Large Market Asymptotics for Differentiated Product Demand Estimators with Economic Models of Supply

Timothy B. Armstrong*
Stanford University

February 15, 2012

Abstract

IO economists often estimate demand for differentiated products using data sets with a small number of large markets. By modeling demand as depending on a small number of product characteristics, one might hope to obtain increasingly precise estimates of demand parameters as the number of products in a single market grows large. In this paper, I address the question of consistency and asymptotic distributions of IV estimates of demand in a small number of markets as the number of products increases in some commonly used demand models under conditions on economic primitives. I show that, under the common assumption of a Bertrand-Nash equilibrium in prices, product characteristics lose their identifying power as price instruments in the limit in many of these models, giving inconsistent estimates in these cases. I find that consistent estimates can still be obtained for many of the cases I consider, but care must be taken in modeling demand and choosing instruments. For cases where consistent estimates can be obtained, I provide sufficient conditions for consistency and asymptotic normality of estimates of parameters and counterfactual outcomes. A monte carlo study confirms that the asymptotic results provide an accurate description of the behavior of estimators in market sizes of practical importance.

*Email: timothya@stanford.edu. Thanks to Han Hong and Liran Einav for guidance and many useful discussions, and Ariel Pakes, John Lazarev and participants at the Stanford demand estimation reading group and third year seminar and the 2010 Econometric Society World Congress for useful comments and criticism. All remaining errors are my own.

1 Introduction

Many empirical studies of markets with differentiated products use data on a relatively small number of markets, each with many products, to estimate demand elasticities. For example, in their application to automobile demand, Berry, Levinsohn, and Pakes (1995) use data on 20 markets, each with about 100 products. For demand models where the number of parameters grows with the number of product characteristics rather than the number of products, one might expect a small number of markets with a large number of products to give good estimates of demand.

In this paper, I examine large market asymptotics for IV estimators for some models commonly used in the differentiated product demand literature under conditions on economic primitives. Whether a particular choice of instruments provides consistent estimates in a particular model depends on its limiting relationship with equilibrium prices, which arises endogenously from a sequence of pricing games. I show that, in several commonly used models, the dependence of prices on product characteristic instruments through markups disappears at a fast enough rate that estimators based on these IVs are inconsistent in a large market setting, even though these estimators are consistent with a large number of small markets. In particular, this is the case with the logit and random coefficients logit models in a large market setting with many firms.

These negative results hold for the random coefficients logit model when a new idiosyncratic logit error is added for each new product without changing the distribution of the other random coefficients. However, the limiting dependence of prices on product characteristics can be restored by more careful modeling of the way the distribution of preferences changes when new products are added. For the nested logit model, I show that the dependence of markups on product characteristics remains in the limit if products are added into new nests, which amounts to adding new random coefficients for new nest indicator variables as the market size grows. In this case, the markup of a given product will depend on the characteristics of products in the same nest in the limit, but not characteristics of other products, so care must still be taken in choosing instruments. The findings described above hold with a large number of small firms. I also examine markets with a small number of large firms and find that, in this case, the identifying power of product characteristic instruments depends on having variation in average product quality across, rather than within, firms.

The results in this paper give clear practical guidelines for addressing identification issues in empirical work that uses these models. First, researchers should be careful to specify a model and a set of instruments for estimating it such that the supply side model being used for policy counterfactuals is consistent with demand being well identified by these instruments with the data at hand. This paper provides conditions for verifying whether this is the case, and provides examples of models used in practice where this is the case, and where it is not. These results can be used as a guide in specifying a demand model and a set of instruments such that the instruments will have good identifying power under the supply side model being used for counterfactual analysis. Second, since identification may be weak, researchers should test for identification, or do inference in a way that is robust to weak identification or lack of identification. Despite the complicated nonlinear nature of some of the models considered in this paper, one can test for identification using, for example, the test proposed by Wright (2003) for nonlinear GMM models.

It is important to emphasize that testing for identification alone without following the first recommendation does not adequately address the issues that this paper brings up. Suppose that a researcher tests whether product characteristic instruments identify a demand model and finds evidence that they strongly identify the model. The researcher then uses a supply side model to perform policy counterfactuals about, say, prices after a merger. If the supply side model is such that, because of issues pointed out in this paper, variation in markups is constrained to be small, the researcher will find that the merger will have little effect on prices. But this is based on a model that is inconsistent with the finding that the product characteristic instruments are strongly correlated with prices that the researcher found when testing for identification. The very fact that the model is well identified means that it must be misspecified in an economically important way that leads to incorrect conclusions in the counterfactuals. The same issues will arise if, rather than testing for identification, the author constructs confidence regions for the policy counterfactual based on tests that are robust to weak identification or lack of identification. The full model of supply and demand constrains such tests to have power close to their size unless it is misspecified in a way that fundamentally biases counterfactual estimates. The results in this paper can be used to specify a model of demand that will not suffer from these problems. See section 8 for more on how the results of this paper can be used to guide empirical work.

While the results in this paper show that the identifying power of product characteristic instruments in large markets is a subtle question, I find that cost shifter instruments have identifying power in large markets in most of the cases I consider. Since the variation in markups settles down in the limit, instrumental variables that shift marginal cost account for a non negligible amount of the variation in prices in the limit, and the sample correlation between IV estimates and covariates converges to a positive definite matrix under suitable conditions on the model primitives.

In addition to deriving asymptotic approximations for estimates of model parameters, I also derive asymptotic approximations for estimates of counterfactual quantities involving large markets. Since the counterfactual quantity is a different function of the parameters for each market size, this does not follow immediately from the delta method and asymptotic distributions of parameter estimates. To overcome this, I use an additional step in which the sequence of estimated counterfactuals is approximated by a fixed function of the parameter estimates.

To examine how well these asymptotic results describe the behavior of these estimators in market sizes of practical importance, I provide a monte carlo study. I find that the asymptotic theory developed in this paper is backed up by the results of the monte carlos. With a large number of small firms, the logit and random coefficients logit models lead to product characteristic instruments performing poorly in large markets, while cost shifter instruments give increasingly precise estimates as the number of products increases. With a small number of large firms, product characteristic instruments do well when some firms have an overall advantage in terms of product quality, but not when firms draw product characteristics from the same data generating process. It is worth noting that, while previous papers have reported monte carlos for these demand models (Berry, 1994; Berry, Linton, and Pakes, 2004), this paper is, to my knowledge, the first to perform monte carlos in which prices are formed by computing an equilibrium in a supply side game. While the previous monte carlos are useful for answering other questions, solving an economic model of supply for prices is necessary for examining the implications of supply side models for the power of price instruments.

To my knowledge, the only other authors to examine the behavior of IV estimates of differentiated product demand models in a small number of large markets are Berry, Linton, and Pakes (2004). They provide high level conditions for consistency and asymptotic normality, but leave open the question of which estimators are consistent and asymptotically normal under which economic models of price setting and conditions on economic primitives. My results complement their paper by providing asymptotic results under conditions on economic primitives, and considering estimation of counterfactual outcomes that change with the size of the market. I also abstract from sampling error in market shares and simulation error in computing IV estimates, while Berry, Linton, and Pakes (2004) focus on these issues. Bajari and Benkard (2005) consider estimation of a class of differentiated product demand models that includes one of the models I consider using data from a small number of large markets, but they propose a different method of estimation.

This paper is also related to the literature on weak instruments (see Stock, Wright, and Yogo, 2002, for a survey of this literature). That literature uses sequences of underlying distributions in which the correlation of instruments with endogenous variables shrinks with the sample size to get asymptotic approximations that better approximate finite sample distributions. This paper shows that such sequences arise endogenously from equlibrium prices when asymptotics are taken in the number of products per market in a certain class of models. To my knowledge, this is the first result in which weak instruments arise endogenously from equilibrium behavior in a sequence of pricing games. The fact that the power of instruments is tied to demand through equilibrium pricing in a supply side model can be used to inform decisions about how to model demand.

Despite these parallels with the weak IV literature, it is important to emphasize that, in the settings considered here that lead to weak instruments, the correlation of IVs with prices goes to zero even more quickly than in the typical case considered in the weak IV literature. In the settings considered here that lead to weak identification, the correlation of the instruments with prices will decrease so quickly that weak instrument robust tests will have trivial power. This contrasts with the typical case considered in the weak IV literature in which the correlation goes to zero, but is still large enough relative to the sample size that the instruments provide some amount of useful exogenous variation, although this case would likely arise if the results of this paper were extended to allow the sample size and market size to increase at the same time.

In addition to the literatures on weak instruments and on estimation of discrete choice models of demand, this paper relates to theoretical results on oligopoly pricing in markets where demand is characterized by a discrete choice model. Existence of equilibrium in some of the pricing games I consider follows from arguments in Caplin and Nalebuff (1991) and Vives (2001) or similar methods. There is also a literature examining how restrictions on demand elasticities in discrete choice models place restrictions on the possible outcomes of empirical applications. Bajari and Benkard (2003), Ackerberg and Rysman (2005), and others have argued that logit type errors lead to a bias toward finding high valuations of new goods. Bajari and Benkard (2003) also point out that logit errors lead to markups being bounded away from zero as the number of firms increases under Bertrand competition, making the logit model a bad choice for modeling markets where goods are homogenous enough that one expects perfect competition with a large number of products. Some of my findings add to this body of knowledge. For example, I show that, while markups are bounded away from zero in the random coefficients logit model, the dependence of the markup on product characteristics disappears unless the distribution or number of random coefficients changes with the number of products. However, the main focus of this paper is on implications for the consistency of instrumental variables estimators, rather than how the models I consider restrict the possible outcomes of empirical applications.

The paper is organized as follows. Section 2 describes the class of models being studied. Section 3 gives a nontechnical discussion of the nature of the results for a special case. Section 4 presents a simple sufficient condition for asymptotic equivalence of IV estimators in different models. Section 5 derives the asymptotic behavior of equilibrium prices and IV estimates in some commonly used demand models by verifying the conditions of section 4. Section 6 addresses the issue of predicting counterfactual outcomes in large markets using demand estimates. Section 7 presents the results of a monte carlo study. Section 8 discusses implications for applied work. Section 9 concludes.

2 The Model

In this section, I describe the class of models and estimators considered in this paper and define some notation that will be used later. The models and much of the notation follow Berry (1994).

The researcher observes data from a single market with J products labeled 1 through J and an outside good labeled 0, and M consumers. Each product j has a price p_j and a vector of other characteristics observed to the researcher, $x_j \in \mathbb{R}^K$, and an unobserved variable ξ_j, which can be interpreted as a combination of unobserved product characteristics and aggregate preference shocks. In addition, each individual consumer i has consumer specific unobserved components of demand ε_{ij} and ζ_i, which are iid across consumers.

Consumer i's utility for the jth product is given by $u_{ij} = u(x_j, p_j, \xi_j, \varepsilon_{ij}, \zeta_i)$ for some function u. Each consumer buys the product for which utility is the highest, and no consumer buys more than one product. Rather than individual purchasing decisions, we observe aggregate market shares, including the proportion of consumers who make no purchase (the share of the outside good). These come from aggregating purchasing decisions over the εs and ζs of all consumers. As is common in the literature, I simplify the analysis by assuming that the number of consumers is large enough to ignore sampling variation in market shares from realizations of the εs and ζs, so that the market share of good j, $s_j(x, \xi, p)$, is equal to the population probability of choosing good j conditional on x, ξ, and

p: $s_j(x,\xi,p) = E_{\varepsilon,\zeta} I(u_{ij} > u_{ik} \text{ all } k \neq j)$.

In the models I consider here, utility can be written in the following form for some parameters (α,β,σ): $u_{ij} = x_j'\beta - \alpha p_j + \xi_j + g_i(\varepsilon_{ij},\zeta_i,x_j,p_j,\sigma)$. Thus, utility can be separated into a linear part that that does not depend on individual preferences, and a function that may depend on individual preferences, and the unobserved aggregate demand shock only enters the first part. This is the case in the models considered in Berry (1994). Following the literature, I denote the linear part by $\delta_j \equiv x_j'\beta - \alpha p_j + \xi_j$. Since shares only depend on ξ through δ, we can write them as $s_j(\delta,x,p,\sigma)$.

On the producer side, there are F firms labeled 1 through F. Firm f produces the set of goods $\mathcal{F}_f \in \{1,\ldots,J\}$. I use the notation $\vec{p}_f$ to denote the vector of prices for products produced by firm f and similar notation for vectors containing other variables produced by a given firm. Firm f has cost function $C_f(\vec{q}_f,\vec{x}_f,\vec{\eta}_f)$ for a vector of cost shocks $\vec{\eta}_f$. I use the notation MC_j to denote the marginal cost of producing good j (the derivative of C_f with respect to q_j). As is common in the literature, I assume a Nash-Bertrand equilibrium in prices. With Nash-Bertrand competition and an interior best response, the first order conditions for firm f choosing the price of good j are

$$\begin{aligned}
&\frac{\partial}{\partial p_j}\left\{\sum_{k\in\mathcal{F}_f} p_k \cdot M s_k(x,p,\xi) - C_f(\vec{q}_f,\vec{x}_f,\vec{\eta}_f)\right\} \\
&= M\cdot\sum_{k\in\mathcal{F}_f} p_k \frac{\partial}{\partial p_j} s_k(x,p,\xi) + M\cdot s_j(x,p,\xi) - \sum_{k\in\mathcal{F}_f} MC_k \cdot M\frac{\partial}{\partial p_j} s_k(x,p,\xi) = 0 \\
&\Leftrightarrow \sum_{k\in\mathcal{F}_f} (p_k - MC_k)\frac{\partial}{\partial p_j} s_k(x,p,\xi) + s_j(x,p,\xi) = 0. \qquad (1)
\end{aligned}$$

For single product firms, this simplifies to

$$p_j = MC_j - \frac{s_j(x,p,\xi)}{\frac{d}{dp_j} s_j(x,p,\xi)}. \qquad (2)$$

When new products are added to the demand system, the equilibrium price will change, so that the equilibrium price and share of good j depend on the size of the market J. That is, even though x, η, and ξ are sequences, prices and market shares will actually be triangular arrays, so that a more precise notation for the equilibrium price of good j would be $p_{j,J}$. To avoid extra subscripts, I use p_j to denote the price of good j when the context is clear, using $p_{j,J}$ when clarification is needed.

