A Pure Characteristic Model of Demand.

by

Steve Berry and Ariel Pakes,
Yale University and Harvard University.

For I.O. purposes, the promise of demand models in characteristic, rather than product space, were in the possiblity of:

1. Analyzing own and cross price elasticities in markets with a relatively large number of goods.

2. Analysis of demand for new, or repositionned old, goods, (a related problem is the analysis of consumer surplus generated by new or repositionned old goods).

Typical Specification.

Variants date back to McFadden (1978) in econometrics, and to Hotelling (1929) and Lancaster(1971) in theory. MicroBLP(2003)

$$U_{ij} = \delta_j + \Sigma_{kr} x_{jk} z_{ir} \beta^o_{rk} + \Sigma_{kl} x_{jk} \nu_{il} \beta^u_{kl} + \epsilon_{ij},$$

where

$$\delta_j \equiv \Sigma_k x_{jk} \lambda_k + \xi_j,$$

and some of the $x's$ may be correlated with the product specific unobservable ξ (usually at least price). In I.O. often only product level data available (BLP).

Conceptual Problem: It endows the utility function with an additive disturbance with the property that its density, conditional on the realizations of the additive components for the other products, is positive on the entire real line.

Counterintuitive implications (troubling for evaluating the demand for and the consumer surplus from new products).

- If we introduce a new product that is virtually identical to an old product the model will insure that the demand for the two products will exceed the old demand.

- As we increase the number of goods each consumer's utility grows without bound (regardless of the observed characteristics of those goods).

Also implies that the product space can never "fill up" and markups cannot go to zero as we increase the number of products.

Restatement of Problem. The dimension of the characteristic space, K, is a function of the number of products (contradicts the spirit of the theory literature which focuses on demand and product location in a *given* characteristic space).

Advantages of Model with "Tastes for Products".

Largely computational. It insures the needed probabilities are; (i) nonzero (at every θ), (ii) have smooth derivatives (requires regularity conditions), (iii) are defined by integrals with simple limits of integration, and (iv) when aggregated into shares can easily be inverted to obtain ξ (BLP).

Current Paper: Provide an estimation algorithm with similar computational properties for a model without the $\epsilon_{i,j}$.

Finite Dimensional Models

Demand is determined by the location or "address" of the products in the characteristics space, and an exogenous distribution of consumer preferences over this space. As the number of products increases, the product space fills up, with products becoming very good substitutes for one another.

General version: the "ideal point" models (Anderson, De Palma and Thisse,1993) where

$$u_{ij} = \|X_j - \nu_i\| - \alpha_i p_j,$$

where $\|\cdot\|$ is some distance metric, which after some rewriting is

$$u_{ij} = x_j \beta_i - \alpha_i p_j + \lambda_i \xi_j.$$

Two Assumptions.

- We only allow for one unobserved product characteristic (opposite "extreme" of standard model where there is a new unobserved characteristic for each product)

- ξ is a "vertical" characteristic in the sense that every individual would prefer more of it; i.e. we assume that $\lambda_i > 0$ for all i.

Normalizations. Utility functions of each individual only identified up to an affine transformation. Chose

- add $-u_{i,0}$ to the utility of each choice so that the utility of the outside option is zero

- divide each $u_{i,j}$ by λ_i (so that the coefficient of ξ is one for all consumers).

Then

$$u_{ij} = x_j\beta_i - \alpha_i p_j + \xi_j,$$

and

$$u_{i,0} = 0,$$

which is identical to the model in BLP *without* their additive component with full support.

Since the ξ is unobserved we still have a single multiplicative normalization. Chose the mean of $\alpha_i = 1$. This fixes the units of ξ (in terms of the mean of the marginal utilities of a dollar).

Estimation Issues.

Need

1. method of calculating the aggregate market share function for fixed θ.

