A Random-Coefficients Discrete-Choice Normal Model of Demand

Margaret M. Cigno, Elena S. Patel and Edward S. Pearsall[1]

1. Introduction

In this paper we provide the technical details of the demand model and econometric method we have used to estimate price elasticities for U.S. postal services in a companion paper Cigno *et al* (2012), "Estimates of U.S. Postal Price Elasticities of Demand Derived from a Random-Coefficients Discrete-Choice Normal Model".

Conventional econometric approaches typically fail to yield useable estimates of price elasticities when they are applied to demand systems with many similar products such as the markets for automobiles, breakfast cereals and postal services. To correctly represent the effects of substitution possibilities among the products, each price should appear in every demand equation of the model. This makes the number of cross-price elasticities large except for models with very few products. Unfortunately, the prices of similar products tend to be highly correlated within a sample. In practice, the equations used to describe demand must be overly-restrictive with respect to the prices in order to avoid near multi-co-linearity when the model is fit. Often, each equation of a conventional model is specified with its own-price but without the entire set of cross prices. The econometrics then yields an incomplete and inconsistent set of parameter estimates including the price elasticities.

A possible solution to the estimation problem is to fit a Random-Coefficients Discrete-Choice Logit model. The past decade of empirical demand research in industrial organization has been dominated by the estimation of such models following the method of Berry, Levinsohn and Pakes (BLP 1995).[2] Their model describes individual behavior yet can be fit with only market-level price and share data in combination with observable product characteristics, population demographics and, when necessary, an effective set of instrumental variables. The model is attractive because demand elasticities derived from the model are capable of representing any demand behavior, yet are sufficiently restricted that the estimates usually conform well to *a priori* expectations of signs and magnitudes. Most recent applications of the BLP methodology have followed the "Practioner's Guide" and have used software developed by Nevo (2000a and 2000b).

However, the BLP/Nevo model is not without drawbacks that seriously limit its use. Although the model has been applied to small purchases (Nevo's example is breakfast cereals), the model is actually designed for "large ticket" items such as automobiles. The literature

---

[1] Margaret M. Cigno is the Director of the Office of Accountability and Compliance (OAC) of the U.S. Postal Regulatory Commission (PRC). Elena S. Patel is an economist and member of the OAC staff. Edward S. Pearsall is an economist and consultant to the PRC. He can be reached at *espearsall@verizon.net.* The views expressed in this paper are those of its authors and do not necessarily represent the opinions of the PRC.

[2] An excellent recent survey of modern econometric models of demand behavior may be found in Nevo (2011). This paper also contains a nearly complete bibliography of the major literature regarding the Random-Coefficients Discrete-Choice Logit model.

provides little justification for the extension to frequent and non-exclusive choices of small items such as a household's annual purchases of postal services. Furthermore, the computational demands of the BLP/Nevo estimation methodology are extreme. To fit a multi-product model to a modest sample can require weeks of computation time using a dedicated personal computer with no assurance of ultimate success.

In this paper we develop a model, similar in most respects to the BLP/Nevo model, which we call the Random-Coefficients Discrete-Choice Normal model. Unlike the BLP/Nevo model, our model describes discrete choices at the margin and is directly applicable to multiple purchases of small items. We replace the demographics variables of the BLP/Nevo model with their principal components. The random elements of the coefficients are then assumed to be independent and identically distributed (i.i.d.) from standardized Normal distributions. With these changes a simple single-variable numerical integration replaces a tedious and far less accurate simulation employed in the BLP/Nevo methodology. In applications involving 15 products and 40-year time series, our model has been successfully fit in about 12 hours using a personal computer and a purpose-written Lotus 123 worksheet.

This paper follows a road map previously traced by BLP and Nevo. We begin by presenting the conceptual indirect utility equation and its distribution assumptions. Next the model is transformed into an observationally equivalent model by replacing the demographics variables with their principal components. We then show that when the household-level disturbances are assumed to be independently drawn from standard Normal distributions, the transformed model may be solved for the market shares of the products by a single-variable numerical integration over the range [-∞,∞]. Solving the shares model yields the mean indirect utilities of the products in each time period (or market). The demand elasticities for the model may also be retrieved by numerical integration. Finally, we supply a summary description of the estimation strategy used to fit the model by least squares. Our strategy is a simplified version of the Generalized Method of Moments methodology proposed by BLP and Nevo. However, our simplifications to the model make it possible to rely on the Newton-Raphson method to make the calculations more efficiently.

2. Indirect Utility

Following BLP/Nevo, we define the indirect utility to a household from the purchase and consumption of an additional unit of one of several named products plus an "outside good". The outside good represents the possibility that the household's best option may be to not purchase an additional unit of any of the named products.[3] Specifically, the indirect utility to household i from consuming one more unit of product j during period t is

$$U_{ij} = \xi_j + \left(Y_i + C_i - P_j\right)\left(\alpha + \Pi_y D_i\right) + X_j(\beta + \Pi_x D_i) + Z_j \gamma_j + \varepsilon_{ij} \qquad \textbf{(1)}$$

Eqn (1) applies to every time period (or market) $t = 1, \ldots, T$ to every household $i = 1, \ldots, I$ and to every product $j = 0,1, \ldots, J$, including the outside good, $j = 0$. The time period index t has

[3] To estimate market shares for the outside good, it is usually necessary to make an assumption about the size of the markets. This can be done in many ways, all of which are somewhat arbitrary.

been dropped from the equation to simplify notation. We also simplify notation by assuming that the same households are present in every time period.

Eqn (1) is a linear utility function with random coefficients and two mean-centered disturbances. The two disturbances are ξ_j, a disturbance to the mean indirect utility for every household that occurs from the purchase of product j, and ε_{ij} , a disturbance to the individual household's indirect utility from the purchase of product j.