In all of the models below, I will assume that the marginal cost of producing good j does not depend on the scale of production and takes a linear form. For the single product case, this amounts to assuming $MC_j = x_j'\gamma + \eta_j$ with $E(\eta_j|x) = 0$. The linearity assumption is made only to avoid identification coming purely from functional form. For this, it does not matter if γ and η have any structural interpretation as long as the linearity holds. If $E(MC_j|x)$ were nonlinear with mean utility assumed linear as above, one could estimate the model by instrumenting with higher moments of x. The assumption that marginal cost does not depend on quantity rules out instruments that only affect prices through changes in marginal cost from changes in production levels. Letting marginal cost depend on quantity could potentially make asymptotics in J more complicated since this requires thinking about how large M is relative to J to find the limiting behavior of $M \cdot s_j$. However, the results here could easily be extended to this case once the limiting behavior of M is decided on.

Finally, in all of the models below, I assume that the vector of unobserved demand shocks ξ is mean independent of observed product characteristics: $E(\xi|x) = 0$. This assumption provides the basis for the instrumental variables estimates considered in this paper. While there are certainly cases where this assumption is questionable, I focus on asking when this assumption allows demand to be consistently estimated, rather than asking in which applications it is likely to hold.

2.1 Estimation

Suppose the share function $s(\delta, x, p, \sigma)$ is invertible in its first argument with inverse $\delta(s, x, p, \sigma)$. Then we might hope to estimate the model using the equation

$$\delta_j(s, x, p, \sigma) = x_j'\beta - \alpha p_j + \xi_j. \tag{3}$$

However, the parameter σ enters into a function with shares, which are endogenous. In addition, prices may be correlated with the unobserved ξ through at least two channels. First, ξ_j enters the markup $s(\delta, x, p, \sigma)/\frac{d}{dp_j}s(\delta, x, p, \sigma)$. Second, we may think that ξ_j is correlated with the unobserved component of marginal cost, η_j, if goods that are more desirable in unobserved ways are also more expensive to make in unobserved ways.

A solution that is commonly used in the literature is to use characteristics of other products as additional instruments. Since market shares depend on the product characteristics of all firms and prices depend on the product characteristics of all firms through the Bertrand markup $s(\delta, x, p, \sigma)/\frac{d}{dp_j}s(\delta, x, p, \sigma)$, it seems reasonable that moments based on characteristics of other products might be used to consistently estimate the parameters of the model. Suppose that we use some vector valued function $h_j(x_{-j})$ as excluded instruments. The parameter estimates minimize the GMM criterion function

$$\left\| \frac{1}{J} \sum_{j=1}^{J} (\delta_j(s,x,p,\sigma) - x_j'\beta + \alpha p_j) z_j \right\|_{W_J} \tag{4}$$

where $z_j = (x_j, h_j(x_{-j}))'$ and W_J is a positive definite weighting matrix. This results in the following linear IV formula for the estimates of β and α as a function of the estimate $\hat{\sigma}$ of σ:

$$\begin{pmatrix} \hat{\beta} \\ \hat{\alpha} \end{pmatrix} = \left(\left[\sum_{j=1}^{J} z_j \begin{pmatrix} x_j & -p_j \end{pmatrix} \right]' W_J \left[\sum_{j=1}^{J} z_j \begin{pmatrix} x_j & -p_j \end{pmatrix} \right] \right)^{-1}$$
$$\left[\sum_{j=1}^{J} z_j \begin{pmatrix} x_j & -p_j \end{pmatrix} \right]' W_J \sum_{j=1}^{J} z_j \delta_j(s,x,p,\hat{\sigma}).$$

3 A Preview of the Results

To get a feel for the results in this paper and how they are shown, consider the logit model with single product firms. Utility is given by

$$u_{ij} = x_j'\beta - \alpha p_j + \xi_j + \varepsilon_{ij} \equiv \delta_j + \varepsilon_{ij}$$

and ε_{ij} is distributed extreme value independently across i and j. The assumptions on utility lead to shares

$$s_j(x,p,\xi) = \frac{\exp(x_j'\beta - \alpha p_j + \xi_j)}{\sum_k \exp(x_k'\beta - \alpha p_k + \xi_k)}$$

which can be inverted to get ξ_j (normalizing the mean utility of the outside good 0 to zero)

$$\log s_j - \log s_0 = x_j'\beta - \alpha p_j + \xi_j. \tag{5}$$

The derivative of firm j's share with respect to j's price is $\frac{d}{dp_j} s_j(x,p,\xi) = -\alpha s_j(x,p,\xi)(1 - s_j(x,p,\xi))$, which gives the Bertrand pricing formula, Equation (2), as $p_j = MC_j + 1/(\alpha(1 - s_j))$.

If each firm's market share goes to zero in the limit (this will be the case if, for example, prices and product characteristics are bounded), inspection of the markup formula shows that it will converge to $1/\alpha$. This suggests that estimators will have the same asymptotic distribution when applied to data generated with Bertrand pricing as they would if the data were generated from a market with constant markups. If this is the case, then estimators that use product characteristics of other firms as instruments for price will be inconsistent, since the excluded instruments are only correlated with prices through the markup term. On the other hand, cost shifters should still provide valid instruments.

In section 4, I provide a way of formally verifying the asymptotic equivalence of IV estimators in different models of price setting. With this machinery in place, the behavior of GMM estimates as the number of products grows in a single market can be examined by showing that they would have the same distribution under a data generating process in which markups are constant or only depend on a particular set of product characteristics. I include equivalence results for unidentified models, so that estimators can be shown to be inconsistent by comparing them to other inconsistent estimators. This provides formal justification for the argument that, since the logit markups converge to a constant in the Bertrand model and product characteristics of other firms cannot be used as instruments in the constant markup model, estimates that use these instruments will not be consistent in the Bertrand model. According to the results in the next section, this will be true as long as markups converge at a $\sqrt{J}$ rate or faster. I show that this convergence will be at a faster than $\sqrt{J}$ rate in section 5.

The convergence of the markup in the logit model to $1/\alpha$ also suggests that estimated counterfactual outcomes involving large markets with logit demand and Bertrand competition with single product firms will be close to estimates that compute the counterfactual outcomes under the assumption of a constant markup. If the estimation error is of a greater order of magnitude than the difference between the two ways of estimating the counterfactual outcome, the two estimates should be roughly the same for the purposes of inference using asymptotic approximations to sampling distributions. While this is problematic for most IO applications, not all of the models I consider have a constant markup in the limit. In section 6, I provide a framework for approximating the sampling distribution of estimates of counterfactual quantities when the counterfactual market has many products, and show how the nested logit model with many nests, a model in which markups do not converge to a constant, fits into this framework. For cases where the counterfactual exercise is still interesting in the limit, this gives asymptotic approximations that are valid as the size of the counterfactual market increases.

4 Equivalence Results for IV Estimators

In this section, I provide sufficient conditions for an IV estimator to have the same asymptotic distribution under two different models of price setting. Since I need these results to show that some estimators are inconsistent, they must apply to models that do not satisfy the rank condition for identification. The following set of assumptions handles this case as well as the case where the parameters are point identified.

Consider a linear IV model with instruments $z_j \in \mathbb{R}^k$, regressors $x_j \in \mathbb{R}^d$, parameter vector $\beta \in \mathbb{R}^d$, and an unobservable error term $\xi_j \in \mathbb{R}$. Let $y_j = x_j'\beta + \xi_j$. For the differentiated product demand models considered in this paper, we will verify these assumptions with $\delta_j(x, p, \sigma)$ in the place of y_j and the vector of prices and covariates ($(x_j', p_j)'$ in the notation of the rest of the paper) playing the role of x_j. The IV estimate with weighting matrix W_J is given by

$$\hat{\beta} \equiv \left(\left[\sum_{j=1}^{J} z_j x_j' \right]' W_J \left[\sum_{j=1}^{J} z_j x_j' \right] \right)^{-1} \left[\sum_{j=1}^{J} z_j x_j' \right]' W_J \left[\sum_{j=1}^{J} z_j y_j \right]. \tag{6}$$

Suppose that a central limit theorem applies to the sample means in this formula so that the following assumptions hold.

Assumption 1. *(i)* $\sqrt{J}\left(\frac{1}{J}\sum_{j=1}^{J} z_j x_j' - M_{zx}\right) \xrightarrow{d} Z_{zx}$ *for some* $k \times d$ *matrix* M_{zx}*, and a* $k \times d$ *random matrix* Z_{zx}*. (ii)* $\frac{1}{\sqrt{J}}\sum_{j=1}^{J} z_j \xi_j \xrightarrow{d} Z_{z\xi}$ *for a multivariate normal random vector* $Z_{z\xi}$*. (iii)* $W_J \xrightarrow{p} W$ *for some positive definite weighting matrix* W*.*

If these assumptions hold with M_{zx} a full rank matrix, $\hat{\beta}$ will be consistent and asymptotically normal. If M_{zx} is not full rank, $\hat{\beta}$ will not be consistent. The following theorem covers both cases. Part (i) is all that is needed to show that an estimator is inconsistent. Part (ii) is the standard consistency and asymptotic normality result for IV estimators in identified models.

Part (iii) derives the asymptotic behavior of possibly inconsistent IV estimates, and can be skipped by readers who are not interested in the behavior of inconsistent estimators. For the logit model discussed in the previous section, it is conceivable that using product characteristics as instruments would give an estimator that is inconsistent, but can be used to form tests with power against fixed alternatives, as in the weak instrument asymptotics of, for example, Stock and Wright (2000). This would be the case if markups converged to a constant at exactly a $\sqrt{J}$ rate. According to part (iii) of the following theorem, this will not be the case if markups converge at a faster than $\sqrt{J}$ rate, which I show to be true in section 5.

Theorem 1. *Under Assumption 1, we have the following.*

(i) $\hat{\beta}$ *is consistent if and only if* M_{zx} *is full rank.*

(ii) If M_{zx} *is full rank, then* $\hat{\beta}$ *is consistent and* $\sqrt{J}(\hat{\beta}-\beta) \xrightarrow{d} (M_{zx}'WM_{zx})^{-1}M_{zx}'WZ_{z\xi}$.

(iii) Let $d_2 = \operatorname{rank} M_{zx}$ *and* $d_1 = d - d_2$. *Let* H *be an invertible* $d \times d$ *matrix such that the first* d_1 *columns of* $M_{zx}H$ *are zero and split* H *into its first* d_1 *and last* d_2 *columns as* (H_1, H_2). *Define* $T_J = H^{-1}(\hat{\beta}-\beta)$ *with* T_{1J} *the first* d_1 *elements and* T_{2J} *the last* d_2 *elements. Then*

$$\begin{pmatrix} T_{1J} \\ \sqrt{J}T_{2J} \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \left((Z_{zx}H_1)'Q_{W,2}'WQ_{W,2}Z_{zx}H_1\right)^{-1}(Z_{zx}H_1)'Q_{W,2}'WQ_{W,2}Z_{z\xi} \\ \left((Ezx'H_2)'Q_{W,1}'WQ_{W,1}Ezx'H_2\right)^{-1}(Ezx'H_1)'Q_{W,1}'WQ_{W,1}Z_{z\xi} \end{pmatrix}$$

where $Q_{W,1}$ *is the* W *inner product projection matrix for the orthogonal complement of the column span of* $Z_{zx}H_1$ *and* $Q_{W,2}$ *is the* W *inner product projection matrix for the orthogonal complement of the column span of* $Ezx'H_2$.

Proof. See appendix. □

Suppose the conditions of the above theorem are known to hold for a model with a particular set of regressors, and we are interested in how the model behaves when we replace the regressors with another set of variables that converge to the regressors in the original model. If the difference between the two sets of regressors disappears quickly enough, the conditions will hold with the new regressors and the same asymptotic distributions, so that IV estimators will have the same asymptotic distribution in both models. This is true even for partially identified or unidentified models, where the IV estimates are not consistent. The following corollary states this formally in a way that will be useful for supply and demand models of differentiated product markets. If, for a particular demand specification, equilibrium prices from two models of supply are close enough to each other asymptotically, IV estimates will have the same asymptotic distribution in both models.

Corollary 1. *Suppose that* x_j, z_j, ξ_j, *and* y_j *satisfy Assumption 1. Let* x_j^* *be any variable such that* $\frac{1}{\sqrt{J}}(\sum_j z_j x_j' - \sum_j z_j x_j^{*\prime}) \xrightarrow{p} 0$ *and let* $y_j^* = x_j^*\beta + \xi_j$. *Then* x_j^*, z_j, ξ_j, *and* y_j^* *satisfy Assumption 1 with the same* M_{zx}, Z_{zx}, *and* $Z_{z\xi}$. *In particular, this will be true if* $\sqrt{J}\max_{j\leq J}\|x_j - x_j^*\| \xrightarrow{p} 0$ *and* $\frac{1}{J}\sum_j \|z_j\|$ *is bounded in probability.*

Proof. The first statement follows from applying Slutsky's theorem to the quantities that converge in distribution under the assumptions of Theorem 1. The second statement follows because

$$\sqrt{J}\|E_J zx' - E_J zx^{*\prime}\| = \sqrt{J}\|E_J z(x' - x^{*\prime})\| \leq \sqrt{J}E_J\|z(x' - x^{*\prime})\| \leq \sqrt{J}E_J\|z\|\|x' - x^{*\prime}\|$$
$$\leq \sqrt{J}\max_{j\leq J}\|x_j' - x_j^{*\prime}\|E_J\|z\|.$$

□

5 Large Market Asymptotics for Some Supply and Demand Models

Using the results from the previous section, I now examine the asymptotic behavior of IV estimates for some common supply and demand models and choices of instruments. I start with a more rigorous treatment of the logit model with single product firms discussed in Section 3.

