Department of Economics
Delhi School of Economics
University of Delhi
Minutes of Meeting

Subject: B.A. (Hons) Economics

Semester-III

Course: **Economics of Climate Change and Natural Resources– ECON032**

Credits: 4 Duration (per week): 4 hours (3Lectures + 1 Tutorial)

Date: 14th August, 2023

Venue: Delhi School of Economics

Convenor: Prof. Surender Kumar

The meeting was attended by the following members:

Sl No.	Teacher Name	College
1.	Animesh Naskar	Hansraj College
2.	Swarup Santra	Satyawati College(Morning)
3.	Manoj Kumar	Shyamlal College(Evening)

Ms. Sujayata Choudhry from Indraprastha College for women was co-opted later.

In order to finalise the readings, internal assessment and evaluation process of the course, the following decisions were taken for the current semester:

1. To cover the course contents in Unit-1 & 2 the following readings were proposed:
   - Alpha C. Chiang. (2000). *Elements of dynamic optimization*. Waveland Press.
   - Hoel, M. (2016). *Optimal control theory with applications to resource and environmental economics* (No. 08/2016).
   - Perman, R., Yu, M., McGilvray, J., & Common, M. (2003). Natural Resource and Environmental. *Economics Pearson Education Limited. Edinburgh Gate Harbor Essex CM20 2JE and Associated Companies throughout the world*.
   - Bhattacharyya, S. C. (2019). *Energy economics: concepts, issues, markets and governance*. Springer Nature.
   - Tietenberg, T., & Lewis, L. (2018). *Environmental and natural resource economics*. Routledge.
2. I. Internal Assessment (IA): 30 marks –
   - two class tests (12 marks each), and
   - 6 marks for attendance.

II. Continuous Assessment (CA): 40 marks –

- Group Project (25 marks).
- Group Discussion (5 marks)
- Problem solving exercises (5 marks)
- Attendance (5 marks)

III. Then End Term Exam: 90 marks will comprise of theoretical and application-based questions.

3. The proposed and tentative pattern for the end-semester final examination-
   Unit-wise suggestive weights:

   Unit-1(30%), Unit-2(30%), Unit-3(25%) and Unit-4(5%)

   The question paper will comprise of 3 sections:

   Section-A: 2 questions-30 marks each

   Section-B: 2 questions-20 marks each

   Section-C: 1 question-20 marks

The detailed unit-wise list of readings are given below:

Discipline Specific Elective 2 (DSE-2): Economics of Climate Change and Natural Resources

Semester	Course title & Code	Credits	Duration (per week)			Eligibility Criteria	Prerequisite
			Lecture	Tutorial	Practical/ Practice		
III/V/VII	**Economics of Climate Change and Natural Resources–ECON032**	**4**	**3**	**1**	**0**	**Class 12th**	**Introductory Microeconomics (ECON001)**

Learning Objectives

The Learning Objectives of this course are as follows:

- The objective of this course is to provide knowledge on the principles of governing and managing natural resources.
- This course introduces the conceptual and theoretical foundations of Resource Economics. In

particular, the efficiency concepts for evaluating natural resource use and policies and potential sources of inefficiency in the context of forestry, fisheries, and exhaustible energy resources will be studied.

- Further, the basics of Economics of Climate change, its implications and policies.

Learning outcomes

The Learning Outcomes of this course are as follows:

- The students get familiarise with basic issues of sustainable resources allocation and economics of climate change.
- It will familiarize students with the Cost-Benefit Analysis, Challenges in estimating costs and benefits of greenhouse gas policies, the Environmental Kuznets curve, and Mitigation of climate change.
- The course will familiarize students with Sustainable development Goals SDGs, History of Convention UNFCCC, India's Intended Nationally Determined Contribution.

Syllabus

UNIT I: Mathematical Prerequisites (12 hours)

*Difference equations; differential equations; phase plane analysis; dynamic optimization

Optimal extraction of a non-renewable resource, Optimal management of renewable resources -Fishery and Forestry

Readings:

*Alpha C. Chiang. (2000). *Elements of dynamic optimization*. Waveland Press.

*Hoel, M. (2016). *Optimal control theory with applications to resource and environmental economics* (No. 08/2016).

Perman, R., Yu, M., McGilvray, J., & Common, M. (2003). Natural Resource and Environmental. *Economics Pearson Education Limited. Edinburgh Gate Harbor Essex CM20 2JE and Associated Companies throughout the world*.

Chapter 15-introduction, 15.1-15.3, 15.6

Chapter 17-Introduction, 17.1-17.3,17.5,17.6(17.6.1 and 17.6.2)

Chapter 18

***Teacher may consult this reading for mathematical prerequisite. The following topics will not be evaluated:** Difference equations; differential equations; phase plane analysis; dynamic optimization

UNIT II: Energy Economics, Energy Transition, and Energy Security (12 hours)
Introduction to Basics of supply, demand, and prices, income elasticities, the eco- nomics of depletable resources, world oil markets, Pathways of energy transition from conventional to renewable energy sources, Policy instruments, Energy security, accessibility and A definition, and Energy poverty

Readings Bhattacharyya, S. C. (2019). *Energy economics: concepts, issues, markets and governance*. Springer Nature.

Chapter 3-3.1-3.5, Annexure-3.1,3.2

Chapter 9

Chapter 11-11.1,11.4 onwards

Chapter 14-14.1,14.3.2-14.3.4

Chapter 20-till 20.6

Chapter 22-22.1-22.6

Tietenberg, T., & Lewis, L. (2018). *Environmental and natural resource economics*. Routledge.

Chapter 7

UNIT III: The Economics of Climate change, Implications, and Policies (12 hours)
Cost-Benefit Analysis, Challenges in estimating costs and benefits of greenhouse gas policies, Environmental Kuznets curve, Mitigation of climate change, Sectoral impact of Climate change, climate change, and inequality, Policy responses, and instruments

Readings:

Harris, J. M., Roach, B., & Environmental, J. M. H. (2007). The economics of global climate change. Global Development and Environment Institute Tufts University

Stern, N.(2007) The Economics of Climate Change: The Stern Review, Cambridge University Press

Part(IV)-Chapter 14,15,16

Stern, D. I. (2004). The rise and fall of the environmental Kuznets curve. World Development, 32(8), 1419-1439.(Till Section 7)

Arnell, N. W., Brown, S., Gosling, S. N., Gottschalk, P., Hinkel, J., Huntingford, C., ... & Zelazowski, P. (2016). The impacts of climate change across the globe: a multi-sectoral assessment. Climatic Change, 134(3), 457-474.

Roberts, J. T. (2001). Global inequality and climate change. Society & Natural Resources, 14(6), 501-509.

UNIT IV: Sustainable Development (09 hours)
Concepts and Measurement, Weak and strong sustainability, Sustainable development Goals SDGs, History of Convention UNFCCC, India’s Intended Nationally Determined Contribution
Readings:

Geoffrey Heal (2012). “Reflections—Defining and Measuring Sustainability” Review of Environmental Economics and Policy Vol. 6, No. 1 (winter 2012), p. 147–163.

Recommended readings

- Harris, J. M., Roach, B., & Environmental, J. M. H. (2007). *The economics of global climate change. Global Development and Environment* Institute Tufts University.
- Pelling, M. (2010). *Adaptation to climate change: from resilience to transformation*.Routledge.
- Callan, Scott, and Janet Thomas. *Environmental Economics and Management: Theory*, Policy and Applications. 4th ed. Florence, KY: South-Western, 2006, chapter 3. Perman, R., Yu, M., McGilvray, J., & Common, M. (2003). Natural Resource and Environmental. *Economics Pearson Education Limited. Edinburgh Gate Harbor Essex CM20 2JE and Associated Companies throughout the world*.
- Bhattacharyya, S. C. (2019). *Energy economics: concepts, issues, markets and governance*. Springer Nature.
- Tietenberg, T., & Lewis, L. (2018). *Environmental and natural resource economics*. Routledge.
- ISBN:9780324320671.
- Barrett, S. (1990) *The problem of global environmental protection,* Oxford Review ofEconomicPolicy6(1):68–79
- Stern,N.(2007) The Economics of Climate Change: The Stern Review, Cambridge UniversityPress.
- Stern, D. I. (2004). *The rise and fall of the environmental Kuznets curve*. World Development, 32(8),1419-1439.
- Babiker, Mustafa, John Reilly, and Henry Jacoby. ”The Kyoto Protocol and DevelopingCountries.”EnergyPolicy28,no.8(2000):525-36.
- IPCC Climate Change 2014: Mitigation of Climate Change (in the press); http://mitigation2014.or draft
- Arnell, N. W., Brown, S., Gosling, S. N., Gottschalk, P., Hinkel, J., Huntingford, C., ... & Zelazowski, P. (2016). *The impacts of climate change across the globe: a multi-sectoral assessment*. Climatic Change, 134(3),457-474.
- Roberts, J. T. (2001). *Global inequality and climate change*. Society & Natural Resources, 14(6),501-509.
- Geoffrey Heal (2012). “Reflections—Defining and Measuring Sustainability” Review of Environmental Economics and Policy Vol. 6, No. 1 (winter 2012), p. 147–163.
- Theenvironmentwrite,2009. ”Definingsustainability: weaksustainability”.

Natural Resource and Environmental Economics

Third Edition

Roger Perman
Yue Ma
James McGilvray
Michael Common

PEARSON
Addison Wesley

Harlow, England • London • New York • Boston • San Francisco • Toronto
Sydney • Tokyo • Singapore • Hong Kong • Seoul • Taipei • New Delhi
Cape Town • Madrid • Mexico City • Amsterdam • Munich • Paris • Milan

CHAPTER 15

The theory of optimal resource extraction: non-renewable resources

Behold, I have played the fool, and have erred exceedingly. **1 Samuel 26:21**

Learning objectives

After the end of this chapter the reader should be able to

- understand the concept of non-renewable resources
- appreciate the distinctions between alternative measures of resource stock, such as base resource, resource potential and resource reserves
- understand the role of resource substitution possibilities and the ideas of a backstop technology and a resource choke price
- construct and solve simple discrete time and continuous time models of optimal resource depletion
- understand the meaning of a socially optimal depletion programme, and why this may differ from privately optimal programmes
- carry out simple comparative dynamic analysis in the context of resource depletion models, and thereby determine the consequences of changes in interest rates, known stock size, demand, price of backstop technology, and resource extraction costs
- compare resource depletion outcomes in competitive and monopolistic markets
- identify the consequences of taxes and subsidies on resource net prices and resource revenues
- understand the concept of natural resource scarcity, and be aware of a variety of possible measures of scarcity

Introduction

Non-renewable resources include fossil-fuel energy supplies – oil, gas and coal – and minerals – copper and nickel, for example. They are formed by geological processes over millions of years and so, in effect, exist as fixed stocks which, once extracted, cannot be renewed. One question is of central importance: what is the optimal extraction path over time for any particular non-renewable resource stock?

We began to answer this question in Chapter 14. There the optimal extraction problem was solved for a special case in which there was one homogeneous (uniform-quality) non-renewable resource. By assuming a single homogeneous stock, the possibility that substitute non-renewable resources exist is ruled out. The only substitution possibilities considered in Chapter 14 were between the non-renewable resource and other production inputs (labour and capital).

But in practice, non-renewable resources are heterogeneous. They comprise a set of resources varying in chemical and physical type (such as oil, gas, uranium, coal, and the various categories of each of these) and in terms of costs of extraction (as a result of differences in location, accessibility, quality and so on). This chapter investigates the efficient and optimal extraction of one component of this set of non-renewable resources where substitution possibilities exist. Substitution will take place if the price of the resource rises to such an extent that it makes alternatives economically more attractive. Consider, for example, the case of a country that has been exploiting its coal reserves, but in which coal extraction costs rise as lower-quality seams are mined. Meanwhile, gas costs fall as a result of the application of superior extraction and distribution technology. A point may be reached where electricity producers will substitute gas for coal in power generation. It is this kind of process that we wish to be able to model in this chapter.

Although the analysis that follows will employ a different (and in general, simpler) framework from that used in Chapter 14, one very important result carries over to the present case. The Hotelling rule is a necessary efficiency condition that must be satisfied by *any* optimal extraction programme. The chapter begins by laying out the conditions for the extraction path of a non-renewable resource stock to be socially optimal. It then considers how a resource is likely to be depleted in a market economy. As you would expect from the analysis in Chapters 5 and 11, the extraction path in competitive market economies will, under certain circumstances, be socially optimal. It is usually argued that one of these circumstances is that resource markets are competitive. We investigate this matter by comparing extraction paths under competitive and monopoly market structures against the benchmark of a 'first-best' social optimum.

The model used in most of this chapter is simple, and abstracts considerably from specific detail. The assumptions are gradually relaxed to deal with increasingly complex situations. To help understanding, it is convenient to begin with a model in which only two periods of time are considered. Even from such a simple starting point, very powerful results can be obtained, which can be generalised to analyses involving many periods. If you have a clear understanding of Hotelling's rule from Chapter 14, you might wish to skip the two-period model in the next section. Then, having analysed optimal depletion in a two-period model, a more general model is examined in which depletion takes place over T periods, where T may be a very large number.

There are two principal simplifications used in the chapter. First, we assume that utility comes directly from consuming the extracted resource. This is a considerably simpler, yet more specialised, case than that investigated in Chapter 14 where utility derived from consumption goods, obtained through a production function with a natural resource, physical capital (and, implicitly, labour) as inputs. Although doing this pushes the production function into the background, more attention is given to another kind of substitution possibility. As we remarked above, other non-renewable resources also exist. If one or more of these serve as substitutes for the resource being considered, this is likely to have important implications for economically efficient resource depletion paths.

Second, we do not take any account of adverse external effects arising from the extraction or consumption of the resource. The reader may find this rather surprising given that the production and consumption of non-renewable fossil-energy fuels are the primary cause of many of the world's most serious environmental problems. In particular, the combustion of these fuels accounts for between 55% and 88% of carbon dioxide emissions, 90% of sulphur dioxide, and 85% of nitrogen oxide emissions (IEA, 1990). In addition, fossil fuel use accounts for significant proportions of trace-metal emissions.

However, the relationship between non-renewable resource extraction over time and environmental degradation is so important that it warrants separate attention. This will be given in Chapter 16. Not surprisingly, we will show that the optimal extraction path will be different if adverse externalities are present causing environmental damage. The depletion model developed in this chapter will be used in Chapter 16 to derive some important results about efficient pollution targets and instruments.

Finally, a word about presentation. A lot of tedious – although not particularly difficult – mathematics is required to derive our results. The main text of this chapter lays emphasis on key results and the intuition which lies behind them; derivations, where they are lengthy, are placed in appendices. You may find it helpful to omit these on a first reading.

For much of the discussion in this chapter, it is assumed that there exists a known, finite stock of each kind of non-renewable resource. This assumption is not always appropriate. New discoveries are made, increasing the magnitude of known stocks, and technological change alters the proportion of mineral resources that are economically recoverable. Later sections indicate how the model may be extended to deal with some of these complications. Box 15.1 – which you should read now – considers several measures of resource stock, and throws some light on the issue of whether it can be reasonable to assume that there are fixed quantities of non-renewable resources.

Box 15.1 Are stocks of non-renewable resources fixed?

Non-renewable resources include a large variety of mineral deposits – in solid, liquid and gaseous forms – from which, often after some process of refining, metals, fossil fuels and other processed minerals are obtained. The crude forms of these resources are produced over very long periods of time by chemical, biological or physical processes. Their rate of formation is sufficiently slow in timescales relevant to humans that it is sensible to label such resources non-renewable. At any point in time, there exists some fixed, finite quantities of these resources in the earth's crust and environmental systems, albeit very large quantities in some cases.

So, in a physical sense, it is appropriate to describe non-renewable resources as existing in fixed quantities. However, that description may not be appropriate in an economic sense. To see why not, consider the information shown in Table 15.1. The final column – *Base resource* – indicates the mass of each resource that is thought to exist in the earth's crust. This is the measure closest to that we had in mind in the previous paragraph. However, most of this base resource consists of the mineral in very dispersed form, or at great depths below the surface. Base resource figures such as these are the broadest sense in which one might use the term 'resource stocks'. In each case, the measure is purely physical, having little or no relationship to economic measures of stocks. Notice that each of these quantities is extremely large relative to any other of the indicated stock measures.

The column labelled *Resource potential* is of more relevance to our discussions, comprising estimates of the upper limits on resource extraction possibilities given current and expected technologies. Whereas the resource base is a pure physical measure, the resource potential is a measure incorporating physical and technological information. But this illustrates the difficulty of classifying and measuring resources; as time passes, technology will almost certainly change, in ways that cannot be predicted today. As a result, estimates of the resource potential will change (usually rising) over time. To some writers, the possibility that resource constraints on economic activity will bite depends primarily on whether or not technological improvement in extracting usable materials from the huge stocks of base resources (thereby augmenting resource potential) will continue more-or-less indefinitely.

However, an economist is interested not in what is technically feasible but in what would become available under certain conditions. In other words, he or she is interested in resource supplies, or potential supplies. These will, of course, be shaped by physical and technological factors. But they will also depend upon resource market prices and the costs of extraction via their influence on exploration and research effort and on expected profitability. Data in the column labelled *World reserve base* consist of estimates of the upper bounds of resource stocks (including reserves that have not yet been discovered) that are economically recoverable under 'reasonable expectations' of future price, cost and technology possibilities. Those labelled *Reserves* consist of quantities that are economically recoverable under present configurations of costs and prices.

In economic modelling, the existence of fixed mineral resource stocks is often used as a simplifying assumption. But our observations suggest that we should be wary of this. In the longer term, economically relevant stocks are not fixed, and will vary with changing economic and technological circumstances.

Box 15.1 *continued*

Table 15.1 Production, consumption and reserves of some important resources: 1991 (figures in millions of metric tons)

	Production	Reserves		World reserve base		Consumption	Resource potential	Base resource
		Quantity	Reserve life (yrs)	Reserve base	Base life (yrs)			Base resource (crustal mass)
Aluminium	112.22	23 000	222	28 000	270	19.46	3 519 000	1 990 000 000 000
Iron ore	929.75	150 000	161	230 000	247	959.6	2 035 000	1 392 000 000 000
Potassium	na	20 000	800	na	>800	25	na	408 000 000 000
Manganese	25	800	32	5 000	200	22	42 000	31 200 000 000
Phosphorus	na	110	na	na	270	na	51 000	28 800 000 000
Fluorine	na	2.5	na	na	12	na	20 000	10 800 000 000
Sulphur	56.87	na	na	na	na	57.5	na	9 600 000 000
Chromium	13	419	32	1 950	150	13	3 260	2 600 000 000
Zinc	7.137	140	20	330	46	6.993	3 400	2 250 000 000
Nickel	0.922	47	51	111	119	0.882	2 590	2 130 000 000
Copper	9.29	310	33	590	64	10.714	2 120	1 510 000 000
Lead	3.424	63	18	130	38	5.342	550	290 000 000
Tin	0.179	8	45	10	56	0.218	68	40 000 000
Tungsten	0.0413	3.5	80	>3.5	>80	0.044	51	26 400 000
Mercury	0.003	0.130	43	0.240	80	0.005	3.4	2 100 000
Silver	0.014	0.28	20	na	na	0.02	2.8	1 800 000
Platinum	0.0003	0.37	124	na	na	0.00029	1.2	1 100 000

Source: Figures compiled from a variety of sources

15.1 A non-renewable resource two-period model

Consider a planning horizon that consists of two periods, period 0 and period 1. There is a fixed stock of known size of one type of a non-renewable resource. The initial stock of the resource (at the start of period 0) is denoted $\bar{S}$. Let R_t be the quantity extracted in period t and assume that an inverse demand function exists for this resource at each time, given by

$$P_t = a - bR_t$$

where P_t is the price in period t, with a and b being positive constant numbers. So, the demand functions for the two periods will be:

$$P_0 = a - bR_0$$

$$P_1 = a - bR_1$$

These demands are illustrated in Figure 15.1.

A linear and negatively sloped demand function such as this one has the property that demand goes to zero at some price, in this case the price a. Hence, either this resource is non-essential or it possesses a substitute which at the price a becomes economically more attractive. The assumption of linearity of demand is arbitrary and so you should bear in mind that the particular results derived below are conditional upon the assumption that the demand curve is of this form.

The shaded area in Figure 15.1 (algebraically, the integral of P with respect to R over the interval $R = 0$ to $R = R_t$) shows the total benefit consumers obtain from consuming the quantity R_t in period t. From a social point of view, this area represents the gross social benefit, B, derived from the extraction and consumption of quantity R_t of the resource.[1] We can express this quantity as

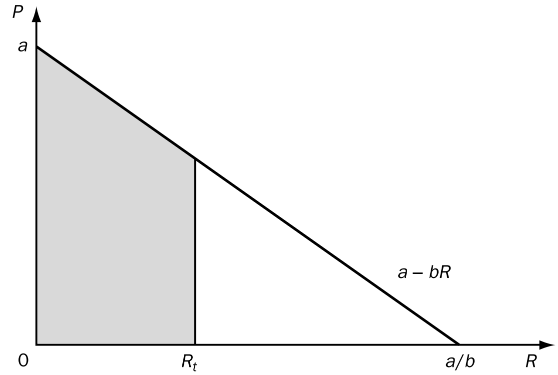


Figure 15.1 The non-renewable resource demand function for the two-period model

$$B(R_t) = \int_0^{R_t} (a - bR)\mathrm{d}R$$

$$= aR_t - \frac{b}{2}R_t^2$$

where the notation $B(R_t)$ is used to make explicit the fact that the gross benefit at time t (B_t) is dependent on the quantity of the resource extracted and consumed (R_t).

However, the gross benefit obtained by consumers is not identical to the net social benefit of the resource, as resource extraction involves costs. In this chapter, we assume that these costs are fully borne by the resource-extracting firms, and so private and social costs are identical.[2] This assumption will be dropped in the following chapter. Let us define c to be the constant marginal cost of extracting the resource ($c \geq 0$).[3] Then total extraction costs, C_t, for the extracted quantity R_t units will be

$$C_t = cR_t$$

The total net social benefit from extracting the quantity R_t is

[1] A demand curve is sometimes taken as providing information about the marginal willingness to pay (or marginal benefit) for successive units of the good in question. The area under a demand curve up to some given quantity is, then, the sum of a set of marginal benefits, and is equal to the total benefit derived from consuming that quantity.

[2] We also assume that benefits represented in the resource demand function are the *only* benefits to society, so there are no beneficial externalities.

[3] Constancy of marginal costs of extraction is a very strong assumption. In the previous chapter, we investigated a more general case in which marginal extraction costs are not necessarily constant. We do not consider this any further here. Later in this chapter, however, we do analyse the consequences for extraction of a once-and-for-all rise in extraction costs.

$$\text{NSB}_t = B_t - C_t$$

where NSB denotes the total net social benefit and B is the gross social benefit of resource extraction and use.[4] Hence

$$\text{NSB}(R_t) = \int_0^{R_t} (a - bR)\mathrm{d}R - cR_t = aR_t - \frac{b}{2}R_t^2 - cR_t \tag{15.1}$$

15.1.1 A socially optimal extraction policy

Our objective in this subsection is to identify a socially optimal extraction programme. This will serve as a benchmark in terms of which any particular extraction programme can be assessed. In order to find the socially optimal extraction programme, two things are required. The first is a social welfare function that embodies society's objectives; the second is a statement of the technical possibilities and constraints available at any point in time. Let us deal first with the social welfare function, relating this as far as possible to our discussion of social welfare functions in Chapters 3 and 5.

As in Chapter 3, the social welfare function that we shall use is discounted utilitarian in form. So the general two-period social welfare function

$$W = W(U_0, U_1)$$

takes the particular form

$$W = U_0 + \frac{U_1}{1+\rho}$$

where ρ is the social utility discount rate, reflecting society's time preference. Now regard the utility in each period as being equal to the net social benefit in each period.[5] Given this, the social welfare function may be written as

$$W = \text{NSB}_0 + \frac{\text{NSB}_1}{1+\rho}$$

Only one relevant technical constraint exists in this case: there is a fixed initial stock of the non-renewable resource, $\bar{S}$. We assume that society wishes to have none of this resource stock left at the end of the second period. Then the quantities extracted in the two periods, R_0 and R_1, must satisfy the constraint:[6]

$$R_0 + R_1 = \bar{S}$$

The optimisation problem can now be stated. Resource extraction levels R_0 and R_1 should be chosen to maximise social welfare, W, subject to the constraint that total extraction of the resources over the two periods equals $\bar{S}$. Mathematically, this can be written as

$$\max_{R_0, R_1} W = \text{NSB}_0 + \frac{\text{NSB}_1}{1+\rho}$$

subject to

$$R_0 + R_1 = \bar{S}$$

There are several ways of obtaining solutions to constrained optimisation problems of this form. We use the Lagrange multiplier method, a technique that was explained in Appendix 3.1. The first step is to form the Lagrangian function, L:

$$\begin{aligned} L = W - \lambda(\bar{S} - R_0 - R_1) &= (\text{NSB}_0) + \left(\frac{\text{NSB}_1}{1+\rho}\right) \\ - \lambda(\bar{S} - R_0 - R_1) &= \left(aR_0 - \frac{b}{2}R_0^2 - cR_0\right) \\ &+ \left(\frac{aR_1 - \frac{b}{2}R_1^2 - cR_1}{1+\rho}\right) - \lambda(\bar{S} - R_0 - R_1) \end{aligned} \tag{15.2}$$

[4] Strictly speaking, social benefits derive from consumption (use) of the resource, not extraction *per se*. However, we assume throughout this chapter that all resource stocks extracted in a period are consumed in that period, and so this distinction becomes irrelevant.

[5] In order to make such an interpretation valid, we shall assume that the demand function is 'quasi-linear' (see Varian, 1987). Suppose there are two goods, X, the good whose demand we are interested in, and Y, money to be spent on all other goods. Quasi-linearity requires that the utility function for good X be of the form $U = V(X) + Y$. This implies that income effects are absent in the sense that changes in income do not affect the demand for good X. In this case, we can legitimately interpret the area under the demand curve for good X as a measure of utility.

[6] The problem could easily be changed so that a predetermined quantity S^* ($S^* \geq 0$) must be left at the end of period 1 by rewriting the constraint as $R_0 + R_1 + S^* = \bar{S}$. This would not alter the essence of the conclusion we shall reach.

in which λ is a 'Lagrange multiplier'. Remembering that R_0 and R_1 are choice variables – variables whose value must be selected to maximise welfare – the necessary conditions include:

$$\frac{\partial L}{\partial R_0} = a - bR_0 - c + \lambda = 0 \tag{15.3}$$

$$\frac{\partial L}{\partial R_1} = \frac{a - bR_1 - c}{1 + \rho} + \lambda = 0 \tag{15.4}$$

Since the right-hand side terms of equations 15.3 and 15.4 are both equal to zero, this implies that

$$a - bR_0 - c = \frac{a - bR_1 - c}{1 + \rho}$$

Using the demand function $P_t = a - bR_t$, the last equation can be written as

$$P_0 - c = \frac{P_1 - c}{1 + \rho}$$

where P_0 and P_1 are gross prices and $P_0 - c$ and $P_1 - c$ are net prices. A resource's net price is also known as the resource rent or resource royalty. Rearranging this expression, we obtain

$$\rho = \frac{(P_1 - c) - (P_0 - c)}{(P_0 - c)}$$

If we change the notation used for time periods so that $P_0 = P_{t-1}$, $P_1 = P_t$ and $c = c_t = c_{t-1}$, we then obtain

$$\rho = \frac{(P_t - c_t) - (P_{t-1} - c_{t-1})}{(P_{t-1} - c_{t-1})} \tag{15.5}$$

which is equivalent to a result we obtained previously in Chapter 14, equation 14.15, commonly known as Hotelling's rule. Note that in equation 15.5, P is a gross price whereas in equation 14.15, P refers to a net price, resource rent or royalty. However, since $P - c$ in equation 15.5 is the resource net price or royalty, these two equations are identical (except for the fact that one is in discrete-time notation and the other in continuous-time notation).

What does this result tell us? The left-hand side of equation 15.5, ρ, is the social utility discount rate, which embodies some view about how future utility should be valued in terms of present utility. The right-hand side is the proportionate rate of growth of the resource's net price. So if, for example, society chooses a discount rate of 0.1 (or 10%), Hotelling's rule states that an efficient extraction programme requires the net price of the resource to grow at a proportionate rate of 0.1 (or 10%) over time.

Now we know how much higher the net price should be in period 1 compared with period 0, if welfare is to be maximised; but what should be the level of the net price in period 0? This is easily answered. Recall that the economy has some fixed stock of the resource that is to be entirely extracted and consumed in the two periods. Also, we have assumed that the demand function for the resource is known. An optimal extraction programme requires two gross prices, P_0 and P_1, such that the following conditions are satisfied:

$$P_0 = a - bR_0$$

$$P_1 = a - bR_1$$

$$R_0 + R_1 = \bar{S}$$

$$P_1 - c = (1 + \rho)(P_0 - c)$$

This will uniquely define the two prices (and so the two quantities of resources to be extracted) that are required for welfare maximisation. Problem 1, at the end of this chapter, provides a numerical example to illustrate this kind of two-period optimal depletion problem. You are recommended to work through this problem before moving on to the next section.

15.2 A non-renewable resource multi-period model

Having investigated resource depletion in the simple two-period model, the analysis is now generalised to many periods. It will be convenient to change from a discrete-time framework (in which there is a number of successive intervals of time, denoted period 0, period 1, etc.) to a continuous-time framework which deals with rates of extraction and use at particular points in time over some continuous-time horizon.[7]

[7] The material in this section, in particular the worked example investigated later, owes much to Heijman (1990).

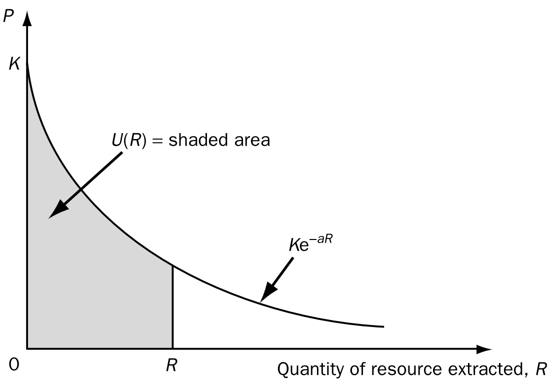


Figure 15.2 A resource demand curve, and the total utility from consuming a particular quantity of the resource

To keep the maths as simple as possible, we will push extraction costs somewhat into the background. To do this, P is now defined to be the net price of the non-renewable resource, that is, the price after deduction of the cost of extraction. Let $P(R)$ denote the inverse demand function for the resource, indicating that the resource net price is a function of the quantity extracted, R. The social utility from consuming a quantity R of the resource may be defined as

$$U(R) = \int_0^R P(R)\mathrm{d}R \tag{15.6a}$$

which is illustrated by the shaded area in Figure 15.2. You will notice that the demand curve used in Figure 15.2 is non-linear. We shall have more to say about this particular form of the demand function shortly.

By differentiating total utility with respect to R, the rate of resource extraction and use, we obtain

$$\frac{\partial U}{\partial R} = P(R) \tag{15.6b}$$

which states that the marginal social utility of resource use equals the net price of the resource.

Assume, as for the two-period model, that the intertemporal social welfare function is utilitarian. Future utility is discounted at the instantaneous social utility discount rate ρ. Then the value of social welfare over an interval of time from period 0 to period T can be expressed as[8]

$$W = \int_0^T U(R_t)\mathrm{e}^{-\rho t}\mathrm{d}t$$

Our problem is to make social-welfare-maximising choices of

(a) R_t, for $t = 0$ to $t = T$ (that is, we wish to choose a quantity of resource to be extracted in each period), and
(b) the optimal value for T (the point in time at which depletion of the resource stock ceases), subject to the constraint that

$$\int_0^T R_t\mathrm{d}t = \bar{S}$$

where $\bar{S}$ is the total initial stock of the non-renewable resource. That is, the total extraction of the resource is equal to the size of the initial resource stock. Note that in this problem, the time horizon to exhaustion is being treated as an endogenous variable to be chosen by the decision maker.

We define the remaining stock of the natural resource at time t, S_t, as

$$S_t = \bar{S} - \int_0^t R_t\mathrm{d}t$$

then by differentiation with respect to time, we obtain

[8] It may be helpful to relate this form of social welfare function to the discrete-time versions we have been using previously. We have stated that a T-period discrete-time discounted welfare function can be written as

$$W = U_0 + \frac{U_1}{1+\rho} + \frac{U_2}{(1+\rho)^2} + \ldots + \frac{U_T}{(1+\rho)^T}$$

We could write this equivalently as

$$W = \sum_{t=0}^{t=T} \frac{U_t}{(1+\rho)^T}$$

A continuous-time analogue of this welfare function is then

$$W = \int_{t=0}^{t=T} U_t\mathrm{e}^{-\rho t}\mathrm{d}t$$

$$\dot{S}_t = -R_t$$

where $\dot{S} = \mathrm{d}S/\mathrm{d}t$, the rate of change of the remaining resource stock with respect to time.

So the dynamic optimisation problem involves the choice of a path of resource extraction R_t over the interval $t = 0$ to $t = T$ that satisfies the resource stock constraint and which maximises social welfare, W. Mathematically, we have:

$$\operatorname{Max} W = \int_0^T U(R_t)\mathrm{e}^{-\rho t}\mathrm{d}t$$

subject to $\dot{S}_t = -R_t$

It would be a useful exercise at this point for you to use the optimisation technique explained in Appendix 14.1 to derive the solution to this problem. Your derivation can be checked against the answer given in Appendix 15.1.

Thinking point

Before moving on to interpret the main components of this solution, it will be useful to pause for a moment to reflect on the nature of this model. It is similar in general form to the model we investigated in Chapter 14, and laid out in full in Appendix 14.2. However, the model is simpler in one important way from that of the previous chapter as utility is derived directly from the consumption of the natural resource, rather than indirectly from the consumption goods generated through a production function. There is a fixed, total stock available of the natural resource, and this model is sometimes called the 'cake-eating' model of resource depletion.

It would also be reasonable to interpret this model as one in which a production function exists implicitly. However, this production function has just one argument – the non-renewable natural resource input – as compared with the two arguments – the natural resource and human-made capital – in the model of Chapter 14.

It is clear that this model can at most be regarded as a partial account of economic activity. One possible interpretation of this partial status is that the economy also produces, or could produce, goods and services through other production functions, using capital, labour and perhaps renewable resource inputs. In this interpretation the non-renewable resource is like a once-and-for-all gift of nature. Using this non-renewable resource provides something over and above the welfare possible from production in its absence. It is this *additional welfare* that is being measured by our term W.

An alternative interpretation is more commonly found in the literature. Here, non-renewable resources consist of a diverse set of different resources. Each element of this set is a particular resource that is fixed and homogeneous. Substitution possibilities exist between at least some elements of this set of resources. For historical, technical or economic reasons, production might currently rely particularly heavily on one kind of resource. Changing technological or economic conditions might lead to this stock being replaced by another. With the passage of time, a sequence of resource stocks are brought into play, with each one eventually being replaced by another. In this story, what our resource depletion model investigates is one stage in this sequence of depletion processes. This interpretation will be used later in the chapter when the concepts of a backstop technology and a choke price are introduced.

These comments raise a general issue about choices that need to be made in doing resource modelling. It is often too difficult to explain everything of interest in one framework. Sometimes, one needs to pick 'horses for courses'. In the previous chapter, we were concerned with substitution between natural resources and physical capital; that required that we explicitly specify a conventional type of production function. In this chapter, that is not of central concern, and so the production function can be allowed to slip somewhat into the background. However, we do wish here to place emphasis on substitution processes between natural resources. That can be done in a simple way, by paying greater attention to the nature of resource demand functions, and to the idea of a choke price for a resource.

Whether or not you have succeeded in obtaining a formal solution to this optimisation problem, intuition should suggest one condition that must be satisfied if W is to be maximised. R_t must be chosen so that the *discounted* marginal utility is equal at each point in time, that is,

$$\frac{\partial U}{\partial R}e^{-\rho t} = \text{constant}$$

To understand this, let us use the method of contradiction. If the discounted marginal utilities from resource extraction were not equal in every period, then total welfare W could be increased by shifting some extraction from a period with a relatively low discounted marginal utility to a period with a relatively high discounted marginal utility. Rearranging the path of extraction in this way would raise welfare. It must, therefore, be the case that welfare can only be maximised when discounted marginal utilities are equal. What follows from this result? First note equation 15.6b again:

$$\frac{\partial U_t}{\partial R_t} = P_t$$

So, the requirement that the discounted marginal utility be constant is equivalent to the requirement that the discounted net price is constant as well – a result noted previously in Chapter 14. That is,

$$\frac{\partial U_t}{\partial R_t}e^{-\rho t} = P_t e^{-\rho t} = \text{constant} = P_0$$

Rearranging this condition, we obtain

$$P_t = P_0 e^{\rho t} \tag{15.7a}$$

By differentiation[9] this can be rewritten as

$$\frac{\dot{P}_t}{P_t} = \rho \tag{15.7b}$$

This is, once again, the Hotelling efficiency rule. It now appears in a different guise, because of our switch to a continuous-time framework. The rule states that the net price or royalty P_t of a non-renewable resource should rise at a rate equal to the social utility discount rate, ρ, if the social value of the resource is to be maximised.

We now know the rate at which the resource net price or royalty must rise. However, this does not fully characterise the solution to our optimising problem. There are several other things we need to know too. First, we need to know the optimal initial value of the resource net price. Secondly, we need to know over how long a period of time the resource should be extracted – in other words, what is the optimal value of T? Thirdly, what is the optimal rate of resource extraction at each point in time? Finally, what should be the values of P and R at the end of the extraction horizon?

It is not possible to obtain answers to these questions without one additional piece of information: the particular form of the resource demand function. So let us suppose that the resource demand function is

$$P(R) = Ke^{-aR} \tag{15.8}$$

which is illustrated in Figure 15.2.[10] Unlike the demand function used in the two-period analysis, this function exhibits a non-linear relationship between P and R, and is probably more representative of the form that resource demands are likely to take than the linear function used in the section on the two-period model. However, it is similar to the previous demand function in so far as it exhibits zero demand at some finite price level.

To see this, just note that $P(R = 0) = K$. K is the so-called *choke price* for this resource, meaning that

---

[9] Differentiation of equation 15.7a with respect to time gives

$$dP_t/dt \equiv \dot{P}_t = P_0\rho e^{\rho t}$$

By substitution of equation 15.7a into this expression, we obtain

$$\dot{P}_t = \rho P_t$$

and dividing through by P_t we obtain

$$\dot{P}_t/P_t = \rho$$

as required.

[10] For the demand function given in equation 15.8, we can obtain the particular form of the social welfare function as follows. The social utility function corresponding to equation 15.6a will be:

$$U(R) = \int_0^R P(R)dR = \int_0^R Ke^{-aR}dR = \frac{k}{a}(1 - e^{-aR})$$

The social welfare function, therefore, is

$$W = \int_0^T U(R_t)e^{-\rho t}dt = \int_0^T \frac{K}{a}(1 - e^{-aR_t})e^{-\rho t}dt$$

the demand for the resource is driven to zero or is 'choked off' at this price. At the choke price people using the services of this resource would switch demand to some alternative, substitute, non-renewable resource, or to an alternative final product not using that resource as an input.

As we shall demonstrate shortly, given knowledge of

- a particular resource demand function,
- Hotelling's efficiency condition,
- an initial value for the resource stock, and
- a final value for the resource stock,

it is possible to obtain expressions for the optimal initial, interim and final resource net price (royalty) and resource extraction rates. What about the final stock level? This is straightforward. An optimal solution must have the property that the stock goes to zero at exactly the same point in time that demand and extraction go to zero.[11] If that were not the case, some resource will have been needlessly wasted. So we know that the solution must include $S_T = 0$ and $R_T = 0$, with resource stocks being positive, and positive extraction taking place over all time up to T. As you will see below, that will give us sufficient information to fully tie down the solution.

Before we proceed to obtain all the details of the solution, one important matter must be reiterated. The solution to a problem of this type will depend upon the demand function chosen. Hence the particular solutions derived below are conditional upon the demand function chosen, and will not be valid in all circumstances. Our model in this chapter assumes that the resource has a choke price, implying that a substitute for the resource becomes economically more attractive at that price. If you wish to examine the case in which there is no choke price – indeed, where there is no finite upper limit on the resource price – you may find it useful to work through some of the exercises provided in the *Additional Materials* for this chapter, which deal with this case among others.

As the mathematics required to obtain the full solution are rather tedious (but not particularly difficult), the derivations are presented in Appendix 15.1. You are strongly recommended to read this now, but if you prefer to omit these derivations, the results are presented in Table 15.2. There it can be seen that all the expressions for the initial, interim and final resource royalty (or net prices) and rate of resource extraction are functions of the parameters of the model (K, ρ and a) and T, the optimal depletion time. As the final expression indicates, T is itself a function of those parameters. Given the functional forms we have been using in this section, if the values of the parameters K, ρ and a were known, it would be possible to solve the model to obtain numerical values for all the variables of interest over the whole period for which the resource will be extracted.

Table 15.2 Optimality conditions for the multi-period model

	Initial ($t = 0$)	Interim ($t = t$)	Final ($t = T$)
Royalty, P	$P_0 = Ke^{-\sqrt{2\rho\bar{S}a}}$	$P_t = Ke^{\rho(t-T)}$	$P_T = K$
Extraction, R	$R_0 = \sqrt{\frac{2\rho\bar{S}}{a}}$	$R_t = \frac{\rho}{a}(T-t)$	$R_T = 0$
Depletion time			$T = \sqrt{\frac{2\bar{S}a}{\rho}}$

Figure 15.3 portrays the solution to our optimal depletion model. The diagram shows the optimal resource extraction and net price paths over time corresponding to social welfare maximisation. As we show subsequently, it also represents the profit-maximising extraction and price paths in perfectly competitive markets. In the upper right quadrant, the net price is shown rising exponentially at the social utility discount rate, ρ, thereby satisfying the Hotelling rule. The upper left quadrant shows the resource demand curve with a choke price K. The lower left quadrant gives the optimal extraction path of the non-renewable resource, which is, in this case, a linear declining function of time.

The net price is initially at P_0, and then grows until it reaches the choke price K at time T. At this point, demand for the resource goes to zero, and the accumulated extraction of the resource (the shaded

[11] In terms of optimisation theory, this constitutes a so-called terminal condition for the problem.

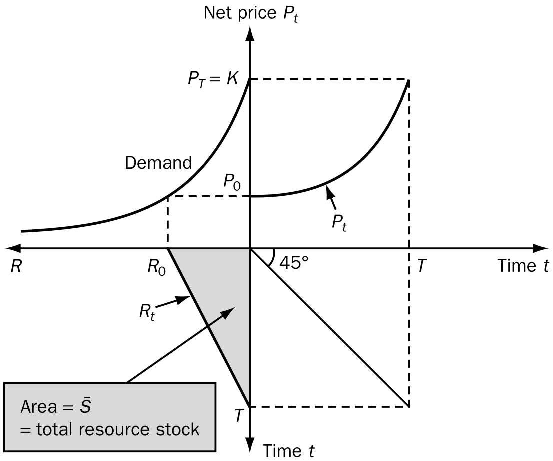


Figure 15.3 Graphical representation of solutions to the optimal resource depletion model

area beneath the extraction path) is exactly equal to the total initial resource stock, $\bar{S}$. The lower right quadrant maps the time axes by a 45° line. A worked numerical example illustrating optimal extraction is presented in Appendix 15.3.

15.3 Non-renewable resource extraction in perfectly competitive markets

Until this point, we have said nothing about the kind of market structure in which decisions are made. It is as if we have been imagining that a rational social planner were asked to make decisions that maximise social welfare, given the constraints facing the economy. The optimality conditions listed in Table 15.2, plus the Hotelling efficiency condition, are the outcome of the social planner's calculations.

How will matters turn out if decisions are instead the outcome of profit-maximising decisions in a perfectly competitive market economy? This section demonstrates that, *ceteris paribus*, the outcomes will be identical. Hotelling's rule and the optimality conditions of Table 15.2 are also obtained under a perfect competition assumption.

Suppose there are m competitive firms in the market. Use the subscript j to denote any one of these m firms. Assume, for simplicity, that all firms have equal and constant marginal costs of extracting the resource. Now as all firms in a competitive market face the same fixed selling price at any point in time, the market royalty will be identical over firms. Given the market royalty P_t, each firm chooses an amount to extract and sell, $R_{j,t}$, to maximise its profits.

Mathematically, the jth firm's objective is to maximise

$$\int_0^T \Pi_{j,t} e^{-it} dt$$

subject to

$$\int_0^T \left(\sum_{j=1}^m R_{j,t}\right) dt = \bar{S}$$

where $\Pi_j = P \cdot R_j$ is firm j's profit and i is the market interest rate. Note that the same stock constraint operates on all firms collectively; the industry as a whole cannot extract more than the fixed initial stock over the whole time horizon. The profit-maximising extraction path is obtained when each firm selects an extraction $R_{j,t}$ at each time, $t = 0$ to $t = T$, so that its discounted marginal profit will be the same at any point in time t, that is,

$$M\Pi_{j,t} e^{-it} = \frac{\partial \Pi_{j,t}}{\partial R_{j,t}} e^{-it} = \frac{\partial PR_{j,t}}{\partial R_{j,t}} e^{-it} = P_t e^{-it}$$
$$= \text{constant, for } t = 0 \text{ to } t = T$$

where $M\Pi_j$ is firm j's marginal profit function. If discounted marginal profits were *not* the same over time, total profits could be increased by switching extraction between time periods so that more was extracted when discounted profits were high and less when they were low. The result that the discounted marginal profit is the same at any point in time implies that

$$P_t e^{-it} = P_0 \text{ or } P_t = P_0 e^{it}$$

Not surprisingly, Hotelling's efficiency rule continues to be a required condition for profit maximisation, so that the market net price of the resource must grow over time at the rate i. The interest rate in this profit maximisation condition is the market rate of interest. Our analysis in Chapter 11 showed that, in perfectly competitive capital markets and in the absence of transactions costs, the market interest rate will be equal to r, the consumption rate of interest, and also to δ, the rate of return on capital.

We appear now to have two different efficiency conditions,

$$\frac{\dot{P}}{P} = \rho \text{ and } \frac{\dot{P}}{P} = i$$

the former emerging from maximising social welfare, the latter from private profit maximisation. But these are in fact identical conditions under the assumptions we have made in this chapter; by assuming that we can interpret areas under demand curves (that is, gross benefits) as quantities of utility, we in effect impose the condition that $\rho = r$. Given this result, it is not difficult to show, by cranking through the appropriate maths in a similar manner to that done in Appendix 15.1, that all the results of Table 15.2 would once again be produced under perfect competition, provided the private market interest rate equals the social consumption discount rate. We leave this as an exercise for the reader.

Finally, note that the appearance of a positive net price or royalty, $P_t > 0$, for non-renewable resources reflects the fixed stock assumption. If the resource existed in unlimited quantities (that is, the resource were not scarce) net prices would be zero in perfect competition, as the price of the product will equal the marginal cost (c), a result which you may recall from standard theory of long-run equilibrium in competitive markets. In other words, scarcity rent would be zero as there would be no scarcity.

15.4 Resource extraction in a monopolistic market

It is usual to assume that the objective of a monopoly is to maximise its discounted profit over time. Thus, it selects the net price P_t (or royalty) and chooses the output R_t so as to maximise

$$\int_0^T \Pi_t e^{-it} dt$$

subject to

$$\int_0^T R_t dt = \bar{S}$$

where $\Pi_t = P(R_t)R_t$.

For the same reason as in the case of perfect competition, the profit-maximising solution is obtained by choosing a path for R so that the discounted marginal profit will be the same at any time. So we have

$$M\Pi_t e^{-it} = \frac{\partial \Pi_t}{\partial R_t} e^{-it} = \text{constant} = M\Pi_0$$

that is,

$$M\Pi_t = M\Pi_0 e^{it} \tag{15.9}$$

Looking carefully at equation 15.9, and comparing this with the equation for marginal profits in the previous section, it is clear why the profit-maximising solutions in monopolistic and competitive markets will differ. Under perfect competition, the market price is exogenous to (fixed for) each firm. Thus we are able to obtain the result that in competitive markets, marginal revenue equals price. However, in a monopolistic market, price is not fixed, but will depend upon the firm's output choice. Marginal revenue will be less than price in this case.

The necessary condition for profit maximisation in a monopolistic market states that the marginal profit (and not the net price or royalty) should increase at the rate of interest i in order to maximise the discounted profits over time. The solution to the monopolist's optimising problem is derived in Appendix 15.2. If you wish to omit this, you will find the results in Table 15.3.

15.5 A comparison of competitive and monopolistic extraction programmes

Table 15.3 summarises the results concerning optimal resource extraction in perfectly competitive and monopolistic markets. The analytical results presented are derived in Appendices 15.1 and 15.2. For convenience, we list below the notation used in Table 15.3.

the Word file *polcos.doc*. These can both be found in the *Additional Materials* for Chapter 15.

15.6 Extensions of the multi-period model of non-renewable resource depletion

To this point, a number of simplifying assumptions in developing and analysing our model of resource depletion have been made. In particular, it has been assumed that

- the utility discount rate and the market interest rate are constant over time;
- there is a fixed stock, of known size, of the non-renewable natural resource;
- the demand curve is identical at each point in time;
- no taxation or subsidy is applied to the extraction or use of the resource;
- marginal extraction costs are constant;
- there is a fixed 'choke price' (hence implying the existence of a backstop technology);
- no technological change occurs;
- no externalities are generated in the extraction or use of the resource.

We shall now undertake some comparative dynamic analysis. This consists of finding how the optimal paths of the variables of interest change over time in response to changes in the levels of one or more of the parameters in the model, or of finding how the optimal paths alter as our assumptions are changed. We adopt the device of investigating changes to one parameter, holding all others unchanged, comparing the new optimal paths with those derived above for our simple multi-period model. (We shall only discuss these generalisations for the case of perfect competition; analysis of the monopoly case is left to the reader as an exercise.)

The reader interested in doing comparative dynamics analysis by Excel simulation may wish to explore the file *hmodel.xls* (together with its explanatory document, *hmodel.doc*) in the *Additional Materials* to Chapter 15. The consequences of each of the changes described in the following subsections can be verified using that Excel workbook.

15.6.1 An increase in the interest rate

Let us make clear the problem we wish to answer here. Suppose that the interest rate we had assumed in drawing Figure 15.3 was 6% per year. Now suppose that the interest rate was not 6% but rather 10%; how would Figure 15.3 have been different if the interest rate had been higher in this way? This is the kind of question we are trying to answer in doing comparative dynamics.

The answer is shown in Figure 15.5. The thick, heavily drawn line represents the original optimal price path, with the price rising from an initial level of P_0 to its choke price, K, at time T. Now suppose that the interest rate rises. Since the resource's net price must grow at the market interest rate, an increase in i will raise the growth rate of the resource royalty, P_t; hence the new price path must have a steeper slope than the original one. The new price path will be the one labelled C in Figure 15.5. It will have an initial price lower than the one on the original price path, will grow more quickly, and will reach its final (choke) price earlier in time (before $t = T$). This result can be explained by the following observations. First, the choke price itself, K, is not altered by the interest rate change. Second, as we have already observed, the new price path must rise more steeply with a higher interest rate. Third, we can deduce that it must begin from a lower initial price level from using the resource exhaustion constraint. The change in interest rate does not alter the

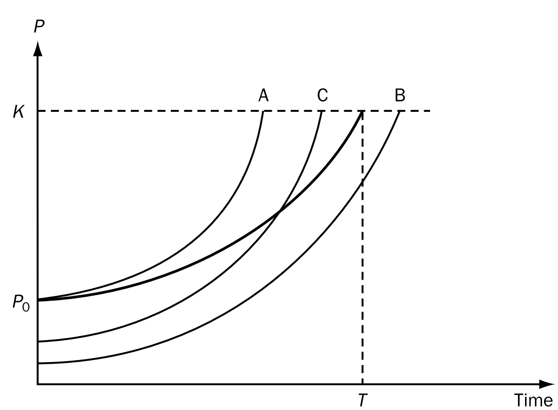


Figure 15.5 The effect of an increase in the interest rate on the optimal price of the non-renewable resource

quantity that is to be extracted; the same total stock is extracted whatever the interest rate might be. If the price path began from the same initial value (P_0) then it would follow a path such as that shown by the curve labelled A and would reach its choke price before $t = T$. But then the price would always be higher than along the original price path, but for a shorter period of time. Hence the resource stock will not be fully extracted along path A and that path could not be optimal.

A path such as B is not feasible. Here the price is always lower (and so the quantity extracted is higher) than on the original optimal path, and for a longer time. But that would imply that more resources are extracted over the life of the resource than were initially available. This is not feasible. The only feasible and optimal path is one such as C. Here the price is lower than on the original optimal path for some time (and so the quantity extracted is greater); then the new price path crosses over the original one and the price is higher thereafter (and so the quantity extracted is lower).

Note that because the new path must intersect the original path from below, the optimal depletion time will be shorter for a higher interest rate. This is intuitively reasonable. Higher interest rate means greater impatience. More is extracted early on, less later, and total time to full exhaustion is quicker. The implications for all the variables of interest are summarised in Figure 15.6.

15.6.2 An increase in the size of the known resource stock

In practice, estimates of the size of reserves of non-renewable resources such as coal and oil are under constant revision. *Proven reserves* are those unextracted stocks known to exist and can be recovered at current prices and costs. *Probable reserves* are stocks that are known, with near certainty, to exist but which have not yet been fully explored or researched. They represent the best guess of additional amounts that could be recovered at current price and cost levels. *Possible reserves* are stocks in geological structures near to proven fields. As prices rise, what were previously uneconomic stocks become economically recoverable.

Consider the case of a single new discovery of a fossil fuel stock. Other things being unchanged, if the royalty path were such that its initial level remained unchanged at P_0, then given the fact that the rate of royalty increase is unchanged, some proportion of the reserve would remain unutilised by the time the choke price, K, is reached. This is clearly neither efficient nor optimal. It follows that the initial royalty must be lower and the time to exhaustion is extended. At the time the choke price is reached, T', the new enlarged resource stock will have just reached complete exhaustion, as shown in Figure 15.7.

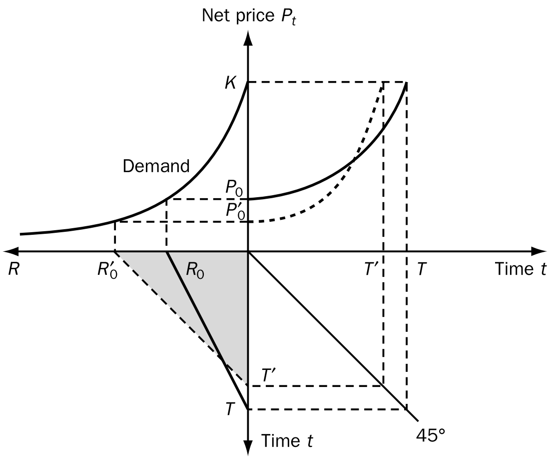


Figure 15.6 An increase in interest rates in a perfectly competitive market

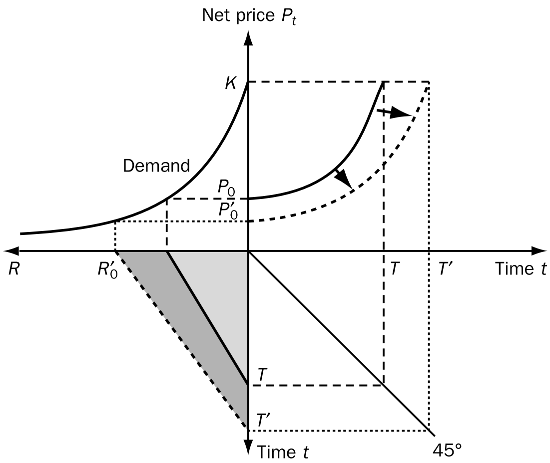


Figure 15.7 An increase in the resource stock

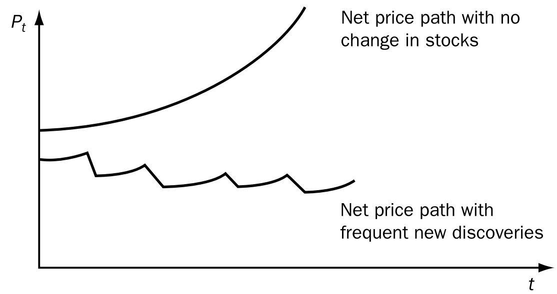


Figure 15.8 The effect of frequent new discoveries on the resource net price or royalty

Now suppose that there is a sequence of new discoveries taking place over time, so that the size of known reserves increases in a series of discrete steps. Generalising the previous argument, we would expect the behaviour of the net price or royalty over time to follow a path similar to that illustrated in Figure 15.8. This hypothetical price path is one that is consistent with the actual behaviour of oil prices.

15.6.3 Changing demand

Suppose that there is an increase in demand for the resource, possibly as a result of population growth or rising real incomes. The demand curve thus shifts outwards. Given this change, the old royalty or net price path would result in higher extraction levels, which will exhaust the resource before the net price has reached K, the choke price. Hence the net price must increase to dampen down quantities demanded; as Figure 15.9 shows, the time until the resource stock is fully exhausted will also be shortened.

15.6.4 A fall in the price of backstop technology

In the model developed in this chapter, we have assumed there is a choke price, K. If the net price were to rise above K, the economy will cease consumption of the non-renewable resource and switch to an alternative source – the backstop source. Suppose that technological progress occurs, increasing the efficiency of a backstop technology. This

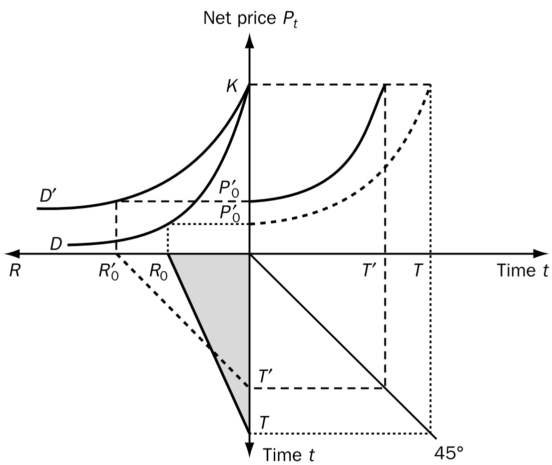


Figure 15.9 The effect of an increase in demand for the resource

will tend to reduce the price of the backstop source, to P_B ($P_B < K$). Hence the choke price will fall to P_B. Given the fall in the choke price to P_B, the initial value of the resource net price on the original optimal price path, P_0, cannot now be optimal. In fact, it is too high since the net price would reach the new choke price before T, leaving some of the economically useful resource unexploited. So the initial price of the non-renewable resource, P_0, must fall to a lower level, P'_0, to encourage an increase in demand so that a shorter time horizon is required until complete exhaustion of the non-renewable resource reserve. This process is illustrated in Figure 15.10. Note that when the resource price reaches the new, reduced choke price, demand for the non-renewable resource falls to zero.

15.6.5 A change in resource extraction costs

Consider the case of an increase in extraction costs, possibly because labour charges rise in the extraction industry. To analyse the effects of an increase in extraction costs, it is important to distinguish carefully between the net price and the gross price of the resource. Let us define:

$$p_t = P_t - c$$

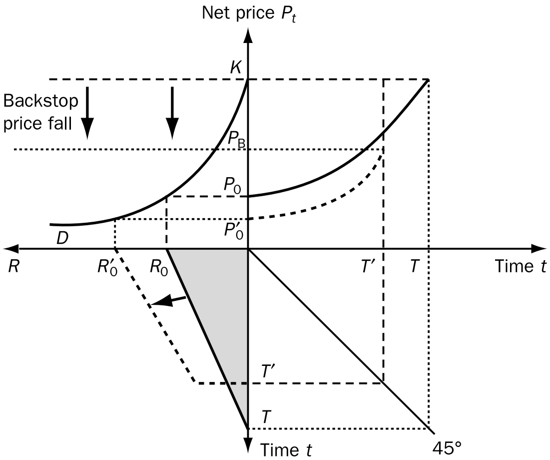


Figure 15.10 A fall in the price of a backstop technology

where p_t is the resource net price, P_t is the gross price of the non-renewable resource, and c is the marginal extraction cost, assumed to be constant. Hotelling's rule requires that the resource *net price* grows at a constant rate, equal to the discount rate (which we take here to be constant at the rate i). Therefore, efficient extraction requires that

$$p_t = p_0 e^{it}$$

Now look at Figure 15.11(a). Suppose that the marginal cost of extraction is at some constant level, c_L, and that the curve labelled *Original net price* describes the optimal path of the net price over time (i.e. it plots $p_t = p_0e^{it}$); also suppose that the corresponding optimal gross price path is given by the curve labelled *Original gross price* (i.e. it plots $P_t = p_t + c_L = p_0e^{it} + c_L$).

Next, suppose that the cost of extraction, while still constant, now becomes somewhat higher than was previously the case. Its new level is denoted c_H. We suppose that this change takes place at the initial time period, period 0. Consider first what would happen if the gross price remained unchanged at its initial level, as shown in Figure 15.11(a). The increase in unit extraction costs from c_L to c_H would then result in the net price being lower than its original initial level. However, with no change having occurred in the interest rate, the net price must *grow* at the same rate as before. Although the net price grows at the same rate as before, it does so from a lower starting value, and so it follows that the new net price p_t would be lower at all points in time than the original net price, and it will also have a flatter profile (as close inspection of the diagram makes clear). This implies that the new gross price will be lower than the old gross price at all points in time except in the original period.

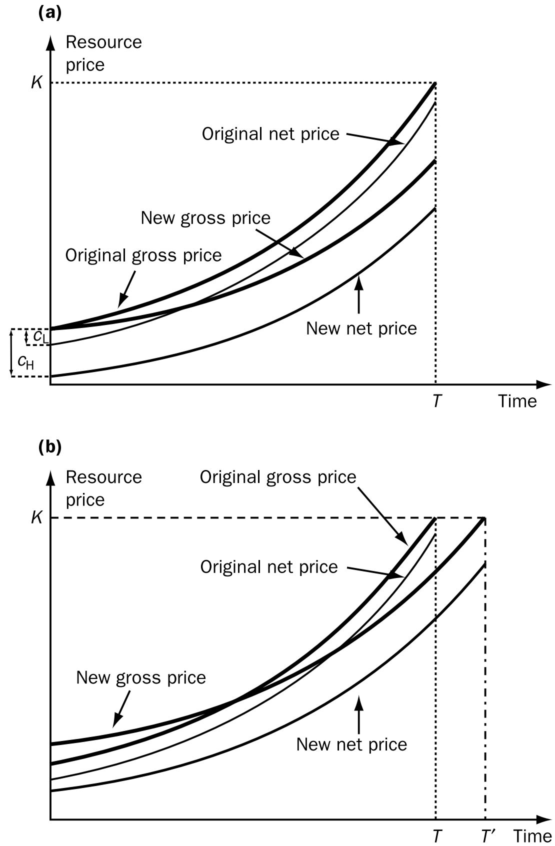


Figure 15.11 (a) An increase in extraction costs: deducing the effects on gross and net prices;
(b) An increase in extraction costs: actual effects on gross and net prices

However, the positions of the curves for the new gross and net prices in Figure 15.11(a) cannot be optimal. If the gross (market) price is lower at all points in time except period 0, more extraction would take place in every period. This would cause the reserve to become completely exhausted before

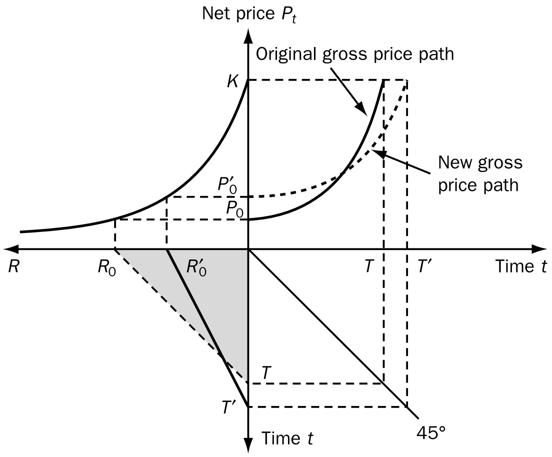


Figure 15.12 A rise in extraction costs

the choke price (K) is reached. This cannot be optimal, as any optimal extraction path must ensure that demand goes to zero at the same point in time as the remaining resource stock goes to zero.

Therefore, optimal extraction requires that the new level of the gross price in period 0, P'_0, must be greater than it was originally (P_0). It will remain above the original gross price level for a while but will, at some time before the resource stock is fully depleted, fall below the old gross price path. This is the final outcome that we illustrate in Figure 15.11(b). As the new gross price eventually becomes lower than its original level, it must take longer before the choke price is reached. Hence the time taken before complete resource exhaustion occurs is lengthened.

All the elements of this reasoning are assembled together in the four-quadrant diagram shown in Figure 15.12. A rise in extraction costs will raise the initial gross price, slow down the rate at which the gross price increases (even though the net price or royalty increases at the same rate as before), and lengthen the time to complete exhaustion of the stock.

What about a fall in extraction costs? This may be the consequence of technological progress decreasing the costs of extracting the resource from its reserves. By following similar reasoning to that we used above, it can be deduced that a fall in extraction costs will have the opposite effects to those just described. It will lower the initial gross price, increase the rate at which the gross price increases (even though the net price increases at the same rate as before), and shorten the time to complete exhaustion of the stock.

If the changes in extraction cost were very large, then our conclusions may need to be amended. For example, if a cost increase were very large, then it is possible that the new gross price in period 0, P'_0, will be above the choke price. It is then not economically viable to deplete the remaining reserve – an example of an economic exhaustion of a resource, even though, in physical terms, the resource stock has not become completely exhausted.

One remaining point needs to be considered. Until now it has been assumed that the resource stock consists of reserves of uniform, homogeneous quality, and the marginal cost of extraction was constant for the whole stock. We have been investigating the consequences of increases or decreases in that marginal cost schedule from one fixed level to another. But what if the stock were not homogeneous, but rather consisted of reserves of varying quality or varying accessibility? It is not possible here to take the reader through the various possibilities that this opens up. It is clear that in this situation marginal extraction costs can no longer be constant, but will vary as different segments of the stock are extracted. There are many meanings that could be attributed to the notion of a change in marginal extraction costs. A fall in extraction costs may occur as the consequence of new, high-quality reserves being discovered. An increase in costs may occur as a consequence of a high-quality mine becoming exhausted, and extraction switching to another mine in which the quality of the resource reserve is somewhat lower. Technical progress may result in the whole profile of extraction costs being shifted downwards, although not necessarily at the same rate for all components.

We do not analyse these cases in this text. The suggestions for further reading point the reader to where analysis of these cases can be found. But it should be evident that elaborating a resource depletion model in any of these ways requires dropping the assumption that there is a known, fixed quantity of the resource. Instead, the amount of the resource that is 'economically' available becomes an endogenous variable, the value of which depends upon resource demand and extraction cost schedules. This also implies that we could analyse a reduction in extraction costs as if it were a form of technological progress; this can increase the stock of the reserve that can be extracted in an economically viable manner. Hence, changes in resource extraction costs and changes in resource stocks become interrelated – rather than independent – phenomena.

15.7 The introduction of taxation/subsidies

15.7.1 A royalty tax or subsidy

A royalty tax or subsidy will have no effect on a resource owner's extraction decision for a reserve that is currently being extracted. The tax or subsidy will alter the present value of the resource being extracted, but there can be no change in the rate of extraction over time that can offset that decline or increase in present value. The government will simply collect some of the mineral rent (or pay some subsidies), and resource extraction and production will proceed in the same manner as before the tax/subsidy was introduced.

This result follows from the Hotelling rule of efficient resource depletion. To see this, define α to be a royalty tax rate (which could be negative – that is, a subsidy), and denote the royalty or net price at time t by p_t. Then the post-tax royalty becomes $(1 - \alpha)p_t$. But Hotelling's rule implies that the post-tax royalty must rise at the discount rate, i, if the resource is to be exploited efficiently. That is:

$$(1 - \alpha)p_t = (1 - \alpha)p_0\mathrm{e}^{it}$$

or

$$p_t = p_0\mathrm{e}^{it}$$

Hotelling's rule continues to operate unchanged in the presence of a royalty tax, and no change occurs to the optimal depletion path. This is also true for a royalty subsidy scheme. In this case, denoting the royalty subsidy rate by β, we have the efficiency condition

$$(1 + \beta)p_t = (1 + \beta)p_0\mathrm{e}^{it} \Rightarrow p_t = p_0\mathrm{e}^{it}$$

We can conclude that a royalty tax or subsidy is neutral in its effect on the optimal extraction path. However, a tax may discourage (or a subsidy encourage) the exploration effort for new mineral deposits by reducing (increasing) the expected pay-off from discovering the new deposits.

15.7.2 Revenue tax/subsidy

The previous subsection analysed the effect of a tax or subsidy on resource royalties. We now turn our attention to the impact of a revenue tax (or subsidy). In the absence of a revenue tax, the Hotelling efficiency condition is, in terms of net prices and gross prices,

$$p_t = p_0\mathrm{e}^{it}$$
$$\Rightarrow (P_t - c) = (P_0 - c)\mathrm{e}^{it}$$

Under a revenue tax scheme, with a tax of α per unit of the resource sold, the post-tax royalty or net price is

$$p_t = (1 - \alpha)P_t - c$$

So Hotelling's rule becomes:

$$[(1 - \alpha)P_t - c] = [(1 - \alpha)P_0 - c]\mathrm{e}^{it} \quad (0 < \alpha < 1)$$

$$\Rightarrow \left(P_t - \frac{c}{1-\alpha}\right) = \left(P_0 - \frac{c}{1-\alpha}\right)\mathrm{e}^{it}$$

Since $c/(1 - \alpha) > c$, an imposition of a revenue tax is equivalent to an increase in the resource extraction cost. Similarly, for a revenue subsidy scheme, we have

$$\left(P_t - \frac{c}{1+\beta}\right) = \left(P_0 - \frac{c}{1+\beta}\right)\mathrm{e}^{it} \quad (0 < \beta < 1)$$

A revenue subsidy is equivalent to a decrease in extraction cost. We have already discussed the effects of a change in extraction costs, and you may recall the results we obtained: a decrease in extraction costs will lower the initial gross price, increase the rate at which the gross price increases (even though the net price or royalty increases at the same rate as before) and shorten the time to complete exhaustion of the stock.

CHAPTER 17 Renewable resources

It will appear, I hope, that most of the problems associated with the words 'conservation' or 'depletion' or 'overexploitation' in the fishery are, in reality, manifestations of the fact that the natural resources of the sea yield no economic rent. Fishery resources are unusual in the fact of their common-property nature; but they are not unique, and similar problems are encountered in other cases of common-property resource industries, such as petroleum production, hunting and trapping, etc.

Gordon (1954)

Learning objectives

After studying this chapter, the reader should be able to

- understand the biological growth function of a renewable resource, and the notions of compensation and depensation in growth processes
- interpret the simple logistic growth model, and some of its variants, including models with critical depensation
- understand the idea of a sustainable yield and the maximum sustainable yield
- distinguish between steady-state outcomes and dynamic adjustment processes that may (or may not) lead to a steady-state outcome
- specify, and solve for its bioeconomic equilibrium outcome, an open-access fishery, a static private-property fishery, and a present value (PV)-maximising fishery
- undertake comparative statics analysis and simple dynamic analysis for open-access and private-property models
- explain under what conditions the stock, effort and harvesting outcomes of private fisheries will not be socially efficient
- describe conditions which increase the likelihood of severe resource depletion or species extinction
- understand the workings, and relative advantages, of a variety of policy instruments that are designed to conserve renewable resource stocks and/or promote socially efficient harvesting

Introduction

Environmental resources are described as renewable when they have a capacity for reproduction and growth. The class of renewable resources is diverse. It includes populations of biological organisms such as fisheries and forests which have a natural capacity for growth, and water and atmospheric systems which are reproduced by physical or chemical processes. While the latter do not possess *biological* growth capacity, they do have some ability to assimilate pollution inputs (thereby maintaining their quality) and, at least in the case of water resources, can self-replenish as stocks are run down (thereby maintaining their quantity).

It is also conventional to classify arable and grazing lands as renewable resources. In these cases reproduction and growth take place by a combination of biological processes (such as the recycling of organic nutrients) and physical processes (irrigation, exposure to wind, etc.). Fertility levels can regenerate naturally so long as the demands made on the soil are not excessive. We may also consider more broadly defined environmental systems (such as wilderness areas or tropical moist forests) as being sets of interrelated renewable resources.

The categories just described are renewable stock resources. A broad concept of renewables would also include flow resources such as solar, wave, wind and geothermal energy. These share with biological stock resources the property that current harnessing of the flow does not mean that the total magnitude of the future flow will necessarily be smaller. Indeed, many forms of energy-flow resources are, for all practical purposes, non-depletable.

Given this diversity of resource types, it will be necessary to restrict what will, and will not, be discussed here. Most of the literature on the economics of renewable resources is about two things: the harvesting of animal species ('hunting and fishing') and the economics of forestry. This chapter is largely concerned with the former; forestry economics is the subject of the following chapter. Agriculture could also be thought of as a branch of renewable resource harvesting. But agriculture – particularly in its more developed forms – differs fundamentally from other forms of renewable resource exploitation in that the environmental medium in which it takes place is designed and controlled. The growing medium is manipulated through the use of inputs such as fertilisers, pesticides, herbicides; temperatures may be controlled by the use of greenhouses and the like; and plant stocks are selected or even genetically modified. In that sense, there is little to differentiate a study of (developed) agricultural economics from the economics of manufacturing. For this reason, we do not survey the huge literature that is 'agricultural economics' in this text, although some of the environmental consequences of agricultural activity are discussed in *Agriculture.doc* in the *Additional Materials*. For reasons of space, we also do not cover the economics of renewable flow resources. Again, a brief outline of some of the main issues is given in *Renewables.doc* in the *Additional Materials*.

It is important to distinguish between stocks and flows of the renewable resource. The stock is a measure of the quantity of the resource existing at a point in time, measured either as the aggregate mass of the biological material (the biomass) in question (such as the total weight of fish of particular age classes or the cubic metres of standing timber), or in terms of population numbers. The flow is the change in the stock over an interval of time, where the change results either from biological factors, such as 'recruitment' of new fish into the population through birth or 'exit' due to natural death, or from harvesting activity.

One similarity between renewable and non-renewable resources is that both are capable of being fully exhausted (that is, the stock being driven to zero) if excessive and prolonged harvesting or extraction activity is carried out. In the case of non-renewable resources, exhaustibility is a consequence of the finiteness of the stock. For renewable resources, although the stock can grow, it can also be driven to zero if conditions interfere with the reproductive capability of the renewable resource, or if rates of harvesting continually exceed net natural growth.

It is evident that enforceable private property rights do not exist for many forms of renewable resource. In the absence of regulation or collective control over harvesting behaviour, the resource stocks are subject to open access. We will demonstrate that open-access resources tend to be over-exploited in both a biological and an economic sense, and that the likelihood of the resource being harvested to the point of exhaustion is higher than where private property rights are established and access to harvesting can be restricted.

As we have said, this chapter is principally about the harvesting of animal resources. Our exposition focuses on marine fishing. With some modifications, the fishery economics modelling framework can be used to analyse most forms of renewable resource exploitation. We begin by setting out a simple model of the biological growth of a fish population. Then the properties of commercial fisheries are examined under two sets of institutional arrangements: an open-access fishery and a profit-maximising fishery in which enforceable private property rights exist.

For the case of the private-property fishery, the analysis proceeds in two steps. First we examine what is usually known as the static private-property fishery. The analysis is kept simple by abstracting from the need to deal explicitly with the passage of time. We do this by focusing attention on steady-state (or equilibrium) outcomes in which variables are taken to be unchanging over time. Some unspecified interval of time is chosen to be representative of all periods in that steady state. The equilibrium is found by solving the model for its profit-maximising solution. By construction that equilibrium would apply to every time period, provided that economic and biological conditions remain unchanged.

The second step involves a generalisation in which the passage of time is modelled explicitly. In this case we investigate a private-property fishery that is managed so as to maximise its present value over an infinite lifetime. All nominal-value flows are discounted at some positive rate to convert to present-value equivalents. Describing the second variant as a generalisation of the first is appropriate because – as we shall show – the static private fishery turns out to be a special case of the present-value-maximising fishery in which owners adopt a zero discount rate. Where discounting takes place at some positive rate, the outcomes of the two models differ.

In common with the practice throughout this text, we also examine the outcomes of the various commercial fishery regimes against the benchmark of a *socially efficient* fishery. We demonstrate that under some conditions the harvesting programme of a competitive fishery where private property rights to the resource stocks are established and enforceable will be socially efficient. However, actual resource harvesting regimes are typically not socially efficient, even where attempts have been made to introduce private property rights. Among the reasons why they are not is the existence of various kinds of externalities. Open-access regimes will almost certainly generate inefficient outcomes. The chapter concludes by examining a set of policy instruments that could be introduced in an attempt to move harvesting behaviour closer to that which is socially efficient.

17.1 Biological growth processes

In order to investigate the economics of a renewable resource, it is first necessary to describe the pattern of biological (or other) growth of the resource. To fix ideas, we consider the growth function for a population of some species of fish. This is conventionally called a fishery. We suppose that this fishery has an intrinsic (or potential) growth rate denoted by g. This is the proportional rate at which the fish stock would grow when its size is small relative to the carrying capacity of the fishery, and so the fish face no significant environmental constraints on their reproduction and survival. The intrinsic growth rate g may be thought of as the difference between the population's birth and natural mortality rate (again, where the population size is small relative to carrying capacity). Suppose that the population stock is S and it grows at a fixed rate g. Then in the absence of human predation the rate of change of the population over time is given by[1]

$$\frac{\mathrm{d}S}{\mathrm{d}t} \equiv \dot{S} = gS \tag{17.1}$$

By integrating this equation, we obtain an expression for the stock level at any point in time:

$$S_t = S_0 \mathrm{e}^{gt}$$

in which S_0 is the initial stock level. In other words, for a positive value of g, the population grows exponentially over time at the rate g and without bounds. This is only plausible over a short span of time. Any population of fish exists in a particular environmental milieu, with a finite carrying capacity, which sets bounds on the population's growth possibilities.

A simple way of representing this effect is by making the actual (as opposed to the potential)

[1] Be careful not to confuse a rate of change with a rate of growth. A rate of change refers to how much extra is produced in some interval of time. A rate of growth is that rate of change divided by its current size (to measure the change in proportionate terms). Note that we shall sometimes refer to an 'amount of growth' (as opposed to a growth rate); this should be read as the population size change over some interval of time.

growth rate depend on the stock size. Then we have what is called density-dependent growth. Using the symbol χ to denote the actual growth rate, the growth function can then be written as

$$\dot{S} = \chi(S)S$$

where $\chi(S)$ states that χ is a function of S, and shows the dependence of the actual growth rate on the stock size. If this function has the property that the proportionate growth rate of the stock ($\dot{S}/S$) declines as the stock size rises then the function is said to have the property of compensation.

Now let us suppose that under a given set of environmental conditions there is a finite upper bound on the size to which the population can grow (its carrying capacity). We will denote this as S_{MAX}. A commonly used functional form for $\chi(S)$ which has the properties of compensation and a maximum stock size is the simple logistic function:

$$\chi(S) = g\left(1 - \frac{S}{S_{MAX}}\right)$$

in which the constant parameter $g > 0$ is what we have called the intrinsic or potential growth rate of the population. Where the logistic function determines the actual population growth rate, we may therefore write the biological growth function as

$$\dot{S} \equiv \frac{dS}{dt} = g\left(1 - \frac{S}{S_{MAX}}\right)S \tag{17.2}$$

The changes taking place in the fish population that we have been referring to so far are 'natural' changes. But as we want to use the notation $\dot{S}$ and dS/dt in the rest of this chapter to refer to the *net* effect of natural changes and human predation, we shall use the alternative symbol G to refer to stock changes due only to natural causes. (More completely, we shall use the notation $G(S)$ to make it clear that G depends on S.) With this change the logistic biological growth function is

$$G(S) = g\left(1 - \frac{S}{S_{MAX}}\right)S \tag{17.3}$$

The logistic form is a good approximation to the natural growth processes of many fish, animal and bird populations (and indeed to some physical systems such as the quantity of fresh water in an underground reservoir). Some additional information on the simple logistic growth model, and alternative forms of logistic growth, is given in Box 17.1. Problem 1 at the end of the chapter also explores the logistic model a little further, and invites you to explore another commonly used equation for biological growth, the Gompertz function.

Box 17.1 The logistic form of the biological growth function

Logistic growth is one example of density-dependent growth: processes where the growth rate of a population depends on the population size. It was first applied to fisheries by Schaefer (1957). The equation for logistic growth of a renewable resource population was given by equation 17.3.

Simple logistic growth is illustrated in Figure 17.1(a), which represents the relationship between the stock size and the associated rate of change of the population due to biological growth. Three properties should be noted by inspection of that diagram.

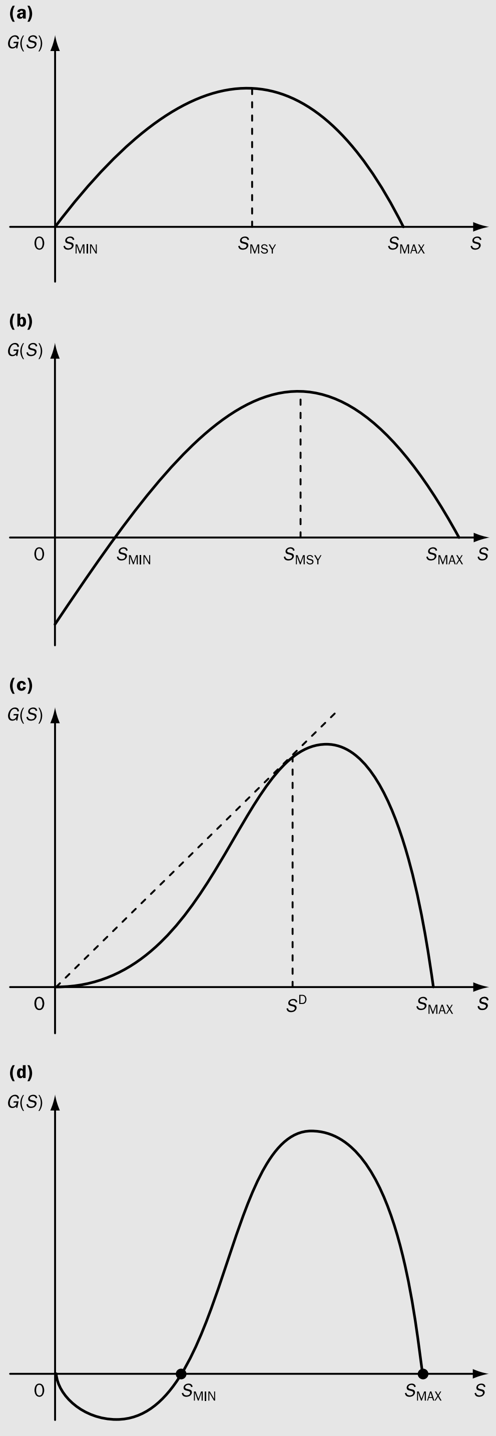
(a) S_{MAX} is the maximum stock size that can be supported in the environmental milieu. This value is, of course, conditional on the particular environment circumstances that happen to prevail, and would change if any of those circumstances (such as ocean temperature or stocks of nutrients) change.

(b) By multiplication through of the terms on the right-hand side of equation 17.3, it is clear that the amount of growth, G, is a quadratic function of the resource stock size, S. For the simple logistic function, the maximum amount of growth (S_{MSY}) will occur when the stock size is equal to half of S_{MAX}.

(c) The amount of biological growth G is zero only at a stock size of zero and a stock size of S_{MAX}. For all intermediate values, growth is positive.

This last property may appear to be obviously true, but it turns out to be seriously in error in many cases. It implies that for any population

Box 17.1 *continued*



size greater than zero natural growth will lead to a population increase if the system is left undisturbed. In other words, the population does not possess any positive lower threshold level.

However, suppose there is some positive population threshold level, S_{MIN}, such that the population would inevitably and permanently decline to zero if the actual population were ever to fall below that threshold. A simple generalisation of the logistic growth function that has this property is:

$$G(S) = g(S - S_{MIN})\left(1 - \frac{S}{S_{MAX}}\right) \qquad (17.4)$$

Note that if $S_{MIN} = 0$, equation 17.4 collapses to the special case of equation 17.3. The generalisation given by equation 17.4 is illustrated in Figure 17.1(b). Several other generalisations of the logistic growth model exist. For example, the modified logistic model:

$$G(S) = gS^{\alpha}\left(1 - \frac{S}{S_{MAX}}\right)$$

has the property that for $\alpha > 1$ there is, at low stock levels, *depensation*, which is a situation where the proportionate growth rate $(G(S)/S)$ is an increasing function of the stock size, as opposed to being a decreasing function (compensation) in the simple logistic case where $\alpha = 1$. A biological growth function exhibiting depensation at stock levels below S^D (and compensation thereafter) is shown in Figure 17.1(c).

Finally, the generalised logistic function

$$G(S) = g\left(\frac{S}{S_{MIN}} - 1\right)\left(1 - \frac{S}{S_{MAX}}\right)S$$

exhibits what is known as *critical depensation*. As with equation 17.4, the stock falls irreversibly to zero if the stock ever falls below S_{MIN}. This function is represented in Figure 17.1(d). It should be evident that if a growth process does exhibit critical depensation, then the probability of the stock being harvested to complete depletion is increased, and increased considerably if S_{MIN} is a large proportion of S_{MAX}.

Figure 17.1 (a) Simple logistic growth; *(b)* Logistic growth with a minimum population threshold; *(c)* Logistic growth with depensation; *(d)* Logistic growth with critical depensation

17.1.1 The status and role of logistic growth models

The logistic growth model is a stylised representation of the population dynamics of renewable resources. The model is most suited to non-migratory species at particular locations. Among fish species, demersal or bottom-feeding populations of fish such as cod and haddock are well characterised by this model. The populations of pelagic or surface-feeding fish, such as mackerel, are less well explained by the logistic function, as these species exhibit significant migratory behaviour. Logistic growth does not only fit biological growth processes. Brown and McGuire (1967) argue that the logistic growth model can also be used to represent the behaviour of a freshwater stock in an underground reservoir.

However, a number of factors which influence actual growth patterns are ignored in this model, including the age structure of the resource (which is particularly important when considering stocks of long-lived species such as trees or whales) and random or chance influences. At best, therefore, it can only be a good approximation to the true population dynamics.

Judging the logistic model on whether it is the best available at representing any particular renewable resource would be inappropriate for our present purposes. One would not expect to find that biological or ecological modellers would use simple logistic growth functions. They will use more complex growth models designed specifically for particular species in particular contexts. But the needs of the environmental economist differ from those of ecological modellers. The former is willing to trade off some realism to gain simple, tractable models that are reasonably good approximations. It is for this reason that much economic analysis makes use of some version of the logistic growth function

Of more concern, perhaps, is the issue of whether it is appropriate to describe a biological growth process by any purely deterministic equation such those given in Box 17.1. Ecological models typically specify growth as being stochastic, and linked in complex ways to various other processes taking place in more broadly defined ecosystems and subsystems. We shall briefly explore these matters later in the chapter.

17.2 Steady-state harvests

In this chapter, much of our attention will be devoted to steady-state harvests. Here we briefly explain the concept. Consider a period of time in which the amount of the stock being harvested (H) is equal to the amount of net natural growth of the resource (G). Suppose also that these magnitudes remain constant over a sequence of consecutive periods. We call this *steady-state* harvesting, and refer to the (constant) amount being harvested as a sustainable yield.

Defining $\dot{S}$ as the actual rate of change of the renewable resource stock, with $\dot{S} = G - H$, it follows that in steady-state harvesting $\dot{S} = 0$ and so the resource stock remains constant over time. What kinds of steady states are feasible? To answer this, look at Figure 17.2. There is one particular stock size (S_{MSY}) at which the quantity of net natural growth is at a maximum (G_{MSY}). If at a stock of S_{MSY} harvest is set at the constant rate H_{MSY}, we obtain a *maximum sustainable yield* (MSY) steady state. A resource management programme could be devised which takes this MSY in perpetuity. It is sometimes thought to be self-evident that a fishery, forest or other renewable resource should be managed so as to produce its maximum sustainable yield. We shall see later that economic theory does not, in general, support this proposition.

H_{MSY} is not the only possible steady-state harvest. Indeed, Figure 17.2 shows that any harvest level between zero and H_{MSY} is a feasible steady-state

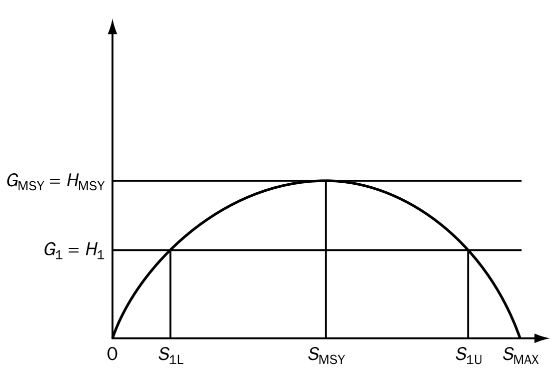


Figure 17.2 Steady-state harvests

harvest, and that any stock between zero and S_{MAX} can support steady-state harvesting. For example, H_1 is a feasible steady-state harvest if the stock size is maintained at either S_{1L} or S_{1U}. Which of these two stock sizes would be more appropriate for attaining a harvest level of H_1 is also a matter we shall investigate later.

Before moving on, it is important to understand that the concept of a steady state is a heuristic device: useful as a way of organising ideas and structuring analysis. But, like all heuristic devices, a steady state is a mental construct and using it uncritically can be inappropriate or misleading. Fisheries and other resource stocks are rarely, if ever, in steady states. Conditions are constantly changing, and the 'real world' is likely to be characterised by a more-or-less permanent state of disequilibrium. For some problems of renewable resource exploitation the analysis of transition processes is more important or insightful than information about steady states. We shall examine some of these 'dynamic' matters later in the chapter. Nevertheless, we will proceed on the assumption that looking at steady states *is* useful, and next investigate their properties under various institutional circumstances.

17.3 An open-access fishery

Our first model of renewable resource exploitation is an open-access fishery model. In conformity with the rest of this book, we study this using continuous-time notation. However, the equations which constitute the discrete-time equivalent of this continuous-time model (and others examined later in the chapter) are listed in full in Appendix 17.1. These will be used at various places in the chapter to give numerical illustrations of the arguments. The numerical values shown in our illustrative examples are computed using an Excel spreadsheet. Should the reader wish to verify that the values shown are correct, or to see how they would change under alternative assumptions, we have made the two spreadsheets used in our computations available to the reader in the *Additional Materials: Comparative statics.xls* and *Fishery dynamics.xls*. While we hope that some readers (or instructors) will find them useful, they are *not* necessary for an understanding of the contents of this chapter.

It is important to be clear about what an open-access fishery is taken to mean in the environmental economics literature. The open-access fishery model shares two of the characteristics of the standard perfect competition model. First, if the fishery is commercially exploited, it is assumed that this is done by a large number of independent fishing 'firms'. Therefore, each firm takes the market price of landed fish as given. Second, there are no impediments to entry into and exit from the fishery. But the free entry assumption has an additional implication in the open-access fishery, one which is *not* present in the standard perfect competition model.

In a conventional perfect competition model, each firm has enforceable property rights to its resources and to the fruits of its production and investment choices. However, in an open-access fishery, while owners have individual property rights to their fishing capital and to any fish that they have actually caught, they have no enforceable property rights to the *in situ* fishery resources, including the fish in the water.[2] On the contrary, any vessel is entitled or is able (or both) to fish wherever its owner likes. Moreover, if any boat operator chooses to leave some fish in the water in order that future stocks will grow, that owner has no enforceable rights to the fruits of that investment. It is as if a generalised 'what one finds one can keep' rule applies to fishery resources. We shall see in a moment what this state of affairs leads to. First, though, we need to set up the open-access fishery model algebraically.

The open-access model has two components:

1. a biological sub-model, describing the natural growth process of the fishery;
2. an economic sub-model, describing the economic behaviour of the fishing boat owners.

[2] This lack of *de facto* enforceability may derive from the fact that the fish are spatially mobile, or from the fact that boats are spatially mobile (or both).

Table 17.1 The open-access fishery model

		General specification	Specific forms assumed
BIOLOGICAL SUB-MODEL:			
Biological growth	(17.5, 17.3)	$\mathrm{d}S/\mathrm{d}t = G(S)$	$G(S) = g\left(1 - \frac{S}{S_{\mathrm{MAX}}}\right)S$
ECONOMIC SUB-MODEL:			
Fishery production function	(17.6, 17.7)	$H = H(E, S)$	$H = eES$
Net growth of fish stock	(17.8)	$\mathrm{d}S/\mathrm{d}t = G(S) - H$	$\mathrm{d}S/\mathrm{d}t = g\left(1 - \frac{S}{S_{\mathrm{MAX}}}\right)S - H$
Fishery costs	(17.9, 17.10)	$C = C(E)$	$C = wE$
Fishery revenue	(17.11)	$B = PH$, P constant	$B = PH$, P constant
Fishery profit	(17.12)	$\mathrm{NB} = B - C$	$\mathrm{NB} = B - C = PeES - wE$
Fishing effort dynamics: open-access entry rule	(17.13)	$\mathrm{d}E/\mathrm{d}t = \delta \cdot \mathrm{NB}$	$\mathrm{d}E/\mathrm{d}t = \delta(PeES - wE)$
BIOECONOMIC EQUILIBRIUM CONDITIONS:			
Biological equilibrium	(17.14)	$G = H$	$G = H$
Economic equilibrium	(17.15)	$E = E^*$ at which NB = 0	$E = E^*$ at which NB = 0

Note: Numbers in parentheses refer to the appropriate equation number in the text.

The model is laid out in full in Table 17.1. Subsequent parts of this section will take you through each of the elements described there. We shall be looking for two kinds of 'solutions' to the open-access model. The first is its equilibrium (or steady-state) solution. This consists of a set of circumstances in which the resource stock size is unchanging over time (a biological equilibrium) *and* the fishing fleet is constant with no net inflow or outflow of vessels (an economic equilibrium). Because the steady-state equilibrium is a joint biological–economic equilibrium, it is often referred to as *bioeconomic* equilibrium.

The second kind of solution we shall be looking for is the adjustment path towards the equilibrium, or from one equilibrium to another as conditions change. In other words, our interest also lies in the dynamics of renewable resource harvesting. This turns out to have important implications for whether a fish population may be driven to exhaustion, and indeed whether the resource itself could become extinct. The properties of such adjustment paths are examined in Section 17.4.

17.3.1 The model described

17.3.1.1 Biological sub-model

In the absence of harvesting and other human interference, the rate of change of the stock depends on the prevailing stock size

$$\mathrm{d}S/\mathrm{d}t = G(S) \tag{17.5}$$

For our worked numerical example, we assume that the particular form taken by this growth function is the simple logistic growth model given by equation 17.3.

17.3.1.2 Economic sub-model

17.3.1.2.1 The harvest function (or fishery production function)

Many factors determine the size of the harvest, H, in any given period. Our model considers two of these. First, the harvest will depend on the amount of resources devoted to fishing. In the case of marine fishing, these include the number of boats deployed and their efficiency, the number of days when fishing is undertaken and so on. For simplicity, assume that all the different dimensions of harvesting activity can be aggregated into one magnitude called *effort*, E.

Second, except for schooling fisheries, it is probable that the harvest will depend on the size of the resource stock.[3] Other things being equal, the larger the stock the greater the harvest for any given level of effort. Hence, abstracting from other determinants of harvest size, including random influences, we may take harvest to depend upon the effort applied and the stock size. That is

$$H = H(E, S) \tag{17.6}$$

This relationship can take a variety of particular forms. One very simple form appears to be a good approximation to actual relationships (see Schaefer, 1954 and Munro, 1981, 1982), and is given by

$$H = eES \tag{17.7}$$

where e is a constant number, often called the catch coefficient.[4] Dividing each side by E, we have

$$\frac{H}{E} = eS$$

which says that the quantity harvested per unit effort is equal to some multiple (e) of the stock size. We have already defined the fish-stock growth function with human predation as the biological growth function less the quantity harvested. That is,

$$\dot{S} = G(S) - H \tag{17.8}$$

17.3.1.2.2 The costs, benefits and profits of fishing

The total cost of harvesting, C, depends on the amount of effort being expended

$$C = C(E) \tag{17.9}$$

For simplicity, harvesting costs are taken to be a linear function of effort,

$$C = wE \tag{17.10}$$

where w is the cost per unit of harvesting effort, taken to be a constant.[5]

Let B denote the gross benefit from harvesting some quantity of fish. The gross benefit will depend on the quantity harvested, so we have

$$B = B(H)$$

In a commercial fishery, the appropriate measure of gross benefits is the total revenue that accrues to firms. Assuming that fish are sold in a competitive market, each firm takes the market price P as given and so the revenue obtained from a harvest H is given by[6]

$$B = PH \tag{17.11}$$

Fishing profit is given by

$$\mathrm{NB} = B - C \tag{17.12}$$

17.3.1.2.3 Entry into and exit from the fishery

To complete our description of the economic submodel, it is necessary to describe how fishing effort is determined under conditions of open access. A crucial role is played here by the level of economic profit prevailing in the fishery. Economic profit is

[3] See Discussion Question 4 for more on this matter and the notion of schooling and non-schooling fisheries.

[4] The use of a constant catch coefficient parameter is a simplification that may be unreasonable, and is often dropped in more richly specified models. Note also that equation 17.7 can be regarded as a special case of the more general form $H = eE^{\alpha}S^{\beta}$ in which the exponents need not be equal to unity. In empirical modelling exercises, this more general form may be more appropriate. Another form of the harvest equation sometimes used is the exponential model $H = S(1 - \exp(-eE))$.

[5] The equation $C = wE$ imposes the assumption that harvesting costs are linearly related to fishing effort. However, Clark *et al.* (1979) explain that this assumption will be incorrect if capital costs are sunk (unrecoverable); moreover, they show that even in a private-property fishery (to be discussed later in the chapter), it can then be privately optimal to have severely depleted fish stocks as the fishery approaches its steady-state equilibrium (although the steady-state equilibrium itself is not affected by whether or not costs are sunk). We return to this matter later.

[6] We could justify this assumption either by saying that the harvesting industry being examined is a small part of a larger overall market, or by arguing that the resource market is competitive, in which case each firm acts as if the market price is fixed (even though price will actually depend *ex post* on the realised total market supply).

the difference between the total revenue from the sale of harvested resources and the total cost incurred in resource harvesting. Given that there is no method of excluding incomers into the industry, nor is there any way in which existing firms can be prevented from changing their level of harvesting effort, effort applied will continue to increase as long as it is possible to earn positive economic profit.[7] Conversely, individuals or firms will leave the fishery if revenues are insufficient to cover the costs of fishing. A simple way of representing this algebraically is by means of the equation

$$\mathrm{d}E/\mathrm{d}t = \delta \cdot \mathrm{NB} \tag{17.13}$$

where δ is a positive parameter indicating the responsiveness of industry size to industry profitability. When economic profit (NB) is positive, firms will enter the industry; and when it is negative they will leave. The magnitude of that response, for any given level of profit or loss, will be determined by δ. Although the true nature of the relationship is unlikely to be of the simple, linear form in equation 17.13, this suffices to capture what is essential.

17.3.1.2.4 Bioeconomic equilibrium

We close our model with two equilibrium conditions that must be satisfied jointly. Biological equilibrium occurs where the resource stock is constant through time (that is, it is in a steady state). This requires that the amount being harvested equals the amount of net natural growth:

$$G = H \tag{17.14}$$

Economic equilibrium requires that the amount of fishing effort be constant through time. Such an equilibrium is only possible in open-access fisheries when rents have been driven to zero, so that there is no longer an incentive for entry into or exit from the industry, nor for the fishing effort on the part of existing fishermen to change. We express this by the equation

$$\mathrm{NB} = B - C = 0 \tag{17.15}$$

which implies (under our assumptions) that $PH = wE$. Notice that when this condition is satisfied, $\mathrm{d}E/\mathrm{d}t = 0$ and so effort is constant at its equilibrium (or steady-state) level $E = E^*$.

17.3.2 Open-access steady-state equilibrium

We can envisage an open-access fishery steady-state equilibrium by means of what is known as the fishery's yield–effort relationship. To obtain this, first note that in a biological equilibrium $H = G$. Then, by substituting the assumed functions for H and G from equations 17.7 and 17.3 respectively we obtain:

$$gS\left(1 - \frac{S}{S_{\mathrm{MAX}}}\right) = eES \tag{17.16}$$

which can be rearranged to give

$$S = S_{\mathrm{MAX}}\left(1 - \frac{e}{g}E\right) \tag{17.17}$$

Equation 17.17 is one equation in two endogenous variables, E and S (with parameters g, e and S_{MAX}). It implies a unique equilibrium stock at each level of effort.[8] Next substitute equation 17.17 into equation 17.7 ($H = eES$), giving

$$H = eES_{\mathrm{MAX}}\left(1 - \frac{e}{g}E\right) \tag{17.18}$$

In an open-access economic equilibrium, profit is zero, so

$$PH = wE \tag{17.19}$$

Equations 17.18 and 17.19 constitute two equations in two unknowns (H and E); these can be solved for the equilibrium values of the two unknowns as functions of the parameters alone. The steady-state stock solution can then be obtained by substituting the expressions for H and E into equation 17.7. We list these steady-state solutions in Box 17.2, together with the numerical values of E, H and S under the

[7] The terms rent, economic rent, royalties and net price are used as alternatives to economic profit. They all refer to a surplus of revenue over total costs, where costs include a proper allowance for the opportunity of capital tied up in the fishing fleet.

[8] This uniqueness follows from the assumption that $G(S)$ is a simple logistic function; it may not be true for other biological growth models.

Box 17.2 Analytical expressions for the open-access steady-state equilibrium and numerical solutions under baseline parameter assumption

The analytical expressions for E^*, S^* and H^* (where an asterisk denotes the equilibrium value of the variable in question) as functions of the model parameters alone are:

$$E^* = \frac{g}{e}\left(1 - \frac{w}{PeS_{\text{MAX}}}\right) \tag{17.20}$$

$$S^* = \frac{w}{Pe} \tag{17.21}$$

$$H^* = \frac{gw}{Pe}\left(1 - \frac{w}{PeS_{\text{MAX}}}\right) \tag{17.22}$$

Derivations of expressions 17.20–17.22 are given in full in Appendix 17.2. Throughout this chapter, we shall be illustrating our arguments with results drawn from fishery models using the parameter values shown in Table 17.2. At various points in the chapter we shall refer to these as the 'baseline' parameter values. It can be easily verified that for this set of parameter values, the steady-state solution is given by $E^* = 8.0$, $S^* = 0.2$ and $H^* = 0.024$.

Table 17.2 Parameter value assumptions for the illustrative numerical example

Parameter	Assumed numerical value
g	0.15
S_{MAX}	1
e	0.015
δ	0.4
P	200
w	0.6

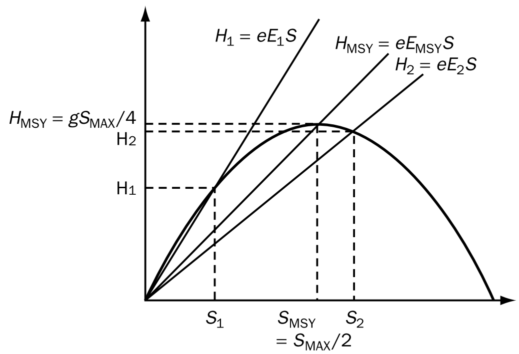


Figure 17.3 Steady-state equilibrium fish harvests and stocks at various effort levels

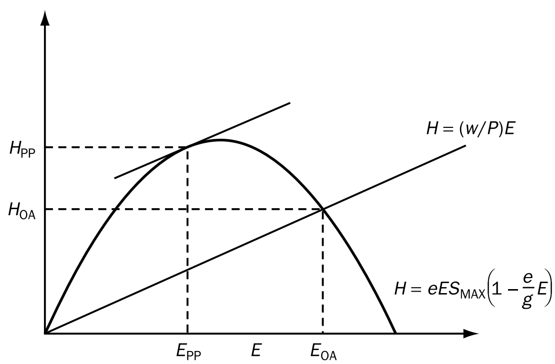


Figure 17.4 Steady-state equilibrium yield–effort relationship

particular set of assumptions about numerical values of the parameters given in Table 17.2.

This solution method can also be represented graphically, as shown in Figures 17.3 and 17.4. Figure 17.3 shows equilibrium relationships in stock–harvest space. The inverted U-shape curve is the logistic growth function for the resource. Three rays emanating from the origin portray the harvest–stock relationships (from the function $H = eES$) for three different levels of effort. If effort were at the constant level E_1, then the unique intersection of the harvest–stock relationship and biological growth function determines a steady-state harvest level H_1 at stock S_1. The lower effort level E_2 determines a second steady-state equilibrium (the pair $\{H_2, S_2\}$). We leave the reader to deduce why the label E_{MSY} has been attached to the third harvest–stock relationship. The various points of intersection satisfy equation 17.17, being equilibrium values of S for particular levels of E. Clearly there is an infinite quantity of possible equilibria, depending on what constant level of fishing effort is being applied.

The points of intersection in Figure 17.3 not only satisfy equation 17.17 but they also satisfy equation 17.18. Put another way, the equilibrium $\{E, S\}$ combinations also map into equilibrium $\{E, H\}$ combinations. The result of this mapping from $\{E, S\}$ space into $\{E, H\}$ space is shown in Figure 17.4. The inverted U-shape curve here portrays the steady-state harvests that correspond to each possible effort level. It describes what is often called the fishery's yield–effort relationship. Mathematically, it is a plot of equation 17.18.

The particular point on this yield–effort curve that corresponds to an open-access equilibrium will be the one that generates zero economic profit. How do we find this? The zero economic profit equilibrium condition $PH = wE$ can be written as $H = (w/P)E$. For given values of P and w, this plots as a ray from the origin with slope w/P in Figure 17.4. The intersection of this ray with the yield–effort curve locates the unique open-access equilibrium outcome.

Alternatively, multiplying both functions in Figure 17.4 by the market price of fish, P, we find that the intersection point corresponds to $PH = wE$. This is, of course, the zero profit condition, and confirms that $\{E_{OA}, H_{OA}\}$ is the open-access effort–yield equilibrium.

17.4 The dynamics of renewable resource harvesting

Our discussion so far has been exclusively on steady states: equilibrium outcomes which, once achieved, would remain unchanged provided that relevant economic or biological conditions remain constant. However, we may also be interested in the *dynamics* of resource harvesting. This would consider questions such as how a system would get to a steady state if it were not already in one, or whether getting to a steady state is even possible. In other words, dynamics is about transition or adjustment paths. Dynamic analysis might also give us information about how a fishery would respond, through time, to various kinds of shocks and disturbances.

A complete description of fishery dynamics is beyond the scope of this book. But some important insights can be obtained relatively simply. In this section, we undertake some dynamics analysis for the open-access model of Section 17.3. Suppose a mature fishery exists that has not previously been commercially exploited. The stock size is, therefore, at its carrying capacity. The fishery now becomes available for unregulated, open-access commercial exploitation. If the market price of fish, P, is reasonably high and fishing cost (per unit of effort), w, is reasonably low, the fact that stocks are high (and so easy to catch) implies that the fishery will be, at least initially, profitable for those who enter it. Have in mind equations 17.7, 17.10, 17.11 and 17.12 when thinking this through.

If a typical fishing boat can make positive economic profit then further entry will take place. How quickly new capacity is built up depends on the magnitude of the parameter δ in equation 17.13. In this early phase, effort is rising over time as new boats are attracted in, and stocks are falling. Stocks fall because harvesting is taking place while new recruitment to stocks is low: the logistic growth function has the property that biological growth is near zero when the stock is near its maximum carrying capacity (equation 17.3). This process of increasing E and decreasing S will persist for some time, but it cannot last indefinitely. As stocks become lower, fish become harder to catch and so the cost per fish caught rises. Profits are squeezed from two directions: harvesting cost per fish rises, and fewer fish are caught.

Eventually, this profit squeeze will mean that a typical boat makes a loss rather than a profit, and so the process we have just described goes into reverse, with stocks rising and effort falling. In fact, for the model we are examining, the processes we are describing here are a little more subtle than this. The changes do not occur as discrete switches but instead are continuous and gradual. We also find that stocks and effort (and also harvest levels) have oscillatory cycles with the stock cycles slightly leading the effort cycles. In some circumstances, these oscillations dampen down as time passes, and the system eventually settles to a steady-state outcome such as that described in the previous section. We illustrate such a transition process in Figure 17.5, where parameter values are given by those shown in Table 17.2. Note that the oscillations shown in this diagram are particularly acute and have massive amplitude; for other combinations of parameter values, the cycles may be far less pronounced. In this case, you should be able to discern that, given enough time, the levels of S and E will settle down to the steady-state values $S^* = 0.2$ and $E^* = 8.0$.

The oscillations exhibited by the dynamic adjustment path in Figure 17.5 – with the variables repeatedly over- and under-shooting equilibrium

Box 17.3 A story of two fish populations

One species of fish – the Peruvian anchovy – and one group of commercial fish – New England groundfish – provide us with case studies of the mismanagement and economic inefficiency which often characterise the world's commercial fisheries. In this box, we summarise reviews of the recent historical experiences of these two fisheries; the reviews are to be found in WR (1994), chapter 10.

Peruvian anchovy are to be found in the Humboldt Upswelling off the west coast of South America. Upswellings of cold, nutrient-rich water create conditions for rich commercial fish catches. During the 1960s and 1970s, this fishery provided nearly 20% of the world's fish landings. Until 1950, Peruvian anchovy were harvested on a small scale, predominantly for local human consumption, but in the following two decades the fishery increased in scale dramatically as the market for fishmeal grew. The maximum sustainable yield (MSY) was estimated as 9.5 million tonnes per year, but that figure was being exceeded by 1970, with harvests beyond 12 million tonnes. In 1972, the catch plummeted. This fall was partially accounted for by a cyclical natural phenomenon, the arrival of the El Niño current. However, it is now recognised that the primary cause of the fishery collapse (with the catch down to just over 1 million tonnes in the 1980s) was the conjunction of overharvesting with the natural change associated with El Niño. Harvesting at rates above the MSY can lead to dramatic stock collapses that can persist for decades, and may be irreversible (although, in this case, anchovy populations do now show signs of recovery).