Two of the terms in Eqn (1) represent the gains and losses to household i when purchasing and consuming an additional unit of product j. The gain in utility occurs when the household consumes the product and is captured by the term

$$X_j(\beta + \Pi_x D_i)$$

X_j is defined as a vector of length x of the properties of product j , with $X_0 = 0$ for the outside good. $(\beta + \Pi_x D_i)$ is an x-vector of random coefficients where D_i is defined as a random vector of length d of the demographic characteristics of household i. D_i includes both observable and unobservable characteristics.[4] D_i is mean-centered over the population allowing us to interpret β as an x-vector of mean responses of indirect utility to changes in the properties of the products. Although D_i has a zero-mean across all time periods (or markets), it can have a non-zero mean, designated $\overline{D}$, in any single time period (or market). Finally, Π_x is an x-by-d matrix of coefficients.

A loss in utility arises because the act of purchasing a good requires expenditures equal to each product's price. This loss is captured by the term

$$(Y_i + C_i - P_j)(\alpha + \Pi_y D_i)$$

Y_i is household income, and C_i is the accumulated consumer surplus derived from the household's current level of consumption of the J produts.[5] C_i is the monetary valuation of the household's current purchases minus their cost, The purchase of an additional unit of product j reduces $Y_i + C_i$ by P_j , the price of product j with $P_0 = 0$ for the outside good. The monetary loss is converted to a utility loss when it is multiplied by the household's marginal utility of income, $(\alpha + \Pi_y D_i)$. This is a scalar random coefficient that combines a mean response to income changes, α, with a component that depends on the d-vector, D_i , of demographic characteristics. Π_y is a 1-by-d vector of coefficients.

Finally, the term $Z_j\gamma_j$ represents exogenous effects that may either increase or decrease indirect utility but do not operate through changes in either the properties of the products or the

[4] This differs somewhat from the BLP/Nevo model which treats observable and unobservable demographics characteristics separately.

[5] The BLP/Nevo model omits consumer surplus because their model is a discrete-choice model describing the indirect utility from the purchase and consumption of a single unit of each of the products. There is no accumulation of consumer surplus from the purchase and consumption of prior units.

demographic characteristics of the household.[6] Z_j is a vector of exogenous effects in the market for product j, with $Z_0 = 0$ for the outside good. γ_j is the associated vector of non-random coefficients.

The assumptions $P_0 = 0, X_0 = 0$, and $Z_0 = 0$ make it unnecessary to separately define an equation for the indirect utility of the outside good as is done by BLP/Nevo.

3. The Demographics Transformation

Eqn (1) differs only in non-essential ways from the indirect utility equation found in BLP/Nevo. As written, it presents all of the same estimation problems and would engage the same tedious estimation methodology. The fundamental source of difficulties is that the vector of demographics variables, D_i , is not a vector of i.i.d. random variables. We alter this property of the model by transforming D_i using principal components and substituting the components into Eqn (1).

To implement this, a suitable estimate of the population variance-covariance matrix of the demographics variables must be available. This covariance matrix is defined as

$$\Omega = \text{cov}(D_i) = \text{E}[(D_i - \bar{D})'(D_i - \bar{D})]$$

Ω is a real, symmetric, positive definite matrix with positive real roots and real characteristic vectors. The roots and characteristic vectors of Ω are defined by the characteristic equation

$$A'\Omega\, A = rI$$

A is a d-by-d matrix whose columns are the characteristic vectors of Ω . The characteristic vectors are orthogonal and normalized so that $A'A = I$. r is the d-vector of the corresponding characteristic roots.

We make the following substitutions into Eqn (1)

$$D_i = A(rI)^{1/2}(\bar{F} + \nu_i)$$

and

$$\bar{F} = (rI)^{-1/2}A'\, \bar{D}$$

$(\bar{F} + \nu_i)$ is the d-vector of principal components of the demographics vector D_i . It contains two terms: a mean vector, $\bar{F}$, and a random vector

$$\nu_i = (rI)^{-1/2}\, A'(D_i - \bar{D})$$

The elements of ν_i are disturbances that are i.i.d. standardized random variables.[7] When substituting for D_i in Eqn (1), we also transform the coefficient matrices Π_x and Π_y as follows:

---

[6] BLP and Nevo do not specifically include terms such as $Z_j\gamma_j$ in their discrete-choice model, but other researchers using their model have included them.

$$\Gamma_x = \Pi_x A\,(rI)^{1/2}$$

and

$$\Gamma_y = \Pi_y A(rI)^{1/2}$$

With these substitutions, the transformed Eqn (1) becomes

$$U_{ij} = \xi_j + \left(Y_i + C_i - P_j\right)\left(\alpha + \Gamma_y(\bar{F} + \nu_i)\right) + X_j\left(\beta + \Gamma_x(\bar{F} + \nu_i)\right) + Z_j\gamma_j + \varepsilon_{ij} \qquad \textbf{(2)}$$

Eqn (2) does not differ from Eqn (1) as a description of the indirect utility derived from the purchase and consumption by household i of the product j. The demographics information contained in the principal components vector, $(\bar{F} + \nu_i)$, is identical to the information contained in the demographics vector, D_i . The two equations differ from one another only by a linear transformation of variables and coefficients. However, Eqn (2) lends itself to a simple representation of its random elements.

The formulas for retrieving the structural coefficient matrices Π_x and Π_y from estimates of the coefficients of Eqn(2) are

$$\Pi_x = \Gamma_x(rI)^{-1/2}A'$$

and

$$\Pi_y = \Gamma_y(rI)^{-1/2}A'$$

We note that Π_x and Π_y remain identified even if Eqn (2) is fit using less than the full subset of principal components. For example, $(\bar{F} + \nu_i)$ might be truncated to include only the principal components corresponding to the largest characteristic roots. The formulas remain workable even when zeros replace the missing columns of Γ_x and Γ_y. Therefore, the collection of demographic variables in D_i can be enlarged without necessarily increasing the dimensions of the estimation problem.

