Econometric Analysis of Discrete Choice

Debopam Bhattacharya
University of Cambridge

August, 2017

Binary Decisions

- Regression analysis concerns effect of X on Y (e.g., effect of study hours on exam results)
- Y continuously distributed in population
- But many important decisions are discrete
- Send daughter to school, ride train/car/bicycle, retire/work
- Want to know effect of covariates (education/income/caste/gender/price) on the probability of decision
- Important for policy analysis (e.g., tuition subsidy), how to target intervention

Example: Enrol teenage daughter in school in India

- School S, no-school N
- Monthly price of school P, monthly hhd expenditure W (No income data in NSS)
- **Economic** Model : Choose school if $U(S, W-P) - U(N, W) \geq 0$
- **Econometric** Model: Choose school if $U(S, W-P, \eta) - U(N, W, \eta) \geq 0$

Model of Choice

- Linear approximation

$$\begin{aligned} & U(S, W-P, \eta) - U(N, W, \eta) \\ \simeq \quad & \beta_0 + \beta_1 P + \beta_2 W + \eta \end{aligned}$$

- Latent desire to enrol $\longrightarrow$ Observed decision

$$\begin{aligned} Y^* &= \beta_0 + \beta_1 P + \beta_2 W + \eta, \\ Y &= \begin{cases} 1, \text{ if } y^* \geq 0 \\ 0, \text{ if } y^* < 0 \end{cases} \end{aligned}$$

Data and Statistical Assumptions

- Each observation in NSS is one teenager, indexed by i, observe P_i, W_i, Y_i
- $Y_i = 1\{\beta_0 + \beta_1 P_i + \beta_2 W_i + \eta_i \geq 0\}$, η_i indep of P_i, W_i
- Observed outcome nonlinear in error term η
- $\eta_i \sim N(0,1)$: Probit Model

Conditional Choice Probability

- Conditional Choice Probability

$$\begin{aligned} q(p,w) &\equiv \Pr(School_i = 1 | P_i = p, W_i = w) \\ &= \Pr(\beta_0 + \beta_1 P_i + \beta_2 W_i + \eta_i \geq 0 | P_i = p, W_i = w) \\ &= \Pr(\eta_i \geq -(\beta_0 + \beta_1 p + \beta_2 w) | P_i = p, W_i = w) \\ &= \Pr(\eta_i \geq -(\beta_0 + \beta_1 p + \beta_2 w)), \text{ by independence} \\ &= 1 - \Phi(-(\beta_0 + \beta_1 p + \beta_2 w)) = \Phi(\beta_0 + \beta_1 p + \beta_2 w) \end{aligned}$$

- Marginal Effect

$$\frac{\partial}{\partial p} \Pr(Y_i = 1 | P_i = p, W_i = w) = \beta_1 \times \phi(\beta_0 + \beta_1 p + \beta_2 w) \neq \beta_1$$

- Predicted value given price=500, income=10000

$$\begin{aligned} & \Phi(\beta_0 + 500 \times \beta_1 p + 10000 \times \beta_2) \\ \neq \quad & \beta_0 + 500 \times \beta_1 p + 10000 \times \beta_2 \end{aligned}$$

Parameter Estimation

- Once we know β's, we can predict the probability of enrolment at different price and income
- General method of estimation: Maximum likelihood
- Intuitively equivalent to minimum distance, analogous to least squares for linear model

$$\min_{\beta_0,\beta_1,\beta_2} \sum_{i=1}^{n} \left(School_i - \Phi\left(\beta_0 + \beta_1 P_i + \beta_2 W_i\right)\right)^2$$

- Get estimates $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$, can predict enrolment at price p and income w by $\Phi\left(\hat{\beta}_0 + \hat{\beta}_1 p + \hat{\beta}_2 w\right)$
- Std error, confidence intervals etc.

