5-1-2002

Limit Theorems for Estimating the Parameters of Differentiated Product Demand Systems

Steven T. Berry

Oliver B. Linton

Ariel Pakes

LIMITED THEOREMS FOR ESTIMATING THE PARAMETERS OF DIFFERENTIATED PRODUCT DEMAND SYSTEMS

By

Steven Berry, Oliver B. Linton and Ariel Pakes

May 2002

COWLES FOUNDATION DISCUSSION PAPER NO. 1372

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

YALE UNIVERSITY

Box 208281

New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

Limit Theorems for Estimating the Parameters of Differentiated Product Demand Systems[1]

Steve Berry[2], Oliver B. Linton[3], and Ariel Pakes[4]
Yale University, London School of Economics, and Harvard University

May 7, 2002

[1]We would like to thank the National Science Foundation for financial support.

[2]Department of Economics, Yale University, 27 Hillhouse Avenue, New Haven, CT 06520–8281, USA. e-mail: steven.berry@econ.yale.edu

[3]Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom. e-mail: lintono@lse.ac.uk

[4]Department of Economics, Harvard University, Cambridge, MA. e-mail: ariel@ariel.fas.harvard.edu

Abstract

We provide an asymptotic distribution theory for a class of Generalized Method of Moments estimators that arise in the study of differentiated product markets when the number of observations is associated with the number of products within a given market. We allow for three sources of error: the sampling error in estimating market shares, the simulation error in approximating the shares predicted by the model, and the underlying model error. The limiting distribution of the parameter estimator is normal provided the size of the consumer sample and the number of simulation draws grow at a large enough rate relative to the number of products. We specialise our distribution theory to the Berry, Levinsohn, and Pakes (1995) random coefficient logit model and a pure characteristic model. The required rates differ for these two frequently used demand models. A small Monte Carlo study shows that the difference in asymptotic properties of the two models are reflected in the models' small sample properties. These differences impact directly on the computational burden of the two models.

1 Introduction

We are often interested in estimating parameters of demand functions from data on the quantity, price, characteristics (and perhaps the production inputs) of a set of products that interact in an imperfectly competitive market. In the simplest case, there is a national market with one observation per product, and the approximations used for the distribution of the estimators are obtained by taking the limit as the number of those products, say J, grows large.

This paper is concerned with issues that arise when we take limits in dimension J. The limiting arguments raise novel econometric issues when interactions between firms are important. Although many of these issues apply to a broader class of models, we will focus our attention on the assumptions made in the literature on the demand for differentiated products. In particular, we will provide consistency conditions and asymptotic distributions for estimators of the parameters of differentiated product discrete-choice demand systems.

Before proceeding to an overview of the paper, we comment on the appropriateness of taking limits in dimension J (or if J is thought of as endogenous, as market size grows large in a framework in which J grows in market size). Our argument here is entirely practical. Industrial organization often has to deal with markets in which both: J is quite large (large enough to think limiting approximations in dimension J are likely to be relevant), and the theory of imperfect competition is clearly relevant (partly because of spatial competition and multi-product firms). It is true that the estimates of parameters of differentiated product demand systems are often obtained from richer data sets than the single cross-section of product level data considered in detail here. For example, micro data which matches individuals to the products they choose, or regional and/or time series variance in the product level data are often also available. However, as discussed below, in most (though not all) of these cases J will still be one of the relevant limiting dimensions, and as a result arguments similar to those given here will have to be used to rationalize the limiting properties of the parameter estimates[1].

Background on the Model and Results

Discrete choice differentiated product demand systems posit that the utility of the consuming unit is a function of: parameters, θ, observed product characteristics, x, random consumer tastes and unobserved (by the econometrician) product characteristics, ξ. Some of the observed characteristics (*e.g.* price) may be correlated with ξ. The consuming unit either chooses one of the J products marketed or it chooses not to spend any money on the goods in this market (in which case we say the consumer chooses the "outside" alternative). Each unit makes the choice that maximizes its utility. The choices of different consumers differ because of their tastes, and the distribution of those tastes is denoted by P^0.

[1] Indeed we do not know of any empirical work on differentiated product demand systems which does not generate their objective function by forming averages over the products in a given market. As a result they all have to worry about the interactions between products that lie at the heart of our analysis.

Our estimate of the model's market shares, say $\sigma(\theta, x, \xi, P)$, are generated by simply adding up over the choices of consuming units with taste distribution P, where P is typically the empirical distribution of tastes from a random sample drawn from P^0. We observe the actual market shares, s. Up to sampling error, these are assumed to be the market shares generated by the model at the true (θ^0, P^0). The true value of the unobservables are implicitly defined by the system

$$\sigma(\theta^0, x, \xi, P^0) = s^0, \tag{1}$$

where θ^0 is the true value of the parameter vector and s^0 is the true value of the market shares (without sampling error).

The equation $\sigma(\theta, x, \xi, P) = s$ can be solved for ξ as a function of (θ, x, s, P). An identifying assumption on the conditional distribution of $\xi(\theta^0, x, s^0, P^0)$ is made and the θ vector is estimated by method of moments. For example, if we assume a zero covariance restriction between some exogenous vector of instruments, z, and the unobserved characteristics, our moment restriction would be

$$E[G_J(\theta)] \equiv E[\frac{1}{J}\sum_{j=1}^{J} z_j \xi_j(\theta, x, s, P)] = 0 \tag{2}$$

at $\theta = \theta_0$, and our estimate of θ would minimize a norm in $\frac{1}{J}\sum_{j=1}^{J} z_j \xi_j(\theta, x, s, P)$.

Several econometric issues arise in this context. First, unlike a traditional microeconomic cross-section, when we add new observations (products) to the market, the shares and prices of the existing products will change. Similar problems arise in other contexts, such as production function estimation, involving interacting firms. To our knowledge no analysis of the limiting properties of parameter estimates as the number of products grow large in an imperfectly competitive market is available, although those properties seem fundamental to empirical work in industrial organization.

In our context the interdependence of firms' decisions implies that away from the true value of θ the observations on $\xi_j(\theta, x, s^0, P^0)$ are not independent from one another. That is since both s_j and p_j are endogneously determined as a function of the characteristics of other products (as well as of own-product characteristics) there is conditional (on the instruments) dependence in the estimate of ξ when $\theta \neq \theta^0$. As a result, consistency proofs that require uniform convergence of objective functions, uniform over all possible values of θ, cannot be used (at least not without a specification for how prices and shares behave as the number of products grow large). Relatedly, efficient instruments are likely to be a function of the characteristics of *all* of the products, and this generates instruments that are not independent over j.

We show how to obtain a consistency proof based on a property of the limiting value of the objective that can frequently be evaluated *a priori*. Given consistency, we then require only local properties of the objective function to characterize the limit distribution of the parameter estimates. As a result we are able to use a "triangular array" argument for the limit distribution of the objective function at $\theta = \theta^0$, together with simpler local convergence results (smoothness assumptions will do), to prove asymptotic normality of the parameter estimates. Our approach to these problems should be broadly applicable to a wide range of models of equilibrium markets.

A further problem turns out to be quite important in estimating demand parameters when either [i] the function $\sigma(\cdot)$ is an integral estimated by Monte Carlo simulation via a number, R, of simulation draws or when [ii] the observed market shares, s, are based on a random sample of consumers of size n and hence are subject to sampling error. In these cases, the disturbances generated by the simulation and sampling processes also impact on the distribution of the estimators. As we shall show, the impact of those disturbances differ markedly depending on which of the available differentiated product demand models are used. The nature of competition in demand space therefore feeds back to the asymptotic limit theory. As a result, the rates of convergence of the estimators differ for different demand models. So our limit theorems are different for the different demand models, and, as illustrated by our Monte Carlo results, this implies that the computational burden induced by simulation is quite different for the different models.

We note that the limit theorems are always developed for rates that allow us to quantify the effects on the limiting distribution of the estimators of all three sources of randomness: the consumer sampling process, the simulation process and the process generating the unobserved product characteristics.

Two Classes of Models

Though we provide results for a quite general class of models, we are particularly concerned with two special cases. The first is the random coefficients logit (or probit) based estimator of demand discussed in Berry, Levinsohn and Pakes (1995; henceforth BLP). Under quite general conditions we show that in the logit and random coefficient logit cases the estimator will be consistent if $J \log J/n$ and $J \log J/R$ converge to zero as J increases. For asymptotic normality at rate $\sqrt{J}$ in these cases we require J^2/n and J^2/R to be bounded. That is, to obtain a consistent and asymptotically normal estimator for the parameters of these models we require the number of simulation draws and the size of the consumer sample to grow as the *square* of the growth in the number of products. So to obtain precise parameter estimates from these models we expect to need to use a relatively large number of simulation draws, especially when the number of products is large.

The second class of models we consider in detail is the "pure characteristic" model. Its theoretical lineage dates back at least to Hotelling's 1929 horizontal model, and it has seen extensive use in the context of the vertical model introduced by Shaked and Sutton (1982). It can be obtained from BLP's specification by simply deleting the independent and identically distributed "logit" errors from each choice alternative. Berry and Pakes (2002) endow the pure characteristics model with an estimation algorithm analogous to the estimation algorithm provided in BLP and discuss the advantages of the pure characteristics framework (focusing on the analysis of the demand for, and the welfare implications of, new goods).

We show that to estimate the parameters of the uni-dimensional (vertical) pure characteristic model consistently we require only that n and R increase at rate $\log J$, while for asymptotic normality we require only that J/n and J/R stay bounded. We also explain why the multidimensional pure characteristic model is likely to obey the same rate restrictions, but do not have a formal proof to that effect. Since the rate at which n and R must grow for asymptotically normal parameter estimates given the pure characteristics model is the square root of the rate at which they must grow to obtain asymptotically normal estimates for BLP's model, we expect to need much smaller numbers of simulation and sampling draws to obtain precise parameter estimates in the pure characteristics case.

The difference in results arises because differences in the nature of competition between the two models imply differences in the properties of the share function; i.e. of $\sigma(\cdot)$ in (1). Equation (1) must be solved for ξ in order to implement our method of moments estimation algorithm. In the models with "diffuse" substitution patterns, such as the random coefficient logit model of BLP, all goods are substitutes for all other goods and the elements of $\partial\sigma(\cdot)/\partial\xi$ go to zero as the number of products increase. It is the elements of the inverse of this partial that determine the impact of simulation and sampling error on the estimate of $\xi(\cdot)$ that satisfies (1). When the partial disappears this inverse grows large. So when J is large a little bit of simulation or sampling error in s causes large changes in the computed value of ξ which, in turn, causes larger variance in our estimate of the objective function (in our estimate of equation (2))[2].

In contrast the pure characteristic model has "local" competition (products are only substitutes with a finite number of other products). The more the number of products the "closer" will a product's nearest competitor tend to be, and the larger will be the market share response to small changes in the quality of the product (i.e. the larger will be $\partial\sigma/\partial\xi$). In the pure characteristic model then, a little bit of simulation or sampling error will have little effect on the computed value of ξ. This suggests that for fixed J we should be able to obtain "well behaved" parameter estimates from the pure characteristic model with fewer simulation draws than we need to use in estimating BLP's model. We provide a Monte Carlo study which indicates that the difference is rather dramatic.

Since the number of simulation draws needed to obtain precise estimates of the objective function is likely to be larger in BLP's model than in the pure characteristic model, the computational burden of simulation in BLP's model is expected to be larger than in the pure characteristics model. Berry and Pakes (2002) show, however, that the computational burden of obtaining the $\xi(\cdot)$ from the system in (1) is typically larger for the pure characteristics model than it is for BLP's model. So there is a trade off to be considered when comparing the computational burden of the two models between the ease of simulation in one model and the ease of computing ξ in the other.

[2]There is an analogy here to the impact of simulation error on the maximum likelihood estimators of discrete choice models when the choice probabilities are simulated. In that case the probabilities that determine the likelihood acts like our $\sigma(\cdot)$ function, and the impact of simulation error on the log-likelihood is larger when the underlying probabilities of a choice are small. If we let the number of choices (our J) grow large all but possibly a small number of probabilities would have to go to zero, so for consistency we would need the simulation error to converge to zero at a faster rate than the rate at which J grows. Unfortunately the analogy to maximum likelihood does not carry over to the pure characteristics model where $\sigma(\cdot)$ has notably different properties, see below.

Generalizations and Limitations.

There are two common ways in which actually estimated differentiated products demand systems differ from our setup. First, the same demand model can be applied to richer types of data. Second, one can add a supply side to the model. For example, one might assume a Nash pricing equilibrium and use the pricing equation together with the demand equation to estimate the demand parameters. While these extensions can greatly aid in obtaining precise parameter estimates, in most cases there is still an interest in how the estimates behave as J becomes large.

On the demand side, richer data could be either [i] observations on multiple markets across time and/or geography or else [ii] direct observations on consumers, matching observed attributes of the consumers to their choices. First consider adding more detailed consumer data within a single market. As explained in Berry, Levinsohn and Pakes (2001), the consumer data can allow one to obtain more precise estimates of parameters governing the interaction between consumer attributes and product characteristics. However, by itself the consumer data does not reveal the mean effects of the product characteristics on demand. That paper shows that in the single-market case with observed consumer choices and unobserved product characteristics, ξ, it is still necessary to take limits in J.

If one has data on multiple independent markets, then the situation is more complicated, and different sets of assumptions might be appropriate. If the same products, or a subset of the same products, appear in every market, as in Nevo's (2001) analysis of breakfast cereals, then the observations on the unobserved quality of the product are not independent across markets. In those cases we are back to requiring limits in dimension J (although as noted by Nevo somewhat different instrumental variables strategies might be available). A similar situation occurs when we have data on a given market over time, and the same, or related products, appear in different time periods (as in BLP's study of auto demand which had data on twenty years with about a hundred products per year).

If, on the other hand, there were a large number of markets with products whose unobservable characteristics are independent over markets, then one may be able to obtain CAN estimators by taking limits solely in the number of markets and not in J[3] . However, if there are a large number of products within each market, the implications from this paper for how simulation and sampling error behave as J grows large will still be useful. That is, our implications for the relationship between J and (n, R) will still be relevant. In other cases, the number of markets and the number of products may each be moderately large so we will require limits as both dimensions grow large but their ratios are bounded; a situation in which our results are also likely to be helpful.

Turning to the supply side, many studies have found that adding a pricing equation and then jointly estimating all parameters from the combined pricing and demand equations can markedly increase the precision of demand-parameter estimates. While the strategy has a cost in additional assumptions, the presence of the demand parameters in the pricing markup equation adds efficiency to the demand estimates. In this case, though, the need for

[3]For a contrasting case with different products in every market, consider broadcast radio stations in different cities.

asymptotics in J does not change and the framework we use here can easily be augmented to include a pricing equation.

Adding the pricing equation does add some clarity to questions about the optimal choice of instruments for our problem (as in Chamberlain, 1987). It makes clear that optimal instruments for price will depend on the characteristics of rival products, rendering semiparametric analysis of optimal instruments (as in Newey (1990) and (1993)) difficult if not impossible. We shall illustrate these problems in the context of our examples and provide some heuristic guidance for the choice of instruments; but we do not currently have a practical answer to the questions of optimal instruments.

Organization and Notation

The paper is organized as follows. In section 2 we present the underlying model. In section 3 we present an overview of the main results and the intuition underlying them. This includes two subsections which introduce our leading examples and explain the differences between them. Section 4 provides the main mathematical details of the arguments (formal proofs are relegated to an appendix). Section 5 explains how to determine rates of convergence given the results of section 4. Section 6 returns to our examples, verifies that they satisfy the conditions set out in section 4, and provides the formal argument leading to their rates of convergence. A small Monte Carlo study is presented in section 7.

We use $\|A\| = \{\mathrm{tr}(A'A)\}^{1/2}$ to denote the Euclidean norm of any $m \times n$ matrix A, $\xrightarrow{P}$ to denote convergence in probability, and $\Longrightarrow$ to mean convergence in distribution. For a matrix $A_{J\times J}$, we say $A = O(g(J))$ if the absolute value of the maximum element of the matrix is of order $g(J)$.

2 The Model and Estimator

We consider a market with J competing products and an outside good. The vectors of product characteristics will be denoted by $(\xi_j,\, x_{1j})$. The $\xi_j \in \mathbb{R}$ are characteristics which are not observed by the econometrician whereas the $x_{1j} \in \mathcal{X}_1 \subset \mathbb{R}^{d1}$ are observed. As noted in BLP (1995) they are analogous to the disturbance in the specification of traditional demand systems and are included to account for the fact that the list of product characteristics used in estimation does not contain all the product characteristics that consumers care about. Note also that without these disturbances the model could not rationalize the data. In large markets, where sampling error in the shares is essentially absent, the model predicts that the estimated shares should fit the observed shares exactly. This would typically be impossible if there were no disturbances.

We assume that the sequence $\{\xi_j\}_{j=1}^{J}$ are independent and identically distributed (i.i.d.) draws, and, for the most part maintain the assumption that

$$E[\xi_j|x_1] = 0 \mathrm{and} E[\xi_j^2|x_1] < \infty \tag{3}$$

with probability one, where $x_1 = (x_{11}, \ldots, x_{1J})$. The role and content of this assumption is discussed in Berry, Levinsohn and Pakes (1995). It can be replaced by other identifying assumptions without changing the logic of the underlying limit theorem.

In addition to the "exogenous" characteristics [those that satisfy $E(\xi_j|x_1) = 0$], we allow products to have additional characteristics, say $x_{2j} \in \mathcal{X}_2 \subset \mathbb{R}^{d2}$, which are "endogenous" (like price) in the sense of being related to the $\{\xi_j\}$. This produces a problem analogous to the traditional simultaneity problem in demand and supply estimation. We let $x_2 = (x_{21}, \ldots, x_{2J})$, $x = (x_1, x_2)$, and $\xi = (\xi_1, \ldots, \xi_J)$. At times we will also need explicit assumptions on the process generating x.

For any given vector of individual characteristics [households of given income, family size, etc.], say $\lambda \in \mathbb{R}^v$, the model determines a map from a parameter vector, $\theta \in \Theta$, where Θ is a compact subset of $\mathbb{R}^k$, and the vectors of product characteristics, (x, ξ), into the market shares purchased by individuals with those characteristics. Let that map be $\omega(x, \xi, \lambda, \theta) : D \to \mathcal{S}_J$, where D is the appropriate product space, and $\mathcal{S}_J$ is the $J+1$ dimensional unit simplex, i.e.,

$$\mathcal{S}_J = \{(s_0, \ldots, s_J)'| \quad 0 \leq s_j \leq 1 \text{ for } j = 0, \ldots, J, \text{ and } \sum_{j=0}^{J} s_j = 1\}.$$

If P is a distribution of λ, then the vector of aggregate market shares predicted by our model, for a given value of θ, and a particular P are

$$\sigma(\xi, \theta, P) = \int \omega(x, \xi, \lambda, \theta) dP(\lambda), \tag{4}$$

where we have suppressed the dependence of σ on x for convenience.