To ease the technical descriptions of the following sections, we simplify Eqn(2) by collecting terms:

$$W_i = (Y_i + C_i)\left(\alpha + \Gamma_y(\bar{F} + \nu_i)\right)$$

and

---

[7] Proof:

$$E[\nu_i] = (rI)^{-1/2}\,A'E[D_i - \bar{D}] = 0$$

$$\begin{aligned}\mathrm{cov}(\nu_i) = E[\nu_i\nu_i'] &= (rI)^{-1/2}\,A'E[(D_i - \bar{D})(D_i - \bar{D})']\,A\,(rI)^{-1/2} \\ &= (rI)^{-1/2}A'\Omega A\,(rI)^{-1/2} \\ &= (rI)^{-1/2}\,rI\,(rI)^{-1/2} = I\end{aligned}$$

$$R_j = \xi_j - P_j\left(\alpha + \Gamma_y\bar{F}\right) + X_j(\beta + \Gamma_x\bar{F}) + Z_j\gamma_j$$

The variable W_i collects the terms that are common to the indirect utility of every product for a single household including the outside good. Because these terms do not contribute to the household's preferences for one product over another, they will eventually drop out of the model as it is developed. The variable R_j is the mean indirect utility of product j for all households. It includes the product-level disturbance, ξ_j, but does not contain any terms that are variable at the household level. Eqn (2) becomes

$$U_{ij} = W_i + R_j + \left(-P_j\Gamma_y + X_j\Gamma_x\right)\nu_i + \varepsilon_{ij} \qquad \textbf{(3)}$$

R_j is not directly observed but is calculated indirectly from the aggregate market shares of the products. It is this calculation, also done by BLP and Nevo, that our transformation of the indirect utility using principal components greatly simplifies. The remaining terms in Eqn (3) are linear combinations of the disturbances ν_i and ε_{ij} . They and their coefficients are entirely responsible for the different preferences exhibited by different households in their purchases.

4. The Shares Model

We assume that the elements of the vector ν_i and the disturbance ε_{ij} are each i.i.d. N(0,1). In contrast, BLP/Nevo assume that ε_{ij} is i.i.d. with an extreme value (Logit) distribution. Under their assumption only ε_{ij} can be eliminated by integration. Their method for calculating mean incremental utilities from market share data is to simulate Eqn (1) by taking random drawings of the demographics variables. An important advantage of our Normal distribution assumption is that it makes it possible to derive the aggregate market shares implied by Eqn (3) by performing a simple single-variable numerical integration over a predetermined range. A less important consequence of this assumption is to establish the standard deviation of the disturbance ε_{ij} as the unit of measurement for indirect utility. This unit of measurement is the same for all households and all products.

Given our normal assumption, the error term in Eqn (3), $\left(-P_j\Gamma_y + X_j\Gamma_x\right)\nu_i + \varepsilon_{ij}$, forms a linear combination of variables that are i.i.d. N(0,1). Because a linear combination of Normally distributed variables is also Normal, the expression $\left(-P_j\Gamma_y + X_j\Gamma_x\right)\nu_i + \varepsilon_{ij}$ is i.i.d N(0,λ_j^2). The standard deviation of $\left(-P_j\Gamma_y + X_j\Gamma_x\right)\nu_i + \varepsilon_{ij}$ is:

$$\lambda_j = +\sqrt{\left(-P_j\Gamma_y + X_j\Gamma_x\right)\left(-P_j\Gamma_y + X_j\Gamma_x\right)' + 1}$$

We may now rewrite Eqn(3) using a single i.i.d. N(0,1) disturbance e_{ij} and the standard deviation λ_j. With this notational simplification, Eqn (3) becomes

$$U_{ij} = W_i + R_j + \lambda_j e_{ij} \qquad \textbf{(4)}$$

We note that the formula for λ_j depends solely on the coefficients of the elements of ν_i and the implicit coefficient of 1 for the disturbance term, ε_{ij} . These coefficients do not relate specifically to household *i*. The standard deviation, λ_j, is the positive square root of the variance so $\lambda_j \geq 1 \; \forall \, j$.

The disturbances e_{ij} are bound to have distributions that approach standardized Normal distributions as a result of the arithmetic of the model even if the distribution of the elements of the demographics vector D_i is not multi-variate Normal. The arithmetic will tend to produce this result because the Normal distribution is the limit distribution for linear combinations of i.i.d. random variables from any distribution. The v_i are linear combinations of the elements of $(D_i - \overline{D})$ as a result of the demographics transformation; the disturbances e_{ij} are themselves linear combinations of the elements of ν_i and the disturbance ε_{ij}. This linear arithmetic will leave the disturbances e_{ij} with distributions that are approximately Normal except, possibly, for models that use only a very few demographics variables with distinctly non-Normal distributions.

The mechanics of preference under discrete-choice are straightforward: product k is preferred to product j if and only if $U_{ik} - U_{ij} \geq 0$.[8] Then, based on Eqn (4), we form the following:

$$\frac{U_{ik} - U_{ij}}{\lambda_j} = \frac{R_k - R_j + \lambda_k e_{ik}}{\lambda_j} - e_{ij}$$

The term W_i drops out because it appears identically in the indirect utility for every product, so the preferences of household i are unaffected by the terms collected in W_i . Moreover, a household's preferences are unaffected by a fixed change that affects all mean indirect utilities equally because the choice depends only on the difference $\left(R_k - R_j\right)$. Therefore, we may arbitrarily assign a value to the mean indirect utility of one of the products without loss of generality. Here, the natural choice is $R_0 = 0$.

