NBER WORKING PAPER SERIES

IDENTIFICATION AND ESTIMATION OF DISCRETE CHOICE DEMAND MODELS WHEN OBSERVED AND UNOBSERVED CHARACTERISTICS ARE CORRELATED

Amil Petrin
Mark Ponder
Boyoung Seo

Working Paper 30778
http://www.nber.org/papers/w30778

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2022

Identification and Estimation of Discrete Choice Demand Models when Observed and Unobserved Characteristics are Correlated
Amil Petrin, Mark Ponder, and Boyoung Seo
NBER Working Paper No. 30778
December 2022
JEL No. C25,L0

ABSTRACT

The standard Berry, Levinsohn, and Pakes (1995) (BLP) approach to estimation of demand and supply parameters assumes that the product characteristic observed by consumers and producers but not the researcher is conditionally mean independent of observed characteristics. We extend BLP to allow all product characteristics to be endogenous, so the unobserved characteristic can be correlated with the observed characteristics. We derive moment conditions based on the assumption that firms choose product characteristics to maximize expected profits given their beliefs at that time about market conditions and that the "mistake" in the amount of the characteristic that is revealed once all products are on the market is conditionally mean independent of the firm's information set. Using the original BLP dataset we find that observed and unobserved product characteristics are highly positively correlated, biasing demand elasticities upward, as average estimated price elasticities double in absolute value and average markups fall by 50%.

Amil Petrin
Department of Economics
University of Minnesota
4-101 Hanson Hall
Minneapolis, MN 55455
and NBER
petrin@umn.edu

Mark Ponder
NERA Economic Consulting
360 Hamilton Ave
White Plains, NY 10601
mark.ponder@nera.com

Boyoung Seo
Kelley Business School
1309 E 10th Street HH3080
Bloomington, IN 47405
seob@indiana.edu

1 Introduction

The identification of discrete choice demand models since Berry, Levinsohn, and Pakes (1995) (BLP) has relied on the assumption that the product characteristic unobserved to the researcher but observed to producers and consumers is conditionally mean independent of all observed product characteristics. Under this identification assumption any function of observed characteristics of all products in the market is a valid instrument for any product's price. Given the abundance of instruments - many of them likely to be very weak - BLP use the structure of their competitive setting to develop product-specific instruments for price that are likely to be highly correlated with that product's price. More recently Gandhi and Houde (2015) show how to extend this logic to develop even more powerful instruments.

Since the inception of its use this assumption has been criticized as being inconsistent with profit-maximizing behavior; it is not clear why firms would choose a level of the unobserved quality for a product independently of the choice of the products' observed characteristics. Empirically we see a high positive correlation among the observed attributes of products, suggesting unobserved product quality is likely to be positively correlated with observed characteristics. If firms do choose to put more unobserved-by-the-researcher quality on products that have more attractive observed characteristics, then instrumenting price with observed product characteristics will not break the positive correlation between price and unobserved quality that BLP are trying to address. Demand elasticities will then continue to be biased in a positive direction because higher prices mean consumers are getting higher unobserved quality, leading consumers to look less price sensitive than they actually are in reality.

In this paper we extend BLP to allow all product characteristics to be endogenous so the unobserved characteristic can be correlated with the other observed characteristics. Spence (1976) formalized the notion that firms' decisions about characteristics' choices are driven by their beliefs about consumer preferences for them and the costs of providing them by showing their first-order conditions for profit-maximization contain terms related to marginal and infra-marginal consumers and costs.[1] We use these first-order conditions for the optimal choice of price and observed and unobserved product characteristics to try to infer firms' beliefs about the distribution of consumers tastes and the structure of costs.

[1] See also the more recent generalization by Veiga and Weyl (2014).

We estimate a model of BLP-type demand and supply under the assumption that firms choose characteristics first given some information set. They do so knowing that once all of the product characteristics and other demand and supply factors have been realized they will compete in prices in a Bertrand–Nash manner. Using the insight from Hansen and Singleton (1982), our identification is based on the assumption that firms' ex-post optimization mistakes are conditionally mean independent of anything the firm knows at the time the firm chooses its product characteristics. This will be true as long as firms do not condition on something that we do not observe that affects their profitability and their characteristics' choices (see Pakes et al. (2015)). An advantage of these setups is that we do not have to completely specify the firm's information set at the time it chooses characteristics; it may include other firms' product lagged or contemporaneous characteristics and demand/cost shocks, signals on all of these, or no information on them at all.

Our approach is complementary to the many papers previous to ours that have exploited Spence's insight that optimization can help with identification of model parameters, including Mazzeo (2002), Sweeting (2007), Crawford and Shum (2007), Lustig (2008), Gramlich (2009), Fan (2013), Eizenberg (2014), and Blonigen et al. (2013).[2] These papers are more general than our approach in the sense that they consider (e.g.) the use of optimization to help with identification of fixed costs, sunk costs, or identification in the face of restricted sets of characteristics from which firms can choose for product characteristics. However, all of these papers maintain some kind of independence between the level or change in the demand or supply shock and the observed product characteristics. Our identification assumption is straightforward to adopt to all of these settings and would allow researchers to sidestep imposing mean independence of observed and unobserved characteristics while at the same time estimating (e.g.) fixed or sunk costs.

The steps necessary to calculate the value of our objective function are identical to the steps in BLP's two-step GMM estimator except we replace the mean independence moments with our optimization moments. The BLP inversion allows us to – for any given parameter value – solve for the unobserved characteristics for every vehicle so we can treat them as another observed characteristic that the firm is choosing optimally. Analogous to the BLP estimation routine, we are able to concentrate out linear parameters leading to an estimator

[2]See the review in Crawford (2012) for a complete list of all papers that use optimization in characteristics for identification.

that is not substantially more computationally burdensome. Using characteristics of competitors vehicles from prior years - which should be known to the firm at the time they made characteristic choices in those prior years - we develop an approximation to the optimal instruments implied by the model's structure. The standard BLP instruments are also valid instruments in our setting and we provide results using these instruments as well. The only other difference with the BLP estimation routine is we include these characteristics in the marginal cost function. Formulating our estimator in the GMM framework means our estimator can easily be supplemented with moments that may further help with identification, as in Petrin (2002) or Berry et al. (2004).

The most striking difference between the BLP estimates and the optimization estimates is that the coefficient on price is much larger under optimization. The impact of this change is that relative to BLP on average elasticities double and estimated markups fall by 50%. We investigate whether a positive correlation between observed and unobserved characteristics is a possible explanation by constructing a "BLP instrumented price", that is, we regress price on the BLP instruments and construct predicted values. We find our unobserved qualities are significantly positively correlated with the instrumented prices with a correlation of approximately 0.5.

A second related difference in model fit relates to the fact that only 10% of U.S. households buy new cars in any given year so both fitted demand models need a way to explain why 90% of households choose the outside good. BLP fits 90% of households choosing the outside good by having consumers view the average unobserved quality of new cars as much worse than the outside good. In contrast, the optimization-fit has consumers strongly desiring new cars relative to the outside good but the significantly higher price elasticity causes 90% not to buy a new car.

Our estimates are almost always much more precisely estimated relative to the BLP-fit model. We also find some of the anomalies in the BLP point estimates are not present in the optimization-fit point estimates. The BLP point estimates imply consumers dislike fuel efficiency but in our setup they strongly and significantly prefer fuel efficiency. The BLP point estimates also imply it cost less to build a bigger and more fuel efficient vehicle while we find the opposite.

The differences we report here between the optimization-fit model and the BLP-fit model have also been found in European automobile data (see Miravete et al. (2015)). They adopt our approach to estimating demand and supply to look at competition in the Spanish automobile market. They report that using the optimization moments on average estimated price elasticities double and estimated markups fall by 50% relative to when they use the BLP moments. Anomalous demand and supply point estimates under the BLP-fit are not present under optimization-fit, the standard errors are much smaller, and their unobserved quality term is positively correlated with observed characteristics.

In Section (2), we specify demand and supply system. In Section (3) we discuss our identifying restrictions and give conditions under which we are locally identified. Section (4) discusses estimation and our choice of instruments. In Section (5), we provide Monte Carlo simulation results for the proposed estimator which demonstrate its properties. Lastly, we apply our approach to the same automobile data BLP used in (6).

2 Demand and Supply

In this section we provide the demand and supply framework for our estimator, which closely follows the model of Berry et al. (1995). We specify a random coefficients system for demand and oligopolistic competition with constant marginal costs for supply. The primary point of departure is that firms compete each period in two stages - a first stage where they optimally choose their product characteristics and a second stage where they compete on price.

2.1 Demand

Each product is defined as a vector of K observed characteristics and price $(X_j, p_j) \in \mathbb{R}^{K+1}$ and an unobserved (to the econometrician) characteristic ξ_j which is observed by both consumers and producers. Product $j = 0$ is the option of not buying a new vehicle and it is standard to normalize its characteristics and price to zero ($X_0 = p_0 = \xi_0 = 0$).

A consumer i is indexed by (D_i, v_i, ε_i), where D_i is a vector of their demographic characteristics, v_i is vector of their K idiosyncratic taste draws $(v_{ik})_{k=1}^{K}$ drawn from a known distribution, one for each of the K characteristics, and ε_i is the vector of their product-specific "tastes" $(\varepsilon_{ij})_{j=1}^{J}$ which are assumed to be independent and identically distributed extreme value across consumers and products. The demand model parameters are given as $\theta^D = (\theta_l, \theta_{nl})$. Utility that consumer i derives from good j is given as

$$u_{ij}\left(\theta^D\right) = \delta(X_j, p_j, \xi_j; \theta^D) + \mu(X_j, p_j, D_i, v_i; \theta^D) + \epsilon_{ij}$$

The term $\delta_j = \delta(X_j, p_j, \xi_j; \theta^D)$ is a product specific utility component and is common to all consumers. The term $\mu(X_j, p_j, D_i, v_i; \theta^D)$ captures the individual specific taste for the characteristics of good j.[3]

Consumer chooses the one and only one product j which yields the highest utility:

$$u_{ij}\left(\theta^D\right) \geq u_{ij'}\left(\theta^D\right), \quad \forall j'.$$

Define s_{ij} as individual i's probability of purchasing good j prior to the realization of ε and denote the vector of market prices as $p = \{p_j\}_j$ (with X and ξ defined analogously). The choice probabilities are given by

$$s_{ij}\left(p, X, \xi, D_i, v_i;\ \theta^D\right) = \frac{\exp\left(\delta(X_j, p_j, \xi_j; \theta^D) + \mu(X_j, p_j, D_i, v_i; \theta^D)\right)}{\sum_{j' \in J} \exp\left(\delta(X_{j'}, p_{j'}, \xi_{j'}; \theta^D) + \mu(X_{j'}, p_{j'}, D_i, v_i; \theta^D)\right)}.$$

Letting $F\left(D_i, v_i\right)$ denote the distribution of consumer characteristics the market share for good j is given as the integral over these consumers:

$$s_j\left(p, X, \xi;\ \theta^D\right) = \int s_{ij}\left(p, X, \xi, D_i, v_i;\ \theta^D\right)\ dF\left(D_i, v_i\right).$$

We turn now to our model of supply.