2. arguments that

   (a) prove existence of a unique ξ conditional for any θ

   (b) computes ξ vector from the market shares

3. a limiting distribution for the parameter vector

The implications from (2b) and (3) imply that the computational properties of the pure-characteristic model are different then those of BLP.

Problem 1: Computing Market Shares.

Straightforward modification to the vertical model. In the vertical model

$$u_{ij} = \delta_j - \alpha_i p_j, \;\; \delta_j = x_j\beta - \alpha_i p_j + \xi_j,$$

and $u_{i,0} = 0$.

Order goods in terms of increasing price. Good j is purchased *iff*

$$\underline{\Delta}_j(\delta, p) \equiv \max_{k>j} \frac{\delta_k - \delta_j}{(p_k - p_j)} < \alpha_i$$

$$< \min_{k<j} \frac{\delta_j - \delta_k}{(p_j - p_k)} \equiv \overline{\Delta}_j(\delta, p)$$

So if the cdf of α is $F(\cdot)$, then the market share of product j is

$$s_j(x, p, \xi; \theta, F) \equiv$$

$$\left(F(\overline{\Delta}_j(x, p, \xi)) - F(\underline{\Delta}_j(x, p, \xi)\right) \times 1\left[\overline{\Delta}_j > \underline{\Delta}_j\right].$$

K Dimensions.

Since $u_{i,j} = x_j\beta_i - \alpha_i p_j + \xi_j$, conditional on β_i the model is once again a vertical model with cut-off points in the space of α_i, but now the

quality levels in those cut-offs depend on the β_i. Hence

$$s_j(x, p, \xi; \theta, F, G) \equiv$$

$$\int \left(F(\overline{\Delta}_j(\delta, p, X, \beta)|\beta) - F(\underline{\Delta}_j(\delta, p, X, \beta)|\beta) \right)$$

$$\times 1\left[\overline{\Delta}_j(x, p, \xi, \beta) > \underline{\Delta}_j(x, p, \xi, \beta) \right] dG(\beta),$$

where $F(\cdot|\beta)$ is the cdf of α given β and $G(\cdot)$ is the cdf of β.

Integral which is typically not analytic. So we use a simulation estimator to approximate it.

Problem 2a: Existence and Uniqueness

Prove that there is in fact a unique ξ associated with each (θ, s, F, G) and provide a way of calculating it.

As in BLP we assume that s^o is in the interior of the J-dimensional unit simplex. Let the model's prediction's be

$$s(\theta, \xi) \equiv s(x, p, \xi; \theta, F, G)$$

for any fixed F and G. Now consider the system of equations

$$s(\theta, \xi) = s^o,$$

Given the normalization $\xi_0 = 0$, our goal is to show that for fixed θ this system has exactly one solution, $\xi(\theta, s^o)$, and to provide a way of finding that solution.

Theorem. Suppose the discrete choice market share function has the following properties:

1. *Monotonicity*. s_j is weakly increasing and continuous in ξ_j and weakly decreasing in ξ_{-j}. Also $\forall \xi_{-j}$ there exists values of ξ_j that

set $s_j(\cdot)$ arbitrarily clost to zero, and values that set it arbitrarily close to one.

2. *Linearity of utility* in ξ. If the ξ for every good (including the outside good) is increased by an equal amount, then no market share changes.

3. *Substitutes with Some Other Good*. Whenever s is strictly between 0 and 1, every product must be a strict substitute with some other good.

Then, for any F, G, θ, and for any market share vector s that is strictly interior to the unit simplex, there is a unique ξ which solves the market share equations.

Comments. It is easy to verify that our conditions will be satisfed for any finite θ as long as

F (the distribution of α) has a density (w.r.t. Lebesgue measure) which is positive on the real line. Note that β need not have a density (and since we will be using simulation it generally will not).

Problem 2b: Computing ξ.

We use 3 different methods for computing ξ.

