Instrument-free Identification and Estimation of Differentiated Products Models using Cost Data

David P. Byrne[a], Susumu Imai[b], Neelam Jain[c] *, Vasilis Sarafidis[d]

[a]Department of Economics, University of Melbourne, Building 110 The University of Melbourne Parkville Victoria 3010 Australia

[b]Department of Economics and Business, Hokkaido University, Kita 9 Nishi 7 Kita ku Sapporo 060-0809 Japan

[c]Department of Economics, City, University of London, Northampton Square London EC1V 0HB United Kingdom

[d]BI Norwegian Business School, Nydalsveien 37 N-0484 Oslo Norway

January 23, 2022

Abstract

We propose a new methodology for identifying and estimating demand in differentiated products models when demand and cost data are available. The method deals with the endogeneity of prices to demand shocks and the endogeneity of outputs to cost shocks by using cost data rather than instruments. Further, we allow for unobserved market size. Using Monte Carlo experiments, we show that our method works well in contexts where commonly used instruments are invalid. We also apply our method to the estimation of deposit demand in the US banking industry.

Keywords: Differentiated products oligopoly, Instruments, Identification, Cost data.

JEL Codes: C13, C18, L13, L41

*Corresponding author

E-mail addresses: byrned@unimelb.edu.au (David Byrne), susimaig@gmail.com (Susumu Imai), Neelam.Jain.1@city.ac.uk (Neelam Jain), vasilis.sarafidis@bi.no. (Vasilis Sarafidis)

1 Introduction

In this paper, we develop a new methodology for identifying and estimating models of differentiated products markets. Our approach requires commonly used demand-side data on products' prices, market shares and observed characteristics, and firm-level cost data. The novelty of our method is that it does not use instrumental variables (IV) to deal with the endogeneity of prices to demand shocks in estimating demand. Instead, we use cost data for identification and estimation of demand parameters. Market-level demand and cost data tend to be available for large industries that are subject to regulatory oversight. Examples include banking, telecommunications, and nursing home care. Such major sectors of the economy represent natural settings for the application of our estimator.

The frameworks of interest for this paper are the logit and the random coefficient logit models of Berry (1994) and Berry et al. (1995) (hereafter, BLP), which have had a substantial impact on empirical research in IO and various other areas of economics.[1] These models of demand incorporate unobserved heterogeneity in product quality, which may lead to price endogeneity issues. Researchers predominantly use instruments to deal with such issues. As Berry and Haile (2014), and others, point out, as long as there are valid instruments available, demand functions can be identified using market-level data.[2]

However, these models of demand have become more complex over time in order to incorporate the rich heterogeneities of agents, thus, requiring more instruments, and their interactions, for identifying and estimating the parameters of interest. It is, in general, a challenge to convincingly argue that all these instruments are valid. Through our Monte Carlo exercises as well as an empirical application, we show how a small subset of invalid instruments can greatly bias parameter estimates in unanticipated directions in nonlinear demand models. In contrast, our approach based on the cost data tends to deliver consistent and reasonable parameter estimates.

Our main theoretical finding is that by combining the demand and cost data, and by using the equilibrium condition that marginal revenue equals marginal cost, one can jointly identify

[1] Leading examples from IO include measuring market power (Nevo (2001)), quantifying welfare gains from new products (Petrin (2002)), and merger evaluation (Houde (2012) and Nevo (2000)). Applications of these methods to other fields include measuring media slant (Gentzkow and Shapiro (2010)), evaluating trade policy (Berry et al. (1999)), and identifying sorting across neighborhoods (Bayer et al. (2007)). For a dynamic extension of the model, see Gowrisankaran and Rysman (2012)

[2] Fox et al. (2012) establish identification of discrete choice models with exogenous regressors. Berry and Haile (2014) prove nonparametric identification of a general market share function when the regressors are endogenous but instruments are available. There has been some research assessing numerical difficulties with the BLP algorithm (Dube et al. (2012) and Knittel and Metaxoglou (2014)), and the use of optimal instruments to help alleviate these difficulties (Reynaert and Verboven (2014)). .

the price coefficients and a nonparametric cost function, without using any instruments. Our methodology uses the inversion procedure developed in Berry (1994), and BLP, not only for the demand side, as they do, but also for the cost side. They use this procedure to express marginal revenue as a function of only observables and parameters. We show that when cost data is available, both marginal revenue and marginal cost can be expressed in this way. This is because we can use the observable cost (or expected cost conditional on other observables) to control for the cost shock. Then, the "market structure" variables, such as the observed characteristics, prices and market shares of rival firms in the market, and market size, which enter in the marginal revenue function, can be used as sources of variation for the identification of price parameters, subject to the exclusion restriction that they do not enter in the cost function. However, these variables do not have to be instruments, that is, they do not need to be orthogonal to the unobserved demand and cost shocks.[3] Further, we show that we do not need any variation in market size to identify and estimate the BLP-demand model but for logit, we do because of the specific nature of the marginal revenue function it implies.

We show that our methodology works even if the true market size is unobservable and possibly correlated with other observed and unobserved variables of the model. We follow Bresnahan and Reiss (1991) partly in that we assume that the variables determining market size do not enter the cost function. However, we do not exclude these variables from the demand function as they do. We believe that our exclusion restriction is more reasonable because typically, the determinants of market size are demographics, which are likely to affect demand but do not affect cost directly.

Our methodology requires relatively weak assumptions. The main requirement we have on the nonparametric cost function is that it is strictly increasing in output and the cost shock. In addition, marginal cost is strictly increasing in the cost shock. We also allow for measurement errors in total cost and fixed cost, as well as, systematic over/under reporting of cost by firms. Further, our identification strategy does not require information on the joint distribution of the demand and cost shocks, as in MacKaye and Miller (2018).

It is important to note that we do not need data on marginal cost or markups. Also, knowledge of the cost function is not necessary. If such information were available, it would be straightforward to use the first order condition to identify the price parameters without any

[3]Petrin and Seo (2019) propose an identification and estimation scheme that allows for observed and unobserved characteristics in their demand equation to be endogenously determined. They skillfully exploit the optimal choice of observed characteristics to create additional moments. However, in the BLP model of demand, the number of first order conditions is less than the number of parameters. Therefore, additional moment restrictions are required.

instruments.[4]

The type of cost data we have in mind comes from firms' income statements and balance sheets, among other sources. Such data has been used extensively in a large parallel literature on estimation of cost functions in empirical IO.[5] Recently, some researchers have also started incorporating cost data as an additional source of variation for identification.[6] However, most of these researchers use instrumental variables (IVs) to identify demand.

For estimation, we propose a two-step Sieve Nonlinear Least Squares (SNLLS) estimator, which avoids having to calculate expected cost conditional on observables, a procedure that is subject to the Curse of Dimensionality. In this estimator, we use marginal revenue, which is a parametric function, to control for the cost shock, rather than the conditional expected cost. We prove that this estimator identifies the true parameters, is consistent and asymptotically normal. This estimator is semiparametric in that it assumes the parametric logit or the BLP demand and a nonparametric cost function. We also show how this estimator can be adapted to accommodate various data and specification issues that arise in practice. These include endogenous product characteristics, imposing restrictions on cost functions such as homogeneity of degree one in input prices, dealing with the difference between accounting cost and economic cost, missing cost data for some products or firms, and multi-product firms.

Through a set of Monte Carlo experiments for the BLP demand model, we illustrate how our estimator delivers consistent parameter estimates even when the demand shock is not only correlated with the equilibrium price and output, but also with the cost shock, input prices, market size and observed characteristics of rival products. Further, we allow the cost shock to be correlated with market size as well. In such a setting, there are no valid instruments to account for price endogeneity. In particular, market size cannot work as an exogenous variation for the supply side, and the orthogonality between the demand and cost shocks cannot be used as a moment restriction for consistent estimation of price parameters. Hence, the IV estimates are shown to be biased. Our numerical exercises also show that variation in market size is needed

[4]See Genesove and Mullin (1998), Wolfram (1999), Clay and Troesken (2003), Kim and Knittel (2003) and Smith (2004), for examples of related research.

[5]Numerous studies (Aigner et al. (1977), Christensen and Greene (1976), Gollop and Roberts (1983) and others) have used such data for various purposes such as to identify inefficiency in production and economies of scale or scope, to measure marginal costs, and to quantify markups for a variety of industries. For identification, researchers either use instruments for output or argue that output is effectively exogenous from firms' point of view in the market they study.

[6]Note that BLP also propose using cost data, if available, for a variety of purposes, including improving their parameter estimates as well as understanding the relationship between prices and marginal costs. Crawford and Yurukoglu (2012) use such data to estimate parameters of a bargaining model. Some researchers have also used demand and cost data to test assumptions regarding conduct in oligopoly models. See, for instance, Byrne (2015). Kutlu and Sickles (2012) use cost data to estimate market power while allowing for inefficiency in production.

for identifying the price parameters of the logit demand but not for the BLP demand model.

We then apply our methodology to the estimation of deposit demand in the US banking industry. We find that our method works well. The magnitude of the estimated deposit interest rate coefficient is similar to the estimates obtained in the existing literature such as Dick (2008) and Ho and Ishii (2012). Further, we find that the IV-based method yields a negative coefficient on the deposit interest rate whereas ours is positive which is what one would expect. These results demonstrate that by comparing the IV-based parameter estimates and those based on our approach, researchers can check the validity of their instruments.

The paper is organized as follows. In Section 2, we specify the standard differentiated products model and review the IV-based estimation approach used in the literature. In Section 3, we explain our identification strategy when demand and cost data are available. In Section 4, we present the two-step SNLLS estimator and analyze its large sample properties. Section 5 contains our main Monte Carlo exercises. In Section 6, we apply our methodology to the estimation of deposit demand in the banking industry. In Section 7, we conclude. The appendix contains several proofs, additional Monte Carlo results and further details of the deposit demand estimation exercise.[7]

2 Differentiated products models and IV estimation

In this section, we describe the standard differentiated products model that we adopt and provide an overview of the IV estimation method. For more details, see Berry (1994), BLP, Nevo (2001) and others. Most features of the model we discuss here are carried over to the next section where we explain our cost data-based identification strategy.

2.1 Differentiated products discrete choice demand models

In the standard model, consumer i in market m gets the following utility from consuming one unit of product j:

$$u_{ijm} = \mathbf{x}_{jm}\boldsymbol{\beta} + p_{jm}\alpha + \xi_{jm} + \epsilon_{ijm},$$

where $\mathbf{x}_{jm}$ is a $1\times K$ vector of observed product characteristics, p_{jm} is price, ξ_{jm} is the unobserved product quality (or demand shock) that is known to both consumers and firms but unknown to researchers, and ϵ_{ijm} is an idiosyncratic taste shock. The demand parameter vector is denoted by $\boldsymbol{\theta} = \left[\alpha, \boldsymbol{\beta}'\right]'$, where $\boldsymbol{\beta}$ is a $K \times 1$ vector.

[7] The data and code that support the findings of this study are available on request from Susumu Imai.

It is assumed that there are $M > 1$ isolated markets.[8] Market m has $J_m + 1 > 2$ products whose aggregate demand across individuals is,

$$q_{jm} = s_{jm} Q_m,$$

where q_{jm} denotes output, Q_m denotes market size and s_{jm} denotes market share. In the case of the Berry (1994) logit demand model, ϵ_{ijm} is assumed to have a logit distribution. Then, the aggregate market share for product j in market m is,

$$s_{jm}(\boldsymbol{\theta}) \equiv s_j\left(\mathbf{p}_m, \mathbf{X}_m, \boldsymbol{\xi}_m; \boldsymbol{\theta}\right) = \frac{\exp\left(\mathbf{x}_{jm}\boldsymbol{\beta} + p_{jm}\alpha + \xi_{jm}\right)}{\sum_{k=0}^{J_m} \exp\left(\mathbf{x}_{km}\boldsymbol{\beta} + p_{km}\alpha + \xi_{km}\right)} = \frac{\exp\left(\delta_{jm}\right)}{\sum_{k=0}^{J_m} \exp\left(\delta_{km}\right)}, \tag{1}$$

where $\mathbf{p}_m = [p_{0m}, p_{1m}, ..., p_{J_m m}]'$ is a $(J_m + 1) \times 1$ vector,

$$\mathbf{X}_m = \begin{bmatrix} \mathbf{x}_{0m} \\ \mathbf{x}_{1m} \\ \vdots \\ \mathbf{x}_{J_m m} \end{bmatrix}$$

is a $(J_m + 1) \times K$ matrix, $\boldsymbol{\xi}_m = [\xi_{0m}, \xi_{1m}, ..., \xi_{J_m m}]'$ is a $(J_m + 1) \times 1$ vector, and

$$\delta_{jm} \equiv \mathbf{x}_{jm}\boldsymbol{\beta} + p_{jm}\alpha + \xi_{jm} \tag{2}$$

is the "mean utility" of product j in market m. Using this definition, we can express the market share in Equation (1) as $s_j\left(\boldsymbol{\delta}(\boldsymbol{\theta})\right) \equiv s_j\left(\mathbf{p}_m, \mathbf{X}_m, \boldsymbol{\xi}_m; \boldsymbol{\theta}\right)$ where $\boldsymbol{\delta}(\boldsymbol{\theta}) = [\delta_{0m}(\boldsymbol{\theta}), \delta_{1m}(\boldsymbol{\theta}), \ldots, \delta_{J_m m}(\boldsymbol{\theta})]'$.

