

LIMIT THEOREMS FOR ESTIMATING THE PARAMETERS OF DIFFERENTIATED PRODUCT DEMAND SYSTEMS*

by

Steve Berry
Yale University

Oliver B. Linton
London School of Economics and Political Science

Ariel Pakes
Harvard University

Contents:

Abstract

1. Introduction

2. The Model

3. Overview of the Main Results

4. Consistency and Asymptotic Normality

5. Examples

6. Monte Carlo Results

Appendix

References

Tables 1 – 3

The Suntory Centre
Suntory and Toyota International Centres
for Economics and Related Disciplines
London School of Economics and Political
Science

Discussion Paper
No. EM/00/400
July 2000

Houghton Street
London WC2A 2AE
Tel.: 020-7405 7686

* We would like to thank the National Science Foundation for financial support.

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. The required rates differ for two frequently used demand models, and a small Monte Carlo study shows that the difference in asymptotic properties of the two models are reflected in the models' small sample properties. The differences impact directly on the computational burden of the two models.

Keywords: Choice models; Method of Moments; Multinomial; Random Coefficients; Vertical Model.

JEL Nos.: C13, C15, C35, L13.

© by the authors. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

1 Introduction

Applied microeconomic studies of markets are often interested in estimating parameters of demand (or production) functions from data on the quantity, price, characteristics and perhaps inputs of a set of products which interact in a given market. In these studies observations are typically product-specific, 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. Of course frequently the analyst has more information available than just product level data on a single market. Often, for example, data on a cross-section of markets, or data that follows a given market over time, will be available. Somewhat less frequently, at least in the study of demand systems, additional micro level data (data that matches individuals to the products they chose) is available. Even in these cases, however, the number of products seems to be a relevant limiting dimension. Thus in studies that have data on many markets, or a given market over time, the number of products at a given spatio-temporal point is frequently at least as large as (and is often *much* larger than) the number of markets or time periods; while in the demand studies where micro level data is available there are parameters whose properties depend on a product-specific disturbance whose impacts do not get averaged out as the number of consumers grow large [see Berry, Levinsohn, and Pakes (1998)].

This paper provides an asymptotic theory for estimators of the parameters of differentiated product demand systems when the number of products in a given market is relatively large. 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, λ , 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 product 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 which maximizes its utility. The model’s estimate of market shares (or choice probabilities), say $\sigma(\theta, x, \xi)$, are generated by simply adding up over the choices of the alternative consuming units. 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 . We also can calculate the model’s prediction for these choice probabilities at alternative values of θ , i.e. $\sigma(\cdot)$. The unobservables are implicitly defined by the system

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

which can be solved for ξ as a function of (θ, x, s) . The θ vector is estimated by method of moments, using the vector of sample moment conditions formed, for example, from a zero covariance restriction between some exogenous vector of instruments, z , and the unobserved characteristics. The moment

restriction is then

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

at $\theta = \theta_0$. Several econometric issues arise in this context. These include the dependence in the observations on $\xi_j(\theta, x, s)$ when θ is not at its true value and the “triangular array” problem that all the elements of s may change when an additional good is added (similarly, all the prices, which are elements of x , may also change.) A further problem turns out to be quite important when either the function $\sigma(\cdot)$ is an integral which is estimated by Monte Carlo simulation based on a finite number (ns) of simulation draws, or when the market shares, s , have sampling error. Then the disturbances generated by the simulation and sampling processes also impact on the distribution of the estimators and (as we shall show) their impacts can be quite large.

We begin with some quite general consistency and asymptotic normality results. Probably most interesting, however, is their specialization to two classes of models which have been used extensively in the literature. The limit theorems for these cases are developed for rates at which all three sources of randomness (the consumer sampling process, the simulation process, and the process generating the product characteristics) contribute to the variance in the limiting distribution of the estimators, as this allows us to evaluate their relative contributions. We consider first the random coefficients logit 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 $J \log J/ns$ converge to zero as J increases. For asymptotic normality at rate \sqrt{J} in these cases we require J^2/n and J^2/ns to be bounded. That is, to obtain a CAN 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. This improves on rates reported in BLP. The other class of model’s we consider in detail is the “pure characteristic” model. Its theoretical lineage dates back at least to Hotelling (1929)’s 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 disturbance “logit” errors from each choice alternative. Berry and Pakes (1999) endow the pure characteristics model with an estimation algorithm analogous to the estimation algorithm provided in BLP for their model, 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 (one characteristic) pure characteristic model consistently we require only that n and ns increase at rate $\log J$, while for asymptotic normality we require only that J/n and J/ns 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. So the rate at which n and ns must grow for asymptotically normal parameter estimates from the pure characteristics model is less than (is the square root of) the rate at which they must grow to obtain asymptotically normal estimates for BLP's model. These results arise because differences in the nature of competition between the two models implies differences in the properties of $\partial\sigma(\cdot)/\partial\xi$. In particular 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 $\partial\sigma(\cdot)/\partial\xi$ goes to zero as the number of products increase. As we will show it is the inverse of this partial that determines 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 causes large changes in the computed value of ξ . 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. In the pure characteristic model then, a little bit of simulation or sampling error will have almost no 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 small monte carlo study which indicates that the difference is rather dramatic. This is one reason to expect the computational burden of the pure characteristic model to be less than the computational burden of BLP's model. Berry and Pakes (1999) show that the computational burden of obtaining the $\xi(\theta, x, s)$ 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. What this paper suggests is that to obtain well-behaved parameter estimates we will have to have much larger consumer samples and a much larger number of simulation draws if one uses BLP's system than if we use the pure characteristics model.

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. Section 4 provides the main mathematical details. In section 5 we discuss our leading examples. Section 6 contains the results of a small Monte Carlo study. The proofs are contained in the appendix. We use $\|A\| = \{\text{tr}(A'A)\}^{1/2}$ to denote the Euclidean norm of any $m \times n$ matrix A , \xrightarrow{p} to denote convergence in probability, and \implies 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

We consider a market with J competing products and an outside good. The vectors of product characteristics will be denoted by (ξ_j, x_{1j}) . The $\xi_j \in \mathbb{R}$ are characteristics which are not observed by the econometrician whereas the $x_{1j} \in \mathcal{X}_1 \subset \mathbb{R}^{d_1}$ are observed. We assume that the sequence $\{\xi_j, x_{1j}\}_{j=1}^J$ are independent and identically distributed (i.i.d.) [or, perhaps, exchangeable] draws, and, for the most part maintain the assumption that $E[\xi_j|x_1] = 0$ and $E[\xi_j^2|x_1] < \infty$ with probability one, where $x_1 = (x_{11}, \dots, x_{1J})$. 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 $E(\xi_j|x_{1j}) = 0$], we allow products to have additional characteristics, say $x_{2j} \in \mathcal{X}_2 \subset \mathbb{R}^{d_2}$, 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 $x_j = (x_{1j}, x_{2j})$, while $x = (x_1, \dots, x_J)$ and $\xi = (\xi_1, \dots, \xi_J)$. At times we will also need explicit assumptions on the process generating x_j .

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 $\omega(x, \xi, \lambda, \theta) : D \rightarrow \mathcal{S}_J$, where D is the appropriate product space, and \mathcal{S}_J is the $J + 1$ dimensional unit simplex, i.e.,

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

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

$$\sigma(\xi, \theta, P) = \int \omega(x, \xi, \lambda, \theta) dP(\lambda), \quad (3)$$

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 (θ^0, P^0) the true value of θ and P . We designate this vector by $s^0 = \sigma(\xi, \theta^0, P^0)$. Although P^0 is known, or rather assumed, 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 $\sigma(\xi, \theta, P^{ns})$, where P^{ns} is the empirical measure of some i.i.d. sample $\lambda_1, \dots, \lambda_{ns}$. For example,

$$\sigma(\xi, \theta, P^{ns}) = \int \omega(x, \xi, \lambda, \theta) dP^{ns}(\lambda) = \frac{1}{ns} \sum_{j=1}^{ns} \omega(x, \xi, \lambda_j, \theta).$$

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 neighbourhood 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 $k, j = 1, \dots, J$). 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 $s^0 > 0$, we have $s_\ell^0 \rightarrow 0$ as $J \rightarrow \infty$ 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 (θ_1, ξ_1) , that the function $\sigma(\xi_1, \theta_1, P^{ns})$ is constructed from ns 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, which are i.i.d. across i , while for any fixed (ξ, θ) , $\sigma_\ell(\xi, \theta, P^{ns}) - \sigma_\ell(\xi, \theta, P^0) = \frac{1}{ns} \sum_{is=1}^{ns} \varepsilon_{\ell, is}(\theta, \xi)$, where $\varepsilon_{\ell, is}(\theta, \xi)$ are independent and have mean zero and the function $\varepsilon_{\ell, is}(\theta, \xi)$ is bounded and continuous. Define the $J \times J$ matrices $V_1 = nE[(s^n - s^0)(s^n - s^0)'] = S - ss'$ and $V_2(\theta, \xi) = nsE[(\sigma(\xi, \theta, P^{ns}) - \sigma(\xi, \theta, P^0))(\sigma(\xi, \theta, P^{ns}) - \sigma(\xi, \theta, P^0))']$.*

We now outline the logic of the estimation procedure. Elsewhere, [Berry, Levinsohn, and Pakes (1995), and Berry and Pakes (1999)] 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

$$s - \sigma(\xi, \theta, P) = 0. \quad (4)$$

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 neighbourhood. The true value of ξ , say $\xi^0 \equiv \xi(\theta^0, s^0, P^0)$, is obtained as the solution to

$$s^0 - \sigma(\xi, \theta^0, P^0) = 0. \quad (5)$$

Define $\xi(\theta, s, P)$ as the unique vector that solves (5). Also define the instrument matrix $z = (z_1, \dots, z_J)$ whose components $z_q = z(x_{11}, \dots, x_{1J})_q \in \mathbb{R}^\ell$, where $z(\cdot)_q : (\mathbb{R}^{d_1})^J \rightarrow \mathbb{R}^\ell$, and $\ell \geq k$

(k is the dimension of θ), for $q = 1, \dots, J$. 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 all 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

$$G_J(\theta, s, P) \equiv \frac{1}{J} \sum_{j=1}^J z_j \xi_j(\theta, s, P). \quad (6)$$

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 $\sigma(\xi, \theta, P^{ns})$. Consequently, what we do is substitute an estimate of ξ , obtained as that value of ξ that sets $s^n - \sigma(\xi, \theta, P^{ns})$ to zero and denoted by $\xi(\theta, s^n, P^{ns})$, into (2) and minimize the resulting objective function. Thus our estimator of θ , say $\hat{\theta}$, is defined as any random variable that satisfies

$$\|G_J(\hat{\theta}, s^n, P^{ns})\| = \inf_{\theta \in \Theta} \|G_J(\theta, s^n, P^{ns})\| + o_p(1/\sqrt{J}). \quad (7)$$

The computation of $\hat{\theta}$ is discussed further in Berry, Levinsohn, and Pakes (1995).