5.1 The Simple Logit

According to the corollary to Theorem 1, the argument in section 3 will go through as long as, letting $p_j^* = MC_j + 1/\alpha$, we have $\sqrt{J}\max_{j\leq J}|p_j - p_j^*| \xrightarrow{p} 0$ and $\frac{1}{\sqrt{J}}\sum_{j=1}^{J}\left(z_j(x_j', p_j^*) - M_{zx}\right) \to Z_{zx}$ for some matrix M_{zx} that is not full rank when z_j only includes characteristics of other products.

To get an idea of the speed of convergence, write the difference between the constant $1/\alpha$ and the markup as

$$p_j - MC_j - 1/\alpha = \frac{1}{\alpha}\frac{1}{1-s_j} - \frac{1}{\alpha} = \frac{1}{\alpha}\frac{1-(1-s_j)}{1-s_j} = \frac{1}{\alpha}\frac{s_j}{1-s_j}$$

If firms have approximately equal market share so that s_j converges to zero at a $1/J$ rate, this expression will converge to zero at a $1/J$ rate as well, and the corollary to Theorem 1 will apply. A sufficient condition for this is for product characteristics and marginal costs to be bounded.

Theorem 2. *In the logit model with single product firms and Bertrand competition, suppose that product characteristics and marginal costs are (almost surely) uniformly bounded. Then*

$\sup_{j\leq J}\sqrt{J}|p_j - MC_j - 1/\alpha| \overset{a.s.}{\to} 0$ *and any IV estimator for this model that satisfies the assumptions of Theorem 1 will satisfy the same assumptions with* p_j *replaced by* $p_j^* = MC_j + 1/\alpha$.

Proof. See appendix. □

Thus, IV estimators for the logit model have the same limiting distribution as the number of products goes to infinity under Bertrand competition as they do with constant markups. Since characteristics of other products are typically only correlated with prices through markups, IV estimators that use these as excluded instruments will be inconsistent. Write $x_j = (1, w_j')'$ and suppose we use some function $h_j(w_{-j})$ of the product characteristics of other firms as the excluded instrument for p_j^* so that the instrument vector is $z_j = (1, w_j', h_j(w_{-j}))'$. The sample correlation of the instruments with the regressors is

$$\frac{1}{J}\sum_{j=1}^{J} z_j(x_j', p_j^*) = \frac{1}{J}\sum_{j=1}^{J}\begin{pmatrix} 1 \\ w_j \\ h_j(w_{-j}) \end{pmatrix}\begin{pmatrix} 1 & w_j' & p_j^* \end{pmatrix}$$

$$= \frac{1}{J}\sum_{j=1}^{J}\begin{pmatrix} 1 \\ w_j \\ h_j(w_{-j}) \end{pmatrix}\begin{pmatrix} 1 & w_j' & MC_j + 1/\alpha \end{pmatrix}$$

$$= \frac{1}{J}\sum_{j=1}^{J}\begin{pmatrix} 1 \\ w_j \\ h_j(w_{-j}) \end{pmatrix}\begin{pmatrix} 1 & w_j' & \gamma_0 + w_j'\gamma_1 + 1/\alpha \end{pmatrix} + \frac{1}{J}\sum_{j=1}^{J}\begin{pmatrix} 1 \\ w_j \\ h_j(w_{-j}) \end{pmatrix}\eta_j.$$

Assuming that a central limit theorem applies to these sample averages with the second sum converging to its expected value of 0, the assumptions of Theorem 1 will hold with M_{zx} equal to the probability limit of the first term, which does not have full rank since the last column is $\gamma_0 + 1/\alpha$ times the first column plus the linear combination of the middle columns given by matrix multiplication with γ_1 on the right (this statement holds for the probability limit as well). This will hold under quite general conditions. For simplicity, I present the result for the case where product characteristics are iid and h_j is a fixed function of finitely many product characteristics, but a mixing condition on the product characteristics would work as well.

Theorem 3. *In the logit model, suppose that the sequence of product characteristics is bounded and iid. Let* $h : \mathbb{R}^{K(L+M)} \to \mathbb{R}^N$ *be a function of the product characteristics such* *that the elements of $h(x_{j-L},\ldots,x_{j+M})(x_j',\eta_j',\xi_j')$ have finite variance. Let $\hat{\beta}$ be the IV estimator given by (6) with $z_j=(1,x_j',h(x_{j-L},\ldots,x_{j+M})')'$ and W_J any weighting matrix with $W_J\xrightarrow{p}W$ for a positive definite matrix W. If prices are set by single product firms with Bertrand competition, $\hat{\beta}$ will be inconsistent and will have the same asymptotic distribution (in the sense of Theorem 1) as if markups were $1/\alpha$ for all firms.*

Proof. The conditions of the theorem imply that central limit theorems apply to the sample means of the discussion above. □

5.2 Nested Logit

The nested logit model generalizes the logit model of the previous section by placing products into groups and allowing for random coefficients on group dummy variables. The J products are split into G mutually exclusive groups. The set of products in a given group $g\in\{1,\ldots,G\}$ is denoted by $\mathcal{J}_g\subseteq\{1,\ldots,J\}$. The share of product j as a fraction of its group g is denoted by $\bar{s}_{j/g}(x,p,\xi)$, and the share of group g as a fraction of all products is given by $\bar{s}_g(x,p,\xi)$.

Consumer i's utility for good j is

$$u_{ij}=x_j'\beta-\alpha p_j+\xi_j+\zeta_{ig}+(1-\sigma)\varepsilon_{ij}\equiv\delta_j+\zeta_{ig}+(1-\sigma)\varepsilon_{ij}$$

where ζ_{ig} is a random coefficient on a dummy variable for group g and ε_{ij} is still extreme value. The distribution of ζ_{ig} depends on σ and is such that $\zeta_{ig}+(1-\sigma)\varepsilon_{ij}$ is extreme value. This leads to the formulas $\bar{s}_{j/g}=\frac{\exp(\delta_j/(1-\sigma))}{D_g}$ and $\bar{s}_g=\frac{D_g^{1-\sigma}}{\sum_h D_h^{1-\sigma}}$ for shares where $D_g=\sum_{j\in\mathcal{J}_g}\exp(\delta_j/(1-\sigma))$. These can be inverted to get

$$\log s_j-\log s_0=x_j'\beta-\alpha p_j+\sigma\log\bar{s}_{j/g}+\xi_j. \tag{7}$$

As with the logit case, p_j will potentially be correlated with ξ_j through unobserved components of marginal cost and the Nash-Bertrand markup. In this case, we have another endogenous variable, $\log\bar{s}_{j/g}$, to worry about as well. The derivative of j's share with respect to j's price is $\frac{ds_j}{dp_j}=\frac{-\alpha}{1-\sigma}s_j(1-\sigma\bar{s}_{j/g}-(1-\sigma)s_j)$, which gives a markup of $\frac{1-\sigma}{\alpha}/(1-\sigma\bar{s}_{j/g}-(1-\sigma)s_j)$. As with the logit case, this suggests that elements of x_{-j} will be correlated with p_j through the markup term, so we might hope that moments based on the exogeneity of product characteristics of other firms would identify the model.

If the number of groups stays constant while J increases, the markup will converge to a constant, and the result will be similar to the logit model, so consider letting the number of groups increase with the number of products per group constant. To get an idea of the asymptotic behavior of equilibrium prices in this model, note that, if the share of each good converges to zero,

$$p_j - MC_j = \frac{1-\sigma}{\alpha}/(1-\sigma\bar{s}_{j/g} - (1-\sigma)s_j) \overset{J\to\infty}{\to} \frac{1-\sigma}{\alpha}/(1-\sigma\bar{s}_{j/g})$$
$$= \frac{1-\sigma}{\alpha}/\left(1-\sigma\frac{\exp(\delta_j/(1-\sigma))}{D_g}\right) = \frac{1-\sigma}{\alpha}\frac{\sum_{k\in\mathcal{J}_g}\exp(\delta_k/(1-\sigma))}{\left[\sum_{k\in\mathcal{J}_g}\exp(\delta_k/(1-\sigma))\right] - \sigma\exp(\delta_j/(1-\sigma))}.$$

Thus, the equilibrium pricing equations for the goods in group g converge to a system of equations that does not depend on goods in other groups. This suggests that prices for goods in group g will converge to prices that solve these equations. If this is the case, and the convergence is at at least a $\sqrt{J}$ rate, then Theorem 1 can be applied to show that the difference between IV estimators in the two models converges in probability to zero. The following theorem shows that this is true for bounded sequences of product characteristics.

Theorem 4. *In the nested logit model described above with $|\mathcal{J}_g|$ constant and the number of groups going to infinity, suppose that, with probability 1, the sequence of marginal costs is bounded away from zero and the sequences of marginal costs and product characteristics are bounded as J approaches infinity. Let $p_1^*,\ldots,p_J^*$ be the unique solution to the system of equations*

$$p_j^* - MC_j = \frac{1-\sigma}{\alpha}\frac{\sum_{k\in\mathcal{J}_g}\exp((x_k'\beta - p_k^*\alpha+\xi_k)/(1-\sigma))}{\left[\sum_{k\in\mathcal{J}_g}\exp((x_k'\beta - p_k^*\alpha+\xi_k)/(1-\sigma))\right] - \sigma\exp((x_j'\beta - p_j^*\alpha+\xi_j)/(1-\sigma))} \tag{8}$$

and let

$$\bar{s}_{j/g}^* = \frac{\exp((x_j'\beta - p_j^*\alpha+\xi_j)/(1-\sigma))}{\sum_{k\in\mathcal{J}_g}\exp((x_k'\beta - p_k^*\alpha+\xi_k)/(1-\sigma))}.$$

Then, with probability 1, any solution $p_1,\ldots,p_J$ to the Nash pricing equations satisfies

$$\sqrt{J}\sup_{j\le J}|p_j - p_j^*| \to 0 \text{ and } \sqrt{J}\sup_{j\le J}|\bar{s}_{j/g}^* - \bar{s}_{j/g}| \to 0.$$

Thus, IV estimators that satisfy the conditions of Theorem 1 will have the same asymptotic *distribution if p_j and $\bar{s}_{j/g}$ are used as they would if they were replaced by p_j^* and $\bar{s}_{j/g}^*$.*

Proof. See appendix. □

Since product characteristics of goods outside of a given good's nest do not enter into Equation 8, product characteristics from other nests will have no identifying power as instruments for price or $\bar{s}_{j/g}$. However, product characteristics from the same nest will have an effect on equilibrium prices even in the limit, suggesting that they will provide valid instruments.

It is useful to think of the asymptotic pricing equation (8) as resulting from a Bertrand pricing game. As noted in the proof of Theorem 4, the asymptotic markups are the same as those that result from Bertrand competition between firms in group g where firm j has demand

$$\frac{\exp((x_j'\beta - \alpha p_j + \xi_j)/(1-\sigma))}{D_g^\sigma} = \frac{\exp((x_j'\beta - \alpha p_j + \xi_j)/(1-\sigma))}{D_g} D_g^{1-\sigma}. \tag{9}$$

Since the first term takes the form of a logit share, this inverse demand function can be thought of as coming from a process where $D_g^{1-\sigma}$ consumers decide to buy one of the products in group g, and then make decisions between goods in group g according to logit preferences. Note the similarity to a nested logit model with two nests, one with all products in group g and one nest with an outside good with mean utility 0. The demand function for this model is given by

$$\frac{\exp((x_j'\beta - \alpha p_j + \xi_j)/(1-\sigma))}{D_g} \frac{D_g^{1-\sigma}}{1 + D_g^{1-\sigma}}.$$

Although the characteristics of products in the same group can be used to form many moments, in the case where firms are symmetric, some of these moments will be redundant, and will provide no additional identifying power. For example, with two products per nest and one observed product characteristic, we have the moment conditions

$$E\left(\log s_{2j} - \log s_0 - x_{2j}'\beta + \alpha p_{2j} - \sigma \log \bar{s}_{2j/g}\right) x_{2j-1} = 0$$

and

$$E\left(\log s_{2j-1} - \log s_0 - x_{2j-1}'\beta + \alpha p_{2j-1} - \sigma \log \bar{s}_{2j-1/g}\right) x_{2j} = 0.$$

However, if firms $2j$ and $2j-1$ have the same distribution of x, ξ, and η, and the same cost functions, these equations will be satisfied by the same set of parameters, so that one of them is redundant. Thus, we have one moment condition to identify the parameters α and σ, so the model does not satisfy the order condition for identification.

5.3 Random Coefficients Logit

Now consider a model with a more general structure for random coefficients, as in Berry, Levinsohn, and Pakes (1995). Consumer i's utility is given by

$$u_{ij} = x_j'\beta - \alpha p_j + \xi_j + \sum_k x_{jk}\zeta_{ik} + \varepsilon_{ij} \equiv \delta_j + \sum_k x_{jk}\zeta_{ik} + \varepsilon_{ij}$$

where ζ_{ik} is a random coefficient on product k. This specification assumes that there is no random coefficient on price. In contrast to the nested logit example, in which we added a new random coefficient for each nest and increased the number of nests for large J, suppose the number of random coefficients is fixed.