Good $j = 0$ is labeled the "outside good" or "no-purchase option" that corresponds to not buying any of the $j = 1, \ldots, J_m$ goods. This good's product characteristics, price, and demand shock are normalized to zero (i.e., $\mathbf{x}_{0m} = \mathbf{0}$, $p_{0m} = 0$, and $\xi_{0m} = 0$ for all m), which implies

$$\delta_{0m}(\boldsymbol{\theta}) = 0. \tag{3}$$

This normalization, together with the logit assumption for the distribution of ϵ_{ijm}, identifies the level and scale of utility.

In BLP, or equivalently, the random coefficient logit model, one allows the price coefficient and coefficients on the observed characteristics to be different for different consumers. Specifically, α

[8] With panel data, the m index corresponds to a market-period.

has a distribution function $F_\alpha(.;\boldsymbol{\theta}_\alpha)$, where $\boldsymbol{\theta}_\alpha$ is the parameter vector of the distribution, and similarly, $\boldsymbol{\beta}$ has a distribution function $F_{\boldsymbol{\beta}}(.;\boldsymbol{\theta}_{\boldsymbol{\beta}})$ with parameter vector $\boldsymbol{\theta}_{\boldsymbol{\beta}}$. The probability with which a consumer with coefficients α and $\boldsymbol{\beta}$ purchases product j is identical to that provided by the market share formula in Equation (1). The aggregate market share of product j is obtained by integrating over the distributions of α and $\boldsymbol{\beta}$:

$$s_j(\mathbf{p}_m, \mathbf{X}_m, \boldsymbol{\xi}_m; \boldsymbol{\theta}) = \int_\alpha \int_{\boldsymbol{\beta}} \frac{\exp(\mathbf{x}_{jm}\boldsymbol{\beta} + p_{jm}\alpha + \xi_{jm})}{\sum_{k=0}^{J_m} \exp(\mathbf{x}_{km}\boldsymbol{\beta} + p_{km}\alpha + \xi_{km})} dF_{\boldsymbol{\beta}}(\boldsymbol{\beta}; \boldsymbol{\theta}_{\boldsymbol{\beta}})\, dF_\alpha(\alpha; \boldsymbol{\theta}_\alpha), \tag{4}$$

where $\boldsymbol{\theta} = [\boldsymbol{\theta}'_\alpha, \boldsymbol{\theta}'_{\boldsymbol{\beta}}]'$. Letting μ_α to be the mean of α and $\boldsymbol{\mu}_{\boldsymbol{\beta}}$ the mean of $\boldsymbol{\beta}$, the mean utility is defined to be

$$\delta_{jm} \equiv \mathbf{x}_{jm}\boldsymbol{\mu}_{\boldsymbol{\beta}} + p_{jm}\mu_\alpha + \xi_{jm}, \tag{5}$$

with $\delta_{0m} = 0$ for the outside good.

2.1.1 Recovering demand shocks

For each market $m = 1, \ldots M$, researchers are assumed to have data on prices $\mathbf{p}_m$, market shares $\mathbf{s}_m = [s_{0m}, s_{1m}, ..., s_{J_m m}]'$ and observed product characteristics $\mathbf{X}_m$ for all the firms in the market. Given $\boldsymbol{\theta}$ and this data, one can solve for the vector $\boldsymbol{\delta}_m$ through market share inversion. That is, if we denote $s_j(\boldsymbol{\delta}_m(\boldsymbol{\theta}); \boldsymbol{\theta})$ to be the market share of firm j predicted by the model, market share inversion involves obtaining $\boldsymbol{\delta}_m$ by solving the following set of J_m equations,

$$s_j(\boldsymbol{\delta}_m(\boldsymbol{\theta}), j; \boldsymbol{\theta}) - s_{jm} = 0, \ \mathit{for}\ j = 0, \ldots, J_m, \tag{6}$$

and therefore,

$$\boldsymbol{\delta}_m(\boldsymbol{\theta}) = \mathbf{s}^{-1}(\mathbf{s}_m; \boldsymbol{\theta}). \tag{7}$$

The vector of mean utilities that solves these equations perfectly aligns the model's predicted market shares to those observed in the data.

IIn the logit model, Berry (1994) shows that we can easily recover mean utilities for product j using its market share and the share of the outside good as $\delta_{jm}(\boldsymbol{\theta}) = \log(s_{jm}) - \log(s_{0m})$, $j = 1, \ldots, J_m$. In the random coefficient model, there is no such closed-form formula for market share inversion. Instead, BLP propose a contraction mapping algorithm that recovers the unique $\delta_{jm}(\boldsymbol{\theta})$ that solves Equation (7) under some regularity conditions. In both cases, δ_{0m} is normalized to 0.

With the mean utilities and parameters in hand, one can recover the structural demand shocks straightforwardly from Equation (2) for the logit demand and Equation (5) for the BLP demand.

2.1.2 IV estimation of demand

A simple regression of Equation (2) or (5) with $\delta_{jm}(\boldsymbol{\theta})$ being the dependent variable and $\mathbf{x}_{jm}$ and p_{jm} being the regressors would yield a biased estimate of the price coefficient. This is because firms likely set higher prices for products with higher unobserved quality, which creates a correlation between p_{jm} and ξ_{jm}, violating the OLS orthogonality condition $E[\xi_{jm}p_{jm}] = 0$. Researchers use a variety of demand instruments to overcome this issue. In particular, they construct a GMM estimator for $\boldsymbol{\theta}$ by assuming that the following population moment conditions are satisfied at the true value of the demand parameters, denoted by $\boldsymbol{\theta}_0$:

$$E[\xi_{jm}(\boldsymbol{\theta}_0)\,\mathbf{z}_{jm}] = \mathbf{0},$$

where $\mathbf{z}_{jm}$ is an $L \times 1$ vector of instruments that is correlated with $\mathbf{x}_{jm}$. Also, instruments are required to satisfy the exclusion restriction that at least one variable in $\mathbf{z}_{jm}$ is not contained in $\mathbf{x}_{jm}$.

2.2 Cost function and supply

For each product j in market m, in addition to the data related to demand explained above, researchers observe output q_{jm} (hence, market size as well), an $L \times 1$ vector of input prices $\mathbf{w}_{jm}$ and cost C_{jm}. The observed cost C_{jm} is assumed to be a function of output, input prices, observed product characteristics and a cost shock υ_{jm}. That is,

$$C_{jm} = C\left(q_{jm}, \mathbf{w}_{jm}, \mathbf{x}_{jm}, \upsilon_{jm}; \boldsymbol{\tau}\right),$$

where $\boldsymbol{\tau}$ is a parameter vector. $C()$ is assumed to be strictly increasing and continuously differentiable in output and the cost shock.

Assuming that there is one firm for each product, firm j's profit function is as follows:

$$\pi_{jm} \quad = \quad p_{jm} \times q_{jm} - C\left(q_{jm}, \mathbf{w}_{jm}, \mathbf{x}_{jm}, \upsilon_{jm}; \boldsymbol{\tau}\right).$$

Let MR_{jm} be the marginal revenue of firm j in market m. BLP assume that firms act as differentiated products Bertrand price competitors. Therefore, the optimal price and quantity of product j in market m are determined by the first order condition (F.O.C.) that equates marginal revenue and marginal cost:

$$\underbrace{MR_{jm} = \frac{\partial p_{jm} q_{jm}}{\partial q_{jm}} = p_{jm} + s_{jm}\left[\frac{\partial s_j\left(\mathbf{p}_m, \mathbf{X}_m, \boldsymbol{\xi}_m; \boldsymbol{\theta}\right)}{\partial p_{jm}}\right]^{-1}}_{MR_{jm}} = MC_{jm} = \underbrace{\frac{\partial C\left(q_{jm}, \mathbf{w}_{jm}, \mathbf{x}_{jm}, \upsilon_{jm}; \boldsymbol{\tau}\right)}{\partial q_{jm}}}_{MC_{jm}}. \tag{8}$$

Note that given the market share inversion in Equation (6), and the specification of mean utility $\boldsymbol{\delta}_m$, $\boldsymbol{\xi}_m$ is a function of $(\mathbf{p}_m, \mathbf{s}_m, \mathbf{X}_m)$ and $\boldsymbol{\theta}$. Therefore, MR_{jm} in Equation (8) can be written as a function of observables and parameters as follows:

$$MR_{jm} \equiv MR_j\left(\mathbf{p}_m, \mathbf{s}_m, \mathbf{X}_m; \boldsymbol{\theta}\right), \tag{9}$$

where $MR_j\left(\mathbf{p}_m, \mathbf{s}_m, \mathbf{X}_m; \boldsymbol{\theta}\right)$ is the j^{th} element of the vector of marginal revenue functions in market m, denoted by $\mathbf{MR}\left(\mathbf{p}_m, \mathbf{s}_m, \mathbf{X}_m; \boldsymbol{\theta}\right)$. Equations (8) and (9) imply that demand parameters can potentially be identified if there is data on marginal cost or even without such data, if the cost function is known or can be estimated and its derivative with respect to output can be taken. BLP assume that marginal cost is log-linear in output and observed product characteristics, i.e., $MC_{jm} = exp\left(\mathbf{w}_{jm}\gamma_w + q_{jm}\gamma_q + \upsilon_{jm}\right)$ (see their Equation 3.1). They then use instruments to deal with the endogeneity of output to cost shocks and of prices to demand shocks. As long as the parametric specification of the supply side is accurate and there are enough instruments for identification, the demand side and the F.O.C.-based orthogonality conditions are sufficient for identifying the demand parameters. We, in contrast, assume cost to be a nonparametric function and use the observed cost to control for the cost shock. This point is further explained in Section 3.

2.2.1 Cost function estimation

As with demand estimation, one can recover unobserved cost shocks through inversion:

$$C_{jm} = C\left(q_{jm}, \mathbf{w}_{jm}, \mathbf{x}_{jm}, \upsilon_{jm}; \boldsymbol{\tau}\right) \Rightarrow \upsilon_{jm} = \upsilon\left(q_{jm}, \mathbf{w}_{jm}, \mathbf{x}_{jm}, C_{jm}; \boldsymbol{\tau}\right). \tag{10}$$

Like demand estimation, there are important endogeneity concerns with standard approaches to estimating cost functions. Specifically, output q_{jm} is endogenously determined by profit-maximizing firms as in Equation (8), and is potentially negatively correlated with the cost shock

υ_{jm}. That is, all else equal, less efficient firms tend to produce less. In dealing with this issue, researchers have traditionally focused on selected industries where endogeneity can be ignored, or used instruments for output.

The IV approach to cost function estimation typically uses excluded demand shifters as instruments. Denoting this vector of cost instruments by $\widetilde{\mathbf{z}}_{jm}$, one can estimate $\boldsymbol{\tau}$ assuming that the following population moments are satisfied at the true value of the cost parameters $\boldsymbol{\tau}_0$: $E\left[\upsilon_{jm}\left(q_{jm}, \mathbf{w}_{jm}, \mathbf{x}_{jm}, C_{jm}; \boldsymbol{\tau}_0\right) \widetilde{\mathbf{z}}_{jm}\right] = \mathbf{0}$.

3 Identification using cost data

In this section, we present our methodology for dealing with the endogeneity issues in identification mentioned above. We propose using cost data in addition to demand data to identify price parameters. We do so by using the control function approach. That is, given output, input prices and observed product characteristics, we use the observed cost to control for the cost shock. We focus primarily on the BLP model but also use logit demand in a simple example to illustrate our identification strategy. Instead of Equation (10), we use a nonparametric cost function. We elaborate on our demand and cost structure further below.

To begin with, we assume that market size is observable. In Subsection 3.3, we demonstrate how our methodology can be modified to the case where market size is not observed, and thus needs to be estimated.

3.1 Main assumptions

We first state all the main assumptions for our methodology. Most of these are standard as discussed in the previous section or simply describe the environment our methodology is applicable to. From now on, we let the subscript 0 on parameters or variables indicate that they are at the true values. For each market in the population, we attach a unique positive real number m as an identifier. Then, we assume $m \in \mathcal{M}$, where $\mathcal{M}$ is the set of all market identifiers, and is an uncountable subset of R_+.

Assumption 1 *Data Requirements: Researchers have data on outputs, product prices, market shares, input prices, observed product characteristics, and total costs of firms.*

Note that market size can be derived from data on outputs and market shares. Thus, we need to assume observability of only two of these three variables. In contrast to BLP, we require data on total costs of firms. But we do not need data on marginal cost.