The seas off the New England coast have been among the most productive, the most intensively studied and the most heavily overfished in the world since 1960. The most important species in commercial terms have been floor-living species including Atlantic cod, haddock, redfish, hake, pollock and flounder. Populations of each are now near record low levels. Although overfishing is not the only contributory factor, it has almost certainly been the principal cause of stock collapses. The New England fisheries are not unusual in this; what is most interesting about this case is the way in which regulatory schemes have failed to achieve their stated goals. In effect, self-regulation has been practised in these fisheries and, not surprisingly perhaps, regulations have turned out to avoid burdening current harvesters. This is a classic example of what is sometimes called 'institutional capture': institutions which were intended to regulate the behaviour of firms within an industry, to conform with some yardstick of 'the common good', have in effect been taken over by those who were intended to be regulated, who then design administrative arrangements in their own interest. The regulations have, in the final analysis, been abysmal failures when measured against the criterion of reducing the effective quantity of fishing effort applied to the New England ground fisheries.

Long-term solutions to overfishing will require strict quantity controls over fishing effort, either by direct controls over the effort or techniques of individual boats, or through systems of transferable, marketable quotas. We investigated some of these instruments in Chapter 7 and do so further later in this chapter.

a bargain could be made and not reneged upon are very unlikely to exist. Each potential bargainer has an incentive to free-ride once a bargain has been struck, by increasing his or her harvest while others reduce theirs. Moreover, even if all existing parties were to agree among themselves, the open-access conditions imply that others could enter the market as soon as rents became positive. Open-access resources thus have one of the properties of a public good – non-excludability – and this alone is sufficient to make it likely that markets will fail to reach efficient outcomes.

17.6 The private-property fishery

In an open-access fishery, firms exploit available stocks as long as positive profit is available. While this condition persists, each fishing vessel has an incentive to maximise its catch. But there is a dilemma here, both for the individual fishermen and for society as a whole. From the perspective of the fishermen, the fishery is perceived as being overfished. Despite each boat owner pursuing maximum profit, the collective efforts of all drive profits down to zero. From a social perspective, the fishery will be economically 'overfished' and the stock level may (but will not necessarily) be driven down to biologically dangerous levels.

What is the underlying cause of this state of affairs? Although reducing the total catch today may be in the collective interest of all (by allowing fish stocks to recover and grow), it is not rational for any fisherman to individually restrict fishing effort. There can be no guarantee that he or she will receive any of the rewards that this may generate in terms of higher catches later. Indeed, there may not be any stock available in the future. In such circumstances, each firm will exploit the fishery today to its maximum potential, subject only to the constraint that its revenues must at least cover its costs.

We shall now discuss a particular set of institutional arrangements that could overcome some of these dilemmas. These arrangements could be described as the private-property fishery. This kind of fishery – and several variants of it – have been explored in depth in the fishery economics literature. However, discussions of the private-property fishery rarely make explicit the institutional assumptions that lie behind it. It is important to do that, however, and so we shall now describe what the private-property fishery is usually taken to mean (and the sense in which we shall be using the term).

The private-property fishery has the following three characteristics:

1. There is a large number of fishing firms, each behaving as a price-taker and so regarding price as being equal to marginal revenue. It is for this reason that the industry is often described as being competitive.
2. Each firm is profit- (or wealth-) maximising.
3. There is a particular structure of well-defined and enforceable property rights to the fishery, such that owners can control access to the fishery and appropriate any rents that it is capable of delivering.

What exactly is this particular structure of property rights? Within the literature there are several (sometimes implicit) answers to this question. We shall outline two of them. One view regards 'the fishery' as an aggregate of a large number of smaller individual fisheries. Each of these sub-fisheries is privately owned by one firm that has property rights to the fish which are there currently and at all points in time in the future.[9] All harvested fish, however, sell in one aggregate market at a single market price. A second view regards the fishery as being managed by a single entity which controls access to the fishery and coordinates the activity of individual operators to maximise total fishery profits (or wealth). Nevertheless, harvesting and pricing behaviour are competitive rather than monopolistic.

Neither of these accounts is satisfactory as a statement of what actually does exist, nor what might realistically exist. The first faces problems in deciding how to specify ownership rights to migratory fish. Moreover, it could only be descriptively accurate if the fishery in question is a huge, highly spatially aggregated, fishery. The researcher does not want to study at this level of aggregation. The second concept – the coordinated fishery – seems problematic in that we rarely, if ever, find examples of such *internally* coordinated fisheries (except in the case of fish farming and the like). And even if one were to find examples, it is difficult to imagine that they would operate as competitive fisheries rather than as monopolies or cartels.[10]

But to label one or both of these views as descriptively unrealistic is to miss the point somewhat. They should be thought of in 'as if' terms. That is, we want a specification such that the industry behaves *as if* each firm has its own 'patch' of fishery that others are not permitted to exploit or *as if* it were coordinated in the way mentioned above. Given either of these *as if* assumptions, the researcher can then reasonably assume that owners undertake economically rational management decisions, and are in a position to make investment decisions confident

[9] The owners of any fishing firm may, of course, lease or sell their property rights to another set of individuals.

[10] In fact, two other variants of the private-property fishery sometimes discussed in the literature are actually these: the monopoly fishery and the cartel fishery. However, given the fact that they are so uncommon in practice, we do not deal with those models in this text, except for a brief reference to monopoly fishery in Appendix 17.3.

in the belief that the returns on any investment made can be individually appropriated. This is what distinguishes a private-property fishery fundamentally from an open-access fishery.

An important benefit from thinking about property rights carefully in this way is the help it gives in developing public policy towards fishery regulation and management. If we are confident that a particular property rights structure would bring about socially efficient (or otherwise desirable) outcomes, then policy instruments can be designed to mimic that structure. We will argue below that an individual transferable quota (ITQ) fishing permit system can be thought of in this way.

17.6.1 The static profit-maximising private-property fishery model

As we explained in the Introduction, our analysis of the private-property fishery proceeds in two steps. The first, covered in this section, develops a simple static model of a private-property fishery in which the passage of time is not explicitly dealt with. In effect, the analysis supposes that biological and economic conditions remain constant over some span of time. It then investigates what aggregate level of effort, stock and harvest would result if each individual owner (with enforceable property rights) managed affairs so as to maximise profits in any arbitrarily chosen period of time. This way of dealing with time – in effect, abstracting from it, and looking at decisions in only one time period (but which are replicated over successive periods) – leads to its description as a *static* fishery model. We shall demonstrate later that this analytical approach only generates wealth-maximising outcomes if fishermen do not discount future cash flows. More specifically, the static private-property fishery turns out to be a special case of a multi-period fishery model: the special case in which owners use a zero discount rate.

The biological and economic equations of the static private-property fishery model are identical to those of the open-access fishery in all respects but one: the open-access entry rule ($dE/dt = \delta \cdot NB$), which in turn implies a zero-profit economic equilibrium, no longer applies. Instead, owners choose effort to maximise economic profit from the fishery.

Table 17.3 Steady-state solutions under baseline parameter value assumptions

	Open access	Static private property
Stock	0.200	0.600
Effort	8.000	4.000
Harvest	0.024	0.036

This can be visualised with the help of Figure 17.4. As we did earlier, multiply both functions by the market price of fish. The inverted U-shape yield–effort equation then becomes a revenue–effort equation. And the ray emerging from the origin now becomes $PH = wE$, with the right-hand side thereby denoting fishing costs. Profit is maximised at the effort level which maximises the surplus of revenue over costs. Diagrammatically, this occurs where the slopes of the total cost and total revenue curves are equal. This in indicated in Figure 17.4 by the tangent to the yield–effort function at E_{PP} being parallel to the slope of the $H = (w/P)E$ line.

An algebraic derivation of the steady-state solution to this problem – showing stock, effort and harvest as functions of the parameters – is given in Box 17.4. It is easy to verify from the solution equations given there that the steady-state values of E, S and H are given by $E^*_{PP} = 4.0$, $S^*_{PP} = 0.60$ and $H^*_{PP} = 0.0360$. To facilitate comparison, the numerical values of the steady-state equilibrium stock, harvest and effort under our baseline parameter assumptions, for both open-access and static private-property fishery, are reproduced in Table 17.3. Under the assumptions we have made about functional forms, the static private-property equilibrium will always lead to a higher resource stock level and a lower effort level than that which prevails under open access. This is confirmed for our particular parameter assumptions, with the private-property stock being three times higher and effort only half as large as in open access.

The steady-state harvest may be higher, lower or identical. This is evident from inspection of Figure 17.4. For the particular set of parameter values used in our illustrative example, private-property harvest is *larger* than open-access harvest, as shown in the diagram. But it will not always be true that private-property harvests exceed those under open access. For example, if P were 80 (rather than 200) and all

Box 17.4 Derivation of the static private-property steady-state equilibrium for our assumed functional forms

The derivation initially follows exactly that given in Section 17.3.2, with equations 17.16 to 17.18 remaining valid here. However, the zero profit condition (equation 17.19) is no longer valid, being replaced by the profit-maximisation condition:

$$\text{Maximise NB} = PH - wE \tag{17.23}$$

Remembering that $H = eES$, and treating E as the instrument variable, this yields the necessary first-order condition,

$$\partial(PeES)/\partial E = \partial(wE)/\partial E \tag{17.24}$$

Substituting equation 17.17 into 17.24 we have

$$\partial\left(PeES_{\text{MAX}}\left(1 - \frac{e}{g}E\right)\right)/\partial E = \partial(wE)/\partial E$$

from which we obtain after differentiation

$$PeS_{\text{MAX}} - 2PES_{\text{MAX}}\left(\frac{e^2}{g}\right) = w \tag{17.25}$$

That is, the marginal revenue of effort is equal to the marginal cost of effort. This can be solved for E^*_{PP} (the subscript denoting 'private property') to give

$$E^*_{\text{PP}} = \frac{1}{2}\frac{g}{e}\left(1 - \frac{w}{PeS_{\text{MAX}}}\right) \tag{17.26}$$

Substitution of E^*_{PP} into 17.17 gives

$$S^*_{\text{PP}} = \frac{1}{2}\frac{PeS_{\text{MAX}} + w}{Pe} \tag{17.27}$$

and then using $H = eES$ we obtain[11]

$$H^*_{\text{PP}} = \frac{1}{4}g\left(S_{MAX} - \frac{w^2}{P^2e^2S_{\text{MAX}}}\right) \tag{17.28}$$

[11] In the Excel spreadsheets, an alternative (but exactly equivalent) version of this expression has been used to generate the Excel formulas, namely

$$H^*_{\text{PP}} = \frac{1}{4}\frac{g(PeS_{\text{MAX}} - w)(PeS_{\text{MAX}} + w)}{P^2e^2S_{\text{MAX}}}$$

other parameter values were those specified in the baseline set (listed in Table 17.2) then an open-access fishery would produce $H = 0.0375$, the maximum sustainable yield of the fishery! In contrast, a private-property fishery would in those circumstances yield only $H = 0.0281$.

The source of this indeterminacy follows from the inverted U shape of the yield–effort relationship. Although stocks will be higher under private property than open access, the quadratic form of the stock–harvest relationship implies that harvests will not necessarily be higher with higher stocks.

17.6.2 Comparative statics

For convenience, we list in Table 17.4 the expressions obtained in earlier sections for the steady-state equilibria of E, H and S. We can use these expressions to make qualitative predictions about the effects of changing a particular parameter on the equilibrium levels of S^*, E^* and H^*. Doing this is known as comparative statics. For example, how will S^* change as w rises or as P increases? Inspection of the formula in the top left cell shows that open-access S^* will increase if w increases and will decrease if P increases. This is also true in the case of the static private-property steady state, as can be seen by inspection of the top right-hand-side cell.

Where the sign of a relationship cannot easily be found by inspection, we may be able to obtain it from the appropriate partial derivative. For example, although it is easy in this case to confirm by inspection that a rise in w will increase open-access S^*, this inference is corroborated by the fact that the partial derivative of S^* with respect to w is $1/Pe$. As P and e are both positive numbers, the partial derivative

Table 17.4 Steady-state open-access and static private-property equilibria compared

	Open access	Static private property
Stock	$S^* = \frac{w}{Pe}$	$S^*_{\text{PP}} = \frac{1}{2}\frac{PeS_{\text{MAX}} + w}{Pe}$
Effort	$E^* = \frac{g}{e}\left(1 - \frac{w}{PeS_{\text{MAX}}}\right)$	$E^*_{\text{PP}} = \frac{1}{2}\frac{g}{e}\left(1 - \frac{w}{PeS_{\text{MAX}}}\right)$
Harvest	$H^* = \frac{gw}{Pe}\left(1 - \frac{w}{PeS_{\text{MAX}}}\right)$	$H^*_{\text{PP}} = \frac{1}{4}g\left(S_{\text{MAX}} - \frac{w^2}{P^2e^2S_{\text{MAX}}}\right)$

Table 17.5 Comparative static results

	P	w	E	g	δ
Open access					
S^*	−	+	−	0	0
E^*	+	−	?	+	0
H^*	?	?	?	+	0
Static private property					
S^*_{PP}	−	+	−	0	0
E^*_{PP}	+	−	?	+	0
H^*_{PP}	+	−	+	+	0

$1/Pe$ is also positive. Sometimes, of course, the direction of an effect cannot be signed unambiguously; this should usually be evident by inspection of the partial derivative.

Table 17.5 lists the signs of these effects from the appropriate partial derivatives. A plus sign means that the derivative is positive, a minus sign means that the derivative is negative, 0 means that the derivative is zero, and ? means that no sign can be unambiguously assigned to the derivative (and so we cannot say what the direction of the effect will be without knowing the actual values of the parameters that enter the partial derivative in question). Note that variations in δ have no effect on any steady-state outcome (although they do affect how fast, if indeed at all, such an outcome may be achieved).

17.6.3 The present-value-maximising fishery model

The present-value-maximising fishery model generalises the model of the static private-property fishery. In doing so it formulates a model that has a more sound theoretical basis and generates a richer set of results. The essence of this model is that a rational private-property fishery will organise its harvesting activity so as to maximise the value of the fishery. We shall refer to this value as the present value (PV) of the fishery. In this section, we outline how a model of a present-value-maximising fishery can be set up, state the main results, and provide interpretations of them. Full derivations have been placed in Appendix 17.3. The individual components of our model are very similar to those of the static private fishery model. However, we now bring time explicitly into the analysis by using an intertemporal optimisation framework. Initially we shall develop results using general functional forms. Later in this section, solutions are obtained for the specific functions and baseline parameter values assumed earlier in this chapter.

As in previous sections of the chapter, we assume that the market price of fish is a constant, exogenously given, number. Moreover, as before, the market is taken to be competitive. However, Appendix 17.3 will also go through the more general case in which the market price of fish varies with the size of total industry catch, and will briefly examine a monopolistic fishery.

It will be convenient to regard harvest levels as the instrument (control) variable. To facilitate this, we specify fishing costs as a function of the quantity harvested and the size of the fish stock. Moreover, it is assumed that costs depend positively on the amount harvested and negatively on the size of the stock.[12,13] That is,

$$C_t = C(H_t, S_t) \quad C_H > 0, C_S < 0$$

The initial population of fish is S_0, the natural growth of which is determined by the function $G(S)$. The fishery owners select a harvest rate for each period over the relevant time horizon (here taken to be

[12] The reader may be confused about our formulation of the harvest cost function. In an earlier section, we wrote $C = C(E)$, equation 17.9. But note that we have also assumed that $H = H(E, S)$, equation 17.6. If 17.6 is written as E in terms of H and S, and that expression is then substituted into 17.9, we obtain $C = C(H, S)$. It is largely a matter of convenience whether we express costs in terms of effort or in terms of harvest and stock. In our discussion of open-access equilibrium, we chose to regard fishing effort as a variable of interest and did not make that substitution. In this section, our interest lies more in the variable H and so it is convenient to make the substitution. But the results of either approach can be found from the other.

[13] There is another issue here that we should mention. The costs of fishing should include a proper allowance for all the opportunity costs involved. For land-based resources, the land itself is likely to have alternative uses, and so its use in any one activity will have a land opportunity cost. For fisheries, however, there is rarely an alternative *commercial* use of the oceans, and so this kind of opportunity cost is not relevant. However, from a social point of view there may be important alternative uses of the oceans (for example, as conserved sources of biodiversity). Hence a difference can exist between costs as seen from a social and a private point of view.

CHAPTER 18 Forest resources

This was the most unkindest cut of all. **William Shakespeare, *Julius Caesar* III.ii (188)**

Learning objectives

Having completed this chapter, the reader should be able to

- understand the various functions provided by forest and other woodland resources
- describe recent historical and current trends in forestation and deforestation
- recognise that plantation forests are renewable resources but natural – particularly primary – forests are perhaps best thought of as non-renewable resources in which development entails irreversible consequences
- explain the key differences between plantation forests and other categories of renewable resource
- understand the concepts of *site value* of land and *land rent*
- use a numerically parameterised timber growth model, in conjunction with a spreadsheet package, to calculate appropriate physical measures of timber growth and yield; and given various economic parameters, to calculate appropriate measures of cost and revenue
- obtain and interpret an expression for the present value of a single-rotation stand of timber
- using the expression for present value of a single rotation, obtain the first-order condition for maximisation of present value, and recognise that this can be interpreted as a modified Hotelling rule
- undertake comparative static analysis to show how the optimal stand age will vary with changes in relevant economic parameters such as timber prices, harvesting costs and interest (or discount) rate
- specify an expression for the present value of an infinite sequence of identical forest rotations, obtain an analytic first-order expression for maximisation of that present value with respect to the rotation age, and carry out comparative static analysis to ascertain how this varies with changes in economic parameters

Introduction

This chapter is concerned with forests and other wooded land. In the first section, the present state of global forest resources is briefly described. We then consider several salient characteristics of forest resources. This draws your attention to some of the particular characteristics of forest resources that differentiate its study from that of fisheries, the principal focus of Chapter 17.

Roughly speaking, forest resources can be divided into three categories: natural forests, semi-natural (disturbed or partly developed) forests, and plantation forests.[1] As we shall see, these are very different in terms of the services that they provide. Our attention

[1] Except where it is necessary to distinguish between the two, we shall use the word ‘forest’ to refer to both forested land and (the less densely stocked) woodland.

in this chapter is largely given to the two 'extreme' cases of natural and plantation forests. Semi-natural forests are a hybrid form, and will share characteristics of the two other cases depending on the extent to which they have been disturbed or managed.

Section 18.3 considers plantation forests. The analysis of plantation forestry is well developed, and it has been the object of an important sub-discipline within economics for well over a century. A plantation forest is a renewable resource, and the techniques we outlined in the previous chapter can be applied to the analysis of it. However, the long span of time taken by trees to reach maturity means that the age at which a stand of trees is cut – the rotation period – is of central (but not exclusive) importance, and is the dimension on which our analysis focuses.

Initially, our emphasis is on the timber yielded by managed forest land. However, all forests – even pure plantation forests – provide a wide variety of other, non-timber, benefits. Forestry policy in many countries is giving increasing weight to non-timber values in forest management choices. Section 18.4 investigates the question of how forests should be managed when they are used, or generate values, in multiple ways.

Not surprisingly, natural (undisturbed) forests are biologically the most diverse and perform a much broader range of ecological, amenity and recreational and other economic services than do plantation forests. We devote the latter part of this chapter, therefore, to looking at deforestation of natural woodland. Particular attention is paid to tropical deforestation, an issue that has become the subject of extensive study within environmental economics in recent decades.

Plantation forests are renewable resources. Does the same hold for undisturbed natural forests? The fact that trees can grow and reproduce suggests that this is so. But a little reflection suggests that matters are not quite so straightforward. If we think about natural forests as ecosystems providing multiple services, and recognise that the ways in which such forests are typically 'developed' or disturbed generate irreversible changes, it becomes clear that they share some of the characteristics of non-renewable resources. Hence it may be preferable, under present conditions at least, to regard natural forests as existing in more-or-less fixed quantities and once 'mined' as being irreversibly lost as natural forests. Although trees may subsequently grow in areas once occupied by natural forest, the gestalt of what constitutes a natural forest cannot be replaced (except over extremely long spans of time). We examine these issues in Section 18.6.

Sections 18.2 and 18.3 make extensive use of economic models of forestry. Several illustrative examples are used in those parts of the chapter. To allow the reader to replicate our results, and to explore the properties of these models a little further, all calculations in this chapter – and all associated diagrams – are performed using Excel workbooks. *Chapter18.xls* contains the calculations and charts used in the main body of the chapter. *Palc18.xls* contains computations used in Appendix 18.2 and some of the problems at the end of this chapter. Details of other associated Excel files are given below at appropriate places. These files can be found on the *Additional Materials* web page.

18.1 The current state of world forest resources

The latest available comprehensive assessment of the state of the world's forest resources is contained in the Global Forest Resources Assessment 2000 (known as 'FRA 2000'), undertaken by the Food and Agriculture Organisation of the United Nations (FAO, 2001). The complete report is available online by searching from the forestry section of the FAO web site at www.fao.org/forestry/index.jsp. Material in this section is largely drawn from that report.

The information found in that report can be usefully summarised by means of two tables. Table 18.1 shows forested and wooded area disaggregated by continents and sub-continental regions; Table 18.2, at a higher but different level of aggregation, shows changes in forested area by forest type for tropical and non-tropical areas. It is evident by inspection of these tables that forested area is in a state of flux, with areas being both won and lost to forest and other woodland. The overall effect, however, is one of falling total forest area, with 9.4 million hectares being lost in net terms in the decade to 2000.

Table 18.1 Global forest resource

Country/Area	Total forest area 1990	Total forest area 2000	2000 forest area as percentage of land area	Forest cover annual rate of change 1990–2000	Forest land area			Other wooded land		Total plantation area, 2000	Annual planting rate, 2000
					Closed	Open	Plantation	Shrubs/ Trees	Forest fallow		
	000 ha	000 ha	%	%	000 ha	000 ha	000 ha	000 ha	000 ha	000 ha	000 ha
TOTAL WORLD	**3 963 429**	**3 869 455**	**29.7**	**–0.2**	**3 334 790**	**444 585**	**112 844**	**1 302 768**	**126 823**	**186 733**	**4458**
Total Africa	**702 502**	**649 866**	**21.8**	**–0.7**	**352 700**	**288 906**	**4 571**	**377 996**	**52 083**	**8 036**	**194**
North Africa	5 930	6 262	1.0	0.6	2 481	2 236	1 207	4 560	0	1 693	60
West Sahelian Africa	46 818	43 570	8.2	–0.7	12 191	34 587	80	45 750	4 284	543	27
East Sahelian Africa	109 271	96 612	19.3	–1.2	27 984	81 362	183	110 897	0	1 156	39
West Africa	51 803	41 594	20.3	–2.0	16 755	33 458	385	26 609	3 963	1 252	33
Central Africa	235 861	227 637	57.1	–0.3	196 648	36 802	30	29 917	10 319	301	6
Tropical Southern Africa	230 306	212 884	38.2	–0.8	79 890	97 707	612	93 931	33 517	989	10
Temperate Southern Africa	9 475	9 453	7.5	0.0	5 084	2 754	1 739	64 812	0	1 729	14
Insular Africa	13 038	11 854	20.1	–0.9	11 667	0	335	1 520	0	374	6
Total Asia	**551 448**	**547 793**	**17.8**	**–0.1**	**416 207**	**58 321**	**53 791**	**122 308**	**20 031**	**115 847**	**3500**
Temperate and Middle East Asia	227 356	248 569	12.0	0.9	172 730	9 645	53 120	86 819	7 234	61 222	1242
South Asia	88 889	87 310	15.3	–0.2	58 219	28 951	405	13 303	120	34 652	1571
Continental South East Asia	87 761	80 896	42.5	–0.8	59 609	19 450	173	18 882	12 018	7 596	351
Insular South East Asia	147 442	131 018	53.3	–1.1	125 649	275	93	3 304	659	12 376	336
Total Oceania	**201 271**	**197 623**	**23.3**	**–0.2**	**196 345**	**1 145**	**2 691**	**423 519**	**451**	**2 848**	**15**
Total Europe	**1 030 475**	**1 039 251**	**46.0**	**0.1**	**1 002 979**		**32 036**	**29 484**		**32 015**	**5**
Total North and Central America	**569 115**	**563 417**	**26.6**	**–0.1**	**508 298**	**27 367**	**16 681**	**308 585**	**26 055**	**17 533**	**234**
Temperate North America	466 684	470 564	25.6	0.1	453 162	0	16 238	250 477	0	16 238	121
Central America and Mexico	84 765	75 680	29.5	–1.1	52 225	27 243	421	58 083	25 295	1 241	112
Caribbean Subregion	17 666	17 173	71.7	–0.3	17 011	124	22	25	760	67	1
Total South America	**908 618**	**871 505**	**50.1**	**–0.4**	**858 261**	**68 846**	**3 074**	**40 876**	**28 203**	**10 455**	**509**
Non-tropical South America	54 029	51 476	14.1	–0.5	21 965	28 733	2 323	14 670	0	3 565	261
Tropical South America	854 589	820 029	59.7	–0.4	822 196	40 113	751	26 206	28 203	6 877	248

Source: Adapted from information drawn from Forest Resource Assessments 2000 (FAO, 2001)

Table 18.2 Forest area changes 1990–2000 in tropical and non-tropical areas (million ha/year)

Domain	Natural forest					Forest plantations			Total forest
	Losses			Gains	Net change	Gains		Net change	Net change
	Deforestation (to other land use)	Conversion (to forest plantations)	Total loss			Conversion from natural forest (reforestation)	Afforestation		
Tropical	−14.2	−1	−15.2	+1	−14.2	+1	+0.9	+1.9	−12.3
Non-tropical	−0.4	−0.5	−0.9	+2.6	+1.7	+0.5	+0.7	+1.2	+2.9
Global	−14.6	−1.5	−16.1	+3.6	−12.5	+1.5	+1.6	+3.1	−9.4

Source: FRA 2000. Table 49–1, p. 334

As Table 18.1 shows, in the year 2000 forests – defined to be areas with at least 10% canopy cover – covered nearly 3.9 billion hectares, of which 95% was natural forest and 5% forest plantations. The former is typically not managed at all (and where it is managed, is not done so primarily for timber production), whereas plantations are commercially operated resources, managed predominantly for timber revenues. While the proportion of plantation forests in total forest land is relatively small, it is growing quickly, at an average of 3.1 million hectares per year during the 1990s. Of this, 1.5 million hectares was converted from natural forest and 1.6 million was on land previously under non-forestry use.

Although the area of plantation forests is relatively small (5% of all forest area), their importance in timber supply is substantially greater (35% of all roundwood – all wood in the rough, for both domestic and industrial purposes – is derived from plantations). Moreover the expansion of plantations has important effects on fuelwood availability, reducing the pressures on natural forests to provide this resource.

Of total forest area, 47% is found in the tropical zone, 9% in the sub-tropics, 11% in the temperate zone and 33% in the boreal zone. Natural forests continue to be lost or converted to other uses at high rates. Between 1990 and 2000, 4.2% of the world's total natural forest area (16.1 million hectares) was lost, with most of this occurring in the tropics (15.2 million hectares). Overall, the picture portrayed in Tables 18.1 and 18.2 shows

- a net loss of world forest area during the 1990s of 2.4%;
- a large loss in tropical forest cover with a much smaller gain in non-tropical forest area;
- a large loss in natural forest area with a much smaller gain in forest plantation area;
- for the broad aggregates considered here, a loss in total forest area in all regions except Europe and temperate North America.

The loss of natural (or primary) forests is a major cause for concern, and one we investigate at length later in the chapter. However, it does appear (see FAO, 2001, p. 343 in web manuscript version) that the net loss of forest land was slower in the 1990s than in the 1980s. This seems to be due to the more rapid expansion of secondary natural forests in the later period, with forest returning to land in which agriculture has been discontinued. Whether the services currently being lost from disappearing primary forests are replaced by the services of maturing secondary natural forests is a moot, but highly important, point.

18.2 Characteristics of forest resources

Let us begin by summarising some of the key characteristics of forest resources, noting several similarities and differences between forest and fish resources.

1. While fisheries typically provide a single service, forests are multi-functional. They directly provide timber, fuelwood, food, water for drinking and irrigation, stocks of genetic resources, and other forest products. Moreover, as ecosystems, forests also provide a wide variety of services, including removal of air pollution, regulation of atmospheric quality, nutrient cycling, soil creation, habitats for

humans and wildlife, watershed maintenance, recreational facilities and aesthetic and other amenities. Because of the wide variety of functions that forests perform, timber managed for any single purpose generates a large number of important external effects. We would expect that the management of woodland resources is often economically inefficient because of the presence of these external effects.
2. Woodlands are capital assets that are intrinsically productive. In this, they are no different from fisheries, and so the techniques we developed earlier for analysing efficient and optimal exploitation should also be applicable (albeit with amendments) to the case of forest resources.
3. Trees typically exhibit very long lags between the date at which they are planted and the date at which they attain biological maturity. A tree may take more than a century to reach its maximum size. The length of time between planting and harvesting is usually at least 25 years, and can be as large as 100 years. This is considerably longer than for most species of fish, but not greatly different from some large animals.
4. Unlike fisheries, tree harvesting does not involve a regular cut of the incremental growth. Forests, or parts of forests, are usually felled in their entirety. It is possible, however, to practise a form of forestry in which individual trees are selectively cut. Indeed, this practice was once common, and is now again becoming increasingly common, particularly where public pressure to manage forests in a multiple-use way is strong. This form of felling is similar to the 'ideal' form of commercial fishing in which adult fish are taken, leaving smaller, immature fish unharvested for a later catch.
5. Plantation forestry is intrinsically more controllable than commercial marine fishing. Tree populations do not migrate spatially, and population growth dynamics are simpler, with less interdependence among species and less dependence on relatively subtle changes in environmental conditions.
6. Trees occupy potentially valuable land. The land taken up in forestry often has an opportunity cost. This distinguishes woodlands from both ocean fisheries (where the ocean space inhabited by fish stocks usually has no value other than as a source of fish) and mineral deposits (where the space occupied by deposits has little or no economic value).
7. The growth in volume or mass of a single stand of timber, planted at one point in time, resembles that illustrated for fish populations in the previous chapter.

To illustrate the assertion made in point 7, we make use of some data reported in Clawson (1977). This refers to the volume of timber in a stand of US Northwest Pacific region Douglas firs. Let S denote the volume, in cubic feet, of standing timber and t the age in years of the stand since planting. (For simplicity, we shall use a year to denote a unit of time.) The age–volume relationship estimated by Clawson for a typical single stand is

$$S = 40t + 3.1t^2 - 0.016t^3$$

Figure 18.1(a) plots the volume of timber over a period up to 145 years after planting. The volume data is listed in the second column in Table 18.3. It is evident from the diagram that an early phase of slow growth in volume is followed by a period of rapid volume growth, after which a third phase of slow growth takes place as the stand moves towards maturity. The stand becomes biologically mature (reaches maximum volume with zero net growth) at approximately 135 years.[2]

How does the amount of annual growth vary with the volume of timber, S? The amount of growth is listed in the third column of Table 18.3, and the growth–volume relationship is plotted in Figure 18.1(b).[3] Although the biological growth function is

[2] Inspection of Clawson's estimated timber growth equation shows that growth becomes negative after (approximately) 135 years. The equation should be regarded as being a valid representation of the growth process only over the domain $t = 0$ to $t = 135$.

[3] Table 18.3 (in a more complete form) and Figures 18.1 (a) and (b) are all generated in the Excel workbook *Chapter18.xls*.

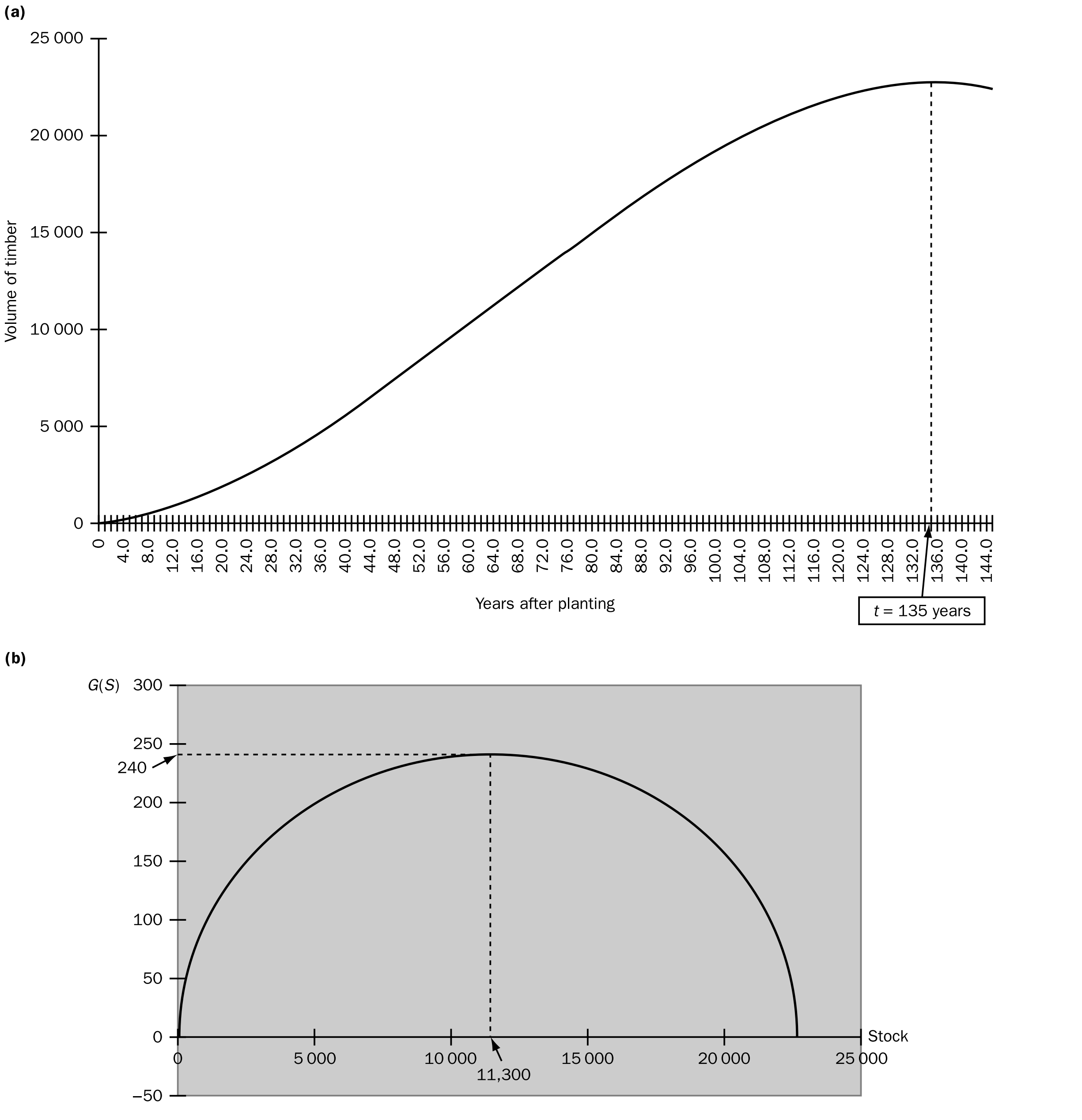


Figure 18.1 *(a)* The volume of timber in a single stand over time; *(b)* Biological growth of a single stand of timber

not logistic in this case, it is very similar in form to simple logistic growth, being a quadratic function (with an inverted U-shaped profile).

Inspection of Figure 18.1(b) or Table 18.3 shows that the biological growth function for one stand reaches a peak annual increment of 240 cubic feet 65 years after planting at a total standing-timber volume of approximately 11 300 cubic feet. When discussing a fishery, we labelled the periodic increment at which the growth function is maximised the

Table 18.3 Present values of revenues, costs and net benefits undiscounted and discounted at 3%

Age of stand t years	Volume of timber S (cu. ft.)	Annual growth $S_t - S_{t-1}$	Interest rate $i = 0.00$			Interest rate $i = 0.03$		
			Revenue R1	Cost C1	Net benefit NB1	Revenue R2	Cost C2	Net benefit NB2
1	43.1	43.1	430.8	5 086.2	–4 655.3	418.1	5083.6	–4665.5
5	275.5	66.9	2 755.0	5 551.0	–2 796.0	2 371.3	5474.3	–3103.0
10	694.0	94.6	6 940.0	6 388.0	552.0	5 141.3	6028.3	–887.0
15	1 243.5	119.8	12 435.0	7 487.0	4 948.0	7 928.9	6585.8	1343.1
20	1 912.0	142.6	19 120.0	8 824.0	10 296.0	10 493.3	7098.7	3394.6
24	2 524.4	159.2	25 244.2	10 048.8	15195.3	12 287.6	7457.5	4830.1
25	2 687.5	163.1	26 875.0	10 375.0	16 500.0	12 694.8	7539.0	5155.9
27	3 025.0	170.6	30 249.7	11 050.0	19 199.8	13 456.8	7691.4	5765.5
30	3 558.0	181.1	35 580.0	12 116.0	23 464.0	14 465.8	7893.2	6572.6
32	3 930.1	187.7	39 301.1	12 860.2	26 441.0	15 048.1	8009.6	7038.5
35	4 511.5	196.8	45 115.0	14 023.0	31 092.0	15 787.4	8157.5	7630.0
39	5 326.0	207.5	53 260.0	15 652.0	37 608.0	16 530.1	8306.0	8224.1
40	5 536.0	210.0	55 360.0	16 072.0	39 288.0	16 674.1	8334.8	8339.3
50	7 750.0	229.3	77 500.0	20 500.0	57 000.0	17 292.6	8458.5	8834.1
60	10 104.0	239.0	101 040.0	25 208.0	75 832.0	16 701.8	8340.4	8361.4
65	11 303.5	240.2	113 035.0	27 607.0	85 428.0	16 082.0	8216.4	7865.6
70	12 502.0	239.0	125 020.0	30 004.0	95 016.0	15 309.5	8061.9	7247.6
80	14 848.0	229.5	148 480.0	34 996.0	113 784.0	13 469.8	7694.0	5775.8
90	17 046.0	210.4	170 460.0	39 092.0	131 368.0	11 455.9	7291.2	4164.7
100	19 000.0	181.7	190 000.0	43 000.0	147 000.0	9 459.5	6891.9	2567.6
110	20 614.0	143.4	206 140.0	46 228.0	159 912.0	7 603.1	6520.7	1082.5
120	21 792.0	95.4	217 920.0	48 584.0	169 336.0	5 954.4	6190.9	–236.5
130	22 438.0	37.9	224 380.0	49 876.0	174 504.0	4 541.9	5908.4	–1366.5
135	22 531.5	5.6	225 315.0	50 063.0	175 252.0	3 925.5	5785.1	–1859.6
140	22 456.0	–29.2	224 560.0	49 912.0	174 648.0	3 367.4	5673.5	–2305.1
145	22 199.5	–66.4	221 995.0	49 399.0	172 596.0	2 865.2	5573.1	–2707.8

'maximum sustainable yield'. But for a stand of trees all planted at one point in time, the concept of a sustainable yield of timber is not meaningful (except for specialised activities such as coppicing). While one can conceive of harvesting mature fish while leaving younger fish to grow to maturity, this cannot happen on a continuous basis in a single-aged forest stand. However, when there are many stands of trees of different ages, it is meaningful to talk about sustainable yields. This is something we shall discuss later.

18.3 Commercial plantation forestry

There is a huge literature dealing with efficient timber extraction. We attempt to do no more than present a flavour of some basic results, and refer the reader to specialist sources of further reading at the end of the chapter. An economist derives the criterion for an efficient forest management and felling programme by trying to answer the following question:

> What harvest programme is required in order that the present value of the profits from the stand of timber is maximised?

The particular aspect of this question that has most preoccupied forestry economists is the appropriate time after planting at which the forest should be felled. As always in economic analysis, the answer one gets to any question depends on what model is being used. We begin with one of the most simple forest models, the single-rotation commercial forest model. Despite its lack of realism, this model offers useful insights into the economics of timber harvesting. However, as we shall see later in the chapter, that which is privately optimal may not be socially efficient. In particular, where private costs and benefits fail to match their social counterparts, a wedge may be driven between privately and socially efficient behaviour. For the moment, we put these considerations to one side.

18.3.1 A single-rotation forest model

Suppose there is a stand of timber of uniform type and age. All trees in the stand were planted at the same time, and are to be cut at one point in time. Once felled, the forest will not be replanted. So only one cycle or rotation – plant, grow, cut – is envisaged. For simplicity, we also assume that

- the land has no alternative uses so its opportunity cost is zero;
- planting costs (k), marginal harvesting costs (c) and the gross price of felled timber (P) are constant in real terms over time;
- the forest generates value only through the timber it produces, and its existence (or felling) has no external effects.

Looked at from the point of view of the forest owner (which, for simplicity, we take to be the same as the landowner), what is the optimum time at which to fell the trees? The answer is obtained by choosing the age at which the present value of profits from the stand of timber is maximised. Profits from felling the stand at a particular age of trees are given by the value of felled timber less the planting and harvesting costs. Notice that because we are assuming the land has no other uses, the opportunity cost of the land is zero and so does not enter this calculation. If the forest is clear-cut at age T, then the present value of profit is

$$(P - c)S_T\,\mathrm{e}^{-iT} - k = pS_T\,\mathrm{e}^{-iT} - k \tag{18.1}$$

where S_T denotes the volume of timber available for harvest at time T, p (in lower-case, note) is the net price of the harvested timber, and i is the private consumption discount rate (which we suppose is equal to the opportunity cost of capital to the forestry firm).

The present value of profits is maximised at that value of T which gives the highest value for $pS_T\mathrm{e}^{-iT} - k$. To maximise this quantity, we differentiate equation 18.1 with respect to T, using the product rule, set the derivative equal to zero and solve for T:[4]

[4] Note from the first of these steps that k does not enter the first derivative, and so immediately we find that in a single rotation model, planting costs have no effect on the efficient rotation length (provided that k is not so large as to make the maximised present value negative).

$$\frac{d}{dT}(pS_T e^{-iT} - k) = \frac{d}{dT}(pS_T e^{-iT})$$

$$= pe^{-iT}\frac{dS}{dT} + pS_T\frac{de^{-iT}}{dT}$$

which, setting equal to zero, implies that

$$pe^{-iT}\frac{dS}{dT} - ipS_T e^{-iT} = 0$$

and so

$$p\frac{dS}{dT} = ipS_T$$

or

$$i = \frac{p\dfrac{dS}{dT}}{pS_T} \tag{18.2}$$

Equation 18.2 states that the present value of profits is maximised when the rate of growth of the (undiscounted) net value of the resource stock is equal to the private discount rate. Note that with the timber price and harvesting cost constant, this can also be expressed as an equality between the proportionate rate of growth of the volume of timber and the discount rate. That is,

$$i = \frac{\dfrac{dS}{dT}}{S_T} \tag{18.2′}$$

We can calculate the optimal, present-value-maximising age of the stand for the illustrative data in Table 18.3. These calculations, together with the construction of the associated graphs are reproduced in the Excel workbook *Chapter18.xls* which can be downloaded from the *Additional Materials* web page. In these calculations, we assume that the market price per cubic foot of felled timber is £10, total planting costs are £5000, incurred immediately the stand is established, and harvesting costs are £2 per cubic foot, incurred at whatever time the forest is felled. The columns labelled R1, C1 and NB1 list the present values of revenues and costs and profits (labelled Net benefit in the table) for a discount rate of zero. Note that when $i = 0$, present values are identical to undiscounted values. The level of the present value of profits (NB1) over time is shown in Figure 18.2. Net benefits are maximised at 135 years, the point at which the biological growth of the stand (dS/dt) becomes zero. With no discounting and fixed timber prices, the profile of net value growth of the timber is identical to the profile of net volume growth of the timber, as can be seen by comparing Figures 18.1(a) and 18.2.

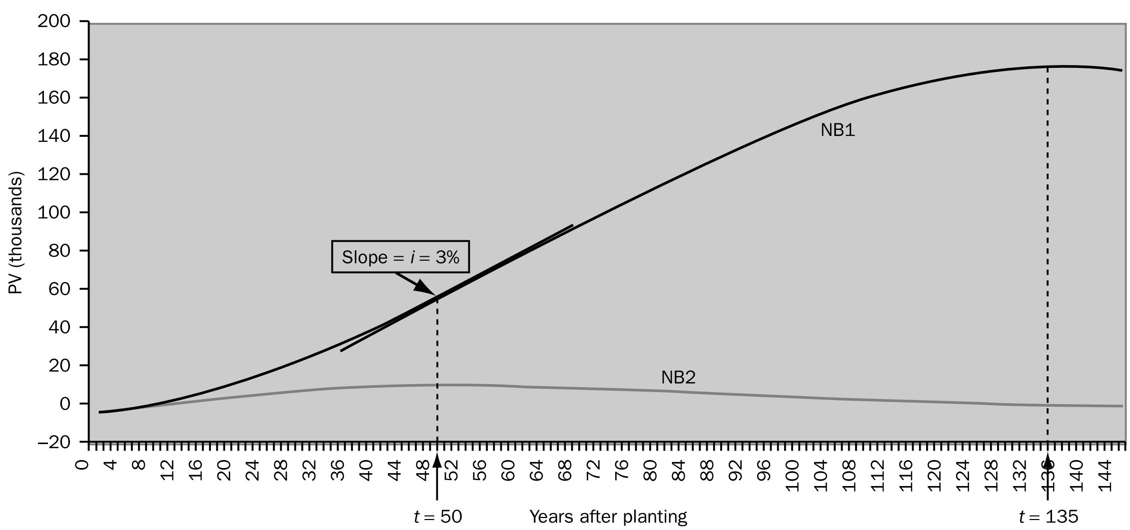


Figure 18.2 Present values of net benefits at $i = 0.00$ (NB1) and $i = 0.03$ (NB2)

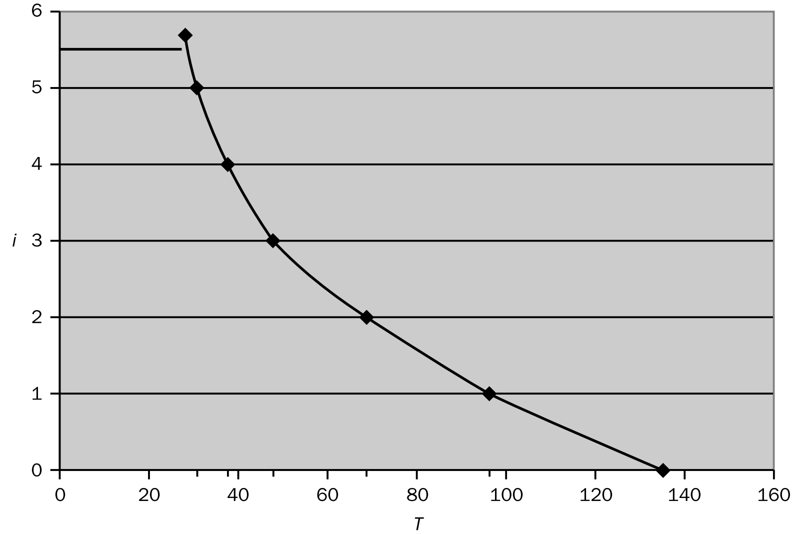


Figure 18.3 The variation of the optimal felling age with the interest rate, for a single-rotation forest

It is also useful to look at this problem in another way. The interest rate to a forest owner is the opportunity cost of the capital tied up in the growing timber stand. When the interest rate is zero, that opportunity cost is zero. It will, therefore, be in the interests of the owner to not harvest the stand as long as the volume (and value) growth is positive, which it is up to an age of 135 years. Indeed, inspection of equation 18.2′ confirms this; given that S is positive, when $i = 0$ $\mathrm{d}S/\mathrm{d}t$ must be zero to satisfy the first-order maximising condition.

Now consider the case where the discount rate is 3%. The columns labelled R2, C2 and NB2 in Table 18.3 refer to the present values of revenues, costs and profits when the interest rate is 3%. The present value of profits at a discount rate of 3% is also plotted in Figure 18.2, under the legend NB2. With a 3% discount rate, the present value of the forest is maximised at a stand age of 50 years.

Expressed in a way that conforms to equation 8.2, the growth of *undiscounted* profits,

$$\frac{p\dfrac{\mathrm{d}S}{\mathrm{d}T}}{pS_T}$$

equals i (at 3%) in year 50, having been larger than 3% before year 50 and less than 3% thereafter. This is shown by the '$i = 3\%$' line which has an identical slope to that of the NB1 curve at $t = 50$ in Figure 18.2. At that point, the growth rate of undiscounted timber value equals the interest rate. A wealth-maximising owner should harvest the timber when the stand is of age 50 years – up to that point, the return from the forest is above the interest rate, and beyond that point the return to the forest is less than the interest rate.

The single-rotation model we have used shows that the optimal time for felling will depend upon the discount rate used. It can be seen from our calculations that this effect can be huge. A rise in the discount rate from zero to 3% not only dramatically lowers the profitability of the forest but also significantly changes the shape of the present-value profile, reducing the age at which the forest should be felled (in our illustrative example) from 135 to 50 years.

More generally, it is clear from our previous arguments that as the interest rate rises the age at which the stand is felled will have to be lowered in order to bring about equality between the rate of change of undiscounted net benefits and the discount rate. In Figure 18.3, we illustrate how the optimal felling age varies with the interest rate for our illustrative data. While the exact relationship shown is only valid under the assumptions used here, it does suggest that small changes in interest rates might dramatically alter privately optimal harvesting programmes.

18.3.2 Infinite-rotation forestry models

The forestry model we investigated in the previous section is unsatisfactory in a number of ways. In particular, it is hard to see how it would be meaningful to have only a single rotation under the assumption that there is no alternative use of the land. If price and cost conditions warranted one cycle then surely, after felling the stand, a rational owner would consider further planting cycles if the land had no other uses? So the next step is to move to a model in which more than one cycle or rotation occurs. The conventional practice in forestry economics is to analyse harvesting behaviour in an infinite time horizon model (in which there will be an indefinite quantity of rotations). A central question investigated here is what will be the optimal length of each rotation (that is, the time between one planting and the next).

When the harvesting of one stand of timber is to be followed by the establishment of another, an additional element enters into the calculations. In choosing an optimal rotation period, a decision to defer harvesting incurs an additional cost over that in the previous model. We have already taken account of the fact that a delay in harvesting has an opportunity cost in the form of interest forgone on the (delayed) revenues from harvesting. But a second kind of opportunity cost now enters into the calculus. This arises from the delay in establishing the next and all subsequent planting cycles. Timber that would have been growing in subsequent cycles will be planted later. So an optimal harvesting and replanting programme must equate the benefits of deferring harvesting – the rate of growth of the undiscounted net benefit of the present timber stand – with the costs of deferring that planting – the interest that could have been earned from timber revenues and the return lost from the delay in establishing subsequent plantings.

Our first task is to construct the present-value-of-profits function to be maximised for the infinite-rotation model. We continue to make several simplifying assumptions that were used in the single-rotation model: namely, the total planting cost, k, the gross price of timber, P, and the harvesting cost of a unit of timber, c, are constant through time. Given this, the net price of timber $p = P - c$ will also be constant.

Turning now to the rotations, we assume that the first rotation begins with the planting of a forest on bare land at time t_0. Next, we define an infinite sequence of points in time that are ends of the successive rotations, t_1, t_2, t_3, ... At each of these times, the forest will be clear-felled and then immediately replanted for the next cycle. The net present value of profit from the first rotation is

$$pS_{(t_1-t_0)}\mathrm{e}^{-i(t_1-t_0)} - k$$

that is, the volume of timber growth between the start and end of the cycle multiplied by the discounted net price of a unit of timber, less the forest planting cost. Notice that because the planting cost is incurred at the start of the rotation, no discounting is required to bring it into present-value terms. But as the timber is felled at the end of the rotation (t_1), the timber revenue has to be discounted back to its present (t_0) value equivalent.

The net present value of profits over this infinite sequence is given by

$$\begin{aligned}\Pi = {} & [pS_{(t_1-t_0)}\mathrm{e}^{-i(t_1-t_0)} - k] \\ & + \mathrm{e}^{-i(t_1-t_0)}[pS_{(t_2-t_1)}\mathrm{e}^{-i(t_2-t_1)} - k] \\ & + \mathrm{e}^{-i(t_2-t_0)}[pS_{(t_3-t_2)}\mathrm{e}^{-i(t_3-t_2)} - k] \\ & + \mathrm{e}^{-i(t_3-t_0)}[pS_{(t_4-t_3)}\mathrm{e}^{-i(t_4-t_3)} - k] \\ & + \ldots\end{aligned} \tag{18.3}$$

Reading this, we see that the present value of profits from the infinite sequence of rotations is equal to the sum of the present values of the profit from each of the individual rotations.

Provided conditions remain constant through time, the optimal length of any rotation will be the same as the optimal length of any other. Call the interval of time in this optimal rotation T. Then we can rewrite the present-value function as

$$\begin{aligned}\Pi = {} & [pS_T\mathrm{e}^{-iT} - k] \\ & + \mathrm{e}^{-iT}[pS_T\mathrm{e}^{-iT} - k] \\ & + \mathrm{e}^{-2iT}[pS_T\mathrm{e}^{-iT} - k] \\ & + \mathrm{e}^{-3iT}[pS_T\mathrm{e}^{-iT} - k] \\ & + \ldots\end{aligned} \tag{18.4}$$

Next, factorise out e^{-iT} from the second term on the right-hand side of equation 18.4 onwards to give

$$\Pi = [pS_T e^{-iT} - k] + e^{-iT}\{[pS_T e^{-iT} - k] + e^{-iT}[pS_T e^{-iT} - k] + e^{-2iT}[pS_T e^{-iT} - k] + \dots\} \quad (18.5)$$

Now look at the term in braces on the right-hand side of equation 18.5. This is identical to Π in equation 18.4. Therefore, we can rewrite equation 18.5 as

$$\Pi = [pS_T e^{-iT} - k] + e^{-iT}\Pi \quad (18.6)$$

which on solving for Π gives[5]

$$\Pi = \frac{pS_T e^{-iT} - k}{1 - e^{-iT}} \quad (18.7)$$

Equation 18.7 gives the present value of profits for any rotation length, T, given values of p, k, i and the timber growth function $S = S(t)$. The wealth-maximising forest owner selects that value of T which maximises the present value of profits. For the illustrative data in Table 18.3, we have used a spreadsheet program to numerically calculate the present-value-maximising rotation intervals for different values of the discount rate. (The spreadsheet is available in *Additional Materials*, Chapter 18, as *Chapter18.xls*, Sheet 2.) Present values were obtained by substituting the assumed values of p, k and i into equation 18.7, and using the spreadsheet to calculate the value of Π for each possible rotation length, using Clawson's timber growth equation. The results of this exercise are presented in Table 18.4 (along with the optimal rotation lengths for a single rotation forest, for comparison). Discount rates of 6% or higher result in negative present values at any rotation, and the asterisked rotation periods shown are those which minimise present-value losses; commercial forestry would be abandoned at those rates. With our illustrative data, at any discount rate which yields a positive net present value for the forest the optimal rotation interval in an infinite-rotation forest is lower than the age at which a forest would be felled in a single rotation model. For example, with a 3% discount rate, the optimal rotation interval in an infinite sequence of rotations is 40 years, substantially less than the 50-year harvest age in a single rotation. We will explain why this is so shortly.

Table 18.4 Optimal rotation intervals for various discount rates

i	Optimal T (years) in infinite-rotation model	Optimal T (years) in single-rotation model
0	99	135
1	71	98
2	51	68
3	40	50
4	33	38
5	29	31
6	26*	26*
7	24*	22*
8	22*	19*
9	21*	17*
10	20*	15*

Notes to table:
1. Data simulated by Excel, using workbook *Chapter18.xls*
2. * For both single- and infinite-rotation models, at interest rates of 6% and above (for the price, cost and growth data used here) the PV is negative even at optimal T, so the land would not be used for commercial forestry. The value of T shown in these cases is that which minimises the PV loss.
3. To simulate the solution for $i = 0$, we used a value of i sufficiently close to (although not exactly equal to) zero so that the optimal rotation, in units of years, was unaffected by a further reduction in the value of i.

It is also useful to think about the optimal rotation interval analytically, as this will enable us to obtain some important comparative statics results. Let us proceed as was done in the section on single-rotation forestry. The optimal value of T will be that which maximises the present value of the forest over an infinite sequence of planting cycles. To find the optimal value of T, we obtain the first derivative of Π with respect to T, set this derivative equal to zero, and solve the resulting equation for the optimal rotation length.

[5] A more elegant method of obtaining equation 18.7 from 18.4 is as follows. Equation 18.4 may be rewritten as

$$\Pi = (pS_T e^{-iT} - k)\,(1 + e^{-iT} + (e^{-iT})^2 + (e^{-iT})^3 + \dots)$$

The final term in parentheses is the sum of an infinite geometric progression. Given the values that i and T may take, this is a convergent sum. Then, using the result for such a sum, that term can be written as $1/(1 - e^{-iT})$, and so

$$\Pi = \frac{(pS_T e^{-iT} - k)}{1 - e^{-iT}}$$

The algebra here is simple but tedious, and so we have placed it in Appendix 18.1. Two forms of the resulting first-order condition are particularly useful, each being a version of the Faustmann rule (derived by the German capital theorist Martin Faustmann in 1849; see Faustmann (1968)). The first is given by

$$\frac{p\dfrac{dS_T}{dT}}{pS_T - k} = \frac{i}{1 - e^{-iT}} \tag{18.8a}$$

and the second, after some rearrangement of 18.8a, is given by

$$p\frac{dS_T}{dT} = ipS_T + i\Pi \tag{18.8b}$$

Either version of equation 18.8 is an efficiency condition for present-value-maximising forestry, and implicitly determines the optimal rotation length for an infinite rotation model in which prices and costs are constant.[6] Given knowledge of the function $S = S(t)$, and values of p, i and k, one could deduce which value of T satisfies equation 18.8 (assuming the solution is unique, which it usually will be). The term Π in equation 18.8b is called the site value of the land – the capital value of the land on which the forest is located. This site value is equal to the maximised present value of an endless number of stands of timber that could be grown on that land.

The two versions of the Faustmann rule offer different advantages in helping us to make sense of optimal forest choices. Equation 18.8b gives some intuition for the choice of rotation period. The left-hand side is the increase in the net value of the timber left growing for an additional period. The right-hand side is the value of the opportunity cost of this choice, which consists of the interest forgone on the capital tied up in the growing timber (the first term on the right-hand side) and the interest forgone by not selling the land at its current site value (the second term on the right-hand side). An efficient choice equates the values of these marginal costs and benefits. More precisely, equation 18.8b is a form of Hotelling *dynamic* efficiency condition for the harvesting of timber. This is seen more clearly by rewriting the equation in the form:

$$\frac{p\left(\dfrac{dS}{dT}\right)}{pS_T} = i + \frac{i\Pi}{pS_T} \tag{18.9}$$

Equation 18.9 states that, with an optimal rotation interval, the proportionate rate of return on the growing timber (the term on the left-hand side) is equal to the rate of interest that could be earned on the capital tied up in the growing timber (the first term on the right-hand side) plus the interest that could be earned on the capital tied up in the site value of the land ($i\Pi$) expressed as a proportion of the value of the growing timber (pS_T).

We can use the other version of the Faustmann rule – equation 18.8a – to illustrate graphically how the optimal rotation length is determined. This is shown in Figure 18.4. The curves labelled 0%, 1%, 2% and 3% plot the right-hand side of equation 18.8a for these rates of interest. The other, more steeply sloped, curve plots the left-hand side of the equation. At any given interest rate, the intersection of the functions gives the optimum T. The calculations required to generate Figure 18.4 are implemented in *Sheet 3* of the Excel file *Chapter18.xls*, together with the chart itself.

The lines plotting the right-hand side of equation 18.8a are generated assuming particular values for P, c, k and i, and also a particular natural growth function describing how timber volume S changes over time. The reader is invited to copy this worksheet, and to study the way in which optimised T varies as p (that is, $P - c$), or k changes, *ceteris paribus*.

18.3.2.1 Comparative static analysis

The results of the previous section have shown that in the infinite-rotation model the optimum rotation depends on:

- the biological growth process of the tree species in the relevant environmental conditions;
- the interest (or discount) rate (i);
- the cost of initial planting or replanting (k);
- the net price of the timber (p), and so its gross price (P) and marginal harvesting cost (c).

[6] Unlike in the case of a single-rotation model, planting costs k do enter the first derivative. So in an infinite-rotation model, planting costs do affect the efficient rotation length.

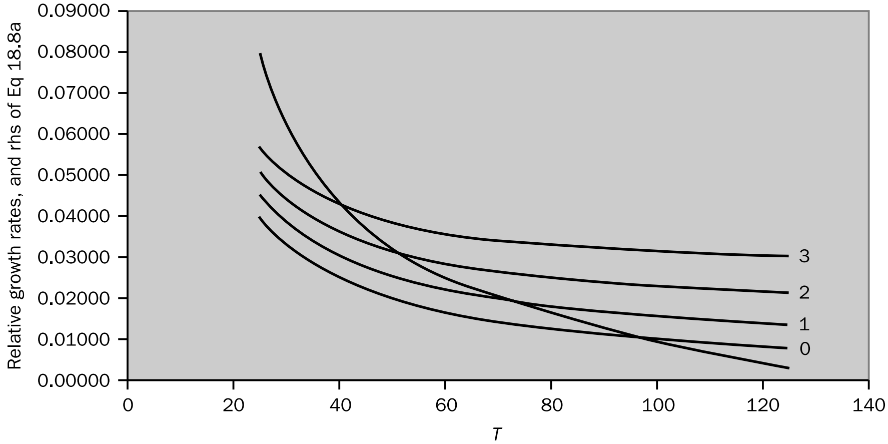


Figure 18.4 Optimal rotation lengths, *T*, as determined by equation 18.8a

Table 18.5 The infinite-rotation model: comparative static results

Change in:	i	k	$p = P - c$
Effect on optimal rotation length	$dT/di < 0$	$dT/dk > 0$	$dT/dp < 0$

Comparative static analysis can be used to make qualitative predictions about how the optimal rotation changes as any of these factors vary. We do this algebraically using equation 18.8b. Derivations of the results are given in Appendix 18.2. Here we just state the results (for convenience, they are tabulated in Table 18.5) and provide some intuition for each of them.

Changes in the interest rate

The result that $dT/di < 0$ means that the interest rate and the optimal rotation period are negatively related. An increase (decrease) in i causes a decrease (increase) in T. Why does this happen? Once planted, there are costs and benefits in leaving a stand unfelled for a little longer. The marginal benefit derives from the marginal revenue product of the additional timber growth. The marginal costs are of two kinds: first, the interest earnings forgone in having capital (the growing timber) tied up a little longer; and second, the interest earnings forgone from not clearing and then selling the bare land at its capital (site) value. If the interest rate increases, the terms of this trade-off change, because the opportunity costs of deferring felling become larger.[7] Foresters respond to this by shortening their forest rotation period.

Changes in planting costs

The result that $dT/dk > 0$ means that a change in planting costs changes the optimal rotation in the same direction. A fall in k, for example, increases the site value of the land, Π. With planting costs lower, the profitability of all future rotations will rise, and so the opportunity costs of *delaying* replanting will rise. The next replanting should take place sooner. The optimal stand age at cutting will fall.

Changes in the net price of timber

The result that $dT/dp < 0$ means that the net price of timber (p) and the optimal rotation length are negatively related. Therefore, an increase in timber prices (P) will decrease the rotation period, and an increase in harvest costs will increase the rotation period. We leave you to deduce the intuition behind this for

[7] There is a trap to watch out for here. An increase in discount rates will increase the opportunity cost of each unit of tied-up capital; but at the same time, it will reduce the magnitude of Π, which you will recall is measured in present-value terms. However, inspection of equation 18.8.4 in Appendix 18.2 confirms that the effect of a change in i on T must be negative.

yourself, in the light of what we have suggested for the two previous cases.

An Excel spreadsheet model (*palc18.xls*) can be used to explore these changes *quantitatively*, for an assumed growth process and particular values of the relevant economic parameters. We recommend that you work through that Excel file, and then experiment further with it. The workbook allows you to reproduce the numbers given in the textbook, to answer the Problems at the end of the chapter, and to see how the comparative static results work out quantitatively.

18.3.2.2 Comparing single and infinite rotations: how does a positive site value affect the length of a rotation?

To see the effect of land site values on the optimal rotation interval, compare equation 18.9 (the Hotelling rule taking into consideration positive site values) with equation 18.10, which is the Hotelling rule when site values are zero (and is obtained by setting $\Pi = 0$ in equation 18.9):

$$\frac{p\dfrac{\mathrm{d}S}{\mathrm{d}T}}{pS_T} = i \tag{18.10}$$

In this case, an optimal rotation interval is one in which the rate of growth of the value of the growing timber is equal to the interest rate on capital alone.

But it is clear from inspection of equation 18.9 that for any given value of i, a positive site value will mean that $(\mathrm{d}S/\mathrm{d}t)/S$ will have to be larger than when the site value is zero if the equality is to be satisfied. This requires a shorter rotation length, in order that the rate of timber growth is larger at the time of felling. Intuitively, the opportunity cost of the land on which the timber is growing requires a compensating increase in the return being earned by the growing timber. With fixed timber prices, this return can only be achieved by harvesting at a point in time at which its biological growth is higher, which in turn requires that trees be felled at a younger age. Moreover, the larger is the site value, the shorter will be the optimal rotation.

It is this which explains why the optimal rotation intervals (for forests that are commercially viable) shown in Table 18.3 are shorter for infinite rotations than for a single rotation. In an infinite-rotation model, land is valuable (because the timber that can be grown on it in the future can yield profits), and the final term in equation 18.9 comes into play.

The reader should note that the way in which bare land is valued by the Faustmann rule – the present value of profits from an infinite sequence of optimal timber rotations – is not the only basis on which one might choose to arrive at land values. Another method would be to value the land at its true opportunity cost basis – that is, the value of the land in its most valuable use other than forestry. In many ways, this is a more satisfactory basis for valuation. This approach can give some insights into forestry location. In remote areas with few alternative land uses, low land prices may permit commercial forest growth even at high altitude where the intrinsic rate of growth of trees is low. In urban areas, by contrast, the high demand for land is likely to make site costs high. Timber production is only profitable if the rate of growth is sufficiently high to offset interest costs on tied-up land capital costs. There may be no species of tree that has a fast enough growth potential to cover such costs. In the same way, timber production may be squeezed out by agriculture where timber growth is slow relative to crop potential (especially where timber prices are low). All of this suggests that one is not likely to find commercial plantations of slow-growing hardwood near urban centres unless there are some additional values that should be brought into the calculus. It is to this matter that we now turn.

18.4 Multiple-use forestry

In addition to the timber values that we have been discussing so far, forests are capable of producing a wide variety of non-timber benefits. These include soil and water control, habitat support for a biologically diverse system of animal and plant populations, recreational and aesthetic amenities, wilderness existence values, and climate control. Where forests do provide one or more of these benefits to a significant extent, they are called multiple-use forests.

Efficiency considerations imply that the choices of how a forest should be managed and how frequently it should be felled (if at all) should take account of the multiplicity of forest uses. If the forest owner is able to appropriate compensation for these non-timber benefits, those benefits would be factored into his or her choices and the forest should be managed in a socially efficient way. If these benefits cannot be appropriated by the landowner then, in the absence of government regulation, we would not expect them to brought into the owner's optimising decisions. Decisions would be privately optimal but socially inefficient.

For the moment we will assume that the owner can appropriate the value generated by all the benefits of the forest: both timber and non-timber benefits. Our first task is to work out how the inclusion of these additional benefits into the calculations alters the optimal rotation age of a forest. Once again we imagine beginning at time zero with some bare land. Let NT_t denote the *undiscounted* value of the flow of non-timber benefits t years after the forest is established. The present value of these non-timber value flows over the whole of the first rotation of duration T is

$$\int_{t=0}^{t=T} NT_t e^{-it} dt$$

Now for simplicity denote this integral as N_T, so that we regard the present value of the stream of non-timber values (N) during one rotation as being a function of the rotation interval (T). Adding the present value of the non-timber benefits to the present value of timber benefits, the present value of all forest benefits for the first rotation is

$$PV_1 = (pS_T - k)e^{-iT} - k + N_T$$

For a single rotation, the optimal age at which the stand should be felled is that value of T which maximises PV_1. Is the rotation age lengthened or shortened? In this special case (a single rotation only) the answer is unambiguous. Provided that non-timber values are positive, the optimal felling age will be increased. This is true irrespective of whether the non-timber values are constant, rising or falling through time. To see why, note that if these values are always positive, the NPV of non-timber benefits will increase the longer is the rotation. This must increase the age at which it is optimal to fell the forest. Problem 5 at the end of this chapter invites you to use an Excel file, *Non timber.xls*, to explore this matter and verify these conclusions.

Matters are more complicated in the case of an infinite succession of rotations of equal duration. Then the present value of the whole infinite sequence is given by

$$\begin{aligned} \Pi^* = & [pS_T e^{-iT} - k + N_T] \\ & + e^{-iT}[pS_T e^{-iT} - k + N_T] \\ & + e^{-2iT}[pS_T e^{-iT} - k + N_T] \\ & + e^{-3iT}[pS_T e^{-iT} - k + N_T] \\ & + \ldots \end{aligned} \qquad (18.11)$$

which is just a generalisation of equation 18.4 including non-timber benefits. Alternatively, we could interpret equation 18.11 as saying that the present value of all benefits from the rotation (Π^*) is equal to the sum of the present value of timber-only benefits from the rotation (Π) and the present value of non-timber-only benefits from the infinite sequence of rotations.

A forest owner who wishes to maximise the net present value of timber and non-timber benefits will choose a rotation length that maximises this expression. Without going through the derivation (which follows the same steps as before), wealth maximisation requires that the following first-order condition is satisfied:

$$p\frac{dS}{dT^*} + N_{T^*} = ipS_{T^*} + i\Pi^* \qquad (18.12)$$

in which asterisks have been included to emphasise the point that the optimal rotation interval when all benefits are considered (T^*) will in general differ from the interval which is optimal when only timber benefits are included in the function being maximised (T). For the same reason, the optimised present value (and so the land site value) will in general be different from their earlier counterparts, and we will denote these as Π^*.

What effect does the inclusion of non-timber uses of forests have on the optimal rotation length? Inspection of equation 18.12 shows that non-timber benefits affect the optimal rotation in two ways:

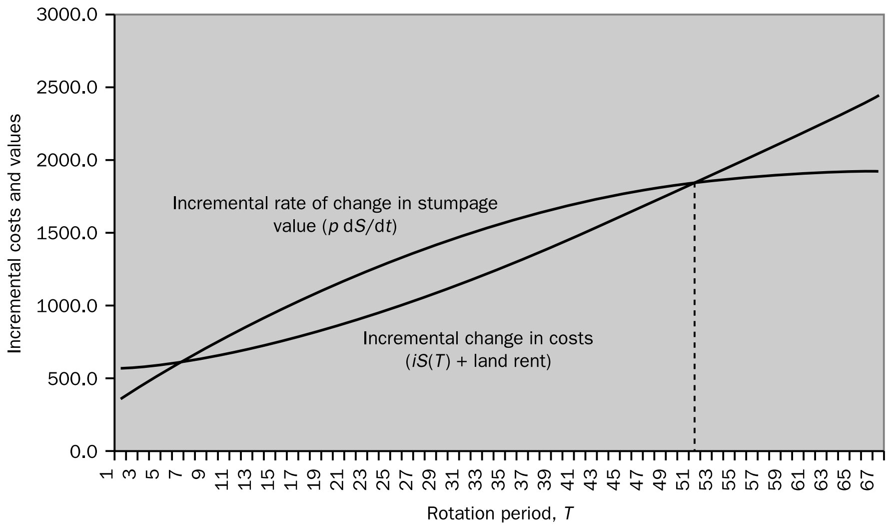


Figure 18.5 Incremental change in value and costs with rotation stand age

- the present value of the flows of non-timber benefits over any one rotation (N_T^*) enter equation 18.12 directly; other things being equal, a positive value for N_T^* implies a reduced value of dS/dT, which means that the rotation interval is lengthened;
- positive non-timber benefits increase the value of land (from Π to Π^*) and so increase the opportunity cost of maintaining timber on the land; this will tend to reduce the rotation interval.

Which of these two opposing effects dominates depends on the nature of the functions $S(t)$ and $N(t)$. Therefore, for infinite-rotation forests it is not possible to say *a priori* whether the inclusion of non-timber benefits shortens or lengthens rotations. However, some qualitative results can be obtained from equation 18.8(b), which for convenience is given again here:

$$p\frac{dS}{dT} = ipS_T + i\Pi$$

Recall that Π is called the site value of the land, and is equal to the maximised present value of an endless number of stands of timber that could be grown on the land. The second term on the right-hand side – often called land rent – is thus the interest forgone by not selling the land at its current site value. The first term on the right-hand side constitutes the interest forgone on the value of the growing timber. Adding these two costs together, we arrive at the full opportunity cost of this choice, the marginal cost of deferring harvesting. The left-hand side is the increase in the net value of the timber left growing for an additional period, and so is the marginal benefit of deferring harvesting. An efficient choice equates the values of these marginal costs and benefits.

This equality is represented graphically in Figure 18.5. The inclusion of non-timber values potentially changes the left-hand side of equation 18.8b. If non-timber values are greater in old than in young forests (are rising with stand age) then non-timber values have a positive annual increment; adding these to the timber values will increase the magnitude of the change in overall (timber + non-timber) benefits, shifting the incremental benefits curve upwards. Its intersection with the incremental costs curve will shift to the right, generating a longer optimal rotation. An equivalent, but opposite, argument shows that falling non-timber benefits will shorten the optimal rotation.

Only if the flow of non-timber benefits is constant over the forest cycle will the optimal rotation interval be unaffected. Hence it is variation over the cycle in non-timber benefits, rather than their existence as such, that causes the rotation age to change.

It is often assumed that NT (the annual magnitude of undiscounted non-timber benefits) increases with the age of the forest. While this may happen, it need not always be the case. Studies by Calish *et al.* (1978) and Bowes and Krutilla (1989) suggest that some kinds of non-timber values rise strongly with forest age (for example, the aesthetic benefits of forests), others decline (including water values) and yet others have no simple relationship with forest age. There is also reason to believe that total forest benefits are maximised when forests are heterogeneous (with individual forests being specialised for specific purposes) rather than being managed in a uniform way (see Swallow and Wear, 1993; Vincent and Blinkley, 1993). All that can be said in general is that it is most unlikely that total non-timber benefits will be independent of the age of forests, and so the inclusion of these benefits into rotation calculations will make some difference.

Note also that in extreme cases the magnitude and timing of non-timber benefits may be so significant as to result in no felling being justified. Where this occurs, we have an example of what is called 'dominant-use' forestry. It suggests that the woodland in question should be put aside from any further commercial forest use, perhaps being maintained as a national park or the like.

18.5 Socially and privately optimal multiple-use plantation forestry

Our discussions of multiple-use forestry have assumed that the forest owner either directly receives all the forest benefits or is able to appropriate the values of these benefits (presumably through market prices). In that case, what is privately optimal will also be what is socially optimal (provided, of course, that there is no divergence between social and private consumption discount rates). But it is most implausible that forest owners can appropriate all forest benefits. Many of these are public goods; even if exclusion could be enforced and markets brought into existence, market prices would undervalue the marginal social benefits of those public goods. In many circumstances, exclusion will not be possible and open-access conditions will prevail.

Where there is a divergence between private and social benefits, the analysis of multiple-use forestry we have just been through is best viewed as providing information about the socially optimal rotation length. In the absence of efficient bargaining (see Chapter 5), to achieve such outcomes would involve public intervention. This might consist of public ownership and management, regulation of private behaviour, or the use of fiscal incentives to bring social and private objectives into line. The fact that forestland often satisfies multiple uses suggests that there are likely to be efficiency gains available where government integrates environmental policy objectives with forestry objectives.

18.6 Natural forests and deforestation

A series of recent studies, including FAO (1995), FAO (2001), and various editions of *World Resources* (by the World Resources Institute), paint a vivid picture of the pattern and extent of natural forest loss and conversion (deforestation). The extent of human impact on the natural environment can be gauged by noting that by 1990 almost 40% of the earth's land area had been converted to cropland and permanent pasture. Most of this has been at the expense of forest and grassland.

Until the second half of the 20th century, deforestation largely affected temperate regions. In several of these, the conversion of temperate forests has been effectively completed. North Africa and the Middle East now have less than 1% of land area covered by natural forest. It is estimated that only 40% of Europe's original forestland remains, and most of what currently exists is managed secondary forest or plantations. The two remaining huge tracts of primary temperate forest – in Canada and Russia – are now being actively harvested, although rates of conversion are relatively slow. Russia's boreal (coniferous) forests are now more endangered by degradation of quality than by quantitative change, and the same is true for all forms of temperate woodland throughout Europe, which appear to be experiencing severe pollution damage, with about a quarter of trees suffering moderate to severe defoliation. The picture is not entirely bleak, however. China has recently undertaken a huge reforestation programme, and the total Russian forest area is currently increasing. And in developed countries, management practices in secondary and plantation forests are becoming more environmentally benign, partly as a result of changing public opinion and political pressure.

Not surprisingly, the extent of deforestation tends to be highest in those parts of the world which have the greatest forest coverage. With the exceptions of temperate forests in China, Russia and North America, it is tropical forests that are the most extensive. And it is tropical deforestation that is now perceived as the most acute problem facing forest resources. In the thirty years from 1960 to 1990 one-fifth of all natural tropical forest cover was lost, and the rate of deforestation increased steadily during that period. FAO (2001) tentatively suggests, though, that this rate may have slightly slowed in the final decade of the 20th century. Box 18.1 contains a

Box 18.1 Tropical deforestation

Tropical deforestation has many adverse consequences. As far as the countries in which the forests are are concerned, valuable timber assets are irretrievably lost, and the loss of tree cover (particularly when it is followed by intensive agriculture or farming) can precipitate severe losses of soil fertility. Indigenous people may lose their homelands (and their distinctive cultures), water systems may be disrupted, resulting in increased likelihood of extreme hydrological conditions (more droughts and more floods, for example), and local climates may be subtly altered. Perhaps most pernicious are the losses in potential future incomes which deforestation may lead to. Tropical forests are immense stores of biological diversity and genetic material, and quasi-option values (see Chapters 12 and 13) are forfeited as this diversity is reduced. With the loss of animal and plant species and the gestalt of a primary tropical forest will go recreational amenities and future tourism potential.

All of this is reinforcing a point made earlier: tropical forests are multiple-service resources *par excellence*. Many of these forest services benefit the world as a whole of course, rather than just local inhabitants. Of particular importance here are the losses of stores of diverse genetic material, the climate control mechanisms that are part of tropical forest systems, and the emission of greenhouse gases when forests are cleared (see Chapter 10 for further details).

Given these adverse consequences, why are tropical forests being lost? There appears to be no single, predominant cause. As with earlier discussions of biodiversity loss, it is convenient to distinguish between proximate (or immediate) causes and fundamental causes. Economists tend to focus on the latter. And important in this latter category – especially for tropical forests – is the absence of clearly defined and enforceable property rights. The lack of access restrictions must at least partially explain the fact that less than 0.1% of tropical logging is currently being done on a sustainable yield basis (WR, 1996).

Many commentators give a large role to population pressure, especially when significant numbers of people in burgeoning populations have no land entitlement or are living close to the margin of poverty. However, it is now being realised that too much weight has been attributed to this cause, and that emphasis has been given to it in part at least because most models of deforestation have been constructed to be population-driven (see FAO, 2001). This reflects a point well worth remembering about economic modelling: what you get out (here the conclusions) depends very much on what you put in (here, modelling structures and assumptions).

Nevertheless, it is not difficult to understand why many governments, faced with growing populations, mounting debt and growing problems in funding public expenditure, will tend to regard tropical forests as capital assets that can be quickly turned into revenues. Moreover, cleared forestland can also provide large additional sources of land for agriculture and ranching, each of which may offer far greater financial returns than are obtainable from natural forests.

This suggests that the conversion of forestland to other uses (principally agriculture) may well be optimal from the point of view of those who make land-use choices in tropical countries.

Box 18.1 *continued*

Of course, it may be the case that the incentive structures are perverse, as a result of widespread market failure. Tropical deforestation is *not* simply the result of ignorance, short-sightedness, or commercial pressure from organised business (although any of these may have some bearing on the matter). It is the result of the patterns of incentives that exist. This way of thinking is important because it suggests ways of changing behaviour, based on altering those incentive structures.

Several writers have developed models of tropical forest conversion arising from optimising rational behaviour. Hartwick (1992) suggests that the use of any single piece of land will be determined by the relative magnitudes of B^F, the net benefits of the land in forestry (which includes both timber and non-timber values) and B^A, the net benefits of the land in agriculture. At the level of the whole economy, there will be many individual natural forest stands, and we can envisage deforestation as a gradual process by which an increasing proportion of these stands is converted to agriculture over time. The socially efficient rate of conversion at any point in time is that at which these benefits are equalised at the margin. That is $MB^F = MB^A$. One might expect MB^F to rise as the remaining area of tropical forest becomes ever smaller. This would tend to slow down forest conversion. However, this effect may be offset by a rise in MB^A which could arise because of population increases or higher incomes. It is not inconceivable that the outcome of this process would be one in which all forestland is converted. That likelihood is increased if MB^F only includes timber benefits, but excludes the non-timber, or environmental, benefits. For the reasons we gave in the text, there are good reasons to believe that the non-timber benefits will be excluded from the optimising exercise.

Barbier and Burgess (1997) develop Hartwick's ideas a little further. Their optimising model specifies a demand-and-supply function for forestland conversion to agriculture. At any point in time, the supply and demand for forestland conversion, taking account of both timber and non-timber benefits, can be represented by the functions labelled S_t^* and D_t^* in Figure 18.6. The price shown on the vertical axis is the opportunity cost of land converted to agriculture: that is, forgone timber and non-timber benefits. The demand function is of the form:

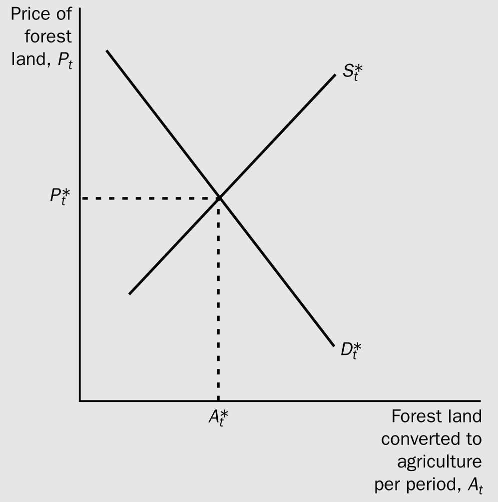


Figure 18.6 The optimal rate of conversion of forested land at time t

$$D = D(P, Y, \text{POP}, Q)$$

where Y is income, POP is the level of population and Q is an index of agricultural yields. Barbier and Burgess expect that $dD/d\text{POP}$ is positive, and so population increases will shift the demand curve rightwards, thus increasing deforestation.

If, however, forest owners are unable to appropriate non-timber benefits, the supply curve will shift to the right relative to that shown in the diagram (which supposes that both timber and non-timber benefits are appropriable by forest owners). Clearly this would also increase the rate of deforestation (by depressing the price of forestland).

We mentioned in the text that the non-timber benefits of tropical forests are received by people throughout the world, not just in the forest vicinities. The benefits are global environmental goods. An interesting attempt to estimate the size of these benefits has recently been made. Kramer and Mercer (1997) used a contingent valuation approach (see Chapter 12) to estimate the size of the one-off monetary payment that US residents would be willing to pay to conserve 5% of tropical forests. Kramer and Mercer's survey responses gave an average value per household of between \$21 and \$31. Aggregated over the US population, this is equivalent to a total single payment of between \$1.9 billion and \$2.8 billion.

summary of the consequences of tropical deforestation and a discussion of its various causes.

It was noted earlier that natural (or primary) forests warrant a very different form of treatment from that used in investigating plantation forestry. Natural forest conversion is something akin to the mining of a resource. These forests represent massive and valuable assets, with a corresponding huge real income potential. While it is conceivable that a forest owner might choose to extract the sustainable income that these assets can deliver, that is clearly not the only possibility. In many parts of the world, as we noted earlier, these assets were converted into income a long time ago. In others, the assets were left almost entirely unexploited until the period after the Second World War. What appears to be happening now is that remaining forest assets are being converted into current income at rates far exceeding sustainable levels.

Where a natural forest is held under private property, and the owner can exclude others from using (or extracting) the forest resources, the management of the resource can be analysed using a similar approach to that covered in Chapter 15 (on non-renewable resources). The basic point is that the owner will devise an extraction programme that maximises the present value of the forest. Whether this results in the forest being felled or maintained in its natural form depends on the composition of the benefits or services the forest yields, and from which of these services the owner can appropriate profits. This explanation is developed further in Box 18.1.

Where private ownership exists, the value of the forest as a source of timber is likely to predominate in the owner's management plans even where the forest provides a multiplicity of socially valuable services. This is because the market mechanism does not provide an incentive structure which reflects the relative benefits of the various uses of the forest. Timber revenues are easily appropriated, but most of the other social benefits of forestry are external to the owner. The signals given to owners by the relative returns to the various forest services lead to a socially inefficient allocation of resources, as we explained in Chapter 5 in discussing the consequences of externalities and public goods. These mechanisms go a long way to explain why the rate of conversion of natural forests is so high, why forestland is often inefficiently converted to other land uses, and why the incentives to replant after clearing are sometimes too low to generate reforestation or to ensure its success.

Our arguments have been premised on the assumption that forestland is privately owned and its use correspondingly controlled. But this analysis is of little relevance in circumstances where forests are not privately owned or where access cannot be controlled. There are two main issues here: the first is the consequence of open-access conditions, and the second is the temptation to 'mine' forests for quick returns.

Many areas of natural forest are *de facto* open-access resources. There is no need to repeat the analysis in Chapter 17 of the consequences of open access for renewable resource exploitation. However, in some ways, the consequences will be more serious in this instance. We argued that open-access fisheries have a built-in defence against stocks being driven to zero: as fish numbers decline to low levels, marginal harvesting costs rise sharply. It usually becomes uneconomic to harvest fish to the point where stock levels have reached critical minimum levels. This does not apply in the case of woodland, however. Trees are not mobile and harvesting costs tend to be affected very little by the stock size. So as long as timber values are high (or the return from other uses of the land is sufficiently attractive), there is no in-built mechanism stopping stock declining to zero. Open access also implies that few individuals are willing to incur the large capital costs in restocking felled timber, particularly when returns are so far into the future.

The second issue we raised above was the temptation of governments and individuals granted tenure of land to convert natural timber assets into current income, or to switch land from forestry to another use which offers quicker and more easily appropriated returns. There is, of course, nothing new about this. It has been happening throughout history, and goes a long way to explaining the loss of natural forest cover in Europe, North Africa and the Middle East. The process is now most acute in tropical forests.

18.7 Government and forest resources

Given the likelihood of forest resources being inefficiently allocated and unsustainably exploited, there are strong reasons why government might choose to intervene in this area. For purely single-use plantation forestry, there is little role for government to play other than guaranteeing property rights so that incentives to manage timber over long time horizons are protected.

Where forestry serves multiple uses, government might use fiscal measures to induce managers to change rotation intervals. It is straightforward to see how this can be done. Well-designed taxes or subsidies can be thought of as changing the net price of timber (by changing either the gross price, P, or the marginal harvest cost, c). We will leave you to deduce what kind of taxes and subsidies would have this effect. In principle, any desired rotation length can be obtained by an appropriate manipulation of the after-tax net price.

Where non-timber values are large and their incidence is greatest in mature forests, no felling may be justified. Government might seek such an outcome through fiscal incentives, but is more likely to do so through public ownership. The most important role for government, though, concerns its policy towards natural forestland. It is by no means clear that public ownership *per se* has any real advantages over private ownership in this case. What matters here is how the assets are managed, and what incentive structures exist.

Finally, we need to give some attention to international issues here. Many of the non-timber values of forest resources are derived by people living not only outside the forest area but also in other countries. Many of the externalities associated with tropical deforestation, for example, cross national boundaries. This implies limits to how much individual national governments can do to promote efficient or sustainable forest use. Internationally concerted action is a prerequisite of efficient or sustainable outcomes. We discussed these issues – including internationally organised tax or subsidy instruments, debt-for-nature swap arrangements and international conservation funds – in Chapter 10.

Summary

- If all markets exist, all the conditions described in Chapter 5 for the efficient allocation of resources are satisfied throughout the economy, and if the interest rate used by private foresters is identical to the social consumption discount rate, privately optimal choices in forestry will be socially efficient, and, given appropriate distributions of initial endowments of property rights, could be socially optimal too.
- These conditions are not likely to be satisfied. Apart from the fact that the 'rest of the economy' is unlikely to satisfy all the necessary efficiency conditions, there are particular aspects of forestry that imply a high likelihood of private decisions not being socially efficient. What are these aspects?
  1. Where forests are privately owned, externalities tend to drive a wedge between privately and socially efficient incentive structures whenever forests serve multiple uses. Forests are multi-functional, providing a wide variety of economic and other benefits. Private foresters are unlikely to incorporate all these benefits into their private net benefit calculations, as they often have very weak or no financial incentives to do so. Non-timber benefits may be very substantial. Where plantation forests are being managed, the presence of these benefits is likely to cause the length of socially optimal rotations to diverge from what is privately optimal.

2. In the case of natural forests, it will also be difficult for whoever has responsibility for land-use decisions to extract appropriate monetary values for these non-timber benefits, particularly when the benefits are received by citizens of other countries. These problems are particularly acute in the case of tropical forests and other open-access woodlands.

- Governments might attempt to internalise externalities by fiscal measures or by the regulation of land use. Alternatively, public ownership of forestland may be used as a vehicle for promoting socially efficient forest management.
- The record of public ownership does not, however, give much cause for confidence that forest policy will be pursued prudently.

Further reading

Excellent reviews of the state of forest resources in the world economy, and experiences with various management regimes, are contained in *World Resources*, published every two years. See, in particular, the sections in WR (1994) and WR (1996). This source also contains an excellent survey concerning trends in biodiversity. Various editions of the *United Nations Environment Programme, Environmental Data Report* also provide good empirical accounts. Extensive references on biodiversity were given in Chapter 17.

A more extensive account of forestry economics (at about the same level as this text), examining the effects of various tax and subsidy schemes, is to be found in Hartwick and Olewiler (1998), chapter 10. Other excellent surveys of the economics of forestry can be found in Anderson (1991), Pearse (1990), Berck (1979) and Johansson and Löfgren (1985). Montgomery and Adams (1995) contains a good account of optimal management, but at a relatively advanced level.

Bowes and Krutilla (1985, 1989) are standard references for multiple-use forestry. Hartman (1976) is an early work in the area, which is also examined in Calish *et al*. (1978), Swallow *et al*. (1990), Swallow and Wear (1993), Pearce (1994) and Vincent and Blinkley (1993).

The value of forests for recreation is analysed by Clawson and Knetsch (1966), Benson and Willis (1991) and Cobbing and Slee (1993), although you should note that these references are primarily concerned with the techniques of valuation of non-marketed goods that we discuss in Chapter 14. Browder (1988) examines the conversion of forest-land in Latin America. The state of tropical and other natural-forest resources, with an emphasis on sustainability and policy, is discussed in Sandler (1993), Barbier and Burgess (1997), Vincent (1992) and Repetto and Gillis (1988). For the effects of acid rain on forests, see CEC (1983) and Office of Technology Assessment (1984).

Tahvonen and Salo (1999) present a synthesis of the Fisher two-period consumption-saving model with the Faustmann model, thereby allowing owner preferences to shape forest management choices.

Discussion questions

1. Is it reasonable for individuals living in Western Europe today to advise others to conserve tropical forests given that the countries in which they live effectively completed the felling of their natural forests centuries ago?
2. Discuss the implications for the harvest rate and possible exhaustion of a renewable resource under circumstances where access to the resource is open, and property rights are not well defined.

3. Discuss the contention that it is more appropriate to regard natural forests as non-renewable than as renewable resources.

4. In what circumstances, and on what criterion, can the conversion of tropical forestry into agricultural land be justified?

5. How will the optimal rotation interval be affected by extensive tree damage arising from atmospheric pollution?

Problems

1. Using a spreadsheet program, calculate the volume of timber each year after planting for a period of up to 130 years for a single unfelled stand of timber for which the age–volume relationship is given by $S = 50t + 2t^2 - 0.02t^3$ (where S and t are defined as in the text of this chapter). Is it meaningful to use this equation to generate stock figures up to this stand age?

   Also calculate:
   (a) The year after planting at which the amount of biological growth, $G(S)$, is maximised.
   (b) The present-value-maximising age for clear felling (assuming the stand is not to be replanted) for the costs and prices used in Table 18.3 and a discount rate of 5%.

   (We suggest that you attempt to construct your own spreadsheet program to answer this question. If you find that this is not possible, you can obtain the answers by adapting *Sheet 4* in *Chapter18.xls*.)

2. Demonstrate that a tax imposed on each unit of timber felled will increase the optimal period of any rotation (that is, the age of trees at harvesting) in an infinite-rotation model of forestry. What effect would there be on the optimal rotation length if the expected demand for timber were to rise?

3. How would the optimal rotation interval be changed as a result of
   (a) an increase in planting costs;
   (b) an increase in harvesting costs;
   (c) an increase in the gross price of timber;
   (d) an increase in the discount rate;
   (e) an increase in the productivity of agricultural land?

4. The following three exercises require that you use the Excel file *palc18.xls*.
   (a) Calculate the optimal rotation lengths for a single-rotation forest for the interest rates 1, 2, 4, 5 and 6%. These should match those shown in Table 18.4.
   (b) Calculate the interest rate above which the PV of the forest becomes negative for *any* rotation length in a single rotation forest. Do the same for an infinite-rotation forest.
   (c) Identify what happens to the gap between the optimal rotation lengths in single- and infinite-rotation models as the interest rate becomes increasingly large (beginning from 0%). Explain the convergence that you should observe. What happens to the PV of the forest at this convergence?

5. The Excel workbook *Non Timber.xls* (see *Additional Materials*) models the consequences of including non-timber values in a single-rotation forest model. The first sheet – *Parameter values* – defines various parameter values, and gives three alternative sets of non-timber present values. Results of the computations are shown in *Sheet 1*. Examine how the inclusion of non-timber benefits alters the optimal stand age at which felling takes place. Does the change vary from one set of non-timber values to another? Do your conclusions differ between the cases where the discount rate is 2% and 4%?

Subhes C. Bhattacharyya

Energy Economics

Concepts, Issues, Markets and Governance

Springer

Chapter 3
Understanding and Analysing Energy Demand

3.1 Introduction

The term "energy demand" can mean different things to different users. Normally it refers to any kind of energy used to satisfy individual energy needs for cooking, heating, travelling, etc., in which case, energy products are used as fuel and therefore generate demand for energy purposes. Energy products are also used as raw materials (i.e. for non-energy purposes) in petrochemical industries or elsewhere and the demand for energy here is to exploit certain chemical properties rather than its heat content.

Similarly, the focus may be quite different for different users: a scientist may focus on equipment or process level energy demand (i.e. energy used in a chemical reaction) while planners and policy-makers would view the aggregate demand from a regional or national point of view. Energy demand can correspond to the amount of energy required in a country (i.e. primary energy demand) or to the amount supplied to the consumers (i.e. final energy demand). Often the context would clarify the meaning of the term but to avoid confusion, it is better to define the term clearly whenever used.

A distinction is sometimes made between energy consumption and energy demand. Energy demand describes a relationship between price (or income or some such economic variable) and quantity of energy either for an energy carrier (e.g. electricity) or for final use (such as cooking). It exists before the purchasing decision is made (i.e. it is an ex ante concept—once a good is purchased, consumption starts). Demand indicates what quantities will be purchased at a given price and how price changes will affect the quantities sought. It can include an unsatisfied portion but the demand that would exist in absence of any supply restrictions is not observable. Consumption on the other hand takes place once the decision is made to purchase and consume (i.e. it is an ex post concept). It refers to the manifestation of satisfied demand and can be measured. However, demand and consumption are used interchangeably in this chapter despite their subtle differences.

S. C. Bhattacharyya, *Energy Economics*, DOI: 10.1007/978-0-85729-268-1_3,

Energy demand is a derived demand as energy is consumed through equipment. Energy is not consumed for the sake of consuming it but for an ulterior purpose (e.g. for mobility, for producing goods and services, or for obtaining a certain level of comforts, etc.). Need is specific with respect to location, technology and user. The derived nature of demand influences energy demand in a number of ways (discussed below), which in turn has influenced the demand analysis by creating two distinct traditions—one following the neoclassical economic tradition while the other focusing on the engineering principles coupled with economic information (Worrel et al. 2004).

This chapter intends to provide a basic understanding of various concepts related to energy demand and show how energy demand could be analysed using simple tools covering both the traditions indicated above.

3.2 Evolution of Demand Analysis

Prior to the first oil shock, the energy sector had a supply-oriented focus where the objective was to meet a given exogenous energy demand by expanding the supply. Since early 1970s, when energy caught the attention of policymakers because of sudden price increases, the research on energy has grown significantly in size. From a level of limited understanding of the nature of demand and demand response due to presence of external shocks (Pindyck 1979) and energy system interactions, there has been a significant build-up of knowledge. Energy models were however not developed for the same purpose—some were concerned with better energy supply system design given a level of demand forecast, better understanding of the present and future demand–supply interactions, energy and environment interactions, energy-economy interactions and energy system planning. Others had focused on energy demand analysis and forecasting.

In the three decades that followed since the first oil shock, the energy sector has experienced a wide range of influences that have greatly influenced energy analysis and modelling activities (Worrel et al. 2004; Laitner et al. 2003):

Firstly, the rise in concerns about global warming which required a very long term understanding of the implications of energy use. This has led to the developments in very long term analysis covering 50–100 years.

Secondly, due to the changes in the market operations with the arrival of competitive market segments in various energy industries, especially in the case of electricity, the focus has shifted to short-term analysis, covering hours or days, essentially for operational purposes.

Thirdly, there are growing concerns about future security of fuel supplies and large capacity expansion needs globally. This is evident from the European decision to create an Energy Market Observatory and the UK decision indicated in the White Paper on Energy in 2007 to establish its own energy data observatory (DTI 2007). The twin concerns of the day, namely that of security of energy supply and environmental concerns of energy use, are contributing to a paradigm shift (Helm, 2005), which in turn is fuelling a closer look at the energy infrastructure development both in developed and developing countries either for replacing age-old, sweated assets or for meeting new demand.

We can add a fourth influence as well—that of vast improvements in computing and communications facilities. The emergence of low cost computing and internet facilities has dramatically changed the data processing and analytical capabilities.

Energy projects tend to be capital intensive and often require long lead time. For example, a thermal power station may need 3–4 years to build, a nuclear or hydro power station requires typically 7–10 years, if not more, and a refinery project can easily take 2–4 years. Given the long gestation period of energy investments and diversity of technologies as well as economic conditions of countries and consequent constraints on the analytical choices, medium to long-term analysis is essential for energy system-related decisions.

Moreover, mobilizing resources for energy projects is not always easy. Hence, correct timing of supply capacity additions is important, for which correct demand projection is a pre-requisite. Lumpiness of investment implies that for such projects huge sums of capital are tied up in advance and no return or output is obtained until the project is completed. Consequently, the decision-makers have to form a view about the future well in advance and plan for new projects and actions. The decision-making depends to a large extent on demand forecasting and misjudgements can lead to costly gaps or equally costly over capacities.

In this respect, developing countries have certain distinct features. Bhatia (1987) indicated a number of difficulties experienced in analyzing the energy demand of developing countries as follows:

(a) Data on traditional energies used widely in rural areas may be lacking and may have to be estimated.
(b) Many poor consumers lacking purchasing power may not enter the commercial energy ladder but over time a shift to commercial energies takes place. This needs to be captured.
(c) Supply shortage in many developing countries implies that consumption may not represent the actual demand due to the existence of unfulfilled demand.
(d) The availability and consumption of commercial energy may be greatly influenced by a few large consumers.
(e) Response to price changes is more difficult to assess due to "difficulties of obtaining complimentary non-energy inputs, the absolute shortages of certain fuels, imperfect product and capital markets."

Moreover, the demand for commercial energies tends to grow faster here compared to the developed countries and generally, the demand for liquid fuels grows faster than other fuels, suggesting a shift from solid to liquid fuels. As an adequate supply of energy is vital for the smooth running of a country, and because of long lead times and capital intensive nature of energy projects, developing countries need to analyze the past trends to forecast the likely paths of energy demand growth.

At the other end of the spectrum, there is need for forecasts for day-to-day operation and management of energy systems. How much electricity needs to be generated next hour or tomorrow? This information forms the basis for unit commitment exercise (i.e. to find out which plants should be used to produce the required electricity so that the operating cost is minimised). Similarly, forecasts for six months to one year are required for business planning purposes, for regulatory approvals, to assess the prospects of the business in the coming year, etc. Thus, short term forecasting is also important in addition to medium to long-term analysis and forecasting.

More recently, in the late 1980s, the emphasis was increasingly on sustainable development. Some of the problems associated with energy use, such as the possibility of global warming, have long-term implications and any strategy to deal with them has to be seen in a long-term context. Very long-term demand forecasts for more than 50 years have also become necessary.

3.3 Overview of Energy Demand Decisions

Energy is not consumed for the sake of it but is used for satisfying some need and is done through use of appliances. Any commercial energy requires monetary exchanges and the decision to switch to commercial energies can be considered as a three-stage decision-making process [see Hartman (1979), Stevens (2000) and Bhattacharyya (2006)].

- First, the household has to decide whether to switch or not (i.e. switching decision).
- Second, it decides about the types of appliances to be used (i.e. appliance selection decision).
- In the third stage, consumption decision is made by deciding the usage pattern of each appliance (i.e. consumption decision). All these stages influence energy demand. This is shown in Fig. 3.1.

As Hartman (1979) indicates, any demand analysis and forecasting should consider this three stage decision-making process and capture related policy variables so that interventions, if required, could be properly designed. There are two decision outcomes at the first stage: to purchase an energy consuming appliance or not to purchase. If appliance is not purchased, demand for that particular use does not arise for that consumer. The switching decision is largely determined by monetary factors: the amount and regularity of money income, alternative uses of money and willingness to spend part of the income to consume commercial energies as opposed to allocating the money to other competing needs.

Once a buying decision is made, two important parameters are to be decided next. If alternative fuel choices are available, which fuel would be used and what type of appliance for this fuel? Once a decision is made to buy an appliance and the appliance is purchased, the only variable leaves in the hand of the user is its

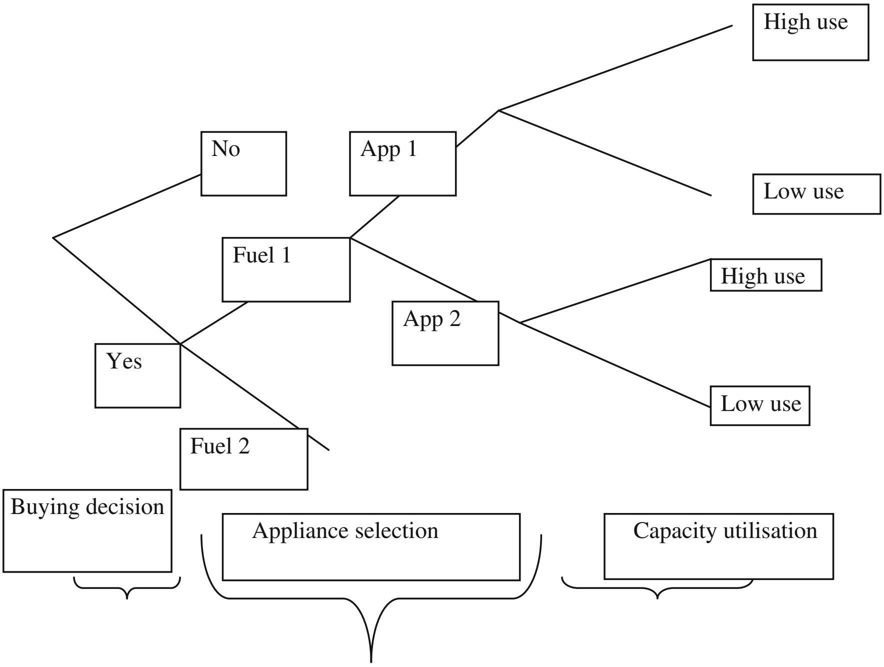


Fig. 3.1 Three-stage decision-making. *Source* Author

utilisation. The level of utilisation varies from consumer to consumer and consumers can adjust utilisation in response to changes in external factors. Box 3.1 provides the implications of each stage of decision making on the demand analysis.

Box 3.1: Implication of the Three-Stage Decision-Making on Energy Demand

For energy demand, information related to appliance holding pattern is important for two reasons:

(1) to understand consumption behaviour: If there is lack of interest in a particular use, it may be that there are important barriers which need to be looked into. These barriers include: cost, financing options, user friendliness, etc.
(2) to understand growth potential: If a particular segment of market is saturated, demand growth from new consumers would be less and vice versa.

Appliance stock and its growth potential are important determinants of demand. For example, in a developing country there is only one car per hundred thousand people. If the government provides cheap gasoline to promote energy access, would it work? The cheap gasoline would go to those having cars and would not benefit the rest. The barriers to owning car need to be looked into first to promote motorized transport.

The second stage has a deciding influence on demand. Often equipment has a long life time (5–10 years) and is costly. Once an appliance is purchased, it will be in operation for sometime. This introduces strong path dependence in energy demand (meaning that the choice of appliance forecloses certain options and influences the demand path). Strong path dependence affects fuel switching possibility and responsiveness of the consumers to external changes. Fuel switching option would be limited by the appliance choice decision and involves capital expenditure, at times of considerable amounts. Limited responsiveness: The rigidity or strong path dependence leaves limited options to consumers in the event of sudden changes in prices or supply conditions in the short run. They have to depend on their existing stock of appliances in any case. The full reaction to external changes is not instantaneous. It is spread over a number of periods because of the rigidity of the system. This process is called lagged reaction (i.e. the reaction lags behind the action) and only over a number of period, the accumulated effect gives the full reaction.

The short term response arises from this factor and its scope is not very broad. Therefore, short-term response is quite limited. This can have a social dimension as low capacity utilisation may lead to deprivation of essential energy services.

The three-stage decision process therefore influences: access to energy services, market growth potential in a particular service or use, path dependence, responsiveness in the short run, reaction response, and consumer's usage behaviour. The above discussion also suggests that technology matters: because energy demand is dependent on technical efficiency, substitution possibility depends on technical options available.

3.4 Economic Foundations of Energy Demand[1]

From the point of view of economics, the principle for estimating and analyzing the demand for energy is not different from that for any other commodity. There

[1] This section relies on Bohi (1981), Chapter 2, Estimating the demand for energy: Issues and Methodologies. Similar treatments are also provided in Hartman (1979), Munasinghe and Meier (1993)

are characteristics of energy demand, institutional features of energy markets, and problems of measurement that require particular attention in analyzing energy markets. But the microeconomic foundation of energy demand is same as for other commodities.

Demand for energy can arise for different reasons. Households consume energy to satisfy certain needs and they do so by allocating their income among various competing needs so as to obtain the greatest degree of satisfaction from total expenditure. Industries and commercial users demand energy as an input of production and their objective is to minimize the total cost of production. Therefore the motivation is not same for the households and the productive users of energy and any analysis of energy demand should treat these categories separately.

From basic microeconomic theory, the demand for a good is represented through a demand function which establishes the relation between various amounts of the good consumed and the determinants of those amounts. The main determinants of demand are: price of the good, prices of related goods (including appliances), prices of other goods, disposable income of the consumer, preferences and tastes, etc. To facilitate the analysis, a convenient assumption (known as ceteris paribus) is made which holds other determinants constant (or unchanged) and the relation between price and the quantity of good consumed is considered. This simple functional form can be written as follows:

$q = f(p)$, where q is the quantity demanded and p is the price of the good. The familiar demand curve is the depiction of the above function.

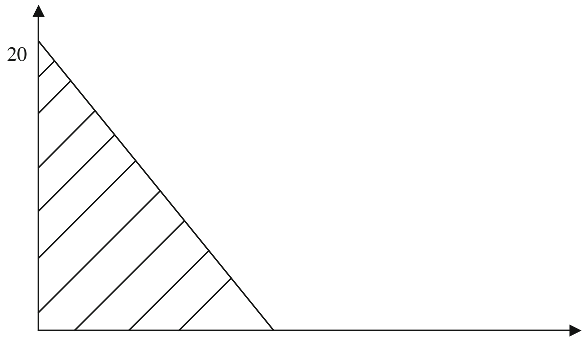
3.4.1 Consumer Demand for Energy: Utility Maximization Problem

The microeconomic basis for consumer energy demand relies on consumers' utility maximization principles. Such an analysis assumes that

- Consumers know their preference sets and ordering of the preferences.
- Preference ordering can be represented by some utility function and
- The consumer is rational in that she will always choose a most preferred bundle from the set of feasible alternatives.

Following consumer theory, it is considered that an incremental increase in consumption of a good, keeping consumption of other goods constant, increases the satisfaction level but this marginal utility (or increment) decreases as the quantity of consumption increases. Moreover, maximum utility achievable given the prices and income requires marginal rate of substitution to be equal to the economic rate of substitution. This in turn requires that the marginal utility per dollar paid for each good be the same. If the marginal utility per dollar is greater for good A than for good B, then transferring a dollar of expenditure from B to A will increase the total utility for the same expenditure. It follows that reduction in the relative price of good A will tend to increase the demand for good A and vice versa.

Fig. 3.2 Budget constraint



We shall use this basic idea in a graphical example to explain how the consumer demand curve for energy could be developed. The mathematical development is provided in Annex 3.1 for interested readers.[2]

Assume that an individual has 100 dollars to allocate between energy E and other goods X. One unit of energy costs 5 dollars while one unit of other goods costs 20 dollars. Accordingly, the individual can buy 20 units of energy or 5 units of other goods or a combination of these goods as shown by the shaded area of Fig. 3.2.

$$\text{In equation form this is written as } 100 = 5E + 20X \tag{3.1}$$

$$\text{Consider a utility function } U = X^{0.5}E^{0.5} \tag{3.2}$$

The combinations of X and E for various levels of utility (e.g. $U = 2$, 3, 4 and 5) can be easily determined for this function (see Fig. 3.3). These curves are called indifference curves. The optimal demand for energy and other commodities could be determined for the given individual from the budget line and the indifference curves (see Fig. 3.3).

The budget line is tangent to the indifference curve ($U = 5$) and the optimal combinations of energy and other goods can be found from this (which turns out to be 10 units of energy and 2.5 units of other goods). Hence, when the energy price is 5 per unit, given the budget constraint, the individual consumes 10 units of energy. This forms one pair of data set for his/her demand curve.

Now consider that the price of energy changes to 10 per unit while the price for other goods remains unchanged. Naturally, the consumer now will be able to consume only 10 units of energy or 5 units of other goods or some combinations of energy and other goods (as shown in Fig. 3.4). Following the method indicated above, the new optimal combination is found and in this particular case, the individual would consume 5 units of energy and 2.5 units of other goods (i.e. just

[2] See also Chapter 2 of Bohi (1981), Munasinghe and Meier (1993) and Medlock III (2009).