2.2 Supply

Define the matrix of product characteristics $Z_j = (X_j, W_j)$ and $Z = (Z_j)_{j \in J}$, where W_j are cost shifters, including X_j itself or the log of it and ω_j is the unobserved cost shock for good j. We follow the literature and specify firm profits as

[3] It is commonly assumed that these functions are linear in the characteristics so that $\delta_j = X_j\beta - \alpha p_j + \xi_j$, and consumer i's taste for characteristic X_k is then given by $\beta_{ik} = \beta_k + \Pi_k D_i + \Sigma_k v_i$. See Nevo (2004).

$$\Pi_f = \sum_{j' \in J_f} \left(p_{j'} - mc_{j'}\left(W_{j'}, \omega_{j'}; \theta\right)\right) \; s_{j'}\left(p, Z, \xi; \theta\right)$$

Here f indexes firms and J_f is the set of products produced by firm f. We assume firms compete with each other every period in two stages. In the first-stage, firms choose their product characteristics to maximize their expected profit. In the second stage, all uncertainty is resolved and firms compete Nash-Bertrand in prices.

Specifically, in the second stage, firms have the following J pricing first-order conditions:

$$s_j + \sum_{j' \in J_f} \left(p_{j'} - mc_{j'}\left(W_{j'}, \omega_{j'}; \theta\right)\right) \frac{\partial s_{j'}}{\partial p_j} \quad = \quad 0, \quad \forall j \in J_f. \tag{1}$$

In the first stage, firms know that they will compete in prices in a Bertrand-Nash manner given the chosen characteristics of all products in the market and that prices are set according to the first-order conditions (1). With this knowledge, firms choose characteristics to maximize expected profits given their information set, denoted I_f for firm f. This information set may differ across firms, and it may include other firms' product characteristics, own- and other-firm cost shifters, some signals on the variables, or no information at all on them.

Let $\theta = \left(\theta^D, \theta^S\right)$ be the vector of all parameters, including the cost-side parameters. In the first step, firm f chooses vectors $X_f = \left(X_j\right)_{j \in J_f}$ and $\xi_f = \left(\xi_j\right)_{j \in J_f}$ to solve:

$$\max_{X_f,\ \xi_f} \qquad E\left[\Pi_f \;\middle|\; I_f\right]$$

with prices determined after characteristics are set in a Bertrand-Nash manner. Given $(Z_f, \xi_f)_f$, the realized value of the first-order condition for characteristic k of product j is given by $\nu_{jk}(\theta)$ and written as

$$\frac{\partial \Pi_f}{\partial X_{jk}} = \quad \textstyle\sum_{j' \in J_f} \Big[\left(p_{j'} - mc_{j'}\left(W_{j'}, \omega_{j'}; \theta\right)\right) \frac{d\ s_{j'}(p,Z,\xi;\theta)}{d\ X_{jk}} + \tag{2}$$

$$s_{j'}\left(p, Z, \xi; \theta\right) \; \frac{\partial\left(p_{j'} - mc_{j'}\left(W_{j'}, \omega_{j'}; \theta\right)\right)}{\partial X_{jk}}\Big] \tag{3}$$

for $k \leq K$, and

$$\frac{\partial \Pi_f}{\partial \xi_{jk}} = \sum_{j' \in J_f} \Bigg[(p_{j'} - mc_{j'}(W_{j'}, \omega_{j'}; \theta)) \frac{d\ s_{j'}(p, Z, \xi; \theta)}{d\ \xi_{jk}} + s_{j'}(p, Z, \xi; \theta)\ \frac{\partial \left(p_{j'} - mc_{j'}\left(W_{j'}, \omega_{j'}; \theta \right) \right)}{\partial \xi_{jk}} \Bigg]$$

for $k = K + 1$. Firms anticipate the change in equilibrium prices that will occur in the second step if they change their product characteristics and this shows up in the first-order condition in the derivative of shares with respect to characteristics X_j (and ξ_j):

$$\frac{d\ s_{j'}(p, Z, \xi; \theta)}{d\ X_{jk}} = \frac{\partial s_{j'}}{\partial X_{jk}} + \sum_{j'' \in J} \frac{\partial s_{j'}}{\partial p_{j''}} \frac{\partial p_{j''}}{\partial X_{jk}}.$$

The first-order condition illustrates that multi-product firms internalize the externality of changing X_j on the profits of its other products $j' \in J_f$. The term in (2) represents the change in profits attributable to *marginal* consumers while the second term in (3) captures those attributable to *inframarginal* consumers. Rational expectations requires that

$$E\left[\left. \frac{\partial \Pi_f}{\partial X_{jk}} \right| I_f \right] = 0,\ \forall k, j \in J_f,\ \forall f \tag{4}$$

In equilibrium, the optimal level of X_f chosen by the firm maximizes expected profits given what the firm knows at the time the characteristics are chosen. We turn now to the identifying restrictions of our model.

3 Identification

We turn now to the issue of identification. Readers interested in the empirical results can skip this section and go directly to Section (5).

3.1 Moment Conditions

Our identification is based on the $(K + 1) \times J$ first-order conditions in (3) coupled with the J first-order conditions with respect to p in (1) and the BLP inversion. Given θ and the data, market shares are inverted to recover mean utility and the level of unobserved quality.

Marginal costs can then be recovered from (1):

$$mc\,(s,p,Z;\theta) = p - \Delta^{-1}\,(s,p,Z;\theta)\,s$$

where

$$\Delta_{ij} = \begin{cases} -\frac{\partial s_j}{\partial p'_j}, & \text{if } j, j' \in J_f \\ 0 & \text{otherwise} \end{cases}$$

The realized value of the firm f's first-order conditions for good j and characteristic k, evaluated at θ, is defined as

$$\nu_{jk}\,(\theta) := \sum_{j' \in J_f} \left[\left(p_{j'} - mc_{j'}\,(s,p,Z;\theta)\right) \frac{d\ s_{j'}\,(p,Z;\theta)}{d\ X_{jk}} + \right. \\ \left. s_{j'}\,(s,p,Z;\theta)\ \ \frac{\partial\left(p_{j'} - mc_{j'}\,(s,p,Z;\theta)\right)}{\partial X_{jk}} \right]$$

We use the insight from Hansen and Singleton (1982) that the first-order conditions from profit maximization, along with rational expectations, implies that if $\theta = \theta_0$ then

$$E\left[\nu_{jk}\,(\theta)\ \mid I_f\right] = 0 \quad \forall k, j \in J_f,\ \forall f \tag{5}$$

Sometimes firms will provide too little of a characteristic and sometimes it will provide too much, but on average these "mistakes" average out. The residual ν_{jk} may include expectational errors that arise due to asymmetric information across competing firms on each others' costs and product characteristics or it may be incomplete information on the outcomes own-firm payoff-relevant variables (like realized cost shocks). ν_{jk} may also include model approximation error or measurement error in the data.[4]

We have more than $K+1$ unknown parameters but only $K+1$ first-order condition conditions. As such, we need instruments that are orthogonal to ν_{jk} in order to estimate the model parameters. Hansen and Singleton (1982) implies that any function of the arguments

[4] As Pakes et al. (2015) note, if there is something known to the firm but not seen by the researcher and if it affects the firm's profits and thus its decisions, the mean of these selected observations will not generally be zero.

of I_f are possible instruments that can be used to identify the model parameters. Letting $H = \{H_l\}_{l=1}^{L} \subset I_f$ be a set of instruments, the moment restrictions become

$$E\left[H_{klj}\nu_{jk}\left(\theta\right) \mid I_f\right] = 0 \quad \forall k, l, j \in J_f, \ \forall f \tag{6}$$

We propose such instruments in the subsequent section. Before doing so, we discuss the local identification of the model.

3.2 Local Identification

Let $\nu_k\left(\theta, s, p, Z\right)$ be the vector of J residuals $\nu_{jk}\left(\theta, s, p, Z\right)$ and $\nu\left(\theta, s, p, Z\right)$ be a vector of the $K+1$ residual vectors $\nu_k\left(\theta, s, p, Z\right)$.[5] Standard rank conditions require that the matrix

$$D = E\left[H \frac{\partial \nu(\theta_0)}{\partial \theta'} \middle| I_f\right]$$

be non-singular for the model to be locally identified. This requires that there are sufficient excluded instruments that are correlated with $\frac{\partial \nu(\theta_0)}{\partial \theta'}$ and that the elements of $\frac{\partial \nu}{\partial \theta'}$ are not collinear. We show in the appendix that it is possible to write $\nu_k(\theta)$ as an affine function of θ_l:

$$\nu_k\left(\theta\right) = \nu_{kc}(\theta_{nl}, X, W, p, s) + \nu_{k\theta}(\theta_{nl}, X, W, p, s)\theta_l = 0, \tag{7}$$

for some new functions ν_{kc} and $\nu_{k\theta}$, neither of which have as arguments θ_l. The linearized residuals help clarify when the model is locally identified. From equation (7), we have

$$\frac{\partial \nu_k}{\partial \theta_l'} = \nu_{k\theta}(\theta_{nl}, X, W, p, s)$$

and

$$\frac{\partial \nu_k}{\partial \theta_{nl}'} = \frac{\partial \nu_{kc}(\theta_{nl}, X, W, p, s)}{\partial \theta_{nl}'} + \frac{\partial \nu_{k\theta}(\theta_{nl}, X, W, p, s)}{\partial \theta_{nl}'}\theta_l$$

[5]For notational simplicity we suppress the dependence of ν_k on the data when the meaning of $\nu_k(\theta)$ is clear.

In general, the expectation of these terms will not be collinear at θ_0. However, in the case where $\nu_{kc} = 0$ it is straightforward to see that local identification fails. When $\nu_{kc} = 0$, then the moment condition $E[H'\nu_k(\theta_0)] = 0$ implies

$$E[H'\nu_{k\theta}(\theta_{nl,0}, X, W, s)]\theta_{l,0} = 0$$

Assuming that the elements of θ_l are non-zero at truth, this implies that the columns of $E[H'\nu_{k\theta}(\theta_{nl}, X, W, p, s)] = E\left[H'\frac{\partial \nu_k}{\partial \theta_l'}\right]$ have rank less than the number of linear parameters, so the rank-condition fails. We now turn to the estimation routine and our proposed instrumental variables.

4 Estimation

There are three sections to this estimation section. Section 4.1 describes how we implement the GMM objective function. Section 4.2 then describes our approximation to the optimal instruments. Section 4.3 show hows to concentrate out all of the "linear" parameters during estimation to reduce the dimensionality of the nonlinear search. Readers only interested in these last details can skip directly there.

4.1 The Estimator

Estimation follows two-stage GMM. For an initial guess at θ_0 we calculate the approximation to the optimal instruments described in the following section. Given those instruments we calculate the optimal weighting matrix and the first stage estimates.[6] At the first stage estimates we recalculate the optimal instruments and the efficient weighting matrix and the re-estimate to get the two-step GMM estimates.