BLP Generalized. The simplest is a generalization of the technique used in BLP. Assume

$$u_{ij} = x_j\beta_i - \alpha_i p_j + \xi_j + \sigma_\epsilon \epsilon_{i,j},$$

$$u_{i,0} = \sigma_\epsilon \epsilon_{i,0}.$$

Note that if $\sigma_\epsilon = 0$ we are back to the pure characteristics model, while if $\sigma_\epsilon = 1$ we have the model in BLP.

If $\sigma_\epsilon > 0$, we can multiply through by $\mu_\epsilon \equiv 1/\sigma_\epsilon$ and

$$s_j = \int_{(\beta,\alpha)} g_j(\beta, \alpha; \theta, \mu_\epsilon) F(d\alpha) G(d\beta).$$

where

$$g_j(\cdot) \equiv \frac{exp\left[(x_j\beta_i - \alpha_i p_j + \xi_j)\mu_\epsilon\right]}{1 + \sum_{q=1}^{J} exp\left[(x_q\beta_i - \alpha_i p_q + \xi_q)\mu_\epsilon\right]}.$$

For fixed μ_ϵ we can compute the inverse to this model using the contraction mapping from BLP (which converges at a geometric rte). As $\mu_\epsilon \to \infty$ these $\xi(\cdot)$ converge to the $\xi(\cdot)$ of the pure characteristics model (uniformly in θ).

Problem. Monte Carlo results in paper. As μ_ϵ grows large we cannot compute the needed exponents, and for μ_ϵ which we can compute the fixed point is not close enough to being solved.

Alternative Contraction. Write the market share as a function of ξ as

$$s_j(\xi_j, \xi_{-j}, 0).$$

Now define the "element-by-element" inverse for product j, $r_j(s_j, \xi_{-j})$, as

$$s_j(r_j, \xi_{-j}, 0) = s_j$$

The vector of element-by-element inverses, say $r(s, \xi)$, when viewed as a function of ξ takes $R^J \to R^J$, i.e.

$$\xi = r(s, \xi^*),$$

is a contraction, but with a modulus of one.

Problem. Because the modulus is one an iterative procedure for calculating ξ is not guaranteed to converge, and in Monte Carlo excercises it often converged too slowly.

Fixed Point Homotopy. For each value of t, we consider the value of ξ that sets $h(\xi, t, \xi_0)$ to zero, where

$$h(\xi, t, \xi_0) = (1 - t) * (\xi - \xi_0) + t * (\xi - r(s, \xi)),$$

where t is a parameter that takes values between zero and one, and ξ_0 is an initial guess for ξ taken from applying BLP's method with a large μ_ϵ.

Call the value of ξ that sets $h(\cdot) = 0$, $\xi(t, \xi_0)$. For $t = 0$, $\xi(t, \xi_0) = \xi_0$. For $t = 1$, $\xi(t, \xi_0)$ is the fixed point that we are looking for. Homotopy: start at $t = 0$ and slowly move t toward one. If t is moved slowly enough, then the new solution "easy" to find (as by a Newton method starting at the prior solution).

In our case we re-write the homotopy equation $h(\xi(t, \xi_0), t, \xi_0) = 0$ as a functional equation, and compute iterations as

$$\xi^k(t, \xi_0) = (1 - t) * \xi_0 + t * r\left(s, \xi^{k-1}(t, \xi_0)\right).$$

This is a strong contraction for $t < 1$. When we converge at a fixed t we increase t and do it again. As $t \to 1$ this stops contracting. If iterations gain you very little when the fixed point is not yet solved to sufficient precision, switch to the homotopy with Newton-type steps to solve the system.

Actual Computations. Start with modification to BLP technique. Then switch to alternative contraction, then to fixed point homotopy.

Problem 3: Limit Theorems.

