Identification and Estimation of Discrete Choice Demand Models when Observed and Unobserved Characteristics are Correlated

Amil Petrin* Boyoung Seo†

January 23, 2019

Abstract

The standard Berry, Levinsohn, and Pakes (1995) (BLP) approach to estimation of demand and supply parameters assumes that the product characteristic unobserved to the researcher but observed by consumers and producers is conditionally mean independent of all characteristics observed by the researcher. We extend BLP to allow all product characteristics to be endogenous, so the unobserved characteristic can be correlated with the other observed characteristics. We derive moment conditions based on the assumption that firms - when choosing product characteristics - are maximizing expected profits given their beliefs at that time about preferences, costs, and competitors' actions with respect to the product characteristics they choose. Following Hansen and Singleton (1982) we assume that the"mistake" in the choice of the amount of the characteristic that is revealed once all products are on the market is conditionally mean independent of anything the firm knows when it chooses its product characteristics. We develop an approximation to the optimal instruments and we also show how to use the standard BLP instruments. Using the original BLP automobile data we find all parameters to be of the correct sign and to be much more precisely estimated. Our estimates imply observed and unobserved product characteristics are highly positively correlated, biasing demand elasticities upward significantly, as our average estimated price elasticities double in absolute value and average markups fall by 50%.

*University of Minnesota and NBER, petrin@umn.edu

†Kelley School of Business, Indiana University, 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 then 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.

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

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

of the product characteristics and other demand and supply factors have been realized they will compete in prices in a Bertrand–Nash manner. 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 (Hansen and Singleton (1982)). 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 (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. 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 this characteristics in marginal

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

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 not buying 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 like 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 and describe strength and weakness of the identification strategy that previous studies used in the discrete choice demand estimation since BLP. In Section (3), we present our approach to identify heterogeneity in tastes and supply parameters by using market-level data. Estimation and choice of instruments are suggested in Section (3.2). Section (4) shows Monte Carlo simulation results. Lastly, we apply our approach to the same automobile data BLP used in (5).

2 Demand and Supply

Our approach to modeling demand and supply, our data, and the steps in our estimation procedure mimic the approach in Berry, Levinsohn, and Pakes (1995) (BLP) up to the identification assumptions. In this section we review their approach. 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 $(y_i,\, v_i,\, \varepsilon_i)$, where y_i is a draw from the distribution of U.S. household incomes for the appropriate year, v_i is vector of their K idiosyncratic normally distributed taste draws $(v_{ik})_{k=1}^{K}$ drawn from a standard normal 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 $\theta^D = (\alpha, \beta, \sigma) \in \mathbb{R}^{1+K+K}$, where α is the marginal utility of income parameter, β is the vector of mean tastes for observed characteristics and includes $(\beta_k)_{k=1}^{K}$, and σ is the vector of K parameters $(\sigma_k)_{k=1}^{K}$ that measure of the extent of heterogeneity in tastes among consumers for each characteristic k. Utility that consumer i derives from good j is given as

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

where the product specific utility component δ_j is common to all consumers and defined as

$$\delta_j = X_j'\beta + \xi_j.$$

Consumer i's taste for characteristic X_k is then given by $\beta_{ik} = \beta_k + \sigma_k v_{ik}$.

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

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

s_{ij} is individual i's probability of purchasing good j prior to the realization of ε and is given by

$$s_{ij}\left(p, X, \xi;\ \theta^D\right) = \frac{exp\left(\alpha\ ln\left(y_i - p_j\right) + X_j'\beta + \xi_j + \sum_k \sigma_k v_{ik} X_{jk}\right)}{\sum_{j' \in J} exp\left(\alpha\ ln\left(y_i - p_{j'}\right) + X_{j'}'\beta + \xi_{j'} + \sum_k \sigma_k v_{ik} X_{j'k}\right)}.$$

Letting $F(y, v, \varepsilon)$ denote the distribution of consumer characteristics the market share for good j is given as the integral over these consumers:

$$s_j\left(p, X, \xi;\ \theta^D\right) = \int_{\{u_{ij}(\theta^D) \geq u_{ij'}(\theta^D),\ \forall j'\}} s_{ij}\left(p, X, \xi;\ \theta^D\right)\ dF(y, v, \varepsilon).$$

BLP develop a method to allow for the possibility that prices are correlated with unobserved characteristics. Ignoring the correlation can result in demand estimates that too inelastic because price is positively correlated with unobserved product quality (see e.g. Trajtenberg (1989))). Market shares are nonlinear functions of prices and observed and unobserved characteristics so standard instrumental variable techniques are not consistent. BLP prove – if the goods are weak substitutes – that at any given set of parameter values (α, σ), there exists a unique vector $\delta(\alpha, \sigma) = \left(\delta_j(\alpha, \sigma)\right)_{j \in J}$ which exactly matches observed shares in data to predicted shares from the model:

$$s_j(\delta(\alpha, \sigma)) = s_j^{\text{data}}, \quad \forall j. \tag{1}$$

Given any set of parameter values (α, σ) their approach solves for this $\delta(\alpha, \sigma)$ and holds it constant during estimation to control for potential correlation between the unobserved characteristics and price. We use this inversion to recover for any set of parameter values (α, σ, β) the unique vector of unobserved characteristics so we can treat it as another observed characteristic the firms choose.[3]

Let $X = (X_j)_{j \in J}$ denote all of the characteristics observed to consumers, producers, and the researcher. For identification BLP allow price to be endogenous but assume that these

[3]Once one has $\delta_j(\alpha, \sigma)$ and β one knows $\xi_j(\theta^D) \equiv \xi_j\left(\delta(\alpha, \sigma), \theta^D\right),\ \forall j.$

X are mean independent of ξ_j $\forall j$:

$$E\left[\ \xi_j\left(\theta_0^D\right)\ \mid X\ \right] = 0 \quad \forall j. \tag{2}$$

Since all X are assumed to be exogenous any function of them can serve as instruments for any vehicle j but many of them may be weak. Using the firm pricing first-order conditions Pakes (1994) provides motivation for using the following as instruments for good j: 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, $\sum_{j' \notin J_f} X_{j'}$.