3 Overview of the Main Results

The objective function we are minimizing, $\|G_J(\theta, s^n, P^{ns})\|$, 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^{ns} . Analogously there are three dimensions in which our sample can grow: as n , as ns , and as J grow large. Our limit theorems will allow various rates of growth for each dimension. Throughout we let $J \rightarrow \infty$ and make n and ns deterministic functions of J , i.e., we write $n(J)$ and $ns(J)$ and let $n(J), ns(J) \rightarrow \infty$ at some specified rate. If $n(J), ns(J) \rightarrow \infty$ at a fast enough rate, then the contribution from simulation and from the distribution of s^n 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, and make assumptions about the rate of growth of n, ns to ensure this (this will allow us to evaluate the contribution of simulation and sampling error to the asymptotic distribution of the estimator). Finally, note also that both s^n and $\sigma(\xi, \theta, P)$ 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 argument which provides the basic logic underlying the proof. The consistency argument is established by showing that:

- (i) $\sup_{\theta \in \Theta} \|G_J(\theta, s^n, P^{ns}) - G_J(\theta, s^0, P^0)\|$ 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 .

To obtain (i), we show that the sequence $\xi(\theta, s^n, P^{ns})$ is close to $\xi(\theta, s^0, P^0)$ uniformly in θ in probability in the sense that

$$\sup_{\theta \in \Theta} \frac{1}{J} \sum_{j=1}^J \{\xi_j(\theta, s^n, P^{ns}) - \xi_j(\theta, s^0, P^0)\}^2 \rightarrow_p 0. \quad (8)$$

Let $\varepsilon^n = s^n - s^0$ and $\varepsilon^{ns}(\theta) = \sigma[\xi(\theta, s^0, P^{ns}), \theta, P^{ns}] - \sigma[\xi(\theta, s^0, P^0), \theta, P^0]$. Since $\xi(\theta, s, P)$ solves (5), to prove (8) we will need that $\|\varepsilon^n + \varepsilon^{ns}(\theta)\| \rightarrow_p 0$ uniformly in θ . In addition, we put conditions on $\sigma(\xi, \theta, P)$ which ensure that the solution to (5) converges to $\xi(\theta, s^0, P^0)$ when $\|\varepsilon^n + \varepsilon^{ns}(\theta)\| \rightarrow_p 0$, and that $s^0 - \sigma(\xi, \theta, P^0)$ is sufficiently far away from zero when ξ is different from $\xi(\theta, s^0, P^0)$. The extra conditions we need here depend on the precise model generating market shares, so will formulate them in a general way, and then specialize to our two leading cases.

To establish (ii), we apply a standard law of large numbers to $G_J(\theta^0, s^0, P^0)$. However, we cannot guarantee convergence of $G_J(\theta, s^0, P^0)$ for $\theta \neq \theta^0$, because at $\theta \neq \theta^0$ equilibrium in the product market will ensure that the $\xi_j(s, \theta, P)$ are correlated. Instead, we use an asymptotic identification condition to bound the function $\|E[G_J(\theta, s^0, P^0)]\|$ uniformly away from zero when θ lies far enough away from θ^0 . The intuition underlying why we expect the identification condition to hold is transparent in the logit example, and generalizes to the other models we consider (a formal verification of this identification condition would require assumptions on the nature of equilibrium in the market being studied).

We turn next to asymptotic normality. We write

$$\xi(\theta, s^n, P^{ns}) = \xi(\theta, s^0, P^0) + \{\xi(\theta, s^n, P^{ns}) - \xi(\theta, s^0, P^{ns})\} + \{\xi(\theta, s^0, P^{ns}) - \xi(\theta, s^0, P^0)\}. \quad (9)$$

By assumption A1, that 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

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

Abbreviate $\sigma(\theta, s, P) = \sigma(\xi(s, \theta, P), \theta, P)$ and $H(\theta, s, P) = H(\xi(s, \theta, P), \theta, P)$. By the mean value theorem the last two terms in equation (9) can be expressed in terms of ε^n and $\varepsilon^{ns}(\theta)$, random variables whose distributions are given by A2; specifically,

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

where $r(\theta, s^n, P^{ns})$ is a remainder term whose properties we will detail below. Write

$$\begin{aligned} G_J(\theta, s^n, P^{ns}) &= \mathcal{G}_J(\theta) + \text{rem}_J(\theta) \\ \mathcal{G}_J(\theta) &= G_J(\theta, s^0, P^0) + \frac{1}{J} z' H^{-1}(\theta, s^0, P^0) \{\varepsilon^n - \varepsilon^{ns}(\theta)\}, \end{aligned} \quad (11)$$

where $\text{rem}_J(\theta) = \frac{1}{J} z' r(\theta, s^n, P^{ns})$. Our limit theorems work from this representation of $G_J(\theta, s^n, P^{ns})$. We proceed as follows. First, we will provide conditions under which $\sqrt{J} \text{rem}_J(\theta)$ converges uniformly to zero in probability. This requires a stochastic equicontinuity condition and a rate restriction on $n(J)$ and $ns(J)$. Next we assume that the function $E[G_J(\theta, s^0, P^0)]$ is approximately linear in θ near θ^0 , i.e., $E[G_J(\theta, s^0, P^0)] \simeq \Gamma(\theta - \theta^0)$ as $\theta \rightarrow \theta^0$, where Γ is full rank. Finally, we assume that $G_J(\theta, s^0, P^0)$ itself is stochastically equicontinuous. Combining these facts, it follows that $\sqrt{J}(\hat{\theta} - \theta^0)$ is the approximate minimizer of the quadratic function $\|\Gamma\sqrt{J}(\theta - \theta^0) + \sqrt{J}\mathcal{G}_J(\theta^0)\|$, which implies that¹

$$\sqrt{J}(\hat{\theta} - \theta^0) = -(\Gamma'\Gamma)^{-1}\Gamma'\sqrt{J}\mathcal{G}_J(\theta^0) + o_p(1).$$

The asymptotic distribution of $\sqrt{J}(\hat{\theta} - \theta^0)$ is therefore determined by the asymptotic distribution of the random vector $\sqrt{J}\mathcal{G}_J(\theta^0)$, which is the sum of three terms: $\sum_{j=1}^J z_j \xi_j / \sqrt{J}$, $z' H^{-1}(\theta^0, s^0, P^0) \varepsilon^n / \sqrt{J}$, and $z' H^{-1}(\theta^0, s^0, P^0) \varepsilon^{ns}(\theta^0) / \sqrt{J}$. These random variables are each asymptotically normal at rates determined by the growth of $n(J)$ and $ns(J)$; they are also mutually independent. 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.

To this end we seek conditions under which the second and third terms are asymptotically normal with zero mean and finite non-zero variance [the first term, $\sum_{j=1}^J z_j \xi_j / \sqrt{J}$, is clearly $O_p(1)$ and asymptotically normal under our conditions]. Note that the variance of $z' H^{-1}(\theta^0, s^0, P^0) \varepsilon^n / \sqrt{J}$ is $\Phi_1(J)$, while the variance of $z' H^{-1}(\theta^0, s^0, P^0) \varepsilon^{ns}(\theta^0) / \sqrt{J}$ is $\Phi_2(J)$, where

$$\Phi_1(J) = \frac{1}{nJ} z' H^{-1} V_1 H^{-1'} z \quad ; \quad \Phi_2(J) = \frac{1}{ns \times J} z' H^{-1} V_2 H^{-1'} z, \quad (12)$$

where $H = H(\theta^0, s^0, P^0)$ and $V_2 = V_2(\theta^0, \xi(s^0, \theta^0, P^0))$.

¹see Pakes and Pollard (1989) for this argument

So, it will be sufficient to find conditions on $n(J)$ and $ns(J)$ that guarantee these matrices are bounded [asymptotic normality follows]. We consider the term $\Phi_1(J)$ [similar comments apply to $\Phi_2(J)$]. The behaviour of the elements of the $J \times J$ matrix $H^{-1}(\theta, s^0, P^0)$ 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 $a' = (a_1, \dots, a_J) = z'H^{-1}(\theta^0, s^0, P^0)$ and suppose, without loss of generality, that z is a $J \times 1$ vector, we have

$$\Phi_1(J) = \frac{1}{nJ} \left[\sum_{j=1}^J a_j^2 s_j - \left(\sum_{j=1}^J a_j s_j \right)^2 \right], \quad (13)$$

since $V_1 = S - ss'$. The magnitude of the matrix Φ_1 depends on the sequence a and the sequence s . The term in square brackets can be considered to be the ‘variance’ of the vector (a_1, \dots, a_J) with respect to the multinomial like measure induced by the sequence of weights (s_1, \dots, s_J) [note that depending on the behaviour of s_0 , these weights do not necessarily sum to one].