Fig. 3.3 Utility maximisation

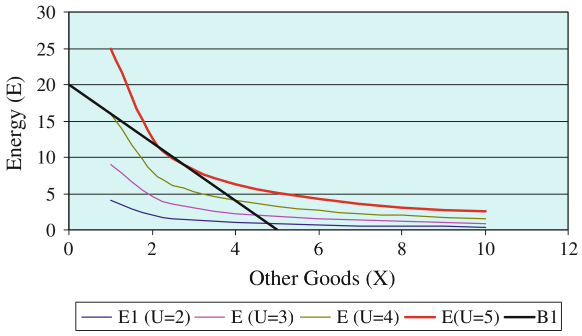


Fig. 3.4 Effect of changes in energy price on the budget line

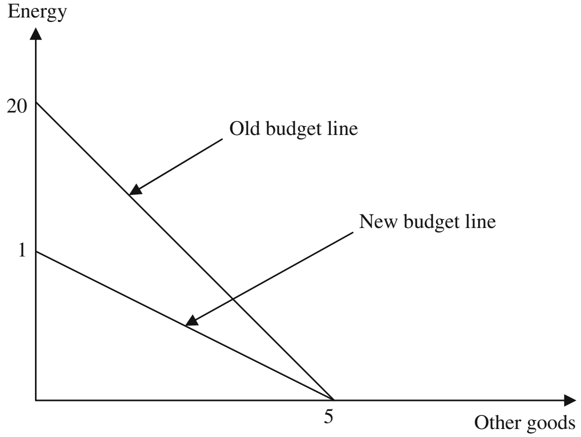
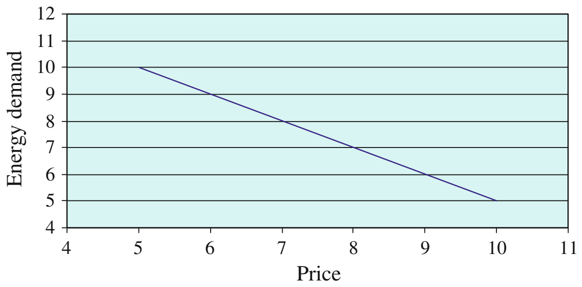


Fig. 3.5 Energy demand curve of an individual



50% reduction of energy demand). This gives another pair of points on the demand curve.

The individual's energy demand schedule can now be drawn using these points (see Fig. 3.5). As you have noticed, in the entire process, we have only changed energy prices while keeping other variables unchanged (i.e. assumed that ceteris paribus condition holds). In Fig. 3.5, the demand curve is downward sloped as is expected.

The market demand function for a particular good is the sum of each individual's demand for that good. The market demand curve for the good is constructed from the demand function by varying the price of the good while holding all other determinants constant.

3.4.2 Cost Minimization Problem of the Producer

In the case of producers, the theory of the producers is used to determine the demand for factors of production. In the production process, it is normally possible to replace one input by the other and the producer would try to find the combination of inputs that would minimize the cost of production. Once again, we use a graphical approach for the general description, while a more mathematical presentation is given in Annex 3.2.

Consider that a producer uses capital and energy to produce her output which follows the production function given in Eq. 3.3.

$$Q = 10K^{0.5}E^{0.5} \tag{3.3}$$

The isoquant map for this production function can be graphed by setting Q at different levels (say 50 or 100) and then finding the combinations of K and E that would produce the given level of outputs (see Fig. 3.6).

Assume that the price of capital and energy per unit is \$1 each. If K units of capital and E units of energy are used in the production process, the total cost will be $K + E$. The cost lines are shown as constraints in Fig. 3.6. As can be seen from the figure, the optimal choice would be at the point where the cost line is tangent to the isoquant. For a given level of output, the demand for input energy can then be determined.

While the above theoretical concepts provide some understanding of energy demand, these theoretical ideas are based on quite restrictive assumptions. While the econometric modelling tradition explicitly follows the economic principles for energy demand analysis and forecasting purposes, this is not the only economic philosophy followed in energy demand modelling. Although price, rationality and optimising behaviour within the neoclassical tradition greatly influence the econometric tradition, others do not always believe in the crucial role of these

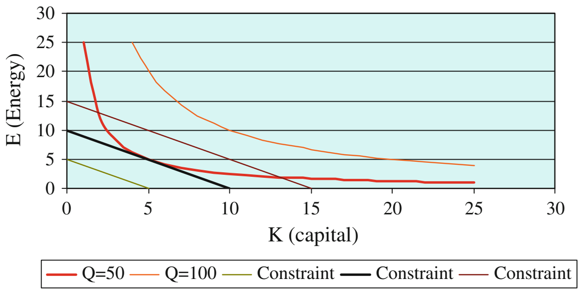


Fig. 3.6 Optimal input selection for the firm

factors. Accordingly, other behavioural assumptions (such as “satisficing” approach in the sense of Herbert Simon or evolutionary approach for technological change) and beliefs are used in some approaches,[3] especially in the “bottom-up” approach or “engineering-economic” approach.

3.5 Alternative Approaches for Energy Demand Analysis

Analysis of the historical evolution of energy demand and its interpretation is an essential part of energy demand analysis. Such an analysis allows identification of the underlying factors affecting energy demand. Various analytical methods are used to analyze energy demand. Three approaches are presented below: simple descriptive analysis, factor (or decomposition) analysis, and econometric analysis.

3.5.1 Descriptive Analysis[4]

Here we present three simple but commonly used indicators that are used to describe the change in demand or its relationship with an economic variable. These are growth rates, demand elasticities and energy intensities.

Any demand analysis starts with a general description of the overall energy demand trends in the past. It enables qualitative characterization of the pattern of energy demand evolution and identification of periods of marked changes in the demand pattern (such as ruptures, inflexions, etc.). This preliminary step could set the scope and the priorities of the analysis (see Fig. 3.7). Such a historical analysis is first based on a graphical presentation of the evolution of demand through time. Two types of graphs are generally used:

- energy demand in absolute value (Mtoe, PJ, etc.) and time;
- energy demand in index and time.

The graph in absolute value provides an indication of the trend while that in index allows comparison with respect to the base year. Index also allows comparison of trends of different fuels and energy groups.

3.5.1.1 Growth Rates

Annual growth rate is another indicator commonly used to describe the trend. This can be on an annual basis or an average over a period. Table 3.1 presents the

[3] See Wilson and Dowlatabadi (2007).

[4] This section is based on UN (1991). See also IEA (1997).

Fig. 3.7 Evolution of global primary energy demand. *Data source* BP Statistical Review of World Energy 2009

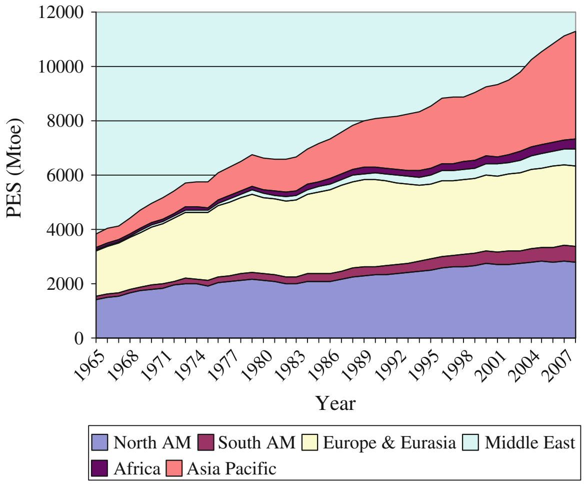


Table 3.1 Mathematical relationships for simple indicators of trend

Indicator	Formula	Parameter description
Year-on-year growth rate	$a = (E_{t+1} - E_t)/E_t$	Where a = annual growth in demand, E_{t+1} = energy consumption in year $t + 1$ and E_t = energy demand in year t
Annual average growth rate over a period	$E_{T1} = E_{T0}\left(1 + a_g\right)^{(T1-T0)}$ $a_g = \left(\frac{E_{T1}}{E_{T0}}\right)^{1/(T1-T0)} - 1$	Where E_{T1} = energy demand in period $T1$ and E_{T0} = energy demand in period $T0$, a_g = annual growth rate
Demand elasticities	$e_t = \frac{(\Delta EC_t/EC_t)}{(\Delta I_t/I_t)}$	Where t is a period given EC is energy consumption I is the driving variable of energy consumption such as GDP, value-added, price, income etc. Δ is the change in the variable
Energy intensity (for a single energy)	$\mathrm{EI}_t = \frac{E_t}{I_t}$	EI_t = energy intensity for year t, E_t = energy consumption in year t and I_t = value of the driving variable (say GDP or value added)
Energy intensity in case of aggregated fuels	$\mathrm{EI}_t = \frac{\sum_{i=1}^{n} E_{it}}{I_t}$	Where E_{it} = energy consumption of ith type of fuel in year t

formula commonly used for this purpose. The year-on-year growth rates are calculated year after year so as to get a historical series. The average growth rate over a period on the other hand provides a picture for the entire period. Although an arithmetic average of the annual year-on-year growth rates can be calculated, this is not done generally. Instead, a geometric average is calculated for the period. Annual growth rates can also be calculated at any level of disaggregation. This is an easily understood indicator capturing the speed of change in demand.

Example According to BP Statistical Review of World Energy, the world primary energy consumption was 9,262.6 Mtoe in 2000. The demand increased to 11,104.4 Mtoe in 2007 and 11,294.9 Mtoe in 2008. Calculate the growth rate of demand between 2007 and 2008. Also calculate the annual average growth rate between 2000 and 2008.

Answer: The primary energy demand increased from 11,104.4 Mtoe in 2007 to 11,294.9 Mtoe in 2008. This amounts to a growth of = (11,294.9 − 11,104.4)/ 11,104.4 = 0.017 or 1.7%.

The annual average growth rate between 2000 and 2008 is = (11,294.9/ 9,262.6)^(1/8) − 1 = 0.0251 or 2.51%.

3.5.1.2 Demand Elasticities

Elasticities measure how much (in percent) the demand would change if the determining variable changes by 1%. In any economic analysis, three major variables are considered for elasticities: output or economic activity (GDP), price and income. Accordingly, three elasticities can be determined. The general formulation is given in Table 3.1. There are two basic ways of measuring elasticities: using annual growth rates of energy consumption and the driving variable, or using econometric relationships estimated from time series data. The first provides a point estimate while the second provides an average over a period, and accordingly, the two will not give the exactly same result.

Output or GDP elasticities of energy demand indicate the rate of change of energy demand for every 1% change in economic output (GDP or value added). Normally the GDP growth is positively related to energy demand but the value of elasticity varies depending on the stage of development of an economy. It is normally believed that the developed countries tend to have an inelastic demand with respect to income (i.e. the elasticity less than 1) while developing countries have an elastic energy demand with respect to income.

Example The primary energy consumption in China increased from 1,970 Mtoe in 2004 to 2,225 Mtoe in 2005. The GDP increased from 14,197 Billion Yuan in 2004 to 15,603 Billion Yuan in 2005 at constant 2,000 prices. What was the GDP elasticity of energy demand in China?

$$\% \text{ change in energy demand} = (2,225 - 1,970)/1,970 = 12.9\%$$

$$\% \text{ change in GDP} = (15,603 - 14,197)/14,197 = 9.9\%$$

$$\text{GDP elasticity} = 12.9/9.9 = 1.31$$

Price elasticities indicate how much demand changes for every percent change in the energy price. Price elasticities are negative numbers, indicating that an increase in price results in a decrease in energy demand. As this elasticity aims to find out the responsiveness of consumers to price changes, the price to be used for elasticity purposes should reflect as closely as possible what consumers really pay (retail price or wholesale price as the case may be). A distinction is normally made between short-term and long-term price elasticities. The short-term price elasticity captures the instantaneous reaction to price changes. In the short run consumers do not have the possibility to change their capital stock and can only change their consumption behaviour and hence only a partial reaction is normally felt. The long-term elasticity would capture the effect of adjustments over a longer period. On the other hand, over the long run, consumers have the possibility of adjusting their capital stock as well as their consumption behaviour. This results in a better reflection of the reaction to price change.

3.5.1.3 Energy Intensities

Energy intensities (also called energy output ratios) measure the energy requirement per unit of a driving economic variable (e.g. GDP, value added, etc.). Energy consumption may refer to a particular energy or to various energy aggregates and is expressed as a ratio of energy demand per unit of economic output (see Table 3.1 for the formula). For an economic driving variable, normally the constant dollar values are used for better comparability across a time scale. Table 3.2 explains the choice of the economic driving variable (I) in different cases. In the productive sectors (industry, agriculture and commercial), the value-added of these sectors should be used to calculate their energy intensity. As for the case of non-productive energy consuming sectors such as household sector, the GDP of the whole country or the private consumption of the households should be used as driving economic variable. As the energy consumption of the transport sector includes the consumption of all vehicles, it is then irrelevant to use only the value-added of the transport companies to calculate the transport sector's energy intensity. Instead, GDP as the economic indicator is more appropriate for calculating transport energy intensity.

Although energy intensity (or energy GDP ratio) is widely used as a measure of relative performance of economies, the ratio is subject to various conceptual and measurement problems. The ratio is highly sensitive to the bases chosen for either of its components and any problem that may distort the size of either the numerator

Table 3.2 Selection of driving economic variable by sector

Sector	Driving economic variable
Whole country	GDP
Industry	Value-added of industry sector
Agriculture	Value-added of agriculture sector
Commercial	Value-added of commercial sector
Transport	GDP
Households	GDP or private consumption

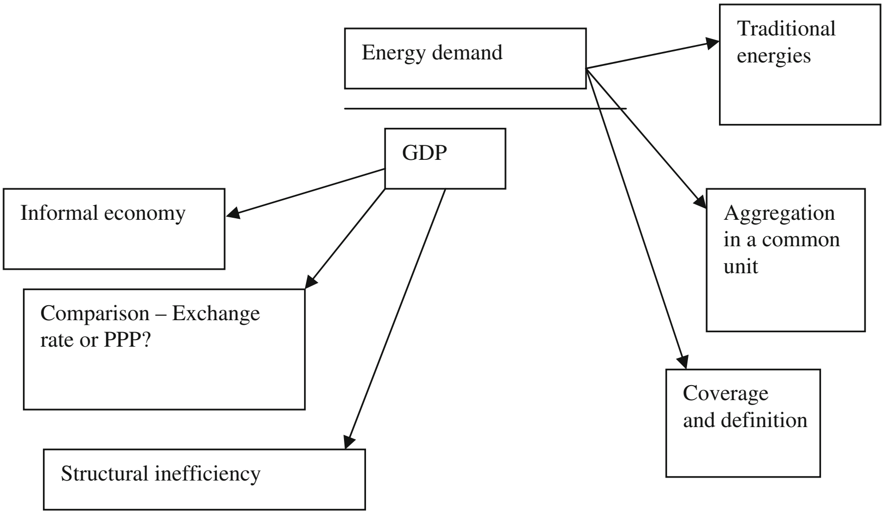


Fig. 3.8 Issues related to energy intensity

(energy consumption) or the denominator (GDP) distorts the picture presented by the ratio (see Fig. 3.8).[5]

The Gross Domestic Product measures the total output of a country's economy. This aggregate statistic represents all goods produced and services rendered within the political boundaries of a country. The GDP can be measured in three standard ways:

(1) by industrial origin, summing up value added by all industries (i.e. gross output minus input);
(2) by summing up the remuneration accruing to all income-producing sectors of the economy; and
(3) by summing up final expenditures to different sectors, that is, presenting aggregated final demand.

The problems related to GDP as a measure of output are:

The measure may be understated by the existence of an underground or informal economy, whose transactions may not be captured by national statistics. This is particularly true of developing countries where many transactions do not get reported in market statistics as they do not enter the market system.

Expenditure on various items may not represent efficient behaviour. In fact, inefficiency would try to increase expenditure and therefore increase GDP when

[5] This discussion is based on Chapter 3, Energy Demand and Economic Growth, Measurement and Conceptual Issues in Policy Analysis, by C. M. Siddayao, West View Press, 1986.

expenditure is used to measure GDP. The GDP statistics may obscure the structural inefficiencies of an economy.

For international comparisons, conversion of the GDP to a common unit is required. The use of foreign exchange rates is the obvious approach. Commonly, exchange rate is used to convert local currency GDP to US$. This faces two problems:

Currency values fluctuate but fluctuations in the exchange rate of a particular currency may not necessarily be related to real changes occurring in the domestic economy.

Exchange rates reflect only the values of internationally traded goods and services and not the entire economic price structure of the reference country.

Depending on the case, the GDP will be understated or overstated in the foreign currency and would lead to distorted intensity.

Studies suggested that purchasing power of low income countries was systematically greater than that suggested by their exchange rates when compared to the purchasing power/exchange rate relationships of high-income countries. For these reasons, various international organizations (World Bank, for example) use another measurement of GDP calculated by converting the national currencies into US dollars with "Purchasing Power Parities (PPP)". The PPP Values are based on a comparison of the purchasing power of a typical "basket" of goods and services, characteristic of each country's consumption pattern.

Problems related to measurement of energy consumption also affect energy intensity estimation. Common issues related to energy measurement are:

- Use of traditional energies in developing countries, data for which is often not accurate and not included in analysis; Exclusion of traditional energies can understate energy consumption and accordingly, energy intensity significantly.
- Aggregation of energies to a common unit can be a problem in itself. Simple summation of heat content of energies does not capture the factors that influence the choice of energy forms. Moreover, such aggregation reflects total energy content rather than available energy. As end-use efficiencies of appliances are different for different forms of energies, such an aggregation is biased towards inefficient technologies.
- Aggregation of hydropower, nuclear power, solar and other renewable energies also poses another problem. The amount of energy is measured only at the output end and not for inputs. But other fossil fuels are measured at the input and output ends. In order to measure hydropower, nuclear power and other such renewable energies on a comparable basis, an assumption has to be made about the amount of fossil fuel input that would be required to provide the same energy. Energy accounts however are not always presented using the production equivalence approach.
- The definition and coverage of energy forms and energy consuming sectors are not same and even within a country can vary from time to time. For comparability, comparable definitions and coverage are required.

Different end-use efficiencies of appliances complicate the problem further. Countries with different fuel mix and appliance use pattern cannot be appropriately compared using toe (oil equivalent values) as the measure of energy consumption. For example, one country relies on coal and traditional energies (e.g. India) to meet its energy demand, while another country is more dependent on natural gas for its needs. If the energy intensity of these two countries is compared using the standard energy intensity approach, the differences in the efficiencies of fuel utilisation will not be captured. A remedy for this problem is to introduce the concept of "oil replacement value", which expresses various fuels in terms of the quantity of oil products that would provide the same amount of useful energy (i.e. same energy service). This approach attempts to measure the effective output of useful work at the downstream end of the energy consumption process and considers the efficiency of the energy utilization equipment. As an example, if the efficiency of the end-use equipment using traditional fuels is 25% of that of oil-using equipment, then 4 toe of traditional fuels would be required to produce the same useful energy provided by1 toe (i.e. 1 ton of oil replacement). Similarly, if the relative efficiency of coal-using equipment is 40% of that of oil-using equipment, 2.5 toe of coal would be equal to 1 tor.

While the replacement value takes care of the differences in energy forms, its principal shortcomings are that it is site-specific (e.g. depends on the type of energy and appliance used) for the resource in question and is time-specific for both resource and technology (i.e. the factors used can change over time, making inter-temporal comparisons difficult). For international comparisons, this approach requires additional information (e.g. appliance efficiency), which may not be readily available.

3.6 Factor (or Decomposition) Analysis

The simple indicators discussed earlier capture the nature of the change in energy demand or use but do not explain the underlying cause. However, for a better understanding of energy use and future energy requirements, it is important to understand the causal factors. A large volume of literature has developed on devising methods and frameworks for explaining the demand. A particular method, known as decomposition method, has been widely used (see Ang and Zhang 2000 for a survey of application of this method).[6] Traditionally, these methods try to identify changes in energy demand arising from a number of factors, the commonly used ones are: changes in economic activity (the activity effect), changes in technological efficiency of energy use at the sector level (the intensity effect) and changes in the economic structure (the structural effect). The

[6] Also see ODYSSEE project for energy efficiency indicators in Europe (http://www.odyssee-indicators.org/). IEA (1997) also presents a large study for IEA Member countries.

two distinct traditions in the energy field: one is known as top-down approach where the focus remains on the aggregate level of analysis and the other is known as bottom-up approach where the overall demand is aggregated from the sector and sub-sector level analysis. The next chapter deals with this aspect in some detail.

Annex 3.1: Consumer Demand for Energy—The Constrained Optimization Problem

Consider that the utility function of a consumer can be written as

$$\text{Utility}\, u = U(X_1, X_2, X_3, \ldots, X_n) \tag{3.31}$$

The consumer has the budget constraint

$$I = p_1X_1 + p_2X_2 + \cdots + p_nX_n \tag{3.32}$$

For maximization of the utility subject to the budget constraint, set the lagrange

$$L = U(X_1, X_2, X_3, \ldots, X_n) - \lambda(I - (p_1X_1 + p_2X_2 + \cdots + p_nX_n)) \tag{3.33}$$

Setting partial derivatives of L with respect to X_1, X_2, X_3,…X_n and λ equal to zero, $n + 1$ equations are obtained representing the necessary conditions for an interior maximum.

$$\begin{aligned} \delta L/\delta X_1 &= \delta U/\delta X_1 - \lambda p_1 = 0; \\ \delta L/\delta X_2 &= \delta U/\delta X_2 - \lambda p_2 = 0; \\ &\vdots \\ \delta L/\delta X_n &= \delta U/\delta X_n - \lambda p_n = 0 \\ \delta L/\delta \lambda &= I - p_1X_1 + p_2X_2 + \cdots + p_nX_n = 0 \end{aligned} \tag{3.34}$$

From above,

$$(\delta U/\delta X_1)/(\delta U/\delta X_2) = p_1/p_2 \text{ or MRS} = p_1/p_2 \tag{3.35}$$

$$\lambda = (\delta U/\delta X_1)/p_1 = (\delta U/\delta X_2)/p_2 = \cdots = (\delta U/\delta X_n)/p_n \tag{3.36}$$

Solving the necessary conditions yields demand functions in prices and income.

$$\begin{aligned} X_1^* &= d_1(p_1, p_2, p_3, \ldots p_n, I) \\ X_2^* &= d_2(p_1, p_2, p_3, \ldots p_n, I) \\ &\vdots \\ X_n^* &= d_n(p_1, p_2, p_3, \ldots p_n, I) \end{aligned} \tag{3.37}$$

An individual demand curve shows the relationship between the price of a good and the quantity of that good purchased, assuming that all other determinants of demand are held constant.

Annex 3.2: Cost Minimization Problem of Producers

Consider a firm with single output, which is produced with two inputs X_1 and X_2. The cost of production is given by

$$\mathrm{TC} = c_1X_1 + c_2X_2 \tag{3.38}$$

This is subject to

$$\mathrm{St}\ q_0 = f(X_1, X_2) \tag{3.39}$$

Write the Lagrangian expression as follows:

$$L = c_1X_1 + c_2X_2 + \lambda(q_0 - f(X_1, X_2)) \tag{3.40}$$

The first order conditions for a constrained minimum are:

$$\delta L/\delta X_1 = c_1 - \lambda\, \delta f/\delta X_1 = 0$$
$$\delta L/\delta X_2 = c_2 - \lambda \delta f/\delta X_2 = 0$$

From above,

$$c_1/c_2 = (\delta f/\delta X_1)/(\delta f/\delta X_2) = \mathrm{RTS}\,(X_1 \text{ for } X_2) \tag{3.41}$$

In order to minimize the cost of any given level of input, the firm should produce at that point for which the rate of technical substitution is equal to the ratio of the inputs' rental prices.

The solution of the conditions leads to factor demand functions.

Annex 3.3: Adaptive Price Expectation Model

Consider that Q_t is related to price expectation and not the actual price level in time t.

$$Q_t = a^* + b^* P_t^* + e_t^* \tag{3.42}$$

where P^* represents expected level of prices, not actual prices

A second relationship defines the expected level of P^*. It is assumed that in each time period, the expectation changes based on an adjustment process between

Chapter 9
Economics of Non-Renewable Resource Supply

9.1 Introduction

Resources like oil, natural gas or coal are non-renewable in nature as they come from finite stock sources. This means once they are used less will be available in the future. This feature of non-renewable energies then introduces an inter-temporal dimension in the use decision: the choice of using it now or later. This chapter presents a brief review of the optimal allocation of non-renewable (or depletable) resources. This is presented using a simple two-period example first, followed by a more formal presentation in the third section. The influence of discount rate, market structure and the effects of changes in the oil market are discussed subsequently. Finally, the link between the theory and the empirical data is considered.

9.2 Depletion Dimension: Now or Later

The inter-temporal aspect can be analysed using a simple framework as shown below (Fig. 9.1).[1] Consider two time periods and the width of the box in the diagram represents the quantity of resource available for consumption. The amount of resource used in time 1 is measured from the left side, while the quantity used in time 2 is measured from the right side. If the resource is consumed now, it is not available for the next period. But the more we use the resource its usefulness reduces to us, implying less marginal utility.

In the absence of time preference for using the resources, the intersection of the two utility curves gives the quantities to be consumed in two periods. But usually we prefer things now than later, because of time value of money. Consequently, we have to discount the marginal utility in period 2 to make it comparable to the

[1] This is based on Hannesson (1998).

S. C. Bhattacharyya, *Energy Economics*, DOI: 10.1007/978-0-85729-268-1_9,

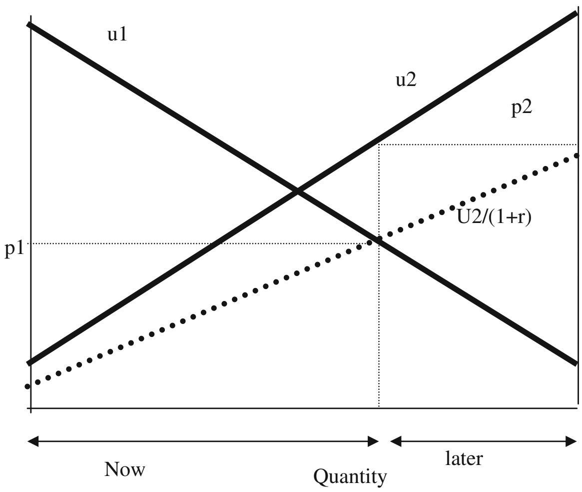


Fig. 9.1 Schematic explaining now or later decision. *Source* Hannesson (1998)

marginal utility in period 1. The dashed line shows the discounted marginal utility in period 2. The intersection of the new line with the marginal utility in period 1 gives the inter-temporal allocation of resources. Thus in mathematical terms,

$$u_1 = u_2/(1+r) \quad \text{or} \quad u_2/u_1 = 1 + r. \tag{9.1}$$

In a market economy, relative prices of goods decide the allocation of resources. If a commodity costs twice as much as another commodity, the marginal utility of the first must be twice as high as that of the second. Thus,

$$p_2/p_1 = u_2/u_1 \tag{9.2}$$

Substituting this in Eq. 9.1, we get,

$$(p_2 - p_1)/p_1 = r. \tag{9.3}$$

That is the price must rise over time at a rate equal to the rate of discount. This is called Hotelling's r percent formula.

The above rule implies that finite resources have a value over and above their cost of production, which is due to their scarcity. This extra value is considered as scarcity rent. Our time preference would require us to consume a bit more in period 1 than in period 2 but for this the price in period 1 has to be somewhat lower than that in period 2.

Why the resource owner will not dump everything now?

If the supplier produces one unit now and invests the money in the capital market he earns rp_1. If he supplies one unit in period 2, he earns p_2, which is $p_1 + (p_2 - p_1)$, which is equal to $p_1 + rp_1$. Thus by waiting for one period, the producer makes the same profit, which makes him indifferent. Thus the expectation of better prices in the future will ensure that not the entire amount is produced in one period.

9.3 A Simple Model of Extraction of Exhaustible Resources

The basic model of the extraction of non-renewable resources was initially proposed by Hotelling (1931). The problem is to find the optimal depletion path of a firm that seeks to extract such resources to maximize its profit. There is a vast body of academic literature on this subject—see Devarajan and Fisher (1981), Fisher (1981) and Krautkraemer (1998) for further details. The basic model is based on the following assumptions: (a) the size of the resource stock is known, (b) the entire reserve is exhausted during the project life, (c) interest rate is fixed.

We define the following terms:

y_t is the quantity of resource extracted in period t;

X_t is the resource stock at the beginning of period t = fixed at $\bar{X}_0$ at time 0;

$C = C(y_t, X_t)$ = total extraction cost;

$P(y_t)$ is the inverse demand function for the resource;

r is the discount rate;

T = time horizon.

The objective is to maximize the net benefit

$$\text{Max}(y_t)\sum_{t=0}^{T}\left[\frac{1}{(1+r)^t}(p_t y_t - c(y_t, X_t))\right] \tag{9.4}$$

S.t.

$$X_0 = \bar{X}_0; \quad X_T = \bar{X}_T \tag{9.5}$$

and

$$\frac{dX_t}{dt} = -y_t \quad \text{or} \quad X_{t+1} - X_t = -y_t. \tag{9.6}$$

The Lagrange function is given by

$$\begin{aligned} L = &\sum_{t=0}^{T}\left[\frac{1}{(1+r)^t}(p_t y_t - c(y_t, X_t))\right] \\ &+ \sum_{t=0}^{T-1}\mu_t(X_t - X_{t+1} - y_t) + \alpha(\bar{X}_0 - X_0) + \beta(\bar{X}_T - X_T). \end{aligned} \tag{9.7}$$

First order condition resulting from differentiation with respect to y_t is:

$$\frac{p_t - (\partial c/\partial y_t)}{(1+r)^t} - \mu_t = 0; \tag{9.8}$$

which can be rewritten as

$$p_t - \frac{\partial c}{\partial y_t} = \mu_t (1 + r)^t = \lambda_t. \tag{9.9}$$

The net price is equal to royalty and in the special case where cost of extraction is negligible the price should grow at the rate of interest. The term on the right hand side of Eq. 9.9 is the user cost, which is directly related to the shadow price of the resource. It suggests that for non-renewable resources, the price should contain an additional element that takes care of the effect of resource depletion. This is the opportunity cost of using the resource now instead of leaving it for the future. In the special case when the cost of extraction is insignificant or zero, the price becomes equal to the rent and hence the rate of price change is just equal to the rate of interest. This is the fundamental result in the economics of exhaustible resources.

9.3.1 Effect of Monopoly on Depletion

Consider the case of pure monopoly—where one producer is functioning in the industry. The problem here is similar to the competitive market. The only difference is in the first condition of optimal depletion because the monopolist will take into account the influence of his output decision on price. The first order condition resulting from differentiation with respect to y_t is given by

$$\frac{p_t + y_t \frac{dp}{dy_t} - (\partial c / \partial y_t)}{(1 + r)^t} - \mu_t = 0; \tag{9.10}$$

or MR − MC = Royalty.

Introducing price elasticity in the above equation we get

$$\frac{p_t \left(1 + \frac{1}{e_p}\right) - (\partial c / \partial y_t)}{(1 + r)^t} - \mu_t = 0; \tag{9.11}$$

which can be re-written as

$$p_t \left(1 + \frac{1}{e_p}\right) - (\partial c / \partial y_t) = \mu_t (1 + r)^t = \lambda_t; \tag{9.12}$$

$$p_t = \lambda_t + (\partial c / \partial y_t) - \frac{\lambda_t + (\partial c / \partial y_t)}{1 + e_p} \tag{9.13}$$

This implies that the price under monopoly would have three components: marginal cost of extraction, royalty and a monopoly rent. This third component is

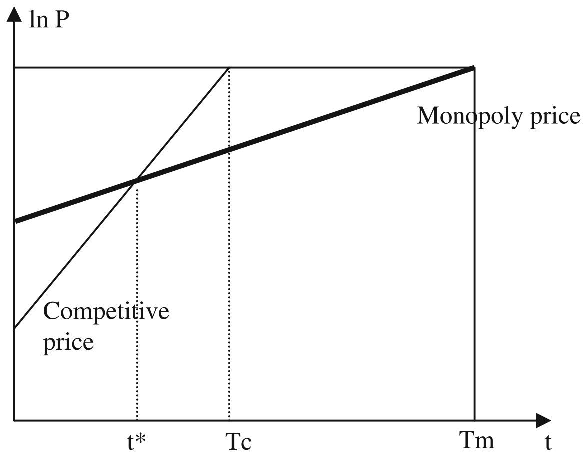


Fig. 9.2 Price path in competitive and monopoly cases

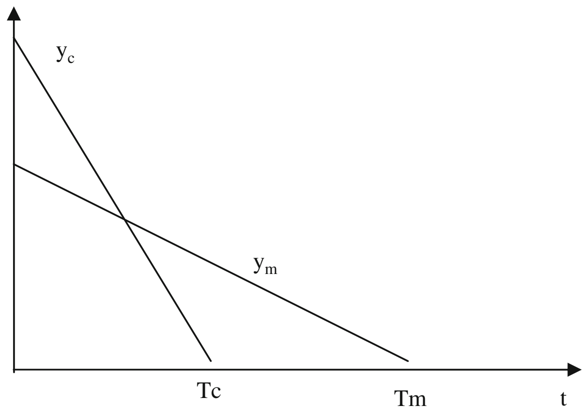


Fig. 9.3 Optimal extraction path

positive for all elasticity values greater than –1.0. In those cases, price under monopoly would be greater than the price under competition.

For a linear demand function, it can be shown that the optimal price path in the case of a monopoly is two times less rapid than that of a competitive market price path. Obviously, the two prices start at different levels and the price charged by the monopolist includes the monopoly rent. This is shown graphically in Fig. 9.2. The optimal extraction path also follows a similar path—under the competitive market situation, the resource is exhausted twice as fast as that under the monopoly in the above case (see Fig. 9.3).

Relating the above idea to the oil market would then suggest that the price change under the OPEC era in the 1970s was an adjustment process where the competitive price path was abandoned in favour of a monopolistic price path. This is shown in Fig. 9.4. Surely, this slows down the extraction and the resource will last longer in this case.

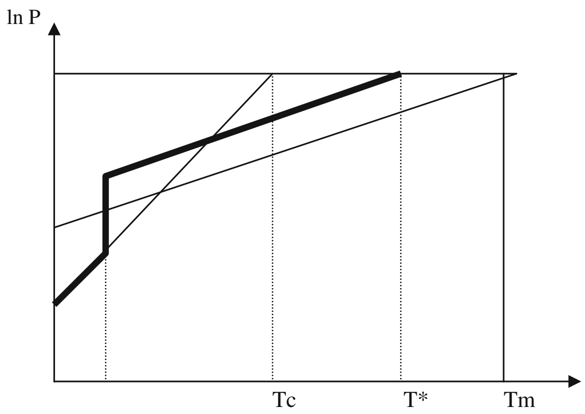


Fig. 9.4 Change in price path under OPEC after the first oil shock

9.3.2 Effect of Discount Rate on Depletion Path

As the discount rate plays an important role in the net worth calculation, the discount rate influences the decision about using non-renewable resources now or in the future. A high discount rate leads to higher rate of extraction initially but the output declines fast and therefore, the resource is exploited quickly (see Fig. 9.5). On the other hand, a lower discount rate prolongs the resource availability through a lower rate of initial extraction and a slower rate of extraction.

The price path for different discount rates again follows the similar pattern (see Fig. 9.6). A high discount rate reduces the initial price but the price path is steeper compared to a low discount rate, which in turn causes to reach the backstop prices earlier.

It needs to be mentioned here that although this application of the Hotelling principles to depletion has given rise to a large volume of academic literature, energy prices do not seem to follow the prescriptions of the theory. As shown in Fig. 9.7, the crude oil price did not follow the price path suggested by the theory, although prices have hardened in recent times. The theory relies on a number of restrictive assumptions and despite much theoretical interest, has not helped much in understanding the fuel price behaviour.

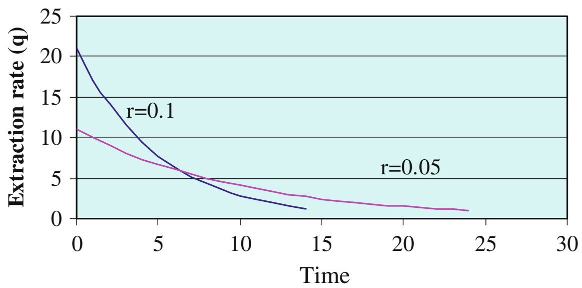


Fig. 9.5 Effect of discount rate on the extraction path

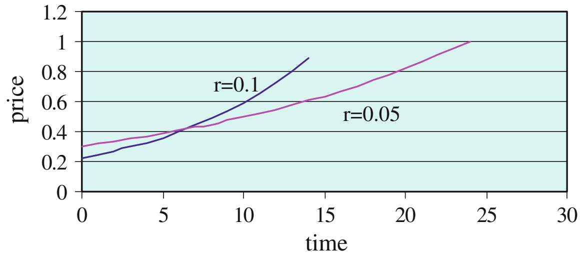


Fig. 9.6 Price path under different discount rates

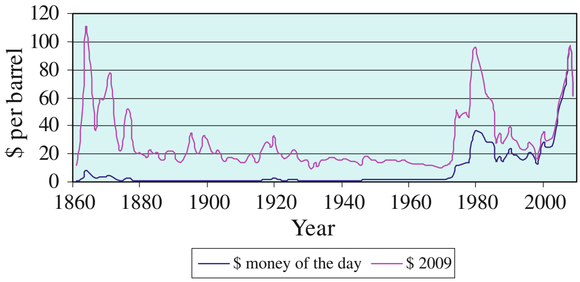


Fig. 9.7 Oil price trend. *Data source* BP Statistical Review of World Energy, 2010

Therefore, from a practical point of view, the relevance and influence of the theory has been quite limited.

9.4 Conclusion

This chapter has provided a simple and a formal introduction to the theory of exhaustible resources. The chapter has restricted itself to the basic model of the theory and did not enter into more elaborate extensions of the theory that has been suggested by various authors to relax some of the restrictive assumptions of the basic model. The outcomes of the model are at odds with the reality of the energy sector and therefore, the practical relevance of the theory remains limited.

References

Devarajan S, Fisher AC (1981) Hotelling's economics of exhaustible resources: fifty years later. J Econ Lit XIX:65–73

Fisher AC (1981) Chapter 2: Resource and environmental economics. Cambridge University Press, London

Hannesson R (1998) Petroleum economics: issues and strategies of oil and natural gas production. Quorum Books, London

Hotelling H (1931) The economics of exhaustible resources. J Polit Economy 39:137–175

Krautkraemer JA (1998) Non-renewable resource scarcity. J Econ Lit XXXVI:2065–2107

Chapter 11
The Economics of Renewable Energy Supply

11.1 Introduction: Renewable and Alternative Energy Background

This chapter focuses on the economics of renewable and alternative energies. The term "alternative energy" refers to any energy forms that are outside the conventional forms of energies we have considered so far. Although conventional energies can be renewable as well (such as hydropower) and can include both renewable and non-renewable sources (such as tar sand, shale gas, etc.), this chapter focuses on modern renewable energies. Most of these energies are available abundantly and the mankind has been using them for various purposes from time immemorial. The direct cost to the consumer remains low in their traditional form of use (such as drying). However, modern ways of using these energies require sophisticated conversion processes, which in turn increase the cost of supply.

The oil price shocks of the 1970s triggered new interest in renewable energy sources. Availability of easy petrodollars facilitated funding of renewable energy research and the field flourished during the periods of high oil prices in the international market. The global concern for climate change and sustainable development provided further impetus to renewable energies. Now renewable energies occupy an important place in any strategy for sustainable development in general and sustainable energy development in particular.

11.1.1 Role at Present

According to IEA,[1] around 13% of global primary energy supply in 2007 came from renewable energies. Out of 12 Giga ton of oil equivalent of primary energy

[1] See also Darmstadter (2003) for a review.

S. C. Bhattacharyya, *Energy Economics*, DOI: 10.1007/978-0-85729-268-1_11,

11.4 Drivers of Renewable Energy

Renewable energies are emerging as alternative fuels as they offer a number of advantages. Following Goldemberg (2004) these are:

a) **Reduction in CO_2 emission and mitigation of climate change**: This is the main driver of renewable energy at present. The concentration of greenhouse gases (GHG) is increasing due to fossil fuel dependence of modern economies. It is believed that the increasing concentration of GHGs has led to warming of our climate. It is forecast that without any mitigation action, the CO_2 concentration in the atmosphere would double the present level by 2050. Renewable energies being carbon free (or neutral) would help reduce the GHG concentration.

b) **Security of energy supply**: Security of energy supply has made a come-back in recent years. This is attributed to recent increases in fossil fuel prices in general and oil prices in particular; concerns for depletion of fossil fuels globally and imminent production decline in the US and UK, and consequent increase in import dependence; increasing competition for supply from emerging consuming countries; political instability in the hydrocarbon resource rich areas; and high economic impacts of energy supply disruption in the developed and rapidly developing countries.

   As fuel diversification is considered as an important strategy for ensuring supply security, developing alternative energies from locally available resources can reduce import dependence and accordingly, renewable energies are being viewed favourably from this perspective.

c) **Improving energy access**: It is now believed that more than 2 billion population worldwide do not have access to clean energies. The problem is more acute in rural areas of poor countries where the supply system may be inexistent. To ensure sustainable development, it is essential to provide clean energy to these people. Renewable energies offer certain advantages in this respect—they reduce environmental and health damages, and save time in fuel collection and improve working conditions. These changes can in turn provide better opportunities for income and reduce poverty.

d) **Employment opportunities**: Renewable energy supply has the potential for employment generation, directly due to decentralised, modular structure of the technologies and local level operation of the systems. And indirectly through improved working conditions or saving in time which would otherwise be used in drudgery.

e) **Other spill-over effects**: Reliance on renewable energies would help improve macro-economic stability. The logic goes as follows: (1) Promotion of renewable energies reduces import dependence; (2) fossil fuel import being the important constituent of the international trade of importing countries, a switch over to the renewable energies is expected to reduce the trade balance; (3) this in turn reduces the possibility of economic shocks due to external factors.

In addition to the above advantages, renewable energy technologies benefited from significant cost reductions over the past decade and such a trend is expected to continue in the future. Cost reductions have made some of these technologies economically feasible and competitive (e.g. wind) with other conventional power generation technologies. In fact, Darmstadter (2003) suggests that the cost reductions were much higher than was expected by many.

Despite enjoying such advantages why are renewable energies unable to capture higher market shares?

This is essentially due to the existence of considerable barriers facing renewable energies. The literature on the subject has identified a number of barriers. Painuly (2001) provides a framework for identifying and analyzing the barriers. He suggests that the barriers can be analysed at a number of levels: first can be grouped in broad categories. Within each category, a number of barriers can then be identified. At a third level, the elements of these barriers can be identified. This disaggregated approach can provide a better clarity on the subject. Neuhoff (2005) has identified four broad categories of barriers and elements within them. These include technological barriers (related to intermittency of supply), uneven playing field (related to failure of the pricing system to internalize externalities of fossil-fuel energies), marketplace barriers (such as access to the grid, regulatory barriers, inappropriate tariffs or incentives for renewable energies, etc.), and non-market barriers (such as administrative difficulties, lack of long-term commitment, lack of information, etc.)

11.5 The Economics of Renewable Energy Supply

This section will first focus on the economics of renewable electricity and then on that of bio-fuel supply.

11.5.1 The Economics of Renewable Electricity Supply

Electricity from renewable resources has a number of technical features:

a) most common forms of renewable energies (such as solar, wind or tidal) are intermittent in nature (i.e. they are not available all the time), and
b) given that electricity cannot be stored in large quantities in a cost effective manner, these energies have to be used when they are available.

As a result of intermittency, a number of issues arise.

(1) Electricity generated from such sources cannot be dispatched following the merit-order dispatch schedule. They have to be used whenever the electricity is available. However, through better forecasting of weather conditions, more accurate assessment of local level generation can be made.

(2) As a consequence of the above, the capacity is used only for a limited time, leading to low capacity utilization. This has been indicated in Fig. 11.6 where it is noted that the average utilization of solar PV systems is less than 10% in Europe, while the average wind capacity utilization is about 20%.
(3) Consequently, such systems cannot provide reliable supplies round the clock and will require back-up capacity (or standby capacity). The standby capacity often relies on non-renewable energies and therefore, the benefits of renewable energies are not available. The standby capacity also increases the cost of supply.

In a study by Gross et al. (2006), it is estimated that the intermittency costs in Britain are of the order of 0.1–0.15 pence/kWh. This is quite substantial compared to the electricity price paid by the consumers.

In addition, renewable electricity often suffers from other biases against it. These include:

a) **Inappropriate valuation**: The value of electricity normally varies depending on whether it is used during the off-peak hours or peak-hours. The peak-period supply should fetch a higher value to the supplier; but as renewable supply is treated outside the wholesale market (being non-dispatchable), the appropriate valuation of its contribution is difficult to make. This would affect the financial and economic viability of the renewable energy projects.
b) **Inappropriate price signals**: Often such units are embedded in the distribution system and rely on net metering (i.e. considers the energy supplied less energy consumed by the unit). But unless the retail tariff is based on time-of-day pricing, the system does not provide proper signal to the consumer and the supplier. This also affects the renewable energy generation and its viability.
c) **Non-internalisation of externalities**: Renewable energies have environmental advantages compared to the fossil fuel-based electricity. Consequently, non-recognition of the external costs[6] in the pricing puts renewable energies at a disadvantage and does not allow two types of energies to be compared on the same level. This acts as a barrier to the renewable energy development.
d) **Fuel risk benefits**: Renewable energies do not face fuel price risks faced by the fossil fuels. In fact, the operating cost of renewable energies is minimal in most cases. However, the market price for fossil-fuel based electricity does not provide the correct signal to the investors and the consumers taking the premium for higher prices for fossil fuels into consideration. This has an adverse effect on the renewable energy development. Awerbuch (2003) suggested that inappropriate fuel risk and financial risk estimation renders renewable electricity costlier, which introduces a systematic policy bias against renewable electricity.

[6] External costs are covered in another chapter where the economics environmental damages from energy use is considered.

Any comparison of electricity supply costs should adequately capture the above differences. The basic indicator—levelised cost—is often used but it may be an in appropriate comparator as it relies only on a specific level of capacity utilisation, which varies widely across electricity generating technologies.

The screening curve approach in conjunction with the load duration curve provides a better picture as this can capture the value of energy at different stages of the load.[7] More complex simulation models are required to capture the differences in costs and technical characteristics of electricity generating techniques and their effects on the supply. This however requires more involved mathematical models, which are beyond the scope of this discussion.

11.5.1.1 Cost Features

The main elements of costs to be considered in the case of electricity supply technologies are:

a) Energy-related costs: Include those costs which are related to energy generation in a facility: costs related to fuels and variable operating and maintenance related costs. Normally, for fossil-fuel based electricity, this component is relatively high while for the renewable fuels, this element tends to be small.
b) Capacity costs: These include the cost of installing the capacity (charges to be paid in relation to installation of a capacity) and the fixed operating and maintenance costs (labour charges, stocks, etc.). For renewable energy based electricity, this is the most important cost element and could be between 50% and 80% of the overall cost of supply.
c) Other related costs: This is a broad category of cost that can include external costs due to environmental damages and climate change, costs related to standby or reserve capacity, and any other costs that should be considered to make the like-for-like comparisons.

   a. Environmental costs are higher for fossil fuels and nearly non-existent for the renewable energies.
   b. On the other hand, standby capacity costs could be important for certain types of renewable energies.
   c. Similarly, fuel price risk (or security risk) could be high for some fossil fuels and should be considered here.

Figure 11.14 presents the comparison of levelised costs of electricity supply for different electricity technologies from the Royal Academy of Engineering (2004) study.[8] Although this figure provides costs relevant for the UK market, it still provides a generic picture.

[7] See the paper by Kennedy (2005) for an application of this method.

[8] Heptonstall (2007) provides a review of unit cost estimates of electricity generation using different technologies.

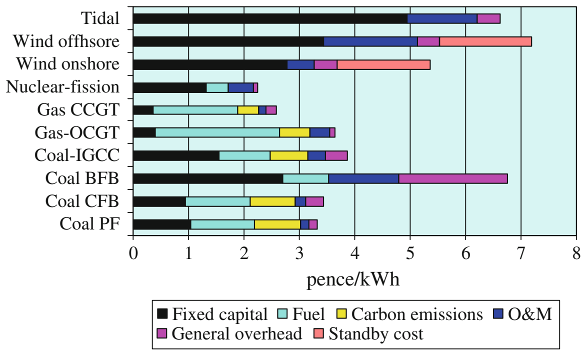


Fig. 11.14 Levelised cost of electricity generation by technologies. *Source* Based on data from Royal Academy of Engineering (2004)

The above figure suggests that most renewable-energies would be cost ineffective solutions for generating electricity even after taking environmental costs into consideration. This is because of high level of standby power costs. If standby power cost is ignored, the onshore wind power becomes quite competitive with commonly used fossil fuels like coal or gas (in an open cycle). However, tidal power and offshore wind power are still not cost effective solutions. The assumptions about fuel prices and capacity utilization rate also affect the outcome significantly. The report assumed full utilization of base load plants and 35% capacity utilisation factor for intermittent sources such as wind. As indicated before, the capacity factor of different technologies varies widely and a uniform assumption does not capture the real situation. Similarly, the fuel price assumptions were quite conservative, making the security of supply insurance premium quite small for fossil fuels.

A study by EPRI (2009) provides the levelised cost of electricity for a future date—2015 and 2025 (see Table 11.1). The message from the above discussion appears to be clear: renewable energies for electricity supply still face cost disadvantages and would require support to ensure their promotion.

Table 11.1 Levelised cost of power generation

Technology description	Cost in 2015 (2008 constant $/MWh)	Cost in 2025 (2008 constant $/MWh)
Super critical pulverized coal	66	86–101
Integrated gasification combined cycle	71	78–92
Combustion turbine combined cycle	74–89	67–81
Nuclear	84	74
Wind	99	82
Biomass circulating fluidised bed	77–90	77
Solar thermal trough	225–290	225–290
Solar PV	456	456

Source EPRI (2009)

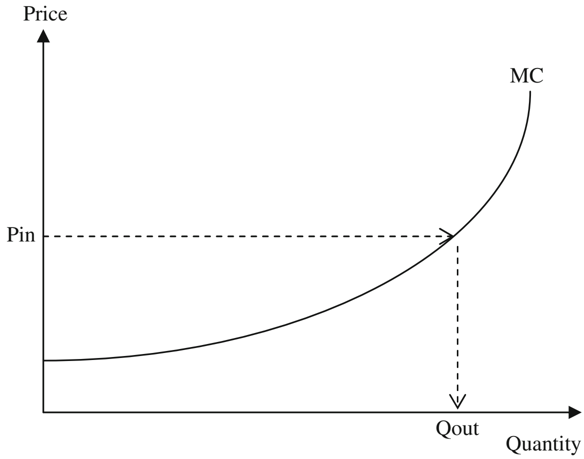


Fig. 11.15 Feed-in tariff principle. *Source* Menanteau et al. (2003)

11.5.1.2 Support Mechanisms[9]

A number of intervention or support mechanisms have been used in practice to promote renewable energy based electricity to overcome barriers arising from market distortions and lack of internalisation of externalities. These include feed-in tariffs, competitive bidding process, renewable obligations, financial incentives, and taxing fossil fuels.

Feed-in Tariffs

This is an intervention by influencing the price. Here the electric utilities are required by law or regulation to buy renewable electricity at fixed prices set normally at higher than the market price. The system has evolved over time: in California, a system of standardised long-term contracts at fixed prices was initiated in the 1980s to promote renewable energies, similar to independent power project contracts. In mainland Europe, the producers were guaranteed a fixed share of the retail price and the contracts lasted for the project life (15–20 years). More recent feed-in tariffs vary by location, by technology and by plant size. The fixed price declines over time and is adjusted periodically but the tariffs are long-term in nature. The basic mechanism is explained in Fig. 11.15.

[9] A well-developed body of literature exists in this area covering alternative support mechanisms and their application to specific technologies or countries. See for example Menanteau et al. (2003), Sawin (2004), Mitchell et al. (2006), del Rio and Gual (2007), Bunter and Neuhoff (2004), Dincia (2006), and World Bank (1997).

In Fig. 11.15, assume that the regulatory or public authorities have fixed the feed-in tariff at P_{in}. All producers whose cost of supply is below this price will enter the market and produce an output Q_{out}. The total cost of support in this case is $P_{in} \times Q_{out}$. The important point to note here is that projects with low cost of production will earn a rent due to their locational or technological advantage. The fixed price system allows the producer to capture this rent, which provides an incentive for further innovation.

Generally the cost of subsidising renewable electricity is passed on to the electricity consumers through the electricity tariff. However, in some cases the tax payers in general or consumers in the area of utility's jurisdiction where the renewable energy development is taking place may bear the cost (Menanteau et al. 2003).

The feed-in tariff system has proved to be a successful instrument. It has been used by those who have successfully developed their renewable electricity market. These countries have often exceeded their national targets. As the producer has tariff certainty over the project life, the system reduces financing risks and facilitates financing. The system is easy to implement and if standardised, the transaction cost can be low. However, the feed-in tariff system through generous payments to producers promotes high cost supply. The long-term nature of the contract can lead to stranded investments, especially in a competitive market. Finally, it is not known in advance how much capacity addition will take place. Therefore, there is no guarantee that a given target will be achieved. If over-supply takes place, the utility has the obligation of purchasing the power, which creates a contingent liability.

Competitive bidding processes

This is a quantity restriction mechanism where the regulator or public authority mandates that a given quantity of renewable electricity would be supported but decides the suppliers of such electricity through a competitive bidding process. Interested producers are asked to submit bids for their proposals, which are ranked in terms of their cost of supply. All proposals are accepted until the target volume is reached. This mechanism is therefore an attempt to discover the supply curve through bids and can be represented in diagrammatic form as shown in Fig. 11.16.

In Fig. 11.16, for the target volume Qt, suppliers up to a marginal cost of P will be selected. However, the price paid to each supplier is limited to the bid price (i.e. pay as per bid) and not the marginal cost of the last qualifying bid. This removes the rent or producer surplus that is available in the case of a feed-in tariff. This reduces the support cost to the area under the supply curve and as a consequence, the burden on the consumers reduces. However, by removing the rent, the incentive to innovate is reduced. As the bidding system decides the quantity to be procured, there is certainty in terms of maximum volume of supply (although whether the target will be reached or not remains unknown). The price to be paid

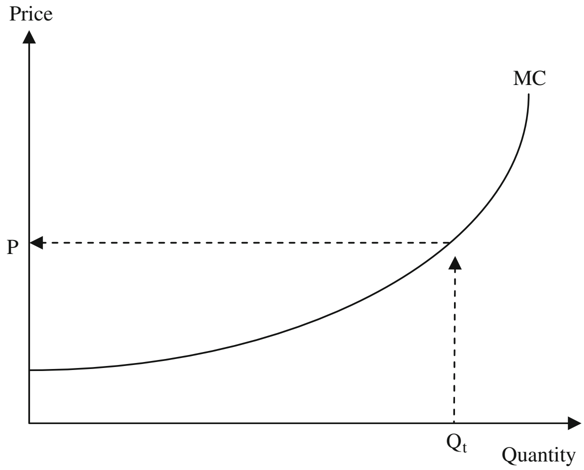


Fig. 11.16 Competitive bidding process principle. *Source* Menanteau et al. (2003)

and therefore the overall cost of support is not known ex-ante (Menanteau et al. 2003).

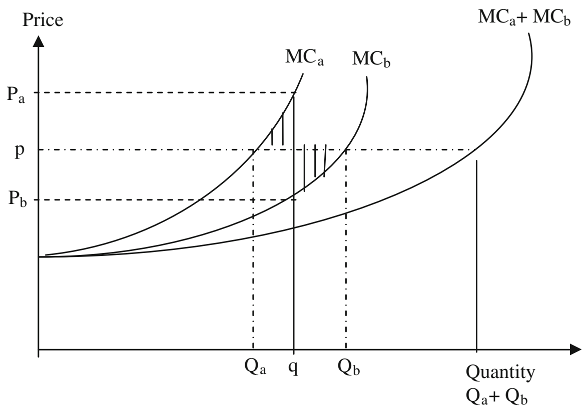
Renewable obligations

Renewable obligations (RO) also work through the quantity restriction mechanism where the government sets the target for renewable electricity supply and lets the price be determined by the market. The obligation is placed on the electricity suppliers to purchase a given percentage of their supply from renewable sources. The target is often tightened over time with the objective of reaching a final level by a target date. The Renewable Portfolio Standard (RPS) used in the United States of America or the Renewable Obligation system in England and Wales are the common examples of this category.

The Renewable Obligation requires the electricity supplier to supply a specific amount of renewable energy in a given year. For example, the RO in England and Wales started in 2002 with a target of 3% for 2003 but the target rises to 15.4% for the year 2015–2016. In theory, the RO is guaranteed to stay at 15.4% level until 2027—thereby guaranteeing a life of 25 years. However, in April 2010, amendments were made to extend the end date to 2037 for new projects.

A number of technologies are recognised as the eligible renewable sources (such as wind, solar energy, biomass, etc.). The producer of renewable electricity receives from the RO administrator a tradable certificate, called the Renewables Obligation Certificate (ROC), for every unit of electricity generation—either at a uniform rate for every unit of renewable electricity produced or at a preferential rate depending on the technology employed (which has been introduced in England and Wales from 1, April 2009).

Fig. 11.17 Economic logic for certificates trading. *Source* Menanteau et al. (2003)



Generators thus have two saleable products[10]: electricity which they sell to electricity suppliers and the ROC that they can sell to electricity suppliers or traders. Certificates are tradeable and trading between suppliers and traders creates a market for these certificates. The economic logic here is that trading of certificates allows electricity suppliers to meet the target at the least cost. This is explained in Fig. 11.17.

Consider two suppliers A and B who are subjected to a renewable target of q. The marginal cost of supply for A is given by MCa while that of B is given by MCb. As A faces a steep cost curve compared to B, if it has to comply with the requirement alone, its cost will be P_a whereas B can meet the target at P_b. However, because of its cost advantage, B could easily expand its renewable supply beyond the required limit and trade the credit with A. This allows both the suppliers to benefit as the system can achieve the target at a lower price p. Thus, B produces up to Q_b while A produces just Q_a and together they still satisfy the 2q requirement set by the regulator at a lower price. This benefits the society as a whole by imposing lesser burden for promoting renewable energies.

In the English system, the suppliers can also pay a buy-out price in lieu of ROCs to meet their obligation or follow a combined approach of buying some ROCs and buy-out the rest. The buy-out price effectively sets the ceiling price for the supplier to buy renewable electricity, and acts as a protective instrument for consumers (Mitchell et al. 2006).

To prove compliance of obligation, suppliers have to redeem their ROCs with the regulator and pay the fine for non-compliance (or buy-out price if available). In England and Wales, the buy-out price is set by the regulator and the revenue so generated is recycled annually to the suppliers presenting the ROCs in proportion to their ROC holding. The market price of ROCs reflects the buy-out price and the recycle payment received by the suppliers.

[10] In England and Wales, the generator can also receive its share of recycled buy-out premium and payment for levy exemption certificates in the consumer is eligible for exemption under the Climate Change Levy agreements (see Mitchell et al. 2006).

11.5.1.3 Performance of Price and Quantity-Based Mechanisms Under Uncertainty and Risk

In ideal conditions of free and cost-less information, the price- and quantity-based mechanisms produce similar results. However, in reality these mechanisms do not yield same results due to incomplete information and uncertainty. Because the supply curve is not known in advance, the shape of the curve would influence the outcome considerably. If the shape of the curve is relatively flat (or elastic), the output in a price-based system will be substantially off the target when the shape in incorrectly estimated (see Fig. 11.18). On the other hand, for steep supply curves, the quantity-based systems face the risk of off-the-mark prices under supply cost uncertainties.

Assume that the regulator assumes the shape of the supply curve as indicated in MC_2 and sets a feed-in-tariff at p, expecting Q_2 as the supply to be supported. But the actual shape turned out to be MC_1, resulting in Q_1 as the supply volume. This results in an increased supply and consequently a higher volume of subsidy for support. On the other hand, for a quantity-based system, assuming the shape as MC_1, the regulator set a quantity q for renewable supply. In reality, the shape turned out to be MC_2. This leads to a significantly higher marginal price to meet the target and would facilitate entry of costly supply options. From above, the following logic can be obtained: when the slope of the marginal cost curve is gentle, the quantity-based system works better in presence of uncertainty whereas the price-based system works better when the slope is steep. In other words, a price-based approach performs poorly when the marginal cost curve is gently sloped and a quantity-based approach works poorly when the slope of the marginal cost curve is steep.

Mitchell et al. (2006) also introduce another set of risks in comparing these mechanisms. They consider price, volume and balancing risks faced by the

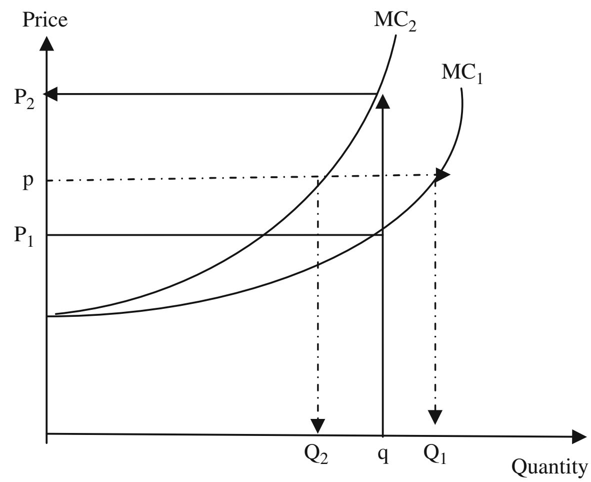


Fig. 11.18 Performance under uncertainty

investors of renewable energies under two broad types of support systems. In the case of feed-in tariffs, the electricity supplier is obligated to buy any amount of renewable electricity produced at the set price. This removes the volume risk. Similarly, the price is known in advance and the contractual arrangement facilitates financing of renewable energy projects. In the context of competitive markets, the renewable generator does not have to worry about the mismatch between predicted and actual supply in a feed-in tariff regime. It is the responsibility of the system operator to take care of the variation. There is no penalty for the mismatch.

On the other hand, the Renewable Obligations do not promise a price—this is decided by the market where supply and demand will determine the outcome. This leaves the investors with a great deal of risk and price uncertainty. Absence of a contract also affects the ability to project finance new capacity additions. Similarly, as the supply volume approaches the target, the generators face the risk that their outputs will not be purchased at the prevalent price. The suppliers would look for cheaper sources and the generator will face the volume risk. Finally, under the British system the renewable generator bears the risk of over or under-performance and faces the balancing risk. Table 11.2 summarises these risks. Accordingly, the RO appears to leave substantial risks to the generators. This can explain the slower growth of renewable electricity capacity in the U.K. However, it is important to indicate here that the British policy aimed at keeping the extra burden on electricity consumers low. The policy has succeeded in achieving this and as the technology matures, the sector and the society are expected to benefit from the prospects of declining costs of future renewable electricity.

11.5.1.4 Financial Incentives

These are fiscal measures used either to reduce the cost of production or increase the payment received from the production. Commonly used incentives include: tax relief (income tax reduction, investment credit, reduced VAT rate, accelerated

Table 11.2 Comparison of performance of support systems under risk (investor's perspective)

Risk type	Feed-in tariff	RO
Price risk	No price risk for generators	Great deal of price risk as price depends on supply–demand interactions
	Generators save money from hedging the price risk	Price likely to fall as supply approaches the target volume
Volume risk	No volume risk—obligation to buy all power produced	Exists
		Individual generators do not have any guarantee of volume
		Once target is met, no security of buying the entire output
Balancing risk	Side-stepped; no penalty for intermittent generation.	Balancing risk exists; penalty imposed for out-of-balance positions.

Source Based on Mitchell et al. (2006)

depreciation, etc.); rebates or payment grants (that refunds a share of the cost of installing the renewable capacity), and low interest loans, etc. Normally these incentives show preferences to particular technologies (hence cherry picking) and may promote capacity but not necessarily energy generation.

11.5.1.5 Taxing Fossil Fuels

The objective here is to reflect the true costs and scarcity of the fossil fuels in the prices paid by the consumers to send a clear signal. Taxing fuels for their environmental and other unaccounted for damages is one way of ensuring the level playing field. The Nordic countries are in the fore-front of such environmentally-oriented taxation. They are the pioneering countries in introducing carbon taxes (i.e. a tax on CO_2 emissions), even before the European Union launched a proposal to introduce community-wide carbon taxes in 1992 (which was never adopted although individual members have introduced some such taxes). Finland was the first country to introduce a CO_2 tax in 1990, followed by Norway and Sweden in 1992 and Denmark in 1992. Besides carbon tax, there are other taxes on energy as well—these include taxes on fuel and electricity and a tax on SO_2 emission. Despite this, it is doubtful whether the polluter is bearing the tax burden as a study by Eurostat (2003) found that the burden is shifted to residential consumers while the industry bears a relatively lower burden.

11.6 The Economics of Bio-fuels

The cost of supply of bio-fuels varies widely depending on the technology, feedstock used and the size of the conversion plant. The energy content of bio-fuels varies significantly and the energy density of bio-fuels is less compared to petrol or diesel. Generally, the plant size and feedstock cost play an important role in the bio-fuel supply cost. However, bio-ethanol and bio-diesel costs do not follow similar patterns and consequently, it is better to analyse them separately.

11.6.1 Bio-Ethanol Cost Features

Two most important cost elements for bio-ethanol production are (OECD 2006):

a) The cost of feedstock: this is the most important cost in bio-ethanol production (accounts for around 41% of the cost of supply). The choice of feedstock explains cost variation across countries to a large extent.
b) Energy and labour costs: These are also quite important in bio-ethanol production and account for about 30% of the costs.

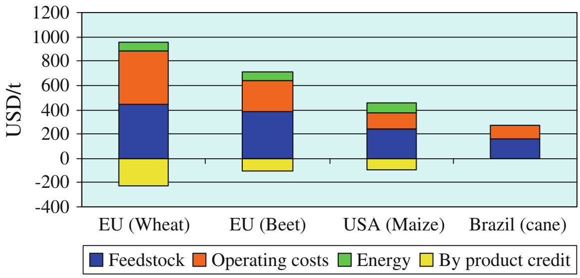


Fig. 11.19 Comparison of bio-ethanol production costs. *Source* OECD (2006)

Capital recovery can be about one-sixth of the costs while the rest is attributed to the cost of chemicals. Some credits are also obtained by selling them and this could change the economics of bio-fuels to some extent.

Brazil is the least cost supplier of bio-ethanol and produces 30% cheaper compared to the US cost and almost 2.5 times cheaper compared to the European production (see Fig. 11.19).

How does bio-ethanol compare with gasoline price? Figure 11.20 provides the comparison. Except Brazil, no other producer is yet able to produce bio-ethanol at a competitive price. The cost of ethanol from maize comes close to gasoline prices in the USA.

The cost of production however falls as the size of the conversion plant increases. In fact, it is reported that the new plants coming up in the USA are exploiting this feature to gain competitive advantage.

As the feedstock demand increases with higher fuel demand, the feedstock price will increase. Higher feedstock price would affect food prices and would encourage diversion of land and agricultural activities towards fuel feedstock supply. This could have adverse consequences for food supply, water use, and for competitiveness of bio-ethanol. In fact, this is one of the main concerns about the first generation bio-fuels.

11.6.2 Bio-Diesel Costs

The feedstock cost plays a much higher role in the case of bio-diesel—almost 80% of the operating costs (Balat and Balat 2008). An example using tallow-based

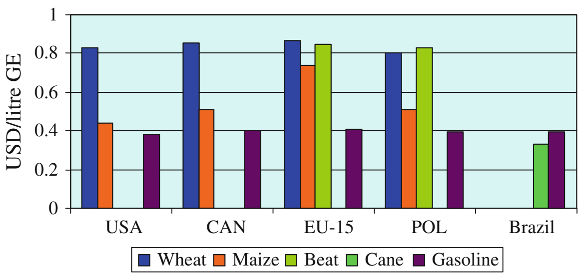


Fig. 11.20 Comparison of bio-ethanol costs with gasoline (ex tax) prices in 2006. *Source* OECD (2006)

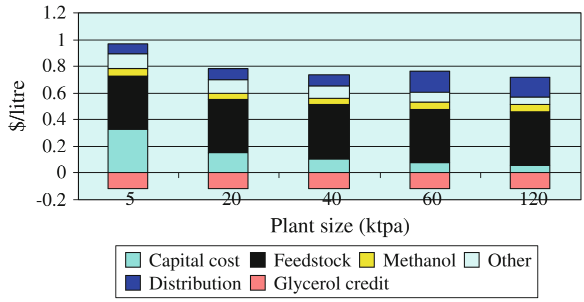


Fig. 11.21 Cost of bio-diesel production from tallow. *Source* Balat and Balat (2008)

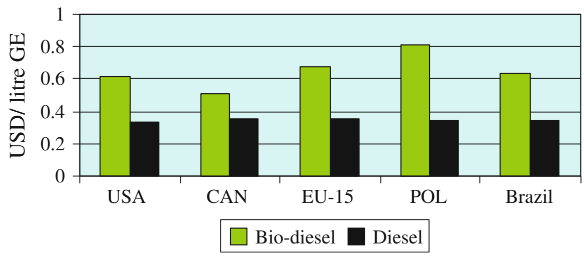


Fig. 11.22 Comparison of bio-diesel cost with diesel prices (ex-tax) in 2006. *Source* OECD (2006)

bio-diesel is provided based on Balat and Balat (2008) in Fig. 11.21. The competition from high value cooking use affects the feedstock price and the cost of production. As a result, nowhere in the world bio-diesel is yet a cost effective solution (see Fig. 11.22).

As bio-diesel or bio-gasoline is not yet competitive, support mechanisms have been developed to promote them.

11.6.3 Support Mechanisms

The generic support mechanisms are quite similar to that used for renewable electricity. The quota system (e.g. EU Directive on Bio-fuels), renewable obligation (UK Renewable Transport Fuel Obligation, RTFO), standards based system and financial incentives are commonly used.[11]

EU Bio-fuels directive: The European Union issued a directive in 2003 requiring members to ensure a minimum level of bio-fuel supply in their markets. The indicative targets set in the Directive were to supply 2% (on energy content basis) of all petrol and diesel used for transport by end of 2005, rising to 5.75% (on energy content basis) by 2010. Most of the members failed to meet the 2005 target

[11] For a brief review of support policies see OECD (2006, pp. 16–21). Also see Chap. 7 of IEA (2004).

and the progress towards 2010 remains limited. In 2009, the Renewable Energy Directive has set a target of 10% share of renewable energy in the transport sector.

RTFO: This is the main instrument being used by the UK to promote bio-fuels in the transport sector.[12] This obligation came in to force in 2008 and the target for 2009/10 is 3.25% renewable fuel use by volume in the transport sector. The mechanism is similar to that of the renewable obligation being used for electricity generation. Each transport fuel supplier (above a certain threshold) has a specific obligation to supply renewable fuels. They can claim certificates for renewable fuel supply and at the end of the compliance period redeem the certificates to demonstrate compliance. The supplier also has a buy-out option in case of non-compliance, set at 15 pence per litre in the first 2 years, rising to 30 p/l from the 2010/11 reporting period.

However, promotion of bio-fuels has raised concerns about food security, water scarcity and adverse effects on the poor. The competition for land for food and fuel production and the limited net energy benefits of the first generation bio-fuels have been highlighted by many, including FAO (2008) and WWI (2006). A careful analysis is therefore required before embarking on a large-scale promotion and supply of bio-fuels.

11.7 Conclusion

This chapter has provided an overview of renewable energy use and has introduced the economic concepts for analysing the developments. The levelised costs for electricity generation from renewable sources are discussed and the cost structure of bio-fuel is presented. The supporting mechanisms used by the government to promote renewable energies are also discussed to bring out the essential features and remaining challenges. Surely, renewable energies will play an important role in the energy mix in the future but many challenges remain before such energies can compete with fossil fuels.

References

Awerbuch (2003) Determining the real cost: why renewable power is more cost-competitive than previously believed, Renewable Energy World (see http://www.awerbuch.com/shimonpages/shimondocs/REW-may-03.doc)

Balat M, Balat H (2008) A critical review of bio-diesel as a vehicular fuel. Energy Convers Manag 49(10):2727–41

Bomb C, McCormick K, Deuwaarder E, Kaberger T (2007) Biofuels for transport in Europe: lessons from Germany and the UK. Energy Policy 35(4):2256–67

[12] See Department for Transport website http://www.dft.gov.uk/pgr/roads/environment/rtfo/.

Chapter 14
International Oil Market

14.1 Introduction

This chapter focuses on the international oil market and presents an overview of the developments in this industry by looking at the resource positions, production and consumption patterns. It also traces the changes in the organisational pattern of this industry over time and highlights the nature of market interactions in these industries. The purpose of the chapter is to capture the essence of the changes in the industry without entering into an elaborate analysis or discussion, which is outside the scope of this chapter. This chapter is organised as follows: first a brief history of the evolution of the oil market is presented by considering two important phases of development—pre-OPEC era and OPEC era. This is followed by an analysis of some key aspects of the market.

14.2 Developments in the Oil Industry

Oil was discovered by Colonel Drake and William A. Smith in 1859. The oil industry has undergone four distinct phases between 1859 and 1960, when OPEC was formed. Here a brief description of the pre-OPEC and post-OPEC era is given. Detailed discussions can be found in, among others, IFP (2007).

14.2.1 Pre-OPEC Era

The four phases of this period are: the period of gold rush, the phase of Standard Oil domination, the internationalization of the industry and the rise of the Seven Sisters. Each phase is described below.

S. C. Bhattacharyya, *Energy Economics*, DOI: 10.1007/978-0-85729-268-1_14,

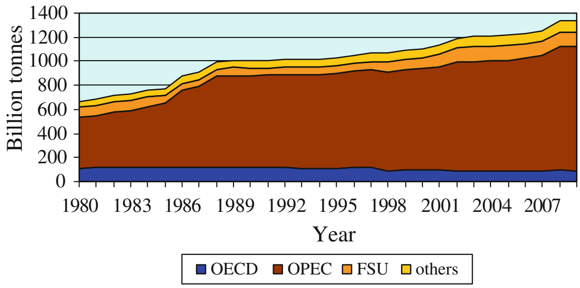


Fig. 14.13 Oil reserve distribution. *Data source* BP Statistical Review of Energy Statistics, 2010

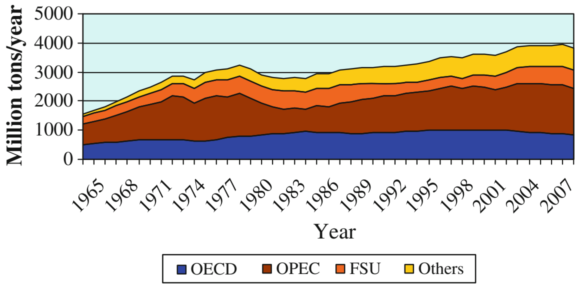


Fig. 14.14 Oil production history. *Data source* BP Statistical Review of Energy Statistics, 2010

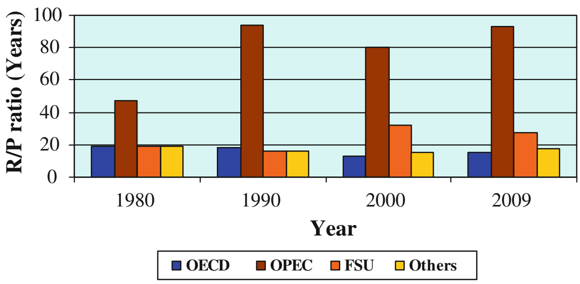


Fig. 14.15 Changes in the oil R/P ratio. *Data source* BP Statistical Review of Energy Statistics, 2010

But this practice of “sweating out” of the reserves in the OECD region leaves the countries vulnerable in terms of their ability to supply in the future. The reserve to production (R/P) ratio for oil is the lowest for the OECD region (see Fig. 14.15). The preference for short-term gains by private companies compared to the societal preference for long-term benefits drives such a development.

As indicated earlier, oil demand traditionally originated from the OECD countries (see Fig. 14.16). More than 70% of oil demand came from this region in 1965 but the share has been falling as the demand from developing countries started to pick up in the 1980s. Although OECD demand still accounts for more than 50% of global oil demand, the developing country share has reached above 40% in 2009.

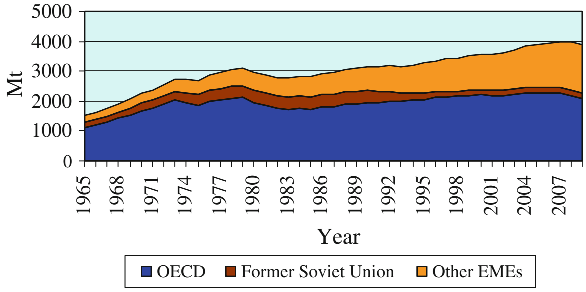


Fig. 14.16 Oil demand trend. *Data source* BP Statistical Review of Energy Statistics, 2010

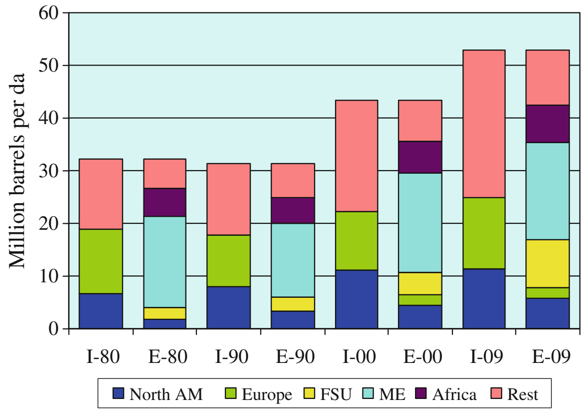


Fig. 14.17 Changes in oil trade movements

China has emerged as the second largest oil consuming country in the world after the USA, while India became the fourth largest oil consuming country in 2009. There are now indications that the oil demand in the industrialized world has past its peak and is in the decline phase. The average growth rate of demand between 2000 and 2009 was −0.7% in the OECD region, as opposed to a growth of 3.5% in the rest of the world (excluding the Former Soviet Union countries). China's oil demand has grown at an average rate of about 7% during this period, showing clear indications of a major shift in the centre of attention in terms of global oil requirements.

As a consequence of the regional demand–supply imbalances, the trade volume has been growing over time (See Fig. 14.17). While Europe and North America (mainly the USA) remain major importers, the growth in trade since 1990 is originating from the other areas (mainly Asia–Pacific). The level of oil import has more than doubled in this region between 1990 and 2009. In terms of sources of supply, the return of the Former Soviet Union supply to its normal level and a greater participation in international trade is clearly evident. The share of the Middle Eastern supply in the trade did not change significantly, which implies that a greater diversification of sources and trading partners has occurred over the past two decades.

14.3.2 Constrained Majors

The history of the 150 year-old oil industry has been dominated by the international oil companies (IOC). Only since the emergence of OPEC and the subsequent nationalization of the oil industry, the national oil companies (NOC) became relevant. Although NOC depended on various services by the IOC and still in many countries the co-operation continues, the canvass has changed quite dramatically. Over time, the role and the power of the NOC became more important and according to Jaffe and Soligo (2007), "14 out of 20 top upstream oil and gas companies in the world are national oil companies" in terms of reserves holdings. In the 1970s and 1980s, when OPEC was pursuing the policy of market control by restricting its output, IOC have invested heavily in non-OPEC countries and benefited from OPEC market regulation which reduced the market uncertainty.

However, as the old fields deplete and concerns for long-term supply security start to bite, IOC start to face stiff competition for acquiring non-OPEC opportunities. The aggressive expansion of China in acquiring overseas reserves, coupled with similar strategies by other developing countries has reduced access for IOC. Chinese success followed a different strategy where the oil company or the Chinese government entered into strategic alliances with the host government for wider economic development of the host country, thereby giving it a special advantage compared to IOC offers. At the same time, the high oil prices of the new millennium have also revitalized resource nationalism especially in Venezuela and Russia. Consequently, IOC while still healthy in financial terms, are finding it difficult to replace their reserve to ensure future sustainability (see Fig. 14.18).

Moreover, the production is already declining in the case of a number of majors and their R/P ratio in recent times is precariously low—ranging between 6 and 13 years (see Fig. 14.19). Contrast this with the average R/P ratio of OPEC of above 90 years in 2009—clearly, the effect of limited access to reserves on the Majors' activities becomes evident.

Oil Major's low R/P ratio along with the precarious R/P ratio of OECD countries clearly justifies their concern for long-term oil supply security. No wonder that the debate about peak oil has a strong developed country bias—and

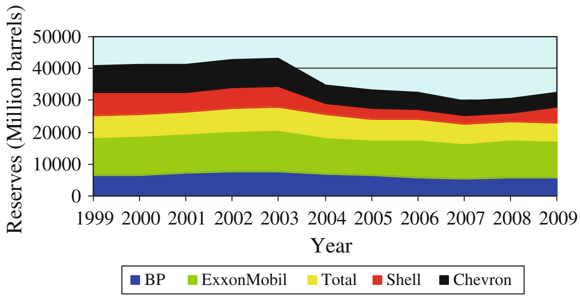


Fig. 14.18 Declining Major's reserve *Source* OPEC Annual Statistical Bulletin 2003, 2008 and 2009

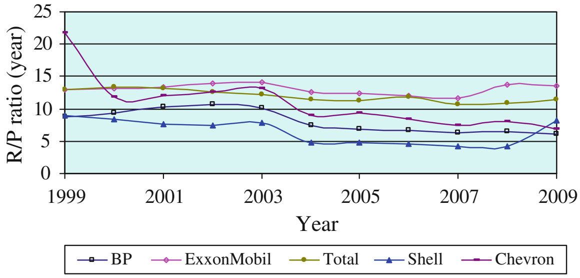


Fig. 14.19 R/P ratio for oil of selected Major oil companies. *Source* OPEC Annual Statistical Bulletin 2003, 2008 and 2009

support of the Majors. After all, the future of major oil companies looks very uncertain if their exclusion from the oil-rich region continues.

14.3.3 Analysis of the OPEC Behaviour

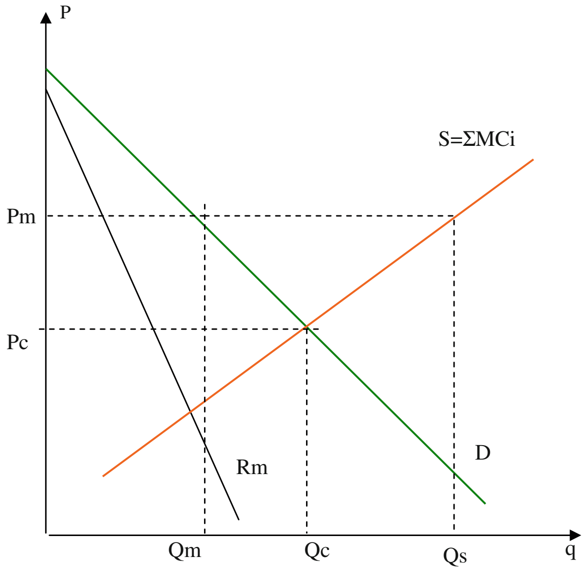
There is a vast literature analyzing the OPEC behaviour and strategies (see for example Slant 1976; Percebois 1989; Greene 1991; MacAvoy 1982; Griffin 1985; Dahl and Yucel 1991, Alhajji and Huettner 2000a, b and Ramcharran 2001). As usual in such an area, there is no consensus about how best OPEC can be described. This difficulty arises because OPEC has followed different strategies at different times to determine prices and production levels (Fattouh 2007). In this section, a simple, diagrammatic presentation of the models analyzing OPEC behaviour is presented.

The models on OPEC behaviour can be categorized into broad groups of models: (a) cartel models such as the dominant firm model or (b) non-cartel models such as target revenue model, and the competitive model. Only a few models were statistically tested and results have been contested by others due to model weaknesses.

14.3.3.1 Cartel Model

A cartel occurs when a group of firms or organizations enter into an agreement to control the market by fixing price and/or limiting supply through production quotas. A cartel may work in a number of ways: as if there is a single monopoly producer, or with market-sharing agreements. The objective is to reduce competition and thereby generate higher profits for the group. In the absence of any agreement, the competitive market conditions will prevail and the price p_c and quantity q_c will be obtained in Fig. 14.20. However, if the producers enter into an agreement to enforce a monopoly price (p_m) in the market, they will have to agree to reduce supply to q_m in such a way that the marginal revenue equals the marginal cost. Each member of the cartel then receives a higher price for the output but any

Fig. 14.20 Cartel as a monopolist



producer will be interested to participate only if it can extract more benefits compared to a competitive environment. As long as this condition is satisfied, members will be happy to support the collusive behaviour.

But, each member would have the tendency to increase its output based on its marginal cost of supply so that its individual profit is maximized. This tendency to cheat will lead to an overproduction (q_s) and the market will see a return of the competitive market price. This represents a natural threat for internal cohesion of any cartel.

Any cartel thus faces a number of problems: in most jurisdictions it is illegal for firms to enter into such a collusive behaviour. OPEC as a group of sovereign nations escapes from this argument. The tendency to cheating by members for individual gains by undermining the collective position is another major threat. Finally, in order to control the market, the group must have accurate information about the shape of the demand and supply curves, elasticity of demand, and actual production by members. Often this information is not readily available although some generic idea may be available.

Stability of any cartel then depends on a number of factors:

a. Group size: A small group is better placed to have a tighter control than a large group.
b. Group characteristics: Homogeneous group members acting on a product with a captive demand or inelastic demand is more likely to succeed than a heterogeneous group.
c. Dispersed, large number of buyers: Widely dispersed consumers will have little chance of colluding with each other. This is an essential requirement for a cartel.

d. Member gains: Each member of the cartel must benefit from the action—otherwise, there is no incentive to join the group.
e. Group discipline: A group that is committed to play by the rules of the cartel is also an important condition.
f. Policing: Any cartel being vulnerable to cheating would need an effective policing mechanism to detect cheating.

Clearly, these requirements are difficult to satisfy in reality.

14.3.3.2 Cartel with a Leader (Dominant Firm Model)

Because of the inherent issue of cohesion, a cartel needs to ensure that the group is able to maintain the market power even if some members are cheating. A cartel with a leader is such a cartel where one of the members can regulate his behavior to maintain the group coherence and can make the group agree to its proposals, thereby protecting leader's interest. A leader should have an important market share, high flexibility in capacity utilization, low financing requirement, and be less sensitive to changes in energy markets.

Consider that the leader knows the market demand D_T well and understands the supply curve of other cartel members (S_O). Based on these, the leader determines its own demand curve D_L such that $D_L = D_T - S_o$. Note that the leader will have no demand when the price rises to $p1$ and it faces the entire demand when the price falls to $p2$ level. The leader then decides its output to maximize its profit and would impose the price p_L (Fig. 14.21) on the cartel. Rest of the members produce $q_T - q_L$. The leader is the price maker and the rest are price takers. The elasticity of

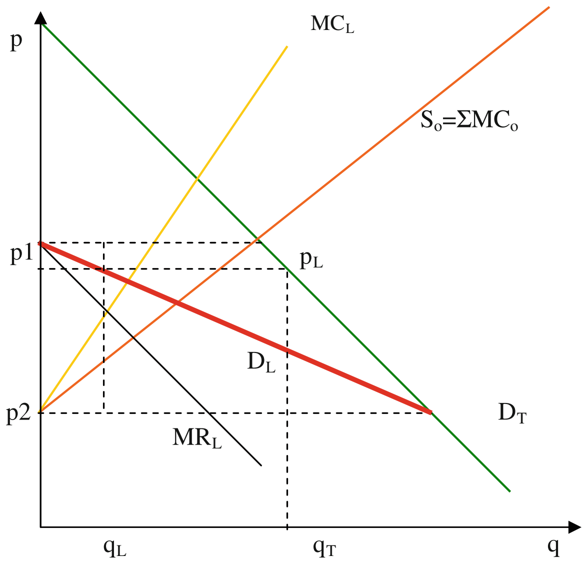


Fig. 14.21 Cartel with a leader

the country is not enough to meet the investment demand, the country would cheat or seek an increase in share. If share is more than that required to meet investment demand, the country may voluntarily reduce output. Only members who are marginal in oil resources would have tendency to cheat. Rich members may not prefer to leave oil to ground as the return may not be remunerative. Small producers may like to defer production. The behaviour of OPEC production in the mid-1970s was consistent with the Target Revenue Model (Ramcharran 2001).

A significant amount of research work has since been done to verify, extend and refine the initial theory. For example, Cremer and Salehi-Isfahani (1980) retained the target revenue hypothesis but modeled OPEC in a competitive framework. Teece (1982) also analysed the OPEC behaviour using the competitive framework with a target revenue constraint. But his analysis differed from Cremer and Salehi-Isfahani (1980) in a number of respects. Teece (1982) considered investment and expenditures as fixed whereas Cremer and Salehi-Isfahani (1980) considered them as endogenous variables. Griffin (1985) used quarterly data for the period between 1973 and 1983 and adopted an ordinary regression model to test the target revenue hypothesis but did not find strong support to the idea. Dahl and Yucel (1991) tested two variants of the competitive model and found no support for the competitive hypothesis. Alhajji and Huettner (2000a) modified Griffin's model by using static and dynamic econometric models but the static models did not give good estimates due to auto-correlation problems. Even the dynamic models did not find support to the target revenue hypothesis. Using a longer set of data Ramcharran (2001) examined the production behaviour of OPEC and non-OPEC countries and found some support for the target revenue hypothesis.

The above shows the differing views on the subject. The results often reflected the choice of the model, data set used and the econometric method used. Empirical evidence did not provide any conclusive outcome on the issue.

Irrespective of the approach used in analyzing the OPEC behaviour, it is important to note that the group represents the interests of major oil producers. Given the size of their reserves, they will remain an important player in the oil market and the group cannot remain idle to any challenge that tries to destroy the captive oil demand arising from the transport sector.

14.3.4 A Simple Analytical Framework of Oil Pricing

To end this chapter, a simple diagrammatic framework is presented based on Stevens (1995), (1996) that combines the supply and demand curves of oil and can be used to explain the oil price movements. The framework is captured in Fig. 14.24.

The demand curve has three segments—highly inelastic for a wide range of prices, with some elastic segment at very low and very high prices (shown as D in the figure). This arises due to capital intensive nature of the appliances used for consuming oil. In the short-run, only some adjustments in the capacity utilization

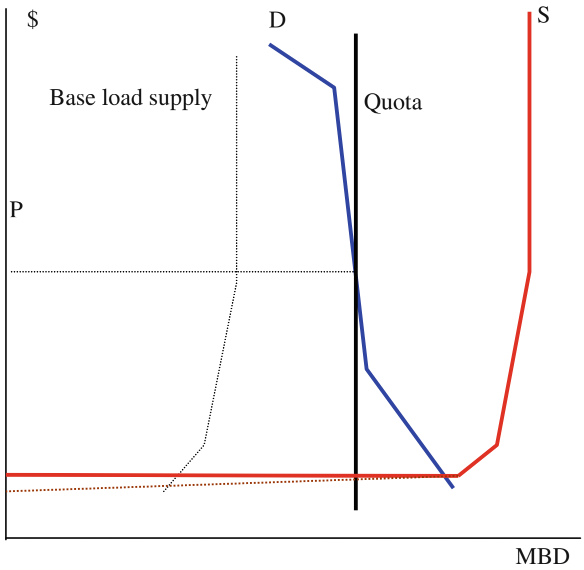


Fig. 14.24 A diagrammatic framework for oil prices

of the appliance is possible, making the demand inelastic over a certain price range. At very high prices, substitutes will appear and make demand elastic. Similarly, at very low prices oil would replace other fuels and therefore would have a greater elasticity of demand.

Similarly, the supply curve has a low cost segment, followed by an increasing cost segment. The horizontal segment represents the low marginal cost of oil supply. This is assumed to be same throughout the world but the argument does not change even if an increasing marginal cost argument is used (as shown by the dotted line). The vertical segment of the supply curve represents the change in the marginal cost due to the fixed capacity (or the capacity constraint) at any given time.

Two groups of suppliers are considered—base load and residual suppliers. The base load suppliers are price takers and supply to capacity for a given price. The residual suppliers are price makers and try to regulate the price by controlling their output. If the price regulation is not used, the supply and demand in the market will decide the price and the market clearing price will be the marginal cost-based. For a target price, the residual producers are then striving to set a quota that yields the desired result. However, this act requires accurate information about the demand and supply. If the information is imperfect, "between a wide range, any price could be regarded as an equilibrium price which 'cleared the market'" (Stevens 1996).

This simple demand–supply based framework can be used to explain the price movements in the oil market. For example, in 2008 when the demand moved outwards, the supply became capacity constrained. Consequently, the prices reached very high levels but the economic consequences of such high prices resulted in demand destruction and the demand curve moved inward to result in a sharp price drop.

Chapter 20
Energy Security Issues

20.1 Introduction

Given the paramount importance of energy for all economic activities around the world, issues related to energy security have gained importance in the wake of recent high oil prices and the fear of supply shortages for natural gas and electricity in many countries. Energy security concerns first emerged in the aftermath of the first oil shock in the 1970s, when oil importing countries were caught unguarded and had to struggle to cope with the adverse effects of oil price rise. Since then countries have followed diverse policies to mitigate the problem. Low oil prices since mid-1980s and the shift of focus in the 1990s to market reform and restructuring meant little attention to the issue of security of supply. It was believed that markets would be able to solve the problems of the energy sector. However, concerns about peaking of oil supply and supply capacity to match the demand have brought back an era of sustained high oil prices. Once again the issue of energy security has become a major policy concern.

This chapter intends to provide an understanding of the concept, its economic dimension and an analysis of various alternative options to deal with it.

20.2 Energy Security: The Concept

"Energy security is commonly defined as reliable and adequate supply of energy at reasonable prices" (Bielecki 2002). Reliable and adequate supply implies uninterrupted supply of energy to meet the demand of the global community. This segment of the definition establishes the link between adequate supply and energy demand at any given time. Supply adequacy and reliability is not a matter of external dependency alone. In many countries (developing and developed) the internal sources of supply could equally be problematic. However, of the literature on energy security focuses on external supply alone as the control over external supply can be limited in most cases.

S. C. Bhattacharyya, *Energy Economics*, DOI: 10.1007/978-0-85729-268-1_20,

Reasonable price on the other hand is a more difficult term as there is no universally accepted benchmark. Economically it would mean market-clearing price in a competitive market where supply and demand balances. But as we shall see below energy security involves externality and therefore internalisation of costs would be essential for efficient resource allocation.

The term is used by different people to mean different things and accordingly, energy security has geopolitical, military, technical and economic dimensions (Bielecki 2002). There is a time dimension of it as well: in the short-term, the main concern relates to the risks of disruption to existing supplies essentially due to act of god, technical or political problem; in the long-term, the risks related to future energy supply also arise.

Like any other concept, this concept is evolving as well. For example, initially, the focus was only on oil and oil products. Now it covers all energies and various types of risks to reliable and adequate supplies (including accidents, terrorist activities, and under investment). The geopolitical, internal and temporal aspects of the issue require a multi-dimensional policy approach to deal with the problem.

The literature has focused on the oil supply security in particular and identifies a number of components of the energy security problem (Toman 2002): (a) exercise of market power by suppliers to raise prices, (b) macroeconomic disruption due to energy price volatility, (c) threats to infrastructure, (d) localised reliability problems, and (e) environmental security. But the problem is not limited to oil supply alone and recent studies focus on the entire gamut of the problem.

20.2.1 Simple Indicators of Energy Security

Two types of indicators are commonly used in the supply security literature: an indicator that expresses the level of exposure in terms of dependence level and an indicator of vulnerability. The level of import dependence of a fuel provides an idea about the price and quantity risks associated with importing the fuel and accordingly, a higher level of imports is generally considered to be a riskier option. Similarly, in the case of an electricity system, high dependence on a single fuel is considered to be a riskier option. But as the risk of supply disruption is associated with the concentration of supply sources and the probability of disruption of supply from each source, a highly import dependent system that is well diversified need not necessarily be a risky one.

20.2.1.1 Indicators of Dependence

Indicators that are relevant for energy diversity and energy security are (IAEA 2005):

(1) Import dependence—this indicator can be used for the overall supply position of a country or a region or for a particular fuel. For example, the ratio of net

energy imports to the primary energy supply in a particular year would provide how reliant the country is on imported supply. If a country consumes 100 Mtoe of primary energy and 90 Mtoe is imported, its import dependence is 90%. High import reliance normally tends to increase the price risk and volume risk related to supply interruption.

Import dependence at the fuel level shows the degree of exposure for each fuel. Often, the import dependence of different fuels varies significantly and a country could have a high import dependence for one fuel but highly self-sufficient in another.

At a more disaggregated level, the import dependence by origin of supply could provide a more accurate picture about the risk. If a country depends on a single country for its imports, the risk is particularly high. On the other hand, a diversified source of imports could reduce the risk of supply disruption.

High import dependence of a fuel does not necessarily mean high risk for a country. It depends on a number of factors: the importance of the fuel in the overall demand; how diversified is the source of supply; and the amount of market power of the suppliers. If all of these factors tend to be adverse for a country, the risk will be high.

The evolution of import dependence of a country can be viewed from a plot of the ratio over a period of time. Similarly, using supply forecasts, the expected changes in the future can be captured.

(2) Fuel Mix—this indicator basically shows the share of a particular fuel in the energy demand of a country or its importance in the energy supply. Depending on the focus of the analysis, this ratio can be determined at different levels:

   (a) The primary energy consumption mix tells how diversified the overall energy demand is. For example, if a country used 90% oil and oil products and 10% gas to meet its primary energy demand, it cannot be said to have a diversified fuel mix.
   (b) The final energy consumption mix gives an indication of fuel diversity at the end-user level.
   (c) The sector level fuel mix provides a similar picture at the end-use sector level. The extension of the analysis at the sector level provides a clearer picture of vulnerability of different sectors. For example, if the industry relies only on electricity and natural gas for its energy needs, and if electricity is dependent on natural gas supply, then the industry is highly exposed to changes in the natural gas supply.
   (d) Electricity generation mix tells which fuels (and technologies) a country uses for its electricity supply.

An analysis of the fuel mix trend can be used to identify any possible adverse changes in the fuel diversity. Corrective policies can then be considered. Similarly, forecasts of future fuel mix can suggest if the country is moving in the right direction or not. For example, the expected closure of coal and nuclear power plants in the UK by 2025 is expected to increase the share of gas in the electricity generation mix. With domestic gas supply declining, such reliance of gas-based power would necessitate gas imports, making the country vulnerable.

(3) Stocks of critical fuels—this indicates the availability of national stocks of a fuel and the length of time that the fuel could be used if supply disruption takes place, assuming current level of consumption. For example, IEA member countries maintain a 90-day stock of critical fuels.

20.2.1.2 Indicators of Concentration and Diversity of Supply

The following indicators are commonly used:

(a) Herfindahl–Hirschman index: The Herfindahl–Hirschman Index (HHI for short) is generally used for market concentration analysis. This is measured by the sum of the squares of the individual market share of each firm in the industry. The HHI ranges from 0 to 10,000, with the lower range obtained when very large number of firms exist in the industry and the higher range reached with a single producer.

The Herfindahl–Hirschman Index is represented as:

$$HHI = \sum_i x_i^2 \tag{20.1}$$

where x_i is the market share.

The level of concentration is high with HHI above 1800. For energy security purposes, the HHI Index can be used to measure the level of concentration of imports from different sources. Thus, by considering x_i to represent the proportion of imports from supply origins, the level of import concentration can be measured.

The HHI has its own shortcomings as it fails to take into account domestic production. It cannot take the political risk into consideration. Percebois (2007) indicated that the HHI of French oil import in 2004 was 2538 and it was 2469 for natural gas. In 2005, the European Union of 25 had the HHI of 2544 for oil imports and 3538 for gas imports. These indices show high levels of import concentrations.

(b) Shannon–Wiener index: The Shannon–Wiener-Index (SWI) is a diversity index. The SW index for the share of imports from different sources is given by:

$$\mathrm{SW} = -\sum_i x_i \ln(x_i) \tag{20.2}$$

where x_i represents the import share from each country (or source). The negative sign at the front of the equation makes sure that the outcome of the SW index is always positive. When all imports come from a single source, the minimum value is reached (which is zero). As the number of countries supplying the fuel increases, the SW index also increases. Therefore, a higher value of the calculated SWI means good situation as regards imports diversification and supply security while a lower value means a worse situation. The main limitations of the HHI remain here also: it cannot take domestic production separately from the imports and the political risk cannot be incorporated.

The UK Energy Digest provides the SW index for the power generation diversity in the country.

(c) Adjusted Shannon–Wiener–Neumann index (SWN index): The adjusted Shannon Wiener Neumann Index (SWNI) removes the limitations of the Shannon–Wiener-Index (SWI). If the political stability factor is included alone, the index takes the form

$$\mathrm{SWN1} = -\sum_i b_i x_i \ln(x_i) \tag{20.3}$$

where b_i is the political stability factor of the country from where imports are coming. The World Bank Report on Governance Matters can be used for the political stability factor. Imports from unstable regions of the world tend to reduce the original Shannon-Wiener-Index and vice versa.

To include the share of indigenous production, the SWN index can be modified as follows:

$$\mathrm{SWN2} = -\sum (b_i x_i \ln(x_i)(1 + g_i)) \tag{20.4}$$

where g_i represents the indigenous production for the country in question.

20.2.2 *Diversity of Electricity Generation in Selected European Countries*

The diversity of fuel-mix of electricity generation in some European countries is considered below. The analysis is presented using two indices: SWI and HHI.

Table 20.1 presents the fuel mix of electricity generation in 5 European countries retained in this study for 1995 and 2005. As can be seen, coal was displaced by natural gas in the UK to a large extent and in Spain and Netherlands to a lesser extent. In Italy, fuel–oil based generation which was the dominant form of power in the mid-1990s was replaced by natural gas. Natural gas consolidated its position as the leader in the Netherlands during this period. Dependence on fossil fuels in electricity generation remained very high in the Netherlands (88%), Italy (79%) and the UK (above 70%). Spain was moderately dependent on fossil fuels in the mid-1990s but its exposure has increased in 2005 to around 60%.

Table 20.1 Fuel-mix of electricity generation in five European countries

		Coal (%)	Natural gas (%)	Oi (%)	Nuclear (%)	Hydro (%)	Others (%)
UK	1995	57.40	15.50		25.20		1.90
	2005	40.85	36.65		19.75		2.75
Germany	1995	54.02	8.05	1.67	28.73	4.51	4.02
	2005	43.46	11.03	1.70	26.29	4.31	13.21
Italy	1995	9.93	19.46	50.03	0.00	17.36	3.22
	2005	14.36	49.15	15.52	0.00	14.13	6.84
Spain	1995	34.95	2.25	8.74	33.14	14.68	6.24
	2005	25.04	26.87	8.30	19.57	7.83	12.39
Netherlands	1995	32.16	51.85	4.77	4.96	0.10	6.16
	2005	23.45	57.73	2.26	3.99	0.08	12.49

Source Bhattacharyya (2009)

Figure 20.1 presents the level of fuel-mix concentration of generation for the period between 1995 and 2005 using HHI. As can be seen, all the countries chosen in the study have HHI above 2000, indicating that the electricity supply in these countries is highly concentrated. The level of concentration has declined in the UK in the early 1990s and then stabilized. Similarly, Spain and Italy have also recorded some improvement in terms concentration in the later half of the 1990s but the improvement in these two cases were over a longer period compared to the UK. Germany did not show any change in the level of concentration of generation fuel mix over the past decade while the situation has deteriorated in the Netherlands. Of the five countries considered here, Spain had the lowest HHI since 1996 while the Netherlands, with an HHI of above 4000, had the highest over the same period. The dominant position of natural gas with a share of above 50% in the fuel mix of electricity generation has adversely affected the concentration in the Netherlands while a well distributed fuel mix of Spain has clearly improved its level of concentration.

Figure 20.2, which provides the trend of SWI of fuel mix for electricity generation in the above five countries between 1995 and 2005, also leads to the same observations as above. In all the five cases, the SWI ranged between 1 and 2, implying that these countries are not dependent on one or two fuels for their

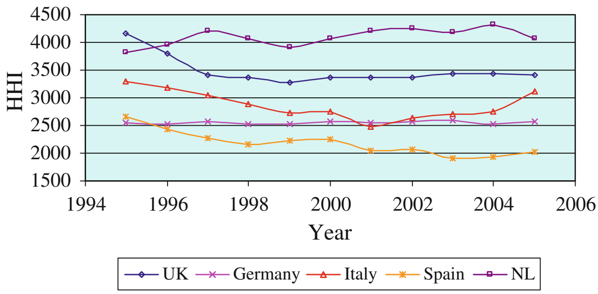


Fig. 20.1 HHI of electricity generation mix in selected European countries. *Source* Bhattacharyya (2009)

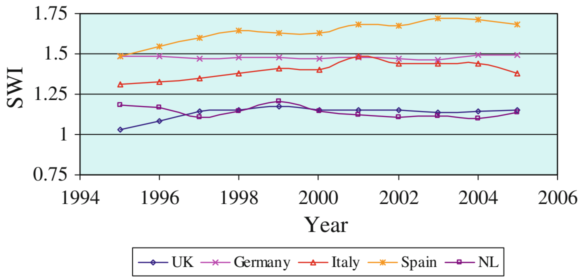


Fig. 20.2 SWI of electricity generation mix in selected European countries. *Source* Bhattacharyya (2009)

electricity generation but their fuel diversity is not highly commendable either. Spain has the most diversified generating system in the sample and the level of diversity has improved during the past decade. Germany and Italy occupy an intermediate position, where the diversity level in the German system has not changed appreciably while the Italian system has recorded an improvement until 2001 followed by a somewhat reduction in the diversity. The liberalised markets of UK and the Netherlands have the least diversified generating systems in the sample and their level of diversity did not change in the past decade.

It is clear that the above five countries rely on fossil fuels to a great extent for their electricity generation. Although their systems are not highly concentrated in terms of fuel mix, they cannot be considered to be in a highly desirable situation either. As the fossil fuel prices have risen in recent times, their electricity system is likely to be vulnerable. It is to this aspect that I now turn to.

20.3 Economics of Energy Security

Energy supply disruptions consider interruptions of supply due to a variety of factors: act of sabotage, failure of a supply technology, breakdown of supply infrastructure, etc. The level of insecurity is reflected by the risk of a physical, real or imaginary supply disruption (Owen 2004). Normally, a high level of insecurity would result in high and unstable prices over a prolonged period.

In order to understand the economics of energy security, it is important to categorise the sources of insecurity. Two types of supply disruption risks could be considered (Markandya and Hunt 2004): strategic and random. A strategic risk would arise due to political instability, market power or even inadequate investments in supply facilities. OPEC deliberately manipulating the supply and prices comes under this category. Random shocks such as terrorist acts on the other hand are more speculative in nature and may not follow any set pattern. Although these risks could affect both domestic and the international markets, the strategic risk has less relevance for the domestic systems. The domestic systems on the other hand could face supply disruption due to insufficient infrastructure, technical failures, social unrest, or due to acts of terrorism (Owen 2004).

This section focuses on the economic aspects of two main components of the energy security issue: the effect of market power on the cost of imported energy and the cost of supply disruption. Oil is used as an example as it is the most traded commodity in the world market and oil imports account for a significant share of imports in many countries. However, the same logic applies to other energies to a great extent.

First, the cost of oil imports is presented. This is followed by a discussion of the cost of supply disruption and analysis of measures to mitigate the risks.

20.3.1 External Costs of Oil Imports[1]

Although oil is a commodity, it has a certain special characteristics:

(a) oil is concentrated in a relatively small area in the Persian Gulf, which allows for monopolistic behaviour in the oil market;
(b) oil has limited (if at all) substitutes in its main uses, which removes the flexibility of users to move away from use of oil;
(c) oil supply shocks may leave nations to serious adjustment problems; and
(d) all stages of oil fuel cycle impose unintended and damaging environmental effects.

Consequently, the market failure argument applies here and the market price of delivered oil to the consumers departs from the full social cost of oil. The social costs may include costs due to non-competitive markets, costs due to environmental damages, and economic losses due to price shocks. Oil consumers do not pay for these costs in the price but the society as a whole pays for them.

One commonly identified externality related to oil import arises due to the monopsony power of certain importers that affect the price of oil in the world market. For a price taker in the international oil market, the price paid by the consumers is equal to the cost of the extra oil to the economy and hence there is no externality here. But if a consumer has a large market share in consumption (say the US), then any extra demand for imports by this consumer would adversely affect the global demand and consequently, the world oil price would increase. This raises the country's total oil import bill—for marginal and infra-marginal imports. While the private cost to consumers is the marginal cost of imports, the society bears the cost higher payments for the infra-marginal quantities, making the social cost higher than the private cost. The difference between the social and private costs is called the monopsony wedge.

The logic of externality would suggest that the market does not convey the correct signal to the consumers and accordingly, the consumption decision would

[1] This section relies on Leiby et al. (1997), Toman (1993), Markandya and Hunt (2004) and Huntington (2009).

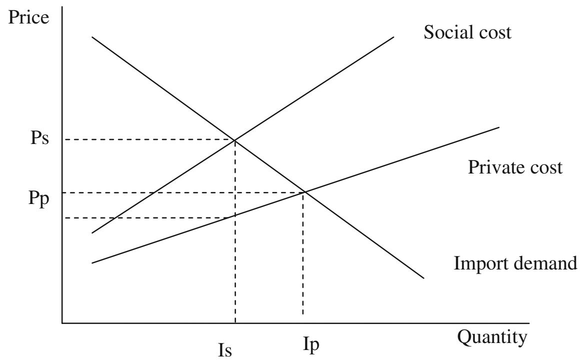


Fig. 20.3 Social cost of oil imports

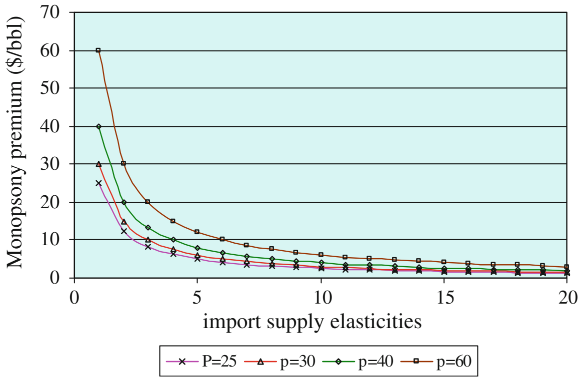


Fig. 20.4 Monopsony premium plot. *Source* Based on Parry and Darmstader (2003)

be based on private costs and not on the social costs. This is shown in Fig. 20.3. While the import based on private cost is Ip, the efficient level of import would be Is based on the social costs.

The effect monopsony power depends on two factors (Parry and Darmstader 2003): the level of import dependence and the effect of monopsony demand on the world oil market. If the country does not depend on import (i.e. import dependence is zero), there is no externality due to monopsony power. With higher level of import dependence, the monopsony wedge increases. Similarly, if the world oil market was perfectly elastic and competitive, the extra import demand from a major consumer would not have any effect on the world oil price and the externality would not exist. But the presence of the OPEC makes the supply non-elastic and the world price is affected by the supply from non-OPEC producers as well.