On the supply side marginal costs are given as

$$ln\left(mc_j\right) = W_j'\gamma + \omega_j$$

where W_j are cost shifters, typically X_j itself or the log of it, and ω_j is the cost shock for good j. As marginal costs are not directly observed BLP assume Bertrand Nash price competition and invert them out using the J pricing first-order conditions:

$$s_j + \sum_{j' \in J_f} \left(p_{j'} - mc_{j'}\right) \frac{\partial s_{j'}}{\partial p_j} \quad = \quad 0, \quad \forall j \in J_f. \tag{3}$$

where firms are indexed by f and J_f is the set of goods that firm f produces. Once marginal costs are recovered BLP identify the cost parameters γ by assuming that the cost shocks are uncorrelated with cost shifters, an assumption analogous to the demand identification condition (2):

$$E\left[\ \omega_j\left(\theta_0^D, \gamma_0\right)\ \mid W\ \right] = 0 \quad \forall j. \tag{4}$$

BLP supply instruments are constructed in the same way as described above by replacing X with W.

The estimation approach used by BLP is optimal two-step Generalized Method of Moments (GMM). Let the model parameter $\theta = \left(\theta^D, \gamma\right)$. Let $G\left(\theta\right)$ be a stacked vector of the mean independence demand and supply moments in (2) and (4), and let $\hat{\Omega}$ denote the inverse of the first-step estimate of variance-covariance matrix of the moments. The GMM estimator is the solution to

$$\min_{\theta} Q(\theta) \quad = \quad G(\theta)' \; \hat{\Omega} \; G(\theta) \, .$$

For any candidate value (θ_s) the steps to calculate $Q(\theta_s)$ include

1. Invert out $\delta(\alpha_s, \sigma_s)$ by matching the model predicted shares to the data and calculate $\xi_j = \delta_j(\alpha_s, \sigma_s) - X_j' \beta_s \; \forall j$.

2. Using equations (3) invert out $mc_j \; \forall j$ and calculate $\omega_j = ln\,(mc_j) - W_j' \gamma_s \; \forall j$.

3. Calculate $G(\theta_s)$ and implied $Q(\theta_s)$.

The main difference between our approach and BLP estimation approach is in Step 3 where we use moment conditions implied by optimization instead of assuming the characteristic not observed by the researcher is mean independent of those that are observed by the researcher.

3 Identification Using Optimization

There is a large and growing literature that uses optimization to identify demand and cost parameters.[4] We follow in this tradition and extend the optimization in prices used in BLP to include all observed and unobserved product characteristics.

3.1 Competitive Setting

As in many papers in this literature we assume firms compete with each other every period in two stages. In the second stage firms know that they will compete in prices in a Bertrand-Nash manner given the chosen characteristics of all products in the market. In the first stage 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. An advantage of using Hansen and Singleton (1982) is that it does not require the researcher to know or specify the entire information set of each firm.

Define $Z_j = (X_j, W_j, \omega_j)$, and $Z = (Z_j)_{j \in J}$. Redefine K be the number of characteristics including ξ and let $\theta = (\alpha, \beta, \sigma, \gamma)$ include the cost parameters. In the first step, firm f

[4] See the introduction.

chooses vectors $X_f = (X_j)_{j \in J_f}$ and $\xi_f = (\xi_j)_{j \in J_f}$ to solve:

$$
\begin{aligned}
\max_{X_f,\ \xi_f} \quad & E\left[\Pi_f \mid I_f\right] \\
\Pi_f \quad = \quad & \sum_{j' \in J_f} \left(p_{j'} - mc_{j'}\left(X_{j'}, \xi_{j'}; \theta\right)\right)\ s_{j'}\left(p, Z, \xi; \theta\right).
\end{aligned}
$$

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

$$
\begin{aligned}
\nu_{jk}(\theta) \quad \equiv \quad & \frac{\partial \Pi_f}{\partial X_{jk}} \\
= \quad & \sum_{j' \in J_f} \left[\left(p_{j'} - mc_{j'}\left(X_{j'}, \xi_{j'}; \theta\right)\right) \frac{d\ s_{j'}\left(p, Z, \xi; \theta\right)}{d\ X_{jk}} \right. \qquad (5) \\
& \left. + s_{j'}\left(p, Z, \xi; \theta\right)\ \frac{\partial\left(p_{j'} - mc_{j'}\left(X_{j'}, \xi_{j'}; \theta\right)\right)}{\partial X_{jk}} \right] \qquad (6)
\end{aligned}
$$

for $k < K$. If $k = K$, the above first-order condition is taken with respect to ξ_j.

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 (5) represents the change in profits attributable to *marginal* consumers while the second term in (6) captures those attributable to *infra marginal* consumers. 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):

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

We follow Fan (2013) and estimate $\frac{\partial p_{j''}}{\partial X_{jk}}$ using the Implicit Function Theorem.

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. Sometimes firms will provide too little of a characteristic and sometimes it will provide too much, but on average these "mistakes" average out. We write it as:

$$E\left[\nu_{jk}\left(\theta_0\right) \mid I_f\right]=0 \quad \forall k, j \in J_f,\ \forall f \tag{7}$$

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. 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. Our identification is based on these K first-order conditions with respect to X coupled with the J first-order conditions with respect to p in (3).