For any given value of the disturbance, e_{ij} , we may find the conditional probability that product k is preferred to product j by evaluating the cumulative distribution function (c.d.f.) of the standard normal distribution, $F(z)$, at $z = \left(R_k - R_j + \lambda_k e_{ik}\right)/\lambda_j$. If $e_{ij} \leq z$ then $\left(U_{ik} - Uij/\lambda k \geq 0\right.$ and product k is preferred to product j. Formally, we have

$$\Pr\left(\frac{U_{ik} - U_{ij}}{\lambda_j} \geq 0 \mid e_{ik}\right) = F\left(\frac{R_k - R_j + \lambda_k e_{ik}}{\lambda_j}\right)$$

We then form the conditional probability that product k is preferred to all other products by multiplying together the individual conditional probabilities for all products except product *k*:

---

[8] Ties are ignored.

$$\Pr\left(\frac{U_{ik} - U_{ij}}{\lambda_j} \geq 0 \;\forall\, j \neq k \mid e_{ik}\right) = \prod_{j \neq k} F\left(\frac{R_k - R_j + \lambda_k e_{ik}}{\lambda_j}\right)$$

Finally, the market share, $S_k(\cdot)$, of product k is equal to the expected value of the probability that product k is preferred to all of the other products:

$$S_k(\cdot) = \int_{-\infty}^{\infty} \prod_{j \neq k} F\left(\frac{R_k - R_j + \lambda_k e_{ik}}{\lambda_j}\right) dF(e_{ik}) \tag{5}$$

Eqn (5) is a single-variable integral of a continuous real-valued function $F(z)$ over the predetermined range $[-\infty, \infty]$. There is no closed-form mathematical expression for $F(z)$, the Normal c.d.f., nevertheless, the integral in Eqn (5) can be evaluated numerically to any necessary degree of accuracy using elementary computational methods such as the trapezoid rule and readily available routines for calculating very accurate values of $F(z)$ and $dF(z)$.

5. Solving the Shares Model

Our shares model predicts the market shares of household purchases at the margin. However, these shares cannot differ noticeably from the market shares for aggregate purchases in a particular time period (or market) when there are a large number of households. A household's purchases of products in any time period are unordered, so any purchase of any product may be selected as the "last", or marginal purchase. Let us select at random a single purchase of each household during the time period and designate this selection as the last unit purchased. The collection of last purchases for all households will then constitute a random sample with average market shares that are closely distributed around means that are the same as the aggregate market shares. Therefore, the expected market shares of the marginal purchases of all households are virtually identical to the aggregate market shares for all purchases made during the period.

Recall that $R_0 = 0$. The shares model is solved for any single time period (or market) by solving for the vector of mean indirect utilities, $R = [R_1, \ldots, R_J]$, that reproduces the vector of observed aggregate market shares, $S^* = [S_1^*, \ldots, S_J^*]$, when we calculate the vector of market shares, $S(R) = [S_1(R), \ldots, S_J(R)]$, by evaluating the integral of Eqn (5) for each named product. The elements of the vectors $\boldsymbol{S}^*$ and $\boldsymbol{S}(\boldsymbol{R})$ corresponding to the outside good are omitted because these shares are determined by the identities: $S_0^* = 1 - \sum_{k=1}^{J} S_k^*$ and $S_0(R) = 1 - \sum_{k=1}^{J} S_k(R)$.

We solve the shares model numerically using an algorithm that employs the matrix of partial derivatives of the market shares, $\boldsymbol{S}(\boldsymbol{R})$, with respect to the elements of the vector of mean indirect utilities, $\boldsymbol{R}$. The elements of this matrix are found by evaluating the integrals formed by differentiating Eqn (5) with respect to the elements of $\boldsymbol{R}$ within the integral. This differentiation results in

$$\frac{\partial S_k(\cdot)}{\partial R_l} = \int_{-\infty}^{\infty} \prod_{j \neq k} -\frac{1}{\lambda_l} F\left(\frac{R_k - R_j + \lambda_k e_{ik}}{\lambda_j}\right) dF\left(\frac{R_k - R_l + \lambda_k e_{ik}}{\lambda_l}\right) dF(e_{ik}) \tag{6}$$

As before, Eqn (6) is a single-variable integration of a continuous real-valued function over the range $[-\infty, \infty]$. The matrix of partial derivatives may be evaluated with the same numerical integration method that is used to evaluate Eqn (5).

The J-by-J matrix of partial derivatives is symmetric and non-singular for $S(R) > 0$. We denote this matrix and its inverse as $\left[\frac{\partial S(R)}{\partial R}\right]$ and $\left[\frac{\partial S(R)}{\partial R}\right]^{-1}$ respectively. To initiate the algorithm, we need a vector $\boldsymbol{R^0}$ that will yield shares $S(R^0) > 0$ when Eqn (5) is evaluated for each of the named products. While the choice $R^0 = 0$ always works, this is a starting point from which the algorithm typically takes many iterations to reach a solution. Our estimation method for the coefficient vectors Γ_x and Γ_y requires solutions to the shares model for a convergent sequence of values. The number of iterations needed to reach these solutions can be considerably reduced by using a previously-computed mean vector R as a starting point.