Parry and Darmstader (2003) suggest a simple relation to capture the monopsony premium or wedge. Generally, if P is the world price of oil and e is the elasticity of import supply, then the monopsony wedge (or premium) is given by P/e. If e is infinite (i.e. the import supply is perfectly elastic), the premium is zero. For various oil prices and import supply elasticities, the premium would vary as shown in Fig. 20.4.

As can be seen from the above plot, the premium could be high for inelastic import supply; otherwise, the premium fall quite sharply and could be low. The literature provides a wide range of estimates for the US, ranging from \$0 to \$14 per barrel, while Parry and Darmstader (2003) prefer to use \$5 per barrel as the premium. However, most of these estimates were based on low oil prices and may not be valid in a high oil price regime. For example, Leiby (2007) estimated the monopsony premium for the US at \$8.9 per barrel (at \$2004 constant prices) considering the conditions prevailing in the new millennium.

20.4 Optimal Level of Energy Independence

Here the marginal cost approach is used to get some idea of optimal dependence. This requires us to construct the curve depicting marginal cost of its import dependence (MDC) and the curve showing marginal cost of security (MSC) as shown in Fig. 20.5 (Percebois 1989).[2]

The marginal import dependence cost (MDC) curve captures the costs of increased energy import dependency. This would include direct and indirect costs to the economy (including military costs, economic disruption costs, etc.). Normally, this curve is expected to be downward sloping with respect to import independence. When a country is fully self-sufficient, the marginal cost of import dependence is zero and it could be very high for 100% import dependence. It is not easy to develop such a curve as the cost depends on many factors such as import diversity, ease of energy substitution, importance given by the society on energy import, etc.

The marginal cost of security curve (MSC) on the other hand is the cost the society is willing to bear for increasing the national energy independence. A country could reduce its import dependence through energy stocks, energy rationing, promoting national supply, etc. The incremental cost of increasing independence would be captured here. It is generally assumed that the marginal cost of security is zero for domestic energy supply (although this need not be true). Costs start to increase at a faster rate with higher levels of independence. So the curve does not start at the origin (there is an offset) and has a steep slope.

The optimal rate of energy independence is given by the intersection of the two marginal curves as shown in Fig. 20.5. The graph suggests that: for an optimal level of energy independence; it is important to consider the costs of ensuring security of supply and the cost of the damage. It is not economically efficient to improve energy independence beyond the optimal level; this is so because the cost of providing the security of supply would be much higher compared to the marginal dependence cost. There is a price (P*) that the society is willing to pay to

[2] This part is based on Percebois (1989).

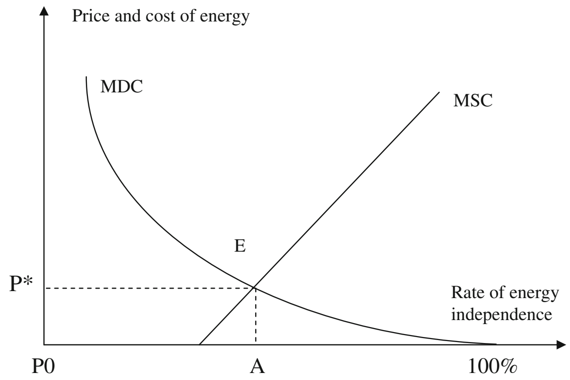


Fig. 20.5 Optimal rate of energy independence. *Source* Percebois (1989)

ensure the optimal level of security of supply—this is the premium that has to be paid to ensure security of energy supply.

20.5 Policy Options Relating to Import Dependence

If oil import imposes external costs to the society, what are the options available to mitigate them? The literature on energy security has considered a number of options and we discuss a few of them in the following paragraphs.

20.5.1 Restraints on Imports

Such a policy aims at imposing import restrictions through tariffs or quotas to mitigate the costs related to import dependence. Alternative policies that would eventually limit energy imports (such as tax on fuels, promotion of domestic supply, fuel substitution, promotion of alternative sources of energies, etc.) could also be considered under this category. We analyse the economic logic of using import quota and import taxes.

20.5.1.1 Effect of Import Tax and Import Restriction

Let us consider an energy importing country whose energy demand and domestic supply are given by schedules D and S respectively in Fig. 20.6. If the country does not participate in international trade, the domestic price would be the market clearing price p1. Assume that the international price p2 and is lower than p1. In an open economy, the supply would be met by a combination of local production and import. The country will produce q3 and import q2–q3. This volume of import would involve a significant foreign exchange outflow for the country.

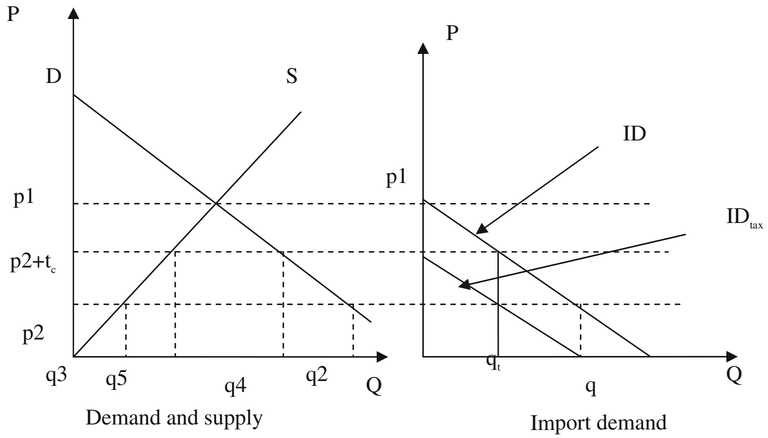


Fig. 20.6 Effect of import tax and import restriction

Consider now that the government is concerned about the energy security and that it imposes an import tax equal to t_c per unit of import. With this tax, importing energy would be costlier which makes import to shrink. The domestic supply would be encouraged at this higher level of price, as more domestic suppliers would be willing to produce. Import volume reduces to q4–q5.

The import demand function for the country is shown in the right hand panel. In absence of any import tax, the import demand is given by ID. At price p1, the import demand is zero but it increases to q when the price is p2. When the tax is imposed, the demand curve shifts to ID_{tax}. At price $(p1 - t_c)$, the demand is zero while with tax t_c, the import volume reduces to q_t. Thus, the import schedule shifts leftwards by $(q - q_t)$.

Now consider the effect of imposing an import quota system. Assume that the government imposes a quota at level q_t (i.e. the imports should not exceed this level). This is shown in the right hand panel. As the imported supply cannot exceed the quota, the price rises to $p2 + t_c$ level, thereby reducing the demand as before. The domestic supply receives encouragement at this price and import remains restricted. In a quota system, the import demand function is represented by $p1Aq_{t.}$ At prices below $p2 + t_c$, the quota is a binding constraint and the level of import remains fixed at q_t.

The tax system is a price-based mechanism. The import demand varies depending on the oil price and the level of tax. The import demand curve is shown by ID_{tax}. The effectiveness of the instrument could be less. The tax revenue accrues to the government. It does not require any additional administrative system. In a quota system, there is no ambiguity about the import level (hence a certain instrument). It requires additional administrative machinery to implement the quota system. It could also lead to corrupt practices (through grant of exemptions) or illegal smuggling of the products. More importantly, the higher revenue goes to the suppliers and not to the government. In fact, as quota is price insensitive beyond a threshold, the exporters have incentives to adjust prices to the higher level.

Thus the two policy options have different economic consequences. In the case of a quota, revenue transfer to the exporting countries would take place, if they are in a position to exploit the situation. It may also involve a higher transaction cost. While in the case of import taxes, the government could earn revenues by reducing import demand.

Therefore, for an importing country it may be beneficial to use an import tax system as long as such a system is compatible with the international trade regimes.

20.5.2 Import Diversification

The logic is simple: do not put all the eggs in one basket. This is because the risk of supply disruption is high when a country relies on a single source for its energy supply (i.e. becomes a captive consumer).

This risk can be mitigated through diversification of the source of supply. From an economic point of view, this implies finding the least-cost supply solution taking supply risks into consideration. However, for oil and to a lesser extent for gas, the global reliance on the Middle East is expected to increase where most of the reserves are located. This coupled with political instability of the region and increasing demand from the developing economies raise concerns for future oil supply security.

Two new developments in the area of import diversification perhaps are worth mentioning.

- The first relates to an increased level of activities and investments in production facilities by importing countries in foreign oil producing regions. Chinese oil companies are now forerunners of this trend and are investing massively around the world. Japan also relied on such a strategy in the 1970s and 1980s although may be less aggressively.
- A second trend appears to be emerging in the form of seeking cooperative solutions rather than relying on competitive outcomes. This trend is noticed in various areas:
  - *Importer-importer co-operation*: China which was engaged in competition with India through rival bidding for acquisition of energy assets elsewhere have now joined hands to jointly develop and acquire such assets. The cooperative strategy is expected to reduce the cost of procurement (and hence the supply) and better use of other resources.
  - *Importer-exporter cooperation*: Joint development by importing and exporting countries would ensure flow of required investments for the development of facilities and could reduce transactional risks.

The framework of cost-benefit analysis plays a vital role in such decisions. A nationalised company can employ a different threshold for decisions compared to a private company (regarding discount rates, profitability ratio, future market conditions, etc.). The long-term nature of these investments and uncertainties about the future as well as risk-averseness of the investors would influence the decisions. However, wrong investment decisions may lead to outgo of significant financial resources and costly supply in the future.

20.5.3 Diversification of Fuel Mix

Diversification of fuel mix in an economy tries to reduce dependence on a particular fuel and to achieve a diversified portfolio of energy supply options. For example, Salameh (2003) indicates that the US has been diversifying its fuel mix for ages to replace oil and coal by natural gas and nuclear. In the future, renewable and other technologies on which it is investing heavily could add more diversity.

The choice is often limited by: the availability of resources, available technological options to exploit such resources, costs and investment requirements, and other considerations including environmental and social concerns.

It is difficult to generalise but a few trends could be indicated.

(a) **Effects of restructuring on fuel diversity in electricity**: Reliance on market forces upon restructuring and reform of the energy industries in the 1990s led to promotion of competitive solutions in the electricity markets. This has resulted in a shift in technology choice for supply as the private investors are now looking for quick recovery of investments. Consequently, low cost options are being preferred compared to capital intensive solutions, reducing supply diversity.
(b) **Come-back fuels**: Coal and nuclear are re-emerging as preferred alternative options for power generation. Stability of coal prices, availability of technological options and higher availability of coal in the demand areas has created a positive mood, although environmental considerations act as a hindrance. Security of supply is forcing many countries to rethink about the nuclear option.
(c) **More renewable energies**: Renewable energies are being promoted for various uses to replace or reduce reliance on fossil fuels, thereby adding diversity and improving security. Various policies such as renewable energy targets or obligations, fixed feed-in tariffs, quicker depreciation and recovery of capital, and fiscal incentives are being used to promote renewable energies.

20.5.4 Energy Efficiency Improvements

Efficient use of energy reduces energy demand, which in turn reduces import requirement. This also reduces environmental damages and resource depletion.

Although significant efforts have gone into energy efficiency improvements and demand-side management programmes, availability of cheap energy has reduced their appeal in the past. With higher energy prices, it could again become easier to pursue some of these objectives.

In this respect, the importance of rational energy pricing needs to be emphasised. If domestic retail prices are maintained at inefficient levels, consumers remain insulated from the price movements and do not appreciate the need for efficient use of energies. Removal of energy subsidies could provide the necessary incentive to consumers, although efforts so far in this direction have yielded little result. The efforts are hindered by non-availability of information, need for sophisticated decision-making, use of non-standard procedures, etc.

20.6 Costs of Energy Supply Disruption

Any supply disruption imposes some costs on the economy due to loss of economic activities, price effects and costs of alternative supply arrangements. For oil, it is considered that the supply interruption will lead to higher import prices, given the dependence of the economy on the imported energy source. This then results in economic loss directly through loss of outputs, unused factors of production, cost of stand-by generation capacities, etc., and indirectly, through increased cost of business due to inefficiencies, misallocation of resources, etc.

The estimation of disruption cost involves the following steps (Razavi 1997):

- formulation of supply interruption scenarios providing information on the volume of supply unavailability over expected disruption periods; The level of insecurity is reflected by the risk of a physical, real or imaginary supply disruption (Owen 2004). Normally, a high level of insecurity would result in high and unstable prices over a prolonged period.
- assessment of how prices would be affected due to such supply interruptions.
- an estimation of GDP loss due to price increases.

Leiby (2007) suggested that the above can be represented as follows:

$$E_{\{\Delta Q\}}[C_d] = \sum \phi_j[C_{Id}(\Delta P(\Delta Q_j)) + C_{GNPd}(\Delta P(\Delta Q_j))] \tag{20.5}$$

where C_d = cost of disruption
C_{Id} = cost due to import disruption
C_{GNPd} = cost of losses due to economic dislocation
ϕ_j = annual probability of supply losses
ΔP = price change
ΔQ = quantity change
$E\ (C_d)$ = Expected cost of disruption

The disruption premium is obtained by considering the marginal change of the above expected cost with respect to import quantity. Leiby (2007) estimated the disruption premium for the US at $4.68 per barrel of oil at ($2004 constant prices). However, as can be imagined, the estimation of such a premium is not easy and involves a large number of assumptions and forecasts about future events. Thus the estimates vary depending quite significantly depending on the choices made.

An understanding of the disruption cost is important for deciding the mitigation strategies. If the cost is high, higher levels of supply reliability could be justified and vice versa.

20.6.1 Strategic Oil Reserves for Mitigating Supply Disruption

The Strategic Petroleum Reserve was a response of the developed countries to the oil price shocks of the 1970s. The objective was to provide a deterrent to deliberate, politically motivated reduction in supplies. This initiative was engineered by the International Energy Agency (IEA) in 1974 under the auspices of the Agreement on an International Energy Program.

Under this agreement, IEA member countries hold a stock of oil equivalent to 90 days of net imports in the previous year. Supply can be released in emergency conditions when the supply disruption exceeds 7% of IEA or any member country supply. Similarly, the EU also has adopted a comprehensive set of measures including the obligation to maintain stocks of three types of petroleum products (namely motor spirit, middle distillates and fuel oil) for at least 90 days of average daily consumption in the preceding calendar year. Although the IEA program and EU measures have some minor variations, the two serve similar purposes and member countries tend to use same stocks for complying with both the obligations (Bielecki 2002).

There are several advantages of such strategic reserves: (a) stock releases pacify markets and dampen price rises; (b) allow time for economies to adjust to the changes, (c) although a few countries are members to the plan, consumers globally benefit from the stock due to market reaction, and (d) they allow room for expanded co-operation among countries. The stockpile can be viewed 'as a publicly provided insurance policy against petroleum market shocks' (Taylor and Van Doren 2005). But what justifies public provision of this service?

Public provision of the stock may be required for a number of reasons (Taylor and Van Doren 2005; Toman 1993):

(a) *non-optimal stockpiling by the private sector*: privately owned inventory may be held at a smaller level than the economically efficient level because:

- the market price may not provide effective signals to investors about the total benefits and costs.
- the presence of externality would create a divergence between the private and social costs and benefits, requiring such an intervention.

- Moreover, the private stockholder may not be able to capture the entire benefit of holding stock when there are significant macroeconomic benefits (Toman 1993; Taylor and Van Doren 2005).
- Finally, changes in the regulatory or fiscal environments could deprive the stockholder some or most of the benefits of holding the stock and thereby discourage non-optimal private stockholding (Taylor and Van Doren 2005).

(b) *Behavioural problem*: private entities guided by profit-maximising behaviour may hold stock rather than releasing it at the time of high prices in the hope of higher profits.
(c) *Cost consideration*: private stockpiling may be costly compared to publicly-owned stockpiling because of technology choice, storage location and size.

However, for any such strategic reserve, a number of issues arise (Toman 1993):

(a) *Reserve sizing*: the sizing of the stock and its use are influenced by the cost of economic disturbance to be mitigated, its probability, size and duration, and the interaction of private and public stocks could also influence the sizing decision.
(b) *Timing and method of stock utilisation*: often the literature on stockpile release profiles provide little help as the models rely on simplified assumptions.
(c) *Arrangements for stock use*: the question of institutional arrangements for using such reserves has been analysed as well. Often it is assumed that the stocks would be sold in the spot market periodically using sealed-bid auctions. Forward sales and sale of options to purchase oil from the reserve at predetermined strike prices are also possible (Toman 1993).

But such reserves also add to the cost (of building and carrying the stock among others) and hence the optimal stock size depends on the costs and benefits derived from the stockpile. Following Razavi (1997) the desired level of stock of strategic reserve (S*) could be determined using a simple framework by comparing the cost of maintaining the reserve and the benefits of avoiding a sudden supply shock (see Fig. 20.7).

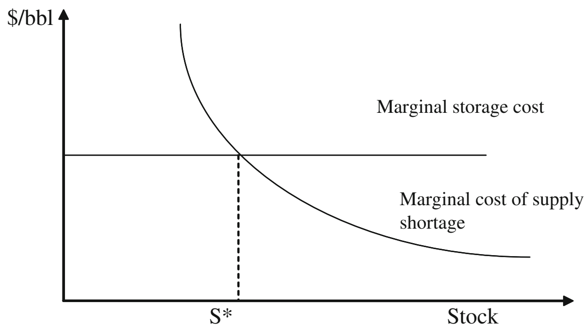


Fig. 20.7 Desired level of strategic stock. *Source* Razavi (1997)

Although strategic reserves are used as a policy option, its costs are not often reflected in the pricing of energy. Taylor and Van Doren (2005) question the economic rationale for maintaining stocks as well for the following reasons:

(a) The cost for maintaining the reserve in the USA was found to be quite high compared to the oil price. They estimate that each barrel of strategic reserve costs the taxpayer between $65 and $80 and maintaining such high cost oil for shortage mitigation does not make economic sense.
(b) The amount of oil stocked is just a fraction of the global oil demand and would not be able to influence the international oil price to any significant level.
(c) The reserves have been used only three times so far in the US history and the timing and volume of stock release did not provide much comfort to the affected population.

20.6.2 International Policy Co-ordination

Security of energy supply has an international public good dimension. This is because measures taken by any country independently would also benefit (or impose costs on) others.

International policy coordination helps avoid free-riding and limit opportunistic behaviour of countries. The crisis-response provisions of the IEA form the essential mechanism for such co-ordination in industrial countries. At a regional level, ASEAN has adopted an Emergency Petroleum Sharing Scheme during shortage and oversupply to assist both importers and exporters of the region (Bielecki 2002).

Any such international mechanism would have to ensure provision of the public good in a fair, cost-sharing programme. Normally larger benefits are expected to accrue to bigger economies. This requires some sort of 'common but differentiated' responsibility approach [adopted for the Climate Change policy coordination] (APERC 2002). Similarly, it may not make sense for smaller countries to go for own strategic reserves due to adverse cost-benefit characteristics and a co-operative solution would be preferable. The possibility of economic and political policy coordination as a group could also be considered.

20.7 Trade-Off between Energy Security and Climate Change Protection

Concerns about the climate change in recent times have imposed an additional consideration in the energy security debate. The diversification of energy supply system to enhance energy security could have a bearing on the climate protection.

Chapter 22
Energy Access

22.1 Problem Dimension

This section presents the gravity of the energy access issue by looking at the present situation and expected future outlook considering the business as usual scenario. Most of the information below is based on IEA reports on the subject.

22.1.1 Current Situation

The most commonly cited figures on the lack of access to energy indicate that there are about 2 billion people without adequate access to clean cooking energy and about 1.7 billion people are without access to electricity (WEA 2000).[1] The origin and genesis of these figures are not easy to find. WEA (2000) does not elaborate on the source of the estimation or the estimation procedure. The World Bank report on Energy Services for the Poor (World Bank 2000) indicates that the estimate of 2 billion people is perhaps outdated. Estimation is difficult due to imprecise definition of the term "access" and lack of good quality data arising from poor understanding of the traditional energy use due to dispersed and distributed nature of this energy and focus on supply of commercial energies in the national energy balances and less focus on where it is used and by whom. Although traditional energies play an important role in many developing countries, the statistics is not reliable and household surveys are not common in all developing countries.

Information on access to electricity is somewhat better. According to WEO (2002), which provided detailed country-wise electricity access information, about 1.64 billion or 27% of the world's population did not have access to electricity in 2000. Since then, IEA has been updating the information on electrification on a

[1] Similar figures are quoted in DfID (2002).

S. C. Bhattacharyya, *Energy Economics*, DOI: 10.1007/978-0-85729-268-1_22,

regular basis and the most recent information suggests that about 1.4 billion population do not have access to electricity. The regional distribution is given in Table 22.1. It shows that two regions have large concentrations of people without access to electricity—South Asia (614 million or 42% of those lacking electricity access globally) and Sub-Saharan Africa (587 million or 40% of those lacking access to electricity).

A closer look at the data shows that about 70% of those lacking access to electricity reside in just 12 countries while the rest 30% is dispersed in all other countries (see Table 22.2). The rural population in most of these countries is lacking access, although in a few countries the urban population also lacks access. While the total number of people without access to electricity is high in South Asian countries, Sub-Saharan Africa fares worse in terms of rate of electricity access. In fact, out of 10 least electrified countries in the world, nine are from sub-Saharan Africa and Myanmar is the only country from Asia (see Table 22.3).

It can also be noted that most of these countries:

(a) have low per capita GDP compared with the world average. Except Indonesia, all countries in Table 22.1 have national average per capita GDP less than 10% of the world average.
(b) have low per capita primary energy consumption, ranging from 8% to 42% of the world average.
(c) Have very low per capita electricity consumption -the national average per capita electricity consumption in these countries ranges between 1% and 15% of the world average.

WEO (2002) provided some details about biomass use in the developing countries and estimated that about 2.39 billion people use biomass for cooking and heating purposes in these countries. This information is available at an aggregated level, which indicates inadequate knowledge about this important source of energy and points to poor quality of information. Subsequently, in 2006, IEA revised the

Table 22.1 Level of electrification in various regions

Region	Population without electricity (Millions)	Electrification rate (%)		
		Overall	Urban	Rural
North Africa	2	98.9	99.6	98.2
Sub-Saharan Africa	587	28.5	57.5	11.9
Africa	589	40.0	66.8	22.7
China and East Asia	195	90.2	96.2	85.5
South Asia	614	60.2	88.4	48.4
Developing Asia	809	77.2	93.5	67.2
Middle East	21	89.1	98.5	70.6
Developing Countries	1453	72.0	90.0	58.4
Transition economies and OECD	3	99.8	100.0	99.5
Global total	1456	78.2	93.4	63.2

Source: WEO (2009)

Table 22.2 Major concentration of population with access to electricity

Country	Rank in terms of population	Population without electricity access (Million)	Share of population without access (%)		
			Urban	Rural	Total
India	2	404.5	6.9	47.5	35.5
Bangladesh	7	94.9	24	72	59
Indonesia	4	81.1	6	48	35.5
Nigeria	8	80.6	31	74	53.2
Pakistan	6	70.4	22	54	42.4
Ethiopia	15	68.7	20	98	84.7
DR Congo	19	57	75	96	88.9
Myanmar	24	42.8	81	90	87
Tanzania	30	36.6	61	98	88.5
Kenya	32	32.8	48.7	95	85
Uganda	37	29.1	57.5	96	91
Afghanistan	44	23.3	78	88	85.6

Source: WEO (2009)

Table 22.3 Reliance on biomass for cooking energy needs in 2004

Region	Total population		Rural		Urban	
	%	Million	%	Million	%	Million
Sub-Saharan Africa	76	575	93	413	58	162
North Africa	3	4	6	4	0.2	0.2
India	69	740	87	663	25	77
China	37	480	55	428	10	52
Indonesia	72	156	95	110	45	46
Rest of Asia	65	489	93	455	35	34
Brazil	13	23	53	16	5	8
Rest of Latin America	23	60	62	59	9	3
Total	52	2528	83	2147	23	461

Source: WEO (2006)

estimate upward to 2.5 billion. This remains the most recent estimate on the use of biomass for cooking purposes. Table 22.3 presents some details about traditional energy consumption in developing countries.

Clearly, such a heavy reliance on traditional energies imposes economic cost on the society. Combustion of household fuels leads to air pollution. As biomass is often used in inefficient stoves, one-fifth of the fuel may be diverted as products of incomplete combustion, thereby creating health hazards. Air pollution is also a concern where coal is used as household energy. Coal smoke contains particulate matters as well as emission of health damaging contaminants. The local level pollution arising from liquid and gas based petroleum products is relatively less due to higher efficiency of cook stoves and better fuel quality.

Combustion of biomass energy indoor is a major source of indoor air pollution. The timing of such pollution (when most of members of the family are present) and the level of exposure due to poorly ventilated houses make poor households vulnerable to serious health effects. Four main health effects are attributed to household use of solid fuels (WEC 2000):

(a) infectious respiratory diseases;
(b) chronic respiratory diseases;
(c) premature deaths
(d) blindness, asthma, heart diseases etc.

As a consequence, 1.5 million premature deaths occur that is directly attributable to high indoor air pollution (WEO 2006), which represents a major heath risk in the developing countries. The regional distribution of these pre-mature deaths follows the biomass use patterns and South Asia and Sub-Saharan Africa suffer the maximum loss in this respect. Many millions also suffer from other lung and respiratory diseases as a result of pollution from burning traditional energies. As women and children are more exposed to such conditions, they are more vulnerable.

22.1.2 Future Outlook

But more importantly, forecasts by IEA suggest that unless policies are implemented to address the access issue, the number of people without access will not decline in the 2030 horizon. Although 75 million is expected to gain access to electricity every year until 2030 (WEO 2002), increases in the population in developing countries of South Asia and Sub-Saharan Africa will mean that electricity access will remain a problem. According to WEO (2002) 680 million in South Asia and 650 million in Sub-Saharan Africa will still live without electricity access. Significant improvements in the rest of the world are expected by this study (see Table 22.4).

The situation will be quite similar in the case of traditional energy use for cooking purposes (see Table 22.5). WEO (2006) suggested that the number of

Table 22.4 Expected future electrification rates (%). *Source* NSSO (2001b)

Region	2002	2015	2030
Sub-Saharan Africa	24	34	51
North Africa	94	98	99
South Asia	43	55	66
China and East Asia	88	94	96
Latin America	89	95	96
Middle East	92	96	99
Total	66	72	78

Source: WEO (2004)

Table 22.5 Outlook for biomass use for cooking in 2015 and 2030 (million)

Region	2004	2015	2030
Sub-Saharan Africa	575	627	620
North Africa	4	5	5
India	740	777	782
China	480	453	394
Indonesia	156	171	180
Rest of Asia	489	521	561
Brazil	23	26	27
Rest of Latin America	60	60	58
Total	2528	2640	2727

Source: WEO (2006)

people using biomass will increase in the 2030 horizon. Most of the population relying on biomass for cooking will live in Asia and Sub-Saharan Africa.

22.2 Indicators of Energy Poverty

As energy is an essential input for economic development of any country, consequently low access to clean energy hinders economic growth and therefore, requires special attention. However, the empirical evidence of energy-poverty link is often presented in simple graphs showing that energy consumption increases with income, or the human development index (HDI) improves with income and higher energy use. Pachauri et al. (2004) indicate that there are three types of measures normally found in the literature to indicate existence of energy poverty: 1) economic measures such as energy poverty line, 2) engineering measures of minimum energy needs, and 3) measures based on access to energy services.

The economic approach tries to find out how much consumers lying below the national poverty line spend on energy and how this expenditure compares with the overall household expenditure. If for example, a consumer spends more than 10% of her expenditure on energy, the consumer may be regarded as lying below the fuel poverty line. Such a definition is used in the U.K. However, the expenditure depends on the fuel mix, level of efficiency of the appliances, size of the household and prices in the market. Therefore, while a large budget share could indicate fuel poverty, it may give wrong signals as well.

The engineering approach uses an estimation of the energy needs to satisfy the basic requirements of any household. These are normative levels often used by government authorities to plan for energy needs of a community or a country. They are based on some assumptions about the types of activities generally performed by households and the energy requirement using available technologies. Clearly, such a norm will vary from one country to another and can vary over time. However, an understanding of the basic needs can help analyse various implications of non-availability of such supplies to the target groups.

Finally, the approach based on access to services departs from the above two in the sense that it tries to find out whether consumers have physical access to the

Table 22.6 Factor goalposts for EDI in 2002

Factor	Maximum	Minimum
Per capita commercial energy consumption	9.4 toe (Bahrain)	0.01toe (Togo)
Share of commercial energy in total final energy	100%	8% (Ethiopia)
Electrification rate	100%	2.6% (Ethiopia)

Source: WEO (2004)

Table 22.7 Example of EDI for India

Factor	Formula	Indicator
Per capita commercial energy consumption	(0.33−0.01)/(9.4−0.10)	0.034
Share of commercial energy in total final energy	(56−8)/(100−8)	0.519
Electrification rate	(46−2.6)/(100−2.6)	0.445
Average index	(0.034+0.519+0.445)/3	0.332

Source: WEO (2004)

supply of energy, and access to markets for equipment. Generally, the poor will have limited choice in terms of access to fuels and equipment choices compared to the well-off consumers. However, this is more data intensive and it may be difficult to compare two situations quantitatively using this approach.

WEO (2004) has presented an index, Energy Development Index (EDI) along the line of HDI. EDI is composed of the following three factors:

- per capita commercial energy consumption,
- share of commercial energy in total final energy use,
- share of population with access to electricity.

An index is created for each factor by considering the maximum value and minimum values observed in the developing world and determining how a particular country has performed. The following formula is used for this index

$$\text{Factor index} = \frac{(\text{Actual value} - \text{minimum value})}{(\text{maximum value} - \text{minimum value})} \tag{22.1}$$

The goalposts (maximum and minimum values) are taken from the observed values within the sample of developing countries considered. For example, for calculating the factor goalposts for EDI in 2002, WEO (2004) used the values shown in Table 22.6.

The simple average of three indicators gives the overall EDI.

For any country, e.g. India, EDI can be calculated using Eq. 22.1 and noting the goalposts as well as actual data for the country. For 2002, India's per capital commercial energy consumption was 0.33 toe, share of commercial energy in the final energy was 55.75% and the rate of electrification was 46%. The individual indicators are shown in Table 22.7.

Although this indicator provides a numerical value, it is not devoid of problems. It perpetuates the idea that higher level of energy consumption is synonymous to

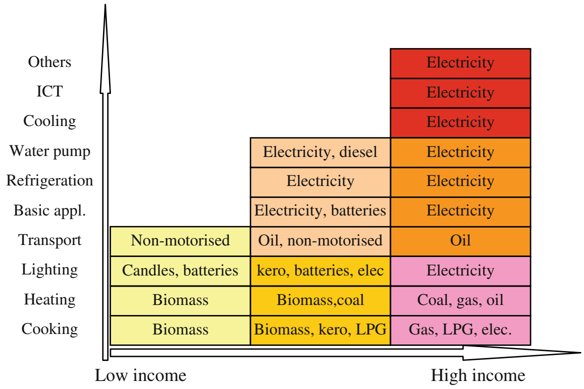


Fig. 22.1 Energy ladder example. *Source* WEO (2002)

economic development. Accordingly, countries in the Middle East with high per capita energy use rank better in this index. It also assumes that biomass energy use represents a symbol of under-development, which need not be the case, depending on how it is used. Finally, it also assumes that the grid-based electrification is essential for development, which is not true either.

22.3 Energy Ladder and Energy Use

It is normally noticed that the energy mix varies significantly among the poor and the rich. Normally, people in the lower income group tend to use more traditional energies to meet their needs. But with higher income people tend to move up the energy ladder and tend to use more commercial energies and less traditional energies. The general idea is presented in Fig. 22.1. As energy is a derived demand, the ability to use any modern fuel is dependent on the affordability of energy-using appliance and the ability to pay for the fuel on a regular basis. This can be an issue with the poor and hence they tend to rely on cheap technology and fuels.

The issue is not restricted to rural areas alone—often the poor in the urban setting are also using traditional energies, but there are urban–rural differences in energy consumption patterns.

An example from India is shown in Fig. 22.2, which shows that 76% of the rural households in 1999–2000 relied on firewood and chips, while only 22% of the urban households used this fuel. Urban households relied more on commercial fuels (LPG, kerosene) and the situation changed quite significantly between 1993–94 and 1999–2000. Indian Census 2001 reported that more than 139 million households in India (72% of all households)[2] rely on traditional energies for their

[2] According to Census 2001, there were around 192 million households in India, of which around 72% reside in rural areas and the rest in urban areas. The average household size was 5.3 persons in 2001.

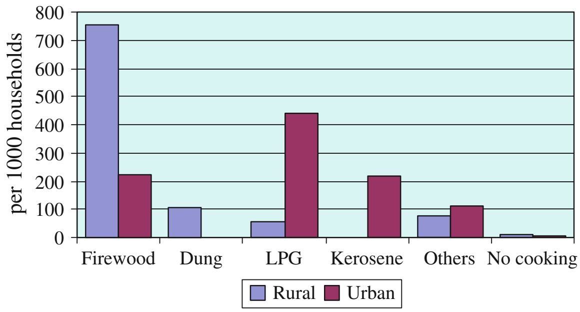


Fig. 22.2 Urban-rural differences in cooking energy use in India. *Source* NSSO (2001a)

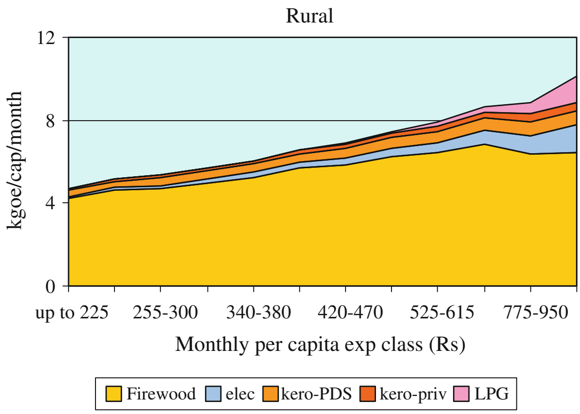


Fig. 22.3 Energy consumption pattern in rural India by expenditure class. *Source* NSSO (2001b)

cooking needs. Out of this, more than 124 million households reside in rural areas, while the remaining 15 million live in urban areas.

Energy consumption pattern by different income groups is more difficult to obtain. Sample data from NSSO (2001b) was used to generate energy consumption[3] by different expenditure classes separately for rural and urban areas (see Figs. 22.3 and 22.4). Figure 22.3 suggests that firewood is the main cooking energy in rural India irrespective of income level, although its share falls from around 90% for the lowest expenditure class to around 64% in the highest expenditure class. Yet, as the higher expenditure classes consume more cooking energy per capita, firewood consumption in absolute terms is more for the higher expenditure classes and the highest expenditure class consumes almost 50% more wood fuel compared to the lowest class. This clearly indicates that the issue of

[3] The data for Figs. 22.3 and 22.4 covers all household energy consumption and does not differentiate between cooking and lighting. However, it is reasonable to assume that electricity is mainly used for lighting while firewood and LPG are used for cooking. Kerosene may be used for both lighting and cooking. NSSO (2001b) provides data in physical units (kg or litres). The following conversion factors were used to arrive at ton of oil equivalent figures: firewood—0.32 kgoe/kg, electricity—0.086 kgoe/kWh, kerosene—0.836 kgoe/l, LPG—1.13 kgoe/kg and coal—0.441 kgoe/kg.

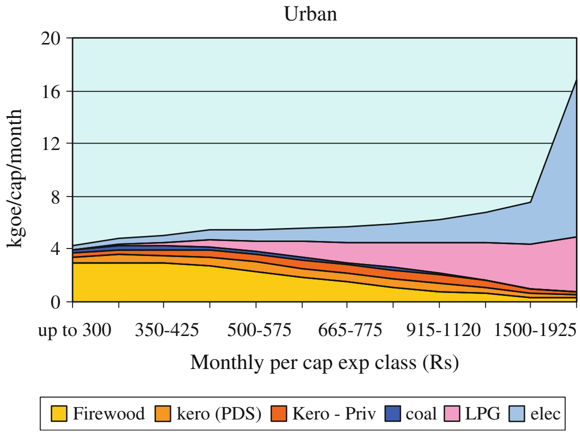


Fig. 22.4 Energy consumption pattern in urban India by income class. *Source* NSSO (2001b)

access to clean cooking energy in rural areas has a much bigger dimension and is not limited to the poor households alone.

The picture changes significantly in urban areas. The use of firewood diminishes quite appreciably with higher expenditure class while the use of cleaner fuels such as LPG or electricity increases. Even at the lowest expenditure class firewood plays a significantly lower role compared to the rural areas (around 70% share compared to 90% in rural areas). High levels of electricity and LPG use by higher expenditure classes suggest that they are unlikely to have affordability problems. Therefore, the problem of access to clean energies in urban areas is a problem faced by the poor households to a large extent. Figures 22.3 and 22.4 also suggest that there is not much difference in the per capita energy consumption in the lower expenditure classes between urban and rural areas. But the highest expenditure class in urban area has a much higher per capita consumption compared to the rest of the households in the country.

This brings us to the drivers influencing the choice decision. This is discussed below.

22.4 Diagnostic Analysis of Energy Demand by the Poor

Energy demand in poor households normally arises from two major end-uses: lighting and cooking (including preparation of hot water).[4] Cooking energy demand is predominant in most cases and often accounts for about 90% of the energy demand by the poor. Such a high share of cooking energy demand arises

[4] In some climatic conditions space heating may also be an important source of energy demand. However, for this discussion space heating demand is not considered.

partly from the low energy efficiency and partly due to limited scope of other end-uses.

Any energy use involves costs and resource allocation problems. Both traditional energies (TE)[5] which play a crucial role in the energy profile of the poor, and modern energies impose private and social costs. The private cost may be in monetary terms or in terms of time spent by the family members to collect the TEs. For collected TEs, the problem of valuation of the cost arises and the collected fuel is considered as free fuel by many, even perhaps by the poor themselves, as no monetary transactions are involved. However, depending on the quantity of collected fuel, its source and the type of labour used in the collection process, the private cost and social cost can be substantial. The social cost arises due to externalities arising from pollution and other socio-economic problems related to particular forms of energy use.

The entire decision-making process for use of any modern energy form (electricity, kerosene or LPG, or renewable energies) as opposed to any other form of traditional energies revolves around monetary transactions. Any commercial energy requires monetary exchanges and the decision to switch to commercial energies can be considered as a three-stage decision-making process. First, the household has to decide whether to switch or not (i.e. switching decision). Second, it decides about the types of appliances to be used (i.e. appliance selection decision). In the third stage, consumption decision is made by deciding the usage pattern of each appliance (i.e. consumption decision).

While the costs do not always lend themselves to monetary-based accounting, the switching decision is largely determined by monetary factors: the amount and regularity of money income, alternative uses of money and willingness to spend part of the income to consume commercial energies as opposed to allocating the money to other competing needs. Appliance selection is affected by similar factors: cost of appliance, the monetary income variables described above and the availability of financing for appliance purchases through formal and informal credit markets. Finally, the consumption decision depends on, among others, family size, activities of the family members, availability of appliances and family income.

This framework of three-stage decision-making (presented in Chap. 3) helps in analysing the problem in a logical manner. The poor normally lack regular money income flows due to unemployment or part-employment, both of which sometimes produce in-kind payments as compensation. Moreover, they often participate in informal sector activities, where barter rather than monetised transactions prevail. It is rational for any household or individual to focus on private monetary costs rather than social and/or non-monetised costs due to the inherent subjectivity and complexity of the valuation problem. Moreover, any modern energy has to compete with other goods and services (including saving for the future) procured by

---

[5] I have preferred to use the term traditional energies to non-commercial energies to avoid any confusion arising out of monetisation or commercialisation of some of such fuels. .

the household for an allocation of monetary resources. Given above characteristics and constraints, it is quite logical for the poor to have a natural preference for the fuel that involves no or minimum money transactions. Reliance on firewood and other traditional energies used for cooking, which constitute the major source of energy demand by the poor, can be explained using this logic.

For any commercial energy to successfully penetrate the energy demand of the poor would then require satisfaction of the following economic factors:

(a) The energy should be suitable and perhaps versatile for satisfying the needs;
(b) It should have a competitive advantage that would place no or little demand for money transactions (in other words, the low cost supplies) in the present circumstances, and/or
(c) the use of modern energy should result in supply of adequate money flows to the poor so that they become willing to spend some part of the money on purchasing commercial energies.

Other supply- and demand-related issues and social factors (such as availability of fuel, social acceptance, ease of use, pollution, etc.) will also affect fuel choice and its use, but they are secondary to economic factors.

The second stage (i.e. appliance selection decision) has a deciding influence on energy demand. Often energy appliance has a relatively long life (5–10 years) and its initial costs are high relative to the income level of the poor. In order to, in a sense, amortise the costs the appliances will likely have to be used for sometime, thereby introducing strong path dependence in energy demand. Strong path dependence affects fuel switching possibility and responsiveness of the consumers to external changes. Fuel switching option will be limited by the appliance choice decision and will involve potentially sizeable capital expenditure. The rigidity or strong path dependence leaves limited options to consumers in the event of sudden changes in prices or supply conditions in the short run, who have to depend on their existing stock of appliances in any case.

The appliance selection decision has important bearings for the poor as well. First, high initial cost of appliances for using modern energy is a major deterrent. Consumers naturally prefer low cost appliances, although they are often energy inefficient. This also results from the difficulty of mental calculations for an economic appliance selection that involve factors such as operating costs, discount rates and appliance life. Second, appliances which the poor consider as essential and affordable will be selected, thereby restricting the choice to a bare minimum. Third, the poor are inherently adverse to experimentation and are unlikely to commit themselves to uncertain and unproven technologies on their own. Fourth, strong path dependence of modern energies is likely to add to the reluctance of the poor to invest in modern energies.

Once a decision is made to switch to a modern fuel and the appliance is purchased, the only variable left in the hand of the user is its utilisation. The short term response of consumers to demand arises from this factor, which is quite limited.

22.5 Evaluation of Existing Mechanisms for Enhancing Access

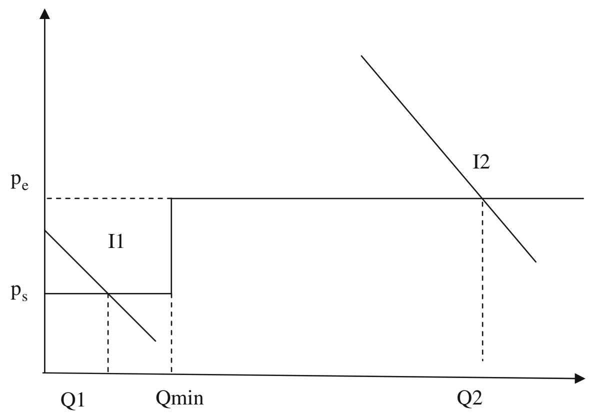
Although a wide range of options are adopted to enhance access to energy, the existing policies rely on the state to provide access by subsidising supply to consumers. A number of energies come under this purview: kerosene for lighting and cooking purposes, LPG for cooking purposes and electricity. Subsidies for such energies could be supported from social considerations: as some minimum amount of energy is required for sustaining livelihood, those who are unable to procure such energies could be supported to procure them. This is essentially the argument behind using lifeline rates for electricity. This is explained in Fig. 22.5 below. If the price is p_e, then consumers with low income will not be able to enter the market as they cannot afford the service. If the consumer surplus of low income consumers multiplied by an appropriate social weight is greater than the social cost of supply, adoption of a lifeline rate could be justified. This does not affect the overall efficiency of the pricing scheme as those having demand above the minimum level of demand Qmin would face the rate at p_e.

The externality argument could also be used to support subsidies: as the use of traditional energies imposes considerable health effects on the population, by switching to clean energies the social cost of health damage could be reduced. As long as the benefits of fuel switching are greater than the social cost, such a subsidy scheme could be followed.

But subsidised energy supply in developing countries has come under scrutiny and the following criticism can be identified:

(a) the subsidy is not targeted, implying that the benefits do not reach the desired group. In many cases all consumers have been given the benefits of subsidy for administrative simplicity, which allows the rich to benefit more as their per capita consumption is higher than the poor. Where the benefit is restricted to

Fig. 22.5 Lifeline rates



the poor, lack of administrative verification and monitoring allows considerable leakage of subsidy, allowing others to benefit.

(b) As energy cannot be used without owning appliances and as subsidies are granted for energy consumption (not for appliance ownership), subsidised supply helps those who can afford appliances. Thus subsidies for LPG and electricity often accrue to the rich.

(c) Continued use of subsidised supply has given rise to a sense of right to this privilege, making subsidy removal politically difficult.

(d) Subsidised supply distorts price signals and increases demand, which in turn requires more investment for supply systems. This can be seen from Fig. 22.6 (also see Chap. 13). As most of the residential consumers contribute to the peak demand, higher consumption requires extra peaking capacity, which is costly but at the same time may not be remunerative for the supplier. Capacity shortage results in absence of new capacities, imposing social costs due to non-availability of supply.

(e) Subsidy imposes revenue burden on the supplier and the state, and when the subsidy is not timely provided, the financial performance of the supplier gets affected.

(f) Inefficient energy use through subsidies adds to pollution and contributes to the climate change problem.

Getting energy prices right essentially means rebalancing the prices by removing subsidies and cross-subsidies. There are two issues involved here: correct prices would make energy supply a commercially attractive proposition but at the same time, commercial energies will become less the competitive as compared to the traditional energies. However, as observed earlier, the subsidy system for petroleum products is not targeted to the poor and such improperly targeted

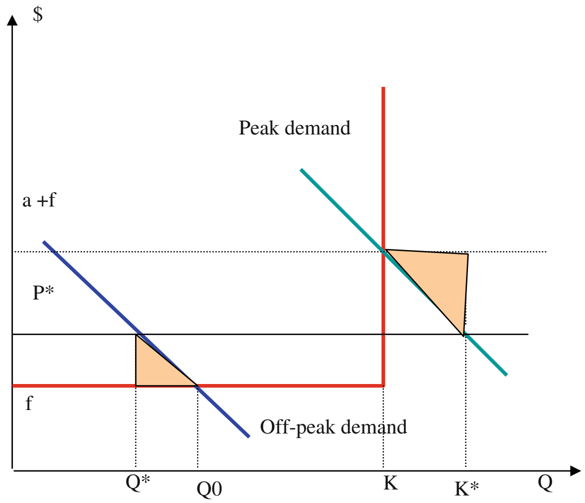


Fig. 22.6 Inefficiency of inappropriate pricing

subsidies could be removed without much effect on the poor. But strategic subsidies would remain a key policy tool of the state to promote commercial energies amongst the poor.

22.6 Effectiveness of Electrification Programmes for Providing Access

To resolve the energy access problem, rural electrification initiatives need to be analysed considering the factors presented in the diagnostic analysis. A number of observations/inferences can be made:

- Electricity is mainly used for lighting purposes and accounts for a minor share of households' energy needs. In order to resolve the energy access problem through electrification, electricity use has to meet the cooking energy requirements of the poor. A number of issues arise in this respect:
  - *Competitiveness*: electricity is unlikely to be competitive when compared with traditional energies used for cooking purposes. Subsidized supply to household belonging to lower income groups normally will allow them to use electricity for lighting. Promotion of electricity supply is unlikely to reduce reliance on traditional energies for cooking per se.
  - *Quality of supply*: As the power supply to rural areas gets low priority, even when access is available, actual supply may be limited, especially during peak demand periods due to prevailing capacity shortage conditions. Lack of adequate supply acts as a hindrance to expansion of electricity use in productive and other activities.
  - *Initial investment*: Use of electricity for cooking entails significant initial investment when compared with traditional energy use. Cash-strapped poor households are unlikely to switch to electric cooking even if quality electricity is available at an affordable rate.

Thus, electricity has a less chance of succeeding in the cost competition with other fuels. This in turn implies that demand for lighting cannot justify the investment in electrification of an area. Consequently, rural electrification alone cannot resolve the problem of energy access in rural areas, as other fuels would be used by the poor to meet cooking demand. It appears that policy makers tend to ignore or forget this simple truth, may be because of better prestige and visibility of electrification projects (and hence for better political mileage).

For economic and financial viability of rural electrification projects, expansion of productive use of electricity is essential. Integrating other rural development programmes with rural electrification could create a synergy for promoting agro-based industrial activities and productive use of electricity in rural areas.

Additionally, countries with a poor record of credible subsidy management system due to resource constraints, sustainability of subsidized schemes is highly doubtful. A credible alternative has to rely on development mechanisms that ensure adequate money supply to the poor on a regular basis. This makes it necessary for rural energy supply issues to be set in a broader canvass of overall development. But experience so far hardly supports the catalytic role of electrification programmes. Rural industry or commerce has not developed as a thriving business proposition so far in many rural areas. Thus, sustainability of subsidized rural electrification system may remain a thorny issue for a long time to come.

Energy sector reform has not been a great success in countries where most of the poor are concentrated and is progressing quite slowly. Electricity reform has not produced the desired results so far and even the progress has been dismal in most areas. Simultaneously, the state funding for electricity has been drastically reduced, without any concomitant participation from the private sector. Private participation in power distribution does not seem to be gaining momentum and it is quite likely that the privately-owned distribution companies will be least interested in undertaking a loss-making activity. Depending on reforms for solving the energy access problem will be synonymous to inaction. This is not suggest that reforms are not required or should not be followed. Energy sector reforms are essential but being a politically sensitive process, making it a pre-condition for providing access to the poor is not a logical approach.

22.7 Renewable Energies and the Poor

Many place great hopes on new technologies for solving the problem (WEC 2000; DfID 2002; World Bank 1996). New technologies that are suitable for distributed energy supplies are now available and can be cost-effective compared to grid-based supplies. Such technologies often have the added advantage of being environment friendly and hence their promotion would be beneficial for the world as a whole. However, despite extensive research and commercialisation efforts over past three decades, these energies are not competitive yet, without subsidies of some sort or other. Using subsidies for creating a market for new technologies has the disadvantage that subsidy removal becomes difficult, as the LPG case demonstrates. The technical fix of the problem does not appear to be an answer.

Consider now the case of renewable energies to analyse whether they meet the above requirements indicated earlier in the diagnostic analysis. As cooking and lighting constitute two major energy demands of poor households (excluding space heating), we consider these two separately. As there are different types of renewable energies (solar, wind, hydro and even sustainable biomass), we focus on solar energy here. Similar arguments can perhaps be advanced for other energies as well.[6]

---

[6] The specific arguments may have to be adjusted in some cases but the generic argument remains valid.

Environmental and Natural Resource Economics

11th Edition

Tom Tietenberg and Lynne Lewis

COMPANION WEBSITE

ROUTLEDGE

Chapter 7

Energy

The Transition from Depletable to Renewable Resources

If it ain't broke, don't fix it!

—Old Maine proverb

Introduction

Energy is one of our most critical resources; without it life would cease. We derive energy from the food we eat. Through photosynthesis, the plants we consume—both directly and indirectly when we eat meat—depend on energy from the sun. The materials we use to build our houses and produce the goods we consume are extracted from the earth's crust, and then transformed into finished products with expenditures of energy.

Currently, many industrialized countries depend on oil and natural gas for the majority of their energy needs. According to the International Energy Agency (IEA), these resources together supply 59 percent of all primary energy consumed worldwide. (Adding coal, another fossil fuel resource, increases the share to 86 percent of the total.) Fossil fuels are depletable, nonrecyclable sources of energy.

According to depletable resource models, oil and natural gas would be used until the marginal cost of further use exceeded the marginal cost of more abundant and/or renewable substitute resources. In an efficient market path, the transition to these alternative sources would be smooth and harmonious. Have the allocations of the last several decades been efficient or not? Is the market mechanism flawed in its allocation of depletable resources? If so, is it a fatal flaw? If not, what caused the inefficient allocations? Are the problems correctable?

In this chapter we shall examine some of the major issues associated with the allocation of energy resources over time and explore how economic analysis can clarify our understanding of both the sources of the problems and their solutions.

Natural Gas: From Price Controls to Fracking

Some Early History

In the United States, during the winter of 1974 and early 1975, serious shortages of natural gas developed. Customers who had contracted for and were willing to pay for natural gas were unable to get as much as they wanted. The shortage (or curtailments, as the Federal Energy Regulatory Commission [FERC] calls them) amounted to 2 trillion cubic feet of natural gas in 1974–1975, which represented roughly 10 percent of the marketed production in 1975. In an efficient allocation, shortages of that magnitude would never have materialized. What happened?

The simple answer is regulation. The regulation of natural gas began in 1938 with the passage of the Natural Gas Act. This act transformed the Federal Power Commission (FPC), which subsequently become FERC, into a federal regulatory agency charged with maintaining "just" prices. In 1954 a Supreme Court decision forced the FPC to extend their price control regulations beyond pipeline companies to include producers as well.

Because the process of setting price ceilings proved cumbersome, the hastily conceived initial "interim" ceilings remained in effect for almost a decade before the Commission was able to impose more carefully considered ceilings. What was the effect of this regulation?

By returning to our models in the previous chapter, we can anticipate the havoc this would raise. The ceiling would prevent prices from reaching their normal levels. Since price increases are the source of the incentive to conserve, the lower future prices would cause an inefficiently large amount of the resource to be used in earlier years. Consumption levels in the earlier years would be higher under price controls than without them.

Effects on the supply side would also be expected to be significant. Producers would produce the resource only when they could do so profitably. Once the marginal cost rose to meet the price ceiling, no more would be produced, regardless of the demand for the resource at that price. Thus, as long as price controls were permanent, less of the resource would be produced with controls than without so production would also be skewed toward the earlier years.

The combined impact of these demand-and-supply effects would be to distort the allocation significantly (see Figures 7.1a and 7.1b). While a number of aspects differentiate this allocation from an efficient one, several are of particular importance: the market would react to price controls by (1) leaving more of the resource in the ground, (2) increasing the rate of consumption, (3) causing the time of transition to be earlier, and (4) creating an abrupt transition, with prices suddenly jumping to new, higher levels. All are detrimental. The first effect means we would not be using all of the natural gas available at prices consumers were willing to pay. Because price controls would cause prices to be lower than efficient, the resource would be depleted too fast. These two effects would cause an earlier and abrupt transition to the substitute, possibly before the technologies to use it were adequately developed.

The discontinuous jump to a new technology, which results from the fact that price controls eliminate price flexibility, can place quite a burden on consumers. Attracted by artificially low prices, consumers would invest in equipment to use natural gas, only to discover—after the transition—that natural gas was no longer available.

One interesting characteristic of price ceilings is that they affect behavior even before they are binding.[1] This effect is clearly illustrated in Figures 7.1a and 7.1b in the earlier years. Even though the price in the first year is lower than the price ceiling, it is not equal to the efficient price. (Can you see why? Think what effect price controls have on the marginal user cost faced

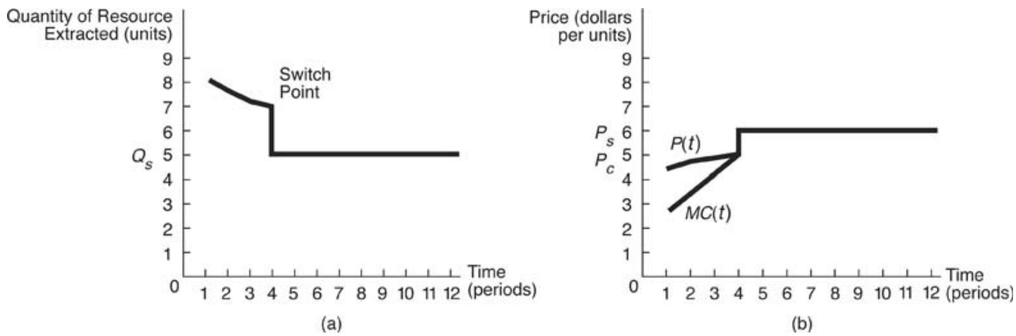


Figure 7.1 (a) Increasing Marginal Extraction Cost with Substitute Resource in the Presence of Price Controls: Quantity Profile (b) Increasing Marginal Extraction Cost with Substitute Resource in the Presence of Price Controls: Price Profile

by producers.) The price ceiling causes a reallocation of resources toward the present, which, in turn, reduces prices in the earlier years.

It seems fair to conclude that, by sapping the economic system of its ability to respond to changing conditions, price controls on natural gas created a significant amount of turmoil. When this kind of political control occurs, the overshoot and collapse scenario can have some validity. In this case, however, it would be caused by government actions rather than any pure market behavior. If so, the adage that opens this chapter becomes particularly relevant!

Politicians may view scarcity rent as a possible source of revenue to transfer from producers to consumers. As we have seen, however, scarcity rent is an opportunity cost that serves a distinct purpose—the protection of future consumers. When a government attempts to reduce scarcity rent through price controls, the result is an overallocation to current consumers and an underallocation to future consumers. Thus, what appears to be a financial transfer from producers to consumers is, in large part, also a transfer of the affected commodity from future consumers to present consumers. Since current consumers mean current votes and future consumers may not know whom to blame by the time shortages appear, price controls are politically attractive. Unfortunately, they are also inefficient; the losses to future consumers and producers are greater than the gains to current consumers. Because controls distort the allocation toward the present, they are also unfair to future consumers. Thus, markets in the presence of price controls are indeed myopic, but the problem lies with the controls, not the market per se.

After long debating the price control issue, Congress passed the Natural Gas Policy Act on November 9, 1978. This act initiated the eventual phased decontrol of natural gas prices. By January 1993, no sources of natural gas were subject to price controls.

Fracking

Natural gas production remained relatively stable from the mid-1970s until the middle of the first decade of the twenty-first century, when a new technology dramatically changed the cost of accessing new sources of natural gas in shale, a type of sedimentary rock.[2] Hydraulic fracturing, or fracking as it is known popularly, is a form of technical progress that combines horizontal drilling with an ability to fracture deep shale deposits using a mixture of high pressure water, sand, and chemicals. Not only does the fractured shale release large quantities of natural gas, but this extraction process also costs less than accessing more conventional sources.

The introduction of this new technology has increased production dramatically in the United States and fracked gas is likely to play an even larger role over the next few decades according to the Energy Information Agency, the statistical arm of the U.S. Department of Energy. If ever there were an example of the profound effect a technical change can have, this is it!

While this production is dramatically changing the energy situation in the United States, that change comes with some controversy (see Debate 7.1).

Oil: The Cartel Problem

Much of the world's oil is currently produced by a cartel called the Organization of the Petroleum Exporting Countries (OPEC). The members of this organization collude to exercise power over oil production and prices. As established in Chapter 2, seller power over resources due to a lack of effective competition leads to an inefficient allocation. When sellers have market power, they can restrict supply and thus force prices higher than otherwise.

Though these conclusions were previously derived for products, they are valid for depletable resources as well. By restricting supply a monopolist can extract more scarcity rent from a depletable resource base than competitive suppliers can. The monopolistic transition results in a slower rate of production and higher prices.[3] The monopolistic transition to a substitute, therefore, occurs later than a competitive transition. While monopolistic exploitation raises the net present value of profits to the sellers, it reduces the net present value of net benefits to society. Although the slower consumption on balance reduces the social costs associated with climate damages, as we shall see in Chapter 17 this outcome is not efficient since many lower-cost ways to reduce these social costs are available.

The cartelization of the oil suppliers has, historically apparently been quite effective (Smith, 2005). Why? Are the conditions that make it profitable unique to oil, or could oil cartelization be the harbinger of a wave of natural resource cartels? What is the outlook for the oil cartel in the future? To answer these questions, we must isolate those factors that make cartelization possible and profitable. Although many factors are involved, four stand out: (1) the price elasticity of demand in both the long run and the short run; (2) the income elasticity of demand; (3) the supply responsiveness of the producers who are not cartel members; and (4) the compatibility of interests among members of the cartel.

Price Elasticity of Demand

The price elasticity of demand is an important ingredient because it determines how responsive demand is to price. When demand elasticities are between 0 and –1 (i.e., when the percentage quantity response is smaller than the percentage price response), price increases lead to increased revenue. Exactly how much revenue would increase when prices increase depends on the magnitude of the elasticity. Generally, the smaller is the absolute value of the price elasticity of demand (the closer it is to 0.0), the larger are the gains to be derived from forming a cartel.

The price elasticity of demand depends on the opportunities for conservation, as well as on the availability of substitutes. As storm windows cut heat losses, the same temperature can

DEBATE 7.1

Does the Advent of Fracking Increase Net Benefits?

While fracking will no doubt lower U.S. dependency on energy imports (the subject of the next section) and provide an economic boost by lowering energy costs in the United States as it displaces more expensive fuels, it also comes with some costs. The main short-term concerns involve water contamination (fracking chemicals leaking into local wells), water depletion (the extraction process uses large quantities of water), air quality issues (some of the toxic fracking chemicals can escape into the surrounding air) and "leakage" (methane, one of the primary components of natural gas and a powerful greenhouse gas that contributes to climate change, can leak into the atmosphere as a result of the fracking process). Further, over the longer run, according to the International Energy Agency, an excessive reliance on natural gas would be incompatible with reaching proposed climate policy goals.

If fracking comes with high benefits *and* high costs, does it make economic sense? The simple answer is that we don't know yet. For one thing, as Chapter 3 reminds us, it would depend on the accounting stance (geographic scope) of the analysis. The geographic regions that benefit may not be the same regions that bear the costs. Different accounting stances could produce different results.

More fundamentally, even if a national benefit-cost analysis could be revealing, many of the components of that analysis are not yet known with sufficient certainty to provide much confidence in the answers this early in the game. To take just one example of our ignorance, if the leakage rate exceeds 3.2 percent, natural gas is apparently no better for the climate than coal or oil. Unfortunately, we have a firm grasp on the leakage rate for only a few specially studied wells. Furthermore, the expected costs from the associated water and air contamination are not yet fully known either. Finally, even if we were able to derive a reasonable answer for the current period prior to much regulation, the answer is likely to be much more favorable to fracking once a regulatory framework to reduce the problems is in place.

Fortunately, studies are underway to fill in the information gaps and regulations that control the most negative net benefit aspects of the industry are likely to follow. Stay tuned.

Sources: Alvarez, R. A., Pacala, S. W., Winebrake, J. J., Chameides, W. L., & Hamburg, S. P. (2012). Greater focus needed on methane leakage from natural gas infrastructure. *PNAS*, *109*, 176435–176440; Jackson, R. B., et al. 2013. The environmental costs and benefits of fracking. *Annual Review of Environment and Resources*, *39*, 327–362.

be maintained with less heating oil. Smaller, more fuel-efficient automobiles reduce the amount of gasoline needed to travel a given distance. The larger the set of these opportunities and the smaller the cash outlays required to exploit them, the more price-elastic the demand. This suggests that demand will be more price-elastic in the long run (when sufficient time has passed to allow adjustments) than in the short run.

The availability of substitutes is also important because it limits the degree to which prices can be profitably raised by a producer cartel. Abundant quantities of relatively inexpensive fuels that could substitute for oil can set an upper limit on the cartel price. Unless the cartel controls those alternative sources as well—and in oil's case it doesn't—any attempts to raise prices above those limits would cause the consuming nations to simply switch to these alternative sources; the cartel would have priced itself out of the market.

Income Elasticity of Demand

The income elasticity of demand is important because it indicates how sensitive demand for the cartel's product is to growth in the world economy. As income grows, demand should grow. This continual increase in demand fortifies the ability of the cartel to raise its prices.

The income elasticity of demand is also important, however, because it registers how sensitive demand is to the business cycle. The higher the income elasticity of demand, the more sensitive demand is to periods of rapid economic growth or to recessions. Economic downturns led to a weakening of the oil cartel in 1983 as well as to a significant fall in oil prices starting in late 2008. Conversely, whenever the global economy recovers, the cartel benefits disproportionately.

Nonmember Suppliers

Another key factor in the ability of producer nations to exercise power over a natural resource market is their ability to prevent new suppliers, not part of the cartel, from entering the market and undercutting the price. Prior to fracking, OPEC produced about 45 percent of the world's oil, but that is changing due to the increase in production from fracked oil. When non-OPEC producers expand their supply dramatically, prices can be expected to fall along with OPEC's market share. If this supply response is large enough, the cartelized allocation of oil would approach the competitive allocation.

Recognizing this impact, the cartel must take the nonmembers into account when setting prices. Salant (1976) proposed an interesting model of monopoly pricing in the presence of a fringe of small nonmember producers that serves as a basis for exploring this issue. His model includes a number of suppliers. Some form a cartel. Others, a smaller number, form a "competitive fringe." The cartel is assumed to set its price so as to maximize its members' profit, taking the competitive fringe production into account. The competitive fringe cannot directly set the price, but, since it is free to choose the level of production that maximizes its own profits, its output does affect the cartel's pricing strategy by increasing the available supply.

What conclusions does this model yield? The model concludes first of all that a resource cartel would set different prices in the presence of a competitive fringe than in its absence. With a competitive fringe, it would set the initial price somewhat lower than the pure monopoly price and allow price to rise more rapidly. This strategy maximizes cartel profits by inducing the competitive fringe to produce more in the earlier periods (in response to higher demand) and eventually to exhaust their supplies. Once the competitive fringe has depleted its reserves, the cartel would raise the price and thereafter prices would increase much more slowly.

Thus, the optimal strategy, from the point of view of the cartel, is to hold back on its own sales during the initial period, letting the other suppliers exhaust their supplies. Sales and profits of the competitive fringe, in this optimal cartel strategy, decline over time, while sales and profits of the cartel increase over time as prices rise and the cartel continues to capture a larger share of the market.

Another fascinating implication of this model is that the formation of the cartel raises the present value of competitive fringe profits by an even greater percentage than it raises the present value of cartel profits. Those without the power gain more in percentage terms than those with the power!

Though this may seem counterintuitive, it is actually easily explained. The cartel, in order to keep the price up, must cut back on its own production level. The competitive fringe, however, is under no such constraint and is free to take advantage of the higher prices caused by the cartel's withheld production without cutting back its own production. Thus, the profits of the competitive fringe are higher in the earlier period, which, in present value terms, are discounted less. All the cartel can do is wait until the competitive suppliers become less of a force in the market. The implication of this model is that the presence of a competitive fringe matters, even if it controls as little as one-third of the production.

The impact of this competitive fringe on OPEC behavior was dramatically illustrated by events in the 1985–1986 period. In 1979, OPEC accounted for approximately 50 percent of world oil production, while in 1986 this had fallen to approximately 30 percent. When the recession cut global demand by 10 percent, the cartel's attempts to keep prices as high as possible by reducing its own production were thwarted by a competitive fringe that simply kept producing. The real cost of crude oil imports in the United States fell from $34.95 per barrel in 1981 to $11.41 in 1986. OPEC simply was not able to hold the line on prices because the necessary reductions in its own production were too large for the cartel members to sustain in the face of continuing supplies from the competitive fringe.

With global economic growth, however, the tide was turned. In the summer of 2008, the price of crude oil soared above $138 per barrel. Strong worldwide demand was coupled with restricted supply from Iraq because of the war. However, these high prices also promoted the major oil companies' search for more unconventional sources of oil including the tar sands in Canada and the use of fracking to extract oil from shale in the United States. Once again the cartel's power was subject to limits, this time imposed by nonmember suppliers.

Compatibility of Member Interests

The final factor we shall consider in determining the potential for cartelization of natural resource markets is the internal cohesion of the cartel. With only one seller, the objective of that seller can be pursued without worrying about alienating others who could undermine the profitability of the enterprise. In a cartel composed of many sellers, that freedom is no longer as wide ranging. The incentives of each member and the incentives of the group as a whole may diverge.

Cartel members have a strong incentive to cheat. A cheater, if undeterred by the other members, could lower its price and capture a larger share of the market. Formally speaking, the price elasticity of demand facing an individual member is substantially higher than that for the group as a whole, because some of the increase in individual sales at a lower individual price represents sales reductions for other members. When producers face markets characterized by high price elasticities, lower prices maximize profits. Thus, successful cartelization presupposes a means for detecting cheating and enforcing the collusive agreement.

In addition to cheating, however, cartel stability is also threatened by the degree to which members fail to agree on pricing and output decisions. Oil provides an excellent example of how these dissensions can arise. Since the 1974 rise of OPEC as a world power, Saudi Arabia has frequently exercised a moderating influence on the pricing decisions of OPEC. Why?

One highly significant reason is the size of Saudi Arabia's oil reserves. Controlling about 22 percent of OPEC's proven oil reserves in 2015 (second only to Venezuela's 24.8 percent) Saudi Arabia has an incentive to preserve the value of those resources. Setting prices too high would undercut the future demand for its oil. As previously stated, the demand for oil in the long run is more price-elastic than in the short run. Countries with smaller reserves, such as Nigeria, know that in the long run their reserves will be gone and therefore these countries are more concerned about the near future. Countries with small reserves want to extract more rent now, but countries with large reserves want to preserve future rent.

This examination of the preconditions for successful cartelization suggests two things: (1) creating a successful cartel is not an easy path for natural resource producers to pursue, and (2) it is quite likely that OPEC's difficulties in exercising control over the market will only increase in the future.

Fossil Fuels: National Security Considerations

Vulnerable strategic imports such as oil have an added cost that is not reflected in the marketplace. National security is a classic public good. No individual corporate importer correctly represents our collective national security interests in making a decision on how much to import. Hence, leaving the determination of the appropriate balance between imports and domestic production to the market generally results in an excessive dependence on imports in terms of both climate change and national security considerations (see Figure 7.2).

In order to understand the interaction of these factors, five supply curves are relevant. Domestic supply is reflected by two options. The first, S_{d1}, is the long-run domestic supply curve without considering the climate change damages resulting from burning more oil, while the second, S_{d2}, is the domestic supply curve that includes these per-unit damages.

While climate change policy is the subject of Chapter 17, we can introduce its relationship to energy choice here. All fossil fuels contain carbon. When these fuels are combusted, unless the resulting carbon is captured, it is released into the atmosphere as carbon dioxide, a contributor to climate change. How much CO_2 is released varies with the type of fuel since energy sources contain different amounts of carbon. As can be seen from Table 7.1, among the fossil fuels, coal contains the most carbon per unit of energy produced and natural gas contains the least.

The upward slopes of these supply curves reflect increasing availability of domestic oil at higher prices, given sufficient time to develop those resources. Imported foreign oil is reflected by three supply curves: P_{w1} reflects the observed world price, P_{w2} includes a "vulnerability premium" in addition to the world price, and P_{w3} adds in the per-unit climate change damages due to consuming more imported oil. The vulnerability premium reflects the additional national security costs caused by imports. All three curves are drawn horizontally to the axis to reflect the assumption that any single importing country's action on imports is unlikely to affect the world price for oil.

As shown in Figure 7.2, in the absence of any correction for national security and climate change considerations, the market would generally demand and receive *D* units of oil. Of this total amount, *A* would be domestically produced and *D* – *A* would be imported. Why?

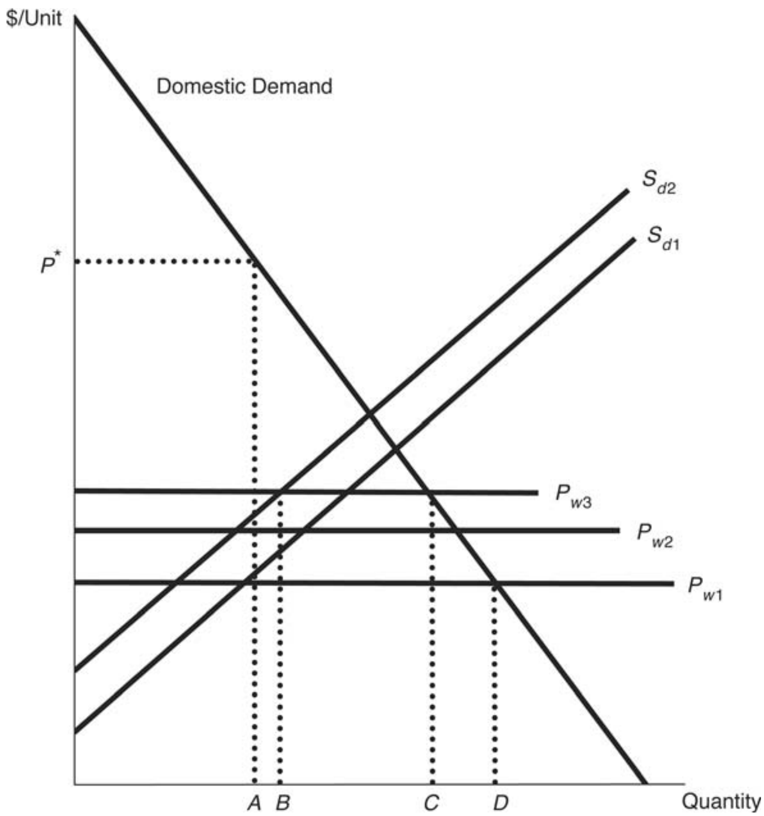


Figure 7.2 The National Security Problem

Table 7.1 Carbon Content of Fuels Pounds of CO_2 Emitted per Million Btu of Energy for Various Fuels

Coal (anthracite)	228.6
Coal (bituminous)	205.7
Coal (lignite)	215.4
Coal (subbituminous)	214.3
Diesel fuel & heating oil	161.3
Gasoline	157.2
Propane	139.0
Natural gas	117.0

Source: Energy Information Administration. Retrieved from https://www.eia.gov/tools/faqs/faq.php?id=73&t=11 (accessed January 28, 2017).

In an efficient allocation, incorporating the national security and climate change considerations, only *C* units would be consumed. Of these, *B* would be domestically produced and *C* – *B* would be imported. Comparing these two outcomes, note that, because national security and climate change are externalities, the market in general tends to consume too much oil and vulnerable imports exceed their efficient level.

What would happen during an embargo? Be careful! At first glance, you would guess that we would consume where domestic supply equals domestic demand, but that is not right. Remember that S_{d1} is the domestic supply curve, *given enough time to develop the resources*. If an embargo hits, developing additional resources cannot happen immediately (multiple-year time lags are common). Therefore, in the short run, the supply curve becomes perfectly inelastic (vertical) at *A*. The price will rise to P^* to equate supply and demand. As the graph indicates, the loss in consumer surplus during an embargo can be very large indeed.

How can importing nations react to this inefficiency? As Debate 7.2 shows, several strategies are available.

The importing country might be able to become self-sufficient, but should it? If Figure 7.2 adequately represents the situation, then the answer is clearly no. The net benefit from self-sufficiency (the allocation where domestic supply S_{d1} crosses the demand curve) is clearly lower than the net benefit from the efficient allocation (C).

DEBATE 7.2

How Should Countries Deal with the Vulnerability of Imported Oil?

Many countries import most of their oil. Since oil is a strategic material, how can the resulting vulnerability to import disruption be addressed?

One vision focuses on a strategy of increasing domestic production, not only of oil, but also of natural gas and coal. This vision includes opening up new oil fields in such places as coastal waters or public lands, as well as expanding the production of newer sources such as tar sands or oil shale. Tax incentives and subsidies could be used to promote domestic production.

Another vision emphasizes energy efficiency and energy conservation. Pointing out that expanded domestic production could exacerbate environmental problems (including climate change), this vision promotes such strategies as mandating standards for fuel economy in automobiles, enacting energy efficiency standards for appliances, and making buildings much more energy efficient.

Using economic analysis, figure out what the effects of these two different strategies would be on the implementing country with respect to (1) oil prices in the short run and the long run, (2) emissions affecting climate change, and (3) imports in the short run and the long run. What strategy or strategies would you like to see chosen by your country? Why?

Why, you might ask, is self-sufficiency so inefficient when embargoes obviously impose so much damage and self-sufficiency could grant immunity from this damage? Why would we want any imports at all when national security is at stake?

The simple answer is that the vulnerability premium is lower than the cost of becoming self-sufficient, but that response merely begs the question, "why is the vulnerability premium lower?" It is lower for three primary reasons: (1) embargoes are not certain events—they may never occur; (2) steps can be taken to reduce vulnerability of the remaining imports; and (3) expanding current domestic production via subsidies would incur a user cost by lowering the domestic amounts available to future users.

The expected damage caused by one or more embargoes depends on the likelihood of occurrence, as well as the intensity and duration. This means that the P_{w2} curve will be lower for imports having a lower likelihood of being embargoed. Imports from countries less hostile to our interests are more secure and the vulnerability premium on those imports is smaller.[4]

For any remaining vulnerable imports, certain contingency programs can be adopted to reduce the damage an embargo would cause. The most obvious measure is to develop a domestic stockpile of oil to be used during an embargo. The United States has taken this route. The stockpile, called the *strategic petroleum reserve*, was originally designed to contain 1 billion barrels of oil (see Example 7.1). A 1 billion barrel stockpile could replace 3 million barrels a day for slightly less than 1 year or a larger number of barrels per day for a shorter period of time. This reserve could serve as a temporary alternative domestic source of supply, which, unlike other oil resources, could be rapidly deployed on short notice. It is, in short, a form of insurance. If this protection can be purchased cheaply, implying a lower P_{w2}, imports become more attractive.

To understand the third and final reason that paying the vulnerability premium would be less costly than self-sufficiency, we must consider vulnerability in a dynamic, rather than static, framework. Because oil is a depletable resource, a user cost is associated with its efficient use. To reorient the extraction of that resource toward the present, as a self-sufficiency strategy would do, reduces future net benefits. Thus, the self-sufficiency strategy tends to be myopic in that it solves the short-term vulnerability problem by creating a more serious one in the future. Paying the vulnerability premium creates a more efficient balance between the present and future, as well as between current imports and domestic production.

We have established the fact that government can reduce our vulnerability to imports, which tends to keep the risk premium as low as possible. Certainly for oil, however, even after the stockpile has been established, the risk premium is not zero; P_{w1} *and* P_{w2} will not coincide. Consequently, the government must also concern itself with achieving both the efficient level of consumption and the efficient share of that consumption borne by imports. Let's examine some of the policy choices.

As noted in Debate 7.2, energy conservation is one possible approach to the problem. One way to accomplish additional conservation is by means of a tax on fossil fuel consumption. Graphically, this approach would be reflected as a shift inward of the after-tax demand curve. Such a tax could reduce energy consumption and emissions of greenhouse gases to an efficient level. It could not, however, achieve the efficient share of imports, since the tax falls on *all* energy consumption, whereas the security problem involves only imports. While energy conservation may increase the net benefit, it cannot ever be the sole policy instrument used or an efficient allocation will not be attained.

Another strategy, the expansion of domestic supply, is already occurring due to fracking in places such as the Bakken Formation. According to the U.S. Geological Service, one of the larger domestic discoveries in recent years of unconventional oil can be found in the Bakken and associated formations in Montana and North Dakota. Parts of these shale formations extend into the Canadian Provinces of Saskatchewan and Manitoba. The introduction of hydraulic fracturing technology to the region in 2008 caused a boom in production and a reduction in imports.

EXAMPLE 7.1

Strategic Petroleum Reserve

The U.S. strategic petroleum reserve (SPR) is the world's largest supply of emergency crude oil. The federally owned oil stocks are stored in huge underground salt caverns along the coastline of the Gulf of Mexico.

Decisions to withdraw crude oil from the SPR are made by the president under the authority of the Energy Policy and Conservation Act. In the event of an "energy emergency," SPR oil would be distributed by competitive sale. In practice, what constitutes an energy emergency goes well beyond embargoes. The SPR has been used rarely and no drawdown involved protecting against an embargo. Some examples of drawdowns include:

- During Operation Desert Storm in 1991, sales of 17.3 million barrels were used to stabilize the oil market in the face of supply disruptions arising from the war.
- After Hurricane Katrina caused massive damage to the oil production facilities, terminals, pipelines, and refineries along the Gulf regions of Mississippi and Louisiana in 2005, sales of 11 million barrels were used to offset the domestic shortfall.
- A series of emergency exchanges conducted after Hurricane Gustav, followed shortly thereafter by Hurricane Ike, reduced the level by 5.4 million barrels.
- During 2011, 30.59 million barrels were sold in response to sustained interruptions in global supplies due to civil unrest in Libya. President Obama authorized the sale as part of a larger coordinated release of petroleum by International Energy Agency countries.

Building up the reserve is accomplished by the Royalty-in-Kind program. Under the Royalty-in-Kind program, producers who operate leases on the federally owned Outer Continental Shelf are required to provide from 12.5 to 16.7 percent of the oil they produce to the U.S. government. This oil is either added directly to the stockpile or sold to provide the necessary revenue to purchase oil to add to the stockpile. In April 2011, however, Congress rescinded all funding for the SPR expansion project.

Subsequently as part of its 2018 budget, the Trump administration proposed to sell off half the oil in the Strategic Petroleum Reserve and use the money to reduce the deficit. Do you think this reduction in size improves efficiency? Why or why not? If you think it is a good idea, do you believe it was always too big or have circumstances changed? If the latter, what circumstances have changed, and how has that lowered the optimal level of the strategic petroleum reserve?

Source: U.S. Department of Energy Strategic Petroleum Reserve website, http://energy.gov/fe/services/petroleum-reserves/strategic-petroleum-reserve (accessed October 20, 2016).

As Example 7.2 points out, however, even a local community that takes advantage of that boom can be in for a rocky ride if its economy depends exclusively or primarily on that single commodity.

Diagrammatically, the effect of the fracking induced expansion in domestic oil production would be portrayed in Figure 7.2 as a shift of the domestic supply curve to the right. Notice

EXAMPLE 7.2

Fuel from Shale: The Bakken Experience

The boom in oil production made possible by fracking resulted in North Dakota becoming the most rapidly growing state in the nation. Population increased in response to rising wages and lots of retail activity and public infrastructure was built to accommodate the rising population.

Then in 2014 dropping oil prices (in part due to an oil glut resulting from increased production from domestic shale) reversed the process. As the prices fell, small rural towns were hit particularly hard.

Williston, North Dakota, for example, experienced job losses both in the oil industry and in the retail sectors built up to serve the influx of new workers. Not only did the population begin to decline, but the investments in public infrastructure made to accommodate the larger population become more difficult to finance. During the boom Williston had built a new $57 million high school, but declining economic activity caused by the lower prices caused a decline in the tax base needed to pay for that school.

Boom-and-bust cycles can be especially devastating to small, resource-dependent communities like Williston. With little diversification of their economic base they become especially vulnerable to swings in the prices of the resources on which they depend so heavily.

Sources: USGS releases new oil and gas assessment for Bakken and Three Forks formations. Retrieved from www2.usgs.gov/blogs/features/usgs_top_story/usgs-releases-new-oil-and-gas-assessment-for-bakken-and-three-forks-formations/; Hydraulic fracturing, www.epa.gov/hydraulicfracturing (accessed February 4, 2016); Oldham, Jennifer, & Philips, Matthew. (2016). The Bakken bust hits North Dakota hard (February 4), www.bloomberg.com/news/articles/2016-02-04/the-bakken-bust-hits-north-dakota-hard (accessed October 11, 2016); Millsap, Adam. (2016). What the boom and bust of Williston, North Dakota teaches us about the future of cities (June 7), www.forbes.com/sites/adammillsap/2016/06/07/williston-nd-and-the-rise-and-fall-of-american-cities/#66ae277a6c81 (accessed October 11, 2016).

that one effect would be to reduce the share of imports in total consumption (an efficient result) and that is already happening. Net energy imports (imports minus exports) peaked in 2005. Since 2005, imports have declined while exports have increased (see Figure 7.3).

The domestic supply expansion would, however, increase climate change emissions (an inefficient result). This strategy also tends to drain domestic reserves faster, which could make the nation more vulnerable in the long run (another inefficient result). The expansion of domestic fossil fuels reduces imports, but the lower prices and resulting increased consumption also tend to intensify the climate change problem.

A third approach would tailor the response more closely to the national security problem. One could use either a tariff on imports equal to the vertical distance between P_{w1} *and* P_{w2} or a quota on imports equal to $C - B$. With either of these approaches, the price to consumers would rise to P_1, total consumption would fall to C, and imports would be $C - B$. This achieves the appropriate balance between imports and domestic production (an efficient result), but it does not internalize the climate change cost from increasing domestic production (an inefficient result). As developed in more detail in Chapter 17, imposing a separate price on carbon would be a necessary component of the package in order to internalize the climate externality.

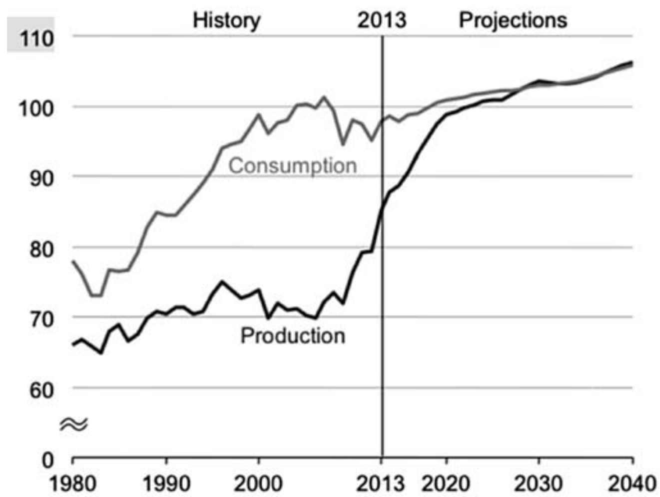


Figure 7.3 Total Energy Production and Consumption 1980–2040

Source: U. S. Energy Information Administration, Annual Energy Outlook 2015, p.17

Electricity: The Role of Depletable Resources

Most observers who think about energy futures see electricity as assuming an increasingly large role in the total energy picture. Evolving technologies for using electricity for both heating and cooling (heat pumps) and transportation (electric vehicles) figure importantly in this perspective.

What energy sources should be used to generate that electricity? While the industrialized world currently depends on conventional sources of oil, coal, and gas for most of our energy, over the long run, in terms of both climate change and national security issues, the obvious solution involves a transition to domestic renewable sources of energy that do not emit greenhouse gases. What role does that leave for the other depletable resources such as natural gas and uranium, which are used to generate electricity?

Although some observers believe the transition to renewable sources of electricity will proceed so rapidly that using these fuels as a bridge will be unnecessary, many others believe that depletable fuels will continue to play a significant transition role.

In the United States, coal, previously a contender, has been losing out to natural gas due mainly to its lower costs resulting from the expansion of fracking. However the increasing focus on reducing greenhouse gases is also a factor since natural gas has a lower carbon content when combusted.

Although other contenders do exist, the fuel other than natural gas receiving the most attention (and controversy) as a transition fuel is uranium. As a potential transition fuel used in nuclear electrical-generation stations, nuclear has its own limitations—safety and economics.

With respect to safety, two sources of concern stand out: (1) nuclear accidents or sabotage, and (2) the storage of radioactive waste. Is the market able to make efficient decisions about the role of nuclear power in the energy mix? In both cases, the answer is no, given the current decision-making environment. Let's consider these issues one by one.

The production of electricity by nuclear reactors involves radioactive elements. If these elements escape into the atmosphere and come in contact with humans in sufficient concentrations, they can induce birth defects, cancer, or death. Although some radioactive elements may also escape during the normal operation of a plant, the greatest risk of nuclear power is posed by the threat of nuclear accidents or deliberate sabotage.

As the accident in Fukushima, Japan, in 2011 made clear, nuclear accidents could inject large doses of radioactivity into the environment. Unlike other types of electrical generation, nuclear processes continue to generate heat long after the reactor is turned off. This means that the nuclear fuel must be continuously cooled, or the heat levels will escalate beyond the design capacity of the reactor shield. If the high heat causes the reactor vessel to fracture, clouds of radioactive gases and particulates will be released into the atmosphere.

An additional concern arises from the need to store nuclear wastes. The waste-storage issue relates to both ends of the nuclear fuel cycle—the disposal of uranium tailings from the mining process and spent fuel from the reactors—although the latter receives most of the publicity. Uranium tailings contain several elements, the most prominent being thorium-230, which decays with a half-life of 78,000 years to a radioactive, chemically inert gas, radon-222. Once formed, this gas has a very short half-life (38 days).

The spent fuel from nuclear reactors contains a variety of radioactive elements with quite different half-lives. In the first few centuries, the dominant contributors to radioactivity are fission products, principally strontium-90 and cesium-137. After approximately 1000 years, most of these elements will have decayed, leaving the transuranic elements, which have substantially longer half-lives. These remaining elements would remain a risk for up to 240,000 years. Thus, decisions made today affect not only the level of risk borne by the current generation—in the form of nuclear accidents—but also the level of risk borne by a host of succeeding generations (due to the longevity of radioactive risk from the disposal of spent fuel).

Nuclear power has also been beset by economic challenges. New nuclear power plant construction has become much more expensive than previously, in part due to the increasing regulatory requirements designed to provide a safer system. In the late twentieth century as its economic advantage over coal dissipated, the demand for new nuclear plants declined. For example, in 1973, in the United States, 219 nuclear power plants were either planned or in operation. By the end of 1998, that number had fallen to 104, a difference due primarily to cancellations.

The transition to lower carbon fuels has created some renewed interest in the nuclear option. The first new nuclear generator in the United States in 20 years entered commercial operation in Tennessee in 2016, a year in which globally nuclear power plants provided a bit over 11 percent of the world's electricity. The World Nuclear Association announced in 2016 that some 440 nuclear power reactors were operating in 31 countries and that over 60 power reactors were currently being constructed in 13 countries. China was constructing eight new reactors a year.

What future role nuclear power will play in other countries after Fukushima remains to be seen.

Electricity: Transitioning to Renewables

Ultimately, our energy needs will have to be fulfilled from renewable energy sources, either because the depletable energy sources have been exhausted or, as is more likely, the environmental costs of using the depletable sources have become so high that renewable sources will be cheaper.

Many of these renewable sources of energy, such as hydroelectric power, wind, photovoltaics, and ocean tidal power are used to generate electricity. These sources not only allow electricity generation to be more sustainable, but they reduce the country's dependence on fossil fuels.

Renewable energy comes in many different forms. Different sources will have different comparative advantages so, ultimately, a mix of sources will be necessary. As Debate 7.3 suggests, the path to greater reliance on renewables is certainly not free of controversy even within the environmental community.

DEBATE 7.3

Dueling Externalities: Should the United States Promote Wind Power?

On the surface the answer seems like a no-brainer, since wind power is a renewable energy source that emits no greenhouse gases, unlike all the fossil fuels it would be likely to replace. Yet some highly visible, committed environmentalists, including Robert F. Kennedy, Jr., have strongly opposed wind projects. Why has this become such a contentious issue?

Opposition to wind power within the environmental community arises for a variety of reasons. Some point out that the turbines can be noisy for those who live, camp, or hike nearby. Others note that these very large turbines can be quite destructive to bats and birds, particularly if they are constructed in migratory pathways. And a number of opponents object to the way the view would be altered by a large collection of turbines on otherwise-pristine mountaintops or off the coast.

Both the benefits from wind power (reduced air pollution including impact on the climate) and the costs (effects on aesthetics, birds, and noise) are typically externalities. This implies that the developers and consumers of wind power will neither reap all of the environmental benefits from reduced impact on the climate, nor will they typically bear all the environmental costs. Making matters even more difficult, some of the environmental costs will be concentrated on a relatively few people (those living nearby), while the benefits will be conferred on all global inhabitants, most of whom will bear absolutely none of these costs. Since the presence of externalities typically undermines the ability of a market to produce an efficient outcome, it is not surprising that the permitting process for new wind power facilities is highly regulated. Regulatory processes generally encourage public participation by holding hearings. The concentrated costs imposed on those living nearby may be an effective motivator to attend the public hearings, which are likely to be held near the proposed site; the diffuse benefits will likely be a less effective motivator for attendance by proponents.

With environmental externalities lying on both sides of the equation and with many of the environmental costs concentrated on a relatively small number of people, it is understandable that these hearings have become so contentious, and that the opposition to wind power is so strong.

Sources: Kennedy Jr., R. F. (December 16, 2005). An ill wind off Cape Cod (op-ed). *The New York Times*; Barringer, F. (June 6, 2006). Debate over wind power creates environmental rift. *The New York Times*.

The extent to which these sources will penetrate the market will depend upon their relative cost and consumer acceptance. New systems are usually initially less reliable and more expensive than old systems. Once they mature, reliability normally increases and cost declines; experience is a good teacher.

Since the early producers and consumers—the pioneers—experience both lower reliability and higher costs, procrastination can be an optimal individual strategy. From an individual point of view, waiting until all the bugs have been worked out and costs come down reduces the risk of making the investment.

From a social point of view, however, if every producer and consumer procrastinates about switching, the industry will never be able to reach a sufficient scale of operation and will not be able to gain enough experience to reach the level of reliability and lower cost that will be necessary to reach the specified renewable goals. How can these initial barriers be overcome?

One strategy involves establishing specific renewable resource goals with deadlines for meeting them. For example, the E.U. Renewable Resource Directive, which establishes an overall policy for the production and promotion of energy from renewable sources in the E.U., requires at least 20 percent of its total energy needs be filled by renewables by 2020. In addition, under the Directive all E.U. countries must also ensure that at least 10 percent of their transport fuels come from renewable sources by 2020.

More recently, E.U. countries have agreed on strengthening their initial renewable target to assure that at least 27 percent of final energy consumption is met from renewables in the E.U. as a whole by 2030. One of the expanding sources is offshore wind. In June 2017 Germany, Denmark, and Belgium backed a pledge to install 60 gigawatts of new offshore wind power next decade, more than five times the world's existing capacity.

Another strategy subsidizes pioneer investments via the tax code.[5] This is commonly done, for example, with production or investment tax credits. Once the market is sufficiently large that it can begin to take advantage of economies of scale and overcome the initial sources of unreliability, the subsidies could be eliminated.

Another common policy approach for overcoming these obstacles involves combining Renewable Portfolio Standards (RPS) for electricity generation with Renewable Energy Credits (RECs). Renewable portfolio standards stipulate a minimum percentage of the total electricity that must be generated by each generator from specified renewable sources such as wind, hydro, or solar. The generating entity can either meet that standard directly by generating the requisite proportion from the specified renewable sources, or indirectly by purchasing renewable energy credits from independent generators.

An independent generator of electricity from a renewable source actually produces two saleable commodities. The first is the electricity itself, which can be sold to the grid, while the second is the renewable energy credit that turns the environmental attributes (such as the fact that it was created by a qualifying renewable source) into a legally recognized form of property that can be sold separately. Generally, renewable generators create one REC for every 1000 kilowatt-hours (or, equivalently, 1 megawatt-hour) of electricity placed on the grid.

Providing this form of flexibility in how the mandate is met lowers the compliance cost, not only in the short run (by allowing the RECs to flow to the areas of highest need), but also in the long run (by making renewable source generation more profitable in areas not under a RPS mandate than it would otherwise be). By 2013 some 29 states and the District of Columbia had a renewable energy standard, with seven more having non-binding goals. Many of those also had REC programs.

How cost-effective have these polices been? Example 7.3 discusses a study that looks specifically at that question.

EXAMPLE 7.3

The Relative Cost-Effectiveness of Renewable Energy Policies in the United States

The United States depends on both renewable portfolio standards, and a suite of production and investment tax credits to promote renewable resources that reduce carbon emissions. It also uses a completely different approach to reduce carbon emissions, one that puts a price directly on those emissions. Although we discuss this carbon-pricing approach in some detail in Chapters 15 and 17, here we simply ask how cost-effective a comprehensive policy such as carbon pricing is relative to policies that are targeted exclusively on promoting renewable resources.

Using a highly detailed model of regional and interregional electricity markets Palmer et al. (2011) examine this question over a time horizon covering the period from 2010 to 2035. The analysis evaluates each of these policy approaches in terms of their relative effectiveness and cost-effectiveness in reducing carbon emissions, their effectiveness in promoting renewable resource electricity generation, and their effects on electricity prices.

Between the two renewable resource policies the tax credit was found to be the least cost-effective, with the renewable portfolio somewhat better. Because it involves a subsidy and the other polices do not, the tax credit leads to relatively lower electricity prices, which supports greater electricity consumption and hence relatively larger emissions. This offsetting increase in emissions diminishes the tax credit's cost-effectiveness.

However the best policy turned out to be the third, a particular form of carbon pricing known as cap-and-trade. As we shall see in more detail in Chapter 17, the price a cap-and-trade policy puts on emissions creates very cost-effective incentives for emissions reduction. The dominance of this approach should therefore not be surprising. Additionally it is the only considered policy that increases the relative cost of using <u>nonrenewable</u> higher carbon sources. Neither the tax credit nor the renewable portfolio standard discourage the use of high-carbon nonrenewable technologies at all; they apply only to renewable sources.

Source: Palmer, K., Paul, A., Woerman, M., & Steinberg. D. C. (2011). Federal policies for renewable electricity: Impacts and interactions. *Energy Policy*, *39*(7), 3975–3991.

Another quite different approach to promoting the use of renewable resources in the generation of electric power is known as a feed-in tariff. Used more commonly in Europe, a feed-in tariff specifies the prices received by anyone who installs qualified renewable capacity that sells electricity to the grid. The level of these prices (typically determined in advance by the rules of the program) is based upon the costs of supplying the power. Specifically they are set sufficiently high so as to assure installers that they will receive a reasonable rate of return on their investment. While in Germany this incentive payment is guaranteed for 20 years for each installed facility, each year the magnitude of the payment for newly constructed generators is reduced (typically in the neighborhood of 1–2 percent per year) in order to reflect expected technological improvements and economies of scale.

A feed-in tariff actually offers two different incentives: (1) it provides a price high enough to promote the desired investment and (2) it guarantees the stability of that price over time rather than forcing investors to face the market uncertainties associated with fluctuating fossil fuel prices or subsidies that come and go.

Of course, when higher prices are paid to renewable investors, these costs must be borne by someone. In Germany the higher costs associated with the feed-in tariffs were typically passed along to electricity ratepayers. German electricity rates have been, as a result, relatively high. In principle these higher costs should be temporary, since rising fossil fuel costs would be expected to rise above the relatively stable prices dictated by feed-in tariffs. Will that prove to be true in practice? Stay tuned.

Spain took a different approach that produced different results. It refused to allow its electric utilities to pass on the increased cost of electricity resulting from the feed-in tariffs to consumers. As a result, its electricity system financial deficit became unsustainable, and in 2013 Spain halted new feed-in tariff contracts for renewable energy.

As we have seen so often in other policy circumstances, the implementation details matter.

Electricity: Energy Efficiency

As the world grapples with creating the right energy portfolio for the future, energy-efficiency policy is playing an increasingly prominent role. An activity is said to be *energy efficient* if it is produced with the minimum amount of energy input necessary to produce a given level of that activity. Activities covered by this definition can be as diverse as heating or lighting a building, driving 100 miles, or producing a ton of paper. In recent years the amount of both private and public money being dedicated to promoting energy efficiency has increased a great deal.

The role for energy efficiency in the broader mix of energy polices depends, of course, on how large the opportunity is. Estimating the remaining potential is not a precise science, but the conclusion that significant opportunities remain seems inescapable.

The existence of these opportunities can be thought of as a necessary, but not sufficient, condition for government intervention. Depending upon the level of energy prices and the discount rate, the economic return on these investments could be too low to justify intervention. In that case the costs of the policy intervention would exceed any gains that would result.

The strongest case for government intervention flows from the existence of externalities. Markets are not likely to internalize these external costs on their own. The natural security and climate change externalities mentioned above, as well as other external co-benefits such as pollution-induced community health effects, certainly imply that the market undervalues investments in energy efficiency.

The analysis provided by economic research in this area, however, makes it clear that the case for policy intervention extends well beyond externalities. Internalizing externalities is a very important, but incomplete, policy response.

Consider just a few of the other foundations for policy intervention. Inadequately informed consumers can impede rational choice, as can a limited access to capital (preventing paying the up-front costs for the more energy-efficient choice even when the resulting energy savings would justify the additional expense in present value terms). Perverse incentives can also play a role, as in the case of someone who lives in a room (think dorm) or apartment where the amount of energy used is not billed directly, resulting in a marginal cost of additional energy use of zero for the occupant. Another related case of perverse incentives arises for rental housing units (Example 7.4).

EXAMPLE 7.4

Energy Efficiency in Rental Housing Markets

Economic analysis can not only help us understand the empirical finding that rental-housing units are typically less energy efficient than owner-occupied units, but also help us to understand the relative efficacy of policies to promote less energy waste in rental units.

To understand the sources of energy waste, consider the incentives. In an owner-occupied unit, the owner bears all the costs and receives all the benefits (the resulting lower energy costs) from an investment in energy efficiency. In a typical rental unit, however, the renter pays for the energy used, while the landlord would pay for any energy efficient investments (such as insulation or an efficient heating system). When prospective renters have no access to credible information on the energy costs associated with this unit (a common case), the rents for various units would not reflect their energy cost differences. Since the costs of investments to reduce energy waste in the rental unit in this case cannot normally be recovered via higher rents, a landlord would underinvest in energy efficiency.

Yet energy efficiency is clearly a cost-effective way not only to reduce waste (by lowering energy costs), but also to lower carbon emissions as well. Can these market barriers be overcome?

A recent experimental economics study addresses this question by examining four policy treatments: (1) mandatory and (2) voluntary energy-efficiency ratings for the unit (similar to energy-efficiency stars for appliances), (3) a performance regulatory standard (similar to energy-efficiency standards for appliances), and (4) a cost-sharing arrangement where landlords would be required to pay a fixed percentage of their tenant's energy bill.

In the baseline treatment (no policies), the authors confirm the theoretical expectation that owners typically invest more in owner-occupied units than landlords invest in rental units.

Among the policy treatments they find that the availability of verified and costless information on rental unit energy costs unequivocally reduces waste, with mandatory information and voluntary information both achieving a high level of efficiency. The regulatory approach was found to result in a higher average investment than the mandatory and voluntary information schemes, but it resulted in fewer properties available in the market; apparently some landlords chose to leave the rental market rather than comply with the regulation. A cost-sharing policy achieves similar efficiency levels as the regulatory standard, but a significantly lower level of efficiency than the voluntary and mandatory information schemes.

The effectiveness of information strategies found by this study is good news indeed, but two caveats must be kept in mind. First, most actual information strategies are not costless to landlords, as they were assumed to be in this study. To the extent that landlords bear some or all of the costs of providing certified information, this study would overestimate (to some unknown degree) the effectiveness of these strategies. Second, experimental economics studies work with participants in a lab, not with data based upon actual market choices. As noted in Chapter 1, lab results are typically informative, but they do not always produce the results drawn from actual field experience.

Source: Burfurd, I., Gangadharan, L., & Nemes, V. (2012). Stars and standards: Energy efficiency in rental markets. *Journal of Environmental Economics and Management*, *64*(2), 153–168.

Could policies to increase energy efficiency (such as subsidizing the cost of weatherizing your home) trigger offsetting responses that reduce their effectiveness? As Example 7.5 points out, in principle they could.

Electricity: Targeted Distributed Energy

One characteristic of distributed energy sources, such as solar, wind, or even energy efficiency, is that they can be located near users. Contrast this with large power plants, which are centrally located. By locating close to users, distributed energy sources can lower the distance (and hence the cost) of transporting electricity from source to user.

Could targeting these distributed sources at areas facing transmission constraints eliminate the expense of building new transmission lines and hence be a cost-effective component in the energy mix needed by that region? As Example 7.6 points out, in the right circumstances, it can.

EXAMPLE 7.5

Energy Efficiency: Rebound and Backfire Effects

Energy efficiency policies can trigger offsetting feedbacks that lower their effectiveness. The literature distinguishes two possible outcomes—the rebound effect and the backfire effect.

Consider an example. A weatherization subsidy lowers both the amount and cost of energy needed to heat or cool the space in your home. Would a homeowner respond to that lower cost by turning up (or down) the thermostat or heating or cooling more rooms? Any increased energy consumed in response to its lower cost is known as a rebound effect. The backfire effect occurs when the rebound effect is so large that a weatherization subsidy actually causes an increase in the amount of energy consumed.

What is the evidence on these effects? A review of the studies seeking to answer this question finds "that the existing literature does not support claims that energy efficiency gains will be reversed by the rebound effect" (Gillingham et al., 2016, p. 85). In other words the existing literature provides little, if any, support for a backfire effect. It does, however, find evidence of rebound effects that can, depending upon the context, range as high as 60 percent.

What does this imply for the effectiveness of energy efficiency policy? The authors conclude that this evidence does imply that energy efficiency policies may be less effective in reducing energy (and reducing carbon emissions) than thought, since rebound effects can offset to some degree the direct energy-reducing effects of the policy. They also, however, note that the welfare effects of the rebound effects are ambiguous—while the increased energy use lowers welfare due to the offsetting increase in damaging climate change impacts, it raises the welfare arising from having more comfortable homes. The existing literature does not provide answers as to which is larger.

Source: Gillingham, Kenneth, Rapson, David, & Wagner, Gernot. (2016). The rebound effect and energy efficiency policy. *Review of Environmental Economics and Policy*, *10*(1) (Winter), 68–88.

EXAMPLE 7.6

Thinking about Cost Reduction Outside of the Box: The Boothbay Pilot Project

The Boothbay Harbor region, a popular summer tourist destination on the Maine coast at the end of a peninsula, has a problem. The existing electricity transmission line serving the area does not have the capacity to handle its large and growing summer electrical demand. The traditional response, upgrading the transmission line, would be very expensive. Could the problem be solved at lower cost in another way?

The Maine Public Utilities Commission decided to discover whether non-transmission alternatives (NTAs)—such as distributed generation, efficiency, storage, and new smart grid technologies—could solve electric grid reliability needs at lower cost and with less pollution than new transmission lines or transmission system upgrades. In 2012, the Commission established the Boothbay Smart Grid Reliability Pilot project to test the NTA hypothesis. In its first 3-year initial phase, the Boothbay Pilot sought to provide experience-based evidence on whether a portfolio of NTAs could reduce electricity load under peak conditions on specific transmission assets in the Boothbay subregion of Central Maine by 2 megawatts (MW), thereby avoiding an estimated $18 million transmission line rebuild.

What did the pilot project show? Based upon the results for the initial phase of this project the evidence suggests that the net cost of the accepted NTAs, together with administrative and operational expenses, is projected to be less than 33 percent of the cost of building a new transmission line and would save ratepayers approximately $18.7 million (including energy savings) over the 10-year project life through 2025.

These results suggest that targeting an integrated package of distributed solutions at those geographic areas facing transmission constraints can produce grid benefits well beyond the direct services they provide to individual customers.

Source: Grid Solar LLC, Final report: Boothbay Sub-Region Smart Grid Reliability Pilot Project (January 19, 2016).

Another new niche for distributed energy sources is to supply remote areas that previously have never had access to the electrical grid. As Example 7.7 points out, townships in Africa are using solar microgrids and novel, technology-based financing models to supply these remote areas.

EXAMPLE 7.7

The Economics of Solar Microgrids in Kenya

Entrepreneurs are constructing solar photovoltaic microgrids in remote rural areas of Kenya. Microgrids in Kenya are small electricity generation and distribution systems that can operate independently of larger grids. Due to their small scale they typically cannot supply electricity as cheaply as the larger grid, but for remote areas that do not have access to the larger grid, the electricity from solar microgrids is typically cheaper than the other local energy alternatives such as producing electricity via diesel generators.

Installing these microgrids requires capital investment and these villages are typically poor and do not have access to this capital. How do they get around this significant barrier? Entrepreneurs supply the capital, own the solar panels, and sell the electricity to local homes and businesses. The product is electricity, not panel installation.

In one financial model, cloud-based software keeps track of consumption and payments via smart meters. The smart meters measure and control power to each customer in town by communicating remotely with payments software. Although power is cut off when the prepaid credit is exhausted, customers can top up their credit when they wish, in amounts as small as a few cents.

One problem is that the greatest demand for power is at night when the sun is not shining, but that problem is overcome with battery storage units that typically hold up to 24 hours of electrical consumption. Storage adds to the cost, but the cost increase apparently is not enough to eliminate the economic advantages of the microgrid to local residents or the profitability to the entrepreneurs. Analysis at the Lawrence Berkeley National Lab suggests that wind and solar can now be economically and environmentally competitive for a large portion of Africa.

Sources: Pearce, Fred. (2015). African lights: Solar microgrids bring power to Kenyan villages (October 27) Yale Environment 360. Available at: http://e360.yale.edu/features/african_lights_microgrids_are_bringing_power_to_rural_kenya; Lawrence Berkeley National Laboratory. (2017). The economic case for wind, solar energy in Africa. *ScienceDaily* (March 27). Available at: www.sciencedaily.com/releases/2017/03/170327172829.htm.

Summary

We have seen that the relationship between government and the energy market is not always harmonious and efficient. In the past, price controls have tended to reduce energy conservation, discourage exploration and supply, cause biases in the substitution among fuel types that penalize future consumers, and create the potential for abrupt, discontinuous transitions to renewable sources. This important example makes a clear case for less, not more, regulation.

This conclusion is not universally valid, however. Other dimensions of the energy problem, such as climate change and national security, suggest the need for some government role. Insecure foreign sources require policies such as tariffs and strategic reserves to reduce vulnerability and to balance the true costs of imported and domestic sources. In addition, government must ensure that the costs of energy fully reflect not only the potentially large environmental costs, including climate change, but also the national security costs associated with a dependence on foreign sources of energy.

Economic analysis reveals that no single strategy is sufficient to solve the national security and climate change problems simultaneously. Subsidizing domestic supply, for example, would reduce the share of imports in total consumption (an efficient result), but it would reduce neither consumption nor climate change emissions (inefficient results). On the other hand, energy conservation (promoted by a tax on energy consumption, for example) would reduce energy consumption and the associated emissions (efficient outcomes) but would not achieve the efficient share of imports (an inefficient result) since an energy tax falls on <u>all</u>

energy consumption, whereas the national security problem involves only imports. An energy tax also would fail to produce a fully efficient resolution for climate impacts since it would focus on energy per se, not the actual emissions emitted by that energy use, a factor that varies widely among fuels. A carbon tax, not an energy tax, would be needed to make this kind of distinction among fuels.

Given the environmental difficulties with all of the depletable transition fuels (tar sands, fracked oil and gas, as well as coal and uranium), energy efficiency and the promotion of renewable sources of energy are now playing (and will presumably continue to play) a larger role.

The menu of energy options as the economy transitions to renewable sources offers a large number of choices. It is far from clear what the ultimate mix will turn out to be, but it is very clear that government policy is a necessary ingredient in any smooth transition to a sustainable-energy future. Since many of the most important costs of energy use are externalities, an efficient transition to these renewable sources will not occur unless the playing field is leveled by internalizing the externalities. The potential for an efficient and sustainable allocation of energy resources by our economic and political institutions clearly exists, even if historically it has not always been achieved.

Discussion Questions

1. Should benefit-cost analysis play the dominant role, a complementary role, or no role in deciding the proportion of electric energy to be supplied by nuclear power? Why or why not?
2. Economist Abba Lerner once proposed a tariff on oil imports equal to 100 percent of the import price. This tariff is designed to reduce dependence on foreign sources as well as to discourage OPEC from raising prices (since, due to the tariff, the delivered price would rise twice as much as the OPEC increase, causing a large subsequent reduction in consumption). Should this proposal become public policy? Why or why not?
3. Does the fact that the strategic petroleum reserve has never been used to offset shortfalls caused by an embargo mean that the money spent in creating the reserve has been wasted? Why or why not?

