + (\partial S / \partial E)(\partial E / \partial A)(\partial A^* / \partial \alpha).$$

By assumption, the direct effect is positive (pollution increases sick time), but the model does not restrict the sign of the indirect effect. However, available empirical evidence indicates that defensive behavior increases with pollution, implying that the indirect effect is negative. Thus, the total effect of pollution on health in equation (25) is less than the direct effect in equation (23). Increases in pollution lead to increased defensive behavior, which partly offsets the adverse health effects of the additional pollution.

Now if you look at the WTP expression in equation (16), you'll see that the correct measure of the effect of pollution on health to use in estimating benefits is the direct effect in equation (23). The terms in equation (23) are partial derivatives; by definition, they are computed while holding other variables, in particular the defensive behaviors, constant. To estimate this direct effect, you need an empirical model that accounts for the effect of behavior on health. With this model, you can counterfactually hold behavior constant to estimate the direct effect of pollution on health for use in the WTP expression. If your model does not account for defensive behavior, then the effects of behavior get mixed in with the direct effect of pollution and you end up estimating the health effect of pollution with equation (25) instead of (23). That would cause you to understate benefits because, based on empirical evidence, the expression in equation (25) is less than the direct effect in equation (23).

That is exactly what happens when the cost of illness method is applied to value changes in pollution. The cost of illness is in effect computed as

$$(26) \quad dC = [dM / dS + w] \times [(\partial S / \partial E)(\partial E / \partial \alpha) + (\partial S / \partial G)(\partial G^* / \partial \alpha) + (\partial E / \partial E)(\partial E / \partial A)(\partial A^* / \partial \alpha)]d\alpha.$$

Subtracting this expression from the willingness-to-pay in equation (14) yields:

$$(27)\quad -(1/\lambda)(\partial U/\partial S)(\partial S/\partial E)(\partial E/\partial \alpha)$$
$$-[(dM/dS)+w][(\partial S/\partial G)(\partial G^{*}/\partial \alpha)$$
$$+(\partial S/\partial E)(\partial E/\partial A)(\partial A^{*}/\partial \alpha)]d\alpha.$$

The difference between the cost of illness and WTP for a reduction in pollution includes the pain and suffering of illness, as well as the impact of the change in defensive behavior. As discussed previously, the behavioral effect cannot be signed theoretically but would appear to be positive based on empirical evidence. Thus, the model suggests that *for a given change in ambient pollution,* WTP probably exceeds the cost of illness because the latter measure neglects both the pain and suffering of illness and the behavioral adjustments people make to reduce harmful effects of pollution. Although the cost of illness is not a utility-constant welfare measure, it is expected to be a lower bound on WTP.

This conclusion rests on empirical evidence concerning how defensive behavior responds to changes in environmental threats to health, so empirical evidence is briefly reviewed. Relatively few studies link protective action to measured concentrations of environmental contaminants, but those that do consistently find that protective action increases with ambient risk.

Akerman, Johnson, and Bergman (1991), Doyle et al. (1991), and Smith, Desvousges, and Payne (1995) all show that the probability of mitigation increases with measured radon concentrations in the home. Dickie and Gerking (1991b) treat medical care as a mitigating input in a health-production function and find that higher ozone concentrations increase doctor visits. Krupnick, Harrington, and Ostro (1990) do not measure averting behavior directly, but present evidence suggesting that persons experiencing acute symptoms attempt to reduce ozone exposure on subsequent days. Bresnahan, Dickie, and Gerking (1997) report that persons who tend to experience symptoms in smoggy conditions reduce time outdoors as ambient ozone concentrations rise above the federal standard. In summary, available empirical evidence consistently indicates that defensive behavior increases with pollution.

2.7 Extensions and Limitations of the Model

2.7.1 Joint Production

The problem of joint production is the major obstacle to using the defensive behavior method to estimate benefits. Joint production occurs when the effect of a defensive behavior on individual welfare does not operate through a single outcome such as length of illness. Defensive behavior might affect more than one health outcome, or it might enter the utility function directly.

Joint production is clearly pervasive (Pollak and Wachter 1975). Substituting bottled water for tapwater affects the taste and odor of drinking water as well as exposure to contaminants. Purchases of safety goods, such as smoke detectors and bicycle helmets, reduce risk of injury as well as risk of death. Staying indoors to avoid air pollution may reduce incidence of several symptoms, while using sun screen lotion reduces risk of sunburn as well as risk of skin cancer.

As discussed by Dickie and Gerking (1991a), joint production presents problems for using an equation such as (14) to estimate benefits. One case considered by Dickie and Gerking involves m defensive behaviors and n health outcomes. If the health production functions are independent and there are at least as many defensive goods as health outcomes ($m \geq n$), then data on prices and marginal products of defensive goods are sufficient to estimate marginal WTP using a simple generalization of equation (14) in which marginal WTP remains equal to the marginal defensive expenditure. On the other hand, if there are fewer defensive goods than health outcomes ($m < n$) or the health production functions are not independent, then the marginal WTP does not equal marginal defensive expenditure, utility ratios cannot be eliminated from the WTP expression, and prices together with marginal products of defensive goods do not encode sufficient estimation to capture WTP. For an early treatment of these issues, see Hori (1975).

In another form of joint production, defensive behaviors enter the utility function directly. This setup is likely appropriate for many applied problems, but unfortunately it hinders application of the defensive behavior method. To illustrate, suppose that both A and G enter the utility function, apart from their influence on exposure and sick time. In this situation, simplifying the expression for WTP in equation (15) to the empirically tractable form in equation (16) is no longer possible. Instead,

$$
\begin{aligned}
(28)\quad & -(\partial V / \partial a) / \lambda \\
& = [-(\partial U / \partial S) / \lambda + dM / dS + w](\partial S / \partial E)(\partial E / \partial \alpha) \\
& = \left[\frac{-p_g + (\partial U / \partial G) / \lambda}{(\partial S / \partial G)}\right](\partial S / \partial E)(\partial E / \partial \alpha) \\
& = \left[\frac{-p_a + (\partial U / \partial A) / \lambda}{(\partial S / \partial E)(\partial E / \partial A)}\right](\partial S / \partial E)(\partial E / \partial \alpha).
\end{aligned}
$$

In this situation, the marginal utilities of the defensive behaviors cannot be eliminated from the WTP expression. Thus, using prices and marginal products for defensive behaviors alone to estimate marginal WTP is impossible.

The situation is somewhat similar if the nonmarket input α indivisibly enters multiple household-production functions, another form of joint production, or enters the utility function directly, a situation sometimes called unavoidable effects. For example, ambient air pollution affects health through the health-production function, but the individual also may have preferences over visibility or the aesthetic quality of the environment. If α enters the utility function of the illustrative model, marginal WTP equals

$$
\begin{aligned}
(29)\quad & -(\partial V / \partial \alpha) / \lambda \\
& = [-(\partial U / \partial S) / \lambda + dM / dS + w](\partial S / \partial E)(\partial E / \partial \alpha) - (\partial U / \partial \alpha) / \lambda \\
& = \partial D / \partial a - (\partial U / \partial a) / \lambda.
\end{aligned}
$$

The marginal defensive expenditure does not incorporate the direct utility effect of reduced pollution and, therefore, provides a lower bound but not an exact measure of the total value of the change in pollution.

2.7.2 Risk

In the illustrative model, ambient pollution and defensive behavior influence the health outcome of sick time with certainty. Yet, environmental and behavioral factors may often be better viewed as influencing the probability or severity of uncertain health outcomes. In this context, it is useful to distinguish discrete and continuous representations of risk according to whether the possible outcomes are countable or uncountable. Results for the discrete case are more important from a practical perspective because estimates of nonmarket values under risk, whether involving defensive behavior or not, almost always assume a small number of discrete states of the world. For example, the most common approach is to assume two possible states of the world: survival and death, or good health and poor health. Pollution decreases the probability of the better health state, while defensive behavior increases it.

Bresnahan and Dickie (1995) show that key results derived under certainty, including Bartik's bounding results and Hori's joint production conditions, extend to discrete state representations of risk with minor modification. Additionally, Freeman's (1991) analysis indicates that these results do not require that preferences accord with expected utility.

In a more general model including a continuum of possible outcomes, Shogren and Crocker (1991) show that expressing marginal WTP independently of preference function parameters is generally not possible, and that defensive expenditures may not bound WTP. However, Quiggin (1993) provides a separability assumption that allows recovery of WTP independent of preferences even in this more general setting.

2.7.3 Dynamic Models, Household Models, and Time Costs

Most empirical applications of the defensive behavior method have employed single-period models similar to the illustrative model presented above, but the method can be applied in dynamic settings (Cropper 1981). Similarly, most applications have applied models of individual behavior, but the method can be applied to household or family settings where defensive behaviors taken by one person, typically a parent, affect the health of another family member such as a child (Agee and Crocker 1996; Dickie 2000). The model also generalizes to account for defensive behaviors requiring time, where the full cost of averting behaviors includes a time cost.

3. CONDUCTING A DEFENSIVE BEHAVIOR STUDY

Table 1 presents a stylized sequence of steps for conducting a defensive behavior study. Of course, every application is different, but the table lists many of the key decisions and actions.

Table 1. Steps in Conducting a Defensive Behavior Study

1. Identify the change to be valued.
   - 1a. Is it a change in outcome, risk, or environment?
   - 1b. Is the change a marginal or a non-marginal change?
2. Describe how defensive behavior affects welfare.
   - 2a. List the defensive behaviors available and the outcomes or risks affected. Describe relationships between behaviors and outcomes.
   - 2b. What is the nature of joint production?
   - 2c. Who makes decisions about defensive behaviors, and who is affected?
3. Decide which defensive behavior approach to use:
   - 3a. Consumer-market study,
   - 3b. Health production function study,
   - 3c. Demand for a necessary defensive input study, or
   - 3d. Defensive expenditure study.
4. Collect data to implement your chosen approach.
5. Estimate your model and its welfare expression.

3.1 Identifying the Change to be Valued

Begin by identifying the change to be valued in terms of the baseline and alternative conditions. Four separate types of values can be computed using the defensive behavior method, and the data requirements and estimation methods differ somewhat among them. Decide whether to value a marginal change in an outcome or risk affected by pollution, a marginal change in pollution, or a non-marginal change in pollution. When valuing a non-marginal change, decide if the aim is to estimate an exact compensating or equivalent variation, or to estimate a bound.

3.2 Describing How Defensive Action Affects Welfare

3.2.1 Defensive Behaviors and Outcomes

Another early step is to describe how defensive behavior affects welfare. Begin by listing the relevant outcomes and defensive behaviors. Most applications fall into one of three categories. One set up frequently used is to consider medical care as a good that mitigates adverse health effects of air pollution exposure. The health effect considered may be risk of infant mortality (Joyce, Grossman, and Goldman 1989), presence of chronic illness (Gerking and Stanley 1986), days of acute illness (Cropper 1981; Dickie 2000), or other effects. A second situation considered is the purchase or use of safety equipment to reduce risk of death or injury in an automobile accident (Carlin and Sandy 1991), bicycle accident (Jenkins, Owens, and Wiggins 2001), or fire (Dardis 1980). A third setting is the use of bottled water, home filtration, boiling of tap water, hauling water from an alternative source, or substituting other beverages as averting behaviors to reduce exposure to contaminated water supplies (Abdalla 1990; Abdalla, Roach, and Epp 1992; Harrington et al. 1986; Laughland et al. 1994). While these examples encompass most applications, the specific valuation problem at hand may lead to other definitions of outcomes and defensive behaviors, such as chelation to reduce body lead burdens (Agee and Crocker 1996) or use of sun screen lotion to reduce risk of skin cancer (Murdoch and Thayer 1990).

It is important to identify the full set of possible defensive behaviors, including both mitigating and averting behaviors. Allowance should be made for possible substitutions between behaviors. For example, multiple defensive behaviors may be available, and people may employ more than one or may choose only one from the set of available options. Obviously the full defensive expenditure would not be captured if some defensive behaviors were not accounted for (e.g., use of home filtration counted but purchases of bottled water ignored).

More subtly, failing to account for all available defensive behaviors may bias estimators of the effectiveness of the actions you do consider. For example, people may use sun screen lotion and then spend more time outdoors in direct sunlight, or use child safety seats in automobiles and then drive less carefully. This kind of offsetting behavior needs to be addressed as it increases risk.

In thinking about substitution opportunities, be aware of related behaviors that you may not consider defensive behaviors. In using prenatal care to value infant health, for example, you probably need to account for maternal consumption of alcohol and tobacco during pregnancy. We may not think of expectant mothers as literally substituting between prenatal care and smoking, but the fact is that both are choices the mother makes that influence health risks to the fetus. Better control for these related behaviors results in more accurate estimates (Grossman and Joyce 1990).

3.2.2 Joint Production

When describing how defensive behavior affects welfare, the issue of joint production naturally arises. Describe the nature of joint production and consider ways of handling the problem. Intuitively, a good way to describe the nature of the problem is to decide whether defensive behaviors indivisibly affect more than one home-produced output or whether defensive behaviors enter the utility function directly. For example, the taste attributes of bottled or home-filtered water relative to tap water are most naturally viewed as arguments of the utility function. Alternately, the joint effects of prenatal care on the health of the fetus and the mother would be a case where a defensive input indivisibly enters two health production functions.

One possible solution for joint production is suggested by Hori's (1975) analysis. If the number of defensive behaviors that do not enter the utility function is at least as great as the number of health outputs, the health-production functions are independent, and the environmental input does not enter the utility function directly, you can estimate a generalized form of equation (16) or (17). See Dickie and Gerking (1991a) for an empirical example.

The Hori solution often will not be applicable, however, because of too few defensive behaviors or because the environmental input enters the utility function. In this case, a standard defensive behavior welfare measure like equation (16) or (17) can be used, but should be interpreted as a bound on WTP using equation (28) or (29). As long as a sign can be asserted for the marginal utility of defensive behavior (equation (28)) or for the environmental input (equation (29)), you know whether the standard welfare measure is too high or too low. If the environmental input is a measure of pollution (a bad rather than a good) and enters the utility function directly, then equation (16) understates benefits of a pollution reduction. If the defensive input is a source of positive utility (e.g., bottled water tastes better than tap water), then equation (16) or (17) overstates benefits. Conversely, if the defensive input is a source of disutility, equation (16) or (17) understates benefits.

A third solution to the joint-production problem is to impose additional structure on the model to support WTP estimation. An example of this approach applies when the environmental input enters the utility function. As discussed in section 2.4, assuming weak complementarity allows application of the Bockstael-McConnell approach and estimation of the compensating variation.

Blomquist (1979) employs a fourth approach of estimating the value of a risk change under the assumption that the joint product has no value. The estimated value of the risk change then is used to recover the implied value of the joint product in equilibrium. A fifth alternative is to simulate WTP under alternative assumptions about the relative marginal benefits of joint products. For example, Dardis (1980) recognized that smoke detectors reduce risk of injury as well as risk of death. She estimated the value of a statistical life under varying assumptions about the relative contribution of injury and fatality avoidance to the benefit measure. You could augment this approach when collecting primary data by asking survey respondents about the relative importance of the joint products.

A sixth solution, which in my opinion would often be the best, is to abandon the use of actual defensive behavior to estimate WTP. Instead, use a stated-preference approach incorporating a contingent defensive behavior, and design the attributes of the contingent defensive behavior to overcome the joint production problem (Dickie and Gerking 1996).

In summary, joint production is a pervasive problem but there are many ways of addressing it. If all else fails, acknowledge the joint production problem and dismiss it. Doing so is a common practice. For example, Dickie and Gerking (1991a) assume that air conditioners are averting goods purchased to reduce exposure to ambient air pollution rather than to provide comfortable temperatures. Agee and Crocker (1996) assume that chelation therapy affects utility only by reduced lead-related health risks, when in fact chelation is a dangerous and painful procedure. Studies of avoidance of water contamination typically do not account for taste differences between bottled water (the avoidance good) and tap water.

3.2.3 Decisionmakers and Affected Persons

A final element of describing how defensive behavior affects welfare is to consider who is making the defensive decisions and whose welfare is affected. The simplest case is an individual taking actions that only affect his or her welfare. In that case only the individual WTP is estimated. But defensive behaviors (e.g., purchasing a water-filtration system, using safe food-preparation practices, or buying a safer car) often affects more than one member of a household. In these cases, household WTP is estimated. A third situation arises when a parent makes defensive decisions on behalf of a child, and the resulting welfare measure is the parent's WTP. The latter two cases call for a model of a household, as opposed to an individual decision maker. The most common approach is the consensus framework developed by Becker (1991). But a discussion of this and similar models is beyond the scope of this chapter; see the Introduction of Behrman, Pollak, and Taubman (1995) for a good overview and references, and Agee and Crocker (1996, 2001), or Dickie (2000) for recent defensive behavior applications.

3.3 Choosing a Defensive Behavior Approach

Before collecting data, choose the defensive behavior study approach. There are four main types. The choice from among them depends partly on the value you decided to estimate in Step 1. Two of the approaches produce a marginal value, while the other two are used to estimate a non-marginal value.

The first approach to marginal valuation in the defensive behavior framework is referred to here as the consumer-market study. This approach involves computing a welfare measure (e.g., equation (17)) using secondary data on the price of a defensive action and its effectiveness in reducing an adverse outcome or risk. An early example is Dardis' (1980) estimate of the value of reducing risk of death in a home fire through the use of smoke detectors. Most subsequent applications also consider purchasing or using a good to reduce risk of death from some type of accident. A recent, well executed example is the study of bicycle safety helmets by Jenkins et al. (2001).

The second approach to conducting a defensive behavior study is to estimate the health-production function, like equation (4), to compute the marginal products needed in equations like (16) or (17) for marginal valuation. Specify the health output on the left-hand side of the production function. This is simply the health outcome you chose to focus on in step 2. Then specify the inputs on the right-hand side of the production function, distinguishing between endogenous and predetermined inputs. Endogenous inputs are behaviors chosen by decision-maker. These are the defensive behaviors (A and G in the illustrative model) and related behaviors you described in Step 2. Predetermined or exogenous inputs are factors that directly affect health, but which are not chosen by the decision-maker, or reflect outcomes of past choices but now can be viewed as fixed from the perspective of the decision maker (Z and α in the illustrative model).

The third approach is to estimate the demand function for a defensive action in order to estimate the compensating or equivalent variation of a non-marginal change in the environmental input (using an equation like (20) for the compensating variation). The fourth approach is to estimate the change in defensive expenditures to bound the value of a non-marginal change, using equation (21), or using the observed change in defensive expenditures. Details on implementing these approaches are given below.

3.4 Collecting Data to Implement the Chosen Approach

3.4.1 Data for Estimating a Health Production Function

The health-production approach has the most stringent data requirements of the four defensive behavior methods. Data needed to estimate demand for a defensive input or defensive expenditures generally are a subset of the data needed to estimate a health-production function, while data requirements for the consumer-market study are the least stringent of all. So the health-production function approach is discussed first and the consumer-market approach last.

The variables involved in the health-production function were specified in step 3: the health output, and endogenous and exogenous inputs. Estimating a health-production function in step 5 requires data on the determinants of defensive decisions. As discussed in connection with equation (19), utility-maximizing choices of defensive behaviors depend on *all* exogenous variables in the model: exogenous inputs in the health production function, income, wages, and prices.

The appropriate set of exogenous inputs to consider will vary somewhat depending on the specific health effect on the left-hand side of the equation. The literature covers at least three types of important exogenous inputs. The first is the individual's pre-existing "stock of health capital." This is an abstraction that can't be measured precisely, but indicators for the individual's overall health status can be used. For example, if the health output is some measure of respiratory symptoms of air pollution, a relevant measure of health capital would be the presence of chronic respiratory diseases like asthma (Dickie and Gerking 1991a). Other researchers have included indicators for health conditions that limit activities or for past experience with the health effect in question. If there are measurable genetic factors that affect the health outcome, include them. For example, Dickie and Gerking (1996, 1997, 2001) treat skin type and complexion as predetermined inputs in producing skin cancer risk.

A second type of predetermined input includes other personal characteristics that affect health or the efficiency of health production, such as age, sex, race, and education. These factors are clearly related to the health outcomes people experience, and the usual approach is to treat them as inputs in the health-production function.

The third set of exogenous inputs is the environmental inputs. These can sometimes be ignored if the aim is to estimate the marginal value of an outcome using equation (17). Otherwise, measures of these inputs, such as the ambient concentrations of air pollutants, are needed. Important consideration include the location of pollution monitors relative to decision-makers' residences or workplaces, whether to consider peak or average concentrations of pollutants, the time interval of pollution measurement (hourly, daily, annual), and whether any lagged effects of pollution on behavior or health exist. Measures of several pollutants and weather conditions (such as temperature, humidity, or precipitation) are commonly needed to isolate the partial effect of any one pollutant (Dickie and Gerking 1991b; Bresnahan, Dickie, and Gerking 1997). The goal is to use the environmental measures that are most closely related to the health effect (check the epidemiological and other health literature) and behaviors you consider.

In addition to the exogenous inputs in the health-production function, the other exogenous variables in the model include income, wages, and prices. Of course, the prices of defensive goods need to be measured. Sometimes everyone faces the same market prices for defensive goods. For example, anyone in my city can buy bottled water or sun screen lotion for the same prices I pay. In these cases, there is no variation in price and the price variable cannot be used as a regressor. But data on price are still needed to compute marginal

WTP. When a defensive good is a durable good, data on capital and operating expenses and on the expected useful life of the good are needed (Dickie and Gerking 1991a). Many times defensive behaviors take time, so you want data on the time required to utilize medical care, to boil tap water, or to use a child safety seat in an automobile.

Prior empirical work on defensive behavior suggests that data on attitudes, beliefs, and perceptions will be helpful in modeling defensive decisions. For example, subjective measures of attitudes toward the relevant pollutant, the health effects of exposure, or the effectiveness of the local water supplier often have significant effects on defensive behavior (Dickie and Gerking 1997; Smith and Desvousges 1986; Smith, Desvousges, and Payne 1995). Previous experience with damage from exposure (Berger et al. 1987; Bresnahan, Dickie, and Gerking 1997; Dickie and Gerking 1997; Jakus 1994), knowledge and information about exposure and its consequences (Abdalla, Roach, and Epp 1992; Dickie and Gerking 1997; Smith and Desvousges 1986; Smith, Desvousges, and Payne 1995), and subjective assessments of risk (Abdalla, Roach, and Epp 1992; Viscusi and Cavallo 1994) all appear to increase the propensity to take defensive action. Thus, it is useful to collect data on variables of this type.

3.4.2 Data for Estimating Demand for a Defensive Input

The data requirements to estimate the demand for a defensive input, to implement the Bockstael-McConnell approach, are similar to if slightly less stringent than those for estimating a health-production function. The aim is to estimate an equation like (19), in which the left-hand side is a measure of demand for a defensive input, and the right-hand side consists of all exogenous variables in the model. Thus, the data requirements are similar to those just discussed for the health-production function. However, to estimate equation (19), you do not need data on the health outcome, and you need data on only one (necessary) defensive input. Estimating the health-production function requires data on the health outcome and on all defensive inputs.

3.4.3 Data for Estimating Changes in Defensive Expenditures

The minimal requirements for a defensive expenditure study are data on expenditures (e.g., spending on bottled water) or on defensive behaviors and their unit costs (e.g., amount of bottled water purchased and its price). As discussed previously, the full cost includes the value of time devoted to defensive behaviors and so the time spent and some measure of the value of time (presumably the wage rate) must be known.

Now, the welfare measure is not based on the level of defensive expenditures but on the change in expenditures in response to a change in environmental conditions. In practice, most researchers have examined changes in defensive expenditures arising from temporary incidents of water contamination (Abdalla 1990; Abdalla, Roach, and Epp 1992; Harrington, Krupnick, and Spofford 1988; Laughland 1996).

With data on the observed change in defensive expenditures, you can bound the compensating or equivalent variation as discussed in section 2 or more fully by Bartik (1988), and that may be enough for some purposes. But the study can be enhanced by showing that defensive expenditure or defensive behavior is plausibly related to key determinants. To do this, data on the determinants of defensive behaviors are needed to estimate an equation like (19) for each defensive action or for overall defensive expenditures.

3.4.4 Data for a Consumer Market Study

The consumer market study requires less data-collection effort than the other three approaches. The reason is that you don't econometrically estimate a model. Instead, you assemble estimates made by others. The typical application is the purchase of a safety good to reduce risk of death in a particular type of accident. The data needs are very context specific, but in general two terms are needed: the cost of the defensive action in the numerator of equation (17), and its effectiveness in the denominator.

Estimates of the effectiveness of safety goods in reducing risk often are available from government, industry, consumer groups or in academic articles or government reports. However, these estimates often are derived by assuming that consumers use the goods in an ideal way (e.g., always having working batteries in the smoke detector, wearing the bicycle helmet every time, and so on). Try to get information on how people actually use the good so the risk reductions may be adjusted accordingly.

Data on prices often are available from government, industry, or consumer groups, but you may need to survey retail sellers yourself. Of course, many goods are sold with different features at different prices and determining the price requires some judgment. A simple and defensible solution usually is to estimate some sort of average price (preferably weighted by market share). But the cause for the price variation should be considered. Do more expensive models provide more safety? If so, then try to match the price variation to the safety variation to get a more refined estimate of the value of reduced risk. If not, then the price variation may arise from other utility-bearing attributes of the product that are unrelated to risk. In that case, I would recommend following Jenkins, Owens, and Wiggins (2001) and using a lower-end price for the good, on the grounds that a person could purchase the risk reduction for this price, and anyone paying more is buying something besides safety. In any event, obtain data on capital and operating expenses if the good is durable and data on any time requirements for use.

3.4.5 Secondary versus Primary Data

The consumer market approach is by definition based on secondary data. Most applications of the health-production and defensive-input demand approaches also have used data originally collected for some other purpose (Agee and Crocker 1996, 2001; Cropper 1981; Dickie 2000; Gerking and Stanley 1986; Joyce, Grossman, and Goldman 1989). Using secondary data avoids the time and expense of conducting a survey. In some cases, nationally representative samples reflecting extensive quality control and high response rates may be available. However, use of secondary data limits the defensive behaviors and health outcomes to those measured in the source. For example, medical care utilization is the only defensive action included in many health data sets. These data often do not include measures of the money prices, let alone time requirements, of medical care. Secondary data often do not include indicators of perceptions and attitudes that are helpful in explaining choices, and confidentiality restrictions can impede matching of survey respondents to measures of local environmental conditions.

Primary data collection has been used in a few health-production or defensive-input demand studies (Dickie and Gerking 1991a, 1991b, 1997) and in most defensive expenditure studies. This is the most costly and most flexible way to acquire the desired data. Conducting a survey allows the researcher to collect data on the specific defensive behaviors and health outcomes of interest, on attitudes and perceptions, and on prices of defensive behaviors. However, this often involves settling for small samples that may not be nationally representative. In collecting primary data, general principles of sampling and survey design apply (see Chapter 3). Some special considerations for defensive behavior studies include measuring the full cost of defensive behaviors (inclusive of time costs) and measuring perceived benefits or costs of pollution and/or defensive behaviors, (e.g., changes in perceived risks).

When developing a survey, I recommend building in a stated-preference or contingent valuation component (see Chapters 5 and 6) involving a contingent defensive good constructed to avoid joint-production problems. Survey respondents already are thinking about defensive behaviors, their costs, and their impacts on health, so a stated-preference component follows quite naturally.

3.5 Estimating the Model

3.5.1 Estimating a Health-Production Function

With your data in hand you are now ready to estimate your model using one of the four defensive behavior approaches. I'll begin with the health production function approach because it is the most complicated. You should consult the large literature on estimating health production functions (Rosenzweig and Schultz 1982, Grossman and Joyce 1990, Gerking and Stanley 1986, and Dickie and Gerking 1997 would be good places to start), but here are the basic ideas.

The main econometric issue in estimating a health production function is the simultaneous equations problem. Intuitively, the health outcomes experienced depend partly on defensive behavior, but defensive behavior is chosen. So you have endogenous variables on the right-hand side of the health production function, and simultaneous equations problems arise. Unobserved factors affecting health outcomes will be correlated with unobserved factors affecting choices of health input. Ignoring this simultaneity results in biased and inconsistent estimators of parameters of the health production function.

This problem has been analyzed extensively by health economists following the influential work of Rosenzweig and Schultz (1983), and the researcher interested in estimating health production functions, or household production functions for other outcomes, should consult this literature (Grossman and Joyce 1990 and the references cited there). Essentially, the remedy is to apply simultaneous-equation estimation methods, such as instrumental variables, two-stage, or systems estimators.

You try to identify parameters of the health production function using exogenous variables that affect the choice of defensive behavior but do not *directly* affect the health output. (The variables indirectly affect health by influencing choices of defensive behaviors.) These variables include income, wages and prices, as well as any attitudinal or demographic variables that you think influence defensive decisions.

You use these variables, together with all of the exogenous inputs in the health production function, as instrumental variables. Usually, you want to use a two-stage estimator such as two-stage least squares. (An alternative is a systems estimator; see Dickie and Gerking 1997). In the first stage, regress the defensive behaviors on the instrumental variables, and use these results to estimate the health production function in the second stage (Gerking and Stanley 1986; Joyce, Grossman, and Goldman 1989).

Unfortunately, theory does not nail down the exact set of instrumental variables you should use, and your results can be sensitive to the choice of instruments. My recommendation is to let theory guide the choice of instruments to the extent you can, test the exogeneity of overidentifying restrictions, and fully document the choices you make.

By using theory to the extent you can, I mean following the principle behind equation (19): choices of defensive behavior depend on all exogenous variables in the model. This gives you a defensible way of choosing instruments and allows you to interpret the first-stage regressions for defensive behaviors as demand equations. Then, at least you have some idea what the first-stage results should look like.

By testing the exogeneity of overidentifying restrictions, I mean looking up this type of test in your econometrics book and implementing it if you can (see Dickie 2000 for an example). Fully documenting the choices you make, means at a minimum listing the instrumental variables used. A better approach is to present the results of the reduced form regressions, perhaps in an appendix, and to describe briefly what happens if you make other choices of instruments.

The estimation problem is further complicated by the fact that defensive inputs and/or health outcomes often will be measured as categorical, ordered or censored variables or as counts (e.g., whether or not a particular defensive behavior was employed, or how many days of illness were experienced). Simultaneous equations estimators are more complex in this setting than in the familiar linear equations, two-stage least squares problem. Relevant estimators are context-specific and a full discussion of this issue lies beyond the scope of this chapter. You need to read what others have done (Agee and Crocker 2001;

Dickie and Gerking 1996; Dickie 2000; Gerking and Stanley 1986; Joyce, Grossman, and Goldman 1989) and consult econometric references.

Once you have estimated the health-production function, estimating marginal WTP is a straightforward application of an equation like (16) or (17). You use your estimated health production function to compute the required marginal products, and use these along with the price (including time costs) of the defensive action to compute WTP.

3.5.2 Estimating Demand for a Necessary Defensive Input

Now, if you are using the Bockstael-McConnell approach and estimating a change in area behind the demand curve for a defensive input, you avoid the simultaneous equations problem. The reason is that the demand equation, like the one specified in equation (19), is a reduced-form equation. It does not include any endogenous variables on the right-hand side. Include as explanatory variables income, prices, wages, and any exogenous inputs in the health production function, as in equation (19). From there, you are estimating a demand equation, which is a standard econometric problem.

Once the demand equation is estimated, compute the welfare change by calculating the area in equation (20). To do this you first estimate the choke price, by setting other explanatory variables at fixed values (e.g., sample means) and solving the estimated demand equation for the price that drives quantity demanded to zero. (It gets a little more complicated if the demand curve does not intersect the price axis; see Dickie and Gerking 1991b). For some functional forms of the demand equation you can solve for a closed-form solution for the integral in equation (20), but in other cases you will have to resort to numerical integration (Dickie and Gerking 1991b).

3.5.3 Estimating Change in Defensive Expenditures

Next consider the problem of estimating the change in defensive expenditures to bound the value of a non-marginal change. Since most studies have used actual defensive expenditures, I will focus on this approach.

The basic computation of the change in defensive expenditures is simple: you multiply the changes in defensive behavior by corresponding unit costs and sum over all the defensive behaviors considered. I recommend following two refinements used by Harrington, Krupnick, and Spofford (1989) concerning time costs. First, conduct sensitivity analyses based on alternative assumptions about the value of time devoted to defensive action. Second, consider whether defensive behaviors are taken jointly with other actions. For example, if someone makes a special trip to haul water from an alternate source, all of the time spent is a cost of defensive action. But if someone buys bottled water while shopping for groceries, the marginal time cost is zero.

As mentioned previously, I think you should conduct some supplementary statistical analysis of the defensive behaviors to bolster the credibility of your expenditure estimate. This involves estimating equations like (19) for each defensive behavior or for overall defensive expenditures to insure that choices of defensive behaviors are plausibly related to determinants.

But I have some reservations about the value of defensive expenditure studies. Granted, you are able to bound WTP but you do not know the accuracy of the bound. You do know that the usual approach of using actual expenditures gives a less accurate bound than the theoretically preferred measure of the change in defensive expenditures that would hold health constant. Moreover, estimates of defensive expenditures are more difficult to interpret from an economic perspective than estimates of the parameters of underlying demand, health production, or preference functions.

3.5.4 Estimating WTP in a Consumer Market Study

Finally, consider the consumer market study. Most of the work is done by now. You have estimated the cost of the defensive good, accounting for capital and operating expenses, time costs, and expected useful life. You have estimated the risk reduction, accounting for how the good actually, rather than ideally, is used. Now get out your calculator and divide the former by the latter and you have your marginal welfare estimate.

Your welfare estimate represents the WTP of the marginal purchaser. Thus it is helpful to know the proportion of individuals or households who purchase good. If this proportion is 50%, you have estimated the WTP of the median individual or household. But your WTP estimate exceeds or falls short of the median WTP (by an indeterminate amount) according to whether the proportion of purchasers is less than or greater than 50%.

The consumer-market approach often can be implemented relatively quickly and at low cost. But unless you are very constrained by time or money, or unless you supplement the basic methodology with substantial statistical analysis (Blomquist 1979), I don't see much point in one of these studies. The method usually involves computation rather than statistical estimation, and therefore no information on the precision of the estimated WTP is available. With no equations being derived from a model of behavior and then estimated econometrically, it is difficult to gauge the plausibility of the results, except to judge whether or not they confirm your prior beliefs.

3.6 Concluding Comments on the Defensive Behavior Method

Previous empirical research supports the basic premise of the defensive behavior method: sizeable fractions of the population appear to engage in defensive action; the propensity to take protective action is plausibly related to determinants suggested by theoretical models, and defensive action increases with the level of pollution. Going from the existence of defensive behavior to the estimation of values, however, has proved difficult.

I'm aware of only two defensive behavior studies using the Bockstael-McConnell approach to estimate welfare based on the demand for a necessary defensive input (Agee and Crocker 1996; Dickie and Gerking 1991b). This is not enough to assess how well the method performs.

Several researchers have estimated marginal WTP using the health-production function approach. But the number of studies is small in light of differences in the health outcomes and pollutants considered, in the persons affected (e.g., adults, school-age children, infants), in conditions at the study sites, and in the changes valued. This makes it difficult to compare results and to draw general conclusions about the validity of the method. It seems clear, however, that much of the research is intended to develop or illustrate methods rather than to compute policy-relevant welfare estimates.

Likewise, a several researchers have used changes in defensive expenditures to bound WTP (Abdalla 1990; Abdalla, Roach, and Epp 1992; Chestnut 1996; Harrington et al. 1986; Jakus 1994; Laughland et al. 1996; Murdoch and Thayer 1990). With the exception of Murdoch and Thayer (1990), these studies use actual defensive expenditures rather than the expenditures that would hold household outputs constant when contamination changes. Most of the studies involve temporary incidents of water contamination. Again, differences between the studies in the adverse outcome avoided and in the duration of contamination complicate comparisons of results.

Interpretation of most of defensive behavior estimates of value is further complicated by incomplete control for joint production. Joint production is the most serious methodological issue affecting validity of the defensive behavior method, and is the major obstacle to its wider application. A related methodological problem concerns measurement of prices of defensive inputs. Measuring the price of medical care, a good often treated as a defensive input, is difficult in light of the (possibly endogenous) variation in insurance coverage. Medical care and many other defensive behaviors require time, and consequently measures of time requirements and of the value of time are necessary to estimate the full price of defensive inputs (see Chapter 9 on the value of time). Furthermore, the cost of some defensive behaviors, such as staying indoors to avoid ambient air pollution, is primarily a matter of disutility rather than monetary expenditure (Freeman 1993).

An additional problem, particularly when secondary data are used, is possible divergence between objective estimates and subjective perception of risks. Decisions are based on perceived costs and benefits of defensive behaviors, but perceptions data are rarely available unless a special survey is conducted.