Assumption 2 *Isolated Markets: Outputs, market shares, prices and costs in market m are functions of variables in market m.*

Assumption 3 *Common Input Prices within Markets: Input price $\mathbf{w}_{jm} = \mathbf{w}_m$ for all j, m.*

We make this assumption to show that we do not need within-market variation in input prices. That is, relaxing it makes it easier for our methodology to work. The assumption is reasonable as usually there is little within-market variation in input prices in the data.

Assumption 4 *BLP demand: Market share s_{jm} is specified as in Equation (4). The distributions of α and each element of $\boldsymbol{\beta}$ are assumed to be independently normal, i.e., $\alpha \sim N\left(\mu_\alpha, \sigma_\alpha^2\right)$, $\beta_k \sim N\left(\mu_{\boldsymbol{\beta}k}, \sigma_{\boldsymbol{\beta}k}^2\right)$, $k = 1, \ldots, K$. Further, $\mu_{\boldsymbol{\beta}k} = 0$, $k = 1, \ldots, K$; $\mu_\alpha < 0$.*

Assumption 5 *Equilibrium Concept: Bertrand-Nash equilibrium holds in each market. That is, for any $j = 1, \ldots, J_m$, $m = 1, \ldots, M$, firm j in market m chooses its price p_{jm} to equalize marginal revenue and marginal cost, given market size Q_m and prices of other firms in the same market $\mathbf{p}_{-j,m}$.*[9]

The next assumption describes the support of variables that determine the equilibrium outcomes in market m. Let the set of these variables be denoted by $\mathbf{V}_m$. Then $\mathbf{V}_m \equiv (Q_m, \mathbf{w}_m, \mathbf{X}_m, \boldsymbol{\xi}_m, \boldsymbol{\upsilon}_m)$, and let $\mathbf{V} \equiv \{\mathbf{V}_m\}_{m \in \mathcal{M}}$. Let $\mathbf{V} \setminus w_{lm}$ be the set $\mathbf{V}$ without the element w_{lm} for any $l = 1, 2, \ldots, L$. For other elements of $\mathbf{V}$, the set $\mathbf{V}$ without the element is similarly defined. The assumption below imposes substantially weaker restrictions on the support of the variables in $\mathbf{V}$ than is typical in the literature. In particular, it imposes minimal restrictions on the joint distribution of these variables.

Assumption 6 *Support of $\mathbf{V}$: The support of Q_m conditional on $\mathbf{V} \setminus Q_m$ can be any nonempty subset of R_+ for all m. The support of w_{ljm} conditional on $\mathbf{V} \setminus w_{ljm}$ is R_+ for all l, j, m; the support of x_{kjm} conditional on $\mathbf{V} \setminus x_{kjm}$ is either R or R_+ for all k, j, m; and the support of ξ_{jm} conditional on $\mathbf{V} \setminus \xi_{jm}$ is R. Finally, the support of υ_{jm} conditional on $\mathbf{V} \setminus \upsilon_{jm}$ is R_+.*

[9] Note that we have assumed this for expositional purposes only. It is not required for identification. As long as MR is a one-to-one function of MC in equilibrium, and not necessarily equal to MC, we can still identify the price parameters. This makes our framework applicable to firms that are under government regulation or organizational incentives or behavioral aspects that prevent them from setting $MR = MC$.

Assumption 6 ensures that the variables in $\mathbf{V}$ are not subject to any orthogonality conditions, which typically restrict the moments of a subset of the unobserved variables $(\boldsymbol{\xi}_m, \boldsymbol{\upsilon}_m)$ conditional on the other variables to be zero. In other words, we do not require them to be econometrically exogenous, and thus, Assumption 6 removes the validity of any conventional instruments.

Note that we do not impose any assumptions on the support of market size other than that it is nonempty and positive. For logit, we require the conditional support to be R_+ since as we show later, market size variation is needed for the identification of the price parameters of logit but not for BLP.

The next two assumptions are about our nonparametric cost function.[10]

Assumption 7 *Properties of the Cost Function: Let C^*_{jm} denote true cost. Then,*

$$C^*_{jm} \equiv C^v\left(q_{jm}, \mathbf{w}_m, \mathbf{x}_{jm}, \upsilon_{jm}\right) + e_f\left(\mathbf{w}_m, \mathbf{x}_{jm}, \upsilon_{jm}\right) + \varsigma_{jm}, \tag{11}$$

where $C^v()$ is the variable cost component, which is a continuous function of q, $\mathbf{w}$, $\mathbf{x}$ and υ, strictly increasing, and continuously differentiable in q and υ, and marginal cost is strictly increasing in υ; $e_f()$ is the deterministic component of fixed cost, a continuous function of $\mathbf{w}$, $\mathbf{x}$ and υ and increasing in υ. The fixed cost shock ς is i.i.d., with mean zero and independent of $\mathbf{V}_m$. Further, for any $q > 0$, $w_l > 0$, $l = 1, \ldots, L$ and $\mathbf{x} \in \mathcal{X}$, where $\mathcal{X}$ is the support of $\mathbf{x}$,

$$lim_{\upsilon \searrow 0} \frac{\partial C^v\left(q, \mathbf{w}, \mathbf{x}, \upsilon\right)}{\partial q} = 0, \quad lim_{\upsilon \nearrow \infty} \frac{\partial C^v\left(q, \mathbf{w}, \mathbf{x}, \upsilon\right)}{\partial q} = \infty.$$

Assumption 8 *Measurement Error in Cost: Let C_{jm} be the observed cost. Then,*

$$C_{jm} = C^*_{jm} + e_{me}\left(q_{jm}, \mathbf{w}_m, \mathbf{x}_{jm}\right) + \nu_{jm}, \tag{12}$$

where $e_{me}()$ is a continuous function and ν_{jm} is i.i.d. with mean 0 and independent of $\mathbf{V}_m$ and the fixed cost shock $\boldsymbol{\varsigma}_m$.[11]

The assumption implies that the measurement error in cost is $e_{me}\left(q_{jm}, \mathbf{w}_m, \mathbf{x}_{jm}\right) + \nu_{jm}$, where $e_{me}()$ is the deterministic component.[12]

[10]Note that we assume that market share s_{jm} does not enter in the cost function. This restriction rules out situations in which firms with high market shares have buying power in the input market.

[11]We can also include $\mathbf{x}_{me,jm}$, a vector of additional variables that determine the deterministic component of the measurement error as well as fixed cost. However, we omit these for the sake of expositional simplicity.

[12]We can include systematic misreporting of true costs. For example, if $\nu\left(C^*\right)$ is the systematic component of the reported true cost, then, if firms report costs truthfully but with an error, then $\nu\left(C^*\right) = C^*$. Alternatively, if firms systematically under-report their true costs, then we could consider a specification like $\nu\left(C^*\right) = \nu C^*$ where

Using Equations (11) and (12), we obtain

$$C_{jm} = C^{v}\left(q_{jm}, \mathbf{w}_{m}, \mathbf{x}_{jm}, \upsilon_{jm}\right) + e_{f}\left(\mathbf{w}_{jm}, \mathbf{x}_{jm}, \upsilon\right) + e_{me}\left(q_{jm}, \mathbf{w}_{jm}, \mathbf{x}\right) + \nu_{jm} + \varsigma_{jm}.$$

From now on, we call $C^{v}\left(q_{jm}, \mathbf{w}_{m}, \mathbf{x}_{jm}, \upsilon_{jm}\right) + e_{f}\left(\mathbf{w}_{jm}, \mathbf{x}_{jm}, \upsilon_{jm}\right) + e_{me}\left(q_{jm}, \mathbf{w}_{jm}, \mathbf{x}\right)$, i.e., the cost data minus the random components of the fixed cost and the measurement error to be the deterministic component of cost or the cost function.

In the next subsection, we show that cost data together with other data and our assumptions enables us to identify the parameters of the distribution of the random coefficients on price $(\mu_{\alpha}, \sigma_{\alpha})$, as well as σ_{β_k}, $k = 1, \ldots, K$. Note that we can only identify $(\mu_{\alpha 0}, \sigma_{\alpha 0}, \boldsymbol{\sigma}_{\boldsymbol{\beta} 0})$ and $\xi_{0jm} + \mathbf{x}_{jm}\boldsymbol{\mu}_{\beta 0}$ in the absence of further restrictions imposed on the model. However, this information is sufficient to identify marginal revenue, and thus, markup, which is of primary interest in most empirical exercises in IO. The additional orthogonality assumption, $E\left(\mathbf{x}_{jm}\xi_{jm}\right) = 0$, identifies $\boldsymbol{\mu}_{\beta 0}$. For the logit model of demand, we identify the price coefficient α_0 and $\xi_{0jm} + \mathbf{x}_{jm}\boldsymbol{\beta}_0$.

From now on, except when noted otherwise, we will denote the vector of true demand parameters we identify to be $\boldsymbol{\theta}_0$. In particular, $\theta_0 = \alpha_0$ for the logit demand specification and $\boldsymbol{\theta}_0 = \left(\mu_{\alpha 0}, \sigma_{\alpha 0}, \boldsymbol{\sigma}'_{\boldsymbol{\beta} 0}\right)$ for the BLP specification.

3.2 The main result

In this subsection, we derive our main theoretical result, namely, that given our assumptions, the parameters $\boldsymbol{\theta}_{c0} = \left(\mu_{\alpha 0}, \sigma_{\alpha 0}, \boldsymbol{\sigma}'_{\boldsymbol{\beta} 0}\right)$ of the BLP model are identified.

Since we allow for measurement error in the cost data and a random component in the fixed cost, the first step in proving our results is to show that the expected cost conditional on observables contains only the deterministic components of the cost function, which is done in Lemma 1. We then provide an intuitive explanation of our identification strategy using the logit demand example and then develop two equivalent definitions of identification using the cost data. The first definition (Definition 1) is based on the first order condition (Equation (8)) and highlights the sources of variation for identification as well as the exclusion restrictions. However, since the control function approach leads to marginal cost becoming an unspecified function of output, input prices, observed product characteristics and the deterministic component of cost, it is difficult to use in proving identification and thus, we develop a second definition of identification (Definition 2) based on pairing of firms that have the same output, input prices and observed

$0 < \nu < 1$. Over-reporting could be captured by the same specification with $\nu > 1$. We omit these for the sake of expositional simplicity.

characteristics. We show that the two definitions are equivalent and then specify a condition on marginal revenue that together with our assumptions ensures identification in the BLP model.

Lemma 1 *Let Assumptions 1-3, 5 and 7-8 be satisfied, and let the marginal revenue function be specified as in Equation (9). Further, let $\mathcal{R} = \{\mathbf{p}, \mathbf{s}, \mathbf{X}, q, \mathbf{w}, Q, j\}$, and let $\widetilde{C}$ be the deterministic component of cost. Then, for any firm j in the population*

$$E\left[C \middle| \left(\widetilde{q} = q, \widetilde{\mathbf{w}} = \mathbf{w}, \widetilde{\mathbf{p}} = \mathbf{p}, \widetilde{\mathbf{s}} = \mathbf{s}, \widetilde{\mathbf{X}} = \mathbf{X}, j\right)\right] = C^v\left(q, \mathbf{w}, \mathbf{x}, \upsilon\right) + e_f\left(\mathbf{w}, \mathbf{x}, \upsilon\right) + e_{me}\left(q, \mathbf{w}, \mathbf{x}\right) \equiv \widetilde{C}.$$

Proof. From Assumption 7, marginal cost is strictly increasing in υ, and given any $(q, \mathbf{w}, \mathbf{x})$ in the population, the support of marginal cost is R_+. Therefore, given Assumption 5, for any observation $\mathcal{R} = \{\mathbf{p}, \mathbf{s}, \mathbf{X}, q, \mathbf{w}, Q, j\}$ in the population, there exists a unique υ such that

$$MR_j\left(\mathbf{p}, \mathbf{s}, \mathbf{X}; \boldsymbol{\theta}_{c0}\right) = MC\left(q, \mathbf{w}, \mathbf{x}, \upsilon\right). \tag{13}$$

Because Equation (13) determines a unique υ given $(q, \mathbf{w}, \mathbf{x})$, υ is a function of $(q, \mathbf{w}, \mathbf{x}, MR_j\left(\mathbf{p}, \mathbf{s}, \mathbf{X}; \boldsymbol{\theta}_{c0}\right))$. Thus, Assumptions 7 and 8 result in

$$\begin{aligned}
&E\left[C \middle| \left(\widetilde{q} = q, \widetilde{\mathbf{w}} = \mathbf{w}, \widetilde{\mathbf{p}} = \mathbf{p}, \widetilde{\mathbf{s}} = \mathbf{s}, \widetilde{\mathbf{X}} = \mathbf{X}, j\right)\right] \\
&= E\Big[C^v\left(q, \mathbf{w}, \mathbf{x}, \upsilon\left(q, \mathbf{w}, \mathbf{x}, MR_j\left(\mathbf{p}, \mathbf{s}, \mathbf{X}; \boldsymbol{\theta}_{c0}\right)\right)\right) + e_f\left(\mathbf{w}, \mathbf{x}, \upsilon\left(q, \mathbf{w}, \mathbf{x}, MR_j\left(\mathbf{p}, \mathbf{s}, \mathbf{X}; \boldsymbol{\theta}_{c0}\right)\right)\right) \\
&+ e_{me}\left(q, \mathbf{w}, \mathbf{x}\right) + \varsigma + \nu \Big| \left(\widetilde{q} = q, \widetilde{\mathbf{w}} = \mathbf{w}, \widetilde{\mathbf{p}} = \mathbf{p}, \widetilde{\mathbf{s}} = \mathbf{s}, \widetilde{\mathbf{X}} = \mathbf{X}, j\right)\Big] \\
&= C^v\left(q, \mathbf{w}, \mathbf{x}, \upsilon\right) + e_f\left(\mathbf{w}, \mathbf{x}, \upsilon\right) + e_{me}\left(q, \mathbf{w}, \mathbf{x}\right)
\end{aligned}$$

■

In this subsection, we call a variable observable if it is directly observable in the population or can be recovered as the expectation of a directly observable variable conditional on other directly observable variables. Thus, $\widetilde{C}$ is observed because it is the conditional expectation of observed cost conditional on other observed data. From now on, we sometimes refer to the deterministic component of cost as cost for convenience.