The actual market shares in the population are given by evaluating this function at $(\theta^0,\ P^0)$ the true value of θ and P. We designate this vector by $s^0 = \sigma(\xi,\ \theta^0,\ P^0)$. Note that though P^0 is assumed to be known, we typically will not be able to calculate $\sigma(\xi,\ \theta,\ P^0)$ analytically and will have to make do with a simulator of it, say $\sigma(\xi,\ \theta,\ P^R)$, where P^R is the empirical measure of some i.i.d. sample $\lambda_1, \ldots, \lambda_R$. For example,

$$\sigma(\xi, \theta, P^R) = \int \omega(x, \xi, \lambda, \theta) dP^R(\lambda) = \frac{1}{R} \sum_{r=1}^{R} \omega(x, \xi, \lambda_r, \theta).$$

Also though the vector s^0 is a random quantity determined by the realization of ξ ($s^0 = \sigma(\xi, \theta^0, P^0),$), we shall at times treat s^0 as if it were a non-random quantity [but all our results are proved with probability one over the distribution of s^0].

We will make the following regularity assumptions on $\sigma(\xi, \theta, P)$.

Assumption A1. (regularity conditions for share function) *For every finite J, for all $\theta \in \Theta$, and for all P in a neighborhood of P^0, $\partial\sigma_j(\xi, \theta, P)/\partial\xi_k$ exists, and is continuously differentiable in both ξ and θ, with $\partial\sigma_j(\xi, \theta, P)/\partial\xi_j > 0$, and for $k \neq j$, $\partial\sigma_j(\xi, \theta, P)/\partial\xi_k \leq 0$ (for k, $j = 1, \ldots, J$). Moreover, $s_j^0 > 0$ for all j.*

Note that although these properties must hold for each finite J, they need not hold in the limit. Thus although we assume that $s_\ell^0 > 0$ for all ℓ, we have $s_\ell^0 \to 0$ as $J \to \infty$ for all but possibly a finite subset of the products. Although we do not explicitly model the process which generates the products with positive market shares, below we require the process that generates the (ξ, x) tuples to satisfy certain regularity conditions.

The observed vector of market shares are denoted by $s^n \in \mathcal{S}_J$. Generally, s^n will be constructed from n i.i.d. draws from the population of consumers. Similarly, we assume that for any fixed (θ, ξ), say $(\theta_1,\, \xi_1)$, that the function $\sigma(\xi_1,\, \theta_1,\, P^R)$ is constructed from R independent, unbiased, simulation draws. This makes it natural to make A2.

Assumption A2. *The market shares* $s_\ell^n = \frac{1}{n}\sum_{i=1}^n 1(C_i = \ell)$, *where* C_i *is the choice of the* i^{th} *consumer, and* C_i *are i.i.d. across* i. *For any fixed* (ξ, θ), $\sigma_\ell(\xi, \theta, P^R) - \sigma_\ell(\xi, \theta, P^0) = \frac{1}{R}\sum_{r=1}^R \varepsilon_{\ell,r}(\theta, \xi)$, *where* $\varepsilon_{\ell,r}(\theta, \xi)$ *are independent across* r *and have mean zero, while the function* $\varepsilon_{\ell,r}(\theta, \xi)$ *is bounded, continuous, and differentiable in* θ. *Define the* $J \times J$ *matrices* $V_2 = nE[(s^n - s^0)(s^n - s^0)'] = diag[s^0] - s^0 s^{0\prime}$ *and* $V_3(\theta, \xi) = RE[(\sigma(\xi, \theta, P^R) - \sigma(\xi, \theta, P^0))(\sigma(\xi, \theta, P^R) - \sigma(\xi, \theta, P^0))']$.

Here $diag[x]$ is notation for a diagonal matrix with x on the principal diagonal. Also we can allow for more general simulators like those based on importance sampling advocated by BLP, by simply replacing the $V_3(\cdot)$ given in $A2$ with the appropriate importance sampling variance covariance matrix in the results that follow.

We now outline the logic of the estimation procedure. Elsewhere, [BLP (1995), and Berry and Pakes (2002)] we provide quite general conditions which insure that for every $(s, \theta, P) \in \mathcal{S}_J^o \times \Theta \times \mathbf{P}$, where $\mathcal{S}_J^o = \{s : 0 < s_\ell < 1 \text{ for all } \ell\}$ and $\mathbf{P}$ is a family of probability measures, there is a unique solution for the $\xi(\theta, s, P)$ that satisfies

$$s - \sigma(\xi, \theta, P) = 0. \tag{5}$$

By the implicit function theorem, Dieudonné (1969, Theorem 10.2.1), and A1, the mapping $\xi(\theta, s, P)$ is continuously differentiable in θ, s, P, in some neighborhood.. The true value of ξ, say $\xi^0 \equiv \xi(\theta^0,\, s^0,\, P^0)$, is obtained as the solution to

$$s^0 - \sigma(\xi,\, \theta^0,\, P^0) = 0. \tag{6}$$

Define the instrument matrix $z = (z_1, \ldots, z_J)$ whose components $z_q = z(x_{11}, \ldots, x_{1J})_q \in \mathbb{R}^\ell$, where $z(\cdot)_q : (\mathbb{R}^{d1})^J \to \mathbb{R}^\ell$, and $\ell \geq k$ (k is the dimension of θ), for $q = 1, \ldots, J$. Note that we allow the value of the instruments for the j^{th} observation to be a function of the values of the characteristics of all the observations. This is because most notions of equilibrium in use [e.g., Nash in prices or quantities] imply that the endogenous variables we are instrumenting [i.e., price] are functions of the characteristics all the products. We will require only weak regularity conditions on the z_q and will introduce them where needed below.

Now let

$$G_J(\theta, s, P) \equiv \frac{1}{J} \sum_{j=1}^{J} z_j \xi_j(\theta, s, P). \tag{7}$$

The assumption that $E(\xi_j | x_1) = 0$ ensures that $E[G_J(\theta^0,\, s^0,\, P^0)] = 0$. If we were able to calculate $\xi_j(\theta,\, s^0,\, P^0)$, then (2) would suggest using as our estimate of θ the method of moments estimator, Hansen (1982), obtained by minimizing the norm of $G_J(\theta, s^0, P^0)$. Unfortunately we observe only s^n and not s^0, and we cannot calculate $\sigma(\xi,\, \theta,\, P^0)$ but only $\sigma(\xi,\, \theta,\, P^R)$. Consequently, what we do is substitute an estimate of ξ, obtained as that value of ξ that sets $s^n - \sigma(\xi,\, \theta,\, P^R)$ to zero and denoted by $\xi(\theta,\, s^n,\, P^R)$, into (2) and minimize the resulting objective function. Thus our estimator of θ, say $\widehat{\theta}$, is defined as any random variable that satisfies

$$\|G_J(\widehat{\theta}, s^n, P^R)\| = \inf_{\theta \in \Theta} \|G_J(\theta, s^n, P^R)\| + o_p(1/\sqrt{J}). \tag{8}$$

The computation of $\widehat{\theta}$ is discussed further in BLP(1995).

3 Overview of the Main Results and Two Examples.

The objective function we are minimizing, $||G_J(\theta,\, s^n,\, P^R)||$, has a distribution determined by three independent sources of randomness: randomness generated from the draws on the vectors $\{\xi_j,\, x_{1j}\}$, randomness generated from the sampling distribution of s^n, and that generated from the simulated distribution P^R. Analogously there are three dimensions in which our sample can grow: as n, as R, and as J grow large. Our limit theorems will allow various rates of growth for each dimension. Throughout we let $J \to \infty$ and make n and R deterministic functions of J, i.e., we write $n(J)$ and $R(J)$ and let $n(J), R(J) \to \infty$ at some specified rate. If $n(J), R(J) \to \infty$ at a fast enough rate, then the contribution from simulation and sampling error will be of smaller order, and the asymptotics will be dominated by the randomness of ξ. We would like to guarantee that all three terms contribute to the asymptotics, so we make assumptions about the rate of growth of n, R to ensure this (this will allow us to evaluate the contribution of simulation and sampling error to the asymptotic distribution of the estimator). Finally, keep in mind that both s^n and $\sigma(\xi, \theta, P^R)$ take values in $\mathbb{R}^J$, where J is one of the dimensions that we let grow in our limiting arguments [although for expositional ease we have not indexed these functions by J in the statement of our assumptions, those assumptions should be interpreted as holding for each finite J].

We begin with a heuristic arguments which explains the steps we take to obtain our proofs. The fact that the dimension of the share function grows with J, makes the proofs required to validate the arguments in each of these steps quite detailed. As a result, we close this section with a sketch of the intuition underlying our results for our two leading examples. The formal arguments required to prove our results are delayed until the next section.

The consistency argument is established by showing that:

(i) $\sup_{\theta\in\Theta} ||G_J(\theta, s^n, P^R) - G_J(\theta, s^0, P^0)||$ converges to zero in probability.

(ii) an estimator that minimized $\|G_J(\theta, s^0, P^0)\|$ over $\theta \in \Theta$ would be consistent for θ^0.

(i) insures that neither simulation nor sampling error impacts on the consistency of our estimator. To establish it we assume that the instruments satisfy regularity conditions and then provide conditions which insure that $||\xi(\theta, s^n, P^R) - \xi(\theta, s^0, P^0)||^2/J$ converges to zero in probability uniformly in $\theta \in \Theta$. This latter point will require convergence of s^n to s^0 and P^R to P^0, and sufficient regularity of the mapping $(s, P) \mapsto \xi(\theta, s, P)$. We need the convergence to be uniform, either in absolute value or relative to the target quantity (convergence relative to the target quantity is relevant because the population quantity itself is shrinking with J). Smoothness conditions on the function $\xi(\cdot)$ allow us to convert closeness of (s^n, P^R) to (s^0, P^0) into closeness of $\xi(\theta, s^n, P^R)$ to $\xi(\theta, s^0, P^0)$. Note that (s^n, P^R) is a "function valued" nuisance parameter, similar to the nuisance parameters used in semiparametric estimation; see Newey (1994) etc.[4]

To establish (ii), we apply a version of Pakes and Pollard (1989, Theorem 3.1). This requires that: (a) $G_J(\theta^0, s^0, P^0)$ converges to zero, and (b) for all θ outside of a neighborhood θ^0, $G_J(\theta, s^0, P^0)$ stays bounded away from zero. Since at $\theta = \theta_0$, the $\xi_j(\theta^0, s^0, P^0)$ are indeed conditionally independent of one another (conditional on all the z_j), standard laws of large numbers can be used to insure (a). The problem in using standard uniform convergence arguments to guarantee (b) is that to verify them we would require a model for how the distribution of product characteristics (including price) evolves as the number of products grows large. What we do instead is provide an asymptotic identification condition which bounds the function $||E[G_J(\theta, s^0, P^0)]||$ uniformly away from zero when θ lies far enough away from θ^0. This condition, which suffices for (b), does not require that $G_J(\theta, s^0, P^0)$ converges at all, and puts only weak restrictions on how the characteristic distribution changes as J grows large. We provide the intuition underlying why we expect the identification condition to hold in the context of our examples presently.

We turn next to the asymptotic normality result. Write

$$\xi(\theta, s^n, P^R) = \xi(\theta, s^0, P^0) + \left\{\xi(\theta, s^n, P^R) - \xi(\theta, s^0, P^R)\right\} + \left\{\xi(\theta,\ s^0, P^R) - \xi(\theta,\ s^0, P^0)\right\}. \tag{9}$$

Next we express the last two terms in this expression in terms of the simulation and sampling errors and the parameters of the model. The simulation and sampling errors are defined by the $J \times 1$ vectors

$$\varepsilon^n = s^n - s^0 \text{and} \varepsilon^R(\theta) = \sigma^R(\theta) - \sigma(\theta).$$

By A2 both $\varepsilon^n$and$\varepsilon^R(\theta)$ is a sum of i.i.d. mean zero random vectors with known covariance.

From equation (5) and the definition of ε^n and $\varepsilon^R(\theta)$,

$$s^0 + \varepsilon^n - \varepsilon^R(\theta) = \sigma[\xi(\theta, s^n, P^R), \theta, P^0].$$

[4]It should be noted here that our problem does not fall into the usual case of semiparametric estimators because the entire vector $s^n = (s_1^n, \ldots, s_J^n)'$ affects each ξ_j. To offset this we have the fact that $E(s^n) = s^0$ for all J so that the 'nonparametric estimator' has zero bias unlike the usual semiparametric case.

We can therefore expand the inverse map from (s^n, θ, P) to $\xi(\cdot)$ around s^0. More formally by assumption A1, for each J, almost every P, almost all ξ, and every $\theta \in \Theta$, the function $\sigma(\xi,\ \theta,\ P)$ is differentiable in ξ, and its derivative has an inverse, say

$$H^{-1}(\xi, \theta, P) = \left\{ \frac{\partial \sigma(\xi, \theta, P)}{\partial \xi'} \right\}^{-1}. \tag{10}$$

Abbreviate $\sigma(\theta, s, P) = \sigma(\xi(s, \theta, P), \theta, P)$, $H(\theta, s, P) = H(\xi(s, \theta, P), \theta, P)$, and $H_0 = H(\theta^0, s^0, P^0)$, and further let $\sigma^R(\theta) = \sigma[\xi(\theta, s^0, P^0), \theta, P^R]$ and $\sigma(\theta) = \sigma[\xi(\theta, s^0, P^0), \theta, P^0]$.

Now two Taylor expansions give us the last two terms in equation (9) in terms of $H^{-1}(\cdot)$, ε^n and $\varepsilon^R(\theta)$. That is, $\xi(\theta, s^n, P^R) \simeq \xi(\theta, s^0, P^0) + H^{-1}(\theta, s^0, P^0)\{\varepsilon^n - \varepsilon^R(\theta)\}$, where the approximation sign indicates that we have omitted the second order terms from the Taylor's expansion.

Substituting our approximation for $\xi(\theta, s^n, P^R)$ into the objective function, we obtain our linear approximation to $G_J(\theta, s^n, P^R)$ as

$$\mathcal{G}_J(\theta) = G_J(\theta, s^0, P^0) + \frac{1}{J} z' H_0^{-1} \left\{ \varepsilon^n - \varepsilon^R(\theta^0) \right\}. \tag{11}$$

Next we provide conditions under which:

(a) $\sup_{||\theta - \theta^0|| \le \delta_J} \sqrt{J}[\mathcal{G}_J(\theta) - G_J(\theta, s^n, P^R)]$ converges to zero in probability for any sequence $\delta_J \to 0$.

(b) An estimator that minimized $\|\mathcal{G}_J(\theta)\|$ over $\theta \in \Theta$ would be: (i) asymptotically normal at rate $\sqrt{J}$; (ii) have a variance-covariance matrix which is the sum three mutually independent terms (one resulting from randomness in the draws on product characteristics, one from sampling error, and one from simulation error).

A consequence of (a) is that the estimator obtained from minimization of the criterion function $\|\mathcal{G}_J(\theta)\|$, has the same limit distribution as our estimator (i.e., as $\widehat{\theta}$ as defined in equation (8), and since the former is easier to analyze, we work with it. The general principles behind showing (a) are well understood: it requires a stochastic equicontinuity condition in the relevant stochastic process and some pointwise convergence. The difficulty here is in applying the conditions to specific models.

To establish (b) we provide a slight generalization to Theorem 3.3 in Pakes and Pollard (1989). The generalization allows for the fact that the underlying distributions of the random variables we are taking averages of may depend on J. The proof of (b) also requires a smoothness condition on the non-random function $E[G_J(\theta, s^0, P^0)]$ at $\theta = \theta^0$, and a further stochastic equicontinuity condition on the stochastic process $G_J(\theta, s^0, P^0)$ similar to condition (iii) of Theorem 3.3 of Pakes and Pollard (1989).

That proof shows that the random vector $\sqrt{J}\mathcal{G}_J(\theta^0)$ is the sum of three terms:

$$T_{J1} = \frac{1}{\sqrt{J}} \sum_{j=1}^{J} z_j \xi_j \quad ; \quad T_{J2} = \frac{1}{\sqrt{J}} z' H_0^{-1} \varepsilon^n \quad ; \quad T_{J3} = \frac{1}{\sqrt{J}} z' H_0^{-1} \varepsilon^R(\theta^0). \tag{12}$$

These random variables are each asymptotically normal at rates determined by the growth of $n(J)$ and $R(J)$; they are also mutually independent so that $\mathrm{var}[\sqrt{J}\mathcal{G}_J(\theta^0)] = \mathrm{var}[T_{J1}] + \mathrm{var}[T_{J2}] + \mathrm{var}[T_{J3}]$. We develop the limit theory so that all three terms are of the same magnitude, i.e., so that the effects of share estimation and simulation are captured by our approximations.[5] Finally, applying the arguments of Pakes and Pollard (1989) we then obtain the asymptotic distribution of $\sqrt{J}(\widehat{\theta} - \theta^0)$ in terms of $\partial E[G_J(\theta^0, s^0, P^0)]/\partial\theta$ and $\mathrm{var}[\sqrt{J}\mathcal{G}_J(\theta^0)]$.

3.1 Two Examples

The main purpose of this paper is to obtain consistency and asymptotic normality results for the empirical analogues of two frequently used models of demand: i) the logit model and its extension to the random coefficients logit as discussed in BLP(1995), and ii) the "pure characteristics model" which first appeared as the horizontal model of Hotelling (1929) [see also Shaked and Sutton's (1982) vertical model], and has recently been endowed with an estimation algorithm by Berry and Pakes (2002).

The next section provides a formal consistency and asymptotic normality argument for a broader class of models which includes the models of interest as special cases. In a subsequent section we work out the implications of these theorems for our two special cases. Before proceeding to the formal sections we provide the intuition underlying the results for the simple logit and pure characteristics models. The discussion here ignores both the conditions required for uniform convergence and the second order terms in the Taylor expansion which produces $\mathcal{G}_J(\cdot)$ from $G_J(\cdot)$.

3.1.1 The Simple Logit

The utility the i^{th} individual derives from consuming product j is

$$u_{ij} = x_j\theta + \xi_j + \epsilon_{ij} \equiv \delta_j + \epsilon_{ij}, \tag{13}$$

where x_j is a vector of observed product characteristics which typically includes price, ξ_j is an unobserved characteristic, and ϵ_{ij} is an i.i.d. (over both products and individuals) extreme value error term. Since we can add an individual specific constant to all utilities without changing the distribution of choices, there is a free normalization in this model. This is customarily resolved by setting the utility of the outside good $u_{i0} = \epsilon_{i0}$.

Individual i chooses the product which maximizes its utility. The market share function is obtained by solving for that maximum and then integrating out over the distribution of ϵ to obtain

$$\sigma_j(x, \xi, \theta) = \frac{e^{x_j\theta+\xi_j}}{1 + \sum_{k=1}^J e^{x_k\theta+\xi_k}}, \quad j = 1, \ldots, J, \tag{14}$$

[5] A simpler asymptotic scheme would be to specify the rates $n(J)$ and $R(J)$ such that $T_{J2}, T_{J3} = o_p(1)$, but this approach provides less information and is likely to understate the asymptotic variance of the estimator.

while $\sigma_0(x,\xi,\theta) = (1+\sum_{k=1}^J e^{x_k\theta+\xi_k})^{-1}$. Note that this is one of the few models which has an analytic form for the market share function. As a result there is no need for simulation and no simulation error in this model (i.e., $\epsilon^R(\theta) \equiv 0$).