Primary data collection can provide improved measures of defensive input prices and perceptions and may help overcome joint production problems. If the time and expense of a special survey are to be taken, however, the researcher should consider whether the defensive behavior method is the best approach to obtaining accurate and defensible valuation estimates. If problems of joint production or input price measurement are severe, or if survey information on defensive expenditures will only bound WTP, then a stated preference approach may be a better way to value health effects of pollution.

4. DAMAGE COSTS AND THE COSTS OF ILLNESS

4.1 Introduction

The damage cost method attempts to estimate the resource cost associated with environmental changes, rather than WTP. The method has been applied to materials damage and other endpoints, but like the defensive behavior method it is most frequently applied to health valuation. The specific damage cost method applied in the health context is known as the cost of illness approach. This method is the most frequently used approach in cost-benefit and cost-effectiveness analysis of health, safety and environmental policy, and similar methods are commonly used to assess damages in civil cases (Link 1992).

The cost of illness is defined as the sum of direct and indirect costs associated with illness, injury, or death. Direct costs reflect the value of resources used to diagnose, treat, rehabilitate, or support ill or injured persons. Indirect costs, on the other hand, are not out-of-pocket expenses but rather reflect the value of output not produced because of morbidity or premature mortality. Foregone earnings are the main component of indirect costs.

The basic cost-of-illness methodology predates Adam Smith, but modern pioneering works include Mushkin and Collings (1959), Weisbrod (1961), and especially Rice (1966), who estimated the total annual costs of illness for the U.S. Her paper, along with its extensions and updates (Cooper and Rice 1976; Rice, Kopstein, and Hodgson 1985), set the standard for aggregate, prevalence-based cost-of-illness studies. As methods have evolved since the 1960s, the definition of direct costs has been broadened to include non-medical expenditures such as rehabilitation and special education; the valuation of foregone household production has become standard in the best studies, and the incidence approach has emerged as an alternative to the prevalence approach as a basis for measuring costs.

4.2 Prevalence and Incidence Approaches

There are two approaches for measuring costs, corresponding to two different measurements of the occurrence of a disease or condition. The *prevalent* population consists of all persons who have a condition at a given time, while the *incident* population consists of new cases diagnosed during a given time. Correspondingly, prevalence-based costs are annual costs associated with cases of a condition present during the year, while incidence-based costs are lifetime costs of new cases diagnosed during the year.

In the prevalence approach, direct costs and indirect morbidity costs are measured as costs incurred during the year. Indirect mortality costs are the present value of lifetime losses in earnings and household production associated with persons who die from the condition during the year. The incidence approach, on the other hand focuses on *lifetime* costs from onset of the condition until recovery or death, measured for persons diagnosed within the year.

In measuring the costs of asthma, for example, the prevalence approach focuses on all persons alive at the beginning of the year who have been diagnosed with asthma (or perhaps those who currently suffer from asthma symptoms). Direct costs of treatment include emergency room visits, physician visits, pharmaceuticals, and similar expenditures during the year. Morbidity costs include lost earnings and foregone household production from days of asthma-related illness during the year. Mortality costs reflect the present value of lost lifetime earnings and household production for the small number of asthma-related deaths during the year.

The incidence approach measures lifetime costs of persons diagnosed with asthma during the year. Medical costs and other expenses are forecast over remaining life expectancy while accounting for survival probabilities and the potential for becoming asymptomatic. Lost earnings and household production from asthma-related morbidity over the remaining lifetime, and from future premature mortality, also are forecast and discounted.

The incidence approach is more complicated because it requires forecasting the course of morbidity and mortality, and the associated treatments and costs, years into the future. However, practitioners agree that the choice between these methods should be dictated not by convenience but by the nature of the policy question. The prevalence approach is appropriate for valuing policies affecting treatment, alleviation, or management of existing diseases, because these policies affect costs in the prevalent population while leaving incidence unchanged. The incidence approach is appropriate for valuing policies that prevent or delay onset of new cases of disease, because prevention avoids the lifetime costs that would otherwise occur.

5. CONDUCTING A COST OF ILLNESS STUDY

Table 2 lists a stylized sequence of steps to take in conducting a cost-of-illness study. Of course, every application is different, but the table lists many of the key decisions and actions.

Table 2. Steps for Conducting a Cost of Illness Study

1.	Define the valuation problem.
	Specify the alternative condition. Determine appropriate specificity of condition valued. Define linkage to environmental conditions. Prevalence or incidence approach?
2.	Estimate direct costs.
	Identify relevant categories of health-care costs. Obtain data on utilization and cost for each category. Estimate costs using prevalence or incidence methodology. Estimate non-medical direct costs.
3.	Estimate indirect costs.
	Estimate morbidity costs. Estimate morality costs using value of statistical life or cost-of-illness methodology. Account for changes in value of home production.
4.	Adjust costs to a common base year.
	Adjust for changes in prices and wages. Discount to present value. Adjust for changes in prevalence or incidence over time.

5.1 Defining the Valuation Problem

Start by defining the valuation problem, including baseline and alternative conditions. Interpretation of cost-of-illness estimates is difficult when the alternative condition, the counterfactual relative to which cost estimates are made, is left implicit (Waitzman, Scheffler, and Romano 1994).

Before measuring costs, determine the specificity of the condition to be valued. The most common approach is to estimate costs for a particular diagnosis or group of diagnoses, such as costs for asthma, low birth weight, or hypertension. This approach is useful for evaluating policies that affect a particular condition and is convenient because conditions are defined to match diagnosis codes used in medical records. An alternative approach may be appropriate when the policy considered affects a causative agent for several conditions. For example, policies reducing motor vehicle accidents would reduce the incidence of many different injuries, and policies to limit tobacco smoking would reduce incidence of many illnesses. You can try to estimate all costs attributable to the causative agent, rather than focusing on a specific condition (Hartunian et al. 1981; Aligne and Stoddard 1997; Rice and Max 1992). A third approach is to estimate costs of particular outcomes that may be associated with several underlying conditions and/or causative agents, such as costs of respiratory hospital admissions. The least specific approach is to estimate aggregate costs for broad categories of disease, such as the total cost of respiratory illness, or the total cost of accidental injuries, in the U.S. (Rice, Hodgson, and Kopstein 1985). This approach has limited value in assessing specific policies but is informative about the economic burden associated with categories of illness.

When estimating costs of illness for environmental policy, another consideration is making the link between costs and environmental quality. Typically, this link is documented by an epidemiological study or other information from a risk assessment. Connecting illness costs with the environment is most straightforward when pollution affects the incidence of a disease. In that situation, costs can be attributed to pollution in proportion to the impact of pollution on incidence. When pollution aggravates existing disease, it may aggravate only some outcomes associated with the disease, making a simple proportionate allocation of prevalence costs to pollution inappropriate.

Having resolved these preliminary issues, consider how to measure the costs of illness. Determine whether to estimate costs using a prevalence or an incidence approach (section 4.2). Then, try to measure incremental rather than total costs. Incremental costs are the additional costs associated with a condition, and clearly are a better guide to policy decisions than the total costs incurred by persons with the condition. For example, the total medical expenses and work loss of persons with hypertension is less informative about the economic burden of the disease than the incremental costs of hypertension alone.

Total costs, however, are more easily calculated. One simple way to estimate incremental costs is to first estimate total costs and then deduct the average costs for the population as a whole (including persons without the condition). Yet even this approach may fail to accurately estimate incremental costs in the presence of comorbidities, the common situation where persons have more than one condition. Isolating incremental costs associated with one of the conditions alone can be quite difficult. Judgment can be used in matching costs to conditions or an arbitrary rule such as assigning all costs to the diagnosis listed first on medical records can be used. Almost all cost-of-illness studies attempt to measure indirect costs incrementally, by focusing on work loss related to the condition in question. But direct costs sometimes are measured as total costs incurred by persons with the condition.

5.2 Estimating Direct Costs

A good starting point is to identify relevant categories of health-care costs, such as hospital care (inpatient, outpatient, emergency room, and surgical services), physician services (inpatient and outpatient), nursing or long-term care facility services, home health care, and drugs and devices. Then, obtain data on utilization and cost for each category, for persons with the condition to be valued. National data sources frequently used for this purpose are described elsewhere (Haddix et al. 1996, Appendix E; Waitzman, Scheffler, and Romano 1994; Agency for Health Care Policy and Research 1996). Persons with the condition are matched to utilization of services using diagnosis codes. Other potential data sources include information provided by specific health plans or by health-care providers. Then estimate costs as the product of utilization times cost per unit, summed over categories.

While the prevalence approach focuses on annual costs of all persons who have the condition, the incidence approach requires a forecast of lifetime costs of persons diagnosed in a given year. Two alternative methods may be used to construct this estimate.

The *empirical*, *actual*, or *effective demand* approach uses current cross-sectional data on the age profile of costs, and assumes the same profile will continue to hold in the future. Thus, the costs of a 65 year-old today are taken to predict the costs to be experienced in 20 years by today's 45 year-old. The *theoretical* or *required* approach simulates the lifetime profile of medical care that would be required for a person of a given age with the condition. Reviews of the medical literature, expert opinion, or physician panels may be used to develop this profile. In either case, adjust the time path of costs by an expected relative price increase for medical care (medical care component of the consumer price index relative to the overall CPI).

Estimate non-medical direct costs similarly. Identify cost categories, such as rehabilitation services, or developmental or special education services.

Estimate utilization rates by persons with the condition, based on expert opinion or reviews of the medical literature, and multiply by cost estimates.

5.3 Estimating Indirect Costs

Morbidity and mortality costs usually are estimated separately. Consider first morbidity, which may affect earnings in three distinct ways. A chronic or acute illness may cause a person to miss regularly scheduled work time. A chronic illness also may limit the amount of time a person regularly works. Finally, by restricting the kind or amount of work a person can do, a chronic impairment may reduce the wage received.

Except in cases of total disability, cost-of-illness studies usually have focused only on reductions in regularly scheduled work time. The estimate is constructed by combining estimates of restricted activity days or work loss days for persons with a condition (from the National Health Interview Survey, for example) with data on the labor force participation and earnings of the general population (from the Current Population Survey for example). Age- and gender-specific participation and earnings data often, but not always, are used. Restricted activity and work loss days in the general population may be subtracted from those experienced by persons with the condition to obtain incremental estimates.

In the case of chronic disability, a more complete accounting of indirect cost includes effects of restrictions on the kind or amount of work impaired individuals can do, provided care is taken to avoid double-counting. Many national surveys inquire whether respondents are limited in the kind or amount of work they can do; the NHIS also inquires whether the limitation is caused by a particular chronic condition. For persons unable to work, the entire earnings are assumed lost. For persons able to do only limited work, an alternative procedure may be used which also accounts for potential differences in wages between chronically impaired persons and the general population. The procedure compares annual earnings (by age and gender) for persons with and without a given condition, thus capturing differences arising from wages and from hours worked. The Survey of Income and Program Participation is useful for this purpose, although its coding of conditions is fairly broad.

Prevalence-based morbidity costs include annual earnings foregone for all persons with the condition, while incidence-based estimates include lifetime costs for newly diagnosed cases. Lifetime losses generally are estimated using a fixed expected worklife based on age, gender, and race or education. Some analysts use an alternative "L-P-E" method based on survival probabilities (L), labor force participation rates (P), and earnings (E) given participation at various stages of the life cycle. Typically, real wages are assumed to rise at a constant rate of productivity growth (one percent per year is a standard assumption).

Consider next the earnings losses from premature mortality. The best approach for valuing mortality losses does not rely on the cost-of-illness method but instead uses estimates of the value of a statistical life derived from the WTP to reduce fatal risks. If the cost-of-illness method of mortality valuation is to be used, however, it would proceed as follows. The prevalence approach includes foregone earnings from persons who died from the condition in a given year, identified from data on cause of death. The incidence approach includes foregone earnings from premature mortality of persons diagnosed during a given year, and requires information on the change in life expectancy (or the change in annual survival probabilities) associated with the condition. Survival information may be obtained from data sets compiled from death certificates, from expert opinion, or from reviews of the medical literature. Aside from the difference in the number and timing of deaths considered, the methodology for estimating mortality costs is essentially the same for prevalence and incidence approaches, and mirrors the incidence approach to estimating morbidity costs. Current earnings estimates are extrapolated forward at a constant rate, usually one percent per year.

Finally consider the value of foregone household production, an important issue when measuring costs experienced by homemakers (usually no estimate of household production is made for persons in the labor force). Two main approaches have been taken to value household production services.

The *market cost* approach is based on wages required to hire out the services performed by a homemaker using either the average wage for domestic help or for individual types of workers for specific household tasks. Clearly, domestic help is not a perfect substitute for homemaker services and if it were, only households who valued homemaker time at less than market cost would choose to keep a homemaker out of the labor force. The *opportunity cost* approach is based on the foregone earnings of the homemaker. Foregone earnings typically are estimated using persons in the labor force, matched by personal characteristics like age and education. The potential self-selection bias is evident to practitioners, and estimates of shadow wages are often discussed but rarely used. In any event, the value of output produced must exceed the opportunity cost of homemaker time, or the homemaker would not remain at home. A problem with both these approaches is the focus on costs of household labor input rather than values of home-produced outputs. (Two approaches have been proposed to value household outputs, but neither is commonly used.)

5.4 Adjusting Costs to a Common Base Year

Three adjustments may be required, all relatively straightforward from a computational perspective: adjustments in prices and wages, discounts to present value, and adjustments in prevalence or incidence rates. Prices and wages often must be adjusted because data were obtained from a previous year, or results are forecast into the future. Adjustment for future growth in real wages has been discussed in connection with indirect cost estimation. Actual historical growth in nominal wages often is used to increase earnings from a previous year to the present. Changes in the real price of medical care over time generally are accounted for using the medical care component of the consumer price index relative to the all-items Consumer Price Index. Conceptual issues of discounting lie well beyond the scope of this chapter, but the choice of a discount rate can substantially affect incidence-based cost estimates. Many cost-of-illness practitioners follow the sensible procedure of presenting results for a range of alternative discount rates. If prevalence or incidence is stable, or if interest centers on per-case estimates of costs, then there is no need to adjust results for changes in prevalence or incidence. But if aggregate estimates are desired, then any substantial temporal changes in prevalence or incidence should be factored into the aggregation of costs.

5.5 Concluding Comments on the Cost of Illness Method

Direct costs (perhaps by major category of cost), morbidity costs, and mortality costs should be presented separately and summed for a total. Resources permitting, perform sensitivity analysis to illustrate effects of changing key assumptions, and compare results to other cost-of-illness studies (the number of cost-of-illness studies is large enough that estimates are almost surely available for the same or similar conditions). Comparing results to estimates of total costs for broad categories of illness is sometimes helpful. For example, compare estimated costs of a particular respiratory disease to the Rice, Hodgson, and Kopstein (1985) estimates of cost for all respiratory diseases.

Cost-of-illness methods have been refined over decades of widespread use. Nonetheless, at least four important methodological issues remain incompletely resolved. First, most data on direct costs report charges rather than costs. Costs are the desired measure, but cost-to-charge ratios vary markedly by source of third-party payment, location, and other factors. Second, direct cost measures usually include fixed overhead expenses like insurance administration, rather than only variable costs. Overhead would unlikely vary in proportion to any particular illness. Third, all available methods of measuring indirect costs from foregone household production have shortcomings, and no satisfactory method is used to measure indirect costs for children or retired persons. Finally, techniques for allocation of costs in the presence of comorbidities are somewhat arbitrary.

Methodological difficulties aside, the validity of the cost-of-illness method can be challenged on more fundamental grounds: it is not a money measure of welfare change. The costs of illness are likely to be a lower bound on WTP, and that may be information enough for some policy issues, such as the cost-benefit analysis of removing lead from gasoline (US EPA 1985). But available evidence suggests that the relationship between WTP and illness costs is not constant, but instead varies by illness.

Two types of empirical comparisons of WTP and costs of illness can be distinguished. One type compares estimates of WTP and costs of illness for the same individuals. For example, Berger et al. (1986) find that WTP to avoid a group of symptoms exceeds the associated cost of illness by a factor of 21. Chestnut et al. (1988, 1996) find ratios of WTP to the cost of illness for angina episodes ranging from 2 to 8. Dickie and Gerking (1991b) and Rowe and Chestnut (1985) each report ratios of 2 to 4, for health effects of ozone and asthma episodes, respectively, although the Dickie and Gerking estimates include only medical expenses (not foregone earnings). A second type of study compares WTP estimated from one sample of individuals to costs of illness from another study. Agee and Crocker (1996) find that WTP to reduce child lead burdens exceeds the cost of illness by a factor of 3 to 20, while Viscusi et al. (1991) and Krupnick and Cropper (1992) report a WTP to cost of illness ratio for chronic bronchitis of 3 to 6. In summary, it appears that WTP exceeds the cost of illness by a factor of at least two, but the divergence between the two measures varies substantially over different health endpoints and research efforts.

6. FINAL WORDS

This chapter discussed implementation of defensive behavior and damage cost methods of nonmarket valuation, with particular emphasis on valuation of human health. The defensive behavior method emphasizes behavioral adjustments individuals make to environmental changes. Joint production and other implementation problems hinder wider use of the defensive behavior method. The damage cost method estimates real resource costs of environmental pollution. Damage costs are not equal to WTP, but are expected to provide a lower bound.

ACKNOWLEDGMENTS

I thank A. Myrick Freeman and the editors for helpful comments on previous drafts.

REFERENCES

Abdalla, C. A. 1990. Measuring Economic Losses from Ground Water Contamination: An Investigation of Household Avoidance Cost. *Water Resources Bulletin* 26:451-463.

Abdalla, C. 1994. Groundwater Values from Avoidance Cost Studies: Implications for Policy and Future Research. *Journal of the American Agricultural Association* 76:1062-1067.

Abdalla, C. A., B. A. Roach, and D. J. Epp. 1992. Valuing Environmental Groundwater Changes Using Averting Expenditures: An Application to Groundwater Contamination. *Land Economics* 68:163-69.

Agee, M. D., and T. D. Crocker. 1996. Parental Altruism and Child Lead Exposure: Inferences from the Demand for Chelation Therapy. *The Journal of Human Resources* 31(3):677-691.

Agee, M. D., and T. D. Crocker. 2001. *Smoking Parents' Valuations of Own and Children's Health.* Association of Environmental and Resource Economists Conference, Assessing and Managing Environmental and Public Health Risks. www.aere.org/.

Agency for Health Care Policy and Research. 1996. *A Compendium of Selected Public Health Data Sources.* Washington, DC: U.S. Department of Health and Human Services.

Akerman, J., F. R. Johnson, and L. Bergman. 1991. Paying for Safety: Voluntary Reduction of Residential Radon Risks. *Land Economics* 67:435-46.

Aligne, C., and J. Stoddard. 1997. Tobacco and Children: An Economic Evaluation of the Medical Effects of Parental Smoking. *American Journal of Diseases of Children* 151:648-653.

Bartik, T. J. 1988. Evaluating the Benefits of Non-marginal Reductions in Pollution Using Information on Defensive Expenditures. *Journal of Environmental Economics and Management* 15:111-27.

Behrman, J. R., R. A. Pollak, and P. Taubman. 1995. *From Parent to Child: Intrahousehold Allocations and Intergenerational Relations in the United States.* Chicago: University of Chicago Press.

Berger, M. C., G. C. Blomquist, D. Kenkel, and G. S. Tolley. 1987. Valuing Changes in Health Risk: A Comparison of Alternative Measures. *Southern Economic Journal* 53:967-83.

Blomquist, G. 1979. Value of Life-Saving: Implications of Consumption Activity. *Journal of Political Economy* 87(3):540-58.

Bockstael, N. E., and K. E. McConnell. 1981. Theory and Estimation of the Household Production Function for Wildlife Recreation. *Journal of Environmental Economics and Management* 8:199-214.

Bockstael, N. E., and K. E. McConnell. 1983. Welfare Measures in the Household Production Framework. *American Economic Review* 73:806-14.

Bresnahan, B. W., and M. Dickie. 1994. Averting Behavior and Policy Evaluation. *Journal of Environmental Economics and Management* 29(3):378-392.

Bresnahan, B. W., M. D., and S. Gerking. 1997. Averting Behavior and Urban Air Pollution. *Land Economics* 73(3):340-357.

Carlin, P. S., and R. Sandy, 1991. Estimating the Implicit Value of a Young Child's Life. *Southern Economic Journal* 58(1):186-202.

Chestnut, L. G., R. Keller, W. E. Lambert, and R. D. Rowe. 1996. Measuring Heart Disease Patients' Willingness to Pay for Changes in Angina Symptoms. *Medical Decision Making* 16:65-77.

Chestnut, L. G., S. D. Colome, L. R. Keller, W. E. Lambert, B. Ostro, R. D. Rowe, and S. L. Wojciechowski. 1988. *Heart Disease Patients' Averting Behavior, Costs of Illness, and Willingness to Pay to Avoid Angina Episodes.* Final Report to the Office of Policy Analysis, US Environmental Protection Agency.

Chestnut, L. G., L. R. Keller, W. E. Lambert, and R. D. Rowe. 1996. Measuring Heart Patients' Willingness to Pay for Changes in Angina Symptoms. *Medical Decision Making* 16: 65-77.

Cooper, B. S., and D. P. Rice. 1976. The Economic Cost of Illness Revisited. *Social Security Bulletin:* 21-36.

Courant, P. N., and R. C. Porter. 1981. Averting Expenditure and the Cost of Pollution. *Journal of Environmental Economics and Management* 8(4):321-329.

Cropper, M. L. 1981. Measuring the Benefits from Reduced Morbidity. *American Economic Review* 71:235-40.

Cropper, M. L., and A. J. Krupnick, 1990. *The Social Costs of Heart and Lung Disease,* Resources for the Future, Washington, DC. June.

Dardis, R. 1980. The Value of a Life: New Evidence from the Marketplace. *American Economic Review* 70(5):1077-1082.

Dickie, M., and S. Gerking. 1991a. Valuing Morbidity: A Household Production Approach. *Southern Economic Journal* 57(3):690-702.

Dickie, M. 1991b. Willingness to Pay for Ozone Control: Inferences from the Demand for Medical Care. *Journal of Environmental Economics and Management* 21:1-16.

Dickie, M. 1996. Formation of Risk Beliefs, Joint Production and Willingness to Pay to Avoid Skin Cancer. *The Review of Economics and Statistics* 78(3):451-464.

Dickie, M. 1997. Genetic Risk Factors and Offsetting Behavior: The Case of Skin Cancer. *Journal of Risk and Uncertainty*. Forthcoming.

Dickie, M. 2000. *Willingness to Pay for Children's Health: A Household Production Approach.* Presented at American Economic Association annual meeting.

Doyle, J. K., G. H. McLelland, W. D. Schulze, S. R. Elliott, and G. W. Russell. 1991. Protective Responses to Household Risk: A Case Study of Radon Mitigation. *Risk Analysis* 11(1):121-34.

Freeman, A. M. III. 1991. Indirect Methods for Valuing Changes in Environmental Risks with Nonexpected Utility Preferences. *Journal of Risk and Uncertainty* 4:153-165.

Freeman, A. 1993. *The Measurement of Environmental and Resource Values: Theory and Methods*. Washington, DC: Resources for the Future.

Gerking, S., and L. R. Stanley. 1986. An Economic Analysis of Air Pollution and Health: The Case of St. Louis. *The Review of Economics and Statistics* 78:115-21.

Gerking, S., and W. Schulze. 1981. What Do We Know about Benefits of Reduced Mortality from Pollution Control? *American Economic Review* 71:228-234.

Gold, M. R., J. E. Siegel, L. B. Russell, and M. C. Weinstein. 1996. *Cost-Effectiveness in Health and Medicine.* New York: Oxford University Press.

Grossman, M. 1972. On the Concept of Health Capital and the Demand for Health. *Journal of Political Economy* 80:223-55.

Grossman, M., and T. J. Joyce. 1990. Unobservables, Pregnancy Resolutions, and Birth Weight Production Functions in New York City. *Journal of Political Economy* 98(5):983-1007.

Haddix, A. C., S. M. Teutsch, P. A. Shaffer, and D. O. Dunet. 1996. *Prevention Effectiveness: A Guide to Decision Analysis and Economic Evaluation.* Oxford University Press.

Harford, J. D. 1984. Averting Behavior and the Benefits of Reduced Soiling. *Journal of Environmental Economics and Management* 11:296-302.

Harrington, W., and P. R. Portney. 1987. Valuing the Benefits of Health and Safety Regulation. *Journal of Urban Economics* 22:101-112.

Harrington, W., A. J. Krupnick, and W. O. Spofford. 1989. The Economic Losses of a

Waterborne Disease Outbreak. *Journal of Urban Economics* 25:116-37.

Hartunian, N. S., C. N. Smart, and M. S. Thompson. 1981. *The Incidence and Economic Costs of Major Health Impairments.* Lexington, MA: Lexington Books.

Hori, H. 1975. Revealed Preferences for Public Goods. *American Economic Review* 65:978-91.

Hu., T., and F. H. Sandifer. 1981. *Synthesis of Cost-of-Illness Methodology: Part 1.* Report to the National Center for Health Services Research, Department of Health and Human Services.

Jakus, P. M. 1994. Averting Behavior in the Presence of Public Spillovers: Household Control of Nuisance Pests. *Land Economics* 70:273-285.

Jenkins, R. R., N. Owens, and L. B. Wiggins. 2001. Valuing Reduced Risks to Children: The Case of Bicycle Safety Helmets. *Contemporary Economic Policy* 19:397-408.

Jenkins, R. R., N. Owens, and L. B. Wiggins. 2001. Valuing a Statistical Child's Life: The Case of Bicycle Safety Helmets. *Contemporary Policy Issues.* Forthcoming.

Joyce, T. J., M. Grossman, and F. Goldman. 1989. An Assessment of the Benefits of Air Pollution Control: The Case of Infant Health. *Journal of Urban Economics* 25:32-51.

Just, R. E., D. L. Hueth, and A. Schmitz. 1982. *Applied Welfare Economics and Public Policy.* Englewood Cliffs, NJ: Prentice-Hall, Inc.

Krupnick, A. J., and M. L. Cropper. 1992. The Effect of Information on Health Risk Valuation. *Journal of Risk and Uncertainty* 5:29-48.

Krupnick, A. J., W. Harrington, and B. Ostro. 1990. Ambient Ozone and Acute Health Effects: Evidence from Daily Data. *Journal of Environmental Economics and Management* 18:1-18.

Laughland, A. S., W. N. Musser, J. S. Shortle, and L. M. Musser. 1996. Construct Validity of Averting Cost Measures of Environmental Benefits. *Land Economics* 72(1):100-112.

Link, A. N. 1992. *Evaluating Economic Damages: A Handbook for Attorneys.* Westport, CT: Quorum Books.

Max, W, D. P. Rice, and E. J. MacKenzie. 1990. The lifetime cost of injury. *Inquiry* 27:332-43.

Murdoch, J. C., and M. A. Thayer. 1990. The Benefits of Reducing the Incidence of Nonmelanoma Skin Cancers: A Defensive Expenditures Approach. *Journal of Environmental Economics and Management* 18(2):107-119.

Mushkin, S. J., and E. A. Collings. 1959. Economic Costs of Disease and Injury. *Public Health Reports* 74(a):795-809.

Pollak, R. A., and M. Wachter. 1975. The Relevance of the Household Production Function and Its Implications for the Allocation of Time. *Journal of Political Economy* 83(2):255-277.

Quiggin, J. 1992. Risk, Self-Protection and Ex Ante Economic Value – Some Positive Results. *Journal of Environmental Economics and Management* 23:40-53.

Rice, D. P. 1966. Estimating the Cost of Illness. *Health Economics Series.* Public Health Service, Washington, DC.

Rice, D. P.,T. A. Hodgson, and A. N. Kopstein. 1985. The Economic Cost of Illness: A

Replication and Update. *Health Care Financing Review* 39:61-81.

Rice, D., T. Hodgson, and A. Kopstein. 1985. The Economic Costs of Illness: A Replication and Update. *Health Car□Financing Review* 7(1):61-80.

Rice, D., and W. Max. 1992. *The Cost of Smoking in California, 1989.* Sacramento, CA: California State Department of Health Services.

Rosenzweig, M. R., and T. P. Schultz. 1983. Estimating a Household Production Function: Heterogeneity, the Demand for Health Inputs, and Their Effects on Birth Weight. *Journal of Political Economy* 91(5):723-46.

Rowe, R. D., and L. G. Chestnut. 1985. *Oxidants and Asthmatics in Los Angeles: A Benefits Analysis.* U.S. Environmental Protection Agency, Washington, DC.

Shogren, J. F., and T. D. Crocker, 1991. Risk, Self-Protection, and Ex Ante Economic Value. *Journal of Environmental Economics and Management* 20(1):1-15.

Smith, V. K. 1991. Household Production Functions and Environmental Benefit Estimation. In *Measuring the Demand for Environmental Quality.* Edited by J. B. Braden, and C. D. Kolstad. Amsterdam: Elsevier.

Smith, V. K., and W. H. Desvousges. 1986. Averting Behavior: Does It Exist? *Economics Letters* 20:291-96.

Smith, V. K., W. H. Desvousges, and J. W. Payne. 1995. Do Risk Information Programs Promote Mitigating Behavior? *Journal of Risk and Uncertainty* 10:203-22.

Tolley, G., D. Kenkel, and R. Fabian. 1994. *Valuing Health for Policy: An Economic Approach.* Chicago: University of Chicago Press.

U.S. Environmental Protection Agency (U.S. EPA), 1985. *Costs and Benefits of Reducing Lead in Gasoline: Final Regulatory Impact Analysis.*

Viscusi, W. K., W. Magat, and J. Huber. 1991. Pricing Environmental Health Risks: Survey Assessments of Risk-Risk and Risk-Dollar Trade-offs for Chronic Bronchitis. *Journal of Environmental Economics and Management* 21(1):32-51.

Viscusi, W. K., and G. O. Cavallo. 1994. The Effect of Product Safety Regulations on Safety Precaution. *Risk Analysis* 14:917-29.

Waitzman, N., R. Scheffler, and P. Romano. 1996. *The Costs of Birth Defects: Estimates of the Value of Prevention*, Lanham, MD: University Press of America.

Watson, W., and J. Jaksch. 1982. Air Pollution: Household Soiling and Consumer Welfare Losses. *Journal of Environmental Economics and Management* 9:248-262.

Weisbrod, B. A. 1961. *Economics of Public Health.* Philadelphia: University of Pennsylvania.

Weiss, K. B., P. J. Gergen, and T. A. Hodgson. 1992. An Economic Evaluation of Asthma in the United States. *The New England Journal of Medicine*, 326(13):862-866.

Chapter 12

BENEFIT TRANSFER

Randall S. Rosenberger and John B. Loomis
Oregon State University and Colorado State University, respectively

1. INTRODUCTION

Previous chapters of this book have described how to conduct an original nonmarket valuation study. However, original research can be time-consuming and expensive. What can you do if your time and/or funding prevents you from conducting an original study? For example, suppose the West Virginia Department of Environmental Protection (WVDEP) is deliberating on whether to grant a permit for a coal mine. Suppose a decision must be made within 60 days. The agency wants to balance the benefits derived from extracting the coal with the environmental damages caused by mining. One impact is the loss of recreational fishing downstream from the mining site. WVDEP wants to know the value of fishing on this stream. How could you provide this information? Even if WVDEP had the necessary funds to conduct a primary study, obviously there is not enough time. In this chapter, we describe how existing data or summary statistics from previous research can be used in different decision contexts. We use the term *benefit transfer* to describe the use of information from previous research to inform decisions.

Benefit transfer is a colloquial term adopted by economics and means the use of existing data or information in settings other than for what it was originally collected. Although most benefit transfers focus on estimating values, their use can be much broader than this. Sometimes we may be interested in the responsiveness of demand for certain goods or services. What is the effect of a three dollar fee increase on the demand for camping at a U.S. Forest Service campground, and what would be the total revenue? Here we are not necessarily interested in the total value of camping, but in the change in use and potential revenue capture of a user fee program. Thus, benefit transfer can be used to inform policy and decision making processes at various stages. It can be used in framing the policy context, evaluating policies (U.S. EPA's (1997) assessment of the Clean Air Act), defining the extent of an affected market, pre-screening of natural resource damage assessment, and even determining whether original research is warranted. So, even though we will be focusing on its use to address valuation needs for nonmarket goods and services, do not lose sight of its broader potential. Before we get into the mechanics of benefit transfer, let us provide a historical context and a formal definition for the method.

1.1 A Historical Context

Rick Freeman's opening chapter in this book does a wonderful job of illustrating the need for value measures. Benefit transfer meets this need. Although people have been learning from each others' experiences for a long time, the formal process of using benefit transfer to obtain estimates for nonmarket goods and services is only a few decades old. This short history reflects the age of nonmarket valuation.

The U.S. Army Corps of Engineers, the U.S. Bureau of Reclamation, and the U.S. Forest Service identified a need for estimates of recreation values for use in formal project evaluations and planning purposes. In 1973, the U.S. Water Resources Council published unit day estimates for recreation activities for use in evaluating water-related projects. They updated these recreation value estimates in 1979 and 1983. In 1980, the U.S. Forest Service began publishing Resources Planning Act (RPA) values (as per person per activity day estimates) for recreation (Rosenberger and Loomis 2001). Other RPA values were published for timber, forage, minerals, and water. The U.S. Forest Service recreation estimates were driven by the Renewable Resources Planning Act of 1974, which required, among other things, a formal analysis of the costs and benefits associated with their programs.[1] Both, the U.S. Water Resources Council's unit day values and the U.S. Forest Service's RPA values, were derived primarily from a combination of past empirical evidence, expert judgment, and political screening.

In the early 1980s, Freeman (1984) began the formal process of evaluating benefit transfer. He defined some specific conditions under which primary data could be transferable. In 1992, a special section on benefit transfer was published in the journal *Water Resources Research.* Many of the top resource economists provided extensive critique of the method in this special section. Collectively, the various articles suggested protocol, defined theory, identified needs, and presented new approaches.

Most benefit transfer exercises prior to the special section in *Water Resources Research* used a value transfer method that either transferred point estimates or measures of central tendency from original research, or used administratively approved estimates. Loomis (1992) proposed that more information, and therefore more robust benefit transfer, could be achieved with the transfer of entire demand (or benefit or willingness-to-pay) functions. Walsh, Johnson, and McKean (1989, 1992) and Smith and Kaoru (1990) conducted meta-regression analyses of recreation economic research outcomes, potentially providing another approach to defining functions that could be used in benefit transfer.

Since 1992, many formal studies have been conducted investigating the application of the valuation methods described in this book and the validity of the various benefit transfer approaches that will be described in this chapter. The trend in research on benefit transfer is toward building models that are more sensitive to underlying nuances of data collected, either from multiple sites in a single study (Rosenberger and Phipps 2001; VandenBerg, Poe, and Powell 2001) or data collected by different sources (Rosenberger and Loomis 2001; Woodward and Wui 2001).

1.2 Benefit Transfer Defined

We can learn about something through our direct experiences and vicariously through others' experiences. Benefit transfer fits the latter. Benefit transfer is the adaptation of information derived from original research in a different context. The context of original research is often referred to as the study site. Let us define measures for the study site as V_S.[2] We also have a policy site, or a context for which we need information (V_P), but have little to none of it. Ultimately, we derive estimates of V_{Pj} for policy site j from the outcomes of original research at study site i (V_{Si}). Study site values (V_{Si}) become transfer values (V_{Ti}) when applied to policy site j:

(1) $$V_{Si} \Rightarrow V_{Ti}$$

Original research provides content and context specific information regarding the policy site. This is because the target of original research is to address a specific need in a specific context. In the case of benefit transfer, $i \neq j$, or information for the policy site is derived from original research conducted at a different site. Benefit transfer provides content and context relevant information. The information transferred should be relevant to the policy site context. Only in rare circumstances will the transferred information be specific to the policy site. Specificity would occur only if the study site and policy site were identical. This chapter discusses how estimates of V_{Si} can be used to estimate V_{Pj}, or the method of benefit transfer.

Now that we have provided a formal definition and a historical context of benefit transfer, the next section will clarify some of the terms used above, such as point estimate transfer, benefit function transfer, and meta-regression analysis transfer.

2. MODELING AND APPLYING BENEFIT TRANSFER

Several benefit transfer methods have been developed to meet the need of defining estimates of V_{Pj}, or the value at policy site j. We broadly classify these approaches as: (1) value transfer and (2) function transfer. Value transfer entails the direct application of summary statistics from original research to a policy context. Function transfer entails the application of a statistical function that relates the summary statistics of original research to the specifics of the study site. We will discuss each of these categories of benefit transfer and identify various approaches for each.

2.1 Value Transfer

Value transfer is the direct application of original research summary statistics (such as per unit measures of willingness to pay (WTP), measures of elasticity, or other measures of marginal effects) to a policy site. There are essentially three approaches to conducting value transfers, including transfers of point estimates, transfers of measures of central tendency, and transfers of administratively-approved estimates.