Before providing formal results, we outline the logic of our identification argument. In particular, we first explain how we remove the need for instruments to deal with the endogeneity of the supply shock. We use the following three equations, for firm j in market m, for identification:

$$s_{jm} = q_{jm}/Q_m, \tag{14}$$

$$MR_j\left(\mathbf{p}_m, \mathbf{s}_m, \mathbf{X}_m; \boldsymbol{\theta}_{c0}\right) = MC\left(q_{jm}, \mathbf{w}_m, \mathbf{x}_{jm}, \upsilon_{jm}\right), \tag{15}$$

$$\widetilde{C}_{jm} = C^v\left(q_{jm}, \mathbf{w}_m, \mathbf{x}_{jm}, \upsilon_{jm}\right) + e_f\left(\mathbf{w}_m, \mathbf{x}_{jm}, \upsilon_{jm}\right) + e_{me}\left(q_{jm}, \mathbf{w}_m, \mathbf{x}_{jm}\right). \tag{16}$$

If we had data on marginal cost, denoted by MC_{jm}, we could just use Equation (15) for identification of $\boldsymbol{\theta}_{c0}$. Then, by substituting MC_{jm} into the RHS, we would have a function of only observables. For example, in the logit model, Equation (15) can then be written as:

$$p_{jm} + \frac{1}{\left(1 - s_{jm}\right)\alpha_0} = MC_{jm},$$

and the price coefficient α_0 can be identified as

$$\alpha_0 = \frac{1}{\left(1 - s_{jm}\right)\left(MC_{jm} - p_{jm}\right)}.$$

However, marginal cost is generally not observable. One could then consider estimating the cost function $\widetilde{C}_{jm} \equiv C\left(q_{jm}, \mathbf{w}_m, \mathbf{x}_{jm}, \upsilon_{jm}\right)$, and taking its derivative with respect to output to derive the marginal cost. However, as discussed in the previous section, this strategy runs into the potential endogeneity issue of output being correlated with the cost shock and thus requires the use of instruments which we propose to avoid with our methodology.[13]

Instead, we use cost data to control for the cost shock. That is, we invert the cost function in Equation (16) to derive

$$\upsilon_{jm} = \upsilon\left(q_{jm}, \mathbf{w}_m, \mathbf{x}_{jm}, \widetilde{C}_{jm}\right). \tag{17}$$

Such inversion is possible because Assumption 7 implies that given output, input price and observed characteristics, the deterministic component of cost is a strictly increasing function of the cost shock. After substituting Equation (17) into Equation (15), the F.O.C. becomes:

$$MR_j\left(\mathbf{p}_m, \mathbf{s}_m, \mathbf{X}_m; \boldsymbol{\theta}_{c0}\right) = \psi\left(q_{jm}, \mathbf{w}_m, \mathbf{x}_{jm}, \widetilde{C}_{jm}\right), \tag{18}$$

where ψ is an unspecified function. Note that there are no unobservables in Equation (18) that may be correlated with the observables and create endogeneity problems.

We now define identification based on this F.O.C, letting firm jm denote firm j in market m.

Definition 1 *Identification by the F.O.C.: Let the marginal revenue function be specified as in*

[13] Such a derivative would also include the derivative of the deterministic component of the measurement error with respect to output, which should not be part of marginal cost.

Equation (9). Then, the true parameter vector $\boldsymbol{\theta}_{c0}$ is identified if the following two statements hold only at $\boldsymbol{\theta}_{c0}$:

1. *For any firm jm in the population, marginal revenue is positive and can be expressed as a function of only $\left(q_{jm}, \mathbf{w}_m, \mathbf{x}_{jm}, \widetilde{C}_{jm}\right)$.*

2. *Given $(q, \mathbf{w}, \mathbf{x})$, this function is one-to-one in $\widetilde{C}$.*

The function of $\left(q_{jm}, \mathbf{w}_m, \mathbf{x}_{jm}, \widetilde{C}_{jm}\right)$ in the above definition corresponds to $\psi()$ in Equation (18). From Assumption 7, we know that given $(q, \mathbf{w}, \mathbf{x})$, $\widetilde{C}$ and marginal cost are strictly increasing in the cost shock υ, which implies that, given $(q, \mathbf{w}, \mathbf{x})$, ψ is strictly increasing and thus, one-to-one in $\widetilde{C}$. Our definition implies that only at the true parameter vector $\boldsymbol{\theta}_{c0}$, given $(q, \mathbf{w}, \mathbf{x})$, each value of $\widetilde{C}_{jm}$ is associated with a unique marginal revenue. Intuitively this means that for firms with the same output, input prices and observed characteristics, having equal observed cost is equivalent to having equal marginal revenue only at the true parameter vector.

The sources of variation we use for identification of $\boldsymbol{\theta}_{c0}$ are similar to the ones in the literature. In the logit demand model, as we explain below, these are market size Q_m and price p_{jm}, and in BLP, additionally, we can use price, market share and observed characteristics of the rival firms. All these sources of variation reflect the "market structure" and appear in the marginal revenue function but not in the marginal cost function ψ (see Equation (18)). The difference from the literature is that these variables do not need to be instruments, i.e. orthogonal to the cost shock υ_{jm}, because we have already controlled for it using $\widetilde{C}_{jm}$. Thus, our identification strategy is based on the exclusion restriction that market structure variables do not enter the cost function directly.

To illustrate how Equation (18) identifies $\boldsymbol{\theta}_{c0}$, we use the logit demand model, where the parameter we identify is the price coefficient, i.e, $\boldsymbol{\theta}_{c0} = \alpha_0$. Then, using Equation (14), Equation (18) can be written as:

$$MR_j\left(\mathbf{p}_m, \mathbf{s}_m, \mathbf{X}_m; \alpha_0\right) = p_{jm} + \frac{1}{\left(1 - q_{jm}/Q_m\right)\alpha_0} = \psi\left(q_{jm}, \mathbf{w}_m, \mathbf{x}_{jm}, \widetilde{C}_{jm}\right). \tag{19}$$

Given Assumption 7, Conditions 1 and 2 of Definition 1 are clearly satisfied for α_0. Next, we consider any $\alpha \neq \alpha_0$. Then,

$$
\begin{aligned}
& MR_j\left(\mathbf{p}_m, \mathbf{s}_m, \mathbf{X}_m; \alpha_0\right) \\
= \; & p_{jm} + \frac{1}{\left(1 - q_{jm}/Q_m\right)\alpha_0} = p_{jm} + \frac{1}{\left(1 - q_{jm}/Q_m\right)\alpha} + \frac{1}{\left(1 - q_{jm}/Q_m\right)}\left(\frac{1}{\alpha_0} - \frac{1}{\alpha}\right) \\
= \; & MR_j\left(\mathbf{p}_m, \mathbf{s}_m, \mathbf{X}_m; \alpha\right) + \frac{1}{\left(1 - q_{jm}/Q_m\right)}\left(\frac{1}{\alpha_0} - \frac{1}{\alpha}\right).
\end{aligned}
$$

Substituting into (19), we obtain

$$
\begin{aligned}
& MR_j\left(\mathbf{p}_m, \mathbf{s}_m, \mathbf{X}_m; \alpha\right) \\
= \; & p_{jm} + \frac{1}{\left(1 - q_{jm}/Q_m\right)\alpha} = \psi\left(q_{jm}, \mathbf{w}_m, \mathbf{x}_{jm}, \widetilde{C}_{jm}\right) - \frac{1}{\left(1 - q_{jm}/Q_m\right)}\left(\frac{1}{\alpha_0} - \frac{1}{\alpha}\right) \qquad (20) \\
\equiv \; & \widetilde{\psi}(q_{jm}, \mathbf{w}_m, \mathbf{x}_{jm}, \widetilde{C}_{jm}, Q_m, \alpha).
\end{aligned}
$$

Note that $\widetilde{\psi}$ includes market size as an argument, violating Condition 1 of Definition 1 for $\alpha \neq \alpha_0$. Thus, α_0 is identified. Note also that if we do not have any variation in market size, i.e., $Q_m = \overline{Q}$, then $\widetilde{\psi}$ remains a function of $\left(q, \mathbf{w}, \mathbf{x}, \widetilde{C}\right)$. Furthermore, if $Q_m = \overline{Q}$, in Equation (20), $\widetilde{C}$ enters in ψ and we know that given $(q, \mathbf{w}, \mathbf{x})$, ψ is a one-to-one function of $\widetilde{C}$, and thus, $\widetilde{\psi}$ satisfies both Conditions 1 and 2 for $\alpha \neq \alpha_0$. Hence, the true price coefficient cannot be identified. Thus, for the logit model, our identification strategy requires variation in market size. Price variation is also needed unless $1/\alpha_0$ is zero, otherwise Equation (19) would fail to hold. This becomes transparent later in this subsection.

However, dealing with unspecified functions ψ and $\widetilde{\psi}$ makes the identification analysis complex and unintuitive. This is because for each parameter $\boldsymbol{\theta}_c$, we need to evaluate whether marginal revenue at $\boldsymbol{\theta}_c$ is a function of only $\left(q, \mathbf{w}, \mathbf{x}, \widetilde{C}\right)$. Instead, in our analysis, we use an alternative equivalent way of proving identification, which we call the pairing approach. This approach lets us focus on the parametric marginal revenue side. The only role of cost data and the marginal cost function is to identify the following two sets of pairs of firms in different markets: those that have the same true marginal revenue, and those that have different true marginal revenues. Then, from these two sets of pairs, we proceed to identify the price coefficient by using only the demand side. We illustrate this approach for the logit model first. As we will see, the pairing approach provides us with the exact sources of variation needed to identify the price parameter in the logit model, namely, market size and price of the firm's product.

More specifically, we "fix" the variables in the marginal cost function by finding a pair of firms

$(jm, j^\dagger m^\dagger)$ in the data that have the same output, same input price, same observed characteristics and the same cost: $\left(q_{jm}, \mathbf{w}_m, \mathbf{x}_{jm}, \widetilde{C}_{jm}\right) = \left(q_{j^\dagger m^\dagger}, \mathbf{w}_{m^\dagger}, \mathbf{x}_{j^\dagger m^\dagger}, \widetilde{C}_{j^\dagger m^\dagger}\right)$ but $(p_{jm}, Q_m) \neq (p_{j^\dagger m^\dagger}, Q_{m^\dagger})$. Then, from Definition 1,

$$\psi\left(q_{jm}, \mathbf{w}_m, \mathbf{x}_{jm}, \widetilde{C}_{jm}\right) = \psi\left(q_{j^\dagger m^\dagger}, \mathbf{w}_{m^\dagger}, \mathbf{x}_{j^\dagger m^\dagger}, \widetilde{C}_{j^\dagger m^\dagger}\right),$$

and thus, using Equation (19), at α_0,

$$\begin{aligned} & p_{jm} + \frac{1}{\left(1 - q_{jm}/Q_m\right)\alpha_0} = \psi\left(q_{jm}, \mathbf{w}_m, \mathbf{x}_{jm}, \widetilde{C}_{jm}\right) \\ = \; & \psi\left(q_{j^\dagger m^\dagger}, \mathbf{w}_{m^\dagger}, \mathbf{x}_{j^\dagger m^\dagger}, \widetilde{C}_{j^\dagger m^\dagger}\right) = p_{j^\dagger m^\dagger} + \frac{1}{\left(1 - q_{j^\dagger m^\dagger}/Q_{m^\dagger}\right)\alpha_0}. \end{aligned} \tag{21}$$

In the above equation, ψ is eliminated and what remains is the equality of the true marginal revenue of the two firms that have different prices and market shares.