The model predicts that market shares are determined by the random variables $x_j\theta + \xi_j$. For now assume this family of random variables has bounded support [because say x_j, ξ_j, and θ have bounded support] and density bounded away from zero on this support. Note that this implies that (with probability one); (a) market shares are all of magnitude $O(1/J)$, and (b) that for all finite J all products have market shares which are strictly positive.

From (14) the model also has an analytic expression for the unobserved product characteristic

$$\xi_j(\theta, s, P^0) = \ln(s_j) - \ln(s_0) - x_j\theta. \tag{15}$$

So our estimator is found by minimizing a norm of

$$G_J(\theta, s^n, P^0) = J^{-1}\sum_{j=1}^J z_j\xi_j(\theta, s^n, P^0) = J^{-1}\sum_{j=1}^J z_j[\ln(s_j^n) - \ln(s_0^n) - x_j\theta],$$

and can be interpreted as a linear instrumental variable estimator.

Assume temporarily that $\sup_{\theta\in\Theta} ||G_J(\theta, s^n, P^0) - G_J(\theta, s^0, P^0)||$ converges to zero in probability. Then all we require for consistency is that for all θ outside of a neighborhood θ^0, $G_J(\theta, s^0, P^0)$ stays bounded away from zero. But

$$\|G_J(\theta, s^0, P^0) - G_J(\theta^0, s^0, P^0)\| = \|J^{-1}\sum_{j=1}^J z_j' x_j(\theta - \theta^0)\|,$$

where z_j is a vector of instruments of dimension at least as large as that of x_j. Thus a sufficient condition for identification is that for J sufficiently large $J^{-1}\sum_{j=1}^J z_j' x_j$ is of full column rank with probability arbitrarily close to one.

Typically z_j will consist of the $x_{1,j}$, or the exogenous product characteristics, and instruments for price (which is allowed to be correlated with the ξ_j). So our identification condition requires the price of the product to be a function of observables which are not collinear with that product's exogenous characteristics. To formally verify whether this condition we would have to specify the nature of the pricing equilibrium. However all assumptions used to approximate equilibria in differentiated product markets imply that a product's price is a function both of its own and its competing product's factor prices, and of the characteristics of competing products (these determine markups). Since none of these variables are likely to be collinear with price, and (at least) the characteristics of competitors products are observable, the identification assumption seems unobjectionable. Indeed the interesting question is not whether our identification condition is satisfied, but which instrument will lead to an efficient estimator. This is a question that does depend on the precise nature of the pricing equilibrium (as well as the structure of ownership of the products), as we illustrate in our discussion of the vertical example below.

We now move on to the asymptotic normality result. In the logit case it is easy to derive the elements of the inverse share matrix analytically. We have

$$\frac{\partial \sigma_j(x,\xi,\theta)}{\partial \xi_k} = \begin{cases} \sigma_j(x,\xi,\theta)(1-\sigma_j(x,\xi,\theta)) & k=j \\ \\ -\sigma_k(x,\xi,\theta)\sigma_j(x,\xi,\theta) & \text{if} k \neq j, \end{cases} \tag{16}$$

Let $H(s,\theta)$ denote the $J \times J$ share matrix derivative evaluated at $\xi = \xi(s,\theta)$, i.e., $H_{jk}(s,\theta) = \partial \sigma_j(x,\xi(s,\theta),\theta)/\partial \xi_k$. Then if $S = \text{diag}(s)$ and $i = (1,\ldots,1)'$ it can easily be verified that

$$H(s,\theta) = S - ss' \quad \text{while} \quad H(s,\theta)^{-1} = S^{-1} + ii'/s_0.$$

Substituting into equation (11), the contribution of sampling error to $\sqrt{J} \times \mathcal{G}_J(\theta^0)$, or $T_{J,2}(\cdot)$ in (12), is $\sqrt{J}^{-1} z' H^{-1}(s^0)\varepsilon^n$. The argument of the last section then implies that to obtain a limiting distribution of the estimator all we require is a rate of growth for $n(J)$ that produces a finite variance for $T_{J,2}(\cdot)$. For simplicity let z contain a single variable, and recall that our conditions imply that, $\underline{c}/J \leq s_j^0 \leq \overline{c}/J$, for $j = 0,1,\ldots$. Then since A2 insures that $\text{var}(\varepsilon^n) = H(s^0)$, if we let $\overline{z}_J$ be the sample average of z, we have

$$\text{var}(\frac{1}{\sqrt{J}} z' H^{-1}(s^0)\varepsilon^n) = \frac{1}{nJ} z' H^{-1}(s^0) H(s^0) H(s^0)^{-1} z = \frac{1}{nJ}\sum_{j=1}^{J}[\frac{z_j^2}{s_j^0} + \frac{(\sum_{j=1}^{J} z_j)^2}{s_0^0}] \leq \frac{1}{\underline{c}n}[\sum_{j=1}^{J} z_j^2 + J^2 \overline{z}_J].$$

Assuming then that $\overline{z}_J$ is bounded, asymptotic normality requires $n(J)$ to grow like J^2. The intuition for this result is as follows. Since the shares must always sum to one, we have

$$\left| \sum_{k \neq j} \frac{\partial \sigma_k}{\partial \xi_j} \right| = \frac{\partial \sigma_j}{\partial \xi_j} < \infty. \tag{17}$$

In the logit model an increase in any particular ξ has a small impact on the shares of all products, and since the sum of those impacts must be finite, as J grows large its impact on the share of any given product goes down like $1/J$. It is the inverse map from changes in s to the implied $\xi(\cdot)$ that determines the influence of sampling and simulation error on our estimator, and as J grows large the derivatives of this inverse map grow large. To counteract this effect we need to increase the number of sampling and simulation draws, i.e., reduce the variance in those errors, at a rate faster than J; in particular we need $n \propto J^2$. Below we provide the formalities that prove this result and show that the same rate conditions hold for the random coefficient logit analyzed in BLP.

3.1.2 The Vertical Model

Perhaps the simplest among the models with a finite set of product characteristics discussed in Berry and Pakes (2002) is the "vertical" model of Shaked and Sutton (1982). In this model the utility function is

$$u_{ij} = \delta_j - \lambda_i p_j,$$

where $\delta_j = x_j\beta + \xi_j$ and we normalize the outside alternative so that $\delta_0 = p_0 = 0$.

Order the products so $0 = \delta_0 < \delta_1 < \delta_2 < \delta_3 < \dots$. Let $F(\cdot)$ denote the distribution of λ (the marginal utility of income), assume it is increasing over its domain and let

$$\Delta_j = (\delta_j - \delta_{j-1})/(p_j - p_{j-1}), \quad \text{for} \quad j = 1, \dots, J.$$

Then necessary and sufficient conditions for all goods to have positive market share in this model are that $0 = p_0 < p_1 < p_2 < \dots$, and $\Delta_j = (\delta_j - \delta_{j-1})/(p_j - p_{j-1})$ are ordered as $\Delta_1 > \Delta_2 > \dots$. In this case the market shares are given by

$$s_0 = 1 - F(\Delta_1), s_j = F(\Delta_j) - F(\Delta_{j+1}), for j = 1, \dots, J-1, s_J = F(\Delta_J). \tag{18}$$

We analyze this model in detail in section 6.1.2. Here we simply want to point out two properties of its share function. First though equation (17) must hold in the vertical as well as the logit model, in the vertical model

$$\frac{\partial \sigma_k}{\partial \xi_j} = 0 \quad \text{for} \quad j \notin \{j-1, j, j+1\}.$$

That is competition is "local" — only a small number of cross partials are nonzero. Consequently as J grows large none of the nonzero elements of $H(\cdot) \equiv \partial\sigma/\partial\xi$ go to zero, and the elements of $H^{-1}(\cdot)$ remain bounded. This implies that both simulation and sampling error are likely to have less impact on estimators of the vertical than on the horizontal model. Indeed it will allow us to prove an asymptotic normality result when both the number of simulation and the number of sampling draws grows at rate J (rather than J^2 as required for the logit model).

The local nature of competition in the vertical model makes it relatively easy to consider questions related to the choice of instruments for this model. If we assume that there is a Nash pricing equilibrium, and that each product is owned by a distinct firm

$$p_j = mc_j + \frac{F(\Delta_j) - F(\Delta_{j+1})}{f(\Delta_j)\frac{\delta_j - \delta_{j-1}}{(p_j - p_{j-1})^2} + f(\Delta_{j+1})\frac{\delta_{j+1} - \delta_j}{(p_{j+1} - p_j)^2}} \tag{19}$$

for $j = 1, \dots, J$, where $f(\cdot)$ is the density for $F(\cdot)$. So the price of product "j" depends directly on the characteristics and factor prices of the products adjacent to j, and indirectly on the factor prices and characteristics of the other products (through the price of the adjacent products), implying that "good" instruments are likely to depend more on adjacent then non-adjacent product characteristics.

4 Consistency and Asymptotic Normality

We need to specify the way in which the large vector $(s^n, \sigma^R(\theta))$ approaches $(s^0, \sigma(\theta))$. Since these are expanding vectors in which almost all of the individual elements of $(s^0, \sigma(\theta))$ are decreasing to zero, it will not suffice to specify how each component $(s_j^n, \sigma_j^R(\theta))$ approaches

$(s_j^0, \sigma_j(\theta))$; we will require stronger, uniform, notions of convergence, as is common in semiparametric estimation problems.

We will work with the product space $\mathcal{S}_J \times \Theta \times \mathbf{P}$, where $\mathbf{P}$ is the set of probability measures, and endow the marginal spaces with (pseudo) metrics: the L_∞ metric on $\mathbf{P}$, $\rho_P(P,Q) = \sup_{B\in\mathcal{B}} |P(B) - Q(B)|$, where $\mathcal{B}$ is the class of all Borel sets on $\mathbb{R}^k$, the Euclidean metric on Θ, $\rho_E(\theta, \theta') = \|\theta - \theta'\|$, and a metric ρ_{α,s^0} on $\mathcal{S}_J$ defined below. We suppose that

$$\rho_{\alpha,s^0}(s, s^*) = \begin{cases} \max_{0\le j\le J} \left| \frac{(s_j)^\alpha - (s_j^*)^\alpha}{(s_j^0)^\alpha} \right| & \text{if} 0 < \alpha \le 1 \\ \\ \max_{0\le j\le J} \left| s_j - s_j^* \right| & \text{if} \alpha = 0. \end{cases} \tag{20}$$

The metric $\rho_{\alpha,s}$ depends on the parameter α; the higher α is, the stronger is the metric.[6] We state the theory for general α, but in the examples we will take different values of this parameter as convenient. In the logit-like case, we use $\alpha = 1$, while in the vertical case we take $\alpha = 0$. We also put a metric on the space where ξ lives and for this we shall just take the averaged Euclidean metric $\rho_\xi(\xi, \xi^*) = J^{-1}||\xi - \xi^*||^2 = J^{-1}\sum_{j=1}^J (\xi_j - \xi_j^*)^2$. Finally, define for each ϵ, the following neighborhoods of θ^0, P^0, and s^0: $\mathcal{N}_{P^0}(\epsilon) = \{P : \rho_P(P, P^0) \le \epsilon\}$ and $\mathcal{N}_{s^0}(\epsilon) = \{s : \rho_{\alpha,s}(s, s^0) \le \epsilon\}$, $\mathcal{N}_{\theta^0}(\epsilon) = \{\theta : \rho_E(\theta, \theta^0) \le \epsilon\}$, and for each θ and any $\epsilon > 0$, define $\mathcal{N}_{\xi^0}(\theta; \epsilon) = \{\xi : \rho_\xi(\xi, \xi(\theta, s^0, P^0)) \le \epsilon\}$.

4.1 Consistency

The consistency result will require several assumptions in addition to A1-A2: conditions controlling the way in which $s^n, \sigma^R(\theta)$ approach $s^0, \sigma(\theta)$; some asymptotic identification conditions, and some fairly mild restrictions on the instruments.

Assumption A3. *The random sequences s^n and $\sigma^R(\theta)$ are consistent with respect to the corresponding metrics, i.e.,*

$$\text{(a)}\rho_{\alpha,s^0}(s^n, s^0) \xrightarrow{P} 0 \quad ; \quad \text{(b)}\sup_{\theta\in\Theta} \rho_{\alpha,\sigma(\theta)}(\sigma^R(\theta), \sigma(\theta)) \xrightarrow{P} 0, \tag{21}$$

where $\sigma^R(\theta) = \sigma[\xi(\theta, s^0, P^0), \theta, P^R]$ and $\sigma(\theta) = \sigma[\xi(\theta, s^0, P^0), \theta, P^0]$. Furthermore, we suppose that the true market shares satisfy [for the α defined in (20)]

$$\text{(c)}\frac{1}{nJ^\alpha}\sum_{j=0}^J \frac{s_j^0(1-s_j^0)}{(s_j^0)^{2\alpha}} \xrightarrow{P} 0 \quad ; \quad \text{(d)}\sup_{\theta\in\Theta}\left| \frac{1}{R\cdot J^\alpha}\sum_{j=0}^J \frac{\sigma_j(\theta)(1-\sigma_j(\theta))}{(\sigma_j(\theta))^{2\alpha}} \right| \xrightarrow{P} 0.$$

Assumption A3(a) is complicated because the dimensions of the vectors s^n and s^0 increase with J. Note that each s_ℓ^n is a sum of independent bounded random variables with expectation

[6] We include the $j = 0$ term because the uniform convergence of all other terms does not imply its convergence.

Note that the space $\mathcal{S}_J$ and metric ρ_s both change with J; nevertheless, the space can be embedded in the limiting space consisting of all infinite sequences.

s_ℓ^0, conditional on the realization of ξ. Therefore, to verify assumption A3 requires restrictions on the growth rates of $n(J)$ and $R(J)$, and on the limiting behavior of the vector s^0. We will focus on the special case defined by the following assumption:

Condition S. *Suppose that exists positive finite constants* $\underline{c}$ *and* $\bar{c}$ *such that with probability one* $\underline{c}/J \leq s_\ell^0 \leq \bar{c}/J$ *for each* $\ell = 0, 1, \ldots, J$.

This condition implies that $\text{var}(s_\ell^n) = O(1/nJ)$ by assumption A2. Therefore, $(s_\ell^n - s_\ell^0)/s_\ell^0 = O_p(\sqrt{J/n})$ for each $\ell = 0, 1, \ldots, J$. This gives the pointwise rate of convergence; to obtain the sup-norm convergence rate [with respect to the pseudo-metric $\rho_s(s^1, s^2) = \max_{0\leq\ell\leq J} |s_\ell^1 - s_\ell^2| / s_\ell^0$], we apply the Bonferroni and Bernstein inequalities [see Pollard (1989)] to obtain

$$\begin{aligned}
\Pr\left[\max_{0\leq\ell\leq J}\left|\frac{s_\ell^n - s_\ell^0}{s_\ell^0}\right| > \epsilon\right] &\leq \sum_{\ell=0}^{J} \Pr\left[\left|\frac{s_\ell^n - s_\ell^0}{s_\ell^0}\right| > \epsilon\right] \\
&\leq \sum_{\ell=0}^{J} \exp\left(-\frac{\epsilon^2}{2\text{var}(s_\ell^n/s_\ell^0) + 2\epsilon/ns_\ell^0}\right) \\
&\leq \sum_{\ell=0}^{J} \exp\left(-\epsilon^2 O(n/J)\right). \qquad (22)
\end{aligned}$$

A sufficient condition for (22) to decrease to zero is that $J^{1+\epsilon}/n \to 0$ for any $\epsilon > 0$, which implies (21)(a). Assumption A3(b) is similar but requires uniformity over θ. Assumption A3(c) is implied by $J^\alpha/n \to 0$ under condition S, likewise A3(d).

Assumption A4 is a fairly mild restriction on the instruments that will be satisfied for example if they are bounded. Note that there is no presumption that a law of large numbers holds since to show that we would need to be more specific about the details of how the instruments are constructed and the nature of the equilibrium.

Assumption A4. *The instruments are such that the matrix* $z'z/J$ *is stochastically bounded, i.e., for all* $\epsilon > 0$ *there exists an* M_ϵ *such that* $\Pr[||z'z/J|| > M_\epsilon] < \epsilon$.

Next we provide an assumption that ensures the uniform mean square convergence for the vector $\xi(\theta, s^n, P^R)$. We reinterpret solving the equations $s = \sigma(\xi, \theta, P)$ as a minimization problem, thus $\xi(\theta, s, P)$ is the unique minimum of $||s - \sigma(\xi, \theta, P)||$. In fact it is convenient to take a monotonic transform of both sides of the equation $s = \sigma(\xi, \theta, P)$. Specifically, we introduce the componentwise transformation $\tau_J : \mathbb{R}^J \to \mathbb{R}^J$ [i.e., $\tau_J(s) = (\tau_J(s_1), \ldots, \tau_J(s_J))'$] and the $J \times 1$ vector $\psi_J(\xi, \theta, s, P) = \tau_J(s) - \tau_J(\sigma(\xi, \theta, P))$. We then define

$$\xi(\theta, s, P) = \arg\min_{\xi\in\mathbb{R}^J} ||\psi_J(\xi, \theta, s, P)|| \qquad (23)$$

for any θ, s, P. For any bijective transform $\tau_J(\cdot)$, (23) has the same solution. We already know that there exists a unique solution $\xi(\theta, s, P)$ to $s = \sigma(\xi, \theta, P)$ for all (θ, s, P); this is equivalent to saying that $\psi_J(\xi, \theta, s, P) = 0$ if and only if $\xi = \xi(\theta, s, P)$. We use the new definition of $\xi(\theta, s, P)$ as an optimization estimator to guarantee its statistical properties; in view of the increasing dimensions of ψ_J, ξ, however, we must refine the concept of uniqueness of

$\xi(\theta, s, P)$. Let $\tau_J(x) = J^{-\alpha/2}\tau_\alpha(x)$ for some fixed function $\tau_\alpha(x)$, and let $\dot{\tau_\alpha}(x) = d\tau_\alpha(x)/dx$. We shall take

$$\tau_\alpha(x) = \begin{cases} \frac{x^{1-\alpha}-1}{1-\alpha} & \text{if} 0 \le \alpha < 1 \\ \\ \log x & \text{if} \alpha = 1. \end{cases}$$

For each α the function $\tau_\alpha(\cdot)$ is monotonic. In the logit-like case, we use $\alpha = 1$ and $\tau_\alpha(x) = \log x$, while in the pure characteristics case we take $\alpha = 0$ and $\tau_\alpha(x) = x$. The next condition is an asymptotic identification condition used in the analysis of the preliminary estimation of ξ.