2.1.1 Point Estimate Transfer

Point estimate transfer uses measure(s) of V_{Si}, given the context of study site i (Q_{Si}), to estimate the needed measure (V_{Pj}) for policy site j, given the context of the policy site (Q_{Pj}):

$$(2) \qquad V_{Pj} | Q_{Pj} = V_{Si} | Q_{Si}$$

Point estimate transfer typically uses a single measure. However, when possible, a range of estimates could be transferred to provide bounds on the probable value at the policy site. In addition, we recommend that confidence intervals be constructed around point estimate transfers when possible. This provides additional information regarding the precision of the study site measures.

Table 1 provides an overview of the steps in conducting a point estimate transfer. To illustrate this approach, we will discuss an actual application that John Loomis conducted for a federal agency. The U.S. Army Corps of Engineers was considering the removal of four dams on the Lower Snake River, from its confluence with the Columbia River upstream to Lewiston, Idaho. One of the benefits of this action would be the restoration of spawning habitat for native salmon populations, increasing the salmon population by about 47,500 fish according to fisheries biologists. While salmon have commercial and sport fishing use values, the agency was interested in the passive use value of the change in native salmon populations due to dam removal. For various reasons, the agency decided not to conduct original research. Therefore, it was decided to use benefit transfer to help inform their decision. You can judge whether the potential gain in accuracy from conducting original research would be worth the cost.

The policy context was defined as the Pacific Northwest, passive use value estimates for native salmon, a population increase of 47,471 fish, and estimates in per household annual WTP. Given the defined policy context, a literature search was conducted. When conducting a literature search for a benefit transfer, you should consider consulting experts, making inquiries on e-mail listserves, or writing to academic departments and government agencies in addition to keyword searches of electronic databases. These personal contacts are targeted at locating gray literature, such as theses and dissertations, unpublished working papers, conference proceedings, and agency reports. This gray literature may help overcome the problem of publication bias (Stanley 2001). Publication bias is the result of journal editors preferring to publish studies that show statistically significant results. The information you seek may not be provided in the published literature.

The literature search for the Snake River Study was conducted through keyword searches in American Economic Association's EconLit database and Environment Canada's (1998) Environmental Values Reference Inventory (EVRI) database. The latter database focuses on valuation studies, but has a cost associated with access. Three studies that provided values for salmon populations were found. Other studies located in the search were discarded due to their not providing value estimates, which is the information sought after in this benefit transfer exercise.[3] A fourth, unpublished study, was made available upon request. Already knowing the sparsity of original research on this subject, it was determined that this literature search pretty well exhausted the literature on passive use values for salmon.

Table 1. Steps in Conducting a Point Estimate Transfer

Step 1	Define the policy context. This definition should include various characteristics of the policy site, what information is needed, and in what units.
Step 2	Locate and gather original research outcomes. Conduct a thorough literature review, and obtain copies of potentially relevant publications.
Step 3	Screen the original research studies for relevance. How well does the original research context correspond to the policy context? Are the point estimates in the right units, or can they be adjusted to the right units? What is the quality of the original research?
Step 4	Select a point estimate or range of point estimates. This point estimate or range of point estimates should have the best fit out of the candidate estimates.
Step 5	Transfer the point estimate or range of point estimates. Aggregate the point estimate to the policy site context by multiplying it by the total number of units, providing a total value estimate for the good or service at the policy site.

Table 2 lists some of the characteristics of these studies. In addition to screening the studies for passive use value estimates, the studies were also screened regarding their overall quality. It was determined that each study was appropriate to the current need and was conducted using the appropriate valuation methodologies. Some of the following characteristics were used to

Table 2. Salmon Passive Use Value Studies

Study	Study Site Context	Increase in Salmon Population	Sample Frame	WTP per Household (in 1998 dollars)	Marginal Passive Use Value per Fish
Olsen, Richards, and Scott (1991)	Doubling of Columbia River Basin salmon and steelhead populations	2,500,000	ID, MT, OR, WA	$32.52	$163
Hanemann, Loomis, and Kanninen (1991)	Increasing salmon populations on the San Joaquin River in CA	14,900	CA	$221.96	$186,829
Loomis (1996a, b)	Increasing salmon populations on the Elwha River in WA due to dam removal	300,000	WA	$76.48	$3,197
Layton, Brown, and Plummer (1999)	Increasing anadromous fish populations for Eastern WA/Columbia River	1,000,000	WA	$119.04	$1,492
Layton, Brown, and Plummer (1999)	Increasing anadromous fish populations for Eastern WA/Columbia River	250,000	WA	$227.64	$11,420

determine whether the passive use values were appropriate to the policy context. All but one of the studies focused on increasing salmon populations in the Pacific Northwest. Each of the studies provided per household annual WTP value estimates.

Now that the necessary information was located, the value estimates to transfer need to be chosen. One approach would be to transfer a range of point estimates. This would result in annual household WTP ranging from a low of about $32 to a high of about $228. This range is quite large, illustrating one of the potential inaccuracies of conducting benefit transfer. Another approach would be to select a single point estimate based on judgments regarding goodness-of-fit between the study site and the policy site. One way of determining the match between the two sites is the change in anadromous fish populations. Obviously the Olsen, Richards, and Scott (1991) study is the least likely candidate study, having a change in fish population of 2.5 million compared to the policy site change of about 47,500 fish. The first estimate provided in the Layton, Brown, and Plummer (1999) study, based on a non-declining baseline change in fish population, is also not appropriate given the change of one million fish. The second Layton, Brown, and Plummer (1999) context (250,000 fish) and the Loomis context (300,000 fish) are closer fits, but still include significantly different population changes. The marginal value per fish may be lower for these population changes than the value for the policy site, resulting in a conservative estimate of the value for the policy site. The Hanemann, Loomis, and Kanninen (1991) study is the closest to the policy site context, with 14,900 fish. But given the smaller change in fish population, the marginal value per fish would tend to overstate the value of fish at the policy site (Table 2).

In this case, it was decided to transfer a range of point estimates in order to bracket the value at the policy site. The potential range in total passive use value for a change of 47,500 salmon due to dam removal on the Lower Snake River was then calculated and the extent of the market or the total number of households affected by the change in salmon population was estimated. Since passive use values can accrue to anyone, and since the agency is at the national level, a national accounting stance is warranted. However, the agency restricted the extent of the market to include only those households in the Pacific Northwest plus California. This is about 12.5 million households. Given this information, the total passive use value for the change in salmon population at the policy site could be calculated. The formula was: total passive use value = {[(annual household WTP * number of households) / (change in fish population at the study site)]*(change in fish population at the policy site)}. For example, using the Loomis value of $76.48 per household, the total passive use value for the change in salmon population at the policy site is about $151 million.[4] The other total passive use value estimates include $542 million using the Layton, Brown, and Plummer (1999) study and $8.9 billion using the Hanemann, Loomis, and Kanninen (1991) study. If the low value of $151 million, when combined with the other recreational benefits of dam removal (e.g., river recreation and gain in fishing values), is larger than the costs, then this benefit transfer exercise has contributed to answering the policy question and no further passive use value analysis would be needed. However, if more precision would change the decision of whether dam removal is economically efficient, then original research would need to be conducted.

2.1.2 Measure of Central Tendency Value Transfer

Another value transfer approach is the measure of central tendency transfer. This approach only differs from the point estimate transfer by taking an average or other measure of central tendency from several studies in the literature (Table 3). A measure of central tendency transfer entails using a mean, median, or other measure of central tendency based on all or a subset of original research outcomes. This approach can be defined as:

$$(3) \qquad V_{Pj} | Q_{Pj} = \overline{V}_S | \overline{Q}_S$$

where V_{Pj} is the measure needed for policy site *j* given the site's characteristics Q_{Pj}, and $\overline{V}_S$ is a measure of central tendency for all or a subset of study site measures reported in the literature given conditions at the study sites($\overline{Q}_s$). All study site measures should be adjusted to a common unit relevant to the policy site.

Continuing with our previous example of estimating total passive use for an increase of 47,500 salmon due to dam removal on the Lower Snake River, note that the first three steps are the same as the point estimate transfer (Table 3). Our next step is to calculate a measure of central tendency for all of the estimates. In this case, we are relying on information gained from all studies, disregarding criteria used in the previous example for inclusion or exclusion. We also want to convert each estimate to a common metric. The common metric we will use is the passive use value per fish. This is the square bracketed part of the equation used to estimate total passive use value above, or value per fish = [(annual household WTP * number of households) / (change in fish population at the study site)], as reported in Table 2.

The average of all estimates is $40,620 in passive use value per fish. The median value for all estimates is $3,197. The disparity between these two measures of central tendency is caused by the large value reported for the Hanemann, Loomis, and Kanninen (1991) study. If we had to choose between these two measures, we would go with the latter value. If we eliminate the

Table 3. Steps in Conducting a Measure of Central Tendency Transfer

Step 1	Define the policy context. This definition should include various characteristics of the policy site, what information is needed, and in what units.
Step 2	Locate and gather original research outcomes. Conduct a thorough literature review, and obtain copies of potentially relevant publications.
Step 3	Screen the original research studies for relevance. How well does the original research context correspond to the policy context? Are the point estimates in the right units, or can they be adjusted to the right units? What is the quality of the original research?
Step 4	Calculate the average value or other measure of central tendency for the point estimates. This average value should be based on those estimates that have the best fit out of the candidate estimates.
Step 5	Transfer the average value estimate. Aggregate the average value to the policy site context by multiplying it by the total number of units, providing a total value estimate for the good or service at the policy site.

Hanemann, Loomis, and Kanninen (1991) measure, the average value for the remaining four estimates is $4,068 in passive use value per fish. The median of these four estimates is $2,344. We will use the average value in this exercise to further illustrate the importance of calculating a confidence interval.

The average value of the four estimates has a standard error of 2,528. Therefore, the 95% confidence interval on this average value ranges from a low of -$887 to a high of $9,023. Not only is this range large, but the lower end is negative. Basically, the confidence interval is telling us that even though we are 95% confident the value lies in this range, the range is so large that it is meaningless. This total passive use value for a change of 47,471 fish is about $193 million. The same arguments that were made regarding the effect of the point estimate benefit transfer measure on the policy process can be made here. That is, if this estimate, when combined with other values associated with dam removal are significantly greater than or smaller than the costs, then the benefit transfer exercise has added information to the policy process. If a more precise estimate may change the implications of the information, then original research is warranted.

Aside from this example, we want you to be aware of another issue related to average value transfers. As we saw above, the confidence interval is very large. This is in part due to the small number of estimates and the broad range of these estimates. Based on large numbers theory, we would expect increased accuracy in average value transfers as the number of estimates reported in the literature increases. This would result in a higher probability that the average value from the literature is close to V_{Pj}. To illustrate this point, we will look at the literature for outdoor recreation use values (Rosenberger and Loomis 2001).

Let us compare confidence intervals for average use values of two outdoor recreation activities, big game hunting and non-motorized boating (Table 4). In the case of big game hunting, a large number of estimates exists in the literature. From 1967 to 1998, 177 use value estimates are reported. For this same time period, only 19 use value estimates are reported for non-motorized boating. The broad range of individual estimates is similar for both activities. The mean and median per person per activity day values for big game hunting are \$43 and \$37, respectively. The mean and median per person per activity day values for non-motorized boating are respectively \$62 and \$36. Estimates for both activities were adjusted to 1996 dollars. The standard error for big game hunting estimates is 2.21, resulting in a 95% confidence interval for the average value of \$39 to \$47. The standard error for non-motorized boating estimates is 13.76, resulting in a 95% confidence interval for the average value of \$35 to \$89. If the known value for each activity lies somewhere in its respective confidence interval based on the literature, then the precision of an average value transfer for big game hunting is probably better than for non-motorized boating.

Table 4. Use Value Estimates for Two Outdoor Recreation Activities, 1967-1998

Activity	Number of Estimates	Mean of Estimates	Median of Estimates	Standard Error of Mean	Range of Estimates
Big Game Hunting	177	\$43.17	\$37.30	2.21	\$15 to \$263
Non-motorized Boating	19	\$61.57	\$36.42	13.76	\$5 to \$209

Source: Rosenberger and Loomis (2001)

2.1.3 Administratively Approved Estimate Transfer

Administratively approved estimate transfer is arguably the simplest approach to benefit transfer. That is, if these measures are available and match your data needs. Federal public land agencies commonly use administratively approved estimates in assessing management and policy actions. The U.S. Forest Service has used RPA values since 1980 (Rosenberger and Loomis 2001). These RPA values are provided for groups of outdoor recreation activities and Forest Service regions of the country. Similarly, the U.S. Bureau of Reclamation and U.S. Army Corps of Engineers have relied upon the U.S. Water Resources Council's unit day values of recreation use for decades (U.S. Water Resources Council 1973, 1979, 1983).

Administratively approved estimates are derived from empirical evidence in the literature, expert judgment, and political screening. There are two main issues with using administratively approved estimates. First, the criteria used in the political screening process is unknown. This process may ignore some empirical evidence or use arbitrary adjustment factors. For example, we believe the 1990 RPA recreation values for the U.S. Forest Service were possibly reduced in the political screening process because of the following criteria. (1) Much of the original research on outdoor recreation activity use values was conducted for unique sites under unique conditions. This is a form of selection bias that tends to inflate estimates. Not all U.S. Forest Service recreation sites are unique or of high quality. The U.S. Water Resources Council attempted to deal with the site quality issue by providing marginal changes in value based on judgment regarding site quality. (2) Some original research studies inadequately accounted for the effects of substitute sites, which tends to lower values for the study site. (3) Special interests were involved in reducing the value of recreation in the planning process due to its link to budget allocations to different program areas.

A second issue is that administratively approved estimates are only updated every so often. Therefore, estimates may not reflect the latest empirical evidence. One distinct advantage of using administratively approved estimates for agency purposes is that the estimates have survived the political screening process.

The administratively approved transfer can be defined as:

$$V_{Pj} = V_{SA} \tag{4}$$

where V_{SA} is the administratively approved measure used to estimate the value at policy site j (V_{Pj}). Table 5 outlines the steps involved in conducting an administratively approved estimate transfer. We caution against using these measures solely on the basis that the U.S. Forest Service or U.S. Army Corp of

Engineers uses them. Understand how and why these values were developed, and then determine whether they meet your needs. Administratively approved estimates are developed to address certain agency needs, which may not match well with your needs.

Table 5. Steps in Conducting an Administratively Approved Estimate Transfer

Step 1	Define the policy context. This definition should include various characteristics of the policy site, what information is needed and in what units.
Step 2	Obtain administratively approved estimate. These estimates are typically published by an agency. Check with the relevant agency's policy or research division.
Step 3	Transfer the administratively approved estimate. Aggregate the estimate to the policy site context by multiplying it by the total number of units, providing a total value estimate for the good or service at the policy site.

2.1.4 How Good are Value Transfers?

The answer to the question, "How good are value transfers?" is largely unknown. This is basically because the actual value for a policy site is unknown; otherwise there would be no need for benefit transfer. If the best approximation of the actual value for a policy site is through original research, but we cannot conduct original research, then how close to the actual value is your benefit transfer? It would be like playing pin-the-tail-on-the-donkey, without the donkey. How can we know how close we were to the target, when no target exists? Therefore, in order to assess the validity or accuracy of benefit transfer, we need to know what the target is. Validity tests of benefit transfer include some original research.

Table 6 provides some empirical evidence on the relative validity of value transfers. Ignore the last column for now. We will return to this column when we compare the validity of value transfers with that of function transfers. The studies listed in Table 6 were conducted to address the validity of benefit transfer and involved an original research effort for the same good at multiple sites. The actual value for each site is estimated from site-specific models. The validity of benefit transfer is then tested by comparing the values across these different sites. That is, since we "know" the actual value at a particular site, how close are the transferred values if used in a benefit transfer?

Table 6. Empirical Tests of the Validity of Benefit Transfers.

Study Reference	Issue	Range of Error for Value Transfers	Range of Error for Function Transfers
Loomis (1992)	Fishing	4 - 39%	1 - 18%
Parsons and Kealy (1994)	Water quality	4 - 34%	1 - 75%
Loomis et al. (1995)	Recreation	----	1 - 475%
Bergland, Magnussen, and Navrud (1995)	Water quality	25 - 45%	18 - 41%
Downing and Ozuna (1996)	Fishing	0 - 577%	----
Kirchhoff, Colby, and LaFrance (1997)	Rafting	35 - 69%	2 - 210%
Kirchhoff (1998)	Recreation	----	1 - 7,028%
Brouwer and Spaninks (1999)	Agricultural land	27 - 36%	22 - 40%
Morrison and Bennett (2000)	Wetlands	4 - 191%	----
Rosenberger and Loomis (2000a)	Recreation	----	0 - 319%
VandenBerg, Poe, and Powell (2001)	Water quality		
Individual Sites		1 - 239%	0 - 298%
Pooled Data		0 - 105%	1 - 56%
Rosenberger and Phipps (2001)	Recreation	4 - 490%	2 - 62%
Shrestha and Loomis (2001)	International recreation	----	1 - 81%

Adapted from and expanded on Brouwer (2000).

The validity measures shown in the table were calculated as the percentage difference between the known value for the good at a particular site and the transfer value for that good. That is, whenever we conduct a benefit transfer, some associated error is bound to occur:

$$V_{Ti} = V_{Pj} + \delta_{ij} \tag{5}$$

where V_{Pj} is the known value for policy site j based on site-specific data, V_{Ti} is the transferred value from study site i, and δ_{ij} is the error associated with the transfer. This error is the difference between the known value and the transferred value. The summary validity statistics presented in Table 6 are the absolute percent error:

$$(6) \qquad \% \delta_{ij} = \left| \left(\left(V_{Ti} - V_{Pj} \right) / V_{Pj} \right) \right| * 100$$

A primary goal of benefit transfer is to provide accurate or valid information for situations in which it is needed. As such, we would like to minimize the generalization error associated with benefit transfer. We will discuss two validity tests of benefit transfer later.

Most of the value transfers indicated in Table 6 were for point estimate transfers. The value transfers can be either very precise or have significant error. This disparity is large, even though the formal tests were conducted in a semi-controlled context. That is, data quality, modeling quality, and analysts' judgments were held constant. Therefore, part of the disparity may be due to differences between the study site and policy site. One of the validity studies discussed later investigates the effect of the correspondence between the study site and policy site on the magnitude of the transfer error.

2.2 Function Transfer

Function transfers are more technically oriented than value transfers. They entail the transfer of functions or statistical models that define relationships between vectors of data collected at a study site. Some of these models were introduced previously in this book. The category of function transfers may be categorized as demand (or benefit or WTP) functions[5] or meta-regression analysis functions.

Function transfers are generally considered to perform better than value transfers. (See Table 6, although this may not be readily apparent.) This increased accuracy is because function transfers may be tailored to fit some of the characteristics of the policy site. Value transfers, on the other hand, are invariant to differences between the study site and the policy site.

2.2.1 Demand or Benefit Function Transfer

Demand or benefit function transfers are based on the premise that the study site estimate for site i (V_{Si}) is a function of characteristics of the study site context (Q_{Si}) (e.g., location, physical features, and climate) and other explanatory variables (X_{Si}) (e.g., sociodemographics, attitudes, and time):

$$(7) \qquad V_{Si} = f(Q_{Si}, X_{Si})$$

This book previously provided the reasons why and how to estimate these functions. This additional information may be used to take advantage of these relationships when conducting benefit transfer. Rather than relying solely on value transfers, precision may be gained from incorporating these relationships in benefit transfer. Basically, this is what John Loomis suggested in his 1992 paper (Loomis 1992). A value transfer requires a strong similarity between study sites and policy sites, which may not always be there. The invariance of value transfer measures to other relevant characteristics of a policy site makes these transfers insensitive or less robust to significant differences between the study site and the policy site. Therefore, we should be able to increase the precision of benefit transfer if we can tailor a function to fit the specifics of a policy site.

The beginning steps to conducting a demand or benefit function transfer are the same as for a value transfer, with the exception that additional information is required from publications. Some form of a function that models the statistical relationships between the summary measures of interest and characteristics of the original research effort, including characteristics of the study site and the study population must be reported if a study is to be used in a benefit function transfer. You will ultimately adjust this function to specific characteristics of the policy site, thereby estimating tailored values for the policy site. A near perfect match between the study site and policy site is not required, since you can potentially compensate for these differences in function transfers.

Table 7. Steps in Conducting a Demand or Benefit Function Transfer

Step 1	Define the policy context. This definition should include various characteristics of the policy site, what information is needed and in what units.
Step 2	Locate and gather original research outcomes. Conduct a thorough literature review and obtain copies of potentially relevant publications.
Step 3	Screen the original research studies for relevance. How well does the original research context correspond to the policy context? What is the quality of the original research? And most importantly, is a demand or benefit function provided?
Step 4	Gather summary data for the policy site. The demand or benefit function provided by original research has several independent or explanatory variables associated with it. Gather summary data on the policy site for as many of the variables in the model as possible.
Step 5	Predict the policy site benefit estimate by multiplying the summary statistics reflecting the policy site by the regression coefficients in the transfer function. This results in a tailored estimate for the policy site.
Step 6	Aggregate the tailored estimate. Aggregate the tailored estimate to the policy site context by multiplying it by the total number of units, providing a total value estimate for the good or service at the policy site.

The demand or benefit function transfer can be defined as:

$$(8) \qquad V_{P_j} = f_S(Q_{S|P_j}, X_{S|P_j})$$

The policy site measure ($V_{P|S}$) is derived from the study site function (f_S) adjusted to the characteristics of the policy site ($Q_{S|P}$ and $X_{S|P}$). This is why, in step 4 (Table 7), summary data are gathered on the policy site for as many of the independent, or explanatory, variables in the model as possible. This information is used to adjust the study site function to the policy site context.

Following is an example of a benefit function transfer. This example is simplistic[6], but it enables us to illustrate several issues: (1) application of benefit function transfers, (2) effect of multi-site data modeling on transfer accuracy, and (3) validity testing between value and function transfer approaches.

Assume that we are interested in obtaining an estimate of the value to a small Northeastern town for improving groundwater quality used for drinking to a very safe level. We conducted a literature search, discovering several groundwater quality improvement studies. The following study seems to provide the best fit for our valuation needs.

A case study by VandenBerg, Poe, and Powell (2001) (hereafter, VPP) estimates the benefits of improving groundwater quality used for drinking to a "very safe" level in twelve towns in New York, Massachusetts, and Pennsylvania. The study also tests the relative accuracy of two benefit transfer approaches. The authors used a contingent valuation survey with a payment card question format. Mean WTP per household per year was calculated for each of the twelve towns using the survey responses. This mean WTP is treated as the benchmark or known estimate (V_{Pj}) for each town j to which the transferred estimate (V_{Ti}) is compared. VPP used the estimate derived for each of the other eleven towns (study sites) as possible point measures to transfer to the twelfth town (the policy site).

To perform the benefit function transfer, a protocol first used by Loomis (1992) is employed whereby all of the survey data except for one town were pooled and a WTP equation was estimated. The independent variables of this function were then set at the levels for the nth, or excluded, town in order to predict mean WTP per household at this excluded town. In all, twelve benefit function models were estimated. Unfortunately, none of these models is reported in this source. You could request these models from the authors, obtain a copy of VandenBerg's thesis, or look for other references listed in the publication. VPP, however, did report a benefit function model based on pooling the data for all twelve towns, which we will use later in this example.

A sufficient number of explanatory variables that can be used to link and control for differences between the study and policy sites must be included. In the VPP example, explanatory variables included demographics such as education and income. This type of data, which is available from secondary sources, allows one to easily compensate for differences in demographics between the study site and the policy site. However, if this type of data is not part of the demand or benefit function at the study site, then you cannot compensate for any known differences between the two sites.

VPP also included a series of responses to risk perceptions, perceptions regarding current safety of water, and perceived likelihood of future contamination. These variables were statistically significant and contributed to the explanatory power of their models. However, they make real world benefit transfer difficult because the analyst would need to know the perception variables for the policy site. Typically, these perception variables are not available from secondary sources and would require a survey of households at the policy site. Benefit function transfer reduces the amount of information, and thus the type of survey that would need to be collected. However, this application illustrates one of the limitations of inferring confidence from benefit transfer validity testing studies. In real-world applications, you would like to have explanatory variables in the benefit transfer function that are available at both the study site and the policy site.

Another drawback of conducting benefit function or demand transfers is the implicit assumption that the statistical relationships (regression coefficients) between the dependent and independent variables at the study site are the same at the policy site. Empirical tests of this assumption show that it may be false, especially for inter-state or inter-region transfers (Downing and Ozuna 1996; Kirchhoff, Colby, and LaFrance 1997; Loomis 1992; Loomis et al. 1995; VandenBerg, Poe, and Powell 2001).

VPP provide a benefit (mean WTP per household per year) function based on pooling all the data for the twelve towns (Table 8). This model contains several dummy variables that would enable compensation for differences between the study and policy populations. No variables that identify physical characteristics of the sites exist. Most of the variables would require conducting a survey at the policy site in order to determine how its population differs from the populations behind the model. However, VPP do provide summary statistics for most of the variables for each town in the data set. Table 8, last two columns, shows our adjustment of this function to reflect characteristics of one of these towns. The policy site for this example is Horsham, PA.

The actual study site estimate of mean WTP per household per year, or V_{Pj}, for Horsham is \$67.45. This estimate is our known target value, which is not available in real-world benefit transfer situations. However, knowing the target value demonstrates how the benefit function transfer works. We will multiply the regression coefficients from the model by measures of the variables for the policy site to derive the partial WTP for each of the variables. See VPP for a full account of the different variables and site measures.

The first variable in the model, the constant or intercept term, is transferred in full to the policy site. This is done by multiplying it by one. The next two variables regarding perceptions of contamination illustrate a type of decision that might be necessary if a difference exists between how the variables are reported in the model and how the summary statistic is reported for the town. That is, the variables are dummy variables in the model for the *No*

Table 8. OLS Regression Model for Groundwater Quality Protection, Dependent Variable = Mean Willingness To Pay per Household per Year, n=667.

Variable	Regression Coefficient	Policy Site Measure	Partial WTP
Constant	-29.68	1	-29.68
Perception of Contamination (0, 1 = No Experience)	-23.48	1	-23.48
Perception of Contamination (0, 1 = Don't Know)	-26.96	0.24	-6.47
Likelihood of Future Contamination (0, 1 = Likely or Very Likely)	17.51	0.5	8.76
Likelihood of Future Contamination (0, 1 = Not Sure)	9.41	0.5	4.70
Interest in Community Water Issues (0, 1 = Mild to No Interest)	-20.66	0.5	-10.33
Interest in Community Water Issues (0, 1 = Interested)	-11.15	0.5	-5.58
Perceived Water Quality (0, 1 = Unsafe or Somewhat Safe)	29.92	0.5	14.96
Perceived Water Quality (0, 1 = Safe)	21.07	0.5	10.54
College Degree (0, 1 = Has College Degree)	-17.51	0.24	-4.20
Some College (0, 1 = Has Some College, But No Degree)	-15.72	0.24	-3.77
Average Risk Perception (3 question composite, 1= Safe to 5 = Unsafe)	9.91	4	39.64
Number of Perceived Potential Contamination Sources (number of sources)	2.56	3.17	8.12
Trust in Government and Organizations (9 question composite, 1 = Do Not Trust to 3 = Trust)	15.67	1.91	29.93
Household Income (dollars/year)	0.0008	45500	36.40
Total Mean WTP = ∑(Column 2 * Column 3) =			**$69.54**

Adjusted $R^2 = 0.15$, adapted from VandenBerg, Poe, and Powell (2001).

Contamination and *Don't Know* categories, whereas the summary statistic for the town is reported as a categorical variable recorded as 1 =*No Contamination*, 2=*Don't Know*, and 3=*Yes Contamination*. Since the summary statistic for the town is reported as 1.24, meaning *No Contamination* is the predominant category, we assume the full effect of the *No Contamination* variable and 24% of the effect of the *Don't Know* variable.

The next six variables in the model regarding Likelihood of Future Contamination, Interest in Community Water Issues, and Perceptions of Water Safety required similar assumptions on our part. These six variables are dummy variables that code for specific categories in the original variable, whereas the summary statistics for the policy site are the mean of a single variable. VPP report the summary statistic for the three variables for the town of Horsham to be 2.63, 2.54, and 2.63, respectively. These numbers are the means of categorical variables based on a 1 to 4 scale or 1 to 5 scale. Since the means are very close to the midpoint of the scales, we decided to keep half the effect of each of the dummy variables in the model. For example, we assume half the town is interested in water quality and half the town is not interested. Thus, in Table 8, each of these six regression coefficients are multiplied by 0.5, transferring half the effect of each dummy variable.

The rest of the variables in the model are relatively more straightforward in applying to the policy site. The college education variables do not perfectly match with the summary statistic reported for the policy site. The policy site reports that 48% of the town is college educated based on the survey responses. What proportion of this 48% has a college degree versus some college is unknown. Since the regression coefficients for the education dummy variables are similar, we decided to split this proportion in half, thus transferring 24% of each education variable for a total 48% effect of college education.

For the policy site, VPP report that the mean average risk perception is 4, the mean number of perceived potential contamination sources is 3.17, the mean trust is 1.91, and mean household income per year is $45,500 for Horsham. Thus, we multiply the respective coefficients by these reported levels for the policy site to obtain the partial WTP effects of these variables.

Summing all of the partial effects listed in the last column of Table 8 results in $69.54 in total mean WTP per household per year for the policy site town of Horsham, PA. The actual mean WTP per household per year for the town of Horsham, PA is $67.45, based on site specific data. Our benefit function transfer estimate is well within a 95% confidence interval for the actual value for the site. The percent error associated with our transfer measure (equation (6)) is about 3%.

VPP also tested the relative accuracy of benefit function transfers as compared to value transfers using individual site data and the n-1 pooled data approaches. The range for value transfers is 1% to 239% error, whereas the error range from 0% to 298% for function transfers when modeling the data at the individual site level (Table 6). The mean percent error is 42% for the value transfers and 44% for the function transfers. However, if the n-1 pooled data modeling approach is used, the range for value transfers and function transfers narrow to 0 - 105% and 1 - 56%, respectively (Table 6). The mean percent error for benefit function transfers falls to 18%, which is significantly different than the mean percent error of 31% for the value transfer approach. This difference indicates that greater variability in the data results in more robust function transfer models and less error. This benefit is basically due to the increased ability to compensate for measured differences between the two sites. VPP also report that function transfers outperformed value transfers less than half of the time using the individual site data modeling approach. However, when the data are pooled, then the function transfers outperformed the value transfers 75% of the time.

An innovative test used by VPP determines if grouping the data by some specific characteristics would improve the accuracy of their benefit transfer. Using the n-1 pooled data approach, which was proven to be superior to the individual site data modeling approach, VPP found some improvements in benefit transfer accuracy. Two ways that they grouped the data were by the state in which the town is located and by whether the town had previous contamination or not. Benefit transfers were then restricted to only within the group. For example, an estimate generated from an n-1 pooled data model for Pennsylvania was not transferred to a New York town. VPP found an improvement in the value transfer from 31% mean error reported above, to 22% mean error when restricted to within state groupings. The contamination groupings resulted in less accuracy for value transfers. Function transfer accuracy, however, was improved from 18% in the pooled data model to 16% mean error when sites were grouped by previous contamination.

2.2.2 Meta-Regression Analysis Function Transfer

Another function transfer approach that is gaining application is the meta-regression analysis function transfer. Demand or benefit function transfers rely on statistical relationships defined for certain variables based on a single study. Meta-regression analysis summarizes and synthesizes outcomes from several studies. There are essentially two approaches to meta-regression analysis: (1) pooling the actual data from multiple studies and (2) using summary statistics, such as value estimates, from multiple studies. We will focus on the more prevalent latter approach.

Meta-regression analysis enables us to overcome some of the issues related to demand or benefit function transfers. Namely, we may be able to statistically explain the variation found across empirical studies (as we found in the salmon passive use example). Some of this variation in value estimates may be due to identifiable characteristics among the different studies themselves, such as valuation method, survey mode, geographic location, and so forth. These characteristics are not explanatory variables in the original studies because they define the context of original research and are, therefore, constant in original research. Meta-regression analysis may be able to discern the individual effects some of these variables may have on the individual research outcomes.

Meta-regression analysis traditionally has been concerned with understanding the influence of methodological and study-specific factors on research outcomes and providing summaries and syntheses of past research (Stanley and Jarrell 1989). The first two meta-analyses on environmental and natural resource economic studies were by Smith and Kaoru (1990) on travel cost studies of recreation benefits and by Walsh, Johnson, and McKean (1989, 1992) on outdoor recreation benefit studies. More recent applications of meta-analysis for similar purposes include groundwater (Boyle, Poe, and Bergstrom 1994), air quality and associated health effects (Smith and Huang 1995; Smith and Osborne 1996; Desvousges, Johnson, and Banzhaf 1998), value of statistical life (Mrozek and Taylor 2002), endangered species (Loomis and White 1996), price elasticities for water (Espey, Espey and Shaw 1997), recreational fishing (Sturtevant, Johnson and Desvousges 1998), wetland values (Woodward and Wui 2001), and outdoor recreation benefit estimates (Rosenberger and Loomis 2001).

The dependent variable in a meta-regression analysis is a summary statistic from each individual study, such as a value estimate, elasticity, or other measure. The independent or explanatory variables are characteristics of the model, survey design, and data of the original studies. We also may be able to explain some of the inter-study variation in research outcomes by modeling the effects of characteristics that are typically held constant within an individual study, such as valuation methodology, survey mode, time, and physical attributes of the study site.

A basic premise of meta-regression analysis is the existence of an underlying valuation function, of which original research studies are independent random draws. Woodward and Wui (2001) note that this premise probably is false. The draws are not random because a reason exists for conducting original research on some sites and not others (selection bias). Peer-review screening for statistically significant results in journals (publication bias) is also an issue. The draws probably are not independent due to multiple estimates from single studies or from researchers who work closely together. There also is the potential for autocorrelation due to learning effects and improvements in methodology over time. However, a recent attempt at estimating this underlying valuation function using meta-regression analysis suggests that these potential biases may be small (Rosenberger and Phipps 2001).

The steps to conducting a meta-regression analysis are adapted from Stanley (2001) (Table 9). We will illustrate the application of meta-regression analysis by summarizing one that we conducted as we updated the U.S. Forest Service RPA values for outdoor recreation (Rosenberger and Loomis 2001). Another excellent example of a meta-regression analysis, including the subjective decisions that must be made, is Woodward and Wui's (2001) analysis of wetland valuation studies.

Our project consisted of two primary objectives: to update the RPA values based on the past decade of empirical evidence and to develop a meta-regression analysis transfer model for outdoor recreation use values. Thus, our policy context was defined as outdoor recreation use value studies conducted in the United States and Canada. Since we were updating previous literature reviews, we focused on outdoor recreation valuation studies conducted from 1988 to 1998.

We conducted keyword searches in a wide range of electronic databases, such as American Economic Association's EconLit, First Search, the University of Michigan's Dissertation and Master's Abstracts, and Water Resources

Table 9. Steps in Conducting a Meta-Analysis Function Transfer.

Step 1	Define the policy context. This definition should include various characteristics of the policy site, what information is needed and in what units.
Step 2	Develop a standard database structure. Conduct a thorough literature review and obtain copies of potentially relevant publications. Develop a master coding strategy that allows you to consistently code as much information as possible regarding each study. This information includes the dependent and independent variables in the original analysis, methodological and other study characteristics, source of the study, and authors of the study.
Step 3	Screen the original research studies for relevance. Reduce the literature search outcomes to include those studies containing relevant empirical estimates, tests, or findings.
Step 4	Choose and reduce the summary statistic to a common metric. The summary statistic would be the primary information need for the policy site. Reduction to a common metric may include reducing all empirical estimates to the same unit (e.g. annual basis). This summary statistic will serve as the dependent variable in the meta-analysis regression.
Step 5	Choose the independent variables. These variables are those characteristics of the individual studies that you hypothesize are important, or consequential to differences in the summary statistics. Forward thinking when designing the master code sheet in step 2 reduces the chances that you will have to code variables as you think of them. Note that if you do not code a variable, then you cannot include it in your regression analysis.
Step 6	Conduct the meta-regression analysis. The summary statistic will serve as the dependent variable, and the independent variables will serve as the explanatory variables. The purpose of meta-regression analysis is to explain the variation in the dependent variable across studies. Standard econometric issues are relevant here.
Step 7	Gather summary data for the policy site. The meta-regression analysis model has several associated independent variables. Gather summary data on the policy site for as many of the variables in the model as possible.
Step 8	Predict the policy site summary statistics from the meta-regression model by multiplying the summary statistics reflecting the policy site by the regression coefficients in the transfer function. This results in a tailored estimate for the policy site.
Step 9	Aggregate the tailored estimate. Aggregate the tailored estimate to the policy site context by multiplying it by the total number of units, providing a total value estimate for the good or service at the policy site.

Adapted from Stanley (2001).