Let

$$P^{ns}(\cdot) = F \times G^{ns}$$

and assume that our estimate of s^o is obtained from the choices of a sample of consumers of size n and therefore is denoted s^n. We use

s^n and a simulation estimator to compute $\xi(\cdot)$ from the equations

$$\mathsf{s}_\mathsf{j}^\mathsf{n} = s_j(\xi, x, p, ; \theta, P^{ns}),$$

and then minimize a norm of

$$G_J(\theta, s^n, P^{ns}) = \frac{1}{J} \sum_{j=1}^{J} z_j \xi_j(\mathsf{s}^\mathsf{n}, x, p, \theta, P^{ns}).$$

The objective function, $\|G_J(\theta,\, s^n,\, P^{ns})\|$, has a distribution determined by three independent sources of randomness:

- from the draws on the vectors $\{\xi_j,\, x_j\}$,
- from the sampling distribution of s^n, and
- from the simulated distribution P^{ns}.

Analogously there are three dimensions in which our sample can grow: as n, as ns, and as J grow large. The limit theorems use rates of growth for each dimension that imply that all sources of randomness effect the limit distribution.

Heuristic argument.

$$\xi(\theta,\, s^n,\, P^{ns}) =$$

$$\xi(\theta,\, s^0,\, P^0) + \left\{\xi(\theta,\, s^n,\, P^{ns}) - \xi(\theta,\, s^0,\, P^{ns})\right\} +$$

$$\left\{\xi(\theta,\, s^0,\, P^{ns}) - \xi(\theta,\, s^0,\, P^0)\right\}.$$

Recall that the $\xi(\cdot)$ are implicitly defined to the solution to the following equation

$$\mathsf{s}_\mathsf{j}^\mathsf{n} = s_j(\xi, x, p, ; \theta, P^{ns}),$$

which can be rewritten as

$$\mathsf{s}_\mathsf{j}^\mathsf{0} + \varepsilon^n - \varepsilon^{ns}(\theta) = s_j(\xi, x, p, ; \theta, P^0),$$

where

$$\varepsilon^n = s^n - s^0,$$

is the sampling error, and

$$\varepsilon^{ns}(\theta) =$$

$$s[\xi(\theta,\, s^0,\, P^{ns}),\, \theta,\, P^{ns}] - s[\xi(\theta,\, s^0,\, P^{ns}),\, \theta,\, P^0],$$

is the simulation error.

Now let

$$H^{-1}(\xi, \theta, P) = \left\{ \frac{\partial s(\xi,\, \theta,\, P)}{\partial \xi'} \right\}^{-1}.$$

Note that $H^{-1}(\cdot)$ is a $J \times J$ matrix, so its dimension is increasing as we increase J (as does s).

Given our previous definitions and sufficient smoothness we have

$$\xi(\theta,\, s^n,\, P^{ns}) = \xi(\theta,\, s^0,\, P^0) +$$

$$H_o^{-1}(\theta,\, s^0,\, P^0)\,\{\varepsilon^n - \varepsilon^{ns}(\theta)\} + r(\theta,\, s^n,\, P^{ns}),$$

where $r(\theta,\, s^n,\, P^{ns})$ is the remainder term resulting from our application of the mean value theorem.

Consequently the limit theorems work from

$$G_J(\theta,\, s^n,\, P^{ns}) = G_J(\theta,\, s^0,\, P^0) +$$

$$\frac{1}{J} z' H_o^{-1}(\theta,\, s^0,\, P^0)\,\{\varepsilon^n - \varepsilon^{ns}(\theta)\} + \frac{1}{J} z' r(\theta,\, s^n,\, P^{ns}).$$

Consistency gotten by providing conditions under which

- the second and third terms in this equation converge to zero in probability uniformly in θ (insured by smoothness and rate conditions on n and ns),

- an estimator which minimized $\|G_J(\theta, s^0, P^0)\|$ over $\theta \in \Theta$ would lead to a consistent estimator of θ^0. Nonstandard because as J increases we change the behavior of each of the $j = 1, \ldots, J$ components of the sum. Provide intuitive conditions which bound $\|G_J(\theta, s^0, P^0)\|$ away from zero for $\|\theta - \theta_0\| \geq \kappa > 0$ (do not require convergence to anything at those θ).