Assumption A5. *For all $\delta > 0$, there exists $C(\delta)$ such that*

$$\lim_{J\to\infty} \Pr\left[\inf_{\theta\in\Theta} \inf_{\xi\notin\mathcal{N}_{\xi^0}(\theta;\delta)} \left\|\tau_J(\sigma(\xi, \theta, P^0)) - \tau_J(\sigma(\xi(\theta, s^0, P^0), \theta, P^0)\right\| > C(\delta)\right] = 1.$$

Our assumptions (3) imply that $G_J(\theta^0, s^0, P^0) = o_p(1)$. Assumption A6 is our "identification" condition [c.f. Theorem 3.1 of Pakes and Pollard (1989)]. Note that it does not require convergence of the objective function $G_J(\theta, s^0, P^0)$ at $\theta \neq \theta^0$ (that would require conditions on the process generating the $x's$ and an equilibrium assumption).

Assumption A6. *For all $\delta > 0$, there exists $C(\delta)$ such that*

$$\lim_{J\to\infty} \Pr\left[\inf_{\theta\notin\mathcal{N}_{\theta^0}(\delta)} \|G_J(\theta,\, s^0,\, P^0) - G_J(\theta^0,\, s^0,\, P^0)\| \ge C(\delta)\right] = 1.$$

Theorem 1 [Consistency] *Suppose that A1-A6 hold for some $\alpha \in [0, 1]$ and some $n(J), R(J) \to \infty$. Then, $\widehat{\theta} \xrightarrow{P} \theta^0$.*
The proof is in the appendix. This result applies to a wide range of models and to growth rates on $n(J), R(J)$ as we will see later.

4.2 Asymptotic Normality

We next establish the asymptotic distribution of $\widehat{\theta}$. We shall give conditions under which $\sqrt{J}\mathcal{G}_J(\theta^0)$ is asymptotically normal with bounded variance, while $\sqrt{J}[G_J(\theta, s^n, P^R) - \mathcal{G}_J(\theta)] = o_p(1)$ uniformly over a shrinking neighborhood of θ^0. Additional standard arguments deliver the asymptotic distribution of $\sqrt{J}(\widehat{\theta} - \theta^0)$ in terms of the variance of $\sqrt{J}\mathcal{G}_J(\theta^0)$. The precise magnitude of the variance of $\sqrt{J}\mathcal{G}_J(\theta^0)$ is determined by the behavior of the matrix H_0^{-1}, an issue we will come back to below.

Assumption B1. *θ^0 is an interior point of Θ.*

Assumption B2. *For all θ in some $\delta > 0$ neighborhood of θ^0*

$$E\left[G_J(\theta,\, s^0,\, P^0)\right] = \Gamma^J(\theta - \theta^0) + o(\|\theta - \theta^0\|)$$

uniformly in J. The matrix $\Gamma^J \to \Gamma$ as $J \to \infty$, where Γ has full (column) rank.

In B2 we require only that the expectation of $G_J(\theta, s^0, P^0)$ be differentiable rather than the function itself. This condition is similar to condition (ii) of Theorem 3.3 in Pakes and Pollard (1989). What is different here is that the expectation of $G_J(\theta, s^0, P^0)$. This is because the derivative of $\xi(\cdot)$ with respect to θ depends on J and the form of the instruments both will, in general, depend on the number and characteristics of the products marketed.

ASSUMPTION B3. *For all sequences of positive numbers δ_J such that $\delta_J \to 0$,*

$$\sup_{\|\theta-\theta^0\|\leq\delta_J} \|\sqrt{J}[G_J(\theta, s^0, P^0) - EG_J(\theta,\ s^0,\ P^0)] - \sqrt{J}[G_J(\theta^0,\ s^0,\ P^0) - EG_J(\theta^0,\ s^0,\ P^0)]\| = o_p(1).$$

This assumption is essentially condition (iii) of Theorem 3.3 in Pakes and Pollard (1989). It insures that provided B1-B3 hold and $\sqrt{J}G_J(\theta^0, s^0, P^0)$ is asymptotically normal, any consistent estimator that minimized $\|G_J(\theta, s^0, P^0)\|$ would be asymptotically normal.

To go further we need to work with the disturbances generated by the expansion in (9) and (10). Define the stochastic process in (ξ, P, θ)

$$\nu_J(\xi, P, \theta) = \frac{1}{\sqrt{J}} z' H^{-1}(\xi, \theta, P) \left\{\varepsilon^n - \varepsilon^R(\theta)\right\}, \tag{24}$$

where $\varepsilon^n = (\varepsilon_1^n, \ldots, \varepsilon_J^n)'$ and $\varepsilon^R(\theta) = (\varepsilon_1^R(\theta), \ldots, \varepsilon_J^R(\theta))'$. This process has the structure of a sum of independent random variables from a triangular array as can be seen after interchanging the order of summation, thus

$$\nu_J(\xi, P, \theta) = \sum_{i=1}^{n} Y_{Ji}(\xi, \theta, P) - \sum_{r=1}^{R} Y_{J,r}^*(\xi, \theta, P),$$

$$Y_{Ji}(\xi, \theta, P) = \frac{1}{n\sqrt{J}} \sum_{j=1}^{J} a_j(\xi, \theta, P)\varepsilon_{ji} \quad ; \quad Y_{J,r}^*(\xi, \theta, P) = \frac{1}{R\sqrt{J}} \sum_{j=1}^{J} a_j(\xi, \theta, P)\varepsilon_{j,r}(\theta), \tag{25}$$

where $z'H^{-1}(\xi, \theta, P) \equiv (a_1(\xi, \theta, P), \ldots, a_J(\xi, \theta, P))$. The random variables Y_{Ji} and $Y_{J,r}^*$ are independent across i and r with mean zero and with a distribution that changes with J. This structure is used to apply laws of large numbers and central limit theorems for triangular arrays of independent random variables.

ASSUMPTION B4. *Let $Y_{Ji} = Y_{Ji}(\xi(\theta^0, s^0, P^0), \theta^0, P^0)$ and $Y_{J,r}^* = Y_{J,r}^*(\xi(\theta^0, s^0, P^0), \theta^0, P^0)$. Suppose that $\lim_{J\to\infty} E(z'\xi\xi'z/J) = \Phi_1$ and that*

$$\text{(a)} \lim_{J\to\infty} nE\left[Y_{Ji}Y_{Ji}'\right] = \Phi_2 \quad ; \quad \text{(b)} \lim_{J\to\infty} RE\left[Y_{J,r}^* Y_{J,r}^{*\prime}\right] = \Phi_3 \tag{26}$$

for finite positive definite matrices Φ_j, $j = 1, 2, 3$ and that for some $\delta > 0$, $E(||z'\xi/\sqrt{J}||^{2+\delta}) = o(1)$ and

$$\text{(c)} nE\left[\|Y_{Ji}\|^{2+\delta}\right] = o(1) \quad ; \quad \text{(d)} RE\left[||Y_{J,r}^*||^{2+\delta}\right] = o(1). \tag{27}$$

Condition B4 guarantees that $\sqrt{J}\mathcal{G}_J(\theta^0)$ is asymptotically normal with variance $\Phi = \sum_{i=1}^{3}\Phi_i$. The reason for condition (27) is that as J increases the distribution of the random variables $Y^*_{J,r}$ and Y_{Ji} changes, so we must use the Lyapunov Central Limit Theorem for triangular arrays of independent but not necessarily identically distributed random variables, which in turn requires moment conditions holding to power $2+\delta$. Our examples will translate these conditions into restrictions on $n(J)$ and $R(J)$. To do so we shall have to make more detailed assumptions about z and H_0. The next section will provide the details for our two leading cases.

Finally, we use a stochastic equicontinuity condition on the stochastic process (24) to handle remainder terms. This approach to asymptotics is now well established in econometrics, see the recent survey of Andrews (1994).

Assumption B5. *The process $\nu_J(\xi, P, \theta)$ is stochastically equicontinuous in (ξ, P, θ) at $(\xi(s^0, P^0, \theta^0), P^0, \theta^0)$, that is, for all sequences of positive numbers ϵ_J with $\epsilon_J \to 0$, we have*

$$\sup_{\|\theta-\theta^0\|\le\epsilon_J}\ \sup_{(\xi,P)\in\mathcal{N}_{\xi^0}(\theta^0;\epsilon_J)\times\mathcal{N}_{P^0}(\epsilon_J)} \left\|\nu_J(\xi, P, \theta) - \nu_J(\xi(s^0, P^0, \theta^0), P^0, \theta^0)\right\| = o_p(1).$$

In B5 we need to insure that $\sqrt{J}[G_J(\theta, s, P) - EG_J(\theta, s^0, P^0)]$ can be made arbitrarily close to $\sqrt{J}[G_J(\theta^0, s, P) - EG_J(\theta^0, s^0, P^0)]$ (with arbitrarily large probability) by making θ close to θ^0. This is stronger than the condition needed to make $\sqrt{J}[G_J(\theta, s^0, P^0) - EG_J(\theta, s^0, P^0)]$ close to $\sqrt{J}[G_J(\theta^0, s^0, P^0) - EG_J(\theta^0, s^0, P^0)]$ (we have also to insure that the consumer sampling and the simulation processes do not cause jumps in the disturbance process at values of θ close to θ^0). The stochastic equicontinuity assumption is sufficient to ensure that the remainder term is of smaller order in probability than $\sqrt{J}\mathcal{G}_J(\theta^0)$. We verify this condition below for the logit case. With these conditions we can give the asymptotic normality of $\widehat{\theta}$. The proof is in the appendix.

Theorem 2. [Asymptotic Normality] *Suppose that A1-A6 and B1-B5 hold for some α. Then, with $\Phi = \Phi_1 + \Phi_2 + \Phi_3$,*

$$\sqrt{J}(\widehat{\theta} - \theta^0) \Longrightarrow N[0,\, (\Gamma'\Gamma)^{-1}\Gamma'\Phi\Gamma(\Gamma'\Gamma)^{-1}].$$

Standard errors can be constructed in the usual way. Specifically, when G_J is differentiable in θ let

$$\widehat{\Gamma} = \frac{\partial G_J}{\partial\theta}(\widehat{\theta},\, s^n,\, P^R),$$

and this will consistently estimate Γ; when G_J is not differentiable in θ we must use numerical derivatives as in Pakes and Pollard (1989). Furthermore, let

$$\widehat{\Phi}_2 = \frac{1}{nJ}z'\widehat{H}^{-1}\widehat{V}_2\widehat{H}^{-1\prime}z \quad ; \quad \widehat{\Phi}_3 = \frac{1}{RJ}z'\widehat{H}^{-1}\widehat{V}_3\widehat{H}^{-1\prime}z,$$

where $\widehat{H} = H(\widehat{\theta}, s^n, P^R)$, $\widehat{V}_2 = S^n - s^n s^{n\prime}$, and $\widehat{V}_3 = V_3(\widehat{\theta}, \xi(s^n, \widehat{\theta}, P^R))$, and these will consistently estimate Φ_2 and Φ_3.

We now turn to the efficiency question. One can improve the efficiency of $\widehat{\theta}$ by taking the weighted norm criterion, i.e.,

$$\|G_J(\theta, s, P)\|^2_{W_J} = G_J(\theta, s, P)' W_J G_J(\theta, s, P)$$

for some weighting matrix W_J. The resulting class of estimators can be treated similarly to above: it suffices for asymptotic normality to make the additional assumption that $W_J \rightarrow_p W$ for some symmetric positive definite matrix W, in which case the asymptotic variance is $(\Gamma' W \Gamma)^{-1} \Gamma' W \Phi W \Gamma (\Gamma' W \Gamma)^{-1}$. The optimal weighting matrix is proportional to Φ^{-1}, and the resulting efficient estimator has asymptotic variance $(\Gamma' \Phi^{-1} \Gamma)^{-1}$.

A few final points on efficiency. First if we make a comparison with the estimator that is optimal when s^0, P^0 are known [and the corresponding moment $G_J(\theta, s^0, P^0)$ can be computed], we find that the variance of our estimator is strictly larger, so an estimator of the variance which ignores sampling and simulation error will be biased downwards. Also, since we are only dealing with the demand subsystem here, our estimator can only be efficient in a limited information sense. That is in virtually all currently used pricing models the pricing equation also depends on the parameters of the demand system. So if we were willing to make an equilibrium assumption on how prices are set, we could also use the pricing equation to help estimate the demand parameters.

A related issue is the question of finding an efficient estimator under the conditional moment restriction $E[\xi_j | x_1] = 0$ in the sense of Chamberlain (1987). The form of the efficient estimator will depend on the nature of the pricing equilibrium, and on the ownership structure of products. Moreover, since, in general, the efficient instrument will depend differently on the characteristics and factor prices of all competing products (see the example in section 3.1.2), the number of dimensions needed for a semiparametric approximation to those instruments [as in Newey (1990,1993)] will grow in J.[7]

5 Determining the Rates of Convergence

We seek conditions under which B4 is true, which is equivalent to finding conditions under which the random variables

$$T_{J2} \equiv \sqrt{J}^{-1} z' H_0^{-1} \varepsilon^n, \quad \text{and} \quad T_{J3} \equiv \sqrt{J}^{-1} z' H_0^{-1} \varepsilon^R(\theta)$$

[7] As noted in Pakes (1992), if we are willing to assume exchangeability of the pricing function in the order of the characteristic vectors of the competing products owned by a given firm, and in the firms themselves, the dimensionality problem can be reduced by using an exchangeable basis in forming the semiparametric estimators. Though this way of looking at the problem provides some intuition for the choice of instruments, see for e.g. BLP, in practical situations even a low order exchangeable basis has typically been found to have too high a dimension to be of much use.

are asymptotically normal with zero mean and finite non-zero variances respectively $\lim_{J\to\infty} \Phi_2(J)$ and $\lim_{J\to\infty} \Phi_3(J)$, where

$$\Phi_2(J) = \frac{1}{nJ} z' H_0^{-1} V_2 H_0^{-1} z \quad ; \quad \Phi_3(J) = \frac{1}{RJ} z' H_0^{-1} V_3 H_0^{-1} z. \tag{28}$$

Keep in mind that the matrix H_0 is dimension $J \times J$ and J grows large in our limiting argument.

Since for fixed J both T_{J2} and T_{J3} are a sum of i.i.d. random variables central limit theorems for triangular arrays imply that it will be sufficient to find conditions on $n(J)$ and $R(J)$ that guarantee that the Φ matrices are bounded. We consider the term $\Phi_2(J)$ [similar comments apply to $\Phi_3(J)$]. The behavior of the elements of $H^{-1}(\theta, s^0, P^0)$ has a key role here, and, consequently, we will consider several different scenarios regarding these quantities as is appropriate for different models [i.e., models that have been used for demand estimation], each of which generates a different limit theorem.

The different limit theorems arise because the different models have different implications for the components of $\partial\sigma(\cdot)/\partial\xi$. In particular in the models with "diffuse" substitution patterns, such as the random coefficient logit model of BLP in which all goods are substitutes for all other goods, that partial goes to zero as the number of products increase, and its inverse grows large. Consequently, when J is large a little bit of sampling error causes large changes in the computed value of ξ. In contrast, in the pure characteristic model, competition is "local", the more the number of products the "closer" will your nearest competitor tend to be and the larger will be the response to small changes in the quality of the product. In these cases a little bit of simulation or sampling error will have almost no effect on the computed value of ξ.

Formally, if we let $a' = (a_1, \ldots, a_J) = z' H_0^{-1}$ and suppose, without loss of generality, that z is a $J \times 1$ vector, we have [conditional on s^0]

$$\Phi_2(J) = \frac{1}{nJ} \left[\sum_{j=1}^{J} a_j^2 s_j^0 - \left(\sum_{j=1}^{J} a_j s_j^0 \right)^2 \right], \tag{29}$$

since $V_2 = diag[s^0] - s^0 s^{0\prime}$. The magnitude of the matrix Φ_2 depends on the vectors a and s^0. Note that the term in square brackets in (29) can be considered to be the 'variance' of the vector $(a_1, \ldots, a_J)$ with respect to the multinomial like measure induced by the sequence of weights $(s_1^0, \ldots, s_J^0)$ [note that depending on the behavior of s_0^0, these weights do not necessarily sum to one even asymptotically].

There are three factors that influence the magnitude of $\Phi_2(J)$. First, the rate at which s_j^0, $j = 0, 1, \ldots, J$ decline with J; for the purposes of this discussion, we shall assume that Condition S holds (roughly, all shares go down like $1/J$). Second, the rate at which the $a_j's$ grow or decline with J. Finally, the variability of the sequence $\{a_j\}$ also has a role to play in some cases.

In general, if for some function $g(\cdot)$, we have $|a_j| \le g(J)$ for $j = 1, \ldots, J$, then for all J,

$$\sum_{j=1}^{J} a_j^2 s_j^0 - \left(\sum_{j=1}^{J} a_j s_j^0\right)^2 \leq \sum_{j=1}^{J} a_j^2 s_j^0 \leq \left(\max_{1\leq j\leq J} |a_j|\right)^2 \sum_{j=1}^{J} s_j^0 \leq g(J)^2. \tag{30}$$

This gives a global bound on the variance matrix $\Phi_2(J)$; it is essentially this bound that was used in BLP to provide sufficient conditions for asymptotic normality.

However, it turns out that in a leading special case (the logit and random coefficient logit), there is further structure that can sometimes be exploited to give tighter bounds on $\Phi_2(J)$. Specifically, when Condition S hold in these cases we have

$$(a_1, \ldots, a_J) = g(J)\{(1, \ldots, 1) + O(1/J)\}$$

for some non-decreasing function g [i.e., the normalized a's have zero sample variability]. Then, we have

$$\begin{aligned} \sum_{j=1}^{J} a_j^2 s_j^0 - \left(\sum_{j=1}^{J} a_j s_j^0\right)^2 &\simeq g(J)^2 \left[\sum_{j=1}^{J} s_j^0 - \left(\sum_{j=1}^{J} s_j^0\right)^2\right] \\ &= g(J)^2 \left[1 - s_0^0 - \left(1 - s_0^0\right)^2\right] \\ &= g(J)^2 s_0^0 (1 - s_0^0). \end{aligned} \tag{31}$$

When condition S holds, the share of the outside alternative s_0^0 is $O(1/J)$, and so (31) is $O(g(J)^2/J)$, and we get a reduction in the magnitude of the variance from the crude bound (30).[8]

In a subsequent section we investigate three examples. Our purpose is to verify the order of magnitude of the covariance matrix $\Phi_2(J)$ and to establish the precise rate of growth on $n(J), R(J)$ required to achieve asymptotic normality. We achieve this by identifying the rate of growth and variability of the sequence $\{a_1, \ldots, a_J\}$.

6 Examples

Section 3.1 introduced two examples and we now provide a detailed analysis of both of them. The first was the logit model. In this specification the utility for any one good, conditional on the utilities of all the other goods, has full support. As we showed this implies "diffuse" substitution patterns which, in turn, make estimators of the parameters of the model quite sensitive to sampling and simulation error. The simple logit only accommodates very restrictive substitution patterns. So after formalizing our results for the simple logit we

[8]Note that when the share of the outside alternative is $O(1)$, (31) is the larger magnitude $O(g(J)^2)$. In this case, there is no gain and (30) is not improved.

move on to the random coefficients logit model of BLP (1995)(analogous results hold for the nested logit, the multinomial probit, and the random coefficients probit).