Abstract Index. Gray literature was searched using conference proceedings, Carson et al.'s (1994) contingent valuation method bibliography, as well as our own collections of working papers and journal article reprints. Papers were screened for relevance (i.e., whether they contained use value estimates for outdoor recreation activities with study sites in the United States or Canada). We did not eliminate any studies based on our subjective judgment regarding the quality of their methods, models, or data. Stanley (2001) recommends that you err on the side of inclusion, thus letting the meta-regression analysis estimate how research choices influence values. By including all studies, we also protect the variability associated with value estimates evident in the empirical literature. Our database consisted of 760 estimates from 163 separate studies conducted from 1967 to 1998 in the United States and Canada.

We developed a master coding sheet containing 126 fields covering six categories. (1) We coded the complete citation for each study. (2) We coded the primary summary statistic that would become the dependent variable in our regression analysis. This summary statistic is the per person per activity day use value estimate from each study. A per person per activity day use value is the value for a person engaged in a recreation activity for any part of a day. We reduced all value estimates to this common metric by adjusting for inflation to 1996 dollars using the implicit price deflator and for disparity in units such as per trip, per season, or per year estimates. Some of the studies did not provide enough information for adjusting the value estimate reported in the study and were subsequently dropped from our analysis.

We also coded all value estimates reported in a single study. Stanley (2001) suggests that an average of all estimates from a single source be calculated. This reduces the potential bias associated with weighting a single study higher than other studies in the database. For example, if you have twenty studies that report one estimate each, and one study that reports twenty estimates, then the single study reporting multiple estimates carries half the weight for the entire sample. To reduce this potential bias, you could take the average of the latter study, reducing its weight to be equal to all other studies included in the database. We choose not to follow this procedure because some of the studies in our database were estimating values for the same activity, but in different regions of the country. Other studies reported multiple value estimates, but for different activities. We choose to statistically test for and model this potential effect, which we will discuss shortly.

(3) We coded different attributes of the value estimates, including if they were WTP or willingness-to-accept (WTA) estimates, and median or mean estimates. Most values coded in the database were mean WTP measures. We did not code the estimates based on ordinary or Hicksian measures because the difference between these measures is expected to be insignificant for recreation (Bergstrom 1990). Unfortunately, since we did not code for this characteristic, we could not test for systematic differences between these types of value estimates. An effect can not be tested if it was not coded. The sheer size of our database prohibited adding the code for each.

(4) We coded all aspects of the valuation methodology employed in each study, including whether it was a travel cost model, random utility model, contingent valuation model, or attribute-based model. Within each category, we coded survey mode (phone, mail, in-person survey, etc.), value elicitation method (dichotomous choice, payment card, etc.), functional form, substitute sites, travel time, and so on. (5) We coded characteristics of the study site including public or private ownership type of public land, environment type (wetland, forest, lake, etc.), and type of recreation activity (twenty two separate activities in all). (6) We coded characteristics of the sample population, including average age, income, education, etc.

Next we conducted our regression analysis, where the valuation estimates adjusted for units and inflation served as our dependent variable, and the other variables served as our independent variables. The purpose of the regression analysis is to explain as much of the variation in the dependent variable across the studies as possible using the independent variables. Issues of multicollinearity, heteroskedasticity, and autocorrelation are relevant here. The procedure we used was ordinary least squares, after testing for and affirming a normal distribution for the dependent variable. Many of the variables could not be included in the regression analysis due to our inability to code them for all studies. For example, most of the studies did not report any summary statistics regarding the characteristics of their sample population such as mean income, age, education, etc. These variables may be important to explaining some of the variation in the value estimates, but we cannot test for this if the information is not provided in the original study. Our sample size was reduced to 701 value estimates. We refer you to Chapter 3 for suggestions regarding reporting of results for original research. The quality of original research and reporting of its results are constraints on the ability to conduct valid benefit transfer.

We used a backward elimination procedure to specify our meta-regression model. This procedure eliminates variables one-by-one that are insignificant in explaining the variation in the dependent variable. The t-statistics were based on White's standard error of the regression coefficients which are corrected for heteroskedasticity and serial correlation. A significance level of 0.20 was used as the cut-off point for inclusion in the final model. The effect of dropping many variables may represent minor modeling choices that remain as part of the random study-to-study background (Stanley 2001). Another econometric issue tested for was within study correlation, a single study accounted for more than one value estimate in the database. This can result in an unequal weighting and a bias in the regression coefficients. This effect was tested for by modeling the data as panel data, where each panel is composed of various commonalities such as source, researcher, etc. No significant effect of this sort was found in the database (Rosenberger and Loomis 2000b).

This model is reported in Table 10 (ignore the last two columns for the moment). More detail regarding the meta-regression analysis model may be found in Rosenberger and Loomis (2001). The first thirteen variables code for the effects of valuation method, modeling choice, and survey mode. Other candidate variables in these categories were not found to be significant in the model. The TREND variable is coded, categorically with one equal to the first year of data (1967), two equal to the second year (1968) up to thirty two for the last year of the data (1998). The next nine variables are code for location effects, such as region of the country (proxied by U.S. Forest Service management regions) and type of environmental surroundings (lake, river, forest). The last 10 variables are dummy variables identifying the type of recreation activity. The model explains 27% of the variation in the dependent variable, which is typical of large meta-analysis databases (Walsh, Johnson, and McKean 1992; Smith and Kaoru 1990; Woodward and Wui 2001). Other site characteristics likely influence the value estimates, such as sample population demographics and site-specific physical attributes. But those variables were either not reported or not identified in most of the studies.

Up to this point, meta-regression analysis provides a statistical summary or quantitative review of the literature. Since the analysis provides the influence of coded variables or study characteristics on value estimates, can we use the meta-analysis regression as a benefit transfer function?[7] Some research has tested the application of meta-regression analyses for the purpose of benefit transfer (Desvousges, Johnson and Banzhaf 1998; Kirchhoff 1998; Sturtevant,

Table 10. OLS Meta-Regression Model for Outdoor Recreation Use Values, Dependent Variable = Consumer Surplus per Person per Activity Day, N = 701.

Variable	Coefficient	Policy Site Measure	Partial Use Value
Constant	81.27	1	81.27
Method (0, 1=CVM)	-21.59	0.636	-13.73
DCCVM (0,1=Dichotomous Choice CVM)	-36.98	0.177	-6.54
OE (0,1=Open Ended CVM)	-51.76	0.354	-18.32
ITBID (0,1=Iterative Bidding CVM)	-46.40	0.096	-4.45
SPRP (0, 1=Combined Stated and Revealed Preference)	-57.80	0.006	-0.35
PAYCARD (0,1=Payment Card CVM)	-83.19	0.006	-0.50
CONJOINT (0,1=Conjoint Method)	-74.03	0.001	-0.07
PHONE (0,1=Phone Survey)	-15.25	0.495	-7.55
INDIVID (0,1=Individual Travel Cost)	-40.15	0.153	-6.14
ZONAL (0,1=Zonal Travel Cost)	-55.70	0.185	-10.30
RUM (0,1=Random Utility Model)	-58.42	0.027	-1.58
SUBS (0,1=Included Substitute Sites)	-17.62	0.264	-4.65
VALUNIT (0,1=Per Activity Day Units)	-9.07	0.412	-3.74
TREND (Year Results Reported; 1=1967...1998=32)	0.98	35	34.30
FSADMIN (0,1=National Forest Site)	-17.82	1	-17.82
FSR1 (0,1=Forest Service Region 1; Montana, Idaho)	11.41	0	0
FSR4 (0,1=Forest Service Region 4; Nevada, Utah, So. Idaho)	5.53	0	0
FSR6 (0,1=Forest Service Region 6; Oregon, Washington)	-10.84	0	0

Variable	Coefficient	Policy Site Measure	Partial Use Value
FSR8 (0,1=Forest Service Region 8; southeastern U.S.)	-5.13	0	0
LAKE (0,1=Lake Site)	-18.29	0	0
RIVER (0,1=River Site)	16.79	0	0
FOREST (0,1=Forested Site)	-9.16	1	-9.16
PUBLIC (0,1=Public Site)	13.31	1	13.31
SWIM (0,1=Swimming)	-15.51	0	0
OFFRD (0,1=Off-Road Vehicle)	-17.34	0	0
NOMTRBT (0,1=Nonmotorized Boating)	13.81	0	0
BIKE (0,1=Bicycling)	-14.31	1	-14.31
XSKI (0,1=Crosscountry Skiing)	-5.94	0	0
SNOWMOB (0,1=Snowmobiling)	-20.92	0	0
BGHUNT (0,1=Big Game Hunting)	15.39	0	0
WATFOWL (0,1=Waterfowl Hunting)	9.89	0	0
FISH (0,1=Fishing)	7.06	0	0
ROCKCL (0,1=Rockclimbing)	62.03	0	0
Total Use Value/Person/Activity Day = ∑(Column2 * Column3)			**$9.67**

Adjusted $R^2 = 0.27$.

Johnson, and Desvousges 1998; Rosenberger and Loomis 2000a, 2001; Rosenberger and Phipps 2001).

We define a meta-regression analysis transfer function as follows:

$$(9) \qquad V_{P_j} = f_S(Q_{S|P_j}, \overline{X}_{S|P_j}, M_{S|P_j})$$

Equation (9) states that the value for policy site j (V_{P_j}) is a function of data included in or distinguishable from each study site i. The other variables can be quantity/quality variables (Q), socio-demographic variables (e.g., income, age, and education) and site characteristics (e.g., presence of water, geographic location, and land type) (X), and methodological variables (e.g., valuation method, modeling format, and functional form) (*M*) for each study (*i*). The application of this model to benefit transfer is similar to adjusting the benefit function discussed in the previous section. Value estimates tailored to a policy site may be obtained by adjusting the meta-regression analysis function to specific characteristics of that site.

Let us review a hypothetical example of using the meta-regression analysis provided in Table 10 (Rosenberger and Loomis 2001). Assume that a local mountain biking club is soliciting the U.S. Forest Service to construct a mountain biking trail in the Allegheny National Forest in Pennsylvania. The U.S. Forest Service is interested in whether the benefits derived from use of this trail would be worth the construction and maintenance costs. The Forest Service is seeking an estimate of the total annual use value for mountain biking on the proposed trail, but it will not fund an original research project because the overall impacts of the trail are expected to be small.

We can start by searching the literature for use value estimates for mountain biking. We find one study that specifically estimated the use value for mountain biking. However, this study focused on mountain biking on a world-class trail in Moab, Utah (Fix and Loomis 1998). These results are probably not transferable to other sites given the unique quality of study site. We decide to use the meta-regression analysis model for outdoor recreation use value estimates provided in Table 10.

Adjustments made to the meta-analysis function are intended to tailor the model to specific characteristics of the policy site (Table 10, third column). Therefore, some measures are required for the policy site. However, other judgments must be made in tailoring this model to the policy site. For example, how do you adjust for methodological influences, such as the first thirteen variables in the meta-analysis function? One way is to hold the effect of these variables constant, by adjusting the variables average value in the database. For example, the effect of METHOD in Table 10 is moderated by accounting for the presence of this variable in the database. Multiplying it by 0.636, which means about 64% of the studies were contingent valuation studies in the database, holds the effect of this variable constant at its mean value. Of course, measures for the variables in the database probably are known as a result of the meta-regression analysis. Any unidentified measures must be found or a judgment must be made regarding the magnitude of their effect on the transfer.

We set TREND to 35, reflecting that the coefficient estimates a $0.98 increase per year from 1967 onward. We chose 35 because the year 2001 would be number 35 in this series. The rest of the variables are dummy variables either coding for certain characteristics or for specific recreation activities. Because the policy site is mountain biking in the Allegheny National Forest (this is in U.S. Forest Service Region 9), we accept the full effect of the FSADMIN, FOREST, PUBLIC, and BIKING variables by multiplying them by one. All other dummy variables do not reflect characteristics of the policy site. Therefore, we do not accept any of their effects and multiply them by zero. Summing all of the partial values reported in Table 10, last column, we estimate the per person per activity day use value for mountain biking in the Allegheny National Forest to be $9.67. In the final step, we multiply this per unit value estimate by the total projected visitor days for the proposed trail to determine the total use value for the trail.

Confidence in this transfer estimate may be illustrated by comparing it to empirical estimates from the literature. Fix and Loomis (1998) estimate the use value for mountain biking on the Slickrock trail in Moab, Utah, to be about $60 per person per activity day. Our transfer measure is not unreasonable because their measure is for such a high profile, unique site. The updated RPA value (Rosenberger and Loomis 2001) for generic biking is about $35 based on a single study estimate for rail-trail use. Therefore, we conclude that our measure is a conservative estimate. If additional precision gained from original research would not significantly change the decision on whether to construct the trail, then our benefit transfer has served its purpose by informing the project decision.

2.2.3 "How Good are Function Transfers?"

Empirical evidence suggests that function transfers are more precise than value transfers (Table 6). The validity measures for function transfers provided in Table 6, last column, were estimated using the same formula as used for value transfer validity measures (Equation (6)). The relative improvement is primarily due to being able to tailor benefit functions to specifics of the policy site. In addition, evidence suggests that meta-regression analysis transfers perform better than demand or benefit function transfers (Brouwer and Spaninks 1999; VandenBerg, Poe, and Powell 2001; Rosenberger and Phipps 2001). However, misapplications of meta-regression analysis transfers can result in significant error (Kirchhoff 1998).

In an innovative experiment, Randy Rosenberger showed how measures of site characteristics (e.g., physical and sample population attributes) that are invariant within a single study, but vary across multiple studies, can account for much of the error associated with traditional benefit transfer (Rosenberger and Phipps 2001). These effects can be modeled by applying meta-regression analysis, thus providing a meta-analysis transfer function that significantly improves the accuracy of benefit transfer. This experiment reduced the error of value transfers and n-1 pooled data function transfers from an average of 140% error to 20% error in the meta-regression analysis benefit transfer.

3. CONCLUDING REMARKS

This chapter was intended to expose you to the world of benefit transfer. The method of benefit transfer is described along with example applications. We hope that your understanding is now sufficient to critique benefit transfer or even conduct your own benefit transfer. Many additional example applications are referenced in Table 6. In conclusion, we want to leave you with some of our thoughts and suggestions on the future direction of benefit transfer methodology.

We need to continue to learn about benefit transfer through the continued use and documentation of benefit transfer. Many benefit transfer protocols are suggested in the literature. Yet no universally accepted systematic set of protocol for the application of benefit transfer has been designed. A systematic set of protocol would potentially reduce the variability in benefit transfer applications.

We, as benefit transfer practitioners, need better access to primary research data. Toward this end, we need to support values databases or data warehouses, similar to the effort of the Environmental Valuation Reference Inventory (Environment Canada 1998) and our recreation database (Rosenberger and Loomis 2001).

We need to develop methods for verifying and calibrating transfer data. Brouwer (2000) suggests using stakeholder involvement methods for verifying transfer data. One way this could be done is to conduct a low cost, broad survey of the affected population. This survey would provide an indicator of the affected population's characteristics, thus confirming or denying benefit transfer data. Feather and Hellerstein (1997) provide a calibration method that can be used to adjust (calibrate) aggregate data based on information from subsets of individual data. Smith, van Houtven, and Pattayanak (1999) illustrate the use of a generalized method of moments technique for calibrating transfer functions.

We need more primary research studies that specifically target benefit transfer (Bergland, Magnussen, and Navrud 1995) or that are specifically designed for future benefit transfer applications. This can include more research on meta-analysis function transfers. These studies would provide information that could increase our ability to critique and improve benefit transfer methodology. We foresee improvement in the quality and an increase in the number of meta-regression analyses being conducted in the future. This will be greatly due to a desire to better understand our ever-increasing body of knowledge as more empirical evidence is recorded on nonmarket values.

We need to explore the use of other technologies in benefit transfer process. For example, Bateman, Lovett, and Brainard (1999); Eade and Moran (1996); and Lovett, Brainard, and Bateman (1997) are investigating the use of Geographic Information Systems (GIS) technology in the benefit transfer process. Rosenberger and Phipps (2001) illustrate the potential accuracy gains from incorporating site characteristics in meta-regression analyses.

We need to improve our ability to transfer information from one census group or culture to another (Krupnick et al. 1996). Can the literature on health values for adults inform the health values for children? Can information derived from one racial group be transferred to another? Can information be transferred internationally or cross-culturally? Shrestha and Loomis (2001) provide evidence for the ability to conduct international benefit transfer, at least for some recreation activities.

NOTES

1 The Government Performance and Results Act of 1993 superceded some previous federal legislation requiring formal cost-benefit analyses of federal programs.

2 *V* is used to denote value information or data and can consist of measures of benefits or costs, resource quantities or qualities, population characteristics, and other relevant information such as elasticities, dose-response effects, regression coefficients, and t-values.

3 Studies that do not report any data or insufficiently report data may not be of use. Other factors can include a poor match between your data needs for your policy site context (what is affected and how impacts are measured) and the context of the study site data. Boyle and Bergstrom (1992) describe how some of the data may not be

relevant for your needs.

4 The calculation is: Total Passive Use Value = {[($76.48 per household * 12,541,700 households)/(300,000 fish at study site)] * (47,461 fish at the policy site)} = $151,743,990.

5 Other functions include dose-response or production functions, especially prevalent in the health sciences literature.

6 Another reason this example is simplified is that it deals with a benefit function, which is a direct estimation method. As such, it directly models the relationship between WTP and independent variables. Other models, such as demand models, may not be as easily adjusted or may not be amenable to adjustment depending on how the models are developed, including functional form (Adamowicz, Fletcher, and Graham-Tomasi 1989).

7 The potential use of meta-regression analysis in defining benefit transfer functions is like the holy grail of benefit transfer: developing a function that can be used to estimate different types of values for different policy contexts. That is, even in conditions where no point estimates or demand functions are reported in the literature, a meta-regression analysis function may be able to provide such estimates or functions.

REFERENCES

Adamowicz, W. L., Fletcher, J. J., and Graham-Tomasi, T. 1989. Functional Form and the Statistical Properties of Welfare Measures. *American Journal of Agricultural Economics* 71(2):414-421.

Bateman, I. J., A. A. Lovett, and J.S. Brainard. 1999. Developing a Methodology for Benefit Transfers Using Geographical Information Systems: Modeling Demand for Woodland Recreation. *Regional Studies* 33(3):191-205.

Bergland, O., K. Magnussen, and S. Navrud. 1995. *Benefit Transfer: Testing for Accuracy and Reliability*. Discussion Paper #D-03/1995. Norway: Dept. of Economics and Social Sciences. 21p.

Berstrom, J. C. 1990. Concepts and Measures of the Economic Value of Environmental Quality: A Review. *Journal of Environmental Management* 31(3):215-228.

Boyle, K. J., and J. C. Bergstrom. 1992. Benefit Transfer Studies: Myths, Pragmatism, and Idealism. *Water Resources Research* 28(3):657-663.

Boyle, K., G. Poe, and J. Bergstrom. 1994. What Do We Know about Groundwater Values?: Preliminary Implications from a Meta Analysis of Contingent-Valuation Studies. *American Journal of Agricultural Economics* 76(Dec):1055-1061.

Brouwer, R. 2000. Environmental Value Transfer: State of the Art and Future Prospects. *Ecological Economics* 32(1):137-152.

Brouwer, R., and F. A. Spaninks. 1999. The Validity of Environmental Benefits Transfer: Further Empirical Testing. *Environmental and Resource Economics* 14:95-117.

Carson, R. T., J. L. Wright, N. Carson, A. Alberini, and N. Flores. 1994. *A Bibliography of Contingent Valuation Studies and Papers.* California Natural Resources Damage Assessment, Inc.

Desvousges, W. H., F. R. Johnson, and H. S. Banzhaf. 1998. *Environmental Policy Analysis with Limited Information: Principles and Applications of the Transfer Method.* Massachusetts: Edward Elgar. 244 p.

Downing, M., and T. Ozuna Jr. 1996. Testing the Reliability of the Benefit Function Transfer Approach. *Journal of Environmental Economics and Management* 30(3):316-322.

Eade, J. D. O., and D. Moran. 1996. Spatial Economic Valuation: Benefits Transfer Using Geographical Information Systems. *Journal of Environmental Management* 48(2):97-110.

Environment Canada. 1998. *Environmental Valuation Resource Inventory homepage.* Available: www.evri.ec.gc.ca/evri.

Espey, M., J. Espey, and W. D. Shaw. 1997. Price Elasticity of Residential Demand for Water: A Meta-analysis. *Water Resources Research* 33(Jun):1369-1374.

Feather, P., and D. Hellerstein. 1997. Calibrating Benefit Function Transfer to Assess the Conservation Reserve Program. *American Journal of Agricultural Economics* 79(1): 151-162.

Fix, P., and J. Loomis. 1998. Comparing the Economic Value of Mountain Biking Estimated Using Revealed and Stated Preference. *Journal of Environmental Planning and Management* 41(2):227-236.

Freeman, A. M., III. 1984. On the Tactics of Benefit Estimation under Executive Order 12291. In *Environmental Policy under Reagan's Executive Order: The Role of Benefit-cost Analysis.* Edited by V. K. Smith. Chapel Hill, NC: The Univ. of North Carolina Press.

Hanemann, M., J. Loomis, and B. Kanninen. 1991. Statistical Efficiency of Double Bounded Dichotomous Choice Contingent Valuation. *American Journal of Agricultural Economics* 73(4):1255-1263.

Kirchhoff, S. 1998. *Benefit Function Transfer Vs. Meta-analysis as Policy-making Tools: A Comparison.* Paper presented at the workshop on Meta-Analysis and Benefit Transfer: State of the Art and Prospects. Amsterdam: Tinbergen Institute. April 6-7, 1998.

Kirchhoff, S., B. G. Colby, and J. T. LaFrance. 1997. Evaluating the Performance of Benefit Transfer: an Empirical Inquiry. *Journal of Environmental Economics and Management* 33(1):75-93.

Krupnick, A., K. Harrison, E. Nickell, and M. Toman. 1996. The Value of Health Benefits from Ambient Air Quality Improvements in Central and Eastern Europe: An Exercise in Benefits Transfer. *Environmental and Resource Economics* 7(4):307-332.

Layton, D., G. Brown, and M. Plummer. 1999. *Valuing Multiple Programs to Improve Fish Populations.* Davis, CA: .Dept. of Environmental Science and Policy, University of California.

Loomis, J. 1992. The Evolution of a More Rigorous Approach to Benefit Transfer: Benefit Function Transfer. *Water Resources Research* 28(3):701-705.

Loomis, J. 1996a. Measuring the Economic Benefits of Removing Dams and Restoring the Elwha River: Results of a Contingent Valuation Survey. *Water Resources Research* 32(2):441- 447.

Loomis, J. 1996b. How Large is the Extent of the Market for Public Goods: Evidence from a Nationwide Contingent Valuation Survey. *Applied Economics* 28:779-782.

Loomis, J., B. Roach, F. Ward, and R. Ready. 1995. Testing the Transferability of Recreation Demand Models Across Regions: a Study of Corps of Engineers Reservoirs. *Water Resources Research* 31(3):721-730.

Loomis, J. B., and D. White. 1996. Economic Benefits of Rare and Endangered Species: Summary and Meta-Analysis. *Ecological Economics* 18(2):197-206.

Lovett, A. A., J. S. Brainard, and I. J. Bateman. 1997. Improving Benefit Transfer Demand Functions: a GIS Approach. *Journal of Environmental Management* 51(4):373-389.

Morrison, M. and J. Bennett. 2000. Choice Modeling, Non-use Values Benefit Transfers. *Economic Analysis and Policy* 30(1):13-32.

Mrozek, J. R., and L. O. Taylor. 2002. What Determines the Value of Life? A Meta- Analysis. *Journal of Policy Analysis and Management* 21(2):253-270.

Olsen, D., J. Richards, and R. D. Scott. 1991. Existence and Sport Values for Doubling the Size of Columbia River Basin Salmon and Steelhead Runs. *Rivers* 2(1):44-56.

Parsons, G. R., and M. J. Kealy. 1994. Benefits Transfer in a Random Utility Model of Recreation. *Water Resources Research* 30(8):2477-2484.

Rosenberger, R. S., and J. B. Loomis. 2000a. Using Meta-analysis for Benefit Transfer: In-Sample Convergent Validity Tests of an Outdoor Recreation Database. *Water Resources Research* 36(4):1097-1107.

Rosenberger, R. S., and J. B. Loomis. 2000b. Panel Stratification in Meta-analysis of Economic Studies: An Investigation of its Effects in the Recreation Valuation Literature. *Journal of Agricultural and Applied Economics* 32(3):459-470.

Rosenberger, R. S., and J. B. Loomis. 2001. *Benefit Transfer of Outdoor Recreation Use Values: a Technical Document Supporting the Forest Service Strategic Plan (2000 Revision)*. General Technical Report RMRS-GTR-72. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 59p.

Rosenberger, R. S., and T. T. Phipps. 2001. *Improving the Validity of Benefit Transfers: A Site Correspondence Model and Meta-Analysis Transfer.* RRI Research Paper 2001-6. Morgantown, WV: West Virginia University, Regional Research Institute and Division of Resource Management.

Shrestha R. K., and J. B. Loomis. 2001. Testing a Meta-Analysis Model for Benefit Transfer in International Outdoor Recreation. *Ecological Economics* 39(1):67-83.

Smith, V. K., and J. Huang. 1995. Can Markets Value Air Quality? A Meta-Analysis of Hedonic Property Value Models. *Journal of Political Economy* 103(Feb):209-227.

Smith, V. K., and Y. Kaoru. 1990. Signals or Noise?: Explaining the Variation in Recreation Benefit Estimates. *American Journal of Agricultural Economics* 72(2):419-433.

Smith, V. K., and L. Osborne. 1996. Do Contingent Valuation Estimates Pass a Scope Test?: A Meta- Analysis. *Journal of Environmental Economics and Management* 31(Nov):287-301.

Smith, V. K., G. van Houtven, and S. Pattanayak. 1999. *Benefit Transfer as Preference Calibration.* Discussion Paper 99-36. Washington, DC: Resources for the Future. 37p.

Stanley, T. D. 2001. Wheat From Chaff: Meta-Analysis as Quantitative Literature Review. *Journal of Economic Perspectives* 15(3):131-150.

Stanley, T. D., and S. B. Jarrell. 1989. Meta-regression Analysis: A Quantitative Method of Literature Surveys. *Journal of Economic Surveys* 3:161-170.

Sturtevant, L. A., F. R. Johnson, and W. H. Desvousges. 1998. *A Meta-Analysis of Recreational Fishing.* Durham, NC: Triangle Economic Research. Unpublished paper.

U.S. EPA. 1997. *The Benefits and Costs of the Clean Air Act, 1970 to 1990.* U.S. Environmental Protection Agency (2000, June 13). Available: www.epa.gov/economics. 350p.

U.S. Water Resources Council. 1973. Principles, Standards, and Procedures for Water and Related Land Resource Planning. *Federal Register* 38(174): Part III, 24,778-24,945.

U.S. Water Resources Council. 1979. Procedures for Evaluation of National Economic Development (NED) Benefits and Costs in Water Resources Planning (Level c). *Federal Register* 44(243): 72,892-72,976.

U.S. Water Resources Council. 1983. *Economic and Environmental Principles And Guidelines for Water and Related Land Resources Implementation Studies.* Washington, DC: U.S. Government Printing Office. 137 p.

VandenBerg, T. P., G. L. Poe, and J. R. Powell. 2001. Assessing the Accuracy of Benefits Transfers: Evidence from a Multi-site Contingent Valuation Study of Groundwater Quality. In *The Economic Value of Water Quality.* Edited by J. C. Bergstrom, K. J. Boyle, and G. L. Poe. Mass: Edward Elgar.

Walsh, R. G., D. M. Johnson, and J. R. McKean. 1989. Issues in Nonmarket Valuation and Policy Application: A Retrospective Glance. *Western Journal of Agricultural Economics* 14(Jul):178-188.

Walsh, R. G., D. M. Johnson, and J. R. McKean. 1992. Benefit Transfer of Outdoor Recreation Demand Studies: 1968-1988. *Water Resources Research* 28(3): 707-713.

Walsh, R. G., D. M. Johnson, and J. R. McKean. 1992. Special Section: Benefit Transfers. *Water Resources Research* 28(3):651-722.

Woodward, R. T., and Y-S Wui. 2001. The Economic Value of Wetland Services: A Meta-Analysis. *Ecological Economics* 37(2):257-270.

Chapter 13

NONMARKET VALUATION IN ACTION

Daniel W. McCollum
USDA Forest Service, Rocky Mountain Research Station

1. INTRODUCTION

If a tree falls in the forest and no one is around to hear it, does it make a sound? Likewise, if a nonmarket valuation study is done and it is not used to affect or inform policy or management, does it serve a purpose? Previous chapters in this book have gone into detail about how to implement specific nonmarket valuation techniques. After reading all those chapters, one might reasonably ask: Who cares? Do these methods make a difference in the real world of policy and management decisions related to resource allocation? The answer is . . . well, sometimes.

Certainly there are studies that contribute to the development and refinement of methods to estimate nonmarket values. Such studies indirectly affect policy and management decisions by providing the technology to inform decisions and debate. Other studies directly address management issues or policy questions. Still other studies, which may or may not be aimed at specific policy or management decisions, can alter the policy debate and influence the decision making process over time. These categories overlap, and some studies could fall into all three.

References in previous chapters in this volume point out many studies that led to the development, refinement, and understanding of ways to estimate the values that people place on nonmarket goods and services. This chapter focuses on studies in the latter two categories, that is, nonmarket valuation studies that have had an identifiable influence on policy or management decisions.

First, I will give a somewhat anecdotal overview of the role of nonmarket valuation studies in policy and management decisions. Then I will turn to four studies that had identifiable effects on policy or management decisions. In each case, I will briefly describe the central research questions and talk about the tradeoffs that needed to be considered. Design and implementation of each study will be described, followed by an overview of study results. Finally, I will point out how results of the study entered into a policy or management decision. Along the way, I will point out some details on the steps, decisions, and general research process aspects of the studies, to provide some intuition on why things were done as they were.

It is not hard to find good nonmarket valuation studies in the published literature. It is a bit harder to ascertain which ones were done in the context of a policy or management decision.[1] It is harder still to learn whether and how the studies entered into the decision making process.[2] In some cases there are agency or consultant reports that give more and broader detail about studies described in journal papers; but even those reports do not tell how results were actually used, and they are often hard to get hold of unless one knows where to look.

The four studies of particular interest here are applications of the methods described in previous chapters. The first two use the contingent valuation method. The first of these was controversial because it sought to estimate increments to passive-use value[3] resulting from alternative levels of fluctuation in river flow associated with operation of Glen Canyon dam. As alluded to by Boyle in Chapter 5, contingent valuation does fine when use values are the subject of study; but questions continue about whether contingent valuation can validly measure passive-use values. Aside from those issues associated with passive-use values, the Glen Canyon study is a fairly straightforward application of the contingent valuation method.

The second study, of anglers in Colorado, is also a straightforward contingent valuation application. Among other things, it illustrates that studies do not always have to be complex and all-encompassing to be useful. But the study departs from the usual approach in that it uses contingent valuation data and results to estimate expected changes in participation resulting from increases in price (the concept of price elasticity of demand). While measures of net economic value were estimated, the emphasis in applying the results was on response (in terms of participation) to changes in price.

Following that is a study that is almost the antithesis of the Colorado angler study. The "Costs and Benefits of Reducing Lead in Gasoline" is a complex and all-encompassing study. It (of necessity) considers benefits from a wide range of sources. It also uses several methods to estimate those diverse benefits. The study is an application of the benefit transfer concept discussed in Chapter 12. Most of the individual studies on which it relied were applications of the defensive behavior and averted damage cost methods discussed in Chapter 11.

The final study is an hedonic study that focused on the contribution of water quality to lakefront property values in Maine. This study was also used to illustrate methodological issues in Chapter 10.

All four studies mentioned above were undertaken to address a rather immediate and specific policy or management question. That is certainly one way that nonmarket valuation contributes to policy and decision making: there is a problem; how can it be solved and what does one need to consider in formulating a solution? But nonmarket valuation might contribute in other ways as well. Following the four studies, I briefly discuss some other ways that nonmarket valuation can contribute to natural resource management and policy.

2. A SOMEWHAT ANECDOTAL OVERVIEW

Many nonmarket valuation studies have been done to address particular issues. Their results suggest various things about how policy could be structured or how management practices might be carried out. The extent to which such studies are actually used and enter into decisions is another matter. Decision makers sometimes regard valuation studies as "academic exercises" that do not really contain information relevant to public policy. They also might not find the study or its results to be credible or believable for one reason or another. Those and other factors, such as discussed by Mitchell (2002) and Whittington (2002), work to the detriment of nonmarket valuation studies being used in policy and management decisions.

2.1 Litigation and Damage Assessment

At the same time, one does hear about studies that have been used, and have made a difference in policy or management decisions. One avenue through which nonmarket valuation studies have been used is litigation and damage assessment. In a letter from the mid-1980s, Gardner Brown expressed to Rich Bishop that the Bishop and Heberlein (1979) study, which compared contingent valuation estimates of value for goose hunting permits in Wisconsin to actual cash transactions, had been instrumental in reaching an out-of-court settlement in a wetlands damage case (Brown 2002, Bishop 2002). Smith (2000) cited the example of litigation over the American Trader Oil Spill in California in 1990, in which the state relied on existing data to estimate lost beach recreation attributable to the spill and used benefit transfer to estimate lost consumer surplus value.[4] Other environmental damage cases have drawn from nonmarket valuation studies that have been accepted and used by the courts. There is even legal precedence for using nonmarket valuation in the context of damages in the form of lost passive-use value (DC Court of Appeals 1989). While more recent controversies (discussed in the next paragraph) have made loss of passive-use value more difficult to justify, its use has not been completely overturned.

Environmental damage litigation has also resulted in controversies and barriers to the use of nonmarket valuation. The 1989 Exxon Valdez oil spill case in Alaska is illustrative here. The magnitude and location of the event, the involvement of several high profile players (U.S. Government, State of Alaska, Exxon), the perception of a defendant with deep pockets, among other factors, put a spotlight on nonmarket valuation, particularly contingent valuation. Boyle talks about the attack on the contingent valuation method and the resulting controversies in his Introduction to Chapter 5 in this volume. Bishop, in Chapter 14, talks about the controversies as the "CV War." Unfortunately, in the Exxon case, an opportunity to demonstrate the usefulness of nonmarket valuation was cut short by an out-of-court settlement and by the subsequent squelching of nonmarket valuation studies that were being designed but had not been implemented.

To what extent did nonmarket valuation, and the studies that were being designed and tested to estimate the levels of damage experienced by various parties, contribute to an understanding of the loss of natural-resource-related benefits that accrued to various parties and influence the settlement? Unknown. On the positive side, the resultant controversy over nonmarket valuation, and contingent valuation in particular, has led to a more rigorous examination and testing of the methods. Those efforts have largely raised the level of quality seen in valuation studies. However, the notion of the use of nonmarket valuation in litigation seems to some people to carry negative connotations—that nonmarket valuation is being used as a club to punish the bad guys or to stand in the way of progress, depending on one's point of view.

2.2 Natural Resource Policy and Decisions

On the other hand, there are indications that nonmarket valuation analyses have been used to inform policy and management related to natural resources over time. Relatively early in the development of contingent valuation, results of a valuation study of waterfowl (Hammack and Brown 1974) were used by the Department of the Interior to argue in support of appropriations for wetlands before Congress. The appropriations were made and participants in the discussions, while doubting that the specific numbers were decisive, thought that the notion of net economic value attributable to waterfowl was a contributing factor (Brown 2002). The State of Montana, Department of Fish, Wildlife, and Parks, has over the years conducted several studies related to the economic value of various uses of fish and wildlife resources in Montana (Loomis, Duffield, and Brooks 1987, Loomis and Cooper 1988, among others). Similar studies have been done in Idaho (Donnelly et al. 1985, Loomis et al. 1985, among others). Those studies have contributed to species and habitat management decisions, and have affected hunting, fishing, and nonconsumptive wildlife use in Montana and Idaho.

Over the past thirty years or so, the USDA Forest Service has moved from being perceived (and perceiving itself) as a timber-producing agency toward being perceived as a provider of recreation opportunities. Evidence of that shift includes the large decrease in timber harvest from national forests and the large increase in recreation use of national forests and grasslands. While driven by changing public values and preferences, that (continuing) change has been both fueled and reinforced by nonmarket valuation studies that showed large net benefits to the public from engaging in recreation activities on public lands (Kaiser 2001). Those studies validated the agency's activities and programs to promote, enhance, and maintain recreational use of national forests and grasslands. They also illustrated the increasing share of the Forest Service contribution to the nation's well-being that is attributable to recreation activities as opposed to timber and other resource extraction uses of the land.

