Estimating the random coefficients logit model of demand using aggregate data

David Vincent

Deloitte Economic Consulting
London, UK

davivincent@deloitte.co.uk

September 14, 2012

Introduction

- Estimation of consumer demand in differentiated product industries plays a central role in applied economic analysis
- The conventional approach is to specify a system of demand functions that correspond to a valid preference ordering, and estimate the parameters using aggregate data
- A popular example is the Almost Ideal Demand System of Deaton(1980), where market shares are linear functions of the logarithm of prices, and real expenditure.
- A major concern in adopting this approach, is the large number of parameters that need to be estimated, even after the restrictions of adding-up homogeneity a symmetry have been imposed
- The dimensionality problem can be solved if preferences are assumed to be separable; however, this places severe restrictions on the degree of substitutability between goods in different sub-groups

Introduction

- The logit-demand model (McFadden 1973) is another way to address the dimensionality problem, by assuming instead that consumers' have preferences over product characteristics
- Although easy to estimate, this model again imposes strong a-prior restrictions over the patterns of substitutability
- The purpose of this presentation is to discuss the random coefficients logit demand model (Berry Levinhson Pakes 1995)
- This framework accommodates consumer heterogeneity, by allowing taste parameters to vary with individual characteristics and requires market level data for estimation
- The model produces cross price elasticities that are more realistic and allows for the case where prices are endogenous
- It is very popular in the Industrial Organization literature and routinely applied by regulatory authorities, yet these is no official BLP Stata command!

The Model

- Following Nevo(2005), assume we observe $t = 1, .., T$ markets consisting of I_t consumers and J products. For each market, data is available on total quantities sold, prices and product characteristics of all J products
- Markets are assumed to be independent and can be cross-sectional (e.g. different cities) or repeated observations
- let u_{ijt} denote the indirect utility that individual i experiences in market t when consuming product j, and assume this depends on a $K \times 1$ vector of product characteristics x_{jt}, price p_{jt} an unobserved component ξ_{jt}, and an idiosyncratic error ϵ_{ijt}. If the utility function is quasi-linear utility, then:

$$u_{ijt} = \alpha_i(y_i - p_{jt}) + x'_{ijt}\beta_i + \xi_{jt} + \epsilon_{ijt} \tag{1}$$

- where y_i is income, β_i is a $K \times 1$ vector of coefficients and α_i is the marginal utility of income.

The Model

- Consumer i also has the choice to buy the outside product $j = 0$ with normalized utility $u_{i0t} = \alpha_i y_i + \epsilon_{i0t}$.
- Both β_i and α_i and assumed to be linear functions of characteristics D_i and v_i of dimensions $d \times 1$ and $(K+1) \times 1$:

$$\begin{pmatrix} \beta_i \\ \alpha_i \end{pmatrix} = \begin{pmatrix} \beta_0 \\ \alpha_0 \end{pmatrix} + \Pi D_i + Lvi \tag{2}$$

- where $\mathrm{v}_i \sim iid(0, I_{K+1}), D_i \sim iid(0, \Sigma_D)$,$\Pi$ is a $K+1 \times d$ matrix of coefficients, and $LL^{'} = \Sigma_v$
- Although both D_i and v_i are unobserved, the distribution of the demographics D_i including Σ_D is assumed to be known
- This is not the case for v_i where a parametric distribution is assumed (e.g. normal)
- In practice $F_D(D)$ is the empirical non-parametric distribution

The Model

- Define the set: $A_{ijt} = \{\epsilon_{it} : u_{ijt} > u_{ikt}, \forall j \neq k\}$, then the probability that individual i selects product j in market t is

$$Pr_{ijt} = \int_{A_{ijt}} dF(\epsilon_{it} \mid D_i, v_i) \qquad (3)$$

- Integrating over the unobserved variables D_i and v_i yields:

$$Pr_{jt} = \int_{D_i} \int_{v_i} Pr_{ijt} dF(D_i \mid v_i) dF(v_i) \qquad (4)$$

- where Pr_{jt} is the same for all i and can be estimated by the product share $s_{jt} = \frac{q_{jt}}{M_t}$ where M_t is the market size
- The error in this approximation is $O(I_t^{-1/2})$ and will be negligible for large I_t which is often the case