There are three factors that influence the magnitude of this quantity. First, the magnitude of the sequence s_j^0 ; we shall assume that these quantities are all of order $O(1/J)$. Second, the rate at which the $a_j s$ grow or decline with J . Finally, the variability of the sequence $\{a_j\}$. In general, if $(a_1, \dots, a_J) = g(J)(a_{o1}, \dots, a_{oJ})$ for some function $g(\cdot)$ and sequence $\{a_{o1}, \dots, a_{oJ}\}$ having uniformly bounded variance, i.e., for all J , $\sum_{j=1}^J a_{oj}^2/J - (\sum_{j=1}^J a_{oj}/J)^2 \leq \bar{m} < \infty$, then for all J ,

$$\sum_{j=1}^J a_j^2 s_j - \left(\sum_{j=1}^J a_j s_j \right)^2 = g(J)^2 \left\{ \sum_{j=1}^J a_{oj}^2 s_j - \left(\sum_{j=1}^J a_{oj} s_j \right)^2 \right\} \leq m^* g(J)^2, \quad (14)$$

where $m^* < \infty$. This gives a global bound on the variance; 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 be

exploited to give tighter bounds on the variance. Specifically, in some cases we have

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

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

$$\begin{aligned} \sum_{j=1}^J a_j^2 s_j - \left(\sum_{j=1}^J a_j s_j \right)^2 &\simeq g(J)^2 \left[\sum_{j=1}^J s_j - \left(\sum_{j=1}^J s_j \right)^2 \right] \\ &= g(J)^2 [1 - s_0 - (1 - s_0)^2] \\ &= g(J)^2 s_0(1 - s_0). \end{aligned} \tag{15}$$

When the share of the outside alternative s_0 is $O(1/J)$, then (15) is $O(g(J)^2/J)$; when the share of the outside alternative is $O(1)$, it is the larger magnitude $O(g(J)^2)$. In the former case, we get a reduction in the magnitude of the variance.

In a subsequent section we investigate three examples. Our purpose is to verify the order of magnitude of the covariance matrix $\Phi_1(J)$ and to establish the precise rate of growth on $n(J), ns(J)$ required to achieve asymptotic normality. We achieve this by identifying the rate of growth and variability of the sequence $\{a_1, \dots, a_J\}$.

3.1 Structure of the Argument

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 (1999).

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 the following section we work out the implications of these theorems for our special cases. The reader who is not interested in the details of the mathematical argument should be able to proceed directly to section 5 and take away what is needed for the examples.

4 Consistency and Asymptotic Normality

4.1 Some Notation

We will work with the product space $\mathcal{S}_J \times \Theta \times \mathbf{P}$, and endow the marginal spaces with (pseudo) metrics ρ_P on \mathbf{P} , ρ_E on Θ , and ρ_s on \mathcal{S}_J . Let ρ be the corresponding metric on the product space. Some leading examples of metrics include: 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 , and the Euclidean metric on Θ , $\rho_E(\theta, \theta') = \|\theta - \theta'\|$. As for the metric on \mathcal{S}_J , we take

$$\rho_s(s^1, s^2) = \begin{cases} \max_{1 \leq j \leq J} \left| \frac{(s_j^1)^\alpha - (s_j^2)^\alpha}{(s_j^0)^\alpha} \right| & \text{if } 0 < \alpha \leq 1 \\ \max_{1 \leq j \leq J} |s_j^1 - s_j^2| & \text{if } \alpha = 0. \end{cases}$$

Note that the space \mathcal{S}_J and metric ρ_s both change with J , but the space can be embedded in the limiting space consisting of all sequences. The metric depends on the parameter α , which will in turn depend on the model being treated. The higher α is, the stronger is the metric. Define for each ϵ , the following neighbourhoods of θ^0, P^0 , and s^0 : $\mathcal{N}_{P^0}(\epsilon) = \{P : \rho_P(P, P^0) \leq \epsilon\}$ and $\mathcal{N}_{s^0}(\epsilon) = \{s : \rho_s(s, s^0) \leq \epsilon\}$, $\mathcal{N}_{\theta^0}(\epsilon) = \{\theta : \rho_E(\theta, \theta^0) \leq \epsilon\}$, and define the product neighbourhood $\mathcal{N}(\epsilon) = \mathcal{N}_{P^0}(\epsilon) \times \mathcal{N}_{s^0}(\epsilon)$. For each θ and any $\epsilon > 0$, define $\Xi(\theta; \epsilon) = \{\xi : \frac{1}{J} \|\xi - \xi(\theta, s^0, P^0)\|^2 \leq \epsilon\}$.

4.2 Consistency

We make several additional assumptions.

ASSUMPTION A3. *The random sequences s^n and P^{ns} are strongly consistent with respect to the corresponding metrics, i.e.,*

$$(a) \Pr \left[\lim_{n(J) \rightarrow \infty} \rho_s(s^n, s^0) = 0 \right] = 1 \quad ; \quad (b) \Pr \left[\lim_{ns(J) \rightarrow \infty} \rho_P(P^{ns}, P^0) = 0 \right] = 1. \quad (16)$$

Furthermore, we suppose that the true market shares satisfy

$$(c) \frac{1}{nJ^\alpha} \sum_{j=1}^J \frac{s_j^0(1-s_j^0)}{(s_j^0)^{2\alpha}} \rightarrow 0 \quad ; \quad (d) \frac{1}{ns \cdot J^\alpha} \sum_{j=1}^J \frac{s_j^0(1-s_j^0)}{(s_j^0)^{2\alpha}} \rightarrow 0.$$

When ρ_P is the uniform metric, A3(b) is the Glivenko-Cantelli result for the uniform convergence of the sample distribution function of identically distributed random variables. 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 s_ℓ^0 . Therefore, to verify this assumption requires restrictions on the growth rate of $n(J)$ and s_ℓ^0 . For example, suppose that $s_\ell^0 = O(1/J)$ for each ℓ , which 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 $\ell = 1, \dots, J$. This gives the pointwise rate of convergence; to obtain the sup-norm convergence rate [with respect to the pseudo-metric $\rho_s(s^1, s^2) = \max_{1 \leq \ell \leq J} |s_\ell^1 - s_\ell^2|/s_\ell^0$], we apply the Bonferroni and Bernstein inequalities [the latter is applicable because each s_ℓ^n is a sum of independent bounded random variables with expectation s_ℓ^0 , see Pollard (1989)] to find

$$\begin{aligned} \Pr \left[\max_{1 \leq \ell \leq J} \left| \frac{s_\ell^n - s_\ell^0}{s_\ell^0} \right| > \epsilon \right] &\leq \sum_{\ell=1}^J \Pr \left[\left| \frac{s_\ell^n - s_\ell^0}{s_\ell^0} \right| > \epsilon \right] \\ &\leq \sum_{\ell=1}^J \exp \left(-\frac{\epsilon^2}{2\text{var}(s_\ell^n/s_\ell^0) + 2\epsilon/ns_\ell^0} \right) \\ &\leq \sum_{\ell=1}^J \exp(-\epsilon^2 O(n/J)). \end{aligned} \quad (17)$$

A sufficient condition for (17) to decrease to zero is that $(J \log J)/n \rightarrow 0$, which implies (16)(a). Assumption A3(c) is implied by $J^\alpha/n \rightarrow 0$ when $s_\ell^0 = O(1/J)$ for each ℓ .

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 $\xi(\theta, s^n, P^{ns})$. In this context it is convenient to introduce the one to one (or bijective) transformation τ_J [the transformation is applied componentwise, i.e., $\tau_J(s) = (\tau_J(s_1), \dots, \tau_J(s_J))'$] and the $J \times 1$ vector $\psi_J(\xi, \theta; s, P) = \tau_J(s) - \tau_J(\sigma(\xi, \theta, P))$. We then define

$$\xi(\theta, s, P) = \arg \min_{\xi} ||\psi_J(\xi, \theta; s, P)||$$

for any θ, s, P . With this definition we are able to draw on the existing stock of asymptotic theory for optimization estimators to describe the properties of $\xi(\theta, s^n, P^{ns})$. 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

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

In the logit-like case, we use $\alpha = 1$ and $\tau_\alpha(x) = \log x$, while in the vertical 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 $\epsilon(\delta)$ such that with probability one*

$$\liminf_{J \rightarrow \infty} \inf_{\theta \in \Theta} \inf_{\|\xi - \xi(\theta, s^0, P^0)\| \geq \delta \sqrt{J}} \|\psi_J(\xi, \theta; s^0, P^0) - \psi(\xi(\theta, s^0, P^0), \theta; s^0, P^0)\| > \epsilon.$$

Assumption A6 is our “identification” condition [see 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*

$$\lim_{J \rightarrow \infty} \Pr \left[\inf_{\theta \notin N_{\theta^0}(\delta)} \|G_J(\theta, s^0, P^0) - G_J(\theta^0, s^0, P^0)\| \geq C(\delta) \right] = 1.$$

THEOREM 1 [Consistency] *Suppose that A1-A6 hold for some $\alpha \in [0, 1]$. Then, $\hat{\theta} \xrightarrow{p} \theta^0$.*

The proof is in the appendix.

4.3 Asymptotic Normality

Premultiplying (11) by \sqrt{J} , we get

$$\sqrt{J}G_J(\theta, s^n, P^{ns}) = \sqrt{J}\mathcal{G}_J(\theta) + \sqrt{J}\text{rem}_J(\theta). \quad (18)$$

We shall give conditions under which $\sqrt{J}\mathcal{G}_J(\theta^0)$ is asymptotically normal with bounded variance, while $\sqrt{J}\text{rem}_J(\theta^0)$ is of smaller order. The precise magnitude of the variance of $\sqrt{J}\mathcal{G}_J(\theta^0)$ is determined by the behaviour of the matrix $H(\theta^0, s^0, P^0)^{-1}$, an issue we will come back to below.