In each stage, estimation has the following steps. Let $\nu(\hat{\theta})$ be a vector created by evaluating $\nu_{jk}(\theta)$ at a guess $\hat{\theta}$. Given a set of instruments H and a weight matrix W, the empirical moment conditions may be written as[7]

[6]As in BLP we use importance sampling to minimize simulation error. We draw importance samples at an initial estimate θ_1, and then evaluate instruments H and optimal weighting matrix Ω at θ_1 for GMM estimation. Once the first step estimates are converged at θ_2, we re-draw importance samples, re-derive instruments, and re-evaluate optimal weighting matrix at θ_2. Then, we repeat the search over θ.

[7]We suppress the dependence of the moment conditions on the data for notational simplicity.

$$g(\theta) = \frac{1}{N} H' \nu(\hat{\theta})$$

and the GMM objective function is given by

$$GMM(\theta) = g(\theta)' H W H' g(\theta)$$

Each stage of estimation then reduces to the following steps:

1. Fix θ_{nl} equal to initial guess θ_{nl}^1
2. Solve for $\delta(\theta_{nl}^1)$ that matches shares
3. Use the first-order conditions for prices to recover marginal costs.
4. Evaluate the GMM objective function.
5. Repeat from Step 1 until the objective function is minimized.

There is a pathological case that needs to be dealt with, where $\alpha \to \infty$ and $\gamma \to 0$ can set the moment conditions exactly equal to zero. We view this case as un-economic as it implies that costs are independent of the characteristics and consumers are infinitely price sensitive. There are several potential solutions to eliminating this case. First, the researcher may re-write the moments conditions to rule out this case. Given our model specification, each term in ν_{kc} is multiplied by $\frac{\sigma_k}{\alpha}$. Pulling this common term out, we can write ν_{kc} as $\nu_{kc}(\theta_{nl}, X, W, s) = \frac{\sigma_k}{\alpha}\tilde{\nu}_{kc}(\theta_{nl}, X, W, p, s)$, for some function $\tilde{\nu}_{kc}(\theta_{nl}, X, W, p, s)$. Importantly, the function $\tilde{\nu}_{kc}(\theta_{nl}, X, W, p, s)$ is always non-zero and varies non-linearly with θ_{nl}. Also,

$$\begin{aligned}
\frac{\alpha}{\sigma_k}\nu_k &= \frac{\alpha}{\sigma_k}\left(\nu_{kc}(\theta_{nl}, X, W, p, s) + \nu_{k\theta}(\theta_{nl}, X, W, p, s)\tilde{\theta}_l\right) \\
&= \tilde{\nu}_{kc}(\theta_{nl}, X, W, p, s) + \nu_{k\theta}(\theta_{nl}, X, W, p, s)\tilde{\theta}_l
\end{aligned}$$

is zero whenever ν_k is zero, but not when $\alpha \to \infty$. It is clear that we can recover the true θ_l given estimates of $\tilde{\theta}_l$ and θ_{nl}. [8]

[8] We note the similarity between this problem and estimating Euler Equations with CES utility. The Euler Equation is given by

Alternatively, the researcher may "compactify" the parameter space, choosing an upper bound on the absolute value of the parameters such that the true parameter values lie in the compact set. In practice, the researcher would set an upper-bound on the α (say 300) and search over local minima that satisfy the bound. This is a similar procedure as is proposed in Newey and McFadden (1994) to deal with a likelihood function that becomes unbounded as the variance parameter approaches 0 (see page 2,136). Assuming that only the true θ_0 satisfies in the first-order conditions in the compact set would then lead to consistency of our proposed estimator.

4.2 Instruments

Chamberlain (1987) shows that the efficient set of instruments are the expected value of the derivatives of the error term with respect to the parameters evaluated at the true parameter θ_0.

In our context this optimal instrument H is a $JK \times |\theta|$ matrix

$$H = E\left[\nu\left(\theta_0\right)\nu\left(\theta_0\right)' \mid I\right]^{-1} E\left[\left.\frac{\partial\nu\left(\theta_0\right)'}{\partial\theta}\right| I\right]'. \tag{8}$$

Letting $E\left[\nu\left(\theta_0\right)\nu\left(\theta_0\right)' \mid I\right] = I_{JK}$ for now, H_{jkl}, the element $(jk,\, l)$ of the derivative, is given as

$$H_{jkl} = E\left[\left.\frac{\partial\nu_{jk}\left(\theta_0\right)}{\partial\theta_l}\right| I_f\right] \quad \forall k, l, j \in J_f,\ \forall f. \tag{9}$$

$$E\left[R_{t+1}\beta_0(1-\alpha_0)c_{t+1}^{-\alpha_0}|I_t\right] = (1-\alpha_0)c_t^{-\alpha_0}$$

which can be set exactly to zero by taking $\alpha_0 \to \infty$ or setting $\alpha_0 = 1$. However, transforming the equation to an equivalent form

$$E\left[R_{t+1}\beta_0\left(\frac{c_{t+1}}{c_t}\right)^{-\alpha_0}\middle| I_t\right] - 1 = 0$$

rules out these pathological cases during estimation.

where the derivative of ν_{jk} with respect to θ_l is given as:

$$\begin{aligned}
\frac{\partial \nu_{jk}(\theta_0)}{\partial \theta_l} &= \sum_{j' \in J_f} \left[\frac{\partial (p_{j'} - mc_{j'})}{\partial \theta_l} \frac{d\ s_{j'}}{d\ X_{jk}} + (p_{j'} - mc_{j'}) \frac{d^2\ s_{j'}}{d\ \theta_l\ d\ X_{jk}} \right. \\
&\quad \left. + \frac{d\ s_{j'}}{d\ \theta_l} \frac{\partial (p_{j'} - mc_{j'})}{\partial X_{jk}} + s_{j'} \frac{\partial^2\ (p_{j'} - mc_{j'})}{\partial \theta_l\ \partial X_{jk}} \right] \\
\text{where } \frac{d\ s_j}{d\ \theta_l} &= \frac{\partial s_j}{\partial \theta_l} + \sum_{j' \in J} \frac{\partial s_j}{\partial X_{j'}} \frac{\partial X_{j'}}{\partial \theta_l} + \sum_{j' \in J} \frac{\partial s_j}{\partial p_{j'}} \frac{\partial p_{j'}}{\partial \theta_l} + \sum_{j' \in J} \sum_{j'' \in J} \frac{\partial s_j}{\partial p_{j'}} \frac{\partial p_{j'}}{\partial X_{j''}} \frac{\partial X_{j''}}{\partial \theta_l}
\end{aligned}$$

for $k \leq K$. If $k = K + 1$, $d\ X$ or ∂X is substituted to $d\ \xi$ or $\partial \xi$. We follow Fan (2013) and recover $\frac{\partial p_{j''}}{\partial X_{jk}}$ using the Implicit Function Theorem. In principle we are exactly identified, i.e. the total number of instruments is equal to the total number of model parameters. These instruments place larger weights on the first-order conditions which are most responsive to changes in the parameters contained in θ.

There are four significant challenges to calculating the optimal instruments. We do not know the true value of parameters θ_0 and we do not know the information set I_f of any firm. Even if we knew I_f we would have to specify the distribution of the remaining unknown random variables conditional on the information set to be able to integrate over it. Finally, $\frac{\partial X}{\partial \theta}$ and $\frac{\partial p}{\partial \theta}$ are complicated unknown equilibrium objects.

We follow Berry et al. (1999) and choose an informed guess θ_g and then approximate the optimal instrument H_{jkl} by using the value of the derivative itself $\frac{\partial \nu_{jk}}{\partial \theta_l}$ calculated under different assumptions about what is known to the firm at time when the characteristics' decisions are made. We set the terms $\frac{\partial X}{\partial \theta}$ and $\frac{\partial p}{\partial \theta}$ to zero because of the difficulties of estimating them so $\frac{d\ s_j}{d\ \theta_l} = \frac{\partial s_j}{\partial \theta_l}$ in our estimation routine.[9]

Let $X_t = (X_{jt})_{j \in J_t}$ be a vector of characteristics of all products available in year t, and define ξ_t, W_t, p_t, and ω_t similarly. Let $X_{f,t} = (X_{jt})_{j \in J_{ft}}$ be the set of firm f's products, and $X_{-f,t} = (X_{jt})_{j \notin J_{ft}}$ be the set of firm f's competitors' products. A natural assumption is that each firms' information sets contain their own contemporaneous costs shocks and their

[9]Leaving out these terms as well as letting $E\left[\nu(\theta_0)\nu(\theta_0)' \mid I\right] = I_{JK}$ is not a consistency issue but instead an efficiency issue.

competitors' characteristics from the previous year:

$$\{X_{-f,\ t-1},\ \xi_{-f,\ t-1},\ W_{-f,\ t-1},\ \omega_{-f,\ t-1},\ \omega_{f,\ t}\} \ \subset \ I_{f,t}^{\text{lagged}}$$

When calculating the derivative for a product characteristic for firm f ones would use observed and unobserved characteristics of products of the firms competing against f from the previous year. At those characteristics and firm f's current observed and unobserved characteristics the researcher would solve for the Bertrand-Nash equilibrium prices and then evaluate the derivative at those prices for firm f.

Given the lagged information set, lagged version of BLP-instruments are valid instruments for ν_{jk} – own product characteristics $X_{j,t}$, other product characteristics within own firm $X_{j'\neq j,f,t}$, and competitor's product characteristics $X_{-f,\ t-1}$ from the previous year. Moreover, BLP type instruments can be applied to price – own price, other prices within own firm, and competitor's prices from the previous year are valid instruments in our setup.

Another natural assumption is that

$$\{X_{-f,\ t},\ \xi_{-f,\ t},\ p_{-f,\ t},\ W_{-f,\ t},\ \omega_t\} \ \subset \ I_{f,t}^{\text{contemporaneous}}$$

so firm f knows its competitors' contemporaneous choices of characteristics and costs at the time of decision. This information set implies that the conditional expectation of the FOCs are taken with respect to approximation or measurement error. In this case, the derivative is evaluated at realized values of X_t, ξ_t, W_t, p_t, and ω_t. In addition, given the contemporaneous information set, BLP-type instruments evaluated at the realized values are valid instruments for ν_{jk}.

4.3 Concentrating Out Linear Parameters

BLP reduce the dimensionality of their parameter search by "concentrating out" parameters $\theta_l = (\beta, \gamma)$, which enter their moment conditions linearly. An implication of (7) is that θ_l enters the moment conditions linearly, conditional on θ_{nl} and the data. As such, we can concentrate out θ_l, thereby allowing us to restrict our search to the space of the non-linear parameters, denoted $\theta_{nl} = (\alpha, \sigma)$. This reduces the dimensionality of the nonlinear search by $\#\beta \times \#\gamma$ parameters and leads to a more robust estimation algorithm. For each guess of θ_{nl}, the value of θ_l that minimizes the GMM objective function for a given guess of θ_{nl} is given by

$$\theta_l(\theta_{nl}, X, W, p, s) = -\left(\nu_\theta' H W^{-1} H' \nu_\theta\right)^{-1} \nu_\theta' H W^{-1} H' \nu_c$$

where ν_c comes from stacking the K ν_{kc} functions and ν_θ comes from stacking the $\nu_{k\theta}$ functions. In general, the terms ν_{kc} and $\nu_{k\theta}$ will have complex functional forms. We provide expressions for ν_{kc} and $\nu_{k\theta}$ in Appendix Section (9) given our specification of marginal costs and demand.