Self-Test Exercises

1. During a worldwide recession in 1983, the oil cartel began to lose market share. Why would a recession make the cartel likely not only to lose sales, but also to lose market share?
2. Assume the demand and marginal cost conditions given in the second self-test exercise in Chapter 2. In addition, assume that the government imposes a price control at $P = \$80/3$. (a) Find the consumer and producer surplus associated with the resulting allocation. (b) Compare this price control allocation to the monopoly allocation in part (c) of that self-test exercise.
3. Some time ago, a conflict between a paper company and a coalition of environmental groups arose over the potential use of a Maine river for hydroelectric power generation. As one aspect of its case for developing the dam, the paper company argued that without hydroelectric power the energy cost of operating some specific paper machines would be so high that they would have to be shut down. Environmental groups countered that the

energy cost was estimated to be too high by the paper company because it was assigning all of the high-cost (oil-fired) power to these particular machines. That was seen as inappropriate because all machines were connected to the same electrical grid and therefore drew power from all sources, not merely the high-cost sources. They suggested, therefore, that the appropriate cost to assign to the machines was the much lower average cost. Revenue from these machines was expected to be sufficient to cover this average cost. Who was right?

4. Peaking plants, those that are only called into service during times of peak demand, are typically cheaper to build (compared to base-load plants, which operate all of the time), but have relatively high operating costs. Explain why it makes sense for utilities to use this lower-capital, high-operating-cost type of plant for peaking and the high-capital, lower-operating-cost type of plant for base load.
5. If OPEC raised the price of oil high enough, would that be sufficient to promote an efficient energy mix?
6. Label the following as *true*, *false*, or *uncertain* and explain your choice. (*Uncertain* means that it can be either true or false depending upon the circumstances.)

    a. All members of a resource cartel share a common objective, namely increase prices as much and as soon as possible.
    b. By holding prices lower than they would otherwise be, placing a price control on a depletable resource increases both the speed with which the resource is extracted over time and the cumulative amount ultimately extracted.
    c. A price control actually has no influence on the extraction path of a depletable resource until such time as the market price actually reaches the level of the price control.
    d. Forcing companies that drill offshore for oil to compensate victims of any oil spill from one of its facilities would be an efficient requirement.

7. Explain why the existence of a renewable energy credit market would lower the compliance costs for utilities forced to meet a renewable portfolio standard.
8. Using Figure 7.2, show how the level of oil imports and the price level would be affected if the country represented in that figure acted to internalize national security issues, but ignored climate change impacts.
9. a. Some new technologies, such as LED light bulbs, have the characteristic that they cost more to purchase than more conventional incandescent alternatives, but they save energy. How could you use the present value criterion to decide how cost-effective these new technologies are? What information would you need to do the calculations? How would the calculations be structured? How would you use the results of these calculations to decide on their cost-effectiveness?
    b. A typical monthly electrical bill has two components: (1) a fixed monthly change (e.g. \$10.00 a month) and (2) a usage component (e.g. \$0.14 per kilowatt-hour consumed). If a utility is planning to raise the amount they charge customers for electricity, would you expect that increase to discourage, encourage, or have no effect on the demand for LED light bulbs? Does it depend on which component they change? Why or why not?

10. Electric heat pumps are technologies that in the right circumstances can be cost-effective sources of heating. In a cold climate they frequently complement more typical energy sources such as oil or natural gas boilers in order to reduce total energy costs. In order to be cost-effective, however, the savings on oil and natural gas from using the heat pumps

must be large enough to justify both their initial costs and the subsequent cost of the additional electricity to run them. Would you expect the number of heat pump sales to be affected by the magnitude of local interest rates? Why or why not?

Notes

1 For a complete early recognition of this point, see Lee (1978).
2 Although we focus here on the role of fracking in natural gas production from shale, as noted in Example 7.1, it is also being used to increase oil production from shale.
3 The conclusion that a monopoly would extract a resource more slowly than a competitive mining industry is not perfectly general. It is possible to construct demand curves such that the extraction of the monopolist is greater than or equal to that of a competitive industry. As a practical matter, these conditions seem unlikely. That a monopoly would restrict output, while not inevitable, is the most likely outcome.
4 It is this fact that explains the tremendous U.S. interest in Canadian and Latin American oil, in spite of the fact that, historically, it has not necessarily been cheaper.
5 While we focus here on renewable technologies used to generate electricity, in general tax credits and other subsides are also used to promote renewable technologies such as biofuels or solar energy installations used to directly heat buildings.

Further Reading

Anthoff, D., & Hahn, R. (2010). Government failure and market failure: On the inefficiency of environmental and energy policy. *Oxford Review of Economic Policy*, *26*(2), 197–224. A selective survey of the literature to highlight what is known about the efficiency of particular kinds of policies, laws, and regulations in managing energy and environmental risk.

Gillingham, K., Newell, R. G., & Palmer, K. (2009). Energy efficiency economics and policy. *Annual Review of Resource Economics*, *1*, 597–620. Reviews economic concepts underlying decision-making in energy efficiency and conservation and the related empirical literature.

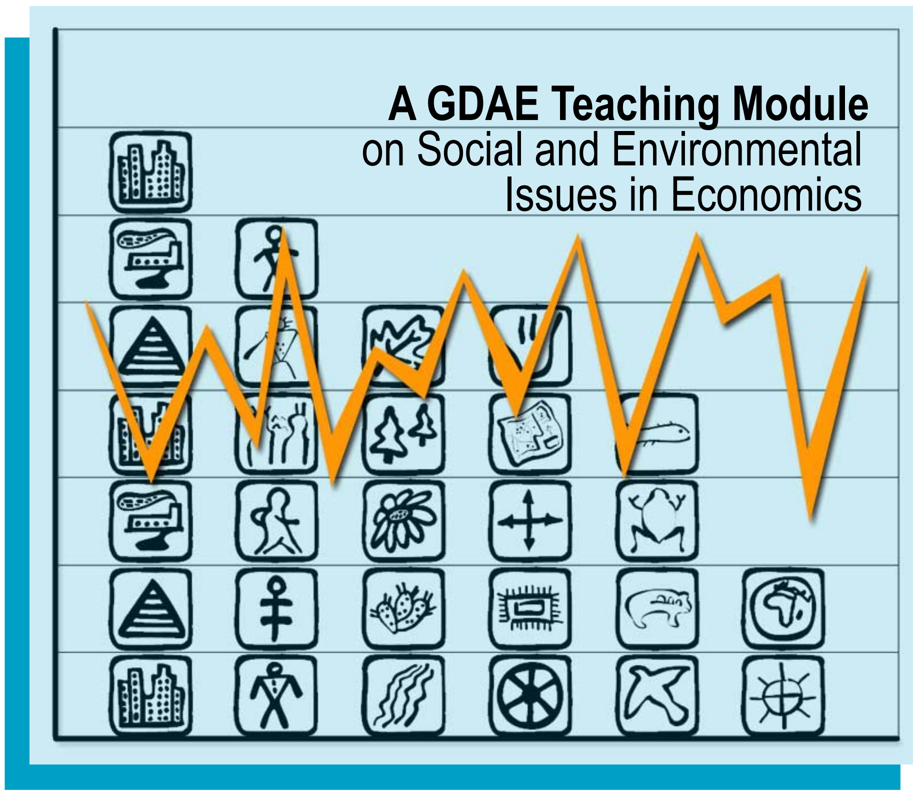
Gillingham, K. Rapsony, D. & Wagner, G. (2016). The rebound effect and energy efficiency policy. *Review of Environmental Economics and Policy*, *10*(1), 68–88. One argument against energy efficiency, called the rebound effect, occurs when the actual energy savings are lower than initially expected due to offsetting consumption increases. This article reviews the outstanding literature on the magnitude and policy implications of the rebound effect.

Schmalensee, R. (2009). Evaluating policies to increase electricity generation from renewable energy. *Review of Environmental Economics and Policy*, *6*, 45–64. Evaluates policies aimed at increasing the generation of electricity from renewable sources based upon on a review of experience in the United States and the European Union (E.U.).

Additional references and historically significant references are available on this book's Companion Website: www.routledge.com/cw/Tietenberg

The Economics of Global Climate Change

By Jonathan M. Harris
and Brian Roach



Global Development And Environment Institute
Tufts University
Medford, MA 02155
http://ase.tufts.edu/gdae

Students may also download the module directly from http://ase.tufts.edu/gdae.

Comments and feedback from course use are welcomed:

Global Development And Environment Institute
Tufts University
Medford, MA 02155
http://ase.tufts.edu/gdae

E-mail: gdae@tufts.edu

THE ECONOMICS OF GLOBAL CLIMATE CHANGE

based on: *Environmental and Natural Resource Economics: A Contemporary Approach* by Jonathan M. Harris (Houghton Mifflin, 2006, http://college.hmco.com)

1. CAUSES AND CONSEQUENCES OF CLIMATE CHANGE

Concern has grown in recent years over the issue of **global climate change**[1]. In terms of economic analysis, **greenhouse gas** emissions, which cause planetary climate changes, represent both an environmental **externality** and the overuse of a **common property resource.**

The atmosphere is a **global commons** into which individuals and firms can release pollution. Global pollution creates a "public bad" born by all **--** a negative externality with a wide impact. In many countries environmental protection laws limit the release of local and regional air pollutants. In these situations, in economic terminology, the negative externalities associated with local and regional pollutants have to some degree been internalized. But few controls exist for carbon dioxide (CO_2), the major greenhouse gas. This global air pollutant has no short-term damaging effects at ground level, but atmospheric accumulations of carbon dioxide and other greenhouse gases will have significant effects on global temperature and weather, although there is uncertainty about the probable scale and timing of these effects (See Box 1).

If indeed the effects of climate change are likely to be severe, it is in everyone's interest to lower their emissions for the common good. If no agreement or rules on emissions exist, actions by individual firms, cities or nations will be inadequate. Climate change can thus be viewed as a **public good** issue, requiring collaborative action. Since the problem is global, only a strong international agreement binding nations to act for the common good can prevent serious environmental consequences.

NOTE – terms denoted in **bold face** are defined in the **KEY TERMS AND CONCEPTS** section at the end of the module.

[1] The issue, often called global warming, is more accurately referred to as global climate change. The phenomenon will produce complex effects – with warming in some areas, cooling in others, and generally increased variability in weather patterns.

BOX 1: WHAT IS THE GREENHOUSE EFFECT?

The sun's rays travel through a greenhouse's glass to warm the air inside, but the glass acts as a barrier to the escape of heat. Thus plants that require warm weather can be grown in cold climates. The global greenhouse effect, through which the earth's atmosphere acts like the glass in a greenhouse, was first described by French scientist Jean Baptiste Fourier in 1824.

Clouds, water vapor, and the natural greenhouse gases carbon dioxide (CO_2), methane, nitrous oxide, and ozone allow inbound solar radiation to pass through, but serve as a barrier to outgoing infrared heat. This creates the natural **greenhouse effect**, which makes the planet suitable for life. Without it, the average surface temperature on the planet would average around -18° C (0ºF), instead of approximately 15°C (60º F).

The possibility of an *enhanced* or *human-induced* greenhouse effect was introduced one hundred years ago by the Swedish scientist Svante Arrhenius. He hypothesized that the increased burning of coal would lead to an increased concentration of carbon dioxide in the atmosphere, and would warm the earth. Since Arrhenius' time greenhouse gas emissions have grown dramatically. Carbon dioxide concentrations in the atmosphere have increased by about 35% over pre-industrial levels. In addition to increased burning of fossil fuels such as coal, oil and natural gas, synthetic chemical substances such as chlorofluorocarbons (CFCs) as well as methane and nitrous oxide emissions from agriculture and industry contribute to the greenhouse effect.

Scientists have developed complex computer models that estimate the effect of current and future greenhouse gas emissions on the global climate. While considerable uncertainty remains in these models, virtually all scientists agree that the human-induced greenhouse effect poses a significant threat to the global ecosystem. The global average temperature has increased by about 0.7°C (1.3°F) during the 20th century. The Intergovernmental Panel on Climate Change (IPCC) concluded in 2001 that humans are already having a discernable impact on the global climate: "most of the observed warming over the last 50 years is likely to have been due to the increase in greenhouse gas concentrations." In 2007 they reaffirmed and strengthened this conclusion.

Current emissions trends will lead to a doubling of greenhouse gas concentration over pre-industrial levels by around 2050. The IPCC projects a global average temperature increase of 1 to 6 degrees Centigrade, or 2 to 10 degrees Fahrenheit, by 2100. This would have significant impacts on climate throughout the world.

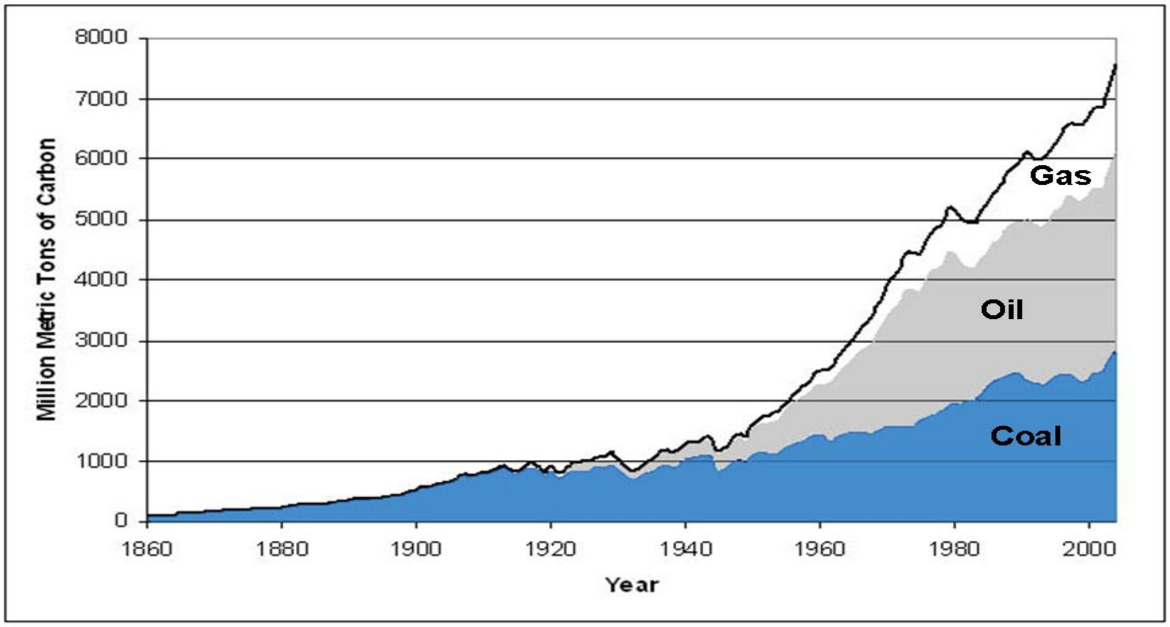
Sources: Cline, 1992; Fankhauser, 1995; IPCC, 2001, 2007.

Because CO_2 and other greenhouse gases continuously accumulate in the atmosphere, stabilizing or "freezing" emissions will not solve the problem. Greenhouse gases persist in the atmosphere for decades or even centuries, continuing to affect the climate of the entire planet long after they are emitted. Greenhouse gases are **stock pollutants**: only major reductions in emissions, to a level consistent with the planet's absorptive capacity (thought to be 50-80% below current emissions levels), will prevent ever-increasing atmospheric accumulations. The development of national and international policies to combat global climate change is a huge challenge, involving many scientific, economic, and social issues.

Trends and Projections for Global Carbon Emissions

Global emissions of carbon dioxide from the combustion of fossil fuels rose dramatically during the 20th century, as illustrated in Figure 1. The use of petroleum is currently responsible for about 42% of global carbon emissions, while coal is the source of another 36%. The United States is presently the world's largest emitter of CO_2 – releasing about one-quarter of the global total while having less than 5% of the world's population. China, the world's second largest source of CO_2 emissions, is likely to surpass the U.S. within the next few years[2].

Figure 1. Global Carbon Dioxide Emissions from Fossil Fuel Combustion, 1860-2004

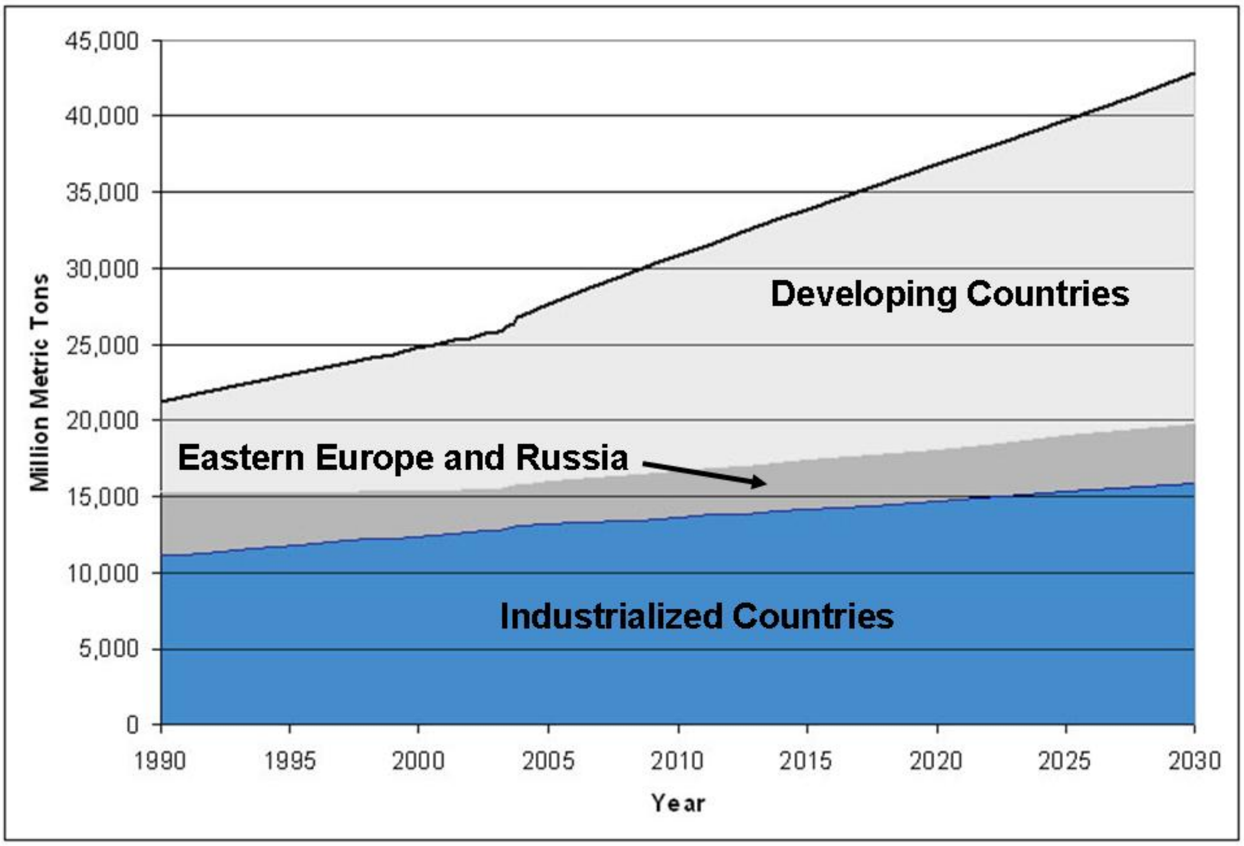


Source: Carbon Dioxide Information Analysis Center (CDIAC), http://cdiac.esd.ornl.gov/trends/emis/em_cont.htm.

[2] In June 2007, a Dutch research group reported that China had surpassed the U.S. in carbon emissions, but as of November 2007 this had not been confirmed. Data on recent carbon emissions are not necessarily precise.

Progress on combating global climate change has been slow, despite three global conferences dealing with the issue – the 1992 United Nations Conference on Environment and Development (UNCED) at Rio de Janeiro, a 1997 meeting in Kyoto, Japan that produced the agreement known as the Kyoto Protocol, and the World Summit on Sustainable Development in 2002 – as well as numerous follow-up negotiating sessions. Current projections show carbon emissions continuing to increase in the future (see Figure 2).

Figure 2. Projected Carbon Dioxide Emissions through 2030, by Region



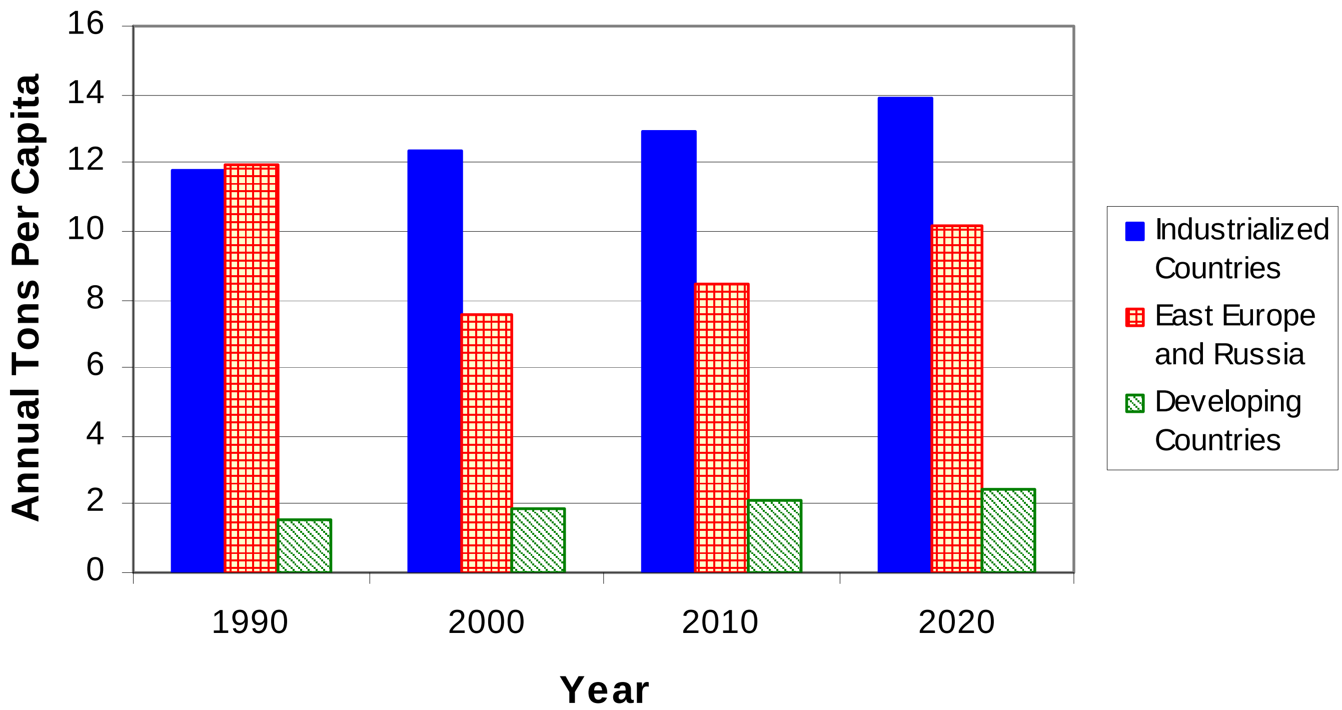
Source: U.S. Department of Energy, 2007. The vertical axis in Figure 2 measures million metric tons of carbon dioxide (The vertical axis in Figure 1 shows million metric tons of carbon; the weight of a given amount of emissions measured in tons of carbon dioxide is about 3.67 times the total weight in carbon)

Figure 2 projects an increase in global carbon dioxide emissions of about 27% between 1990 and 2004. The growth in carbon emissions is expected to continue in the coming decades. According to the U.S. Energy Information Administration, global CO_2 emissions are projected to increase by approximately 59% between 2004 and 2030. These projections are for the U.S.E.I.A.'s "reference case", which assumes business as usual, with no major efforts to reduce carbon emissions. As we will see, strong polices to shift away from carbon-based fuels could alter these projections.

As of 2004, the industrialized countries were responsible for just over half of global carbon emissions. However, as seen in Figure 2 most of the growth in future carbon emissions is expected to come from rapidly expanding developing economies such as China and India. For example, CO_2 emissions in China are projected to grow by 140% between 2004 and 2030.

Although carbon emissions are projected to grow fastest in developing nations, per-capita emissions in 2020 will still be much higher (about six times higher) in the industrialized countries, as shown in Figure 3. The developing nations argue that they should not be required to limit their emissions while the industrial nations continue to emit so much more on a per-capita basis. The global imbalance in per-capita emissions is a critical issue that has yet to be adequately addressed in the policy debate on global climate change.

Figure 3. Per-Capita Emissions of Carbon Dioxide by Region, with projections to 2020



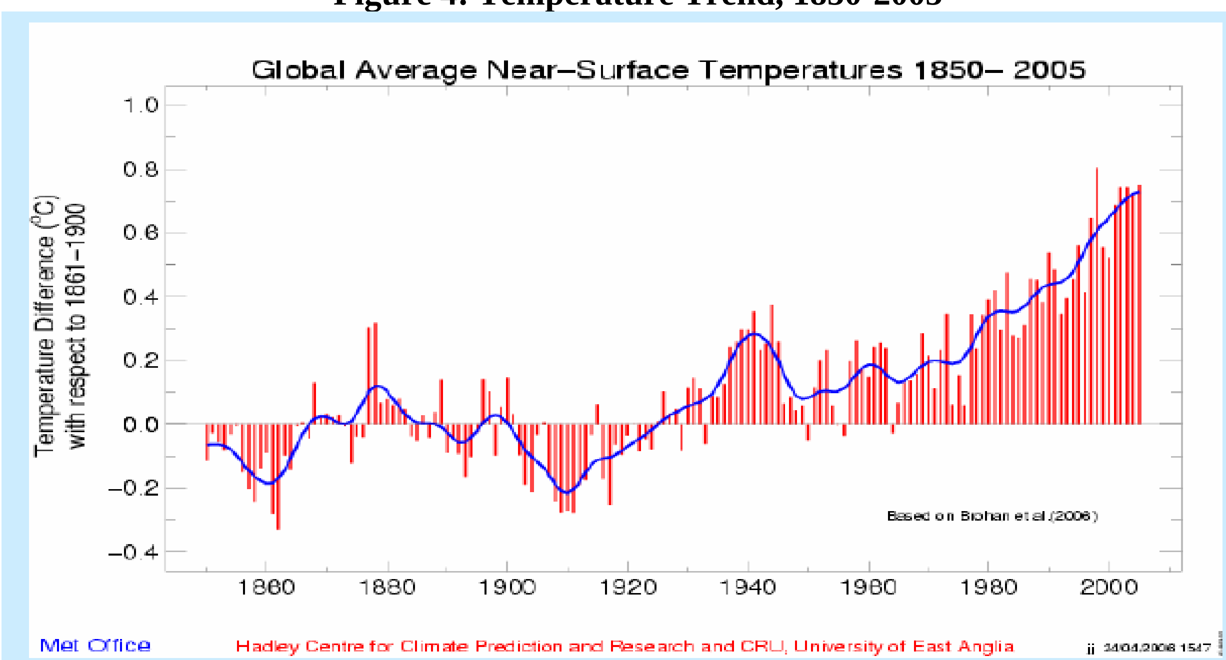
Source: U.S. Department of Energy, 2004.

Trends and Projections for Global Climate

The earth has warmed significantly since reliable weather records have been kept (Figure 4). Over the last 100 years the global average temperature has risen about 0.7°C, or about 1.3°F. Global temperatures since 2000 have been particularly warm – six of the seven warmest years on record have occurred since 2000. There is also evidence that the rate of warming, currently about 0.13°C per decade, is increasing. Not all areas are warming equally. The Arctic and Antarctic regions have been warming at about double the global rate.[3]

Warmer temperatures have produced noticeable effects on ecosystems. In most regions of the world, glaciers are retreating. For example, Glacier National Park in Montana had 150 glaciers when the park was established in 1910. As of 2005 it had only 27 glaciers remaining and by 2030 it is estimated that the park will no longer have any of its namesake glaciers. Climate change is also leading to rising sea levels. Sea-level rise is attributed to the melting of glaciers and ice sheets, and to the fact that water expands when it is heated. The oceans warmed, on average, about 0.1°C between 1961 and 2003. The combination of warmer oceans and melting ice has led to sea levels rising at about two millimeters per year.

Figure 4: Temperature Trend, 1850-2005



Note: This graph compares annual temperatures to the average for the years 1861-1900. Temperatures during 1850-1860 and most of the years from 1900-1925 were below that average. Since 1925 the trend has been a strong increase in temperature compared to the late nineteenth century.

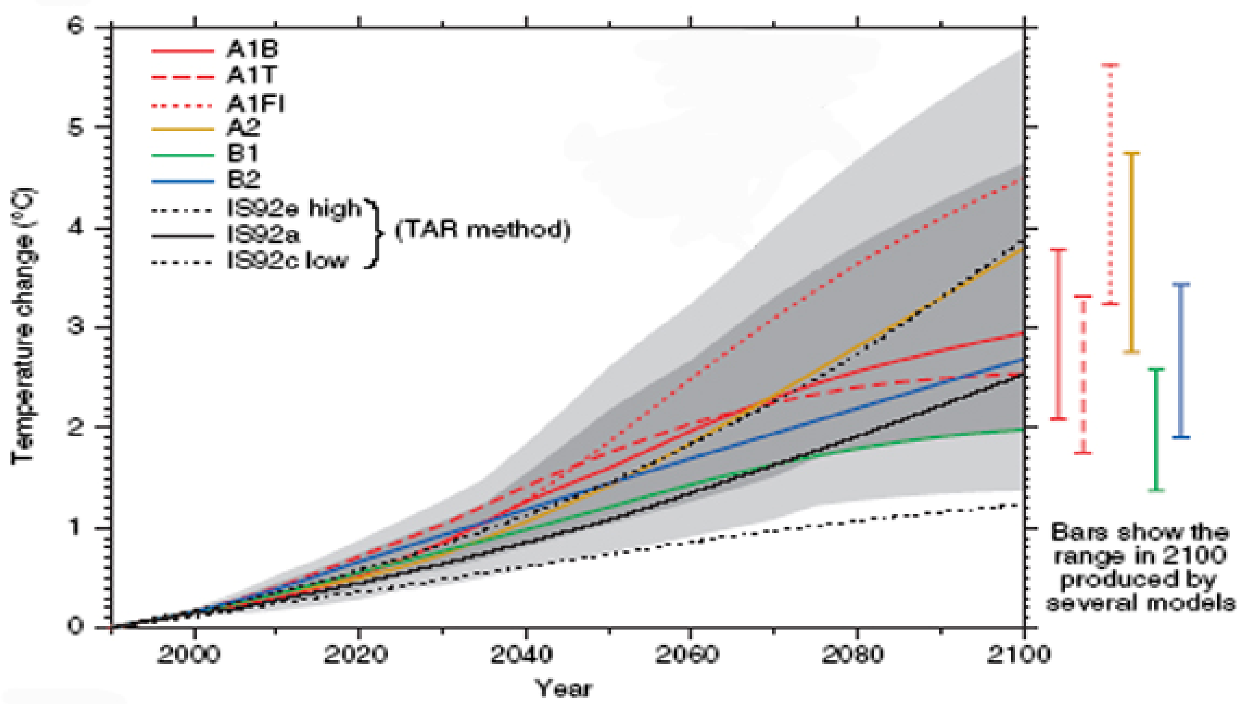
[3] IPCC, 2007.

Although some warming may be a natural trend, the Intergovernmental Panel on Climate Change (IPCC) in 2007 concluded that:

> Most of the observed increase in global average temperatures since the mid-20th century is very likely due to the observed increase in anthropogenic greenhouse gas concentrations. Discernable human influences now extend to other aspects of climate, including ocean warming, continental-average temperatures, temperature extremes, and wind patterns. (IPCC, 2007, Summary for Policymakers, p. 10)

Future projections of climate change depend upon the path of future emissions. Even if all emissions of greenhouse gases were ended today, the world would continue warming over the next few decades because the ultimate environmental effects of emissions are not realized immediately. Based on a wide range of models with different assumptions about future emissions, the IPCC estimates that during the 21st century global average temperatures will rise between 1.1°C (2°F) and 6.4°C (11°F), with the range more likely to be between 1.8°C (3°F) and 4°C (7°F). The range of possible temperature increases is shown in Figure 5.

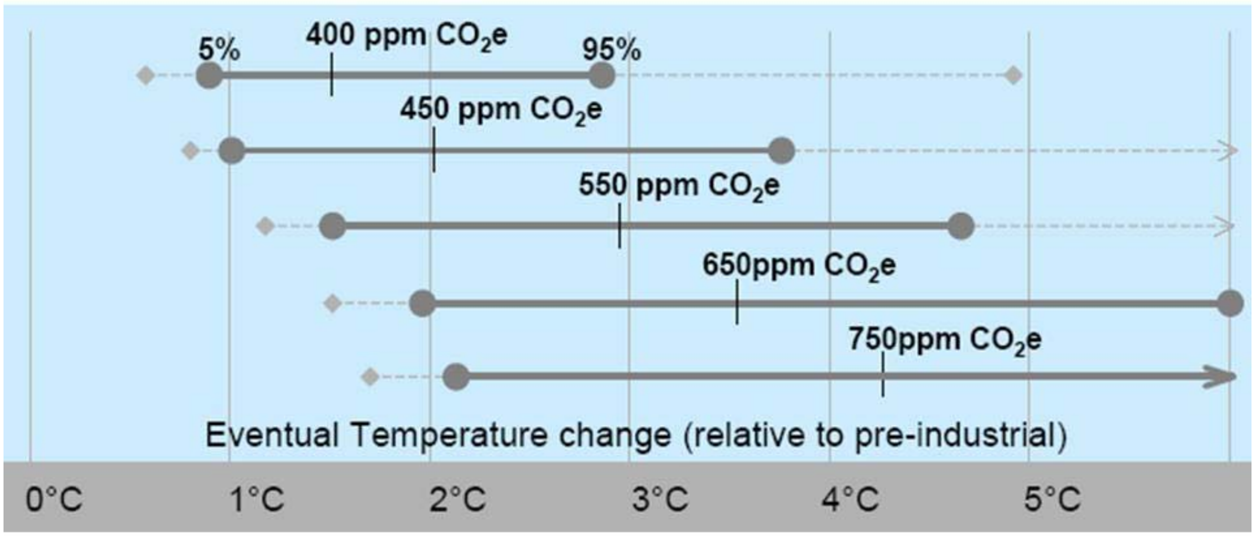
Figure 5. Global Temperature Trends Projected to 2100



Source: IPCC, 2001. IPCC 2007 projections are substantially similar, but have a slightly greater range, from 1.1ºC to 6.4ºC.
Note: The IPCC used a variety of assumptions about economic growth and energy policies to construct the specific numbered projections. The gray areas represent "envelopes" showing the possible range of estimates for the various projections.

The magnitude of actual warming and other effects will depend upon the level at which atmospheric concentrations of CO_2 and other greenhouse gases are ultimately stabilized. The current atmospheric CO_2 concentration is around 380 ppm. When we consider the contribution of other greenhouse gases, the overall effect is equivalent to a concentration of 430 ppm of CO_2, referred to as CO_2e. Figure 6 below relates the stabilization level of greenhouse gases, measured in CO_2e, to the resulting rise in global average temperatures, incorporating the degree of uncertainty. The solid bar at each level of CO_2e represents a range of temperature outcomes that is likely to occur with a 90% probability. The dashed line extending beyond this interval at either end represents the full range of predicted results from the major existing climate models. The vertical line around the middle of each bar represents the mid-point of the different predictions.

Figure 6. The Relationship between the Level of Greenhouse Gas Stabilization and Eventual Temperature Change



Source: Stern, 2007.

This projection suggests that stabilizing greenhouse gas concentrations at 450 ppm CO_2e would be 90% likely to eventually result in a temperature increase between 1.0 and 3.8°C, with a small probability that the rise could be significantly more than this. With current greenhouse gas concentrations in the atmosphere at 430 ppm CO_2e, stabilization at 450 ppm would be extremely challenging. As we will see later, even stabilization at 550 ppm CO_2e would require dramatic policy action.

BOX 2: PACIFIC ISLANDS DISAPPEAR AS OCEANS RISE

Veu Lesa, a 73-year old villager in the Pacific island nation of Tuvalu, does not need scientific reports to tell him the sea is rising. The evidence is all around him. The beaches of his childhood are vanishing. The crops that used to feed his family have been poisoned by salt water. In April 2007 he was evacuated when a high tide flooded his home, showering it with rocks and debris.

For Tuvalu, a string of nine picturesque atolls and coral islands, global warming is not an abstract danger; it is a daily reality. The tiny South Pacific nation, only 4m above sea level at its highest point, may not exist in a few decades. Its people are already in flight; more than 4000 have moved to New Zealand, and many of the remaining 10,500 are planning to join the exodus. Neighboring islands have already disappeared as a result of rising sea level. So far the seas have completely engulfed only uninhabited, relatively small islands, but the crisis is growing all along the shores of the world's atolls.

Almost the entire coastline of the 29 atolls of the Marshall Islands is eroding. Second World War graves on its main Majuro atoll are being washed away, roads and sub-soils have been swept into the sea and the airport has been flooded several times despite being supposedly protected by a high sea wall.

The people of Tuvalu are finding it difficult to grow their crops because the rising seas are poisoning the soil with salt. In both Kiribati and the Marshall Islands families are desperately trying to keep the waves at bay by dumping trucks, cars and other old machinery in the sea and surrounding them with rocks. The story is much the same in the Maldives. The Indian Ocean is sweeping away the beaches of one-third of its 200 inhabited islands. "Sea-level rise is not a fashionable scientific hypothesis," says President Gayoom. "It is a fact."

The seas are rising partly because global warming is melting glaciers and nibbling away at the polar ice caps, but mainly because the oceans expand as their water gets warmer. Scientists have estimated that these processes will raise sea levels by a foot or more over the next century, quite enough to destroy several island nations.

The higher the seas rise, the more often storms will sweep the waves across the narrow atolls, carrying away the land - and storms are expected to increase as the world warms up. Moreover, many islands will become uninhabitable long before they physically disappear, as salt from the sea contaminates the underground freshwater supplies on which they depend.

Adapted from: Lean, Geoffrey, “They’re Going Under: Two Islands Have Disappeared Beneath the Pacific Ocean - Sunk by Global Warming.” *The Independent*, June 13, 1999, p. 15; “A Vanishing Pattern of Islands,” *The Canberra Times*, July 21, 2007.

2. ECONOMIC ANALYSIS OF CLIMATE CHANGE

Scientists have modeled the effects of a projected doubling of accumulated carbon dioxide in the earth's atmosphere. Some of the predicted effects are:

- Loss of land area, including beaches and wetlands, to sea-level rise
- Loss of species and forest area, including coral reefs and wetlands
- Disruption of water supplies to cities and agriculture
- Health damage and deaths from heat waves and spread of tropical diseases
- Increased costs of air conditioning
- Loss of agricultural output due to drought

Some beneficial outcomes might include:

- Increased agricultural production in cold climates
- Lower heating costs
- Less deaths from exposure to cold

In addition to these effects, there are some other, less predictable but possibly more damaging effects, including:

- Disruption of weather patterns, with increased frequency of hurricanes and other extreme weather events
- A possible rapid collapse of the Greenland and West Antarctic Ice Sheets, which would raise sea levels by 12 meters or more, drowning major coastal cities
- Sudden major climate changes, such as a shift in the Atlantic Gulf Stream, which could change the climate of Europe to that of Alaska
- Positive **feedback effects,**[4] such as an increased release of carbon dioxide from warming arctic tundra, which would speed up global warming

The IPCC projects that with increasing emissions and higher temperatures, negative effects will intensify and positive effects diminish (Table 1). As shown in Figure 5, there is considerable uncertainty about the expected global warming in the next century. We need to keep such uncertainties in mind as we evaluate economic analyses of global climate change.

[4] A feedback effect occurs when an original change in a system causes further changes that either reinforce the original change (positive feedback) or counteract it (negative feedback).

Table 1. Possible Effects of Climate Change

Type of Impact	Eventual Temperature Rise Relative to Pre-Industrial Temperatures				
	1°C	**2°C**	**3°C**	**4°C**	**5°C**
Freshwater Supplies	Small glaciers in the Andes disappear, threatening water supplies for 50 million people	Potential water supply decrease of 20-30% in some regions (Southern Africa and Mediterranean)	Serious droughts in Southern Europe every 10 years 1-4 billion more people suffer water shortages	Potential water supply decrease of 30-50% in Southern Africa and Mediterranean	Large glaciers in Himalayas possibly disappear, affecting ¼ of China's population
Food and Agriculture	Modest increase in yields in temperature regions	Declines in crop yields in tropical regions (5-10% in Africa)	150-550 million more people at risk of hunger Yields likely to peak at higher latitudes	Yields decline by 15-35% in Africa Some entire regions out of agricultural production	Increase in ocean acidity possibly reduces fish stocks
Human Health	At least 300,000 die each year from climate-related diseases Reduction in winter mortality in high latitudes	40-60 million more exposed to malaria in Africa	1-3 million more potentially people die annually from malnutrition	Up to 80 million more people exposed to malaria in Africa	Further disease increase and substantial burdens on health care services
Coastal Areas	Increased damage from coastal flooding	Up to 10 million more people exposed to coastal flooding	Up to 170 million more people exposed to coastal flooding	Up to 300 million more people exposed to coastal flooding	Sea level rise threatens major cities such as New York, Tokyo, and London
Ecosystems	At least 10% of land species facing extinction Increased wildfire risk	15-40% of species potentially face extinction	20-50% of species potentially face extinction Possible onset of collapse of Amazon forest	Loss of half of Arctic tundra Widespread loss of coral reefs	Significant extinctions across the globe

Sources: Stern, 2007; IPCC, 2007.

Given these uncertainties, some economists have attempted to place the analysis of global climate change in the context of **cost-benefit analysis**. Others have criticized this approach as an attempt to put a monetary valuation on issues with social, political, and ecological implications that go far beyond dollar value. We will first examine economists' efforts to capture the impacts of global climate change through cost-benefit analysis, and then return to the debate over how to implement greenhouse gas reduction polices.

Cost-Benefit Studies of Global Climate Change

Without policy intervention, carbon emissions can be expected to continue to rise approximately as projected in Figure 2. Aggressive and immediate policy action would be required first to stabilize and then to reduce total CO_2 emissions in the coming decades. In performing a cost-benefit analysis, we must weigh the consequences of the projected increase in carbon emissions versus the costs of current policy actions to stabilize or even reduce CO_2 emissions. Strong policy action to prevent climate change will bring benefits equal to the value of damages that are avoided[5]. These benefits must be compared to the costs of taking action. Various economic studies have attempted to estimate these benefits and costs. The results of one such study for the U.S. economy are shown in Table 2.

The study is based on an estimated doubling of CO_2 over pre-industrial levels. When the monetized costs are added up, the total annual U.S. damages are estimated at approximately $60 billion (1990 dollars). This is about 1% of U.S. GNP. Although different economic studies come up with different estimates, most of them are in the range of 1-2% GNP. Cost estimates for larger temperature change over the longer term rise to around 5% of GNP (the far-right column of Table 2).

Note, however, that there are also some "Xs" and "Ys" in the totals – unknown quantities that cannot easily be measured. The damages from species extinctions, for example, are difficult to estimate in dollar terms: the estimates used here show a cost of at least $4 billion in the short term and $16 billion in the long term, with additional unknown costs in both the short and long term.

[5] These benefits of preventing damage can also be referred to as **avoided costs**.

Table 2. Estimates of Annual Damages to the U.S. Economy from Global Climate Change (billions of 1990 dollars)

Type of Damage	Short-term warming based on doubling CO_2 levels (+2.5 degrees C)	Very long-term warming (+10 degrees C)
Agriculture	17.5	95.0
Forest loss	3.3	7.0
Species extinctions	4.0 + X_1	16.0 + Y_1
Sea-level rise		35.0
Building dikes, levees	1.2	
Wetlands loss	4.1	
Drylands loss	1.7	
Electricity requirements	11.2	64.1
Non-electric heating	-1.3	-4.0
Human amenity	X_2	Y_2
Human life loss	5.8	33.0
Human morbidity	X_3	Y_3
Migration	0.5	2.8
Increased hurricanes	0.8	6.4
Construction costs	+/- X_4	+/- Y_4
Loss of leisure activities	1.7	4.0
Water supply costs	7.0	56.0
Urban infrastructure costs	0.1	0.6
Air pollution		
Tropospheric ozone	3.5	19.8
Other air pollution	X_5	Y_5
Total	**61.1 + X_1 + X_2 + X_3 +/- X_4 + X_5**	**335.7 + Y_1 + Y_2 + Y_3 +/- Y_4 + Y_5**

Source: Cline, 1992.

In addition to the Xs and Ys, other monetized estimates could also be challenged on the grounds that they fail to capture the full value of potential losses. For example, oceanfront land is more than just real estate. Beaches and coastal wetlands have great social, cultural, and ecological value. The market value of these lands fails to capture the full scope of the damage society will suffer if they are lost. Valuing human health and life is very controversial – this study follows a common cost-benefit practice of assigning a value of about $6 million to a life, based on studies of the amounts people are willing to pay to avoid life-threatening risks, or are willing to accept (for example in extra salary for dangerous jobs) to undertake such risks.[6]

[6] For more discussion on the controversy over valuation of life, see Harris, 2006, Chapter 6.

In addition, these estimates omit the possibility of the much more catastrophic consequences that *could* result if weather disruption is much worse than anticipated. Hurricane Katrina in August 2005, for example, caused over $80 billion in damage, in addition to loss of over 1800 lives. If climate changes cause severe hurricanes to become much more frequent, the estimate given in Table 2 of less than one billion annual losses could be much too low. Another of the unknown values – human morbidity, or losses from disease – could well be enormous if tropical diseases extend their range significantly due to warmer weather conditions.

Clearly, these damage estimates are not precise, and are open to many criticisms. But suppose we decide to accept them – at least as a rough estimate. We must then weigh the estimated benefits of policies to prevent climate change against the costs of such policies. To estimate these costs, economists use models that show how economic output is produced from factor inputs such as labor, capital, and resources.

To lower carbon emissions, we must cut back the use of fossil fuels, substituting other energy sources that may be more expensive. In general, economic models predict that this substitution would reduce GNP growth. One major study showed GNP losses ranging from 1 to 3 percent of GNP for most countries, with higher potential long-term losses for coal-dependent developing nations such as China[7].

How can we weight the costs of taking action on global warming against the benefits in terms of avoided damage? Much depends on our evaluation of future costs and benefits. The costs of taking action must be born today or in the near future. Many of the benefits of taking action (the avoided costs of damages) are further in the future. How can we decide today how to balance these future costs and benefits?

Economists evaluate future costs and benefits by the use of a **discount rate**. Costs and benefits in the future are considered to have a lower dollar value than the same costs and benefits today, with the size of the difference depending on the choice of discount rate (see Box 3). The problems and implicit value judgments associated with discounting add to the issues of ethics and judgment that we have already noted in valuing costs and benefits. This suggests that we should consider some alternative approaches – including techniques that incorporate ecological as well as economic costs and benefits.

[7] Manne and Richels, 1992.

BOX 3: DISCOUNTING

Economists calculate the present value of a cost or benefit of $X that occurs in years in the future using the equation:

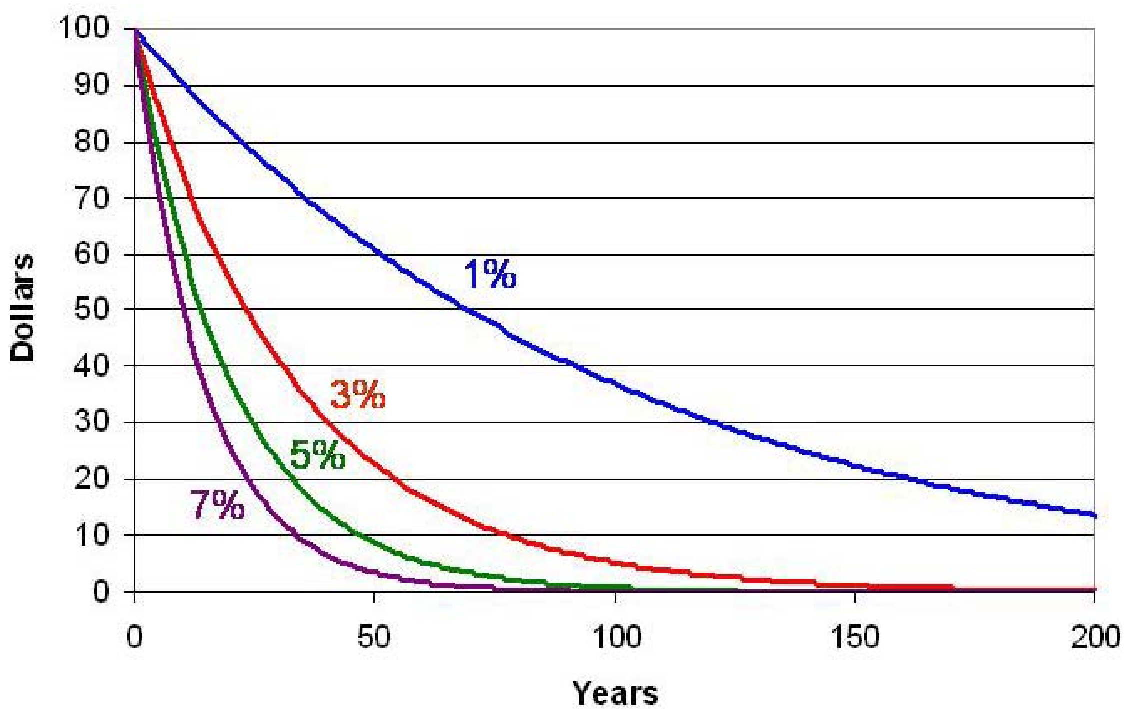
*Present Value ($X) = $X / (1 + r)*^{n}

where *r* is the discount rate. So, for example, if we want to determine the present value of a benefit of $50,000 received 25 years from now with a discount rate of 5%, it would be:

$\$50,000 / (1 + 0.05)^{25} = \$14,765$

The choice of a discount rate becomes more important the further out in time one goes. Figure 7 below shows the present value of $100 for different time periods into the future using several discount rates that have been used in climate change cost-benefit analyses. We see that when a discount rate of 5% or 7% is used, costs or benefits 100 years into the future are negligible – worth only $0.76 and $0.12 respectively. Even with a discount rate of 3%, the value of $100 is only $5.20 after 100 years. But when the discount rate is 1%, impacts 100 years into the future are still significant – worth about $37.

Figure 7. The Effects of Different Discount Rates



3. ANALYZING LONG-TERM EFFECTS OF CLIMATE CHANGE

Economic studies dealing with benefit-cost analysis of climate change have come to very different conclusions about policy. According to a study by William Nordhaus and Joseph Boyer[8], the "optimal" policy strategy would be only a small reduction in greenhouse gas emissions below current projections. This would require few changes in the carbon-based energy path typical of current economic development.

Until recently, most economic studies of climate change reached conclusions similar to those of the Nordhaus and Boyer study, although a few recommended more drastic action. The debate on climate change economics altered in October 2006, when Nicholas Stern, a former chief economist for the World Bank, released a 700-page report, sponsored by the British government, titled "The Stern Review on the Economics of Climate Change".[9] Publication of the Stern Review generated significant media attention and has intensified the debate about climate change in policy and academic circles. While most previous economic analyses of climate change suggested relatively modest policy responses, the Stern Review strongly recommends immediate and substantial policy action:

> The scientific evidence is now overwhelming: climate change is a serious global threat, and it demands an urgent global response. This Review has assessed a wide range of evidence on the impacts of climate change and on the economic costs, and has used a number of different techniques to assess costs and risks. From all these perspectives, the evidence gathered by the Review leads to a simple conclusion: the benefits of strong and early action far outweigh the economic costs of not acting.
>
> Using the results from formal economic models, the Review estimates that if we don't act, the overall costs and risks of climate change will be equivalent to losing at least 5% of global GDP each year, now and forever. If a wider range of risks and impacts is taken into account, the estimates of damage could rise to 20% of GDP or more. In contrast, the costs of action – reducing greenhouse gas emissions to avoid the worst impacts of climate change – can be limited to around 1% of global GDP each year.[10]

What explains the dramatic difference between these two approaches to economic analysis of climate change? One major difference is the choice of the discount rate to use in valuing future costs and benefits.

The present value (PV) of a long-term stream of benefits or costs depends on the discount rate. A high discount rate will lead to a low present valuation for benefits that are mainly in the longer-term, and a high present valuation for short-term costs. On the

---

[8] Nordhaus and Boyer, 2000.

[9] Now available in book form (Stern, 2007). The full Stern Review is available online at http://www.hm-treasury.gov.uk/independent_reviews/stern_review_economics_climate_change/sternreview_index.cfm, including both a 4-page and 27-page summaries.

[10] Stern Review, Short Executive Summary, page vi.

other hand, a low discount rate will lead to a higher present valuation for longer-term benefits. The estimated net present value of an aggressive abatement policy will thus be much higher if we choose a low discount rate (see Box 3).

While both the Stern and Nordhaus/Boyer studies used standard economic methodology, Stern's approach gives greater weight to long-term ecological effects. The Stern Review uses a low discount rate of 1.4% to balance present and future costs. Thus even though costs of aggressive action appear higher than benefits for several decades, the high potential long-term damages sway the balance in favor of aggressive action today. These are significant both for their monetary and non-monetary impacts. In the long term, damage done to the environment by global climate change will have significant negative effects on the economy too. But the use of a standard discount rate of in the 5-10% range has the effect of reducing the present value of significant long-term future damages to relative insignificance.

Another difference between the two studies concerns their treatment of uncertainty. Stern's approach gives a heavier weighting to uncertain, but potentially catastrophic impacts. This reflects the application of a **precautionary principle**: if a particular outcome could be catastrophic, even though it seems unlikely, strong measures should be taken to avoid it. This principle, which has become more widely used in environmental risk management, is especially important for global climate change because of the many unknown but potentially disastrous outcomes possibly associated with continued greenhouse gas accumulation (see Box 4).

A third area of difference concerns the assessment of the economic costs of action to mitigate climate change. Measures taken to prevent global climate change will have economic effects on GDP, consumption, and employment, which explains the reluctance of governments to take drastic measures to reduce significantly emissions of CO_2. But these effects will not all be negative.

The Stern Review conducted a comprehensive review of economic models of the costs of carbon reduction. These cost estimates are very much dependent on the modeling assumptions that are used. The predicted costs of stabilizing atmospheric accumulations of CO_2 at 450 parts per million range from a 3.4 percent decrease to a 3.9.percent *increase* in GDP. The outcomes depend on a range of assumptions including:

- The efficiency or inefficiency of economic responses to energy price signals
- The availability of non-carbon "backstop" energy technologies
- Whether or not nations can trade least-cost options for carbon reduction
- Whether or not revenues from taxes on carbon-based fuels are used to lower other taxes

- Whether or not external benefits of carbon reduction, including reduction in ground-level air pollution, are taken into account[11]

Depending on which assumptions are made, policies for emissions reduction could range from a minimalist approach of slightly reducing the rate of increase in emissions to a dramatic CO_2 emissions reduction of 50% - 80%.

Climate Change and Inequality

The effects of climate change will fall most heavily upon the poor of the world. For example, analysis by the IPCC found that a 2.5°C temperature increase would result in a loss of 1.0 to 1.5% of GDP in developed countries but a 2–9% loss of GDP in developing countries.[12] While the richer countries may have the economic resources to adapt to many of the effects of climate change, poorer countries will be unable to implement preventative measures, especially those that rely on the newest technologies.

The way in which economists incorporate inequality into their analyses can have a significant impact on their policy recommendations. If all costs are evaluated in dollars, a loss of, say, 10% of GDP in a poor country is likely to be much less than a loss of 3% of GDP in a rich country. Thus the damages from climate change in poor countries, which may be large as a percentage of GDP, would receive relatively little weight because the losses are relatively small in dollar terms. The Stern Review asserts that the disproportionate effects of climate change on the world's poorest people should increase the estimated costs of climate change. Stern estimates that, without the effects of inequity, the costs of a BAU scenario will be 11-14% of global GDP. Weighing the impacts on the world's poor more heavily gives a cost estimate of 20% of global GDP.

Thus we see that assumptions about the proper way to evaluate social and environmental costs and benefits can make a big difference to policy recommendations. Most economists who have analyzed the problem agree that action is necessary (see Box 5) but there is a wide scope of opinion on how drastic this action should be, and how soon it should occur.

[11] Stern Review, Chapter 10: "Macroeconomic Models of Costs".
[12] IPCC, Second Assessment Report, 1996.

BOX 4: CLIMATE TIPPING POINTS AND SURPRISES

Much of the uncertainty in projections of climate change relates to the issue of feedback loops. A feedback loop occurs when an initial change, such as warmer temperatures, produces changes in physical processes which then amplify or lessen the initial effect (a response that increases the original effect is called a positive feedback loop; a response that reduces it is a negative feedback loop). An example of a positive feedback loop would be when warming leads to increased melting of arctic tundra, releasing carbon dioxide and methane, which add to atmospheric greenhouse gas accumulations and speed up the warming process.

As a result of various feedback loops associated with climate change, recent evidence suggests that warming is occurring faster than most scientists predicted just 5 or 10 years ago. This is leading to increasing concern over the potential for "runaway" feedback loops which could result in dramatic changes in a short period. Some scientists suggest that we may be near certain climate tipping points which, once exceeded, pose the potential for catastrophic effects.

Perhaps the most disturbing possibility would be the rapid collapse of the Greenland and West Antarctic Ice Sheets. While the IPCC forecasts a sea level of rise of 0.2 to 0.6 meters by 2100, the melting of these two ice sheets would raise sea levels by 12 meters or more. Such a scenario is still controversial, and considered unlikely to occur in the 21st century, but new research suggests that changes can occur much faster than originally expected. Scientists used to think that ice melting on the top of an ice sheet would take 10,000 years to penetrate to the bottom of the ice sheet, where it can lubricate the ice sheet and cause it to slide more rapidly towards the ocean. But in 2006 scientists observed the rapid draining of several lakes of melted ice on the Greenland Ice Sheet which resulted in changes in the movement of the surface ice in a matter of hours – a phenomenon that could lead to much more rapid collapse of the whole ice sheet.

Rapid climate change has occurred before. During the last ice age, sea levels rose at a rate of about 5 meters per century. Ice core data indicate that about 11,000 years ago temperatures in the Arctic rose 16°F or more within a decade, perhaps within a single year. Some of the feedback loops that produce such changes are only starting to be understood.

Source: "Melting Ice Turns up the Heat," Fred Pearce, *Sydney Morning Herald*, November 18, 2006.

4. POLICY RESPONSES TO CLIMATE CHANGE

Two types of measures can be used to address climate change; **preventive measures** tend to lower or mitigate the greenhouse effect, and **adaptive measures** deal with the consequences of the greenhouse effect and trying to minimize their impact.

Preventive measures include:

- Reducing emissions of greenhouse gases, either by reducing the level of emissions-related economic activities or by shifting to more energy-efficient technologies that would allow the same level of economic activity at a lower level of CO_2 emissions.

- Enhancing **carbon sinks**.[13] Forests recycle CO_2 into oxygen; preserving forested areas and expanding reforestation have a significant effect on net CO_2 emissions.

Adaptive measures include:

- Construction of dikes and seawalls to protection against rising sea level and extreme weather events such as floods and hurricanes.

- Shifting cultivation patterns in agriculture to adapt to changed weather conditions in different areas, and relocating people away from low-lying coastal areas.

- Creating institutions that can mobilize the needed human, material, and financial resources to respond to climate-related disasters.

For any particular preventive or adaptive measure, an economic approach suggests that we should apply **cost-effectiveness analysis** in considering which policies to adopt. The use of cost-effectiveness analysis avoids many of the controversies associated with cost-benefit analysis. While cost-benefit analysis attempts to offer a basis for deciding whether or not a policy should be implemented, cost-effectiveness analysis accepts a goal as given by society, and uses economic techniques to evaluate the most efficient way to reach that goal.

In general, economists favor approaches that work through market mechanisms to achieve their goals (see Box 5). Market-oriented approaches are considered to be cost-effective; rather than attempting to control market actors directly, they shift incentives so that individuals and firms will change their behavior to take account of external costs and

[13] Carbon sinks are areas where excess carbon may be stored. Natural sinks include the oceans and forests. Human intervention can either reduce or expand these sinks through forest management and agricultural practices.

benefits. Examples of market-based policy tools include **pollution taxes** and **transferable, or tradable, permits**. Both of these are potentially useful tools for greenhouse gas reduction. Other relevant economic policies include measures to create incentives for the adoption of renewable energy sources and energy-efficient technology.

BOX 5: ECONOMISTS' STATEMENT ON CLIMATE CHANGE

1. The review conducted by a distinguished international panel of scientists under the auspices of the Intergovernmental Panel on Climate Change has determined that "the balance of evidence suggests a discernible human influence on global climate." As economists, we believe that global climate change carries with it significant environmental, economic, social, and geopolitical risks, and that preventive steps are justified.

2. Economic studies have found that there are many potential policies to reduce greenhouse-gas emissions for which the total benefits outweigh the total costs. For the United States in particular, sound economic analysis shows that there are policy options that would slow climate change without harming American living standards, and these measures may in fact improve U.S. productivity in the longer run.

3. The most efficient approach to slowing climate change is through market-based policies. In order for the world to achieve its climatic objectives at minimum cost, a cooperative approach among nations is required -- such as an international emissions trading agreement. The United States and other nations can most efficiently implement their climate policies through market mechanisms, such as carbon taxes or the auction of emissions permits. The revenues generated from such policies can effectively be used to reduce the deficit or to lower existing taxes.

This statement has been endorsed by over 2,500 economists, including eight Nobel laureates.

Source: Redefining Progress, http://www.rprogress.org/publications/2001/econstatement.htm

Policy Tools: Carbon Taxes

The release of greenhouse gases in the atmosphere is a clear example of a negative externality that imposes significant costs on a global scale. In the language of economic theory, the current market for carbon-based fuels such as coal, oil, and natural gas takes into account only private costs and benefits, which leads to a market equilibrium that does not correspond to the social optimum. From a social perspective the market price for fossil fuels is too low and the quantity consumed too high.

A standard economic remedy for internalizing external costs is a per-unit tax on the pollutant. In this case, what is called for is a **carbon tax**, levied exclusively on carbon-based fossil fuels in proportion to the amount of carbon associated with their production and use. Such a tax will raise the price of carbon-based energy sources, and so give consumers incentives to conserve energy overall, as well as shifting their demand to alternative, non-carbon sources of energy (which are not taxed). Demand may also shift from carbon-based fuels with a higher proportion of carbon, such as coal, to those with relatively lower carbon content, such as natural gas.

> "Carbon taxes would appear to consumers as energy price increases. But since taxes would be levied on primary energy, which represents only one part of the cost of delivered energy (such as gasoline or electricity) and more important, since one fuel can in many cases be substituted for another, overall price increases may not be jolting. Consumers can respond to new prices by reducing energy use and buying fewer carbon-intensive products (those that require great amounts of carbon-based fuels to produce). In addition, some of these savings could be used to buy other less carbon-intensive goods and services.
>
> "Clearly, a carbon tax creates an incentive for producers and consumers to avoid paying the tax by reducing their use of carbon-intensive fuels. Contrary to other taxed items and activities, this avoidance has social benefits – reduced energy use and reduced CO_2 emissions. Thus, declining tax revenues over time indicate policy success – just the opposite of what happens when tax policy seeks to maintain steady or increasing revenues."[14]

Table 3 shows the impact that different levels of a carbon tax would have on the prices of coal, oil, and natural gas. A $10/ton carbon tax, for example, raises the price of a barrel of oil by about a dollar, equivalent to only about two cents per gallon.[15] A $100/ton carbon tax would equate to an increase in gasoline prices of about 24 cents per gallon.

---

[14] Dower and Zimmerman, 1992.
[15] There are 42 gallons in a barrel of oil.

Table 3. Alternative Carbon Taxes on Fossil Fuels

	Coal	Oil	Natural Gas
Tons of carbon per unit of fuel	0.574/ton	0.102/barrel	0.015/Mcf (thousand cubic feet)
Average price (2007)	$25.16/ton	$88.79/barrel	$5.90/Mcf
Carbon tax amount per unit of fuel:			
$10/ton of carbon	$5.74/ton	$1.02/barrel	$0.15/Mcf
$100/ton of carbon	$57.42/ton	$10.15/barrel	$1.49/Mcf
$200/ton of carbon	$114.85/ton	$20.31/barrel	$2.98/Mcf
Carbon tax as a percent of fuel price:			
$10/ton of carbon	23%	1%	3%
$100/ton of carbon	228%	11%	25%
$200/ton of carbon	456%	23%	51%

Source: Carbon emissions calculated from carbon coefficients and thermal conversion factors available from the U.S. Department of Energy. Oil price is mid-November 2007 world average. Natural gas price is August 2007 average U.S. wellhead price. Coal price 2006 U.S. average. All price data from the U.S. Energy Information Administration.

Will these taxes affect people's driving or home heating habits very much? This depends on the **elasticity of demand** for these fuels. Elasticity of demand is defined as:

$$Elasticity\ of\ demand = \frac{Percent\ change\ in\ demand}{Percent\ change\ in\ price}$$

Economists have measured the elasticity of demand for different fossil fuels, particularly gasoline. One study[16] surveyed all the available research on the elasticity of demand for motor fuels and found that within the short-term (about one year or less) elasticity estimates averaged -0.25.[17] This means that a 10% increase in the price of gasoline would be expected to decrease gasoline demand in the short term by about 2.5%.

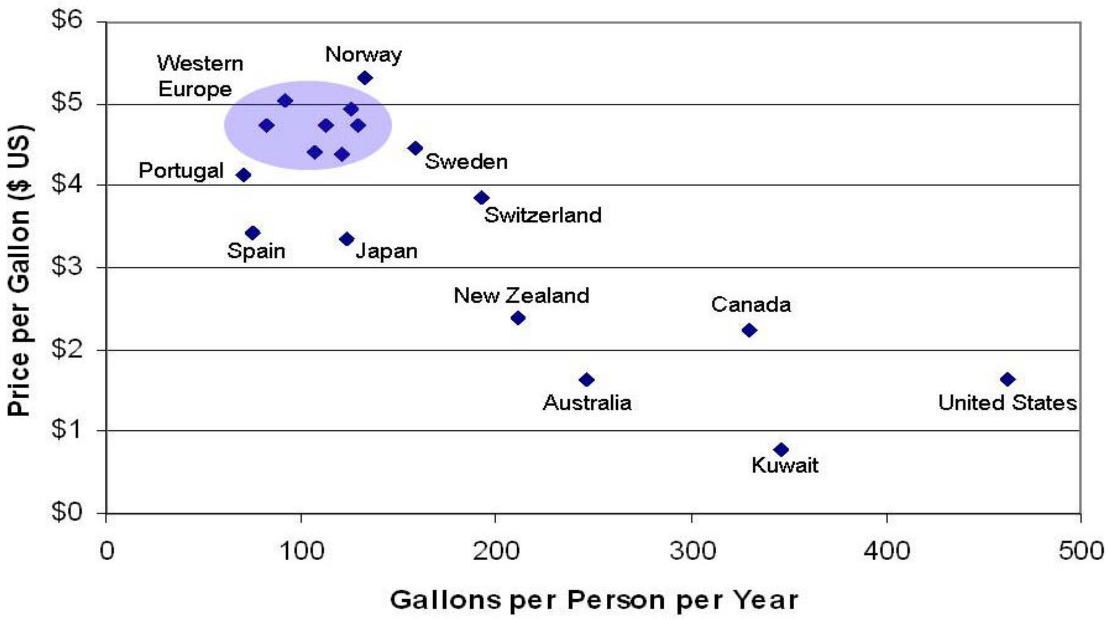
In the long-term (about 5 years or so) people are more responsive to gasoline price increases as they have time to purchase different vehicles and adjust their driving habits. The average long-term elasticity of demand for motor fuels was -0.64. According to Table 3, a $200 carbon tax would increase the price of gasoline by 48 cents per gallon. Assuming a retail price of $3 per gallon, this would translate to a 16% price increase. A

[16] Phil Goodwin, Joyce Dargay, and Mark Hanly. "Elasticities of Road Traffic and Fuel Consumption with respect to Price and Income: A Review," *Transport Reviews*, 24(3):275-292, May 2004.
[17] A 2006 paper by Jonathan E. Hughes, Christopher R. Knittel, and Daniel Sperling ("Evidence of a Shift in the Short-Run Price Elasticity of Gasoline Demand," NBER Working Paper No. W12530, September 2006) indicates that the short-run price elasticity of demand for gasoline may have significantly declined in recent years. They estimate an elasticity of demand for 2001-2006 of -0.03 to -0.08, compared with their estimate of an elasticity of demand for 1975-1980 of -0.21 to -0.34.

long-term elasticity of -0.64 suggests that once people have time to fully adjust to this price change, we would expect the demand for gasoline to decline by about 10%.

Figure 8 shows a cross-county relationship between gasoline prices and per capita consumption. (Since the cost of producing a gallon of gasoline varies little across countries, variations in the price of gallon in different countries is almost solely a function of differences in taxes.) Notice that this relationship is similar to that of a demand curve: higher prices are associated with lower consumption, lower prices with higher consumption. The relationship shown here, however, is not exactly the same as a demand curve; since we are looking at data from different countries, the assumption of "other things equal", which is needed to construct a demand curve, does not hold. Differences in demand may, for example, be partly a function of differences in income levels rather than prices. Also, people in the United States may drive more partly because travel distances (especially in the Western U.S.) are greater than in many European countries. But there does seem to be a clear price/consumption relationship. The data shown here suggest that it would take a fairly big price hike – in the range of $0.50- $1.00 per gallon or more – to affect fuel use substantially.

Figure 8. Gasoline Price versus Use in Industrial Countries, 2003



Note: Shaded area represents price/consumption range typical of West European countries.
Source: U.S. Department of Energy, 2004. Adapted from Roodman, 1997, with updated data.

Would such a tax ever be politically feasible? Especially in the United States, high taxes on gasoline and other fuels would face much opposition, especially if people saw it as infringing on their freedom to drive. As Figure 8 shows, the U.S. has by far the highest consumption per person and the lowest prices outside of the Middle East. But let's note two things about the proposal for substantial carbon taxes:

- First, revenue recycling could redirect the revenue from carbon and other environmental taxes to lower other taxes. Much of the political opposition to high energy taxes comes from the perception that they would be an *extra* tax – on top of the income, property, and social security taxes that people already pay. If a carbon tax was matched, for example, with a substantial cut in income or social security taxes, it might be more politically acceptable. The idea of increasing taxes on economic "bads" such as pollution, while reducing taxes on things we want to encourage, such as labor and capital investment, is fully consistent with principles of economic efficiency[18]. Rather than a net tax increase, this would be **revenue-neutral tax shift** - the total amount which citizens pay to the government in taxes is unchanged. Some of the tax revenues could also be used to provide relief for low-income people to offset the burden of higher energy costs.

- Second, if such a revenue-neutral tax shift did take place, individuals or businesses whose operations were more energy-efficient would actually save money overall. The higher cost of energy would also create a powerful incentive for energy-saving technological innovations and stimulate new markets. Economic adaptation would be easier if the higher carbon taxes (and lower income and capital taxes) were phased-in over time.

Policy Tools: Tradable Permits

An alternative to a carbon tax is a system of tradable carbon permits. A carbon trading scheme could be national in scope, or include several countries. An international permit system would work as follows:

- Each nation would be allocated a certain permissible level of carbon emissions. The total number of carbon permits issued would be equal to the desired national goal. For example, if carbon emissions for a particular country are currently 40 million tons and the policy goal is to reduce this by 10%, then permits would be issued to emit only 36 million tons. Note that different nations could be obliged to meet different targets, which is the case under the Kyoto Protocol.

- Permits are allocated to individual carbon-emitting sources in each nation. Including all carbon sources (e.g., all motor vehicles) in a trading scheme is clearly not practical. Instead, under most proposals permits would be allocated to the largest carbon emitters, such as power companies and manufacturing plants, or else to the suppliers through which carbon fuels enter the country – oil importers, coal mines, etc. These permits could initially be allocated for free on the basis of

[18] To encourage higher investment, carbon tax revenues could be used to lower capital gains or corporate taxes.

past emissions, or could be auctioned to the highest bidders. Economic theory indicates that the effectiveness of the trading system should be the same regardless of how the permits are allocated. However, there is a significant difference in the distribution of costs and benefits: giving permits out for free essentially amounts to a government subsidy to the polluters, while auctioning permits imposes real costs upon firms and generates public revenues.

- Firms are able to trade permits freely among themselves. Firms whose emissions exceed the number of permits they hold must purchase additional permits or else face penalties. Meanwhile firms that are able to reduce their emissions below their allowance at low cost will seek to sell their permits for a profit. Firms will settle upon permit prices through free market negotiations. It may also be possible for environmental groups or other organizations to purchase permits and retire them – thus reducing overall emissions.

- Nations and firms could also receive credit for financing carbon reduction efforts in other countries. For example, a German firm could get credit for installing efficient electric generating equipment in China, replacing highly polluting coal plants.

From an economic point of view, the advantage of a tradable permit system is that it would encourage the least-cost carbon reduction options to be implemented. Depending on the allocation of permits, it might also mean that developing nations could transform permits into a new export commodity by choosing a non-carbon path for their energy development. They would then be able to sell permits to industrialized nations who were having trouble meeting their reduction requirements.

To demonstrate the economic impacts of a tradable carbon permit system, we can use the analytical concept of **marginal net benefits**. Figure 9 shows the marginal net benefit of carbon emissions to producers and consumers.[19] We would expect the marginal net benefit curve to slope downward because the initial carbon emissions are used to produce those goods and services which are most valued by producers and consumers. Subsequent carbon emissions are used to produce goods and services of lower net value.

The emissions level Q_E will result if there are no limits on emissions – this is the market equilibrium, where consumers and producers maximize net benefits. We can see that the marginal benefits of the last units of carbon emissions are rather small. However, producers and consumers interacting in a market do not take into account environmental externalities. Thus the overall level of carbon emissions is too high from the perspective of maximizing social welfare.

---

[19] The marginal net benefit curve is derived from the demand and supply curve (in this case for carbon-based fuels), showing the marginal benefits of the product minus the marginal costs of the supply.

Under a permit system, Q* represents the total number of permits issued. The equilibrium permit price will then be P*, reflecting the marginal net benefit of carbon emissions at Q*. It is advantageous for emitters who gain benefits greater than P* from their emissions to purchase permits, while those with emissions benefits less than P* will do better to reduce emissions and sell any excess permits.

Figure 9. Determination of a Carbon Permit Price

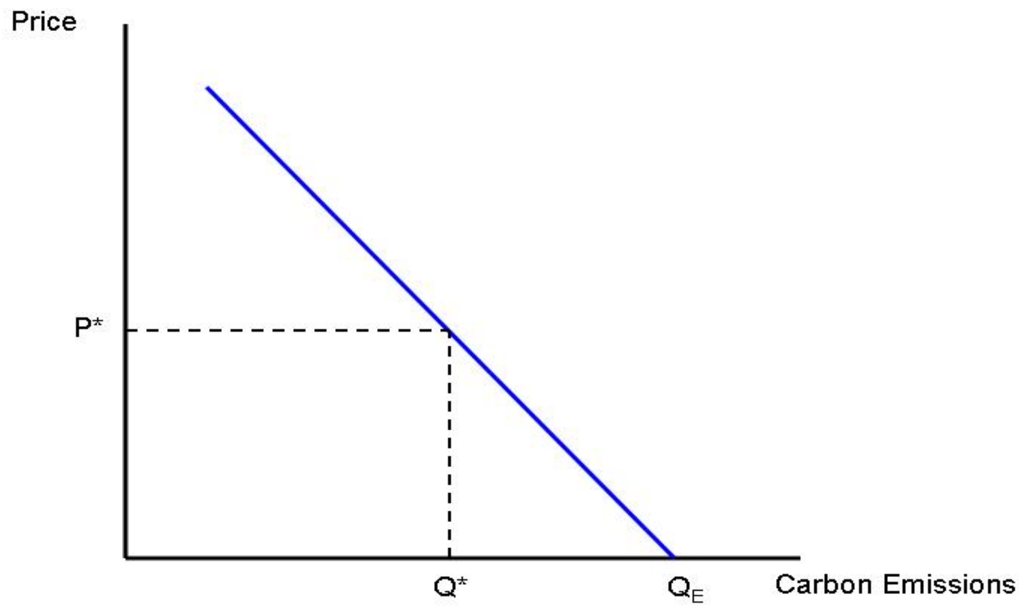
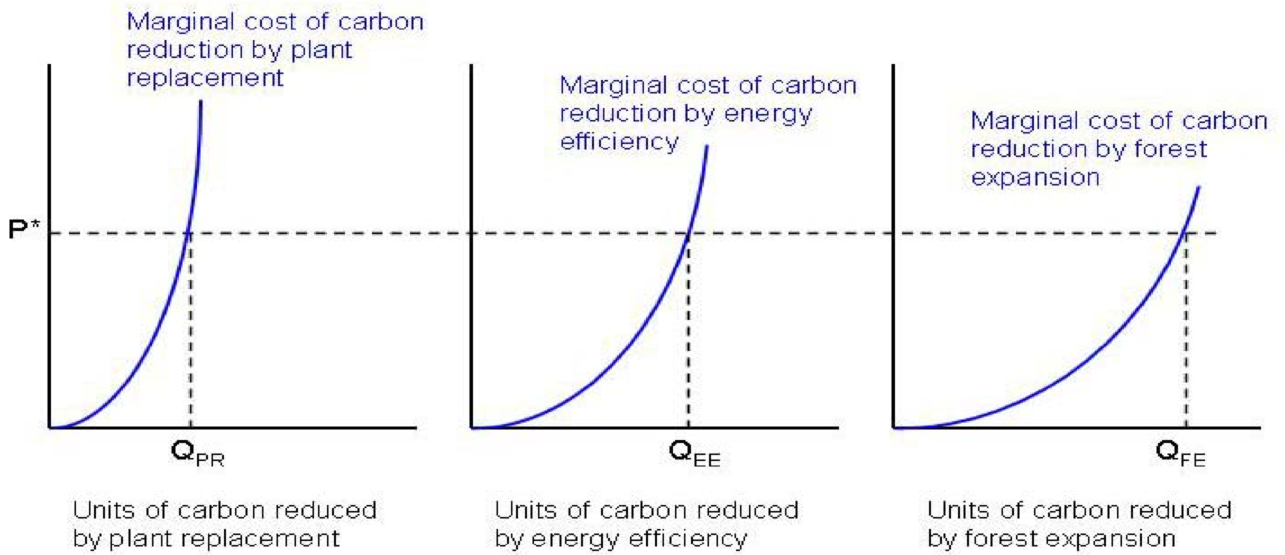


Figure 10 shows how this system affects carbon reduction strategies. Three possibilities are shown. In each case, the graph shows a *marginal cost* of reducing carbon emissions through a particular policy or technology. These marginal costs generally rise as more units of carbon are reduced, but they may be higher and increase more rapidly for some options than others. In this example, replacement of power plants using existing carbon-emitting technologies is possible, but will tend to have high marginal costs – as shown in the first graph in Figure 10. Reducing emissions through greater energy efficiency has lower marginal costs, as seen in the middle graph. Finally, carbon storage through forest area expansion has the lowest marginal costs. The permit price P* (as determined in Figure 9) will govern the relative levels of implementation of each of these strategies. Firms will find it profitable to reduce emissions with a given policy option so long as the costs of that option are lower than the cost of purchasing a permit. In this example, we see that forest expansion would be used for the greatest share of the reduction, while plant replacement would be used for the lowest share.

Nations and corporations who participate in such a trading scheme can decide for themselves how much of each control strategy to implement, and will naturally favor the least-cost methods. This will probably involve a combination of different approaches. Suppose one nation undertakes extensive reforestation. It is then likely to have excess permits, which it can sell to a nation with few low-cost reduction options. The net effect will be the worldwide implementation of the least-cost reduction techniques.

This system combines the advantages of economic efficiency with a guaranteed result: reduction in the overall emissions level Q*. The problem, of course, is to achieve agreement on the initial allocation of permits. There may also be measurement problems, and issues such as whether to count only commercial carbon emissions, or to include emissions changes resulting from land use patterns.

Figure 10. Carbon Reduction Options with a Permit System



Note: Marginal costs shown here are hypothetical.

Other Policy Tools: Subsidies, Standards, R&D, and Technology Transfer

Although political problems may prevent the adoption of sweeping carbon taxes or transferable permit systems, there are a variety of other policy measures which have potential to lower carbon emissions. These include:

- *Shifting subsidies from carbon-based to non-carbon-based fuels.* Many countries currently provide direct or indirect subsidies to fossil fuels. The elimination of these subsidies would alter the competitive balance in favor of alternative fuel sources. If these subsidy expenditures were redirected to renewable sources, especially in the form of tax rebates for investment, it could promote a boom in investment in solar photovoltaics, fuel cells, biomass and wind power – all technologies which are currently at the margin of competitiveness in various areas.

- *The use of efficiency standards to require utilities and major manufacturers to increase efficiency and renewable content in power sources.* A normal coal-fired generating plant achieves about 35% efficiency, while a high-efficiency gas-fired co-generation facility achieves from 75% to 90% efficiency. Current automobile fuel-efficiency standards in the United States do not exceed 27.5 miles per gallon, while efficiencies of up to 50 miles per gallon are achievable with proven technology. Tightening standards over time for plants, buildings, vehicles, and appliances would hasten the turnover of existing, energy-inefficient capital stock.

- *Research and development (R&D) expenditures promoting the commercialization of alternative technologies.* Both government R&D programs and favorable tax treatment of corporate R&D for alternative energy can speed commercialization. The existence of a non-carbon "backstop" technology significantly reduces the economic cost of measures such as carbon taxes, and if the backstop became fully competitive with fossil fuels carbon taxes would be unnecessary.

- ***Technology transfer** to developing nations.* The bulk of projected growth in carbon emissions will come in the developing world. Many energy development projects are now funded by agencies such as the World Bank and regional development banks. To the extent that these funds can be directed towards non-carbon energy systems, supplemented by other funds dedicated specifically towards alternative energy development, it will be economically feasible for developing nations to turn away from fossil-fuel intensive paths, achieving significant local environmental benefits at the same time.

Climate Change Policy in Practice

Climate change is an international environmental issue. Each individual nation has little incentive to reduce its emissions if other nations do not agree to similar reductions, because unilaterally reducing emissions could impose significant costs while having a negligible effect on overall emissions. Thus a binding international agreement is necessary, especially if the policy goal is to reduce emissions by 50-80%.

The most comprehensive international agreement on climate change has been the Kyoto Protocol. Under the treaty industrial countries agreed to emission reduction targets by 2008-2012 compared to baseline emissions in 1990. For example, the United States agreed to a 7% reduction, France to an 8% reduction, and Japan to a 6% reduction. Developing nations such as China and India are not bound to emissions targets under the treaty, an omission that the United States and some other countries objected to. As of October 2007, the Kyoto Protocol has been ratified by 176 countries. The United States signed the treaty in 1998 but has never ratified it. In 2001, the Bush administration rejected the Kyoto Protocol, arguing that negotiations had failed and that a new approach was necessary. While this has dealt a serious blow to efforts to control global greenhouse gas emissions, the Kyoto Protocol nonetheless entered into force in early 2005 after Russia ratified the treaty in November 2004.

To achieve the goals of the Protocol in a cost effective manner, the treaty includes three “flexibility mechanisms.” One is the trading of emissions permits among nations that are bound by specific targets. Thus one nation unable to meet its target could purchase permits from another nation that reduces its emissions below its requirements. The European Union has set up a carbon trading system which went into effect in 2005. (see Box 6).

Another flexibility mechanism is **joint implementation**, whereby an industrial nation receives credit for financing emission-reducing projects in other countries bound to emissions targets, mainly in transitional countries such as Russia and Lithuania. The third is the **clean development mechanism**, whereby industrial nations can obtain credit for financing emission-reducing or emission-avoiding projects in developing nations not bound to specific emissions targets, including China and India.

As the Kyoto Protocol approaches its 2012 expiration date, will the treaty meet its objectives? The overall goal was a 5% reduction (compared to the 1990 baseline) in greenhouse gas emissions among participating countries. As of 2004, total emissions among countries that have signed the treaty, including countries that haven’t ratified it such as the U.S. Australia, have declined about 3% compared to the 1990 baseline.[20] This appearance of success is largely illusory, for much of the decline is a result of economic collapse in the former Soviet Union and other Eastern European countries. For these transitional nations, overall emissions have declined by 37%. For the remaining participating countries, overall emissions have increased by 11%. Canada, for example, agreed to a 6% reduction but its emissions had increased 27% over the baseline as of 2004. Negotiations are currently underway to draft a successor to the Kyoto Protocol when it expires in 2012. A central question in these negotiations is whether it will be possible to obtain agreements from the United States and developing nations to meet emissions targets.

[20] United Nations Framework Convention on Climate Change, “National Greenhouse Gas Inventory Data for the Period 1990-2004 and Status of Reporting,” October 19, 2006.

BOX 6: THE EUROPEAN UNION CARBON TRADING SYSTEM

In 2005 the European Union launched its Emissions Trading Scheme (ETS), which covers about 12,000 facilities that collectively emit about 40% of the EU's carbon emissions. Under the ETS, each nation develops a National Allocation Plan to determine the overall number of permits available in the country, and the number of permits to allocate to each facility. So far, permits have been allocated to firms. for free. Any unneeded permits can be sold on the open market.

The effectiveness of the ETS has proven to be disappointing. The problem is that national governments have been too generous in allocating the free permits, resulting in a declining price for carbon permits. While the going price of a permit to emit a ton of carbon gradually rose during the first year or so of the program, up to about 30 euros, the price crashed in May 2006 to under €10 per ton and then continued to fall, going below €1 per ton in early-2007. EU statistics indicate that 93% of the facilities included in the ETS emitted less carbon than allowed by their permit allocation in 2006. With a glut of permits available, those firms emitting above their allocation were able to purchase permits at very low prices. Statistics also show that the ETS has had little effect on overall carbon emissions in the EU.

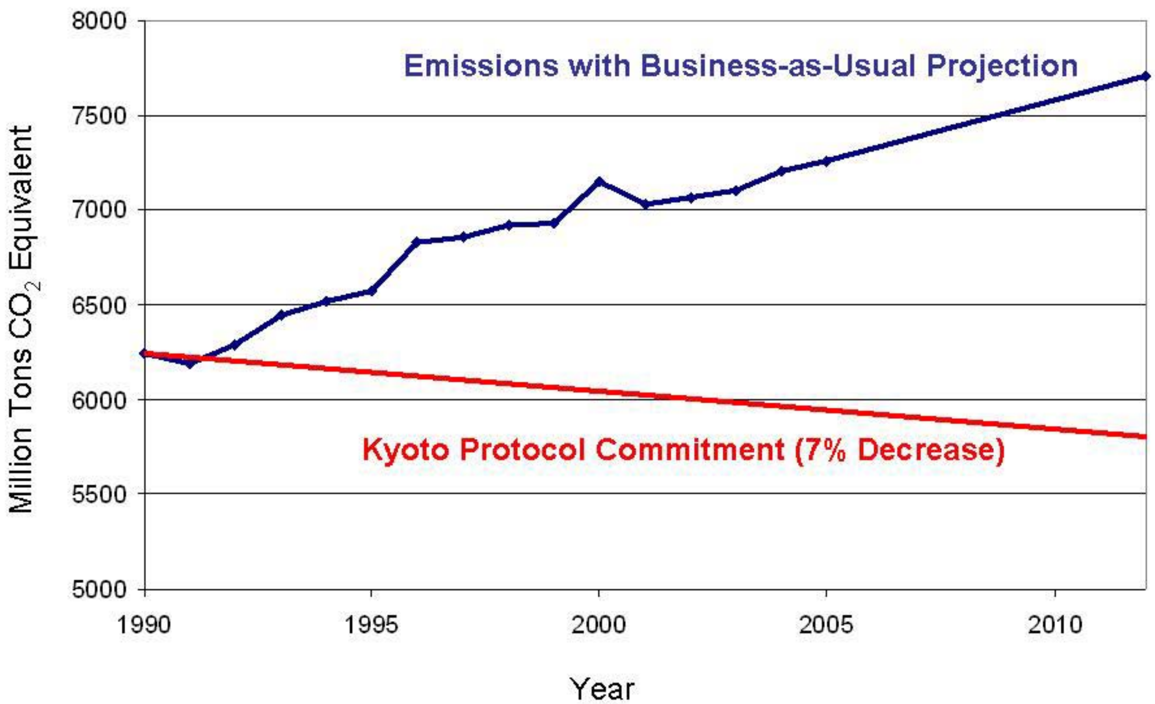
The EU is currently moving towards the second phase of the ETS, which will cover the 2008-2012 period. EU nations have recognized that permits were over-allocated in the initial phase and intend to reduce the number of permits available in the second phase, with the goal of increasing the permit price in the future and thus creating greater incentives for firms to reduce emissions.

Sources: "Q&A: Europe's Carbon Trading Scheme," *BBC News*, December 20, 2006; "Smoke Alarm: EU Shows Carbon Trading is not Cutting Emissions," *The Guardian*, April 3, 2007.

While the United States has dropped out of the Kyoto Protocol, it has set its own climate change goals. Unlike the Kyoto Protocol, these goals are voluntary rather than binding. In 2002 President Bush set a goal of reducing **greenhouse gas intensity** by 18% between 2002 and 2012. Greenhouse gas intensity is defined as the quantity of greenhouse gas emissions per unit of GDP. Thus even if actual emissions remain constant greenhouse gas intensity will decline as long as the economy is growing. In fact, it is possible for greenhouse gas intensity to decline even while actual emissions increase.

Does the goal of reducing greenhouse gas intensity by 18% represent an ambitious goal? It does not. Consider that greenhouse gas intensity fell by 21% during the 1980s and 16% in the 1990s, without any active policies to limit greenhouse gas emissions. Thus the plan essentially amounts to a business-as-usual approach. We see in Figure 11 that greenhouse gas emissions in the U.S. are increasing. Even if the Bush Administration's goal is met, greenhouse gas emissions in the U.S. are projected to be about 24% higher in 2012 as compared to the 1990 baseline for the Kyoto Protocol (see Figure 11). Contrast this emissions increase to the 7% *decrease* that would have been required for the U.S. under the Kyoto Protocol.

Figure 11. Historical and Projected Greenhouse Gas Emissions in the United States



Sources: Historical emissions from the *U.S. Inventory of Greenhouse Gas Emissions and Sinks*, U.S. Environmental Protection Agency. Projected emissions from *Fourth Climate Action Report to the UN Framework Convention on Climate Change*, U.S. Department of State, 2007.

The Future of Climate Change Policy

Will the limited policy measures now being taken to control greenhouse emissions be sufficient? Recent evidence of increased rapidity of climate change suggests that the cumulative impact of emissions may be more severe than anticipated. Arctic ecosystems have shown clear signs of breakdown as temperatures rise, raising the possibility of feedback effects from tundra melting, which would further accelerate global warming.[21]

[21] Richard B. Alley, "Abrupt Climate Change," *Scientific American* November 2004; Clifford Kraus, "Eskimos Fret as Climate Shifts and Wildlife Changes," *New York Times*, September 6, 2004.

A report prepared for the U.S. Department of Defense cited the possibility of large-scale drought in critical agricultural regions; a collapse of the North Atlantic Gulf Stream, causing an abrupt shift to much colder temperatures in Europe and the Northern U.S.; and widespread civil unrest and mass migration caused by disruption of water and food supplies.[22] The costs of such developments would be clearly be very high, well into the higher range of estimates in Table 2, amounting to hundreds of billions of dollars per year.

The Intergovernmental Panel on Climate Change has estimated that the stabilization of atmospheric CO_2 levels would require reduction of CO_2 emissions to a small fraction of current levels. This goal is far beyond the Kyoto Protocol targets, and would require major policy intervention to redirect the world's economies towards non-carbon energy sources. The IPCC also finds, however, that opportunities for reductions of 30-70% in greenhouse gas emissions are available at a net cost below $100 per ton of carbon equivalent; a substantial portion of these cuts would have low or even zero marginal cost. According to these figures, the IPCC's maximum estimated reduction, of 5 billion tons, could be achieved at a net cost of several hundred billion dollars – a large amount, but probably less than the cost of the high-scenario damages, even using standard discount rates.[23] Certainly the low-cost cuts look like a good investment.

Economic analysis could thus justify much more aggressive climate change policy, but significant political barriers stand in the way of such policies, especially in the U.S. As the ratifying nations move to implant the Kyoto Protocol, and as attention focuses on future policies "beyond Kyoto", the economic policy measures discussed in this chapter will certainly become increasingly important. Political leaders and the public will determine how strongly we will respond to this major issue of the twenty-first century, but economic policies will be central to accomplishing the goals we choose.

[22] Peter Schwartz and Doug Randall, "An Abrupt Climate Change Scenario and Its Implications for U.S. National Security," October 2003, available at http://www.ems.org/climate/pentagon_climate_change.html.
[23] IPCC 2001, 2007.

SUMMARY

Climate change, arising from the greenhouse effect of heat-trapping gases, is a global problem. All nations are involved in both its causes and consequences. Currently developed nations are the largest emitters of greenhouse gases, but emissions by developing nations will grow considerably in coming decades. The most recent scientific evidence indicates that effects during the twenty-first century may range from a global temperature increase of1.1ºC (2ºF) to as much as 6.4ºC (11.5ºF). In addition to simply warming the planet, other predicted effects include disruption of weather patterns and possible sudden major climate shifts.

One approach to economic analysis of climate change is cost/benefit analysis. The benefits in this case are the damages potentially averted through action to prevent climate change; the costs are the economic costs of shifting away from fossil fuel dependence, as well as other economic implications of greenhouse gas reduction. Cost-benefit studies have estimated both costs and benefits in the range of several percent of GDP. However, the relative evaluation of costs and benefits depends heavily on the discount rate selected. Since the damages are expected to increase with time, the use of a high discount rate leads to a lower evaluation of the benefits of avoiding climate change. In addition, some effects such as species loss and effects on human life and health are difficult to measure in monetary terms. Also, depending on the assumptions used in economic models, the GDP impacts of policies to avoid climate change could range from a 3.4% decrease to a 3.9% increase in GDP.

Policies to respond to global climate change could be preventive or adaptive. One of the most widely discussed policies is a carbon tax, which would fall most heavily on fuels causing the highest carbon emissions. The revenues from such a tax could be recycled to lower taxes elsewhere in the economy, or they could be used to assist people in lower income brackets, who will suffer most from higher costs of energy and goods. Another policy option is tradable carbon emissions permits, which could be bought and sold by firms or nations, depending on their level of carbon emissions. Both these policies have the advantage of economic efficiency, but it has been difficult to obtain the political support necessary to implement them. Other possible policy measures include shifting subsidies away from fossil fuels and towards renewable energy, strengthening energy efficiency standards, and increasing research and development on alternative energy technologies.

The Kyoto Protocol mandating reductions of greenhouse gases by industrialized nations went into force in 2005, but the U.S. refused to participate. Effective climate change policy in the future will require involvement of the U.S. as well as China, India, and other developing nations. Much more ambitious reduction targets will be needed to avoid the costs associated with long-term climate change.

KEY TERMS AND CONCEPTS

Adaptive measures: policies intended to adapt to adverse environmental impacts.

Avoided costs: costs avoidable through environmental preservation or improvement.

Carbon sinks: portions of the ecosystem with the ability to absorb certain quantities of carbon dioxide, such as forests, soils and oceans.

Carbon tax: a per-unit tax on goods and services based on the quantity of carbon dioxide emitted during the production or consumption process.

Clean development mechanism: a component of the Kyoto Protocol that allows industrial countries to receive credits for helping developing countries to reduce their carbon emissions.

Common property resource: a resource not subject to private ownership and available to all, such as a public park, or the oceans, or the capacity of the Earth and its atmosphere to absorb carbon.

Cost-benefit analysis: a tool for policy analysis that attempts to monetize all the costs and benefits of a proposed action, in order to determine the net benefits.

Cost-effectiveness analysis: a policy tool that determines the least-cost approach for achieving a given goal.

Discount rate: the annual rate at which future benefits or costs are discounted relative to current benefits or costs.

Elasticity of demand: the sensitivity of the quantity demanded to prices.

Externality: an effect of a market transaction on individuals or firms other than those directly involved in the transaction.

Feedback effects: the process of changes in a system leading to other changes that either counteract or reinforce the original change.

Global climate change: the changes in global climate, including temperature, precipitation, and storm frequency and intensity, that result from changes in greenhouse gas concentrations in the atmosphere.

Global commons: global common property resources such as the atmosphere and the oceans.

Greenhouse effect: the effect of certain gases in the earth's atmosphere trapping solar radiation, resulting in an increase in global temperatures and other climactic changes.

Greenhouse gas: gases such as carbon dioxide and methane whose atmospheric concentrations influence global climate by trapping solar radiation.

Greenhouse gas intensity: the amount of greenhouse gas emissions per unit of economic output.

Joint implementation: a component of the Kyoto Protocol whereby industrial nations can obtain credit for financing carbon-reducing projects in other industrial nations.

Marginal net benefit: the net benefit of the consumption or production of an additional unit of a resource; equal to marginal benefit minus marginal cost.

Pollution taxes: a per-unit tax based on the pollution associated with the production of a good or service.

Public goods: goods available to all, whose use by one person does not reduce their availability to others.

Precautionary principle: the principle that policies should take steps to avoid outcomes damaging to health or environment, even if the damaging outcomes cannot be predicted with certainty, and especially when such outcomes are potentially catastrophic or irreversible.

Preventive measures: policies intended to prevent adverse environmental impacts.

Revenue-neutral tax shift: policies designed to balance tax increases on certain products or activities with reductions in other taxes, such as a reduction in income taxes that offset a carbon-based tax.

Stock pollutant: a pollutant that accumulates in the environment, such as carbon dioxide and chlorofluorocarbons (CFCs).

Technology transfer: the process of sharing technological information or equipment, particularly among nations.

Transferable (tradable) permits: permits that allow a certain quantity of pollution and that may be traded among firms or nations.

REFERENCES

Cline, William R. *The Economics of Global Warming*. Washington D.C.: Institute for International Economics, 1992.

Dower, Roger C. and Zimmerman, Mary. *The Right Climate for Carbon Taxes, Creating Economic Incentives to Protect the Atmosphere*. Washington D.C.: World Resources Institute, 1992.

Fankhauser, Samuel. *Valuing Climate Change: the Economics of the Greenhouse*. London: Earthscan Publications, 1995.

Harris, Jonathan M. *Environmental and Natural Resource Economics: A Contemporary Approach*. Boston, Massachusetts: Houghton Mifflin, 2006 http://college.hmco.com

Intergovernmental Panel on Climate Change (IPCC). *Climate Change 2001, Volume 1: The Scientific Basis.* Cambridge, UK: Cambridge University Press, 2001.

Intergovernmental Panel on Climate Change (IPCC). *Climate Change 2007, Volume 1: The Physical Science Basis*. Available at http://www.ipcc.ch/

Manne, Alan S. and Richels, Richard G. *Buying Greenhouse Insurance: The Economic Costs of CO2 Emissions Limits*. Cambridge, Mass: The MIT Press, 1992.

Nordhaus, William D., and Joseph Boyer, *Warming the World: Economic Models of Global Warming*. Cambridge, Mass: The MIT Press, 2000.

Roodman, David M. *Getting the Signals Right: Tax Reform to Protect the Environment and the Economy*. Worldwatch Paper # 134. Washington, D.C.: Worldwatch Institute, 1997.

Stern, Nicholas, *The Economics of Climate Change: The Stern Review*. New York and Cambridge, U.K.: Cambridge University Press, 2007. Available at http://www.hm-treasury.gov.uk/independent_reviews/stern_review_economics_climate_change/stern_review_report.cfm

U.S. Department of Energy, Energy Information Administration. *International Energy Outlook 2004*, Report No. DOE/EIA-0484 (2004).

U.S. Department of Energy, Energy Information Administration. *International Energy Outlook 2007*. Report No. DOE/EIA-0484 (2007).

DISCUSSION QUESTIONS

1. Do you consider cost-benefit a useful means of addressing the problem of climate change? How can we adequately value things like the melting of arctic ice caps and inundation of island nations? What is the appropriate role of economic analysis in dealing with questions that affect global ecosystems and future generations?

2. Which policies to address climate change would be most effective? How can we decide which combination of policies to use? What kinds of policies would be especially recommended by economists? What are the main barriers to effective policy implementation?

3. The process for formulating and implementing international agreements on climate change policy has been plagued with disagreements and deadlocks. What are the main reasons for the difficulty in agreeing on specific policy actions? From an economic point of view, what kinds of incentives might be useful to induce nations to enter and carry out agreements? What kinds of "win-win" policies might be devised to overcome negotiating barriers?

EXERCISES

1. Suppose that under the terms of an international agreement, U.S. CO_2 emissions are to be reduced by 200 million tons, and those of Brazil by 50 million tons.

Here are the policy options that the U.S. and Brazil have to reduce their emissions:

USA:

Policy options	Total emissions reduction (million tons carbon)	Cost ($ billion)
A: Efficient machinery	60	12
B: Reforestation	40	20
C: Replace coal fueled power plants	120	30

Brazil:

Policy options	Total emissions reduction (million tons carbon)	Cost ($ billion)
A: Efficient machinery	50	20
B: Protection of Amazon forest	30	3
C: Replace coal fueled power plants	40	8

a) Which policies are most efficient for each nation in meeting their reduction targets? How much will be reduced using each option, at what cost, if the two nations must operate independently? Assume that any of the policy options can be partially implemented at a constant marginal cost. For example, the U.S. could choose to reduce carbon emissions with efficient machinery by 10 million tons at a cost of $2 billion. (Hint: start by calculating the average cost of carbon reduction in dollars per ton for each of the six policies).
b) Suppose a market of transferable permits allows the U.S. and Brazil to trade permits to emit CO_2. Who has an interest in buying permits? Who has an interest in selling permits? What agreement can be reached between the U.S. and Brazil so that they can meet the overall emissions reduction target of 250 million tons at the least cost? Can you estimate a range for the price of a permit to emit one ton of carbon? (Hint: use your average cost calculations from the first part of the question.)

2. Suppose that the annual consumption of an average American household is 2000 gallons of oil in heating and transportation and 300 ccf (hundred cubic feet) of natural gas. Using the figures given in Table 2 on the effects of a carbon tax, calculate how much an average American household would pay per year with an added tax of $10 per ton of carbon. (One barrel of oil contains 42 gallons.) Assume that this relatively small tax initially causes no reduction in the demand for oil and gas. Figuring 100 million households in the United States, what would be the revenue to the U.S. Treasury of such a carbon tax?

What would be the national revenue resulting from a tax of $200 per ton of carbon? Consider the issue of the impact of increased prices on consumption – a reasonable assumption about consumption elasticity might be that a $200 per ton tax would cause the quantity of oil and gas consumed to decline by 20%. How might the government use such revenues? What would the impact be on the average family? Consider the difference between the short-term and long-term impacts.

WEB LINKS

1. http://epa.gov/climatechange/index.html The global warming web site of the U.S. Environmental Protection Agency. The site provides links to information on the causes, impact, and trends related to global climate change.

2. http://www.ipcc.ch/ The web site for the Intergovernmental Panel on Climate Change, a United Nations-sponsored agency "to assess the scientific, technical, and socioeconomic information relevant for the understanding of the risk of human-induced climate change." Their web site includes assessment reports detailing the relationships between human actions and global climate change.

3. http://climate.wri.org/ World Resource Institute's web site on climate and atmosphere. The site includes several articles and case studies, including research on Clean Development Mechanisms.

4. http://www.unfccc.de/ Home page for the United Nations Framework Convention on Climate Change. The site provides data on the climate change issue and information about the ongoing process of negotiating international agreements related to climate change.

5. http://www.weathervane.rff.org/ A web site sponsored by Resources for the Future devoted to climate change issues. The site includes several research papers on the trading of greenhouse gas emissions permits.

ELSEVIER
www.elsevier.com/locate/worlddev

World Development Vol. 32, No. 8, pp. 1419–1439, 2004

Printed in Great Britain
0305-750X/$ - see front matter

doi:10.1016/j.worlddev.2004.03.004

The Rise and Fall of the Environmental Kuznets Curve

DAVID I. STERN *
Rensselaer Polytechnic Institute, Troy, NY, USA

Summary. — This paper presents a critical history of the environmental Kuznets curve (EKC). The EKC proposes that indicators of environmental degradation first rise, and then fall with increasing income per capita. Recent evidence shows however, that developing countries are addressing environmental issues, sometimes adopting developed country standards with a short time lag and sometimes performing better than some wealthy countries, and that the EKC results have a very flimsy statistical foundation. A new generation of decomposition and efficient frontier models can help disentangle the true relations between development and the environment and may lead to the demise of the classic EKC.

Key words — environmental Kuznets curve, pollution, economic development, econometrics, review, global

1. INTRODUCTION

The environmental Kuznets curve (EKC) is a hypothesized relationship between various indicators of environmental degradation and income per capita. In the early stages of economic growth degradation and pollution increase, but beyond some level of income per capita, which will vary for different indicators, the trend reverses, so that at high income levels economic growth leads to environmental improvement. This implies that the environmental impact indicator is an inverted U-shaped function of income per capita. Typically, the logarithm of the indicator is modeled as a quadratic function of the logarithm of income. An example of an estimated EKC is shown in Figure 1. The EKC is named for Kuznets (1955) who hypothesized that income inequality first rises and then falls as economic development proceeds.

The EKC concept emerged in the early 1990s with Grossman and Krueger's (1991) path-breaking study of the potential impacts of NAFTA and the concept's popularization through the 1992 World Bank Development Report (IBRD, 1992). If the EKC hypothesis were true, then rather than being a threat to the environment, as claimed by the environmental movement and associated scientists in the past (e.g., Meadows, Meadows, Randers, & Behrens, 1972), economic growth would be the means to eventual environmental improvement. This change in thinking was already underway in the emerging idea of sustainable economic development promulgated by the World Commission on Environment and Development (1987) in *Our Common Future*. The possibility of achieving sustainability without a significant deviation from business as usual was an obviously enticing prospect for many—letting

* I thank Cutler Cleveland, Quentin Grafton and two anonymous referees, for comments on drafts of this manuscript. Participants at seminars at the University of Wollongong, Clark University, University of York, University of Cambridge, Rensselaer Polytechnic Institute, and the Australian National University and at conferences organized by the Association of American Geographers in Los Angeles and the International Society for Ecological Economics in Sousse, Tunisia aided in developing my arguments. Additionally, I am grateful to my collaborators Mick Common, Roger Perman, and Kali Sanyal for their contributions discussed in this paper. Final revision accepted: 23 March 2004.

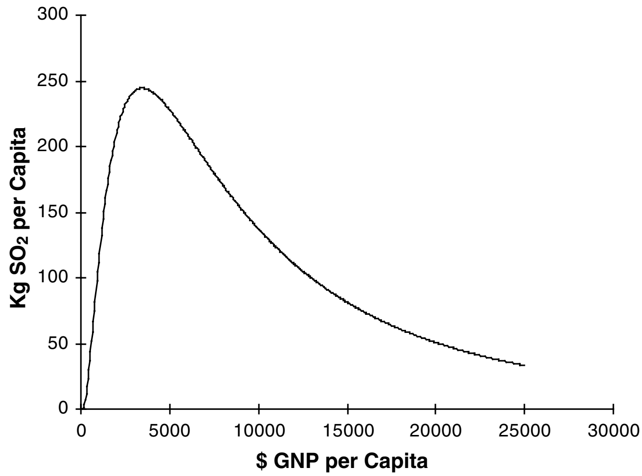


Figure 1. *Environmental Kuznets curve for sulfur emissions. Source: Panayotou (1993) and Stern, Common, and Barbier (1996).*

humankind "have our cake and eat it" (Rees, 1990, p. 435).

The EKC is an essentially empirical phenomenon, but most of the EKC literature is econometrically weak. In particular, little or no attention has been paid to the statistical properties of the data used—such as serial dependence or stochastic trends in time-series [1]—and little consideration has been paid to issues of model adequacy such as the possibility of omitted variables bias. [2] Most studies assume that, if the regression coefficients are nominally individually or jointly significant and have the expected signs, then an EKC relation exists. However, one of the main purposes of doing econometrics is to test which apparent relationships, or "stylized facts," are valid and which are spurious correlations.

When we do take diagnostic statistics and specification tests into account and use appropriate techniques, we find that the EKC does not exist (Perman & Stern, 2003). Instead, we get a more realistic view of the effect of economic growth and technological changes on environmental quality. It seems that emissions of most pollutants and flows of waste are monotonically rising with income, though the "income elasticity" is less than one and is not a simple function of income alone. Income-independent, time-related effects reduce environmental impacts in countries at all levels of income. The new (post-Brundtland) conventional wisdom that developing countries are "too poor to be green" (Martinez-Alier, 1995) is, itself, lacking in wisdom. In rapidly growing middle-income countries, however the scale effect, which increases pollution and other degradation, overwhelms the time effect. In wealthy countries, growth is slower, and pollution reduction efforts can overcome the scale effect. This is the origin of the apparent EKC effect. The econometric results are supported by recent evidence that, in fact, pollution problems are being addressed and remedied in developing economies (e.g., Dasgupta, Laplante, Wang, & Wheeler, 2002).

This paper follows the development of the EKC concept in approximately chronological order. I do not attempt to review or cite all of the rapidly growing number of studies. The next two sections of the paper review in more detail the theory behind the EKC and the econometric methods used in EKC studies. The following sections review some EKC analyses and their critique. Sections 6 and 7 discuss the more important recent developments that have changed the picture that we have of the EKC. The final sections discuss alternative approaches—decomposition of emissions and efficient frontiers—and summarize the findings.

2. THEORETICAL BACKGROUND

The EKC concept emerged in the early 1990s with Grossman and Krueger's (1991) pathbreaking study of the potential impacts of NAFTA and Shafik and Bandyopadhyay's

(1992) background study for the 1992 *World Development Report*. The EKC theme was popularized by the World Bank's *World Development Report 1992* (IBRD, 1992), which argued that: "The view that greater economic activity inevitably hurts the environment is based on static assumptions about technology, tastes and environmental investments" (p. 38) and that "As incomes rise, the demand for improvements in environmental quality will increase, as will the resources available for investment" (p. 39). Others have expounded this position even more forcefully with Beckerman (1992) claiming that "there is clear evidence that, although economic growth usually leads to environmental degradation in the early stages of the process, in the end the best—and probably the only—way to attain a decent environment in most countries is to become rich." (p. 491). In his highly publicized and controversial book, *The Skeptical Environmentalist*, Lomborg (2001) relies heavily on the 1992 *World Development Report* (Cole, 2003a) to argue the same point, while many environmental economists take the EKC as a stylized fact that needs to be explained by theory. All this is despite the fact that the EKC has never been shown to apply to all pollutants or environmental impacts and recent evidence, [3] discussed in this paper, challenges the notion of the EKC in general. The remainder of this section discusses the economic factors that drive changes in environmental impacts and may be responsible for rising or declining environmental degradation over the course of economic development.