Shares can be obtained by integrating the logit shares for fixed ζ, which gives, letting P_ζ be the probability measure of the random coefficients,

$$s_j = \int \sigma_j(\delta, \zeta)\, dP_\zeta(\zeta)$$

where

$$\sigma_j(\delta, \zeta) = \frac{\exp(\delta_j + \sum_k x_{jk}\zeta_k)}{\sum_\ell \exp(\delta_\ell + \sum_k x_{\ell k}\zeta_k)}.$$

Differentiating under the integral and using the formulas for logit elasticities for fixed ζ gives

$$\frac{ds_j}{dp_j} = -\alpha \int \sigma_j(\delta, \zeta)(1 - \sigma_j(\delta, \zeta))dP_\zeta(\zeta)$$

so that the Bertrand markup is

$$\frac{\int \sigma_j(\delta, \zeta)\, dP_\zeta(\zeta)}{\alpha \int \sigma_j(\delta, \zeta)(1 - \sigma_j(\delta, \zeta))dP_\zeta(\zeta)}.$$

Assuming that, for some c_J with $\sqrt{J}c_J$ converging to 0, $\sup_\zeta \sigma_j(\delta, \zeta) \leq \bar{c}_J$ for almost all sequences of x_j and ξ_j, we will have

$$\frac{1}{\alpha} \leq \frac{\int \sigma_j(\delta,\zeta)\,dP_\zeta(\zeta)}{\alpha \int \sigma_j(\delta,\zeta)(1-\sigma_j(\delta,\zeta))dP_\zeta(\zeta)} \leq \frac{\int \sigma_j(\delta,\zeta)\,dP_\zeta(\zeta)}{\alpha \int \sigma_j(\delta,\zeta)(1-\bar{c}_J))dP_\zeta(\zeta)} = \frac{1}{\alpha(1-\bar{c}_J)}$$

so that markups will converge to $1/\alpha$ at a faster than $\sqrt{J}$ rate. A sufficient condition for this is for ζ to have bounded support and for product characteristics and prices to be bounded.

If the distribution of the vector ζ of random coefficients is not bounded but has thin tails, a more involved argument is necessary. The following theorem covers the commonly used case where ζ has a normal distribution. The proof involves applying an argument similar to the one above after truncating the distribution of ζ at an increasing sequence of points.

Theorem 5. *In the random coefficients model with no random coefficient on price, suppose that product characteristics and prices are bounded with probability one. If the random coefficients ζ are normally distributed, then $\sqrt{J}\max_{j\leq J}|p_j - MC_j - 1/\alpha|$ converges to zero as J goes to infinity.*

Proof. See appendix. □

In addition to the assumptions used for the logit and nested logit models, Theorem 5 requires the assumption that prices do not increase without bound as products are added. Although this restriction on equilibrium prices is not derived from conditions on economic primitives as in the logit and nested logit cases, it is likely to hold for most specifications of the random coefficients.

Applying the corollary to Theorem 1, this theorem implies that using characteristics of other products will not provide consistent estimates in this model even if the true σ is known and used to estimate α and β. This holds under essentially the same conditions as for the logit model.

5.4 The Vertical Model

In contrast to the above models in which consumers have an idiosyncratic preference term ε_{ij} for each item, consider a model in which consumers agree on the ranking of goods, but differ in their willingness to pay for product quality, as in Bresnahan (1987). Utility of an individual consumer is given by

$$u_{ij} = x_j'\beta - \zeta_{ip}p_j + \xi_j \equiv \delta_j - \zeta_{ip}p_j$$

where ζ_{ip} represents consumer i's preference for product quality. A small value of ζ_{ip} means that consumer i has a high value for the quality of the inside goods relative to the numeraire good. The outsize good 0 has $p_0 = 0$ and δ_0 normalized to 0.

Arrange the goods in order of product quality so that $\delta_1 < \ldots < \delta_J$. If all products have positive market share, this will imply that prices satisfy $p_1 < \ldots < p_J$ as well. Consumer i will prefer good j to $j-1$ if

$$\delta_j - \zeta_{ip} p_j > \delta_{j-1} - \zeta_{ip} p_{j-1} \Leftrightarrow \Delta_j \equiv \frac{\delta_j - \delta_{j-1}}{p_j - p_{j-1}} > \zeta_{ip}.$$

Combining this with the expression for $j+1$, consumer i will prefer j to its neighbors if $\Delta_j > \zeta_{ip} > \Delta_{j+1}$. In order for all products to have positive market share, this must hold for some ζ_{ip} for all j, so we must have $\Delta_1 > \ldots > \Delta_J$. If this is the case, consumers who prefer j to its neighbors will also prefer j to all other products, so, letting F be the cdf of ζ_{ip}, market shares will be given by

$$s_j = F(\Delta_j) - F(\Delta_{j+1}). \tag{10}$$

If we define $\Delta_0 = \infty$ and $\Delta_{J+1} = -\infty$, this will hold for good J and the outside good 0 as well.

This can be inverted to give

$$F^{-1}\left(\sum_{k=j}^{J} s_k\right)(p_j - p_{j-1}) = (x_j - x_{j-1})'\beta + \xi_j - \xi_{j-1}.$$

If F is known, this equation can be estimated using OLS. If F is allowed to depend on an unknown parameter, more instruments will be needed, so it will be useful to study the identifying power of moment conditions based on characteristics of other products.

Differentiating the formula for shares with respect to p_j gives, letting f be the pdf of ζ_{ip},

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

This gives markups in an interior Bertrand equilibrium as

$$p_j - MC_j = \frac{F(\Delta_j) - F(\Delta_{j+1})}{f(\Delta_j)\frac{\Delta_j}{p_j - p_{j-1}} + f(\Delta_{j+1})\frac{\Delta_{j+1}}{p_{j+1} - p_j}}. \tag{11}$$

Suppose that, for some $\underline{\zeta} > 0$, $\underline{\zeta} \leq \zeta_{ip}$ for all consumers. That is, willingness to pay for product quality is bounded from above. In this case, if all products have positive market share, we will have $\Delta_j > \underline{\zeta}$ for all j. Thus, the denominator in Equation 11 will be bounded from below as J increases, so, if market shares all converge to zero, markups will converge to zero at the same rate or faster. If firms have approximately equal market shares asymptotically, they will converge to zero at a $1/J$ rate, fast enough for Theorem 1 to hold.

One set of primitive conditions under which markups will converge to zero at a fast rate is the following. In addition to assuming that ζ_{ip} is bounded from below, suppose that the density f of the random coefficient is bounded from above by $\overline{f}$ and from below by $\underline{f}$. Suppose that product characteristics are added in such a way that $\sqrt{J}\max_{j\leq J}\delta_j - \delta_{j-1} \to 0$ and that all products have positive market share in equilibrium. Then

$$p_j - MC_j = \frac{F(\Delta_j) - F(\Delta_{j+1})}{f(\Delta_j)\frac{\Delta_j}{p_j - p_{j-1}} + f(\Delta_{j+1})\frac{\Delta_{j+1}}{p_{j+1} - p_j}} \leq \frac{\overline{f}}{\underline{f}} \frac{\Delta_j - \Delta_{j+1}}{\frac{\Delta_j}{p_j - p_{j-1}} + \frac{\Delta_{j+1}}{p_{j+1} - p_j}} \leq \frac{\overline{f}}{\underline{f}}(p_j - p_{j-1}).$$

In order for product j to have positive market share, we must have

$$\underline{\zeta} < \frac{\delta_j - \delta_{j-1}}{p_j - p_{j-1}} \Rightarrow p_j - p_{j-1} < \frac{\delta_j - \delta_{j-1}}{\underline{\zeta}}.$$

Thus,

$$\sqrt{J}\max_{j\leq J} p_j - MC_j \leq \sqrt{J}\frac{\overline{f}}{\underline{f}\cdot\underline{\zeta}}\max_{j\leq J}\delta_j - \delta_{j-1} \to 0.$$

5.5 Multi Product Firms

I now consider extending the results for the models considered so far to Bertrand competition with multi product firms. If the number of products sold by each firm is fixed and the number of firms grows large, the results are similar the single product case, although, due to the difficulty of proving existence and uniqueness of equilibrium for these models with multi product firms, these results place some conditions directly on equilibrium prices. In particular, these results require prices to be bounded as the number of products increases, and the nested logit model requires the existence of an equilibrium in a limiting form of the game in which price is a differentiable function of costs and characteristics.

If the number of firms is held constant and the number of products per firm increases, the outcome is less clear. Assuming that prices satisfy a mixing condition for a law of large numbers, logit markups will converge to a nonrandom value that is constant within each firm. In addition, the commonly used method of instrumenting with the sample average of product characteristics from the same firm will fail under this condition if the xs are iid. These results suggest that, in the case with a small number of large firms, ex ante asymmetry in the process generating the product characteristics for each firm, rather than variation in characteristics of the individual products, will be necessary for the limiting markup to be affected by product characteristics. In the logit model, if each firm draws product characteristics from a different data generating process then, in the limit, each firm will charge different markups, but all products owned by the same firm will have the same markup. Thus, the identity of the firm producing a given product would potentially be a valid instrument.

First consider keeping the number of products per firm fixed and taking asymptotics in the number of firms. For the logit model, we have $\frac{\partial s_j}{\partial p_j} = -\alpha s_j(1 - s_j)$ and, for $k \neq j$, $\frac{\partial s_j}{\partial p_k} = \alpha s_j s_k$. Substituting this into the first order conditions for p_j (equation 1) and dividing by $-\alpha s_j$ gives

$$(p_j - MC_j)(1 - s_j(x, p, \xi)) - \sum_{k \in \mathcal{F}_f, k \neq j} (p_k - MC_k) s_k(x, p, \xi) - \frac{1}{\alpha} = 0. \tag{12}$$

Assuming that prices and product characteristics are bounded as J increases, shares will go to zero at a faster than $\sqrt{J}$ rate. In this case, markups will converge to $1/\alpha$ at a faster than $\sqrt{J}$, as in the single product case.

For the nested logit model, it can be checked that, for $k \neq j$ and k and j in the same nest, $\partial s_k / \partial p_j = \frac{\alpha}{1-\sigma} s_k(\sigma \bar{s}_{j/g} + (1 - \sigma) s_j)$. For k in some other nest ℓ, we have $\partial s_k / \partial p_j = \alpha s_k s_j$. Plugging these into the first order conditions for firm f setting p_j gives

$$\begin{aligned} 0 = &-\frac{\alpha}{1 - \sigma}(p_j - MC_j) s_j (1 - \sigma \bar{s}_{j/g} - (1 - \sigma) s_j) \\ &+ \sum_{k \in \mathcal{F}_f \cap \mathcal{J}_g, k \neq j} (p_k - MC_k) \frac{\alpha}{1 - \sigma} s_k (\sigma \bar{s}_{j/g} + (1 - \sigma) s_j) + \sum_{k \in \mathcal{F}_f \setminus \mathcal{J}_g} (p_k - MC_k) \alpha s_k s_j + s_j. \end{aligned}$$

Rearranging gives

$$\begin{aligned} 0 = &\frac{1 - \sigma}{\alpha} - (p_j - MC_j)(1 - \sigma \bar{s}_{j/g} - (1 - \sigma) s_j) \\ &+ \sum_{k \in \mathcal{F}_f \cap \mathcal{J}_g, k \neq j} (p_k - MC_k) \frac{\bar{s}_{k/g}}{\bar{s}_{j/g}} (\sigma \bar{s}_{j/g} + (1 - \sigma) s_j) + \sum_{k \in \mathcal{F}_f \setminus \mathcal{J}_g} (p_k - MC_k)(1 - \sigma) s_k \end{aligned}$$

This can be written as, for $\tilde{r}_J$ a term that converges to zero at faster than a $\sqrt{J}$ rate as long as prices and product characteristics are bounded as J increases,

$$0 = \frac{1-\sigma}{\alpha} - (p_j - MC_j)(1 - \sigma \bar{s}_{j/g}) + \sum_{k \in \mathcal{F}_f \cap \mathcal{J}_g, k \neq j} (p_k - MC_k)\sigma \bar{s}_{k/g} + \tilde{r}_J. \tag{13}$$

If this system of equations has a unique solution, and the function that takes marginal costs and product characteristics of nest g and the remainder term to the vector of prices for nest g that solves this system of equations for nest g has an invertible derivative for marginal costs and product characteristics in a compact set that contains them by assumption, then an argument similar to the one used for Theorem 4 will show that prices in the nested logit game converge uniformly at a faster than $\sqrt{J}$ rate to those that solve these equations. As with the single product firm case, equilibrium prices do not depend on characteristics of goods in other nests asymptotically. This holds even for products in other nests owned by the same firm.