Then, α_0 can be identified straightforwardly from the above marginal revenue equality as follows:

$$\alpha_0 = -\frac{1}{p_{jm} - p_{j^\dagger m^\dagger}}\left[\frac{1}{\left(1 - q_{jm}/Q_m\right)} - \frac{1}{\left(1 - q_{j^\dagger m^\dagger}/Q_{m^\dagger}\right)}\right]. \tag{22}$$

Note that since $q_{jm} = q_{j^\dagger m^\dagger}$, if we assume constant market size, the term in the bracket is always zero, and thus, $\alpha_0 \neq 0$ cannot be identified. Furthermore, without variation in price, RHS is either not bounded or not well-defined. Therefore, identification using pairing requires variation in both price and market size. We show next that given $q_{jm} = q_{j^\dagger m^\dagger}$, $\mathbf{w}_m = \mathbf{w}_{m^\dagger}$ and $\mathbf{x}_{jm} = \mathbf{x}_{j^\dagger m^\dagger}$, there exist (Q_m, ξ_{jm}) and $\left(Q_{m^\dagger}, \xi_{j^\dagger m^\dagger}\right)$ that generate the above prices $\left(p_{jm}, p_{j^\dagger m^\dagger}\right)$ and market shares $(s_{jm},\ s_{j^\dagger m^\dagger})$. First, we choose $Q_m = q_m/s_{jm}$, $Q_{m^\dagger} = q_{m^\dagger}/s_{j^\dagger m^\dagger}$. This is feasible because from Assumption 6, the conditional support of market size Q is R^+. Also, using Equation (1), and the normalization in Equation (3), we can choose ξ_{jm}, $\xi_{j^\dagger m^\dagger}$ from the conditional support of R as:

$$\begin{aligned} \xi_{jm} &= -\mathbf{x}_{jm}\boldsymbol{\beta}_0 - p_{jm}\alpha_0 + ln\left(s_{jm}\right) - ln\left(s_{0m}\right), \\ \xi_{j^\dagger m^\dagger} &= -\mathbf{x}_{j^\dagger m^\dagger}\boldsymbol{\beta}_0 - p_{j^\dagger m^\dagger}\alpha_0 + ln\left(s_{j^\dagger m^\dagger}\right) - ln\left(s_{0m^\dagger}\right). \end{aligned}$$

It is clear from the discussion above that in logit, identification of the price coefficient is based on the true marginal revenue equality for a pair of firms that has the same output, input prices, observed product characteristics and cost. Indeed, Equations (21) and (22) indicate that we only need to consider a single pair of firms to prove identification. In BLP however, marginal revenue is a complex, nonlinear function of parameters and thus, we cannot show analytically that the demand parameters $\boldsymbol{\theta}_{c0}$ are identified or that the marginal revenue equality for one pair of firms generates a unique set of true parameters. Instead, we exploit the information contained in the data about the set of pairs whose two firms have the same output, input prices and observed characteristics. We divide these firm-pairs into two subsets of pairs according to whether the two firms within a pair have equal observed cost or not. The group of pairs whose firms have the same observed cost must have the same true marginal revenue while the opposite is true for the other group. We use this insight to formulate a condition on the marginal revenue function that is sufficient for identifying $\boldsymbol{\theta}_{c0}$.

We now reformulate our identification definition in terms of pairing.

Definition 2 *Identification by Pairing: Let the marginal revenue function be specified as in Equation (9). We say that* $\boldsymbol{\theta}_{c0}$ *is identified if the following holds only for* $\boldsymbol{\theta}_c = \boldsymbol{\theta}_{c0}$*:*

1 *For any firm* jm *in the population,*

$$MR_j\left(\mathbf{p}_m, \mathbf{s}_m, \mathbf{X}_m, \boldsymbol{\theta}_c\right) > 0. \tag{23}$$

2 *Given any two firms* $jm \neq j^\dagger m^\dagger$ *in the population with* $\left(q_{jm}, \mathbf{w}_m, \mathbf{x}_{jm}\right) = \left(q_{j^\dagger m^\dagger}, \mathbf{w}_{m^\dagger}, \mathbf{x}_{m^\dagger}\right)$*,*

$$MR_j\left(\mathbf{p}_m, \mathbf{s}_m, \mathbf{X}_m, \boldsymbol{\theta}_c\right) = MR_{j^\dagger}\left(\mathbf{p}_{m^\dagger}, \mathbf{s}_{m^\dagger}, \mathbf{X}_{m^\dagger}, \boldsymbol{\theta}_c\right),$$

if and only if

$$\widetilde{C}_{jm} = \widetilde{C}_{j^\dagger m^\dagger}.$$

It is straightforward to see that the two definitions are equivalent. Both require positivity of true marginal revenue and a one-to-one relationship between the observable cost and the true marginal revenue given output, input prices and observed characteristics, and violation of at least one of these conditions if the parameter vector is not the true one. The pairwise approach to identification is simpler than the first definition because we no longer have to deal with the unspecified marginal cost function. Instead, we identify the true parameter vector by examining all the pairs in which the two firms have the same output, input prices and observed characteristics. We check if marginal revenue is strictly positive and whether the within-pair equality between the two observed costs and between the two marginal revenues holds simultaneously at a candidate parameter vector. If any marginal revenue is nonpositive or if simultaneity does not hold for some of these pairs, i.e., if for some pairs, only the costs are equal but not the marginal revenues, or vice versa, then the candidate parameter vector cannot be the true one. We can also state it more formally by defining the set $\mathcal{S}$ to be the set of pairs whose two firms have the same output, input prices and observed characteristics, and letting $\mathcal{C} \subset \mathcal{S}$ be the subset of $\mathcal{S}$ whose two firms have the same cost, and $\mathcal{MR}(\boldsymbol{\theta}_c) \subset \mathcal{S}$ be the subset of $\mathcal{S}$ whose two firms have the same marginal revenue. Then, Definition 2 states that $\boldsymbol{\theta}_{c0}$ is identified if for any $\boldsymbol{\theta}_c \neq \boldsymbol{\theta}_{c0}$, either positivity in Equation (23) is violated for some firm jm in the population, or

$$\mathcal{C} = \mathcal{MR}(\boldsymbol{\theta}_{c0}) \neq \mathcal{MR}(\boldsymbol{\theta}_c),$$

or both.

Lemma 2 *Definitions 1 and 2 are equivalent.*

The proof is in the appendix.

We next state a condition on the demand model that together with our assumptions is sufficient for identification of the demand parameters. We need this condition because the information we can use from the data on cost and the assumptions on the cost function are not sufficient to identify the true marginal revenue. Among the pairs of firms in $\mathcal{S}$, the cost data allows us to identify the subset of pairs (subset 1), whose two firms have the same true marginal revenue (i.e. have the same cost), and the subset of pairs, (subset 2), whose two firms have different true marginal revenues (i.e. different costs). The additional source of information that identifies the true marginal revenue and $\boldsymbol{\theta}_{c0}$ needs to come from the functional form of the demand model, such as logit or BLP.

Condition 1 *Let the marginal revenue function be specified as in Equation (9). Let* $\mathcal{D} = \{\mathbf{p}, \mathbf{s}, \mathbf{X}\}$ *and* $\mathcal{D}^\dagger = \{\mathbf{p}^\dagger, \mathbf{s}^\dagger, \mathbf{X}^\dagger\}$ *be two sets of vectors of prices, market shares and observed product characteristics, and let* $\boldsymbol{\theta}_{c0}$ *be the true parameter vector. Then, for any given* $\boldsymbol{\theta}_c \neq \boldsymbol{\theta}_{c0}$*, either positivity in Equation (23) is violated for a firm* jm *in the population; or the following statement holds, or both: There exist* $\mathcal{D}$ *and* $\mathcal{D}^\dagger$ *that satisfy the following properties: 1) for any reordering of the rows in set* $\mathcal{D}^\dagger$*,* $\mathcal{D} \neq \mathcal{D}^\dagger$*, and 2) there exists a row* j *in* $\mathcal{D}$ *and a row* $j^\dagger$ *in* $\mathcal{D}^\dagger$*, such that*

2a. $p_l > 0$, $0 < s_l < 1$ *for* $l = 1, ..., J$ *and* $p_l^\dagger > 0$, $0 < s_l^\dagger < 1$, *for* $l = 1, .., J^\dagger$, *and* $0 < \sum_{l=1}^{J} s_l < 1$, $0 < \sum_{l=1}^{J^\dagger} s_l^\dagger < 1$.

2b. $\mathbf{x}_j = \mathbf{x}^{\dagger}_{j^{\dagger}}$.

2c. *Either* $MR_j(\mathbf{p}, \mathbf{s}, \mathbf{X}, \boldsymbol{\theta}_c) = MR_{j^{\dagger}}\left(\mathbf{p}^{\dagger}, \mathbf{s}^{\dagger}, \mathbf{X}^{\dagger}, \boldsymbol{\theta}_c\right)$ *or* $MR_j(\mathbf{p}, \mathbf{s}, \mathbf{X}, \boldsymbol{\theta}_{c0}) = MR_{j^{\dagger}}\left(\mathbf{p}^{\dagger}, \mathbf{s}^{\dagger}, \mathbf{X}^{\dagger}, \boldsymbol{\theta}_{c0}\right)$ *but not both.*

Condition 1 is the restatement of Definition 2 of identification without any restrictions on the cost side. That is, Condition 1 applied to the population is equivalent to requiring that for any $\boldsymbol{\theta}_c \neq \boldsymbol{\theta}_{c0}$, either for some firm in the population, marginal revenue is nonpositive at $\boldsymbol{\theta}_c$ or within $\mathcal{S}$, the set of pairs whose two firms have the same positive marginal revenue under $\boldsymbol{\theta}_{c0}$ and the corresponding set of pairs under $\boldsymbol{\theta}_c$ cannot be equal (that is, $\mathcal{MR}(\boldsymbol{\theta}_{c0}) \neq \mathcal{MR}(\boldsymbol{\theta}_c)$), or both. Thus, the demand specification is such that only the true parameter can exactly replicate subsets 1 and 2 identified by the data. In the lemma below, we show that given our assumptions, Condition 1 is sufficient for identification because for any pairs of observed demand variables $(\mathbf{p}, \mathbf{s}, \mathbf{X}, j)$ and $\left(\mathbf{p}^{\dagger}, \mathbf{s}^{\dagger}, \mathbf{X}^{\dagger}, j^{\dagger}\right)$ satisfying $\mathbf{x}_j = \mathbf{x}_{j^{\dagger}}$, we can always find two firms in the population that have the same output, input prices and observed characteristics and these demand variables.

Lemma 3 *Suppose Assumptions 1-3, 5-8 and Condition 1 are satisfied. Then,* $\boldsymbol{\theta}_{c0}$ *is identified according to Definition 2 of identification.*

Proof. First, consider $\boldsymbol{\theta}_{c0}$. Then, given Assumptions 5 and 7, in the population, marginal revenue is positive at $\boldsymbol{\theta}_{c0}$. Also, given Assumptions 2-3 and 5-8, Equation (18) holds, and from Equation (18), for any $\left(\mathcal{D}, q, \mathbf{w}, \mathbf{x}, \widetilde{C}, j\right)$ and $\left(\mathcal{D}^{\dagger}, q, \mathbf{w}, \mathbf{x}, \widetilde{C}^{\dagger}, j^{\dagger}\right)$ in the population with $\widetilde{C} = \widetilde{C}^{\dagger}$,

$$MR_j(\mathbf{p}, \mathbf{s}, \mathbf{X}, \boldsymbol{\theta}_{c0}) = \psi\left(q, \mathbf{w}, \mathbf{x}, \widetilde{C}\right) = \psi\left(q, \mathbf{w}, \mathbf{x}, \widetilde{C}^{\dagger}\right) = MR_{j^{\dagger}}\left(\mathbf{p}^{\dagger}, \mathbf{s}^{\dagger}, \mathbf{X}^{\dagger}, \boldsymbol{\theta}_{c0}\right).$$

Similarly, if $MR_j(\mathbf{p}, \mathbf{s}, \mathbf{X}, \boldsymbol{\theta}_{c0}) = MR_{j^{\dagger}}\left(\mathbf{p}^{\dagger}, \mathbf{s}^{\dagger}, \mathbf{X}^{\dagger}, \boldsymbol{\theta}_{c0}\right)$, then, given $(q, \mathbf{w}, \mathbf{x})$, from Assumptions 5-7, there exists a unique cost shock υ such that

$$MR_j(\mathbf{p}, \mathbf{s}, \mathbf{X}, \boldsymbol{\theta}_{c0}) = MR_{j^{\dagger}}\left(\mathbf{p}^{\dagger}, \mathbf{s}^{\dagger}, \mathbf{X}^{\dagger}, \boldsymbol{\theta}_{c0}\right) = MC(q, \mathbf{w}, \mathbf{x}, \upsilon).$$

Therefore, $\widetilde{C} = \widetilde{C}^{\dagger}$. Hence, for any pair of firms with the same $(q, \mathbf{w}, \mathbf{x})$, $\widetilde{C} = \widetilde{C}^{\dagger}$ if and only if $MR_j(\mathbf{p}, \mathbf{s}, \mathbf{X}, \boldsymbol{\theta}_{c0}) = MR_{j^{\dagger}}\left(\mathbf{p}^{\dagger}, \mathbf{s}^{\dagger}, \mathbf{X}^{\dagger}, \boldsymbol{\theta}_{c0}\right)$. Therefore, $\boldsymbol{\theta}_{c0}$ satisfies the conditions for identification of Definition 2.