Studies looking at economic impact and the contributions of recreation to local and regional economies added fuel to the change as well. While some would argue that those economic impact and expenditure studies had more influence than the nonmarket valuation studies, both played a role and, in fact, were often components of the same studies (McCollum et al. 1990, Cordell and Bergstrom 1991, USDA Forest Service 1995, for example, all used the same database). That change in the Forest Service was noted by a National Research Council panel in their discussion of ("the few") comprehensive studies of the total value of forest products: "Recent work on goods and services produced on public lands managed by the U.S. Forest Service indicates that more forestland value is due to recreational and wildlife services than to timber, mineral, and range goods" (Nordhaus and Kokkelenberg 1999, p. 135).

Some of the studies alluded to above also played specific policy and management roles. McCollum et al. (1990) used a travel-cost-based model to estimate net economic values for several types of primary purpose recreation trips in each of the nine Forest Service regions for the 1990 RPA Program Analysis.[5] Other recreation values used in the 1990 Program Analysis came from a literature review by Walsh, Johnson, and McKean (1988), which collected values that had been estimated with a variety of methods in a variety of places. Those "RPA Values" (which have been updated over time) continue to be used by the national forests and grasslands for a variety of planning and policy evaluations.

2.3 Use by the EPA

The Environmental Protection Agency (EPA) has been a prominent user of nonmarket valuation over the past thirty years. Nonmarket valuation analyses have been used to estimate benefits (in benefit-cost analyses) accruing from environmental regulations related to the Clean Air Act, the Clean Water Act, and the Resource Conservation and Recovery Act, among others. In several cases, nonmarket valuation has contributed directly to the environmental regulations adopted (Morgenstern 1997). A report (Economic Studies Branch 1987) analyzing EPA's use of benefit-cost analysis found:

- Analysis improves environmental regulation. In several cases, benefit-cost analysis resulted in increased net benefits to society from environmental regulations.
- Benefit-cost analysis often provides the basis for stricter environmental regulations. Environmentalists often fear that economic analysis will lead to less strict environmental regulations in an effort to save costs.

In fact, EPA's analysis revealed that the opposite is just as often the case.

- Many environmental statutes prevent EPA from considering costs and even some benefits when setting environmental standards . . . EPA's experience shows, however, that some of the traditional statutory decision criteria, such as "health effect thresholds" and "technical feasibility," frequently do not provide clear distinctions for decision making. Being able to consider the full range of benefits and costs associated with alternative standards would enhance the information available in making these decisions.

The report (Economic Studies Branch 1987) quotes an Assistant Administrator at EPA:

> Over the years since EPA was founded, EPA's use of benefit-cost analysis in environmental rulemaking has increased considerably. While recognizing the limitations of benefit-cost analyses, we are finding these analyses to be increasingly useful tools in helping to provide the balance required in complex regulatory decisions.

3. THE GLEN CANYON PASSIVE-USE VALUE STUDY

This section draws heavily from Welsh et al. (1995) and Bishop and Welsh (1998).

3.1 Research Issues/Questions

Glen Canyon Dam, located on the Colorado River just upstream from the Grand Canyon, backs up water for some 200 miles stretching from northern Arizona into southern Utah to form Lake Powell. The dam and lake are part of a system of dams and reservoirs along the Colorado River designed to store water and provide for electric power generation and outdoor recreation. When the dam was completed and went into operation in the early 1960s, it changed the downstream environment of the Colorado River. Floods that had regularly raged through the Grand Canyon during spring runoff from the Rocky Mountains were eliminated or greatly moderated. The sediment that had scoured some parts of the canyon during those floods and got deposited in other parts became trapped in Lake Powell. Water released from the dam was substantially colder than it had been before the dam, especially during the summer, because it came from deep within the lake instead of from the shallower river.

Because output of electricity can be changed quickly and relatively inexpensively, hydropower is valuable to an electric utility for meeting peak demand. Peak-load electricity generation at Glen Canyon Dam, however, caused dramatic fluctuations in downstream flows and water levels over the course of a day as more water was released to generate more power. The result was changes in riverine and riparian environments.

An environmental impact statement (EIS) released in 1995 (U.S. Bureau of Reclamation 1995) identified several adverse impacts resulting from the fluctuating releases from Glen Canyon Dam:

- Native fishes, some of which were listed as endangered or were candidates for such listing, were adversely affected.
- Riverbed sand and hundreds of acres of sandbar beaches along the river were eroding and not being replenished by the flow of sediment now trapped in Lake Powell. Daily fluctuation of flow and resulting water levels due to peak-load electricity generation exacerbated the problem. Loss of sandbars was of environmental concern because they supported vegetation and backwaters that provided habitat for fish and wildlife. Those habitats were being eroded away.
- Archeological sites and American Indian traditional cultural sites were being lost to erosion.
- The dam created conditions (water temperature, etc.) that allowed introduction of non-native trout species and establishment of a sport fishery. Daily fluctuation in water releases from the dam, however, adversely affected food supplies and spawning success of those trout. As a result, the fishery remained heavily dependent on stocking of hatchery-reared fish.
- The enjoyment and safety of white-water boating in the Grand Canyon and on fishing downstream were adversely affected.

The research question, then, was: "What are the likely effects of alternative patterns of dam operation, using dam operations existing at the time of the EIS as the baseline?" Pattern of dam operation means the level and fluctuation in water releases from the dam used to produce electricity.

3.2 Tradeoffs

Changes in dam operation that would benefit the downstream environment and the quality of recreation use of the river and its surrounding areas would result in a loss of peak-load electric power generation capability and a reduction in total value of power produced at the dam. The conflicts or tradeoffs are thus between the type, level, and availability of environmental amenities and recreational opportunities along the Colorado River on one hand, and electric power generation capability on the other. Those tradeoffs can be evaluated by:

a. measuring the loss in economic value associated with reduced hydropower (specifically this would be the cost of acquiring peak-load power from an alternative source,[6] which would likely translate to an increase in electricity rates for people served by the Glen Canyon Dam power plant), and
b. comparing that to: (1) positive effects on the use values of recreation downstream, particularly whitewater boating and fishing; and (2) positive effects on passive-use values due to mitigation of the environmental and archeological/cultural degradations identified above.

3.3 Design and Implementation

The suite of Glen Canyon Environmental Studies included studies of the effects of alternative conditions of dam operation on the value of electric power generated and on the value of downstream recreational use of the Colorado River. Changes in the value of electric power produced were estimated using a power system model that predicted the cost of supplying power over a large area of the West depending on constraints on Glen Canyon Dam operation. Changes in recreation values were estimated using a contingent valuation study. The estimated benefits of improved recreation were small compared to the cost of reducing peak load electricity generation at Glen Canyon Dam. The decision about whether and how to change dam operation turned instead on comparing changes in power values and passive-use values. Therefore, the passive-use value study is the focus here.

The Glen Canyon passive-use value study (as described by Welsh et al. 1995) was guided by extensive input from managers and user groups. Various public agencies with responsibility for Grand Canyon resources, Glen Canyon Dam, and marketing of the power generated at the dam participated in each step of the research, from conception and survey design through analysis and presentation of results. This included federal agencies, tribal governments, and the State of Arizona. Private environmental groups and power interest groups were also included. Early versions of survey components were tested and refined in focus groups, and a small number of personal interviews were conducted during which subjects were debriefed about their reactions to information material, individual survey questions, and sections of the survey. Champ, in Chapter 3 of this volume, talks about those elements of study and survey design. A peer review panel of four nationally prominent resource economists reviewed research plans and results at each key stage in the research.

Because releases from Glen Canyon Dam affect resources in Grand Canyon National Park, it was decided that the relevant study population was all residents of the United States. Within that national sampling frame, two samples were identified as needing to be studied: (1) a national sample, drawn from all residents of the U.S.; and (2) a marketing area sample, drawn from the subset of the national population whose energy needs were serviced by the Salt Lake City Area Integrated Projects, which includes the power plant at Glen Canyon Dam. That design ensured that estimates of passive-use value would reflect both the values held by the larger population of U.S. residents and the values held by the subpopulation that would be directly affected by the increased cost of electric power.

The passive-use value study evaluated three of the several alternatives assessed in the environmental impact statement. The baseline to which the three alternatives were compared was the "no-action" alternative, or dam operating conditions existing at the time of the EIS. The assessed alternatives were: (1) **Moderate fluctuating flow**—featuring moderate reductions in the magnitude of daily fluctuations in water released from the dam; (2) **low fluctuating flow**—featuring large reductions in the magnitude of daily fluctuations (further reductions from the moderate flow alternative); and (3) **seasonally adjusted steady flow**—providing steady flows on a seasonally adjusted or monthly basis. Those alternatives were selected because they covered most of the range of alternative dam operations, and it was considered most likely that the eventual "preferred alternative" would be within that range.

Respondents were asked to evaluate **one of the three** alternatives against the no-action alternative; i.e., they were asked to choose the specified alternative (assigned to respondents at random) or reject that alternative in favor of the baseline condition. The two samples of interest, the national sample and the marketing area sample, resulted in six versions of the questionnaire. An additional (seventh) version was added to the national sample to examine whether and how estimated values were affected by including specific information about the impacts that water-flow alternatives would have on power costs, even though the increased cost would not be borne by respondents. In other words, did people respond differently when reminded how the change in dam operation would affect other people?

The design process for the survey instrument included in-depth interviews with key managers and other people involved with Glen Canyon Dam. It also included fifteen focus groups held in the Southwest and around the country. Several things were learned from those interviews and focus groups. People seemed to be able to distinguish between the aquatic and riparian resources of the Colorado River and the much larger set of resources of the Grand Canyon as a whole; i.e, they could focus on the relevant issues. It was clearly confirmed that interest in those resources of the Colorado River extended across the country. Issues of concern included beaches and vegetation, archeological sites, American Indian traditional use areas, native fish and endangered species, trout, and the impacts of dam operation alternatives on electricity prices.

Background information was provided to all respondents along with the final questionnaire. That information consisted of four pages describing the dam and the study area downstream, and outlining dam operations using both verbal description and a map. It then described the affected resources and explained why concerns had arisen about the effects of dam operation on those resources. Finally, it pointed out that changing dam operations could reduce adverse environmental effects, but that the cost of electricity to households and farmers in the region could increase. As discussed by Boyle (Chapter 5), it is often a challenge to know how much information to present to respondents. Four pages is a lot, but in this case the focus groups indicated it was not too much and it did not appear to be a problem in the final pretest. People would take the time to read the information. Their willingness to read was borne out in responses to a set of true-false questions designed to encourage respondents to read the background information and to test how well they understood it.

A referendum format was used for the contingent valuation exercise. Respondents were first asked whether they would vote for the proposed alternative if it cost them nothing. Response categories were "No," "Yes," and "I would choose not to vote." That first question, would you vote for the proposal if it cost you nothing, was designed to control for "yea-saying" (discussed in Chapter 5). That is, the investigators wanted to give respondents a chance to express approval (or disapproval) for the stated alternative in the absence of any price effects. That allowed respondents the psychological freedom to base their response to the subsequent willingness to pay (WTP) question on the dollar amount specified without seeming to be against a perceived need to take action to mitigate a damaging situation.

Respondents who said "yes," they would vote for the alternative at no cost, were asked whether they would pay a specified amount every year for the foreseeable future. Response categories were "Definitely No," "Probably No," "Not Sure," "Probably Yes," and "Definitely Yes."

Following the WTP question, respondents were asked what they would choose to spend less money on if the proposal passed and they had to pay the specified amount, followed by an opportunity to change their WTP vote after that additional consideration. The purpose was to encourage respondents to explicitly consider their own budget.

The contingent valuation section of the questionnaire was followed by questions about whether respondents had been to or heard of Glen Canyon Dam; their views on environmental issues, American Indian issues, hydroelectric issues, national parks and the Grand Canyon in particular; and social/demographic information. Such questions aid in the interpretation of valuation results and in understanding what might be driving the values expressed in the contingent valuation. Responses to those questions were used to assess construct validity[7] of study results. Documenting validity makes the results more credible to decision makers and, hence, more likely to be used. In Chapter 14 of this volume, Bishop discusses the need for more attention to validity in the future.

Questionnaires administered to the national sample and to the marketing area sample differed in the payment vehicle used for the contingent valuation referendum. In the national sample, respondents were told that higher electric rates could not make up for all the revenue lost as a result of the proposal and that taxpayers would have to make up the difference. The payment vehicle for the national sample, then, was an increase in taxes every year for the foreseeable future. For the marketing area sample, the payment vehicle was an increase in electric bills for the foreseeable future.

The description of resources affected by the dam operation alternative also differed between the two samples. Both environmental and electricity price impacts of the dam operation alternative were described to the national sample, while only the environmental impacts were described to the marketing area sample. For the latter, the impact on electricity price was provided as the amount specified in the WTP question.

The initial sample in each of the seven subsample groups consisted of 850 individuals—a total initial sample of 5,950. A mail survey was used, achieving overall response rates of 66% for the national sample and 75% for the marketing area sample. An attempt was made to contact all nonrespondents by telephone. The phone interview consisted of questions to learn why the person had not responded, and some questions from the original questionnaire (not the WTP question) that would allow tests of whether and how nonrespondents differed from respondents (i.e., the existence of nonresponse bias). Again, the purpose was to establish validity of the sample and give credence to the study results.

3.4 Results

While the exact percentage varied among groups to which particular dam operation alternatives were posed, more than 70% of respondents said that, if there were no cost to them, they would vote in favor of the dam operation alternative they were asked to consider. The increase in electric rates was viewed as a negative attribute by many respondents, but the damage imposed by the existing level of dam operation was viewed as more negative, hence their support for the alternatives in the absence of personal cost.

Analyses of the WTP responses yielded estimated positive increments to passive-use value as shown in Table 1. All increments were relative to the

Table 1. Increments to Passive-Use Value vs. Losses in Power Generation Capability

Glen Canyon Dam Operation Alternative	Positive Annual Increment to Passive-Use Value	Estimated Annual Loss Due to Less Peak Load Generation Capability
Moderate Fluctuating Flow	[a] $2.29 billion ($13.65)	$36.7 million to $54 million
Low Fluctuating Flow	$3.38 billion ($20.15)	$15.1 million to $44.2 million
Seasonally Adjusted Steady Flow	$3.44 billion ($20.55)	$88.3 million to $123.5 million

[a]Numbers in parentheses indicate the increment to passive-use value expressed as a mean annual amount per household. Totals are aggregated across all households in the U.S.

baseline level of dam operations. Those values represented average household value aggregated across all households in the U.S. They were ***additions*** to passive-use value expressed in annual amounts for the foreseeable future, and they were large amounts so they caught policy makers' attention .[8]

The estimated national level increments to passive-use values were more than an order of magnitude greater, billions compared to millions, than the estimated losses due to less peak-load electric power generation capability at the Glen Canyon Dam power plant, also shown in Table 1. Estimated losses from constrained electric power generation were given as ranges to reflect a range of modeling assumptions.

As mentioned above, the Glen Canyon passive-use value survey instrument included questions to allow examination of content and construct validity in an attempt to answer the question: "Can decision makers believe the results of contingent valuation studies in general, and this study in particular?" That validity assessment exercise points out several desirable features of the study. First, the survey instrument was developed based on interaction and testing with focus groups and a pilot study; refinements were made to the survey instrument as a result of both. Second, respondents understood the background information, as indicated by the results of the true-false test at the beginning of the questionnaire, so they had some basis on which to form their values of the alternatives. Third, respondents were asked whether they thought the study results should be used in deciding how Glen Canyon Dam should be operated in the future. Ninety-five percent said yes, indicating that respondents overwhelmingly felt they were engaging in a meaningful process. Fourth, the pilot study showed that responses were unaffected by minor changes in question wording, indicating a robustness of results. Fifth, the sample appeared representative of the U.S. population. High response rates to the mail survey combined with telephone follow up of nonrespondents helped minimize any response bias. Finally, statistical analyses showed responses to the WTP questions to be strongly related to factors such as income, education, environmental attitudes, and expectations about future visits to the Grand Canyon in ways that were consistent with prior expectations. All those checks on the study process and data lend credence to the study's results and findings.

3.5 Use of Results and Their Range of Influence

As the final step in the EIS process, Secretary of the Interior Babbitt issued a Record of Decision announcing his choice of a slightly modified version of what Welsh et al. (1995) called the low fluctuating flow alternative as the "preferred alternative." Among the reasons cited, Secretary Babbitt noted:

> This alternative would create conditions that promote the protection and improvement of downstream resources while maintaining some flexibility in hydropower production. Although there would be a significant loss of hydropower benefits due to the selection of the preferred alternative (between $15.1 and $44.2 million annually), **a recently completed nonuse value study indicates that the American people are willing to pay much more than this loss to maintain a healthy ecosystem in the Grand Canyon.** (Quoted in Bishop and Welsh 1998, p. 192, bolding added.)

Summing up their application of contingent valuation, Bishop and Welsh observed that:

> Evaluating alternatives for Glen Canyon Dam operations was a quintessential opportunity to apply passive-use values in the analysis of a proposed regulatory change. Clearly power and recreational values are potentially important, yet to base the decision solely on such use values neglects any stake that the public at large may have in the environmental resources of Grand Canyon National Park and nearby areas along the Colorado River. Contingent valuation offered the possibility of measuring these broader environmental values and including them in the analysis of a major regulatory decision (Bishop and Welsh, 1998, p.184).

4. THE COLORADO ANGLER STUDY[9]

This section draws heavily from McCollum, Haefele, and Rosenberger (1999).

4.1 Research Issues/Questions

Whirling disease is a parasitic condition affecting trout, primarily rainbows and cutthroats. Other species of trout and salmon are affected to lesser degrees. Warm-water fishes such as bass, walleyes, and catfish are not affected. Young fish are most at risk from the parasite, which attacks the cartilage, resulting in deformities. Infected fish can display a distinctive rapid whirling, or swimming in circles, hence the name. The disease can be fatal to very young fish if they are infected heavily enough, which likely explains why younger age classes of trout have been observed to be missing in sections of several rivers infested with whirling disease. Larger infected trout generally do not die from whirling disease, but they can be carriers. In most cases, infected adult fish show no signs of whirling disease and live normal life spans. The disease does not infect humans. The full effects of whirling disease in the wild are still unknown. One problem, however, is that the parasite's spores are hardy and are even known to survive passage through the intestinal tract of birds; so the parasite may persist indefinitely and keep spreading.

The incidence and spread of whirling disease resulted in increased costs to the Colorado Division of Wildlife (CDOW) as the fishery management program tried to clean up and protect state fish hatcheries from the parasite and at the same time continue its stocking program, often purchasing fish from private hatcheries. Even before whirling disease was found in Colorado, fishing license sales were not completely covering the cost of the state's fish program. With the need to deal with whirling disease the gap between revenues and costs was even greater. CDOW needed information about what management and funding options would be acceptable to anglers.

4.2 Tradeoffs

In any decision regarding user fees for public resources, one tradeoff involves participation. Fees can be set to cover the cost of providing a public good or service, but doing so might exclude particular users or groups of users, thereby creating a question of social equity. A balance must be achieved between covering costs and making the good or service accessible to all. In this case, another tradeoff involved alternative means of potentially achieving the desired result of restoring a healthy trout population to Colorado waters. License fees could be increased to give CDOW more money to clean up infested fish hatcheries, buy fish from other (non-infected) sources, and fund research on ways to mitigate the effects of whirling disease. Alternatively, regulations could be imposed or changed to reduce fish harvest and leave more (healthy) trout in the water.

4.3 Design and Implementation

A fee increase could result in either an increase or decrease in total revenue, depending on participation. Therefore, it was important to learn about angler values, preferences, and probable responses to an increase in license fees under several potential fee structures. Basic design of the contingent valuation scenarios was accomplished through a lengthy series of interactions with CDOW managers. It was important for the research team to fully understand both the management situation regarding whirling disease and the potential options for dealing with the problem, as well as the political realities facing CDOW. Wildlife- and fish- related regulations and policies are established by the governor-appointed Wildlife Commission, with input and advice from CDOW. License fees are set by the state legislature and had changed only four times in the previous thirty-five years. One manifestation of those political realities was the infeasibility of eliminating or substantially cutting back—even temporarily—the fish stocking program. Such a change could have provided one way to save costs, especially in light of the need to purchase higher cost fish from hatcheries not infested with whirling disease, but the political sideboards eliminated that option. In fact, one option offered by some within CDOW was actually to increase stocking to offset the loss of trout available to anglers due to whirling disease.

It was also important for CDOW managers to consider how specific fee increases and changes in management policies and regulations might be structured, and how those changes could be described and woven into valuation scenarios. Beyond the importance of manager involvement to accurately depict technical aspects of the valuation scenarios, active participation by managers in the study design increased their "buy-in" to the study and acceptance of the results at the end, which provided further credibility to the study.

It was decided to pose four fee increase options. Two of the options retained the existing fee structure under which all anglers within each license class (resident annual, resident senior, resident combination fishing/small game hunting, etc.) pay the same license fees, regardless of where they fish and what they fish for. Option 1 was an across-the-board fee increase that kept fishing conditions, including level of stocking, the same as currently existed. Option 2 had the same across-the-board fee increase but it was accompanied by a 25% increase in stocking; other conditions stayed as they currently existed.

Fee increase options 3 and 4 represented changes in the fee structure by requiring stamps in addition to a fishing license for certain kinds of fishing. Under Option 3 the "Stocked Trout Water Stamp" would be required (in addition to a fishing license) for anglers fishing for trout in streams or lakes where catchable size trout were stocked, regardless of whether any trout were kept. Option 4 required the "Trout Stamp" for anglers fishing for trout (wild or stocked) anywhere in Colorado, again regardless of whether any trout were kept. For both stamp options, all other fishing conditions, including level of stocking, remained as currently existed. The stamp options were based on a "user pay" notion: because whirling disease affects trout, only those Colorado anglers who fish for trout would bear the increased cost of the fish program due to whirling disease. The issue was how narrowly to focus the incidence of the fee increase—on anglers who fished directly for stocked trout or on all anglers who fished for any trout at all.

The study population was Colorado residents age 18 or older who bought regular annual fishing licenses, annual senior fishing licenses, or combination fishing and small game hunting licenses in 1997. A sample stratified by license type was drawn from the CDOW license files.

The research team was concerned that asking each respondent to consider four fee increase scenarios might be too much burden. Yet, it was desirable to have respondents consider more than one scenario to allow for some "within respondent" comparisons of alternatives. It was decided to ask respondents to consider two scenarios. Focus groups indicated that people could and would consider two scenarios without loss of interest; the groups suggested more than two valuation scenarios would be too much and response quality would suffer.

There was also concern that responses to questions based on the first scenario encountered within a questionnaire might affect responses based on subsequent scenarios, particularly if respondents did not consider the second option to be completely independent and distinct from the first option they encountered. To test hypotheses regarding any such effects, the study was structured to accommodate all combinations of fee increase options (1 with 2, 1 with 3, etc.). Further, to test (and control) for any order effects on responses, each subsample was split and the options were reversed in the order they appeared on the questionnaire. For example, the subsample containing Options 1 and 3 was split so that half the respondents saw Option 1 followed by Option 3, and half saw Option 3 followed by Option 1. In that way order effects would be randomized, but their existence could be tested for.

After the final focus group, a revised questionnaire draft was mailed to a random sample of 650 Colorado anglers as a pretest. Half the sample received Options 1 and 2. The other half received Options 3 and 4. Only one mailing was used; the response rate was 38%. Those data were analyzed with two goals in mind. One was a final opportunity to detect any questions for which responses indicated respondents misunderstood the question or had trouble answering. Respondents were told in the instructions to "please write in the margins and tell us if something was hard to understand or answer." The second goal was to obtain preliminary estimates of value (or WTP) for each fee increase scenario to provide information with which to set the "offer amounts" for the dichotomous choice questions in the final questionnaire.

The final questionnaire contained four parts. Part One asked about anglers' fishing characteristics, behaviors, and experiences. Part Two asked anglers to agree or disagree with thirty-seven statements related to fishing and fishery management to get at attitudes, beliefs, and preferences. As was discussed in the context of the Glen Canyon study above, one reason for collecting those data was to help interpret the valuation results and evaluate construct validity. Part Three included the fee-increase scenarios (two scenarios in each questionnaire), posing both a dichotomous choice willingness to pay (WTP) question and an open-ended WTP question for each scenario. The questions asked whether respondents would purchase a license or stamp under the conditions posed, so the payment vehicle was user fees. Part Four collected demographics and socioeconomic characteristics.

The initial sample size was 8,209. One reason for that sample size was the unusually high number of undeliverables (around 30%) observed in previous mail-survey-based studies that had drawn their samples from CDOW license files. Indeed, the Colorado Angler study ended up with 32% of the initial mailing being returned as undeliverable. As discussed by Champ in Chapter 3, adequate sample size is a function of the degree of statistical power desired in the smallest subsample that one wants to analyze; but more is better.

The questionnaire was administered by mail. An advance letter was sent to introduce the study and let people know to expect the survey. The survey packet was sent to those whose introductory letter had not been returned as undeliverable; that packet contained a cover letter, a sheet of questions and answers about the study, and the survey booklet. A postcard was sent to everyone who received a survey packet, thanking those who had responded and reminding those who had not that we needed to hear from them. Two additional rounds of survey packets were sent to nonrespondents. The survey achieved a 71% response.

4.4 Results

Comparison of responses to the across-the-board-fee-increase scenario with the existing level of stocking and that with a 25% increase in stocking indicated minimal willingness to pay for a substantial increase in stocking. That result answered one of CDOW's questions. Were anglers willing to pay for increased stocking? No.

Price elasticity of demand was calculated for the three types of licenses. All were found to be price inelastic. CDOW would increase revenue if license fees were increased. There appeared to be a revenue maximum from license fees at relatively modest fee increases of \$4 for regular annual licenses, \$2 for senior licenses, and \$6 for combination licenses. Revenue increase at those fee levels, however, would be less than \$1 million over current revenue. That figure included the loss of some license sales: at those fee levels, 87% of current annual license holders, 94% of seniors, and 85% of combination license holders would still be expected to buy licenses. How to balance the increased revenue and decreased participation is a policy decision.

The relative insensitivity of license demand to fee level was supported by the attitude/preference data. Only about a third of anglers said the price of the license affects whether they buy a fishing license.

Regarding how to handle the problem of restoring a healthy trout population to Colorado waters, the study found that stricter bag limits were preferred to season closures (even if that meant a limit of two trout per day, which is a strict limit compared to the existing limit of eight per day), and that both bag limits and closures were preferred to an increase in fees as a mechanism for protecting trout populations. Furthermore, in the event that a fee increase could not be avoided, the results suggested it would be more acceptable to combine a fee increase with stricter bag limits for trout than to combine a fee increase with season closures for trout.

This study emphasized using different kinds of data to look at different aspects of the issue, and to provide evidence of validity. Data regarding attitudes and preferences might allow comparison of management alternatives in ways that strict contingent valuation data might not; they also confirm and reinforce the valuation results. Furthermore, it was stressed that one must look at the entire distribution of economic values resulting from the contingent valuation analyses, rather than focusing only on the mean and/or median. Depending on the analytical question(s), answers might be found in different parts of the range—the 90th percentile, for example.

4.5 Use of Results and Their Range of Influence

Based on results of the Colorado Angler study, along with other considerations, CDOW decided against proposing an increase in license fees to the State Legislature. Streams in certain areas of western Colorado went to a two-trout-per-day bag limit. Regulations were proposed to the Wildlife Commission to decrease the statewide limit for coldwater species from eight fish per day to four in areas not under more restrictive "special regulations," such as those western Colorado streams that had gone to a limit of two fish per day. Additional areas of western Colorado were proposed to come under the special regulation of two trout per day. No season closures for trout were proposed, and the level of stocking, in fact, decreased slightly from the 1997 baseline used in the study.

The Colorado Angler study was one in a series of studies based on economics and social science that provided input to CDOW from a broader spectrum of wildlife users and the general public than have traditionally had a voice in policy and decision-making.

5. COSTS AND BENEFITS OF REDUCING LEAD IN GASOLINE

This section draws heavily from Schwartz et al. (1984).

5.1 Research Issues/Questions

Since 1973, the U.S. Environmental Protection Agency (EPA) has regulated the use of lead as an additive to gasoline. Section 211 of the Clean Air Act gave the EPA Administrator authority to control or prohibit any fuel or fuel additive that:

- causes, or contributes to, air pollution which may reasonably be anticipated to endanger the public health or welfare, or
- will impair to a significant degree the performance of any emission control device or system which is in general use. . . .

To avoid adverse effects of lead in the environment and protect emission control equipment rendered ineffective by lead additives, EPA required that cars, beginning with model year 1975, meet tighter emissions limits. To meet those stricter emissions standards, automobile manufacturers installed catalytic converters, which required unleaded gasoline. Pre-1975 cars, without catalytic converters, were still able to use leaded gasoline. In several stages between 1976 and 1982, EPA mandated that the lead content of leaded gasoline be reduced from over 2.0 grams per gallon to 1.1. EPA projected that demand for leaded gasoline would drop to zero as the automotive fleet aged and cars designed to use leaded fuel were replaced by those designed for unleaded fuel, making further restriction of lead in gasoline unnecessary.

Studies by EPA and others, however, found that widespread "misfueling" (the use of leaded gasoline in vehicles designed for unleaded gasoline) was occurring. Use of leaded fuel affected the catalyst in catalytic converters, rendering the converters ineffective. The result of the misfueling was a significant slowing of the decline in lead emissions from cars, compared to EPA projections made when lead content of gasoline was reduced to 1.1 grams per gallon. It also challenged the assumption that leaded gasoline would soon be eliminated due to lack of demand. At the same time, health research was producing a substantial literature about the adverse effects of even low levels of lead exposure. Several studies strengthened the identification of lead from gasoline as a major source of blood lead, and new information on the effects of lead on physiological functions had become public. So, while reductions in lead emissions were slowing down, evidence was mounting that exposure to even low levels of lead was more injurious than previously realized.

Why would a person use leaded gasoline in a car designed for unleaded fuel? For one thing, leaded gasoline was cheaper. At the time, the price differential between regular leaded gasoline and regular unleaded gasoline averaged about seven cents per gallon. Second, petroleum refiners added lead to gasoline as an inexpensive way to increase octane level. Higher octane levels boosted performance, particularly in larger engines.

Beyond the difference in price and the performance advantage of a higher octane level, lead additives added some lubricating properties to gasoline that reduced valve wear in engines not specifically designed for unleaded fuels.[10] Especially for use in trucks, those lubricating properties of lead buoyed the demand for leaded gasoline.

The problem that misfueling caused in cars with catalytic converters, the resulting lower-than-expected decline in demand for leaded gasoline, and the increased recognition of serious health effects from even low levels of lead in the blood, coupled with the fact that gasoline had been identified as the major source of environmental exposure to lead, caused EPA to reexamine its policies. Two alternatives were considered: (1) an outright ban on the use of lead in gasoline. (2) a "low-lead" option designed to eliminate misfueling. Under the second option, lead in gasoline would be further restricted (from the then current restriction of 1.1 grams per gallon) to 0.10 grams per gallon. That would be enough lead to prevent excessive valve wear, but it would minimize environmental contamination; it also would actually cost more to produce such low-lead gasoline than unleaded, thus eliminating the price incentive for misfueling. Furthermore, retail availability of leaded gasoline would be restricted (such as by being limited to full service stations) to avoid misfueling.[11]

The research question, then, was: "What costs and benefits would result if one or the other of the alternative restrictions on lead in gasoline were implemented?

5.2 Tradeoffs

Under either alternative, drivers of older cars and trucks designed to run on leaded fuel would pay more, as would those drivers who had previously misfueled their cars, because both unleaded and the alternative low-lead fuel were more expensive to produce than the leaded fuel in use at the time. On that same side of the equation, there would be some decrease in performance as octane-increasing benefits of lead would be lost under both the ban on lead and the low-lead alternative, and the lubricating properties of lead would be lost in the total ban alternative. Loss of those lubricating properties could cause severe damage to some older vehicles. On the other side of the equation, there would be benefits from both alternatives in the form of vehicle maintenance savings because the corrosive effects of lead on engines would be eliminated or reduced. Furthermore, health benefits, particularly in children, would be realized from elimination or reduction of lead emissions from automobiles. The benefits are described in detail below following a short discussion of costs.

5.3 Design and Implementation

The costs and benefits of the two alternative levels of increased restriction on lead in gasoline were compared to a base case of continuing the existing level of restriction (1.1 grams of lead per gallon). Under existing regulations, the lead standard had to be met on a quarterly average rather than for every gallon produced. In addition, refiners were able to average their own production or sell off-sets to each other. Two refiners could agree, for example, that one would produce gasoline with 1.0 grams of lead per gallon and the other would produce an equal amount of gasoline with 1.2 grams per gallon (and pay the first refiner some amount of money). That ability allowed the refining industry as a whole to minimize the cost of meeting the restrictions on lead.

Costs of the additional restrictions on lead additives were estimated by projecting the demand for gasoline, and the leaded/unleaded split, for the base (i.e., currently existing) scenario and for each of the alternative levels of restriction. Those projected demands were then used in a linear program model of the refining industry to calculate the incremental costs of production. The study looked only at the cost of production. Because factors other than production cost contribute to the retail price of gasoline—marketing strategies, etc.—EPA did not attempt to estimate incremental changes in gasoline prices to consumers. Because this book is primarily about estimating benefits, that is all I will say about cost estimation. Schwartz et al. (1984) contains much more detailed information about the cost estimation procedures.

Several categories of benefits (relative to the base scenario) were identified and considered in the analysis:

1. *Benefits from reduced vehicle maintenance requirements, which accrued to vehicle owners.*

Lead compounds and their associated scavengers[12] foul and corrode the engines and exhaust systems of all vehicles using leaded gasoline. Operation and maintenance savings from lower lead or no lead in gasoline would result primarily from: less frequent tuneups, less frequent exhaust system replacements, and less frequent oil changes. Subtracted from those benefits would be the possibility of valve damage to vehicles designed for leaded gasoline but using unleaded gasoline (which would occur under the complete ban on lead but not under the low-lead alternative).

2. *Benefits of avoiding excess hydrocarbon, carbon monoxide, and nitrous oxide emissions, which accrued to the population as a whole.*

Hydrocarbon (HC), carbon monoxide (CO), and nitrous oxide (NO_X) are the conventional pollutants from automobiles. Using leaded gasoline in engines designed for unleaded (misfueling) rendered their catalytic converters ineffective, which then caused a substantial increase in HC, CO, and NO_X emissions. Furthermore, HC and NO_X combined to form ozone, thus increasing ozone concentrations. EPA (and others) estimated that over 12% of all vehicles equipped with catalytic converters were being misfueled. Because the low-lead alternative assumed no misfueling occurred (because low-lead gasoline would be more expensive to produce than unleaded, thus removing the incentive for misfueling), both alternatives produced the same levels of reduction in HC, CO, and NO_X emissions compared to the base scenario.

3. *Health benefits of reducing lead in children with high blood lead levels, which accrued directly to affected children and their families, and indirectly to the population in general.*

The analyses of health benefits focused on children. While adults experienced adverse effects from lead, those effects generally occurred at higher lead levels than in children. When children had blood levels of lead over 30 micrograms per deciliter (μg/dl), they required follow-up and/or medical treatment for associated conditions such as anemia, kidney damage, hypertension, and other pathophysiological consequences. Some children had blood lead levels high enough to reduce cognitive performance, including the loss of several IQ points. Researchers found those cognitive deficits to remain after three years, even after medical attention. Such adverse cognitive effects required compensatory education to overcome the additional learning disabilities. The analyses were based on reduced exposure to lead resulting from reduced or eliminated lead in gasoline, and not on exposure from lead-based paint or other sources that would not be affected by the restriction of lead in gasoline.

4. *Health benefits of reducing lead in children with lower blood lead levels, which accrued to affected children and their families.*

As measurement tools improved, researchers detected pathophysiological effects at blood lead levels previously thought to be safe (i.e., under 30 μg/dl). Those results warranted concern about even small changes in the total body lead burden of children, especially those children who were subject to sources of lead exposure in addition to lead from gasoline. Effects included inhibition of some enzymatic processes, changes in brain activity, increased risk of abnormally small red blood cells, and inhibitions or disruptions of several other metabolic processes. Such effects became detectable at blood lead levels ranging from 10 μg/dl to 25 μg/dl. Indications of mild cognitive effects were also found at low lead exposure levels.

The comparisons of benefits and costs between the base scenario and the alternatives were made for the (one) year, 1988, when the ban on lead or requirement for low-lead fuel would become effective. The analyses (actually done in 1983) projected demand for gasoline, the leaded/unleaded split, and the costs and benefits five years into the future. Such an approach is not uncommon in policy analyses. A decision has to be made and implemented, both of which take time, so effects of a program must be estimated well before they actually occur.

I will not discuss the details of how all those benefits were estimated, but I would encourage those interested in benefit transfer to read those details in the Schwartz et al. (1984) report and appendices. To illustrate what was done and how, I will briefly describe benefit estimation from reduced vehicle maintenance and health benefits from reduced lead. The former are tangible and relatively easy to identify. The latter are harder to identify and even harder to estimate. Both estimates used methods that are described in this book as benefit transfer (Chapter 12). Results of studies done at other times, in other places, and for other purposes were applied to estimate benefits of the proposed restrictions on lead additives. The studies underlying the benefit transfer relied on methods described by Dickie in Chapter 11 of this book.