In the full random coefficients model with multi product firms, the first order conditions for product j are

$$\begin{aligned}
&-\alpha(p_j - MC_j) \int \sigma_j(\delta, \zeta)(1 - \sigma_j(\delta, \zeta))\, dP_\zeta(\zeta) \\
&+ \alpha \sum_{k \in \mathcal{F}_j, k \neq j} (p_k - MC_k) \int \sigma_j(\delta, \zeta)\sigma_k(\delta, \zeta)\, dP_\zeta(\zeta) + s_j = 0.
\end{aligned}$$

This can be rearranged to give

$$\begin{aligned}
&(p_j - MC_j)\frac{\int \sigma_j(\delta, \zeta)(1 - \sigma_j(\delta, \zeta))\, dP_\zeta(\zeta)}{\int \sigma_j(\delta, \zeta)\, dP_\zeta(\zeta)} \\
&= \sum_{k \in \mathcal{F}_j, k \neq j} (p_k - MC_k)\frac{\int \sigma_j(\delta, \zeta)\sigma_k(\delta, \zeta)\, dP_\zeta(\zeta)}{\int \sigma_j(\delta, \zeta)\, dP_\zeta(\zeta)} + \frac{1}{\alpha}.
\end{aligned}$$

Under the assumptions of Theorem 5, the left hand side converges to $(p_j - MC_j)$ at faster than a $\sqrt{J}$ rate. Assuming prices are bounded, the first term on the right hand side is bounded by a constant times $\frac{\int \sigma_j(\delta,\zeta)\sigma_k(\delta,\zeta)\, dP_\zeta(\zeta)}{\int \sigma_j(\delta,\zeta)\, dP_\zeta(\zeta)}$. This term goes to zero at the required rate using the same argument as for $\frac{\int \sigma_j^2(\delta,\zeta)\, dP_\zeta(\zeta)}{\int \sigma_j(\delta,\zeta)\, dP_\zeta(\zeta)}$, since, using the notation of the proof of Theorem 5 in the appendix, $\int \sigma_j(\delta, \zeta)\sigma_k(\delta, \zeta)\, dP_\zeta(\zeta) \leq \bar{s}_J^2 + P_\zeta(\|\zeta\| > k_J)$, giving the same bound on the numerator.

5.6 Multi Product Firms with Many Products

Now consider fixing the number of firms and letting the number of products per firm grow. First, consider the logit model. Substituting the expression for shares into equation 12 and rearranging gives

$$p_j - MC_j = (1 - s_j)^{-1} \left(\frac{1}{\alpha} + \sum_{k \in \mathcal{F}_f, k \neq j} \frac{(p_k - MC_k) \exp(\delta_k)}{\sum_\ell \exp(\delta_\ell)} \right).$$

Let $r_k = \exp(\delta_k)$, $m_k = (p_k - MC_k)\exp(\delta_k)$, $\overline{r}_f = \frac{1}{|\mathcal{F}_f|} \sum_{k \in \mathcal{F}_f} r_k$, and $\overline{m}_f = \frac{1}{|\mathcal{F}_f|} \sum_{k \in \mathcal{F}_f} m_k$. Suppose that, for each firm f, a law of large numbers applies to the sample averages $\overline{r}_f$ and $\overline{m}_k$ so that $\overline{r}_f \xrightarrow{p} \mu_{rf}$ for some μ_{rf} and $\overline{m}_f \xrightarrow{p} \nu_{rf}$ for some ν_{rf}. This will hold if equilibrium prices satisfy a mixing condition. Suppose that $|\mathcal{F}_f|/J \to \pi_f$ for some π_f for all f. That is, the proportion of products owned by each firm converges to a constant. Then, if shares converge to zero,

$$p_j - MC_j = (1 - s_j)^{-1} \left(\frac{1}{\alpha} + \frac{\overline{m}_f - \frac{1}{|\mathcal{F}_f|} m_j}{\sum_{h=1}^F \frac{|\mathcal{F}_h|}{|\mathcal{F}_f|} \overline{r}_h} \right) \xrightarrow{p} \frac{1}{\alpha} + \frac{\nu_{rf}}{\sum_{h=1}^F \frac{\pi_h}{\pi_f} \mu_{rh}}.$$

Thus, markups will converge to a constant that does not vary within a firm.

Convergence to a constant markup might be expected for the logit model since substitution patterns are determined by the independence of irrelevant alternatives property, so that all products compete with all other products in a roughly symmetrical way. Since the effect of any one x_k on p_j is not that large, the effects of each x_k on p_j cancel out in the limit if the xs are not too correlated. This will not be the case in all models. However, certain choices of instruments will lead to inconsistent estimates under more general conditions. Suppose that, for each firm f, $\frac{1}{|\mathcal{F}_f|} \sum_{j \in \mathcal{F}_f} x_j \xrightarrow{p} \mu$ for the same μ (for example, if each firm's sequence of product characteristics are draws of the same stationary process) and $\frac{1}{|\mathcal{F}_f|} \sum_{j \in \mathcal{F}_f} p_j \xrightarrow{p} \mu_{pf}$ for some μ_{pf}. Again, the law of large numbers for prices will follow if a mixing condition holds on equilibrium prices.

Suppose that, for good j produced by firm f, we use $\frac{1}{|\mathcal{F}_f|} \sum_{k \in \mathcal{F}_f, k \neq j} x_k$ as the excluded instrument for price. Suppose that x_j contains a constant and write $x_j = (1, w_j')$. The sample covariance of the instruments with the covariates is

$$\frac{1}{J}\sum_{f=1}^{F}\sum_{j\in\mathcal{F}_f}\begin{pmatrix}\frac{1}{|\mathcal{F}_f|}\sum_{k\in\mathcal{F}_f,k\neq j}w_k\\ w_j\\ 1\end{pmatrix}\begin{pmatrix}1 & w_j' & p_j\end{pmatrix}.$$

The first $K-1$ rows are

$$\frac{1}{J}\sum_{f=1}^{F}\sum_{j\in\mathcal{F}_f}\frac{1}{|\mathcal{F}_f|}\sum_{k\in\mathcal{F}_f}w_k\begin{pmatrix}1 & w_j' & p_j\end{pmatrix}-\frac{1}{J}\sum_{f=1}^{F}\frac{1}{|\mathcal{F}_f|}\sum_{j\in\mathcal{F}_f}\begin{pmatrix}w_j & w_jw_j' & w_jp_j\end{pmatrix}.$$

If a law of large numbers applies to rows K through $2K-1$, the second term will converge in probability to zero. The first term is

$$\sum_{f=1}^{F}\frac{|\mathcal{F}_f|}{J}\left\{\frac{1}{|\mathcal{F}_f|}\sum_{k\in\mathcal{F}_f}w_k\right\}\left\{\frac{1}{|\mathcal{F}_f|}\sum_{j\in\mathcal{F}_f}\begin{pmatrix}1 & w_j' & p_j\end{pmatrix}\right\}.$$

Under the assumptions above, this will converge in probability to

$$\mu\sum_{f=1}^{F}\pi_f\begin{pmatrix}1 & \mu' & \mu_{fp}\end{pmatrix}.$$

The last row of the sample covariance matrix is

$$\sum_{f=1}^{F}\frac{|\mathcal{F}_f|}{J}\frac{1}{|\mathcal{F}_f|}\sum_{j\in\mathcal{F}_f}\begin{pmatrix}1 & w_j' & p_j\end{pmatrix}\xrightarrow{p}\sum_{f=1}^{F}\pi_f\begin{pmatrix}1 & \mu' & \mu_{fp}\end{pmatrix}.$$

Thus, the sample covariance of the instruments with the regressors converges in probability to a rank deficient matrix, since the first $K-1$ rows are multiples of the last row.

Although these arguments for the case of finitely many firms with an increasing number of products rely on a mixing condition for prices that is not derived from economic primitives, they are suggestive of which instruments will be valid for which models in this setting. In the logit model, markups converge to a value that is constant within a firm. The limiting markup will differ between firms if the proportion of products owned by each firm or the data generating process for product characteristics differs between firms. In this case, instrumental variables must exploit variation in average markups across firms, rather than variation in markups within a firm. Average product characteristics within a firm, weighted by the proportion of products owned by a firm, may provide valid instruments. Since the number of firms is bounded, firm indicator variables could be used as well. These instruments rely on variation in markups across firms caused by variation in the number and average characteristics of products owned by each firm. If firms are roughly symmetric, average product characteristics will not vary across firms in the limit, so that this instrument will not have identifying power.

5.7 Discussion

Although simple necessary and sufficient conditions for a demand system to have a limiting form in which product characteristics have a strategic effect appear to be difficult to obtain, the examples in this section shed some light on which models lead to product characteristics affecting markups in the limit. When the number of random coefficients is kept finite while the number of firms grows, as in the logit and random coefficients logit examples, the idiosyncratic error term wipes out all other components of demand, and product characteristics play no role in the limiting form of competition. When the number of random coefficients increases as well, as in the nested logit model, it may happen in such a way that, in the limit, small groups of products cater to different sets of consumers with different coefficients. If this is the case, observing a large market is like observing a large number of niche markets for consumers with different values of the random coefficients.

While the nested logit example shows that it is possible to make product characteristics matter in the limit by increasing the number of random coefficients or changing their distribution, the researcher must be careful to do this without increasing the number of parameters to be estimated. In the nested logit specification, the random coefficient for each nest is restricted to have the same distribution, so that, even though the number of random coefficients increases, the number of parameters stays finite. This way of adding random coefficients will be appropriate in some applications, but there will certainly be cases where the researcher will want want to specify demand in a different way. The demand specifications in this section should not be thought of as a rigid set of models from which to choose, but rather as a set of examples that can guide the researcher towards a demand model that is flexible enough to allow the data to answer a particular question while imposing enough restrictions that it can be estimated with a small number of large markets. This could mean taking the nested logit model considered here and adding random coefficients to some continuous variables to allow for more consumer heterogeneity, or perhaps using a more complicated nesting structure or allowing some serial correlation between nests. While the results in this paper do not apply immediately to such models, the intuition and techniques for deriving asymptotic distributions from the examples I consider here will be useful in many applications.

In the logit model with a small number of large firms, the power of product characteristic instruments depends on some firms having a line of products that is, on average, more desirable than products of other firms. In this context, the cannibalization effect of characteristics of a single product on markups of similar products diminishes too quickly to aid identification. Rather, product characteristics shift a firm's markups only if they make that firm's products more desireable on average. This lends support to the practice of using averages of charcteristics of products of the same firm as instruments, proposed by Berry, Levinsohn, and Pakes (1995). Since there is no exogenous variation driven by product characteristics within a firm, there is no additional variation that a researcher could take advantage of using subsets of a firm's product characteristics.

6 Counterfactuals in Large Markets

In most applications, the parameters of a discrete choice demand system are not of interest purely for their own sake. Rather, the researcher is interested in equilibrium outcomes under some policy change. For example, prices after a hypothetical merger are often of interest. If we think of the counterfactual as happening in a fixed market, the estimated counterfactual outcome will be a fixed function of the estimated parameters of the original model, so that, if this function is differentiable, the delta method can be combined with the asymptotic distributions derived in this paper to give approximations to the finite sample distribution of the estimated counterfactual quantity. However, the counterfactual quantity often involves a market of comparable size to the market used for estimation. For example, outcomes after entry of a new firm or a merger of existing firms in the market used for estimation are often of interest. In this case, an approximation to the distribution of the estimated counterfactual outcome that is asymptotically valid when the size of the counterfactual market is also increasing seems desirable, given that the approximation of the distribution of the parameters used to compute the counterfactual quantity comes from this asymptotic exercise.

This is not just a technical issue. If large market approximations to the distributions of the estimated counterfactual quantities restrict the possible outcomes of counterfactual exercises, the researcher will want to use this information to choose which demand specifications are appropriate for answering which questions. Consider, for example, the simple logit model with single product firms. As the number of firms increases, the markup approaches a constant, so that using the "limiting" model with a constant markup to predict the price change from a merger will always give zero as the answer. This suggests that using this model with data from a single market for this type of merger analysis is not a good idea even if cost side instruments are available because, once the market is large enough for asymptotics to give a good approximation, the model is restricting the predicted price change to be close to zero. On the other hand, if the researcher thinks that a nested logit specification with many small nests and the products in the potential merger in the same nest is appropriate, the answer will depend on the primitives of the model even in the limit, so that estimating this model is a sensible approach to this question. See Bajari and Benkard (2003) for derivations of other implications of models with logit style idiosyncratic terms for each new product that restrict the possible outcomes of welfare analysis and estimation of markups in ways that are undesirable for many applications.

In the remainder of this section, I propose a framework for addressing these issues. Suppose we are interested in some counterfactual quantity τ_J that, for a given market size J, is a function of the primitives of the model. The primitives of the model include the parameters $\theta \equiv (\alpha, \beta, \sigma, \gamma)$ and a sequence of covariates and error terms, for which I use the notation $\tilde{x}$, $\tilde{\xi}$, and $\tilde{\eta}$, since I allow them to be different from those in the market used to estimate the original parameters. I label $w \equiv (\tilde{x}, \tilde{\xi}, \tilde{\eta})$ for notational convenience, so that $\tau_J = \tau_J(\theta, w)$. Here, to avoid extra subscripts, $\tilde{x}$, $\tilde{\xi}$, and $\tilde{\eta}$ contain the entire sequence of covariates, so that if a counterfactual outcome in a market with the first J products is of interest, τ_J is defined to be a function of the parameters and the first J elements of w only. The sequence w could contain variables from the market used for estimation, or it could be a vector of covariates for a completely different market.

Suppose that, for all θ, $\tau_J(\theta, w)$ converges in probability to some $\tau_\infty(\theta, w)$. Typically, $\tau_\infty(\theta, w)$ is a differentiable function of the parameters θ and a finite subset of the elements of w, so that its asymptotic distribution can be obtained using the delta method. If $\tau_J(\theta, w)$ converges to $\tau_\infty(\theta, w)$ at a faster than $\sqrt{J}$ rate with an appropriate condition on uniformity in θ, $\sqrt{J}(\tau_J(\hat{\theta}, w) - \tau_J(\theta, w))$ will have the same asymptotic distribution as $\sqrt{J}(\tau_\infty(\hat{\theta}, w) - \tau_\infty(\theta, w))$, so that this asymptotic distribution can be used as an approximation for inference for $\tau_J(\theta, w)$. The following theorem gives sufficient conditions for this to hold.