We now consider any $\boldsymbol{\theta}_c \neq \boldsymbol{\theta}_{c0}$. Suppose there exists $\left(\mathcal{D}, q, \mathbf{w}, \mathbf{x}, \widetilde{C}, j\right)$ in the population satisfying $MR_j(\mathbf{p}, \mathbf{s}, \mathbf{X}, \boldsymbol{\theta}_c) \leq 0$. Then, $\boldsymbol{\theta}_c$ violates the first condition of Definition 2. Next, we consider the case where any $\left(\mathcal{D}, q, \mathbf{w}, \mathbf{x}, \widetilde{C}, j\right)$ in the population satisfies $MR_j(\mathbf{p}, \mathbf{s}, \mathbf{X}, \boldsymbol{\theta}_c) > 0$.

We analyze the two cases of Condition 1 separately.

Case 1: Suppose $(\mathcal{D}, j)$ and $\left(\mathcal{D}^\dagger, j^\dagger\right)$ satisfy $MR_j(\mathbf{p}, \mathbf{s}, \mathbf{X}, \boldsymbol{\theta}_c) = MR_{j^\dagger}\left(\mathbf{p}^\dagger, \mathbf{s}^\dagger, \mathbf{X}^\dagger, \boldsymbol{\theta}_c\right)$ but $MR_j(\mathbf{p}, \mathbf{s}, \mathbf{X}, \boldsymbol{\theta}_{c0}) \neq MR_{j^\dagger}\left(\mathbf{p}^\dagger, \mathbf{s}^\dagger, \mathbf{X}^\dagger, \boldsymbol{\theta}_{c0}\right)$. Then, for any $(q, \mathbf{w}, \mathbf{x})$, from Assumptions 6 and 7, there exist $\left(\upsilon, \upsilon^\dagger\right)$ in the population such that $MR_j(\mathbf{p}, \mathbf{s}, \mathbf{X}, \boldsymbol{\theta}_{c0}) = MC(q, \mathbf{w}, \mathbf{x}, \upsilon)$ and $MR_{j^\dagger}\left(\mathbf{p}^\dagger, \mathbf{s}^\dagger, \mathbf{X}^\dagger, \boldsymbol{\theta}_{c0}\right) = MC\left(q, \mathbf{w}, \mathbf{x}, \upsilon^\dagger\right)$. Because marginal cost is strictly increasing in the cost shock, this implies that $\upsilon \neq \upsilon^\dagger$, and since the deterministic component of cost is increasing in the cost shock, $\widetilde{C} \neq \widetilde{C}^\dagger$. Therefore, in this case, $MR_j(\mathbf{p}, \mathbf{s}, \mathbf{X}, \boldsymbol{\theta}_c) = MR_{j^\dagger}\left(\mathbf{p}^\dagger, \mathbf{s}^\dagger, \mathbf{X}^\dagger, \boldsymbol{\theta}_c\right)$ but for $(q, \mathbf{w}, \mathbf{x}) = \left(q^\dagger, \mathbf{w}^\dagger, \mathbf{x}^\dagger\right)$, $\widetilde{C} \neq \widetilde{C}^\dagger$.

Case 2: Suppose $MR_j(\mathbf{p}, \mathbf{s}, \mathbf{X}, \boldsymbol{\theta}_{c0}) = MR_{j^\dagger}\left(\mathbf{p}^\dagger, \mathbf{s}^\dagger, \mathbf{X}^\dagger, \boldsymbol{\theta}_{c0}\right)$ but $MR_j(\mathbf{p}, \mathbf{s}, \mathbf{X}, \boldsymbol{\theta}_c) \neq MR_{j^\dagger}\left(\mathbf{p}^\dagger, \mathbf{s}^\dagger, \mathbf{X}^\dagger, \boldsymbol{\theta}_c\right)$. Then, for any $(q, \mathbf{w}, \mathbf{x})$, from Assumptions 6 and 7, there exists υ such that $MR_j(\mathbf{p}, \mathbf{s}, \mathbf{X}, \boldsymbol{\theta}_{c0}) = MR_{j^\dagger}\left(\mathbf{p}^\dagger, \mathbf{s}^\dagger, \mathbf{X}^\dagger, \boldsymbol{\theta}_{c0}\right) = MC(q, \mathbf{w}, \mathbf{x}, \upsilon)$. Therefore, both firms have the same cost shock, and thus, $\widetilde{C} = \widetilde{C}^\dagger$. Therefore, in this case, for $(q, \mathbf{w}, \mathbf{x}) = \left(q^\dagger, \mathbf{w}^\dagger, \mathbf{x}^\dagger\right)$, $\widetilde{C} = \widetilde{C}^\dagger$ but $MR_j(\mathbf{p}, \mathbf{s}, \mathbf{X}, \boldsymbol{\theta}_c) \neq MR_{j^\dagger}\left(\mathbf{p}^\dagger, \mathbf{s}^\dagger, \mathbf{X}^\dagger, \boldsymbol{\theta}_c\right)$.

If marginal revenue is positive at $\boldsymbol{\theta}_c$ for all the firms in the population, either Case 1 or Case 2 holds.

Together, we have shown that $\boldsymbol{\theta}_{c0}$ is identified. ■

Note that in Condition 1 or in proving the lemma above, we did not need to make any assumptions about independence of any of the variables from each other, except for the fixed cost shock and the random component of the measurement error of cost. That is, $\boldsymbol{\theta}_{c0}$ is identified regardless of possible correlation across input prices, variable cost shock, observed characteristics, unobserved product characteristics and market size, within markets or across markets. For example, a positive correlation between market size and the demand/cost shock can arise as a larger market size may induce firms to invest in quality improvements or more advertising, which improves unobserved product quality but increases cost. But this does not break our identification strategy.[14]

Also, note that in the market share specification, there are no moment restrictions on the unobserved characteristics, and thus, they can contain market-level fixed effects. In particular,

[14]Notice that any violation of the F.O.C. (Equation (15)) may result in $\boldsymbol{\theta}_{c0}$ not being identified. An example would be if higher prices and more advertising spending signal product quality, as in the model of Milgrom (1986). It is also important to note that each pair of observations satisfying Condition 1 can be generated from different equilibria. Since the observables $\left\{q, \mathbf{w}, \mathbf{x}, \widetilde{C}\right\}$ uniquely determine the pair of firms that have the same cost shock υ, and the marginal cost, the above procedure identifies the true price coefficient even when multiple equilibria exist.

consider the BLP specification with market-level fixed effects:

$$\begin{aligned} u_{ij} &= \mathbf{x}_j \boldsymbol{\beta}_i + p_j \alpha_i + \xi_{jm} + \epsilon_{ij}, \\ \xi_{jm} &= \xi_{f,m} + \widetilde{\xi}_{jm}, \\ E\left[\widetilde{\xi}_{jm} \middle| \mathbf{x}_{jm}, \xi_{f,m}\right] &= 0, \end{aligned}$$

where $\xi_{f,m}$ denotes market m-specific heterogeneity. Because we do not use such moment conditions for identification, these fixed effects do not prevent us from identifying and consistently estimating the BLP parameters $\boldsymbol{\theta}_{c0} = (\mu_{\alpha 0}, \sigma_{\alpha 0}, \boldsymbol{\sigma}_{\beta 0})$.

Our main identification result is stated in the following proposition, with the proof in the appendix:

Proposition 1 *Suppose Assumptions 1-8 are satisfied. Then, the BLP coefficients* $\boldsymbol{\theta}_{c0} = (\mu_{\alpha 0}, \sigma_{\alpha 0}, \boldsymbol{\sigma}_{\beta 0})$ *are identified.*

Note that this result holds even without any variation in market size across markets. To see why, suppose we find a pair firms jm and $j^\dagger m^\dagger$ that have the same $(q, \mathbf{w}, \mathbf{x})$ and the same marginal revenue. Then, if the demand specification is logit, without market size variation, they have the same market share, and Equation (22) tells us that we cannot identify the price coefficient. On the other hand, under BLP, even though the same market size leads to the pair of firms with the same $(q, \mathbf{w}, \mathbf{x})$ having the same market share, these firms can have different price effects on own market share due to differences in prices, market shares and observed product characteristics of rival firms, and thus, different prices in the relationship.

More formally, those two firms satisfy

$$s_{jm} = s_{j^\dagger m^\dagger},\ MR_{jm} = MR_{j^\dagger m^\dagger},\ \frac{\partial s_{jm}}{\partial p_{jm}} \neq \frac{\partial s_{j^\dagger m^\dagger}}{\partial p_{j^\dagger m^\dagger}}.$$

Therefore,

$$p_{jm} = MR_{jm} - \left[\frac{\partial s_{jm}}{\partial p_{jm}}\right]^{-1} s_{jm} \neq MR_{j^\dagger m^\dagger} - \left[\frac{\partial s_{j^\dagger m^\dagger}}{\partial p_{j^\dagger m^\dagger}}\right]^{-1} s_{j^\dagger m^\dagger} = p_{j^\dagger m^\dagger}.$$

Thus, the relationship

$$p_{jm} - p_{j^\dagger m^\dagger} = -s_{jm} \left[\left(\frac{\partial s_{jm}}{\partial p_{jm}}\right)^{-1} - \left(\frac{\partial s_{j^\dagger m^\dagger}}{\partial p_{j^\dagger m^\dagger}}\right)^{-1}\right]$$

identifies the parameters.[15]

3.3 Identification of unobserved market size

We now consider the case where market size Q_m is not observed, and thus, needs to be estimated. This is an important issue in the empirical IO literature. Because market participation is unobserved, it is often hard for researchers to measure the total number of participants of a market without any arbitrariness.

We follow Bresnahan and Reiss (1991) and specify the market size as follows:

$$ln\left(Q_m\right) = \lambda_{c0} + \mathbf{z}_m \boldsymbol{\lambda}_{\mathbf{z}0}, \tag{24}$$

where $\mathbf{z}_m$ is a $1 \times K_{\mathbf{z}}$ vector of observables in market m, and $\boldsymbol{\lambda}_{\mathbf{z}0} = (\lambda_{\mathbf{z}01}, \ldots, \lambda_{\mathbf{z}0K_{\mathbf{z}}})$.

Then, the true market share of firm j in market m, denoted by

$$s^*_{jm} \equiv q_{jm}/exp\left(\lambda_{c0} + \mathbf{z}_m \boldsymbol{\lambda}_{\mathbf{z}0}\right) \tag{25}$$

is unobservable. Bresnahan and Reiss (1991) and other literature on this issue assume that variables that determine market size are not included in the market share equation. However, we do not impose such a restriction since one can convincingly argue that demographic variables determine not only market size but also consumer demand. Thus, the modified utility function for individual i in market m consuming product j is

$$u_{ijm} = \mathbf{x}_{jm}\boldsymbol{\beta}_{\mathbf{x}} + \mathbf{z}_m \boldsymbol{\beta}_{\mathbf{z}} + p_{jm}\alpha + \xi_{jm} + \epsilon_{ijm}. \tag{26}$$

On the other hand, following the literature, we assume that the variables determining market size are not included in the cost function. This assumption is reasonable as demographic variables usually do not enter the production function. Then, it follows that market structure variables only enter in the marginal revenue function but not in the marginal cost function. Therefore, the identification procedure is the same as before.

[15] The exclusion restriction for the logit model is that marginal revenue only depends on own price and own market share. That is, unobserved product characteristics of firms and prices of rival firms in a market do not enter directly in the marginal revenue equation of any given firm: these variables only enter indirectly through the market share function. For the BLP demand, we have similar exclusion restrictions at high prices. That is, if we let p_{jm} be own price, the exclusion restriction we use is that at high prices, in the 2^{nd} term of the marginal revenue function, own price only enters through $p_{jm} - p_{j-1,m}$ (where $p_{j-1,m}$ is the next highest price in the market), and $p_{jm} - p_{j+1,m}$ (where $p_{j+1,m}$ is the next lowest price in the market). For details, see the appendix.

We prove identification for the logit demand model here, and for the BLP model in the appendix.