Some product characteristics may not be update every period, implying that firms maximize the sum of the future stream of profits at the time of decision. We discuss the robustness of our approach in Section 6.

3.2 Instruments

We have more than K unknown parameters but only K first-order condition conditions. We use the insight by Hansen and Singleton (1982) that any function of the arguments of I_f are possible 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 $JK\times$ the number of parameters 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

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

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

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

for $k < K$. If $k = K$, $d\ X$ or ∂X is substituted to $d\ \xi$ or $\partial \xi$. 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.[5] 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. In the benchmark setup we assume firms' information sets contain their own contemporaneous costs shocks and their 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 we use observed and

[5]Leaving out these terms is not a consistency issue but instead an efficiency issue. In our monte carlos it is possible to calculate these terms we compare the monte carlo results with and without them to check on the importance of these terms for precision.

unobserved characteristics of products of the firms competing against f from the previous year. At those characteristics we solve for the Bertrand-Nash equilibrium prices and evaluate the derivative at those prices for firm f.

In the second case we assume

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

so firm f knows its competitors' contemporaneous choices of characteristics and costs at the time of decision. In this case, the derivative is evaluated at realized values of X_t, ξ_t, W_t, p_t, and ω_t.[6]

Estimation

Our estimation approach mimics the steps in the discussion of the BLP estimation approach up to two differences. We replace the moments in Step 3 with our optimization moments. Using the instruments suggested above, the moment condition is given by

$$\begin{aligned} G_{kl}(\theta) &\equiv E\left[\hat{H}_{jkl}\ \nu_{jk}(\theta)\right] \quad \forall k, l \\ &= 0 \end{aligned} \tag{10}$$

where $\hat{H}_{jkl}$ is our approximation to the optimal instrument H_{jkl}. We also include ξ_j in the set of marginal cost shifters and we estimate a parameter associated with it.[7] For an initial guess at θ_0 – we use the BLP estimates – we calculate the approximation to the optimal instruments. Given those instruments we calculate the optimal weighting matrix and the first stage estimates. At the first stage estimates we recalculate the optimal instruments and the efficient weighting matrix and the reestimate to get the two-step GMM estimates.[8]

[6] This case implies that the conditional expectation of the FOCs are taken with respect to approximation or measurement error.

[7] 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.

[8] Using the GMM approach allows one to augment the setup with additional moment conditions in a straightforward manner as in Petrin (2002) or Goolsbee and Petrin (2004).

4 Monte Carlo Simulation

We investigate the properties of our estimator with two simple and transparent monte carlos. Readers not interested in these details can skip directly to the BLP application although

4.1 Monopoly

We first consider a single-product monopolist who chooses a single product characteristic to maximize expected profits. At the time she chooses the characteristic she knows her cost shock but she does not know the demand shock, although she knows the distribution from which it will be drawn. She also knows the profit maximizing price for any given demand-cost-characteristic tuple, so she can calculate expected profits for any chosen level of product characteristic. The specifics follow.

We consider both logit and random coefficients discrete choice demand setups. Let β_{i0} denote the base-level of utility consumer i derives from purchased 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

$$\begin{aligned} u_i &= \beta_{i0} + X\beta_i - \alpha p + \xi + \varepsilon_i \\ \beta_{i0} &\sim N(0, \sigma_1) \\ \beta_i &\sim N(\beta, \sigma_X), \end{aligned}$$

with β defined as the mean taste for X and (σ_1, σ_X) characterizing the heterogeneity in taste. In the logit case $(\sigma_1, \sigma_X) = (0, 0)$ and in the random coefficient case they are non-negative.

We write the demand parameters together as $\theta^D = (\alpha, \beta, \sigma_1, \sigma_X)$. Market share in the logit case is given by

$$s\left(p, X, \xi; \theta^D\right) = \frac{exp\left(X\beta - \alpha p + \xi\right)}{1 + exp\left(X\beta - \alpha p + \xi\right)}$$

and in the random coefficients' case, which integrates over the distribution of consumers $G(i)$, share is given by

$$s\left(p, X, \xi; \theta^D\right) = \int_i \frac{exp\left(\beta_{i0} + X\beta_i - \alpha p + \xi\right)}{1 + exp\left(\beta_{i0} + X\beta_i - \alpha p + \xi\right)} dG(i).$$

The monopolist's marginal cost is given as

$$ln\left(mc\right)=\gamma lnX+\omega,$$

with γ the elasticity of marginal costs with respect to the characteristic X and cost shock ω. All parameters together are denoted $\theta=(\alpha,\beta,\sigma_1,\sigma_X,\gamma)$.

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

$$\Pi\left(Z;\theta\right) \;=\; \left(p-mc\left(X,\omega;\;\theta\right)\right)\;s\left(p,X,\xi;\;\theta\right)\;.$$

The timing is as follows. The cost shock ω^* is realized. Let $\tilde{Z}=(X,\xi,\omega^*)$. The monopolist knows the demand parameters and the distribution of the demand shock $F\left(\xi\right)$ but she does not see the realized demand shock ξ^* before the characteristic choice is made. She solves

$$\begin{aligned}\max_X \quad & E\left[\Pi\left(\tilde{Z};\theta\right)\middle|\,\omega^*,F(\xi)\right]\\ = & \int\left(p-mc\left(X,\omega^*;\;\theta\right)\right)\;s\left(p,X,\xi;\;\theta\right)\;dF\left(\xi\right),\end{aligned}$$

knowing p will be set to maximize profits once the characteristic is set and ξ^* is realized.