Endogenous Income

- W and η correlated, eg, parents' education is omitted variable
- Then probit regression does not give us what we want to know

$$q(p, w): \text{ what is the probability of school if given wealth } w$$

- If linear model then can use IV and two-stage least squares
- In nonlinear models need to use **Control Functions** – active research area
- Smith-Blundell (Ecta86), Blundell-Powell (ReStud04), Imbens-Newey (Ecta, 09)

Control function

- Find IV for income, eg., avg neighborhood income, Z
- Z indep of individual's own taste, Z correlated with W

1. Regress W_i on Z_i, P_i, residual $\hat{V}_i$
2. Add residual as additional regressor in Probit

$$\Pr(S_i = 1) = \Phi\left(\beta_0 + \beta_1 P_i + \beta_2 W_i + \underbrace{\beta_3 \hat{V}_i}_{\text{Correction term}}\right)$$

3. Finally average across $\hat{V}_i$

$$\hat{q}(500, 10000) = \frac{1}{n}\sum_{i=1}^{n} \Phi\left(\beta_0 + 500 \times \beta_1 + 10000 \times \beta_2 + \beta_3 \hat{V}_i\right)$$

Going deeper

- Simple Question

  What is the expected enrolment if govt offers a tuition subsidy of 100 rupees?

$$\frac{1}{n}\sum_{i=1}^{n}\Phi\left(\hat{\beta}_0+\hat{\beta}_1(P_i-100)+\hat{\beta}_2 W_i\right)$$

- Deeper Welfare Question

  What is the "Cash-Equivalent" (Equivalent Variation) of this subsidy, what is the deadweight loss?

My Research

- Because η varies across people, same subsidy has different cash-equivalents for different people
- **Average** EV of a Rs 100/month subsidy
- Challenge: EV based on utilities, not observable directly
- How to infer welfare-relevant features of utilities from choice data?

EV of a subsidy

- Recall $Y = 1\{U(S, W-P, \eta) - U(N, W, \eta) \geq 0\}$
- Price fall from P to $P-100$ due to subsidy
- For people with income w and facing price p_0, EV is defined by

$$\begin{aligned} & \max\{U(S, w+EV-p_0, \eta), U(N, w+EV, \eta)\} \\ = \quad & \max\{U(S, w-(p_0-100), \eta), U(N, w, \eta)\} \end{aligned}$$

- For fixed w, p_0, the EV is a function of η
- Distbn of $\eta \longrightarrow$ distbn of EV $\longrightarrow$ AverageEV

Bhattacharya (Ecta 2015)

$$
\begin{aligned}
E\left(EV\right) &= \int_{p_1}^{p_0} q\left(p, w + p - \left(p_0 - 100\right)\right) dp \\
&= \int_{p_0 - 100}^{p_0} \Phi\left(\hat{\beta}_0 + \hat{\beta}_1 p + \hat{\beta}_2 \left(w + p - \left(p_0 - 100\right)\right)\right) dp
\end{aligned}
$$

Change in Marshallian consumer surplus: Area under dd curve

$$
\begin{aligned}
&\int_{p_0 - 100}^{p_0} \Phi\left(\hat{\beta}_0 + \hat{\beta}_1 p + \hat{\beta}_2 w\right) dp \\
\neq\ & E\left(EV\right), \text{ unless } \textbf{zero income-effect}.
\end{aligned}
$$

Deadweight Loss

Efficiency cost of subsidy

$$\begin{aligned} E\left(DWL\right) &= 100 \times q\left(p_0 - 100, w\right) - \int_{p_0-100}^{p_0} q\left(p, w + p - p_1\right) dp \\ &= 100 \times \Phi\left(\hat{\beta}_0 + \hat{\beta}_1\left(p_0 - 100\right) + \hat{\beta}_2 w\right) \\ &\quad - \int_{p_0-100}^{p_0} \Phi\left(\hat{\beta}_0 + \hat{\beta}_1 p + \hat{\beta}_2\left(w + p - \left(p_0 - 100\right)\right)\right) dp \end{aligned}$$

These results bring together "empirical program evaluation" and "classical welfare theory" of public finance.