First, note that since market share $\mathbf{s}$ and market size Q are unobserved, and market size is a function of $\mathbf{z}$ in Equation (25), marginal revenue is a function of $(\mathbf{p}, \mathbf{z}, \mathbf{q}, \mathbf{X})$ instead of $(\mathbf{p}, \mathbf{s}, \mathbf{X})$. Therefore, the first order condition is modified to be

$$MR_j\left(\mathbf{p}, \mathbf{z}, \mathbf{q}, \mathbf{X}; \boldsymbol{\theta}_{c0}\right) = MC\left(q, \mathbf{w}, \mathbf{x}, \upsilon\right),$$

where $\boldsymbol{\theta}_{c0}$ now includes the price coefficient and the parameters of the market size equation, i.e. $\boldsymbol{\theta}_{c0} = (\alpha_0, \lambda_{c0}, \boldsymbol{\lambda}_{\mathbf{z}0})$. Then, the cost shock can be expressed as follows:

$$\upsilon = \upsilon\left(q, \mathbf{w}, \mathbf{x}, MR_j\left(\mathbf{p}, \mathbf{z}, \mathbf{q}, \mathbf{X}; \boldsymbol{\theta}_{c0}\right)\right).$$

Furthermore, instead of $\mathcal{R} = \{\mathbf{p}, \mathbf{s}, \mathbf{X}, \mathbf{q}, \mathbf{w}, j\}$, now we have $\mathcal{R} = \{\mathbf{p}, \mathbf{X}, \mathbf{q}, \mathbf{w}, \mathbf{z}, j\}$ with which we derive the deterministic component of cost given below:

$$E\left[C\middle|\left(\widetilde{\mathbf{q}} = \mathbf{q}, \widetilde{\mathbf{w}} = \mathbf{w}, \widetilde{\mathbf{p}} = \mathbf{p}, \widetilde{\mathbf{z}} = \mathbf{z}, \widetilde{\mathbf{X}} = \mathbf{X}, j\right)\right] = C^v\left(q, \mathbf{w}, \mathbf{x}, \upsilon\right) + e_f\left(\mathbf{w}, \mathbf{x}, \upsilon\right) + e_{me}\left(q, \mathbf{w}, \mathbf{x}\right) \equiv \widetilde{C}.$$

Then, as before, we form pairs of firms that have the same $(q, \mathbf{w}, \mathbf{x})$ and the same $\widetilde{C}$. Through these pairs, we identify the parameters $\boldsymbol{\theta}_{c0}$ using the following two restrictions on the cost function and the marginal revenue function of the logit demand model: 1) p and $\mathbf{z}$ do not enter in the cost function. 2) Given q, variation in $\mathbf{z}$ changes the market share only through the market size equation (24), not through the utility function in Equation (26).

Furthermore, in this set of pairs, we choose the ones whose within-pair prices are equal, i.e., $p_{jm} = p_{j^\dagger m^\dagger}$. Then, since the two firms in the pair have the same marginal revenue, we derive

$$p_{jm} + \frac{1}{\left(1 - s^*_{jm}\right)\alpha_0} = p_{j^\dagger m^\dagger} + \frac{1}{\left(1 - s^*_{j^\dagger m^\dagger}\right)\alpha_0}.$$

It follows that within each pair, the true market shares must be equal. Thus, using Equation (25), we obtain

$$\begin{aligned} ln\left(s^*_{jm}\right) &= ln\left(q_{jm}\right) - ln\left(Q_m\right) = ln\left(q_{jm}\right) - \lambda_{c0} - \mathbf{z}_m \boldsymbol{\lambda}_{\mathbf{z}0} \\ &= ln\left(s^*_{j^\dagger m^\dagger}\right) = ln\left(q_{j^\dagger m^\dagger}\right) - \lambda_{c0} - \mathbf{z}_{m^\dagger} \boldsymbol{\lambda}_{\mathbf{z}0}, \end{aligned}$$

which results in

$$\left(\mathbf{z}_m - \mathbf{z}_{m^\dagger}\right)\boldsymbol{\lambda}_{\mathbf{z}0} = 0. \tag{27}$$

By using Equation (27), we first show that the vector $\boldsymbol{\lambda}_{\mathbf{z}0}$ is identified up to a multiplicative constant. That is, $\widehat{\boldsymbol{\lambda}}_{\mathbf{z}0} \equiv \boldsymbol{\lambda}_{\mathbf{z}0}/\lambda_{\mathbf{z}01}$ is identified. We do so by finding $k = 1, \ldots, K_{\mathbf{z}}$ pairs of firms in different markets $j^{(k)}m^{(k)}$ and $j^{\dagger(k)}m^{\dagger(k)}$ satisfying $q_{j^{(k)}m^{(k)}} = q_{j^{\dagger(k)}m^{\dagger(k)}}$, $\mathbf{w}_{m^{(k)}} = \mathbf{w}_{m^{\dagger(k)}}$, $\mathbf{x}_{j^{(k)}m^{(k)}} = \mathbf{x}_{j^{\dagger(k)}m^{\dagger(k)}}$, $p_{j^{(k)}m^{(k)}} = p_{j^{\dagger(k)}m^{\dagger(k)}}$ and $\widetilde{C}_{j^{(k)}m^{(k)}} = \widetilde{C}_{j^{\dagger(k)}m^{\dagger(k)}}$, but $\mathbf{z}_{m^{(k)}} \neq \mathbf{z}_{m^{\dagger(k)}}$. If we assume that conditional on $(\boldsymbol{\xi}_m, \boldsymbol{\upsilon}_m, \mathbf{w}_m, \mathbf{X}_m)$, the support of $\mathbf{z}_m$ is $R^{K_{\mathbf{z}}}$, then the space spanned by $\mathbf{z}_{m^{(k)}} - \mathbf{z}_{m^{\dagger(k)}}$, $k = 1, \ldots K_{\mathbf{z}}$, subject to the restriction of Equation (27), has rank $K_z - 1$. Therefore, under our normalization, the equations

$$\left(\mathbf{z}_{m^{(k)}} - \mathbf{z}_{m^{\dagger(k)}}\right)\widehat{\boldsymbol{\lambda}}_{\mathbf{z}0} = 0,\ k = 1, \ldots, K_{\mathbf{z}}$$

identify $\widehat{\boldsymbol{\lambda}}_{\mathbf{z}0} \equiv \boldsymbol{\lambda}_{\mathbf{z}0}/\lambda_{\mathbf{z}01}$.

Next, we focus on the identification of λ_{c0} and $\lambda_{\mathbf{z}01}$ by setting $z_{mk} = 0$ for $k = 2, \ldots K_{\mathbf{z}}$. That is, only one variable z_{m1} determines market size. We then proceed by considering two pairs of firms $k = 1, 2$ where $\mathbf{w}_{m^{(k)}} = \mathbf{w}_{m^{\dagger(k)}} = \mathbf{w}$, $\mathbf{x}_{m^{(k)}} = \mathbf{x}_{m^{\dagger(k)}} = \mathbf{x}$, $\widetilde{C}_{j^{(k)}m^{(k)}} = \widetilde{C}_{j^{\dagger(k)}m^{\dagger(k)}}$ and for small $\Delta z > 0$, $z_{m^{(k)}1} = 0$, $z_{m^{\dagger(k)}1} = \Delta z$. Output is different across the two pairs, that is, $q_{j^{(1)}m^{(1)}} = q_{j^{\dagger(1)}m^{\dagger(1)}} = q$, $q_{j^{(2)}m^{(2)}} = q_{j^{\dagger(2`)}m^{\dagger(2)}} = q'$ for $q' \neq q$. Note we do not put any restrictions on prices within the pairs. Then, these two pairs identify λ_{c0} regardless of the value of α_0. More concretely, using

$$s^*_{j^{(1)}m^{(1)}} = q/exp\left(\lambda_{c0}\right),\ s^*_{j^{\dagger(1)}m^{\dagger(1)}} = q/exp\left(\lambda_{c0} + \Delta z \lambda_{\mathbf{z}01}\right),$$

and Equation (22), we have for pair 1,

$$p_{j^{(1)}m^{(1)}} - p_{j^{\dagger(1)}m^{\dagger(1)}} = -\frac{1}{\alpha_0}\left[\frac{q}{exp\left(\lambda_{c0}\right) - q} - \frac{q}{exp\left(\lambda_{c0} + \Delta z \lambda_{\mathbf{z}01}\right) - q}\right] \equiv \Delta p\left(q, 0, \Delta z\right).$$

Note that

$$\frac{q}{exp\left(\lambda_{c0} + \Delta z \lambda_{\mathbf{z}01}\right) - q} \approx \frac{q}{exp\left(\lambda_{c0}\right) - q} - \frac{q}{\left(exp\left(\lambda_{c0}\right) - q\right)^2}\left[exp\left(\lambda_{c0} + \Delta z \lambda_{\mathbf{z}01}\right) - exp\left(\lambda_{c0}\right)\right],$$

and since we can find q such that $q \neq exp\left(\lambda_{c0}\right)$,

$$\Delta p\left(q, 0, \Delta z\right) = p_{j^{(1)}m^{(1)}} - p_{j^{\dagger(1)}m^{\dagger(1)}} \approx -\frac{1}{\alpha_0}\frac{q}{\left(exp\left(\lambda_{c0}\right) - q\right)^2} exp\left(\lambda_{c0}\right)\Delta z \lambda_{\mathbf{z}01} \neq 0$$

holds and is bounded.

Next, we do the same with the second pair with $q' \neq q$ where $q' \neq exp(\lambda_{c0})$ as well. Letting $B(q, q', 0, \Delta z) \equiv \Delta p(q, 0, \Delta z) / \Delta p(q', 0, \Delta z)$, we have

$$B\left(q, q', 0, \Delta z\right) \equiv \frac{\Delta p(q, 0, \Delta z)}{\Delta p(q', 0, \Delta z)} \approx \frac{q}{q'}\left[\frac{exp(\lambda_{c0}) - q'}{exp(\lambda_{c0}) - q}\right]^2 = \frac{q}{q'}\left[1 + \frac{q - q'}{exp(\lambda_{c0}) - q}\right]^2,$$

which identifies λ_{c0}. Then, to identify $\lambda_{\mathbf{z}01}$, we do the same with two new pairs having $z_{m1} = z$, $z_{m^\dagger 1} = z + \Delta z$, and everything else defined in the same manner as for the first two pairs, and, given λ_{c0}, we do similar calculations as before to derive

$$B\left(q, q', z, \Delta z\right) \equiv \frac{\Delta p(q, z, \Delta z)}{\Delta p(q', z, \Delta z)} \approx \frac{q}{q'}\left[\frac{exp(\lambda_{c0} + z\lambda_{\mathbf{z}01}) - q'}{exp(\lambda_{c0} + z\lambda_{\mathbf{z}01}) - q}\right]^2 = \frac{q}{q'}\left[1 + \frac{q - q'}{exp(\lambda_{c0} + z\lambda_{\mathbf{z}01}) - q}\right]^2.$$

Since λ_{c0} is already identified, the above equation identifies $\lambda_{\mathbf{z}01}$.

4 Estimation

In practice, an estimator that directly applies the parametric identification results in Subsection 3.2 will likely suffer from the Curse of Dimensionality. To implement such an estimator, one would need to derive an estimator of the deterministic component of cost $\widetilde{C}_{jm}$. Furthermore, we would need to find pairs with $q_{jm} \approx q_{j^\dagger m^\dagger}$, $\mathbf{x}_{jm} \approx \mathbf{x}_{j^\dagger m^\dagger}$, $\mathbf{w}_m \approx \mathbf{w}_{m^\dagger}$ and $\widetilde{C}_{jm} \approx \widetilde{C}_{j^\dagger m^\dagger}$. For most markets of interest, $\mathbf{X}_m$ would contain some product characteristics across a non-negligible number of firms. This makes the dimensionality problem potentially quite severe. Further, this estimator may not converge at the parametric rate and/or may not be asymptotically normal (see Khan and Tamer (2010) for a discussion of similar concerns). Because of these reasons, we construct an estimator that exploits the parametric marginal revenue in such a way that the above derivation is no longer required. This estimator conditions on marginal revenue, which is a parametric function of the data, rather than the conditional expected cost.

We propose to embed the estimation of demand parameters in the estimation of the deterministic component of cost $\widetilde{C}_{jm}$. To overcome the problem of a possible correlation between the cost shock υ_{jm} and output q_{jm}, we argue below that given q_{jm}, $\mathbf{w}_m$, and $\mathbf{x}_{jm}$, we can use marginal revenue $MR_j(\mathbf{p}_m, \mathbf{s}_m, \mathbf{X}_m, \boldsymbol{\theta}_c)$ to control for υ_{jm} as long as the demand parameter vector $\boldsymbol{\theta}_c$ equals the vector of true values $\boldsymbol{\theta}_{c0}$. The lemma below formalizes this control function idea.

Lemma 4 *Suppose that Assumptions 5-8 are satisfied. Then,* $\widetilde{C}_{jm} = \varphi(q_{jm}, \mathbf{w}_m, \mathbf{x}_{jm}, MR_{jm}(\boldsymbol{\theta}_{c0}))$

for firm j in market m with observables $\{\mathbf{p}_m, \mathbf{s}_m, \mathbf{X}_m, q_{jm}, \mathbf{w}_m, j\}$, where φ is a function that is strictly increasing and continuous in marginal revenue.