The second example introduced in section 3.1, and dealt with in detail below, was the vertical model of Shaked and Sutton (1982). This and the horizontal model of Hotelling (1929) are uni-dimensional examples of a class of models Berry and Pakes (2002) call the pure characteristics model, and we consider in more detail below. In these models individual's preferences are defined on a finite dimensional space of product characteristics, and substitution patterns are "local" in the sense that cross price and characteristic elasticities are only non-zero for a finite number of products.

The asymptotic behavior of our estimator is likely to be different in the two models. In the first, more traditional class, our examples indicate that the variance in both the simulation and the sampling error must decline at a rate faster than J increases for consistency and at the rate J^2 for asymptotic normality. For our example of the second class of models, the variance in the sampling and the simulation error can decline at any rate for consistency and must decline at rate J for asymptotic normality.

6.1 The logit model

Recall from equations (13) and (14) that the market shares predicted by the logit model are

$$\sigma_j(x,\xi,\theta) = \frac{e^{x_j\theta+\xi_j}}{1+\sum_{k=1}^J e^{x_k\theta+\xi_k}}, \quad j=1,\ldots,J \quad \text{while} \quad \sigma_0(x,\xi,\theta) = \frac{1}{(1+\sum_{k=1}^J e^{x_k\theta+\xi_k})},$$

and from equation (16)

$$\frac{\partial\sigma}{\partial\xi} \equiv H(s,\theta) = S - ss', \quad \text{while} \quad H(s,\theta)^{-1} = S^{-1} + ii'/s_0,$$

where $S = \mathrm{diag}[s]$ and $i = (1,\ldots,1)'$. $H(s,\theta)$ is the $J\times J$ share matrix derivative evaluated at $\xi = \xi(s,\theta)$, and does not depend on the parameter vector θ.

We now verify the conditions of our theorem assuming the random variables $x_j\theta+\xi_j$ have bounded support and density bounded away from zero on this support.[9] This implies market shares are all of magnitude $O(1/J)$ with probability one, i.e., that Condition S holds.

[9] However, when $x_j\theta+\xi_j$ has unbounded support, then a variety of outcomes are possible. For example, suppose that $x_j\theta+\xi_j$ is standard normal, then $\sigma_j(x,\xi,\theta) = O_p(1/J)$, $j=0,1,\ldots,J$, but

$$\min_{1\le j\le J}\sigma_j(x,\xi,\theta) = O_p(J^{-1}\exp(-\sqrt{2\log J})),$$

that is, $\min_{1\le j\le J}\sigma_j(x,\xi,\theta)$ is of smaller order than J^{-1} in probability. In fact, it is only slightly smaller and it can be shown to be larger than $O_p(J^{-(1+\eta)})$ for any $\eta>0$. Thus although 'most' market shares must be $O(1/J)$, the extreme values predicted by the model are a little bit wider. This model never predicts zero market share for any product unless parameters take extreme values like $\theta=\pm\infty$, and this is precluded by the technical requirement that the parameter space be compact. However, one can find even more extreme behaviour of $\min_{1\le j\le J}\sigma_j(x,\xi,\theta)$ when the distribution of $x_j\theta+\xi_j$ has heavy tails like the Cauchy. Depending on this distribution one can have arbitrarily small shares predicted by the model at the extreme outcome of $x_j\theta+\xi_j$. When $x_j\theta+\xi_j$ has finite variance, the consequence of having small shares for our estimation

It is straightforward to verify that condition S implies A3. We simply assume that the instruments are stochastically bounded thus satisfying A4. Taking $\tau_\alpha(x) = \log(x/\sigma_0)$, we have

$$\tau_J(\sigma(\xi, \theta, P^0)) - \tau_J(\sigma(\xi(\theta, s^0, P^0), \theta, P^0) = \xi - \xi(\theta, s^0, P^0)$$

and so the identification condition A5 is also satisfied.

As noted a sufficient condition for A6 is that for each $\epsilon > 0$ there is a $J(\epsilon)$ such that for any $J > J(\epsilon)$, $J^{-1}\sum_j z_j x_j'$ has full column rank with probability $1-\epsilon$, since then

$$\inf_{\theta \notin \mathcal{N}_{\theta^0}(\delta)} \|G_J(\theta, s^0) - G_J(\theta^0, s^0)\| = \inf_{\theta \notin \mathcal{N}_{\theta^0}(\delta)} \| \left(\frac{1}{J} \sum_{j=1}^{J} z_j x_j' \right) (\theta - \theta^0)\| \geq \inf_{\theta \notin \mathcal{N}_{\theta^0}(\delta)} C\|\theta - \theta_0\| \geq C\delta,$$

with probability $1-\epsilon$. In terms of the pricing problem this requires that the price of a product not be a linear function of that product's demand side attributes. However, we know that the solution to the pricing problem generates a pricing function which depends on the characteristics of competitor's, as well as on its own characteristics.

We have just verified the conditions for consistency and we move on to the conditions needed for asymptotic normality; in particular, (26) and (27) when condition S is true. Without loss of generality assume z is a vector, and recall from section 5 that to prove (26) it suffices to find a rate of growth for n that makes the limit, as J grows large, of $\Phi_2(J)$ finite (element by element), where

$$\Phi_2(J) = (nJ)^{-1}[\sum_{j=1}^{J} a_j^2 s_j^0 - (\sum_{j=1}^{J} a_j s_j^0)^2] \quad \text{and} \quad a_k = z' H(\cdot)^{-1} e_k.$$

The formula for $H^{-1}(\cdot)$, and condition S (i.e., all $s_j > \underline{c}/J$) implies

$$a_k = \frac{z_k}{s_k} + \frac{\sum_{j=1}^{J} z_j}{s_0} = \frac{J^2 \overline{z}_J}{\underline{c}}[1 + O(1/J)], \tag{32}$$

where $\overline{z}_J$ is the sample mean of z, which is bounded by assumption. From (28) if

$$(a_1, \ldots, a_J) = g(J)[(1, \ldots, 1) + O(1/J)]$$

then the components of $\Phi_2(J)$ are $O_p[g(J)^2/J^2 n]$. Equation (32) implies we satisfy this condition with $g(J) = J^2 O_p(1)$. Thus the components of $\Phi_2(J)$ are $O_p(J^2/n)$; i.e., n must grow like J^2 for asymptotic normality.

We now verify (27). Note that $|\sum_{j=1}^{J} a_j \varepsilon_{ji}| \leq \max_{1 \leq j \leq J} |a_j| \sum_{j=1}^{J} |\varepsilon_{ji}| \leq cJ^2$ for some constant c, because $\sum_{j=1}^{J} |\varepsilon_{ji}| \leq \sum_{j=1}^{J} 1(C_i = j) + E1(C_i = j) \leq 2$, and (32) is true. Therefore,

$$E\left[\left|\frac{1}{n\sqrt{J}} \sum_{j=1}^{J} a_j \varepsilon_{ji}\right|^{2+\delta}\right] \leq \left(\frac{\overline{c}J^2}{n\sqrt{J}}\right)^{2+\delta}$$

problem is not so severe: it only affects the argument for normality through the remainder term magnitudes, so that normality is preserved although you may need stronger restrictions on the rates of growth of $n(J)$ and $R(J)$.

for any δ. Thus $nE[|Y_{Ji}|^{2+\delta}] = O(J^{3+3\delta/2}n^{-(1+\delta)}) = o(1)$, which, after substituting $n(J) = J^2$, satisfies our condition provided $3 + 3\delta/2 - 2(1+\delta) < 0$. That is condition (27) is satisfied for any $\delta > 2$.

Finally, we turn to the stochastic equicontinuity condition B5. In the logit case, there is no simulation, i.e., P is known exactly, and there is only the sampling error to consider. Furthermore, since the equation for ξ is explicit, we can equivalently work with the process in s,

$$\nu_J(s) = \frac{1}{J} z' H^{-1}(s) \varepsilon^n,$$

where $H^{-1}(s) = S^{-1} - ii'/(1 - i's)$. In the appendix we show that

$$\nu_J(s^n) - \nu_J(s^0) = O_p(J^{3/2}/n), \tag{33}$$

so that the remainder terms are of smaller order than the leading variance terms.

In conclusion, the asymptotic variance of $\sqrt{J}(\widehat{\theta} - \theta_0)$ is $(E(zx'))^{-1}(\Phi_1 + \Phi_2)(E(xz'))^{-1}$, where

$$\begin{aligned} \Phi_2 &= \lim_{J\to\infty} \frac{1}{nJ} z' H_0^{-1} z \\ &= \lim_{J\to\infty} \left[\frac{1}{nJ} \sum_{j=1}^{J} z_j z_j' s_j^{-1} + \frac{J}{n} \frac{E(z)E(z')}{s_0} \right]. \end{aligned} \tag{34}$$

The first term is

$$\frac{1}{nJ} \sum_{j=1}^{J} z_j z_j' e^{-(x_j\theta + \xi_j)} \times \left(1 + \sum_{k=1}^{J} e^{x_k\theta + \xi_k}\right) = O_p(J/n)$$

provided $E[||z_j z_j' e^{-(x_j\theta+\xi_j)}||] < \infty$. The second term in Φ_2 is $O_p(J^2/n)$, and is dominant in this case.[10] Therefore,

$$\Phi_2 = \lim_{J\to\infty} \frac{J^2}{n} \times \frac{E(z)E(z')}{\lim_{J\to\infty}(Js_0)}.$$

6.1.1 The Random Coefficients Logit

The logit model is not very suited to empirical work; as is well-known, it implies odd substitution patterns between products. However, the random coefficients logit, given by

$$u_{ij} = \delta_j + x_j \lambda_i + \epsilon_{ij},$$

is known to give more reasonable substitution patterns because of the random coefficients on the x vector. Our notation is intended to separate out the terms with interactions between

[10] One could also consider the case in which the outside alternative is $O(1)$. Then the two terms in (34) are of equal magnitude and we must include both.

individual and product, $x_j\lambda_i$; usually, the product characteristics that interact are the prices. The systematic utility δ_j depends on the parameters θ and on the product characteristics x_j, ξ_j, i.e., $\delta_j = x_j\theta + \xi_j$.

The market share for this model is given by

$$\sigma_j(x, \xi, \theta) = \int \frac{e^{\delta_j + x_j\lambda}}{1 + \sum_k e^{\delta_k + x_k\lambda}} dP(\lambda) \equiv \int \omega_j(\lambda) dP(\lambda) \equiv E[\omega_j(\lambda)], \tag{35}$$

where P is a given probability measure. Note that the integrand, $\omega_j(\lambda)$, is just the logit market share function evaluated at a particular value of the random coefficients [we have supressed its other arguments x, θ, ξ]. The derivatives of the market share function are

$$\frac{\partial\sigma_j}{\partial\xi_k} = \begin{cases} \int \omega_j(\lambda)\left\{1 - \omega_j(\lambda)\right\} dP(\lambda) & j = k \\ \\ -\int \omega_j(\lambda)\omega_k(\lambda) dP(\lambda) & \text{if} k \neq j. \end{cases}$$

In matrix terms we can write the share matrix

$$H = E[\mathcal{H}(\lambda)], \text{where} \mathcal{H}(\lambda) = W(\lambda) - w(\lambda)w(\lambda)'$$

in which $W(\lambda) = diag(\omega_1(\lambda), \ldots, \omega_J(\lambda))'$ and $w(\lambda) = (\omega_1(\lambda), \ldots, \omega_J(\lambda))'$. Unfortunately, there is no analytic inverse for this model and no easy expression (that we know of) for the inverse matrix H_0^{-1} in the general continuous case. However, we can still characterize its properties sufficiently well to ensure that property (31) holds.

By the convexity of the matrix inverse [Groves and Rothenberg (1969)] we have

$$H^{-1} = [E\mathcal{H}(\lambda)]^{-1} \leq E[\mathcal{H}(\lambda)^{-1}]$$

in the positive definite sense. The inverse of any given logit matrix is $W(\lambda)^{-1} + ii'/\omega_0(\lambda)$. If we assume that $\omega_j(\lambda) \geq \underline{\omega}_j$ for all $j = 0, 1, \ldots, J$ for some nonrandom sequence of constants $\underline{\omega}_j$ that obey condition S, then

$$[E\mathcal{H}(\lambda)]^{-1} \leq \underline{W}^{-1} + \frac{ii'}{\underline{\omega}_0} \equiv \overline{H^{-1}}, \tag{36}$$

where $\underline{W} = diag(\underline{\omega}_1, \ldots, \underline{\omega}_J)'$. Furthermore, $H^{-1}V_2H^{-1} \leq \overline{H^{-1}}V_2\overline{H^{-1}}$ by the properties of positive definite symmetric matrices [Anderson (1984, Theorem A1.1)]. We can now apply the results from the previous subsection. Under condition S, the variance term (29) is of order J^2/n as in the fixed coefficient logit case. The remaining arguments of the previous subsection hold here too so that the condition for the central limit theorem is satisfied in the random coefficient case. In fact, we are able to prove in this case that

$$\Phi_2 = \lim_{J\to\infty} \frac{J^2}{n} \times \frac{E(z)E(z')}{\lim_{J\to\infty}(J \int \omega_0(\lambda) dP(\lambda))} \tag{37}$$

$$\Phi_3 \leq \lim_{J\to\infty} \frac{J^2}{R} \times \frac{E(z)E(z')}{\lim_{J\to\infty}(J \int \omega_0(\lambda) dP(\lambda))}. \tag{38}$$

We could provide more detailed formalizations of both the identification and stochastic equicontinuity conditions, but we really have nothing substantive to say that we have not already said in the context of the fixed coefficient logit model.

6.2 The Vertical Model

Recall from equation (18) that the market shares are then given by

$$s_0 = 1 - F(\Delta_1), s_j = F(\Delta_j) - F(\Delta_{j+1}), for j = 1, \ldots, J-1, s_J = F(\Delta_J),$$

where $\Delta_j = (\delta_j - \delta_{j-1})/(p_j - p_{j-1})$, and $\Delta_1 > \Delta_2 > \ldots$, while $\delta_j = x_j\beta + \xi_j$ and $\delta_1 > \delta_2 > \ldots$ (recall $\delta_0 = p_0 = 0$).

Since the simple vertical model only require integration over one dimension of heterogeneity, we assume there is no simulation error. Further for this model the inversion from shares to ξ is obtained from the recursive system $\delta_j - \delta_{j-1} = (p_j - p_{j-1})F^{-1}(1 - \sum_{r=1}^{j-1} s_r)$. So our requirement for consistency, i.e., that $J^{-1}||\xi(s^n) - \xi(s^0)||^2 \to_p 0$, is expressed in terms of

$$\xi_j(s^n) - \xi_j(s^0) = \sum_{l=1}^{j}(p_l - p_{l-1})[F^{-1}(1 - \sum_{l=0}^{j-1} s_l) - F^{-1}(1 - \sum_{l=0}^{j-1} s_l^n)].$$

For simplicity we assume that the distribution of λ (i.e., $F(\cdot)$) has bounded support and is strictly increasing (so its inverse satisfies a Lipschitz condition), and that whatever equilibrium is established $\max_{j \le J}(p_j - p_{j-1}) = c < \infty$.[11] Then for any $\epsilon > 0$

$$\begin{aligned} \Pr\left[\frac{1}{J}\sum_{j=1}^{J}\left\{\xi_j(s^n) - \xi_j(s^0)\right\}^2 > \epsilon\right] &\le \max_{j\le J}\Pr\left[\left\{\xi_j(s^n) - \xi_j(s^0)\right\}^2 > \epsilon\right] \\ &\le J\max_{j\le J}\Pr\left[\left\{\sum_{l=0}^{j-1} s_l - \sum_{l=0}^{j-1} s_l^n\right\}^2 > \epsilon/c\right] \\ &\le J\exp(-\epsilon n/c), \end{aligned}$$

by Bernstein's inequality (since $\sum_{l=0}^{j-1} s_l^n$ can be expresses as a sum of n independent random variables each bounded by one). The last term goes to zero provided $n \to \infty$ faster than $\log J$.

[11] Other assumptions might be relevant here, but would require more detailed analysis of the pricing equilibrium for the model to make sense. For e.g. if the support of λ were $\mathcal{R}^+$ and $max_{j\le J}\Delta_j f(\Delta_j) \to 0$, which might happen, for example if the distribution of $1/\lambda$ were lognormal or Pareto, and we assumed a standard Nash in prices equilibrium, we would have to insure that difference in prices of adjacent products would go to zero for their to be positive markups (and hence for entry) as market size grows larger; see the pricing equation for this model in (19).

Recall that to find the rate at which we need n to grow for the consistency and asymptotic normality results we need the elements of the matrix H^{-1}, where $H = \partial\sigma/\partial\xi$. Letting $\alpha_1 = f(\Delta_1)/p_1$, $\alpha_2 = f(\Delta_2)/(p_2 - p_1), \ldots, \alpha_J = f(\Delta_J)/(p_J - p_{J-1})$, it can be shown that

$$H = \begin{pmatrix} \alpha_1+\alpha_2 & -\alpha_2 & 0 & \cdots & 0 \\ -\alpha_2 & \alpha_2+\alpha_3 & \ddots & 0 & 0 \\ 0 & \ddots & \ddots & -\alpha_{J-1} & 0 \\ \vdots & 0 & -\alpha_{J-1} & \alpha_J+\alpha_{J-1} & -\alpha_J \\ 0 & 0 & 0 & -\alpha_J & \alpha_J \end{pmatrix}. \tag{39}$$

The matrix H is a band matrix with all elements more than one place from the diagonal being zero. Note also that all row and columns sums are zero apart from the first row and column, and so the matrix is not diagonal dominant. Furthermore, it can be verified that

$$H^{-1} = \left(\sum_{r=1}^{\min(i,j)} \frac{1}{\alpha_r}\right)_{i,j} = \begin{pmatrix} \frac{1}{\alpha_1} & \frac{1}{\alpha_1} & \frac{1}{\alpha_1} & \cdots & \frac{1}{\alpha_1} \\ \frac{1}{\alpha_1} & \frac{1}{\alpha_1}+\frac{1}{\alpha_2} & \frac{1}{\alpha_1}+\frac{1}{\alpha_2} & \cdots & \frac{1}{\alpha_1}+\frac{1}{\alpha_2} \\ \frac{1}{\alpha_1} & \frac{1}{\alpha_1}+\frac{1}{\alpha_2} & \frac{1}{\alpha_1}+\frac{1}{\alpha_2}+\frac{1}{\alpha_3} & \cdots & \frac{1}{\alpha_1}+\frac{1}{\alpha_2}+\frac{1}{\alpha_3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{1}{\alpha_1} & \frac{1}{\alpha_1}+\frac{1}{\alpha_2} & \cdots & \cdots & \frac{1}{\alpha_1}+\cdots+\frac{1}{\alpha_J} \end{pmatrix}.$$

Notice that any fixed i, j element of the inverse matrix is of order one as $J \to \infty$ (this is in contrast to the logit models where the individual elements of the inverse were all of order J).