4.4 Dynamic Moments

To this point, the discussion has assumed that product characteristics are readily adjustable so that firms update these characteristics each period. However, if firms face fixed adjustment costs then there may be periods of inaction where firms do not find it profitable to make adjustments. This implies that when a firm chooses to make an adjustment to its portfolio, it maximizes the sum of a future stream of profits rather than per-period profit. We follow the discussion and spirit of Pakes (1994) in proposing a method for the case when periods of non-adjustment are possible.

Let Φ be the scrap value of exiting and let ζ represent a fixed cost incurred by each firm when it adjusts its portfolio. Let $\chi(X_t)$ be 1 if the firm remains in the market given its optimal policy and 0 otherwise. Given per-period profits of $\Pi_{ft} = \sum_{j \in J_f} (p_j - \text{mc}_j(X_{jt})) M s_j(p_t, X_t)$ for firm f, the firm's expected profit from adjusting some subset of characteristics at time t is given by

$$V(X_{ft}) = \max\left\{\Phi, E\left[\max\left\{\Pi_{ft}(X_{ft}) + \beta V(X_{ft}), \max_{X'_{ft}}\left\{\Pi_{ft}(X'_{ft}) + \beta V(X'_{ft}) - \zeta\right\}\right\} \Big| I_{ft}\right]\right\}$$

Firms compare the value of adjusting their product attributes against holding their portfolio fixed. The maximum value is then compared to the scrap value and the firm decides whether or not to exit. Following Pakes (1994), define the random variable τ^* as the number of periods or inaction, i.e. the number of periods until at least one of the characteristics changes, a product is dropped or introduced, or the firm exits the market.

$$\tau^* = \min_{\tau \geq 1} \{\chi(X_{t+\tau}) = 0 \text{ or } X_{ft+\tau} \neq X_{ft}\}$$

If a product's characteristics are changed, then the first-order conditions are satisfied for that good. In the case where τ^* is known to be 1, then estimation reduces to using the moments (5). Otherwise, one can construct the moment conditions[10]

$$E\left[\sum_{\tau=0}^{\tau^*-1} \beta^\tau \frac{\partial \Pi_{f,t+\tau}}{\partial X_{jt}} + \beta^{\tau^*} \chi(X_{t+\tau^*}) \frac{\partial \Pi_{f,t+\tau^*}}{\partial X_{jt+\tau^*}} \middle| I_{ft}\right] = 0$$

and use the sample analog during estimation for given realizations of τ^* and the appropriate subset of observations. Given estimates of the cost and demand parameters, the researcher can use a bounds estimator approach to recover an estimate of the sunk adjustment cost as in Sweeting (2007).

One issue with this approach is that firm's need to account for the "reaction" functions of their competitors when solving these first-order conditions, which are in general complex equilibrium objects that are not known a-priori. As discussed in Pakes (1994), either additional work or additional assumptions are necessary to operationalize this approach. For instance, the researcher can recover a nonparametric estimate of the reaction functions and use this in the Euler Equation to arrive at a semi-parametric estimator. However, this approach fails with many products as the nonparametric estimator depends on all state variables, e.g. the entire vector of product characteristics in a given market. The problem can also be ameliorated by specifying the firm's beliefs during estimation. For instance, the researcher could assume some form of bounded rationality where firms do not anticipate a change in their competitors characteristics, meaning that they are held fixed at time t levels. Or, the researcher could assume that firms forecast the response of their competitors based on history of some aggregate market variables up to time t.

[10] This Euler Equation is analogous to Lemma 2 in Pakes (1994) without the continuous adjustment cost and no depreciation.

5 Monte Carlo Simulation

We investigate the properties of our estimator with a simple Monte Carlo that closely resembles our empirical analysis. Readers not interested in these details can skip directly to the BLP application. We consider single-product oligopolists who choose a single, observed characteristic and an unobserved quality level to maximize expected profits. Production is subject to a marginal cost shock, the distribution of which is known to all firms in the industry when they choose their optimal bundle of characteristics. They also knows the profit maximizing equilibrium price for any given demand-cost-characteristic tuple, so they can calculate expected profits for any chosen level of product characteristic. The specifics follow.

5.1 Demand

Demand takes a random coefficients, discrete-choice functional form. Let β_{i0} denote the base-level of utility consumer i derives from purchasing the good. β_i is the taste for the single good characteristic X. Both ε_i and ε_{i0} are distributed i.i.d. extreme value. Consumer i purchases the good if u_i is greater than or equal to $u_{i0} = \varepsilon_{i0}$ where

$$
\begin{aligned}
u_{ij} &= X_j\beta_i - \alpha_i p_j + \xi_j + \varepsilon_{ji} \\
\alpha_i &= \frac{\alpha}{y_i} \\
\ln(y_i) &\sim N\left(\mu_y, \sigma_y\right), \\
\beta_i &\sim N\left(\beta, \sigma_X\right)
\end{aligned}
$$

with β defined as the mean taste for X and σ_X characterizing the heterogeneity in taste, both of which need to be estimates. Following the empirical set-up of BLP, the price elasticity depends on the distribution of income, assumed to follow a known log-normal distribution with mean μ_y and standard deviation σ_y. These parameters are known to the researcher. The coefficient α, which governs the degree of price sensitivity, is unknown and needs to be estimated.

We write the demand parameters together as $\theta^D = (\alpha, \beta, \sigma_X)$. Individual choice probabilities have the standard logit form

$$s_{ij}\left(p, X, \xi; \theta^D\right) = \frac{\exp\left(X_j\beta_i - \alpha_i p_j + \xi_j\right)}{1 + \sum_m \exp\left(X_m\beta_i - \alpha_i p_m + \xi_m\right)}$$

and market shares come from integrating over the distribution of consumers $G(i)$,

$$s_j\left(p, X, \xi; \theta^D\right) = \int_i s_{ij}\left(p_j, X_j, \xi_j; \theta^D\right) dG(i).$$

As is standard in the literature, the utility from the outside good is normalized to 0.

5.2 Supply

Firms play a two-stage game. In the first stage, they optimally choose their level of X and ξ to maximize expected profits, where the expectation is taken over a known distribution of cost shocks. In the second stage Marginal cost is given by the following quadratic in the log characteristic and the unobserved quality

$$\ln\left(\mathrm{mc}_j\right) = \gamma_0 + \gamma_x \ln\left(X_j\right) + \gamma_{x^2} \ln\left(X_j\right)^2 + \gamma_\xi \xi_j + \gamma_{\xi^2}\xi_j^2 + \gamma_{x\xi} \ln\left(X_j\right)\xi + \omega_j$$

with cost shock ω_j.[11] The cost shock ω_j is split into two parts: ω_{1j} which is known in stage 1 and ω_{2j} which is realized in stage 2. All parameters together are denoted $\theta = (\alpha, \beta, \sigma_X, \gamma)$.

Let $Z = (X, \xi, \omega)$. Profits for firm j are given as

$$\Pi\left(Z; \theta\right) \;\; = \;\; \left(p_j - mc\left(X_j, \omega_j;\; \theta\right)\right)\; s_j\left(p, X, \xi;\; \theta\right)\; .$$

The timing is as follows. The vector of cost shocks ω_1 is realized. The oligopolist knows the demand parameters and the distribution of the stage 2 cost shocks $F\left(\omega_2\right)$ but she does not

[11] When characteristics are optimally chosen, a linear index in demand with linear marginal costs will lead to either a corner solution or a continuum of solutions. We use a quadratic form in X and ξ for log-marginal cost to ensure a unique, interior solution.

see the realized shock ω_2 before the characteristic choice is made. She solves for X

$$
\begin{aligned}
\max_{X_j,\xi_j} \quad & E\left[\Pi\left(Z;\theta\right)\Big|\,\omega_1, F(\omega_2)\right] \\
= & \int \left(p_j - mc\left(X_j,\xi_j,\omega_j;\ \theta\right)\right)\ s_j\left(p,X,\xi;\ \theta\right)\ dF\left(\omega_2\right),
\end{aligned}
$$

knowing p_j will be set to maximize profits once the characteristic is set.

Let $\hat{Z} = (X^*, \xi^*, \omega)$ with X_j^* and ξ_j^* are the optimal amounts of X_j and ξ_j chosen before ω_2 is realized. For ease of notation, let $\tilde{X}_j = (X_j, \xi_j)$ and $\Delta\left(\tilde{X}^*, \omega; \theta\right) = p_j\left(\tilde{X}^*, \omega; \theta\right) - mc\left(\tilde{X}_j^*, \omega;\ \theta\right)$. Then X_j^* satisfies

$$
\begin{aligned}
E\left[\frac{\partial \Pi\left(\hat{Z};\theta\right)}{\partial X_j}\Bigg|\,\omega_1, F(\omega_2)\right] &= \int \left(\Delta\left(\tilde{X}^*,\omega;\theta\right)\frac{d\ s_j\left(p_j\left(\tilde{X}^*,\omega;\theta\right),\tilde{X}_j^*;\ \theta\right)}{d\ X_j}\right) dF\left(\omega_2\right) + \\
&\quad \int \left(\frac{\partial \Delta\left(\tilde{X}^*,\omega;\theta\right)}{\partial X_j} s_j\left(p_j\left(\tilde{X}^*,\omega;\theta\right),\tilde{X}_j^*;\ \theta\right)\right) dF\left(\omega_2\right) \\
&= 0
\end{aligned}
$$

where $p_j\left(\tilde{X}^*, \omega; \theta\right)$ is the optimal price given $\tilde{X}^*$ and ω, maximizing the expected profits where expectation is taken over $F\left(\omega_2\right)$. The firms also solve an analogous equation for the unobserved characteristic ξ_j. Letting $Z^* = (X^*, \xi^*, \omega^*)$, where ω^* corresponds to the realized cost shocks, the value of the derivative is given as

$$
\begin{aligned}
\nu_j(Z^*;\theta) &= \frac{\partial \Pi\left(Z^*;\theta\right)}{\partial X} \\
&= \Delta\left(\tilde{X}^*,\omega;\theta\right)\frac{d\ s_j\left(p_j\left(\tilde{X}^*,\omega^*;\theta\right),\tilde{X}^*;\ \theta\right)}{d\ X_j} + \\
&\quad \frac{\partial \Delta\left(\tilde{X}^*,\omega;\theta\right)}{\partial X_j} s_j\left(p_j\left(\tilde{X}^*,\omega^*;\theta\right),\tilde{X}^*;\ \theta\right). \qquad (10)
\end{aligned}
$$

This derivative will sometimes be positive and sometimes be negative depending upon whether "too much" or "too little" of X_j and ξ_j was chosen prior to the realized demand shock. By the way the data are constructed on average these “mistakes” will average out:

$$E\left[\nu_j(Z^*;\theta)\right|\omega_1, F(\omega_2)] = 0,$$

and this moment, along with the marginal cost equation, is our source of identification.