ASSUMPTION B1. θ^0 is in the topological interior of Θ .

ASSUMPTION B2. *For all θ in some $\delta > 0$ neighbourhood of θ^0*

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

where $\Gamma^J \rightarrow \Gamma$ as $J \rightarrow \infty$, and Γ has full (column) rank.

Define the stochastic process in (ξ, P, θ)

$$\nu_J(\xi, P, \theta) = \frac{1}{\sqrt{J}} z' H^{-1}(\xi, \theta, P) \{\varepsilon^n - \varepsilon^{ns}(\theta)\}, \quad (19)$$

where $\varepsilon^n = (\varepsilon_1^n, \dots, \varepsilon_J^n)'$ and $\varepsilon^{ns}(\theta) = (\varepsilon_1^{ns}(\theta), \dots, \varepsilon_J^{ns}(\theta))'$. 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

$$\begin{aligned} \nu_J(\xi, P, \theta) &= \sum_{i=1}^n Y_{Ji}(\xi, \theta, P) - \sum_{is=1}^{ns} Y_{J, is}^*(\xi, \theta, P), \\ Y_{Ji}(\xi, \theta, P) &= \frac{1}{n\sqrt{J}} \sum_{j=1}^J a_j(\xi, \theta, P) \varepsilon_{ji} \quad ; \quad Y_{J, is}^*(\xi, \theta, P) = \frac{1}{ns\sqrt{J}} \sum_{j=1}^J a_j(\xi, \theta, P) \varepsilon_{j, is}(\theta), \end{aligned} \quad (20)$$

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

ASSUMPTION B3. Let $Y_{Ji} = Y_{Ji}(\xi(\theta^0, s^0, P^0), \theta^0, P^0)$ and $Y_{J, is}^* = Y_{J, is}^*(\xi(\theta^0, s^0, P^0), \theta^0, P^0)$, and suppose that $\lim_{J \rightarrow \infty} E(z' \xi \xi' z / J) = \Phi_0$ and that

$$(a) \lim_{J \rightarrow \infty} ns E [Y_{J, is}^* Y_{J, is}^{*'}] = \Phi_1 \quad ; \quad (b) \lim_{J \rightarrow \infty} n E [Y_{Ji} Y_{Ji}'] = \Phi_2 \quad (21)$$

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

$$(a) ns E \left[\|Y_{J, is}^*\|^{2+\delta} \right] = o(1) \quad ; \quad (b) n E \left[\|Y_{Ji}\|^{2+\delta} \right] = o(1). \quad (22)$$

The reason for condition (22) is that as J increases the distribution of the random variables $Y_{J, is}^*$ 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 $ns(J)$. To do so we shall have to make more detailed assumptions about z and H . The next section will provide the details of our two leading cases.

Finally, we use a stochastic equicontinuity condition on the stochastic process (19) to handle remainder terms. This approach to asymptotics is now well established in econometrics, see the recent survey of Andrews (1994).

ASSUMPTION B4. The process $\nu_J(\xi, P, \theta)$ is stochastically equicontinuous at (s^0, P^0) , that is, for all $\varsigma > 0$ and $\eta > 0$, there exists an $\epsilon > 0$ such that

$$\limsup_{J \rightarrow \infty} \Pr \left[\sup_{(\theta, \xi, P) \in \Theta \times \Xi(\theta; \epsilon) \times \mathcal{N}_{P^0}(\epsilon)} \|\nu_J(\xi, P, \theta) - \nu_J(\xi(s^0, P^0), P^0, \theta)\| > \varsigma \right] < \eta.$$

In B4 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). Condition B2 requires only that the expectation of $G_J(\theta^0, s^0, P^0)$ be differentiable. This condition is similar to condition (iii) of Theorem 3.3 in Pakes and Pollard (1989). 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 $\hat{\theta}$.

THEOREM 2. [Asymptotic Normality] *Suppose that A1-A6 and B1-B4 hold for some α . Then,*

$$\sqrt{J}(\hat{\theta} - \theta^0) \implies N \left\{ 0, (\Gamma' \Gamma)^{-1} \Gamma' \left(\sum_{i=0}^2 \Phi_i \right) \Gamma (\Gamma' \Gamma)^{-1} \right\}.$$

The proof is in the appendix. Standard errors can be constructed in the usual way. One can improve efficiency by taking the weighted norm with weights proportional to the inverse of the covariance matrix $\sum_{i=0}^2 \Phi_i$. However, since we are only dealing with the demand subsystem here, such an estimator would only be efficient in a limited information sense.

5 Examples

We consider two classes of models. In the first class substitution patterns are relatively diffuse. Specifically, the distribution of utilities for one good, conditional on the utilities of all the other goods, has full support. A leading example here is the logit model. The logit is much criticized as an empirical model, and so we also consider the random coefficients logit model of Berry, Levinsohn, and Pakes (1995). In the second class of models substitution patterns are more concentrated and depend on preferences for a finite number of product characteristics. Leading examples here are the horizontal model of Hotelling (1929) and the vertical model of Shaked and Sutton (1982). Berry and Pakes (1999) provide an estimation algorithm for this class of models.

We show that the asymptotic behaviour of our estimator can be quite 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. We explain how these rate conditions are related to the underlying assumptions about the nature of substitution between products. In the sequel all calculations will be done at true values for s, P and so we will dispense with the unnecessary superscripts wherever possible.

5.1 Diffuse Substitution Patterns

We give a quite general result which covers pretty much all cases with “diffuse” substitution patterns (as in the logit and random coefficient logit examples introduced below). Our bounds on n and ns for the logit based cases are sharper than those given in this general result.

ASSUMPTION C1. *Suppose that the instruments z_i are bounded and*

$$0 < \underline{c} \leq \inf J \frac{\partial \sigma_j(\xi, \theta, P)}{\partial \xi_j} \leq \sup J \frac{\partial \sigma_j(\xi, \theta, P)}{\partial \xi_j} \leq \bar{c} < \infty,$$

where the infimum and supremum are taken over $j = 1, \dots, J$, $\xi \in \Xi(\theta; \epsilon)$ and $(s, P) \in \mathcal{N}(\epsilon)$ for some positive shrinking sequences of constants $\epsilon(J)$. Also, suppose that for some $d > 0$,

$$J[1 - \Delta(J)] \geq d, \tag{23}$$

where

$$\Delta(J) = \sup \sum_{k \neq j, k=1}^J \frac{|\partial \sigma_k(\xi, \theta, P) / \partial \xi_j|}{\partial \sigma_j(\xi, \theta, P) / \partial \xi_j},$$

where the supremum is taken over $j = 1, \dots, J$, $\xi \in \Xi(\theta; \epsilon)$, $\theta \in \Theta$, and $(s, P) \in \mathcal{N}(\epsilon)$.

The quantity $\Delta(J)$ measures the degree of diagonal dominance in the sensitivity of the share matrix to the ξ , i.e., the fraction of people who will shift to the “outside alternative” or “zero option” as ξ decreases.

COROLLARY. *Suppose that A1-A6 hold for some α . Then, sufficient conditions for (21) and (22) to hold are given by*

$$\limsup_{J \rightarrow \infty} \frac{J^3}{n(J)} < \infty \quad ; \quad \limsup_{J \rightarrow \infty} \frac{J^3}{ns(J)} < \infty. \quad (24)$$

For consistency, the weaker rate condition

$$\lim_{J \rightarrow \infty} \frac{J^2}{n(J)} = 0 \quad ; \quad \lim_{J \rightarrow \infty} \frac{J^2}{ns(J)} = 0. \quad (25)$$

suffices. These are essentially the conditions reported in BLP. As we will see our logit case (even our random coefficients logit) specializes the assumptions in C1 further and in turn requires weaker rate conditions on n and ns .

5.1.1 The logit model

This is the most familiar of the models with diffuse substitution patterns. Partly as a result, it is quite easy to illustrate the roles of the assumptions in deriving the limit properties of the estimator for this model. Consequently we provide more detail on verification of our assumptions for this case.

Consider the utility function

$$u_{ij} = \delta_j + \epsilon_{ij} = x_j \theta + \xi_j + \epsilon_{ij}, \quad (26)$$

where x_j is a vector of observed product characteristics, ξ_j is the unobserved characteristic and ϵ_{ij} is an i.i.d. extreme value error term. The systematic utility is given by δ_j .

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 $u_{i0} = \epsilon_{i0}$. Taking expectations with respect to the distribution of ϵ , we obtain that the market share function is

$$\sigma_j = \frac{e^{x_j \theta + \xi_j}}{1 + \sum_k e^{x_k \theta + \xi_k}}, \quad j = 1, \dots, J.$$

One consequence of this is that when $s_j = O(1/J)$ for all j , the outside alternative is $O(1/J)$ also. Note also that there is an analytic expression for the observed product characteristic

$$\xi_j = \ln(s_j) - \ln(s_0) - x_j \theta. \quad (27)$$

This relationship holds for all s, ξ . The resulting estimator of θ is just a linear instrumental variable regression of $\ln(s_j^n) - \ln(s_0^n)$ on x_j .