At iteration n the algorithm calculates a new mean vector, R^{n+1} , from the previous mean vector R^n. Before we can make this calculation, we must numerically integrate Eqn (5) to compute the elements of $S(R^n)$. Moreover, we must numerically integrate Eqn (6) to compute the elements of $\left[\frac{\partial S(R^n)}{\partial R}\right]$ and invert the matrix to obtain $\left[\frac{\partial S(R^n)}{\partial R}\right]^{-1}$. R^{n+1} is computed from

$$R^{n+1} = R^n + \mu \left[\frac{\partial S(R^n)}{\partial R}\right]^{-1} \left(S^* - S(R^n)\right)$$

The scalar parameter, μ, is chosen in the range $[0,1]$. Values of μ close to zero increase the dynamic stability of the algorithm but slow its rate of convergence. Values too close to one can result in unstable oscillations in the calculated shares when R^n is not close to $\boldsymbol{R^0}$. The iterations are repeated until the calculated shares approximate the observed market shares, The termination rule that we have used in our applications is:

$$\sqrt{\sum_{k=0}^{J} \frac{\left(S_k^* - S_k(R)\right)^2}{J+1}} < \tau$$

where τ is a preset tolerance level. The left-hand side of this termination rule is the average distance of the calculated market shares from the observed market shares.

Our computational experience with the algorithm is limited at this time to our postal applications. This experience indicates that the algorithm can be made both robust and efficient by varying μ to converge on 1 as the average distance in the termination rule shrinks. We set $\mu = 0.3$ when the average distance is large and allow μ to converge on 1 as average distance approaches τ. When $R^0 = 0$ is used as the starting point, the algorithm typically requires only 8 to 10 iterations to solve the shares model. In the latter stages of our estimation methodology, the shares model is usually solved to an accuracy of $\tau = 1e^{-6}$ in just two iterations using a previous solution as the starting point.[9]

6. The Demand Elasticities

The BLP model was originally conceived to strike a delicate balance between traditional demand models, such as a set of translog demand equations, and models that fit logit equations to market shares data. Econometric fits of traditional demand models encounter difficulties because of the large number of highly correlated price terms in the equations. This difficulty can be overcome by fitting a straight-forward logit model to market shares; however, the estimated price and cross-price elasticities derived from a logit model are so severely restricted that they fail to represent an acceptable range of demand behavior (see BLP 1995 or Nevo 2000a). The BLP model is an effective compromise because the demand elasticities derived from the fitted model are capable of representing any demand behavior, yet they are sufficiently restricted so that the estimates usually conform well to *a priori* expectations of signs and magnitudes.

Demand elasticities derived from our variant of the BLP model are also a successful compromise. This is demonstrated, following BLP and Nevo, by exhibiting the formulas for the elasticities and observing that there is nothing about them that prevents the demand elasticities from taking any credible values.

The general formula for the demand elasticity for product k with respect to the price P_l of product l is

$$\frac{P_l}{S_k(\cdot)}\frac{\partial S_k(\cdot)}{\partial P_l} = \frac{P_l}{S_k(\cdot)}\left[\frac{\partial S_k(\cdot)}{\partial R_l}\frac{\partial R_l}{\partial P_l} + \frac{\partial S_k(\cdot)}{\partial \lambda_l}\frac{\partial \lambda_l}{\partial P_l}\right] \quad \textbf{(7)}$$

Moreover, from Eqn (4) we can derive

$$\frac{\partial R_l}{\partial P_l} = -\left(\alpha + \Gamma_y \bar{F}\right)$$

and

$$\frac{\partial \lambda_l}{\partial P_l} = \frac{\Gamma_y\left[P_l\Gamma_y' - X_l\Gamma_y'\right]}{\lambda_l}$$

The general formula for the demand elasticity for product k with respect to the element h of the vector of properties of product l, X_l , is

$$\frac{P_l}{S_k(\cdot)}\frac{\partial S_k(\cdot)}{\partial X_l^h} = \frac{P_l}{S_k(\cdot)}\left[\frac{\partial S_k(\cdot)}{\partial R_l}\frac{\partial R_l}{\partial X_l^h} + \frac{\partial S_k(\cdot)}{\partial \lambda_l}\frac{\partial \lambda_l}{\partial X_l^h}\right] \quad \textbf{(8)}$$

---

[9] BLP and Nevo also use an algorithm to solve a shares model that is similar to ours. However, their algorithm uses a simulation to estimate product shares while we perform a numerical integration. The BLP/Nevo algorithm does not rely on an estimate of the matrix of partial derivatives, $\left[\frac{\partial S(R)}{\partial R}\right]$. Instead, their algorithm exploits a less efficient contraction mapping of R^n into R^{n+1} discovered by BLP. Consequently, their algorithm can take hundreds of iterations to solve their shares model while ours typically requires fewer than ten. BLP/Nevo also use a different termination rule.

Again, from Eqn (4) we can derive the following

$$\frac{\partial R_l}{\partial X_l^h} = \beta^h + \Gamma_x^h \bar{F}$$

and

$$\frac{\partial \lambda_l}{\partial P_l} = \frac{-\Gamma_x^h [P_l \Gamma_x' - X_l^h \Gamma_x']}{\lambda_l}$$

Here, we define β^h and X_l^h to be elements of the vector β and the row of the matrix Γ_x corresponding to X_l^h.

The right-hand sides of Eqn (7) and Eqn (8) contain two components. The first component captures the effect on demand of a change in mean indirect utility, R_l . This effect is identical for all households. The second component reflects the effect of changes in the standard deviation of the disturbance from Eqn (4), λ_l . This term captures the effect of the variability in household utility on aggregate demand.

The demand elasticities are calculated by inserting values for the partial derivatives appearing in the right-hand sides of Eqn (7) and Eqn (8). In particular, the partial derivative $\left[\frac{\partial S_k(\cdot)}{\partial R_l}\right]$ is obtained in the numerical evaluation of Eqn (6). Likewise, we can derive the partial derivative $\left[\frac{\partial S_k(\cdot)}{\partial \lambda_l}\right]$:

$$\frac{\partial S_k(\cdot)}{\partial \lambda_l} = \int_{-\infty}^{\infty} \prod_{j \neq k} - \frac{(R_k - R_l + \lambda_k e_{ik})}{\lambda^2} F\left(\frac{R_k - R_j + \lambda_k e_{ik}}{\lambda_j}\right) dF\left(\frac{R_k - R_l + \lambda_k e_{ik}}{\lambda_l}\right) dF(e_{ik}) \qquad \textbf{(9)}$$

This partial derivative is also a single-variable integral of a continuous function over the range $[-\infty, \infty]$ and can be evaluated using the same methods used in calculating the market shares.