Asymptotic Normality requires, in addition, local regularity conditions and a limiting distribution for

$$H_o^{-1}(\theta,\, s^0,\, P^0)\left\{\varepsilon^n - \varepsilon^{ns}(\theta)\right\}.$$

The rate needed for this limit distribution depends on how the elements of the $J \times J$ matrix $H_o^{-1}(\theta, s^0, P^0)$ grow, as J gets large. This differs for the two classes of models, as will become clear from two examples.

In the simple logit model, $u_{i,j} = \delta_j + \epsilon_{i,j}$ and the required inverse is

$$\xi_j(\theta) = (ln[s_j] - ln[s_0]) - x_j\beta - \alpha p_j.$$

So

$$\frac{\partial \xi}{\partial s_j} = \frac{1}{s_j}.$$

Now consider how randomness in our s^n samples effects the estimate of $\xi_j(\theta)$. If $s^n - s^o = \epsilon^n$ the first order impact of this randomness on the value of our objective function at any θ will be given by

$$H_o^{-1}(\theta, s^0) \times \epsilon_n = \frac{\partial \xi}{\partial s}|_{s=s^0} \times \epsilon^n,$$

which from above contains expressions like

$$\epsilon_j^n \frac{1}{s_j}.$$

As $J \to \infty$, $s_j \to 0$ and the impact of any given sampling error grows without bound. Thus to insure that sampling error stays bounded as $J \to \infty$ we will need strict requirements on how the variance in measurement error goes to zero, i.e. how n increases, as J increases.

BLP is similar but there are two sources or randomness whose impacts increase as J grows large, sampling error and simulation error. Consequently to obtain an asymptotically normal estimator of the parameter vector from this model both n and ns must grow at rate J^2.

Practical Implication. In data sets with large J one will have to use many simulation draws before one can expect to obtain an accurate

estimator whose distribution is approximated well by a normal with finite variance.

Vertical Model. Then $u_{ij} = \delta_j - \alpha_i p_j$, , $\delta_j = x_j\beta + \xi_j$ and $u_{i0} = 0$. Recall then that if $F(\cdot)$ is the distribution of α_i, the market share of good j, $j = 1, \ldots, J-1$ is

$$s_j = F(\Delta_j) - F(\Delta_{j+1}), \quad s_J = F(\Delta_J).$$

Now the derivative matrix needed to analyze the effect of the simulation and sampling errors is of the form

$$\frac{\partial s_j}{\partial \xi_p} = f(\Delta_j)\frac{\partial \Delta_j}{\partial \xi_p} - f(\Delta_{j+1})\frac{\partial \Delta_{j+1}}{\partial \xi_p},$$

$$= f(\Delta_j)\frac{1}{p_j - p_{j-1}} + f(\Delta_{j+1})\frac{1}{p_{j+1} - p_j}.$$

None of these elements tend to zero as $J \to \infty$ (if anything we expect these elements to grow large as J grows large). Consequently to obtain an asymptotically normal estimate of the parameter vector in the vertical model both n and ns need only grow (at most) at rate J (and maybe slower than that depending on the price process). That is we should not need as many simulation draws, to obtain reasonable parameter estimates from the vertical model.

Computational Comparison. Trade off

- The number of simulation draws needed to get accurate estimates of the inverse, and hence of the moment conditions, must grow at rate J^2 in the model with a taste for products, while it need grow at most at rate J in the pure characteristics model.

- Second the contraction mapping used to compute the inverse is expected to converge at a geometric rate for the model with tastes for products, but not for the pure characteristics model, and we have good reason to worry that the inversion for the pure-characteristic model will take more time.

Initial Monte Carlos.