When evaluating $\nu_j(Z^*;\theta)$ in Equation (10), we mimic the standard empirical settings where econometricians cannot observe ω_2 and ξ. At each θ we invert demand to find the realized demand shock $\xi\left(X^*, p^*;\ \theta^D\right)$. Then, we apply the FOC with respect to price in (1) to find $mc\left(p^*, s\left(p^*, X^*, \xi\left(X^*, p^*;\ \theta^D\right)\right);\ \theta^D\right)$. Then, $\nu_j(Z^*;\theta)$ in Equation (10) is evaluated at the realized price, characteristic, recovered demand and cost shocks and marginal cost at each θ.

The optimal instruments are given by

$$H \quad = \quad E\left[\frac{\partial\nu\left(Z^*;\theta\right)}{\partial\theta}, \frac{\partial\omega_2\left(Z^*;\theta\right)}{\partial\theta}\right|\omega_1, F(\omega_2)\Bigg] T(Z^*)$$

where $T(Z^*)$ serves to normalize the error matrix and is equal to the expected inverse second-moment matrix of the residuals. As we are interested in the properties of our estimator in standard empirical settings where the researcher could not compute the optimal instruments we mimic our proposed empirical approach to approximating the optimal instruments by evaluating the derivative at the realized values Z^*. The derivative is approximated at some initial value $\hat{\theta}$, which we take to be a random draw centered at the true parameter value.

Note that the second stage objects are correlated with the random shocks to marginal cost and therefore cannot be used when constructing the optimal instruments. This will in general be true when violations of the first-order conditions are due to optimization error, rather than approximation error. We follow suggestions in Berry et al. (1999) and Gandhi and Houde (2015) and approximate these second stage objects using variation from the first-stage variables. Specifically, we non-parametrically regress prices on a flexible functional form in X_j and X_{-j} to approximate the pricing function. Let P_x be a potentially high-order polynomial in X_j and X_{-j}. We estimate $\hat{p}$ as

$$\hat{p} = \exp(P_x(P_x'P_x)^{-1}P_x' \ln(p))$$

and approximate the market shares as

$$\hat{s}_j = \frac{1}{N}\sum_i \frac{\exp(X_j\hat{\beta}_i + \hat{\xi}_j - \hat{\alpha}_i\hat{p}_j)}{1 + \sum_k \exp(X_k\hat{\beta}_i + \hat{\xi}_k - \hat{\alpha}_i\hat{p}_k)}$$

where $\hat{\xi}_j$ is a consistent estimate of ξ_j using the parameter estimates. The approximate optimal instruments can then be constructed using $\hat{p}$ and $\hat{s}$ in place of their counterparts. The derivative of the residual is then evaluated at the observed level of X and the approximate values of $\hat{\xi}$, $\hat{p}$, and $\hat{s}$.

The moment condition we use is given by

$$\begin{aligned} G_l\left(Z^*;\theta\right) &\equiv E\left[\hat{H}_l \ \cdot [\nu(Z^*;\theta), \omega_2(Z^*;\theta)]\right] \quad \forall l \\ &= 0 \end{aligned}$$

By applying two step GMM with moment condition $G_l\left(\theta\right)$, we estimated the parameters.[12]

[12]We provide a linearized version of these moment conditions in the appendix.

Table 1: Data Generating Process Summary Statistics

	X	ξ	Price	Shares
Mean	1.41	1.82	4.98	0.01
Std Dev.	0.16	0.18	2.12	0.01
Max	2.62	2.42	45.88	0.08
Min	1.10	1.34	1.48	0.00
J	3000			

We simulate $M = 1,500$ markets with two firms for each of $N = 100$ times. Table 1 shows summary statistics for our data generating process, which was calibrated to closely resemble the automobile data used in in Berry et al. (1995), and that we use in our empirical exercise. We chose a value of 80 for α, leading to price elasticities similar to our empirical estimates. The parameters governing income were chosen to generate a dispersion in prices similar to those in the BLP automobile data. Each market has only two firms so the market shares are notably larger, however the share of consumers choosing the outside good is approximately the same as the BLP data. Finally, by choosing $1,500$ markets, we have a similar number of observations as the BLP automobile data.

Table 2: Monte Carlo Simulation - Duopoly, J=3000

Parameter	Truth	IV at non-truth		IV at truth	
		Mean	RMSE	Mean	RMSE
α	80	79.39	1.50	79.82	0.37
β	2	1.84	0.33	1.92	0.11
γ_c	1.25	1.25	0.06	1.25	0.02
γ_X	-0.15	-0.14	0.16	-0.15	0.04
γ_{X^2}	1	0.94	0.10	0.97	0.04
γ_ξ	-0.25	-0.24	0.05	-0.25	0.02
γ_{ξ^2}	0.12	0.11	0.02	0.11	0.01
$\gamma_{X\xi}$	0.10	0.10	0.08	0.10	0.03
σ	0.50	0.50	0.01	0.51	0.01

Table 2 shows the estimated results under two assumptions. First, we assume that we have a consistent estimate of the model parameters, which we generate by taking the true values and adding noise to them. Second, we consider the case where the optimal instruments are constructed at the true parameter values. This allows us to compare the loss in efficiency from using an approximation. In both cases, the parameter estimates are close to the population values. The price coefficient is estimated accurately and precisely, which is encouraging as price elasticities are a common object of interest. The mean parameters for X are estimated with the most error and the largest RMSE, relative to the true population value. When the true parameter values are used to construct the optimal instruments, we see the RMSE decrease by a factor of 3 for most parameters. This shows that the approximation does introduce noise into our estimates and this primarily impacts the mean cost and utility coefficients. However, in most cases the additional noise has little impact on inference, and the estimates are not statistically significantly different from the true parameters values. We turn now to our empirical application

6 Application to BLP Data

6.1 Empirical Framework

We use the exact same data used in BLP. There are twenty new U.S. automobile markets - one for each year from 1971 to 1990 - for a total of 2217 observations on prices, quantities, and characteristics of different vehicle models. We assume the firms set the same $K = 5$ characteristics as those that enter into the BLP utility function, including the ratio of horsepower to weight, interior space (length times width), miles per dollar, whether air conditioning is standard (a proxy for luxury), and the unobserved quality. The five cost shifters (W) are the unobserved quality, the log of ratio of horsepower to weight, the log of interior space, air conditioning, and the log of miles per gallon.[13] In a market with J products there are J observations on the K realized first-order conditions. The outside good quality ξ_0 is normalized to zero and we do not separately estimate the mean utility for new vehicles (i.e. constant term) instead letting it remain in the unobserved quality so in our setup $\beta \in R^{K-1}$. A $\xi_j > 0$ implies that new car on average is preferred to not purchasing a new good. Parameter $\theta = (\beta, \sigma, \alpha, \gamma)$ consists also of $\sigma \in \mathbb{R}^K$, $\alpha \in \mathbb{R}$, and $\gamma \in \mathbb{R}^K$ for a total of 3K=15 parameters to be identified.

Following the base specification in BLP, we assume that utility is given by

$$u_{ij}(\theta) \quad = \quad \alpha \; ln\,(y_i - p_j) + \delta_j + \sum_{k=1}^{K} \sigma_k v_{ik} X_{jk} + \varepsilon_{ij}$$

where $\delta_j = X_j'\beta + \xi_j$. Income draws y_i follow the same log-normal distribution estimated in BLP and v_{ik} are normally distributed. Additionally, we assume that marginal costs are independent of the output level and are comprised of two terms: one term that is log-linear

[13] Air conditioning is an indicator variable which raises the issue of differentiability. We estimate the model both with and without the air conditioning first-order condition as we remain overidentified even when we do not use this condition. At the cost of complicating the estimator by having to combine moment equalities with moment inequalities we could add an inequality related to air conditioning or any other indicator-type characteristic.

in the product characteristics and a second term which is unobserved by the econometrician. Specifically,

$$\ln(mc_j) = \gamma_\xi \xi_j + W_j' \gamma + \omega_j$$

The only difference between our marginal cost specification and that of BLP is that the level of unobserved quality, ξ_j, impacts marginal costs.

BLP assume that all X are exogenous and so any function of them can serve as instruments for any vehicle j. Using the firm pricing first-order conditions Pakes (1994) provides motivation for using the following as instruments for good j (which we call the BLP instruments): own product characteristic X_{jk}, $\forall k$, the sum of characteristic across own-firm products $\sum_{j' \neq j, j' \in J_f} X_{j'}$, and the sum of all characteristics across competing firms, $\sum_{j' \notin J_f} X_{j'}$. These instruments approximate the equilibrium pricing function where markup of a product depends on other products' characteristics. These instruments remain valid in our approach when $\{X_{jk}\} \in I_f$, and so can provide the basis for estimation.

To estimate the model parameters BLP impose that unobserved quality, ξ_j, is orthogonal to observed product characteristics. Formally, let $X = (X_j)_{j \in J}$ denote all of the characteristics observed to consumers, producers, and the researcher. Additionally, they assume that unobserved cost shocks are orthogonal to observed characteristics. BLP then use the following identifying restrictions,

$$E\left[H_{jl}^D \xi_j \left(\theta_0^D\right) \mid X\right] = 0 \quad \forall j, l. \tag{11}$$

$$E\left[H_{jl}^S \omega_j \left(\theta_0^D, \gamma_0\right) \mid W\right] = 0 \quad \forall j, l. \tag{12}$$

where H_{jl}^D and H_{jl}^S are functions of X and W, respectively. These conditions rule out correlation between observed and unobserved product characteristics. Ignoring this correlation can result in demand estimates that too inelastic when price is positively correlated with unobserved product quality (see e.g. Trajtenberg (1989))). Therefore, we replace these moment conditions with our moment conditions

$$E\left[H_{jkl}\nu_{jk}(\theta)\ \mid I_f\right],\ \forall\ k,l \tag{13}$$

where H_{jkl} is the approximation to the optimal instruments. We estimate the model under two different choices for I_f. The "lagged" information set, I^{lagged}, includes only last years observed and unobserved characteristics to construct $\nu_{jk}(\theta)$. In this case when firm f chooses her characteristics she does so using the configuration of competitors' last years products and characteristics to forecast her best guesses at profit maximizing characteristics' choices. In doing so she calculates the Bertrand-Nash prices that would be realized given her choices of observed and unobserved characteristics and the realized characteristics of her competitors products in the previous year. On the other hand, "contemporaneous" information set, I^{cont}, uses contemporaneous characteristics to construct $\nu_{jk}(\theta)$. We approximate $E\left[\nu(\theta)\nu(\theta)'\,|I\right] = I_{JK}$.[14] We transform the instrument $\hat{H}_{jkl}$ into a block diagonal matrix so that we have $K * 15 = 75$ instruments as a benchmark specification.

[14]This simplification does not affect consistency, only the efficiency. Another approximation of $E\left[\nu(\theta)\nu(\theta)'\,|I\right]$ can be done by a block diagonal matrix where a block is a K by K variance-covariance matrix of $\nu_{jk}(\theta)\,|_{k=1,\ldots,K}$ for each firm f and year t.