The Model: Distributional Assumptions

- To evaluate the integral in (3) first assume that ϵ_{ijt} are *iidd* and have a Type I extreme value distribution. Then:

$$Pr_{ijt} = \frac{\exp(x_{ijt}^{'}\beta_i - \alpha_i p_{jt} + \xi_{jt})}{1 + \sum_k \exp(1 + x_{ijt}^{'}\beta_i - \alpha_i p_{jt} + \xi_{jt})} \qquad (5)$$

- To evaluate (4), it is necessary to specify the distributions of D_i and v_i. At one extreme, we could assume $\Sigma_D = \Sigma_v = 0$
- Although appealing, consider the price elasticities:

$$e_{jkt} = \begin{cases} -\alpha_0 p_{jt}(1 - s_{jt}) & \text{if } j = k; \\ -\alpha_0 p_{kt} s_{kt} & \text{if } \jmath \neq k. \end{cases}$$

- As shares are often small, the own price elasticities will be proportional to price. This is unrealistic
- Furthermore, the cross price elasticities restrict proportionate increases to be identical for all goods

The Model: Distributional Assumptions

- When preferences are allowed to differ, the elasticities will be:

$$e_{jkt} = \begin{cases} -\frac{p_{jt}}{s_{jt}} \int \alpha_i Pr_{ijt}(1 - Pr_{ijt}) dF(D_i, v_i) & \text{if } j = k; \\ \frac{p_{kt}}{s_{jt}} \int \alpha_i Pr_{ijt} Pr_{ikt} dF(D_i, v_i) & \text{if } j \neq k. \end{cases}$$

- The price sensitivity is now a probability weighted average, and can differ over products. As such the model allows for flexible substitution patterns
- To continue, assume $v_i \sim iidn(0, I_{K+1})$, let $F(D_i)$ be the EDF, and denote $\delta_{jt} = x_{jt}^{'}\beta_0 - \alpha_0 + \xi_{jt}$ as the mean-utility. Then the integral in (4) can be approximated by simulation:

$$s_{jt} = \frac{1}{R} \sum_{r=1}^{R} \frac{\exp(\delta_{jt} + [p_{jt}, x_{jt}^{'}](\Pi D_r + L v_i))}{1 + \sum_k \exp(\delta_{jt} + [p_{kt}, x_{kt}^{'}](\Pi D_r + L v_i))} \qquad (6)$$

Estimation

- As prices p_{jt} may be correlated with error ξ_{jt}, the parameters β_0, α_0, L and Π are estimated by GMM.
- This is carried out in three steps excluding an initial step
  0. Draw R individuals $v_1, ..., v_R$, and $D_1, ..., D_R$ from v_i an D_i
  1. For a given value of Π, L, solve for the vector $\delta = [\delta_{11}, ..., \delta_{JT}]^{'}$ such that predicted shares using (4) equals observed shares
  2. Compute the sample-moment conditions $T^{-1} \sum_{t=1}^{T} Z_t \xi_t$ where Z_t is a $J \times I$ set of instruments, and form the GMM-objective function.
  3. Search for the values $\beta_0, \alpha_0, \Pi, L$ that minimize the objective function in step 4
- To simplify the notation, let x_{jt} contain all variables, assume L is diagonal and define $\theta = (\theta_1^{'}, \theta_2^{'})^{'}$, where:

$$\theta_1 = \begin{pmatrix} \alpha_0 \\ \beta_0 \end{pmatrix} \quad , \theta_2 = \begin{pmatrix} vec(\Pi^{'}) \\ diag(L) \end{pmatrix} \tag{7}$$

Estimation: Step 1

- For each market $t = 1, ..., T$, we need the $J \times 1$ vector δ_t such that:

$$s(\delta_t, \theta_2) = s_t \tag{8}$$

- where $s_t = [s_{1t}, ..., s_{Jt}]'$. This system of J equations can be solved using the contraction mapping suggested by BLP.
- For a given vector δ_t^n, this involves computing:

$$\delta_t^{n+1} = \delta_t^n + \log s_t - \log\left(s(\delta_t^n, \theta_2)\right) \tag{9}$$