Extensions (Bhattacharya, working paper)

- Price of multiple options (school/work/stay-at-home) change simultaneously, merger analysis
- Eliminate Option (e.g. Child Labor legislation)
- Although illustrated with probit, results do not require assumptions on heterogeneity distribution

$$\begin{aligned} E(EV) &= \int_{p_1}^{p_0} q(p, w + p - (p_0 - 100))\, dp \\ q(p, w) &= \Pr(School | P = p, W = w) \end{aligned}$$

- **Nonparametric** Identification

Illustration

- Indian girls aged 15-18, 25000 households from NSS 2004
- Choice between stay-at-home/work/school
- Rs 1= \$0.02

Data Summary

	Mean	Std. Dev.	Min	Max
home	0.19	0.39	0	1
working	0.20	0.40	0	1
inschool	0.61	0.49	0	1
school-price	168.44	383.41	0	41666
local wage	123.59	170.75	0	1800
income	4063.04	6030.37	1	583192
female	0.44	0.49	0	1
age	16.13	1.70	13	18
hhsize	5.84	2.53	1	36
adult_literacy	0.34	0.23	0	1
muslim	0.15	0.35	0	1
scst	0.30	0.46	0	1

Welfare Effects

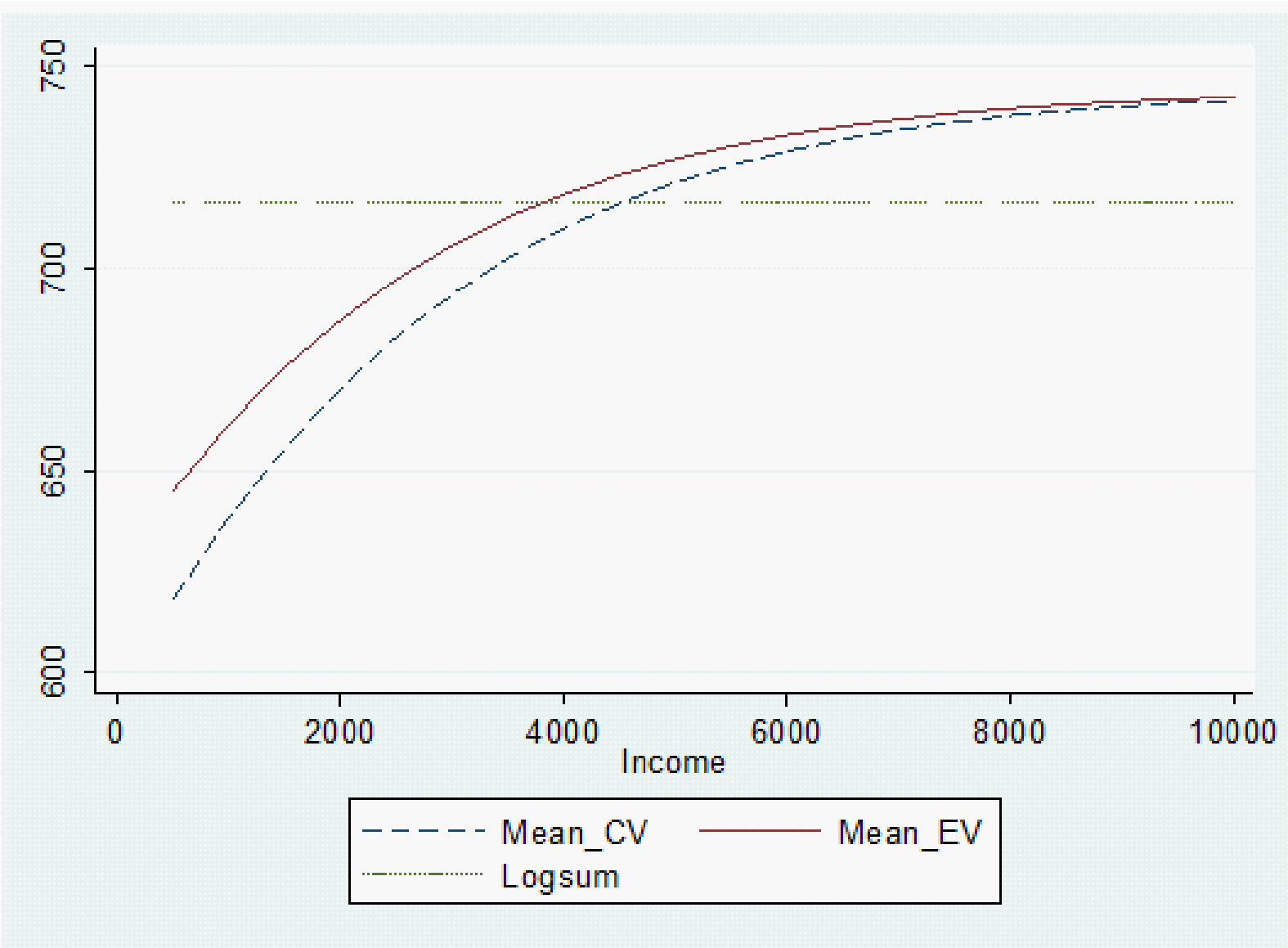
- Policy: Tuition subsidy plus wage rise $\mathbf{p}_0 = (0, -25, 790)$, $\mathbf{p}_1 = (0, -256, 45)$,
- Welfare calculation for Hindu, 15 yr old, family literacy 60%, hhsize 5, by gender and income (MPCE)