Theorem 6. *Suppose that $\hat{\theta} \xrightarrow{p} \theta$ and, for some $\varepsilon > 0$, $\sqrt{J} \sup_{\|\theta' - \theta\| < \varepsilon} (\tau_J(\theta', w) - \tau_\infty(\theta', w)) \xrightarrow{p} 0$ for some $\tau_\infty(\theta, w)$ and $\sqrt{J}(\tau_\infty(\hat{\theta}, w) - \tau_\infty(\theta, w)) \xrightarrow{d} Z$ for some random variable Z. Then $\sqrt{J}(\tau_J(\hat{\theta}, w) - \tau_J(\theta, w)) \xrightarrow{d} Z$.*

Proof. It suffices to show that $\sqrt{J}(\tau_J(\theta,w)-\tau_\infty(\theta,w)) \xrightarrow{p} 0$ and $\sqrt{J}(\tau_J(\hat\theta,w)-\tau_\infty(\hat\theta,w)) \xrightarrow{p} 0$. Then, Slutsky's theorem will give the result. The first statement follows immediately from the assumptions of the theorem. The probability that the quantity in the second statement is greater than some $\eta > 0$ is bounded by $P(\|\hat\theta-\theta\| \geq \varepsilon) + P(\sqrt{J}\sup_{\|\theta'-\theta\|<\varepsilon}(\tau_J(\theta,w) - \tau_\infty(\theta,w)) > \eta)$, and both probabilities converge to zero. □

One special case of this theorem that is of interest is when $\tau_\infty(\theta,w)$ is a constant that does not depend on θ or w. This corresponds to cases discussed in the introduction to this section, such as price changes after mergers in the simple logit, where the limiting counterfactual outcome does not depend on the data. In this case, Theorem 6 shows that this convergence happens at a fast enough rate that approximations to sampling distributions based on $\sqrt{J}$ asymptotics will never give a prediction for the counterfactual outcome that depends on the data, since the limiting distribution for the estimated counterfactual will be degenerate when scaled up by $\sqrt{J}$. Although higher order approximations may give nondegenerate results, this suggests focusing empirical work on cases where $\tau_\infty(\theta,w)$ depends on θ in an interesting way.

With a bit of work, the results in section 5 can be used to verify the conditions of this theorem for many of the models considered in that section. As an example, consider the nested logit model where the counterfactual outcome of interest is the price of some good j in a market with $J+L$ single product firms where the market used for estimation has J products, L is a fixed, possibly negative, integer, and the covariates and errors in the counterfactual market are given by the first $J+L$ elements of some bounded sequence $(\tilde x,\tilde\xi,\tilde\eta)$ for which the number of elements in a nest is bounded and the true marginal cost is bounded away from zero. This covers entry, exit, and changes in characteristics of existing products. The multi product case, which covers post merger prices, will follow from similar arguments, but requires assumptions on the equilibrium of the limiting form of the game. Formally, I define $p_{j',J'}(\theta',x',\xi',\eta')$ to be the Bertrand equilibrium price for product j' in a market with J' single product firms each selling a product given by one of the first J' elements of (x',ξ',η') when demand is nested logit with parameters given by θ' (here, x' and $\tilde x$ will be understood to contain the group indicator variables). We are interested in $\tau_J(\theta,\tilde x,\tilde\xi,\tilde\eta) \equiv p_{j,J+L}(\theta,\tilde x,\tilde\xi,\tilde\eta)$, which we estimate using $\tau_J(\hat\theta,\tilde x,\tilde\xi,\tilde\eta)$. Let g be the group containing the product of interest j. Let $\tau_\infty(\theta',x',\xi',\eta') \equiv p_j^*(\theta',x',\xi',\eta')$ be defined as the price of product j in the solution to the asymptotic pricing equations given by equation 8 with primitives given by θ' and the elements of (x,ξ,η) in group g.

The first assumption of the theorem, that, for some ε, $\sqrt{J}\sup_{\|\theta'-\theta\|<\varepsilon}|p_{j,J+L}(\theta',\tilde x,\tilde\xi,\tilde\eta) -$

$p_j^*(\theta', \tilde{x}, \tilde{\xi}, \tilde{\eta})| \xrightarrow{p} 0$ follows almost immediately from theorem 4. The only difference is that we now need the convergence to be uniform in a neighborhood of the true θ. In fact, the proof of the theorem can be extended so that this is true, which I do in the proof of the theorem in the appendix. Now, if $\sqrt{J}(\hat{\theta} - \theta) \xrightarrow{d} N(0, A)$ for some matrix A (this can be verified for a given set of instruments using theorems 1 and 4), then, by the delta method, the second assumption of theorem 6, that $\sqrt{J}(p_j^*(\hat{\theta}, \tilde{x}, \tilde{\xi}, \tilde{\eta}) - p_j^*(\theta, \tilde{x}, \tilde{\xi}, \tilde{\eta}))$ converges in distribution to some Z, will hold with Z distributed $N(0, p_{j\theta}^*(\theta, \tilde{x}, \tilde{\xi}, \tilde{\eta})' A p_{j\theta}^*(\theta, \tilde{x}, \tilde{\xi}, \tilde{\eta}))$ where $p_{j\theta}^*(\theta, \tilde{x}, \tilde{\xi}, \tilde{\eta})$ is the derivative of $p_j^*(\theta, \tilde{x}, \tilde{\xi}, \tilde{\eta})$ with respect to θ if we can show that p_j^* is in fact differentiable with respect to θ. This follows from the implicit function theorem since, as shown in the proof of theorem 4, the function that takes p^* and the model primitives to the difference between the two sides of the pricing equation 8 is differentiable with respect to p^* with an invertible derivative matrix.

Thus, estimates of prices in the counterfactual market will be consistent and asymptotically normal as the size of the counterfactual market and the market used for estimation increase. Given a consistent estimate $\hat{A}$ of A, we can estimate the variance of the limiting normal distribution using $p_{j\theta}^*(\hat{\theta}, \tilde{x}, \tilde{\xi}, \tilde{\eta})' \hat{A} p_{j\theta}^*(\hat{\theta}, \tilde{x}, \tilde{\xi}, \tilde{\eta})$. Consistency of this estimate of the limiting variance follows from continuity of $p_{j\theta}^*(\theta, \tilde{x}, \tilde{\xi}, \tilde{\eta})$ as a function of theta, which can be verified by computing this derivative using the implicit function as in the proof of theorem 4. In addition, the above argument shows that the point estimate of the price of good j in the counterfactual pricing market computed using the asymptotic pricing equation 8 will be asymptotically equivalent to the estimate computed using the full Bertrand equilibrium. This may be useful if the full equilibrium is difficult to compute.

As discussed at the beginning of the section, this also has implications for which types of questions a researcher would want to answer by estimating a nested logit demand system in a large market. If we are interested in the price of good j when another product is added, p_j^* will only be different from the old price if the new product is in the same nest. Thus, estimating a nested logit model in a large market may be useful for predicting the price change of a good that can be modeled as being in the same nest as the new product, but modeling the new product as entering a different nest will constrain the estimated price change to be close to zero. Theorem 6 provides a way of formalizing the intuition that, once the market size is large enough that large market approximations can be used for estimation, the counterfactual outcome is close enough to its limiting value that one would not want to estimate the model unless the counterfactual quantity is still interesting in the "limiting model." When the assumptions of the theorem hold, using the "limiting model" for computing counterfactuals gives the same result as using the model with finite J for the purposes of first order asymptotics.

7 Monte Carlo

With asymptotic results, there is always the question of whether these results give good approximations in sample sizes (or, in this case, market sizes) of practical importance. In this section, I examine this question with a monte carlo study of some of the models considered in section 5.

7.1 Logit with Symmetric Firms

According to the analysis in section 5, using characteristics of other products as instruments will give inconsistent estimates in the simple logit model in a single market with a bounded number of products per firm under general conditions, while cost shifters will provide consistent estimates. If firms draw product characteristics from the same data generating process, the arguments in section 5.6 suggest that product characteristic instruents will still perform poorly. Tables 1 and 2 give the results of a monte carlo study of IV estimators in the logit model with data from a single market with firms drawing product characteristics from the same data generating process. Table 1 gives the median bias and median absolute deviation of the IV estimates of the price coefficient $-\alpha$ in this model using characteristics of other products as instruments. Table 2 gives the median bias and median absolute deviation of the IV estimates of of the estimated $-\alpha$ using cost shifters as instruments on the same monte carlo data sets. I report median bias and median absolute deviation rather than bias and mean squared error since IV estimators may lack moments, and these reported measures of central tendency toward the true value and dispersion from the true value that can be interpreted in a similar way. For the product characteristic instruments, I use average product characteristics from other products produced by the same firm, $\sum_{k\in\mathcal{F}_f,k\neq j} x_k$ to instrument for p_j in the logit regression (5). Each row of each table reports results for a different market structure (number of firms and products per firm), but the same parameters α, β, and γ and the same process generating x, η, and ξ.

The data generating process for the monte carlo data sets is as follows. x_j contains a constant and a uniform $(0,1)$ random variable. I generate the cost shifter, z_j, as another uniform random variable independent of x. To generate η and ξ, I generate three independent uniform $(0,1)$ random variables u_{1j}, u_{2j}, and u_{3j}, and set $\xi_j = u_{1j} + u_{3j} - 1$ and $\eta_j =$

$u_{1j} + u_{2j} - 1$ so that η and ξ are correlated and bounded. x_j, ξ_j, and η_j are independent across products j. The parameters used to generate this table are $\alpha = 1$, $\beta = (3,6)'$, and $\gamma = (2,1,1)'$, where the last element of γ is the coefficient of the excluded cost instrument. Note that with these parameter values, the variance of the observed portion of utility $x'\beta$ across products and the idiosyncratic component ε_{ij} across consumers is of the same order of magnitude (the variance of the extreme value error ε_{ij} is $\pi^2/6$, while the nonconstant element of $x'\beta$ has a variance of 3). For a small number of monte carlo draws (fewer than 1%), the equation solver did not converge to a solution for equilibrium prices, and these were discarded. The corresponding rows from each table use the same monte carlo data set of 1000 monte carlo replications.

The monte carlo distributions of the price coefficient estimates reflect the asymptotic results for this model. The estimates with cost instruments are not very dispersed, and are centered close to the true value of -1. As one would expect for an estimator that is consistent as J goes to infinity, the estimates become more precise as more products are added either by increasing the number of firms or the number of products per firm.

The estimates using product characteristic instruments perform much worse. In addition to being more dispersed and centered further from the true value than the cost instrument estimates for fixed sample sizes, these estimates do not become more precise as the number of products increases. Thus, the inconsistency results of section 5 provide a good description of the product characteristic IV estimates for the range of market sizes, firm sizes, and distributions of the model primitives in the monte carlo.

According to the analysis of section 5, product characteristics of other firms have no identifying power in this model as the number of products goes to infinity and the number of firms stays fixed because markups converge to a constant. Table 3 illustrates this for the monte carlo data. The first column gives the average sample variance of prices within a market, and the second gives the average sample variance of markups within a market. For a fixed number of products, there tends to be more variation in the markup with a smaller number of firms, but, as the number of products increases with a fixed number of products per firm, variation in the markup decreases, and is swamped out by variation in marginal costs. For the larger market sizes in the monte carlo, variation in markups accounts for about 1/100th of the variation in prices within a market or less.

Products	Products per firm	Median bias of $-\hat{\alpha}$	Median absolute deviation of $-\hat{\alpha}$
20	2	0.50676	0.80569
20	5	0.43779	0.89869
20	10	0.67988	0.89654
60	2	0.33801	0.80091
60	5	0.29468	0.85239
60	10	0.38019	0.84481
100	2	0.32209	0.84566
100	5	0.42134	0.90229
100	10	0.37298	0.84535

Table 1: Monte Carlo Results for $-\hat{\alpha}$ in Logit Model with Product Characteristic Instruments

Products	Products per firm	Median bias of $-\hat{\alpha}$	Median absolute deviation of $-\hat{\alpha}$
20	2	-0.0042745	0.26522
20	5	-0.0058366	0.30762
20	10	0.033014	0.40284
60	2	-0.0025152	0.12628
60	5	0.0072365	0.14028
60	10	0.0039809	0.14887
100	2	-0.0000040	0.096628
100	5	-0.0083498	0.098462
100	10	0.0096867	0.10476

Table 2: Monte Carlo Results for $-\hat{\alpha}$ in Logit Model with Cost Instruments

Products	Products per firm	Average sample variance of prices	Average sample variance of markups
20	2	0.398	0.022518
20	5	0.54993	0.12252
20	10	1.0105	0.46788
60	2	0.34705	0.0025035
60	5	0.37425	0.013976
60	10	0.44604	0.051802
100	2	0.33904	0.0008845
100	5	0.35364	0.0052182
100	10	0.38597	0.019302

Table 3: Sample Variance of Prices and Markups within a Market

7.2 Random Coefficients Logit with Symmetric Firms

If the number of products per firm increases with a new extreme value error added for each product, but no changes in the random coefficients, the results from section 5 for the random coefficients model are similar to those for the logit model. Product characteristics do not have identifying power in the limit. Tables 4 and 5 present the results of a monte carlo study of the random coefficients logit model with data from a single market. Table 4 gives results with product characteristic instruments and table 5 gives results for cost instruments. For both tables, I treat the variance of the random coefficients, σ, as known and solve for the IV estimates of the other parameters with σ fixed at its true value. Since the resulting IV estimator is the closed form solution to a system of linear equations, this eliminates potential concerns that negative results may come from a failure to minimize a nonlinear GMM objective function. Estimators that perform poorly can be expected to do even worse when σ needs to be estimated as well.

The random coefficients are generated as 10 draws from a normal distribution. For each monte carlo run, these same 10 draws are used both in solving for equilibrium prices and generating shares and in inverting the shares to estimate the model. Another way of putting this is that the random coefficients for each replication come from a discrete distribution that places equal mass on 10 points, where these these 10 points are drawn from a normal distribution. I set the variance of the normal distribution from which the mass points for the random coefficient is drawn from to 9, and generate all other variables according to the same data generating process as for the logit monte carlo of section 7.1.