The product demands also have elasticities with respect to the elements of the mean demographic vector, $\bar{D}$. The general formula for the elasticity of demand for product k with respect to the element m of the mean vector of demographic characteristics, $\bar{D}$ is

$$\frac{\bar{D}^m}{S_k(\cdot)} \frac{\partial S_k(\cdot)}{\partial \bar{D}^m} = \frac{\bar{D}^m}{S_k(\cdot)} \left[\sum_{l=1}^{J} \frac{\partial S_k(\cdot)}{\partial R_l} \frac{\partial R_l}{\partial \bar{D}^m} \right] \qquad \textbf{(10)}$$

where[10]

$$\frac{\partial R_l}{\partial \bar{D}^m} = [-P_l \Gamma_x + X_l \Gamma_x](rI)^{-1/2} A'$$

[10] This can be seen by substituting for $\bar{F}$ in the equation for mean utility and collecting terms on $\bar{D}$:

$$R_l = \xi_j - P_l \alpha + X_l \beta + [-P\Gamma_x + X_l \Gamma_x](rI)^{-\frac{1}{2}} A' \bar{D}$$
$$= \xi_j - P_l \alpha + X_l \beta + [-P_l \Pi_x + X_l \Pi_x] \bar{D}$$

Unlike Eqn (7) and Eqn (8), Eqn (10) shows that the demographic changes only affect demand by changing the mean indirect utility of the named products, R_l .

Finally, we note that Eqns(7-10) all employ information that is specific to a time period or market. Therefore, the elasticities all depend upon values of the demographic variables, product characteristics, and external effects used to compute them from the formulas.

7. The Estimation Method

Our method of estimation follows BLP/Nevo, however our description here tracks our application of the method to postal services in Cigno *et al* (2012). We omit the use of instrumental variables and assume that the mean disturbances, ξ_j, are i.i.d. in order to simplify the presentation. The GMM estimator described by BLP and Nevo reduces to a two part application of ordinary and non-linear least squares.

The coefficients that express the demand behavior of the households are the same in every period (or market), for every household, and for every product. This allows the model to be fit as a single all-encompassing equation for the mean incremental utility of the named products:

$$R_j = \xi_j - P_j\left(\alpha + \Gamma_y\bar{F}\right) + X_j(\beta + \Gamma_x\bar{F}) + Z_j\gamma_j \;\; \forall\, j = 1, \dots, J \qquad \textbf{(11)}$$

The outside good is omitted because $R_0 = 0$ in every period. The only disturbance present in Eqn (11) is the mean disturbance, ξ_j . The disturbance is assumed to be i.i.d. across all periods and named products. This assumption can be made tenable for products with widely differing prices, such as postal services, by scaling units so that a single purchase of any product entails a roughly equivalent average expenditure.[11]

At first glance, Eqn (11) appears to satisfy the requirements for estimation by ordinary least squares. The equation is linear in its parameters; the disturbance term is additive and spherical; and, for many applications, the exogenous variables appearing on the right-hand side of Eqn (11) are predetermined and measured with negligible error.[12] In fact, the only complication that prevents us from fitting the equation with elementary methods is that the dependent variable, R_j, cannot be independently measured. Instead, it is calculated, as we have described, by solving a shares model that conspicuously depends upon the estimates of the coefficients Γ_x and Γ_y .

Our method of estimation, in brief, is to minimize the sample variance of the mean disturbance, ξ_j . Following Nevo, we redefine the coefficient vectors and observation matrices to describe the estimation methodology. The coefficients are assembled into two column vectors:

[11] Alternatively, the variance-covariance matrix of the disturbances can be estimated from a preliminary fit of the model as we describe here. Our methodology can then easily be extended using Generalized Least Squares. This extension should improve the efficiency of the estimation method.

[12] The latter condition is generally assumed to apply to U.S. postal rates. Between 1970 and 2006 the nominal rates were predetermined by a regulatory process; since 2006, they have been linked to the Consumer Price Index under a formula stipulated by Congress. All econometric demand studies of U.S. postal volumes to date have treated the postal rates as exogenous.

$$\Theta_1 = \begin{bmatrix} \alpha \\ \beta \\ \gamma_1 \\ \vdots \\ \gamma_J \end{bmatrix} \text{ and } \Theta_2 = \begin{bmatrix} \Gamma_{y1} \\ \Gamma_{x1} \\ \vdots \\ \Gamma_{yd} \\ \Gamma_{xd} \end{bmatrix}$$

where the d columns of Γ_x and Γ_y are stacked to create the vector Θ_2. This arrangement segregates the model's coefficients into a subset, Θ_1, that only occurs in Eqn (9) and a subset, Θ_2, that is also needed to solve the shares model. The observation matrices, X_1 and X_2 , are defined with each row holding observed values of the exogenous variables conforming to Θ_1 and Θ_2 . A single row of X_1 is the vector

$$X_1 = \left[-P_j, X_j, 0 , \dots , Z_j, \dots , 0\right]$$

for a single period/product pair. A single row of X_2 holds cross-products formed by multiplying together the elements of $\bar{F}$ and either the negative of the price P_j or an element of the vector X_j . In particular, for any single period/product pair we have

$$X_2 = \left[-P_j\overline{F_1}, X_j\overline{F_1}, \dots , -P_j\overline{F_d}, X_j\overline{F_d}\right]$$