- Iteration continues using (8) and (9) until $\|\delta_t^n - \delta_t^{n-1}\|$ is below a specified tolerance level.
- In the Stata command `blp`, iteration is over $w_t = \exp(\delta_t)$ and δ_t is recovered at convergence. This saves considerable time

Estimation: Step 2

- Let Z_t be a $J \times I$ matrix of instruments that satisfies $E[Z_t^{'}\xi_t] = 0$ and define the GMM-objective function as:

$$Q = \bar{h}^{'}(\theta) W_T \bar{h}(\theta) \tag{10}$$

- where $\bar{h} = T^{-1} Z^{'} \xi$ are the sample moments based on δ and W_T is a positive definite weighting matrix. If the errors are homoskedastic, a consistent estimator of W is:

$$\hat{W} = (T^{-1} \hat{\sigma}_{\xi}^{2} Z^{'} Z)^{-1} \tag{11}$$

- If instead the errors are assumed to be correlated over J and heteroskedastic over t, then:

$$\hat{W} = (T^{-1} \sum_{t=1}^{T} Z_t^{'} \hat{\xi}_t \hat{\xi}_t^{'} Z_t) \tag{12}$$

- Estimation using (12) is carried out from an initial estimate of $\hat{\theta}$ based on (11). This is often referred to as the two-step method

Estimation: Step 3

- The GMM-estimator $\hat{\theta}$ is the vector that minimizes (10), and is the solution to the following first order conditions:

$$\frac{\partial Q}{\partial \theta_1} = X^{'} ZWZ^{'} \xi = 0 \quad (13)$$

$$\frac{\partial Q}{\partial \theta_2} = D_{\theta_2} \delta^{'} ZWZ^{'} \xi = 0 \quad (14)$$

- To reduce search-time, θ_1 can be written as a function of θ_2

$$\hat{\theta}_1 = (X^{'} ZWZ^{'} X)^{-1} X^{'} ZWZ^{'} \delta(\theta_2) \quad (15)$$

- The search is now limited to θ_2, but to employ a Newton method, the analytical derivatives $D_{\theta_2} \delta_t^{'}$ are required.
- By the implicit function theorem applied to $s(\delta_t(\theta_2), \theta_2) = s_t$:

$$D_{\theta_2} \delta_t^{'} = -(D_{\delta_t} s_t)^{-1} D_{\theta_2} s_t \quad (16)$$

Estimation: Step 3

- The elements inside the matrices of (16) are:

$$D_{\theta_2}\delta_t' = -\begin{pmatrix} \frac{\partial s_{1t}}{\partial \delta_{1t}} & .. & \frac{\partial s_{1t}}{\partial \delta_{Jt}} \\ \vdots & \vdots & \vdots \\ \frac{\partial s_{Jt}}{\partial \delta_{1t}} & .. & \frac{\partial s_{Jt}}{\partial \delta_{Jt}} \end{pmatrix}^{-1} \begin{pmatrix} \frac{\partial s_{1t}}{\partial \sigma_1} & .. & \frac{\partial s_{1t}}{\partial \sigma_{K_1}}, & \frac{\partial s_{1t}}{\partial \pi_{11}} & .. & \frac{\partial s_{1t}}{\partial \pi_{K_1 d}} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial s_{Jt}}{\partial \sigma_1} & .. & \frac{\partial s_{Jt}}{\partial \sigma_{K_1}}, & \frac{\partial s_{Jt}}{\partial \pi_{11}} & .. & \frac{\partial s_{Jt}}{\partial \pi_{K_1 d}} \end{pmatrix}$$

- From equation (6), the derivatives are:

$$\frac{\partial s_{jt}}{\partial \delta_{jt}} = R^{-1}\sum_{r=1}^{R} Pr_{rjt}(1 - Pr_{rjt})$$

$$\frac{\partial s_{jt}}{\partial \delta_{mt}} = R^{-1}\sum_{r=1}^{R} Pr_{rjt} Pr_{rmt}$$

$$\frac{\partial s_{jt}}{\partial \sigma_{k}} = R^{-1}\sum_{r=1}^{R} Pr_{rjt} v_{rk}(x_{jtk} - \sum_{m=1}^{J} x_{mtk} s_{rmt})$$

$$\frac{\partial s_{jt}}{\partial \pi_{kd}} = R^{-1}\sum_{r=1}^{R} Pr_{rjt} D_{rd}(x_{jtk} - \sum_{m=1}^{J} x_{mtk} s_{rmt})$$