The results are similar to those of the logit model presented in section 7.1. Estimates that use the product characteristic instruments perform poorly, and do not become more precise as the size of the market increases, while cost shifter instruments do not suffer from these problems. Here, σ is held fixed so that the price coefficient and the coefficients of product characteristics are the only parameters that have to be estimated. Estimates that use product characteristics as instruments will likely perform even worse when σ needs to be estimated as well.

7.3 Logit with Asymmetric Firms

The results so far use the same data generating process for product characteristics for each firm. With a large number firms, product characteristic instruments should do poorly regardless of whether the data generating process for product characteristics varies across firms

Products	Products per firm	Median bias of $-\hat{\alpha}$	Median absolute deviation of $-\hat{\alpha}$
20	2	0.71992	1.6982
20	5	0.53011	1.2564
20	10	0.39139	1.055
60	2	0.69935	1.7864
60	5	0.71713	1.539
60	10	0.63133	1.4198
100	2	0.77397	1.6783
100	5	0.64665	1.6274
100	10	0.73668	1.5371

Table 4: Monte Carlo Results for $-\hat{\alpha}$ in Random Coefficients Logit Model with Product Characteristic Instruments

Products	Products per firm	Median bias of $-\hat{\alpha}$	Median absolute deviation of $-\hat{\alpha}$
20	2	0.029988	0.48535
20	5	0.034285	0.45894
20	10	-0.038265	0.50143
60	2	0.011275	0.26557
60	5	0.0080281	0.24548
60	10	0.002217	0.26156
100	2	-0.0099175	0.20699
100	5	0.0012001	0.18459
100	10	-0.0080423	0.18609

Table 5: Monte Carlo Results for $-\hat{\alpha}$ in Random Coefficients Logit Model with Cost Instruments

Products	Products per firm	Median bias of $-\hat{\alpha}$	Median absolute deviation of $-\hat{\alpha}$
20	2	0.28316	0.74258
20	5	0.083707	0.43282
20	10	0.26423	0.60083
60	2	0.34542	0.80526
60	5	0.17064	0.5633
60	10	0.011654	0.32352
100	2	0.29983	0.87659
100	5	0.21537	0.72778
100	10	0.086334	0.42442

Table 6: Monte Carlo Results for $-\hat{\alpha}$ in Logit Model with Product Characteristic Instruments with Asymmetric Firms

according to the analysis in section 5. However, with a small number of firms and a large number of products per firm, the arguments in section 5.6 suggest that product characteristic instruments will typically have power if the data generating processes for product characteristics is different for each firm. To investigate whether this gives a good description of finite sample behavior, I run a monte carlo for the logit model similar to the one presented in section 7.1, but with firms drawing product characteristics from different data generating processes. Everything is the same as in the logit monte carlo presented in section 7.1 except for the data generating process for the product characteristic x_j. Rather than drawing x_j from the same distribution for each product, I draw x_j from a uniform $(-1.5, -.5)$ distribution for half of the firms, and a uniform $(.5, 1.5)$ distribution for the other half of the firms.

As predicted, while the product characteristic instruments still tend to perform poorly with a large number of products per firm, these estimates do better with a small number of large firms. With 60 products and 10 products per firm, the IV estimate with product characteristic instruments is centered near zero. This contrasts with the large median bias that occurs with product characteristic instruments in the same situation with symmetric firms, and with a larger number of smaller asymmetric firms. Note that even with 10 firms and 10 products per firm, the product characteristic estimator performs much better than with symmetric firms. With 20 firms, each with 5 products, however, the asymmetry in product characteristics does not appear to help much in this case. Note that the cost shifters still perform well.

Products	Products per firm	Median bias of $-\hat{\alpha}$	Median absolute deviation of $-\hat{\alpha}$
20	2	0.030595	0.2525
20	5	0.05201	0.3152
20	10	0.058311	0.3775
60	2	0.011253	0.13973
60	5	0.0063074	0.1352
60	10	0.0010869	0.14376
100	2	0.0072848	0.09454
100	5	0.00876	0.10094
100	10	0.010869	0.10398

Table 7: Monte Carlo Results for $-\hat{\alpha}$ in Logit Model with Cost Instruments with Asymmetric Firms

8 Implications for Empirical Work

Determining whether asymptotic approximations can be applied to a particular estimator, or whether a parameter can even be consistently estimated, is clearly important for any empirical study that uses it. For the IV estimates considered in this paper, this depends on how strongly the instruments are correlated with prices. Since prices are determined by equilibrium in a supply side model, a careful researcher can use this information to determine which instruments will have a strong enough correlation with prices to give good IV estimates. In most IO applications, the researcher must take a stand on the supply side anyway, so it makes sense to use this information in choosing instruments and specifying a demand model as well. The results in this paper show how to do this.

Another approach to determining whether the correlation between instruments and prices is strong enough to give good estimates is through a first stage test for identification. While such a test should certainly be performed and will complement an analysis based on results in this paper, it is important to emphasize that testing for identification alone does not sufficiently address these issues. This paper shows that certain models of supply and demand used in practice can constrain product characteristic instruments to have poor identifying power in the data set at hand. If a researcher estimates such a model after performing a first stage test that finds evidence that product characteristics strongly identify the model, the researcher is ignoring evidence that the model is misspecified in ways that are economically important for policy counterfactuals. The very fact that the instruments are strongly correlated with price means that the demand model and pricing game fail to fit the data in ways that are likely to lead to incorrect policy counterfactuals.

This paper shows how to construct models that avoid such issues. After constructing such a model, one should still test whether it is strongly identified with the data at hand. If the test finds evidence for identification, the researcher can proceed confidently, knowing that this result confirms, rather than rejects, the supply side model that will later be used in policy counterfactuals.

The results in section 5 cover many demand models used in practice. However, the best approach to modeling demand in a particular application may not fall neatly into one of the examples considered in this paper. In this case, these examples can be used as a guide toward a demand specification that meets the researcher's needs. For example, one might want a demand specification in which product characteristics matter for competition with many firms, but restricting all of the random coefficients to indicator variables as in the nested logit model may be too stringent for a particular application. For such applications, the researcher could specify a nested logit model with some additional random coefficients on continuous variables or some other model with a large number of random coefficients. Rather than taking the results of section 5 as a rigid set of models from which to choose, these results should be thought of as a guide toward building a demand specification that is flexible enough for a particular application, but can be estimated with data from a small number of large markets.

9 Conclusion

Thinking about the implications of a complete economic model for identification is always important. The results of this paper show that the issue takes a different form when the number of products is large relative to the number of markets. Instruments that can be used to consistently estimate a model with data on many markets may not even satisfy the order condition under large market asymptotics. A monte carlo study confirms that these asymptotic approximations provide a good description of the logit and random coefficients logit models in market sizes of practical importance. The IV estimates that are inconsistent as the number of products increases perform poorly. The consistent estimators do better, and become increasingly accurate as the size of the market increases.

Once the issues involved with large market asymptotics are taken into account, one can use the results in this paper in much the same way one would use identification results to guide empirical work in the small market case. The results in this paper can be used to decide what types of models can be consistently estimated, what testable restrictions they imply, and how they restrict the possible outcomes of counterfactual analysis.

Appendix

Proof of Theorem 1

Proof. Part (i) follows from (ii) and (iii), and (ii) follows by the usual combination of a law of large numbers, a central limit theorem, and Slutsky's theorem. Part (iii) essentially follows from applying results for partially identified IV (see, for example, Stock and Wright (2000)) to a version of the model that is reparameterized so that the parameter of interest is $H^{-1}\beta$. We have, letting A_J be the $d \times d$ diagonal matrix with the first d_2 diagonal entries equal to 1 and the last d_1 equal to $\sqrt{J}$,

$$\begin{aligned}
\hat{\beta} - \beta &= \left(\left[\sum_{j=1}^{J} z_j' x_j \right]' W_J \left[\sum_{j=1}^{J} z_j' x_j \right] \right)^{-1} \left[\sum_{j=1}^{J} z_j' x_j \right]' W_J \left[\sum_{j=1}^{J} z_j (y_j - x_j' \beta) \right] \\
&= \arg\min_{\gamma} \| E_J z\xi - E_J z x' \gamma \|_{W_J}
\end{aligned}$$

so that

$$\begin{aligned}
&\begin{pmatrix} T_{1J} \\ \sqrt{J} T_{2J} \end{pmatrix} = A_J H^{-1} (\hat{\beta} - \beta) = \arg\min_{\gamma} \left\| E_J z\xi - E_J z x' H A_J^{-1} \gamma \right\|_{W_J} \\
&= \arg\min_{\gamma} \left\| \sqrt{J} E_J z\xi - \sqrt{J} E_J z x' H A_J^{-1} \gamma \right\|_{W_J} \\
&= \arg\min_{\gamma} \left\| \sqrt{J} E_J z\xi - (\sqrt{J} E_J z x' H_1, E_J z x' H_2) \gamma \right\|_{W_J}.
\end{aligned}$$

By the continuous mapping theorem, this converges to

$$\arg\min_{\gamma} \| Z_{z\xi} - (Z_{zx} H_1, M_{zx} H_2) \gamma \|_{W_J}.$$

The result follows from applying the partitioned least squares formula to this expression. □

Proof of Theorem 2

Proof. Define $m_j = x_j'\beta + \xi_j$. First, note that

$$p_j - MC_j = \frac{1}{\alpha}\frac{\sum_{i\neq j}\exp(m_i - \alpha p_i) + \exp(m_j - \alpha p_j)}{\sum_{i\neq j}\exp(m_i - \alpha p_i)} = \frac{1}{\alpha} + \frac{1}{\alpha}\frac{\exp(m_j - \alpha p_j)}{\sum_{i\neq j}\exp(m_i - \alpha p_i)}$$
$$\leq \frac{1}{\alpha} + \frac{1}{\alpha}\exp(m_j - \alpha p_j)$$

(the inequality follows because, with an outside good, the summand in the denominator is 1 for $i = 0$). Let $p_{mon}(MC_j, m_j)$ be the value of p_j that solves $p_j - MC_j = \frac{1}{\alpha} + \frac{1}{\alpha}\exp(m_j - \alpha p_j)$ (this is the price that firm j would set if it were a monopoly with only the outside good as a competitor). Since $p_j - MC_j - 1/\alpha\exp(m_j - \alpha p_j) - 1/\alpha$ is an increasing function of p_j, the Bertrand value of p_j, for which this expression is nonpositive by the above display, is no greater than $p_{mon}(MC_j, m_j)$, which sets this expression equal to zero. If MC_j and m_j are bounded by some B as assumed, then, since $p_{mon}(MC_j, m_j)$ is nondecreasing in MC_j and m_j by comparative statics arguments, $0 \leq p_j \leq p_{mon}(MC_j, p_j) \leq p_{mon}(B, B)$. Thus, for some M, $\max_{j\leq J}|m_j - \alpha p_j| \leq M$, so

$$\max_{j\leq J}\left|p_j - MC_j - \frac{1}{\alpha}\right| = \max_{j\leq J}\frac{1}{\alpha}\left|\frac{\exp(m_j - \alpha p_j)}{\sum_{i\neq j}\exp(m_i - \alpha p_i)}\right| \leq \left|\frac{\exp(M)}{\sum_{i\neq j}\exp(-M)}\right| = \frac{\exp(2M)}{J}$$

which converges to zero when scaled up by $\sqrt{J}$. The statement about IV estimates then follows from the corollary to Theorem 1. $\square$

Proof of Theorem 4

Proof. For use in section 6, I prove the slightly stronger result that the convergence is also uniform in the parameters over a neighborhood of any fixed value of the parameters for which the conditions of the theorem hold, that is, for any $\theta_0 \equiv (\alpha, \beta, \gamma, \sigma)$ for which the conditions of the theorem hold

$$\sqrt{J}\sup_{\|\theta-\theta_0\|<\varepsilon, j\leq J}\|p_j^*(x,\xi,\eta,\theta) - p_j(x,\xi,\eta,\theta)\| \to 0$$

where $\theta \equiv (\alpha, \beta, \gamma, \sigma)$ and $p_j^*(x,\xi,\eta,\theta)$ and $p_j(x,\xi,\eta,\theta)$ are defined as the solutions to the system of equations given by 8 and the nested logit equilibrium prices respectively.

Define $f : \mathbb{R}^{(4+d)|\mathcal{J}_g|+2d+2} \to \mathbb{R}^{|\mathcal{J}_g|}$ by

$$
\begin{aligned}
&f_j(p, x, \xi, \eta, \theta, r) \\
&= p_j - MC_j(x, \eta, \theta) \\
&- \frac{1-\sigma}{\alpha} \frac{\sum_k \exp((x_k'\beta - p_k\alpha + \xi_k)/(1-\sigma))}{[\sum_k \exp((x_k'\beta - p_k\alpha + \xi_k)/(1-\sigma))] - \sigma \exp((x_j'\beta - p_j\alpha + \xi_j)/(1-\sigma))} + r_j.
\end{aligned}
$$

Then p_g^* satisfies $f(p_g^*, x_g, \xi_g, \eta_g, 0) = 0$ and any solution p to the Nash pricing equations satisfies $f(p_g^*, x_g, \xi_g, \eta_g, \tilde{r}) = 0$ for

$$
\tilde{r}_j = \frac{1-\sigma}{\alpha} \frac{(1-\sigma)s_j(p,x)}{(1-\sigma\bar{s}_{j/g}(p,x))(1-\sigma\bar{s}_{j/g}(p,x) - (1-\sigma)s_j(p,x))}
$$

where the functions s_j and $\bar{s}_{j/g}$ take prices and product characteristics to the expressions for nested logit shares defined earlier in the section.