5.4 Results

5.4.1 Benefits from Reduced Vehicle Maintenance

Estimation of benefits from reduced vehicle maintenance used: existing on-road vehicle studies, engine tests on road simulators, theoretical and engineering calculations, and changes in manufacturer recommendations for vehicle maintenance under different conditions (including reasons for the changes). Results and findings of those studies were applied to the alternative restrictions of lead levels in gasoline to assess changes in vehicle maintenance costs compared to the base scenario. Evaluation of those studies and data was required to determine their applicability to the case of imposing tighter restrictions on lead in gasoline. The need for such evaluation in benefit transfer was discussed by Rosenberger and Loomis in Chapter 12 of this book. Some studies were excluded: some of the effects shown in the studies could not be related exclusively to differences between leaded and unleaded gasoline, some studies used data bases with sample sizes judged to be too small to provide meaningful conclusions, and some were not considered reasonable to extrapolate to vehicles operating in 1988.

For the purposes of this book, it is instructive to look more closely at the benefit estimates for less frequent exhaust system replacements. The studies used for that part of the analysis are summarized in Table 2. All of the studies found differences in expected lifetimes of exhaust systems, as measured in miles, between matched pairs of unleaded and leaded fueled vehicles. The range of estimated differences between leaded and unleaded replacement rates, however, was very broad, from only 20% fewer muffler changes for unleaded vehicles (based only on theoretical calculations) to, more commonly, virtually no muffler replacements for unleaded vehicles in four of the eight distinguishable fleets (in on-road vehicle studies). Averaging the results of all those studies results in a projection of one exhaust system replacement every 56,000 miles for cars using leaded fuel, and essentially none for vehicles using unleaded fuels during the test periods.

Those studies were conducted on fleets of vehicles over several years, but for less than the lifetimes of the vehicles. It is possible, therefore, that the studies ended shortly before many of the unleaded vehicles required exhaust system replacement. Perhaps the replacement rates would have increased significantly had the fleets traveled another 10,000 to 20,000 miles. Thus, the reported findings might overestimate the differences between leaded and unleaded fueled vehicles in terms of their need for exhaust system replacement.

It is useful to look most closely at the Wintringham et al. (1972) findings, because their vehicles had the greatest mileage and there is a clear geographic distinction between the fleets. Their Baton Rouge fleet, after over 84,000 miles of travel per car (compared to a projected vehicle lifetime of 100,000 miles), had essentially zero exhaust system repairs for unleaded vehicles, but about one per 31,000 miles for leaded vehicles. By comparison, the Detroit fleet, after 73,000 miles of travel per vehicle, had a rate for unleaded fueled vehicles of one exhaust system repair per 46,000 miles, but a rate of one per 24,000 miles

Table 2. Summary of Studies on Maintenance Differences Between Leaded and Unleaded Fueled Vehicles (Exhaust

	Replacement rates per 11,000 miles (or per yr.)						
	Unleaded fueled	Leaded fueled	Avg. MPV/Yr.[a]	Accumulated MPV[b]	Type of Service	# of Vehicles/ Length of Test	Location
Pahnke & Conte (1969)	.0033	.187	11,400	65,000	Personal use	59 pairs / 4.7 yrs.	South NJ & Wilmington, DE
Pahnke & Bettoney (1971)	.220	.275	10,000	---	Theoretical calculations	---	---
Gray & Azhari (1972)					Commuting & business use	12 pairs / 2 and 3 yrs.	Chicago & suburbs
MY[c] 1967:	.149	.535	7,500	24,000			
MY 1968:	0	.217	7,500	17,000			
Gray & Azhari (1972)	.004	.071	Not Reported	1 to 6 yrs.	Personal use (consumer panel)	151 pairs / 1 - 5 yrs.	Eastern states Concentrated Mid-Atlantic
Wintringham, et. al., (1972)					Employee fleet (business & personnel)		
Detroit:	.155	.289	14,575	72,883		31 pairs	Detroit
Baton Rouge:	.004	.358	16,850	84,260		33 pairs / 5 yrs.	Baton Rouge
Hickling Partners (1981)	2.4 times as many for leaded vehicles		Unknown	23,810 lead 24,990 unlead	Municipal service	835 / 5 yrs.	Montreal Edmonton

[a] miles per vehicle per year, [b] miles per vehicle, and [c] model year. Source: Schwartz et al.

for leaded vehicles. The main reason for the different experiences in Baton Rouge and Detroit, the authors concluded, was the greater external corrosion due to road salts in the colder climate. This was consistent with findings by Environment Canada (Hickling Partners 1982) for two municipal fleets in cold climates, which had 42% fewer exhaust system replacements, at equivalent mileage, for cars using unleaded fuel.

Findings by Pahnke and Conte (1969) and Gray and Azhari (1972), conducted in the mid-Atlantic region, Chicago, and the eastern U.S., were closer to the Baton Rouge findings by Wintringham et al. (1972): there were virtually no exhaust repairs for vehicles using unleaded fuel. (The average muffler replacement rates for leaded fuel cars among the different studies also varied greatly, ranging from one per 20,500 miles to one per 154,900 miles.)

Weighting the Wintringham et al. (1972) findings for Detroit and Baton Rouge according to the proportion of registered cars in Sunbelt versus Snowbelt states in 1982 (43% and 57%, respectively), mufflers nationally would last an average of three times longer on unleaded vehicles than on leaded ones. However, because of concern that those limited duration studies may have underestimated muffler replacements over the lives of vehicles using unleaded fuel, Schwartz et al. (1984) conservatively concluded that mufflers on vehicles using unleaded fuel would last twice as long (in miles) as those on vehicles using leaded fuel. Given a 100,000 mile projected lifetime of a car, that meant about two exhaust system replacements per leaded vehicle versus one per vehicle using unleaded fuel. To supplement their conclusions based on the fleet studies, Schwartz and his colleagues consulted several automotive specialists to confirm the reasonableness of their assumptions.

Exhaust system replacement savings were calculated as follows: for leaded exhaust systems replaced once every 50,000 miles, each mile accounts for .00002 of the system replacement; for unleaded vehicles replaced once every 100,000 miles (doubled exhaust lifetime), the system replacement figure is .00001 replacements per mile. The difference is .00001 exhaust system replacements per mile. At $120 per repair (muffler, tailpipe, and exhaust pipe), that was 0.12 <u>cents</u> per mile or 1.68 <u>cents</u> per gallon of gasoline (at the average 14 miles per gallon achieved by cars in the late 1960s). A savings of 1.68 cents per gallon times light-duty vehicle demand (22.8 billion gallons of leaded gasoline in 1988) yielded exhaust system savings of $383 million for the complete-ban-of-lead alternative (in 1983 dollars).[13]

Now, turn to the low-lead alternative restriction. Most of the studies used to estimate maintenance savings involved fleets of vehicles, half of which used commercially available leaded gasoline. In the late 1960s, when those studies were conducted, the weighted average lead content of gasoline (weighted by proportions of premium and regular grades) was about 2.3 grams of lead per gallon (g. lead/gal). Unfortunately, because the discussion at the time focused on relatively high lead levels versus "zero" lead, there were very few data with which to define the relationship between low lead concentrations and exhaust system corrosion between 2.3 and 0 g. lead/gal.

Gray and Azhari (1972) was the only study that examined exhaust system corrosion rates at levels as low as 0.5 g. lead/gal of gasoline. They found no difference between corrosion rates at 2.3 g. lead/gal and 0.5 g. lead/gal, with corrosion rates at both lead levels being 10 to 20 times higher than those for vehicles using unleaded gasoline. That suggested a threshold at or below 0.5 g. lead/gal, below which lead levels must fall before any savings can be achieved by fewer exhaust system replacements. It also suggested that no savings were achieved from previous "lead phase-downs," since the current (1983) concentration was 1.1 g. lead/gal–well above Gray's and Azhari's upper bound threshold of 0.5 g. lead/gal. With no information to the contrary, Schwartz et al. (1984) assumed a linear relationship between lead levels and exhaust corrosion below 0.5 g. lead/gal, and calculated the benefits from less frequent exhaust system replacement on those vehicles <u>designed to use **leaded** gasoline</u> to be 80% of the benefits calculated for those vehicles in the complete ban on lead alternative (a drop from 0.5 to 0.1 is 80% of the drop from 0.5 to 0). Benefits to those vehicles designed to use <u>unleaded</u> gasoline were assumed to be the same as calculated in the complete ban on lead alternative. The reasoning for the latter assumption was that fuel with a lead content of 0.1 g. lead/gal would be more expensive to produce than unleaded fuel. Coupled with restrictions on leaded fuel availability, unleaded fuel would become the low cost alternative and be easier to obtain. Hence, no one with a car designed to run on unleaded gasoline would use leaded fuel, i.e., misfuel.

Similar to the process followed for estimating benefits from less frequent exhaust system replacement, Schwartz et al. (1984) estimated benefits from less frequent tuneups, as measured by longer spark plug change intervals, and less frequent oil changes for unleaded fueled vehicles. The oil change benefit raises an additional interesting and relevant point in the area of benefit estimation. The fleet studies investigating differences in maintenance costs between leaded and unleaded fueled vehicles tended either to ignore effects on engine oil or find very small savings. In general, those studies were not conducted in a manner to easily determine the effects of using leaded or unleaded gasoline on oil change intervals or engine wear. Hence, the analysis of oil change benefits relied more heavily on experimental studies of engine wear with leaded and unleaded gasoline at varying oil change intervals than on fleet studies. One exception was a fleet study (Pless 1974) specifically designed to examine oil change effects. Schwartz et al. (1984) hypothesized that consumers were not aware of the potential decrease in oil change requirements when using unleaded gasoline, and/or did not tend to change their habitual maintenance behavior (and therefore continue to change oil on the shorter intervals). Even if consumers did not realize the possible short-term cost savings of fewer oil changes, they would benefit from better engine durability over the longer term with unleaded gasoline. The point is that people can benefit from a policy, even if they do not realize they are benefitting, and it is appropriate to include such benefits in the calculation.

5.4.2 Health Benefits of Reduced Lead

The monetized benefits, estimated by Schwartz et al. (1984), of reducing the number of children with blood lead levels above 30 μg/dl (micrograms per deciliter) fall into two categories: (1) avoided costs of testing for and monitoring children with elevated blood lead levels, and medically treating children with very elevated blood lead levels; and (2) costs associated with the cognitive effects of lead exposure resulting in blood lead above 30 μg/dl. Those benefits were estimated on a per-child basis and applied to the estimated number of children affected in the one year covered by the benefit-cost analysis.

Using regression analysis based on survey data collected for the National Center for Health Statistics, Schwartz et al. (1984) estimated the reduction in numbers of children having blood lead levels over 30 μg/dl to be 43,000 for the low-lead alternative and 45,000 for the complete ban on lead alternative. The data contained health and demographic information for a representative sample of the U.S. non-institutionalized civilian population aged 6 months through 74 years. Sixty percent of the sample were asked to provide a blood sample, which was tested for lead content. Other data included total lead used in gasoline production. Further details can be found in Schwartz et al. (1984). Even though the discussion focused on airborne lead coming from vehicle emissions, airborne lead is eventually deposited on land, water, buildings, and other surfaces where it can be ingested in addition to being inhaled.

It was assumed that all children with blood lead levels above 30 μg/dl would receive follow-up medical attention and/or immediate medical treatment. Unfortunately, however, many–maybe most–children with elevated blood lead levels are not detected, although their lives are adversely affected. Those children who go untreated bear a burden that Schwartz et al. (1984) valued equal to the cost of follow-up and treatment. The Schwartz et al. estimates of 43,000 and 45,000 fewer children with high blood lead, given above, included those who would go undetected.

Three categories of basic follow-up and treatment were defined based on severity of lead poisoning. The least severe category (30% of children with blood lead levels above 30 μg/dl) was assumed to receive one follow-up blood test. Children in the middle category (47%) were assumed to receive six regularly scheduled blood tests, and about half those children would also have a county sanitarian visit their homes to evaluate possible sources of lead exposure. They would also be examined and evaluated by a physician. The most severe category (23% of children with blood lead levels above 30 μg/dl) was assumed to receive a three-day hospital stay for testing, and have a county sanitarian inspect their homes. Furthermore, they would receive six monthly follow-up blood tests, followed by six quarterly follow-up blood tests. Finally, all children in the most severely affected category were assumed to receive a neurological examination, and one-third would undergo further testing and chelation treatment to remove lead from the body. The assumed courses of treatment were based on recommendations from the Centers for Disease Control (CDC). Using EPA data and data from the Department of Health and Human Services, Schwartz et al. (1984) estimated the weighted average medical cost per child with blood lead level over 30 μg/dl to be $950 (across all levels of severity) for the one-year period being evaluated.

Evidence for cognitive effects of lead in children with blood lead levels above 30 μg/dl was fairly strong. Furthermore, research showed that these effects remained three years later. Schwartz et al. (1984) estimated that about one-third of children with blood lead levels above 30 μg/dl (those falling into their most severe category above, and some in the middle category) would be affected severely enough so as to make them seven times more likely to be forced to repeat a grade in school. It would not be possible to completely restore those children's performance; hence, lifetime work and production could be affected. However, tutoring, reading teachers, school psychologists, and the like could help improve their achievement in school. As a result, Schwartz et al. used the cost of such supplemental education as a proxy for the avoided cost resulting from reduced lead poisoning.

Given the finding of at least a three-year persistence in cognitive effects of high lead levels, the cost of correcting those cognitive effects (or cost avoided if the high lead levels were reduced) would be at least three years of compensatory education. An average of one year of compensatory education per child with blood lead level over 30 μg/dl was assumed (one third of children would need three years of such education) as the proxy for avoided cost resulting from reduced lead in gasoline. Data from the U.S. Department of Education, Office of Special Education Programs, were used to estimate that cost ($4,290 per child per year). That figure was quite close to an independent estimate of the special education cost for non-retarded lead exposed children, cited by Schwartz et al. (1984).

The estimates of benefits (avoided cost) did not include adverse effects in adults or infants under six months, although emerging data at the time suggested that fetuses and newborn infants may be most vulnerable to lead effects. They did not include non-neurological effects such as kidney damage, anemia, and other medical problems. They did not include such welfare losses as parents' lost work time; adverse health effects of chelation (such as removal of beneficial minerals along with the lead for those who undergo that treatment); or behavioral problems that could alter attention span or in more overt forms, such as serious behavioral abnormalities, affect the education of other children in the classroom. Finally, non-quantifiable effects such as pain from the treatments were not evaluated. Hence, it was argued by Schwartz et al. (1984), that their benefit estimates were conservative.

5.4.3 Results in Summary

The cost and benefit estimations calculated by Schwartz et al. (1984) are summarized in Table 3. Note that some benefits were not estimated in monetary terms and are represented as letters. The low-lead level restriction showed a net benefit of $786 million, and the complete ban on lead showed a net benefit of $704 million, both were for the year 1988 expressed in 1983 dollars.

Subsequent to the draft benefit-cost analysis (Schwartz et al. 1984) some re-analyses and recalculations were done (Nichols 1997), resulting in a net benefit for the low-lead alternative restriction of $803 million, compared to $786 million. Furthermore, the final Regulatory Impact Analysis issued by EPA

Table 3. Comparison of Benefits and Costs of Lead Reduction Options in 1988 (millions of 1983 dollars)

COSTS AND BENEFITS	Low-lead Option*	All Unleaded**
COSTS		
Manufacturing Costs	$503	$691
Non-monetized Valve Damage to Engines that Need Lead	----	D
TOTAL COSTS	$503	$691+D
BENEFITS		
Maintenance Benefits	$660	$755
Environmental and Health Benefits		
Conventional pollutants		
Reduced damage by eliminating misfueling	$404	$43
Non-monetized health benefits†	H_1	H_1
Lead		
Reduced medical care costs	$41	$43
Reduced cognitive damage	$184	$193
Non-monetized health benefits††	H_2	H_3
TOTAL BENEFITS	$1,289+$H_1$+$H_2$	$1,395+$H_1$+$H_3$
NET BENEFITS	$786+$H_1$+$H_2$	$704+$H_1$+$H_3$-D

* The option would make a low lead gasoline (0.10 grams of lead per gallon) available only for those few vehicles that required some lead. It assumes no misfueling.

** All lead in gasoline would be banned by 1988.

† These include chronic health effects of ozone and CO, and any effects of reduced sulfate particulates.

†† Since medical costs and cognitive damage were monetized only for children with high blood lead (>30 μg/dl), H_2 and H_3 represent other benefits for this group (pain, lost work time by parents, etc.) for the lower lead group (<30 μg/dl). H_2 and H_3 differ because the numbers of children at risk under the two options differ.

Source: Schwartz et al. (1984, p. I.9)

included estimates for the period 1986–1992 rather than just 1988 (Nichols 1997), which further increased net benefits for both alternatives.

Also subsequent to the draft, new research was published that linked blood lead levels in adults with blood pressure. Schwartz, a coauthor on one of those studies (Pirkle et al. 1985), used the results to estimate additional benefits from the two alternative restrictions on lead. The calculations considered reduced incidence of hypertension but, more importantly, they also estimated reduced incidence of mortality and morbidity due to cardiovascular disease associated with elevated blood pressure. The blood pressure results dramatically increased the estimated net benefits of the restrictions under both alternatives; even with fairly conservative assumptions, those blood pressure benefits were roughly three times larger than all the other benefit categories combined.

Because they had not been subjected to the level of peer review to which the other components of the benefit-cost analysis had been, EPA decided that the blood pressure benefits could be included in the Regulatory Impact Analysis, but that net benefits would be calculated both with and without the blood pressure benefits. The case for the restrictions had to stand even if blood pressure benefits were excluded (Nichols 1997).

5.5 Use of Results Their Range of Influence

Nichols (1997) summarized the outcome of EPA's consideration of the alternative restrictions on lead in gasoline.

> In March 1985, EPA promulgated a rule to slash the use of lead in gasoline by over an order of magnitude (from 1.1 g. lead/gal to 0.1 g. lead/gal) in less than a year. . . . EPA's decision was supported by an extensive benefit-cost analysis. That analysis indicated that lowering lead in gasoline would reduce serious health effects related to elevated lead levels in blood; such effects were clearest for children, but new evidence suggested that adults might reap even larger benefits related to reduced blood pressure levels. The rule was also expected to reduce "misfueling"—the use of leaded gasoline in cars required to use unleaded gasoline—and its associated damage to pollution-control catalysts in new motor vehicles. Protecting these catalysts would reduce emissions of three other pollutants: hydrocarbons, nitrogen oxides, and carbon monoxide. . . . The RIA (Regulatory Impact Analysis) prepared by EPA estimated that the monetized benefits of the

> rule would exceed its cost by roughly a factor of three-to-one, even if one attached no weight to recent studies linking elevated lead levels to high blood pressure and consequent cardiovascular disease. If those benefits were included, the benefit-cost ratio rose to more than ten-to-one. (Nichols 1997, p. 49)

The 1985 rule proved even more effective than anticipated in reducing lead in gasoline, despite the fact that lead was not completely banned. The combination of the phase down and declining numbers of engines on the road that were allowed to use lead rapidly reduced the availability of leaded gasoline and its sales. Many gas stations stopped carrying leaded gasoline, using the freed up tanks and pumps to offer an intermediate grade of unleaded (Nichols 1997). In the Clean Air Act Amendments of 1990, Congress formalized the demise of leaded gasoline by banning lead as a fuel additive effective December 31, 1995.

6. LAKEFRONT PROPERTY VALUES IN MAINE

This section draws heavily from Boyle et al. (1998) and Boyle (2002).

6.1 Research Issues/Questions

Although Maine is known for clear and high quality lakes, lake water quality was threatened. At the time of the study (1997–98), 260 lakes and ponds totalling more than 238,188 acres did not meet federal and state standards for swimming, aquatic life support, or trophic status (which indicates the level of photosynthetic activity, i.e., algae and aquatic plant growth). Another 44,004 acres were considered to be unimpaired but threatened. The threat to lake water quality was due mostly to nonpoint source pollution originating as excess runoff from development, silviculture, and agriculture. The general symptom of the resulting eutrophication (or increased nutrient loading) was manifested as increased photosynthetic productivity, primarily in the form of algal growth. Excess algae leads to decreased water transparency and reduced oxygen content in the water, often causing changes in the lake's biological community such as in the distribution of fish species. "Cultural eutrophication,"—eutrophication that does not occur naturally but is induced by human activity—was the most preventable cause of poor water quality in Maine's lakes.

Although eutrophication was characterized by changes in water quality measures such as dissolved oxygen and chlorophyll levels, water clarity was the manifestation most observable to the public. Water clarity was measured using a Secchi disk.[14] Reduced water clarity associated with eutrophication reduced a lake's aesthetic appeal, decreased recreational benefits, and lowered the prices of properties around the lake.

Protecting lake water quality was not without costs, and monies to prevent or reduce eutrophication had been allocated with little or no information about the economic benefits of protecting lake water clarity. During the 1980s, $80,000 to $250,000 per year were allocated by the state for lake protection and restoration. Information about the economic effects and benefits of protecting lake water clarity would have been useful in prioritizing lake management efforts and in developing public education programs to reduce eutrophication.

Lakefront property owners were likely to receive the greatest economic gains from improved lake water clarity because the benefits could be capitalized into the prices of lakefront property. Those same lakefront property owners could also directly affect lake water clarity by actions they took on their properties.

The research questions, then, involved: What variables explain variations in sales prices for lakefront property, and where does water clarity fit within those variables, i.e. what are the implicit prices of water clarity? What actions can landowners take on their properties to protect water clarity and directly affect the values of their properties? Finally, can a demand equation be estimated to portray the marginal amounts that people are willing to pay for cleaner lakes? Such a demand equation could be used to measure the benefits (losses) from improved (degraded) water clarity for benefit-cost analysis of lake protection efforts.

6.2 Tradeoffs

Trophic status of a lake affected the aesthetic appeal of the lake, but it also affected the uses one could make of the lake. Swimming, boating, and fishing, among other uses, could be adversely affected or precluded entirely by high levels of algae and other plant growth. Lakes with high levels of plant growth could be restored, but restoration was technically difficult and expensive.

Preventing a lake *from becoming* eutrophic (characterized by high plant growth) was less expensive, but had its own difficulties. About 12% of Maine's lakes were classified as eutrophic; another 9% of the lakes were classified as oligotrophic (characterized by a low level of plant growth); the remaining 79% were mesotrophic—somewhere in between. Trophic status of a lake was affected by age and shape of the lake, geology of the watershed, ratio of watershed area to lake area, the flushing rate of water through the lake, human impact, and other factors. Some elements could be controlled by landowners and other citizens, others could not. The more informed property owners were about the nonpoint source pollution causes of lake eutrophication and the benefits they could enjoy by protecting lake water quality from cultural eutrophication, the greater their incentive would be to take voluntary action to prevent nonpoint source pollution and support lake protection regulations. Effects of water clarity on the price of lakefront property could provide a substantial incentive for individual property owners to take actions to protect lake water clarity.

6.3 Design and Implementation

As discussed by Taylor in Chapter 10 of this book, hedonic models can be used to estimate the share of property prices attributable to specific characteristics, such as water clarity. Individuals act to select the property with the most desirable bundle of characteristics. All other characteristics equal, one would expect people to pay more for a property on a lake with high water clarity than they would for a property on a lake with lower water clarity.

Hedonic price equations were estimated to express property price as a function of property characteristics, characteristics of structures on the property, characteristics of the property location, and the natural log of water clarity multiplied by surface area of the lake. Water clarity was expressed as a natural log to reflect the expected nonlinearity of the relationship between property price and water clarity. While a modest change in water clarity is quite noticeable at low levels of water clarity, a similar change in water clarity on a very clear lake can be imperceptible.

Water clarity was measured as the minimum clarity during the summer months (June 1 through September 30) of the year the property sold. The interaction between water clarity and lake surface area was used because of collinearity between those variables. That specification implied that water clarity was more important to consumers of lakefront property on larger lakes. Boyle et al. (1998) supposed that people may be willing to lower their standard for water clarity to locate on a small lake to avoid boat traffic and other activities that occur on a larger lake.

In the second stage of the hedonic model, the demand for water clarity was expressed as a function of implicit prices for water clarity, square feet of living area, feet of lake frontage, income, and dummy variables for whether: (1) the person visited the lake before purchasing the property, (2) the person expected lake water clarity to improve in the future, (3) the person expected lake water clarity to decline in the future, and (4) friends or relatives of the person owned property on the lake. Data on income and the dummy variables came from a mail survey of property purchasers. Living area was presumed to be a complement to water clarity, so that as the implicit price of living area increased, people would choose a lower level of water clarity. Feet of lake frontage was presumed to be a substitute for water clarity, so as the implicit price of frontage increased, people would choose a higher level of water clarity, i.e., people will choose to purchase property on a lake with a higher level of water clarity but with less lake frontage. Expected signs on the first three dummy variables were indeterminate, while Boyle et al. (1998) supposed people would be more likely to purchase property on a lake where friends or family owned property.

Data were collected on sale prices and property characteristics from local communities selected for the study and the Maine Bureau of Taxation. Characteristics included: feet of frontage on the lake, square feet of living area, number of bathrooms, source of water, presence of central heating, lot size, distance to nearest city, the lake the property was on, and lake area.

Data on water clarity came from the Maine Department of Environmental Protection (DEP) as Secchi disk readings of the minimum clarity in the lake during the summer months of the year the property was sold. Because some lakes were not monitored as closely as others, a regression equation was estimated for each market area to predict minimum water clarity for observations without water clarity data.

The data for the study covered 36 lakes encompassing 64 organized towns. Lakes were grouped into seven market areas.[15] Data for each market area comprised 41 to 258 individual property sales. The purpose of the grouping was to allow estimated implicit prices for water clarity to vary across real estate markets. All the market areas but one contained one or more lakes that had undergone restoration projects that involved media coverage of water quality problems and causes.

6.4 Results

Property values were highest in the area of Auburn, Maine and lowest in the northern Maine market. Average minimum water clarity was also highest for the Auburn group of lakes and lowest for the northern Maine group.

Separate hedonic equations were estimated for each market area. That allowed the implicit price for water clarity to vary across groups of lakes so as to reflect differences in market conditions. Water clarity was a significant predictor of variations in property prices for four of the seven market groups.

Boyle et al. (1998) noted that policy questions, most often consider the effects of incremental changes in water clarity, not marginal changes. For example, how much would property prices increase if water clarity improved from 2 meters of transparency to 4 meters? Based on their model, Boyle et al. were able to estimate implicit prices for (or the share of property price attributable to) water clarity on each lake, under existing conditions and under hypothesized changes in conditions of water clarity. For example, about 4.9% of the purchase price in the China Lake area was attributable to water clarity; that amounted to $4,456 on average at the existing water clarity level of 1.76 m. of transparency. At Pushaw Lake, with water clarity of 3.32 m. of transparency, about 45% ($25,542) of the purchase price was attributable to water clarity. Property value effects of changes in water clarity were the differences in implicit prices based on the before and after levels of water clarity. If water clarity in Pushaw Lake improved to 4.32 m. of transparency, property prices would be expected to increase by $5,604. A degradation to 2.32 m. transparency led to an expected decrease of $7,629. Aggregating such calculations indicated that a change in water clarity could result in millions of dollars in gained or lost property values on Maine's major lakes.

The estimated demand function for water clarity indicated a hyperbolic shape. The Maine DEP staff suggested the threshold of recognition of a lake with compromised water clarity is about 3 meters. Buyers' preferences followed that threshold quite closely. At water clarity levels less than 3 meters of transparency, the demand function was very steep, indicating people were willing to pay substantial amounts for improved water clarity. At clarity levels above 3 meters of transparency, the demand function was relatively flat, indicating people were willing to pay less for additional units of water clarity. The shape of the demand function was consistent with the fact that it is more difficult for people to observe changes in water clarity at higher thresholds of clarity. It also indicated that economic losses from reduced water clarity were greater than corresponding gains from increased clarity.

While the hedonic price equations revealed the effect of water clarity on property prices, the area below the demand function measured the economic benefit to property owners from changes in water clarity. Such areas under the demand function between different levels of water clarity are useful for benefit-cost analyses of lake protection efforts. Both the hedonic price equations and the estimated demand function suggested that, from an economic perspective, it was better to protect water clarity than to allow clarity to decline and then try to reverse the negative trend; each increment of decline resulted in increasing economic losses and greater cost of lake rehabilitation. The findings indicated that it was in property owners' best interest to take actions to protect water quality in their lakes. Such actions included: creating buffer zones, controlling drainage, landscaping so as to minimize nutrient runoff into the lake and intercept nonpoint source pollution as it ran through their properties to the lake, and encouraging their neighbors to do the same.

6.5 Use of Results and Their Range of Influence

To help balance their budget during the economic downturn in the early 1990s, the Maine DEP cut their lakes program. There were questions about whether the program was really accomplishing enough to justify its cost. Restoration of lake water quality to mitigate eutrophication was both technically difficult and expensive, so such questions were easy to ask. However, the lakefront property owners study documented a relationship between water quality and property values, showing real effects of decrements to lake water quality on real people and their pocketbooks. The matter came before the state legislature, where the lakes program was reinstated with increased funding.

Rather than trying to restore water quality in eutrophic lakes, emphasis was more effectively placed on protection and prevention—protecting lakes that do not already show the deleterious effects of advanced eutrophication and preventing further deterioration in lakes that do show those effects. Actions to protect and prevent are most effective at local levels. Results of the study were used at the town level to show local citizens that high water quality was in their own interest and that actions they took to protect lake water quality and prevent further degradation were personally beneficial. Support was gained for programs to educate people about what they could do to help protect water quality and how they could help bring about larger changes. Likewise, recognition of personal benefit from water quality garnered support for zoning regulation and other means of modifying land use to protect lake water quality.

Because lakes often border more than one town, efforts to protect lake water quality helped bring towns together to cooperate on projects. Peer pressure was brought to bear on towns that were lagging in their standards to bring them up to the levels of neighboring towns.

A question that naturally arose was: since lakefront property owners are often wealthy people, why should the state spend a lot of effort to increase or maintain private property values? One answer pointed to secondary benefits from increased (or at least non-decreased) property values, whereby benefits actually accrued to a broader group than just the property owners. Local tax revenue, for example, came largely from property taxes, which were based on property values. The DEP did some calculations illustrating the extent to which higher lake water quality indirectly protected the tax base in local communities. Other studies documented benefits accruing to other segments of the population from maintained or enhanced lake water quality. Recreation users made up one such group; beyond the benefit that accrued directly to recreation users, communities benefitted from the influx of money from nonresidents who used the lake for recreational purposes (Schuetz 1998).

Finally, the Maine Lakefront Property Owners study has been used in the context of benefit transfer. Both the quantitative and qualitative results of the study have been used in Wisconsin and Minnesota to estimate benefits from lakes programs in those states.

7. BUT WAIT, THERE'S MORE . . .

The studies discussed so far have dealt with rather specific management or policy issues and pending decisions, and they have shown how nonmarket valuation can play a direct role. Nonmarket valuation can also affect policy in a more indirect way. Policy and management decision making is a process. As pointed out in the introduction to this chapter, information is a critical input to that process. Nonmarket valuation information is an important component of that information input. It is not, however, the only information input and it is not necessarily the determining input.

The studies described above were cases in which nonmarket valuation analyses were identifiable contributors to the outcome. The valuation analyses showed that benefits greatly outweighed costs, so the policy change should be implemented in the case of lead content of gasoline. The examination and evaluation of tradeoffs contributed to the choice of a preferred alternative in the case of Glen Canyon Dam operation. Angler preferences and the likelihood of an insufficient increase in revenue contributed to fishing license fees not being increased in Colorado. Demonstrated benefits to lakefront property owners resulted in restored funding and focused efforts on education and outreach to local areas to encourage local and personal actions to prevent further degradation of lake water quality in Maine. Those studies have a rather clear trail between analysis and decision.

In other instances, information on nonmarket values is brought to bear on the question, considered by decision makers, and over-ruled in favor of other factors. Does that mean the nonmarket valuation analyses done to address those questions were less than successful, or that they did not make a difference? No! In such cases, information on nonmarket values entered the process and a decision was made involving multiple considerations and criteria. The key is that the decision was informed. Decision makers knew what the tradeoffs were, and who and what would be affected how; and they made their decision. That's how the process is supposed to work.

Not only do nonmarket valuation analyses contribute to the decision making process by helping to clarify tradeoffs and identify magnitudes and directions of effects, sometimes they affect the process itself. Nonmarket valuation, and insights resulting from such analyses, can inform, or even change, the debate about an issue. They might identify new and relevant players or parties to a decision. Bringing those new parties to the decision making table might alter the process well beyond a specific decision. Nonmarket valuation analyses might affect the mindset of decision makers such that they broaden the scope of their consideration of issues that come before them. Decision makers might have been unaware of tradeoffs brought to light in nonmarket valuation analyses, or they might not have recognized the magnitude or reach of particular effects. Again, the process could be altered beyond any specific decision. If one moves the track by only a few degrees, the train ends up at a different destination.

One example of nonmarket valuation studies that altered the process by which decision making takes place is a series of studies conducted in Alaska over a period of years (Swanson, Thomas, and Donnelly 1989, Peterson et al. 1992, McCollum and Miller 1994, Miller and McCollum 1994, Miller, Sinnott, and McCollum 1994, Clayton and Mendelsohn (1993), Miller and McCollum 1997, McCollum, Haefele, and Miller 1998, Miller, Miller, and McCollum 1998, Miller and McCollum 1999, and others).[16] The project began as an effort to assess the net economic value of wildlife-related activities in Alaska and their contribution to the state's economy. It included both Alaska residents and nonresident visitors to Alaska. Tangential efforts included a study linking wildlife to quality and net economic value of visitor experiences at Denali National Park, and one on bear viewing at McNeil River. Both the content and the volume of studies were important in building a cumulative effect on the Alaska Department of Fish and Game (ADFG) and its client groups—some traditional client groups such as hunters, and some nontraditional client groups such as wildlife viewers and ecotourism operators. The studies were done in a collaborative process that involved interagency steering committees with memberships that included: all the federal land management agencies in Alaska; all the state-level resource management agencies; and, later in the process, representatives from the tourism industry, Alaska Native groups, and the environmental community. The broad-based and participatory nature of the process further added to the cumulative (and ongoing) effects of the studies.

Results from that series of nonmarket valuation studies have played a role in revising policies for managing wolves and bears in Alaska, both in terms of suspending some regulations and designing succeeding regulations. They have also played a role in establishing a nonconsumptive wildlife program within ADFG. More importantly, the studies have contributed to an awareness within ADFG and other agencies that nonmarket valuation and social-based analyses have an important role in providing decision makers complete information on effects, tradeoffs, and consequences of alternative management actions. The following is taken from a letter, dated July 25, 1995, by Alaska Governor Tony Knowles to Bruce Alberts, president of the National Academy of Sciences, requesting the Academy undertake a scientific and economic review of management of wolves and bears in Alaska:

> When I took office late last year, I suspended the state's current wolf control program after it was revealed that the program was unacceptable in its treatment of both wolves and nontargeted species. As Governor, I believe I have a responsibility to see that Alaska's game management program meets three tests: It must be based on solid science, a full cost-benefit analysis must show that the program makes economic sense for Alaskans, and it must have broad public support. I will not reinstitute predator control measures unless and until they meet those three tests.
>
> There will always be disagreement about predator control from ethical and other perspectives. **However, public confidence in the science and socioeconomics upon which management is based can go a long way toward public acceptance of a management program** . . . (Quoted in National Research Council 1997, p. 196, bolding added.)

8. CONCLUSIONS

Nonmarket valuation can affect policy or management in several ways. A study can be methodological in nature and contribute to the technology by developing new, or refining existing, methods of valuation as was illustrated in previous chapters of this book. Or, a nonmarket valuation study can address a particular management or policy issue and contribute to specific decisions, such as the four studies discussed in this chapter. Finally, nonmarket valuation studies can affect the process by which policy/management decisions are made; the Alaska studies just discussed are an example. At the same time, there are several reasons, discussed in the anecdotal overview in this chapter, why nonmarket valuation analyses might not be considered. One job of economists is to conduct nonmarket valuation analyses in ways that maximize credibility of the study and results. That's where this book comes in.

All the chapters previous to this one presented information and guidelines aimed at accomplishing a high quality analysis using the variety of methods discussed. Freeman, in Chapter 1, set the stage and talked about reasons why one might go through a nonmarket valuation exercise in the first place. To frame the questions and build an analytical framework, one must be anchored in relevant theory. Flores, in Chapter 2, laid out the basics of economic theory relating to and underlying nonmarket valuation. The foundation of any nonmarket valuation analysis is data. Without valid and reliable data, the results will be neither valid nor reliable. In the absence of validity and reliability, the analysis will never be taken seriously. Champ, in Chapter 3, laid out the process one would go through to design a data collection effort or to evaluate the quality and relevance of an existing data set. Those components of theoretical framework and valid, reliable data are common to any analysis. Chapters 4 through 11 laid out the menu of methods available in the state-of-the-art nonmarket valuation toolbox. The various authors talked about basics of the methods and how to use them. They discussed issues one needs to be aware of and analytical decisions that need to be made along the way.