If there were no change in the structure or technology of the economy, pure growth in the scale of the economy would result in growth in pollution and other environmental impacts. This is called the scale effect. The traditional view that economic development and environmental quality are conflicting goals reflects the scale effect alone. Proponents of the EKC hypothesis argue that

> at higher levels of development, structural change towards information-intensive industries and services, coupled with increased environmental awareness, enforcement of environmental regulations, better technology and higher environmental expenditures, result in leveling off and gradual decline of environmental degradation (Panayotou, 1993, p. 1).

Thus there are both proximate causes of the EKC relationship—scale, changes in economic structure or product mix, changes in technology, and changes in input mix, as well as underlying causes such as environmental regulation, awareness, and education, which can only have an effect via the proximate variables. Let us look in more detail at the proximate variables:

(a) *Scale* of production implies expanding production at given factor-input ratios, output mix, and state of technology. It is normally assumed that a 1% increase in scale results in a 1% increase in emissions. This is because if there is no change in the input–output ratio or in technique there has to be a proportional increase in aggregate inputs. However, there could, in theory, be scale economies or diseconomies of pollution (Andreoni & Levinson, 2001). Some pollution control techniques may not be practical at a small scale of production and vice versa or may operate more or less effectively at different levels of output.

(b) Different industries have different pollution intensities. Typically, over the course of economic development the *output mix* changes. In the earlier phases of development there is a shift away from agriculture toward heavy industry which increases emissions, while in the later stages of development there is a shift from the more resource intensive extractive and heavy industrial sectors toward services and lighter manufacturing, which supposedly have lower emissions per unit of output. [4]

(c) Changes in *input mix* involve the substitution of less environmentally damaging inputs for more damaging inputs and vice versa. Examples include: substituting natural gas for coal and substituting low sulfur coal in place of high sulfur coal. As scale, output mix, and technology are held constant, this is equivalent to moving along the isoquants of a neoclassical production function.

(d) Improvements in the *state of technology* involve changes in both:

- *Productivity* in terms of using less, *ceteris paribus* of the polluting inputs per unit of output. A general increase in total factor productivity will result in lower emissions per unit of output even though this is not necessarily an intended consequence.
- *Emissions specific changes in process* result in lower emissions per unit of input. These innovations are specifically intended to reduce emissions. [5]

This framework applies most directly to emissions of pollutants. For concentrations of pollutants, decentralization of economic activity with development is also important (Stern *et al.*, 1996). Deforestation is also a flow of environmental degradation. Improved technology would imply more replanting, selective cutting, wood recovery etc. that reduces deforestation per unit wood produced. Stock pollutants or impacts need a different, dynamic framework.

Though any actual change in the level of pollution must be a result of change in one of the proximate variables, those variables may be driven by changes in underlying variables that also vary over the course of economic development. A number of papers have developed theoretical models of how preferences and technology might interact to result in different time paths of environmental quality. The various studies make different simplifying assumptions about the economy. Most of these studies can generate an inverted U-shape curve of pollution intensity but there is no inevitability about this. The result depends on the assumptions made and the values of particular parameters. Lopez (1994) and Selden and Song (1995) assume infinitely lived agents, exogenous technological change and that pollution is generated by production and not by consumption. John and Pecchenino (1994), John, Pecchenino, Schimmelpfennig, and Schreft (1995), and McConnell (1997) develop models based on overlapping generations where pollution is generated by consumption rather than by production activities. In addition, Stokey (1998) allows endogenous technical change and Lieb (2001) generalizes Stokey's (1998) model, arguing that satiation in consumption is needed to generate the EKC. Finally, Ansuategi and Perrings (2000) incorporate transboundary externalities. Magnani (2001) discusses how individual preferences are converted into public policy. Andreoni and Levinson (2001) argue that none of these special assumptions is needed and economies of scale in abatement are sufficient to generate the EKC. Most studies model the emission of pollutants. Lopez (1994) and Bulte and van Soest (2001), among others, develop models for the depletion of natural resources such as forests or agricultural land fertility. It seems easy to develop models that generate EKCs under appropriate assumptions. None of these theoretical models has been tested empirically. Furthermore, if, in fact, the EKC for emissions is monotonic, as more recent evidence suggests, the ability of a model to produce an inverted U-shaped curve is not a particularly desirable property.

3. ECONOMETRIC FRAMEWORK

The earliest EKCs were simple quadratic functions of the levels of income. But, economic activity inevitably implies the use of resources and, by the laws of thermodynamics, use of resources inevitably implies the production of waste. Regressions that allow levels of indicators to become zero or negative are inappropriate except in the case of deforestation where afforestation can occur. A logarithmic dependent variable will impose this restriction. Some studies, including the original Grossman and Krueger (1991) paper, used a cubic EKC in levels and found an N-shape EKC. This might just be a polynomial approximation to a logarithmic curve. The standard EKC regression model is, therefore:

$$\ln(E/P)_{it} = \alpha_i + \gamma_t + \beta_1 \ln(\mathrm{GDP}/P)_{it} + \beta_2(\ln(\mathrm{GDP}/P))^2_{it} + \varepsilon_{it}, \qquad (1)$$

where E is emissions, P is population, and ln indicates natural logarithms. The first two terms on the RHS are intercept parameters which vary across countries or regions i and years t. The assumption is that, though the level of emissions per capita may differ over countries at any particular income level, the income elasticity is the same in all countries at a given income level. The time specific intercepts account for time-varying omitted variables and stochastic shocks that are common to all countries. The "turning point" income, where emissions or concentrations are at a maximum, is given by:

$$\tau = \exp(-\beta_1/(2\beta_2)). \qquad (2)$$

Usually the model is estimated with panel data. Most studies attempt to estimate both the fixed and random-effects models. The fixed-effects model treats the α_i and γ_t as regression parameters. The random-effects model treats the α_i and γ_t as components of the random disturbance. If the effects α_i and γ_t and the explanatory variables are correlated, then the random-effects model cannot be estimated consistently (Hsiao, 1986). Only the fixed-effects model can be estimated consistently. A Hausman (1978) test can be used to test for inconsistency in the random-effects estimate by comparing the fixed-effects and random-effects slope parameters. A significant difference indicates that the random-effects model is estimated inconsistently, due to correlation between the explanatory variables and the error components. Assuming that there are no other statistical problems, the fixed-effects model can be estimated consistently, but the estimated parameters are conditional on the country and time effects in the selected sample of data (Hsiao, 1986). Therefore, they cannot be used to extrapolate to other samples of data. This means that an EKC estimated with fixed-effects using only developed country data might say little about the future behavior of developing countries. Many studies compute the Hausman statistic and, finding that the random-effects model cannot be consistently estimated, estimate the fixed-effects model. But few have pondered the deeper implications of the failure of this orthogonality test.

GDP may be an integrated variable. If the EKC regressions do not cointegrate, then the estimates will be spurious. Until recently, very few studies have reported any diagnostic statistics for integration of the variables or cointegration of the regressions. Therefore, it is unclear what we can infer from the majority of EKC studies. Testing for integration and cointegration in panel data is a rapidly developing field. Perman and Stern (2003) employ some of these tests and find that sulfur emissions and GDP per capita may be integrated variables. The unit root hypothesis could be rejected for sulfur (but not GDP) using the Im, Pesaran, and Shin (2003) (IPS) test when the alternative was trend stationarity. But alternative hypotheses and tests result in acceptance of the unit root hypothesis. Heil and Selden (1999) find the same result for carbon dioxide emissions and GDP using the IPS test. But they prefer results that allow for a structural break 1974, in which allows them to strongly reject the unit root hypothesis for both GDP and carbon. Coondoo and Dinda (2002) yield similar results to Perman and Stern (2003) for carbon dioxide emissions. de Bruyn (2000) and Day and Grafton (2003) carry out time-series unit root tests for the Netherlands, the United Kingdom, the United States, West Germany, and Canada for a variety of pollutants with very similar results.

4. RESULTS OF EKC STUDIES

Many basic EKC models relating environmental impacts to income without additional explanatory variables have been estimated. But the key features differentiating the models for different pollutants, data etc. can be displayed by reviewing a few of the early studies and examining a single impact in more detail. I review the contributions of Grossman and Krueger (1991), Shafik (1994), and Selden and Song (1994) and then look in more detail at studies for sulfur pollution and emissions. Finally, I briefly discuss studies that estimate an EKC for energy use.

Many EKC studies have also been published that include additional explanatory additional explanatory variables, intended to model underlying or proximate factors, such as "political freedom" (e.g., Torras & Boyce, 1998) or output structure (e.g., Panayotou, 1997), or trade (e.g., Suri & Chapman, 1998). Stern (1998) reviews several of these. In general, the included variables turn out to be significant at traditional significance levels. Testing different variables individually is however subject to the problem of potential omitted variables bias. Further, these studies do not report cointegration or other statistics that might tell us if omitted variables bias is likely to be a problem or not. Therefore, it is not clear what we can infer from this body of work. Given these problems, I do not review these studies systematically here.

To some (e.g., Lopez, 1994) the early EKC studies indicated that local pollutants were more likely to display an inverted U-shape relation with income, while global impacts such as carbon dioxide did not. This picture fits environmental economics theory—local impacts are internalized within a single economy or region and are likely to give rise to environmental policies to correct the externalities on pollutees before such policies are applied to globally externalized problems. But as we will see, the picture is not quite so clear cut even in the early studies. Furthermore, the more recent evidence on sulfur and carbon dioxide emissions shows there may be no strong distinction between the effect of income per capita on local and global pollutants. Stern *et al.* (1996) determined that higher turning points were found for regressions that used purchasing power parity (PPP) adjusted income compared to those that used market exchange rates and for studies using emissions of pollutants relative to studies using ambient concentrations in urban areas. In the initial stages of economic development urban and industrial development tends to become more concentrated in a smaller number of cities which also have rising central population densities. Many developing countries have a "primate city" that dominates a country's urban hierarchy and contains much of its modern industry—Bangkok is one of the best such examples. In the later stages of economic development, urban and industrial development tends to decentralize. Moreover, the high population densities of less-developed cities are gradually reduced by suburbanization. So, it is possible for peak ambient pollution concentrations to fall as income rises even if total national emissions are rising.

The first empirical EKC study was the NBER working paper by Grossman and Krueger (1991) [6] that estimated EKCs for SO_2, dark matter (fine smoke), and suspended particles (SPM) using the GEMS dataset as part of a study of the potential environmental impacts of NAFTA. The GEMS dataset is a panel of ambient measurements from a number of locations in cities around the world. Each regression involved a cubic function in levels (not logarithms) of PPP per capita GDP and various site-related variables, a time trend, and a trade intensity variable. The turning points for SO_2 and dark matter were at around \$4,000–5,000 while the concentration of suspended particles appeared to decline even at low income levels. At income levels over \$10,000–15,000, Grossman and Krueger's estimates show increasing levels of all three pollutants.

The results of Shafik and Bandyopadhyay's (1992) [7] study were used in the 1992 *World Development Report* (IBRD, 1992) and were, therefore, particularly influential. They estimated EKCs for 10 different indicators using three different functional forms: log-linear, log-quadratic and, in the most general case, a logarithmic cubic polynomial in PPP GDP per capita as well as a time-trend and site-related variables. In each case, the dependent variable was untransformed. They found that lack of clean water and lack of urban sanitation decline uniformly with increasing income, and over time. Both measures of deforestation were found to be insignificantly related to the income terms. River quality tended to worsen with increasing income. The two air pollutants, however, conform to the EKC hypothesis. The turning points for both pollutants were at income levels of between \$3,000 and \$4,000. Finally, both municipal waste and carbon emissions per capita increased unambiguously with rising income.

Selden and Song (1994) estimated EKCs for four emissions series: SO_2, NO_x, SPM, and CO using longitudinal data primarily from developed countries. The estimated turning points were all very high compared to the two earlier studies. For the fixed-effects version of their model they were (converted to US\$1,990 using the US GDP implicit price deflator): SO_2, \$10,391; NO_x, \$13,383; SPM, \$12,275; and CO, \$7,114. This study showed that the turning point for emissions was likely to be higher than that for ambient concentrations.

Table 1 summarizes several studies of sulfur emissions and concentrations, listed in order of estimated income turning point. Panayotou (1993) used cross-sectional data, nominal GDP, and the assumption that the emission factor for each fuel is the same in all countries—this study has the lowest estimated turning point of all. With the exception of the Kaufmann *et al.* (1998) estimate, all turning point estimates using concentration data are less than \$6,000. Kaufmann *et al.* (1998) used an unusual specification that includes GDP per area and GDP per area squared variables.

Among the emissions based estimates, both Selden and Song (1994) and Cole *et al.* (1997) used databases dominated by, or consisting solely of, emissions from OECD countries. Their estimated turning points are \$10,391 and \$8,232 respectively. List and Gallet (1999) used data for the 50 US states from 1929–94. Their estimated turning point is the second highest in the table. Income per capita in their sample ranges from \$1,162 to \$22,462 in 1987 US dollars. This is a greater range of income levels than is found in the OECD-based panels for recent decades. This suggests that including more low-income data points in the sample might yield a higher turning point. Stern and Common (2001) estimated the turning point at over \$100,000. They used an emissions database produced for the US Department of Energy by ASL (Lefohn, Husar, & Husar, 1999) that covers a greater range of income levels and includes more data points than any of the other sulfur EKC studies.

We see that the recent studies that used more representative samples find that there is a monotonic relation between sulfur emissions and income just as there is between carbon dioxide and income. Interestingly, Dijkgraaf and Vollebergh (1998) estimate a carbon EKC for a panel data set of OECD countries finding an inverted U-shape EKC in the sample as a whole (as well as many signs of poor

Table 1. *Sulfur EKC studies*

Authors	Turning point 1990 USD	Emissions or concentrations	PPP	Additional variables	Data source for sulfur	Time period	Countries/cities
Panayotou (1993)	$3,137	Emissions	No	–	Own estimates	1987–88	55 developed and developing countries
Shafik (1994)	$4,379	Concentrations	Yes	Time trend, locational dummies	GEMS	1972–88	47 cities in 31 countries
Torras and Boyce (1998)	$4,641	Concentrations	Yes	Income inequality, literacy, political and civil rights, urbanization, locational dummies	GEMS	1977–91	Unknown number of cities in 42 countries
Grossman and Krueger (1991)	$4,772–5,965	Concentrations	No	Locational dummies, population density, trend	GEMS	1977, '82, '88	Up to 52 cities in up to 32 countries
Panayotou (1997)	$5,965	Concentrations	No	Population density, policy variables	GEMS	1982–84	Cities in 30 developed and developing countries
Cole, Rayner, and Bates (1997)	$8,232	Emissions	Yes	Country dummy, technology level	OECD	1970–92	11 OECD countries
Selden and Song (1994)	$10,391–10,620	Emissions	Yes	Population density	WRI—primarily OECD source	1979–87	22 OECD and 8 developing countries
Kaufmann, Davidsdottir, Garnham, and Pauly (1998)	$14,730	Concentrations	Yes	GDP/Area, steel exports/GDP	UN	1974–89	13 developed and 10 developing countries
List and Gallet (1999)	$22,675	Emissions	N/A	–	US EPA	1929–94	US States
Stern and Common (2001)	$101,166	Emissions	Yes	Time and country effects	ASL	1960–90	73 developed and developing countries

econometric behavior). The turning point is at only 54% of maximal GDP in the sample. A study by Schmalensee, Stoker, and Judson (1998) also finds a within sample turning point for carbon. In this case, a 10-piece spline was fitted to the data such that the coefficient estimates for high-income countries are allowed to vary from those for low-income countries. All these studies suggest that the differences in turning points that have been found for different pollutants may be due, at least partly, to the different samples used. I will discuss the econometric reasons for this sample-dependent behavior below.

In an attempt to capture all environmental impacts of whatever type, a number of researchers (e.g., Suri & Chapman, 1998; Cole *et al.*, 1997) have estimated EKCs for a proxy total environmental impact indicator—total energy use. In each case, they found that energy use per capita increases monotonically with income per capita. This result does not preclude the possibility that energy intensity—energy used per dollar of GDP produced—declines with rising income or even follows an inverted U-shaped path (e.g., Galli, 1998).

The only robust conclusions from the EKC literature appear to be that concentrations of pollutants may decline from middle income levels, while emissions tend to be monotonic in income. As we will see below, emissions may decline simultaneously over time in countries at widely varying levels of development. Given the poor statistical properties of most EKC models, it is hard to come to any conclusions about the roles of other additional variables such as trade. Too few quality studies have been done of other indicators apart from air pollution to come to any firm conclusions about those impacts either.

5. THEORETICAL CRITIQUE OF THE EKC

A number of critical surveys of the EKC literature have been published (e.g., Ansuategi, Barbier, & Perrings, 1998; Arrow *et al.*, 1995; Copeland & Taylor (2004); Dasgupta *et al.*, 2002; Ekins, 1997; Pearson, 1994; Stern, 1998; Stern *et al.*, 1996). This section discusses the criticisms raised against the EKC in the earlier surveys on theoretical (rather than methodological) grounds. The more recent surveys raise similar points but have more evidence to marshal.

The key criticism of Arrow *et al.* (1995) and others was that the EKC model, as presented in the *1992 World Development Report* and elsewhere, assumes that there is no feedback from environmental damage to economic production as income is assumed to be an exogenous variable. The assumption is that environmental damage does not reduce economic activity sufficiently to stop the growth process and that any irreversibility is not so severe that it reduces the level of income in the future. In other words, there is an assumption that the economy is sustainable. But, if higher levels of economic activity are not sustainable, attempting to grow fast in the early stages of development when environmental degradation is rising may prove counterproductive. [8]

It is clear that emissions of many pollutants per unit of output have declined over time in developed countries with increasingly stringent environmental regulations and technical innovations. But the mix of residuals has shifted from sulfur and nitrogen oxides to carbon dioxide and solid waste so that aggregate waste is still high and per capita waste may not have declined. [9] Economic activity is inevitably environmentally disruptive in some way. Satisfying the material needs of people requires the use and disturbance of energy flows and materials stocks. Therefore, an effort to reduce some environmental impacts may just aggravate other problems. [10]

Both Arrow *et al.* (1995) and Stern *et al.* (1996) argued that, if there was an EKC type relationship, it might be partly or largely a result of the effects of trade on the distribution of polluting industries. The Hecksher-Ohlin trade theory suggests that, under free trade, developing countries would specialize in the production of goods that are intensive in the factors that they are endowed with in relative abundance: labor and natural resources. The developed countries would specialize in human capital and manufactured capital intensive activities. Part of the reduction in environmental degradation levels in the developed countries and increases in environmental degradation in middle income countries may reflect this specialization (Hettige, Lucas, & Wheeler, 1992; Lucas, Wheeler, & Hettige, 1992; Suri & Chapman, 1998). Environmental regulation in developed countries might further encourage polluting activities to gravitate toward the developing countries (Lucas *et al.*, 1992). These effects would exaggerate any apparent decline in pollution intensity with rising income along the EKC. In our finite world the poor countries of today would be unable to find further countries from which to import resource-intensive products as they, themselves, become wealthy. When the poorer countries apply similar levels of environmental regulation they would face the more difficult task of abating these activities rather than outsourcing them to other countries (Arrow *et al.*, 1995; Stern *et al.*, 1996). Copeland and Taylor (2004) conclude that, in contrast to earlier work (e.g., Jaffe, Peterson, Portney, & Stavins, 1995), recent research shows that increased regulation does tend to result in more decisions to locate in less regulated locations. On the other hand, there is no clear evidence that trade liberalization results in a shift in polluting activities to less-regulated countries.

Furthermore, Antweiler, Copeland, and Taylor (2001) and Cole and Elliott (2003) argue that the capital-intensive activities that are concentrated in the developed countries are more polluting and hence developed countries have a natural comparative advantage in polluting goods in the absence of regulatory differences. There are no clear answers on the impact of trade on pollution from the empirical EKC literature.

Stern *et al.* (1996) argued that early EKC studies showed that a number of indicators: SO_2 emissions, NO_x, and deforestation, peak at income levels around the current world mean per capita income. A cursory glance at the available econometric estimates might have lead one to believe that, given likely future levels of mean income per capita, environmental degradation should decline from the present onward. This interpretation is evident in the 1992 *World Development Report* (IBRD, 1992). Income is not however, normally distributed but very skewed, with much larger numbers of people below mean income per capita than above it. Therefore, it is median rather than mean income that is the relevant variable. Selden and Song (1994) and Stern *et al.* (1996) performed simulations that, assuming that the EKC relationship is valid, showed that global environmental degradation was set to rise for a long time to come. Figure 2 presents projected sulfur emissions using the EKC in Figure 1 and UN and World Bank forecasts of economic and population growth. Despite this and despite recent estimates that indicate higher or nonexistent turning points, the impression produced by the early studies in the policy, academic, and business communities seems slow to fade (e.g., Lomborg, 2001).

6. RECENT DEVELOPMENTS

Significant developments, since my last general survey of the EKC in 1998, fall into three classes: (a) Empirical case study evidence on environmental performance and policy in developing countries that is discussed in this

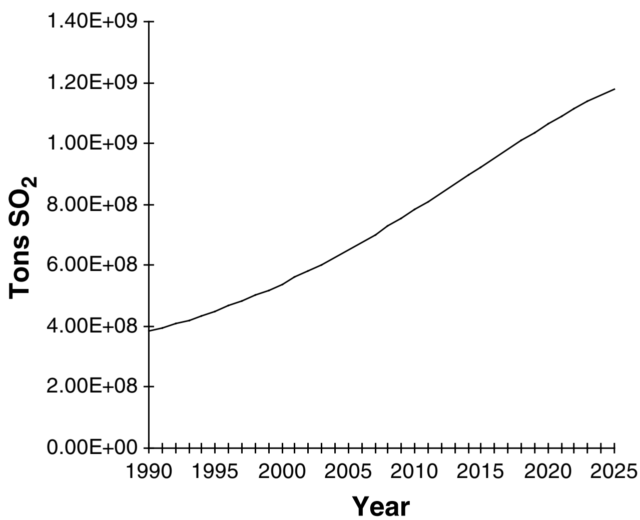


Figure 2. *Projected sulfur emissions. Source: Stern et al. (1996).*

section; (b) improved econometric testing and estimates discussed in the following section; and (c) a new wave in the investigation of environment-development relations using decomposition analysis and efficient frontier methods, discussed in Section 8.

Dasgupta *et al.* (2002) wrote a critical review of the EKC literature and other evidence on the relation between environmental quality and economic development in the *Journal of Economic Perspectives*. Figure 3 illustrates four alternative viewpoints discussed in the article regarding the nature of the emissions and income relation. The conventional EKC needs no further discussion. Two viewpoints argue that the EKC is monotonic. The "new toxics" scenario claims that while some traditional pollutants might have an inverted U-shape curve, the new pollutants that are replacing them do not. These include carcinogenic chemicals, carbon dioxide, etc. As the older pollutants are cleaned up, new ones emerge, so that overall environmental impact is not reduced. The "race to the bottom" scenario posits that emissions were reduced in developed countries by outsourcing dirty production to developing countries. These countries will find it harder to reduce emissions. But the pressure of globalization may also preclude further tightening of environmental regulation in developed countries and may even result in its loosening in the name of competitiveness.

The revised EKC scenario does not reject the inverted U-shape curve but suggests that it is shifting downward and to the left over time due to technological change. But this argument is already present in the *1992 World Development Report* (IBRD, 1992). Dasgupta *et al.* also review the theoretical literature and some of the econometric specification issues. But their main contribution is presenting evidence that environmental improvements are possible in developing countries and that peak levels of environmental degradation will be lower than in countries that developed earlier.

According to Dasgupta *et al.* (2002), regulation of pollution and enforcement increase with income but the greatest increases happen from low to middle income levels and increased regulation is expected to have diminishing returns. There is also informal or decentralized regulation in developing countries—Coasian bargaining. Further, liberalization of developing economies over the last two decades has encouraged more efficient use of inputs and less subsidization of environmentally damaging activities—globalization is in fact good for the environment. The evidence seems to contradict the "race to the bottom" scenario. Multinational companies respond to investor and consumer pressure in their home countries and raise standards in the countries in which they invest. Further, better methods of regulating pollution such as market instruments are having an impact even in developing countries. Better information on pollution is available, encouraging government to regulate and empowering local communities. Those that argue that there is no regulatory capacity in developing countries seem to be wrong.

Much of the Dasgupta *et al.* evidence is from China. Other researchers of environmental and economic developments in China come to similar conclusions. Gallagher (2003) finds that China is adopting European Union standards for pollution emissions from cars with an

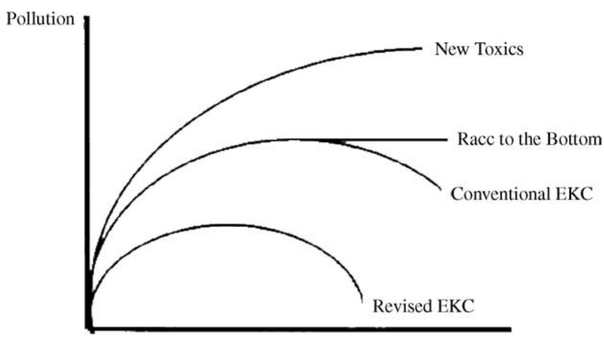


Figure 3. *Environmental Kuznets curve: alternative views. Source: Dasgupta et al. (2002) and Perman and Stern (2003).*

approximately 8–10-year lag. Clearly, China's income per capita is far more than 10 years behind that of Western Europe. Streets *et al.* (2001), Zhang (2000), Jiang and McKibbin (2002), and Wang and Wheeler (2003) all report on substantial reductions of pollution intensities and levels in recent years.

7. ECONOMETRIC CRITIQUE OF THE EKC

Econometric criticisms of the EKC fall into four main categories: heteroskedasticity, simultaneity, omitted variables bias, and cointegration issues.

Stern *et al.* (1996) raised the issue of heteroskedasticity that may be important in the context of regressions of grouped data (see Maddala, 1977). Schmalensee *et al.* (1998) found that regression residuals from OLS were heteroskedastic with smaller residuals associated with countries with higher total GDP and population as predicted by Stern *et al.* (1996). Stern (2002) estimated a decomposition model using feasible GLS. Adjusting for heteroskedasticity in the estimation significantly improved the goodness of fit of globally aggregated fitted emissions to actual emissions.

Cole *et al.* (1997) and Holtz-Eakin and Selden (1995) used Hausman tests for regressor exogeneity to directly address the simultaneity issue. They found no evidence of simultaneity. In any case, simultaneity bias is less serious in models involving integrated variables than in the traditional stationary econometric model (Perman & Stern, 2003). Coondoo and Dinda (2002) test for Granger Causality between CO_2 emissions and income in various individual countries and regions. As the data are differenced to ensure stationarity, this test can only address short-run effects. The overall pattern that emerges is that causality runs from income to emissions or there is no significant relationship in developing countries, while in developed countries causality runs from emissions to income. This suggests that simultaneity is not important.

Stern and Common (2001) use three lines of evidence to suggest that the EKC is an inadequate model and that estimates of the EKC in levels can suffer from significant omitted variables bias: (a) Differences between the parameters of the random-effects and fixed-effects models, tested using the Hausman test; (b) differences between the estimated coefficients in different subsamples, and (c) tests for serial correlation. Table 2 presents the key results from an EKC model estimated with data from 74 countries (in the World sample) over 1960–90. For the non-OECD and World samples, the Hausman test shows a significant difference in the parameter estimates for the random-effects and fixed-effects model. This indicates that the regressors—the level and square of the logarithm of income per capita—are correlated with the country effects and time effects. As these effects model the mean effects of omitted variables that vary across countries or across time, this indicates that the regressors are likely

Table 2. *Stern and Common (2001) key results*

Region	Model	Levels				First differences	
		Turning points	Hausman test	Chow F-test	ρ	Turning points	Mean income elasticity
OECD	FE	\$9,239			0.9109	\$55,481	0.67
	RE	\$9,181	0.3146 (0.8545)		0.9070		
Non-OECD	FE	\$908,178			0.8507	\$18,039	0.50
	RE	\$344,689	14.1904 (0.0008)		0.8574		
World	FE	\$101,166		10.6587 (0.0156)	0.8569	\$33,290	
	RE	\$54,199	10.7873 (0.0045)	4.0256 (0.0399)	0.8624		

All turning points in real 1990 purchasing power parity US dollars.

correlated with omitted variables and the regression coefficients are biased. [11] The OECD results pass this Hausman test but this result turned out to be very sensitive to the exact sample of countries included in the subsample.

As expected, given the Hausman test results, the parameter estimates are dependent on the sample used, with the non-OECD estimates showing a turning point at extremely high-income levels and the OECD estimates a within sample turning point (Table 2). As mentioned above, these results exactly parallel those for developed and developing country samples of carbon emissions. The Chow *F*-test tests whether the two subsamples can be pooled, and therefore that there is a common regression parameter vector, a hypothesis that is rejected.

The parameter ρ is the first order autoregressive coefficient of the regression residuals. This level of serial correlation indicates misspecification either in terms of omitted variables or missing dynamics.

Harbaugh *et al.* (2002) carry out a sensitivity analysis of the original Grossman and Krueger (1995) results. They use an updated and larger version of the ambient pollution data set and test a number of alternative specifications. Using the new extended dataset with Grossman and Krueger's original cubic specification results in the coefficients changing sign and peak and trough levels altering wildly. Altering the specification in various ways—adding explanatory variables, using time dummies instead of a time trend, using logs, removing outliers, and averaging the observations across monitors in each country—changes the shape of the curve. The final experiment they carry out is to include only countries with GDP per capita above $8,000. In contrast to Stern and Common (2001), this results in a monotonic curve. The authors comment:

> This may seem counterintuitive. SO_2 concentrations in Canada and the United States have declined over time at ever decreasing rates... the regressions... include... a linear time trend... after detrending the data with the time function, pollution appears to increase as a function of GDP (p. 548).

There are several differences between the Harbaugh *et al.* (2002) model and the Stern and Common (2001) model that may explain the different results obtained for high income countries. Harbaugh *et al.* (2002) use concentrations data, a linear time trend and a dynamic specification, while Stern and Common (2001) use emissions data, individual time dummies, and a static specification. Stern and Common's (2001) first differences results (Table 2) are very similar to the Harbaugh *et al.*'s (2002) results, which suggests that the dynamic specification could be important.

Millimet, List, and Stengos (2003) use a different strategy to test the robustness of the parametric EKC—comparing it to semi-parametric curves estimated using the same dataset for US states used by List and Gallet (1999). But they claim that parametric models are too pessimistic—finding high turning points—while their alternative semi-parametric models result in U-shaped curves with lower turning points. In addition, they reject the parametric specification in favor of the semi-parametric. But neither parametric nor semi-parametric curves seem to fit the observed data very well in the figures presented in the paper. Furthermore, results for individual states are varied, with the nitrogen dioxide curves mostly rising throughout the income range and many of the sulfur dioxide curves falling—the reverse of the national panel data results. These results could, therefore, be further evidence of the fragility of the EKC rather than evidence for a low turning point semi-parametric specification.

In contrast to Harbaugh *et al.* (2002), Cole (2003b) claims that the EKC model is fairly robust. But his basic levels sulfur emissions EKC has a significant Hausman statistic for a test of whether the random and fixed-effects parameters differ. Adding trade variables to the model results in an insignificant Hausman statistic and a somewhat higher turning point. Using logarithms increases this turning point further (Stern, 2004). The sulfur series cannot be tested for unit roots, but other series in the dataset he uses do show unit root behavior and results using first differences indicate a higher turning point than the levels results. I conclude that the model with trade variables performs better but the basic EKC is misspecified and appropriate econometric techniques appear to raise the turning point.

Perman and Stern (2003) test Stern and Common's (2001) data and models for unit roots and cointegration respectively. Panel unit root tests indicate that all three series—log sulfur emissions per capita, log GDP capita, and its square—have stochastic trends. Results for cointegration are less clear cut. Around half the individual country EKC regressions cointegrate, but many of these have parameters with "incorrect signs." Some panel cointegration tests indicate cointegration in all countries and some accept the noncointegration hypothesis. But even when cointegration is found, the form of the EKC relationship varies radically across countries with many countries having U-shaped EKCs. A common cointegrating vector for all countries is strongly rejected. Koop and Tole (1999) similarly found that random and fixed-effects specifications of a deforestation EKC were strongly rejected in favor of a random coefficients model with widely varying coefficients and insignificant mean coefficients.

In the presence of possible noncointegration, we can estimate a model in first differences. The estimated turning points indicate a largely monotonic EKC relationship and are more similar across subsamples, though the parameters are still significantly different (Table 2). The estimated income elasticity is less than one—there are factors that change with income which offset the scale effect, but they are insufficiently powerful to overcome fully the scale effect.

Figure 4 presents the time effects from the first difference estimates. The OECD saw declining emissions holding income constant over the entire period, though the introduction of the LRTAP agreement in the mid-1980s in Europe resulted in a larger decline. Developing countries saw rising emissions in the 1960s and declining emissions since 1973, *ceteris paribus*. Similarly, Lindmark (2002) uses the Kalman filter to extract a technological change trend for carbon dioxide emissions in Sweden from 1870–1997. This trend had a positive growth rate until about 1970 and a negative growth rate since. [12]

Day and Grafton (2003) test for cointegration of the EKC relation using Canadian time-series data on a number of pollutants using the Engle-Granger and Johansen methods. They fail to reject the noncointegration hypothesis in almost every case. de Bruyn's (2000) time-series Engle-Granger tests for the Netherlands, the United Kingdom, the United States, and West Germany for SO_x, NO_x and CO_2 finds cointegration for the CO_2 EKC in the Netherlands and West Germany, but not in any other case.

Using various lines of evidence, the majority of studies have found the EKC to be a fragile model suffering from severe econometric misspecification. Use of more appropriate methods tends to indicate higher turning points and possibly a monotonic curve for emissions of major pollutants. A better model may result from including additional variables to represent either proximate or underlying causes of change in emissions. I next turn to a consideration of

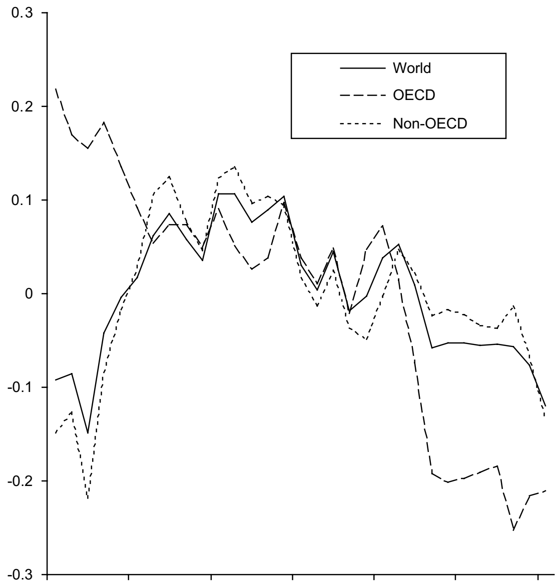


Figure 4. *Time effects: first differences sulfur EKC. Source: Stern and Common (2001).*

Part IV
Policy Responses for Mitigation

The first half of this Review has considered the evidence on the economic impacts of climate change itself, and the economics of stabilising greenhouses in the atmosphere. Parts IV, V and VI now look at the policy response.

The first essential element of climate change policy is carbon pricing. Greenhouse gases are, in economic terms, an externality: those who produce greenhouse gas do not face the full consequences of the costs of their actions themselves. Putting an appropriate price on carbon, through taxes, trading or regulation, means that people pay the full social cost of their actions. This will lead individuals and businesses to switch away from high-carbon goods and services, and to invest in low-carbon alternatives.

But the presence of a range of other market failures and barriers mean that carbon pricing alone is not sufficient. Technology policy, the second element of a climate change strategy, is vital to bring forward the range of low-carbon and high-efficiency technologies that will be needed to make deep emissions cuts. Research and development, demonstration, and market support policies can all help to drive innovation, and motivate a response by the private sector.

Policies to remove the barriers to behavioural change are a third critical element. Opportunities for cost-effective mitigation options are not always taken up, because of a lack of information, the complexity of the choices available, or the upfront cost. Policies on regulation, information and financing are therefore important. And a shared understanding of the nature of climate change and its consequences should be fostered through evidence, education, persuasion and discussion.

The credibility of policies is key; this will need to be built over time. In the transitional period, it is important for governments to consider how to avoid the risks that long-lived investments may be made in high-carbon infrastructure.

Part IV is structured as follows:

- **Chapter 14** looks at the principles of carbon pricing policies, focusing particularly on the difference between taxation and trading approaches.

- **Chapter 15** considers the practical application of carbon pricing, including the importance of credibility and good policy design, and the applicability of policies to different sectors.

- **Chapter 16** discusses the motivation for, and design of, technology policies.

- **Chapter 17** looks at policies aimed at removing barriers to action, particularly in relation to the take-up of opportunities for energy efficiency, and at how policies can help to change preferences and behaviour.

14 Harnessing Markets for Mitigation – the role of taxation and trading

Key Messages

- **Agreeing a quantitative global stabilisation target range for the stock of greenhouse gases (GHGs) in the atmosphere is an important and useful foundation for overall policy.** It is an efficient way to control the **risk** of catastrophic climate change in the long term. Short term policies to achieve emissions reductions will need to be consistent with this long-term stabilisation goal.

- **In the short term, using price-driven instruments (through tax or trading) will allow flexibility in how, where and when emission reductions are made, providing opportunities and incentives to keep down the cost of mitigation.** The price signal should reflect the marginal damage caused by emissions, and rise over time to reflect the increasing damages as the stock of GHGs grows. For efficiency, it should be common across sectors and countries.

- **In theory, taxes or tradable quotas could establish this common price signal across countries and sectors.** There can also be a role for regulation in setting an implicit price where market-based mechanisms alone prove ineffective. In practice, tradable quota systems – such as the EU's emissions-trading scheme – may be the most straightforward way of establishing a common price signal across countries. To promote cost-effectiveness, they also need flexibility in the timing of emissions reductions.

- **Both taxes and tradable quotas have the potential to raise public revenues.** In the case of tradable quotas, this will occur only if some firms pay for allowances (through an auction or sale). Over time, there are good economic reasons for moving towards greater use of auctioning, though the transition must be carefully managed to ensure a robust revenue base.

- **The global distributional impact of climate-change policy is also critical.** Issues of equity are likely to be central to securing agreement on the way forward. Under the existing Kyoto protocol, participating developed countries have agreed binding commitments to reduce emissions. Within such a system, company-level trading schemes such as the EU ETS, which allow emission reductions to be made in the most cost-effective **location** – either within the EU, or elsewhere – can then drive financial flows between countries and promote, in an equitable way, accelerated mitigation in developing countries.

- At the **national – or regional** – level, governments will want to choose a policy framework that is suited to their specific circumstances. Tax policy, tradable quotas and regulation can all play a role. In practice, some administrations are likely to place greater emphasis on trading, others on taxation and possibly some on regulation.

14.1 Introduction

This chapter focuses on the first and key element of a mitigation strategy – how best to ensure GHG emissions are priced to reflect the damage they cause.

This chapter focuses on the principles of policy and, in particular, on the efficiency, equity and public finance implications of tax and tradable quotas. Chapter 15 follows with a detailed discussion of the practical issues associated with the implementation of tax and trading schemes.

Section 14.2 begins by setting out the basic theory of externalities as this applies to climate change. Based on this, Section 14.3 sets out two overarching principles for reducing GHG emissions efficiently. First, abatement should occur just up to the point where the costs of going any further would outweigh the extra benefits. Second, a common price signal is needed across countries and sectors to ensure that emission reductions are delivered in the most cost-effective way.

Section 14.4 explores the policy implications of the significant risks and uncertainties surrounding both the impacts of climate change, and the costs of abatement. It concludes that a long-term quantity ceiling – or stabilisation target – should be used to limit the total stock of GHGs in the atmosphere. In the short term, to keep down the costs of mitigation, the amount of abatement should be driven by a common price signal across countries and sectors, and there should be flexibility in how, where and when reductions are made. Over time, the price signal should trend upwards, as the social cost of carbon is likely to increase as concentrations rise towards the long-term stabilisation goal.

These sections conclude that both taxes and tradable quotas have the potential to deliver emission reductions efficiently. The other key dimensions of climate change policy – tackling market failures that limit the development low carbon technologies, and removing barriers to behavioural change are discussed in Chapter 16 and Chapter 17 respectively.

The penultimate section of the chapter considers the public-finance aspects of taxes and tradable quotas. Finally, Section 14.8 briefly considers the international dimension of carbon-pricing policy. These international issues are treated in greater depth in Part VI of this Review – in particular, the challenge of how national action can be co-ordinated and linked at the international level to support the achievement of a long-run stabilisation goal is considered in Chapter 22.

14.2 Designing policy to reduce the impact of the greenhouse-gas externality

As described in Chapter 2, the climate change problem is an international and intergenerational issue.

Climate change is a far more complicated negative externality than, for example, pollution (such as smog) or congestion (such as traffic jams). Key features of the greenhouse-gas externality are:

- it is a global externality, as the damage from emissions is broadly the same regardless of where they are emitted, but the impacts are likely to fall very unevenly around the world;
- its impacts are not immediately tangible, but are likely to be felt some way into the future. There are significant differences in the short-run and long-run implications of greenhouse-gas emissions. It is the stock of carbon in the atmosphere that drives climate change, rather than the annual flow of emissions. Once released, carbon dioxide remains in the atmosphere for up to 100 years;
- there is uncertainty around the scale and timing of the impacts and about when irreversible damage from emission concentrations will occur;
- the effects are potentially on a massive scale.

These characteristics have implications for the most appropriate policy response to climate change. In the standard theory of externalities[1], there are four ways in which negative externalities can be approached:

- a tax can be introduced so that emitters face the full social cost of their emissions[2] ie. a carbon price can be established that reflects the damage caused by emissions;
- quantity restrictions can limit the volume of emissions, using a 'command and control' approach;

[1] Developed mainly in the first half of the last century.
[2] Pigou (1920) showed how taxes can establish a marginal cost to polluters equal to the marginal damage caused by their pollution.

- a full set of property rights can be allocated among those causing the externality and / or those affected (in this case including future generations), which can underpin bargaining or trading[3];
- a single organisation can be created which brings those causing the externality together with all those affected[4].

In practice, cap-and-trade systems tend to combine aspects of the second and third approach above. They control the overall quantity of emissions, by establishing binding emissions commitments. Within this quantity ceiling, entities covered by the scheme – such as firms, countries or individuals – are then free to choose how best – and where – to deliver emission reductions within the scheme. The largest example of a cap-and-trade scheme for GHG emissions is the EU's Emissions Trading Scheme, and there are a range of other national or regional emissions trading schemes, including the US Regional Greenhouse Gas Initiative and the Chicago Climate Exchange.

The Kyoto Protocol established intergovernmental emissions trading for those countries that took quantified commitments to reduce GHG emissions, as well as other mechanisms to increase the flexibility of trading across all Parties to the Protocol. The Kyoto Protocol and its flexible mechanisms are discussed in detail in Chapter 22.

Whatever approach is taken, the key aim of climate-change policy should be to ensure that those generating GHGs, wherever they may be, face a marginal cost of emissions that reflects the damage they cause. This encourages emitters to invest in alternative, low-carbon technologies, and consumers of GHG-intensive goods and services to change their spending patterns in response to the increase in relative prices.

14.3 Delivering carbon reductions efficiently

Where markets are well functioning, two conditions must hold to reduce GHG emissions efficiently[5]:

- Abatement should take place up to the point where the benefits of further emission reductions are just balanced by the costs. Or – put another way – abatement should occur up to the point where the marginal social cost of carbon is equal to the marginal cost of abatement. This is a necessary condition for choosing the appropriate level of emissions, and hence setting a long-term stabilisation target (and is explained fully in Chapter 13).

- To deliver reductions at least cost, a common price signal is required across countries and different sectors of their economies at a given point in time. For example, if the marginal cost of reduction is lower in country A than in country B, then abatement costs could be reduced by doing a little more reduction in country A, and a little less in country B.

In ideal conditions – perfectly competitive markets, perfect information and certainty, and no transaction costs – both taxes and quantity controls, if correctly designed, can meet these criteria, and be used to establish a common price signal across countries and sectors. Taxes can set the global price of greenhouse gases, and emitters can then choose how much to emit. Alternatively, a total quota (or ceiling) for global emissions can be set and tradable quotas can then determine market prices.[6]

Without market imperfections and uncertainty, and with an appropriate specification of taxes and quotas (entailing an allocation of property rights), both approaches would produce the

[3] Coase (1960)
[4] Meade (1951). This is not discussed further, as it is clearly not a practical option in relation to climate change.
[5] These conditions abstract from uncertainty and market imperfections.
[6] Continuous trading is necessary to ensure a common price between auctions/ allocations.

same price level and quantity of emissions[7]. The remainder of this chapter, and Chapter 15, consider how the considerable uncertainties and imperfections that exist in the real world affect the choice and design of policy.

14.4 Efficiency under uncertainty – the implications for climate-change policy

Substantial uncertainty exists around the timing and scale of impacts, as well as the costs of abatement. In such circumstances, prices and quantity controls are no longer equivalent and policy instruments will need to be chosen with care to reduce GHG emissions efficiently.

Weitzman (1974) examined how price (here tax) and quota or quantity-control instruments compare where there is uncertainty about the costs and benefits of action, and how this affects the comparative efficiency of the two instruments[8]. A price instrument sets a price for a required service or good and lets markets determine its supply. In contrast, a quota instrument specifies a particular level of supply. Applying the Weitzman analysis to pollution:

- Prices are preferable where the benefits of making further reductions in pollution change less with the level of pollution than do the costs of delivering these reductions i.e. when the marginal damage curve – or the marginal social cost of carbon - is relatively flat, compared with the marginal abatement cost curve, as pollution rises.
- Quantity controls are preferable where the benefits of further reductions increase more with the level of pollution than do the costs of delivering these reductions i.e. there are potentially large and sharply rising costs associated with exceeding a given level of pollution.

Box 14.1 sets out these economic arguments in detail[9].

Box 14.1 Prices versus quantities in the short term and long term.

Figure (A) illustrates how Weitzman's analysis is applied in the climate-change case. If the emissions reductions are measured over a short period, say a year, the expected marginal benefits of abatement are flat or gently decreasing as the quantity of emission reduction increases (from left to right). This reflects the fact that variations in emissions in any single year are unlikely to have a significant effect on the ultimate stock of greenhouse gases. The expected marginal costs of abatement (MAC_E), however, are steeply increasing as abatement activity intensifies; firms find it progressively more difficult to reduce emissions, unless they can adjust their capital stock and choice of technology (assumed by definition to be impossible in the short term).

If it were known with certainty that the marginal costs of abatement were given by the schedule MAC_E, the policy-maker should set the rate of the emission tax to equal T_E, given by the intersection of the schedule with the marginal benefits of abatement, also assumed to be known. The optimal quantity of emission quotas or allowances allocated (Q_E) would also be given by this intersection, giving rise to an equilibrium price in a perfectly competitive allowance market of P_E. The choice of quota or tax would not matter in this case.

However, following the exposition in Hepburn (2006), suppose that the real marginal costs of abatement in the period are not known with certainty in advance and turn out to be higher at every point, as represented by the curve MAC_{REAL}, and that the policy-maker cannot adjust the policy instrument in anticipation. In this case, the optimal quantity of allowances to be allocated would in fact turn out to have been Q_{REAL}. In Figure 14.1, the efficiency loss caused by issuing Q_E instead of Q_{REAL} allowances is given by the large blue triangle. If instead a tax had been set at T_E, the efficiency loss resulting from having set a slightly lower

[7] But it is worth noting that even if these ideal conditions were to hold, the nature of the climate-change problem means there are limitations to the applicability of some of the policy options set out above. In particular, a **full set** of property rights cannot be allocated, because many of those affected by the impacts of climate change are yet to be born. It is not possible for them to bargain with the current emitters for the impacts that they will have to endure.
[8] Weitzman (1974)
[9] This box draws on the exposition in Hepburn (2006).

tax rate than turns out to have been warranted is given by the small red triangle. Thus it is often argued that a tax is superior to a quota as an instrument of climate-change policy[10] in the short run. As Chapter 2 explains, however, diagrams like that in Figure (A) need to be interpreted with great care, as the positions of both the curves may depend on policy settings in earlier and later periods.

(A) The efficiency of taxes and tradable allowances in climate-change mitigation in the short term.

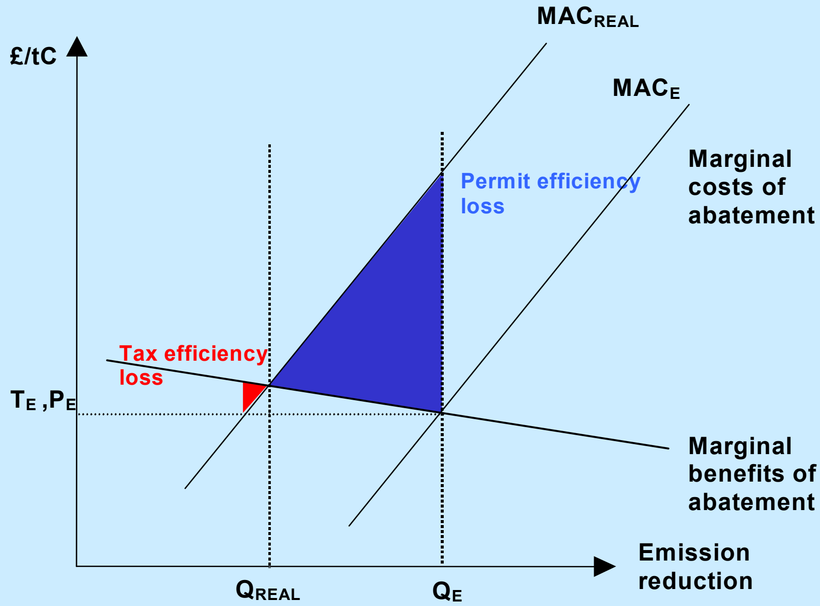
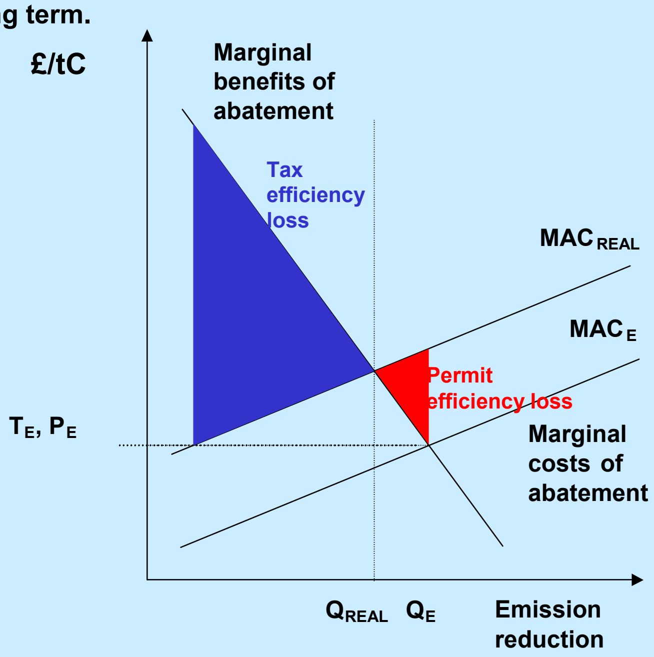


Figure (B) illustrates the situation in the long term, with the cumulative emissions reductions required to reach the ultimate stabilisation target on the x-axis now, instead of annual emissions reductions as in Figure (A). The curve representing the marginal benefits of abatement is steeply decreasing, as more and more abatement effort is put in (put another way, the costs of the impacts of climate change increase steeply as cumulative emissions increase). But the marginal costs of abatement are only gently increasing as a function of abatement effort, since in the long run there is more flexibility. In the certainty case with MAC_E as the true cost of abatement curve, Q_E is the appropriate cumulative quota, while T_E is the equivalent tax[11]. But if MAC_E represents the expected costs of abatement and MAC_{REAL} the higher ex post actual costs, the efficiency loss implied by setting the tax at T_E (the blue triangle) is now much larger than that implied by setting the quantity of tradable allowances at Q_E. Of course, if the policy-maker is able to revise the tax or quota schedule as information comes in about the marginal abatement costs function, s/he can do better than keeping either schedule fixed.

[10] The direct allocation of non-tradable allowances requires information about relative costs across firms, as well as total costs, and so is likely to be even less efficient, given the uncertainties in the real world, than promoting perfect competition in the market for allowances.
[11] Strictly, there is an intertemporal tax schedule that generates cumulative emissions reductions Q_E

(B) The efficiency of taxes and tradable allowances in climate-change mitigation in the long term.



This contrast between short-term and long-term marginal cost and marginal benefit curves gives rise to the problem of how to combine a tax-like regime in the short term with a quantitative constraint in the long term. A rule is needed for updating the tax in the light of new information about costs over the long term and the ex post quantity of emissions.

In the case of climate change, these arguments indicate that the most efficient instrument – over a particular time horizon – will depend on:

- how the total costs of abatement change with the level of emissions;
- how the total benefits of abatement change with the level of emissions;
- the degree of uncertainty about both costs and benefits of abatement.

Chapter 8 explains that it is the total stock of GHGs in the atmosphere that drives the damage from climate change. In economic terms, this means that the marginal damage associated with emitting one more unit of carbon is likely to be more or less constant over short periods of time. Thus, in the short-term, the marginal damage curve is likely to be fairly flat. But over the long term, as the stock of GHGs grows, marginal damages are likely to rise and – as the stock reaches critical levels – marginal damages may rise sharply. In other words, the damage function is likely to be strongly convex (as discussed in Part Two and Chapter 13)[12].

On the other side of the equation, many uncertainties remain about the marginal costs of abatement. Many new technologies that could be used to reduce carbon emissions are not yet in widespread use. Trying to abate rapidly in the short term – when the capital in industries emitting greenhouse gases is fixed and technologies are given – can quickly become costly for firms, as the marginal cost of abatement is likely to rise sharply[13]. In particular, if the costs of abatement prove to be unexpectedly high, then setting a fixed quantity target in the short term could prove unexpectedly costly. Over the long term – as the capital stock is replaced and new lower-carbon technologies become available – the

[12] To the extent that damages may relate to the *rate* of climate change, the relationship is more complex, but it remains true that the damage curve is likely to respond most to cumulative emissions over several years or even decades.

[13] For a discussion of the relative abatement costs and marginal benefits of climate change see, for example, Lydon (2002) and Pizer (2002). Both conclude that the marginal damage curve is relatively flat – at least in the short term – and, as such, there are strong arguments for flexibility in the quantity of abatement in the short term, subject to a fixed carbon price.

marginal costs of abating in the long term are likely to be broadly flat, or, put another way, bounded relative to incomes. The implications are explained more fully in Box 14.1.

These characteristics of the costs and benefits of abatement and damage from emissions suggest three things:

- Policy instruments should distinguish between the short term and long term, ensuring that short-term policy outcomes are consistent with achieving long-term goals[14];
- The policy-maker should have a clear long-term goal for stabilising concentrations of greenhouse gases in the atmosphere. This reflects, first, the likelihood that marginal damages (relative to incomes) will accelerate as cumulative emissions rise and, second, that the marginal costs of abatement (relative to incomes) are likely to be relatively flat in the long term once new technologies are available.
- In the short term, the policy-maker will want to choose a flexible approach[15] to achieving this long-term goal, reflecting the likelihood that marginal damages will be more or less constant, and there will be risks of sharply rising costs from forcing abatement too rapidly.

In practical terms, this means that a long-term stabilisation target should be used to establish a quantity ceiling to limit the total stock of carbon over time. Short-term policies (based on tax, trading or in some circumstances regulation) will then need to be consistent with this long-term stabilisation goal. In the short term, the amount of abatement should be driven by a common price signal across countries and sectors, and should not be rigidly fixed[16].

This common price signal could – in principle – be delivered through taxation or tradable quotas. A country can levy taxes without consultation with another, but harmonisation requires agreement. In practice, therefore, it may prove difficult to use taxes to deliver a common price signal in the absence of political commitment to move towards a harmonised carbon tax across different countries. In contrast, to the extent that a tradable quota scheme embraces both different countries and sectors, it may be an effective way of delivering a consistent price signal across a wide area – though this, of course, requires agreement on the mechanics of the scheme. International co-ordination issues are fully discussed in Chapter 22 – here it is sufficient to note that building consensus on the best way forward will be critical to achieving a long-run stabilisation goal.

14.5 Setting short term policies to meet the long term goal

The key question that arises from the previous section is how to combine a price instrument that allows flexibility about where, when and what emissions are reduced in the short term, with a long-term quantity constraint. In particular, the challenge is how to ensure that the short-term policy framework remains on track to deliver the long-term stabilisation goal.

There are two important aspects to this:

- having established the long-term stabilisation goal, the price of carbon is likely to rise over time, because the damage caused by further emissions at the margin-the social cost of carbon- is likely to increase as concentrations rise towards this agreed long-term quantity constraint;

- short-term tax or trading policies will then need to be consistent with delivering this long-term quantitative goal.

In the short-term, applying these principles to tax and trading, this means that:

[14] The short term is defined as the period during which the capital stock is essentially fixed. This will vary from sector to sector.
[15] With respect to the size of emission reductions.
[16] One option is to combine price controls within a quota trading system in the short term. This is discussed more in Chapter 15.

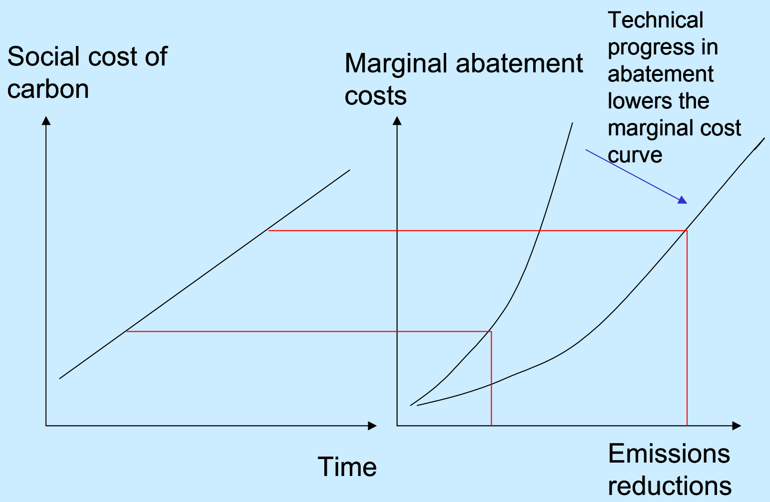
- In a tax-based regime, the tax should be set to reflect the marginal damage caused by emissions. Abatement should then occur up to the point where the marginal cost of abatement is equal to this tax. See Box 14.2.

- In a tradable-quota scheme, the parameters of the scheme – notably the total quota allocation – should be set with a view to generating a market price that is consistent with the social cost of carbon (SCC). In practice – and within the time period between allocations in a tradable-quota system – the market price may be higher or lower than the SCC. This is because the actual market price will reflect both the quota-driven demand for carbon reductions and the marginal cost of delivering reductions in the most cost-effective location. Ex-post, the trading period will therefore deliver abatement up to where the marginal abatement cost equals the actual market price.

In the case of either tax or trading, clear revision rules are therefore necessary to ensure that short-term policies remain on track to meet the long-term stabilisation goal. In particular, the short-term policy framework should be able to take systematic account of the latest scientific information on climate change, as well as improved understanding of abatement costs.

The framework within which any principles for revisions apply must be clear, credible, predictable and set over long time horizons, say 20 years, with regular points, say every five years, to review new evidence, analysis and information[17]. Chapter 22 discusses the challenge of achieving this at an international level.

Box 14.2: The social cost of carbon and the carbon price



Up to the long-term stabilisation goal, the social cost of carbon will rise over time, because marginal damage costs also rise. This is because atmospheric concentrations are expected to rise, so that temperatures are likely to rise; marginal damage costs are expected to rise with temperature. These effects are assumed to outweigh the declining marginal impact of the stock of gases on global temperature at higher temperatures.

As GHG concentrations move towards the stabilisation goal, the price of carbon should reflect the social cost of carbon. In any given year, abatement should then occur up to where the marginal cost is equal to this price, as set out in the right-hand part of the diagram above. If, over time, technical progress reduces the marginal cost of abatement, then at any given price level there should be more emission reductions.

[17]Newell et al (2005)

Revision rules for climate-change policies can be compared to setting interest rates within a well-specified inflation-targeting regime[18]. The stabilisation target is analogous to the inflation target. In the UK, the Monetary Policy Committee each month sets a short-term policy instrument, the interest rate on central-bank money, until their next meeting, in order to keep inflation on track to hit its target. The analogy with climate-change policy would be the setting of a tax rate or an emissions trading quota for, say, a five year period, with firms and households making their own decisions about emissions reductions subject to that carbon-price path and their expectations about policy-makers' commitment to the long-term stabilisation goal.

The analogy is not, however, exact. First, there is widespread agreement about the appropriate long-term goal for monetary policy – price stability, which corresponds to a small positive measured inflation rate. In the climate-change case, there is not yet agreement about the stabilisation level at which that stability should be achieved. Second, the stabilisation objective is likely to have to be revised intermittently – possibly by a large amount – to reflect improved scientific and economic understanding of the climate-change problem, whereas the definition of price stability in terms of a specific inflation measure is less problematic. And third, the locus of decision-making in monetary policy clearly lies with the monetary authority of the country for which the inflation rate is measured, whereas climate change requires international collective action.

Nevertheless, the comparison with an inflation-targeting regime draws attention to the importance of building the credibility of policy-makers. This requires clarity about the ultimate objective of policy and giving policy-makers control over an instrument that can change private-sector behaviour. It also means announcing the principles governing changes in the policy instrument in advance, giving policy-makers incentives to keep aiming at the ultimate target, and holding policy-makers accountable for their actions.

14.6 The interaction between carbon pricing and fossil fuel markets

Imperfections in the markets for exhaustible resources and energy could have important interactions with carbon-pricing policy that should also be considered.

Carbon emissions come from energy production and use across various sectors (see Chapter 7). Much of this energy is generated using exhaustible resources such as oil. In the face of climate change policy, the owners of the natural resource may be willing to reduce producer prices substantially in order to sell off the commodity before it becomes obsolete or of a much lower value. Thus any carbon-pricing policy would need to be carefully designed to ensure it does not accelerate the pace with which carbon-intensive exhaustible resources are used up. The policy implications of this – as well as market imperfections more generally – are explored in Box 14.3.

[18] This analogy has been explored by Helm et al (2005).

Box 14.3 Efficiency market structure and exhaustible resources

Energy and related markets have pervasive market imperfections that will affect the efficiency of a given policy instrument[19]. For example, the collusive behaviour of the OPEC cartel can make it difficult to predict what the final impact on market prices will be from either a tax or a quota-driven carbon price. Thus, on the one hand, OPEC might respond to a carbon tax by further restricting supply, pushing up producer prices and retaining most of their rents. On the other hand, they may choose to retain market share and extract a lower rent[20] with little change in carbon emissions[21].

Where the input prices concerned relate to fossil fuels, the policy must also take account of the fact that such fuels are exhaustible natural resources. Prices to consumers will reflect both the marginal costs of extraction and a scarcity rent (which reflects the stock of the natural resource relative to the expected demand schedule over time). In these circumstances, attempts to reduce carbon emissions through tax measures (imposing the social cost on polluters) may simply lead to a fall in producer prices, with little change in consumption and therefore carbon emissions. In some models, the incidence of the tax would fall wholly on the resource owner's rent. For the same reason, the introduction of new renewable-energy technologies may simply accelerate the use of carbon-intensive energy sources[22] – as the owners of the natural resource try to sell them off before they become obsolete or fall sharply in value. In these circumstances – for some market structures, and in the absence of carbon capture and storage – optimal tax theory can suggest that a declining ad valorem[23] tax rate over time may eventually be desirable, to delay fossil-fuel consumption and push back in time the impacts of climate change[24]. In this case, the tax rate through time reflects more than the social cost of carbon, as it is also takes account of these other market dynamics. The key point here is that there are many complexities that should be considered.[25]

Under a tradable quota system, the price associated with an emissions quota may be much higher than expected if exhaustible-resource pricing is ignored. In effect, rent may be transferred from the owners of fossil fuels to the owners of the allowances (or issuers, if allowances are auctioned). More generally, if trading creates rents, it may undermine the acceptability of policy and lead to gaming, wasted resources in rent-seeking, and possibly corruption. Where incumbent firms enjoy rents, they may also discourage competition and new entry.

14.7 Public finance issues

Both taxes and tradable quotas can be used to raise public funds. Carbon taxes automatically raise public revenues, but tradable-quota systems only have the potential to raise public revenue if firms have to purchase the quotas from government through a sale or auction.

Carbon taxes automatically transfer funds from emitting industries to the public revenue. This transfer may be used to:

- enhance the revenue base[26];
- limit the overall tax burden on the industry affected through revenue recycling[27];

[19] See Blyth and Hamilton (2006) for background discussion on the nature of electricity markets, interaction with fossil fuel markets and issues to consider for policy approaches to introducing climate policy to electricity systems.
[20] This would shift rents from OPEC to Kyoto countries.
[21] Hepburn (2006)
[22] The economic theory of exhaustible natural resources is exposited in Hotelling (1931) and Dasgupta & Heal (1979).
[23] Ad valorem taxes are based on the value or price of a good or service. The alternative to ad-valorem taxation is a fixed-rate tax, where the tax base is the quantity of something, regardless of its price.
[24] There is a debate about whether the tax rate should first rise and then fall. See Ulph & Ulph (1994) and Sinclair (1994).
[25] For a more detailed discussion, see Newbery (2005).
[26] In practice, the overall impact on the revenue base may be limited, if taxes are reduced elsewhere in the economy.

- reduce taxes elsewhere in the economy;

Revenue recycling to the industry can encourage emitters to reduce GHG emissions, without increasing their overall tax burden relative to other parts of the economy[28]. The advantage of this approach is that it can ease the initial impact of the scheme for those industries facing the greatest increase in costs, and therefore ease the transition where carbon taxes are introduced. As the introduction of carbon pricing through taxation is a change to the rules of the game (which will affect shareholders in the short run), there is a case for some transitional arrangements. Over time, however, recycling may discourage or slow the necessary exit of firms from the polluting sectors. Monitoring and protecting the position of incumbents in this way could also reduce competition.

Alternatively, revenue from carbon taxes can be used to reduce taxes elsewhere in the economy. In such circumstances, the revenue from the carbon tax is sometimes argued to generate a so-called 'double dividend' by allowing other distortionary taxes to be reduced.

But this argument needs some care. There is no doubt that environmental taxes have the special virtue of reducing 'public bads', at the same time as they generate revenue. Reducing the 'bad' is indeed central to any assessment of this type of tax. But arguments invoking the so-called 'double dividend' as sometimes advanced in general terms (i.e. that there is always a double dividend), can be incorrect. Putting the reduced public bad to one side for a moment, there is a 'dead-weight' loss to the economy from raising any tax on the margin. Whether it is greater or less with goods associated with carbon (compared with other goods or services) is unclear and depends on the circumstances. For example, where energy is subsidised, reducing the subsidy (equivalent to raising the tax) will probably be a gain in terms of reducing deadweight losses. Note, however, that where other taxes have been optimally set - and abstracting from the externality – then the deadweight loss on the margin from increasing any one tax will be exactly the same as the loss on another and there will clearly be no 'double dividend' in this context.

This is not an argument against raising revenue through pricing GHG emissions. On the contrary: there are strong benefits from ensuring that GHG emissions are properly priced to reflect the damage they cause. Thus GHG taxes have the clear additional benefit relative to other ways of raising revenue of reducing a 'bad'. Where that benefit has not been adequately recognised, they will be underused relative to other forms of taxation.

In contrast, a quota-based system will not automatically raise revenue unless firms must initially purchase some or all quotas from the government in either an auction or a direct sale. In contrast, if quotas are allocated for free, then the asset is passed to the private sector and the benefits ultimately accrue to the owners and shareholders of the firms involved[29]. In the short term, there may be reasons for introducing auctioning slowly – to ease the transition to a new policy environment. Equally, finance ministries will want to ensure that the overall tax revenue base is reliable and predictable: revenues from auctioning may be less predictable than those from taxation. In the long term, however, there is little economic justification for such transfers from the public sector to individual firms and their shareholders[30].

Free allocation of quotas to business also has a number of other potential drawbacks. These are discussed in more detail in the next chapter, which focuses on practical issues associated with the implementation of tax and trading schemes.

In summary, a tax-based approach will automatically generate public revenues, whereas a tradable-quota approach will only generate revenues if quotas are sold. Requiring firms to pay for the right to pollute is consistent with a move to raise revenue via the taxation of 'bads'

[27] The ultimate incidence of the tax is on the industries' customers and – in the absence of perfect competition – shareholders.
[28] Although, as already noted, in a competitive industry the tax will ultimately fall on the consumer.
[29] To the extent that firms are able to pass on to consumers the increase in marginal production costs, a system with free quotas may be regressive (because shareholders tend to be wealthier than the general population).
[30] Where the ultimate incidence of the tax falls on customers, they pay a price of carbon, but there is no benefit to the wider revenue base.

rather than 'goods'[31]. In the case of climate change, where understanding of the potential damage caused by emissions continues to improve, there is a strong argument for shifting the balance of taxation. In the case of tradable quotas, there are good economic reasons for moving towards greater use of auctioning over time, though the transition will need to be carefully managed **–** in particular, to ensure a robust revenue base.

14.8 Co-ordinating action across countries

The mitigation of climate change requires co-ordinated action across different countries. In thinking about the differences between tax and tradable quotas, it is therefore important to recognise the different implications they have for market-driven financial flows between countries.

Chapter 22 will explore the challenges in building up broadly similar price signals for carbon around the world. Issues of equity – as discussed in Chapter 2 – are likely to be central to creating frameworks that support this goal. It is therefore important to consider how taxes and tradable quota systems may differ in the relative ease with which they can drive financial flows between countries.

In theory, either a tax or a tradable quota system could drive financial flows from the developed to developing countries. Under a tax-based system, revenues raised will in the first instance flow to national governments. An additional mechanism would need to be put in place to transfer resources to developing countries.

Under a tradable-quota system, there are a number of ways that governments in rich countries can drive flows, either through direct purchase of quotas allocated to developing countries or through the creation of company-level trading where companies have access to credits for emissions reductions created in developing countries. In this case, financial flows between sectors and/or countries can occur automatically as carbon emitters search for the most cost-effective way of reducing emissions. The opportunities and challenges in these areas are discussed in detail in Chapters 22 and 23.

In summary, financial flows from developed to developing countries can occur under either a tax or tradable-quota system. However, market-driven financial flows will only occur automatically under the latter route, and only at sufficient scale if national quotas are set appropriately.

14.9 The performance of taxation and trading against principles of efficiency, equity and public finance considerations

In terms of the criteria discussed above – efficiency, equity and public finance – carbon taxes perform well against the efficiency and public finance criteria, as they:

- can contribute to establishing a consistent price signal across regions and sectors. However, this may prove difficult if a country perceives that it is acting in isolation, and – as discussed in chapter 22 – there are many reasons why achieving a common price signal through harmonising taxes across countries is likely to be difficult to achieve;
- raise public revenues;
- can be kept stable, and thus do not risk fluctuations in the marginal costs that could increase the total costs of mitigation policy.

[31] Were auctioning to substitute in whole or in part for taxation, it would be important to manage the revenue base to underpin the sustainability of the public finances.

However,

- they do not automatically generate financial flows to developing countries in search of the most efficient carbon reductions.

In terms of the criteria discussed above – efficiency, equity and the impact on public finances – the strengths of a tradable quota scheme are:

- to the extent that the scheme embraces different sectors and countries, it will establish a common price signal and therefore have the potential to drive carbon reductions efficiently;

- to the extent that inter-country trading is allowed, it will ensure carbon reductions are made in the most cost-effective location, and automatically drive private-sector financial flows between regions;

- if allowances are sold or auctioned, then the scheme also has the potential to generate public revenues.

Some countries may make substantial use of tax measures to reduce GHG emissions. Others may place greater emphasis on participation in emissions trading schemes or, indeed, regulation. Some countries may choose a mix of all three depending on the sector, other policies, market structures, and political and constitutional opportunities and constraints.

The effectiveness of any tax or emissions trading scheme depends on its credibility and on good design. Investors need a credible and predictable policy framework on which to base their investment decisions; and good design is important to ensure effectiveness and efficiency. This is discussed in detail in the next chapter.

Carbon-pricing policy is only one element of a policy response to climate change. There are a range of other market failures and barriers to action which must be tackled. For this reason, carbon pricing policy should sit alongside technology policies, and policies to remove the behavioural barriers to action. These two further objectives are discussed in Chapter 16 and Chapter 17 respectively.

14.10 Conclusion – building policies for the future

A shared understanding of the long-term goals for stabilisation is a crucial guide to climate change policy-making: it narrows down strongly the range of acceptable emissions paths, and establishes a long-term goal for policy. But, from year to year, flexibility in when, where and how reductions are made will reduce the costs of meeting these goals. Policies should adapt to changing circumstances as the costs and benefits of climate change become clearer over time. This means that short-term policy may be revised periodically to take account of information, as and when it comes, so as to keep on track towards meeting a long-term goal.

This need for both a long-term goal, and consistent short-term policy to meet this, should guide action at the international and national level to price carbon.

At the international level, this means that the key policy objectives for tackling climate change should include:

- Choosing a policy regime that:

  i. in the long term, will stabilise the concentration of greenhouse gases in the atmosphere, and establish a long-term quantity goal to limit the risk of catastrophic damage;
  ii. in the short term, uses a price signal (tax or trading) to drive emission reductions, thus avoiding unexpectedly high abatement costs by setting short-term quantity constraints that are too rigid.

- Establishing a consistent price signal across countries and sectors to reduce GHG emissions. This price signal should reflect the damage caused by carbon emissions.

In theory, either taxes or tradable quotas – and in specific circumstances regulation – can play a role in establishing a common price signal. Chapter 22 discusses the potential difficulties of co-ordinating national policies to achieve this.

Both taxes and tradable quotas can contribute to raising public revenues. Under a tradable quota scheme, this depends on using a degree of auctioning and, over time, there are sound economic reasons for doing so. However, this would need to be well managed, understanding fully the implications for governments' revenue flows, and ensuring that these remain predictable and reliable.

Taxes and tradable quotas can both support the financing of carbon reductions across different countries. However, only a tradable-quota system will do this automatically, provided there is an appropriate initial distribution of quotas and structure of rules.

At the national – or regional level – governments will want to tailor a package of measures that suits their specific circumstances, including the existing tax and governance system, participation in regional initiatives to reduce emissions (eg. via trading schemes), and the structure of the economy and characteristics of specific sectors.

Some may choose to focus on regional trading initiatives, others on taxation and others may make greater use of regulation. The factors influencing this choice are discussed in the following chapter.

15 Carbon Pricing and Emissions Markets in Practice

Key Messages

Both tax and trading can be used to create an explicit **price for carbon;** and regulation can create an implicit price.

For all these instruments, **credibility, flexibility and predictability are vital to effective policy design.**

A lack of credible policy may undermine the effectiveness of carbon pricing, as well as creating uncertainties for firms considering large, long-term investments.

To establish the credibility of carbon pricing globally will take time. During the transition period, governments should consider how to deal with investments in long-lived assets which risk locking economies into a high-carbon trajectory.

To reap the benefits of emissions trading, deep and liquid markets and well designed rules are important. Broadening the scope of schemes will tend to lower costs and reduce volatility. Increasing the use of auctioning is likely to have benefits for efficiency, distribution and potentially the public finances.

Decisions made now on the third phase of the EU Emissions Trading Scheme pose an opportunity for the scheme to influence, and be the nucleus of, future global carbon markets.

The establishment of common incentives across different sectors is important for efficiency. The overall structure of incentives, however, will reflect other market failures and complexities within the sectors concerned, as well as the climate change externality.

The characteristics of different sectors will influence the design and choice of policy tool. Transaction costs of a trading scheme, for instance, will tend to be higher in sectors where there are many emission sources. The existing framework of national policies in these sectors will be an important influence on policy choice.

15.1 Introduction

This chapter considers how markets for emission reductions can be built on the principles considered in Chapter 14. The application of these principles requires careful analysis of the context of specific economies and institutional structures– at the national, international, regional or sectoral levels.

Section 15.2 discusses the importance of designing policies in a way which creates confidence in the future existence of a robust carbon price, so that businesses and individuals can plan their investment decisions accordingly. The current use of emissions trading schemes is discussed in Section 15.3, and 15.4 focuses particularly on the issues around creating a credible carbon price in emissions trading schemes.

The choice and design of such policy instruments also depends on the specific sectoral context. Policies which work for one sector may be inappropriate for another, although a common price is still needed across sectors for efficiency in the costs of mitigation. The relationship between climate change policy and other objectives, such as energy security and local air pollution, is also important. These issues are discussed in 15.5.

Carbon pricing is only one part of a strategy to tackle climate change. It must be complemented by measures to support the development of technologies, and to remove the barriers to behavioural change, particularly around take-up of energy efficiency. These two elements are discussed in Chapters 16 and 17.

15.2 Carbon pricing and investment decisions

Investors need a predictable carbon policy

Businesses always have to take uncertainties into account when making investment decisions. Factors such as the future oil price, changes in consumer demand, and even the weather can affect the future profitability of an investment. Business decision-makers make judgements on how these factors are likely to evolve over time.

But unlike many other uncertainties that firms face, climate change policy is created solely by governments. To be successful, a carbon pricing policy must therefore be based on a framework that enables investors to have confidence that carbon policy will be maintained over sequential periods into the future.

Serious doubt over the future viability of a policy, or its stringency, risks imposing costs without having a significant impact on behaviour, so increasing the cost of mitigation. Creating an expectation that a policy is very likely to be sustained over a long period is critical to its effectiveness.

Credibility, flexibility and predictability are key to effective policy

Three essential elements for an effective policy framework are credibility (belief that the policy will endure, and be enforced); flexibility (the ability to change the policy in response to new information and changing circumstances); and predictability (setting out the circumstances and procedures under which the policy will change). These apply to any type of policy, including the technology and regulatory measures set out in the following chapters, but are particularly pertinent to carbon pricing.

A key issue for credibility is whether the policy commands support from a range of interest groups. Public opinion is particularly important: sustained pressure from the public for action on climate change gives politicians the confidence to take measures which they might otherwise deem too risky or unpopular. It must also make sense within an international context: if there are good prospects for a robust international framework, this will greatly enhance the credibility of national goals for emissions reductions.

As Chapter 14 has discussed, the flexibility to adjust policy in the short term is an important principle for efficient pricing under conditions of uncertainty. Policy must be robust to changing circumstances and changing knowledge. If policy is seen to be excessively rigid, its credibility may suffer, as people perceive a risk that it will be dropped altogether if circumstances change.

Building in predictable and transparent revision rules from the start is the best way to maintain confidence in the policy, whilst also allowing flexibility in its application.

Issues of credibility are particularly important for investments in long-lived capital stock

Taking a long-term view on the carbon price is particularly important for businesses investing in long-lived assets[1]. Assets such as power stations, industrial plant and buildings last for many decades, and businesses making investment decisions on these assets often have longer time horizons than many governments.

If businesses believe that carbon prices will rise in the long run to match the damage costs of emissions over time, this should lead them to invest in low-carbon rather than high-carbon assets. But in the transitional period, where the credibility of carbon pricing is being

[1] See Helm et al (2005) which argues that credibility problems in recent UK energy and carbon policy have costs for meeting objectives on energy and climate change. The irreversibility of energy investments and the risk of governments reneging on commitments to carbon commitments imply a need for a more consistent policy framework.

established worldwide, there is a risk that future carbon prices are not properly factored into business decision-making, and investments may be made in long-lived, high-carbon assets.

This could lock economies into a high-carbon trajectory, making future mitigation efforts more expensive. Governments should take careful account of this: as well as providing as much clarity as possible about future carbon pricing policies, they should also consider whether any additional measures may be justified to reduce the risks[2].

Uncertainty about the long-term future framework for carbon pricing is also a reason why additional measures to encourage the development of low-carbon technologies are important. This is discussed in Chapter 16.

Policy uncertainty not only undermines climate change policy – it can also undermine security of supply, by creating an incentive to delay investment decisions.

Uncertainty about the future existence or overall direction of policy creates difficulties for how businesses respond. There is a risk that businesses will adopt a 'wait and see' attitude, delaying their investment decisions until the policy direction becomes clearer.

Blyth and Yang (2006) look at the incentives for a company faced with a decision on whether to invest in high-carbon or low-carbon infrastructure. If a decision is expected at some point in the future about whether or not a new climate change policy will be introduced, a company which makes its investment decision now, risks a loss later if it makes the wrong call on policy. If it waits until the policy is agreed, it can make a more informed choice. Given this uncertainty, a much higher expected profit level would be required to trigger the investment now[3].

In the energy sector, such delays in investment could create serious problems for a country's security of supply. Modelling work by Blyth and Yang (2006) indicates that an increase in the period of relative carbon price stability from 5 to 10 years (which could equate to increasing the length of an allocation period in a trading scheme) could reduce the size of the investment thresholds arising from uncertainty by a factor of 2 or more[4].

Credibility may also vary between policy instruments

Credibility may vary between different types of policy instrument. For instance, taxation provides governments with a revenue stream, and there tends to be an expectation that it will not be in a government's interests to abolish it. Regulation may be more effective in countries with a culture of using command and control methods, or where there are political or administrative problems with raising taxes or with tax collection. Specific national circumstances, including constitutional structures, the stability of political institutions and the quality of legal infrastructures and enforcement, play a key role in determining what credible policy is.

Another important element is the level at which policy takes place. Regulation or trading schemes which are agreed at the EU level, for instance, are difficult to reverse, and hence may be seen as more credible than some national policies.

The issues surrounding credibility in trading schemes are discussed in detail in the following section.

15.3 Experience in emissions trading

As outlined in Chapter 14, emissions trading has several benefits. Emissions trading schemes can deliver least-cost emission reductions by allowing reductions to occur wherever they are cheapest. A key corollary benefit to this is that it generates automatic transfers between

[2] Grubb et al (1995), Lecocq et al (1998).
[3] See Blyth and Sullivan (2006)
[4] See Blyth and Yang (2006)

countries, while delivering the least-cost reductions. In many instances, introducing trading schemes is also an easier mechanism through which to achieve a common carbon price across countries than attempts to harmonise taxes. As such, trading schemes can be used to introduce carbon pricing, without risking carbon leakage and competitiveness implications between participating countries. Emissions trading is therefore a very powerful tool in the framework for addressing climate change at an international level.

Emissions trading is not new to environmental policy. Trading in emissions has been used to reduce sulphur dioxide and nitrous oxide emissions that cause acid rain in the US since 1995[5]. The experience of this scheme increased interest in the potential use of emissions trading to tackle climate change – particularly due to its potential cost effectiveness compared to the use of regulation. Burtaw (1996) estimated that emissions trading under the US Acid Rain Program saved 50% of the costs compared to command and control.

The use of carbon trading schemes is expanding

During the 1990s, as experience of emissions trading for air pollution grew in the US, the EU began to consider the potential of using trading to help meet its Kyoto target emission reduction obligations. The European Commission presented a 'Green Paper' in 2000 that proposed the use of emissions trading. It showed that a comprehensive trading scheme could reduce compliance costs of meeting Kyoto by a third, compared to a scenario with no trading instrument[6].

The EU has since gone on to implement a trading scheme in major energy intensive and energy generation sectors, and in so doing, established the world's largest greenhouse gas emissions market. Launched in January 2005, the EU emissions trading scheme (EU ETS) is still in its infancy. The scheme will enter a second, longer phase in 2008, with a major review on the scheme's design from 2013 to be launched in 2007. Box 15.1 describes how the EU ETS works, and discusses the experience of the scheme to date.

Box 15.1 The European Union Emissions Trading Scheme (EU ETS)

The EU ETS is the first international emissions trading scheme.It established a uniform price of carbon for greenhouse gas emissions from specific heavy industry activities in the 25 EU member states. Phase One of the scheme was launched on 1 January 2005 and runs to the end of 2007. Phase Two runs from 2008-12, and the scheme will continue with further phases beyond 2012. Participation is mandatory for emissions from industrial sectors specified in the scheme. These currently include energy generation, metal production, cement, bricks, and pulp and paper[7].

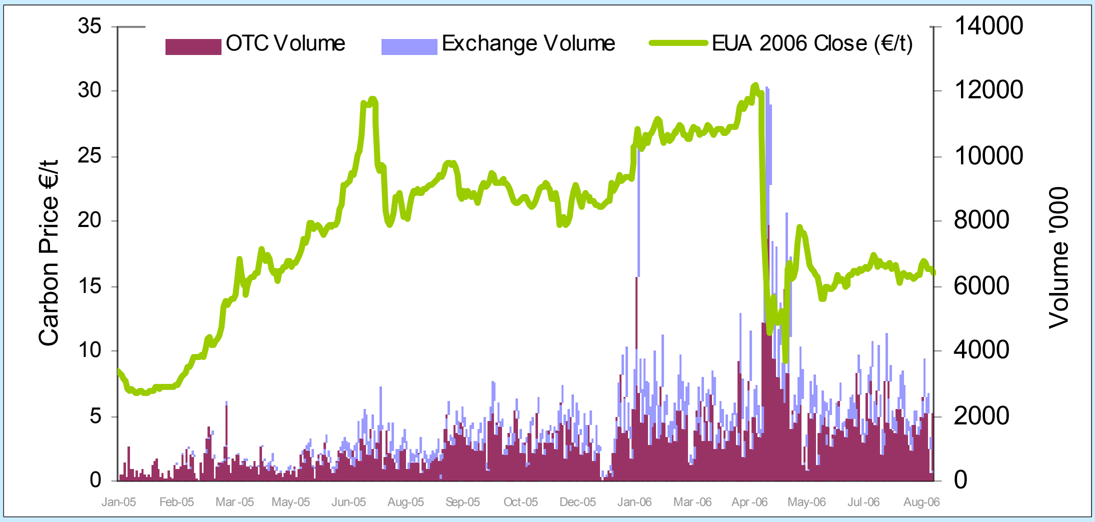
Member states decide, through their National Allocation Plans (NAPs), on the quota or total allocation of allowances for each phase within their country, and on how these are distributed between companies. The plans are subject to approval by the European Commission. They must demonstrate that allocation levels will not exceed expected emission levels in sectors, and are in line with broader plans to make reductions to meet Kyoto targets[8]. Allowances are then issued to all firms on the basis of the NAP. Firms in the scheme must provide an annual report on their emissions, which is audited by a third party.

In Phase One, the scheme covers less than 40% of all EU25 GHG emissions[9], with the permit market over the three-year period worth around US $115 billion[10]. The majority of permits are

[5] See www.epa.gov/airmarkets/arp/index.html for more detail on the US Acid Rain Program.
[6] The 2000 Green Paper estimated the cost of meeting Kyoto as €9 billion euros without trading, €7.2 billion with trading amongst energy producers only, €6.9 billion with trading among energy producers and energy intensive industry and €6 billion with trading among all sectors. See EC (2000).
[7] The scheme covers emissions from heat and energy use from installations of a particular size in these sectors. See EC (2003) for more detail on the scope of the EU ETS
[8] Articles 9 to 11 and Annex III of EU (2003) outline the criteria for allocation in the NAP
[9] Based on emission estimates for EU25 countries in WRI (2005)
[10] This assumes around 2 billion tonnes of allowances are allocated each year for three years, and that the average allowance price is $19 (€15)

currently allocated for free to installations included in the scheme (only 0.2% of all allowances will be auctioned in Phase One[11]), and most member states have prevented the banking of allowances between the two phases. An allowance market has developed through trade exchanges and brokers, with the City of London emerging as an important location for trading. Traded volumes have grown steadily (see below). The price of allowances has been in the range of €10 to €25 per tonne of CO_2 for most of the period, with a steep price drop in April 2006.

The market for EU allowances (EUAs) –prices and volumes



Source: Data taken from Point Carbon, www.pointcarbon.com

Early experience in the scheme has highlighted a number of important issues:

- **The potential for emissions trading schemes to generate demand for emissions reductions in developing countries**: the Linking Directive has enabled EU-based industry to purchase carbon reductions from the cheapest source, including projects and programmes being implemented in the developing world through the use of the Clean Development Mechanism[12]. This has driven growing interest of EU firms in the CDM market, particularly as CDM credits can be used in either phase of the scheme. The CDM market volume grew threefold between 2005 and 2006, to 374 million tonnes (CO_2e), much of this driven by demand from the EU ETS[13].

- **The importance of long term confidence in the future of the scheme:** the EU ETS will continue with a third phase beyond 2012. But companies would like greater clarity over what the EU ETS will look like in Phase III and beyond in order to help judge the impact on their investment decisions. A survey to discover the issues that need to be considered in the review of the EU ETS put the need for certainty on future design issues in the scheme as a top priority[14]. The majority of those surveyed also stated they would prefer allocation decisions to be made a few years in advance of trading periods, and trading periods be lengthened to around 10 years.

[11] Schleich and Betz (2005)
[12] The Clean Development Mechanism is one of the flexible mechanisms under the Kyoto Protocol. Its operation is discussed in detail in Chapter 23.
[13] Capoor and Ambrosi (2006) state that European and Japanese private entities dominated the buy-side of the CDM market in 2005 and 2006, taking up almost 90% of transacted project emissions credits.
[14] See McKinsey et al (2005) for details of the survey of governments, companies and NGO views on issues for the Review of the EU ETS. For UK companies, see also UKBCSE and The Climate Group (2006)
[15] Grubb et al (2006)
[16] Grubb et al (2006)
[17] EC (2005)
[18] See Kruger and Egenhofer (2005). Also, some countries such as the UK went further asking firms to provide verification of data submitted by firms on historic emissions which werebaselines for initial allocations.
[19] See EC (2004) for details of these guidelines.
[20] See Egenhofer and Fujiwara (2005)

- **The impact of imperfect information on prices:** at the start of trading in January 2005, traders had limited information on supply and demand for emission allowances. In particular, the NAPs did not contain clear data on the assumptions lying behind the projections of emissions used as the basis for allocations. The release of the first data on actual emissions from the scheme's participants in April 2006 led to a sharp downward correction in prices (see figure above), as the data showed that the initial NAP allocations exceeded emissions in most sectors of the scheme[15]. The volatility that this caused demonstrates the importance of transparency in initial allocation plans.

- **The difficulties of ensuring scarcity in the market:** overall allocation in the EU ETS market is not set centrally. Rather, it is the sum of 25 individual member state decisions, subject to approval by the Commission. As such, total EU allocation is an outcome of many decisions at various levels, with a risk of gaming on allocation levels between member states if they make their decisions expecting allocation levels will be higher elsewhere in Europe. It has therefore been difficult to ensure scarcity in the EU ETS market. As a result, the total EU wide allocation in Phase One is estimated to be only 1% below projected "business as usual" emissions[16],[17]. This underlines the need for stringent criteria on allocation levels for member states, and robust decisions by the European Commission on NAPs to ensure scarcity in the scheme.

- **The need for robust administrative systems:** the methods used to determine allocations placed considerable demands on companies to collect, verify and submit historical data on emissions.In addition, to ensure confidence in compliance standards across the EU on measuring emissions[18], companies had to set up monitoring, reporting and verification systems in line with EU guidelines[19]. Costs were high for small firms that had low annual emissions included in the EU ETS; requests to reconsider the minimum size of plants included in the scheme have subsequently been made by both member states and business.[20]

The growing importance of the use of emissions trading markets to price carbon is also illustrated by the scope of trading schemes planned or already operating across the world. Norway introduced emissions trading in January 2005 for major energy plants and heavy industry. New South Wales (Australia) already operates a mandatory baseline-and-credit scheme for electricity retailers. Japan and South Korea are also running pilot programmes for a limited number of companies.

Elsewhere, the biggest plans for new emissions trading markets are in the USA, through the Regional Greenhouse Gas Initiative (RGGI) from January 2009[21], and California's plans for using a cap and trade scheme from 2008[22]. Switzerland and Canada also plan to implement trading schemes as part of their programmes to meet Kyoto commitments.The voluntary market for carbon reductions is also growing, driven by demand from both companies and individuals looking to reduce or offset their emissions[23]. The CCX (Chicago Climate Exchange) is an example of a voluntary carbon market. Since December 2003, US based companies that take on voluntary targets to reduce GHG emissions have used this market to achieve their targets.

The following section outlines the design issues that impact on trading scheme efficiency and market effectiveness.

[21] RGGI covers Connecticut, Delaware, Maine, New Hampshire, New Jersey, New York and Vermont. See www.rggi.org for more details.
[22] See announcements by the Governor of the State of Calitornia, www.climatechange.ca.gov
[23] See Butzengeiger (2005) and Taiyab (2006) for more on markets for voluntary carbon offsets

15.4 Designing efficient and well-functioning emissions trading schemes

To reap the benefits of emissions trading, deep and liquid markets and well-designed rules are important.

Emissions trading schemes will, necessarily, deliver carbon prices that vary over time. But a degree of price stability through the emergence of a predictable average price within the emissions trading mechanism is important, particularly for businesses planning long-term investments. And the efficient operation of the scheme, including its impact on incentives, is important to achieve least-cost reductions.

One option to limit the bounds of price movements is to supplement the market instrument itself with price controls, such as formal price caps and price floors[24]. Although this approach has some attractions in principle, there are significant problems with its practical implementation and effectiveness, including the implications for the feasibility of linking with other schemes. These are set out in Box 15.2.

Box 15.2 Price caps and floors in emission trading schemes

As explained in Hepburn et al (2006), a hybrid instrument can in principle be tailored to ensure that in the long term, an overall quantity ceiling is achieved, but that in the short term there is sufficient flexibility to avoid temporarily very high marginal abatement costs. This would help to achieve the balance of long-term certainty and short-term flexibility discussed in Chapter 14.

Price caps (or ‘safety valves')[25] supply allowances on demand if the agreed ceiling price is hit, and would eliminate the risk of price spikes. Price floors would stop the carbon price from falling below a minimum level. They can be implemented in a number of ways, including through a levy that only becomes operational once the floor is breached, or by guaranteeing a minimum future quota price to emitters, by entering a contract to buy permits (which the government can then sell back to the market)[26] – although the risks to the public finances from this latter route should be taken seriously.

However, people would still have to believe that the caps and floors themselves will not be changed. There are also risks that the imposition of a cap alone would damage incentives for investing in low carbon technologies as it sets an upper limit on the future expected price, lowering potential returns to low carbon technology[27].

Importantly, the use of different price caps and floors in different schemes would compromise the efficiency of regional trading schemes- there are risks of carbon leakage and unintended transfers across jusrisdictions with different carbon price ranges. As such, to operate efficiently, price caps and floors would need to be the same across all participating countries. Agreeing a common price cap or floor across countries is likely to suffer from the same difficulties as any attempt to harmonise carbon taxes more generally. Even if countries within a single scheme could agree a cap or floor, this would present an obstacle to linking to other schemes with different rules. This is a drawback to the practical applicability of these methods.

Fundamentally, to ensure confidence in a stable long-term carbon price, and to realise the full efficiency benefits of any trading scheme, the creation of deep, liquid and efficient markets is essential. Several factors can facilitate this:

[24] See, for instance, Pizer (2002) and Pizer (2005)
[25] See Jacoby and Ellerman (2004)
[26] Helm and Hepburn (2005)
[27] Blyth and Yang (2006) modelling shows that in principle, price caps and floors would reduce uncertainty on future prices, but as people need to believe that caps will stay the impact is limited.Stronger effects on reducing uncertainty come from lengthening the period of price stability from 5 to 10 years as discussed above.

- **Broadening the scope** of the scheme, to include more gases, more countries, and international credits;
- Ensuring appropriate **scarcity** in the system;
- **Lengthening the trading periods**, to provide longer-term confidence;
- Designing appropriate **allocation schemes**; and
- Promoting **transparency**.

The following sections discuss these in more detail.

Broadening the scope of the scheme will tend to lower costs and reduce volatility

In general, the deeper and more liquid a market, the harder it is for any individual trade to affect the overall price level, and hence the less volatile the market will tend to be. Introducing different economic sectors or countries to a market can also reduce the impact of a shock in any one sector on the scheme as a whole. In addition, the greater the degree of flexibility about what type of emissions reductions are made and where they are made, the lower the cost will be.

There are a number of ways to widen the scope of trading schemes. One is to widen the number of sectors and activities covered by an individual scheme. Some of the practical issues associated with this are discussed in Section 15.5 below.

Another is to offer access to flexible mechanisms such as Joint Implementation (JI) or the Clean Development Mechanism (CDM)[28]. This expands the options for generating credits for emissions reductions to most parts of the world, maximising the opportunities for efficiency. The environmental benefits of using these credits will depend on the credits representing a real reduction on what emission levels would otherwise have been (the 'business as usual' level of emissions). Countries that can generate CDM credits do not have binding caps on emissions, and are often fast changing economies; as such, establishing a credible estimate of what a business as usual baseline is, and whether reductions would have taken place in the absence of the CDM project, can be complex[29]. Chapter 23 examines this in more detail.

Linking different national or regional cap and trade schemes is also desirable on efficiency grounds, but, to reap the efficiency benefits, the schemes should be broadly similar in design. The practical issues of linking are discussed in Chapter 22.

The introduction of new sectors, and linking to new regions, can cause some short-term price instability, as there is uncertainty over the net impacts of newly included sectors and their response to the scheme. But the impact on long-term stability should still be positive.

As well as bringing extra depth and liquidity into markets, commonality or linking of schemes avoids the leakage, confusion and inefficiency of parallel schemes with different carbon prices. In any one area or country, a single or unified scheme is better than a proliferation of schemes.

The degree of scarcity in the market is important in determining prices

To facilitate more stable carbon markets, allocation levels should be consistent with overall national, regional or multilateral emissions reductions targets, and be clearly below expected 'business as usual' (BAU) emissions. This is complicated by the uncertainties in predicting future emissions over an entire trading period.

The first phase of the EU ETS illustrates this. Allocation decisions were based on projections of BAU emissions of the sectors in the scheme, many of which appear to have been overestimated, meaning that total EU allocation was just 1% under projections of BAU of the

[28] These mechanisms are discussed fully in Chapter 23.
[29] The CDM Executive Board approves methodologies for baseline setting in CDM projects. See Chapter 23.

whole EU ETS. In contrast, earlier emissions trading schemes such as the US Sulphur Dioxide trading programme, had allocation levels at around 50% below baseline emissions[30].

The degree of scarcity in a scheme depends not just on the cap which is set for the scheme itself, but also on whether or not companies are permitted to use credits for emission reductions that are generated in areas without a cap, such as those from the CDM. As long as these credits represent real emission reductions, there is little reason to limit their use, as cost-efficiency demands that emissions reductions are made wherever this is cheapest.

If allowing the use of mechanisms such as the CDM turns out to deliver large quantities of low-cost reductions into a trading scheme, then, at the time when allocations for subsequent periods for the scheme are set, the cap may need to be tightened to ensure that the carbon price continues to reflect the social cost of carbon, and is consistent with the achievement of the long-term goal for stabilisation. The impact of CDM credits on the price should be considered alongside other emerging information on the costs and benefits, as part of the revision process for allocations.

Greater certainty on the evolution of prices over future trading periods, and banking and borrowing between periods, can help to smooth compliance over time and investment cycles

Longer trading periods in trading schemes can help to smooth compliance over time and investment cycles, as they allow the private sector to have greater control over the timing of the response to carbon policy. They also reduce policy risk to the extent that they suggest a deeper commitment to carbon policy. However, excessively long commitment periods limit policymakers' flexibility in responding to changing information and circumstances.As the previous chapter discussed, this is important in order to keep down the overall costs of carbon pricing to the economy[31], and to readjust targets as more information on climate change itself is gathered.

The key issues for investor confidence are a commitment to the long-term future of the scheme and predictability in its overall shape and rules. This predictabilty can be achieved through establishing revision rules for future allocation periods. For instance, governments may announce that future allocations will be contingent on factors such as the price of permits in the preceding period. They could also announce a target range for prices[32] (which should be in line with the expected trajectory for the social cost of carbon – see Chapter 13). Setting out expectations on issues such as expansion to new sectors, or the use of CDM, could also be important. These principles could be set over a very long time period of perhaps 10 to 20 years, with allocations made at more regular intervals.

Within this framework, banking, and possibly borrowing, can be used to create links between different phases of a trading scheme. Banking is the ability to carry over unused quotas from one period to another, and borrowing the ability to use or purchase quotas from a future period in the current period. This allows trading to take place across commitment periods, as well as across sectors and countries. This can improve flexibility, as well as reducing the risk of price spikes or crashes at the end of trading periods discussed above.

Some existing emission trading schemes already allow banking. Banking should help to encourage early emission reductions where this is more cost effective[33]. For example, the heavy use of banking in the US Acid Rain Program has been seen by some as a success in terms of delivering early reductions and improving efficiency. Ellerman and Pontero (2005) found that 30% of allowances were banked between 1995-99 (Phase One of the programme). Firms made efficient decisions to make earlier reductions and bank allowances forward, due to the expectation of tighter caps in future phases. As a result, in total, emissions reduced in Phase One were twice that required to the meet the cap.

[30] See Grubb and Neuhoff (2006) for a discussion of the use of projections and price volatility in the EU ETS.
[31] Helm and Hepburn (2006)
[32] See Newell et al (2005) for an example of how such revision rules could work.
[33] However, unrestricted banking can also allow emissions to be concentrated in time (Tietenberg,1998) – and such hoards of emissions could have high associated damage costs compared to dispersed emissions.

In contrast, very few existing emissions trading schemes have made use of borrowing. The main reason why borrowing has been restricted in existing trading schemes is credibility and compliance, including the risk of borrowing simply being offset by compensating increases in allocations in future periods. In theory, unrestricted borrowing could delay emissions reductions indefinitely, thus raising the risk of 'overshooting' a long run quantity ceiling. A credible enforcement strategy, and long-term principles for allocation, are therefore essential to ensure that reductions borrowed from the future are real and delivered.

Where there are longer periods within which compliance is possible, and a clearer view of the longer term direction of carbon policies, liquid futures markets in carbon are more likely to emerge, and hedging instruments will be developed that allow firms to manage price uncertainty more systematically.

The choice and design of allocation methodology is an important determinant of both efficiency and distributional impact

Permits in an emissions trading scheme can be allocated for free, or sold (usually, though not necessarily, through auction[34]). It is possible to combine these – for instance, the EU ETS allowed for up to 5% of permits to be auctioned in Phase One, and 10% in Phase Two.

In principle and assuming perfect competition, free allocation and auctioning should both be equally efficient. In both cases, businesses face the same marginal costs arising from the emission of an extra tonne of carbon dioxide, and should therefore make the same decision on whether or not to emit in either case.

But this argument is static, ignores the structure of markets and takes no account of distributional or public finance issues. In reality the methods differ in two important respects. First, free allocation methodologies can dampen incentives to incorporate the cost of carbon into decision making consistently, and distort competition. Thus they slow adjustment and potentially raise the overall cost of compliance.

Second, they differ in their distributional impact. Free allocations give companies lump sum transfers in the form of carbon allowances; depending on market structure and demand.Such transfers may result in windfall profits. Not surprisingly, free permits are generally favoured by existing players in an industry. Auctioning leads to financial transfers to governments, which may have benefits for the public finances, depending on whether this is a new revenue flow or a substitute for other sources of finance.

These issues were raised in the preceding chapter, and are explored in the next two sections.

Free allocations can significantly distort incentives

There are a number of reasons why emissions trading schemes based on free allocation may distort incentives for emissions reductions:

- If there is an expectation that the baseline year upon which free allocations are based will be updated, participants have incentives to invest in dirty infrastructure and emit more now to get more free allowances in the future[35]. A one-off allocation based on past emissions (or grandfathering) over all trading periods is one way of avoiding this. However, as a trading scheme matures, the relevance of past emission levels may become a less and less relevant basis for the likely emissions of each plant, say ten or more years later.

[34] The discussion in this section assumes that the sale of permits to industry would happen through auctioning. Other methods are also possible, such as direct sales; these are not discussed fully here, but would be subject to some of the same arguments.
[35] Neuhoff et al (2006) also find that in an international emissions trading scheme, if updating is used in one country but not others, it equates to free riding by the country that uses updating.

- Free allocations can act as a disincentive to new entry to a market, restricting competition and reducing efficiency. If incumbents receive free allowances, but new plants must purchase allowances, free allocations directly create barriers to entry, meaning that the provision of free allocations for new plants may be required[36]. In turn, the rules for free allocations to new plants may indirectly distort incentives: if allocations are given in proportion to the expected emissions from the new plant, they may reward higher-carbon technologies[37].

- There may also be disincentives to exit from markets. The existence of 'use it or lose it' closure rules, which mean a plant must be open in order to receive free allowances, may prevent the closure of inefficient plants.This would mean emission levels are higher than if plants could keep allowances if they shut down, or had no free allowances to begin with[38].

- Under auctioning, with no lump sum of free allowances, businesses will face upfront costs in buying permits to cover their emissions. This will tend to bring management attention to the importance of making efficient decisions that fully account for the cost of carbon. Free allocations may not have the same behavioural impact, delaying adjustments to making effective decisions on carbon compliance[39].

Free allocation methodologies can therefore seriously reduce the dynamic efficiency of a trading scheme, making the cost of reductions higher in the longer term than would otherwise be the case.

Benchmarking the emissions needed for efficient low carbon technologies for both existing and new plants is an alternative basis for issuing free allocations. It offers the opportunity to more clearly 'reward' clean technologies, and penalise carbon intensive technology by developing an average 'rate' of emissions for particular fuels, technologies or plant sizes. The more standardised a benchmark is, the more effective benchmarking is likely to be[40]. Benchmarking can also be used specifically for new entrants, by allocating on the basis of the most efficient technologies available[41].

Auctioning can avoid many of the incentive problems associated with free allocation, although good design is necessary to avoid introducing new inefficiencies. Small, frequent auctions may be more effective in limiting any market power that may exist in the permit market[42]. In principle, to ensure an efficient outcome, the auction method should promote competition and participation for small as well as larger emitters. While one auction at the beginning of the permit period may minimise administration costs, it may also carry a risk of larger players buying the majority of permits and extracting oligopoly rents in the secondary permit market. More frequent auctions also allow for all players to adjust bids and learn from experience of early auctions, and may be helpful in promoting price stability[43]. Given the administrative costs of the data required for free allocation methodologies, auctioning may also offer lower administrative costs to both firms and governments.

[36] In an international trading scheme, if one country has free allowances for new plants, there are compeitiveness implications if other countries do not. This logic drove all 25 EU member states chose to set aside of allowances for new entrant plants that total around 5% of all EU allowances.

[37] Modelling of the UK electricity sector in Neuhoff et al (2006), demonstrates that free allowances for new plants using high carbon technologies could increase overall emissions. The existence of a 'use it or lose it ' closure rule for EU ETS allocations will reduce plant retirement rates and reduce investment in new plants, causing higher emission levels.

[38] In the EU ETS, most member states had 'use it or lose it' closure rules, mainly due to the rules for free allocation to new plants. In Germany, a 'transfer rule' allowed allowances from old plants to be retained if a new plant was built. This still risks new plants receiving higher allocation levels than needed.

[39] Hepburn et al (2006a)

[40] Neuhoff et al (2006) show that for generation plants in the EU ETS, benchmarks based on plant capcity as opposed to fuel and technology specific benchmarks are the least distorting.

[41] The use of benchmarking on the basis of low carbon technology emission rates is an option and has been used in the EU ETS NAPs of some member states. See DTI (2005) for an example of the use of benchmarks for 'new entrant' plants in the UK

[42] Hepburn et al (2006a) considers auction design in the EU ETS

[43] Hepburn et al (2006a)

Using free allocation has benefits for managing the transition to emissions trading, but risks creating substantial windfall profits

Free allocations and auctioning have very different distributional impacts. This has led to a debate over whether allocation methods will affect the profitability of firms, as well as the implications for competitiveness. Carbon pricing will most affect the operating costs of energy intensive industries that compete in international markets, such as non-ferrous metals and some chemicals sectors (see Chapter 11). In the first instance, as auctioning and free allocation both impose the same marginal cost on emissions (as the carbon price is the same), the profit maximising quantity and price for any company should be the same in each case, and there should be no impact on the fundamental risks to competitiveness from the choice of allocation method.

There is, however, an important difference in terms of the impact on companies' balance sheet, which may have competitiveness implications[44]. A firm with free allocations that competes against other firms who face the cost of carbon but do not have free allowances, would be in an advantageous direct position in the sense that it receives a subsidy. It could for example, use this to capture market share by a period of low prices. However, if a firm competes against other firms who do not face a cost of carbon, the 'subsidy' of free allowances may be used to maintain its competitiveness, rather than gain competitive advantage over other firms.

This subsidy effect means that free allocations may have an important role to play in managing the transition to carbon pricing. Full auctioning imposes an immediate hit on companies' balance sheets equivalent to the full cost of all their emissions, whereas free allocation means that companies only have to pay for the cost of any additional permits they need to purchase. This difference in upfront costs may be important, particularly for firms that have significant sunk costs in existing assets and need to invest in lower-carbon assets in response.

In terms of the impact on firms' profits, free or purchased allowances are one factor influencing whether firms face profit or losses from the introduction of a trading scheme. Emissions trading increases the marginal costs of production, but the extent to which firms have to internalise these costs and therefore suffer reduced profits, will depend on:

- whether they can pass on costs to consumers (which depends on market structure and the shape of the demand curve for the good);
- whether they have ways of reducing emissions themselves which are cheaper than buying allowances (cost effective abatement); and
- whether they have some free allowances that can compensate for increased marginal costs

A firm that receives free allowances equal to its existing emissions can make the same profits as before from unchanged production activities, provided the market price for its output is unchanged – or do still better by responding to the new price for carbon. What happens to the market price for its product will depend on industrial structure.

If firms are in perfectly competitive markets, the increase in marginal costs from emissions trading will be fully reflected in prices to consumers, and (in the absence of abatement) profits will stay the same as before the scheme's introduction. Any free allowances they receive equate to windfall profits[45]. But where firms operate in markets where there is international competition and/or very elastic demand and so are unable to pass on costs, free allowances

[44] Smale et al (2006) show that marginal cost increases from the EU ETS most affects the competitiveness of the aluminium sector as it competes in a very global market, and does not get free allowances to compensate-the aluminium sector is currently not directly covered by the scheme, but still faces higher electricity prices.
[45] Sijm et al (2006) show that in the EU ETS, free allocation to electricity generation companies has created substantial windfall profits while consumers have faced increased electricity prices to reflect allowance costs.

can act to maintain profitability by compensating for the increasing operating costs and reduced revenue that may be necessary to maintain market share[46].

Nevertheless, whatever the market structure, it is important that free allocations are only temporary. They may be necessary to manage a transition, but if permanently used, they would distort competition and emission reductions will be below their efficient levels.

The creation of robust institutions, and the collection and provision of reliable information, are important for efficiency

Price stability can also be encouraged by the provision of robust information. In particular, transparent and regular information on actual emissions of scheme participants, as well as on the intial allocations, will help to reveal the basis of market demand and supply.