Let $\hat{Z}=(X^*,\xi,\omega^*)$ with X^* the optimal amount of X. X^* solves

$$\begin{aligned}E\left[\frac{\partial\Pi\left(\hat{Z};\theta\right)}{\partial X}\middle|\,\omega^*,F(\xi)\right] \;=\; & \int\left(\left(p\left(X^*,\omega^*;\theta\right)-mc\left(X^*,\omega^*;\;\theta\right)\right)\frac{d\;s\left(p\left(X^*,\omega^*;\theta\right),X^*,\xi;\;\theta\right)}{d\;X}\right)dF\left(\xi\right)\\ & +\int\left(\frac{\partial\left(p\left(X^*,\omega^*;\theta\right)-mc\left(X^*,\omega^*;\;\theta\right)\right)}{\partial X}s\left(p\left(X^*,\omega^*;\theta\right),X^*,\xi;\;\theta\right)\right)dF\left(\xi\right)\\ = \;& 0\end{aligned}$$

where $p\left(X^*,\omega^*;\theta\right)$ is the optimal price given X^* and ω^*, maximizing the expected profits where expectation is taken over $F\left(\xi\right)$. Letting $Z^*=(X^*,\xi^*,\omega^*)$ the realized value of the derivative is given as $\nu(Z^*;\theta)=\frac{\partial\Pi(Z^*;\theta)}{\partial X}$, given as

$$\begin{aligned}\nu(Z^*;\theta) \;=\; & \left(p\left(X^*,\omega^*;\theta\right)-mc\left(X^*,\omega^*;\;\theta\right)\right)\frac{d\;s\left(p\left(X^*,\omega^*;\theta\right),X^*,\xi^*;\;\theta\right)}{d\;X}\\ & \frac{\partial\left(p\left(X^*,\omega^*;\theta\right)-mc\left(X^*,\omega^*;\;\theta\right)\right)}{\partial X}s\left(p\left(X^*,\omega^*;\theta\right),X^*,\xi^*;\;\theta\right).\end{aligned}$$

This derivative will sometimes be positive and sometimes be negative depending upon whether "too much" or "too little" of X was chosen prior to the realized demand shock. $\nu(Z^*;\theta)$ evaluated at the realized $p^* = p(Z^*)$, the profit maximizing price given Z^*, is zero by construction of p^*. By the way the data are constructed on average these "mistakes" will average out:

$$E\left[\nu(Z^*;\theta)\right|\omega^*, F(\xi)] = 0,$$

and this moment is our source of identification. When evaluating ν, we mimic the standard empirical settings where econometricians cannot observe cost and demand shocks. At each θ we apply Berry-inversion in (1) to find $\xi\left(X^*, p^*;\ \theta^D\right)$ and FOC with respect to price in (3) to find $mc\left(p^*, s\left(p^*, X^*, \xi\left(X^*, p^*;\ \theta^D\right)\right);\ \theta^D\right)$. Then, ν is evaluated at $p^*, X^*, mc\left(p^*, s\left(p^*, X^*, \xi\left(X^*, p^*;\ \theta^D\right)\right);\ \theta^D\right)$, and $s\left(p^*, X^*, \xi\left(X^*, p^*;\ \theta^D\right);\ \theta^D\right)$ at each θ. This is what I did in mc exercise but it sounds wrong as ν evaluated at realized p^* automatically become zero without an error at the true θ. [Maybe the right way to evaluate ν in estimation is to derive $p\left(X^*, \omega;\ \theta\right)$ where ω is found from inverted mc minus $\gamma ln X^*$]

In this monte carlo setup we could construct the optimal instruments, which are given by

$$H = \frac{1}{E\left[\nu\left(Z^*;\theta\right)^2 \middle| \omega^*, F(\xi)\right]} E\left[\frac{\partial \nu\left(Z^*;\theta\right)}{\partial \theta}\middle| \omega^*, F(\xi)\right].$$

Letting $E\left[\nu\left(Z^*;\theta\right)^2 \middle| \omega^*, F(\xi)\right]$ any scalar, the l element of the instrument is given as

$$H_l = E\left[\frac{\partial \nu\left(Z^*;\theta\right)}{\partial \theta_l}\middle| \omega^*, F(\xi)\right] \quad \forall l$$

because we know the distribution of Z^*. 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^*. That is, H_l is approximated at at some initial value $\hat{\theta}$ and $p\left(X^*, \omega^*;\ \hat{\theta}\right) \approx p\left(X^*, \omega\left(X^*, p^*, \xi\left(X^*, p^*;\ \hat{\theta}^D\right);\ \hat{\theta}\right);\ \hat{\theta}\right)$ where ω^* is approximated by $mc\left(p^*, s\left(p^*, X^*, \xi\left(X^*, p^*;\ \theta^D\right)\right);\ \theta^D\right) - \hat{\gamma} ln X^*$. We also check to see whether including $\frac{\partial X}{\partial \theta}$ and $\frac{\partial p}{\partial \theta}$ makes a big difference in efficiency as these terms that are very difficult to approximate in any empirical application (we can numerically approximate them). The moment condition, analogous to (10) is

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

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

We simulate $M = 1000$ markets for each of $N = 200$ times. Table 1 and 2 show the estimated results with logit and random coefficient setup, respectively. In both cases the average of the point estimates across the 200 simulations is very close to the truth and the standard deviations of the estimates across monte carlos are small relative to the magnitude of their respective coefficients.