The observation vector for R_j is denoted $R(\Theta_2)$ to indicate its dependence on the coefficients in Θ_2 . The vector of disturbances for the sample is ξ_j . With these definitions, the model and the sample are simply described by the matrix equation

$$R(\Theta_2) = \xi + X_1\Theta_1 + X_2\Theta_2 \qquad \textbf{(12)}$$

Our model is fit by finding estimates of $\widehat{\Theta}_1$ and $\widehat{\Theta}_2$ that minimize the sum of the squared residual disturbances:

$$\xi'\xi = [R(\Theta_2) - X_1\Theta_1 - X_2\Theta_2]'[R(\Theta_2) - X_1\Theta_1 - X_2\Theta_2]$$

This problem can be reduced to a minimization over just the vector Θ_2 by making the following substitution:

$$\widehat{\Theta}_1 = (X_1'X_1)^{-1}X_1'[R(\Theta_2) - X_2\Theta_2]$$

This is simply the least-squares estimate of Θ_1 for a fit of Eqn (12) in the form

$$R(\Theta_2) + X_2\Theta_2 = \xi + X_1\Theta_1$$

Substituting back into Eqn (12) and rearranging, we get

$$\xi = [I - X_1(X_1'X_1)^{-1}X_1'][R(\Theta_2) - X_2\Theta_2]$$

from which

$$\xi'\xi = [R(\Theta_2) - X_2\Theta_2]'[I - X_1(X_1'X_1)^{-1}X_1']'[I - X_1(X_1'X_1)^{-1}X_1'][R(\Theta_2) - X_2\Theta_2]$$

Because the matrix $[I - X_1(X_1'X_1)^{-1}X_1']$ is idempotent, the sum of squares further reduces to

$$\xi'\xi = [R(\Theta_2) - X_2\Theta_2]'[I - X_1(X_1'X_1)^{-1}X_1'][R(\Theta_2) - X_2\Theta_2] \qquad \textbf{(13)}$$

Eqn (13) is a function of only the one vector, Θ_2 . In addition, we can calculate the sum of squares using $\widehat{\Theta}_1$ without actually computing the matrix $[I - X_1(X_1'X_1)^{-1}X_1']$ because

$$\xi'\xi = [R(\Theta_2) - X_2\Theta_2]'\left[R(\Theta_2) - X_1\widehat{\Theta}_1 - X_2\Theta_2\right]$$

Eqn (13) is real-valued, continuous, and strictly convex so long as $R(\Theta_2) - X_2\Theta_2$ is not fixed with respect to any element of Θ_2 . Therefore, for any case of interest, we locate its minimum at the point where its gradient vanishes, *i.e.*, where $\nabla(\xi'\xi) = 0$.

There exist a number of numerical methods for solving the vector equation $\nabla(\xi'\xi) = 0$. One of the more effective is the Newton-Raphson algorithm.[13] We begin iteration n with an estimate, Θ_2^n , for which we have calculated the gradient ∇^n. The method also requires an estimate of the matrix of second-order partial derivatives of Eqn (13). This is the Jacobian matrix of $\nabla(\xi'\xi)$ evaluated for Θ_2^n. The Jacobian matrix is denoted J_∇^n. It is non-singular with inverse $[\,J_\nabla^n\,]^{-1}$. The Newton-Raphson method employs a linear approximation to $\nabla(\xi'\xi)$ in the region of Θ_2^n:

$$\nabla(\xi'\xi) \approx \nabla^n + J_\nabla^n[\Theta_2 - \Theta_2^n]$$

The basic idea of the method is to set $\nabla(\xi'\xi) = 0$ and solve the linear approximation for a new vector,

$$\Theta_2^{n+1} = \Theta_2^n - [J_\nabla^n]^{-1}\nabla^n$$

Then repeat the process until $\nabla^n \approx 0$. The algorithm requires a starting vector, Θ_2^0 and a termination rule for judging when $\nabla \approx 0$. For our applications to postal products we have used $\Theta_2^0 = 0$ and have terminated when every element of ∇^n is absolutely less than 0.005.

Differentiating Eqn (13) yields the formula for the gradient of $\xi'\xi$:

$$\nabla(\xi'\xi) = 2[\nabla R(\Theta_2) - X_2]'\,[I - X_1(X_1'X_1)^{-1}X_1'][R(\Theta_2) - X_2\Theta_2]$$

which can also be calculated as

$$\nabla(\xi'\xi) = 2[\nabla R(\Theta_2) - X_2]'\left[R(\Theta_2) - X_1\widehat{\Theta}_1 - X_2\Theta_2\right]$$

where $\nabla R(\Theta_2)$ is the gradient of $R(\Theta_2)$. The formula for the approximation to the Jacobian matrix is obtained by differentiating again with $\nabla R(\Theta_2)$ treated as fixed

---

[13] Although the Newton-Raphson method is effective, it is problematic for the BLP/Nevo model because it requires accurate estimates at each iteration of the gradient ∇^n and the Jacobian matrix J_∇^n . The simulation used to solve the shares model in the BLP/Nevo methodology is a slow and inaccurate way to estimate ∇^n and J_∇^n . With our revisions, the calculation of ∇^n and an approximation to J_∇^n are made easier and much more accurate.

$$J_{\nabla} \approx 2[\nabla R(\Theta_2) - X_2]'[\nabla R(\Theta_2) - X_2]$$

The gradient vector $\nabla R(\Theta_2)$ includes a sub-vector $\nabla R_t(\Theta_2)$ holding elements corresponding to the named products for each time period (or market). By the Implicit Function Theorem:

$$\nabla R_t(\Theta_2) \approx \left[\frac{\partial R_t}{\partial \lambda_t}\right]\left[\frac{\partial \lambda_t}{\partial \Theta_2}\right] = \left[\frac{\partial S_t(R_t)}{\partial R_t}\right]^{-1}\left[\frac{\partial S_t(R_t)}{\partial \lambda_t}\right]\left[\frac{\partial \lambda_t}{\partial \Theta_2}\right]$$