The importance of information of this kind is illustrated by the experience of the EU ETS when the first verified emissions data of installations included in the scheme were published in March 2006. As Box 15.2 showed, prices dropped sharply in response, as it was clear that, for many firms, actual emissions were well below the number of allowances given to them at the start of the scheme. Revealing information on actual emissions more regularly through the trading period would help limit this volatility. Such requirements for more frequent information releases would, however, impose additional costs on emitters, implying that these requests may need to be limited to the largest emitters.

The quality of monitoring, reporting and verification standards is integral to confidence in a trading scheme. A transparent and well enforced system of measuring and reporting emissions is crucial for securing the environmental credibility of a scheme as well as free trade across plants. Monitoring, reporting and verification (MRV) rules ensure that a tonne of carbon emitted or reduced in one plant is equal to a tonne of carbon emitted or reduced in a different plant[47].

Just as these issues are important in national and regional emissions trading schemes, the emergence of a liquid and efficient global carbon market has similar requirements. Indeed, to facilitate such a market, the EU and others wanting to develop global emissions trading will need to build on existing institutions to develop trading infrastructure. The World Bank emphasises that this includes ensuring strong legal bases to enforce compliance in the jurisdictions of participating firms and agreeing on minimum standards for monitoring, reporting and verification of emissions. Institutions that can deliver predictable and transparent information for emissions markets will also be vital, as will general oversight on the transparency of financial services that support trading such as securities, derivative products or hedge funds[48].

Drawing out implications for the future of the EU emissions trading scheme

The EU ETS will continue beyond 2012 with a third phase. The details of Phase III have yet to be determined, and will be considered in the European Commission's review of the EU ETS in 2007. The review will propose developments in the scheme, drawing on the experience of the EU ETS to date. In particular, it will consider the expansion of the scheme to other sectors (including transport) and links to other trading schemes.

Decisions made now on the third phase of the scheme that will run post 2012, pose an opportunity for the EU ETS – the most important emissions trading market – to influence other emerging markets, as well as to be the nucleus of future global carbon markets. Based on the analysis in this section, there are certain key principles to consider in taking the EU ETS scheme forward. These are set out in Box 15.3.

[46] To maintain profits, commentators state various levels of free allocation as necessary, they need not be 100%. See, for instance, work by Bovenberg and Goulder (2001), Smale et al (2006), Vollebergh et al (1997), Quirion (2003) on allocation and profitability. Also Hepburn et al (2006b) provide a generalised theoretical framework, including an analysis of asymmetric market structure and apply this to four EU ETS sectors.
[47] Kruger and Egenhofer (2005)
[48] Capoor and Ambrosi (2006)

Box 15.3 Principles for the future design of the EU ETS

A credible signal

- Setting out a **credible long-term vision** for the overall scheme over the next few decades could boost investor's confidence that carbon pricing will exist in the EU going forward
- The overall EU limit on emissions should be set at a level that **ensures scarcity** in the allowance market. Stringent criteria for allocation volumes across all EU sectors are necessary.
- To realise efficiency in the scheme, and minimise perverse incentives, there should be a move to **greater use of auctioning** in the longer term, although some free allocation may be important to manage short-term transitional issues[49].
- Where free allocation is necessary, standardised **benchmarking** is a better alternative to grandfathering and updating.

A deep and liquid market

- **Clear and frequent information on emissions** during the trading period would improve the efficient operation of the market, reducing the risks of unnecessary price spikes.
- Clear and predictable **revision rules** for future trading periods, with the possibility of **banking** between periods, would help smooth prices over time, and improve credibility
- **Broadening participation** to other major industrial sectors, and to sectors such as aviation, would help deepen the market[50].
- Enabling the EU ETS to **link with other emerging trading schemes** (including in the USA and Japan) could improve liquidity as well as establish the ETS scheme as the nucleus of a global carbon market.
- **Allowing use of emission reductions from the developing world** (such as the CDM or its successor) can continue to benefit both the efficiency of the EU scheme as well as the transfer of low carbon technology to the developing world

15.5 Carbon pricing across sectors of the economy

Abatement costs are minimised when the carbon price is equalised across sectors

As discussed in Chapter 9, sectors vary widely in terms of the current availability and average cost of abatement options. The cost of avoiding deforestation, for instance, appears to be relatively low compared with the cost of many low-carbon power generation options; by contrast, in aviation, although there are some opportunities for efficiency gains, options for technology switching are currently very limited.

As discussed in the previous chapter, to minimise the total cost of abatement, the carbon price (whether explicit via a tax or trading instrument, or implicit via regulation) should be equalised across sectors. When the carbon price is applied to sectors with cheap abatement options, initially, emissions will tend to decline more; when applied to sectors with more expensive abatement options, the degree of abatement will be less than in cheaper abatement sectors. At the same time, the price increase for the output of the latter sectors will be, and should be, greater.

This means that from an efficiency perspective, sectors with expensive abatement options should not be excluded from carbon pricing; but neither should they be subject to a different higher carbon price in that sector in order to achieve abatement.

[49] See Neuhoff et al (2006) for more on free allocation and perverse incentives in the EU ETS
[50] See Environment Agency (2006) for more detail on expansion options in the EU ETS.

As well as carbon pricing, governments should also look at the use of technology policies and efficiency policies across sectors – these are considered in the following two chapters. It is also important to consider climate change policy within the context of meeting other policy objectives within sectors, including its interaction with the treatment of externalities such as local air pollution and congestion.

The overall structure and scale of policy incentives will therefore reflect other market failures and complexities within the sectors concerned, as well as the climate change externality. As economies make the transition to full carbon pricing, they may in practice use a mix of instruments.

How the characteristics of different sectors affect choice and design of instrument

The characteristics of sectors may influence the choice and design of the carbon pricing instrument. The underlying economic structures in which the emitters operate in sectors will differ, with implications for the attractiveness of using tax, trade or regulation instruments.

Some of the relevant features of different sectors include:

- Transaction costs: this may be affected by the number and dispersion of emitters, and the institutional arrangements for monitoring and pricing.
- Carbon leakage: this is the risk that emissions-intensive activity moves to an area not subject to a carbon constraint. The choice and design of an instrument may have implications for carbon leakage and competitiveness.
- Distributional impacts: depending on the market structure of the sector, the choice of policy instrument may have different implications for who bears the cost.
- Existing frameworks: policy choices will be influenced by existing national policy frameworks and regulatory structures.

It is also important to consider where in the value chain to price carbon. If “upstream” emissions are priced (for instance, at the power station or oil refinery), it is not necessary to price “downstream” emissions as well (for instance, in domestic buildings or individual vehicles). However, Chapter 17 focuses particularly on policies to enable investments in energy efficiency by the end-user, which are not discussed separately here.

The following sections analyse how these factors influence policy choice in power and heavy industry, road transport and aviation, and agriculture.

Power and heavy industry

At a global level, power and heavy industry (such as iron and steel, cement, aluminium, paper industries and chemical and petrochemicals) are large emitters. Because of their high carbon intensity, these sectors are likely to be very sensitive to carbon pricing. They typically invest in very long-lived capital infrastructure such as power plant or heavy machinery, so a clear indication of the future direction of carbon pricing policy is particularly important to them.

Power markets in particular are characterised by imperfect market structures, including state monopolies, regulatory constraints, and often large-scale subsidies. The interaction of carbon pricing with these imperfections is complex. Other industries such as paper and chemicals are more decentralised and deregulated. But overall, sources of emissions are concentrated amongst a relatively few, large, stationary installations, where emissions can be effectively measured and monitored.

The concentrated nature of emissions from these sources make them, in principle, well suited to emissions trading. As already discussed, the first and second phases of the EU ETS cover emissions from these sectors. Other trading schemes have a similar focus – the Regional Greenhouse Gas Initiative in the north-east of the USA, for instance, will cover only the power sector.

However, trading is not the only option. Tax could also be an effective mechanism, and would have the advantage of providing greater price predictability. Examples of countries using taxation to meet climate change goals in these sectors include the UK, which has used the Climate Change Levy, a revenue-neutral mechanism which encourages emissions reductions across sectors including industry; and Norway, which introduced a carbon tax in the early 1990s, covering much of its heavy industry as well as the transport sector (Box 15.4).

Box 15.4 A carbon tax in practice: Norway[51]

Like other Scandinavian countries, Norway introduced a carbon tax in the early 1990s. The tax was to form part of substantial shift in fiscal policy as Norway aimed to use the revenue generated by environmental taxes to help reduce distorting labour taxes.

The Norwegian carbon tax initially covered 60 percent of all Norwegian energy related CO_2 emissions. There are several sectors that were exempted from the tax, including cement, foreign shipping, and fisheries. Natural gas and electricity production are also exempt, although virtually all Norway's electricity production is from carbon-free hydroelectric power. Partial exemptions apply to sectors including domestic aviation and shipping, and pulp and paper.

The tax generates substantial revenues; in 1993 the tax represented 0.7 percent of total revenue, which by 2001 had increased to 1.7 percent. The tax is estimated to have reduced *CO_2* emissions by approximately 2.3% between 1990 and 1999[52]. Overall in Norway, between 1990-1999 GDP grew by approximately 23 percent, yet emissions only grew by roughly 4 percent over the same period, indicating a decoupling of emissions growth from economic growth.

There is also some evidence that the tax helped to provide incentives for technological innovation. The Sleipner gas field is one of the largest gas producers in the Norwegian sector of the North Sea. The gas it produces contains a higher CO_2 content than is needed for the gas to burn properly. With the imposition of a carbon tax the implied annual tax bill to Statoil, the state oil company, was approximately $50m for releasing the excess CO_2. This induced Statoil researchers to investigate the storing of excess carbon dioxide in a nearby geological formation. After several years of study, a commercial plant was installed on the Sleipner platform in time for the start of production in 1996. Experience with this plant has has made an important contribution to the understanding of carbon capture and storage technology.

However, there have been some difficulties in the implementation of the tax:

- The impact of the tax on industry was weakened because of numerous exemptions put in place because of competitiveness concerns. This created a complex scheme, and blunted the incentive for industry to modify or upgrade existing plants.
- The carbon tax did not reflect the actual level of carbon emitted from fuels. For instance, low and high-emission diesel fuels are taxed at the same level, despite causing different levels of environmental damage.
- Although Norway, Sweden, Finland and Denmark all put carbon taxes in place in the early 1990s, they have not been able to harmonise their approaches – demonstrating the difficulties of co-ordinating tax policy internationally, even amongst a relatively small group of countries.

Heavy industries compete in international markets, and as Chapter 11 illustrated, there are some risks to competitiveness and of carbon leakage from the use of carbon policy in such sectors. In terms of tax and trading instruments, there may be a difference in impact if taxes cannot be harmonised globally. This is because an international trading scheme imposes a

[51] This draws on Ekins and Barker (2001)
[52] Bruvoll and Larsen (2002)

uniform carbon price across countries, minimising competitiveness implications for countries within the scheme, whereas taxes may impose different costs in different countries.

Regulatory measures have not played a major role in these sectors, although these have been used for other pollutants in the power sector, the EU's Large Combustion Plants Directive being one example. The concentrated number of companies and sources of emissions may make formal or informal sectoral agreements on best practice an effective complement to carbon pricing – this is discussed in Chapter 22.

Road transport

Although the production of fuel for road transport is centralised at oil refineries, most of the emissions from road transport come from a very large number of individual cars and other vehicles. Demand for transport tends to rise with income. There is considerable scope to improve efficiency in the sector, although the responsiveness of demand to price is low, and breakthrough technologies such as hydrogen are still some years away.

Many countries currently levy a road transport fuel tax. Fuel taxes are a close proxy for a carbon tax because fuel consumption closely reflects emissions. They are frequently aimed at other externalities at the same time (discussed further below), and have the advantage of providing a steady revenue stream to the government. Another example is taxes on purchase or annual car taxes, which can be calibrated by the efficiency of the vehicle.

However, it is also possible to use emissions trading in the road transport sector (see Box 15.5). A possible risk of including road transport in an emissions trading scheme is that permit prices and oil prices might move in tandem, thus exacerbating the extent of oil price fluctuations facing the motorist (in contrast to taxes, which are levied as a fixed amount rather than a percentage of fuel price charged, meaning that the fuel price is prone to less variation).

Box 15.5 Ways to include road transport in an emissions trading scheme

There are three main ways in which emissions from road transport could be included in an emissions trading scheme; they differ according to whom the permits are allocated to.

- Motorists. Individual motorists would have to surrender permits whenever they purchased fuel. Quantity instruments might be better than prices at encouraging motorists to reduce their consumption of fuel. However, there would probably be high transaction costs associated with this approach.
- Refineries. Refineries located in the region of the scheme, would have to buy permits to cover the emissions generated when the fuel that they produce is used in vehicles. It would probably be necessary to couple this approach with border adjustments to the price of imported fuel to avoid carbon leakage. Border adjustments are discussed in detail in Chapter 22.
- Manufacturers. Vehicle manufacturers would be faced with a target for fuel efficiency of the average vehicle sold and, to the extent that they exceeded this target, they would have to buy permits to cover the excess expected lifetime carbon emissions from fuel inefficient vehicles. However, future emissions from these vehicles would be uncertain, making this hard to reconcile with trading schemes based on actual emissions.

The European Commission is currently reviewing the operation of the EU ETS, including whether it should be extended to include other sectors such as road transport.

The inclusion of aviation, road, rail and maritime could increase the size of the EU ETS by up to 50% (such that the EU ETS would cover around 55% of total EU 25 greenhouse emissions, and a larger proportion of total CO_2 emissions), with benefits for liquidity[53].

[53] Estimates based on emission estimates for EU 25 in 2000 from WRI (2006).

Regulatory measures play an important role in the transport sectors in many countries. Vehicle standards – which may be mandatory or voluntary – can put an implicit value on carbon, by restricting the availability of less efficient vehicles. These measures are discussed in more detail in Chapter 17.

In practice, a combination of policies may be justified. Existing policy frameworks and institutional structures in countries will be an important determinant of policy choice. Countries with a history of high fuel taxes, for instance, would need to think very carefully about the public finance implications of switching to trading with free allocations; voluntary standards might be very effective in countries with a strong tradition of co-operation between government and business, but much less so in countries with a different culture.

As in other sectors, climate change is not the only market failure in the transport sector and there are important interactions with other policy goals. Congestion, for instance, imposes external costs on other motorists by increasing their journey time. Congestion pricing and carbon pricing are very similar approaches from an economic point of view - they both price for an externality. Congestion charging could have a positive or negative impact on carbon emissions from transport, depending on how the instrument is designed and level at which the charge is set.

Aviation

Aviation faces some difficult challenges. Whilst there is potential for incremental improvements in efficiency to continue, more radical options for emissions cuts are very limited. The international nature of aviation also makes the choice of carbon pricing instrument complex. Internationally coordinated taxes are difficult to implement, since it is contrary to International Civil Aviation Organisation (ICAO) rules to levy fuel tax on fuel carried on international services[54]. The majority of the many bilateral air service agreements that regulate international air services also forbid taxation of fuel taken on board. Partly for this reason, levels of taxation in the aviation sector globally are currently low relative to road transport fuel taxes. This contributes to congestion and capacity limits at airports – a form of rationing, which is an inefficient way of regulating demand.

While either tax or trading would, in principle, be effective ways to price emissions from this sector, the choice of tax, trading or other instruments is likely to be driven as much by political viability as by the economics. Chapter 22 will discuss further the issues of international co-ordination of policy in this area (as well as in shipping, which faces similar issues). A lack of international co-ordination could lead to serious carbon leakage issues, as aircraft would have incentives to fuel up in countries without a carbon price in place.

The level of the carbon price faced by aviation should reflect the full contribution of emissions from aviation to climate change. As outlined in Box 15.6, the impact of aviation on the global warming (radiatiive forcing) effect is expected to be two to four times higher than the impact of the CO_2 emissions alone by 2050. This should be taken into account, either through the design of a tax or trading scheme, through both in tandem, or by using additional complementary measures.

[54] Article 24 of Chicago Convention exempts fuel for international services from fuel duty. See ICAO (2006).

Box 15.6 The impact of aviation on climate change

Aviation CO_2 emissions currently account for 0.7 Gt CO_2[55] (1.6% of global GHG emissions). However the impact of aviation on climate change is greater than these figures suggest because of other gases released by aircraft and their effects at high altitude. For example, water vapour emitted at high altitude often triggers the formation of condensation trails, which tend to warm the earth's surface. There is also a highly uncertain global warming effect from cirrus clouds (clouds of ice crystals) that can be created by aircraft.

In 2050 under 'business as usual' projections, CO_2 emissions from aviation would represent 2.5% of global GHG emissions[56]. However taking into account the non-CO_2 effects of aviation would mean that it would account for around 5% of the total warming effect (radiative forcing) in 2050[57].

The uncertainties over the overall impact of aviation on climate change mean that there is currently no internationally recognised method of converting CO2 emissions into the full CO2 equivalent quantity.

Agriculture and land use

Agricultural emissions come from a large number of small emitters (farms), over three quarters of which are in developing and transition economies. Emissions from agriculture depend on the specific farming practices employed and the local environment conditions. Since the sources tend to be distributed, there would be high transaction costs associated with actual measurement of GHG at the point of emission.

An alternative approach in this sector would be to focus on pricing GHG emission 'proxies'. For example, excessive use of fertiliser or high nutrient livestock feeds is associated with high emissions, but by appropriate pricing, emissions can be reduced. However in practice, in many developing countries fertiliser is actually subsidised, largely to support the incomes of farmers. In many countries it is a somewhat regressive subsidy, as it is the richer farmers or agribusinesses who gain most.

Difficulties associated with measuring emissions are also the reason why it is difficult to incorporate GHG emissions from agriculture into a trading scheme. However there are examples of projects that have overcome these problems and enabled farmers who adopt sustainable agriculture practices, to sell their emission savings on to others via voluntary schemes; this issue is discussed further in Chapter 25.

Inadequate water pricing can intensify the problems of weak fertiliser pricing, since water and fertiliser are complementary inputs – additional fertiliser works much better with stronger irrigation.

Many countries have adopted regulation of agricultural practices. For example, regulations for the use of water in growing rice, the quantity and type of fertiliser used in crop production, or the treatment of manure. Regulations are often location specific, because local conditions influence best practice. However, in developing countries, enforcement of regulations can be difficult because they may not have the institutional structures or resources to allocate to this task. Better pricing of inputs is generally a preferable route: income support to poor farmers or agricultural workers can be organised in much better ways than subsidised inputs.

[55] WRI (2005).
[56] Aviation BAU CO2 emissions in 2050 estimated at 2.3 GtCO2, from WBCSD (2004). Total GHG emissions in 2050 estimated at 84 GtCO2e (for discussion of how calculated, see Chapter 7).
[57] IPCC (1999). This assumes that the warming effect (radiative forcing) of aviation is 2 to 4 times greater than the effect of the CO2 emissions alone. This could be an overestimate because recent research by Sausen et al (2005) suggests the warming ratio is closer to 2. It could be an underestimate because both estimates exclude the highly uncertain possible warming effects of cirrus clouds.

There are complex challenges involved with the inclusion of deforestation, the major cause of land use emissions, in carbon trading schemes. These are discussed in detail in Chapter 25.

15.6 Conclusions

Chapter 14 discussed how, at the global level, policymakers need both a shared understanding of a long-run stabilisation goal, and the flexibility to revise short-run policies over time.

At the national – or regional level – policy makers will want to achieve these goals in a way that builds on existing policies, and creates confidence in the future existence of a carbon price. In particular, they will want to assess how carbon pricing (through either taxation, tradable quotas or regulation) will interact with existing market structures, and existing policies (for instance, to encourage the development of renewable energy or petrol taxes).

Governments will want to tailor a package of measures that suits their specific circumstances. Some may choose to focus on regional trading initiatives, others on taxation and others may make greater use of regulation. The key goal of policy should be to establish common incentives across different sectors, using the most appropriate mechanism for a particular sector. With market failures elsewhere, other objectives, and the costs of adjustment associated with long-lived capital, it will be important to look at both the simple price or tax options as well as quotas and regulation to see what incentives in particular sectors really work.

Carbon pricing is only one element of a policy approach to climate change. The following two chapters discuss the role of technology policy, and policies to influence attitudes and behaviours, particularly in regard to energy efficiency. All three elements are important to achieve lowest cost emissions reductions.

16 Accelerating Technological Innovation

Key Messages

Effective action on the scale required to tackle climate change requires a widespread shift to new or improved technology in key sectors such as power generation, transport and energy use. Technological progress can also help reduce emissions from agriculture and other sources and improve adaptation capacity.

The private sector plays the major role in R&D and technology diffusion. But closer collaboration between government and industry will further stimulate the development of a broad portfolio of low carbon technologies and reduce costs. Co-operation can also help overcome longer-term problems, such as the need for energy storage systems, for both stationary applications and transport, to enable the market shares of low-carbon supply technologies to be increased substantially.

Carbon pricing alone will not be sufficient to reduce emissions on the scale and pace required as:

- Future pricing policies of governments and international agreements should be made as credible as possible but cannot be 100% credible.
- The uncertainties and risks both of climate change, and the development and deployment of the technologies to address it, are of such scale and urgency that the economics of risk points to policies to support the development and use of a portfolio of low-carbon technology options.
- The positive externalities of efforts to develop them will be appreciable, and the time periods and uncertainties are such that there can be major difficulties in financing through capital markets.

Governments can help foster change in industry and the research community through a range of instruments:

- **Carbon pricing**, through carbon taxes, tradable carbon permits, carbon contracts and/or implicitly through regulation will itself directly support the research for new ways to reduce emissions;
- **Raising the level of support for R&D** and demonstration projects, both in public research institutions and the private sector;
- **Support for early stage commercialisation investments in some sectors.**

Such policies should be complemented by tackling institutional and other non-market barriers to the deployment of new technologies.

These issues will vary across sectors with some, such as electricity generation and transport, requiring more attention than others.

Governments are already using a combination of market-based incentives, regulations and standards to develop new technologies. These efforts should increase in the coming decades.

Our modelling suggests that, in addition to a carbon price, **deployment incentives for low-emission technologies should increase two to five times globally** from current levels of around $33billion.

Global public energy R&D funding should double, to around $20 billion, for the development of a diverse portfolio of technologies.

16.1 Introduction

Stabilisation of greenhouse gases in the atmosphere will require the deployment of low-carbon and high-efficiency technologies on a large scale. A range of technologies is already available, but most have higher costs than existing fossil-fuel-based options. Others are yet to be developed. Bringing forward a range of technologies that are competitive enough, with a carbon price, for firms to adopt is an urgent priority.

In the absence of any other market failures, introducing a fully credible carbon price path for applying over the whole time horizon relevant for investment would theoretically be enough to encourage suitable technologies to develop. Profit-maximising firms would respond to the creation of the path of carbon prices by adjusting their research and development efforts in order to reap returns in the future. This chapter sets out why this is unlikely to be sufficient in practice, why other supporting measures will be required, and what form they could take.

This chapter starts by examining the process of innovation and how it relates to the challenge of climate change mitigation, exploring how market failures may lead to innovation being under-delivered in the economy as a whole. Section 16.3 looks more closely at the drivers for technology development in key sectors related to climate change. It finds that clean energy technologies face particularly strong barriers – which, combined with the urgency of the challenge, supports the case for governments to set a strong technology policy framework that drives action by the private sector.

Section 16.4 outlines the policy framework required to encourage climate related technologies. Section 16.5 discusses one element of this framework – policies to encourage research, development and demonstration. Such policies are often funded directly by government, but it is critical that they leverage in private sector expertise and funding.

Investment in Research and Development (R&D) should be complemented by policies to create markets and drive deployment, which is discussed in Section 16.6. A wide range of policies already exist in this area; this section draws together evidence on what works best in delivering a response from business.

A range of complementary policies, including patenting, regulatory measures and network issues are also important; these issues are examined in Section 16.7. Regulation is discussed in the context of mitigation more generally, and in particular in relation to energy efficiency in Chapter 17.

Overall, an ambitious and sustained increase in the global scale of effort on technology development is required if technologies are to be delivered within the timescales required. The decline in global public and private sector R&D spending should be reversed. And deployment incentives will have to increase two to five-fold worldwide in order to support the scale of uptake required to drive cost reductions in technologies and, with the carbon price, make them competitive with existing fossil fuel options. In Chapter 24, we return to the issue of technological development, considering what forms of international co-operation can help to reduce the costs and accelerate the process of innovation.

16.2 The innovation process

Innovation is crucial in reducing costs of technologies. A better understanding of this complex process is required to work out what policies may be required to encourage firms to deliver the low-emission technologies of the future.

Defining innovation

Innovation is the successful exploitation of new ideas[1]. Freeman identified four types of innovation in relation to technological change[2]:

- Incremental innovations represent the continuous improvements of existing products through improved quality, design and performance, as has occurred with car engines;
- Radical innovations are new inventions that lead to a significant departure from previous production methods, such as hybrid cars;
- Changes in the technological systems occur at the system level when a cluster of radical innovations impact on several branches of the economy, as would take place in a shift to a low-emission economy;
- Changes of techno-economic paradigm occur when technology change impacts on every other branch of the economy, the internet is an example.

[1] DTI (2003)
[2] Freeman (1992)

Many of the incentives and barriers to progress for these different types of technological change are very different from each other.

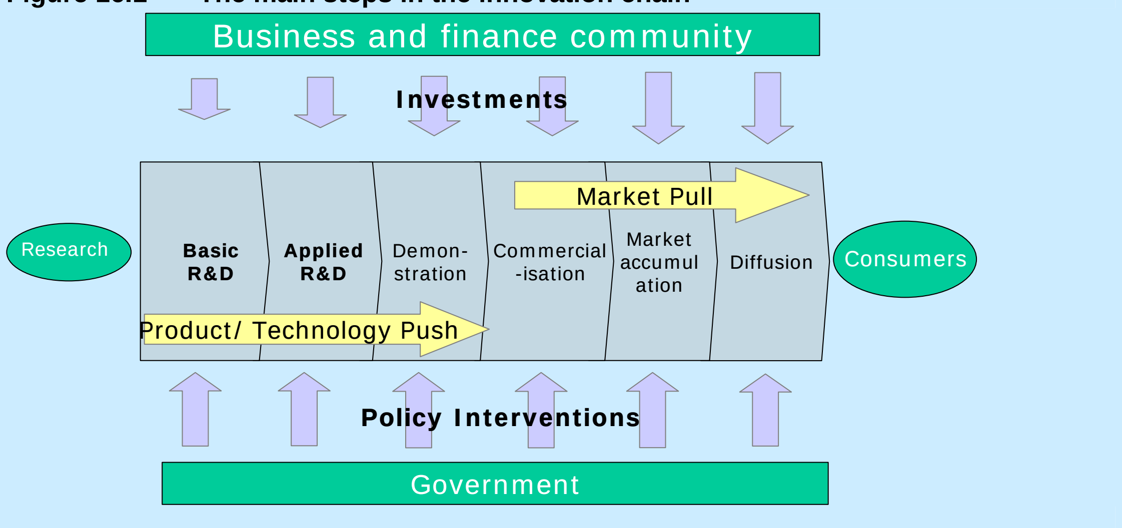
Innovation is about much more than invention: it is a process over time

Joseph Schumpeter identified three stages of the innovation process: invention as the first practical demonstration of an idea; innovation as the first commercial application; and diffusion as the spreading of the technology or process throughout the market. The traditional representation of the diffusion process is by an S-shaped curve, in which the take-up of the new technology begins slowly, then 'takes off' and achieves a period of rapid diffusion, before gradually slowing down as saturation levels are reached. He proposed the idea of 'creative destruction' to describe the process of replacement of old firms and old products by innovative new firms and products.

There is an opportunity for significant profits for firms as the new product takes off and this drives investment in the earlier stages. High profits, coupled with the risk of being left behind, can drive several other firms to invest through a competitive process of keeping up. As incumbent firms have an incentive to innovate in order to gain a competitive advantage, and recognising that innovation is typically a cumulative process that builds on existing progress, market competition can stimulate innovation[3]. As competition increases, and more firms move closer to the existing technological frontier of incumbents, the expected future profits of the incumbents are diminished unless they innovate further. Such models imply a hump-shaped relationship between the degree of product market competition and innovation, as originally suggested by Schumpeter.

An expanded version of this 'stages' model of innovation that broadens the invention stage into basic R&D, applied R&D and demonstration is shown in the subsequent figure. In this chapter the term R&D will be used but this will also cover the demonstration stage[4]. The commercialisation and market accumulation phases represent early deployment in the market place, where high initial cost or other factors may mean quite low levels of uptake.

Figure 16.1 The main steps in the innovation chain[5]



This model is useful for characterising stages of development, but it fails to capture many complexities of the innovation process, so it should be recognised as a useful simplification. A more detailed characterisation of innovation in each market can be applied to particular markets using a systems approach[6]. The transition between the stages is not automatic; many products fail at each stage of development. There are also further linkages between

[3] Aghion et al (2002): Monopolists do not have competitive pressures to innovate while intense competition means firms may lack the resource or extra profit for the innovator may be competed away too quickly to be worthwhile.
[4] R,D&D (Research, Development and Demonstration) can be used for this but it can lead to confusion over the final D as some of the literature uses deployment or diffusion in the same acronym.
[5] Grubb (2004)
[6] For an excellent overview of innovation theory see Foxon (2003)

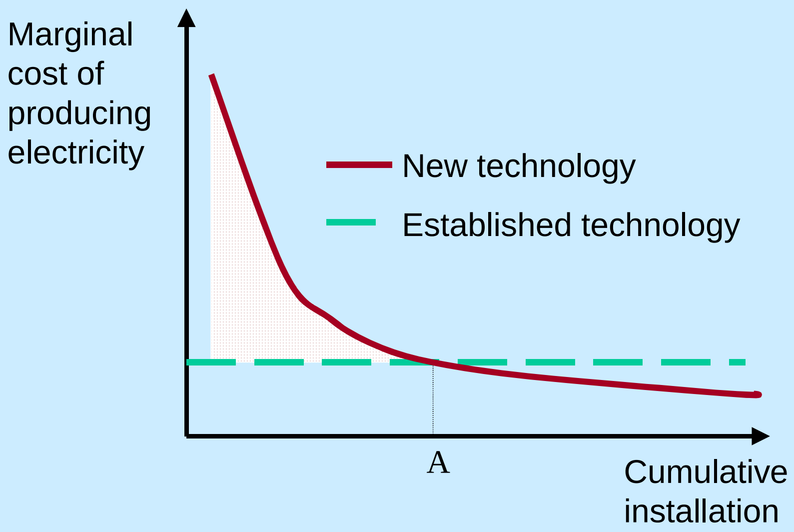
stages, with further progress in basic and applied R&D affecting products already in the market and learning also having an impact on R&D.

Experience curves can lead to lock-in to existing technologies

As outlined in Section 9.7 dynamic increasing returns, such as economies of scale and learning effects, can arise during production and lead to costs falling as production increases. These vary by sector with some, such as pharmaceuticals, experiencing minimal cost reductions while others fall by several orders of magnitude. These benefits lead to experience curves as shown in Box 9.4.

Experience curves illustrate that new technologies may not become cost effective until significant investment has been made and experience developed. Significant learning effects may reduce the incentive to invest in innovation, if companies wait until the innovator has already proven a market for a new cost effective technology. This is an industry version of a collective action problem with its associated free-rider issues.

Figure 16.2 Illustrative experience curve for a new technology



Dynamic increasing returns can also lead to path dependency and 'lock-in' of established technologies. In this diagram, the market dominant technology (turquoise line) has already been through a process of learning. The red line represents a new technology, which has the potential to compete. As production increases the cost of the new technology falls because of dynamic increasing returns, shown by the red line above. In this case, the price of the new technology does ultimately fall below the level of the dominant technology. Some technological progress can also be expected for incumbent dominant technologies but existing deployment will have realised much of the learning[7].

The learning cost of the new technology is how much more the new technology costs than the existing technology; shown by the dotted area where the red line is above the blue. During this period, the incumbent technology remains cheaper, and the company either has to sell at a loss, or find consumers willing to pay a premium price for its new product. So, for products such as new consumer electronics, niche markets of "early adopters" exist. These consumers are willing to pay the higher price as they place a high value on the function or image of the product.

The learning cost must be borne upfront; the benefits are uncertain, because of uncertainty about future product prices and technological development, and come only after point A when, in this case, the technology becomes cheaper than the old alternative. If, as is the case in some sectors, the time before the technology becomes competitive might span decades and the learning costs are high, private sector firms and capital markets may be unwilling to

[7] The learning rate is the cost reduction for a doubling of production and this requires much more deployment after significant levels of investment.

take the risk and the technology will not be developed, especially if there is a potential free-rider problem.

Innovation produces benefits above and beyond those enjoyed by the individual firm ('knowledge spillovers'); this means that it will be undersupplied

Information is a public good. Once new information has been created, it is virtually costless to pass on. This means that an individual company may be unable to capture the full economic benefit of its investment in innovation. These knowledge externalities (or spillovers) from technological development will tend to limit innovation.

There are two types of policy response to spillovers. The first is the enforcement of private property rights through patenting and other forms of protection for the innovator. This is likely to be more useful for individual products than for breakthroughs in processes or know-how, or in basic science. The disadvantage of rigid patent protection is that it may slow the process of innovation, by preventing competing firms from building on each others' progress. Designing intellectual property systems becomes especially difficult in fields where the research process is cumulative, as in information technology[8]. Innovation often builds on a number of existing ideas. Strong protection for the innovators of first generation products can easily be counterproductive if it limits access to necessary knowledge or research tools for follow-on innovators, or allows patenting to be used as a strategic barrier to potential competitors. Transaction costs, the equity implications of giving firms monopoly rights (and profits) and further barriers such as regulation may prevent the use of property rights as the sole incentive to innovate. Also much of value may be in tacit knowledge ('know-how' and 'gardeners' craft') rather than patentable ideas and techniques.

Another broad category of support is direct government funding of innovation, particularly at the level of basic science. This can take many forms, such as funding university research, tax breaks and ensuring a supply of trained scientists.

Significant cross-border spillovers and a globalised market for most technologies offer an incentive for countries to free-ride on others who incur the learning cost and then simply import the technology at a later date[9]. The basic scientific and technical knowledge created by a public R&D programme in one country can spillover to other countries with the capacity to utilise this progress. While some of the leaning by doing will be captured in local skills and within local firms, this may not be enough to justify the learning costs incurred nationally.

International patent arrangements, such as the Trade Related International Property Rights agreement (TRIPs[10]), provides some protection, but intellectual property rights can be hard to enforce internationally. Knowledge is cheap to copy if not embodied in human capital, physical capital or networks, so R&D spillovers are potentially large. A country that introduces a deployment support mechanism and successfully reduces the cost of that technology also delivers benefits to other countries. Intellectual property right issues are discussed in more detail in Section 23.4.

International co-operation can also help to address this by supporting formal or informal reciprocity between RD&D programmes. This is explored in Chapter 24.

Where there are long-term social returns from innovation, it may also be undersupplied

Government intervention is justified when there is a departure between social and private cost, for example, when private firms do not consider an environmental externality in their investment decisions, or when the benefits are very long-term (as with climate change mitigation) and outside the planning horizons of private investments. Private firms focus on private costs and benefits and private discount rates to satisfy their shareholders. But this can lead to a greater emphasis on short-term profit and reduce the emphasis on innovations and other low-carbon investments that would lead to long-term environmental improvements.

---

[8] Scotchmer (1991)
[9] Barreto and Klaassen (2004)
[10] The agreement on Trade Related Intellectual Property Rights (TRIPs) is an international treaty administered by the World Trade Organization which sets down minimum standards for most forms of intellectual property regulation within all WTO member countries.

16.3 Innovation for low-emission technologies

The factors described above are common to innovation in any sector of the economy. The key question is whether there are reasons to expect the barriers to innovation in low-emission technologies to be higher than other sectors, justifying more active policies. This section discusses factors specific to environmental innovation and in particular two key climate change sectors – power generation and transport.

Lack of certainty over the future pricing of the carbon externality will reduce the incentive to innovate

Environmental innovation can be defined[11] as innovation that occurs in environmental technologies or processes that either control pollutant emissions or alter the production processes to reduce or prevent emissions. These technologies are distinguished by their vital role in maintaining the 'public good' of a clean environment. Failure to take account of an environmental externality ensures that there will be under-provision or slower innovation[12].

In the case of climate change, a robust expectation of a carbon price in the long term is required to encourage investments in developing low-carbon technologies. As the preceding two chapters have discussed, carbon pricing is only in its infancy, and even where implemented, uncertainties remain over the durability of the signal over the long term. The next chapter outlines instances in which regulation may be an appropriate response to lack of certainty. This means there will tend to be under-investment in low-carbon technologies. The urgency of the problem (as outlined in Chapter 13) means that technology development may not be able to wait for robust global carbon pricing. Without appropriate incentives private firms and capital markets are less likely to invest in developing low-emission technologies.

There are additional market failures and barriers to innovation in the power generation sector

Innovation in the power generation sector is key to decarbonising the global economy. As shown in Chapter 10, the power sector will need to be at least 60% decarbonised by 2050[13] to keep on track for greenhouse gas stabilisation trajectories at or below 550ppm CO_2e.

For reasons that this section will explore the sector is characterised by low levels of research and development expenditure by firms. In the USA, the R&D intensity (R&D as a share of total turnover) of the power sector was 0.5% compared to 3.3% in the car industry, 8% in the electronics industry and 15% in the pharmaceutical sector[14]. OECD figures for 2002 found an R&D intensity of 0.33% compared to 2.65% for the overall manufacturing sector[15]. Unlike in many other sectors, public R&D represents a significant proportion, around two thirds of the total R&D investment[16].

The available data[17] on energy R&D expenditure show a downward trend in both the public and private sector, despite the increased prominence of energy security and climate change. Public support for energy R&D has declined despite a rising trend in total public R&D. In the early 1980s, energy R&D budgets were, in real terms, twice as high as now, largely in response to the oil crises of the 1970s.

[11] Taylor, Rubin and Nemet (2006)
[12] Anderson et al (2001); Jaffe, Newell and Stavins (2004) and (2003)
[13] This is consistent with the ACT scenarios p86 IEA, 2006 which would also require eliminating land use change emissions to put us on a path to stabilising at 550ppm CO_2e
[14] Alic, Mowery and Rubin (2003)
[15] Page 35: OECD, (2006)
[16] There are doubts as to the accuracy of the data and the IEA's general view is that private energy R&D is considerably higher than public energy R&D (though this still represents a significant share).
[17] Page 33-37: OECD (2006)

Figure 16.3 Public energy R&D investments as a share of GDP[18]

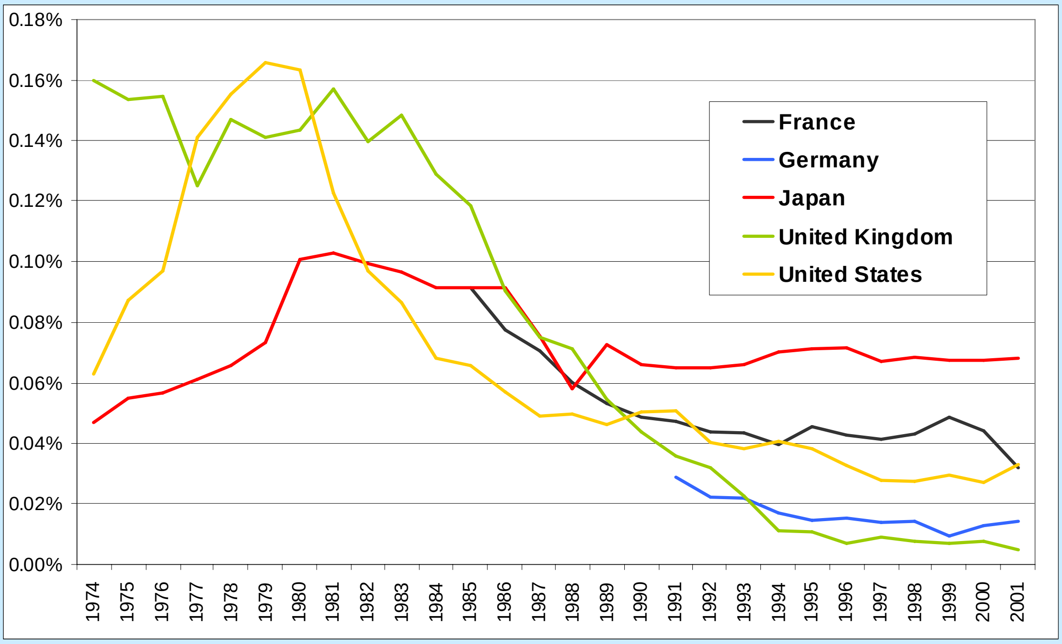
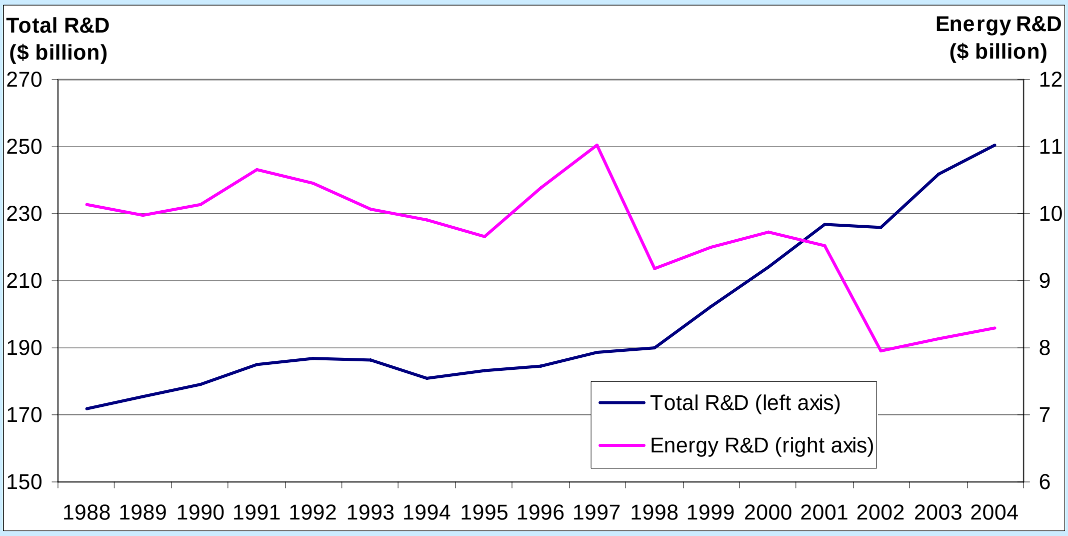


Figure 16.4 Public R&D and public energy R&D investments[19]

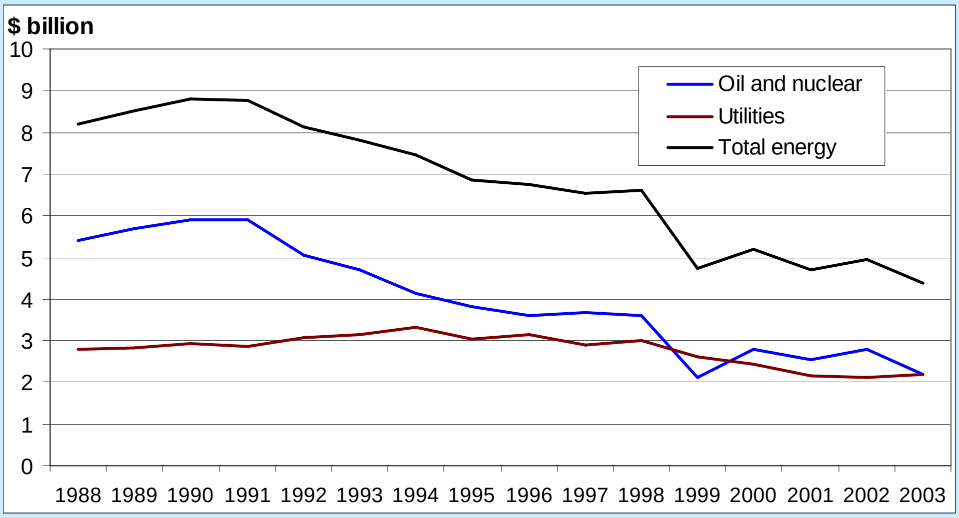


Private energy R&D has followed a similar trend and remains below the level of public R&D. The declines in public and private R&D have been attributed to three factors. *First*, energy R&D budgets had been expanded greatly in the 1970s in response to the oil price shocks in the period , and there was a search for alternatives to imported oil. With the oil price collapse in the 1980s and the generally low energy prices in the 1990s, concerns about energy security diminished, and were mirrored in a relaxation of the R&D effort. Recent rises in oil prices have not, yet, led to a significant increase in energy R&D. *Second*, following the liberalisation of energy markets in the 1990s, competitive forces shifted the focus from long-term investments such as R&D towards the utilisation of existing plant and deploying well-developed technologies and resources - particularly of natural gas for power and heat, themselves the product of R&D and investment over the previous three decades. *Third*, there

[18] Source: IEA R&D database http://www.iea.org/Textbase/stats/rd.asp Categories covered broken down in IEA total Figure 16.8
[19] OECD countries Page 32: OECD (2006)

were huge declines in R&D expenditures on nuclear power following the experiences of many countries with cost over-runs, construction delays, and the growth of public concerns about reactor safety, nuclear proliferation and nuclear waste disposal. In 1974, electricity from nuclear fission and fusion accounted for 79% of the public energy R&D budget; it still accounts for 40%. Apart from nuclear technologies, energy R&D budgets decreased across the board (Figure 16.8).

Figure 16.5 Trends in private sector energy R&D[20]



The sector's characteristics explain the low levels of R&D

There are a number of ways to interpret these statistics, but they suggest that private returns to R&D are relatively low in the sector. There are four distinct factors which help explain this.

The first factor is the nature of the learning process. Evidence from historical development of energy-related technologies shows that the learning process is particularly important for new power generation technologies, and that it typically takes several decades before they become commercially viable. Box 9.4 shows historical learning curves for energy technologies.

If early-stage technologies could be sold at a high price, companies could recover this learning cost. In some markets, such as IT, there are a significant number of 'early adopters' willing to pay a high price for a new product. These 'niche markets' allow innovating companies to sell new and higher-cost products at an early profit. Later, when economies of scale and learning bring down the cost, the product can be sold to the mass market. Mobile phones are a classic example. The earliest phones cost significantly more but there were people willing to pay this price.

In the absence of niche markets the innovating firm is forced to pay the learning cost, as a new product can be sold only at a price that is competitive with the incumbent. This may mean that firms would initially have to sell their new product at a loss, in the hope that as they scale up, costs will reduce and they can make a profit. If this loss-making period lasts too long, the firm will not survive.

In the power sector, niche markets are very limited in the absence of government policy, because of the homogeneous nature of the end-product (electricity). Only a very small number of consumers have proved willing to pay extra for carbon-free electricity. As cost reductions typically take several decades this leaves a significant financing gap which capital markets are unable to fill. Compounding this, the power generation sector also operates in a highly regulated environment and tends to be risk averse and wary of taking on technologies that may prove costlier or less reliable. Together, these factors mean that energy generation

[20] Source Page 35 OECD (2006); For US evidence see Kammen and Nemet (2005)

technologies can fall into a 'valley of death', where despite a concept being shown to work and have long-term profit potential they fail to find a market.

For energy technologies, R&D is only the beginning of the story. There is continual feedback between learning from experience in the market, and further R&D activity. There is a dependence on tacit knowledge and a series of incremental innovations in which spillovers play an important role and reduce the potential benefits of intellectual property rights. This is in strong contrast with the pharmaceutical sector. For a new drug, the major expense is R&D. Once a drug has been invented and proven, comparatively little further research is required and limited economies of scale and learning effects can be expected.

The second factor is infrastructure. National grids are usually tailored towards the operation of centralised power plants and thus favour their performance. Technologies that do not easily fit into these networks may struggle to enter the market, even if the technology itself is commercially viable. This applies to distributed generation as most grids are not suited to receive electricity from many small sources. Large-scale renewables may also encounter problems if they are sited in areas far from existing grids. Carbon capture and storage also faces a network issue, though a different one; the transport of large quantities of CO_2, which will require major new pipeline infrastructures, with significant costs.

The third factor is the presence of significant existing market distortions. In a liberalised energy market, investors, operators and consumers should face the full cost of their decisions. But this is not the case in many economies or energy sectors. Many policies distort the market in favour of existing fossil fuel technologies[21], despite the greenhouse gas and other externalities. Direct and indirect subsidies are the most obvious. As discussed in Section 12.5 the estimated subsidy for fossil fuels is between $20-30 billion for OECD countries in 2002 and $150-250 billion per year globally[22]. The IEA estimate that world energy subsidies were $250 billion in 2005 of which subsidies to oil products amounted to $90 billion[23]. Such subsidies compound any failure to internalise the environmental externality of greenhouse gases, and affect the incentive to innovate by reducing the expectations of innovators that their products will be able to compete with existing choices.

Finally, the nature of competition within the market may not be conducive to innovation. A limited number of firms, sometimes only one, generally dominate electricity markets, while electricity distribution is a 'natural' monopoly. Both factors will generally lead to low levels of competition, which, as outlined in Section 16.1, will generally lead to less innovation as there is less pressure to stay ahead of competitors. The market is also usually regulated by the government, which reduces the incentive to invest in innovation if there is a risk that the regulator may prevent firms from reaping the full benefits of successful innovative investments.

These barriers will also affect the deployment of existing technologies

The nature of competition, existing infrastructure and existing distortions affect not only the process of developing new technologies; these sector-specific factors can also reduce the effectiveness of policies to internalise the carbon externality. They inhibit the power of the market to encourage a shift to low-carbon technologies, even when they are already cost-effective and especially if they are not. The generation sector usually favours more traditional (high-carbon) energy systems because of human, technical and institutional capacity. Historically driven by economies of scale, the electricity system becomes easily locked into a technological trajectory that demonstrates momentum and is thereby resistant to the technical change that will be necessary in a shift to a low-carbon economy[24].

---

[21] Neuhoff (2005).
[22] Source: REN21 (2005) which cites; UNEP & IEA. (2002). Reforming Energy Subsidies. Paris. www.uneptie.org/energy/publications/pdfs/En-SubsidiesReform.pdf Also Johansson, T. & Turkenburg, W. state in (2004). Policies for renewable energy in the European Union and its member states: an overview. *Energy for Sustainable Development* 8(1): 5-24.that "at present, subsidies to conventional energy are on the order of $250 billion per year" and $244 billion per annum between 1995 and 1998 (34% OECD) in Pershing, J. and Mackenzie (2004) Removing Subsidies.Leveling the Playing Field for Renewable Energy Technologies. Thematic Background Paper. International Conference for Renewable Energies, Bonn (2004)
[23] WEO, (in press)
[24] Amin (2000)

Despite advances in the transport sector, radical change may not be delivered by the markets

Transport currently represents 14% of global emissions, and has been the fastest growing source of emissions because of continued growth of car transport and rapid expansion of air transport. Innovation has been dominated by incremental improvements to existing technologies, which depend on oil. These, however, have been more than offset by the growth in demand and shift towards more powerful and heavier vehicles. The increase in weight is partly due to increased size and partly to additional safety measures. The improvements in the internal combustion engine from a century of learning by doing, the efficiency of fossil fuel as an energy source and the existence of a petrol distribution network lead to some 'lock-in' to existing technologies. Behavioural inertia compounds this 'lock-in' as consumers are also accustomed to existing technologies.

Certain features of road transport suggest further innovative activity could be delivered through market forces. Although there is no explicit carbon price for road fuel, high and stable fuel taxes[25] in most developed countries provide an incentive for the development of more efficient vehicles. Niche markets also exist which help innovative products in transport markets to attract a premium. These factors together help to explain how hybrid vehicles have been developed and are now starting to penetrate markets, with only very limited government support: some consumers are content to pay a premium for what can be a cleaner and more fuel-efficient product. There is also a small number of large global firms in this sector, each of which have the resources to make significant innovation investments and progress. They can also be less concerned about international spillovers as they operate in several markets.

Incremental energy efficiency improvements are expected to continue in the transport sector. These will be stimulated both by fuel savings and, as they have been in the past, by government regulation. Both the hybrid car, and later, the fuel cell vehicle, are capable of doubling the fuel efficiency of road vehicles, whilst behavioural changes - perhaps encouraged, for example, by congestion pricing or intelligent infrastructure[26] - could lead to further improvements.

Markets alone, however, may struggle to deliver more radical changes to transport technologies such as plug-in hybrids or other electrical vehicles. Alternative fuels (such as biofuel blends beyond 5-10%, electricity or hydrogen) may require new networks, the cost of which is unlikely to be met without incentives provided by public policy. The environmental benefit of alternative transport fuels will depend on how they are produced. For example, the benefit of electric and hydrogen cars is limited if the electricity and hydrogen is produced from high emission sources. Obstacles to the commercial deployment of hydrogen cell vehicles, such as the cost of hydrogen vehicles and low-carbon hydrogen production, and the requirement to develop hydrogen storage further, ensure it is unlikely that such vehicles will be widely available commercially for at least another 15 to 20 years.

In Brazil policies to encourage biofuels over the past 30 years through regulation, duty incentives and production subsidies have led to biofuels now accounting for 13% of total road fuel consumption, compared with a 3% worldwide average in 2004. Other countries are now introducing policies to increase the level of biofuels in their fuel mix. Box 16.1 shows how some governments are already acting to create conditions for hydrogen technologies to be used. Making hydrogen fuel cell cars commercial is likely to require further breakthroughs in fundamental science, which may be too large to be delivered by a single company, and are likely to be subject to knowledge spillovers.

The development of alternative technologies in the road transport sector will be important for reducing emissions from other transport sectors such as the aviation, rail and maritime sectors. The local nature of bus usage allows the use of a centralised fuel source and this has led to early demonstration use of hydrogen in buses (see Box 16.1). In other sectors, such as aviation where weight and safety are prominent concerns, early commercial development is unlikely to take place and will be dependent on development in other areas first. The capital stock in the aviation, maritime and rail sectors (ships, planes and trains) lasts several times

[25] There are exceptions in the case of biofuels with many countries offering incentives through tax incentives.
[26] Intelligent infrastructure uses information to encourage efficient use of transport systems.
http://www.foresight.gov.uk/Intelligent_Infrastructure_Systems/Index.htm

longer than road vehicles so this may result in a slower rate of take-up of alternative technologies. The emissions associated with rail transport can be reduced through decarbonising the fuel mix through biofuels or low carbon electricity generation. In the aviation sector improved air traffic management and reduced weight, through the use of alternative and advanced materials, can add to continued improvements in the efficiency of existing technologies.

Box 16.1 Hydrogen for transport

Hydrogen could potentially offer complete diversification away from oil and provide very low carbon transport. Hydrogen would be best suited to road vehicles. The main ways of producing hydrogen are by electrolysis of water, or by reforming hydrocarbons. Once produced, hydrogen can be stored as a liquid, a compressed gas, or chemically (bonded within the chemical structure of advanced materials). Hydrogen could release its energy content for use in powering road vehicles by combustion in a hydrogen internal combustion engine or a fuel cell. Fuel cells convert hydrogen and oxygen into water in a process that generates electricity. They are almost silent in operation, highly efficient, and produce only water as a by-product. Hydrogen can produce as little as 5% of the emissions of conventional fuel if produced by low-emission technologies.[27]

There are several hydrogen projects around the world including:

- Norway: plans for a 580km hydrogen corridor between Oslo and Stavanger in a joint project between the private sector, local government and non-government organisations. The first hydrogen station opened in August 2006
- Denmark and Sweden: interested in extending the Norwegian hydrogen corridor
- Iceland: home to the first hydrogen fuelling station in April 2003 and it is proposed that Iceland could be a hydrogen economy by 2030
- EU: trial of hydrogen buses
- China: hydrogen buses to be used at the Beijing Olympics in 2008
- California: plans to introduce hydrogen in 21 interstate highway filling stations

Innovation will also play a role by addressing emissions in other sectors, reducing demand and enabling adaptation to climate change.

Innovation has enabled energy efficiency savings, for example, through compact fluorescent and diode based lights and automated control systems. Furthermore, innovation is likely to continue to increase the potential for energy efficiency savings. Energy efficiency innovation has often been in the form of incremental improvements but there is also a role for more radical progress that may require support. Some markets (such as the cement industry in some developing countries including China and building refurbishment in most countries) are made up of small local firms not large multinationals, which are less likely to undertake research since their resources and potential rewards are smaller. In addition, R&D, for example, in building technologies and urban planning could have a profound impact on the emissions attributed to buildings and increase climate resilience. Chapter 17 discusses energy efficiency in more detail.

[27] E4tech, (2006)

Box 16.2 The scope for innovation to reduce emissions from agriculture

Research into fertilisers and crop varieties associated with lower GHG emissions could help fight climate change[28]. In some instances it may be possible to develop crops that both reduce emissions and have higher yields in a world with more climate change (see Box 26.3).

Another important research area in agriculture will be how to enhance carbon storage in soils, complementing the need to understand emissions from soils (see Section 25.4). The economic potential for enhanced storage is estimated at 1 GtCO2e in 2020, but the technical potential is much greater (see Section 9.6).

Research into sustainable farming practices (such as agroforestry) suitable to local conditions could lead to a reduction in GHG emissions and may also improve crop yields. It could reduce GHG emissions directly by reducing the need to use fertilisers, and indirectly by reducing the emissions from industry and transport sectors to produce the fertiliser[29].

Research into livestock feeds, breeds and feeding practices could also help reduce methane emissions from livestock.

In addition to using biomass energy (see Box 9.5), agriculture, and associated manufacturing industries, have the potential to displace fossil-based inputs for sectors such as chemicals, pharmaceuticals, manufacturing and buildings using a wide range of products made from renewable sources.

Direct emissions from industrial sectors such as cement, chemical and iron and steel can also benefit from further innovation, whether it is in these sectors or in other lower-carbon products that can be substitutes. Innovation in the agricultural sector, discussed in a mitigation context in Box 16.2 above, can also help improve the capacity to adapt to the impacts of climate change. New crop varieties can improve yield resilience to climate change[30]. The Consultative Group on International Agricultural Research (CGIAR) will have a role to play in responding to the climate challenge through innovation in the agricultural sector (see Box 24.4). The development and dissemination of other adaptation technologies is examined in Chapter 19.

16.4 Policy implications for climate change technologies

Policy should be aimed at bringing a portfolio of low-emission technology options to commercial viability

Innovation is, by its nature, unpredictable. Some technologies will succeed and others will fail. The uncertainty and risks inherent in developing low-emission technologies are ideally suited to a portfolio approach. Experience from other areas of investment decisions under uncertainty[31] clearly suggests that the most effective response to the uncertainty of returns is to develop a portfolio. While markets will tend to deliver the least-cost short-term option, it is possible they may ignore technologies that could ultimately deliver huge cost savings in the long term.

As Part III set out, a portfolio of technologies will also be needed to reduce emissions in key sectors, because of the constraints acting on individual technologies. These constraints and energy security issues mean that a portfolio will be required to achieve reductions at the scale required. There is an option value to developing alternatives as it enables greater and potentially less costly abatement in the future. The introduction of new options makes the marginal abatement cost curve (see Section 9.3) more elastic. Early development of economically viable alternatives also avoids the problem of 'locking in' high-carbon capital stock for decades, which would also increase future marginal abatement costs. Policies to encourage low-emission technologies can be seen as a hedge against the risk of high abatement costs.

[28] Norse (2006).
[29] Box 25.4 provides further examples of sustainable farming practices.
[30] IRRI (2006).
[31] Pindyck and Dixit (1994)

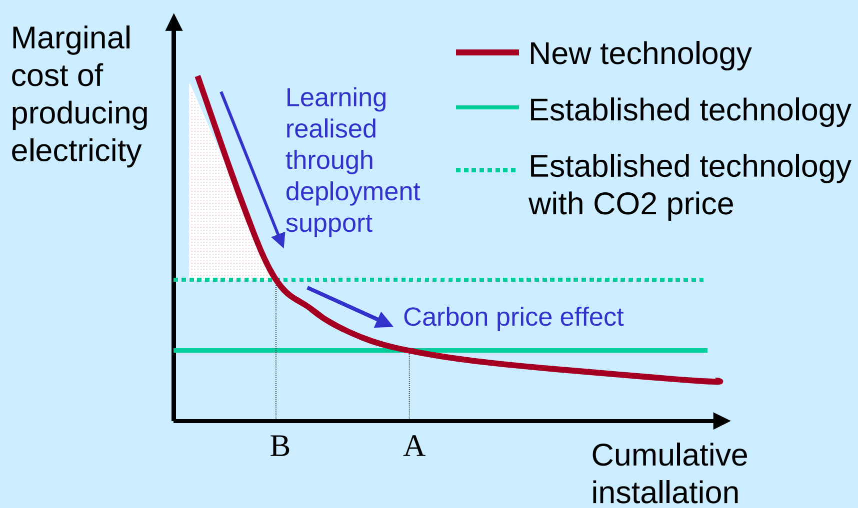
There are costs associated with developing a portfolio. Developing options involves paying the learning cost for more technologies. But policymakers should also bear in mind links to other policy objectives. A greater diversity in sources of energy, for instance, will tend to provide benefits to security of supply, as well as climate change. There is thus a type of externality from creating a new option in terms of risk reduction as well as potential cost reduction. Firms by themselves do not have the same perspective and weight on these criteria as broader society. The next section looks at how the development of a suitable portfolio can be encouraged

Developing a portfolio requires a combination of government interventions including carbon pricing, R&D support and, in some sectors, technology-specific early stage deployment support. These should be complemented by policies to address non-market barriers.

Alongside carbon pricing and the further factors identified in Chapter 17, supporting the development of low-emission technologies can be seen as an important element of climate policy. The further from market the product, given some reasonable probability of success, the greater the prima facie case for policy intervention. In the area of pure research, spillovers can be very significant and direct funding by government support is often warranted. Closer to the market, the required financing flows are larger, and the private returns to individual companies are potentially greater. The government's role here is to provide a credible and clear policy framework to drive private-sector investment.

The area in the innovation process between pure research and technologies ready for commercialisation is more complex. Different sectors may justify different types of intervention. In the electricity market, in particular, deployment policies are likely to be required to bring technologies up to scale. How this support is delivered is important and raises issues about how technology neutral policy should be, which will be discussed later in this chapter in Section 16.6.

Figure 16.6 Interaction between carbon pricing and deployment support[32]



This diagram summarises the links between two of the elements of climate policy. The introduction of the carbon price reduces the learning cost since the new technology, for example a renewable, in this illustrative figure becomes cost effective at point B rather than point A, reducing the size of the learning cost represented by the dotted area. Earlier in the learning curve, deployment support is required to reduce the costs of the technology to the point where the market will adopt the technology. It is the earlier stages of innovation, research, development and demonstration which develop the technology to the point that deployment can begin.

[32] In this figure the policy encourage learning but firms may be prepared to undertake investments in anticipation of technological progress or carbon price incentives.

Across the whole process, non-market barriers need to be identified and, where appropriate, overcome. Without policy incentives when required, support will be unbalanced, and bottlenecks are likely to appear in the innovation process[33]. This would reduce the cost effectiveness at each other stage of support, by increasing the cost of the technology and delaying or preventing its adoption.

Uncertainties, both with respect to climate change and technology development, argue *for* investment in technology development. Uncertainties in irreversible investments argue for postponing policies until the uncertainties are reduced. However, uncertainties, especially with respect to technology development, will not be reduced exogenously with the 'passage of time' but endogenously through investment and the feedback and experience it provides.

Most of the development and deployment of new technologies will be undertaken by the private sector; the role of governments is to provide a stable framework of incentives

Deployment support is generally funded through passing on increased prices to the consumers. But it should still be viewed, alongside public R&D support, as a subsidy and should thus be subject to close scrutiny and, if possible, time limited. The private sector will be the main driver for these new technologies. Deployment support provides a market to encourage firms to invest and relies on market competition to provide the stimulus for cost reductions. Both public R&D and deployment support are expected to have a positive impact on private R&D.

In some sectors the benefits from innovation can be captured by firms without direct support for deployment, other than bringing down institutional barriers and via setting standards. This is particularly so in sectors that rely on incremental innovations to improve efficiency rather than a step change in technology, since the cost gap is unlikely to be so large. In these sectors firms may be comfortable to invest in the learning cost of developing low-emission technologies.

Firms with products that are associated with greenhouse gas emissions are increasingly seeking to diversify in order to ensure their long-run profitability. Oil firms are increasingly investing in low-emission energy sources. General Electric's Ecomagination initiative has seen the sale of energy efficient and environmentally advanced products and services rise to $10.1 billion in 2005, up from $6.2 billion in 2004 - with orders nearly doubling to $17 billion. GE's R&D in cleaner technologies was $700m in 2005 and expected to rise to $1.5 billion per annum by 2010.[34] Indeed in a number of countries the private sector is running ahead of government policy and taking a view on where such policy is likely to go in the future which is in advance of what the current government is doing.

R&D and deployment support have been effective in encouraging the development of generation technologies in the past

Determining the benefits of both R&D and deployment is not easy. Studies have often successfully identified a benefit from R&D but without sufficient accuracy to determine what the appropriate level of R&D should be. Estimating the appropriate level is made more difficult by the broad range of activities that can be classed as R&D. Ultimately the benefits of developing technologies will depend on the amount of abatement that is achieved (and thus the avoided impacts) and the long-term marginal costs of abating across all the other sectors within the economy (linked to the carbon price), both of which are uncertain.

However, some evidence provides indications of the effectiveness of policy in promoting the development of technologies:

- **Estimates of R&D benefits**. Private returns from economy-wide R&D have been estimated at 20-30% whilst the estimated social rate of return was around 50%[35].

[33] Weak demand-side policies risk wasting R&D investments see Norberg-Bohm and Loiter (1999) and Deutch (2005)
[34] Source GE press release May 2006: http://home.businesswire.com/portal/site/ge/index.jsp?ndmViewId=news_view&newsId=20060517005223&newsLang=en&ndmConfigId=1001109&vnsId=681
[35] Kammen and Margolis (1999)

While it is private-sector not public-sector R&D that has been positively linked with growth, the public-sector R&D can play a vital role in stimulating private spending up to the potential point of crowding out[36]. It also plays an important role in preserving the 'public good' nature of major scientific advances. Examples of valuable breakthroughs stimulated by public R&D must be weighed up alongside examples of wasteful projects.

- **Historical evidence**. Examining the history of existing energy technologies and the prominent role that public R&D and initial deployment have played in their development illustrates the potential effectiveness of technology policy. Extensive and prolonged public support and private markets were both instrumental in the development of all generating technologies. Military R&D, the US space programme and learning from other markets have also been crucial to the process of innovation in the energy sector. This highlights the spillovers that occur between sectors and the need to avoid too narrow an R&D focus. This experience has been mirrored in other sectors such as civil aviation and digital technologies where the source has also been military. Perhaps this is related to the fact that US public defence R&D was eight times greater than that for energy R&D in 2006 (US Federal Budget Authority). Historical R&D and deployment support has delivered the technological choices of the present with many R&D investments that may have seemed wasteful in the 1980s, such as investments in renewable energy and synfuels, now bearing fruit. The technological choices of the coming decades are likely to develop from current R&D.

Box 16.3 Development of existing technology options[37]

Nuclear: From the early stages of the Cold War, the Atomic Energy Commission in the US, created primarily to oversee the development of nuclear weapons, also promoted civilian nuclear power. Alic et al[38] argue that by exploiting the 'peaceful atom' Washington hoped to demonstrate US technological prowess and perhaps regain moral high ground after the atomic devastation of 1945. The focus on weapons left the non-defence R&D disorganised and starved of funds and failed to address the practical issues and uncertainties of commercial reactor design. The government's monopoly of nuclear information, necessary to prevent the spreading of sensitive information, meant state R&D was crucial to development.

Gas: The basic R&D for gas turbine technology was carried out for military jet engines during World War II. Since then developments in material sciences and turbine design have been crucial to the technological innovation that has made gas turbines the most popular technology for electricity generation in recent years. Cooling technology from the drilling industry and space exploration played an important role. In the 1980s improvements came from untapped innovations in jet engine technology from decades of experience in civil aviation. Competitive costs have also been helped by low capital costs, reliability, modularity and lower pollution levels.

Wind: The first electric windmills were developed in 1888 and reliable wind energy has been available since the 1920s. Stand-alone turbines were popular in the Midwestern USA prior to centrally generated power in the 1940s. Little progress was made until the oil shocks led to further investment and deployment, particularly in Denmark (where a 30% capital tax break (1979-1989) mandated electricity prices (85% of retail) and a 10% target in 1981 led to considerable deployment) and California where public support led to extensive deployment in the 1980s. Recent renewable support programmes and technological progress have encouraged an average annual growth rate of over 28 % over the past ten years[39].

Photovoltaics: The first PV cells were designed for the space programme in the late 1950s. They were very expensive and converted less than 2% of the solar energy to electricity. Four decades of steady development, in the early phases stimulated by the space programme, have seen efficiency rise to nearly 25% of the solar energy in laboratories, and costs of commercial cells have fallen by orders of magnitude. The need for storage or ancillary power

[36] When public expenditure limits private expenditure by starving it of potential resources such as scientists OECD (2005)
[37] Alic, Mowery and Rubin (2003)
[38] Alic, Mowery and Rubin (2003)
[39] Global Wind Energy Council http://www.gwec.net/index.php?id=13

sources have held the technology back but there have been some niche markets in remote locations and, opportunities to reduce peak demand in locations where solar peaks and demand peaks coincide.

Public support has been important. A study by Norberg-Bohm[40] found that, of 20 key innovations in the past 30 years, only one of the 14 they could source was funded entirely by the private sector and nine were totally public. Recent deployment support led the PV market to grow by 34% in 2005. Nemet[41] explored in more detail how the innovation process occurred. He found that, of recent cost reductions, 43% were due to economies of scale, 30% to efficiency gains from R&D and learning-by-doing, 12% due to reduced silicon costs (a spillover from the IT industry).

- **Learning curve analysis**. Learning curves, as shown in Box 9.4 and in other studies[42], show that increased deployment is linked with cost reductions suggesting that further deployment will reduce the cost of low-emission technologies. There is a question of causation since cost reductions may lead to greater deployment; so attempts to force the reverse may lead to disappointing learning rates. The data shows technologies starting from different points and achieving very different learning rates. The increasing returns from scale shown in these curves can be used to justify deployment support, but the potential of the technologies must be evaluated and compared with the costs of development.

16.5 Research, development and demonstration policies

Government has an important role in directly funding skills and basic knowledge creation for science and technology

At the pure science end of the spectrum, the knowledge created has less direct commercial application and exhibits the characteristics of a 'public good'. At the applied end of R&D, there is likely to be a greater emphasis on private research, though there still may be a role for some public funding.

Governments also fund the education and training of scientists and engineers. Modelling for this review suggests that the output of low-carbon technologies in the energy sector will need to expand nearly 20-fold over the next 40-50 years to stabilise emissions, requiring new generations of engineers and scientists to work on energy-technology development and use. The prominent role of the challenge of climate change may act as an inspiration to a new generation of scientists and spur a wider interest in science.

R&D funding should avoid volatility to enable the research base to thrive. Funding cycles in some countries have exhibited 'roller-coaster' variations between years, which have made it harder for laboratories to attract, develop, and maintain human capital. Such volatility can also reduce investors' confidence in the likely returns of private R&D. Kammen[43] found levels changed by more than 30% in half the observed years. Similarly it may be difficult to expand research capacity very quickly as the skilled researchers may not be available. Governments should seek to avoid such variability, especially in response to short-term fuel price fluctuations. The allocation of public R&D funds should continue to rely on the valuable peer review process and this should include post-project evaluations and review to maximise the learning from the research. Research with clear objectives but without over-commitment to narrow specifications or performance criteria can eliminate wasteful expenditures[44] and allow researchers more time to apply to their research interests and be creative.

Governments should seek to ensure that, in broad terms, the priorities of publicly funded institutions reflect those of society. The expertise of the researchers creates an information asymmetry with policymakers facing a challenge in selecting suitable projects. Arms-length

[40] Norberg-Bohm (2000)
[41] Source: Nemet, in press
[42] For an example Taylor, Rubin and Nemet (2006)
[43] Kammen (2004)
[44] Newell and Chow (2004)

organisations and expert panels such as research-funding bodies may be best placed to direct funding to individual projects.

Three types of funding are required for university research funding.

- Basic research time and resources for academic staff to pursue research that interests them.
- Research programme funding (such as research councils) that directs funding towards important areas.
- Funding to encourage the transfer of knowledge outside the institution. The dissemination of information encourages progress to be applied and built on by other researchers and industry and ensures that it not be unnecessarily duplicated elsewhere.

Research should cover a broad base and not just focus on what are currently considered key technologies, including basic science and some funding to research the more innovative ideas[45] to address climate change. Historical examples of technological progress when the research was not directed towards specific economic applications (such as developments in nanotechnology, lasers and the transistor) highlight the importance of open-ended problem specification. There must be an appropriate balance between basic science and applied research projects[46]. Increases in energy R&D (as discussed in the final section of this chapter) can be complemented by increased funding for science generally. The potential scale of increase in basic science will vary by country depending on their current level and research capabilities[47].

There may also be a case for demonstration funding to prove viability and reduce risk. An example of this is the UK DTI's 'Wave and Tidal Stream Energy Demonstration Scheme' that will support demonstration projects undertaken by private firms. This has many features to encourage the projects and maximise learning through provision of test site and facilities and systematic comparison of competing alternatives. Governments can help such projects through providing infrastructure. Demonstration projects are best conducted or at least managed by the private sector.[48]

Energy storage is worthy of particular attention

Inherent uncertainty on fruitful areas of research ensures governments should be cautious against picking winners. However, some areas of research suggest significant potential through a combination of probability of success, lead-times and global reward for success. Priorities for scientific progress in the energy sector should include PV (silicon and non-silicon based), biofuel conversion technologies, fusion, and material science.

As markets expand, all the key low carbon primary energy sources will run into constraints. Nuclear power will be confined to base-load electricity generation unless energy storage is available to enable its energy to follow loads and contribute to the markets for transport fuels. Intermittent renewable energy forms with backup generation will face the same problem. Electricity generation from fossil fuels with carbon capture and storage will likewise be unable to enter the transport markets unless improved and lower cost forms of hydrogen storage or new battery technology are developed. Solar energy can in theory meet the world's energy needs many times over, but will, like energy from wind, waves and tides, eventually depend on the storage problem being solved.

The analysis of the costs of climate change mitigation in Chapter 9 provides further confirmation of the need for an expansion of RD&D activities in energy storage technologies. A failure to develop such technologies will inevitably increase the costs of mitigation once low-emission options for electricity generation are exploited. In contrast, success in this area will

[45] For some examples, see Gibbs (2006)
[46] Newell and Chow (2004)
[47] In 2004 the UK Government published a ten-year Science and Innovation Investment Framework, which set a challenging ambition for public and private investment in R&D to rise from 1.9% to 2.5% of UK GDP, in partnership with business; as well as the policies to underpin this. An additional £1 billion will be invested in science and innovation between 2005-2008, equivalent to real annual growth of 5.8% and to continue to increase investment in the public science base at least in line with economic growth. http://www.dti.gov.uk/science/science-funding/framework/page9306.html
[48] Newell and Chow (2004)

allow low-emission sources to provide energy in other sectors, such as transport. Current R&D and demonstration efforts on hydrogen production and storage along with other promising options for storing energy (such as advanced battery concepts) should be increased. This should include research on devices that convert the stored energy, such as the fuel cell.

In the case of applied energy research, partnership between the public and private sectors is key

It is important that public R&D leverages private R&D and encourages commercialisation. Ultimately the products will be brought into the market by private firms who have a better knowledge of markets, and, so it is important that public R&D maintains the flow of knowledge by ensuring public R&D complements the efforts of the private sector.

The growth and direction of private R&D efforts will be a product of the incentives for low-emission investments provided by the structure of markets and public policies. Public R&D should aim to complement, not compete, with private R&D, generally by concentrating on more fundamental, longer-term possibilities, and by sharing in the risks of some larger-scale projects such as CCS. In many areas the private sector will make research investments without public support, as has been the case recently on advanced biofuels (see Box 16.4).

Box 16.4 Second generation biofuels

Cellulosic ethanol is a not-yet-commercialized fuel derived from woody biomass. In his 2006 State of the Union address, Bush praised the fuel's potential to curb the nation's "addiction to foreign oil". A joint study by the Departments of Agriculture and Energy[49] concludes that U.S. biomass feedstocks could produce enough ethanol to displace 30 percent of the nation's gasoline consumption by 2030.

In May 2006, Goldman Sachs & Co became the first major Wall Street firm to invest in the technology. Goldman Sachs & Co invested more than $26 million in Iogen Corp., an Ottawa-based company that operates the world's first and only demonstration facility that converts straw, corn stalks, switchgrass and other agricultural materials to ethanol. Iogen hopes to begin construction on North America's first commercial cellulosic ethanol plant next year.

In September 2006 Richard Branson announced plans to invest $3 billion in mitigating climate change. Some of this will be invested in Virgin Fuels, which will develop biofuels including cellulosic ethanol.

The OECD[50] found that economic growth was closely linked to general private R&D, not public R&D, but that public R&D plays a vital role in stimulating private spending. There is evidence[51] from the energy sector that patents do track public R&D closely, which suggests that they successfully spur innovation and private sector innovation. R&D collaboration between the public and private-sector is one way of reducing the cost and risks of R&D.

The public sector could fund private sector research through competitive research funding, with private sector companies bidding for public funds as public organisations currently do from research councils. Prizes to reward innovation can be used to encourage breakthroughs. Historically they have proved very successful but defining a suitable prize can be problematic[52]. An alternative approach, as suggested for the pharmaceutical sector, is to commit to purchase new products to reward those that successfully innovate.[53]

[49] US Departments of Agriculture and Energy (2005)
[50] OECD (2005)
[51] Kammen and Nemet (2005)
[52] Newell and Wilson (2005)
[53] Kremer and Glennerster (2004)

Box 16.5 Public-private research models - UK Energy Technologies Institute[54]

In 2006, the UK launched the Energy Technologies Institute (ETI). It will be funded on a 50:50 basis between private companies and the public sector with the government prepared to provide £500 million, creating the potential for a £1 billion institute over a minimum lifetime of ten years.

The institute will aim to accelerate the pace and volume of research directed towards the eventual deployment of the most promising research results. ETI will work to existing UK energy policy goals including a 60% reduction in emissions by 2050.

The ETI will select, commission, fund, manage and, where appropriate, undertake research programmes. Most investment will focus on a small number of key technology areas that have greatest promise for deployment and contributing to low-emission secure energy supplies.

16.6 Deployment policy

A wide range of policies to encourage deployment are already in use.

In addition to direct emissions pricing through taxes and trading and R&D support, there are strong arguments in favour of supporting deployment in some sectors when spillovers, lock-in to existing technologies, or capital market failures prevent the development of potentially low-cost alternatives. Without support the market may never select those technologies that are further from the market but may nevertheless eventually prove cheapest. Policies to support deployment exist throughout the world including many non-OECD countries[55]. China and India have both encouraged large-scale renewable deployment in recent years and now have respectively the largest and fifth largest renewable energy capacity worldwide[56].

There is some deployment support for clean technologies in most developed countries. The mechanism of support takes many forms though the costs are generally passed onto the consumer. The presence of a carbon price reduces the cost and requirement for deployment support. Deployment support is generally a small component of price when spread across all consumption (see Box 16.7) but does add to the impact of carbon pricing on electricity prices. Policymakers should consider the impact of deployment support on energy prices over time. Consumers will be paying for the development of technologies that benefit consumers in the future.

[54] http://www.dti.gov.uk/science/science-funding/eti/page34027.html
[55] Page 20 REN 21 Renewables global status report 2005 - See page 20 REN 21 (2005)
[56] Figures from 2005 - excluding large scale hydropower. Page 6 REN 21 (2006)

Box 16.6 Examples of existing deployment incentives

- **Fiscal incentives**: including reduced taxes on biofuels in the UK and the US; investment tax credits.
- **Capital grants** for demonstrator projects and programmes: clean coal programmes in the US; PV 'rooftop' programmes in the US, Germany and Japan; investments in marine renewables in the UK and Portugal; and numerous other technologies in their demonstration phase.
- **Feed-in tariffs** are a fixed price support mechanism that is usually combined with a regulatory incentive to purchase output: examples include wind and PVs in Germany; biofuels and wind in Austria; wind and solar schemes in Spain, supplemented by 'bonus prices'; wind in Holland.
- **Quota based schemes**: the Renewable Portfolio Standards in twenty three US States; the vehicle fleet efficiency standards in California
- **Tradable quotas**: the Renewables Obligation and Renewable Transport Fuels Obligation in the UK.
- **Tenders for tranches of output** (the former UK Non Fossil Fuel Obligation) with increased output prices subsidised out of the revenues from a general levy on electricity tariffs.
- **Subsidy** of the infrastructure costs of connecting new technologies to networks.
- **Procurement policies of public monopolies:** This was the approach historically of the public monopolies in electricity for purchase of nuclear power throughout the OECD; it is currently the approach in China. It is often combined with regulatory agreements to permit recovery of costs, soft loans by governments, and, in the case of nuclear waste, government assumption of liabilities.
- **Procurement policies of national and local governments**: these include demonstrator projects on public buildings; use of fuel cells and solar technologies by defence and aerospace industries; hydrogen fuel cell buses and taxis in cities; energy efficiency in buildings.

The deployment mechanisms described in Box 16.6 can be characterised as price or quantity support, with some tradable approaches containing elements of both. The costs of these policies are generally passed directly on to consumers though some are financed from general taxation. When quantity deployment instruments are not tradable, the policymaker should consider whether there are sufficient incentives to strive for cost reductions and whether the supplier can profit from passing an excessive cost burden onto the consumer. If the level of a price deployment instrument is too low no deployment will occur, while if it is too high large volumes of deployment will occur with financial rewards for participants which are essentially government created rents. With tradable quantity instruments, the market is left to determine the price, usually with tradable certificates between firms. This does lead to price uncertainty. If the quantity is too high, bottlenecks may lead to a high cost. If the quantity is too low, there may not be sufficient economies of scale to reduce the cost.

Both sets of instruments have proved effective but existing experience favours price-based support mechanisms. Comparisons between deployment support through tradable quotas and feed-in tariff price support suggest that feed-in mechanisms achieve larger deployment at lower costs[57]. Central to this is the assurance of long-term price guarantees. The German scheme, as described in Box 16.7 below, provides legally guaranteed revenue streams for up to twenty years if the technology remains functional. Whilst recognising the importance of planning regimes for both PV and wind, the levels of deployment are much greater in the German scheme and the prices are lower than comparable tradable support mechanisms (though greater deployment increases the total cost in terms of the premium paid by consumers). Contrary to criticisms of the feed-in tariff, analysis suggests that competition is greater than in the UK Renewable Obligation Certificate scheme. These benefits are logical as the technologies are already prone to considerable price uncertainties and the price uncertainty of tradable deployment support mechanisms amplifies this uncertainty. Uncertainty discourages investment and increases the cost of capital as the risks associated with the uncertain rewards require greater rewards.

[57] Butler and Neuhoff (2005); EC (2005); Ragwitz, and Huber (2005); Fouquet et al (2005)

Box 16.7 Deployment support in Germany

Feed-in tariffs have been introduced in Germany to encourage the deployment of onshore and offshore wind, biomass, hydropower, geothermal and solar PV[58]. The aim is to meet Germany's renewable energy goals of 12.5% of gross electricity consumption in 2010 and 20% in 2020. The policy also aims to encourage the development of renewable technologies, reduce external costs and increase the security of supply.

Each generation technology is eligible for a different rate. Within technologies the rate varies depending on the size and type. Solar energy receives between €0.457 to 0.624 per kWh while wind receives €0.055 to 0.091per kWh. Once the technology is built the rate is guaranteed for 20 years. The level of support for deployment in subsequent years declines over time by 1% to 6.5% each year with the rate of decline derived from estimated learning curves[59].

In 2005 10.2% of electricity came from renewables (70% supported with feed-in tariffs) the Federal Environment Ministry (BMU) estimate that the current act will save 52 million tonnes on CO_2 in 2010. The average level of feed-in tariff was €0.0953 per kWh in 2005 (compared to an average cost of displaced energy of €0.047 kWh). The total level of subsidy was €2.4 billion Euro at a cost shared all consumers of €0.0056 per kWh (3% of household electricity costs)[60]. There are an estimated 170,000 people working in the renewable sector with an industry turnover of €8.7 billion.[61]

The 43.7 TWh of electricity covered by the feed in tariffs was split mostly between wind (61%), biomass (19%) and hydropower (18%). It has succeeded in supporting several technologies. Solar accounted for 2% (0.2% of total electricity) with an average growth rate of over 90% over the last four years. Despite photovoltaic's low share Germany has a significant proportion of the global market with 58% of the capacity installed globally in 2005 (39% of the total installed capacity) and 23% of global production.[62]

Regulation can also be used to encourage deployment, for example by reducing uncertainty and accelerating spillover effects, and may be preferable in certain markets (see Chapter 17 for details). Performance standards encourage uptake and innovation in efficient technologies by establishing efficiency requirements for particular goods, in particular encouraging incremental innovation Alternatively, technology specific design standards can be targeted directly at the cleanest technologies by mandating their application or banning alternatives.

There are already considerable sums of money spent on supporting technology deployment. It is estimated that $10 billion[63] was spent in 2004 on renewable deployment, around $16 billion is spent each year supporting existing nuclear energy and around $6.4billion[64] is spent each year supporting biofuels. The total support for these low-carbon energy sources is thus $33 billion each year. Such sums are dwarfed by the existing subsidies for fossil fuels worldwide that are estimated at $150 billion to 250 billion each year. All these costs are generally paid by the consumer.

Technology-neutral incentives should be complemented by focused incentives to bring forward a portfolio of technologies

Policy frameworks can be designed to treat support to all low-carbon technologies in a 'technology-neutral' way. The dangers of public officials 'picking winners' should point to this

[58] Originally introduced in 1991 with the Electricity Feed Act this was replaced in 2000 with the broader Act on Granting Priority to Renewable Energy Sources (Renewable Energy Sources Act) and amended in 2004 http://www.ipf-renewables2004.de/en/dokumente/RES-Act-Germany_2004.pdf
[59] Small hydropower does not decline and is guaranteed for 30 years and large hydropower only 15 years.
[60] BMU (2006a)
[61] BMU (2006b)
[62] http://www.iea-pvps.org/isr/index.htm
[63] Deployment share of figure page 16 REN 21, 2005 grossed up to global figure based on IEA deployment figures. Nuclear figure from same source.
[64] Based on global production of 40 billion litres and on an average support of £0.1 per litre and a PPP exchange rate of $1.6 to £1

as the starting point in most sectors. Markets and profit orientated decisions, where the decision maker is forced to look carefully at cost and risk are better at finding the likely commercial successes. However, the externalities, uncertainties and capital market problems in some sectors combine with the urgency of results and specificity of some of the technological problems that need to be solved when tackling climate change, all point to the necessity to examine the issues around particular technologies and ensure that a portfolio develops.

The policy framework of deployment support could differentiate between technologies, offering greater support to those further from commercialisation, or having particular strategic or national importance. This differentiation can be achieved several ways, including technology-specific quotas, or increased levels of price support for certain technologies. Policies to correct the carbon externality (taxes / trading) are, and should continue to be, technology neutral. Technology neutrality is also desirable for deployment support if the aim is to deliver least cost reductions to meet short-term targets, since the market will deliver the least-cost technology.

However, as has already been discussed, the process of learning means that longer-established technologies will tend to have a price advantage over newer technologies, and untargeted support will favour these more developed technologies and bring them still further down the learning curve. This effect can be seen in markets using technology-neutral instruments: in the USA, onshore wind accounts for 92% of new capacity in green power markets[65].

This concentration on near-to-market technologies will tend to work to the exclusion of other promising technologies, which means that only a very narrow portfolio of technologies will be supported, rather than the broad range which Part III of this report shows are required. This means technology neutrality may be cost efficient in the short term, but not over time.

Most deployment support in the electricity generation sector has been targeted towards renewable and nuclear technologies. However, significant reductions are also expected from other sources. As highlighted in Box 9.2 carbon capture and storage (CCS) is a technology expected to deliver a significant portion of the emission reductions. The forecast growth in emissions from coal, especially in China and India, means CCS technology has particular importance. Failure to develop viable CCS technology, while traditional fossil fuel generation is deployed across the globe, risks locking-in a high emissions trajectory. The demonstration and deployment of CCS is discussed in more detail in Chapter 24. Stabilising emissions below 550ppm CO_2e will require reducing emissions from electricity generation by about 60%[66]. Without CCS that would require a dramatic shift away from existing fossil-fuel technologies.[67]

Policies should have a clear review process and exit strategies, and governments must accept that some technologies will fail.

Uncertainty over the economies of scale and learning-by-doing means that some technological failures are inevitable. Technological failures can still create valuable knowledge, and the closing of technological avenues narrows the investment options and increases confidence in other technologies (as they face less alternatives). The Arrow-Lind theorem[68] states that governments are generally large enough to be risk neutral as they are large enough to spread the risk and thus have a role to play in undertaking riskier investments. It is not a mistake per se to buy insurance or a hedge that later is not needed and that is in many ways a suitable analogy for fostering a wider portfolio of viable technologies than the market would do by itself[69].

Credibility is also important to policy design. Policies benefit from providing clear, bankable, signals to business. There is a role for monitoring and for a clear exit strategy to prevent excessive costs and signal the ultimate goal of these policies: competition on a level playing

[65] Bird and Swezey (2005)
[66] This is consistent with the IEA ACT scenarios see Box 9.7
[67] For more on CCS see Boxes 9.2 and 24.8 and Section 24.3
[68] Arrow and Lind (1970)
[69] Deutch (2005)

field. A good example has been the Japanese rebates in the 'Solar Roofs' programme, which have declined gradually over time, from 50% of installed cost in 1994 to 12% in 2002 when the scheme ended.

Alternative approaches can also help spur the deployment of new innovations. For example, extension services, the application of scientific research and new knowledge to agricultural practices through farmer education, had a significant impact on the deployment of new crop varieties during the Green Revolution. Also, organisations such as the Carbon Trust in the UK, Sustainable Development Technologies Canada, established by governments but independent of them to allow the application of business acumen, have proved successful in encouraging investment in the development and demonstration of clean technologies. They can play an important role at each stage of the technology process, from R&D to ensuring their widespread deployment once they have become cost effective. They have proved especially successful in acting as a "stamp of approval" that spurs further venture capital investment. Finding niche markets and building these into large-scale commercialisation opportunities is a key challenge for companies with promising low carbon technologies. These organisations are at the forefront of identifying niche markets for commercialisation of new technologies and promoting public-private investment in deployment.

16.7 Other supporting policies

***Other policies have an important impact on the viability of technologies*.**

There are many other policy options available to governments that can affect technology deployment and adoption. Governments set policies such as the planning regime and building standards. How these are set can have an important impact on the adoption of new technologies. They can constrain deployment either directly or indirectly by increasing costs. Regulations can stifle innovation, but if well designed they can drive innovation. Depending how these are set, they can act as a subsidy to low-emission alternative technologies or to traditional fossil fuels. Setting the balance is difficult, since their impacts are hard to value. But they must be considered since they can have an important effect on the outcome.

- The intellectual property regime can act as an incentive to the innovator, but the granting of the property right can also slow the dissemination of technological progress and prohibit others from building on this innovation. Managing this balance is an important challenge for policymakers.

- Planning and licensing regulations have proven a significant factor for nuclear, wind and micro-generation technologies. Planning can significantly increase costs or, in many cases, prevent investments taking place. Local considerations must be set against wider national or global concerns.

- It is important how governments treat risks and liabilities such as waste, safety or decommissioning costs for nuclear power or liabilities for CO_2 leakage from CCS schemes. Governments can bear some of these costs but, unless suppliers and ultimately consumers are charged for this insurance, it will be a subsidy.

- Network issues are particularly important for energy and transport technologies. The existing transport network and infrastructure, especially fuel stations, is tailored to fossil fuel technologies.

- Intermittent technologies such as wind and solar may be charged a premium if they require back-up sources. How this is treated can directly affect economic viability, depending on the extent of the back-up generation required and the premium charged.

- Micro-generation technologies can sell electricity back to the grid and do not incur the same distribution costs and transmission losses as traditional much larger sources. The terms under which such issues are resolved has an important impact on the economics of these technologies. Commercially proven low-carbon technologies require regulatory frameworks that recognise their value, in terms of flexibility and modularity[70], within a distributed energy system. Regulators should innovate in response to the challenge of integrating these technologies to exploit their potential, and unlock the resultant opportunities that arise from shifting the generation mix away from centralised sources.

- Capacity constraints may arise because of a shortage in a required resource. For example, there may be a shortage of skilled labour to install a new technology.

- There are other institutional and even cultural barriers that can be overcome. Public acceptability has proven an issue for both wind and nuclear and this may also be the case for hydrogen vehicles. Consumers may have problems in finding and installing new technologies. Providing information of the risks and justification of particular technologies can help overcome these barriers.

16.8 The scale of action required

Extending and expanding existing deployment incentives will be key

Deployment policies encourage the private sector to develop and deploy low-carbon technologies. The resulting cost reductions will help reduce the cost of mitigation in the future (as explained in Chapter 10). Consumers generally pay the cost of deployment support in the form of higher prices. Deployment support represents only a proportion of the cost of the technology as it leverages private funds that pay for the market price element of the final cost.

It is estimated that existing deployment support for renewables, biofuels and nuclear energy is $33 billion each year (see Section 16.6). The IEA's Energy Technology Perspectives[71] looks at the impact of policies to increase the rate of technological development. It assumes that $720billion of investment in deployment support occurs over the next two to three decades. This estimate is on top of an assumed carbon price (whether through tax, trading or implicitly in regulation) of $25 per tonne of CO_2. If the IEA figure is assumed to be additional to the existing effort, it suggests an increase of deployment incentives of between 73% and 109%, depending on whether this increase is spread over two or three decades.

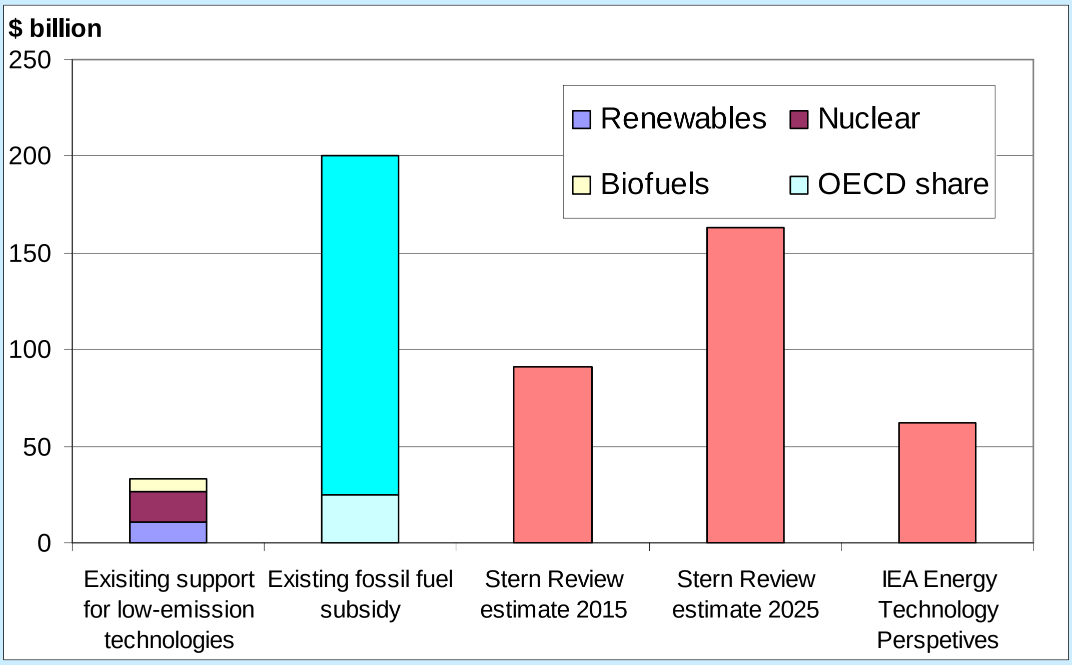
The calculations shown in Section 9.8 include estimates of the level of deployment incentives required to encourage sufficient deployment of new technologies (consistent with a 550ppm CO_2e stabilisation level). The central estimates from this work are that the level of support required will have to increase deployment incentives by 176% in 2015 and 393% in 2025[72]. These estimates are additional to an assumed a carbon price at a level of $25 per tonne of CO_2.

At this price the abatement options are forecast to become cost effective by 2075 so the level of support tails off to zero by this time. If policies lead to a price much higher than this before the technologies are cost effective then less support will be required. Conversely if no carbon price exists the level of support required will have to increase (by a limited amount initially but by much larger amounts in the longer term). While most of this cost is expected to be passed on to consumers, firms may be prepared to incur a proportion of this learning cost in order to gain a competitive advantage.

Such levels of support do represent significant sums but are modest when compared with overall levels of investment in energy supply infrastructure ($20 trillion up to 2030[73]) or even estimates of current levels of fossil-fuel subsidy as shown in the graph below.[74]

[70] Small-scale permits incremental additions in capacity unlike large technologies such as nuclear generation.
[71] Page 58, IEA (2006)
[72] See papers by Dennis Anderson available at www.sternreview.org.uk
[73] IEA (in press)
[74] In this graph mid points in the fossil fuel subsidy range is used in and the IEA increase made over a 20 year period.

Figure 16.7 Estimated scale of current and necessary global deployment support



The level of support required to develop abatement technologies depends on the carbon price and the rate of technological progress, which are both uncertain. It is clear from these numbers that the level of support should increase in the decades to come, especially in the absence of carbon pricing. Based on the numbers above, an increase of 2-5 times current levels over the next 20 years should help encourage the requisite levels of deployment though this level should be evaluated as these uncertainties are resolved.

The scale is, however, not the only issue. It is important that this support is well structured to encourage innovation at low cost. A diverse portfolio of investments is required as it is uncertain which technologies will prove cheapest and constraints on individual technologies will ensure that a mix is necessary. Those technologies that are likely to be the cheapest warrant more investment and these may not be those that are the currently the lowest cost. This requires a reorientation of public support towards technologies that are further from widespread diffusion.

Some countries are already offering significant support for new technologies but globally this support is patchy. Issues on coordinating deployment support internationally to achieve the required diversity and scale are examined in Chapter 24.

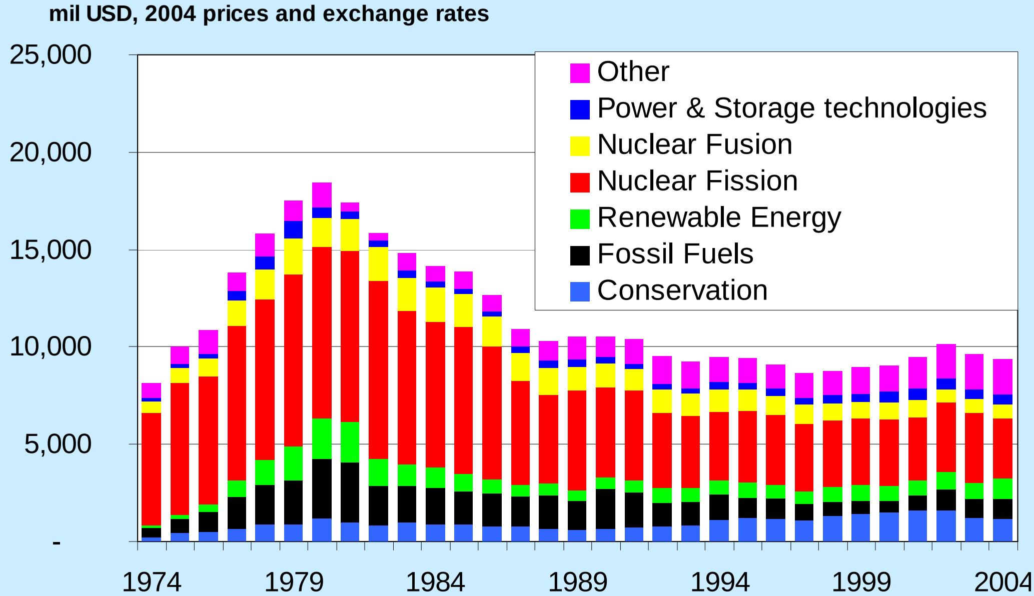
Global energy R&D funding is at a low level and should rise

Though benefits of R&D are difficult to evaluate accurately a diverse range of indicators illustrate the benefits of R&D investments. Global public energy R&D support has declined significantly since the 1980s and this trend should reverse to encourage cost reductions in existing low-carbon technologies and the development of new low-carbon technological options. The IEA R&D database shows a decline of 50% in low-emission R&D[75] between 1980 and 2004. This decline has occurred while overall government R&D has increased significantly[76]. A recent IEA publication on RD&D priorities[77] strongly recommends that governments consider restoring their energy RD&D budgets at least to the levels seen, in the early 1980s. This would involve doubling the budget from the current level of around $10

[75] For countries available includes renewables, conservation and nuclear. The decline is 36% excluding nuclear.
[76] OECD R&D database shows total public R&D increasing by nearly 50% between 1988 and 2004 whilst public energy R&D declined by nearly 20% over the same period.
[77] Page 19 OECD (2006)

billion[78]. This is an appropriate first step that would equate to global levels of public energy R&D around **$20 billion** each year.

Figure 16.8 Public energy R&D in IEA countries[79]



The directions of the effort should also change. A generation ago, the focus was on nuclear power and fossil fuels, including synthetic oil fuels from gas and coal, with comparatively few resources expended on conservation and renewable energy. Now the R&D efforts going into carbon capture and storage, conservation, the full range of renewable energy technologies, hydrogen production and use, fuel cells, and energy storage technologies and systems should all be much larger.

A phased increase in funding, within established frameworks for research priorities, would allow for the expansion in institutional capacity and increased expertise required to use the funding effectively. A proportion of this public money should target be designed to encourage private funds, as is proposed for the UK's Energy Technology Institute (see Box 16.5).

Private R&D should rise in response to market signals. Private energy R&D in OECD countries fell in recent times from around $8.5bn at the end of the 1980s to around $4.5bn in 2003[80]. Significant increases in public energy R&D and deployment support combined with carbon pricing should all help reverse this trend and encourage an upswing in private R&D levels.

This is not just about the total level of support. How this money is spent is crucial. It is important that the funding is spread across a wide range of ideas. It is also important that it is structured to provide stability to researchers while still providing healthy competition. There should be rigorous assessment of these expenditures to ensure that they maintained at an appropriate level. Approaches to encourage international co-operation to achieve these goals are explored in Chapter 24.

16.9 Conclusions

This chapter explores the process of innovation and discovers that externality from the environmental impact of greenhouse gas emissions exacerbates existing market imperfections, limiting the incentive to develop low-carbon technologies. This provides a

[78] 2005 figure Source: IEA R&D database http://www.iea.org/Textbase/stats/rd.asp
[79] Source: IEA Energy R&D Statistics
[80] Page 35, OECD (2006)

strong case for supporting the development of new and existing low-carbon technologies, particularly in a number of key climate change sectors. The power of market forces is the key driver of innovation and technical change but this role should be supplemented with direct public support for R&D and, in some sectors, policies designed to create new markets. Such policies are required to deliver an effective portfolio of low-carbon technologies in the future.

Climatic Change (2016) 134:457–474
DOI 10.1007/s10584-014-1281-2

The impacts of climate change across the globe: A multi-sectoral assessment

N. W. Arnell · S. Brown · S. N. Gosling · P. Gottschalk · J. Hinkel · C. Huntingford · B. Lloyd-Hughes · J. A. Lowe · R. J. Nicholls · T. J. Osborn · T. M. Osborne · G. A. Rose · P. Smith · T. R. Wheeler · P. Zelazowski

Received: 30 August 2013 / Accepted: 16 October 2014 / Published online: 11 November 2014

Abstract The overall global-scale consequences of climate change are dependent on the distribution of impacts across regions, and there are multiple dimensions to these impacts. This paper presents a global assessment of the potential impacts of climate change across several sectors, using a harmonised set of impacts models forced by the same climate and

This article is part of a Special Issue on "The QUEST-GSI Project" edited by Nigel Arnell.

Electronic supplementary material The online version of this article (doi:10.1007/s10584-014-1281-2) contains supplementary material, which is available to authorized users.

N. W. Arnell (✉) · B. Lloyd-Hughes · T. M. Osborne · G. A. Rose · T. R. Wheeler
Walker Institute, University of Reading, Reading, UK
e-mail: n.w.arnell@reading.ac.uk

S. Brown · R. J. Nicholls
University of Southampton and Tyndall Centre for Climate Change Research, Southampton, UK

S. N. Gosling
University of Nottingham, Nottingham, UK

P. Gottschalk · J. Hinkel
PIK Potsdam, Potsdam, Germany

J. Hinkel
Global Climate Forum, Berlin, Germany

C. Huntingford
Centre for Ecology and Hydrology, Wallingford, UK

J. A. Lowe
Met Office Hadley Centre, Exeter, UK

T. J. Osborn
Climatic Research Unit, School of Environmental Sciences, University of East Anglia, Norwich, UK

P. Smith
University of Aberdeen, Aberdeen, UK

P. Zelazowski
Environmental Change Institute, University of Oxford, Oxford, UK

socio-economic scenarios. Indicators of impact cover the water resources, river and coastal flooding, agriculture, natural environment and built environment sectors. Impacts are assessed under four SRES socio-economic and emissions scenarios, and the effects of uncertainty in the projected pattern of climate change are incorporated by constructing climate scenarios from 21 global climate models. There is considerable uncertainty in projected regional impacts across the climate model scenarios, and coherent assessments of impacts across sectors and regions therefore must be based on each model pattern separately; using ensemble means, for example, reduces variability between sectors and indicators. An example narrative assessment is presented in the paper. Under this narrative approximately 1 billion people would be exposed to increased water resources stress, around 450 million people exposed to increased river flooding, and 1.3 million extra people would be flooded in coastal floods each year. Crop productivity would fall in most regions, and residential energy demands would be reduced in most regions because reduced heating demands would offset higher cooling demands. Most of the global impacts on water stress and flooding would be in Asia, but the proportional impacts in the Middle East North Africa region would be larger. By 2050 there are emerging differences in impact between different emissions and socio-economic scenarios even though the changes in temperature and sea level are similar, and these differences are greater in 2080. However, for all the indicators, the range in projected impacts between different climate models is considerably greater than the range between emissions and socio-economic scenarios.

1 Introduction

The assessment reports of the Intergovernmental Panel on Climate Change (IPCC) review hundreds of studies into the potential impacts of climate change (e.g. IPCC 2007, 2014). Two key conclusions can be drawn from these assessments. First, the distribution of impacts across space and between regions is as relevant as the global aggregate impact when assessing the global-scale impacts of climate change; the distribution of impacts is highlighted in IPCC reports as one of the five integrative ‘reasons for concern’ about climate change alongside aggregate impacts. Second, impacts occur across many dimensions of the environment, economy and society and therefore need to be expressed in terms of multiple indicators. However, there have still so far been few consistent studies of the impact of climate change across sectors and the global domain. Most global studies have concentrated on one sector, and different studies have used different climate and socio-economic scenarios. The few multi-sectoral studies (Hayashi et al. 2010; van Vuuren et al. 2011; Piontek et al. 2014) have used few climate models and a small number of indicators. It has therefore been difficult to produce consistent assessments not only of the global-scale impacts of climate change, but also of the potential for multiple impacts across several sectors. Such assessments are of value not only to global-scale reviews of the potential consequences of climate change, but also to organisations concerned with the distribution of impacts across space. These include development, disaster management and security agencies, together with businesses or organisations with international coverage or supply chains.

This paper presents for the first time an assessment of the multi-dimensional impacts of climate change across the global domain for a wide range of sectors and indicators, using consistent climate and socio-economic scenarios and a harmonised methodology. Impacts are estimated under four different future world scenarios using up to 21 different climate model patterns to characterise the spatial pattern of climate change. The assessment constructs a set of coherent narratives of impact across regions and sectors, and also includes a representation of some of the major sources of uncertainty in potential regional impacts. It complements other global-scale assessments that used the same methodology and models to identify the relationship between amount of climate forcing and impact (Arnell et al. 2014) and the impacts avoided by climate mitigation policy (Arnell et al. 2013).

2 Methodology

2.1 Overview of the approach

The assessment involves the application of a suite of spatially-explicit impacts models run with scenarios describing a range of emissions and socio-economic futures. These emissions and socio-economic futures are here represented by the A1b, A2, B1 and B2 SRES storylines (IPCC Intergovernmental Panel on Climate Change 2000). Scenarios characterising the spatial and seasonal distribution of changes in climate and sea level around 2020, 2050 and 2080 are constructed from up to 21 global climate models (Meehl et al. 2007a) in order to assess the climate-driven uncertainty in the projected impacts for a given future. The period 1961-1990 is used as the climate baseline.

The impact sectors and indicators are summarised in Table 1 (see Supplementary Information for details of the impact models). They span a range of the biophysical and socio-economic impacts of climate change, but do not represent a fully comprehensive set covering all impact areas which may be of interest; they represent an 'ensemble of opportunity' based on the availability of models. All the land-based impact models use the same baseline climatology, and all the indicators relating to socio-economic conditions use the same socio-economic data. The impact assessment is therefore harmonised, but is not a fully integrated assessment because interactions between sectors are not represented. Only one impact model is used in each sector, so the uncertainty associated with impact model structure and form is not considered.

The socio-economic impacts of climate change in a given year are expressed relative to the situation in that year in the absence of climate change (i.e. assuming that the climate remains the same as over the baseline period 1961-1990). For the 'pure' biophysical indicators—crop productivity, suitability of land for cropping, terrestrial ecosystems and soil organic carbon—impacts are compared with the 1961-1990 baseline. Impacts are presented at the regional scale (Supplementary Table 1).

Most of the indicators represent change in some measurable impact of climate change, such as the average annual number of people flooded in coastal floods or crop productivity. Three of the indicators (water scarcity, river flooding and crop suitability), however, represent change in *exposure* to impact. The extent to which exposure translates into impact depends on the water management and agricultural practices in place, but these are so locally diverse and dependent on local context that it is currently not feasible to represent them numerically in global-scale impacts models. The indicators do not incorporate the effects of adaptation to climate change, with the exception of crop productivity where the crop variety planted varies with climate (see Supplementary Information).