4.2 Duopoly

Suppose two single-product firms, $j = 1, 2$. Utility of consumer i is identical to above except she now has two choices,

$$u_{ij} = \beta_{i0} + X_j\beta_i - \alpha p_j + \xi_j + \varepsilon_{ji}.$$

Each firm has marginal cost, $ln(mc_j) = \gamma\ lnX_j + \omega_j$. Let $Z_j = (X_j, \xi_j, \omega_j)$. Each firm j's profits are given as

$$\Pi_j\left(Z_1, Z_2;\ \theta\right) = \left(p_j - mc_j\left(X_j, \omega_j\right);\ \theta\right)\ s_j\left(p_1, p_2, X_1, X_2, \xi_1, \xi_2;\ \theta\right)$$

where p_js are determined by competing in Bertrand Nash after X_js are set.

The timing is specified identically to the monopoly case but the information set of each firm I_j now includes own cost and demand shocks. First, cost and demand shocks $(\xi_1^{*}, \xi_2^{*}, \omega_1^{*}, \omega_2^{*})$ are realized, and each firm observes own ξ_j^{*} and ω_j^{*} but not the competitor's. Both firms know the demand parameters and the distributions of the each shock, $F_\xi\left(\xi_1\right) = F_\xi\left(\xi_2\right)$ and $F_\omega\left(\omega_1\right) = F_\omega\left(\omega_2\right)$. Each firm j simultaneously chooses X_j given

(ξ_j^*, ω_j^*):

$$\begin{aligned}
\max_{X_j} \quad & E\left[\Pi_j\left(Z_1, Z_2; \theta\right) \middle| \xi_j^*, \omega_j^*, F_\xi, F_\omega\right] \\
= \quad & \int\int \left(p_j - mc_j\left(X_j, \omega_j^*;\ \theta\right)\right)\ s_j\left(p_1, p_2, X_1, X_2, \xi_1, \xi_2;\ \theta\right)\ dF_\xi\left(\xi_{-j}\right)\ dF_\omega\left(\omega_{-j}\right),
\end{aligned}$$

knowing p_1 and p_2 will be set optimally once the both firms' characteristics are set.

Let $\hat{Z}_j = \left(X_j^*, \xi_j, \omega_j\right)$ and $Z_j^* = \left(X_j^*, \xi_j^*, \omega_j^*\right)$ with X_j^* the optimal choice of X_j: X_j^* solves

$$\begin{aligned}
& E\left[\left.\frac{\partial \Pi_j\left(Z_j^*, \hat{Z}_{-j}; \theta\right)}{\partial X_j}\right| \xi_j^*, \omega_j^*, F_\xi, F_\omega\right] \\
= \ & \int\int \left(\left(p_j - mc_j\left(X_j^*, \omega_j^*;\ \theta\right)\right) \frac{d\ s_j\left(p_1, p_2, X_1^*, X_2^*, \xi_j^*, \xi_{-j};\ \theta\right)}{d\ X_j}\right)\ dF_\xi\left(\xi_{-j}\right)\ dF_\omega\left(\omega_{-j}\right) \\
+ \ & \int\int \left(\frac{\partial\left(p_j - mc_j\left(X_j^*, \omega_j^*;\ \theta\right)\right)}{\partial X_j} s_j\left(p_1, p_2, X_1^*, X_2^*, \xi_j^*, \xi_{-j};\ \theta\right)\right)\ dF_\xi\left(\xi_{-j}\right)\ dF_\omega\left(\omega_{-j}\right) \\
= \ & 0 \quad \forall j = 1, 2
\end{aligned}$$

where $p_j = p_j\left(X_1^*, X_2^*, \xi_j^*, \omega_j^*;\ \theta, F_\xi, F_\omega\right)$, the optimal price maximizing the expected profits under Bertrand Nash where expectation is taken over F_ξ and F_ω.

The realized value of the derivative is given as $\nu_j\left(Z_1^*, Z_2^*;\ \theta\right) = \frac{\partial \Pi_j\left(Z_1^*, Z_2^*; \theta\right)}{\partial X_j}$:

$$\begin{aligned}
\nu_j\left(Z_1^*, Z_2^*;\ \theta\right) \quad = \quad & \left(p_j - mc_j\left(X_j^*, \omega_j^*;\ \theta\right)\right) \frac{d\ s_j\left(p_1, p_2, X_1^*, X_2^*, \xi_1^*, \xi_2^*;\ \theta\right)}{d\ X_j} \\
& \frac{\partial\left(p_j - mc_j\left(X_j^*, \omega_j^*;\ \theta\right)\right)}{\partial X_j} s\left(p_1, p_2, X_1^*, X_2^*, \xi_1^*, \xi_2^*;\ \theta\right).
\end{aligned}$$

Notice that $\nu_j\left(Z_1^*, Z_2^*;\ \theta\right)$ is evaluated at $p_j = p_j\left(X_1^*, X_2^*, \xi_j^*, \omega_j^*;\ \theta, F_\xi, F_\omega\right)$.[9] This derivative will not be exactly zero because X is chosen prior to the competitors' realized demand and supply shocks. By the way the data are constructed on average these "mistakes" average out:

$$E\left[\nu_j(Z_j^*, Z_j^*; \theta) \middle|\ \xi_j^*, \omega_j^*, F_\xi, F_\omega\right] = 0, \quad \forall j = 1, 2$$

[9] $\nu_j\left(Z_1^*, Z_2^*;\ \theta\right)$ evaluated at $p_j^* = p(Z_1^*, Z_2^*)$ is zero by construction of p_j^*.

and this moment is our source of identification. When evaluating ν, we pretend that we do not observe ω_js. At each θ, we "inver out" $\xi_j\left(X_1^*, X_2^*, p_1^*, p_2^*;\ \theta^D\right)$s and derive $mc_j\left(p_j^*, s_j\left(p_1^*, p_2^*, X_1^*, X_2^*, \xi_1\left(X_1^*,\right.\right.\right.$ according to (1) and (3), respectively.