Table 3: Estimated Parameters of the Demand and Supply

Parameter	Characteristic	BLP	FOC			
			Dynamic	Full IV	Part IV	Contemp.
Term on price α	$ln(y-p)$	43.501	162.919	144.131	155.661	137.032
		(6.427)	(48.834)	(34.054)	(81.672)	(17.737)
Means (β's)	Constant	-7.061				
		(0.941)				
	HP/weight	2.883	0.937	1.168	1.527	0.886
		(2.019)	(0.229)	(0.192)	(0.884)	(0.158)
	Size	3.460	0.133	0.108	0.316	0.618
		(0.610)	(0.034)	(0.036)	(0.422)	(0.062)
	Air	1.521	1.319	1.722	0.975	1.324
		(0.891)	(0.276)	(0.331)	(0.812)	(0.140)
	MP$	-0.122	1.678	2.442	2.401	2.644
		(0.320)	(0.345)	(0.414)	(1.633)	(0.295)
Std. Dev. (σ's)	Constant	3.612	3.330	3.190	3.093	2.276
		(1.485)	(1.886)	(1.093)	(6.304)	(0.471)
	HP/weight	4.628	2.986	3.007	2.818	2.963
		(1.885)	(0.662)	(0.587)	(2.351)	(0.389)
	Size	2.056	0.641	0.934	0.919	0.371
		(0.585)	(0.128)	(0.150)	(0.747)	(0.169)
	Air	1.818	2.009	1.773	1.607	1.286
		(1.695)	(0.394)	(0.257)	(1.247)	(0.150)
	MP$	1.050	0.846	0.859	1.612	0.771
		(0.272)	(0.356)	(0.286)	(0.984)	(0.150)

Continued on next page

Table 3: Estimated Parameters of the Demand and Supply

Parameter	Characteristic	BLP	FOC			
			Dynamic	Full IV	Part IV	Contemp.
Cost params. (γ's)	Constant	0.952				
		(0.194)				
	Mean charac. (ξ)		0.113	0.122	0.121	0.134
			(0.033)	(0.028)	(0.056)	(0.017)
	HP/weight	0.477	0.049	0.059	0.067	0.053
		(0.056)	(0.016)	(0.015)	(0.041)	(0.011)
	Size	-0.046	0.026	0.030	0.054	0.118
		(0.081)	(0.007)	(0.006)	(0.066)	(0.013)
	Air	0.619	0.170	0.226	0.135	0.188
		(0.038)	(0.042)	(0.042)	(0.078)	(0.024)
	MPG	-0.415	0.423	0.551	0.538	0.636
		(0.055)	(0.086)	(0.081)	(0.386)	(0.075)
J		2,217	902	2,125	2,125	2,217
Num. IVs		15	59	43	19	72
Median net util.	$u_{ij} - \alpha \ln(y_i - p_j)$	-0.915	4.856	5.432	7.289	3.520

6.2 Results

Table 3 shows the demand and supply estimates. The first column restates the original BLP results and the second column reports the "dynamic" results, our base specification. For these estimate, we use the optimization conditions given the information set $I_{f,t}^{\text{lagged}}$. That is, we estimate the parameters under the assumption that firms only know last year's characteristics of their competitors' cars when choosing their characteristics. The dataset is restricted to observations of new models or existing car models where at least one characteristic is changed more than 10%. Although this reduces the number of observations to approximately half, the model is estimated on observations where the first-order conditions are most likely to hold. Columns three and four are estimated using the full sample.[15]

Column three uses the full set of instruments (Full IV) of which there are 43, one for each parameter-FOC pair after dropping one instrument due to high correlation. It is well known that while additional instruments always improve standard errors, if many of them are weak bias can be introduced into the estimates.[16] For this reason we also use a subset of these instruments that we think are likely to be the most informative (Partial IV). For each characteristic X_k (except ξ) we use only the derivatives with respect to $(\alpha, \beta_k, \gamma_k, \sigma_k)$. For ξ we use the derivatives with respect to $(\alpha, \gamma_k, \sigma_k)$ giving us a total of 19 instruments. The last column uses the "contemporaneous" information set where firms know the contemporaneous product characteristics of competitors when choosing their characteristic levels.

The four FOC estimators produce similar results, despite the different dataset sizes and assumptions on the information set of firms. The dynamic estimates tend to have higher standard errors due to using half the relevant observations. However, the estimates are similar to the other specifications indicating that these observations convey a significant amount of the relevant variation during estimation. The contemporaneous results are highly similar

[15] We drop the first year observations due to the lagged information set, resulting in the total number of models 2,125.

[16] For example see Bekker (1994), Newey and Smith (2004), or Hansen et al. (2008).

to the lagged results, indicating that the lagged state of the market is a good predictor of the present state as well.

The most striking difference between the BLP estimates in column one and optimization estimates in columns two through five is that the coefficient on price is much larger in the latter cases; consumers are significantly more price sensitive when optimization conditions are used for identification. Table 4 investigates the impact of this difference on estimated elasticities and markups.[17] On average elasticities increase by 31% in absolute value in response to the increase in price sensitivity. This causes estimated markups to fall by on average around 22%.

Table 4: Implied Elasticities and Markups

	Elasticities			Markups ($)		
	BLP	FOC		BLP	FOC	
		Full IV	Part IV		Full IV	Part IV
Lexus LS400	-3.027	-4.836	-5.008	9,214.54	5,754.35	5553.02
Lincoln Towncar	-3.030	-5.708	-5.973	8,310.82	4,633.86	4435.15
Nissan Maxima	-4.124	-7.867	-8.155	3,385.84	1,780.48	1716.92
Ford Taurus	-3.952	-8.205	-8.966	2,679.14	1,363.66	1244.59
Chevy Cavalier	-5.899	-10.284	-11.668	1,327.75	755.42	654.93
Nissan Sentra	-6.304	-10.751	-12.420	909.79	533.02	459.87
Mean	-4.087	-7.796	-8.482	4,051.87	2,393.99	2,280.53
Median	-3.975	-8.219	-8.789	2,751.77	1,397.99	1,324.34
Std. Deviation	1.120	2.130	2.577	3,905.32	2,821.63	2,712.61

"Full IV" uses the full set of instruments while "Part IV" uses a subset of instruments.

One potential explanation for these changes across identification conditions is that observed and unobserved characteristics are positively correlated because firms put more unobserved

[17] Here we use Full IV and Partial IV to ensure that limiting the dataset for the Dynamic model does not skew the results.

quality into cars with high observed quality to the researcher. In this case the instrumented price in the BLP setup will be positively correlated with unobserved quality and this may be leading to an upward bias in the price coefficient. Table 5 explores whether ξ is positively correlated with X by regressing estimated ξ's on all of the BLP demand instruments. Consistent with the price coefficient changes, the BLP instruments explain 50% of the variation in ξ across vehicles and except for miles per dollar – which is negatively correlated with ξ – all other characteristics are positively correlated with ξ. The negative correlation between miles per dollar and ξ might be the reason that the coefficient in the BLP setup of miles per dollar is negative, that is, why people appear not to like fuel efficiency. In reality they like fuel efficiency but it is negatively correlated with other unobserved features of the vehicle that consumers' value.

Table 5: $E\left[\xi_j \mid X\right] \neq 0$

ξ	Full IV		Part IV	
Constant	0.898	-3.102	1.700	-1.686
	(0.594)	(1.122)	(0.673)	(1.241)
HP/weight	6.693	6.123	7.506	5.919
	(0.593)	(0.613)	(0.672)	(0.678)
Size	5.463	4.239	5.607	3.888
	(0.294)	(0.335)	(0.334)	(0.371)
Air	1.397	0.863	2.503	1.693
	(0.135)	(0.136)	(0.154)	(0.150)
MP$	-2.808	-3.952	-3.122	-4.844
	(0.0996)	(0.143)	(0.113)	(0.158)
Other BLP instruments	No	Yes	No	Yes
R-squared	0.621	0.736	0.624	0.751

"Full IV" uses the full set of instruments while "Part IV" uses a subset of instruments.

The last step is to check whether the BLP instrumented price is positively correlated with ξ. We construct the instrumented price by regressing price on the BLP instruments to get a predicted price for each vehicle. Table 6 reports the estimates of the regression of these instrumented prices on an intercept and ξ. The coefficient is significant and positive and the correlation between the instrumented price and ξ is approximately 0.15. Thus the hypothesis that the price coefficient is biased up under the assumption of mean independence because observed and unobserved product characteristics are positively correlated is consistent with all of our findings from the model estimated with the optimization conditions. Similarly, Table 7 reports the results of the regression of the cost shocks on the BLP instruments for the cost function. Observed cost characteristics explain almost half of the movement in the unobserved cost shock, implying that exogeneity between W and ω does not hold. This would naturally further bias the estimated coefficients, although the direction is ambiguous.

Table 6: Correlation Between Instrumented price ($\hat{p}(IV_X)$) and ξ

$\hat{p}(IV_X)$	Full IV	Part IV
Constant	7.734	7.048
	(0.206)	(0.202)
ξ	0.778	0.773
	(0.031)	(0.026)
R-squared	0.220	0.281

$\hat{p}(IV_X)$ is predicted price on BLP demand instruments.

"Full IV" uses the full set of instruments while "Part IV" uses a subset of instruments.

Table 7: $E\left[\omega_j \mid W\right] \neq 0$

ω	Full IV		Part IV	
Constant	0.090	0.993	-0.165	0.655
	(0.136)	(0.164)	(0.148)	(0.184)
ln(HP/weight)	0.326	0.048	0.294	0.031
	(0.035)	(0.033)	(0.038)	(0.036)
ln(Size)	-0.800	-0.418	-0.871	-0.467
	(0.064)	(0.060)	(0.069)	(0.067)
Air	0.341	0.155	0.299	0.134
	(0.017)	(0.0160)	(0.019)	(0.018)
ln(MPG)	0.049	-0.136	0.123	0.025
	(0.043)	(0.0441)	(0.047)	(0.049)
Other BLP instruments	No	Yes	No	Yes
R-squared	0.314	0.578	0.283	0.533

"Full IV" uses the full set of instruments while "Part IV" uses a subset of instruments.

A second related difference is in how the two demand models fit the data. Both models exactly match market shares of products using the BLP inversion. Only 10% of U.S. households buy new cars in any given year so both fitted demand models need a way to explain why 90% of households choose the outside good. The way they do so is quite different and the difference can be found in the final row of Table 3, which reports the average utility of purchasing a new vehicle net of price, $u_{ij} - \alpha \ln(y_i - p_j)$, from each model's fit: BLP predicts it at -4 while our approach predicts it at 7.5. BLP fits 90% of households not buying by having consumers derive strong negative utility from the act of buying a car relative to the outside option of no new car (excluding negative utility from price). In contrast, the optimization-fit has consumers strongly desiring new cars relative to the outside good but the significantly higher price elasticity causes 90% not to buy a new car.