Assume that the z are bounded. Then, for $k = 1, \ldots, J$,

$$\begin{aligned} a_k \equiv z'H^{-1}e_k &\leq \max|z_l| \times \left[J\sum_{\ell=1}^{k}\frac{1}{\alpha_\ell} + \sum_{j=1}^{k-1}\left(\sum_{\ell=1}^{j}\frac{1}{\alpha_\ell} - \sum_{\ell=1}^{k}\frac{1}{\alpha_\ell}\right)\right] \\ &= \max|z_l| \times \left[J\sum_{\ell=1}^{k}\frac{p_\ell - p_{\ell-1}}{f(\Delta_\ell)} - \sum_{j=1}^{k-1}\sum_{\ell=j+1}^{k}\frac{p_\ell - p_{\ell-1}}{f(\Delta_\ell)}\right], \end{aligned}$$

which gives the individual elements in the vector $z'H^{-1}$. It now follows that because $p_j < p_k$ when $j < k$,

$$a_k \leq \max|z_l| \times J\sum_{\ell=1}^{k}\frac{p_\ell - p_{\ell-1}}{f(\Delta_\ell)}, \tag{40}$$

which is of order J for any fixed k.

For Theorems 1 and 2 we must determine the magnitude of the sample variance of the sequence $(a_1, \ldots, a_J)$ with respect to the multinomial$(s_1, \ldots, s_J)$ or equivalently the multinomial$(1/J, \ldots, 1/J)$. In fact,

$$\frac{1}{J}\sum_{k=1}^{J}a_k^2 - \left(\frac{1}{J}\sum_{k=1}^{J}a_k\right)^2 \leq \frac{1}{J}\sum_{k=1}^{J}a_k^2$$

$$
\begin{aligned}
&\leq \quad J^2 \left(\max|z_l|\right)^2 \times \frac{1}{J}\sum_{k=1}^{J}\left(\sum_{\ell=1}^{k}\frac{p_\ell - p_{\ell-1}}{f(\Delta_\ell)}\right)^2 \\
&\leq \quad J^2 \left(\max|z_l|\right)^2 \times \frac{\frac{1}{J}\sum_{k=1}^{J}\left(\sum_{\ell=1}^{k}(p_\ell - p_{\ell-1})\right)^2}{\left\{\min_{1\leq\ell\leq J} f(\Delta_\ell)\right\}^2} \\
&\leq \quad J^2 \left(\max|z_l|\right)^2 | \times \underline{m}^{-2}\frac{1}{J}\sum_{k=1}^{J} p_k^2,
\end{aligned}
$$

which is finite provided $\frac{1}{J}\sum_{k=1}^{J} p_k^2$ is finite and $\min_{1\leq\ell\leq J} f(\Delta_\ell) \geq \underline{m} > 0$. Both of these conditions are assumed (although one can use a similar argument to obtain different rates when they are not).

Given the lower bound $\underline{m}$ and that the price sequence has a finite second moment, (30) holds with $g(J) = J$. Therefore, the covariance matrix $\Phi_2(J)$ is of order J/n. That is, in this case, we obtain consistency if n increases at any rate faster than $\log J$, while the asymptotic normality result holds with all three terms contributing provided n grows like J. We do not know whether one can improve on our inequality.

Note the contrast to the logit-type models, where n must increase at rate J for consistency and rate J^2 for the asymptotic normality result [when all shares are the same magnitude]. The difference between the models is due to the difference between localized and diffuse competition. In the models with sampling and simulation errors, the derivative of market share with respect to product quality is declining at the same rate as the shares. Therefore, the elements of the inverse derivative matrix $(d\sigma\,/d\xi)^{-1}$ are growing in J, and the number of simulation draws must increase at a faster rate to offset this. In the vertical model, competition is localized and the derivative of market share with respect to product quality does not decline in J, and so the elements of the inverse derivative matrix stay bounded. As a result our limit theorems can suffice with a lower rate of growth for n in the vertical model.

7 Monte Carlo Results

In this section we discuss Monte Carlo results for simple versions of our models. We start with logit-type models. In particular we present results for a simple logit where the market shares are observed with sampling error, and then for a random coefficients logit with simulation error in the computed shares. Next we turn to the pure characteristic models. Here we start with a simple vertical model where market shares are observed with sampling error and then move to a multi-dimensional pure characteristics model with simulation error. The monte-carlo results reinforce the theoretical discussion in the previous sections. That is to obtain a "well-behaved" estimator for the first class of models sample sizes and simulation draws must be quite large and increase rapidly in J. The sample sizes and number of simulation draws which seem to be necessary for estimating the versions of the pure-characteristic model can be much smaller, and, do not increase nearly as rapidly in J.

All of our examples here involve data on a single-cross section of markets, to fit with our theoretical discussion of how estimates behave as J varies within a market. In practice there are several reasons to prefer to estimate off data that features a cross section or time series of different market equilibria.

For the logit model, the deterministic part of utility is drawn as

$$\delta_j = x_j\beta + \xi_j, \tag{41}$$

where ξ_j is drawn from the standard normal distribution. The x's are a constant and a standard normal, with a β coefficient on the constant of 3 and a slope coefficient of 1. (Except as noted, all random variables in the Monte Carlo exercises are i.i.d. draws.)

Table 1 gives the mean estimated value of β_2 across 1000 Monte Carlo datasets. Each column gives results for a different value of J, the number of products, while the panels running down the table vary the number of consumer draws used to calculate the market share of the sample (n). Note that zero shares are discarded from the dataset. The fourth panel gives results for n set proportional to J, while in the fifth panel n is set equal to J^2 . The last column uses the true expected shares (i.e., "$n = \infty$").

In the second row of each panel is the simulated standard deviation (the standard error of the estimate across the simulated samples) and the third row gives the standard error of the mean (the simulated standard error divided by $\sqrt{1000}$). Apart from the inversion, the simple logit model is linear in parameters. Thus, given no sampling error in the shares, we should get unbiased results even for small J. This is consistent with the results for $n = \infty$.

We see that the results are particularly bad for small n relative to J, with a large apparent bias. This is in large part due to the sample selection bias that comes from throwing small share products out of the market.[12] A good with a low value of x will tend to have a positive market share only if it has a large value of ξ while a good with a high value of x will tend to have a positive share even for small ξ . This generates a negative correlation between x and ξ among goods with postive market shares.

Table 2 gives Monte Carlo results for a random coefficients logit. In this case (as in most of the empirical literature on aggregate data), we assume that observed market shares have no sampling error.[13] We can always simulate positive predicted shares and so there is no sample selection problem. In this case, we can consider small values of R, but because the computational burden is higher we do not include a set of results for J^2.

Our random coefficients logit example once again sets $\delta_j = x_j\beta + \xi_j$, but now $\beta = (-5, 1)$. Utility of consumer i for product j is

$$u_{ij} = \delta_j + \theta_x \lambda_i x_{j2} + \epsilon_{ij}, \tag{42}$$

where λ is standard normal, the standard deviation of the random "taste for x", θ_x, is set to one and x_{j2} is the non-constant element of x. As usual, the ϵ's are i.i.d. extreme-value

[12] We did not deal with this problem in our theoretical analysis above, but it is likely to be a problem for datasets built from small samples of consumers.

[13] I.e. we are assuming that the observed shares are aggregated over a very large number of consumers.

draws. The market shares are calculated by taking R draws from the distribution of the random coefficient λ. The "observed" market shares are set to their expected value at the true parameter values (i.e., we are assuming that the observed shares are aggregated over a very large number of consumers.)

Computation of the inverse shares follows BLP, but we do not use a variance reduction (importance sampling) scheme of sort used in that paper.

Table 2 summarizes the estimates of θ_x, the standard deviation of the random coefficient on the non-constant x. The results are consistent with the theory that suggest that the estimation routine will perform badly when the number of simulation draws is "small" relative to the number of products. In particular there seems to be a bias that increases in J holding R fixed (at least at low values of R). Also when J gets large (take our $J = 100$), we need fairly large value of R for that bias to go away (for J=100 we probably need $R > 500$). As the theory predicts the variance goes down in the number of products (in J). Recall that for this model consistency requires both J and R to increase and R must increase at a faster rate than J.

Table 3 has results for the vertical model. As in Table 1, the variance in observed shares is generated by small samples of consumers rather than from simulation error in the predicted shares. Once again, this can produce zero observed market shares, but in the vertical model the zero share products can be included in the estimation routine at little cost.[14]

The exact vertical model considers a utility function of

$$u_{ij} = \delta - \theta_p \lambda_i p_j, \tag{43}$$

where δ is "quality", λ_i is consumer-specific part of the the marginal disutility of a price increase and θ_p is a parameter of the model. To keep the random coefficient in an easy one-parameter family, we assume that 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. In fact, θ_p is not separately identified from demand-side data and so is held fixed at one in the Monte Carlo experiments (this is just a normalization.)

Quality is modeled as $\delta_j = x_j\beta + \xi_j$, where the two components of x are a constant and a uniform drawn from $(0, 2)$. β is set equal to (1.5,1). The "unobserved" ξ_j is uniform on $(-1, 1)$. To insure that the expected shares are all positive, price is set equal to δ^2.[15]

The results in Table 3 summarize the estimates of β_2, the slope coefficient on x in the quality equation. These results are very different than those for the logit-type models in table 2. Indeed when we use the vertical model it is striking that there is no apparent inconsistency in the estimates anywhere in the table (even when $J = 200$ and $n = 50$). As expected, for fixed n the variance decreases in J. However for small n the decrease is almost

[14] In practice, the inversion for δ simply sets the δ of zero share products to the δ of the next lowest-priced good. Since zero shares occur in the vertical model when δ's are "close together", this creates little bias. Note the contrast to the logit model, where zero share products have systematically low δs and where the inversion routine cannot handle zero shares. We should note that this is the choice of δ for a zero market share product produced by our estimation algorithm; any δ below this value would also be consistent with a zero market share.

[15] In the vertical model, all shares will be positive if price increases "fast enough" in quality.

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.

We conclude with an example of the computation of δ in a multidimensional pure characteristics- model. To the vertical model of Table 3, we add a random coefficient on the observed x,

$$u_{ij} = \delta + \theta_x \lambda_{i1} x_j - \theta_p \lambda_{i2} p_j, \tag{44}$$

with

$$\delta = \beta_0 + \beta_1 x_j + \xi. \tag{45}$$

There are now two dimensions of the unobserved consumer tastes, related to x and p.

In the vertical model of Table 3, computation was not an issue and so we focused on small consumer samples as the source of "simulation" error. In the pure random coefficients model, the market shares must be simulated and so we focus on simulation error. We assume that the consumer sample is very large. This is fairly realistic in many datasets and avoids the problem of sample selection that arises in the small sample case.

As discussed in Berry and Pakes (2002), estimation of the parameters is computationally cumbersome as the number of products increases, making it difficult to estimate the model with many repetitions in a Monte Carlo exercise. For computational tractability, we therefore focus on the computation of δ at fixed values of parameters. In particular, we hold the parameters (θ_x, θ_p) of the random coefficients at their true values and then compute δ from the constructed data on x, p and expected market shares. The method of solving for ξ is the "exact" homotopy method of Berry and Pakes (2002). To summarize the relevant error in the computation, we regress the computed δ on x to obtain an estimate of (β_0, β_1). In Table 4, we report the mean estimates of β_1 we obtained from repeating this procedure for different number of products and simulation draws.

The data for Table 4 were created via the following assumptions. The observed x_j is drawn as 1.5 times a random uniform on (0,1). The unobserved ξ_j is drawn as a a random uniform on (0,1). (Note that somewhat more of the variance in δ comes from x as oppose to ξ, which will aid the estimation procedure.) The term δ is then constructed via the parameters $(\beta_0, \beta_1) = (2, 1)$. To ensure positive market shares, price is set equal to a convex function of δ, $e^{\delta}/10$. The random "taste" for x is standard normal, while the random term on price is modeled as a standard log-normal (with $\mu = 1$.)

The results in Table 4 are consistent with our conjecture that the multi-dimensional pure characteristics model behaves much as the single-dimendional (vertical) model. In particular there is no obvious bias in the estimates even when there are only a small number of simulation draws. The "∞" row of Table 4 uses the true δ that created the data (as this is the δ that would be recovered if both n and R were infinite). This row therefore gives the results from the model without any simulation error. It is apparent that at low values of J and high values of R very little of the standard error of the estimate is attributable to simulation error, but that fraction is still quite large when $J = R$ (note that throughout we keep R fairly small as that keeps the computational burden of estimating the model repeatedly on different simulated data sets manageable). Overall, however, the table seems consistent with the conjecture that the multidimensional pure characteristics model behaves similar to the unidimensional characteristic model; in particular we do not need R to grow faster than J for consistency and fairly precise estimates can be obtained from relatively small values of R.

A Appendix

Proof of Theorem 1. We first show that the estimator defined as any sequence that satisfies

$$\|G_J(\widehat{\theta},\, s^0,\, P^0)\| = \inf_{\theta\in\Theta} \|G_J(\theta,\, s^0,\, P^0)\| + o_p(1)$$

is consistent. Note that assumption A1 together with the law of large numbers for triangular arrays [see, for example, Billingsley (1986, Theorem 6.2)] imply that $\|G_J(\theta^0, s^0, P^0)\| = o_p(1)$. Therefore, by Theorem 3.1 of Pakes and Pollard (1989) it will suffice to show that for every $(\delta, \varepsilon) > (0, 0)$ there exists a $C^*(\delta) > 0$ and an $J(\varepsilon)$ such that for $J \geq J(\varepsilon)$

$$\Pr\left[\inf_{\theta\notin\mathcal{N}_{\theta^0}(\delta)} \|G_J(\theta)\| \geq C^*(\delta)\right] \geq 1-\varepsilon,$$

where we have omitted indexing G_J by $(s^0,\, P^0)$ for notational convenience. From the triangle inequality $\inf_{\theta\notin\mathcal{N}_{\theta^0}(\delta)} \|G_J(\theta) - G_J(\theta^0)\| \geq C(\delta)$ implies that

$$\inf_{\theta\notin\mathcal{N}_{\theta^0}(\delta)} \|G_J(\theta)\| \geq C(\delta) - \|G_J(\theta^0)\|.$$

Fix $\varepsilon > 0$, and let $\varepsilon^* = \min\{\varepsilon,\, C(\delta)\}$, so that $0 < \varepsilon^* \leq \varepsilon$. Since $\|G_J(\theta^0)\| = o_p(1)$, there exists $J_1(\varepsilon^*)$ such that for any $J \geq J_1(\varepsilon^*)$, $\Pr\{\|G_J(\theta^0)\| \geq \varepsilon^*/2\} \leq \varepsilon^*/2$. By assumption A1, there exists $J_2(\varepsilon^*)$ such that for $J \geq J_2(\varepsilon^*)$, $\Pr\{\inf_{\theta\notin\mathcal{N}_{\theta^0}(\delta)} \|G_J(\theta) - G_J(\theta^0)\| \geq C(\delta)\} \geq 1-\varepsilon^*/2$. Consequently, (2) implies that for $J \geq \max\{J_1(\varepsilon^*),\, J_2(\varepsilon^*)\}$

$$\Pr\left[\inf_{\theta\notin\mathcal{N}_{\theta^0}(\delta)} \|G_J(\theta)\| \geq C(\delta) - \varepsilon^*/2\right] \geq 1-\varepsilon^* \geq 1-\varepsilon.$$

To complete the proof let $C^*(\delta) = C(\delta) - \varepsilon^*/2 > 0$.

We now return to the actual estimator $\widehat{\theta}$ and show that

$$\|G_J(\widehat{\theta},\, s^n,\, P^R)\| = \inf_{\theta\in\Theta} \|G_J(\theta,\, s^0,\, P^0)\| + o_p(1). \tag{46}$$

We show that

$$\sup_{\theta\in\Theta} \frac{1}{J}||\xi(\theta, s^n, P^R) - \xi(\theta, s^0, P^0)||^2 = o_p(1), \tag{47}$$

which implies that

$$\begin{aligned} \sup_{\theta\in\Theta} ||\frac{1}{J}z'\{\xi(\theta, s^n, P^R) - \xi(\theta, s^0, P^0)\}||^2 &\leq \frac{1}{J}||z'z||^2 \times \frac{1}{J}\sup_{\theta\in\Theta}||\xi(\theta, s^n, P^R) - \xi(\theta, s^0, P^0)||^2 \\ &= o_p(1), \end{aligned}$$

i.e., that $\sup_{\theta\in\Theta}\|G_J(\theta,\, s^n,\, P^R)-G_J(\theta,\, s^0,\, P^0)\| = o_p(1)$. This in turn implies (46) by the triangle inequality.