The optimal instruments are given as

$$H = E\left[\nu\left(Z_1^*, Z_2^*; \theta\right)\nu\left(Z_1^*, Z_2^*; \theta\right)' \mid \xi_j^*, \omega_j^*, F_\xi, F_\omega\right]^{-1} E\left[\frac{\partial \nu\left(Z_1^*, Z_2^*; \theta\right)'}{\partial \theta} \mid \xi_j^*, \omega_j^*, F_\xi, F_\omega\right]'$$

where $\nu\left(Z_1^*, Z_2^*; \theta\right)$ is a 2 by 1 vector of $(\nu_j)_{j=1,2}$. Letting $E\left[\nu\left(Z_1^*, Z_2^*; \theta\right)\nu\left(Z_1^*, Z_2^*; \theta\right)' \mid \xi_j^*, \omega_j^*, F_\xi, F_\omega\right] = I_2$, the lth column of the instrument is given as

$$H_l = E\left[\frac{\partial \nu\left(Z_1^*, Z_2^*; \theta\right)'}{\partial \theta_l} \mid \xi_j^*, \omega_j^*, F_\xi, F_\omega\right] \quad \forall l.$$

Approximation of the optimal instruments is done by solving the equilibrium $p_j(X_1^*, X_2^*, \xi_j\left(X_1^*, X_2^*, p_1^*, p_2^*;\ \hat{\theta}\right.$ at the initial value $\hat{\theta}$ and realized $X_1^*, X_2^*, \xi_1\left(X_1^*, X_2^*, p_1^*, p_2^*;\ \theta^D\right), \xi_2\left(X_1^*, X_2^*, p_1^*, p_2^*;\ \theta^D\right), \omega_1^*$, and ω_2^* where ω_j^*s are approximated by $mc_j\left(p_j^*, s_j\left(p_1^*, p_2^*, X_1^*, X_2^*, \xi_1\left(X_1^*, X_2^*, p_1^*, p_2^*;\ \theta^D\right), \xi_2\left(X_1^*, X_2^*, p_1^*, p_2^*;\right.\right.\right.$ $\hat{\gamma} ln X_j^*$.

The moment condition, analogous to (10) is

$$\begin{aligned} G_l\left(Z_1^*, Z_2^*; \theta\right) &\equiv E\left[\hat{H}_l'\ \nu(Z_1^*, Z_2^*; \theta)\right] \quad \forall l \\ &= 0 \end{aligned}$$

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

5 Application to BLP Data

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 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.[10] In a market with J products there are J observations on the K realized first-order conditions. The outside good qualilty ξ_0 is normalized to zero and we do not separately estimate the mean utility for new vehicles 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.

"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, $I^{\text{contemporaneous}}$, uses contemporaneous characteristics to construct $\nu_{jk}(\theta)$. We approximate $E\left[\nu(\theta)\nu(\theta)'|I\right] = I_{JK}$.[11] We transform the instrument $\hat{H}_{jkl}$ into a block diagonal matrix so that we have $K * 15 = 75$ instruments as a benchmark specification.

Table 4 shows the demand and supply estimates. The first column restates the original BLP results and columns two and three labeled with FOC (first-order condition) are estimated using 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. Column two uses the full set of instruments (Full IV) of which there are 43, one for each parameter-FOC pair after dropping 32 instruments 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.[12] 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

[10] 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.

[11] 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.

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

to $(\alpha, \gamma_k, \sigma_k)$ giving us a total of 19 instruments. In the robustness section we estimate parameters based on the assumption that firms know their competitors' contemporaneous characteristics when making choices.

The most striking difference between the BLP estimates in column one and optimization estimates in columns two and three 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 5 investigates the impact of this difference on estimated elasticities and markups. On average elasticities double in absolute value in response to the increase in price sensitivity. This causes estimated markups to fall by on average around 50%.

One possible explanation for these changes across identification conditions is that observed and unobserved characteristics are positively correlated because firms put more unobserved 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 6 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 70% 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.

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 7 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 0.5. 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.

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 4, which reports the average ξ's from each model's fit. BLP fits 90% of households not buying by having consumers derive strong negative unobserved quality on average from the act of buying a car relative to the outside option of no new car ($\bar{\xi} = -7.10$ versus $\xi_0 = 0$). In contrast, the optimization-fit has consumers strongly desiring new cars relative to the outside good – ($\bar{\xi} = 3.52$) – 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 4 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.

Before turning to the rest of our results we note that 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 European automobile market. Using the optimization moments estimated price elasticities double on average and estimated markups fall relative to when they use the BLP moments, They find some anomalous demand and supply point estimates under the BLP-fit that are not present under optimization-fit. Under the optimization-fit their standard errors are much smaller and their unobserved quality term is positively correlated with observed characteristics.

Table 8 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.

6 Robustness

Characteristics of automobile do not change every year. We allow for the "dynamic" optimization FOCs, by restricting the observations to when at least one characteristic is changed or a new model is introduced. It is equivalent to approximating the sum of the future stream of profits at the time of changing the characteristics to $E\left[\Pi_f \mid I_{f,t}\right] = E\left[\sum_{\tau=t}^{T-1} \pi_{f,\tau} \mid I_{f,t}\right] \approx E\left[\pi_{f,t} \mid I_{f,t}\right]$. T refers to the year when at least one characteristic is changed for more than 10%. The first two columns in Table 9 compares the benchmark results and the "dynamic" results. Column 1 is the Full IV estimates presented in Table 4 with the benchmark "lagged" information set. Column 2 is the Full IV estimates with the "lagged" information set, only using the restricted observations. Although this reduces the number of observations to approximately half the standard errors do not increase much as most of the variation in moments originate from the restricted observations.