The proof proceeds by first showing that $\sqrt{J}\max_{j\leq J}\tilde{r}_j$ converges to zero, and then using the implicit function theorem and the mean value theorem to get a linear approximation to the p that solves $f(p, x, \xi, \eta, r) = 0$ as a function of r. The first statement follows since

$$
|\tilde{r}_j| \leq \frac{1-\sigma}{\alpha} \frac{s_j(p,x)}{1-\sigma-(1-\sigma)s_j(p,x)}.
$$

so that $\sqrt{J}\max_{j\leq J}\tilde{r}_j$ will converge to zero as long as $\sqrt{J}\max_{j\leq J}s_j$ converges to zero. Inspection of the formula for s_j shows that this will hold as long as equilibrium prices are bounded.

For r small and $MC(x, \theta, \eta)$ bounded away from zero, the equation $f(p, x, \theta, \eta, r) = 0$ has a unique solution for p. To see that a solution exists, note that this equation is equivalent to the first order condition for setting prices in the Bertrand pricing game with demand given by $q_j(p) \equiv \exp((x_j'\beta - \alpha p_j)/(1-\sigma))/D_g^\sigma$ and marginal cost equal to $MC_j + r_j$. An equilibrium exists in this game, since it is log supermodular (see pp. 151-152 of Vives (2001)):

$$\begin{aligned}
\frac{\partial^2 \log \pi_j}{\partial p_j \partial p_k} &= \frac{\partial^2 \log q_j(p)}{\partial p_j \partial p_k} \\
&= \frac{\partial^2}{\partial p_j \partial p_k} \left\{ \log \exp((x_j'\beta - \alpha p_j)/(1-\sigma)) - \sigma \log \sum_\ell \exp((x_\ell'\beta - \alpha p_\ell)/(1-\sigma)) \right\} \\
&= -\frac{\partial}{\partial p_j} \sigma \frac{\frac{-\alpha}{1-\sigma} \exp((x_k'\beta - \alpha p_k)/(1-\sigma))}{\sum_\ell \exp((x_\ell'\beta - \alpha p_\ell)/(1-\sigma))} \\
&= \frac{\alpha\sigma \exp((x_k'\beta - \alpha p_k)/(1-\sigma))}{1-\sigma} \frac{\frac{\alpha}{1-\sigma} \exp((x_j'\beta - \alpha p_j)/(1-\sigma))}{\left(\sum_\ell \exp((x_\ell'\beta - \alpha p_\ell)/(1-\sigma))\right)^2} > 0.
\end{aligned}$$

Uniqueness follows from verifying a dominant diagonal condition for f (see p. 47 of Vives (2001)). We have

$$\begin{aligned}
\frac{\partial f_j}{\partial p_j} &= 1 - \frac{1-\sigma}{\alpha} \sigma \frac{1}{\left(1 - \sigma \bar{s}_{j/g}(p)\right)^2} \frac{\partial}{\partial p_j} \bar{s}_{j/g}(p) \\
&= 1 - \frac{1-\sigma}{\alpha} \sigma \frac{1}{\left(1 - \sigma \bar{s}_{j/g}(p)\right)^2} \frac{-\alpha}{1-\sigma} \bar{s}_{j/g}(p)(1 - \bar{s}_{j/g}(p)) = 1 + \sigma \frac{\bar{s}_{j/g}(p)(1 - \bar{s}_{j/g}(p))}{\left(1 - \sigma \bar{s}_{j/g}(p)\right)^2}
\end{aligned}$$

and, for $k \neq j$,

$$\begin{aligned}
\frac{\partial f_j}{\partial p_k} &= -\frac{1-\sigma}{\alpha} \sigma \frac{1}{\left(1 - \sigma \bar{s}_{j/g}(p)\right)^2} \frac{\partial}{\partial p_k} \bar{s}_{j/g}(p) \\
&= -\frac{1-\sigma}{\alpha} \sigma \frac{1}{\left(1 - \sigma \bar{s}_{j/g}(p)\right)^2} \frac{\alpha}{1-\sigma} \bar{s}_{j/g}(p) \bar{s}_{k/g}(p)) = -\sigma \frac{\bar{s}_{j/g}(p) \bar{s}_{k/g}(p))}{\left(1 - \sigma \bar{s}_{j/g}(p)\right)^2}.
\end{aligned}$$

Thus,

$$\frac{\partial f_j}{\partial p_j} - \sum_{k \neq j} \left| \frac{\partial f_j}{\partial p_k} \right| = 1 + \frac{\sigma \bar{s}_{j/g}(p)}{\left(1 - \sigma \bar{s}_{j/g}(p)\right)^2} \left(1 - s_{j/g}(p) - \sum_{k \neq j} s_{k/g}(p) \right) = 1 > 0.$$

Since a unique p solves $f(p, x, \xi, \eta, \theta, r) = 0$ for the elements of (x, ξ, η) in the given bounded set, θ in the given neighborhood of θ_0, and r close to zero, this defines p as a function $\phi(x, \xi, \eta, \theta, r)$ of the remaining variables. By the implicit function theorem, the derivative matrix of ϕ is given by

$$D\phi(x, \xi, \eta, \theta, r) = (D_p f(\phi(x, \xi, \eta, \theta, r), x, \xi, \eta, \theta, r))^{-1} D_{x,\xi,\eta,\theta,r} f(\phi(x, \xi, \eta, \theta, r), x, \xi, \eta, \theta, r)$$

where subscripts denote blocks of the derivative matrix corresponding to derivatives with respect to given variables (the derivative matrix of f with respect to p is invertible since it is diagonally dominant). Since $p = \phi(x, \xi, \eta, \theta, \tilde{r})$ and $p^* = \phi(x, \xi, \eta, \theta, 0)$, by the mean value theorem, for every index j, there is a $\overline{r}$ between 0 and $\tilde{r}$ such the difference between p_j and p_j^* is given by the jth row of

$$(D_p f(\phi(x, \xi, \eta, \theta, \overline{r}), x, \xi, \eta, \theta, \overline{r}))^{-1} D_r f(\phi(x, \xi, \eta, \theta, \overline{r}), x, \xi, \eta, \theta, \overline{r})\tilde{r}.$$

Since the elements of $(D_p f(\phi(x, \xi, \eta, \theta, r), x, \xi, \eta, \theta, r))^{-1} D_r f(\phi(x, \xi, \eta, \theta, r), x, \xi, \eta, \theta, r)$ are continuous functions of x, ξ, η, θ, and r, the function that maps t to the maximum of the absolute values of the elements of $(D_p f(\phi(x, \xi, \eta, \theta, r), x, \xi, \eta, \theta, r))^{-1} D_r f(\phi(x, \xi, \eta, \theta, r), x, \xi, \eta, \theta, r)t$ takes a maximum M as x, ξ, η, θ, and r range over the compact set that contains them and t ranges over the unit sphere in $\mathbb{R}^{|\mathcal{J}_g|}$. This gives

$$\sqrt{J} \max_{j \le J, \|\theta - \theta_0\| < \varepsilon} |p_j^* - p_j| \le \sqrt{J} \max_{j \le J} M \|\tilde{r}_j\| \to 0.$$

The rate of uniform convergence for $\bar{s}_{j/g}$ follows since $\bar{s}_{j/g}$ is equal to $\bar{s}^*_{j/g}$ with p_k^* replaced by p_k in the definition, and the formula in the definition has a derivative with respect to the vector of prices in group g that is bounded in an open set containing all values of $(x, \xi, \eta, \theta, p)$ that can be taken under the assumptions of the theorem. Thus, by the mean value theorem, for some finite B, $\sqrt{J} \max_{j \le J, \|\theta - \theta_0\| < \varepsilon} |\bar{s}^*_{j/g} - \bar{s}_{j/g}| \le \sqrt{J} B \max_{j \le J, \|\theta - \theta_0\| < \varepsilon} |p_j^* - p_j| \to 0.$

$\square$

Proof of Theorem 5

Proof. The markup can be written as

$$\alpha^{-1} \left(1 - \frac{\int \sigma_j^2(\delta, \zeta)\, dP_\zeta(\zeta)}{\int \sigma_j(\delta, \zeta)\, dP_\zeta(\zeta)} \right)^{-1}.$$

Thus, it suffices to show that $\left(\int \sigma_j^2(\delta, \zeta)\, dP_\zeta(\zeta) \right) / \left(\int \sigma_j(\delta, \zeta)\, dP_\zeta(\zeta) \right)$ converges to zero at a faster than $\sqrt{J}$ rate. Suppose that product characteristics are bounded and fix a sequence $k_J \to \infty$. Let $\overline{s}_J$ and $\underline{s}_J$ be the supremum and infimum respectively of $\sigma_j(\delta, \zeta)$ with $\|\zeta\| \le k_J$,

$j \leq J$, and the elements of δ ranging over the given bounded set. Then

$$\max_{j \leq J} \frac{\int \sigma_j^2(\delta,\zeta)\, dP_\zeta(\zeta)}{\int \sigma_j(\delta,\zeta)\, dP_\zeta(\zeta)} \leq \frac{\int \sigma_j^2(\delta,\zeta) I(\|\zeta\| \leq k_J)\, dP_\zeta(\zeta) + P_\zeta(\|\zeta\| > k_J)}{\int \sigma_j(\delta,\zeta) I(\|\zeta \leq k_J\|)\, dP_\zeta(\zeta)}$$
$$\leq \frac{\int \overline{s}_J^2 I(\|\zeta\| \leq k_J)\, dP_\zeta(\zeta) + P_\zeta(\|\zeta\| > k_J)}{\int \underline{s}_J I(\|\zeta \leq k_J\|)\, dP_\zeta(\zeta)} \leq \frac{\overline{s}_J^2 + P_\zeta(\|\zeta\| > k_J)}{\underline{s}_J(1 - P_\zeta(\|\zeta > k_J\|))}.$$

If we can choose k_J so that $\sqrt{J} P_\zeta(\|\zeta > k_J\|)/\underline{s}_J$ and $\sqrt{J}\overline{s}_J^2/\underline{s}_J$ both go to zero, we will have the desired result. Since product characteristics are bounded, there exists some B such that, for all j,

$$\left| \sum_k x_{jk}\zeta_k \right| \leq B\|\zeta\|.$$

Letting M be a bound for δ_j, this gives the following bounds for $\overline{s}_J$ and $\underline{s}_J$:

$$\overline{s}_J \leq \frac{\exp(M + Bk_J)}{\sum_\ell \exp(-M - Bk_J)} = \frac{\exp(2M + 2Bk_J)}{J}$$
$$\underline{s}_J \geq \frac{\exp(-M - Bk_J)}{\sum_\ell \exp(M + Bk_J)} \geq \frac{\exp(-2M - 2Bk_J)}{J}.$$

This gives $\overline{s}_J^2/\underline{s}_J \leq \exp(6M + 6Bk_J)/J$. If the distribution of ζ is joint normal, then, for some constants K_1 and K_2,

$$P_\zeta(\|\zeta\| > k_J) \leq K_1 \exp(-K_2 k_J^2).$$

If this holds, then

$$\frac{P_\zeta(\|\zeta\| > k_J)}{\underline{s}_J} \leq \frac{K_1 \exp(-K_2 k_J^2) J}{\exp(-2M - Bk_J)} = K_1 \exp(-K_2 k_J^2 + Bk_J + 2M) J.$$

For $k_J = (\log J)^{2/3}$, we will have $\sqrt{J}\overline{s}_J^2/\underline{s}_j \to 0$ and $\sqrt{J} P_\zeta(\|\zeta\| > k_J)/\underline{s}_J \to 0$ as desired. □

References

Ackerberg, D. A., and M. Rysman (2005): "Unobserved Product Differentiation in Discrete-Choice Models: Estimating Price Elasticities and Welfare Effects," *The RAND Journal of Economics*, 36(4), 771788, ArticleType: primary_article / Full publication date:

Winter, 2005 / Copyright 2005 The RAND Corporation.

Bajari, P., and C. L. Benkard (2005): “Demand estimation with heterogeneous consumers and unobserved product characteristics: A hedonic approach,” *Journal of Political Economy*, 113(6), 12391276.

Bajari, P., and L. Benkard (2003): “Discrete Choice Models as Structural Models of Demand: Some Economic Implications of Common Approaches,” .

Berry, S., J. Levinsohn, and A. Pakes (1995): “Automobile Prices in Market Equilibrium,” *Econometrica*, 63(4), 841890.

Berry, S., O. B. Linton, and A. Pakes (2004): “Limit Theorems for Estimating the Parameters of Differentiated Product Demand Systems,” *The Review of Economic Studies*, 71(3), 613654.

Berry, S. T. (1994): “Estimating Discrete-Choice Models of Product Differentiation,” *The RAND Journal of Economics*, 25(2), 242262.

Bresnahan, T. F. (1987): “Competition and Collusion in the American Automobile Industry: The 1955 Price War,” *The Journal of Industrial Economics*, 35(4), 457482.

Caplin, A., and B. Nalebuff (1991): “Aggregation and Imperfect Competition: On the Existence of Equilibrium,” *Econometrica*, 59(1), 2559.

Stock, J. H., and J. H. Wright (2000): “GMM with Weak Identification,” *Econometrica*, 68(5), 10551096.

Stock, J. H., J. H. Wright, and M. Yogo (2002): “A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments,” *Journal of Business and Economic Statistics*, 20(4), 518–529.

Vives, X. (2001): *Oligopoly pricing*. MIT Press.

Wright, J. H. (2003): “Detecting Lack of Identification in Gmm,” *Econometric Theory*, 19(02), 322–330.