The result (47) follows from the following argument. We show below that

$$\sup_{\theta\in\Theta} ||\psi_J(\xi(\theta, s^n, P^R), \theta, s^0, P^0)|| = o_p(1). \tag{48}$$

Then, by Assumption A5: when $||\xi - \xi(\theta, s^0, P^0)|| \geq \delta\sqrt{J}$, we have $\inf_{\theta\in\Theta} ||\psi_J(\xi, \theta, s^0, P^0)|| \geq \varepsilon$. This implies that $||\xi(\theta, s^n, P^R) - \xi(\theta, s^0, P^0)||^2/J = o_p(1)$ by contradiction, which concludes the proof of (47) and hence (46). The result (48) follows because:

$$\begin{aligned}
\sup_{\theta\in\Theta} ||\psi_J(\xi(\theta, s^n, P^R), \theta, s^0, P^0)|| &\leq \sup_{\theta\in\Theta} ||\psi_J(\xi(\theta, s^n, P^R), \theta, s^0, P^0) - \psi_J(\xi(\theta, s^n, P^R), \theta, s^n, P^R)|| \\
&\leq \sup_{\theta\in\Theta}\sup_{\xi} ||\psi_J(\xi, \theta, s^0, P^0) - \psi_J(\xi, \theta, s^n, P^R)|| \\
&\leq \left\|\tau_J(s^n) - \tau_J(s^0)\right\| \\
&\quad + \sup_{\theta\in\Theta}\sup_{\xi} \left\|\tau_J(\sigma(\xi, s^0, P^0) + \varepsilon^R(\theta)) - \tau_J(\sigma(\xi, s^0, P^0))\right\|.
\end{aligned}$$

For some intermediate values $\overline{s}_j$ we have by the mean value theorem

$$\begin{aligned}
\left\|\tau_J(s^n) - \tau_J(s^0)\right\|^2 &= \frac{1}{J^\alpha}\sum_{j=1}^{J}\left[\dot{\tau_\alpha}\,(\overline{s}_j)(s_j^n - s_j^0)\right]^2 \\
&\leq \max_{1\leq j\leq J}\left|(s_j^0)^\alpha\,\dot{\tau_\alpha}\,(\overline{s}_j)\right|^2 \frac{1}{J^\alpha}\sum_{j=1}^{J}\left[\frac{s_j^n - s_j^0}{(s_j^0)^\alpha}\right]^2 \\
&\leq \max_{1\leq j\leq J}\left|(s_j^0)^\alpha\,\dot{\tau_\alpha}\,(\overline{s}_j)\right|^2 \times \frac{1}{nJ^\alpha}\sum_{j=1}^{J}\frac{s_j^0(1-s_j^0)}{(s_j^0)^{2\alpha}} \times (1+o_p(1)) \\
&= o_p(1),
\end{aligned}$$

by assumption A3, while $\max_{1\leq j\leq J}|(s_j^0)^\alpha\,\dot{\tau_\alpha}\,(\overline{s}_j)| \leq M$ with probability tending to one by assumptions A3. This is because

$$\begin{aligned}
M &\geq \max_{1\leq j\leq J}\left|(\overline{s}_j)^\alpha\,\dot{\tau_1}\,(\overline{s}_j)\right| \\
&= \max_{1\leq j\leq J}\left|\{(s_j^0)^\alpha + \overline{s}_j^\alpha - (s_j^0)^\alpha\}\,\dot{\tau_\alpha}\,(\overline{s}_j)\right| \\
&\geq \max_{1\leq j\leq J}\left|(s_j^0)^\alpha\tau_\alpha(\overline{s}_j)\right| - \max_{1\leq j\leq J}\left|\frac{\overline{s}_j^\alpha - (s_j^0)^\alpha}{\overline{s}_j^\alpha}\right| \max_{1\leq j\leq J}|\overline{s}_j^\alpha\,\dot{\tau_\alpha}\,(\overline{s}_j)| \\
&= \max_{1\leq j\leq J}\left|(s_j^0)^\alpha\tau_\alpha(\overline{s}_j)\right| + o_p(1),
\end{aligned}$$

where the $o_p(1)$ term follows from A3(a) and (c). The result

$$\sup_{\theta\in\Theta}\left\|\tau_J(\sigma(\xi, s^0, P^0)+\varepsilon^R(\theta))-\tau_J(\sigma(\xi, s^0, P^0))\right\| = o_p(1)$$

follows by similar arguments using A3(b) and (d) ■

Proof of Theorem 2. As discussed in section 3, this will follow from Pakes and Pollard (1989, Theorem 2) provided our remainder terms are $o_p(1)$ and the leading terms satisfy a central limit theorem.

Leading Term Argument. We show that

$$\left[\operatorname{var}\left(c'\sqrt{J}\mathcal{G}_J(\theta^0)\right)\right]^{-1/2} c'\sqrt{J}\mathcal{G}_J(\theta^0) \tag{49}$$

is asymptotically normally distributed with mean zero and variance one for any vector c. Since the three terms in $\sqrt{J}\mathcal{G}_J(\theta^0)$, denoted T_{J1}, T_{J2}, and T_{J3}, say, are mutually independent it suffices to show that $\operatorname{var}\left(c'T_{J\ell}\right)^{-1/2} c'T_{J\ell}$, $\ell = 1, 2, 3$, converge to standard normal random variables. Then, by the Cramér-Wold device [the fact that a multivariate random variable is normal if any linear combination of its elements are], we have the result.

A standard central limit theorem for mutually uncorrelated random variables establishes that

$$\left(c'E\{\operatorname{var}(\xi|z)zz'\}c\right)^{-1/2} c'J^{-1/2}z'\xi(\theta^0, s^0,\, P^0) \Longrightarrow N(0,\, 1).$$

Condition (27) enables us to apply the Lyapunov central limit theorem for triangular arrays [see for example, Billingsley (1986, Theorem 27.3)], which says that the random variables $c'\sum_{i=1}^n Y_{Ji}$ and $c'\sum_{r=1}^R Y_{J,r}^*$ are asymptotically normal.

We now turn to the remainder terms. For each fixed θ, we use a Taylor series approximation to $\xi(\theta,\, s^n,\, P^R)-\xi(\theta,\, s^0,\, P^R)$ and to $\xi(\theta,\, s^0,\, P^R)-\xi(\theta,\, s^0,\, P^0)$. Specifically, by the intermediate value theorem

$$\begin{aligned} 0 &= \sigma(\xi(\theta, s^n, P^R), \theta, P^R) - s^n \\ &= \sigma(\xi(\theta, s^0, P^R), \theta, P^R) - s^n + \frac{\partial\sigma(\overline{\xi}, \theta, P^R)}{\partial\xi'}\left\{\xi(\theta, s^n, P^R)-\xi(\theta, s^0, P^R)\right\}, \end{aligned} \tag{50}$$

where $\overline{\xi}$ is intermediate between $\xi(\theta, s^n, P^R)$ and $\xi(\theta, s^0, P^R)$. In fact, there are different vectors $\overline{\xi}$ for each row, but we suppress this for notational convenience. Thus using the facts that $\sigma(\xi(\theta, s^0, P^R), \theta, P^R) = s^0$ and that for any $\xi \in \mathcal{N}_{\xi^0}(\theta; \epsilon)$ the matrix $\partial\sigma(\xi, \theta, P^R)/\partial\xi'$ is invertible with probability tending to one, we can write

$$\xi(\theta, s^n, P^R) - \xi(\theta, s^0, P^R) = -\left\{\frac{\partial\sigma(\overline{\xi}, \theta, P^R)}{\partial\xi'}\right\}^{-1}\varepsilon^n \tag{51}$$

with probability tending to one. Likewise,

$$\begin{aligned} 0 &= \sigma(\xi(\theta, s^0, P^R), \theta, P^R) - s^0 \\ &= \sigma(\xi(\theta, s^0, P^0), \theta, P^R) - s^0 + \frac{\partial\sigma(\underline{\xi}, \theta, P^R)}{\partial\xi'} \left\{\xi(\theta, s^0, P^R) - \xi(\theta, s^0, P^0)\right\}, \end{aligned}$$

where $\underline{\xi}$ are intermediate between $\xi(\theta, s^0, P^R)$ and $\xi(\theta, s^0, P^0)$ as before. Then we use the fact that $\sigma(\xi(\theta, s^0, P^0), \theta, P^R) - s^0 = \sigma(\xi(\theta, s^0, P^0), \theta, P^R) - \sigma(\xi(\theta, s^0, P^0), \theta, P^0) = \varepsilon^R(\theta)$ to obtain that with probability tending to one

$$\xi(\theta, s^0, P^R) - \xi(\theta, s^0, P^0) = -\left\{\frac{\partial\sigma(\underline{\xi}, \theta, P^R)}{\partial\xi'}\right\}^{-1} \varepsilon^R(\theta). \tag{52}$$

Therefore,

$$\begin{aligned} \sqrt{J}[\mathcal{G}_J(\theta) - G_J(\theta, s^n, P^R)] &= -\frac{1}{\sqrt{J}} z' \left[H(\overline{\xi}, \theta, P^R)^{-1} - H(\theta, s^0, P^0)^{-1}\right] \varepsilon^n \\ &\quad -\frac{1}{\sqrt{J}} z' \left[H(\underline{\xi}, \theta, P^R)^{-1} - H(\theta, s^0, P^0)^{-1}\right] \varepsilon^R(\theta). \end{aligned} \tag{53}$$

We must establish that $\sqrt{J}[\mathcal{G}_J(\theta) - G_J(\theta, s^n, P^R)] = o_p(1)$ uniformly in θ in a shrinking neighborhood of θ^0. We just show that

$$\sup_{||\theta-\theta^0|| \leq \epsilon_J} ||\frac{1}{\sqrt{J}} z'\{H(\overline{\xi}, \theta, P^R)^{-1} - H(\theta^0, s^0, P^0)^{-1}\}\varepsilon^n|| = o_p(1), \tag{54}$$

from which the result follows. The proof for the term (53) is similar and is omitted. Since $\overline{\xi}$ is intermediate between $\xi(\theta, s^n, P^R)$ and $\xi(\theta, s^0, P^R)$ it is also consistent in mean square, i.e., there exists a sequence $\epsilon_J \to 0$ such that $\Pr[\overline{\xi} \notin \mathcal{N}_{\xi^0}(\theta^0; \epsilon_J)] \to 0$. Furthermore, for this ϵ_J we have $\Pr\{\rho_P(P^R, P^0) \geq \epsilon_J\} \to 0$ by the Glivenko-Cantelli theorem. Then, notice that for any $\eta > 0$,

$$\begin{aligned} &\Pr\left[\sup_{||\theta-\theta^0|| \leq \epsilon_J} ||\frac{1}{\sqrt{J}} z'\{H(\overline{\xi}, \theta, P^R)^{-1} - H(\theta, s^0, P^0)^{-1}\}\varepsilon^n|| > \eta\right] \\ \leq\; &\Pr\left[\sup_{||\theta-\theta^0|| \leq \epsilon_J} ||\nu_J(\overline{\xi}, P^R, \theta) - \nu_J(\xi(s^0, P^0, \theta), P^0, \theta)|| > \eta\right] \\ \leq\; &\Pr\left[\sup_{||\theta-\theta^0|| \leq \epsilon_J} \sup_{(\xi,P) \in \mathcal{N}_{\xi^0}(\theta^0;\epsilon_J) \times \mathcal{N}_{P^0}(\epsilon_J)} ||\nu_J(\xi, P, \theta) - \nu_J(\xi(s^0, P^0, \theta), P^0, \theta)|| > \eta\right] \\ &+ \Pr\left[\overline{\xi} \notin \mathcal{N}_{\xi^0}(\theta^0; \epsilon_J)\right] + \Pr\left[P^R \notin \mathcal{N}_{P^0}(\epsilon_J)\right] \\ =\; &\Pr\left[\sup_{||\theta-\theta^0|| \leq \epsilon_J} \sup_{(\xi,P) \in \mathcal{N}_{\xi^0}(\theta^0;\epsilon_J) \times \mathcal{N}_{P^0}(\epsilon_J)} ||\nu_J(\xi, P, \theta) - \nu_J(\xi(s^0, P^0, \theta), P^0, \theta)|| > \eta\right] + o(1) \\ =\; &o(1) \end{aligned}$$

by the stochastic equicontinuity condition B5. ■

PROOF OF (33). It suffices to show that for any random sequence $s(n)$ converging to s^0 we have $||\nu_J(s(n)) - \nu_J(s^0)|| \to_p 0$. We shall take $s(n) = s^n$ and show that $R_n = \nu_J(s^n) - \nu_J(s^0) = o_p(1)$, where

$$R_n = \frac{1}{\sqrt{J}} z'\{(S^n)^{-1} - S^{-1}\}(s^n - s) + \frac{1}{\sqrt{J}} z'ii'(s^n - s)\{\frac{1}{1 - i's^n} - \frac{1}{1 - i's}\} \equiv R_{n21} + R_{n22}.$$

The following argument shows that under our conditions $R_{n21} = O_p(J^{3/2}/n)$ and $R_{n22} = O_p(J^{3/2}/n)$. We deal first with R_{n21}, which can be rewritten using a geometric series expansion as

$$|R_{n21}| \le \max \|z_\ell\| \times \frac{1}{\sqrt{J}} \sum_{\ell=1}^{J} \frac{\delta_\ell^2}{1 + \delta_\ell},$$

where $\delta_\ell = (s_\ell^n - s_\ell)/s_\ell$. For any $\epsilon > 0$,

$$\begin{aligned}
\Pr\left[|R_{n21}| > \epsilon\right] &\le \Pr\left[|R_{n21}| > \epsilon \,\text{and} \max_{1\le\ell\le J} |\delta_\ell| \le 1/2\right] + \Pr\left[\max_{1\le\ell\le J} |\delta_\ell| > 1/2\right] \\
&\le \Pr\left[|R_{n21}| > \epsilon \text{ and } \max_{1\le\ell\le J} |\delta_\ell| \le 1/2\right] + o(1)
\end{aligned}$$

by the uniform convergence of δ_ℓ assumed in A3. When $\max_{1\le\ell\le J} |\delta_\ell| \le 1/2$, $|R_{n21}| \le \frac{2}{\sqrt{J}} \sum_{\ell=1}^{J} \delta_\ell^2$, and by the Markov inequality

$$\begin{aligned}
\Pr\left[\frac{2}{\sqrt{J}} \sum_{\ell=1}^{J} \delta_\ell^2 > \epsilon\right] &\le \frac{\frac{2}{\sqrt{J}} \sum_{\ell=1}^{J} E(\delta_\ell^2)}{\epsilon} \\
&= \frac{\frac{2}{n\sqrt{J}} \sum_{\ell=1}^{J} \frac{(1-s_\ell)}{s_\ell}}{\epsilon} = O(J^{3/2}/n).
\end{aligned}$$

Similar calculation applies to R_{n22}. ■

PROOF OF (38) AND (37). We show that for any vector z,

$$\frac{z'H^{-1}e_k}{J^2} = \frac{\mu_z}{\overline{s}_0} + O(1/J), \quad k = 1, \dots, J, \tag{55}$$

where $\mu_z = \lim_{J\to\infty} J^{-1} \sum_{j=1}^{J} z_j$ and $\overline{s}_0 = \lim_{J\to\infty} J \int s_0(\lambda) dP(\lambda)$. The variance formula then follows from (31). Note that the matrix Γ is the same as in the fixed coefficient logit case.

The proof of (55) is quite long because we can't directly calculate the inverse of H in this case. Instead we approximate the continuous mixture by a sequence of finite mixture, each of whose inverse we can compute. Let $T, T_J : \mathbf{P} \to \mathbb{R}$, where

$$T_J(P) = \frac{z'H(P)^{-1}e_k}{J^2} \quad ; \quad T(P) = \frac{\mu_z}{\overline{s}_0(P)},$$

where the notation $H(P)$ emphasizes the dependence of the matrix H on the probability measure P. We must show that for all $\epsilon > 0$, there exists J_0 such that for all $J \geq J_0$,

$$|T_J(P) - T(P)| < \epsilon.$$

We shall work with a discrete mixture of fixed coefficient models indexed by m. By the triangle inequality

$$\begin{aligned} |T_J(P) - T(P)| &\leq |T_J(P) - T_J(P_m)| + |T_J(P_m) - T(P_m)| + |T(P_m) - T(P)| \\ &= I + II + III \end{aligned}$$

for any m. The proof that III is small follows directly from our assumptions and the strong law of large numbers. We show below that II converges to zero uniformly in m, J. What remains is to show that I is small, which follows from the crude inequality

$$\frac{1}{J^2}\left|z'H(P)^{-1}e_k - z'H(P_m)^{-1}e_k\right| \leq \frac{1}{J^2}\left\|z'H(P_m)^{-1}\right\| \|H(P) - H(P_m)\| \left\|H(P)^{-1}e_k\right\| \tag{56}$$

and the following bounds (obtained below)

$$\begin{aligned} \left\|z'H(P_m)^{-1}\right\| &\leq O(J^{5/2}) & (57) \\ \left\|H(P)^{-1}e_k\right\| &\leq O(J^2) & (58) \\ \|H(P) - H(P_m)\| &\leq O(1/m^{(1-\eta)/2}J^{1/2}), & (59) \end{aligned}$$

provided $J^{4+\eta}/m \to 0$.

Proof of (58). Writing $H(P)^{-1} = C^{-1}B$, we have that $H(P)^{-1}e_k = \left(\frac{b_{1k}}{c_1}, \ldots, \frac{b_{Jk}}{c_J}\right)$ whose (squared) norm is

$$\sum_{j=1}^{J} \frac{b_{jk}^2}{c_j^2} \leq \frac{1}{\min_{1\leq j\leq J} c_j^2}\left(\sum_{j=1}^{J} b_{jk}\right)^2 \leq \frac{\text{cons tan t}}{J^2\left\{1 - \Delta(J)\right\}^2} = O(J^4)$$

because the elements of B and C are known to be positive. This establishes (58). The verification of (57) is given below.

Proof of (59). Specifically, we show that the matrix $H(P) = \int S(\lambda)dP(\lambda) - \int s(\lambda)s(\lambda)'dP(\lambda)$ can be well approximated by the matrix $H(P_m) = \int S(\lambda)dP_m(\lambda) - \int s(\lambda)s(\lambda)'dP_m(\lambda)$, where P_m is an empirical distribution of size m from the population governed by P, that is,

$$H(P_m) = \frac{1}{m}\sum_{\ell=1}^{m}\left\{S(\lambda_\ell) - s(\lambda_\ell)s(\lambda_\ell)'\right\}.$$

We work element by element. Since $Js_j(\lambda)$ is bounded away from both zero and infinity, we have that for positive finite constants c_1 and c_2,

$$\begin{aligned}
\Pr\left[\left|J^2\int s_j(\lambda)s_k(\lambda)\left\{dP_m(\lambda)-dP(\lambda)\right\}\right|>\frac{\kappa}{m}\right] &\leq \exp\left[-2\kappa^2/mc_1\right]\\
\Pr\left[J\left|\int s_j(\lambda)(1-s_j(\lambda))\left\{dP_m(\lambda)-dP(\lambda)\right\}\right|>\frac{\kappa}{m}\right] &\leq \exp\left[-2\kappa^2/mc_2\right],
\end{aligned}$$

by Hoeffding's exponential inequality, see Pollard (1984, p191). Therefore taking $\kappa = cm^{1/2}(\log m)^r$, we have by the Bonferroni inequality,

$$\begin{aligned}
&\Pr\left[\max_{1\leq j\neq k\leq J}\left|J^2\int s_j(\lambda)s_k(\lambda)\left\{dP_m(\lambda)-dP(\lambda)\right\}\right|>\frac{c(\log m)^r}{m^{1/2}}\right]\\
\leq\ &\sum\sum_{j\neq k}\Pr\left[J^2\left|\int s_j(\lambda)s_k(\lambda)\left\{dP_m(\lambda)-dP(\lambda)\right\}\right|>\frac{c(\log m)^r}{m^{1/2}}\right]\\
=\ &O(J^2)\exp[-c^*(\log m)^{2r}] \qquad (60)
\end{aligned}$$

for some constant c^*. Taking $m = J^\alpha$ for any $\alpha > 0$, we get that

$$\sum_{m=1}^{\infty}\Pr\left[\max_{1\leq j\neq k\leq J}\left|J^2\int s_j(\lambda)s_k(\lambda)\left\{dP_m(\lambda)-dP(\lambda)\right\}\right|>\frac{c(\log m)^r}{m^{1/2}}\right]<\infty$$

provided $r > 3/2c^*\alpha$, so that by the Borel-Cantelli lemma, we have for any $\eta > 0$,

$$m^{(1-\eta)/2}\max_{1\leq j\neq k\leq J}\left|J^2\int s_j(\lambda)s_k(\lambda)\left\{dP_m(\lambda)-dP(\lambda)\right\}\right|\longrightarrow 0 \qquad (61)$$

with probability one. Similarly,

$$m^{(1-\eta)/2}\max_{1\leq j\leq J}\left|J\int s_j(\lambda)(1-s_j(\lambda))\left\{dP_m(\lambda)-dP(\lambda)\right\}\right|\longrightarrow 0 \qquad (62)$$

with probability one. In conclusion, the discrete mixture of logits well approximates any random coefficient logit matrix. Specifically, (59) follows because

$$\begin{aligned}
\|H(P)-H(P_m)\|^2 &= \sum_{j=1}^{J}\{H(P)-H(P_m)\}_{j,j}^2+\sum_{\substack{j=1\\ j\neq k}}^{J}\sum_{k=1}^{J}\{H(P)-H(P_m)\}_{j,k}^2\\
&\leq J\max_{1\leq j\leq J}\{H(P)-H(P_m)\}_{j,j}^2+J^2\max_{1\leq j\neq k\leq J}\{H(P)-H(P_m)\}_{j,k}^2\\
&= O(1\Big/\sqrt{Jm^{1-\eta}})
\end{aligned}$$

with probability one for large m, J by (61) and (62).