Then, in Chapter 12, Rosenberger and Loomis talked about benefit transfer and applying the results of nonmarket valuation analyses done at one place and time "off the shelf" to other situations. Nonmarket valuation studies can be expensive and time consuming. Benefit transfer hints that there might be less expensive and more timely options under certain conditions. Can benefit transfer provide relevant and useful information, and bring that to bear on a pending policy or management decision? Absolutely. Can benefit transfer be applied in all cases? No. One needs to consider the individual situation, and systematically evaluate the relevance and applicability of studies to be potentially transferred.

Finally, this chapter brings things together a little and demonstrates that yes, nonmarket valuation analyses do get used and do contribute to policy and management decisions. In some cases, they even contribute to altering the courses of resource management agencies and other institutions. In order to achieve those effects, studies must be anchored in the principles discussed in previous chapters of this book. Another lesson one can glean from the preceding chapters is that the methods are not static. They continue to develop and evolve over time. In the concluding chapter, Bishop discusses where nonmarket valuation is headed, and issues to be considered in the future.

ACKNOWLEDGMENTS

Randy Rosenberger and John Loomis reviewed this chapter as part of the process that created this book. They provided several helpful suggestions. Patty Champ, Tom Brown, and Kevin Boyle also reviewed the chapter and provided many helpful suggestions. Mary Carr's editorial comments improved the clarity of the text in several places. All their contributions are gratefully acknowledged, but remaining errors or lapses in clarity are mine alone.

NOTES

1 For example, Krutilla and Fisher (1975) describe some of the relatively early studies related to nonmarket valuation–Hell's Canyon Dam, skiing in the Mineral King Valley, the Trans-Alaska pipeline, among others. Those are classic studies that are used to teach graduate students, but details are sparse as to how they entered the decision process.

2 One such example is the Adamowicz et al. study (1994). The only hint of the larger study, out of which the journal paper came, is in the data section of the paper, which tells readers that data came from a "multiphase survey sponsored by Alberta Environment." The larger report on the "Little Bow River Project/Highwood River Diversion Plan" that was cited in the paper could not be easily found. In other communications (Boxall 2001), I learned more about that larger study. Both the revealed and stated preference models that were reported in the journal paper were part of a benefit-cost analysis of an irrigation expansion project. Results of that analysis indicated net benefits to recreational fisheries. Ultimately, the irrigation expansion was approved (according to the Natural Resources Conservation Board web site in Canada). This is not uncommon for published studies—it is difficult to draw a line from published study to use in a decision process unless one has a specific connection to the study.
(An epilogue: Later on, the larger Little Bow River report was located on a web site: http://www.nrcb.gov.ab.ca/dec9601.html.)

3 I should point out that "passive-use" value is also referred to as "nonuse" value by some economists. That is nonuse in the sense of no active or direct use, it does not mean useless.

4 Smith also pointed to an exchange between Chapman, Hanemann, and Ruud (1998) and Dunford (1999) in the AERE (Association of Environmental and Resource Economists) newsletter as "an interesting description of the case from the perspectives of the plaintiff and defense witnesses."

5 RPA refers to the Forest and Rangeland Renewable Resources Planning Act of 1974 which, among other provisions, directs the Forest Service to periodically prepare a "Recommended Program" laying out a management strategy for the agency to follow in managing national forests and grasslands over the next period of years. In the planning

process underlying the Recommended Program, alternative strategies and their associated outputs of forest management are evaluated. One of the outputs is recreation. The "RPA recreation values" are a set of net economic values expressing the benefit received by participants in a variety of recreation activities. Peterson et al. (1987) provide more detail about the RPA values.

6 Electrical power is on a grid system such that power produced at one point can be transmitted over a wide area. A reduction in electricity generated at one point, e.g., Glen Canyon Dam, can be compensated by increasing generation at some other point on the grid. But different costs are associated with generation at different power plants.

7 Validity of a study refers to its being relevant and meaningful. Borrowing from Bishop in Chapter 14, content validity is about procedures. Were the procedures and methods used in the study conducive to measuring respondents' "true" value? Construct validity is about measures and relationships between measures. Theoretical construct validity focuses on theoretical and intuitive expectations about relationships that should exist between variables, and whether empirical measures are consistent with theory. Convergent construct validity focuses on the convergence between a measure and other measures of the same theoretical construct; the extent to which different measures of the same thing agree or are consistent. Questions can be designed into surveys to provide data with which to compare variables and test hypotheses related to validity, thus giving added credence to study results.

8 The average household values used to aggregate up to the national totals were based on the conservative approach of accepting only responses of "definitely yes" in favor of the alternative as positive responses to the WTP question when estimating the probability models from which values were calculated. Those who responded "probably yes" and those who were unsure were counted among the "no" votes. Values of those who said they would vote against the alternative level of dam operation or would not vote, even if changing dam operation would cost them nothing, were assumed to be zero. Those people were skipped around the WTP question and not included in the estimated model of the probability of a yes vote as a function of cost imposed by the alternative.

9 With apologies to Laura Taylor for using yet another fishing study.

10 The deposits from leaded gasoline formed a coating on exhaust valve seats. On pre-1971 cars and non-diesel trucks, that thin layer protected against the abrasive and adhesive wear that can occur between the exhaust valve face and valve seat. Drivers and mechanics had grown up with this knowledge that lead additives performed that lubricating function, so it was part of the conventional wisdom. The fact was that by 1971, several major engine manufacturers were building engines with valve seat metallurgy that eliminated or minimized such valve wear with unleaded gasoline. Yet, perceptions are sometimes hard to change.

11 Having gas pumped by attendants at full service stations would ensure that cars designed for unleaded did not get leaded fuel. Another incentive to avoid misfueling was to make gas pump nozzles for unleaded fuels smaller than those for leaded fuels and place a restrictor in the inlet tube on cars' gas tanks such that pump nozzles for leaded gasoline would not fit into the tank inlet on cars designed for unleaded gasoline.

12 Scavengers, primarily ethylene dibromide (EDB) and ethylene dichloride (EDC), were necessary to remove lead deposits from engines. In the absence of these scavengers engine performance would rapidly deteriorate to complete inoperability. While scavengers effectively removed lead deposits from the combustion chamber, significant amounts of lead were still deposited on internal engine and exhaust system surfaces.

Such deposits were very corrosive in the warm and humid conditions within engines and exhaust systems. Hence, reduction or removal of lead would result in a reduction in maintenance requirements.

13 Changing to savings per gallon, then extrapolating to 1988 via changes in leaded gasoline demand, automatically adjusted for changes in fuel economy and changes in miles per year among different cohorts of vehicles. Vehicles traveling fewer miles burned fewer gallons and, hence, acquired fewer savings. Likewise, vehicles with better fuel economy achieved lower savings than average. It should be noted that the benefit estimates assumed that savings were a function of fuel use. Given the trend (in the early 1980s) towards more fuel efficient cars, such an assumption may considerably underestimate actual benefits, as the age of the vehicle becomes an important variable in determining the life of a muffler. Implicit in the Schwartz et al. (1984) model was an assumption that an automobile getting 28 miles per gallon (mpg) would need a new muffler every 200,000 miles, or at 42 mpg, 300,000 miles. To the extent that these muffler lifetimes are overestimated, benefits are underestimated. Schwartz et al. were unable to find data related to the effects of time on muffler life.

14 Secchi disks are round disks that are white and black on alternating quadrants. The disks are lowered into the water on a metered line. The depth at which the disk disappears from sight is a measure of water clarity.

15 A market area was defined as a group of lakes in close proximity to each other and near a large community. The seven markets selected were: Lewiston/Auburn, Augusta, Waterville, Newport, Ellsworth, northern Aroostook county, and Camden.

16 Certainly this is not the only example. It is just one I know about.

REFERENCES

Adamowicz, W., J. Louviere, and M. Williams. 1994. Combining Revealed and Stated Preference Methods for Valuing Environmental Amenities. *Journal of Environmental Economics and Management* 26(3):271-292.

Bishop, Richard C. 2002. Personal communication.

Bishop, R. C., and T. A. Heberlein. 1979. Measuring Values of Extramarket Goods: Are Indirect Measures Biased? *American Journal of Agricultural Economics* 61:926-930.

Bishop, R. C., and M. P. Welsh. 1998. Contingent Valuation: Incorporating Nonmarket Values. In *Better Environmental Decisions: Strategies for Governments, Businesses, and Communities*, pp. 177-193. Edited by Sexton, K., A. A. Marcus, K. W. Easter, and T. D. Burkhardt. Washington, DC: Island Press.

Boxall, Peter. 2001. Personal communication.

Boyle, Kevin J. 2002. Personal communication.

Boyle, K. J., S. R. Lawson, H. J. Michael, and R. Bouchard. 1998. Lakefront Property Owners' Economic Demand for Water Clarity in Maine Lakes. Miscellaneous Report 410 (ISSN 1070-1516). Orono, ME: University of Maine, Maine Agricultural and Forest Experiment Station.

Brown, Gardner M., Jr. 2002. Personal communication.

Chapman, D. J., W. M. Hanemann, and P. Ruud. 1998. The American Trader Oil Spill: A View from the Beaches. AERE Newsletter 18(2):12-25.

Clayton, C., and R. Mendelsohn. 1993. The Value of Watchable Wildlife: A Case Study of McNeil River. *Journal of Environmental Management* 39:101-106.

Cordell, H. K., and J. C. Bergstrom. 1991. A Methodology for Assessing National Outdoor Recreation Demand and Supply Trends. *Leisure Sciences* 13(1):1-20.

District of Columbia Court of Appeals. 1989. State of Ohio v. U.S. Department of the Interior, 880 F.2d 432 (DC Cir.).

Donnelly, D. M., J. B. Loomis, C. F. Sorg, and L. J. Nelson. 1985. Net Economic Value of Recreational Steelhead Fishing in Idaho. Resource Bulletin RM-9. Fort Collins, CO: USDA Forest Service, Rocky Mountain Forest and Range Experiment Station.

Dunford, R. W. 1999. The American Trader Oil Spill: An Alternative View of Recreation Use Damages. AERE Newsletter 19(1):12-20.

Economic Studies Branch, Office of Policy Analysis, Office of Policy, Planning and Evaluation, U.S. Environmental Protection Agency. 1987. EPA's Use of Benefit-Cost Analysis: 1981-1986. Report number: EPA-230-05-87-028.

Gray, D. S., and A. G. Azhari. 1972. Savings and Maintenance Dollars with Lead-Free Gasoline. SAE (Society of Automotive Engineers) Paper #720084 (January).

Hammack, J., and G. M. Brown, Jr. 1974. *Waterfowl and Wetlands: Toward Bioeconomic Analysis.* Baltimore: Johns Hopkins University Press for Resources for the Future.

Hickling Partners, Inc. 1981. Final Report on the Assessment of the Economic Impact on the Automotive Parts/Service Industry of Proposed Gasoline Lead Content Regulations. Submitted to Policy, Planning, and Assessment Directorate, Environment Canada (March).

Kaiser, H. Fred, Jr.. 2001. Director (retired), USDA Forest Service, Resource Valuation and Use Research. Personal communication.

Krutilla, J. V., and A. C. Fisher. 1975. *The Economics of Natural Environments.* Baltimore: Johns Hopkins University Press for Resources for the Future.

Loomis, J. B., J. Cooper. 1988. The Net Economic Value of Antelope Hunting in Montana. Bozeman, MT: Montana Department of Fish, Wildlife and Parks.

Loomis, J. B., and D. M. Donnelly, C. F. Sorg, L. Oldenburg. 1985. Net Economic Value of Hunting Unique Species in Idaho: Bighorn Sheep, Mountain Goat, and Antelope. Resource Bulletin RM-10. Fort Collins, CO: USDA Forest Service, Rocky Mountain Forest and Range Experiment Station.

Loomis, J. B., J. Duffield, and R. Brooks. 1987. The Net Economic Value of Fishing in Montana. Bozeman, MT: Montana Department of Fish, Wildlife and Parks.

McCollum, D. W., M. A. Haefele, and S. M. Miller. 1998. Attributes and the Value of a Recreation Experience: A Preliminary Analysis of Wildlife Viewing in Denali National Park. In: Jakus, P. M. (compiler), Western Regional Research Publication, W-133, Benefits and Costs of Resource Policies Affecting Public and Private Land, Eleventh Interim Report, pp. 179-200. Knoxville: Dept. of Agricultural Economics and Rural Sociology, University of Tennessee.

McCollum, D. W., M. A. Haefele, and R. S. Rosenberger. 1999. A Survey of 1997 Colorado Anglers and Their Willingness to Pay Increased License Fees (Project Report No. 39). Project Report for the Colorado Division of Wildlife. Fort Collins, CO: Colorado State University, Human Dimensions in Natural Resources Unit, and USDA Forest Service, Rocky Mountain Research Station. 105 pp. (Also available at http://www.fs.fed.us/rm/value.)

McCollum, D. W., and S. M. Miller. 1994. Alaska Voters, Alaska Hunters, and Alaska Nonresident Hunters: Their Wildlife Related Trip Characteristics and Economics. Anchorage: Alaska Dept. of Fish and Game, Division of Wildlife Conservation.

McCollum, D. W., G. L. Peterson, J. R. Arnold, D. C. Markstrom, and D. M. Hellerstein. 1990. The Net Economic Value of Recreation on the National Forests: Twelve Types of Primary Activity Trips Across Nine Forest Service Regions. Research Paper RM-289. Fort Collins, CO: USDA Forest Service, Rocky Mountain Research Station.

Miller, S. M., and D. W. McCollum. 1994. Alaska Voters, Alaska Hunters, and Alaska Nonresident Hunters: Their Characteristics and Attitudes Towards Wildlife. Anchorage, Alaska: Dept. of Fish and Game, Division of Wildlife Conservation.

Miller, S. M., and D. W. McCollum. 1997. Alaska Nonresident Visitors: Their Attitudes Towards Wildlife and Wildlife Related Trip Characteristics and Economics. Anchorage: Alaska Dept. of Fish and Game, Division of Wildlife Conservation.

Miller, S.M., and D.W. McCollum. 1999. Less May Mean More: Maximizing the Economic, Environmental, and Social Benefits from Alaska's Visitor Industry. In *Sustainable Future Conference Proceedings* pp. 54-56. Compiled by Brown, K. Anchorage: Alaska Conservation Alliance. Also available on Alaska Division of Tourism Web Site (http://www.dced.state.ak.us/cbd/toubus/lessmay.htm)

Miller, S. M., S. D. Miller, and D. W. McCollum. 1998. Attitudes Toward and Relative Value of Alaskan Brown Bears To Resident Voters, Resident Hunters, and Nonresident Hunters. *Ursus* 10:357-376.

Miller, S. M., R. Sinnott, and D. W. McCollum. 1994. Differentiating Among Wildlife-related Attitudinal Groups in Alaska. *Transactions of the 59th North American Wildlife and Natural Resources Conference* 59:512-523.

Mitchell, R. C. 2002. On Designing Constructed Markets in Valuation Surveys. *Environmental and Resource Economics* 22(1-2):297-321.

Morgenstern, R. D., ed. 1997. *Economic Analyses at EPA: Assessing Regulatory Impact.* Washington, DC: Resources for the Future.

National Research Council. 1997. *Wolves, Bears, and Their Prey in Alaska.* Washington, DC: National Academy Press.

Nichols, A. L. 1997. Lead in Gasoline. In *Economic Analyses at EPA: Assessing Regulatory Impact*, pp. 49-86. Edited by Morgenstern, R. D. Washington, DC: Resources for the Future.

Nordhaus, W. D., and E. C. Kokkelenberg, eds. 1999. *Nature's Numbers*. Washington, DC: National Academy Press.

Panke, A. J., and W. E. Bettoney. 1971. Role of Lead Antiknocks in Modern Gasoline. SAE (Society of Automotive Engineers) Paper #710842.

Panke, A. J., and J. E. Conte. 1969. Effects of Combustion Chamber Deposits and Driving Conditions on Vehicles' Exhaust Emissions. SAE (Society of Automotive Engineers) Paper #690017.

Peterson, G. L., T. C. Brown, and D. H. Rosenthal. 1987. An Improved Framework for Estimating RPA Values. Gen. Tech. Rep. RM-138. Fort Collins, CO: USDA Forest Service, Rocky Mountain Forest and Range Experiment Station.

Peterson, G. L., C. S. Swanson, D. W. McCollum, and M. H. Thomas, eds. 1992. *Valuing Wildlife Resources in Alaska*. Boulder, CO: Westview Press.

Pirkle, J. L., J. R. Schwartz, J. R. Landis, and W. R. Harlen. 1985. The Relationship Between Blood Lead Levels and Blood Pressure and Its Cardiovascular Risk Implications. *American Journal of Epidemiology* 121(2):246-258.

Pless, L. G. 1974. A Study of Lengthened Engine Oil-Change Intervals. SAE (Society of Automotive Engineers) Paper #740139.

Schuetz, J. F. 1998. Economic Values and Impacts Associated with Access Users' Recreational Use of Maine's Great Ponds. Unpublished masters thesis. University of Maine, Orono.

Schwartz, J., J. Leggett, B. Ostro, H. Pitcher, and R. Levin. 1984. Costs and Benefits of Reducing Lead in Gasoline. Draft final report (EPA-230-03-84-005). Washington, DC: U.S. Environmental Protection Agency, Office of Policy, Planning, and Evaluation, Office of Policy Analysis.

Smith, V. K. 2000. JEEM and Non-market Valuation: 1974-1998. *Journal of Environmental Economics and Management* 39(3):351-374.

Swanson, C. S., M. Thomas, and D. M. Donnelly. 1989. Economic Value of Big Game Hunting in Southeast Alaska. Resource Bulletin RM-16. Fort Collins, CO: USDA Forest Service, Rocky Mountain Forest and Range Experiment Station.

U.S. Bureau of Reclamation. 1995. Operation of Glen Canyon Dam Environmental Impact Statement. Denver: U.S. Bureau of Reclamation.

USDA Forest Service. 1995. Socioeconomic Effects and Implications of the Proposed Program. In: Draft RPA Program, http://www.fs.fed.us/land/RPA/chapt4.htm.

Walsh, R. G., D. M. Johnson, and J. R. McKean. 1988. Review of Outdoor Recreation Economic Demand Studies with Nonmarket Benefit Estimates: 1968-1988. Technical Report 54. Ft. Collins, CO: Colorado State University, Colorado Water Resources Research Institute.

Welsh, M. P., R. C. Bishop, M. L. Phillips, and R. M. Baumgartner. 1995. GCES Non-Use Value Study, Final Report. Prepared for the Glen Canyon Environmental Studies Non-Use Value Committee, by Hagler Bailly Consulting, Inc., Madison, WI 53711-1058.

Whittington, D. 2002. Improving the Performance of Contingent Valuation Studies in Developing Countries. *Environmental and Resource Economics* 22(1-2):323-367.

Wintringham, J. S., A. E. Felt, W. J. Brown, and W. E. Adams. 1972. Car Maintenance Expense in Owner Service with Leaded and Non-leaded Gasolines. SAE (Society of Automotive Engineers) Paper #720499 (May).

Chapter 14

WHERE TO FROM HERE?

Richard C. Bishop
University of Wisconsin–Madison

1. INTRODUCTION

While our editors were too polite to call it that, I was assigned to write the Old Professor's chapter. It's up to me to wrap up this volume by taking stock and looking into the future. Actually, being 30 years past the Ph.D., I rather like this role. Taking the long view suits me fine.

Our editors have given us fine historical perspectives on valuation methods in Chapters 4 and 8, and several of the authors provide historical background in their chapters. Hence, I do not need to spend much time on the past, but let's pause to consider how far nonmarket valuation has come. What you have seen in this volume represents the culmination of a great deal of successful research. Most of this progress has occurred during my professional lifetime.

Consider the theoretical side first. In graduate school 30 years ago, we puzzled over the theory, wondering what the Hicksian measures really meant. Hicks, you will recall, dealt with the welfare effects of price changes. As students, we wondered how his concepts ought to be applied to environmental quality and publicly provided outdoor recreation. Now we know. Over the years, such concepts as weak complementarity and household production functions have evolved and matured.

When I started out, empirical tools for valuation were primitive at best. Most travel cost studies used zonal models. Only a few models using individual level data had been published. RUMs had not even been dreamt of by those of us in the trenches. Only a very few contingent valuation studies existed.

Referendum formats were unheard of and the broader stated preference approaches presented in Chapters 6 and 7 were at a very early stage. Hedonics also was in its infancy. Possible uses of averting cost and other such potential measures of value were being considered on an intuitive level, but were not well understood or widely applied. Against this backdrop, what you have seen here is very impressive. And all this ground has been gained in a mere 30 years.

For this chapter, it is more interesting to ask, where to from here? What can be said about the future directions of nonmarket valuation? In considering this question, let us not belabor the obvious. For example, it goes almost without saying that we should leave behind the debate over whether revealed preference or stated preference measures are better. It is obvious that we need both. Revealed preference measures are, and should continue to be, our bread and butter so long as we have the data to do the job. As will become clearer later in this chapter, revealed preference data have high "content validity," which will continue to give them an edge when they are workable. But at every turn, there are not enough revealed preference data, or there is not sufficient variation in the data, or it is necessary to extrapolate beyond the range of the data, or there are no revealed preference data at all.

This is why contingent valuation is by far the most widely applied nonmarket valuation technique the world over. Either we must continue to make progress on stated preference methods or we might as well pack up our puny tool kits and go home so far as valuation of environmental resources is concerned. Without stated preference methods, there would still be an environmental economics, but its usefulness would be much diminished. We economists could still mumble on about externalities, optimal abatement, transferable discharge permits, property rights, etc. We could do a bit of work on recreation. Hedonic analysis and the other revealed preference tools you have read about would still be available to provide partial measures of the value of some environmental quality changes in some cases. But without stated preference methods, we would be very limited in the number of such changes we could fully value. There are simply not enough of the right kinds of revealed preference data to do what needs to be done. And, without values, we will be hard put to set optimal pollution taxes, determine how many transferable discharge permits to issue, estimate natural resource damages, and apply our other policy ideas.

I could go many different directions in the Old Professor's chapter. I have chosen to elaborate on a theme that came up in Chapter 4 and occasionally elsewhere. I want to make an appeal for devoting more attention in the future to the *validity* of our measures; this chapter is about what I call *validity assessment*. As a researcher who designs and executes valuation studies, what do I need to think about and do to enhance the validity of the results? Or, as a reader or an official reviewer of the work of others, what should I look for to assess a study's validity?

Many think validity is an issue associated only with contingent valuation,[1] but as we shall see later on, it is a fundamental concern for all valuation methods whether they involve stated preference or revealed preference data.

When you think about it, a whole lot of attention has been devoted to validity across most of the chapters of this book. As just one example, when Laura Taylor told us, in Chapter 10, about possible measurement errors in the dependent variable in hedonic property value studies, she was talking about a threat to the validity of study results. Beyond describing the theory of various valuation techniques and laying out the nuts and bolts needed for applications, most of this book has involved advice of one kind or another about how to make your applications as valid as possible. The authors may not have used the term, but they were talking about validity.

So I am not trying to argue that economists have neglected this issue. Far from it. But I am arguing that, in the future, we need to strive to be more thorough and systematic in assessing the validity of valuation studies, methods, and approaches. I will assert up front that economics as a whole has serious shortcomings in this regard and that the way we approach valuation is symptomatic of the problem. We lack a clearly defined, widely accepted theory of economic measurement.

The starting point for this chapter can be succinctly stated: true economic values are unobservable. As we saw in earlier chapters, in applied welfare economics, the value of something is measured in terms of either willingness to pay (WTP) or willingness to accept compensation (WTA). When some economic parameter changes, we ask how much compensation, paid or received, would make economic agents *indifferent* about the change.[2] This compensation (paid or received) is what we try to measure in valuation studies. It is what should be referred to as the "true value" of whatever change is being considered. Here, and without fully realizing it, valuation researchers encounter a basic problem, an instance of what I like to refer to as the fundamental conundrum of the social sciences. Economists and other social scientists cannot see inside people's heads. Because indifference is a state of mind, true values are unobservable.

If indifference-producing amounts of compensation could be observed directly, then validation of our empirical results would be easy. We could just use standard empirical measures to estimate values and compare results to corresponding true values.[3] The closer an empirical measure comes, the more valid it would be. But since true values cannot be observed, all we can do is interpret real-world evidence—e.g., market purchases and answers to survey questions—as observable clues about unobservable true values. This makes validity assessment much more complex. How do we judge the validity of our measures of value when we cannot observe true values?

In Section 2 of this chapter, I want to propose my first cut of a theory of economic measurement. I will start with the "Three Cs" of validity assessment, content, construct, and criterion validity. These terms were introduced by Tom Brown in Chapter 4 and also used by Kevin Boyle in Chapter 5. But the Three Cs have been brought into economics very much as they evolved in psychology. While they can form a skeleton for a theory of economic measurement, we need to put a lot of economic flesh on the bones.

To do this I will need some terminology. The term "method" can mean an overall approach to valuation such as the travel cost method or it can refer to a more specific set of procedures within an approach, such as the nested logit travel cost method. To distinguish, let us use the term *method* in the narrower sense in this chapter. Methods are specific procedures used in valuation studies. Examples of methods are mail surveys, mixed logit travel cost models, hedonic models using property sales prices as the dependent variable, and referendum contingent valuation questions. In keeping with the section headings in this book, an *approach* is a set of methods characterized by a common underlying theme such as the travel cost approach, the contingent valuation approach, the hedonic price approach, and the cost of illness approach. Finally, let's distinguish a third level, the individual *application*. In an application, an approach is applied using a set of methods to obtain one or more value estimates for environmental amenities or other public goods or services. The Three Cs apply in research at all three levels, but in different ways.

Section 3 will illustrate in some detail how validity assessment would work for an individual application of the contingent valuation approach. We will assume that we are assigned to serve as reviewers of a new contingent valuation application and see how we would apply the Three Cs. I want to make validity assessment seem more concrete by summarizing the full set of criteria I would use in reviewing an actual application.

This will be a useful exercise for another reason. More is known about how validity assessment should work in the case of contingent valuation than for the other approaches. Much can be learned from experience with contingent valuation that will inform efforts to assess validity for other approaches, methods, and applications, as we will begin to see in Section 4. There, I will draw some conclusions from the discussion of contingent valuation that are applicable to other valuation approaches. Whether a contingent valuation application or a study using some other valuation approach is at issue, there will be few if any definitive validity tests. Hence, validity assessment will most often involve judgments of professionals in an area of research. Such judgments need to be informed by a careful appraisal of the application's strengths and weaknesses across a full set of validity criteria.

In Section 5, I will change the focus. Up to that point, all the attention will have been on validity assessment of individual applications. Section 5 will consider how to assess the validity of whole valuation approaches such as contingent valuation or hedonic property value analysis. Basically, in addressing the validity of approaches, we are interested in the principles for judging whether a state-of-the-art application is likely to produce valid results. It turns out that theoretical analysis is important in assessing the content validity of approaches. We will see how this works for defensive behavior and hedonic methods. Construct validity is also important. I will illustrate the pitfalls of uninformed application of construct validity tests by recounting some history. In the 1990s, some prominent economists launched a broadside attack on the contingent valuation approach based on construct validity criteria. Recounting what happened in the "CV War" will illustrate how, when it comes to evaluating approaches, the weight of evidence across a full range of studies must be considered. We will also see how criterion validity tests can add potent evidence for or against the validity of approaches.

Section 6 turns the focus again, this time onto the validity of methods. Much of the valuation research that economists do is aimed at improving methods. Validity assessment is particularly important here because the credibility of methodological conclusions depends on the quality of the underlying studies designed to test methods and devise new ones. A balanced research agenda involving all Three Cs is needed.

Just before concluding the chapter, I give some attention to revealed preference applications. Many economists seem to presume at least implicitly that if one uses revealed preference data and does a satisfactory job with the econometrics, the results are valid. This is a mistake. Section 7 will explain why and show how full application of the Three Cs strategy can help sort things out. A full set of validity criteria will need to be developed in order to assess the validity of travel cost, hedonic, and other revealed preference studies. This will require substantial effort on the part of researchers who work with revealed preference approaches. I will suggest some directions that may help get the ball rolling.

Section 8 provides some concluding thoughts. The Old Professor will tell you how he got involved in nonmarket valuation in the first place and why he is optimistic about its future, especially if we devote more effort to validity.

2. TOWARDS A THEORY OF ECONOMIC MEASUREMENT

Mitchell and Carson (1989) were among the first in the environmental valuation literature to take a formal look at this problem. They drew on the classical theory of testing from psychology. Psychologists are in the business of measuring unobservable mental phenomena. WTP and WTA seem simple compared to many of the concepts they deal with; consider examples such as racial prejudice, competence in mathematics, and human intelligence. Hence, it is not surprising that Mitchell and Carson drew on psychology. Although they focused on contingent valuation, I believe that the overall strategy they describe is applicable whenever we are trying to assess the validity of economic value measures.

As you have seen in earlier chapters, assessing the validity of a given valuation study would take three mutually reinforcing directions, like three legs of a stool. Let's take a much closer look at the Three Cs—content, construct, and criterion validity. I want to begin by proposing a broader concept of content validity than others have used. Content validity is really about procedures. One asks whether the approach chosen and the methods used in implementing the approach were fully conducive to measuring the true value. For a study valuing a market good, for example, when analysts or reviewers examine the procedures used in modeling supply and demand, they are considering the content validity of the study as I want to use that term. If the econometric model seems poorly conceived or other questionable procedures are identified, then the ability of the study to measure true WTP or WTA would become suspect.

Construct validity is traditionally broken down into two parts. *Theoretical validity* involves tests of theoretical and intuitive prior expectations about relationships that should exist among variables. For example, other things held constant, one would expect the price of a dwelling unit to be related to its square footage. Should a hedonic price function fail to show that relationship, this would be grounds for questioning its theoretical validity. *Convergent validity* involves "the correspondence between a measure and other measures of the same theoretical construct" (Mitchell and Carson 1989, p. 294). For example, suppose that an outdoor recreational opportunity has been valued using contingent valuation and a travel cost model. Comparing the value estimates would be a test of convergent validity. If the two values are close, then this is taken as evidence of validity. If they differ, then at least one of them would seem to lack validity.

Criterion validity is the third leg of the stool. In some circumstances, it may be possible to value an amenity using an approach that is already widely accepted as valid by practitioners of the discipline. If such an approach is available and can be applied in some circumstance at reasonable cost, then the results can be used as criteria for assessing the criterion validity of the approaches that are normally available to us. For example, most economists would be satisfied if contingent valuation yielded values comparable to values from a well-functioning market. An experiment might be set up to yield market values that could serve as criteria to judge the validity of contingent values.

Confusion may arise about the relationship between convergent and criterion validity since both involve the comparison of values derived by applying different valuation approaches or methods. In testing for convergent validity, the two measures have roughly equal status. When I illustrated convergent validity by suggesting comparison of contingent and travel cost values, I was implicitly assuming that neither is considered inherently more valid than the other. In other words, neither has the status to be considered a criterion against which to measure the validity of the other. If you were convinced that the travel cost approach is inherently more valid than contingent valuation, then such a comparison would constitute a criterion validity test of contingent valuation.

The Three Cs—content, construct, and criterion validity—are mutually reinforcing. For example, a study with high content validity may fall because theoretical priors are not met. Likewise, a study that seems questionable on content and construct validity grounds might nevertheless stand if it uses methods that have previously performed well in criterion validity testing.

While the Three Cs are intuitively appealing, psychology can take us only part way to an economic theory of measurement. We economists need to adapt these concepts to our own needs. We deal with different phenomena using theories that are fundamentally different from what psychologists use. Also, we tend to be more interested in aggregates (e.g., mean WTP) while they are more interested in individuals (e.g., IQ scores). For these and other reasons, the Three Cs as they evolved in psychology are only a starting point for us.

We do have some bits and pieces to start with in constructing our own theory of measurement. The criteria commonly applied in econometrics are probably the most fully developed. None of us would try to publish a regression model without checking out the coefficients for statistical significance and correctness of signs, for example. In the terminology used here, these are construct (theoretical) validity tests. And we have more complicated rules as well, such as "Thou shalt not mine the data!" This is an issue of content (procedural) validity. We are also doing a pretty good job of motivating our empirical measures with appropriate theory, another issue of content validity. What we need to do is integrate what we are already doing into the broader framework and fill in the gaps .

As noted in the introduction, the process of adapting the Three Cs to nonmarket valuation is most advanced in the case of contingent valuation. In the spirit of other chapters in this book that focus on how to do things, the next section will show how I would proceed in assessing the validity of a contingent valuation study.

3. HOW TO ASSESS VALIDITY OF A CONTINGENT VALUATION STUDY

Let us suppose that we have before us a brand new contingent valuation study and that we have been assigned to assess its validity. We will play the role of reviewers and see how we would apply the Three Cs to get the job done.

Let us assume that we have a fairly complete report of what was done and of the results.

Dan McCollum and I have attempted to formulated a complete list of the questions we would ask in a content validity assessment of a contingent valuation study (Bishop and McCollum 1997). Space will not allow me to go into much depth, but a brief review of questions will help the reader gain an understanding of what one looks for in the content validity phase of an assessment. In the process, we will learn some important lessons about validity assessment more generally. Here are the questions McCollum and I would ask:

1. *Was the true value clearly and correctly defined?* Since theory should guide the research questions that are asked, the data that are gathered, and the way those data are analyzed, a valid study will have had its theoretical act together at the outset. As reviewers, we would want to look for a theoretical model in the study report that defines the true value, and check out the contingent valuation scenario and eventual econometric analysis for consistency with that model. A study lacks content validity if it does everything else right but seeks to measure the wrong thing.

2. *Were the environmental attributes relevant to potential subjects fully identified?* In the process of designing the study, the researchers need to have had a fairly complete understanding of what they were setting out to value. Information on which attributes of an environmental resource are important to study subjects may be gained through conferring with policy makers and scientists, and through applying focus groups and other qualitative tools to study subjects drawn from the population that will be the focus of the final survey. As reviewers, we would look at reported procedures to see if this was done.

3. *Were the potential effects of the intervention on environmental attributes and other economic parameters adequately documented and communicated?* Typically, the valuation study will have focused on an intervention where public or private actions affect the environmental resources in question. Once respondent-relevant attributes of those resources were identified (question 2), the potential effects of the proposed intervention should have been documented and effectively communicated to study subjects. To verify that this was done, we would look carefully at the survey instrument to inform our judgement about its effectiveness in this regard. Also, we would examine descriptions of research procedures with this issue in mind. If draft survey materials were pretested in focus groups, this would be taken as positive evidence. Other procedures might have involved administering the draft survey

to a small sample and then debriefing the subjects individually to see whether they interpreted the survey as the researchers intended. The more casual and ad hoc the procedures seem to have been in this regard, the less confidence we, as reviewers, will have in the results.

4. *Were respondents aware of their budget constraints and of the existence and status of environmental and other substitutes?* Some scholars, such as those of the NOAA Blue Ribbon Panel on Contingent Valuation (U.S. Department of Commerce 1993), which was also discussed in Chapter 5 of this book, believe that respondents need to be reminded about such things if they are to give valid responses to contingent valuation questions. If the survey in question failed to do this, its content validity will have suffered in the eyes of many reviewers.

5. *Was the context for valuation fully specified and incentive compatible?* Context refers to all dimensions of the proposed transaction that study subjects feel are relevant to their answers to the contingent valuation question. If important parts of the context for valuation were missing or poorly communicated, respondents may have had trouble giving valid responses to the contingent valuation question (Fischhoff and Furby 1988). Simply asking respondents what they would be willing to pay for the intervention would not be considered be sufficient in this regard. When valuing a public good, many think that framing a scenario as a referendum provides respondents with a context that they can readily understand. It is also desirable, all else equal, to use a context that is incentive compatible. See Chapter 5 for further discussion of incentive compatible response formats. Here again, we would want to carefully examine the survey instrument and associated discussion in the report, this time to evaluate the context of the contingent valuation question.

6. *Did survey participants accept the scenario? Did they believe the scenario?* A study subject accepts the scenario when he or she implicitly agrees to proceed with the valuation exercise based on the information and context provided. Scenario rejection can lead either to poor quality data or to non-response to contingent valuation questions. Content validity is enhanced if respondents not only accept the scenario, but believe it. Belief is achieved when respondents consider it likely that the supply of the amenities in question and what they actually will pay depend on their answers. As reviewers, we would look for results from the qualitative phase of the research (e.g., focus groups) and responses to the questions in the survey that indicate what subjects considered when they answered the contingent valuation question. If such

results indicate that subjects accepted the scenario and possibly even believed it, this would support the validity of the contingent valuation results. Studying the scenario itself for probable credibility in the eyes of respondents might also help inform our judgment of acceptance and belief.