Will deal with two issues

- Compare implications of different ways of computing pure characteristics model

- Compare impact of simulation error in BLP to that in the pure characteristics model.

Computing the Pure Characteristic Model.

Vertical model has analytic inverse (or almost so). So here we use

$$u_{ij} = \delta_j + \sigma_x \nu_{ix} x_j - (\alpha_i . * p_j)$$

where

$$ln(\alpha_i) = \sigma_p \nu_{ip}; \;\; \delta_j = \bar{\beta}_0 + \bar{\beta}_x x_j + \xi_j$$

The consumer-specific random terms (ν_{ix}, ν_{ip}) are distributed standard normal (so that α_i is log normal, with a normalized mean.) Base case: x variable is drawn as twice a uniform (0,1) draw that is specific to the firm plus 1/2 a uniform (0,1) draw that is common to all firms in a market. ξ is drawn as a uniform on (-0.5,0.5). Price, p, is set as $p_j = e^{\delta_j}/20$. Note that p is a function of δ and δ is a function of ξ, so that p and ξ are correlated in the simulated datasets.

Calculating δ

Table provides example markets of 5 products. The first column of each table gives the "true" answer, the second column gives the "exact" (non-approx.) inverse and the remaining columns give the random coefficients logit approximation at different values of μ. The first two panels of the table show the results without simulation error – each calculated inverse and share uses exactly the same 5000 draws used to create the orginal shares. The last panel uses a simulation of 250 draws to calculate shares.

- the "exact" inverse is able to nearly exactly reproduce the true δ's, and this conclusion is independent of whether we have simulation error (see panel C)

- the approximation with scale factor $\mu = 1$ are unable to even reproduce the order of the $\delta' s$, but take only just over 1% of the computer time the exact calculations took.

- we have to move to $\mu_\epsilon = 50$ before we reproduce the order of the $\delta' s$ but then the approximation does do quite well; at least for the first two panels. However

  - $exp[\mu_\epsilon] \approx 5.2 \times 10^{21}$ and the compute time for this approximation is now just one-half of what it would be under the exact calculation.

  - Once simulation error is introduced it has more of an impact on the $\mu_\epsilon = 50$ approximations then on the homotopy (as we should have expected).

Examples of Calculating δ Using Different Methods

	True	"Exact"	Approx* $\mu = 1$	Approx $\mu = 10$	Approx $\mu = 50$
A. Results With No Simulation Error.					
δ_1	2.9896	2.9896	3.0818	3.0150	3.0360
δ_2	3.2142	3.2142	2.0859	3.1869	3.2637
δ_3	3.5618	3.5618	3.1296	3.5968	3.6201
δ_4	4.0433	4.0433	4.1301	4.1010	4.1048
δ_5	4.1013	4.1013	2.6780	4.0552	4.1577
Relative Time:		1.0000	0.0112	0.0987	0.4456
B. An Example Which Uses 250 Simulation Draws.					
δ_1	2.7627	2.7703	3.4262	2.9119	2.8984
δ_2	3.4540	3.4773	3.5727	3.6775	3.7288
δ_3	3.5703	3.5943	3.0197	3.7750	3.7958
δ_4	3.7310	3.7592	-1.4011	3.4720	3.9165
δ_5	3.7644	3.7938	4.1594	3.9948	4.0194

* This is BLP.

- Tables 2,3, & 4 compare exact/approximate pure characteristic and BLP parameter estimates. Table 2 & 3 data consists of 20 markets with a small number of products (distributed uniformly between $[2, ..., 10]$). Table 4 has a large number of products (100) and a small number of markets (5). All tables use 500 simulation draws. Still need to do comparisons in terms of elasticities and welfare measures.

- Table 2. For this data design the RCL does quite well, though the μ_ϵ one needs to use depends on the variance in ξ.

- Table 3. Use of BLP with an estimated σ does much worse, at least in terms of coefficients.

- Table 4. The RCL application does noticeable worse, and BLP does quite badly.