Since this estimator has an explicit form one can derive directly the conditions needed for consistency. We look for a sufficient condition for the mean square consistency of $\xi_j(s^n, \theta)$. We have

$$\frac{1}{J} \sum_{j=1}^J \{\xi_j(s^n, \theta) - \xi_j(s^0, \theta)\}^2 = \frac{1}{J} \sum_{j=1}^J \left\{ \ln\left(1 + \left(\frac{s_\ell^n - s_\ell}{s_\ell}\right)\right) - \ln\left(1 + \left(\frac{s_0^n - s_0}{s_0}\right)\right) \right\}^2.$$

It is straightforward to verify that this quantity is $O_p(J/n)$ under our conditions. Therefore, the consistency condition here is that $J \log J/n \rightarrow 0$.

In the logit case it is easy to derive the elements of the inverse share matrix analytically. Let H denote the $J \times J$ share matrix derivative evaluated at the truth, then H has elements

$$\frac{\partial \sigma_j}{\partial \xi_k} = \begin{cases} s_j(1 - s_j) & k = j \\ -s_k s_j & \text{if } k \neq j, \end{cases} \quad (28)$$

i.e., $H(s) = S - ss'$, where $S = \text{diag}(s)$, with $s = (s_1, \dots, s_J)'$. Therefore, $H^{-1} = S^{-1} + ii'/(1 - i's)$, where $i = (1, \dots, 1)'$. Letting e_k be the vector with k^{th} element equal to one and all other elements zero, we get

$$a_k = z'H^{-1}e_k = \left(\frac{z_1}{s_1}, \dots, \frac{z_J}{s_J} \right) e_k + \frac{\sum_{j=1}^J z_j}{s_0} i' e_k = J^2 \left\{ \frac{\bar{z}}{J s_0} + O(1/J) \right\}, \quad (29)$$

where $\bar{z} = J^{-1} \sum_{\ell=1}^J z_\ell$, that is, $g(J) = J^2$ and $a_{oj} = a_{ok}$ for all j, k in (14).

We now verify the conditions needed for asymptotic normality; in particular, (21), (22), and B4. Note that (29) shows that $nE[|Y_{Ji}|^2] = O(J^2/n)$, which implies that (21(a)) is satisfied provided $\sup_J J^2/n < \infty$. We now verify (22). Note that

$$\left| \sum_{j=1}^J a_j \varepsilon_{ji} \right| \leq \max_{1 \leq j \leq J} |a_j| \sum_{j=1}^J |\varepsilon_{ji}| \leq cJ^2$$

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 (29) is true. Therefore,

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

for any δ . In conclusion, $nE[|Y_{Ji}|^{2+\delta}] = O(J^{3+3\delta/2} n^{-(1+\delta)}) = o(1)$, provided $3 + 3\delta/2 - 2(1 + \delta) < 0$, i.e., any $\delta > 2$ suffices. In conclusion, condition (22) is satisfied for any $\delta > 2$. In the fixed coefficient logits case there is no need for simulation and so we do not have a $\varepsilon^{ns}(\theta)$.

We next provide some guidance regarding the identification conditions A6 and B2. In the logit case we have

$$G_J(\theta) - G_J(\theta^0) = \left(\frac{1}{J} \sum_{j=1}^J z_j x_j' \right) (\theta - \theta^0).$$

Consequently, a sufficient condition for A6 is that for each $\varepsilon > 0$ there is an $J(\varepsilon)$ such that for any $J \geq J(\varepsilon)$ the matrix, $J^{-1} \sum_{j=1}^J z_j x_j'$ has rank k with probability at least $1 - \varepsilon$, because then

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

as required. 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.

Finally, we turn to the stochastic equicontinuity condition B4. 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 work with the process

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

where $H^{-1}(s) = S^{-1} - i i' / (1 - i' s)$. We verify assumption B4 in the appendix. Specifically, we show that

$$\nu_J(s^n) - \nu_J(s^0) = O_p(J^{3/2}/n), \quad (30)$$

so that the remainder terms are of smaller order than the leading variance terms, and B4 is satisfied when J^2/n is bounded.

5.1.2 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 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 the product characteristics x_j, ξ_j .

The market share for this model is given by

$$\sigma_j = \int \frac{e^{\delta_j + x_j \lambda}}{1 + \sum_k e^{\delta_k + x_k \lambda}} dP(\lambda) \equiv \int s_j(\lambda) dP(\lambda) \equiv E[s_j(\lambda)]. \quad (31)$$

Note that the integrand, $s_j(\lambda)$, is just the logit market share function evaluated at a particular value of the random coefficients. The derivatives of the market share function are

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

Unfortunately, there is no analytic inverse for this model and no easy expression (that we know of) for the inverse matrix H^{-1} . However, we can still characterize its properties sufficiently well to

ensure that the property (15) holds. We state the result for the leading case that $s_j = O(1/J)$ for $j = 0, 1, \dots, J$.

THEOREM 3. *For any vector z ,*

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

where $\mu_z = \lim_{J \rightarrow \infty} J^{-1} \sum_{j=1}^J z_j$ and $\bar{s}_0 = \lim_{J \rightarrow \infty} J \int s_0(\lambda) dP(\lambda)$.

This result is proved in the appendix. It says that each element of the sequence a in (13) has magnitude $O(J^2)$ and that the elements are identical to this order of magnitude. Therefore, the variance term (13) 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.

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.

5.2 Non Diffuse Substitution: The Vertical Model

We now consider models with a finite set of product characteristics as in Berry and Pakes (1999) (they drop the i.i.d. ε_{ij} terms in the logit and random coefficients logit). The simplest example 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 $\delta_j = 0$. Without loss of generality we take $\beta = 0$.

Suppose that the products are ordered as $0 = \delta_0 < \delta_1 < \delta_2 < \delta_3 < \dots$. Necessary and sufficient conditions for all goods to have positive market share are: that $0 = p_0 < p_1 < p_2 < \dots$, that $\Delta_j = (\delta_j - \delta_{j-1}) / (p_j - p_{j-1})$ are ordered as $\Delta_1 > \Delta_2 > \dots$, and that $F(\cdot)$ (the distribution of λ_i) is strictly increasing over the domain of the Δ 's. The market shares are then given by

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

Since the integrals defining the shares in this model are uni-dimensional we shall suppose that there is no simulation error. Still we need to show that $J^{-1} \|\xi_j(s^n) - \xi_j(s^0)\|^2 \rightarrow_p 0$. To this end note that

$$\xi_j(\Delta(s)) = \sum_{l=1}^j (p_l - p_{l-1}) \Delta_l(s),$$

where it is understood that $p_0 = 0$. Since $\Delta_j(s) = F^{-1}(1 - \sum_{l=0}^{j-1} s_l)$, we have

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

Recall that $F(\cdot)$ is strictly increasing (so its inverse satisfies a Lipschitz condition) and assume that whatever equilibrium is established $\max_{j \leq J} (p_j - p_{j-1}) = c < \infty$. Then for any $\epsilon > 0$

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

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

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 $\alpha_1 = f(\Delta_1)/p_1$, $\alpha_2 = f(\Delta_2)/(p_2 - p_1), \dots, \alpha_J = f(\Delta_J)/(p_J - p_{J-1})$, it can be shown that

$$H = \begin{pmatrix} \alpha_1 + \alpha_2 & -\alpha_2 & 0 & \dots & 0 \\ -\alpha_2 & \alpha_2 + \alpha_3 & \ddots & 0 & 0 \\ 0 & \ddots & \ddots & -\alpha_{J-1} & 0 \\ \vdots & 0 & -\alpha_{J-1} & \alpha_J + \alpha_{J-1} & -\alpha_J \\ 0 & 0 & 0 & -\alpha_J & \alpha_J \end{pmatrix}. \quad (32)$$

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 column sums are zero apart from the first row and column. Therefore,

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 $J \rightarrow \infty$ (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 $k = 1, \dots, J$,

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

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

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

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, \dots, a_J) with respect to the multinomial (s_1, \dots, s_J) or equivalently the multinomial $(1/J, \dots, 1/J)$. In fact,

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

which is finite provided $\frac{1}{J} \sum_{k=1}^J p_k^2$ is finite and $\min_{1 \leq \ell \leq J} f(\Delta_\ell) \geq \underline{m} > 0$. We assume both of these conditions in what follows (although one can use a similar argument to obtain different rates when

they do not.² Note that the requirement that $\frac{1}{J} \sum_{k=1}^J p_k^2 < \infty$ does impose conditions both on the way we add products as J increase and on the nature of the pricing equilibrium, but those conditions seem reasonable.

Given the lower bound \underline{m} and that the price sequence has a finite second moment, (14) holds with $g(J) = J$. Therefore, the covariance matrix $\Phi_1(J)$ is of order J/n . That is, in this case, we obtain consistency if n and ns increase at any rate faster than $\log J$, while the asymptotic normality result holds with all three terms contributing provided n and ns grow like J . We do not know whether one can improve on our inequality.

Now consider the multifactor model

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

where x_j is a vector of dimension $k - 1$. Note that for each given β_i , we have a one factor model as before except that some goods will not be chosen depending on the value of β_i . The resulting H is a symmetric band matrix with positive diagonal elements and negative off-diagonals. In fact, it is irreducible and diagonally dominant: its first k_0 row sums are $\Delta_1, \dots, \Delta_{k_0} > 0$, while the remaining row sums are zero. This implies that H is strictly positive definite and is an M-matrix, i.e., all elements of H^{-1} are positive just like in the one factor case [see Fiedler (1986, Chapter 5)]. We conjecture that the magnitude of $\Phi_1(J)$ is also J/n for any k , so that the rate of convergence in the k -factor case is the same as in the one factor model, but we *do not know* this for a fact.

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 idiosyncratic 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.