Another difference is that some of the anomalies in the BLP point estimates are not present in the optimization-fit point estimates. The BLP point estimates imply consumers dislike fuel efficiency but in our setup they strongly and significantly like fuel efficiency. They also find costs are *decreasing* as interior space and fuel efficiency increases. We find costs increasing in all of the characteristics, including the unobserved characteristic ξ which enters our cost function but does not enter the BLP cost function.

Table 3 also shows that our estimates are almost always much more precisely estimated relative to the BLP-fit model whether we use the full or partial set of instruments. With the full set of instruments our standard errors are on average a fourth of the standard errors from BLP. The data is exactly the same data so the optimization moments appear to contain more information on both the distribution of consumer preferences and on the cost parameters.

7 Conclusion

Traditional identification since BLP in discrete choice demand model has been to assume no correlation between observed and unobserved characteristics. The major concern of this identification assumption is that it may lead to biased price elasticities if observed and unobserved characteristics are correlated with one other. We avoid this mean independence assumption and infer the distribution of consumer tastes in demand and supply estimation by exploiting optimal choices of product characteristics and prices by firms. We allow firms' information sets at the time they choose characteristics to potentially include competitors' product characteristics, demand, and cost shocks, signals on all of these, or no information at all on them. Following Hansen and Singleton (1982), our identification is based on the assumption that firms are correct in their choices on average even though firms may wish they had made different decisions ex-post.

Using the same automobile data from BLP, we find elasticities double and markups fall by 50%. We also find significantly more precise estimates given the same exact data and some of the slightly puzzling parameter estimates of BLP go away as all of our parameter estimates are of the correct sign.

References

Bekker, Paul A., "Alternative Approximations to the Distributions of Instrumental Variable Estimators," *Econometrica*, 1994, *62* (3), 657–681.

Berry, Steven, James Levinsohn, and Ariel Pakes, "Automobile Prices in Market Equilibrium," *Econometrica*, 1995, *63* (4), 841–890.

_ , _ , and _ , "Voluntary export restraints on automobiles: Evaluating a trade policy," *American Economic Review*, 1999.

_ , _ , and _ , "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market," *Journal of Political Economy*, 2004, *112* (1), 68–105.

Blonigen, Bruce A, Christopher R Knittel, and Anson Soderbery, "Keeping it Fresh: Strategic Product Redesigns and Welfare," 2013.

Chamberlain, Gary, "Asymptotic efficiency in estimation with conditional moment restrictions," *Journal of Econometrics*, 1987, *34* (3), 305–334.

Crawford, Gregory S., "Endogenous product choice: A progress report," *International Journal of Industrial Organization*, 2012, *30* (3), 315–320.

_ and Matthew Shum, "Monopoly Quality Degradation and Regulation in Cable Television," *The Journal of Law and Economics*, 2007, *50* (1), 181–219.

Eizenberg, Alon, "Upstream Innovation and Product Variety in the United States Home PC Market," *The Review of Economic Studies*, 2014, *81*, 1003–1045.

Fan, Ying, "Ownership Consolidation and Product Characteristics: A Study of the US Daily Newspaper Market," *American Economic Review*, 2013, *103* (5), 1598–1628.

Gandhi, Amit and Jean-Francois Houde, "Measuring Substitution Patterns in Differentiated Products Industries,," *University of Wisconsin Working Paper*, 2015.

Gramlich, Jacob, "Gas Prices, Fuel Efficiency, and Endogenous Product Choice in the U.S. Automobile Industry." PhD dissertation, Georgetown University 2009.

Hansen, Christian, Jerry Hausman, and Whitney Newey, "Estimation With Many Instrumental Variables," *Journal of Business & Economic Statistics*, 2008, *26* (4), 398–422.

Hansen, Lars Peter and Kenneth J Singleton, "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica*, 1982, *50* (5), 1269–1286.

Lustig, Joshua, "The Welfare Effects of Adverse Selection in Privatized Medicare." PhD dissertation, UC Berkeley 2008.

Mazzeo, Michael J, "Product choice and oligopoly market structure.," *RAND Journal of Economics*, 2002, *33* (2), 221–242.

Miravete, Eugenio J., Maria J. Moral, and Jeff Thurk, "Innovation, Emissions Policy, and Competitive Advantage in the Diffusion of European Diesel Automobiles," *Working Paper*, 2015.

Nevo, Aviv, "A Practitioner's Guide to Estimation of Random-Coefficients Logit Models of Demand,," *Journal of Economics and Management Strategy*, 2004, *9* (4), 513–548.

Newey, Whitney K and Richard J. Smith, "Higher order properties of GMM and generalized empirical likelihood estimators," *Econometrica*, 2004, *72* (1), 219–255.

Pakes, Ariel, "Dynamic Structural Models: Problems and Prospects. Mixed Continuous Discrete Controls and Market Interactions," in J.J. Laffont and C. Sims, eds., *Advances in Econometrics: Proceedings of the 6th World Congress of the Econometric Society, Vol. 2*, New York: Cambridge University Press, 1994, chapter 5, pp. 171–259.

_ **, Jack Porter, Kate Ho, and Joy Ishii**, "Moment Inequalities and Their Application," *Econometrica*, 2015, *83* (1), 315–334.

Petrin, Amil, "Quantifying the Benefits of New Products: The Case of the Minivan," *Journal of Political Economy*, 2002, (110), 705–729.

Spence, Michael, "Product Selection, Fixed Costs, and Monopolistic Competition," *The Review of Economic Studies*, 1976, *43* (2), 217–235.

Sweeting, Andrew, "Dynamic Product Repositioning in Differentiated Product Markets: The Case of Format Switching in the Commercial Radio Industry," 2007.

Trajtenberg, Manuel, "The Welfare Analysis of Product Innovations, with an Application to Computed Tomography Scanners," *Journal of Political Economy*, 1989, *94*, 444–479.

Veiga, André and E. Glen Weyl, "Product Design in Selection Markets," 2014.

Tables

8 Appendix

8.1 Proof of equation (7)

To show that (7) holds, we start by noting that the J residuals for any characteristic k can be written in matrix notation as

$$\nu_k = \left(\frac{\partial(p - \mathrm{mc})}{\partial X_k'} \circ T\right) s + \left(\frac{ds}{dX_k'} \circ T\right)(p - \mathrm{mc}), \;\; k = 1, ..., K, \tag{14}$$

where $\circ$ denotes element-wise multiplication and T, the ownership matrix, is a block diagonal matrix that identifies products owned by the same firm, so if $i \in J_f$ and $j \in J_f$ for some f, then $t_{ij} = 1$; otherwise, $t_{ij} = 0$. Conditional on a given value of θ_{nl}, several of the terms in ν_k are held constant, including $s(\delta(\theta_{nl}))$ and $(p - \mathrm{mc})$. $\frac{\partial s_i}{\partial p_j'} = \frac{\alpha}{y_i} s_{ij}(1 - s_{ij})$, and with s_{ij} also not a function θ_l, conditional on θ_{nl}, implying the matrix $\frac{\partial s}{\partial p'}$ is also constant. Since $\frac{ds}{dX_k} = \frac{\partial s}{\partial X_k'} + \frac{\partial s}{\partial p'}\frac{\partial p}{\partial X_k'}$, (14) is comprised of three terms that vary with θ_l: $\frac{\partial \mathrm{mc}}{\partial X_k'}$, $\frac{\partial s}{\partial X_k'}$ and $\frac{\partial p}{\partial X_{jk}'}$. None of these terms multiply one another, so if each is established to be affine in θ_l then the entire expression will be as well. Given our specification of utility and marginal cost, it is clear that the first two terms are affine in θ_l, so ν_k will have the required form if $\frac{\partial p}{\partial X_{jk}'}$ is affine as well, which Proposition 2 proves below.

We begin with Proposition 1, which establishes that equation (14) may be written as (7) if the matrix $\frac{\partial p}{\partial X_{jk}'}$ can be written as an affine function of θ_l.

Proposition 1. *Assume that mean utility is linear in β and the derivative of marginal costs with respect to X_{jk} is affine in γ_k, conditional on observables. If $\frac{\partial p}{\partial X_{jk}'}$ can be written as an affine function of θ_l, conditional on observed market shares and the nonlinear parameters, then ν_k can be written as an additive function of two non-linear functions of (θ_{nl}, X, s, p), one of which is linear in θ_l.*

Proof We start by observing that linearity of β implies that the derivative of market shares can be written as an affine function of β, conditional on θ_{nl} and data. The derivative of market shares with respect to the product characteristics are given by

$$\begin{aligned}\frac{ds_j(\delta, X, p, \theta_{nl})}{dX'_k} &= \frac{\partial s_j(\delta, X, p, \theta_{nl})}{\partial X'_k} + \frac{\partial s_j}{\partial \delta'}\frac{\partial \delta}{\partial X'_k} \\ &= \frac{\partial s_j(\delta, X, p, \theta_{nl})}{\partial X'_k} + \frac{\partial s_j}{\partial \delta'}(I \circ \beta_k) \\ &= g_1(\delta, X, p, \theta_{nl}) + g_2(\delta, X, p, \theta_{nl})\beta_k\end{aligned}$$

for functions g_1 and g_2 that do not depend on θ_l, conditional on δ. Additionally,

$$\frac{ds_j(\delta, X, p, \theta_{nl})}{dp'} = \frac{\partial s_j(\delta, X, p, \theta_{nl})}{\partial p'} = g_p(\delta, X, p, \theta_{nl})$$

We use the total derivative to emphasize that X_k impacts shares both directly and through its impact on δ. Similar expressions are straightforward to derive for $\frac{\partial^2 s_j}{\partial p \partial p'}$ and $\frac{\partial^2 s_j}{\partial p \partial X'_k}$.
We focus on each term ν_k in (14) in turn. Market shares, $s_{j'}$, are set equal to their observed values during estimation and so can be treated as data. Markups are inferred from the data and θ_{nl} by $(p - \text{mc}) = \Delta^{-1}s$, where

$$\Delta_{ij} = \begin{cases} -\frac{\partial s_j}{\partial p'_j}, & \text{if } j, j' \in J_f \\ 0 & \text{otherwise} \end{cases}$$

Because the derivative of $s_{j'}$ with respect to price does not directly depend on θ_l, Δ and markups are a function of θ_{nl} and observed data alone. By assumption, the derivative of marginal cost with respective to X_k is affine in θ_l. We have assumed that $\frac{\partial p}{\partial X'_{jk}}$ is linear in θ_l, so all that remains is are the terms $\frac{ds_{j'}}{dX_{jk}}$. Again, we can write these out as

$$\frac{ds_j}{dX_{jk}} = \frac{\partial s_j}{\partial X_{jk}} + \sum_{j' \in J} \frac{\partial s_j}{\partial p_{j'}} \frac{\partial p_{j'}}{\partial X_{jk}}$$

so that the total derivative is given as

$$\frac{ds_j}{dX_{jk}} = g_1(\delta, X, p, \theta_{nl}) + g_2(\delta, X, p, \theta_{nl})\beta_k + g_p(\delta, X, p, \theta_{nl})' \frac{\partial p}{\partial X'_{jk}}$$