Price and Income Effects for Schooling: Semi-Elasticities

Covariate	Male	Female
School-fees	-0.02	-0.04
	(-11.41)	(-11.7)
Local wage	-0.006	-0.01
	(-3.08)	(-2.7)
Income	0.11	0.19
	-14.09	-14.6

Mean EV by Income

Female youth; tuition subsidy of Rs 745 and wage rise of Rs 251



Some related research

- Discrete Choice
  - McFadden and Domencik (1975), Small and Rosen (Ecta81), Dagsvik and Karlstrom (ReStud05)
  - Only under strong and unverifiable assumptions, e.g. no income effect
  - Can we do welfare analysis w/o making strong assumptions on preferences?
- Continuous Choice, eg., deadwt loss of taxing gasoline (petrol) consumption
  - Hausman (AER81), Hausman and Newey (Ecta15)

Continuous Choice (Hausman-Newey, Ecta 2016

$$\begin{aligned}\max_q U(q, w - pq, \eta) &\rightarrow q^*(p, w, \eta) \\ &\rightarrow U(q^*(p, w, \eta), w - pq^*(p, w, \eta), \eta) \\ &\equiv V(p, w, \eta), \text{ indirect utility function}\end{aligned}$$

For tax t on initial price p_0, EV solves

$$V(p_0, w - S, \eta) = V(p_0 + t, w, \eta) \rightarrow S(p_0, w, \eta)$$

In typical datasets, observe $\{q_i^*, P_i, W_i\}$, $i = 1, ..., n$

Unobservable from observable

Using Shephard's lemma

$$\begin{aligned} \frac{\partial S(p,w,\eta)}{\partial p} &= -q^*(p,w-S,\eta), \\ S(p_0+t,w,\eta) &= 0. \end{aligned}$$

- But average EV $\bar{S}(p,w) = \int S(p_0,w,\eta)\,dF(\eta)$ cannot be identified without any restriction on preferences: why?

$$\frac{\partial \bar{S}(p,w)}{\partial p} \neq -\underbrace{\bar{q}^*(p,w-\bar{S})}_{Estimable}$$

- Hausman-Newey show that we can get bounds
- Contrast with discrete case where average EV is point identified without any restriction on preferences.

Bibliography (Discrete Choice)

1. Amemiya, T. Advanced Econometrics, 1985.
2. Wooldridge, Jeffrey M. Econometric analysis of cross section and panel data. MIT press, 2010.
3. Blundell, R. W., and J. L. Powell. Endogeneity in semiparametric binary response models. ReStud, 2004.
4. Imbens, G. W., and W. K. Newey. Identification and estimation of triangular simultaneous equations models without additivity. Ecta 2009.
5. **Imbens-Wooldridge Lecture Notes "Control Functions", Lecture 6, NBER.
6. Smith, R. J., and R. W. Blundell. An exogeneity test for a simultaneous equation Tobit model with an application to labor supply. Ecta 1986.

Bibliography (Empirical Welfare Analysis)

1. Bhattacharya, D. Nonparametric Welfare Analysis for Discrete Choice, Ecta, 2015.
2. Bhattacharya, D. Empirical Welfare Analysis for Discrete Choice under General Heterogeneity, mimeo.
3. Hausman, J. A. Exact consumer's surplus and deadweight loss. AER 1981
4. Hausman, J. and W. Newey. Individual Heterogeneity and Average Welfare", forthcoming, Ecta.
5. Dagsvik, J, and A. Karlström. Compensating variation and Hicksian choice probabilities in random utility models that are nonlinear in income. ReStud 2005.
6. Rosen, H. S., and K. A. Small. Applied welfare economics with discrete choice models. Ecta, 1981