In the last two columns we estimate the "lagged" model by applying the instruments suggested in BLP – own characteristics and the sum of own and competitors' characteristics. As long as the BLP instruments are evaluated at the values given by the "lagged" information set, they are mean independent from the moment ν_{jk}. Standard BLP instruments are 15 but we include three additional BLP instruments with respect to prices as they are a function of the given information. Therefore, there are 18 BLP instruments, evaluated at the competitors' previous characteristics and the simulated equilibrium prices at those values, for each moment. Coefficients are similar to the benchmark specification. Standard errors tend to be lower as there are more instruments.

Table 10 presents robustness results for two additional specifications. Column 1 is the Full IV estimates presented in Table 4 with the benchmark "lagged" information set. Given the precision of the estimates in Table 4 (under the "lagged" information set assumption) we estimate the model allowing for different tastes and cost structures between 1971-1980 and 1981-1990 each of which use 834 and 1175 observations, respectively, in Column 2 and 3. Estimates are still precisely estimated but the standard errors with half samples are almost always higher than those with entire samples, and the magnitude of difference is close to $\frac{1}{\sqrt{2}}$. Most model parameters are similar in the 1970s relative to 1980s.

Column 4 uses the "contemporaneous" information set. While most of the coefficients in Column 4 are similar to Column 1 some are different although none significantly. The point estimate of the price coefficient decreases 5% but again the difference is not significant. Standard errors for some parameters go down relative to the "lagged" information set assumption.

7 Conclusions

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, "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.

Goolsbee, Austan and Amil Petrin, "The Consumer Gains from Direct Broadcast Satellites and the Competition with Cable TV," *Econometrica*, 2004, *72* (2), 351–381.

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.

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, 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

Table 1: Monte Carlo Simulation - Monopoly, Logit

Parameter	Truth	J=1000		J=10000	
		IV at truth	IV at non-truth	IV at truth	IV at non-truth
β_X	3	3.026	2.995	3.026	3.012
		(0.006)	(0.034)	(0.002)	(0.011)
γ	2	2.047	2.019	2.048	2.029
		(0.008)	(0.052)	(0.003)	(0.016)
α	10	9.998	10.001	9.997	9.998
		(0.001)	(0.004)	(0.000)	(0.001)
J		1000	1000	10000	10000

Table 2: Monte Carlo Simulation - Monopoly, Random Coefficient

Parameter	Truth	J=1000		J=10000	
		IV at truth	IV at non-truth	IV at truth	IV at non-truth
β_X	3	3.079	3.090	3.081	3.081
		(0.408)	(0.045)	(0.129)	(0.014)
γ	2	2.028	2.020	2.027	2.029
		(0.772)	(0.065)	(0.244)	(0.021)
α	10	9.982	9.987	9.983	9.980
		(0.104)	(0.069)	(0.033)	(0.022)
σ_1	0.1	0.103	0.104	0.104	0.103
		(0.495)	(0.230)	(0.157)	(0.073)
σ_X	0.5	0.509	0.511	0.509	0.512
		(1.809)	(0.750)	(0.572)	(0.237)
J		1000	1000	10000	10000

Table 3: Monte Carlo Simulation - Duopoly, J=2000

Parameter	Truth	IV at truth
β_X	3	2.9866
		(0.4021)
γ	2	1.96
		(0.1476)
α	10	10.1068
		(1.3635)
σ_1	1.5	1.5008
		(0.1051)
σ_X	1	0.9935
		(0.0681)
J		2000

Table 4: Estimated Parameters of the Demand and Supply

	Variable	BLP	FOC	
			Full IV	Part IV
Term on Price α	$ln(y-p)$	43.501	144.131	155.661
		(6.427)	(34.054)	(81.672)
Means (β's)	Constant	-7.061		
		(0.941)		
	HP/weight	2.883	1.168	1.527
		(2.019)	(0.192)	(0.884)
	Size	3.460	0.108	0.316
		(0.610)	(0.036)	(0.422)
	Air	1.521	1.722	0.975
		(0.891)	(0.331)	(0.812)
	MP$	-0.122	2.442	2.401
		(0.320)	(0.414)	(1.633)
Std. Deviations (σ's)	Constant	3.612	3.190	3.093
		(1.485)	(1.093)	(6.304)
	HP/weight	4.628	3.007	2.818
		(1.885)	(0.587)	(2.351)
	Size	2.056	0.934	0.919
		(0.585)	(0.150)	(0.747)
	Air	1.818	1.773	1.607
		(1.695)	(0.257)	(1.247)
	MP$	1.050	0.859	1.612
		(0.272)	(0.286)	(0.984)
Cost Side Parameters	Constant	0.952		
		(0.194)		
	Mean Charac (ξ)		0.122	0.121
			(0.028)	(0.056)
	ln(HP/weight)	0.477	0.059	0.067
		(0.056)	(0.015)	(0.041)
	ln(Size)	-0.046	0.030	0.054
		(0.081)	(0.006)	(0.066)
	Air	0.619	0.226	0.135
		(0.038)	(0.042)	(0.078)
	ln(MPG)	-0.415	0.551	0.538
		(0.055)	(0.081)	(0.386)
Mean Unobserved	$\bar{\xi}$	-7.10	5.18	6.10
Num. IVs	15	43	19	
J	2217	2125	2125	

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

Table 5: 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	9214.54	5754.35	5553.02
Lincoln Towncar	-3.030	-5.708	-5.973	8310.82	4633.86	4435.15
Nissan Maxima	-4.124	-7.867	-8.155	3385.84	1780.48	1716.92
Ford Taurus	-3.952	-8.205	-8.966	2679.14	1363.66	1244.59
Chevy Cavalier	-5.899	-10.284	-11.668	1327.75	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	4051.87	2393.99	2280.53
Median	-3.975	-8.219	-8.789	2751.77	1397.99	1324.34
Std. Deviation	1.120	2.130	2.577	3905.32	2821.63	2712.61

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

Table 6: $E\left[\xi_j \mid X\right] \neq 0$

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.