²When the argument can be extended to cases where f is not bounded away from zero, the crude bounding argument given above will imply a slower rate of convergence [i.e., $g(J)$ is larger] because now $\min_{1 \leq \ell \leq J} f(\Delta_\ell)$ converges to zero at some rate. We can think of Δ_j as being the order statistics from a sample produced by F . For example, if F is the folded normal, then $\max_j \Delta_j$ grows like $\sqrt{2 \log J}$ and $\min_j f(\Delta_j) = f(\max_j \Delta_j)$ shrinks to zero like $\exp(-(\sqrt{2 \log J})^2) = J^{-2}$. If F is Pareto with $1 - F(x) = x^{-\theta}$ for $x \geq 1$ and $\theta > 0$, then $\max_j \Delta_j$ grows like $J^{1/\theta}$ and $f(\max_j \Delta_j)$ shrinks to zero like $J^{-(\theta+1)/\theta}$.

6 Monte Carlo Results

In this section we discuss some Monte Carlo results for very simple versions of our models. First, we start with a simple logit, where the market shares are observed with sampling error. Then we discuss a random coefficients logit (with simulation error in the computed shares) and finally a pure vertical model.

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

$$\delta_j = x_j\beta + \xi_j, \tag{33}$$

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 draws used to calculate the “observed market share”. 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 empirical standard deviation (not the standard deviation of the mean) while the third row gives the mean squared error. 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.³ Note that the zero-share products are likely to have particularly low values of δ , so we expect a downward bias in the estimated slope coefficient.

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.⁴ We can always simulate positive predicted shares and so there is no sample selection problem. In this case, we can consider smaller values of n , but because the computational burden is higher we do not include a set of results for J^2 .

³We did not deal with this problem in our theoretical analysis above, but it might be a problem for some datasets built from small samples of consumers.

⁴I.e. we are assuming that the observed shares are aggregated over a very large number of consumers.

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

$$u_{ij} = \delta_j + \sigma\nu_i x_{j2} + \epsilon_{ij},$$

where ν is standard normal, σ 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 draws. The market shares are calculated by taking n draws from the distribution of the random coefficient ν . 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 gives summarizes the estimates of σ , the variance of the random coefficient on the non-constant x . The results are consistent with the idea that the estimation routine performs badly at very low values of simulation draws and that variances decrease in the number of products J . Note that the bias seems to increase in J holding n fixed – the bias is high when n is small relative to J . Remember that our consistency results above require both J and n to increase and that in the “logit-like” case, n must increase fairly rapidly.

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 included in the estimation routine at little cost.⁵

The exact vertical model considers a utility function of

$$u_{ij} = \delta - \mu\nu_i p_j, \tag{34}$$

where δ is “quality”, ν_i is the marginal disutility of a price increase and μ is a parameter of the model. To keep the random coefficient in an easy one-parameter family, we assume that ν_i is drawn from the unit exponential distribution, so that μ (set equal to one in the experiments) is the mean disutility of a price increase. In fact, μ 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 modelled 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 .⁶

⁵In 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 δ s and where the inversion routine cannot handle zero shares.

⁶In the vertical model, all shares will be positive if price increases “fast enough” in quality.

The results in Table 3 summarize the estimates of β_2 , the slope coefficient on x in the quality equation. These results are strikingly consistent with the idea that, given the vertical model, even a small number of simulation draws will produce decent estimates. Even with share data constructed from 50 consumers, the estimation gives decent results, although for fixed, small n the variance is not decreasing much in J . In contrast, with large n the variance declines at very close to the expected rate of \sqrt{J} .

In practice, Table 3 gives some hope to the practical implementation of more complicated “pure random coefficient” models (i.e., those without i.i.d. ϵ 's.) While such models may be computational more difficult than traditional random coefficients models, they may be relatively robust with respect to simulation error and so a relatively small number of simulation draws may be necessary when computing market shares.

A Appendix

PROOF OF THEOREM 1. We show that

$$\sup_{\theta \in \Theta} \frac{1}{J} \|\xi(\theta, s^n, P^{ns}) - \xi(\theta, s^0, P^0)\|^2 = o_p(1), \quad (35)$$

which implies that

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

i.e., that

$$\sup_{\theta \in \Theta} \|G_J(\theta, s^n, P^{ns}) - G_J(\theta, s^0, P^0)\| = o_p(1). \quad (36)$$

The result (35) follows from the following argument. First,

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

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

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

because

$$\frac{1}{nJ^\alpha} \sum_{j=1}^J \frac{s_j^0(1 - s_j^0)}{(s_j^0)^{2\alpha}} = o(1)$$

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

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

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

$$\sup \|\tau_J(\sigma(\xi, s^0, P^0) + \varepsilon^{ns}(\theta)) - \tau_J(\sigma(\xi, s^0, P^0))\| = o_p(1)$$

follows by similar arguments.

When $\|\xi - \xi(\theta, s^0, P^0)\| \geq \delta\sqrt{J}$, we have $\inf_{\theta \in \Theta} \|\psi_J(\xi, \theta; s^0, P^0)\| \geq \varepsilon$ by assumption A6. This implies that $\|\xi(\theta, s^n, P^{ns}) - \xi(\theta, s^0, P^0)\|^2/J = o_p(1)$ by standard arguments. Consequently, we now assume that (36) holds and hence

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

Now 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 $C^*(\delta) = C(\delta) - \varepsilon^*/2 > 0$. ■

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.

For each fixed θ , we use a Taylor series approximation to $\xi(\theta, s^n, P^{ns}) - \xi(\theta, s^0, P^{ns})$ and to $\xi(\theta, s^0, P^{ns}) - \xi(\theta, s^0, P^0)$. Specifically, by the intermediate value theorem

$$\begin{aligned} 0 &= \sigma(\xi(\theta, s^n, P^{ns}), \theta, P^{ns}) - s^n \\ &= \sigma(\xi(\theta, s^0, P^{ns}), \theta, P^{ns}) - s^n + \frac{\partial \sigma(\bar{\xi}, \theta, P^{ns})}{\partial \xi'} \{\xi(\theta, s^n, P^{ns}) - \xi(\theta, s^0, P^{ns})\}, \end{aligned} \quad (37)$$

where $\bar{\xi}$ are intermediate between $\xi(\theta, s^n, P^{ns})$ and $\xi(\theta, s^0, P^{ns})$. Thus using the facts that $\sigma(\xi(\theta, s^0, P^{ns}), \theta, P^{ns}) = s^0$ and that for any $\xi \in \Xi(\theta; \varepsilon)$ the matrix $\partial \sigma(\xi, \theta, P^{ns}) / \partial \xi'$ is invertible, we can write

$$\xi(\theta, s^n, P^{ns}) - \xi(\theta, s^0, P^{ns}) = - \left\{ \frac{\partial \sigma(\bar{\xi}, \theta, P^{ns})}{\partial \xi'} \right\}^{-1} \varepsilon^n. \quad (38)$$

Likewise

$$\xi(\theta, s^0, P^{ns}) - \xi(\theta, s^0, P^0) = - \left\{ \frac{\partial \sigma(\underline{\xi}, \theta, P^{ns})}{\partial \xi'} \right\}^{-1} \varepsilon^{ns}(\theta), \quad (39)$$

where $\underline{\xi}$ are intermediate between $\xi(\theta, s^0, P^{ns})$ and $\xi(\theta, s^0, P^0)$. Therefore,

$$\sqrt{J} \text{rem}_J(\theta) = - \frac{1}{\sqrt{J}} z' [H(\bar{\xi}, \theta, P^{ns})^{-1} - H(\theta, s^0, P^0)^{-1}] \varepsilon^n - \frac{1}{\sqrt{J}} z' [H(\underline{\xi}, \theta, P^{ns})^{-1} - H(\theta, s^0, P^0)^{-1}] \varepsilon^{ns}(\theta).$$

We must establish that $\sqrt{J} \text{rem}_J(\theta) = o_p(1)$ uniformly. We just show that

$$\left\| \frac{1}{\sqrt{J}} z' \{H(\bar{\xi}, \theta, P^{ns})^{-1} - H(\theta, s^0, P^0)^{-1}\} \varepsilon^n \right\| = o_p(1). \quad (40)$$

The proof for the second part is similar and is omitted.. By assumption A3 there exists a sequence ϵ_J such that $\Pr\{\rho_s(s^n, s^0) \geq \epsilon_J\} \rightarrow 0$ and $\Pr\{\rho_P(P^{ns}, P^0) \geq \epsilon_J\} \rightarrow 0$. Then, notice that for any $\eta > 0$,

$$\begin{aligned} \Pr[\|\nu_J(\bar{\xi}, P^{ns}, \theta)\| > \eta] &\leq \Pr[\|\nu_J(\bar{\xi}, P^{ns}, \theta)\| > \eta, (s^n, P^{ns}) \in \mathcal{N}(\epsilon_J)] + \Pr[s^n \notin \mathcal{N}_{s^0}(\epsilon_J)] \\ &\quad + \Pr[P^{ns} \notin \mathcal{N}_{P^0}(\epsilon_J)] \\ &\leq \Pr[\|\nu_J(\bar{\xi}, P^{ns}, \theta)\| > \eta, (s^n, P^{ns}) \in \mathcal{N}(\epsilon_J)] + o(1) \\ &\leq \Pr[\sup_{(\xi, P) \in \Xi(\theta; \epsilon) \times \mathcal{N}_{P^0}(\epsilon)} \|\nu_J(\xi, P, \theta)\| > \eta] + o(1) \\ &\leq \Pr[\|\nu_J(\xi(s^0, P^0), P^0, \theta)\| > \eta/2] + o(1) \\ &\quad + \Pr[\sup_{(\xi, P) \in \Xi(\theta; \epsilon) \times \mathcal{N}_{P^0}(\epsilon)} \|\nu_J(\xi, P, \theta) - \nu_J(\xi(s^0, P^0), P^0, \theta)\| > \eta/2] \\ &= \Pr[\|\nu_J(\xi(s^0, P^0), P^0, \theta)\| > \eta/2] + o(1), \end{aligned}$$

by the stochastic equicontinuity condition B4. Finally, for each fixed θ , $\xi \in \Xi(\theta; \epsilon)$, and $P \in \mathcal{N}_{P^0}(\epsilon)$, by Markov's inequality we have for any $\eta > 0$,

$$\begin{aligned} \Pr \left[\left\| \frac{1}{J} z' H(\xi, \theta, P)^{-1} \varepsilon^n \right\| > \eta \right] &\leq \frac{E \left\| \frac{1}{J} z' H(\xi, \theta, P)^{-1} \varepsilon^n \right\|^2}{\eta^2} \\ &= \frac{n E [\|Y_{Ji}(\xi, \theta, P)\|^2]}{\eta^2} = o(1) \end{aligned}$$

by assumption B3 in Theorem 2.