Table 2
Monte Carlo Results: Base Specifications

	(1)	(2)	(3)	(4)
Method	RCL	HTopy	RCL	HTopy
Scale (μ):	30	∞	50	∞
ξ_j =	$U(-\frac{1}{2},\frac{1}{2})$	$U(-\frac{1}{2},\frac{1}{2})$	U(-1.5,1.5)	U(-1.5,1.5)
σ_x (= 1)	1.04	1.03	1.24	1.26
	(0.04)	(0.03)	(0.06)	(0.06)
σ_p (= 1)	1.00	0.98	1.02	1.02
	(0.01)	(0.01)	(0.03)	(0.03)
β_0 (= 2)	2.06	2.00	2.34	2.33
	(0.05)	(0.05)	(0.10)	(0.09)
β_x (= 1)	0.99	1.00	1.04	1.05
	(0.01)	(0.01)	(0.02)	(0.03)

The estimates are means across 100 simulated datasets. Estimated standard deviations of the mean estimates are given in parentheses. The homotopy estimates took on the order of 10 times as long to compute.

Table 3: Results from RC-Logit Estimates

$\xi_j =$	True	(1) $U(-\frac{1}{2}, \frac{1}{2})$	(2) $U(-1.5, 1.5)$
σ_x	1	1.14	1.64
		(0.04)	(0.08)
σ_p	1	1.03	1.09
		(0.01)	(0.03)
β_0	2	2.19	2.79
		(0.06)	(0.12)
β_x	1	1.00	1.03
		(0.01)	(0.03)
Scale, μ	∞	34.08	15.50
		(3.31)	(1.98)
μ (Median)	∞	17.81	4.67

Table 4: Example of Runs on a Dataset with a Large Number of Products

Parameter	Homotopy	RCL Approx	BLP
σ_x (= 1)	0.931	0.976	0.819
σ_p (= 1)	1.006	0.926	0.885
β_0 = (2)	2.265	1.548	2.694
β_x = (1)	0.957	0.985	-0.173
Scale, μ	∞	20*	0.858

*In the random coefficients logit approximation, the scale was initially set to 20. For some combinations of parameter values and markets, this caused numeric problems and the scale in those cases was halved until the numeric problems went away. In a few cases, a scale as low as 2.5 was necessary.

Second Issue: Number of Simulation Draws.

For monte carlo tractability use for BLP

$$u_{ij} = \delta_j + \theta_x \lambda_i x_{j2} + \epsilon_{ij},$$

where

$$\delta_j = x_j \beta + \xi_j,$$

$\beta = (-5, 1)$, λ is standard normal. R is the number of simulation draws, and the "observed" market shares are set to their expected value at the true parameter values. Computation of the inverse shares follows BLP, but we do not use a variance reduction (importance sampling) scheme of sort used in that paper.

Use for pure characteritic model the vertical model with variance in observed shares generated by small samples of consumers rather than from simulation error in the predicted shares.

$$u_{ij} = \delta - \theta_p \lambda_i p_j, \qquad (1)$$

where

$$\delta_j = x_j\beta + \xi_j$$

where the two components of x are a constant and a uniform drawn from $(0, 2)$, $\beta = (1.5, 1)$, ξ_j is uniform on $(-1, 1)$, $p = \delta^2$, λ_i is drawn from the unit exponential distribution, so that θ_p (set equal to one in the experiments) is the mean disutility of a price increase.

Table 5:

- Bias at low values of R relative to J. If $J = 100$ the bias does not seem to go away even when $R = 1000$.

- At low values of R (say $R = 50$) the variance does not seem to go down as J increases. The sampling variance should go down, but for a fixed number of simulation

draws we expect the impact of simulation error to be larger the larger is J. Hence it seems like the simulation variance dominates the model variance for low numbers of simulation draws.

Table 6.

- no apparent inconsistency in the estimates anywhere in the table (even when $J = 200$ and $n = 50$).