Table 7: 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 8: $E\left[\omega_j \mid W\right] \neq 0$

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.

Table 9: Robustness to Dynamic Optimization and BLP IVs

	Variable	FOC IVs		BLP IVs	
		Lagged	Dynamic	Lagged	Dynamic
Term on Price α	$ln(y-p)$	144.131	162.919	161.879	162.565
		(34.054)	(48.834)	(18.299)	(11.774)
Means (β's)	HP/weight	1.168	0.937	0.77	0.996
		(0.192)	(0.229)	(0.091)	(0.073)
	Size	0.108	0.133	0.497	0.266
		(0.036)	(0.034)	(0.058)	(0.022)
	Air	1.722	1.319	1.957	1.220
		(0.331)	(0.276)	(0.188)	(0.089)
	MP$	2.442	1.678	2.227	1.871
		(0.414)	(0.345)	(0.217)	(0.114)
Std. Deviations (σ's)	Constant	3.19	3.33	2.64	3.912
		(1.093)	(1.886)	(0.607)	(0.440)
	HP/weight	3.007	2.986	1.963	2.442
		(0.587)	(0.662)	(0.199)	(0.140)
	Size	0.934	0.641	1.25	0.828
		(0.150)	(0.128)	(0.093)	(0.040)
	Air	1.773	2.009	1.199	1.518
		(0.257)	(0.394)	(0.131)	(0.087)
	MP$	0.859	0.846	0.767	0.815
		(0.286)	(0.356)	(0.120)	(0.075)
Cost Side Parameters	Mean Charac (ξ)	0.122	0.113	0.12	0.114
		(0.028)	(0.033)	(0.013)	(0.008)
	ln(HP/weight)	0.059	0.049	0.042	0.049
		(0.015)	(0.016)	(0.005)	(0.004)
	ln(Size)	0.03	0.026	0.097	0.050
		(0.006)	(0.007)	(0.011)	(0.004)
	Air	0.226	0.17	0.246	0.150
		(0.042)	(0.042)	(0.025)	(0.009)
	ln(MPG)	0.551	0.423	0.543	0.438
		(0.081)	(0.086)	(0.053)	(0.028)
Mean Unobserved	$\bar{\xi}$	5.18	8.929	7.208	8.40
Num. IVs		43	59	75	75
J		2125	902	2125	902

"FOC IVs" are the benchmark results, using FOCs as moments and the optimal instruments H in (9). Lagged results are the same as in FOC Full IV in Table 4.
"BLP IVs" still use FOCs as moments but use BLP IVs as instruments.
"Lagged" specifications evaluate instruments conditional on previous year's competitors' portfolios are included in firm's information set.
"Dynamic" only use observations if at least one characteristic is changed by at least 10% from the previous year or the new product is introduced.

Table 10: Robustness to Firm Information and Subsets of Data

	Variable	Lagged			Contemporaneous
		Pooled	70s	80s	Pooled
Term on Price α	$ln(y-p)$	144.131	140.154	161.219	137.032
		(34.054)	(59.361)	(25.078)	(17.737)
Means (β's)	HP/weight	1.168	1.473	1.472	0.886
		(0.192)	(0.590)	(0.197)	(0.158)
	Size	0.108	0.201	0.307	0.618
		(0.036)	(0.088)	(0.041)	(0.062)
	Air	1.722	1.489	1.552	1.324
		(0.331)	(0.543)	(0.344)	(0.140)
	MP$	2.442	2.214	1.985	2.644
		(0.414)	(0.837)	(0.252)	(0.295)
Std. Deviations (σ's)	Constant	3.19	2.259	2.232	2.276
		(1.093)	(1.355)	(0.739)	(0.471)
	HP/weight	3.007	4.407	3.146	2.963
		(0.587)	(1.355)	(0.422)	(0.389)
	Size	0.934	0.576	0.664	0.371
		(0.150)	(0.201)	(0.095)	(0.169)
	Air	1.773	1.634	1.612	1.286
		(0.257)	(1.007)	(0.181)	(0.150)
	MP$	0.859	0.697	0.68	0.771
		(0.286)	(0.411)	(0.155)	(0.150)
Cost Side Parameters	Mean Charac (ξ)	0.122	0.124	0.123	0.134
		(0.028)	(0.050)	(0.018)	(0.017)
	ln(HP/weight)	0.059	0.076	0.072	0.053
		(0.015)	(0.035)	(0.013)	(0.011)
	ln(Size)	0.03	0.045	0.049	0.118
		(0.006)	(0.015)	(0.008)	(0.013)
	Air	0.226	0.203	0.228	0.188
		(0.042)	(0.092)	(0.041)	(0.024)
	ln(MPG)	0.551	0.544	0.565	0.636
		(0.081)	(0.181)	(0.081)	(0.075)
Mean Unobserved	$\bar{\xi}$	5.18	6.25	6.89	3.52
Num. IVs		43	71	71	72
J		2125	834	1175	2217

"Lagged" are the benchmark results, using instruments conditional on previous year's competitors' portfolios are included in firm's information set. Pooled results are the same as in FOC Full IV in Table 4.
"Contemporaneous" columns use instruments assuming that competitors' contemporaneous product portfolios are known to a firm.
"70s" and "80s" are estimated only using subsamples during 1970s and 1980s, respectively.