We next show that

$$\left[\text{var} \left(c' \sqrt{J} \mathcal{G}_J(\theta^0) \right) \right]^{-1/2} c' \sqrt{J} \mathcal{G}_J(\theta^0) \quad (41)$$

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 $\text{var} (c' T_{J\ell})^{-1/2} c' T_{J\ell}$, $\ell = 1, 2, 3$, 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

$$(c' E \{ \text{var}(\xi|z) z z' \} c)^{-1/2} c' J^{-1/2} z' \xi(\theta^0, s^0, P^0) \implies N(0, 1).$$

Condition (22) 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 $c' \sum_{is=1}^{ns} Y_{Jis}^*$ are asymptotically normal. ■

COROLLARY. The proof of this corollary uses the following lemma:

LEMMA 1. *Suppose that A1 and C1 hold. Let $\partial\sigma(\xi, \theta, P)/\partial\xi' \equiv C(\xi, \theta, P) - A(\xi, \theta, P)$, where $C(\xi, \theta, P) = \text{diag}[\partial\sigma_j(\xi, \theta, P)/\partial\xi_j]$. Assumption A1 insures that both $A(\xi, \theta, P)$ and $C(\xi, \theta, P)$ are nonnegative for all $\xi \in \Xi(\theta; \epsilon)$, θ , and P . Then for each finite J , $[\partial\sigma(\xi, \theta, P)/\partial\xi']^{-1}$ exists, is differentiable in ξ and θ , and satisfies*

$$H^{-1}(\xi, \theta, P) = C^{-1}B,$$

where $B = (b_{rj}) = [I - AC^{-1}]^{-1}$. (We have eliminated the dependence of these matrices on (ξ, θ, P) for notational convenience). Furthermore,

$$B \geq 0 \text{ (element by element)}$$

$$\sum_{r=1}^J b_{rj} \leq [1 - \Delta(J)]^{-1}, \text{ for } j = 1, \dots, J.$$

PROOF OF LEMMA 1. By assumption A1, C^{-1} exists, so if $[I - AC^{-1}]^{-1}$ exists the fact that $(C - A)C^{-1}[I - AC^{-1}]^{-1} = I$ proves that $C - A$ inverse exists. Let $Q = AC^{-1}$ and $B = \lim_{t \rightarrow \infty} \sum_{p=0}^t Q^p$ (since $Q \geq 0$, the limit exists and is nonnegative). If B is finite, then: $B(I - Q)$ is finite and equals I ,

i.e., $[I - AC^{-1}]^{-1}$ exists and is nonnegative. Let $q_j^t = \sum_{r=1}^J q_{rj}^t$ and $q^t = \max_{1 \leq j \leq J} q_j^t$, where $Q = (q_{rj})$. Then, if we can show that $q^t \leq \Delta(J)^t$, since $\Delta(J) < 1$, we will have $\sum_{t=0}^{\infty} q^t \leq [1 - \Delta(J)]^{-1}$. That is, if $q^t \leq \Delta(J)^t$, then B satisfies all conditions of the lemma.

The proof of $q^t \leq \Delta(J)^t$ is by induction on t . It is true by assumption for $t = 1$. Assume it is true for t . Then

$$q_j^{t+1} = \sum_{r=1}^J \sum_{v=1}^J q_{rv}^t q_{vj} = \sum_{v=1}^J \left\{ \sum_{r=1}^J q_{rv}^t \right\} q_{vj} \leq \sum_{v=1}^J \Delta(J)^t q_{vj} \leq \Delta(J)^{t+1}.$$

■

We now return to the proof of the corollary. Write

$$a_j = (z'H^{-1})_j = (z'C^{-1}B)_j = \sum_{\ell=1}^J z_\ell \frac{b_{\ell j}}{c_{\ell \ell}},$$

and therefore,

$$|a_j| \leq \frac{\max_{1 \leq \ell \leq J} |z_\ell|}{\min_{1 \leq \ell \leq J} c_{\ell \ell}} \sum_{\ell=1}^J b_{\ell j} \leq \frac{z_{\sup}}{cd} J^2.$$

Therefore, applying the argument of section 3, we obtain the bound

$$\Phi_1(\xi, \theta, P) \leq c \frac{J^3}{n}$$

for some finite constant c . In conclusion, the covariance matrix of $J^{-1}z'H^{-1}\varepsilon^n$ is bounded by some constant matrix times J^3/n . This bound holds for all $\xi \in \Xi(\theta; \epsilon)$, $\theta \in \Theta$, and $(s, P) \in \mathcal{N}(\epsilon)$. Similar calculations apply to the term $J^{-1/2}z'H^{-1}\varepsilon^{ns}(\theta)$. Therefore, (25) is established. For the asymptotic normality result, we take $\delta = 2$ in (22). Recall that the moment generating function of the centred multinomial vector $(\varepsilon_1, \dots, \varepsilon_J)$ is

$$\varphi(t_1, \dots, t_J) = \left(\sum_{j=1}^J s_j e^{t_j} \right) \left(e^{-\sum_{j=1}^J s_j t_j} \right),$$

where t_1, \dots, t_J are dummy arguments. We have to calculate

$$E \left[\left(\frac{1}{n\sqrt{J}} \sum_{j=1}^J a_j \varepsilon_j \right)^4 \right] = \frac{1}{n^4 J^2} \sum_{j=1}^J a_j^4 E[\varepsilon_j^4] + \dots + \frac{1}{n^4 J^2} \sum_{j=1}^J \sum_{l=1}^J \sum_{k=1}^J \sum_{i=1}^J a_j a_l a_k a_i E[\varepsilon_j \varepsilon_l \varepsilon_k \varepsilon_i],$$

4 distinct indices

where a_j satisfy (29). The first and last terms are

$$\begin{aligned} \frac{1}{n^4 J^2} \sum_{j=1}^J a_j^4 E[\varepsilon_j^4] &= O\left(\frac{1}{n^4 J^2} \times J \times J^8 \times \frac{1}{J}\right) = O\left(\frac{J^6}{n^4}\right) \\ -\frac{3}{n^4 J^2} \sum_{j=1}^J \sum_{l=1}^J \sum_{k=1}^J \sum_{i=1}^J a_j a_l a_k a_i s_j s_l s_k s_i &= O\left(\frac{1}{n^4 J^2} \times J^4 \times J^8 \times \frac{1}{J^4}\right) = O\left(\frac{J^6}{n^4}\right), \end{aligned}$$

4 distinct indices

while the other terms are no larger than this in magnitude. In conclusion,

$$nE[|Y_{Ji}|^4] = O\left(\frac{J^6}{n^3}\right) = o(1),$$

so that condition (22) is satisfied. ■

PROOF OF (30). It suffices to show that for any random sequence $s(n)$ converging to s^0 we have $\|\nu_J(s(n)) - \nu_J(s^0)\| \rightarrow_p 0$. We shall take $s(n) = s^n$ and show that $R_n = \nu_J(s^n) - \nu_J(s^0) = o_p(1)$, where

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

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

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

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

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

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

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

Similar calculation applies to R_{n22} . ■

PROOF OF THEOREM 3. Let $T, T_J : \mathbf{P} \rightarrow \mathbb{R}$, where

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

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

$$\frac{1}{J^2} |z'H(P)^{-1}e_k - z'H(P_m)^{-1}e_k| \leq \frac{1}{J^2} \|z'H(P_m)^{-1}\| \|H(P) - H(P_m)\| \|H(P)^{-1}e_k\| \quad (42)$$

and the following bounds (obtained below)

$$\|z'H(P_m)^{-1}\| \leq O(J^{5/2}) \quad (43)$$

$$\|H(P)^{-1}e_k\| \leq O(J^2) \quad (44)$$

$$\|H(P) - H(P_m)\| \leq O(1/m^{(1-\eta)/2} J^{1/2}), \quad (45)$$

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

PROOF OF (44). Writing $H(P)^{-1} = C^{-1}B$, we have that $H(P)^{-1}e_k = \left(\frac{b_{1k}}{c_1}, \dots, \frac{b_{Jk}}{c_J}\right)$ whose (squared) norm is

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

because the elements of B and C are known to be positive. This establishes (44). The verification of (43) is given below.

PROOF OF (45). 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,

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

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

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

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

$$\begin{aligned} &\Pr \left[\max_{1 \leq j \neq k \leq J} \left| J^2 \int s_j(\lambda) s_k(\lambda) \{dP_m(\lambda) - dP(\lambda)\} \right| > \frac{c(\log m)^r}{m^{1/2}} \right] \\ &\leq \sum_{j \neq k} \Pr \left[\left| J^2 \int s_j(\lambda) s_k(\lambda) \{dP_m(\lambda) - dP(\lambda)\} \right| > \frac{c(\log m)^r}{m^{1/2}} \right] \\ &= O(J^2) \exp[-c^*(\log m)^{2r}] \end{aligned} \tag{46}$$

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

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

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

$$m^{(1-\eta)/2} \max_{1 \leq j \neq k \leq J} \left| J^2 \int s_j(\lambda) s_k(\lambda) \{dP_m(\lambda) - dP(\lambda)\} \right| \longrightarrow 0 \tag{47}$$

with probability one. Similarly,

$$m^{(1-\eta)/2} \max_{1 \leq j \leq J} \left| J \int s_j(\lambda)(1 - s_j(\lambda)) \{dP_m(\lambda) - dP(\lambda)\} \right| \longrightarrow 0 \tag{48}$$

with probability one. In conclusion, the discrete mixture of logits well approximates any random coefficient logit matrix. Specifically, (45) follows because

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

with probability one for large m, J by (47) and (48).