By assumption, $\frac{\partial p}{\partial X'_{jk}} = h_1(\delta, X, p, \theta_{nl}) + h_2(\delta, X, p, \theta_{nl})\theta_l$ for some matrices h_1 and h_2. The parameter β_k is an element of θ_l, and so this reduces to

$$\frac{ds_j}{dX_{jk}} = \tilde{d}_1(\delta, X, p, \theta_{nl}) + \tilde{d}_2(\delta, X, p, \theta_{nl})\theta_l$$

for some matrices $\tilde{d}_1$ and $\tilde{d}_2$. This establishes that all of the terms in ν_{jk} are nonlinear functions of θ_{nl} and either do not depend on θ_l or are linear in θ_l. Plugging these terms back into the residual equation yields

$$\nu_{jk} = \sum_{j' \in J_f} \left(\left(h_1(\delta, X, p, \theta_{nl}) + h_2(\delta, X, p, \theta_{nl})\theta_l - \frac{\partial \mathrm{mc}_j}{\partial X_{jk}} \right) s_{j'} + \left(\tilde{d}_1 + \tilde{d}_2\theta_l \right) (p_{j'} - \mathrm{mc}_{j'}) \right) \tag{15}$$

Pulling out the components of θ_l, we end up with an expression of the form

$$\nu_k = \nu_{kc}(\delta, X, p, \theta_{nl}) + \nu_{k\theta}(\delta, X, p, \theta_{nl})\theta_l$$

which completes the proof. Turning to Proposition 2, we now establish that $\frac{\partial p}{\partial X'_{jk}}$ is affine in θ_l. By the Implicit Function Theorem, the derivative of prices with respect to characteristics has the following form

$$\frac{\partial p}{\partial X'_{jk}} = -\left(\frac{\partial R}{\partial p'}\right)^{-1} \frac{\partial R}{\partial X_{jk}} \tag{16}$$

where R is a vector of residual equations defined by the pricing FOCs. Proposition 2 shows that $\frac{\partial R}{\partial p}$ does not depend on θ_l and that $\frac{\partial R}{\partial X_{jk}}$ is affine in θ_l, thereby establishing that $\frac{\partial p}{\partial X'_{jk}}$ is affine in θ_l.

Proposition 2. *Assume that mean utility is linear in β and the derivative of marginal costs with respect to X_{jk} is affine in γ_k, conditional on observables. Then the derivative of price with respect to each product characteristic can be written as the sum of two nonlinear functions of θ_{nl}, one of which does not depend on θ_l and the other that is a linear function of θ_l.*

Proof Let R be the residual equation for the first-order condition with respect to price. That is,

$$R = s + \left(T \circ \frac{\partial s}{\partial p'}\right)(p - \mathrm{mc})$$

so that the derivative of price with respect to characteristic X_{jk} is given by (16). We show $\frac{\partial R}{\partial p'}$ does not depend on θ_l and that $\frac{\partial R}{\partial X'_{jk}}$ is affine in θ_l. The derivative of the residual equation with respect to price

$$\frac{\partial R}{\partial p'} = \frac{\partial s}{\partial p'} + \sum_j \left(e_j \left(T_j \circ \frac{\partial s_j}{\partial p'}\right) + (p_j - \mathrm{mc}_j)\left(T_j \circ \frac{\partial^2 s_j}{\partial p \partial p'}\right)\right) \tag{17}$$

where e_j is the jth column of the identity matrix. We previously noted that $\frac{\partial s}{\partial p'}$ does not depend on θ_l by equation (15) and that $(p_j - \mathrm{mc}_j)$ does not depend on θ_l as $(p - \mathrm{mc}) = \Delta^{-1}s$, where Δ is defined as before. The second derivative $\frac{\partial^2 s_j}{\partial p \partial p'}$ does not depend directly on θ_l, so $\frac{\partial R}{\partial p'}$ is a nonlinear function of θ_{nl} and the data, and independent of θ_l. Taken together, this means that θ_l does not appear in $\frac{\partial R}{\partial p'}$.

Turning to $\frac{\partial R}{\partial X'_k}$, we can write this derivative as

$$\frac{\partial R}{\partial X'_k} = \frac{\partial s}{\partial X'_k} + \sum_j \left(-\frac{\partial \mathrm{mc}_j}{\partial X'_k}\left(T_j \circ \frac{\partial s_j}{\partial p'}\right) + (p_j - \mathrm{mc}_j)\left(T_j \circ \frac{\partial^2 s_j}{\partial p X'_k}\right)\right)$$

By assumption, $\frac{\partial \mathrm{mc}_j}{\partial X'_k}$ is affine in θ_l. We have previously shown that the terms $\frac{\partial s}{\partial X'_k}$ and $\frac{\partial^2 s_j}{\partial p X'_k}$ are affine in θ_l and that the values of $\frac{\partial s_j}{\partial p'}$ and $(p_j - \mathrm{mc}_j)$ do not depend on θ_l. All terms that are affine in θ_l multiply terms that are fixed in θ_l. As such, the resulting expression will be affine in θ_l and that establishes the result.

9 Application to Mixed Logit Models

Assume the random coefficients are normally distributed and let s_i denote the vector of choice probabilities. Then the derivative of market shares with respect to characteristic k is given by

$$\begin{aligned}\frac{ds(\delta, X, p, \theta_{nl})}{dX_k'} &= \frac{1}{ns}\sum_{i=1}^{ns}\sigma_k\eta_{ik}\left(s_i \circ I - s_i s_i'\right) + \frac{1}{ns}\sum_{i=1}^{ns}\left(s_i \circ I - s_i s_i'\right)\beta_k \\ &= g_1(\delta, X, p, \theta_{nl}) + g_2(\delta, X, p, \theta_{nl})\beta_k\end{aligned}$$

with an analogous expression for $\frac{\partial s}{\partial p'}$.

We now provide the expressions for the residuals when demand has a mixed logit specification. Let P denote the matrix of choice probabilities, which is a $J \times n_s$ matrix. Next, define the following $n_s \times 1$ weight vectors $w_\alpha = \left\{-\frac{\alpha}{y_i}\frac{1}{n_s}\right\}_{i=1}^{n_s}$, $w_{\alpha,2} = \left\{\left(\frac{\alpha}{y_i}\right)^2\frac{1}{n_s}\right\}_{i=1}^{n_s}$, $w_\sigma = \left\{\sigma_k\eta_{ik}\frac{1}{n_s}\right\}_{i=1}^{n_s}$, and $w_{\alpha,\sigma} = \left\{-\left(\frac{\alpha}{y_i}\right)\sigma_k\eta_{ik}\frac{1}{n_s}\right\}_{i=1}^{n_s}$. We denote the $J \times n_s$ weighted matrix of choice probabilities by $P_{w_j} = P \circ w_j'$, for each w_j above. Finally, define the $J \times n_s$ matrix $\zeta = P \circ \left[T\left(P \circ (p - \mathrm{mc})\right)\right]$. Recall that the residuals are given by

$$\nu_k = \left(\frac{\partial(p - \mathrm{mc})}{\partial X_k'} \circ T\right)s + \left(\frac{ds}{dX_k'} \circ T\right)(p - \mathrm{mc}),\ \ k = 1, ..., K,$$

The first term is given by

$$\left(\frac{\partial(p - \mathrm{mc})}{\partial X_k'} \circ T\right)s = -\underbrace{\left(T \circ G^{-1}H_{k,c}\right)s}_{\nu_{k,c,1}} - \underbrace{\left(T \circ G^{-1}H_{k,\beta}\right)s}_{\nu_{k,\beta,1}}\beta_k - \underbrace{\left(T \circ \left(\frac{\mathrm{mc}}{X_k} \circ I\right)\right)}_{\nu_{k,\gamma,1}} s\,\gamma_k$$

and the second term is given by

$$\left(\frac{ds}{dX_k'}\circ T\right)(p-\mathrm{mc}) = \underbrace{\left(T\circ\left(Pw_\sigma\circ I - PP'_{w_\sigma} - \frac{\partial s}{\partial p'}G^{-1}H_{k,c}\right)\right)(p-\mathrm{mc})}_{\nu_{k,c,2}} + \underbrace{\left(T\circ\left(s\circ I - \frac{1}{n_s}PP' - \frac{\partial s}{\partial p'}G^{-1}H_{k,\beta}\right)\right)(p-\mathrm{mc})}_{\nu_{k,\beta,2}}\beta_k - \underbrace{\left(T\circ\frac{\partial s}{\partial p'}G^{-1}H_{k,\gamma}\right)(p-\mathrm{mc})}_{\nu_{k,\gamma,2}}\gamma_k$$

where $G = \frac{\partial R'}{\partial p}$ and $H_k = \frac{\partial R}{\partial X_k'} = H_{k,c} + H_{k,\beta}\beta_k + H_{k,\gamma}\gamma_k$, with their terms are defined below. We then the expressions for $\nu_{k,\theta_l} = (\nu_{k,\beta}, \nu_{k,\gamma})$ and $\nu_{k,c}$ are given by $\nu_{k,\beta} = \nu_{k,\beta,1} + \nu_{k,\beta,2}$, $\nu_{k,\gamma} = \nu_{k,\gamma,1} + \nu_{k,\gamma,2}$, and $\nu_{k,c} = \nu_{k,c,1} + \nu_{k,c,2}$. Finally, after some tedious algebra, the expressions for G and H_k can be shown to be given by

$$\begin{aligned} G = &\left((Pw_\alpha\circ I) - PP'_{w_\alpha}\right) + \left((Pw_{\alpha,2}\circ I) - PP'_{w_{\alpha,2}}\right)\circ(p-\mathrm{mc}) + (\zeta w_{\alpha,2})\circ I + \\ &\left(PP'_{w_{\alpha,2}}\right)\circ T\circ(p-\mathrm{mc})' - 2\zeta P'_{w_{\alpha,2}} + \frac{\partial s}{\partial p'}\circ T, \end{aligned}$$

and

$$\begin{aligned} H_{k\mathrm{c}} = &\left((Pw_\sigma\circ I) - PP'_{w_\sigma}\right) + \left((Pw_{\alpha,\sigma}\circ I) - PP'_{w_{\alpha,\sigma}}\right)\circ(p-\mathrm{mc}) + (\zeta w_{\alpha,\sigma})\circ I + \\ &\left(PP'_{w_{\alpha,\sigma}}\right)\circ T\circ(p-\mathrm{mc})' - 2\zeta P'_{w_{\alpha,\sigma}} \\ H_{k\beta} = &\left((s\circ I) - \frac{1}{n_s}PP'\right) + \left((Pw_\alpha\circ I) - PP'_{w_\alpha}\right)\circ(p-\mathrm{mc}) + \zeta w_\alpha\circ I + \\ &\left(PP'_{w_\alpha}\right)\circ T\circ(p-\mathrm{mc})' - 2\zeta P'_{w_\alpha} \\ H_{k\gamma} = &-\frac{\partial s}{\partial p'}\circ T\circ\frac{\mathrm{mc}}{X_k}, \end{aligned}$$