- for fixed n the variance decreases in J. However for small n the decrease is almost imperceptible, while with large n the variance declines at very close to the rate of $\sqrt{J}$, which is the rate we would expect if simulation had no impact on the estimates at all.

To Come.

- Comparisons of BLP to pure-characteristics when pure-characteristics is truth.

- Comparisons of BLP to pure-characteristics model when both are wrong – the truth is generated from a characteristics model with two unobserved characteristics with random coefficients.

Both sets of comparisons compare

- own and cross price elasticities.

- estimates of consumer surplus

- price indices.

Table 5:
Monte Carlo Estimates for the Random Coefficients Logit
True Value of the Parm is 1
1000 Monte Carlo Repetitions

# Simulation draws (R)	# of Products (J) 10	50	100
10	1.194 (0.982) [.031]	1.218 (0.512) [0.016]	*
50	1.025 (0.645) [0.020]	1.039 (0.311) [0.010]	1.241 (0.495) [0.016]
100	0.982 (0.674) [0.021]	1.013 (0.271) [0.009]	1.037 (0.209) [0.007]
500	0.998 (0.633) [0.002]	1.008 (0.255) [0.008]	1.015 (0.181) [0.006]
10J	0.982 (0.674) [0.014]	1.008 (0.255) [0.008]	1.018 (0.181) [0.006]
J^2	0.982 (0.674) [0.021]	1.008 (0.244) [.008]	1.018 (0.163) [.016]

Notes: Simulated Standard Errors (empirical standard deviations across the repititions) in $(\cdot)$ and Simulated Standard Error of the Estimated Mean in $[\cdot]$.

Table 6 (Vertical Model;
True Value of the Parm is 1)
1000 Monte Carlo Repetitions

# Consumer Draws (n)	# of Products (J)				
	10	25	50	100	200
50	1.023	1.022	1.011	0.997	1.013
	(0.494)	(0.373)	(0.349)	(0.321)	(0.302)
	[0.016]	[0.012]	[0.011]	[0.010]	[0.010]
100	1.005	1.010	1.005	1.002	1.009
	(0.426)	(0.303)	(0.257)	(0.244)	(0.217)
	[0.014]	[0.010]	[0.008]	[0.008]	[0.007]
500	0.993	0.998	1.001	1.005	1.007
	(0.371)	(0.223)	(0.176)	(0.142)	(0.123)
	[0.012]	[0.007]	[0.006]	[0.005]	[0.004]
1000	1.01	0.99	1.00	1.00	1.00
	(0.361)	(0.227)	(0.162)	(0.118)	(0.097)
	[0.011]	[0.007]	[0.006]	[0.004]	[0.003]
$10J$	1.018	1.014	1.008	0.998	0.996
	(0.440)	(0.253)	(0.175)	(0.120)	(0.085)
	[0.014]	[0.008]	[0.006]	[0.004]	[0.003]
J^2	0.998	0.998	1.000	1.002	1.000
	(0.423)	(0.227)	(0.153)	(0.105)	(0.074)
	[0.014]	[0.007]	[0.005]	[0.003]	[0.002]
∞	0.997	0.999	0.999	1.001	0.997
	(0.364	(0.214)	(0.141)	(0.101)	(0.072)
	[0.011]	[0.007]	[0.005]	[0.003]	[0.002]

Notes: Simulated Standard Errors (empirical standard deviations across the repititions) in $(\cdot)$ and Simulated Standard Error of the Estimated Mean in $[\cdot]$.

The Bootstrap vs. the Asymptotic Variance

Not much difference in variance estimates,

Model	Mean Parm	Monte Carlo Std. Dev.	Mean Asymp. Std. Dev.
R.C. Logit	1.010	0.2574	0.2201
Pure Vert	1.002	0.1720	0.1719

Nor is there much difference in asymptotic distribution. See pictures.