7. *How adequate and complete were survey questions other than those designed to elicit values?* Contingent valuation surveys typically include many questions other than those intended to elicit values. Such questions may be useful in validity assessment and serve other purposes. We would want to see what was done in this regard and why.

8. *Was the survey mode appropriate?* For reasons explored in Chapters 3 and 5, it may matter to validity whether the survey was done by mail, telephone, personal interview, or some other mode. We would want to form an opinion about the appropriateness of the survey mode based on the purpose of the study, the complexity of the scenario, and other issues.

9. *Were qualitative research procedures, pretests, and pilots sufficient to find and remedy identifiable flaws in the survey instrument and associated materials?* Depending on the study goals, the complexity of the valuation exercise, and other factors, substantial qualitative research (e.g., focus groups) and preliminary survey work may be needed before the survey instrument and associated materials are ready to go into the field. Neglecting these steps where they would likely have been helpful would reduce the validity of the final results.

10. *Given study objectives, how adequate were the procedures used to choose study subjects, assign them to treatments (if applicable), and encourage high response rates?* The most carefully planned survey instrument still needs to be administered properly in light of study objectives. In Chapter 3, Patty Champ gives sound advice about how to do surveys that will enhance the content validity of study results. We would want to examine the description of the research procedures and resulting response rates to see how well the study dealt with these issues.

11. *Was the econometric analysis adequate?* It goes almost without saying that the econometric analysis needs to be solid in order to get valid results.

12. *How adequate are the written materials from the study?* Readers of studies need to be able to gain a clear understanding of what was done and why, and what the results are. As reviewers, we have had to refer repeatedly to the study report in addressing the first 11 questions. The burden of proof is on the researchers to make the case that their study has validity. If the report is ambiguous or incomplete, the study's content validity will remain in doubt.

Once the content validity assessment is completed, our next task as reviewers would be to assess the construct validity of the study before us. Researchers will typically have estimated one or more regression models where some expression of WTP (e.g., voting yes or no in a referendum question or stated dollar values from open-ended questions) is the dependent variable and various possible explanatory variables are included on the right-hand side. In terms of measurement theory, this is theoretical validity testing. Basically, researchers estimate these sorts of equations in search of evidence confirming that expressed values "look, smell, and feel" like economic values. Remember those latent true values hiding in people's heads? Theoretically, they should be rooted in preferences and past experiences. They also should normally be related to income and availability of substitutes and complements. As we review the study before us, the more often we can see relationships in the data that seem to make sense based on theory and intuition, the more confident we can be that study results have construct validity. The fewer such relationships turn up, the less likely it will seem that the study is getting at people's true values.

Construct validity might also be evaluated through a so-called scope test as you learned in Chapter 5. If one intervention provides more of the environmental amenities in question than another, the first one should have a higher value. If we see that the study before us performed a scope test and passed it, our opinion of the study's validity would be enhanced. Failure of the scope test would ordinarily raise doubts.

We might also find that the study under review included one or more convergent validity tests. For example, an air quality study might include a hedonic property value analysis for comparison (see, for example, Brookshire et al. 1982). Or, a contingent valuation study of groundwater quality might use the cost of bottled water for comparison (see, for example, Poe and Bishop 1999). If such comparisons prove favorable, this adds to the evidence of construct validity. If not, then seeds of doubt will take root.

In sum, the construct validity part of the assessment would look for as many tests of prior expectations (including convergence of alternative measures) as can reasonably be expected to have been performed. In the end, after reviewing this evidence, we, as reviewers, would formulate a judgment of how solidly the study conclusions are supported in the data. The more of our priors that were fulfilled, the stronger our conviction would be that the study in question has succeeded in measuring true values. The more such tests were failed, the weaker the study would be judged to be. If few or no construct validity tests were performed, then the validity of the results would remain questionable.

Criterion validity would enter into our review only indirectly. The overall validity of an individual contingent valuation study depends to some extent on how well the contingent valuation approach has fared in criterion validity tests. Also, criterion validity tests may inform judgments about what constitutes best practice in doing contingent valuation studies.

Application of the Three Cs to methods and approaches will be considered in a moment. First, let's draw some general conclusions about how the Three Cs work in assessing the validity of individual applications.

4. SOME CONCLUSIONS ABOUT VALIDITY ASSESSMENT FOR INDIVIDUAL APPLICATIONS

Let us maintain our focus on contingent valuation a bit longer. Notice, first, that it would be very hard for any contingent valuation study to fully satisfy the criteria I have just described. But it is also true that some studies will rate better than others. Validity is a matter of degree. Most studies will rate well on some criteria and not so well on others. The more positive signals reviewers see, the more positive their final conclusions about a study will be. The opposite also holds.

I can make this point even more strongly. There are few, if any, definitive tests of validity. During content validity assessment, reviewers may discover procedures that make them uneasy. For example, the qualitative research that was done in developing a survey instrument may seem superficial, or a telephone survey may have been done when reviewers feel that personal interviews would have been highly desirable. But most of the time, reviewers are not likely to have misgivings sufficient to cause them to conclude immediately that a study totally failed.

The same holds true for construct validity. Take the most venerable of economic variables, income, as an example. Reviewers would expect the regression analysis associated with a contingent valuation study to include income as a possible predictor variable. During their review, they would note the sign and level of significance of its coefficient. If the sign is positive and significant, they would take this as positive evidence for the validity of the results. But what if the coefficient on income turned out to be insignificant, or negative and significant? Would they immediatelyconclude that the study is invalid? Hardly. Perhaps income was not measured very well, or maybe over the range of incomes represented among study participants, values were not very sensitive to income. Perhaps the environmental amenities in question are truly inferior goods. Likewise, measures of preference or experience that seem really promising may not pan out as predictor variables for a variety of reasons. So, as reviewers look at a study, their doubts will grow as prior expectations go unfulfilled, but they also will normally tolerate some failures when forming their final judgments.

This is true for scope tests. The logic of the scope test is sound: if people desire more of a given environmental amenity or other good, they ought to be willing to pay more for the additional amount. But that is a big "if." I may value my one American flag highly but be willing to pay little or nothing for a second. Similarly, I may place a large value on having a wild population of 400 wolves in northern Wisconsin, yet have a lower value for 800. I may, for example, worry about the larger population being ecologically disruptive or unduly harmful to the deer population. A recent study at the University of Wisconsin (Wilson 2000) concluded that sensitivity to scope is neither a necessary nor a sufficient condition for the validity of contingent valuation studies. Reviewers of a study would normally take passing a scope test as evidence of construct validity; not passing raises questions that may lead to doubts about construct validity. However, it is also possible that there are good or at least plausible explanations for the failure.

All this goes to prove once again that assessing efforts to measure something that is inherently unobservable is a messy business. Validity assessments of valuation studies end up involving large amounts of judgment and, for the foreseeable future, there will always be plenty of room for disagreement. Get two or more of the leading students of contingent valuation together and they are likely to disagree about the specific criteria to be applied and how much weight each criterion should receive in forming final conclusions. We should expect similar disagreements to hold for all the other approaches to valuation discussed in this book.

Thinking in terms of "reviewers" of studies has been a useful starting point. Indeed, there are reviewers of many sorts in the real world. They consider whether articles submitted to journals should be published, whether available studies are good enough to be used in policy analyses for government agencies, and whether studies entered as evidence in litigation are valid or not. But it is also usefulto make theswitch from reviewers to researchers. When researchers embark on a valuation study, they need the kinds of criteria I have described for contingent valuation in front of them from the beginning, since the burden of proof for the validity of their final results will lie first and foremost with them.

It should also begin to be clear that the basic issues I am proposing here are not limited to contingent valuation. A version of the Three Cs is needed for each of the valuation approaches discussed in this book. Some of the criteria that we applied to contingent valuation will carry over to the other approaches. High quality econometrics, for example, will be a sine qua non across the board. But other criteria will need to be tailored to the issues relevant to the particular study and approach at hand within the overall framework of the Three Cs. In a hedonic study, for example, it would not make much sense to ask if the survey instrument reminded respondents of their budget constraints.

Later on, I will explore validity criteria for hedonic and travel costs studies in a preliminary way, but the treatment will necessarily be sketchy. Criteria that are as well integrated and articulated as those for contingent valuation do not yet exist for other approaches. Still, the overall conclusions that have been derived here from looking at contingent valuation will continue to hold for the other approaches. Regardless of approach, validity will, for the foreseeable future, be a matter of degree. There are likely to be few if any definitive validity criteria to go by. Assessments of validity will continue to call for the professional judgment of those knowledgeable in the field, and researchers will need clearly described validity criteria from the outset of their studies.

At this point, I will turn from validity assessment of individual applications to studies designed to test approaches and methods. The Three Cs still apply, but not necessarily in the same way.

5. ON THE VALIDITY OF APPROACHES

The question we want to ask in this section needs to be carefully posed. Clearly, we want to judge an approach with an assumption that it is applied using the best available methods. What constitutes state-of-the-art methods at any point in time will depend on the judgment of leading researchers in the field. The same researchers will have formed judgments about the boundaries within which the approach can successfully be applied. What we want to know is whether the approach in question, if applied within these methodological and circumstantial boundaries, can be expected to give valid results.

5.1 Content Validity

Interestingly, a lot of the action at the approach level involves content validity. For example, in Chapter 11, Mark Dickie says some rather discouraging things about the defensive behavior approach. In essence, he is questioning the content validity of the approach. Once such issues as joint production are factored in, the relationship between a change in defensive expenditures and the true value of a change in pollution becomes ambiguous. No matter how well you estimate the costs of defensive behavior, your results will be less compelling than would be the case if such problems didn't exist. In other words, questions exist about the validity of even the best studies.[4]

This point forces me into a slight digression, but one where a lesson about validity of valuation approaches can be learned. Given this problem with the defensive behavior approach, some readers may wonder why it was included in this book at all. Why include an approach that is known to lack validity? Part of the answer is that the results are nevertheless sometimes useful. In considering a proposal to reduce pollution, for example, knowing something about defensive costs may at least inform one's judgment about whether true benefits are large or small relative to costs. Also, we keep the defensive behavior approach on the table in the hope that more research may succeed in broadening the cases where one can at least say whether defensive expenditures are a lower bound on true values. A biased (i.e., relatively invalid) value can still be useful if one is confident about the direction of the bias.

Another example where content validity has come to the fore is the classic article by Rosen (1974) on the hedonic property value approach. One of Rosen's contributions was to clarify what a hedonic price equation and its associated implicit price function really tell us and what they do not tell us. In particular, the implicit price function should not, except under very special conditions, be interpreted as an expression of marginal willingness to pay across a range of environmental quality levels. Hence, the implicit price function cannot be used to measure the value of infra-marginal changes in environmental quality. Stated differently, if the hedonic price function is properly estimated, then it is valid to use the implicit price function to measure marginal values, but using it to measure infra-marginal values is of questionable validity. To gain a valid estimate of the latter requires solving the associated identification problem, as explained in Chapter 10.

5.2 Construct Validity

Construct validity also plays a role at the approach level. My favorite example here is the study by Carson et al. (1996). Their work examined the relationship between contingent and revealed preference values from studies that estimated both for the same environmental amenities. They found 83 studies that supported 616 such comparisons (studies often contain more than one value estimate per approach). Several statistical tests were performed with results that were consistent. The ratios of contingent values to revealed preference values averaged 0.89 with a 95 percent confidence interval of 0.81–0.96. Additional statistical tests confirmed the robustness of this result. I take this to be strong support for the validity of the approaches involved in the comparisons, including contingent valuation.

A bit of recent history, what I will refer to as the "CV War," will illustrate further how economists view evidence relating to the construct validity of approaches. First in preparation for litigation on the Exxon Valdez oil spill and then in the context of writing general rules for natural resource damage assessment under the Oil Pollution Act of 1990, some contingent valuation studies were conducted by Exxon's consultants, several of whom are prominent economists (Hausman 1993; Diamond and Hausman 1994; McFadden 1994; Diamond 1996). They found that their studies failed some construct validity tests. The conclusions of Diamond and Hausman (1993, p. 4) were typical of the group:

> We conclude that the CV [contingent valuation] method does not measure an economic value that conforms with economic preference concepts. Thus, we also conclude that it is not appropriate to include CV measures of stated willingness to pay (WTP) in either benefit-cost analysis or compensatory damage measurement.

Most of us who have been involved over the long run in research on contingent valuation did not find these arguments very convincing. Exxon's experts had shown that *their* studies lacked validity. But we were not inclined to turn our backs on the evidence, most of which was considered in Chapter 5, that has accumulated over the years supporting the validity of the contingent valuation approach if properly applied. Furthermore, we suspected that some of the reasons the Exxon team's studies failed construct validity tests are rooted in content validity failures (see, for example, Hanemann 1994). Going into a lot of detail would be beyond the scope of what I am trying to accomplish here. It will suffice here to say that we questioned whether their surveys had been designed and executed in ways that were conducive to high content validity. Furthermore, at the center of their attack was the failure of their results in scope tests, yet we were quick to point out that most studies performing scope tests have found their value estimates to be sensitive to scope (see Carson 1997 for one summary).

Of course, those of us who have devoted much of our careers to contingent valuation are not without our vested interests. One might wonder whether we were overly sanguine about contingent valuation because we were unwilling to set aside our past accomplishments. But environmental economists other than those working on contingent valuation were not bowled over by the Exxon-supported studies either. Consider Paul R. Portney, President of Resources for the Future, who certainly did not have a vested interest in contingent valuation. Although what he heard in the debate following release of the Exxon studies gave him cause for concern, he concluded that the method was sufficiently promising to warrant "further research and lively intellectual debate" (Portney 1994, p. 16). As is well-known, the NOAA Blue Ribbon Panel on Contingent Valuation, which included Portney and several other distinguished economists, heard the arguments of the Exxon Team, yet still concluded (U.S. Department of Commerce 1993, p. 4610):

> CV studies convey useful information. We think it is fair to describe such information as reliable by the standards that seem to be implicit in similar contexts, like market analysis of new and innovative products and the assessment of other damages normally allowed in court proceedings. Thus, the Panel concludes that CV studies can produce estimates reliable enough to be the starting point of a judicial process of damage assessment, including lost passive-use [non-use] values.

It should be added that the NOAA Panel was sufficiently concerned about possible biases in contingent valuation that it specified a long list of procedures for doing valid studies. This was the NOAA Panel's attempt to explicitly define methods to enhance the content and construct validity of contingent valuation applications.

I have gone into this matter partly because Old Professors ought to be allowed to tell a few war stories. But there is an important lesson to be learned here about criteria to be used in assessing the validity of valuation approaches. When one considers the construct validity of a whole approach (such as contingent valuation), judgements must be based on the full weight of accumulated evidence, and not on how well it worked in a few specific applications. Individual studies can encounter problems for many reasons.[5] Such failures can hold clues for method improvements, as discussed in the next section. Still, researchers are not willing to give up methods that have previously been judged to have at least minimal validity without a great deal of evidence. In fact, as a general principle of applied economics, even approaches that do not work particularly well often continue to be used until better approaches are discovered (Paris, Caputo, and Holloway 1993). When the CV War broke out, the broader community of scholars was not willing to throw contingent valuation in the trash bin despite the professional status of the attackers. The whole debate is symptomatic of the problem that motivated this chapter: we need to gain consensus about an explicitly stated theory of economic measurement and the rules of evidence for assessing validity.

5.3 Criterion Validity

Now let's turn to the third of the Three Cs. Criterion validity can play a powerful role in assessing the validity of approaches. Again contingent valuation provides examples. Both laboratory and field experiments involving comparisons of contingent values to values from actual cash transactions have attracted a lot of attention.[6] Research of this kind, which is sometimes referred to as "simulated market experiments," should continue to be a high priority. If contingent values compare well with values from actual cash transactions in experimental markets, most economists would take this as potent evidence of validity and rightly so. Likewise, if contingent valuation performs poorly, this is a large cause for concern about the approach. Nevertheless, here as elsewhere, definitive results from one or a few studies are and will likely continue to be elusive. Any simulated market will involve unique circumstances that make the wholesale generalization of results less than straightforward. Furthermore, such experiments can have their own problems in meeting content and construct validity criteria. In the end, such studies will become part of the evidence considered when examining the merits of approaches based on the weight of all the evidence. When he was much younger, the Old Professor once presented at a major national conference a paper drawing on some of his criterion validity studies under the title, "Does Contingent Valuation Work?" (Bishop and Heberlein 1986). He is no longer so bold as to think that one study can answer such a question.

I will soon propose some ways that criterion validity tests might be conducted for valuation approaches other than contingent valuation. First, though, let us consider how the Three Cs apply in evaluating the validity of valuation methods.

6. ON THE VALIDITY OF METHODS

The goal of much day-to-day research on nonmarket valuation is to improve the methods used in applied studies. Pick up an issue of any of the journals that regularly publish articles on nonmarket valuation and chances are good that you will see one or more articles offering new insights on methods. Such papers are continually contributing to the evolution of thinking on best practice in nonmarket valuation studies. In the terminology of this chapter, the goal of methods-oriented research is to improve the content (procedural) validity of later studies.

Sponsors and money for basic research on valuation are scarce. This is unfortunate and may change in the future. In the meantime, applications and methods research most often go hand in hand. The typical study will involve an application designed to produce values of potential relevance to policy, but in the process, researchers are often able to add methodological tests at low cost.

An example will illustrate. Poor et al. (2001) described an application of a hedonic property value model to lakes in Maine. The goal was to estimate values for changes in water quality. This topic has relevance to policy debates over what should be done about nonpoint source water pollution. In the process of applying the model, the researchers encountered an interesting methodological question: is it better to use objective or subjective measures of water quality in such a model? The water quality variable they used was clarity. The objective measure was Secchi disk readings, while the subjective measure was water clarity as perceived by property purchasers who had been surveyed by mail. The researchers ran separate regressions for objective and subjective measures of clarity for four housing markets. They then performed a convergent validity test: do dollar measures of the implicit price of water clarity based on subjective and objective measures converge to roughly the same values? They found evidence to reject the hypothesis of convergence. Furthermore, they found that the objective measures were stronger than subjective ones in predicting variation in housing prices.

The Three Cs apply to methods research with a vengeance. One cannot expect to convincingly compare how well alternative procedures perform or to develop and convincingly test new methods unless the whole study is designed and executed in ways that inspire the confidence of other researchers. In addition, construct validity tests, if successfully passed, help shore up the final conclusions in the minds of other researchers. Criterion validity tests, when available, help to identify which methods work best and could conceivably be fertile ground for finding new methods.

At the same time, it should be recognized that some validity criteria may be relaxed when the only study goals are methodological. For example, in addressing the validity of contingent valuation methods, survey modes and sampling techniques may be acceptable even though they would be rejected if the study objectives involved actual policy analysis or preparation for litigation. Some might object to telephone interviews when scenarios are complex, for instance. But in testing alternative methods, researchers might be able to argue that if the methods work over the phone, surely they would work in personal interviews. Methodological studies could use so-called "convenience samples," samples that are easy and inexpensive to work with even though they may not be representative of all citizens.

Before drawing some final conclusions, I want to turn our attention directly toward revealed preference methods of valuation. As I noted in the introduction, the Three Cs are now commonly referred to in thinking only about stated preference methods. I have already suggested that this is a mistake. Let's take a closer look at revealed preference measures.

7. VALIDITY AND REVEALED PREFERENCE MEASURES OF VALUE

Let me begin with a straightforward economic maxim. Revealed preference data have great credibility as evidence about true values. This is certainly true of data gathered when people go into a market and buy or sell something. It also holds when they decide to take a recreational trip or not take it, or vote in a referendum on an issue that will affect their taxes, or engage in other behaviors involving actual commitments of economic resources. Hence, using revealed preference data contributes in a major way to the content validity of valuation studies. This is all well and good.

At the same time, though, I fear that the confidence we economists have in the logic of revealed preference has lulled us into complacency about validity issues. Starting with revealed preference data is no guarantee of validity. Let me illustrate this problem using a travel cost study.

Hausman, Leonard, and McFadden (1995) applied the travel cost approach to estimate recreational losses to Alaska residents from the Exxon Valdez oil spill in 1989. They considered four kinds of recreation: fishing; boating; hunting; and a composite of hiking, camping, and nature viewing. They developed what George Parsons (Chapter 9 of this book) and others have called a "linked model." That is, models of participation and site choice were estimated in tandem. For each individual, the participation model predicted the demand for trips as a function of several explanatory variables including a composite "price" of trips, which was derived from the site choice model (again, refer to Chapter 9 for further explanation). The spill caused the "price" of trips to change, thus causing a shift in the participation (demand for trips) function and a loss of consumer surplus. Add up this loss across recreation types and individuals in the sample, extrapolate to the population, and you have an estimate of the damages done by the spill.

For the site choice model, they estimated both multinomial logit (MNL) and nested multinomial logit (NMNL) models (see Chapter 9). Because of the problems with the independence of irrelevant alternatives assumption of the MLN models and based on other criteria of a statistical nature, they chose the NMNL models as superior and used them, along with count-data models of trip-taking behavior, to estimate damages.

Hausman, Leonard, and McFadden (1995) did conduct some validity tests of their work. In particular, they noted, more or less in passing, that inclusive values in nested models (again see Chapter 9) must be between zero and unity for the results to be consistent with random utility maximization. There were something like 29 estimated inclusive values, about half of which exceeded unity by more than two standard deviations, a point that they did mention. As explained in Chapter 9, this is symptomatic of a model misspecification problem. The interpretation is that the models are not adequately capturing the substitutability of sites. Hence, the results are not consistent with utility theory. Hausman, Leonard, and McFadden (1995, p. 23) acknowledge this flaw in the analysis and then note, "In general, we might want to respecify the nesting structure and reestimate the model. However, such respecification is not guaranteed to be successful since inclusive value coefficients are often greater than one in NMNL models."

Another worrisome result has to do with catch rates in the fishing model, which have coefficients that are negative and significant. This result is true in both the MNL and NMNL models of fishing site choice. Both models predict that, other things being equal, anglers will be attracted to sites where they will catch less. They explain this result away by saying that it "could be due to the fact that more popular sites have lower catch rates simply because they are more popular" (Hausman, Leonard, and McFadden 1995, p. 21). This explanation begs the question. I would rather think that this wrong sign indicates a failure to adequately model the site choice behavior of anglers. After brief consideration, these issues are not mentioned again in the paper. Without additional caveats of any sort, Hausman, Leonard, and McFadden (1995, p.1) conclude, "These results may provide useful input to government agencies attempting to estimate the appropriate level of taxes, fines, or other regulations for deterring damage to the environment."

Now let's consider what an odd situation this is. Both the wrong inclusive values and the wrong sign on catch constitute failures in construct validity testing. These are not failures of merely peripheral interest. The validity of the final damage estimates depends very directly on having adequately modeled recreational choices.

I am not going to suggest that this study should have been tossed in the waste basket when these problems were identified. It is actually a very interesting study. Here as elsewhere validity is a matter of degree. But I would suggest that in light of problems noted, Hausman, Leonard, and McFadden

(1995) are overly cavalier about the validity of their results.[7] Furthermore, they are not alone in this regard. The article must have undergone a normal peer review process supervised by the editors of a reputable economics journal; presumably reviewers and editors had no serious qualms about the validity of the final conclusions. And it would not be hard to find other examples. The mainstream economics profession evidently believes that if you start with revealed preference data and do an adequate job with the statistical analysis, your results are sufficiently valid to be useful in the policy and legal arenas.

Surely this is wrongheaded. Given the econometric problems we have seen, the results of Hausman, Leonard, and McFadden (1995) could still be good enough for prime time, or they could be badly biased. We have no way of knowing. It is even possible that our current econometric know-how is too limited to obtain valid results. The current state of the art in recreation demand estimation, even in the hands of leading researchers, may not be up to the task of modeling demand for the whole state of Alaska.

Furthermore, in studies like this, the data may be flawed. It is fine to say that, in principle and other things being equal, revealed preference data should yield estimates of true values with strong content validity. In practice, lots of things can go wrong with data. To the their credit, Hausman, Leonard, and McFadden did at least discuss why they think their survey was a good one, but only briefly and in very general terms.[8] Again, I think that what is true for the leaders of our profession permeates the rank and file, where too little attention is being given to possible data problems.

So, my bottom line is that revealed preference studies need to give more attention to validity. They are no different than stated preference studies in this regard. The Three Cs can provide a starting point, but the details will need to be dealt with by researchers in the area. I have no adequate list of content validity questions or construct validity criteria for travel cost, hedonic, or other revealed preference studies to pull off the shelf. Nevertheless, a few thoughts can be stated.

A tremendous amount of effort has gone into improving the econometric methods used by those doing nonmarket valuation studies, and rightly so. Progress there should and will continue to be made. However, I would venture to state that, compared to econometrics, work on data quality is stunted. Sure, there are occasional passing references to data issues, but they are most often ad hoc and lacking in depth.

The data for much of what we do in the future will come from surveys. I cannot imagine doing a state-of-the-art travel cost study without survey data. One can quickly run into the limitations of secondary data for hedonic studies. The analyst may find secondary data on costs of things such as health care and averting strategies, but the data are likely to be from somebody else's survey. Future researchers who are truly committed to assessing the validity of their results will want to know a lot about the quality of their data. I would say that survey methodology for revealed preference studies is a neglected area.

Perhaps we need to start with our students. Given the importance of surveys in nonmarket valuation, one wonders why courses in survey methodology are not part of our regular curricula for Ph.D. students in environmental and resource economics.

More could also be done with convergent validity. In cases where a survey must be done, why not include some stated preference questions for later comparison with revealed preference results? This may also lead to new opportunities to do analyses combining revealed preference and stated preference data.

Criterion validity studies to test revealed preference approaches and methods may not be easy to do, but creative researchers who can come up with opportunities will be able to reap rich dividends. So far as I know, no one has been able to secure permission to take over a recreation site and vary the admission fee so as to estimate a demand function to use as a criterion for travel cost estimates of demand. If a recreation site is not workable, how about using movie patrons as subjects? It might be possible to raise enough money to gather very high quality data on recreationists or home buyers and sellers in order to test the methods we normally use in travel cost and hedonic studies.

Clearly I have only scratched the surface of what needs to be done to raise the quality and hence the credibility of revealed preference valuation studies. It will remain for leading researchers who work with revealed preference approaches to expand on this theme.

8. CONCLUDING THOUGHTS

When I first got involved in nonmarket valuation, my immediate goal was to discredit contingent valuation. My colleague, Tom Heberlein, and I planned and executed what I would now call a criterion validity study (Bishop and

Heberlein 1979). Contingent valuation surprised us and did okay, at least on the WTP side. Rather than fighting one skirmish with nonmarket valuation and getting out, I re-enlisted and became part of a groundswell of researchers working on the topic. Still, I felt that nonmarket valuation was a fad. Battles on that front would rage for a while and then subside, and most of us would reassign ourselves to other sectors of environmental and resource economics. Twenty-five years have passed and interest in nonmarket valuation only continues to grow.

There are many reasons why nonmarket valuation has flourished. For one thing, when applied to environmental issues, benefit-cost analysis is seductive in principle yet woefully incomplete in practice if only market values are counted. If we are going to make benefit-cost analysis work right in environmental policy analysis, then we must have nonmarket values.

Then, too, many of us are not only environmental economists but also environmentalists. I think this has contributed to long-term interest in nonmarket valuation. I have watched many times as economic interests try to take the high ground in policy debates based on very narrow concepts about what is, and is not, economically valuable. This offends my sense of what is right, both as an economist and as an environmentalist. Nonmarket valuation is a way to try to level the playing field.

Furthermore, as an environmentalist, I have much sympathy with the ethical arguments that noneconomists are bringing to bear on environmental issues. I would be among the last to argue that environmental choices should be made entirely on the basis of dollars and cents. Debates over environmental policies would be much the poorer without the ethical arguments that many environmentalists rely on. However, for me, many of the arguments based entirely on environmental ethics miss the mark as badly as narrowly conceived economic arguments. Economics promotes social health by reminding us environmentalists that there is more to life than the environment, essential as it is. Many points of view are needed in the quest for proper balance, with economics not the least important among them.

One other reason why nonmarket valuation has flourished deserves mention. Academics love to be heard outside the academy. We can, of course, persist in our cubicles for years pursuing arcane truths. But even when we do that, is it not in the hope that eventually the world outside will notice what we have done and tell us that it is important? Interest in nonmarket valuation has grown because those outside of economics have noticed it and shown enthusiasm and support for our efforts. In the U.S., our work has been pulled along by federal agencies such as the EPA, the Forest Service, the Bureau of Reclamation, and the Army Corps of Engineers. To an increasing degree, state agencies are showing interest as well. Although less consistent and more ambivalent at times, environmental organizations occasionally tune in. Many other countries have seen similar trends. Europe has the most examples; there, scholars working on nonmarket valuation are now second to none. Other countries and even such venerable international organizations as the World Bank are following suit as well. The days when nearly all nonmarket valuation studies carried an implicit label "Made in the USA" are far behind us. Interest from those outside economics pulls those of us in the discipline along.

So, I think the future looks bright for nonmarket valuation. The quest to adequately account for environmental values in policy analysis will only expand as we and our children and grandchildren continue the struggle to balance environmental quality against other goals.

Research on a broad front will be needed if economists are to meet these needs. As I have said here, I hope work on a theory of economic measurement will be high on the agenda. We are suffering from the lack of such a theory. One symptom was the CV War, which would have been much more productive if all the protagonists had started from widely agreed-upon first principles of measurement. The Three Cs are a place to start in remedying this problem, but they need to be expanded and adapted to economics. The goal should be a comprehensive set of criteria for evaluating the validity of nonmarket valuation studies, methods, and overall approaches.

ACKNOWLEDGMENTS

The work reported here was funded by the Agricultural Experiment Station, College of Agricultural and Life Sciences, University of Wisconsin–Madison, and the University of Wisconsin Sea Grant Institute under grants from the National Sea Grant Program, National Oceanic and Atmospheric Administration, U.S. Department of Commerce, and the State of Wisconsin. Federal grant no. NA46RG0481, projects R/PS-51 and R/PS-52. The ideas described in this chapter evolved from my collaboration with Thomas A. Heberlein over many years. It would not be an exaggeration to say that Tom taught me most of what I know on the topic of reliability and validity of measurement in the social sciences. Of course, all errors, omissions, and misconceptions are the sole responsibility of the author. Shann Fox deserves special thanks for her help in preparing the reference section and proofreading the manuscript. Remaining errors are the sole responsibility of the author.

NOTES

1 This view is so pervasive that it even appears in government manuals on valuation. See U.S. Environmental Protection Agency (2000) where validity is clearly addressed in the section of contingent valuation, but is not mentioned elsewhere.

2 As is well known, whether the pre-change or post-change level of welfare is used as a baseline for valuation is a matter of property rights and/or value judgments. I mean what is said here to apply regardless of which baseline is used.

3 Of course, if it were sufficiently inexpensive to do so, we could just use the true value in policy analysis. We would not need alternative, "practical" measures. But true values, if they were observable, could still be very costly to determine. Then we might occasionally, for scientific purposes, measure them in order to assess the validity of the less expensive methods we more routinely apply.

4 This somewhat awkward way of stating the point is forced on me by the fact that validity here as elsewhere is a matter of degree. In this odd endeavor of trying to estimate unobservables, neat statements such as "the approach is invalid" are often impossible to make.

5 The converse is also true. A proposed approach could conceivably look quite promising in early applications and yet not pan out in the end.

6 Several studies from this literature have already been cited in Chapter 5.

7 Some of my colleagues may be ready to take me to task at this point because I am unwilling to accuse Hausman, Leonard, and McFadden (1995) of ginning up damage estimates simply to serve the interests of the sponsor of their research, Exxon USA. I prefer to keep the discussion focused on the work itself in search of deeper issues in economics.

8 For example, data on the number of trips taken in 1988 and 1989 were gathered through telephone interviews over some unspecified period later on. The authors do raise the issue of recall bias but assure us that the survey was subjected to "extensive pretesting which indicated that the retrospective survey was capable of obtaining accurate information for both 1988 and 1989" (Hausman, Leonard, and McFadden 1995, p. 15). No more is said on the topic. In fact, not very much is known about recall problems going back this far and I know of no standard "pretest" procedures for testing recall. Hence, such vague references to pretesting are not very helpful to someone trying to evaluate the validity of the study. Introspection regarding my own ability to remember how many recreational trips I took a one to two years ago make me think that recall could be a big problem here.

REFERENCES

Bishop, R. C., and T. A. Heberlein. 1979. Measuring Values of Extramarket Goods: Are Indirect Measures Biased? *American Journal of Agricultural Economics* 61(5):926-930.

Bishop, R. C., and T. A. Heberlein. 1986. Does Contingent Valuation Work? In *An Assessment of the Contingent Valuation Method.* Edited by Cummings, R. G., D. S. Brookshire, and W. D. Shulze. Totowa, NJ: Rowman and Allanheld.

Bishop, R. C., and D. W. McCollum. 1997. Assessing the Content Validity of Contingent Valuation Studies. Unpublished paper, Department of Agricultural & Applied Economics, University of Wisconsin-Madison.

Brookshire, D. S., W. D. Schulze, M. A. Thayer, and R. C. d'Arge. 1982. Valuing Public Goods: A Comparison of Survey and Hedonic Approaches. *American Economic Review* 72:165-177.

Carson, R. T. 1997. Contingent Valuation Surveys and Tests of Insensitivity to Scope. In *Determining the Value of Non-Market Goods: Economics, Psychology, and Policy Relevant Aspects of Contingent Valuation Methods.* Edited by R. J. Kopp, W. Pommerhene, and N. Schwartz. Boston: Kluwer, pp. 127-163.

Carson, R. T., N. E. Flores, K. M. Martin, and J. L Wright. 1996. Contingent Valuation and Revealed Preference Methodologies: Comparing Estimates for Quasi-Public Goods. *Land Economics* 72:80-99.

Diamond, P. A. 1996. Testing the Internal Consistency of Contingent Valuation. *Journal of Environmental Economics and Management* 30:337-347.

Diamond, P. A., and J. A. Hausman. 1993. On Contingent Valuation Measurement of Nonuse Values. In *Contingent Valuation: A Critical Assessment*, Edited by Hausman, J. A., Amsterdam: North-Holland, pp. 3-38.

Diamond, P. A., and J. A. Hausman. 1994. Contingent Valuation: Is Any Number Better Than No Number? *Journal of Economic Perspectives* 8:45-64.

Fischhoff, B., and L. Furby. 1988. Measuring Values: A Conceptual Framework for Interpreting Transactions with Special Reference to Contingent Valuation of Visibility. *Journal of Risk and Uncertainty* 1:147-184.

Hanemann, W. M. 1994. Valuing the Environment through Contingent Valuation. *Journal of Economic Perspectives* 8:19-43.

Hausman, J. A., ed. 1993. *Contingent Valuation: A Critical Assessment.* Amsterdam: North-Holland.

Hausman, J. A., G. K. Leonard, and D. McFadden. 1995. A Utility-Consistent, Combined Discrete Choice and Count Data Model Assessing Recreational Use Losses Due to Natural Resource Damage. *Journal of Public Economics* 56:1-30.

McFadden, D. 1994. Contingent Valuation and Social Choice. *American Journal of Agricultural Economics* 76:689-708.

Mitchell, R. C., and R. T. Carson. 1989. *Using Surveys to Value Public Goods: The Contingent Valuation Method.* Washington DC: Resources for the Future.

Paris, Q., M. R. Caputo, and G. J. Holloway. 1993. Keeping the Dream of Rigorous Hypothesis Testing Alive. *American Journal of Agricultural Economics* 75(75th Anniversary Issue):25-40.

Poe, G. L., and R. C. Bishop. 1999. Valuing Incremental Benefits of Groundwater Protection when Exposure Levels Are Known. *Environmental and Resource Economics* 13:241-367.

Poor, P. J., K. J. Boyle, L. O. Taylor, and R. Bouchard. 2001. Objective versus Subjective Measures of Water Clarity in Hedonic Property Value Models. *Land Economics* 77:482-493.

Portney, P. R. 1994. The Contingent Valuation Debated: Why Economists Should Care. *Journal of Economic Perspectives* 8:3-17.

Rosen, S. 1974. Hedonic Prices and Implicit Markets: Product Differentiation. *Journal of Political Economy* 82:34-55.

U.S. Department of Commerce, National Oceanic and Atmospheric Administration. 1993. Natural Resource Damage Assessments Under the Oil Pollution Act of 1990. [Report of the NOAA Blue Ribbon Panel on Contingent Valuation] *Federal Register* 58:4601-4614.

U.S. Environmental Protection Agency. 2000. *Guidelines for Preparing Economic Analyses.* EPA 240-R-00-003. Washington, DC: U.S. Environmental Protection Agency.

Wilson, M. A. 2000. Rethinking Scope Sensitivity and Contingent Valuation Surveys: Strong Environmental Attitudes and Contingent Economic Values. Ph.D. dissertation, Department of Rural Sociology, University of Wisconsin - Madison.

Subject Index