Stata Command: `blp`

`blp` *depvar* $\left[\textit{indepvars} \right]$ $\left[\textit{if} \right]$ $\left[\textit{in} \right]$ *(endogvars=instruments),*
$(stochastic_1 = varlist_1, ., stochastic_K = varlist_K)$
markets(string) draws(#) ,[vce(string),demofile(string),twostep]

1. *endogvars* - specify endogenous variables and instruments
2. *stochastic* - specify variable with random coefficient and demographic variables (if used)
3. *markets* - input the market variable
4. *draws* - specify the number of simulations
5. *vce* - robust if one-step else standard demotfile - specify the path and name of the demographics data (optional)
6. *twostep* - uses optimal weighting matrix

Syntax Example

```
blp s cons x1 x2, stochastic(x1=d1,x2,p) endog(p=p
x12 x22 expx1 p2 z1 z2) demofile(demodata)
markets(mkt) draws(100)
```

Random coefficients logit model estimates

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Mean Utility						
cons	6.5911	1.7905	3.68	0.000	3.081747	10.10048
x1	.21846	.75414	0.29	0.772	-1.259628	1.696563
x2	.95836	.061	15.47	0.000	.8369611	1.079762
p	-.85106	.04698	-18.11	0.000	-.943158	-.7589756
x1						
d1	.57208	.26406	2.17	0.030	.054534	1.089643
SD	.49639	.09599	5.17	0.000	.3082529	.6845402
x2						
SD	.94895	.26694	3.55	0.000	.4257588	1.472146
p						
SD	.26390	.36210	0.73	0.466	-.445815	.9736282

.

Monte Carlo Experiments: DGP

- To exmaine the properties of the estimator, data is generated from the following DGP where $J = 25$ and $T = 30$

$$
\begin{aligned}
u_{ijt} &= 10 + \beta_{1i} x_{1jt} + \beta_{2i} x_{2jt} + \alpha_i p_{jt} + \xi_{jt} + \epsilon_{ijt} \\
\alpha_i &\sim N(-1, 0.5) \\
\beta_{1i} &\sim N(1, 1) \\
\beta_{2i} &\sim N(1, 1) \\
\epsilon_{ijt} &\sim EV \\
\begin{pmatrix} x_{1jt} \\ x_{2jt} \end{pmatrix} &\sim \left[\begin{pmatrix} 10 \\ 10 \end{pmatrix}, \begin{matrix} 2 & 0.2 \\ 0.2 & 2 \end{matrix} \right] \\
p_{jt} &\sim N(10, 1) \\
x_{it} &\sim U(0, 1)
\end{aligned}
$$

Monte Carlo Experiments: Parameter Estimates

- The following table sets out the mean and standard deviation of the parameter estimates σ_{β_1}, σ_{β_2}, σ_{α} from 50 replications using 500 draws for each.

Monte Carlo Results

	X1	X2	Price
True Parameter	1	1	-0.5
Mean	1.046	1.027	-0.538
Standard deviation	0.162	0.146	0.121

Monte Carlo Experiments: Logit Elasticities

- The following table sets out the price elasticities from the logit model assuming homogeneous preferences

Logit Price Elasticities

Product	1	2	3	4	5
1	-7.94469	0.139519	0.124729	0.17224	0.049108
2	0.061907	-7.53451	0.124729	0.17224	0.049108
3	0.061907	0.139519	-7.67698	0.17224	0.049108
4	0.061907	0.139519	0.124729	-7.56353	0.049108
5	0.061907	0.139519	0.124729	0.17224	-7.84014

Monte Carlo Experiments: Random Parameter Logit Elasticities

- The following table set out the price elasticities from the random-parameters logit model.

Logit Price Elasticities

Product	1	2	3	4	5
1	-4.24647	0.018311	0.001294	0.009764	0.161759
2	0.338069	-1.25824	0.145928	0.108114	0.504705
3	0.167242	1.021352	-0.23844	0.089002	0.247704
4	0.749224	0.449351	0.052852	-0.06635	0.291488
5	1.733515	0.292954	0.020543	0.040708	-0.76139