The matrix $\left[\frac{\partial S_t(R_t)}{\partial R_t}\right]$ is the same as the matrix that is calculated by numerically integrating Eqn (6) and then inverting to solve the shares model for period t. The elements of the matrix $\left[\frac{\partial S_t(R_t)}{\partial \lambda_t}\right]$ are computed by numerically integrating Eqn (9). The elements of $\left[\frac{\partial \lambda_t}{\partial \Theta_2}\right]$ are obtained by differentiating λ_t for the period t from Eqn(4). The vectors of the derivatives with respect to the elements of Γ_x and Γ_y are

$$\left[\frac{\partial \lambda_l}{\partial \Gamma_y}\right] = \frac{(-P_l^2\Gamma_y' + P_l\Gamma_x' X_l')}{\lambda_l} \text{ and } \left[\frac{\partial \lambda_l}{\partial \Gamma_x}\right] = \frac{P_l\Gamma_y X_l' - X_l\Gamma_x' X_l'}{\lambda_l}$$

It should be noted that reversing the signs of Γ_x and Γ_y leaves the same standard deviation λ_l. Therefore, the signs of the vectors of the derivatives of λ_l should be verified experimentally.

Estimates of the asymptotic variance-covariance matrices of $\widehat{\Theta}_1$ and $\widehat{\Theta}_2$ can be retrieved from the last iteration of the Newton-Raphson method. Let $\hat{\sigma}^2 = \frac{\hat{\xi}'\hat{\xi}}{N}$ be an estimate of the variance of ξ_j. In our applications, we have used $N = T \times J$ minus the combined number of coefficients in the vectors Θ_1 and Θ_2 minus the number of omitted observations. This replicates the least-squares formula for an unbiased estimate of σ^2 .

At termination, $\widehat{\Theta}_1$ is a result of a least-squares regression of $R(\widehat{\Theta}_2) - X_2\widehat{\Theta}_2$ on X_1. The vector of residuals from this fit is $\hat{\xi}$. The variance-covariance matrix of $\widehat{\Theta}_1$is

$$\text{cov}(\widehat{\Theta}_1) = \hat{\sigma}^2(X_1'X_1)^{-1}.$$

Similarly, $\widehat{\Theta}_2$ is the result of a least-squares fit. The Newton-Raphson equation $\Theta_2^{n+1} = \Theta_2^n - [J_{\nabla}^n]^{-1}\nabla^n$ is the estimator for a least-squares regression of $R(\Theta_2) - X_1\widehat{\Theta}_1 - X_2\Theta_2$ on $\nabla R(\widehat{\Theta}_2) - X_2$ to fit the difference vector $(\Theta_2^n - \Theta_2^{n+1})$. This vector approaches zero at termination so the vector of residuals from this fit also approaches $\hat{\xi}$. The variance-covariance matrix for $(\Theta_2^n - \Theta_2^{n+1})$ is the asymptotic variance-covariance matrix of $\widehat{\Theta}_2$, *i.e.*,

$$\text{cov}(\widehat{\Theta}_2) = 2\hat{\sigma}^2[J_v^n]^{-1} .$$

8. Conclusion

Conventional econometric methods have difficulty fitting demand models incorporating a large number of similar products with highly correlated prices. In order to make the econometrics work the equations in such models must be overly-restrictive with respect to the prices. When fit by conventional methods the estimates can seriously miss-represent price elasticities as well as other characteristics of demand. This is precisely what occurs with conventional econometric models of demand for postal services. In our companion paper (Cigno *et al* 2012) we show that the existing models, which mostly omit cross-price terms, commonly under-estimate the magnitudes of the own-price elasticities by a factor of two or more.

The Random-Coefficients Discrete-Choice Logit model is an attractive alternative but is designed for discrete purchases of large items. The BLP/Nevo methodology for fitting it is also computationally quite challenging. Our Random-Coefficients Discrete-Choice Normal model retains all of the conceptual advantages of the BLP/Nevo model, extends its scope to multiple purchases of small items, simplifies the computations of mean incremental utility and increases their accuracy. Our model can be fit with just a small fraction of the effort typically required by the BLP/Nevo methodology.

The major contribution of our approach is to provide a practical method for estimating price elasticities for demand systems involving many similar products such as U.S. postal services. The applications to postal services in our companion paper (Cigno *et al* 2012) demonstrate that our model and estimation method are effective.

References

Berry, S., J. Levinsohn and A. Pakes, 1995, "Automobile Prices in Market Equilibrium", Econometrica, 63, 841-890.

Cigno, Margaret M., Elena S. Patel and Edward S. Pearsall, 2012, "Estimates of U.S. Postal Price Elasticities of Demend Derived from a Random-Coefficients Discrete-Choice Normal Model", " in M. A. Crew and P. R. Kleindorfer (eds) (*title?*), Northampton, MA, Edward Elgar.

Nevo, A., 2000a, "A Practitioner's Guide to Estimation of Random-Coefficients Logit Models of Demand", Journal of Economics and Management Strategy, 8, 513-548.

Nevo, A., 2000b, "Appendix to 'A Practitioner's Guide to Estimation of Random-Coefficients Logit Models of Demand' – Estimation: the Nitty-Gritty", manuscript available at faculty.wcas.northwestern.edu/~ane686/supplements/Ras_guide_appendix.pdf.

Nevo, A., 2011, "Empirical Models of Consumer Behavior", Annual Review of Economics 2011, 3, 51-75.