PROOF OF II. Consider the discrete mixture

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

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

$$\frac{1}{J^2} z' H^{-1} e_k = \frac{\frac{1}{J} \sum_{j=1}^J z_j}{J \frac{1}{m} \sum_{\ell=1}^m s_0^\ell} + O(1/J), \quad k = 1, \dots, J, \quad (49)$$

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

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

$$z' H^{-1} e_k = m \{z' D^{-1} e_k - z' D^{-1} U (I + V' D^{-1} U)^{-1} V' D^{-1} e_k\} \quad (50)$$

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

$$[(I + V' D^{-1} U)^{-1}]_{ij} = \frac{1 + O(1/J)}{\sum_{\ell=1}^m s_0^\ell [1 + O(1/J)]} \quad (51)$$

for all $i, j = 1, \dots, m$. 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 (49).

We have

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

and

$$I + V' D^{-1} U = \begin{pmatrix} 1 - \sum_{j=1}^J \frac{(s_j^1)^2}{d_j} & - \sum_{j=1}^J \frac{s_j^1 s_j^2}{d_j} & \cdots & - \sum_{j=1}^J \frac{s_j^1 s_j^m}{d_j} \\ - \sum_{j=1}^J \frac{s_j^2 s_j^1}{d_j} & 1 - \sum_{j=1}^J \frac{(s_j^2)^2}{d_j} & & - \sum_{j=1}^J \frac{s_j^2 s_j^m}{d_j} \\ \vdots & & \ddots & \vdots \\ - \sum_{j=1}^J \frac{s_j^m s_j^1}{d_j} & - \sum_{j=1}^J \frac{s_j^m s_j^2}{d_j} & \cdots & 1 - \sum_{j=1}^J \frac{(s_j^m)^2}{d_j} \end{pmatrix}. \quad (52)$$

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

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

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

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

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

$$a = \begin{bmatrix} -\{1 - \sum_{\ell=1}^{m-1} a_{1\ell}\} \\ \vdots \\ -\{1 - \sum_{\ell=1}^{m-1} a_{m-1, \ell}\} \end{bmatrix} ; \quad \delta = \begin{bmatrix} J s_0^1 \\ \vdots \\ J s_0^{m-1} \end{bmatrix},$$

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

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

$$\det\left(X + \frac{E}{J}\right) = \frac{1}{J} \sum_{j,k=1}^m \frac{\partial \det(X)}{\partial x_{jk}} e_{jk} + \frac{1}{2J^2} \sum_{j,k,l,r=1}^m \frac{\partial^2 \det(X)}{\partial x_{jk} \partial x_{lr}} e_{jk} e_{lr} + \dots \quad (53)$$

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

$$x_{jk}^{Adj} = \det(A), \quad (54)$$

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 (53). Also, note that

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

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 (53) is of order m/J^2 and

$$\det(I + V'D^{-1}U) = \det(A) \sum_{\ell=1}^m s_0^\ell [1 + O(1/J)]. \quad (55)$$

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

$$z_{jk}^{Adj} = \det(A)[1 + O(1/J)], \quad j \neq k, \quad (56)$$

which implies (51) holds. ■

PROOF OF (54). We use the fact that determinants are invariant to certain linear transformations and also that the matrix X has the following property

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

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 (56). 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., AND A. PAKES, (1999): "The Pure Characteristics Model of Demand," Mimeo, Yale University.
- [5] BERRY, S., LEVINSOHN, J., AND A. PAKES, (1995): "Automobile prices in market equilibrium," *Econometrica* **63**, 841-890.
- [6] BILLINGSLEY, P. (1986): *Probability and Measure*. 2nd Edition. John Wiley. New York.
- [7] DIEUDONNE, J. (1969): *Foundations of Modern Analysis*. Academic Press: New York.
- [8] FIEDLER, M., (1986): *Special Matrices and their applications in numerical mathematics*. Kluwer: Dordrecht.
- [9] GOLUB, G.H., AND C.F. VAN LOAN (1989): *Matrix Computations*. The Johns Hopkins University Press: Baltimore.
- [10] HANSEN, L. (1982): "Large Sample Properties of Generalized Method of Moments Estimators", *Econometrica* 50, 1029-1054.

- [11] 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.
- [12] HOTELLING, H. (1929): "Stability in Competition," *Economic Journal* 39, 41-57.
- [13] HOTZ, V.J. AND R. MILLER (1993): "Conditional choice probabilities and the estimation of dynamic models," *Review of Economic Studies* 60, 497-529.
- [14] MAMMEN, E. (1989): "Asymptotics with increasing dimension for robust regression with applications to the bootstrap," *The Annals of Statistics* 17, 382-400.
- [15] MARCUS, M. AND H. MINC (1964): *A survey of matrix theory and matrix inequalities*. Dover, N.Y.
- [16] ORTEGA, J.M. AND W.C. RHEINBOLDT (1970): *Iterative solution of nonlinear equation in several variables*. Academic Press, London.
- [17] 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.
- [18] PAKES, A., AND S. OLLEY (1995): "A Limit Theorem for a Smooth Class of Semiparametric Estimators," *Journal of Econometrics* 65, 295-332.
- [19] PAKES, A., AND D. POLLARD (1989): "Simulation and the asymptotics of optimization estimators," *Econometrica* 57, 1027-1057.
- [20] POLLARD, D. (1984): *Convergence of Stochastic Processes*. Springer-Verlag. New York.
- [21] 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.
- [22] 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, 403-417.
- [23] SHAKED, A. AND J. SUTTON (1982). "Relaxing price competition through product differentiation," *The Review of Economic Studies*, 49, 3-13.
- [24] 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

# Sim Draws	# of Products				
	10	25	50	100	200
500	0.941 (0.362) [0.133]	0.798 (0.209) [0.084]	0.778 (0.137) [0.068]	0.633 (0.086) [0.142]	0.518 (0.076) [0.238]
1000	0.997 (0.426) [0.180]	1.013 (0.255) [0.065]	0.974 (0.149) [0.023]	0.934 (0.120) [0.019]	0.882 (0.077) [0.020]
2000	1.023 (0.500) [0.248]	1.046 (0.224) [0.052]	0.998 (0.138) [0.019]	0.976 (0.123) [0.015]	0.923 (0.089) [0.014]
10 <i>J</i>	0.685 (0.406) [0.262]	0.728 (0.214) [0.120]	0.768 (0.132) [0.071]	0.921 (0.110) [0.018]	0.916 (0.088) [0.015]
<i>J</i> ²	0.615 (0.358) [0.275]	0.857 (0.200) [0.060]	1.021 (0.139) [0.020]	1.022 (0.101) [0.010]	1.015 (0.077) [0.006]
∞	1.027 (0.376) [0.141]	0.997 (0.242) [0.058]	0.995 (0.133) [0.017]	1.007 (0.094) [0.009]	1.008 (0.073) [0.005]

Notes: Standard Errors in (·) and Mean Squared Errors in [·].

Table 2:
Monte Carlo Estimates for the Random Coefficients Logit
True Value of the Parm is 1
1000 Monte Carlo Repetitions

# Sim Draws	# of Products		
	10	50	100
10	1.194 (0.982) [1.002]	1.218 (0.512) [0.310]	*
50	1.025 (0.645) [0.416]	1.039 (0.311) [0.098]	1.241 (0.495) [0.303]
100	0.982 (0.674) [0.454]	1.013 (0.271) [0.073]	1.037 (0.209) [0.045]
500	0.998 (0.633) [0.400]	1.008 (0.255) [0.065]	1.015 (0.181) [0.033]
10J	0.982 (0.674) [0.454]	1.008 (0.255) [0.065]	1.018 (0.181) [0.033]

Notes: Standard Errors in (·) and Mean Squared Errors in [·].
*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

# Sim Draws	# of Products				
	10	25	50	100	200
50	1.023 (0.494) [0.150]	1.022 (0.373) [0.294]	1.011 (0.349) [0.010]	0.997 (0.321) [0.013]	1.013 (0.302) [0.044]
100	1.005 (0.426) [0.061]	1.010 (0.303) [0.018]	1.005 (0.257) [0.009]	1.002 (0.244) [0.013]	1.009 (0.217) [0.000]
500	0.993 (0.371) [0.120]	0.998 (0.223) [0.097]	1.001 (0.176) [0.024]	1.005 (0.142) [0.006]	1.007 (0.123) [0.032]
1000	1.01 (0.361) [0.006]	0.99 (0.227) [0.094]	1.00 (0.162) [0.001]	1.00 (0.118) [0.007]	1.00 (0.097) [0.009]
10 <i>J</i>	1.018 (0.440) [0.304]	1.014 (0.253) [0.043]	1.008 (0.175) [0.031]	0.998 (0.120) [0.002]	0.996 (0.085) [0.041]
<i>J</i> ²	0.998 (0.423) [0.002]	0.998 (0.227) [0.026]	1.000 (0.153) [0.044]	1.002 (0.105) [0.000]	1.000 (0.074) [0.055]
∞	0.997 (0.364) [0.088]	0.999 (0.214) [0.054]	0.999 (0.141) [0.004]	1.001 (0.101) [0.000]	0.997 (0.072) [0.000]

Notes: Standard Errors in (·) and Mean Squared Errors in [·].