Impacts can be expressed in either absolute or relative terms, and there are advantages and disadvantages in both when comparing impacts across regions. Large percentage impacts in a region may represent small absolute numbers and therefore make a small contribution to the global impact, but may indicate substantial impacts in the region itself. In contrast, a small percentage impact in another region may produce large absolute impact—and thus contribute substantially to the global total—but the implications for the region itself may be smaller. Most

Table 1 Summary of the impact indicators

Indicator	Description	Drivers of change	Further details
Water			
Population exposed to a change in water resources stress	A change in exposure to stress occurs where runoff in water-stressed watersheds changes significantly, or watersheds move into or out of the stressed class. Water-stressed watersheds have less than 1000 m^3/capita/year. Runoff is estimated using the MacPDM.09 hydrological model.	Change in runoff due to climate change Change in population	Gosling and Arnell (2013)
River flooding			
Flood-prone population exposed to a substantial change in frequency of flooding	A substantial change occurs when the frequency of the baseline 20-year flood doubles or halves. River flows are estimated using the MacPDM.09 hydrological model.	Change in runoff due to climate change Change in population	Arnell and Gosling (2014)
Coastal			
Change in coastal wetland extent	Calculated using DIVA v2.04	Change in relative sea level rise	Brown et al. (2013)
Additional average annual number of people flooded from extreme water levels	Calculated using DIVA v2.04. It is assumed that the level of coastal protection increases with population and wealth	Change in relative sea level rise Change in population Change in income	Brown et al. (2013)
Agriculture			
Cropland exposed to change in crop suitability	A substantial change occurs where a crop suitability index changes by more than 5 %	Change in climate	Index defined in Ramankutty et al. (2002)
Change in spring wheat, soybean and maize productivity	Productivity is simulated using GLAM. Adaptation is incorporated by selecting the variety (from three) with the greatest productivity and varying planting dates with climate	Change in climate Change in CO_2 concentration	Osborne et al. (2012)
Environment			
Proportion of (non-cropped) region with a substantial change in Net Primary Productivity (NPP)	Calculated using JULES/IMOGEN. A substantial change is greater than 10 %.	Change in climate Change in CO_2 concentration	Model summarised in Huntingford et al. (2010)
Change in total regional forest extent	Calculated using JULES/IMOGEN. Change in area under forest plant function types.	Change in climate Change in CO_2 concentration	Model summarised in Huntingford et al. (2010)
Change in soil organic carbon (SOC) in mineral soils	Calculated using RothC, and aggregated over all land cover types.	Change in climate Change in CO_2 concentration	Gottschalk et al. (2012)
Infrastructure			
Change in regional residential heating and cooling energy demands	Energy requirements are based on heating and cooling degree days, population size and assumptions about heating and cooling technologies	Change in climate Change in population Change in income Change in energy efficiency	Model based on Isaac and van Vuuren (2009)

of the impacts in this paper are expressed in absolute terms, but relative changes can be calculated from the data in the tables.

The distribution of impacts between regions and across sectors varies with different spatial patterns of change in climate, as represented by different climate models. One possible way of summarising the global and regional impacts of climate change would be to show the ensemble mean (or median) impact for a given sector and region across all climate model patterns, perhaps with some representation of uncertainty through identifying consistency between the different models (as is often done for climatic indicators such as temperature and precipitation). However, this is problematic when the concern is with multiple indicators of impact and comparisons between regions for two main reasons. The calculation of an ensemble mean makes assumptions about the relative plausibility of different climate models, but more importantly the ensemble mean impact does not necessarily represent a plausible future world. Calculating the average reduces the variability between regions and the relationships between sectors and indicators.

An alternative approach is therefore to treat each climate model as the basis for a separate narrative or story, describing a plausible future world with its associations between indicators and regions. Uncertainty in potential impacts is then characterised for each region and indicator by comparing the range in impacts across different climate models, but recognising that aggregated uncertainty—across regions or indicators—is not equivalent to the sum of the individual uncertainty ranges.

2.2 Climate and sea level rise scenarios

Climate scenarios were constructed (Osborn et al. 2014) by pattern-scaling output from 21 of the climate models in the CMIP3 set (Meehl et al. 2007a: Supplementary Table 2) to match the changes in global mean temperature projected under the four SRES emissions scenarios A1b, A2, B1 and B2. These global temperature changes were estimated using the MAGICC4.2 simple climate model with parameters appropriate to each climate model (Meehl et al. 2007b: Supplementary Fig. 1a). Pattern-scaling was used rather than simply constructing climate scenarios directly from climate model output partly to better separate out the effects of underlying climate change and internal climatic variability, and partly to allow scenarios to be constructed for all combinations of climate model and emissions scenario. Rescaled changes in mean monthly climate variables (and year to year variability in monthly precipitation) were applied to the CRU TS3.0 0.5×0.5^{o} 1961-1990 climatology (Harris et al. 2014) using the delta method to create perturbed 30-year time series representing conditions around 2020, 2050 and 2080 (Osborn et al. 2014). The terrestrial ecosystem and soil carbon impact models require transient climate scenarios, and these were produced by repeating the CRU 1961-1990 time series and rescaling to construct time series from 1991 to 2100 using gradually increasing global mean temperatures (Osborn et al. 2014). Pattern-scaling makes assumptions about the relationship between rate of forcing and the spatial pattern of change, which have been demonstrated to be broadly appropriate for the averaged climate indicators used here (e.g. Tebaldi and Arblaster 2014), but which do constitute caveats to the quantitative interpretation of results.

Sea level rise scenarios were constructed for 17 climate models. Spatial patterns of change in sea level due to thermal expansion were available for 11 of the models, and for the other six globally-uniform thermal expansion scenarios were calculated using MAGICC4.2. Uniform projections of the contributions of ice melt were added to these patterns, and the patterns were rescaled to correspond to specific global temperature changes using the same methods as applied in Meehl et al. (2007b). Ice melt contributions from Greenland and Antarctica, as well as ice caps and glaciers were calculated following the methodology of Meehl et al. (2007b), with additional data to calculate ice sheet dynamics from Gregory and Huybrechts (2006) (see Brown et al. 2013). Global average sea level rise scenarios are shown in Supplementary Fig. 1b; note that the highest change is produced by one model which is considerably higher—by around 20 cm in 2100—than the others. The effects of changes in the Greenland and Antarctic ice sheet dynamics are not incorporated, but the range in sea level rise across the models is large compared with the potential magnitude of the dynamic effect.

Table 2 Regional and global exposure to impact in 2050 in the absence of climate change

	Total population (millon)	Water-stressed population (million)	Flood-prone people (million)	Cropland (thousand km^2)	Average spring wheat yield (kg/ha)	Average soybean yield (kg/ha)	Average maize yield (kg/ha)	Total Soil Organic Carbon (SOC) (Pg C)	Regional average NPP (kg C m^{-2} $years^{-1}$)	Regional forest area (thousand km^2)	Coastal wetland (thousand km^2)	Average annual people flooded in coastal floods (thousand/year)	Heat energy demand (PJ)	Cool energy demand (PJ)
W. Africa	454	42	33	832	793	119	584	13	1.0	1,166	42	5	0	12
C. Africa	179	5	12	216	525	518	910	20	1.2	3,445	10	2	7	4
E. Africa	314	97	13	268	891	702	1,191	9	0.8	432	4	2	103	1
Sn Africa	230	11	14	422	979	1,484	1,460	23	0.8	2,651	33	101	128	22
S. Asia	2,085	1,466	357	2,238	757	699	805	17	0.8	363	38	94	2,588	420
SE Asia	724	0	103	980	732	579	1,341	18	1.4	2,579	114	241	24	2,331
E Asia	1,533	673	147	1,473	1,775	1,545	2,995	39	1.0	601	10	37	16,130	1,769
Central Asia	86	2	7	311	665	477	2,548	11	0.2	6	n/a	n/a	1,086	24
Australasia	47	0	2	308	1,642	1,284	2,914	19	0.8	639	156	19	212	59
N. Africa	266	206	21	363	1,596	400	467	8	0.8	144	6	5	575	24
W. Asia	350	236	9	362	1,136	n/a	2,654	4	0.3	29	5	5	1,427	411
W. Europe	422	160	32	773	3,102	2,018	4,427	19	0.8	872	23	7	6,706	196
C. Europe	118	6	12	504	2,432	885	2,651	9	0.9	110	4	0	1,803	8
E. Europe	202	5	23	1,688	1,145	1,001	2,519	123	0.5	3,163	7	27	4,023	16
Canada	40	7	1	402	922	2,000	2,826	59	0.4	2,884	69	22	1,295	8
US	404	76	10	1,770	995	1,171	2,454	39	0.8	1,931	156	17	5,167	593
Meso-America	251	53	10	485	1,239	1,167	1,417	9	1.0	663	69	2	360	379
Brazil	224	0	16	490	1,835	1,776	2,616	34	1.1	6,013	54	10	79	722
South America	266	18	21	561	1,386	1,527	2,501	39	1.1	3,693	58	11	1,007	373
Global (A1b)	8,196	3,064	843	14,447	1,493	1,346	2,204	513	0.8	31,383	857	606	42,716	7,375
Global (A2)	10,387	4,792	1,083									2,800	38,876	2,083
Global (B1)	8,196	3,064	843									910	40,719	4,696
Global (B2)	9,021	3,652	935									1,150	40,297	3,380
Global (2000)	6,122	1,555	637									3,100	30,447	857

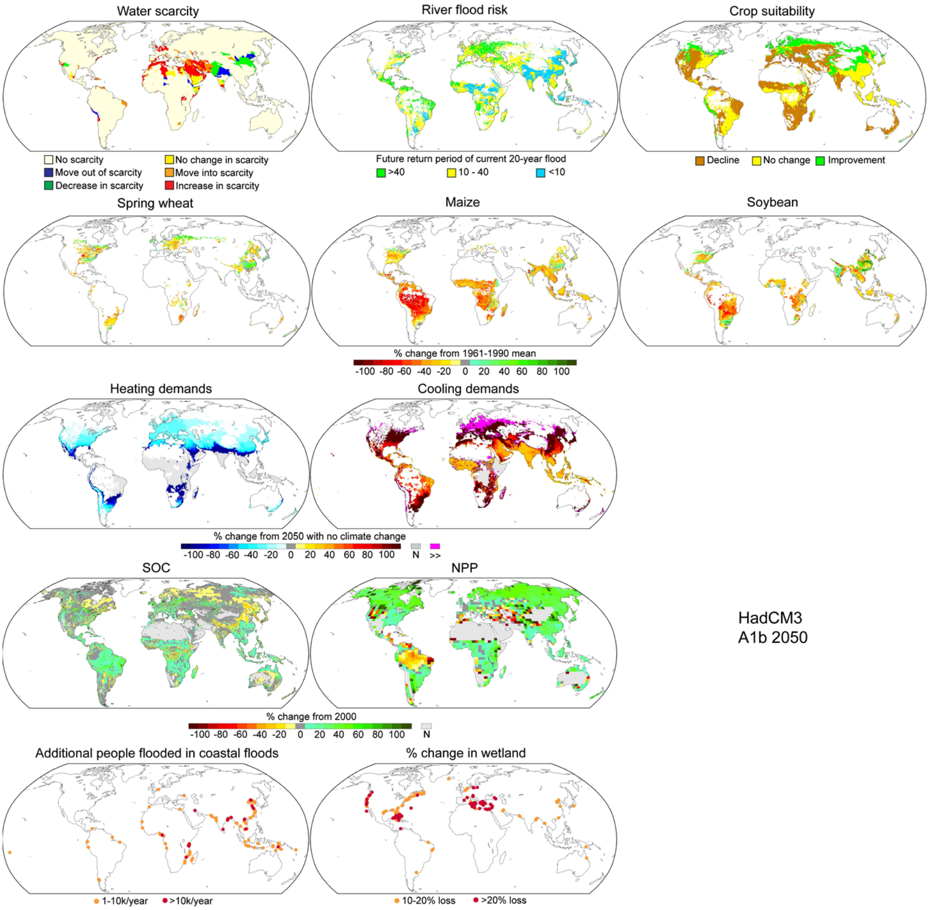


Fig. 1 The geographic distribution of impacts under the A1b 2050 scenario: one plausible model (HadCM3). For river flood risk, white areas indicate that the grid cell floodplain population is less than 1000 people. For crop productivity, white areas indicate that the crop is not currently grown. For heating and cooling demands, white areas indicate that grid cell population is less than 10,000, light grey indicates no heating / cooling demands in either the present or the future, and magenta indicates no demand in the present but some demand in the future. For SOC and NPP, light grey denotes zero values in 2000

2.3 Socio-economic scenarios

Future population and gross domestic product at a spatial resolution of 0.5×0.5^{o} were taken from the IMAGE v2.3 representation of the SRES storylines (van Vuuren et al. 2007). The population living in inland river floodplains was estimated by combining high resolution gridded population data for 2000 (Center for International Earth Science Information Network CIESIN 2004) with flood-prone areas defined in the UN PREVIEW Global Risk Data Platform to estimate the proportions of grid cell population currently living in flood-prone areas. Cropland extent was taken from Ramankutty et al. (2008). It is assumed that river floodplain extent, cropland extent and the proportion of grid cell population living in flood-plains do not change over time.

3 Exposure in the absence of climate change

The impacts of climate change in the future depend on the future state of the world. Table 2 shows the regional exposure to water resources scarcity, river and coastal flooding and residential energy demand in 2050 under the A1b socio-economic scenario, together with (modelled) average regional crop yields and ecosystem indicators, assuming climate and sea level remain at the 1961-1990 level. The table also shows global totals for some of the indicators under the other three socio-economic scenarios, alongside global totals for 2000.

The vast majority of people living in water-stressed watersheds, river floodplains and flooded in coastal floods are in south and east Asia (including India, Bangladesh and China). By 2050 east Asia (predominantly China), with Europe and North America, account for the vast bulk of heating energy requirements. However, the absolute numbers hide regional variations in the proportions of people living in exposed conditions; more than 75 % of North African people would be living in water-stressed watersheds in 2050 (a slightly higher proportion than in 2000), along with two-thirds of people in west Asia (up from 35 % in 2000).

4 The regional impacts of climate change in 2050 in an A1b world

4.1 Introduction

By 2050, global average temperature under A1b emissions would be between 1.4 and 2.9 °C above the 1961-1990 mean, with an average increase across climate models of around 1.9 °C. Global average sea level would be 12 to 32 cm higher than over the period 1961-1990, with an average increase of 18 cm (note that changes in temperature under A1b are between changes under RCP6.0 and RCP8.5: IPCC 2013). However, the spatial patterns of changes in temperature, precipitation, sea level and other relevant climatic variables vary between climate models, so the projected potential impacts also vary. This section first describes the potential impacts across the world and across sectors under one example plausible climate story, and then assesses the uncertainty in impacts by region and sector.

4.2 A coherent story: Impacts under one plausible climate future

Figure 1 and Table 3 show the impacts in 2050 under one illustrative climate model (HadCM3); this particular model has an increase in global mean temperature of 2.2 °C (relative to 1961-1990) in 2050 under A1b emissions, and a global mean sea level rise of 16 cm.

Under this plausible story, approximately 1 billion people are exposed to increased water resources stress due to climate change, relative to the situation in 2050 with no climate change, and almost 450 million people are exposed to a doubling of flood frequency. In contrast, around 1.9 billion water-stressed people see an increase in runoff, and around 75 million flood-prone people are exposed to flooding half as frequently as in the absence of climate change. Approximately 1.3 million additional people are flooded in coastal floods each year. Around a half of all cropland sees a decline in suitability, but about 15 % sees an improvement. Global residential heating energy demands are reduced by 30 % (bringing them back to approximately the 2000 level) but cooling demands rise by over 70 %. The net effect is a reduction in total heating and cooling energy demands of around 15 %. There are, however, considerable regional variations in impact.

Under this story, increases in water scarcity are most apparent in the Middle East, north Africa and western Europe, whilst increases in exposure to river flooding is largest in south

Table 3 Regional and global impacts in 2050, under the A1b emissions and socio-economic scenario, under one plausible climate story. The socio-economic impacts are relative to the situation in 2050 in the absence of climate change (Table 2)

	Pop. exposed to increased water resources stress (millons)	Pop. with decreased water resources stress (millions)	Pop. exposed to doubled flood freq. (millions)	Pop. exposed to halved flood freq. (millions)	Cropland with decline in suitability (thousand km^2)	Croplandwith increase in suitability (thousand km^2)	Change in average spring wheatyield (%)	Change in average soybean yield (%)	Change in average maize yield(%)	Change in Soil Organic Carbon (SOC) content (%)	Change in regional NPP (%)	Change in forest area (%)	Change in coastal wetland (%)	Additional people flooded in coastal floods (thousands/year)	Change in heat energy demand (%)	Change in cool energy demand (%)
W. Africa	26	16	18	3	552	0	40	−15	−32	2	22	6	−7	29	−33	35
C. Africa	5	0	5	1	68	4	−30	−44	−42	1	13	2	−13	4	−88	37
E. Africa	79	12	4	1	102	14	−17	−40	−14	3	34	6	−12	6	−85	93
Sn Africa	17	0	1	3	380	0	−25	−23	−33	8	19	4	−17	312	−70	139
S. Asia	188	1,209	290	5	1,077	436	28	23	−16	−3	22	7	−10	132	−43	38
SE Asia	0	0	32	2	78	0	31	−15	−30	8	20	3	−12	406	−87	31
E Asia	0	636	80	0	130	419	7	1	−16	−3	27	5	−22	222	−28	97
Central Asia	3	0	0	0	289	20	−3	99	−31	8	9	12	n/a	n/a	−24	127
Australasia	0	0	1	0	278	2	−25	−13	−31	2	21	6	−12	65	−39	95
N. Africa	117	0	3	6	292	16	−39	−39	−43	3	27	5	−21	14	−53	78
W. Asia	185	0	0	5	349	3	−17	n/a	−21	3	4	0	−22	39	−35	54
W. Europe	192	0	1	8	448	214	2	−12	−18	3	17	7	−17	17	−27	178
C. Europe	9	0	0	6	249	145	−14	38	−12	5	7	4	−20	2	−26	821
E. Europe	14	0	1	16	1,110	435	−15	33	−14	−1	31	11	−19	5	−23	519
Canada	7	0	0	0	45	286	−3	5	−13	−2	36	8	−6	5	−21	365
US	83	6	1	2	1,068	44	−14	−8	−20	5	21	5	−24	4	−28	134
Meso-America	56	1	0	7	354	0	−9	−25	−30	5	−3	2	−18	9	−71	74
Brazil	28	0	6	6	166	0	−25	−30	−32	14	−8	−1	−9	11	−93	69
South America	20	13	5	7	182	45	−6	−13	−25	5	8	3	−19	28	−44	78
Global	1,025	1,893	447	76	7,215	2,083	1	−13	−22	2	16	4	−15	1,309	−30	73

and east Asia. The suitability of land for cropping declines in most regions, but increases at the northern boundary of cropland and along some margins in east Asia. Spring wheat yields show a mixed pattern of change, maize yields decline everywhere except in parts of north America and eastern China, and soybean yields tend to increase in parts of south and east Asia, north America and small parts of south America, but decrease elsewhere. Increases in coastal flood risk are concentrated in Asia and eastern Africa, whilst wetland losses focus around the Mediterranean and north America. Cooling energy demands increase particularly in regions where there is currently little demand for cooling, but increase only slightly in some warm regions—because the relative change in requirements is smaller. Heating energy demands decrease most in the warmest regions.

Many regions are exposed to multiple overlaying impacts. For example, under this plausible climate story river flood risk increases across much eastern Asia, coastal flood risk increases substantially, and cooling energy demands increase by more than 70 %. At the same time, the productivity of the three example crops increases in parts of eastern Asia, but decreases across much of northern China. The suitability for agriculture appears to increase in northern and western China, although soil organic carbon contents decline (in this case because conversion of forest to arable land reduces the inputs of carbon from vegetation).

In southern Asia, crop suitability declines, productivity of maize declines but soybean productivity increases (in some parts). River flood risk increases and some coastal megacities see increased flood risk. Cooling energy demands rise by around 30–40 %, but there is little change in heating demands. Water scarcity reduces under this story across many water-scarce parts of southern Asia.

The suitability of cropland for crop cultivation declines across much of sub-Saharan Africa, primarily due to reductions in available moisture; more than 90 % of cropland in southern Africa would see a reduction in suitability for crop production. Maize yields reduce by 20–40 %. River flood risk increases substantially in parts of western Africa, and coastal flood risk increases in particular for many east African coastal cities. Across the Middle East and North Africa crop suitability declines and large populations are exposed to increased water scarcity and increased cooling energy demands; NPP also reduces in many parts of the region.

Within western and central Europe, river flood risk is little affected under this story, but around 200 million people are exposed to increased water resources stress. Crop suitability increases in the north of the region but declines elsewhere, and spring wheat productivity declines across much of central and eastern Europe. Cooling energy demands are increased very significantly—from close to zero in northern Europe—but heating energy demands fall by at least 40 %.

Under this story, the main potential impacts in North America appear to be reductions in crop suitability across much of western and central North America, but increases at the northern margins of agriculture, and mixtures of increases and decreases in crop yields. Cooling energy demands increase very significantly in the eastern parts of North America, where heating energy demands fall. Coastal wetland loss is particularly large along the west coast.

Across South America, maize and soybean yields fall and NPP decreases substantially across the Amazon basin; the suitability for cropping declines in the drier parts of eastern south America, but increases along parts of the west coast.

The impacts plotted in Fig. 1 and tabulated in Table 3 would arise under one particular plausible climate future. In principle it is possible to produce similar stories under other climate models. Table 4 shows the global aggregated impacts for each indicator under another six climate models (and they should be compared with the global row in Table 3). Supplementary Figs. 2-7 show the distribution of impacts under six more climate model patterns, and Supplementary Table 3 presents regional impacts under all 21 climate model patterns used.

4.3 Uncertainty in projected regional impacts

The uncertainties in regional impacts, by sector, are given in Table 5, which shows the range in estimated impacts across the climate models used (which range from 21 for most indicators to 7 for SOC). Fig. 2 summarises the regional uncertainty in impacts.

For most impact sectors, the projected ranges are very large. In some cases—specifically the water and river flooding sectors—this is because of very large uncertainty in projected changes in regional rainfall (in south and east Asia, for example). In some other cases, the large uncertainty is because the sector in a region is particularly sensitive to change (for example where the baseline values in the absence of climate change are small—see forest and NPP change in west and central Asia). In other cases, the uncertainty range is dominated by individual anomalous regional changes. For example, the large range in estimated additional people exposed to coastal flooding is due to one particular climate model producing very considerably higher sea level rises in some regions than the others. There is least uncertainty in projected reductions in heating energy requirements and, for most regions, increases in cooling energy requirements; the greatest uncertainty here is in those regions where requirements are currently low—Europe and Canada—but the percentage changes are sensitive to small changes in temperature.

The considerable uncertainty in each region and sector needs to be interpreted carefully. It is not correct simply to add up the extremes of each range across regions and use this to characterise the global range; the global range will be smaller than the sum of the extremes because no one climate model produces the most extreme response in every region. Similarly, it is not appropriate to define the maximum impact across all sectors in a region as the sum of the maximum impacts for each sector, because again no one single climate model produces the maximum impact in all sectors. Indeed, there are some associations between impacts in different sectors between climate models. For example, models which produce the greatest increase in exposure to water resources stress tend to be those which produce the smallest increase in exposure to river flooding, and the greatest area of cropland with a decline in suitability (see Supplementary Fig. 8 for an example).

5 Impacts under different worlds and over time

Figure 3 shows how global impacts vary in 2020, 2050 and 2080 between the four SRES scenarios, across all climate models. There is little difference in impact between either the emissions or socio-economic scenarios in 2020, when temperature differences between the emissions scenarios (Supplementary Figure 1) are very small. By 2050 the differences in temperature between the A1b, A2 and B2 emissions scenarios remain small, but B1 produces a lower increase in temperature so in many sectors impacts are smaller with this scenario. B2 has a lower CO_2 concentration than A1b or A2, so produces a smaller increase in NPP and forest area despite the temperature changes being similar. Socio-economic impacts under A2 are higher than under the other scenarios despite little difference in temperature, and this is because of increased exposure under the A2 world. More people live in water-stressed or flood-prone areas and, in the coastal zone, there is less investment in coastal protection. By 2080 the difference between the emissions and socio-economic scenarios becomes greater. The greatest impacts are under A2, primarily because exposure is greater, and the lowest impacts tend to be under B1 with the lowest increase in temperature. However, for all indicators, the range between climate model patterns is considerably greater than the range between the emissions or socio-economic scenarios.

Table 4 Global impacts in 2050, under the A1b emissions and socio-economic scenario, under six plausible stories. Compare with the global impacts in Table 3

	Pop. exposed to increased water resources stress (millions)	Pop. with decreased water resources stress (millions)	Pop. exposed to doubled flood frequency (millions)	Pop. exposed to halved flood frequency (millions)	Cropland with decline in suitability (thousand km^2)	Cropland with increase in suitability (thousand km^2)	Change in average spring wheat yield (%)	Change in average soybean yield (%)	Change in average maize yield(%)	Change in Soil Organic Carbon (SOC) content (%)	Change in regional NPP (%)	Change in forest area (%)	Change in coastal wetland (%)	Additional people flooded in coastal floods (thousands/year)	Change in heat energy demand (%)	Change in cool energy demand (%)
HadGEM1	1,385	1,385	202	94	7,203	2,150	9	0	−12	4	23	5	−14	1,023	−23	46
ECHAM5	1,369	508	302	55	7,631	1,672	8	−4	−17	3	23	5	−17	1,568	−31	65
CGCM3.1 (T47)	746	1,844	316	64	3,963	4,062	17	−1	−12	3	25	5	−15	1,215	−25	53
CCSM3	639	1,680	321	46	4,362	2,619	8	10	−6	5	27	5	−14	1,003	−20	37
IPSL-CM4	2,221	418	95	264	8,882	1,563	6	−4	−15	3	22	5	−16	1,503	−30	62
CSIRO-Mk3.0	1,820	213	41	130	6,722	2,012	13	5	−10	4	23	5	−14	892	−19	38

Table 5 The range in regional and global impacts in 2050, under the A1b emissions and socio-economic scenario, across all climate models

	Pop. exposed to increased water resources stress (millions)	Pop. with decreased water resources stress (millions)	Pop. exposed to doubled flood freq. (millions)	Pop. exposed to halved flood freq. (millions)	Cropland with decline in suitability (thousand km^2)	Croplandwith increase in suitability (thousand km^2)	Change in average spring wheatyield (%)	Change in average soybean yield (%)	Change in average maize yield(%)	Change in Soil Organic Carbon (SOC) content (%)	Change in regional NPP (%)	Change in forest area (%)	Change in coastal wetland (%)	Additional people flooded in coastal floods (thousand/year)	Change in heat energy demand (%)	Change in cool energy demand (%)
W. Africa	0 to 161	0 to 34	0 to 21	0 to 27	191 to 656	0 to 287	4 to 51	−28 to−2	−45 to−15	−3 to 8	10 to 22	3 to 10	−12 to−7	26 to 233	−67 to−33	16 to 40
C. Africa	0 to 6	0 to 5	0 to 10	0 to 6	4 to 76	0 to 32	−33 to−6	−48 to−14	−46 to−17	1 to 5	13 to 30	2 to 4	−17 to−12	3 to 32	−90 to−63	17 to 46
E. Africa	0 to 167	0 to 97	0 to 10	0 to 6	0 to 139	1 to 124	−36 to−5	−49 to−18	−22 to 3	3 to 7	23 to 59	4 to 13	−16 to−9	2 to 29	−92 to−60	43 to 107
Sn Africa	0 to 17	0 to 11	0 to 8	0 to 4	229 to 380	0 to 17	−46 to−16	−26 to−8	−36 to−11	8 to 11	19 to 50	4 to 10	−20 to−17	142 to 1,909	−71 to−43	57 to 148
S. Asia	52 to 1,460	0 to 1,397	9 to 290	1 to 161	215 to 2,049	17 to 1,621	12 to 41	−17 to 23	−28 to−1	−6 to−1	5 to 30	4 to 9	−13 to−9	93 to 716	−52 to−29	20 to 45
SE Asia	0	0	1 to 71	0 to 40	0 to 159	0 to 2	18 to 33	−32 to−9	−42 to−15	7 to 9	15 to 23	2 to 3	−15 to−11	349 to 643	−87 to−62	15 to 35
E Asia	0 to 506	0 to 648	1 to 113	0 to 21	15 to 662	108 to 491	−2 to 16	0 to 15	−19 to−4	−3 to−2	17 to 32	4 to 7	−24 to−18	47 to 441	−31 to−15	33 to 97
Central Asia	0 to 40	0 to 1	0 to 1	0 to 3	72 to 294	16 to 191	−3 to 50	−100 to 168	−38 to−6	7 to 12	9 to 49	12 to 23	n/a	n/a	−31 to−12	59 to 127
Australasia	0	0	0 to 1	0 to 1	131 to 278	2 to 111	−26 to 5	−13 to 39	−31 to−6	2 to 6	17 to 36	6 to 9	−15 to−10	31 to 110	−46 to−26	45 to 109
N. Africa	109 to 226	0 to 44	0 to 7	5 to 21	185 to 350	0 to 128	−60 to 26	−44 to−10	−44 to−21	1 to 5	9 to 42	5 to 12	−28 to 4	1 to 28	−61 to−33	40 to 87
W. Asia	135 to 308	0 to 134	0 to 1	1. to 9	296 to 357	1 to 22	−17 to 16	n/a	−21 to 11	3 to 14	−16 to 30	−13 to 10	−25 to−16	2 to 47	−39 to−22	30 to 59
W. Europe	24 to 211	0 to 143	0 to 12	1 to 15	249 to 448	214 to 262	2 to 33	−12 to 34	−18 to−1	2 to 4	15 to 39	4 to 8	−23 to−14	6 to 42	−32 to−16	57 to 178
C. Europe	2 to 32	0 to 6	0 to 1	1 to 9	90 to 249	143 to 236	−14 to 17	−25 to 87	−12 to 8	4 to 8	0 to 38	3 to 7	−35 to−6	0 to 6	−30 to−15	189 to 821
E. Europe	0 to 25	0 to 4	0 to 4	3 to 17	358 to 1,144	388 to 893	−15 to 22	10 to 87	−14 to 7	−1 to 1	17 to 36	7 to 11	−23 to−12	1 to 58	−27 to−11	116 to 519
Canada	0 to 7	0 to 7	0	0 to 1	0 to 177	183 to 385	−3 to 35	−40 to 45	−13 to 12	−2 to 0	14 to 41	5 to 9	−33 to−3	3 to 23	−29 to−9	71 to 365
US	21 to 116	0 to 49	0 to 3	0 to 6	300 to 1,545	25 to 270	−14 to 7	−10 to 28	−20 to−1	4 to 6	9 to 37	3 to 5	−28 to−20	3 to 17	−35 to−13	47 to 137
Meso-America	0 to 112	0 to 53	0 to 6	0 to 10	188 to 415	0 to 2	−30 to 4	−34 to 7	−43 to−3	2 to 11	−9 to 30	−1 to 6	−25 to−18	8 to 37	−71 to−34	28 to 77
Brazil	0 to 28	0	0 to 11	0 to 10	21 to 193	0 to 56	−25 to 13	−30 to 2	−32 to−7	14 to 21	−8 to 33	−1 to 4	−13 to−8	11 to 48	−93 to−52	27 to 69
South America	0 to 31	0 to 18	0 to 12	0 to 12	52 to 405	20 to 185	−13 to 15	−21 to 0	−31 to−8	5 to 9	8 to 30	3 to 6	−22 to−18	23 to 142	−48 to−23	31 to 79
Global	533 to 3,098	172 to 2,196	31 to 449	41 to 264	3,783 to 8,882	1,378 to 4,062	1 to 22	−13 to 10	−22 to−6	2 to 5	16 to 27	4 to 5	−18 to−13	763 to 4,101	−34 to−17	29 to 73

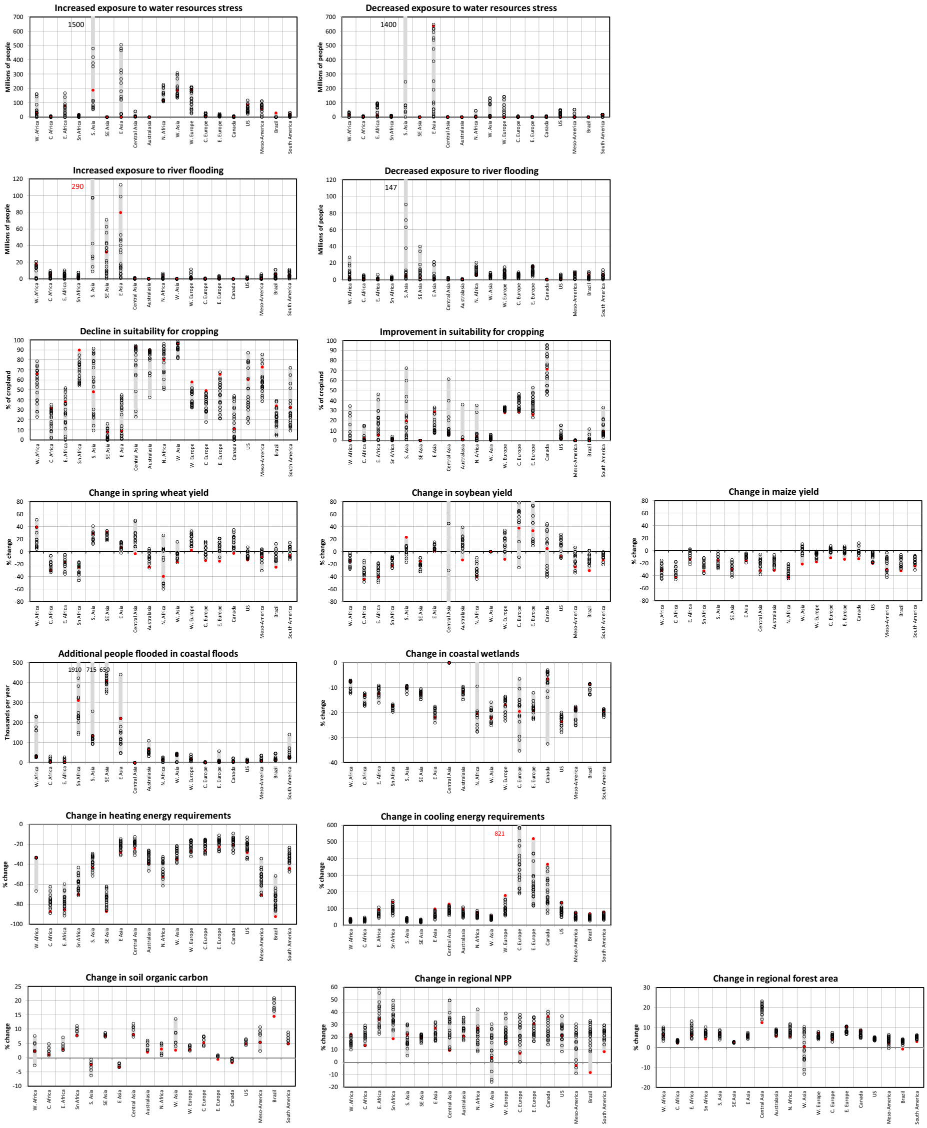


Fig. 2 Uncertainty in regional impacts in 2050, under A1b emissions and socio-economic scenarios. Impacts under individual climate models are shown as open circles; the red circle shows impacts under one specific model (HadCM3)

6 Conclusions

This paper has presented a high-level assessment of the global and regional impacts of climate change across a range of sectors. The assessment used a harmonised set of assumptions and data sets, four scenarios of future socio-economic development and emissions, and climate scenarios constructed from 21 climate models. The distribution of impacts between regions and the relationship between different impact indicators are important, so the assessment first

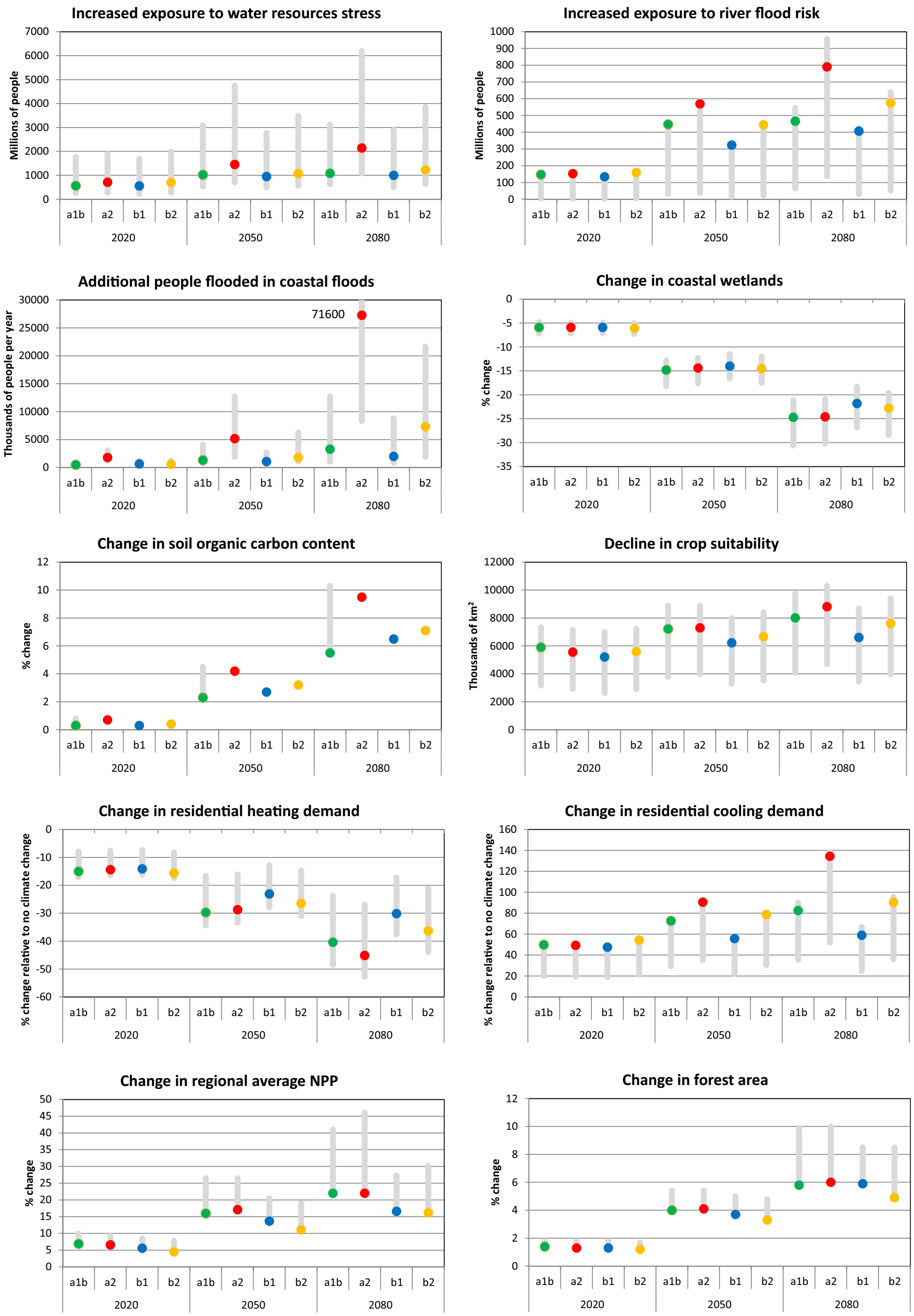


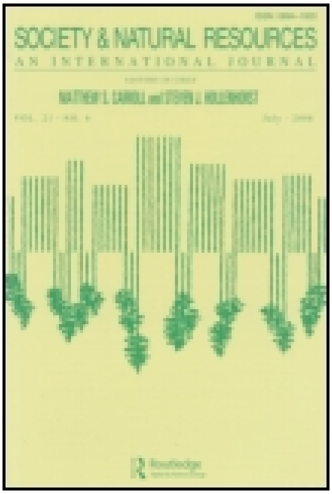
Fig. 3 Global-scale impacts of climate change in 2020, 2050 and 2080 under A1b, A2, B1 and B2 emissions and socio-economic scenarios. The grey bars represent the range across the climate models, the impacts under one specific model (HadCM3) are shown by the solid circle

describes impacts under a set of discrete ‘stories’ based on different climate models, and then considers uncertainty in regional impacts separately. The paper has therefore demonstrated a method for assessing multi-dimensional, regionally-variable impacts of climate change for a global assessment.

With A1b emissions and socio-economics, one plausible climate future (based on one climate model pattern) would result in 2050 in 1 billion people being exposed to increased water resources stress, around 450 million people exposed to increased frequency of river flooding, and an additional 1.3 million people flooded each year in coastal floods. Approximately half of all cropland would see a reduction in suitability for cropping, and the productivity of three major crops—spring wheat, soybean and maize—would be reduced in most regions. Global residential cooling energy requirements would increase by over 70 % globally, but heating energy requirements would decrease so total global heating and cooling energy requirements would reduce globally. The productivity of terrestrial ecosystems would be increased, and soil organic carbon contents would generally increase, leading to improved soil productivity and increased carbon storage. However, there is strong regional variability. Under this one climate model pattern, most of the global impacts on water stress and flooding would be in south, southeast and east Asia, but spring wheat productivity increases across much of Asia. In proportional terms, impacts on water stress and crop productivity are very large in the Middle East and North Africa region, which is exposed to multiple impacts.

There is considerable uncertainty in the projected regional impacts under a given emissions and socio-economic scenario, largely due to differences in the spatial pattern of climate change simulated by different climate models; this uncertainty varies between regions and sectors. Large increases in exposure to water resources stress, for example, are associated with large reductions in crop suitability but small increases in exposure to river flooding. The full richness of relationships between impacts in different places, and in different sectors, can therefore only be understood by comparing narrative stories constructed separately from different climate model scenarios.

There are, of course, a number of caveats with the approach. The climate scenarios used here are based on SRES emissions assumptions, and not on more recent RCP forcings or the climate models reviewed in the most recent IPCC assessment (IPCC 2013). However, the spatial patterns of change in climate under the latest generation of climate model simulations are broadly similar to those used here (Knutti and Selacek 2013). The climate scenarios are constructed by pattern scaling, and whilst this allows a direct comparison between different emissions scenarios and time periods, it does assume a particular relationship between the amount of global temperature change and the spatial pattern of change in climate. The indicators used represent an ‘ensemble of opportunity’, and do not necessarily span the full range of impacts of interest; there are also alternative indicators for many of the sectors considered here. The indicators do not (with the notable exception of crop productivity) explicitly incorporate the effects of adaptation in reducing the consequences of climate change. Comparisons with other single-sector global-scale impact assessments are made difficult by the use of different impact indicators (e.g. in the water sector) and different climate model scenarios. Insofar as it is possible to make comparisons, impacts as estimated in these other assessments are within the ranges presented here, but nevertheless the impacts presented here are best interpreted as indicative only. Finally, the indicators are calculated using only one impact model per sector. It is increasingly recognised that impact model uncertainty may make a substantial contribution to total impact uncertainty in some regions (e.g. Hagemann et al. 2013), and several initiatives are currently under way (for example ISI-MIP: Warszawski et al. 2014) to systematically evaluate the effects of impact model uncertainty.



Society & Natural Resources: An International Journal
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/usnr20

Global Inequality and Climate Change

J. Timmons Roberts [a]
[a] Department of Sociology/Program in Latin American Studies Tulane University New Orleans, Louisiana, USA

Published online: 15 Dec 2010.

Society and Natural Resources, 14:501–509, 2001

0894-1920/2001 $12.00 + .00

Global Inequality and Climate Change

J. TIMMONS ROBERTS

Department of Sociology/Program
in Latin American Studies
Tulane University
New Orleans, Louisiana, USA

Global warming is all about inequality, both in who will suffer most its effects and in who created the problem in the first place. This article describes the inequality empirically in broad strokes and then describes how it has led to the current deadlock in dealing with the problem of global climate change. Regarding bargaining positions in the Kyoto round of negotiations, two factions among rich nations and at least five distinct bargaining positions among poor nations are described and explained. The factional divisions are attributable to the differential influence of "polluting elites" across nations. The article concludes that the only way out of the conundrum of inequity and warming is by both addressing inequality and delinking carbon and development.

Keywords climate change, environment, greenhouse, inequality, stratification, world-systems theory

Global warming is all about inequality, both in who will suffer most its effects, and in who created the problem in the first place. Certainly global warming threatens everyone on the planet, but some places and some people in those places will suffer much sooner and much more than others. For example, many poor nations, especially island nations and those with large populations in low-lying areas, are facing ecological disasters "of biblical proportions" if the sea level rises as much as is predicted. The 1995 Intergovernmental Panel on Climate Change (IPCC) report, bringing together over 2000 scientists from around the world, predicted that Africa will face devastating droughts, which will destabilize governments and bring strife and suffering to the region; there will be flooding in Bangladesh, which will kill millions. There has been no serious refutation of these predictions, and a series of other impacts are already being felt, such as the wrath of hurricanes, droughts, heat waves, and floods (IPCC 1995; Gelbspan 1997; UNEP/WMO 1999). These poor nations are least able to handle the massive dislocations that come with "natural" disasters, which can set their development back decades. Within the poor nations, poor classes often never fully recover from devastating disasters brought on by the increasing climate instability.

Received 17 February 2000; accepted 13 September 2000.

The author thanks Tom Rudel, Gene Rosa, Paul Baer, and three anonymous reviewers for their helpful comments on an earlier draft of this article. The content of the article, however, is the sole responsibility of the author.

Address correspondence to J. Timmons Roberts, Department of Sociology/Program in Latin American Studies, Tulane University, New Orleans, LA 70118, USA.
E-mail: timmons@mailhost.tcs.tulane.edu

While effects of and the ability to handle climate change are unequally distributed, responsibility for the problem is even more unequally distributed. Poor nations remain far behind us in terms of emissions per person. For example the average U.S. citizen dumps as much greenhouse gas into the atmosphere as 8 Chinese and as much as 20 citizens of India (WRI 1998). Overall, the richest 20% of the world's population is responsible for over 60% of its current emissions of greenhouse gases. That figure surpasses 80% if our past contributions to the problem are considered. They probably should be considered, since carbon dioxide, the main contributor to the greenhouse effect, remains in the atmosphere for 120 years. This is simple, true injustice: The innocent are suffering the effects of something (our consumption) from which they drew little or no benefit. As members of the small island states whose cultures are likely to be decimated have pointed out, to understand the links and yet willfully allow the destruction of cultures and people seems plainly immoral.

Some will point out that there is another very different side to this story. China and India, with massive populations and rapid industrialization, will in the next twenty years surpass the wealthy countries in total emissions of greenhouse gases to the atmosphere (GCC 1998). It is true that environmentally speaking, one cannot handle this problem of global warming without addressing the boom of emissions in the developing countries. This point has been seized upon by the United States–headquartered oil and coal industries, who have mobilized think tanks, journalists, scientists, and senators to block any progress on the Kyoto treaty until the poor nations also agree to limits on their carbon emissions (McCright and Dunlap 1999). However, to ask these nations to stop development at a level we would never consider returning to seems hypocritical (Shue 1992). What's more, by their importance in the problem and their sheer numbers in negotiating efforts, the poor nations hold a veto power over efforts to enact a global climate treaty.

In this article I first describe the inequality empirically in broad strokes and then describe how it has led to the current deadlock in dealing with the problem of global climate change. Regarding bargaining positions in the Kyoto round of negotiations, I describe two factions among rich nations and at least five distinct bargaining positions among poor nations. My approach is utilizing some insights from the school of political economic thought labeled world-systems theory (see Shannon 1996 for a review) to attempt to understand the broad patterns of inequality in carbon emissions and differences between nations in their positions on a climate treaty. World-systems theory points us to pay close attention to the "polluting elites" who direct leading sectors of their economies (especially exports) and exercise disproportionate control over the national and foreign policy of nations on the environment (Roberts 1996a; Roberts 1996b; Roberts and Grimes 1999). The perspective, I argue, provides fresh insights into policy directions to address climate change. By way of conclusion, I argue here that the only way out of the conundrum of inequity and warming is by both addressing inequality and delinking carbon and development. Equity and ecology must be dealt with together: in this case, address inequality and decarbonize.

Inequality of Economy and Emissions

Of the world's 6 billion people, almost 5 billion live in countries where the average income is less than $3 a day (World Bank 1997). At the same time, on average, the people in the high-income countries get to live on 23 times that much, and the gap between the two groups is widening (World Bank 1995). About one out of four people in the world lives in absolute poverty, defined as "too poor to afford an adequate diet and other necessities." Malnutrition is said to stunt the growth and development of 40% of all 2-year-olds in poorer countries (UNRISD 1997; CARE 1996).

There are two pollutions, of course, that of wealth and that of poverty (Redclift and Sage 1998). In carbon terms, the worst is unquestionably the pollution of wealth. In the poorer countries, the greater environmental problems are simply survival: having enough to eat, a safe place to sleep, a way to take care of children. So there is great inequality of wealth and greenhouse gas emissions between nations and also within them. We lack much data on intracountry variation in carbon emissions (especially in the poor nations), but it can be said with confidence that the world's richest people cause emissions thousands of times that of the world's poorest.[1]

Still, the "pollution of poverty" must be considered. Many of the world's poor continue to gather firewood or animal waste for fuel. Both create important environmental damage and can contribute to land-use change, biodiversity loss, and climate change. But both are using essentially "renewable" energy sources, and the main threats to our atmosphere come from adding new carbon to the biosphere by burning fossil fuels (Kasting 1998). Traveling by foot, bicycle, and bus, and cooking local foods in unheated, un-air-conditioned tiny homes built of simple, local materials, the fossil fuel use by the world's poor on a per capita basis is almost negligible. Twenty percent of the world's population is responsible for 63% of the emissions, while the bottom 20% of the world's people are only releasing 3% (Population Action International 1998). Per capita, the analysis is striking: The high-income countries (using World Bank categories) are leveling off around 2.5 metric tons of annual carbon emissions per person, while the middle income nations are around 0.6 mT and the poorest around 0.02 mT (Roberts and Grimes 1997; Dietz and Rosa 1997).

It is a basic rule of civil justice, Superfund, and kindergarten ethics that those who create a mess should be responsible for cleaning up their share of the mess. Since carbon dioxide burned now stays in the atmosphere for over 100 years, shouldn't accounting for all the damage the rich nations have done in the past be considered (Neumeyer 2000)? This is a highly contentious issue indeed, but one that we have to consider if we are to address inequality and climate change. When emissions since 1950 are summed,[2] not surprisingly, the gap between rich and poor nations is much higher and is not narrowing or going away anytime soon. The summed emissions from the high-income nations amount to 900 trillion tons of carbon, from the 28% of the world who live in middle income nations only 500 trillion tons, and the poorest majority of the world have dumped only 200 trillion tons.

Roots of a Divisive Debate

A major sticking point in the Kyoto round is precisely this: how to calculate who is polluting how much, and how much they should be required to reduce their emissions. One alternative that has been proposed is to look at total carbon emissions per country and seek a reasonable international level per capita. Another is grandfathering emissions—that is, in the name of political expediency, national reductions are based on a baseline of 1990 levels. There are many other ways to calculate national targets, but this "grandfathering" approach is the one considered most feasible for negotiations (see, e.g., Torvanger and Godal 1999). This is the strategy that was undertaken at Kyoto, with different voluntary targets for 2010 or 2012 agreed to by nations, varying from 8% for European Union nations to smaller reductions and even limited increases for some nations. While there is agreement that rich nations must take the lead, the world does not in this case break down neatly along the lines of core, semiperiphery, and periphery (rich, middle, and poor). We need to look further into the energy intensity of nations, and the nature of their "polluting elites," to understand their bargaining positions. Space is limited, so this will be only a suggestive first crack at this.

Divisions in the Core

There are some surprising but well-known divisions within the wealthy core nations in how much they think they need to take the lead on reductions of carbon emissions. Leading the way on addressing the climate change problem aggressively have been the European Union (EU) nations, some of whom wanted 20% reductions and binding limits on carbon emissions at Kyoto. Densely populated and well aware of the environmental limits to economic growth, Europe's environmental movement is the strongest in the world, with viable political parties in many nations and even a few which have gained real political power.

After repeated foot-dragging by the United States and other nations, much lower, self-chosen emissions limits between 5 and 8% below 1990 levels were settled upon at Kyoto in 1997. Still, several of these nations are showing that they are willing to take unilateral steps to reduce their carbon emissions. This EU coalition appears to be facing a crucial test now, as recent data on how nations are doing meeting the Kyoto targets have pitted Spain against northern European nations (who are most enthusiastic about carbon emissions limitations). Britain is on its way toward reducing emissions by 20%, with the hopes of selling "permits" to other nations which are having more trouble reaching their goals.

On the other side are the foot-draggers, led by Japan, the United States, Switzerland, Canada, Australia, Norway, and New Zealand (called the JUSCANNZ nations). Iceland, Russia, and several others take a similar position but have not been asked for deep binding limits. After succeeding in getting less ambitious targets for emissions reductions at Kyoto, the group moved on to its next negotiating goals. At the top of the list was their desire to "buy" an unlimited amount of "permits," that is, purchase whatever reductions they fail to meet by the deadlines in the treaty (around 2012; Sissell 2000). Such purchases are called "flexibility mechanisms," and they can be in the form of buying reductions that were made in other rich nations (called "Joint Implementation") or in poorer nations (called the "Clean Development Mechanism"). Some projections on how much money could be flowing between nations if trading takes off (and it already has begun) suggest that trading could soon overshadow foreign aid and military aid. A fractious issue is that the EU has pushed for limits on trading permits, based on the concern that with unlimited trading there will be no real change in "business-as-usual" in the JUSCANNZ nations while the Europeans make real sacrifices and risk losing their economic competitiveness. The stage for the standoff is being set, since rather than going to 7% below 1990 levels, the United States is heading in the other direction, with 1997 figures showing an 11.1% increase in carbon dioxide emissions (U.S. EPA 1999).

A critical moment in the struggle over inequality and climate change came in the summer of 1997 before the Kyoto meeting in December. In the U.S. Senate, the Byrd–Hagel resolution passed 95–0 on 25 July 1997, making it impossible for the administration to sign any treaty that did not include the poor countries. Senator Robert C. Byrd declared on the Senate floor that "I do not think the Senate should support a treaty that requires only ... developed countries to endure the economic costs of reducing emissions, while developing countries are free to pollute the atmosphere, and, in so doing, siphon off American industries" (Dewar 1997). Representative

Henry Bonilla stated that Kyoto "is anti-American because it imposes strict, costly penalties on Americans while allowing many Third World countries to pollute our environment at will" (Congressional Record 23 July 1998). At the other end of the limited political spectrum on Capitol Hill, Representative John Dingell said in 1998 that "India, a mass emitter of CO-2, is not going to be bound. Our friends in China have told me that they will never be bound. So that leaves Uncle Sam, the United States, which proposes to be bound by a treaty which is going to cause enormous economic hardship" (Congressional Record 23 July 1998). To understand the position of the representatives, one need look no further than the hard-lobbied positions of the industries that support them (see, e.g., the American Chemistry Council web site). Byrd is a Democrat from the coal-dependent extractive periphery, West Virginia; Bonilla is a Republican from oil-dominated Texas; Dingell represents the auto-industry-dependent state of Michigan. World-system theory suggests that these positions are bald efforts by privileged economic groups at the top of the world inequality system defending their position there, while making outlandishly ironic claims about and impossible demands on those at the bottom of the pyramid. The "polluting elites" are flexing their political muscle. However, there are even more factions outside the core of the world system, tied largely to the nature of their own economic and political elites.

Divisions Outside the Core

Redclift and Sage (1998) argued that climate change still is not a top-priority issue for much of the world's population, because there are more pressing problems. This is true for many nations, but there is not one Third World. Though the poorer nations have often negotiated together, as the "G-77 and China" (which together actually represent 134 nations), there are at least five distinct positions of these nations on the issue of climate change treaties (Dunn 1998).

First, when it comes to obstructing binding limits on emissions of greenhouse gases, not surprisingly the hardest line has been taken by Organization of Petroleum Exporting Countries (OPEC). It has adopted a position very similar to that of the U.S. oil and coal lobbying arm, the "Global Climate Coalition," and other related think tanks, saying "go slow," "do more research," "don't hurt the economy," etc. (McCright and Dunlap 1999). At the COP-4 round of negotiations on the Framework Convention on Climate Change (FCCC) in Buenos Aires in late 1998, OPEC Secretary General Rilwanu Lukman gave a defining speech and released a press statement. It said that "as the group of developing countries that is most dependent upon fossil fuel sales, [OPEC] is concerned that the legitimate right to economic development, shared by all developing countries, is under threat from the mitigation measures that may come under discussion" (OPEC 1999). More concretely, "OPEC's research, for example, suggests that OPEC Member Countries could collectively suffer losses in revenue flows of the order of US$20 billion each year as a result of the proposed mitigation measures being implemented." He continued: "Compensation is a necessary concomitant of a balanced agreement." Finally, and bluntly, "how can fossil fuel producers be expected to give their wholehearted blessing to measures that could wreak havoc with their economies?" (OPEC 1998). On the trading of emissions permits, the OPEC statement in 1998 at Buenos Aires asked, "Could these mechanisms result in the surreptitious inclusion of additional commitments for developing countries?" Further, "The picking of 'low-hanging fruit' [trading for easy, cheap carbon reductions] in developing countries will be largely to the benefit of a far too narrow selection of countries elsewhere in the world."

At the other extreme are the AOSIS nations (Alliance of Small Island States), like Nauru, Marshall Islands, and Fiji. These nations have begun to loudly argue that they are "among the most vulnerable to impacts of climate and sea-level changes ... the inhabitable land tends to be on the coastal fringe ... climate change could mean changes in storm frequencies and intensity and lead to increased risk of flooding. It could upset sediment balances on the islands, leading to beach erosion and displacement of settlements and infrastructure" (AOSIS 1999). These nations could all suffer "a terrifying, rising flood of biblical proportions ... the willful destruction of entire countries and cultures," as described by Kinza Clodumar, president of Nauru at Kyoto in 1997. They call for immediate, drastic action to curb global warming. Their argument is that "(a) Pacific Island Countries make a small or negligible contribution to GHG; (b) They are among the countries which are most impacted; and (c) Knowledge of relevant parameters is very low." Currently 84 countries have signed the Kyoto protocol, but only 9 have ratified it and therefore agreed formally to its binding limits. Almost all of these are the small island states (Depledge 1999).

A third group is India and China, both with huge populations and both industrializing quickly. They call for equity based on carbon emissions per capita, by which both are light-years behind the industrial nations. As mentioned above, one U.S. citizen releases 8 times as much carbon as a Chinese citizen, and 20 times an average Indian. By these measures, both nations could grow and pollute more for a long, long time without reaching a globally standard per capita amount of carbon emissions. China has said that it will not commit to reduce any emissions before 2020. China's lead negotiator said, "In the developed world only two people ride in a car, and yet you want us to give up riding on a bus." India's Centre for Science and Environment pointed out that even when the poor nations emit as much as the wealthy ones, 20% of the world's population will still be responsible for 50% of its carbon (Dunn 1998).

There is also a silent majority of nations in the developing world. Among the 134 nations in the G-77, many are still unheard from on climate. Guyana's delegate, speaking at the 1 June 1999 meeting in Bonn as representative of the group, made clear that the G-77 would "play ball" on climate change only if the redistribution of resources was immediately forthcoming. "Transfer and access to technology, including information technology and enhancement of endogenous technologies, and the provision of financial resources, in particular for the full participation of developing countries in the implementation of these decisions, are necessary" (Drayton 1999). Some argue that this group of 134 has gained "the upper hand" and a series of concessions since Kyoto because most developing countries face no domestic environmental movements pushing them to act on the issue (Depledge 1999). Their states are therefore able to play the negotiations for all the concessions they can get (see Miller 1995; Young et al. 1996).

Some countries have taken a "middle ground" approach, such as the Philippines and especially Argentina, attempting to do some brokering between north and south. Under strong pressure from the United States, Argentina, at the COP-4 round in Buenos Aires in December 1998, proposed accepting emissions targets but only if they applied to *expected* growth, rather than current levels. Kazakhstan also expressed an intention to accept voluntary targets. Korea's role has been fairly protreaty, while Brazil's position has been tougher, not far from those of India and China. Under pressure from President Clinton, Brazil's Chancellor Luiz Felipe Lampreia said flatly, "We cannot accept limitations that interfere with our economic development" (Rossi 1997; A-15). *The Earth Times* called the U.S. attempt to find some allies among the poor nations a "classic tactic of British colonialism—divide and rule" (Gupte 1998). G-77 and China responded with revulsion to the proposal of any binding limits on noncore nations, which “poisoned the atmosphere of negotiations throughout the Conference” (La Rovere 1999). The Clean Development Mechanism remains “extremely controversial,” with some 142 issues needing to be worked out in the next 2 years (Depledge 1999). Poorer nations who are willing to talk about accepting some binding limits on carbon emissions have used the CDM to begin a new round of economic redistribution, since foreign aid has dropped to post World War II lows.

Finally, and quite importantly, there are the “emissions entrepreneurs,” states where projects are already underway to gain some of the potential investments from trading of carbon permits or technology transfer. These countries include most of Central America, Ecuador, and are being led by Costa Rica. That nation has already established an emissions-trading program, certifying reforestation and preservation projects. These nations can see clearly what many others do not: how they can capture a large share of this new redistribution if they act quickly.

Only One Way Out: Decarbonize Development

Because of the way the Kyoto treaty is structured, without the United States or Russia the treaty will not go into effect. In effect, polluting elites in the United States have successfully achieved a stranglehold over any progress. The lack of progress in the U.S. in reducing carbon emissions increases the likelihood that the “greenhouse coalition” lobby and the Senate will resist even meeting the original Kyoto agreement. And given the U.S. lack of progress meeting its own target for the end of the next decade, developing nations are provided a ready excuse for not making cuts. As Brazil’s leading newspaper put it, “Numbers like these [the U.S. emissions] reinforce the disposition of the Brazilian government to reject the idea of taking on additional costs to do its part in reducing the greenhouse effect” (Rossi 1997).

Now some authors are calling for abandoning the Kyoto process entirely and moving on to voluntary initiatives or starting again with only those nations who really want to do something about the problem. They hope to use the politics and psychology of shame and marginalization, to move eventually to a second-generation protocol (e.g., Flavin 1998).

Dunn (1998) and others make a lot of the idea that to resolve the deadlock over inequality and climate change, that development needs to be “decarbonized,” that is, delinked from fossil fuel consumption. The “Clean Development Mechanism” (CDM), which is still emerging from the Kyoto round of talks, is one way that could take place.

> The objective of the CDM is to help the South further its development goals in a less-carbon intensive fashion, while offering the North some flexibility in meeting its Kyoto commitments. As envisioned, the fund would channel Northern investment, technologies, and practices into developing country projects such as solar installations, wind farms, efficient industry boilers, and tree-planting programs. . . . A share of the proceeds from the mechanism will be used to help particularly vulnerable developing countries, such as island states and Bangladesh, cover the costs of climate disruptions. (Dunn 1998, 25)

Much more work needs to be done for political economy or sociology more broadly to meet that potential, and world-system theory may be one way to begin (see, e.g., Roberts and Grimes 1999).

This brings the discussion back to the original point about how issues of equity will have to be dealt with at the same time as the environment (Shue 1992). Richard Benedick, chief U.S. negotiator on the Montreal Protocol, said that "The North's interests in maintaining a healthy planet can only be achieved through aggressive efforts to support national economic advancement in the South" (quoted in Dunn 1998). The point deserves remembering; it's nearly identical to the lesson environmentalists are learning from the persistent demands of environmental justice advocates. Equity and ecology must be dealt with together.

Notes

1. Loren Lutzenheizer's 1996 analysis shows that U.S. citizens with incomes over $75,000 emit nearly four times the amount of carbon as those whose income is under $10,000. We lack analysis on this inequality within other nations, but if the average American emits 16,000 times that of the average Somali, 100,000 or more poor Somalis probably emit as much as one millionaire in the United States.

2. This is a legitimate thing to do since virtually all the carbon emitted since 1945 is still in the atmosphere. Carbon dioxide has an atmospheric lifetime of 120 years, methane 12 years (CDIAC 1999).

References

AOSIS (Alliance of Small Island States). 1999. AOSIS home page: The impacts of climate change on Pacific Island countries. <http://www.sidsnet.org/about/aosis.html, http://chacmool.sdnp.org/sidsdocs/cc.htm>

CARE Development Facts. 1996. 2 September. <http://www.care.org/world/devfact.html>

Carbon Dioxide Information and Analysis Center (CDIAC). 1999. Current greenhouse gas concentrations. <http://cdiac.esd.ornl.gov/pns/current_ghg.html> accessed 8 April 1999.

Depledge, J. 1999. Coming of age at Buenos Aires. The climate change regime after Kyoto. *Environment* 41(7):15–20.

Dewar, H. 1997. Senate advises against emissions treaty that lets developing nations pollute. *Washington Post* 26 July:A11.

Dietz, T. and E. A. Rosa. 1997. Effects of population and affluence on CO_2 emissions. *Proc. Natl. Acad. Sci. USA* 94:175–179.

Drayton, A. 1999. Statement on behalf of the Group of 77 and China by Ms. Alison Drayton of the Delegation of Guyana, at the Opening of the 10th Session of the Subsidiary Bodies of the United Nations Framework Convention on Climate Change. Bonn, Germany, 1 June. <http://www.g77.org/Speeches/060199B.htm>

Dunn, S. 1998. Dancing around the climate issue. Can the North and South get in step? *World Watch* November/December:19–27.

Flavin, C. 1998. Slow give-and-take, the climate itself is running amok. In Buenos Aires. *World Watch* November/December:11–18.

GCC (Global Climate Coalition). 1998. What others are saying. <http://www.globalclimate.org/others.htm> accessed 24 July 1999.

Gelbspan, R. 1997. *The heat is on.* Reading, MA: Addison-Wesley.

Gupte, P. 1998. Developing countries need to support limits on global emissions. *Earth Times News Service.* <http://editor@earthtimes.org/nov/climatechangedevelopingnov4_98htm> accessed 24 July 1999.

Intergovernmental Panel on Climate Change (IPCC) Working Group II. 1995. Climate Change. Second Report of the IPCC, Bonn.

Kasting, J. F. 1998. The carbon cycle, climate, and the long-term effects of fossil fuel burning. *Consequences* 4(1):15–27.

Reflections—Defining and Measuring Sustainability

Geoffrey Heal*

Introduction

There is a saying in business schools that "what gets measured, gets managed." This is bad news for managing our natural capital and the interactions between our economic activity and the environment, which are key to sustainability: these are almost entirely unmeasured and often unmanaged too.

None of the usual measures of economic performance—gross domestic product (GDP), unemployment, inflation—tell us anything about the state of our natural capital. In fact, they can be downright misleading.[1] For example, some parts of India are running out of water, and the water table is falling. Farmers have to drill deeper wells to find water, using more labor and energy. But because this extra spending raises GDP, water shortages appear to be raising India's GDP and making the country better off. While the water shortages are indeed raising economic activity in a macroeconomic sense, more important is the fact that they are a threat to growth in the future: when it is no longer possible to find more water by drilling deeper, agricultural output will collapse and welfare will drop.

So we are missing a warning sign here—the falling water table—and wrongly interpreting it as contributing to growth.

In fact, there is little in our normal economic statistics that would warn us of an impending environmental crisis. We might be warned we are about to run out of oil by high oil prices or out of soil by high food prices, but for a broad range of environmental goods and services, there are no markets, no prices, and hence no signals in the market system to warn us of potential problems.

Can we improve on this? That is the main question I address in this review of the literature on sustainability. This requires not only a discussion of the alternatives to GDP but also an examination of the concept of sustainability and the possibility of quantifying it.

*Professor, Columbia School of Business; e-mail: gmh1@columbia.edu. This article is based on a chapter of my forthcoming book, *Whole Earth Economics*.

[1]For a more general critique of GDP, see Stiglitz, Sen, and Fitoussi (2010) and the pioneering work of Nordhaus and Tobin (1972).

Review of Environmental Economics and Policy, volume 6, issue 1, winter 2012, pp. 147–163
doi:10.1093/reep/rer023
Advance Access publication on January 3, 2012

Alternatives to GDP

What changes could we make in our measures of economic performance that would improve on GDP so that it provides an accurate signal about our environmental performance?

Net Domestic Product

One relatively simple move, at least conceptually, would be to move from GDP to net domestic product (NDP). The difference between these two measures is that the depreciation of capital is subtracted from GDP to arrive at NDP. However, as the national accounts currently exist, the capital whose depreciation is subtracted is physical capital, the only type of capital currently measured.

So subtracting the depreciation of physical capital converts GDP to NDP, which is really a better measure of what the economy is producing and making available to its members. The reason we currently use GDP rather than NDP is purely pragmatic: it is hard to measure depreciation of capital accurately. Our knowledge of depreciation comes largely from how assets are written down for tax purposes, and depreciation rules in tax codes are arbitrary, reflecting more the outcome of political lobbying than economic analysis. As a result of being depreciated over relatively short time frames, many assets are carried on corporate balance sheets at almost no value when, in fact, they are still nearly as valuable as when they were new. True economic depreciation, of course, is the loss of economic value, which for an asset is the decline in the net present value of the services it can provide in the remainder of its life.

From an environmental perspective, the difference between GDP and NDP is important because much of the impact of human activity on natural capital can be thought of as depreciating this capital, reducing it in amount or value. A simple example is Saudi Arabia, which makes its living by extracting and selling oil. Oil is a form of natural capital, so Saudi Arabia is running down its natural capital. Each year its stocks of oil are lower than they were the previous year, with the decrease representing the depreciation of the country's natural capital. The depreciation of this capital is the value of the oil that Saudi Arabia sells. This means that if we were to calculate its NDP by subtracting the depreciation of natural capital from GDP, it would more or less cancel out Saudi Arabia's income from the sale of oil, which is most of its income, and would thus leave the Saudis poor.[2]

The point here is that NDP is a better measure of output than GDP, particularly if we want to measure and value changes in natural capital. However, while NDP is clearly an improvement on what we use today, it by no means provides an answer to the question "What should we measure?"

Human Development Index

Some people have sought to address this question by moving away from a money-based income measure altogether and instead constructing something that tries to measure the well-being of members of a society more directly. The best known of these measures is

[2]For an early discussion of this point, see Repetto et al. (1989) and Dasgupta and Heal (1979).

the human development index (HDI), developed and published by the United Nations Development Program (UNDP). The HDI is based on data in three areas—health, education, and income—which UNDP sees as the key dimensions of welfare. More specifically, for each country, UNDP collects data on life expectancy at birth, mean years of schooling, and average income per capita, and, by taking the equal-weighted geometric mean, combines them into a single number, the country's HDI score. The HDI does not address environmental issues but could provide a model for nonmonetary measures that include environmental data, some of which have been developed recently.[3]

Gross National Happiness

An intriguing variant on the conventional approaches to measuring economic performance is found in the small Himalayan kingdom of Bhutan, where economic and social performance is measured by the index of gross national happiness (GNH).[4] The country's Buddhist culture influenced its decision to move toward a more spiritual measure than GDP with its exclusive emphasis on material possessions. GNH tries to take into account performance across nine dimensions: psychological well-being, time use, community vitality, culture, health, education, environmental diversity, living standard, and governance. Environmental diversity, intended to be a measure of the health of natural capital and ecosystems, is measured by the level of afforestation or deforestation and some other measures of environmental degradation. Time use is a measure of how much time is available for nonwork activities such as recreation and time with family and friends. It also measures the time devoted to volunteer activities that help the community. Community vitality is an attempt to measure trust, reciprocity, how safe people feel, and how closely connected they feel to others.

The intentions behind GNH are clearly excellent; it's difficult to fault them. If there are problems, they lie in the execution of the GNH: Is it in fact possible to measure these concepts? How do we combine the results of measurement into a single number? And, in particular, could this measure be calculated for a country the size and complexity of the United States?

Adjusted Net Savings

Adjusted net savings (ANS) is one of the better measures of sustainability and may provide the kinds of warnings of impending environmental crisis that, as we noted earlier, conventional economic statistics fail to deliver. To calculate ANS, we start with a conventional measure of net investment in plant and equipment, that is, investment net of depreciation. We add to this investment in human capital through education and investment in intellectual capital through research and development (R&D), and then we subtract the depreciation or degradation of natural capital. The World Bank (World Bank 2006, 2010) produces figures for the ANS for each country, but the data are also available on the UNDP web site (http://hdr.undp.org/en/statistics/). I explore and discuss ANS in more detail later in connection with sustainability issues.

[3]See, for example, the CIESIN Environmental Sustainability Index (CIESIN n.d.).
[4]For more information, see the Bhutan government web site at http://www.gnhc.gov.bt/.

Relative Performance of Selected Countries

To offer some sense of what these different measures of economic performance look like when they are applied to individual countries, figures 1, 2, and 3 show GDP per capita, the HDI, and the ANS for six countries: the United States and Germany (two leading industrial countries), China and India (the two preeminent emerging economies and leaders in the BRIC group [Brazil, Russia, India, and China]), and Botswana and Papua New Guinea (two very different small developing countries).

Figure 1 indicates that for all six countries GDP per capita has risen over the last thirty years. However, the implications of this income growth for the well-being of the average citizen are far from unambiguous (see Stiglitz, Sen, and Fitoussi 2010). Figure 1 also indicates that the United States and Germany are far richer than the rest, and it also shows very clearly that Botswana is much richer—in GDP terms at least—than either China or India.

Moving to figure 2, although the HDI measure is totally different from GDP, it actually tells a rather similar story. All countries have again improved their performance over the thirty-year period. The two richest countries are still at the top, although they are much closer together than for the GDP measure. Here China ranks on a par with Botswana, and India and Papua New Guinea are again at the bottom.

Moving to figure 3, which shows trends in ANS as a percentage of gross national income, we see a completely different picture: Botswana dominates the top ranking, with China second. The United States and Germany do not fare well, indicating that being rich is perhaps not the same as being sustainable. This is a good lead-in to the issue of sustainability, which is the focus of the remainder of this Reflections.

The Notion of Sustainability

When it comes to human well-being, we might ask two distinct questions: first, how well off are people now, and how is this level changing over time? And second, can current levels of

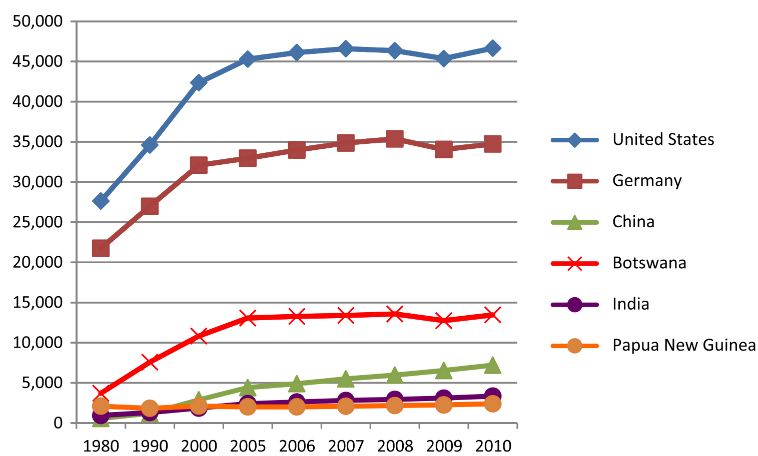


Figure 1 GDP per capita for selected countries, 1980–2010 ($US 2010)
Source: United Nations Environment Program web site, UNDP.

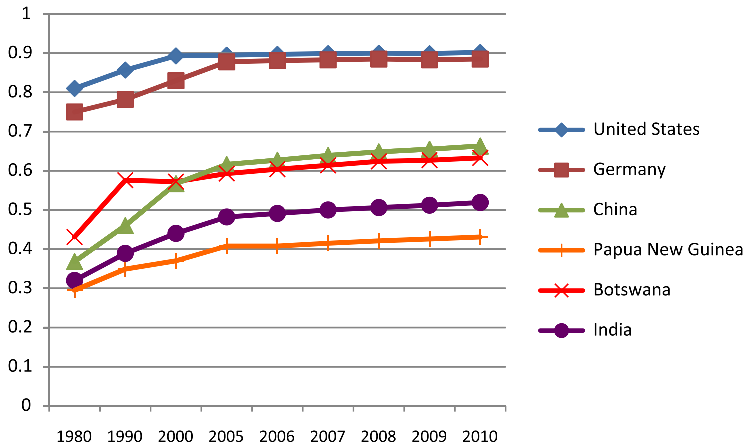


Figure 2 HDI for selected countries, 1980–2010
Source: United Nations Environment Program web site, UNDP.

well-being be sustained over time? That is, will our successors be able to live as well as we do? GDP (and its various refinements) and the HDI (as shown in figures 1 and 2) are attempts to answer the first question. The second question leads to a discussion of sustainability, which, to some degree, is captured by ANS (figure 3).

Sustainability is a buzzword these days. Everyone and everything wants to be perceived as sustainable. However, analytically, the issue of whether or under what circumstances growth can continue, or whether it will be limited by environmental and resource constraints, is not new. It was addressed extensively in the 1970s in the literature on exhaustible resources (see Dasgupta and Heal 1974, 1979; Hartwick 1977; Solow 1974; Stiglitz 1974) and indeed dates all the way back to Malthus (1798/1993). But it's especially topical today because it is becoming more evident that the scale of human activity is now sufficiently great as to affect the operations of global biogeochemical systems, with the carbon and nitrogen cycles obvious examples, and thus the future of life on earth.

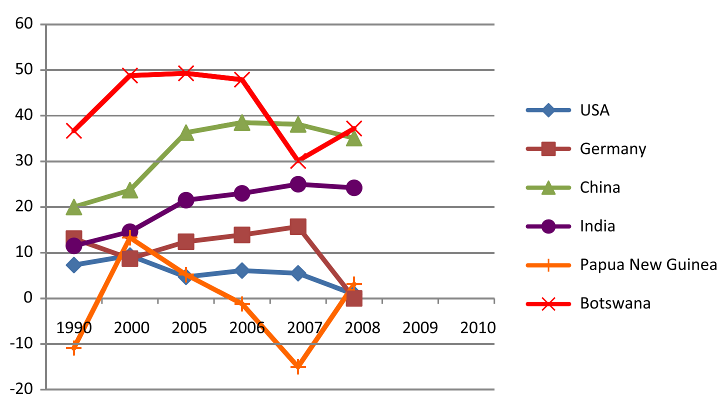


Figure 3 ANS as percentage of gross national income for selected countries, 1990–2008
Source: United Nations Environment Program web site, UNDP.

Some country examples will help to illustrate the issue of sustainability. I focus here on Botswana, Namibia, and Saudi Arabia, which are good countries to think about when trying to understand sustainability.

Botswana and Namibia

Botswana and its neighbor Namibia share a long common border and are similar in many ways: both are arid or semiarid and in the southern cone of Africa. Botswana's population is about 1.8 million, and its land area is 600,000 square kilometers, slightly smaller than Texas. Namibia has a population of 2 million and a larger land area, 825,000 square kilometers. Namibia has a long coastline on the Atlantic Ocean and a major fishery, whereas Botswana is landlocked. Both are largely desert, Botswana being dominated by the Kalahari Desert and Namibia by the Namib.

Botswana has one truly remarkable feature, the Okavango Delta in the north. The Okavango River is the only large river in the world not flowing into a sea: instead it flows into the Kalahari Desert. A huge river rising in the Central African highlands and flowing into one of the hottest deserts on earth is a stunning phenomenon. It creates a unique environment, amazingly rich in species. The environment varies, from aquatic where the river first reaches the desert, home of crocodiles and hippos and a range of fishing birds, to semiarid scrubland on the fringes of the areas irrigated by the river. The Okavango Delta is a natural asset that provides the basis for Botswana's immensely successful ecotourism industry, with thousands of high-paying visitors each year. Namibia has some remarkable landscapes too, especially the skeleton coast area in the north, and a growing ecotourism industry. Both are well worth a visit!

In addition to having the kinds of natural environments that support ecotourism, both Botswana and Namibia are rich in minerals. Botswana has huge deposits of gem-quality diamonds. Namibia also has diamonds, although fewer than Botswana, but in addition it is rich in uranium, lead, tin, zinc, silver, and tungsten. The long Atlantic coast also provides large fishing grounds.

Both countries are generating income by depleting natural capital (their diamonds, uranium, lead, fish, etc.), but they are also generating ecotourism income from their unique biodiversity. Thus they could offset their depletion of natural capital by building up holdings of other forms of capital, or they could let their total stock of capital assets fall. As it happens, Botswana is doing the former while Namibia is doing the latter.[5] As shown in figure 3, Botswana is a paragon of sustainability, building up both its human and physical capital. In fact, Botswana is one of the developing world's success stories, a much needed Africa success story. Its living standards have grown consistently and rapidly since independence in 1966, making it a middle-income country rather than a poor one, and it is also a very successful democracy. As we will see later, Namibia has been less successful in both accumulating capital and raising income levels.

[5]Namibia's unsustainable practices began when it was ruled by South Africa under a UN mandate dating from the end of the Second World War and the defeat of Germany, the former colonizer. The preintegration South African regime treated Namibia as a storeroom to be emptied rather than a country to be developed.

Saudi Arabia

The ultimate country example of unsustainability is Saudi Arabia. It is rich, certainly, but not sustainable, again illustrating that being rich and being sustainable are two very different concepts. Saudi Arabia makes its living by selling oil reserves, in effect selling off the family silver. Eventually, although not in the near future, as its reserves are huge, it will run out of oil and gas. Then there will be nothing to pull from the ground and sell, and unless Saudi Arabia has built up some other forms of capital, its living standards will suddenly collapse. Clearly, its living standards are not sustainable in the long run.

Defining the Concept of Sustainability

A lifestyle, a way of doing things, is sustainable if most of the world's population could continue it for a long time without major adverse consequences. That is, it is a potential dynamic equilibrium of some type.[6] Our current patterns of energy use are not sustainable: they produce greenhouse gases, changing the climate and leading to threats to our lifestyles and even our civilization. Our current patterns of agricultural production are probably not sustainable either: they lead to loss of soil and massive pollution of waterways by fertilizers, are threatened by a changing climate (see Schlenker, Fisher, and Hanemann 2005), and probably depend on levels of water availability that will not continue. Our current patterns of water use are not sustainable: we are depleting underground water faster than it recharges and polluting surface water. And our current levels of fish catch are manifestly unsustainable: they will destroy key fish populations within decades. In analytical models, sustainability is generally defined, as we will see later, in terms of the potential to maintain current living standards well into the future.

The Brundtland Report (1987),[7] written for the United Nations by a committee chaired by Dr. Gro Harlem Brundtland, the former prime minister of Norway, defines sustainable development as "development that meets the needs of the present without compromising the ability of future generations to meet their own needs." Although vague, this does make one point clear: there is an intergenerational aspect to sustainability. That is, we could live well now but ruin the earth in the process and pass on to our successors a world that is greatly diminished, as the current patterns of resource use discussed in the previous paragraph suggest we may be doing. The Brundtland idea of sustainability urges us not to do this. This idea of responsibility to later generations is often interpreted as meaning that we should leave them enough assets to be as well off as we are. This goal is built into the idea of ANS. Thus I consider ANS in more detail toward the end of this article because it can tell us whether we are actually leaving enough for future generations to be as well off as we are. In the remainder of this section I talk about natural capital and the role it plays in sustainability.

[6]For analytical discussions of the concept of sustainability, see Heal (1998) and Neumayer (2010).
[7]Formally the report of the World Commission on Environment and Development (1987).

Trading Natural for Physical and Intellectual Capital

There is another dimension of sustainability, omitted by the Brundtland definition, that has to do with the natural environment. The natural environment—or natural capital—is of immense value to human societies. We depend on it in many ways, and there are services it provides that are likely irreplaceable (see Daily 1997 and Heal 2000). A lot of our activities, those that damage the environment, are depleting this natural capital, running it down so that future generations will inherit less and poorer natural capital: a world with a less stable and hospitable climate, fewer species, less water, and less of many other environmental assets. It is this environmental damage and depletion of natural capital that may be making our activities unsustainable and perhaps condemning our successors to an impoverished lifestyle.

To be fair to ourselves, while we are leaving future generations less natural capital than we inherited, we are leaving them more than we inherited in terms of built capital: more freeways, airports, buildings, and infrastructure. We are also leaving them more intellectual capital than we inherited: our R&D programs are developing cures for diseases, new products, and new ways of doing things. In only the last twenty years the Internet and wireless communications have come from nowhere to dominate our lifestyles: we will hand these on to our successors, together with other things not yet invented, perhaps offsetting or compensating for the depleted environment that we are also leaving them.[8] Thus the composition of humanity's portfolio of capital stocks is changing, away from natural capital and toward other forms of capital.

Will this compensation for the loss of natural capital be adequate? Can we compensate for a depleted natural environment by more of the fruits of human labor and ingenuity? So far we certainly have: we are by general consent far better off than our predecessors a century ago. But what has happened in the interim is that we have built up our intellectual and physical capital massively while at the same time we have run down our natural capital. We have lost forests, rangelands, and quite a number of species, but we have gained cures for common diseases, acquired central heating and air conditioning, domestic appliances, cell phones, laptops, and the Internet. We have traded Spix's macaw, the Chinese freshwater dolphin (the Baiji), and other unique species for the iPhone and other (not so unique) gadgets. Most of us are probably not unhappy with this deal. So it appears that to date we have been able to compensate ourselves for declining natural capital by amassing more of the fruits of human labor and ingenuity. Can we continue this way? Can we maintain or increase our well-being if we continue trading natural environments and endangered species for better technology and infrastructure? This is the crux of the sustainability issue.[9]

It seems likely that matters are changing and that in the future we will not in fact be able to compensate for the loss of natural capital through the accumulation of other forms of capital. Climate change is a new phenomenon, not something we were aware of a century ago. It has grown to prominence with the massive expansion of fossil fuel use in the twentieth century,

[8] For a very clear documentation of these trends in the composition of capital stocks, see World Bank (2006, 2010).

[9] It is also possible that future people will value these natural assets differently from us—perhaps more as they will be scarcer, perhaps less as they will be used to a more synthetic world. On this see Beltratti, Chichilnisky, and Heal (1998) and LeKama and Schubert (2004).

will lead to changes that are qualitatively quite different from those resulting from past economic activity, and within the relatively short time frame of the next few decades could inflict substantial costs on the world. The current rate at which we are losing species is also without historical precedent: while we have driven species to extinction in the past (the dodo and the passenger pigeon are examples), we have never threatened as many species with extinction as we are doing today. Forests are being cleared at a rate unprecedented even in the times of the Industrial Revolution when wood was the principal fuel. We are also having a dramatic and negative effect on the oceans: the populations of large fish in the sea—the ones we eat—are claimed to be down to about 10 percent of what they were only half a century ago.[10]

The bottom line here, then, is that we are depleting our natural capital faster than ever before, and we have already depleted it to a greater extent than ever before. We depend on it, we need it, and our current lifestyle will not survive without it. So it is not clear that the old trade-offs will continue to work, that we can compensate for the loss of natural capital as we have in the past by adding more and more intellectual and physical capital.

Compensating for the Loss of Natural Capital: Some Case Studies

It is interesting to look in detail at this issue of compensating for the loss of natural capital with the buildup of other forms of capital by examining some specific cases. Again, Saudi Arabia is a good example. Let's look at some numbers to get a rough idea of the magnitudes involved. Saudi Arabia produces roughly 10 million barrels of oil per day and has a population of about 25 million. At its peak, oil was selling for about $130 per barrel, although this is much higher than the average price for the last few years, which is below $100. In any case, at the high price of $130 per barrel, Saudi annual oil revenues amount to just under $19,000 per capita. That is to say, if the total oil revenues were divided equally among all Saudis, then each would receive about $19,000 per year. A family of four would have just under $80,000. Not superrich, but not bad for not working. At a price of $60, much more typical of prices over the last few years, each person would get $8,700, making about $35,000 for a family of four. When Saudi Arabia's oil runs out, this income will suddenly cease, and there will be nothing to replace it, unless some of the revenue from oil has been invested in a way that can replace oil as a source of revenues. For example, a fraction of the oil revenues could have been invested in shares and bonds from around the world, yielding a flow of dividends and capital gains to replace the income from oil when the time comes. If enough were invested to replace all the oil income, then Saudi Arabia would be running its economy sustainably. Alternatively, Saudi Arabia could invest income in productive assets such as factories and in the education of its people. This too could generate a source of income that could replace oil revenues in due course.[11]

In contrast to Saudi Arabia, there are oil-producing countries or regions that are trying to run their economies sustainably. Two good examples are Norway and Alaska, through the Norwegian State Petroleum Fund and the Alaska Permanent Fund, respectively. Both of these funds are set up to take revenues from the sale of oil and invest them to provide a long-run income source that will continue even after the oil reserves are depleted. In the case of the

[10]There is some dispute about the accuracy of this claim: see Myers and Worm (2003) and Sibert et al. (2006).
[11]For more discussion of these issues, see Heal (2007).

Norwegian State Fund, revenues for investment come from the government's 80 percent share in Statoil, the Norwegian oil company that develops the country's North Sea oil fields. This fund now has $147 billion invested. The Alaska fund receives about 25% of oil and gas royalties, and it now has accumulated about $28 billion. It pays an annual dividend to all Alaska residents, averaging over $1,000 per year per person and peaking at $1,800. In both cases, what we are seeing is the conversion of natural capital—oil is a form of natural capital—into financial capital. The financial capital can continue and yield dividends after the natural capital is fully depleted. So these are both examples of countries or states compensating for the loss of natural capital by the accumulation of another form of capital.

It's worth thinking briefly about this from an accounting perspective. Think of the country's statement of assets and liabilities. Initially its assets consist largely of natural capital. Then over time this is depleted and the value of the asset falls. If this were all that happened, then the total value of the country's assets would fall. But if the revenues generated by the depletion of natural capital are invested in financial capital, a new asset appears on the balance sheet, the financial capital assets of the investment fund, and the buildup of these assets offsets to some degree the rundown of the natural capital assets. If well managed, this approach could keep the total value of the country's assets constant. Alaska and Norway may be doing this, but Saudi Arabia clearly is not.

Botswana offers another example of a country trying to run its economy sustainably, with the government deliberately investing a significant part of the revenues from diamond mining in physical and human capital (see Lange 2004 and Lange and Wright 2004). Both wealth per person and income per person have roughly tripled in Botswana in the last twenty years. In contrast, both wealth and income per person have declined in neighboring Namibia. Much of Botswana's success is due to its policy of consciously using revenues from natural capital to build up physical and other forms of capital, following the Hartwick (1977) rule. Lacking any explicit policy of using revenues from natural capital to build up other forms of capital, Namibia has seen declines in its total capital stock, the total value of its assets, and its per capita income. Figure 4 illustrates this contrast between Botswana and Namibia.

Back to Trade-offs

What do these cases tell us about sustainability? In the case of mineral resources, they suggest that basing an economy on running down natural capital need not imply unsustainable income levels. A country can compensate for the depletion of this type of natural capital by investing in other forms of capital, keeping its balance sheet intact and replacing one asset with another. The key question that this discussion raises is whether this is also true for the depletion of forms of natural capital other than mineral resources. That is, can we expect to compensate for the loss of aspects of the climate system, or the hydrological cycle, or our biodiversity or tropical forests, by building up more of the kinds of assets that we produce—physical or intellectual capital? Can the accumulation of physical and intellectual capital enable us to adapt to climate change, in the sense of maintaining our living standards in the face of an altered climate?

This is a controversial question. The basic issue here is the extent to which the services of capital constructed by humans can replace the services provided by living natural capital. Can we create substitutes for what we get from nature? In the limit, the answer has to be no. We

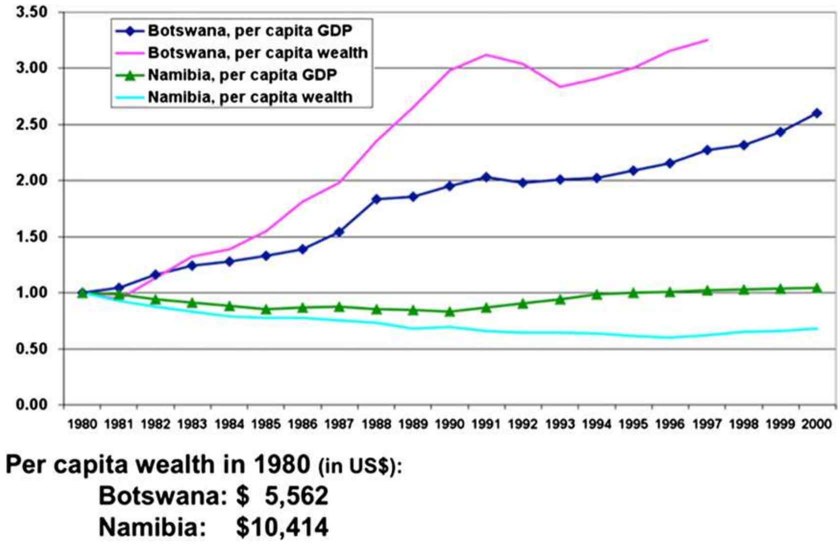


Figure 4 Index of real, per capita GDP and wealth in Botswana and Namibia, 1980–2000
Source: Lange (2004).

need oxygen—it's what powers our bodies. Oxygen is produced by photosynthesis, carried out by plants and by photosynthetic algae in the oceans, so we can't replace them as a source of oxygen. Food is also something whose production depends on the services of natural ecosystems: it depends on the productivity of soil, a complex ecosystem easily damaged by overuse; on the climate, determined in part by the complex worldwide carbon cycle; and on the actions of agricultural pests that attack food crops, as well as their natural predators, such as birds and bats, that keep these pests under control.

Clearly, we cannot replace all aspects of natural capital with physical or financial or intellectual capital. Mineral resources are just wealth: they provide their owners no other services than the wealth they generate in the market. So we can compensate for their depletion by building up our wealth, along the lines of Alaska, Norway, and Botswana. In the case of forests and coral reefs, and ecosystems in general, however, which provide more than just wealth, they cannot be fully replaced by financial assets or physical capital. New York City came to this conclusion when it chose to conserve the Catskill watershed, and it's what the Chinese government concluded when it chose to stop deforesting watersheds and instead moved to an aggressive program of reforestation: there was no real cost-effective substitute for natural capital in the form of forests and riverine ecosystems (see Heal 2000). What matters here are the elasticities of substitution between the various types of capital, both in production and in welfare functions. I am suggesting that for certain types of natural capital, this elasticity is less than one, implying that some minimum quantity is needed to maintain well-being, an idea modeled in a preliminary way in Heal (2009).

So sustainability requires that we keep some of our natural capital intact, as it provides services that matter to us and that we cannot replace. But there are other parts of natural capital that we can safely deplete because we can replace them with money or other assets we can produce. It is largely the living aspects of natural capital that are in the first category, exemplified by species and forests. And it is the inanimate natural capital that we can do without because we can replace it with something else. Ironically, from the way markets are working at the moment, one would think we had reached the opposite conclusion, since mineral resources, and oil in particular, are valued very highly, and biodiversity and forests almost not at all.

Weak or Strong Sustainability?

Sustainability comes in two varieties, weak and strong (see Neumayer 2010). So far, I've implicitly been talking about the former, weak sustainability. We are weakly sustainable if what we are doing will let future generations achieve our living standards or better—if we are not compromising the ability of future generations to meet their needs—which is the core of the Brundtland definition of sustainability. This is a common and totally plausible interpretation of the idea of stewardship and responsibility to future generations. But it does not incorporate any concept of stewardship of and responsibility toward the natural world and the other species with whom we share it, which is at the core of strong sustainability.[12]

As a consequence, not everyone is convinced we should be choosing the "weak" definition of sustainability. In particular, there are environmentalists who believe that natural capital itself, or at least the animate part of it, should be sustained, and that the constancy of this form of natural capital should be our criterion of sustainability. Sustainability for them means sustaining all forms of life on our planet, not just maintaining our own living standards, which they see as a narrow-minded and parochial goal. They see us as having an ethical responsibility to all life-forms on earth and not just to our own life-form. As the dominant species on earth, their argument goes, we owe it to other species, whose destinies are in our hands, to allow them to survive and prosper too.

In analytical terms, the relationship between weak and strong sustainability depends on the elasticity of substitution between animate natural capital and other forms of capital: if this elasticity is small enough, then the two concepts are not that different, whereas if it is large, then they are very different, and maintaining welfare levels does not require maintaining species.

Ultimately, the choice between weak and strong sustainability is a personal one: should we be seeking to conserve human living standards or to conserve all life-forms, or perhaps both? The former implies that we value the animate part of natural capital, biodiversity in essence, only insofar as it contributes to human welfare, whereas the latter implies that we value other life-forms in their own right.

This is an important distinction. The U.S. Endangered Species Act specifically seeks to save species from extinction even if there is no economic benefit to doing so: it reflects the belief that species have a right to exist independent of their value to us, a position taken by an increasing number of people who feel this is an issue on which they have to take a moral stance.[13] Personally, I sympathize with them, agreeing that we do not have the right to

[12]Economists have studied both concepts of sustainability and have called maintaining animate natural capital intact "strong sustainability" and maintaining human living standards "weak sustainability."
[13]Goble (2006) and Callicott (2006) develop this point further.

condemn other species to oblivion. I also believe that it's economically unwise to do so, so that to me, the two arguments move in the same direction.

Clearly, the world is not currently sustainable in the strong or ethical sense: this requires maintaining animate natural capital intact, which in turn requires not causing the extinction of other species. We are failing on this criterion. Whether we are succeeding or failing on the other criterion, namely weak sustainability, or keeping total capital intact and maintaining human welfare, is more of an open question.

Measuring Sustainability

Recall the adage "what gets measured, gets managed." So we need to be able to measure how sustainable or unsustainable our policies and institutions are.

There is, unfortunately, no way that GDP can indicate whether an income level or lifestyle is sustainable. For that we need to look at something quite different: how total wealth is evolving, where total wealth means the total value of *all* capital stocks—natural capital, physical capital, intellectual capital, and any other forms of capital that are relevant.

Measuring Income

We think of income as the return on wealth or on accumulated assets; it's the flow of payments or services from our wealth. Back in the 1930s, John Hicks defined income as "The maximum you can spend this month, consistent with spending the same in all subsequent months." This is a clever definition: it has an element of weak sustainability built in. According to this definition, Saudi Arabia's oil revenues are not income, as the oil will run out, but the earnings of the Norwegian sovereign wealth fund are income because they will be there in all future periods.

There is a subtle point here that we need to examine carefully. I noted earlier that there are some aspects of natural capital that we probably cannot replace, particularly its animate aspects, such as species and rain forests. If this is true, in what sense can we be sure we have a sustainable economy if the total value of the capital stock remains constant? After all, this constancy of the total could conceal a falling stock of animate natural capital and growing stocks of physical and intellectual capital, with the latter two replacing the former. The falling stock of animate natural capital could be compromising our ability to produce foods and medicines and stabilize the climate. I come back to this point later.

Measuring Wealth and Establishing Prices

When we talk about wealth, the total value of the stock of all types of capital, we are talking about a monetary or dollar value. For physical capital, this is relatively easy: we can find prices for items of capital equipment and then use these to value them. Intellectual capital is harder to value, although there are instances in which a value is clearly placed on ideas. For example, a firm may buy a patent from another: that is buying intellectual capital and is a process that puts a price on the intellectual capital that is transferred.

There are also situations in which natural capital is bought and sold; mineral rights, for example, can be traded. Soil is a form of natural capital that is partly mineral and partly animate—there are complex microbial and invertebrate ecosystems in soil that account for its productivity—and it is also bought and sold when land or farms are traded. So there are forms of natural capital, even living natural capital, in which there is a market and for which we can find prices. But there are certainly other types of natural capital for which there are no prices. Biodiversity, for example, is both an important component of natural capital and one for which there is typically no market and no prices. In a case like this, we need to calculate shadow prices, reflecting the value of the resources to society. This is, in fact, exactly what prices will reflect in a well-functioning competitive market. So, in computing a society's total wealth, we can value its natural capital either by market prices, if there are active and competitive markets, or by shadow prices otherwise.[14]

Impact of Scarcity on Prices

It's important to understand the effect of scarcity on the price or shadow price of a critically important form of natural capital. If some form of animate natural capital is truly essential to us, such as watersheds and the services they provide, and is becoming very scarce, then its shadow price will rise sharply. This is the possible resolution to the difficulty noted earlier, that constancy of the value of total capital could mask a decline in important types of natural capital that are essential to our continuing well-being. If one unit of this essential natural capital has a huge shadow price—say $1 billion—and we lose one unit of it, then to keep the total value of wealth constant, we would need to compensate with $1 billion of built or intellectual capital. And if the essential natural capital were to become even scarcer, then we might need even more of other forms of capital to make up for the loss of one unit. So eventually, the essential natural capital will be so scarce and its value so high that it will be impossible to compensate for its loss by adding more of other forms of capital, and the total value of wealth will fall. This would indicate nonsustainability.

More on ANS

Figure 3 illustrated a widely used measure of sustainability that attempts to measure the change in a nation's total wealth, ANS. Developed by the World Bank (see World Bank 2006, 2010; Hamilton and Hartwick 2005), ANS is, in principle, a measure of the total change in the value of all of a nation's capital stocks—physical, natural, intellectual. If it is positive, the country is sustainable, and if negative, it is not sustainable, at least in the sense of weak sustainability.[15] The formal proposition behind my interest in ANS is the following. Along a time path that satisfies the first-order conditions for dynamic optimality, an economy's state valuation function, which is the present value of all future welfare levels, is nondecreasing at time t if and only if ANS is positive at time t. This means that the economy's future is getting better, or no worse, if ANS is nonnegative.[16]

[14]For a more extensive discussion of the valuation of natural capital and the services that it provides, see National Research Council of the National Academies (2005).
[15]There is some complex theory behind this statement. See, for example, Heal and Kristrom (2008), Hamilton and Hartwick (2005), and Arrow, Dasgupta, and Mäler (2003).
[16]See Heal and Kristrom (2008) and the earlier references therein for more details.

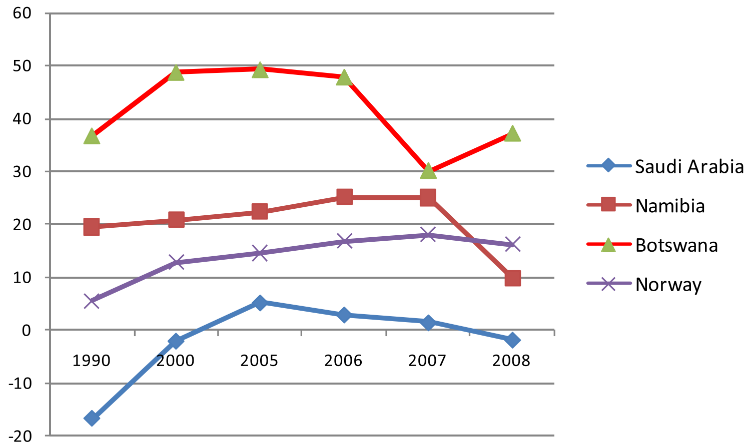


Figure 5 ANS for Botswana, Namibia, Norway, and Saudi Arabia, 1990–2008
Source: UNDP web site.

Figure 5 presents more ANS data, showing the movement of ANS as a percentage of GDP over time for Botswana, Namibia, Saudi Arabia, and Norway. Botswana's ANS is always the highest of the four, whereas Saudi Arabia actually has a negative ANS for much of the time. From the earlier discussion, we know why this is the case. Namibia, very similar to Botswana in many ways, has been much less effective in building up its capital stocks, although its ANS is still quite positive. Norway has a steadily positive ANS, in spite of the fact that its economy, like that of Saudi Arabia, is based on depletion of natural capital.

Let me end by emphasizing that when we discuss the constancy of total wealth as indicating sustainability, we are referring to the sustainability of living standards (i.e., weak sustainability). This was the focus of the Brundtland definition, which spoke of the ability of the present generation to meet its needs without compromising the ability of future generations to do likewise. However, in this context, we are not indicating whether living natural capital is being maintained at a reasonable level.

Conclusions

Where does this leave us in our overall assessment of sustainability and its measurement? First, the concept of sustainability clearly matters: we need to know whether we are compromising the abilities of future generations to live as we live. Second, we need to know whether we are sentencing a large fraction of life on earth to death to the power two, that is, to extinction and complete obliteration. Both issues matter and on the surface appear to reflect somewhat different interpretations of sustainability. But some researchers, particularly in biology and ecology, claim there is in fact no difference between these weak and strong interpretations of sustainability, that any world in which many species are obliterated will be one in which humans suffer too. In economic terms, these researchers are saying there is limited substitutability between some types of natural capital and other forms of capital (see Ehrlich and Goulder 2007).

Where are we when it comes to measuring sustainability and judging whether we are managing our affairs in a sustainable fashion? The report of the Commission on the Measurement of Economic Performance and Social Progress, appointed by President Nicolas Sarkozy of France to consider these issues and chaired by Joseph Stiglitz, concluded that currently we do not have the data to produce a single number that can tell us convincingly whether we are sustainable (Stiglitz, Sen, and Fitoussi 2009). What we would really like to have is the ANS measure, but we are not yet able to construct this accurately because we do not have good quantitative measures of some aspects of wealth, nor do we have measures of the economic values of several important types of wealth. Prominent among the categories of wealth that we cannot measure or value fully are some types of natural capital. The commission suggested that we measure ANS as best we can and continue to improve our measures until we have good ones, and that in the meantime we supplement the ANS with some additional data that indicate the physical state of some of the more important environmental threats that cannot be captured by a wealth measure, such as the concentration of greenhouse gases in the atmosphere, the number of species close to extinction, and the acidity of the oceans.

The bottom line here is that we are not doing well in measuring the sustainability of our economic activities, and that while we do not yet have the data to do a great job, we can do a lot better than we are currently doing. The question of whether we are behaving sustainably is sufficiently important that we need to be working harder to find the answer.

References

Arrow, K. J., P. S. Dasgupta, and K-G Mäler. 2003. Evaluating projects and assessing sustainable development in imperfect economies. *Environmental and Resource Economics* 26 (4): 647–85.

Beltratti, A., G. Chichilnisky, and G. M. Heal. 1998. Uncertain future preferences and conservation. In *Sustainability: Dynamics and uncertainty,* ed. Graciela Chichilnisky Alessandro Vercelli, and Geoffrey Heal, 257–76. Rotterdam, The Netherlands: Kluwer Academic Publishers and Fondazione Eni Enrico Mattei.

Callicott, J. Baird. 2006. Explicit and implicit values. In *The Endangered Species Act at thirty: Conserving biodiversity in human-dominated landscapes,* Vol 2, ed. J. M. Scott, D. D. Goble, and F. W. Davis, 36–48. Washington, DC: Island Press.

CIESIN Environmental Sustainability Index. http://sedac.ciesin.columbia.edu/es/esi/.

Daily, G. 1997. *Nature's services: Societal dependence on natural ecosystems*. Washington, DC: Island Press.

Dasgupta, P. S., and G. M. Heal. 1974. The optimal depletion of exhaustible resources. *Review of Economic Studies,* Symposium on the Economics of Exhaustible Resources, 3–28.

———. 1979. *Economic theory and exhaustible resources*. Cambridge, UK: Cambridge University Press.

Ehrlich, P., and L. Goulder. 2007. Is current consumption compatible with sustainability? A general framework and initial indicators for the United States. *Conservation Biology* 21 (5): 1145–54.

Hamilton, Kirk, and John M. Hartwick. 2005. Investing exhaustible resource rents and the path of consumption. *Canadian Journal of Economics* 38 (2): 615–21.

Hartwick, J. M. 1977. Intergenerational equity and investing the rents from exhaustible resources. *American Economic Review* 66 (December): 972–74.

Heal, G. M. 1998. *Valuing the future: Economic theory and sustainability*. New York: Columbia University Press.

———. 2000. *Nature and the marketplace*. Washington, DC: Island Press.

———. 2007. Are oil-producers rich? In *Escaping the resource curse,* ed. M. Humphreys, J. Sachs, and J. Stiglitz. New York: Columbia University Press.

———. 2009. The economics of climate change: A post-Stern perspective. *Climatic Change* 96 (3): 275–97.

Heal, G. M., and Bengt Kristrom. 2008. A note on national income in a dynamic economy. *Economics Letters* 98 (1): 2–8.

Goble, D. 2006. Evolution of at-risk species protection. In *The Endangered Species Act at thirty: Conserving biodiversity in human-dominated landscapes*. vol. 2, ed. J. M. Scott, D. D. Goble, and F. W. Davis, 6–23. Washington, DC: Island Press.

Lange, G. 2004. Wealth, natural capital and sustainable development: Contrasting examples from Botswana and Namibia. *Environmental and Resource Economics* 29 (3): 257–83.

Lange, G., and M. Wright. 2004. Sustainable development in mineral economies: The example of Botswana. *Environment and Development Economics* 9 (4): 485–505.

LeKama, Alain, and Katheline Schubert. 2004. Growth environment and uncertain future preferences. *Environmental and Resource Economics* 28 (1): 31–53.

Malthus, Thomas. 1798. *An essay on the principle of population*. Reprint, Oxford, UK: Oxford World Classics, 1993.

Myers, Ransom A., and Boris Worm. 2003. Rapid worldwide depletion of predatory fish communities. *Nature* 423 (May): 280–83.

National Research Council of the National Academies. 2005. *Valuing ecosystem services: Toward better environmental decision-making*. Washington, DC: The National Academies Press.

Neumayer, Eric. 2010. *Weak versus strong sustainability: Exploring the limits of two opposing paradigms*. Cheltenham, UK: Edward Elgar.

Nordhaus, W. D., and J. Tobin. 1972. Is growth obsolete? *Economic research: Retrospect and prospect*, vol. 5, *Economic growth*, 1–80. Cambridge, MA: National Bureau of Economic Research, Inc.

Repetto, R., W. Magrath, M. Wells, C. Beer, and F. Rossinni. 1989. *Wasting assets: Natural resources in the national income accounts*. Washington, DC: World Resources Institute. Also A. Markandya and J. Richardson, eds. 1992. *Environmental economics: A reader,* Chapter 25. New York: St. Martin's Press.

Schlenker, W., A. Fisher, and M. Hanemann. 2005. Will U.S. agriculture really benefit from global warming? Accounting for irrigation in the hedonic approach. *American Economic Review* 95 (1): 395–406.

Sibert, John, J. Hampton, P. Kleiber, and M. Maunder. 2006. Biomass, size and trophic status of top predators in the Pacific Ocean. *Science* 314 (5806): 1773–76.

Solow, R. M. 1974. Intergenerational equity and exhaustible resources. *Review of Economic Studies, Symposium on the Economics of Exhaustible Resources,* 29–48.

Stiglitz, J. E. 1974. Growth with exhaustible natural resources: Efficient and optimal growth paths. *Review of Economic Studies,* Symposium on the Economics of Exhaustible Resources, 123–38.

Stiglitz, Joseph E., Amartya K. Sen, and Jean-Paul Fitoussi. 2009. *Report by the Commission on the Measurement of Economic Performance and Social Progress*. Paris: Commission on the Measurement of Economic Performance and Social Progress. www.stiglitz-sen-fitoussi-fr. Also available as Joseph E. Stiglitz, Amartya Sen, and Jean-Paul Fitoussi. 2010. *Mis-measuring our lives: Why GDP doesn't add up*. New York: The New Press.

World Bank. 2006. *Where is the wealth of nations? Measuring capital for the 21st century*. Washington, DC: World Bank.

———. 2010. *The changing wealth of nations: Measuring sustainable development in the new millennium*. Washington, DC: World Bank.

World Commission on Environment and Development. 1987. *Our common future* [also known as the Brundtland Report]. http://www.un-documents.net/wced-ocf.htm.