Proof of II. Consider the discrete mixture

$$H = \frac{1}{m}\sum_{\ell=1}^{m}(S^\ell - s^\ell s^{\ell\prime}),$$

where $s^\ell = (s_1^\ell, \ldots, s_J^\ell)'$, $\ell = 1, \ldots, m$. We show that

$$\frac{1}{J^2} z' H^{-1} e_k = \frac{\frac{1}{J}\sum_{j=1}^J z_j}{J\frac{1}{m}\sum_{\ell=1}^m s_0^\ell} + O(1/J), \quad k = 1, \ldots, J, \tag{63}$$

where $s_0^\ell = 1 - \sum_{j=1}^J s_j^\ell = O(1/J)$, $\ell = 1, \ldots, m$.

Write $H = (D + UV')/m$, where $D = \sum_{\ell=1}^m S^\ell$ and $U = (s^1, \ldots, s^J)$ and $V = -(s^1, \ldots, s^J)$. We have

$$z' H^{-1} e_k = m\left\{z' D^{-1} e_k - z' D^{-1} U (I + V' D^{-1} U)^{-1} V' D^{-1} e_k\right\} \tag{64}$$

by the Sherman-Morrison-Woodbury formula [Golub and Van Loan (1989, p51)]. First note that

$$z' D^{-1} e_k = \frac{z_k}{d_k} = O(J/m),$$

where $d_j = \sum_{\ell=1}^m s_j^\ell = O(m/J)$, $j = 1, \ldots, J$, so this term is of smaller order. We are going to establish that

$$\left[(I + V' D^{-1} U)^{-1}\right]_{ij} = \frac{1 + O(1/J)}{\sum_{\ell=1}^m s_0^\ell\,[1 + O(1/J)]} \tag{65}$$

for all $i, j = 1, \ldots, m$. In this case,

$$\frac{m}{J^2} z' D^{-1} U (I + V' D^{-1} U)^{-1} V' D^{-1} e_k = \frac{1}{J^2 \frac{1}{m}\sum_{\ell=1}^m s_0^\ell} z' D^{-1} U i i' V' D^{-1} e_k + O(1/J),$$

where $i' V' D^{-1} e_k = 1$ and $z' D^{-1} U i = \sum_{j=1}^J z_j$, so we get the required result (63).

We have

$$z' D^{-1} U_{1\times m} = \left(\ \sum_{j=1}^J \frac{z_j s_j^1}{d_j},\quad \cdots \quad ,\sum_{j=1}^J \frac{z_j s_j^m}{d_j}\ \right) \quad ; \quad V' D^{-1} e_k = -\begin{pmatrix} \frac{s_k^1}{d_k} \\ \vdots \\ \frac{s_k^m}{d_k} \end{pmatrix}$$

and

$$I + V' D^{-1} U = \begin{pmatrix} 1 - \sum_{j=1}^J \frac{(s_j^1)^2}{d_j} & -\sum_{j=1}^J \frac{s_j^1 s_j^2}{d_j} & \cdots & -\sum_{j=1}^J \frac{s_j^1 s_j^m}{d_j} \\ -\sum_{j=1}^J \frac{s_j^2 s_j^1}{d_j} & 1 - \sum_{j=1}^J \frac{(s_j^2)^2}{d_j} & & -\sum_{j=1}^J \frac{s_j^2 s_j^m}{d_j} \\ \vdots & & \ddots & \vdots \\ -\sum_{j=1}^J \frac{s_j^m s_j^1}{d_j} & -\sum_{j=1}^J \frac{s_j^m s_j^2}{d_j} & \cdots & 1 - \sum_{j=1}^J \frac{(s_j^m)^2}{d_j} \end{pmatrix}. \tag{66}$$

Substitute $s_j^m = d_j - \sum_{\ell=1}^{m-1} s_j^\ell$ and use the fact that $\sum_{j=1}^J s_j^\ell = 1 - s_0^\ell$, to obtain

$$\begin{aligned}
\sum_{j=1}^{J}\frac{s_j^m s_j^k}{d_j} &= 1-s_0^k-\sum_{\ell=1}^{m-1}\left(\sum_{j=1}^{J}\frac{s_j^k s_j^\ell}{d_j}\right)\equiv 1-s_0^k-\frac{1}{m}\sum_{\ell=1}^{m-1}a_{\ell k}\\
\sum_{j=1}^{J}\frac{(s_j^m)^2}{d_j} &= \sum_{j=1}^{J}d_j+\sum_{\ell=1}^{m-1}\sum_{k=1}^{m-1}\sum_{j=1}^{J}\frac{s_j^\ell s_j^k}{d_j}-2\sum_{\ell=1}^{m-1}\sum_{j=1}^{J}s_j^\ell\\
&\equiv \sum_{\ell=1}^{m-1}\sum_{k=1}^{m-1}a_{\ell k}+\sum_{\ell=1}^{m-1}s_0^\ell-s_\ell^m-(m-2),
\end{aligned}$$

where $a_{\ell k}=\sum_{j=1}^{J}\frac{s_j^k s_j^\ell}{d_j}$. Therefore, we can write

$$I+V'D^{-1}U=\begin{bmatrix} A & a\\ a' & b\end{bmatrix}+\frac{1}{J}\begin{bmatrix} 0_{m-1,m-1} & \delta\\ \delta' & \phi\end{bmatrix}=X+\frac{E}{J},$$

where the $m-1\times m-1$ matrix A is

$$A=\begin{pmatrix} 1-a_{11} & -a_{12} & \cdots & -a_{1,m-1}\\ -a_{12} & 1-a_{22} & \cdots & -a_{2,m-1}\\ \vdots & & \ddots & \vdots\\ -a_{1,m-1} & -a_{2,m-1} & \dots & 1-a_{m-1,m-1}\end{pmatrix},$$

while the $m-1\times 1$ column vectors

$$a=\begin{bmatrix} -\left\{1-\sum_{\ell=1}^{m-1}a_{1\ell}\right\}\\ \vdots\\ -\left\{1-\sum_{\ell=1}^{m-1}a_{m-1,\ell}\right\}\end{bmatrix}\quad;\quad \delta=\begin{bmatrix} Js_0^1\\ \vdots\\ Js_0^{m-1}\end{bmatrix},$$

and the scalars $b=(m-1)-\sum_{\ell=1}^{m-1}\sum_{k=1}^{m-1}a_{\ell k}$ and $\phi=J(-\sum_{\ell=1}^{m-1}s_0^\ell+s_\ell^m)$.

Note that the matrix $X=(x_{jk})$ is singular, in fact the last column (row) is equal to minus the sum of the preceding $m-1$ columns (rows). Therefore, by Taylor expansion

$$\det\left(X+\frac{E}{J}\right)=\frac{1}{J}\sum_{j,k=1}^{m}\frac{\partial\det(X)}{\partial x_{jk}}e_{jk}+\frac{1}{2J^2}\sum_{j,k,l,r=1}^{m}\frac{\partial^2\det(X)}{\partial x_{jk}\partial x_{lr}}e_{jk}e_{lr}+\dots \tag{67}$$

First, we have that $\partial\det(X)/\partial x_{jk}=x_{jk}^{Adj}$, where x_{jk}^{Adj} is the adjoint [i.e., the determinant of the matrix X_{jk} formed by deleting the j'th row and k'th column from X, see Anderson (1984, p598)] of x_{jk}. In fact, for all j,k

$$x_{jk}^{Adj}=\det(A), \tag{68}$$

as we show below. Since most of the matrix $E=(e_{jk})$ is zero, we only need the adjoints corresponding to the outer (right) border of the matrix X, which means there are only order m terms in the first summation in (67). Also, note that

$$\frac{\partial^2 \det(X)}{\partial x_{mj}\partial x_{mk}} = \frac{\partial^2 \det(X)}{\partial x_{jm}\partial x_{km}} = 0 \quad j,k = 1,\ldots,m,$$

so there are only order m^2 terms in the second summation. Furthermore, since

$$\frac{\partial^2 \det(X)}{\partial x_{mj}\partial x_{km}} = \det(A_{jk}) = O(\det(A)/m),$$

the second term in (67) is of order m/J^2 and

$$\det(I + V'D^{-1}U) = \det(A)\sum_{\ell=1}^{m} s_0^{\ell}[1 + O(1/J)]. \tag{69}$$

Finally, we must show that the adjoints of the matrix $Z = X + E/J$ satisfy

$$z_{jk}^{Adj} = \det(A)[1 + O(1/J)], \quad j \neq k, \tag{70}$$

which implies (65) holds. ■

Proof of (68). We use the fact that determinants are invariant to certain linear transformations and also that the matrix X has the following property

$$x_{jm} = -\sum_{\ell=1}^{m-1} x_{j\ell} \quad ; \quad x_{mk} = -\sum_{\ell=1}^{m-1} x_{m\ell}, \quad j,k = 1,\ldots,m,$$

to show that the determinant of the matrix

$$X_{mj} = \begin{bmatrix} x_{11} & \cdots & x_{1,j-1} & x_{1,j+1} & \cdots & x_{1,m} \\ \vdots & & \vdots & \vdots & & \vdots \\ x_{m-1,1} & \cdots & x_{m-1,j-1} & x_{m-1,j+1} & \cdots & x_{m-1,m} \end{bmatrix}$$

is the same as the determinant of the matrix A. Specifically, add columns 1 to $m-2$ to the $m-1$'th column and one gets the matrix A. For general X_{jk} a sequence of such transformations gives the result. ■

Proof of (70). Essentially the same as above.

■

References

[1] Anderson, T.W. (1984): *An introduction to multivariate analysis.* Wiley.

[2] Andrews, D.W.K. (1994): "Empirical process methods in econometrics," in The Handbook of Econometrics volume 4, pp 2247-2294, Eds R.F. Engle and D.L. McFadden.

[3] Berry, S., (1994): "Estimating discrete choice models of product differentiation," *RAND Journal of Economics*, **25** , 242-262.

[4] Berry, S., Levinsohn, J., and A. Pakes, (1995): "Automobile prices in market equilibrium," *Econometrica* **63** , 841-890.

[5] Berry, S., Levinsohn, J., and A. Pakes, (2001): "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market," *Cowles Foundation Working Paper #1337*

[6] Berry, S. and A. Pakes, (2002): "The Pure Characteristics Discrete Choice Model," *Harvard University Working Paper*

[7] Billingsley, P. (1986): *Probability and Measure* . 2nd Edition. John Wiley. New York.

[8] Chamberlain, G. (1987): "Asymptotic efficiency in estimation with conditional moment restrictions," *Journal of Econometrics* 34, 305-334.

[9] Court, A.T. (1939): "Hedonic Price Indexes with Automotive Examples," *The Dynamics of Automobile Demand*, pp.99-117, the General Motors Corporation.

[10] Dieudonne, J. (1969): *Foundations of Modern Analysis.* Academic Press: New York.

[11] Fiedler, M., (1986): *Special Matrices and their applications in numerical mathematics.* Kluwer: Dordrecht.

[12] Golub, G.H., and C.F. van Loan (1989): *Matrix Computations.* The Johns Hopkins University Press: Baltimore.

[13] Griliches, Z. (1961): "Hedonic Price Indices for Automobiles: An Econometric Analysis of Quality Change", *The Price Statistics of the Federal Government*, New York, the Natinal Bureau of Economic Research.

[14] Groves, T., and T.J. Rothenberg (1969). A note on the expected value of an inverse matrix. Biometrika , 690-691.

[15] Hansen, L. (1982): "Large Sample Properties of Generalized Method of Moments Estimators", *Econometrica* 50, 1029-1054.

[16] Hausman, J.A. and D. Wise, (1978). "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences," *Econometrica* 46, 403-426.

[17] Hotelling, H. (1929): "Stability in Competition," *Economic Journal* 39, 41-57.

[18] HOTZ, V.J. AND R. MILLER (1993): “Conditional choice probabilities and the estimation of dynamic models,” *Review of Economic* Studies 60, 497-529.

[19] MAMMEN, E. (1989): “Asymptotics with increasing dimension for robust regression with applications to the bootstrap,” *The Annals of Statistics* **17**, 382-400.

[20] MARCUS, M. AND H. MINC (1964): *A survey of matrix theory and matrix inequalities*. Dover, N.Y.

[21] NEVO, A. (2001). “Measuring Market Power in the Ready to Eat Cereal Industry,” *Econometrica*, 69(2), 307-342.

[22] NEWEY, W.K., (1990): “Efficient instrumental variables estimation of nonlinear models,” *Econometrica* 58, 809-837.

[23] NEWEY, W.K., (1993): “Efficient estimation of models with conditional moment restrictions”, in *Handbook of Statistics, vol. 11*, G.S. Maddala, C.R. Rao, and H.D. Vinod eds., Amsterdam: North-Holland.

[24] ORTEGA, J.M. AND W.C. RHEINBOLDT (1970): *Iterative solution of nonlinear equation in several variables*. Academic Press, London.

[25] PAKES, A., (1992): “Dynamic Structural Models; Problems and Prospects”, in J.J. Laffont and C. Sims (ed.) *Advances in Econometrics; Proceedings of the Sixth World Congress of the Econometric Society*, Cambridge University Press.

[26] PAKES, A. AND S. BERRY AND J. LEVINSOHN, (1993): “Some Applications and Limitations of Recent Advances in Empirical Industrial Organization: Price Indexes and the Analysis of Environmental Change,” *American Economic Review Papers and Proceedings*, 83, 240–246.

[27] PAKES, A., AND S. OLLEY (1995): “A Limit Theorem for a Smooth Class of Semiparametric Estimators,” *Journal of Econometrics* 65, 295-332.

[28] PAKES, A., AND D. POLLARD (1989): “Simulation and the asymptotics of optimization estimators,” *Econometrica* **57**, 1027-1057.

[29] POLLARD, D. (1984): *Convergence of Stochastic Processes*. Springer-Verlag. New York.

[30] PORTNOY, S. (1984): “Asymptotic behavior of M-estimators of p regression parameters when p^2/n is large. I. Consistency,” *The Annals of Statistics* **12**, 1298-1309.

[31] PORTNOY, S. (1985): “Asymptotic behavior of M-estimators of p regression parameters when p^2/n is large. II. Normal approximation,” *The Annals of Statistics* **13**, 403-1417.

[32] SHAKED, A. AND J. SUTTON (1982). “Relaxing price competition through product differentiation,” *The Review of Economic Studies*, 49, 3–13.

[33] SHAKED, A. AND J. SUTTON (1990). “Multiproduct Firms and Market Structure,” *Rand Journal of Economics*, 21, 45-62.

Table 1:
Monte Carlo Estimates for the Simple Logit Model
True Value of the Parm is 1
1000 Monte Carlo Repetitions

# Consumer Draws	# of Products (J)				
(n)	10	25	50	100	200
500	0.941	0.798	0.778	0.633	0.518
	(0.362)	(0.209)	(0.137)	(0.086)	(0.076)
	[0.011]	[0.007]	[0.068]	[0.004]	[0.002]
1000	0.997	1.013	0.974	0.934	0.882
	(0.426)	(0.255)	(0.149)	(0.120)	(0.077)
	[0.014]	[0.008]	[0.005]	[0.004]	[0.002]
2000	1.023	1.046	0.998	0.976	0.923
	(0.500)	(0.224)	(0.138)	(0.123)	(0.089)
	[0.016]	[0.007]	[0.004]	[0.004]	[0.004]
$10J$	0.685	0.728	0.768	0.921	0.916
	(0.406)	(0.214)	(0.132)	(0.110)	(0.088)
	[0.013]	[0.007]	[0.004]	[0.004]	[0.004]
J^2	0.615	0.857	1.021	1.022	1.015
	(0.358)	(0.200)	(0.139)	(0.101)	(0.077)
	[0.011]	[0.006]	[0.004]	[0.003]	[0.002]
∞	1.027	0.997	0.995	1.007	1.008
	(0.376)	(0.242)	(0.133)	(0.094)	(0.073)
	[0.012]	[0.008]	[0.004]	[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 $[\cdot]$.

Table 2:
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	1.218	*
	(0.982)	(0.512)	
	[.031]	[0.016]	
50	1.025	1.039	1.241
	(0.645)	(0.311)	(0.495)
	[0.020]	[0.010]	[0.016]
100	0.982	1.013	1.037
	(0.674)	(0.271)	(0.209)
	[0.021]	[0.009]	[0.007]
500	0.998	1.008	1.015
	(0.633)	(0.255)	(0.181)
	[0.002]	[0.008]	[0.006]
10J	0.982	1.008	1.018
	(0.674)	(0.255)	(0.181)
	[0.014]	[0.008]	[0.006]

Notes: Simulated Standard Errors (empirical standard deviations across the repititions) in $(\cdot)$ and Simulated Standard Error of the Estimated Mean in $[\cdot]$.
*With 100 products and only 10 draws, we had numeric problems computing the estimates.

Table 3:
Monte Carlo Estimates for the Pure 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 $[\cdot]$.

Table 4:
Monte Carlo Estimates for a Pure Characteristics Model
True Value of the Parm is 1
100 Monte Carlo Repetitions

# Simulation	# of Products, (J)			
Draws (R)	10	25	50	100
10	1.039	0.999	1.016	1.021
	(0.370)	(0.332)	(0.311)	(0.325)
	[0.037]	[0.033]	[0.031]	[0.033]
25	1.043	0.993	0.999	1.010
	(0.279)	(0.268)	(0.235)	(0.214)
	[0.028]	[0.027]	[0.024]	[0.021]
50	1.040	1.006	0.992	1.024
	(0.243)	(0.215)	(0.187)	(0.161)
	[0.024]	[0.021]	[0.019]	[0.016]
100	1.036	1.023	0.987	1.012
	(0.224)	(0.182)	(0.143)	(0.136)
	[0.022]	[0.018]	[0.014]	[0.014]
J	1.039	0.993	0.992	1.012
	(0.370)	(0.268)	(0.187)	(0.136)
	[0.037]	[0.027]	[0.019]	[0.014]
∞	1.030	1.013	0.986	1.002
	(0.207)	(0.164)	(0.103)	(0.061)
	[0.021]	[0.016]	[0.010]	[0.006]

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