Proof. By Assumption 5 and 7, we can invert the the marginal cost function with respect to cost shock υ such that $\upsilon = \upsilon(q, \mathbf{w}, \mathbf{x}, MC)$. Using this function to control for υ, the deterministic component of cost becomes

$$\widetilde{C} = C^{v}(q, \mathbf{w}, \mathbf{x}, \upsilon) + e_f(\mathbf{w}, \mathbf{x}, \upsilon) + e_{me}(q, \mathbf{w}, \mathbf{x}) = \varphi(q, \mathbf{w}, \mathbf{x}, MC),$$

where φ is an increasing and continuous function of MC by Assumptions 7 and 8. Because Equation (15) holds, i.e., $MR = MC$, at the true parameter vector $\boldsymbol{\theta}_{c0}$,

$$\widetilde{C} = \varphi(q, \mathbf{w}, \mathbf{x}, MC) = \varphi(q, \mathbf{w}, \mathbf{x}, MR). \tag{28}$$

Thus, the claim holds. ■

We call the function $\varphi(q, \mathbf{w}, \mathbf{x}, MR)$ the pseudo-cost function.

The essence of our estimation strategy is to invert the first order condition given in Equation (18), which is the basis of our identification strategy, with respect to MR and $\widetilde{C}$, given q,w and x. Because Definition 1 guarantees invertibility at the true parameter $\boldsymbol{\theta}_0$, the pseudo-cost function is derived from our identification definition.

4.1 Two-step Sieve Nonlinear Least Squares (SNLLS) estimator

Using the above lemma, we construct an estimator that is based on the control function approach mentioned above, using a nonparametric sieve regression (see Chen (2007) and Bierens (2014)). The following assumption formalizes this:

Assumption 9 *φ can be expressed as a linear function of an infinite sequence of polynomials:*

$$\varphi(q_{jm}, \mathbf{w}_m, \mathbf{x}_{jm}, MR_{jm}(\boldsymbol{\theta}_{c0})) = \sum_{l=1}^{\infty} \gamma_{l0} \psi_l(q_{jm}, \mathbf{w}_m, \mathbf{x}_{jm}, MR_{jm}(\boldsymbol{\theta}_{c0})), \tag{29}$$

where $\psi_1(\cdot), \psi_2(\cdot), \ldots$ are the basis functions for the sieve, which are uniformly bounded on a compact domain, and $\gamma_1, \gamma_2, \ldots$ is a sequence of their coefficients, satisfying $\sum_{l=1}^{\infty} |\gamma_{l0}| < \infty$.

Our estimator is based on Equation (29). It is useful to introduce some additional notation before formally defining the estimator and its sample analog. Let M be the number of markets in the sample, and L_M an integer that increases with M. For some bounded but sufficiently large constant $T > 0$, let $\Gamma_k(T) = \{\pi_k \boldsymbol{\gamma} : \|\pi_k \boldsymbol{\gamma}\| \leq T\}$ where π_k is the operator that applies to an infinite sequence $\boldsymbol{\gamma} = \{\gamma_n\}_{n=1}^{\infty}$, replacing γ_n, $n > k$ with zeros. That is, for $n \leq k$, $\pi_k \gamma_n = \gamma_n$, and for $n > k$, $\pi_k \gamma_n = 0$. The norm $\|\mathbf{x}\|$ is defined as $\|\mathbf{x}\| = \sqrt{\sum_{k=1}^{\infty} x_k^2}$.

We now present our main result on estimation. The proof is in the appendix.

Proposition 2 *Suppose Assumptions 1-9 are satisfied. Then*

$$[\boldsymbol{\theta}_{c0}, \boldsymbol{\gamma}_0] = argmin_{(\boldsymbol{\theta},\gamma)\in\Theta_c\times\Gamma} E\left[C_{jm} - \sum_{l=1}^{\infty} \gamma_l \psi_l \Big(q_{jm}, \mathbf{w}_m, \mathbf{x}_{jm}, MR_{jm}(\boldsymbol{\theta}_c)\Big)\right]^2, \qquad (30)$$

where $\Gamma = lim_{M\to\infty} \Gamma_{L_M}(T)$. *Equation (30) identifies* $\boldsymbol{\theta}_{c0}$.

In the proof, rather than using Equation (30) directly which would require joint identification of $\boldsymbol{\theta}_{c0}$ and $\boldsymbol{\gamma}$, we use the pairing approach to identification discussed in Subsection 3.2. By forming pairs of firms along similar lines as described in that subsection, we eliminate the cost side and thus, the need for identifying $\boldsymbol{\gamma}$. Thus, the pairing approach plays a central role in our identification proofs. In the Appendix, we provide another way of expressing our pairing strategy so that it can be linked better with our estimation strategy. We show there that Definition 2 and Condition 1 can be expressed in terms of conditional variances of the deterministic component of cost as well as of marginal revenue, thereby leading to our estimator.[16] Further, note that we do not require the sieve function $\sum_{l=1}^{\infty} \gamma_l \psi_l (q, \mathbf{w}, \mathbf{x}, MR_{jm}(\boldsymbol{\theta}_c))$ to be one-to-one with respect to $MR_{jm}(\boldsymbol{\theta}_c)$ given $(q, \mathbf{w}, \mathbf{x})$, for $\boldsymbol{\theta}_c \neq \boldsymbol{\theta}_{c0}$. However, in the proof, we show that if the demand function is BLP, the sieve function satisfies this property for all $\boldsymbol{\theta}_c$.

Our SNLLS (Sieve-NLLS) approach deals with issues of endogeneity by adopting a control function approach for the unobserved cost shock v_{jm}. With our estimator, the right hand side of Equation (30) is minimized only when parameters are at their true value $\boldsymbol{\theta}_{c0}$, so that the computed marginal revenue equals the true marginal revenue, and thus, works as a control function for the supply shock v_{jm}. If $\boldsymbol{\theta}_c \neq \boldsymbol{\theta}_{c0}$, then using the false marginal revenue adds noise, which increases the sum of squared residuals in Equation (30). This can be seen from the

[16]We are grateful to an anonymous referee for suggesting this.

following: because ν_{jm} and ς_{jm} are independent to the other observed variables,

$$
\begin{aligned}
& E\left[\left(C_{jm}-\sum_{l=1}^{\infty}\gamma_l\psi_l\left(q_{jm},\mathbf{w}_m,\mathbf{x}_{jm},MR_{jm}\left(\boldsymbol{\theta}_c\right)\right)\right)^2\right] \\
= \; & E\left[\left(C_{jm}-\widetilde{C}_{jm}+\widetilde{C}_{jm}-\sum_{l=1}^{\infty}\gamma_l\psi_l\left(q_{jm},\mathbf{w}_m,\mathbf{x}_{jm},MR_{jm}\left(\boldsymbol{\theta}_c\right)\right)\right)^2\right] \\
\geq \; & E\left[\left(\nu_{jm}+\varsigma_{jm}\right)^2\right]=\sigma_\nu^2+\sigma_\varsigma^2. \qquad (31)
\end{aligned}
$$

Inequality (31) holds with equality if and only if $\boldsymbol{\theta}_c=\boldsymbol{\theta}_{c0}$, by definition (see Equation (28)). Thus, the true demand parameter $\boldsymbol{\theta}_{c0}$ can be obtained as a by-product of this control function approach.[17]

We assume the sample of M markets to be the M random draws of the population, and denote market m to be the m^{th} random draw. Then, the sample analog of Equation (30) is:

$$
\left[\widehat{\boldsymbol{\theta}}_{cM},\widehat{\boldsymbol{\gamma}}_M\right]=argmin_{(\boldsymbol{\theta}_c,\gamma)\in\Theta_c\times\Gamma_{L_M}(T)}\frac{1}{\sum_{m=1}^M J_m}\sum_{m=1}^{M}\sum_{j=1}^{J_m}\left[C_{jm}-\sum_{l=1}^{L_M}\gamma_l\psi_l\Big(q_{jm},\mathbf{w}_{jm},\mathbf{x}_{jm},MR_{jm}\left(\boldsymbol{\theta}_c\right)\Big)\right]^2. \qquad (32)
$$

The set $\Gamma_{L_M}(T)$ makes clear the fact that the complexity of the sieve is increasing in the number of markets.

In the actual estimation exercise, the objective function can be constructed in the following two steps.

Step 1: Given a candidate parameter vector $\boldsymbol{\theta}_c$, derive marginal revenue $MR_{jm}(\boldsymbol{\theta}_c)$ for each jm, $j=1,...,J_m$, $m=1,...,M$.

Step 2: Derive the estimates of $\widehat{\gamma}_l$, $l=1,...,L_M$ by OLS, where the dependent variable is C_{jm} and the RHS variables are $\psi_l\left(q_{jm},\mathbf{w}_m,\mathbf{x}_{jm},MR_{jm}\left(\boldsymbol{\theta}_c\right)\right)$, $l=1,...,L_M$. Then, construct the objective function, which isthe average of squared residuals

$$
Q_M\left(\boldsymbol{\theta}_c\right)=\frac{1}{\sum_{m=1}^M J_m}\sum_{m=1}^{M}\sum_{j=1}^{J_m}\left[C_{jm}-\sum_{l=1}^{L_M}\widehat{\gamma}_l\psi_l\left(q_{jm},\mathbf{w}_m,\mathbf{x}_{jm},MR_{jm}\left(\boldsymbol{\theta}_c\right)\right)\right]^2.
$$

We choose $\boldsymbol{\theta}_c$ that minimizes the objective function $Q_M\left(\boldsymbol{\theta}_c\right)$. In sum, we search for the demand parameters in the outer loop and find the best-fitting cost function in the inner loop for each

[17] After estimating the marginal revenue function, we can recover the cost function. The details are discussed in the appendix.

candidate set of parameters. We use the Newton search algorithm to find the solution.[18]

In the second step, to identify $\boldsymbol{\beta}$ for the logit model and $\boldsymbol{\mu}_{\boldsymbol{\beta}}$ for BLP, we include additional moment conditions in our estimator that leverage the (commonly-used) assumption that $E\left[\boldsymbol{\xi}_{jm}|\mathbf{x}_{jm}\right] = 0$. That is, after obtaining $\widehat{\boldsymbol{\theta}}_{cM}$, we can recover $\widehat{\boldsymbol{\delta}}_M$ by inversion, and then estimate $\widehat{\boldsymbol{\beta}}_M$ for logit or $\widehat{\boldsymbol{\mu}}_{\beta M}$ for BLP simply by OLS as follows:

$$\widehat{\boldsymbol{\beta}}_M = \left(\sum_{m=1}^{M} \mathbf{X}_m'\mathbf{X}_m\right)^{-1} \sum_{m=1}^{M} \mathbf{X}_m'\left(\widehat{\boldsymbol{\delta}}_m - \mathbf{p}_m\widehat{\alpha}_M\right) \text{ or } \widehat{\boldsymbol{\mu}}_{\beta M} = \left(\sum_{m=1}^{M} \mathbf{X}_m'\mathbf{X}_m\right)^{-1} \sum_{m=1}^{M} \mathbf{X}_m'\left(\widehat{\boldsymbol{\delta}}_m - \mathbf{p}_m\widehat{\mu}_{\alpha M}\right). \tag{33}$$

Equations (32) and (33) constitute our two-step SNLLS estimator for parameters $\boldsymbol{\theta} = (\alpha, \boldsymbol{\beta})$ for logit demand and $\boldsymbol{\theta} = \left(\mu_\alpha, \sigma_\alpha, \boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\sigma}_{\boldsymbol{\beta}}\right)$ for BLP demand.[19]

4.2 Further specification and data issues

We have thus far worked with the standard differentiated products model of Berry (1994) and BLP. Depending on the empirical context, however, a number of specification and data-related issues can potentially arise. In this subsection, we list some empirical settings in which our estimator can be adapted by modifying the SNLLS part of the objective function in Equation (32). The details are in the appendix. They are:

1. *Economic versus accounting cost*: With only minor modifications to our estimation procedure, we can consistently estimate the parameters even if the cost data in accounting statements do not reflect the economic cost.

2. *Endogenous product characteristics*: We can deal with the case where firms also choose product characteristics by including the additional first order conditions in our estimator. For more details, see Chu (2010), Fan (2013), Crawford (2012), and Byrne (2015).

3. *Cost function restrictions*: We can incorporate the restriction that the cost function satisfies homogeneity of degree one in input prices. Doing so has the benefit of reducing the dimensionality of the nonparametric pseudo-cost function.

[18] Note that practitioners need to be careful in the nonparametric estimation of the pseudo-cost function if $\mathbf{x}_{jm}$ includes discrete variables.

[19] Note that if the restriction $E\left[\xi_{jm}|X_{jm}\right] = 0$ is not met, we can still obtain consistent parameter estimates of $\boldsymbol{\beta}$ (or $\boldsymbol{\mu}_{\boldsymbol{\beta}}$) if we assume that the observed characteristics of other products $\mathbf{X}_{-jm}$ and ξ_{jm} are uncorrelated. Then, we can use $\mathbf{X}_{-jm}$ as instruments for $\mathbf{x}_{jm}$. In contrast, the literature uses these variables as instruments for both p_{jm} and $\mathbf{x}_{jm}$. For example, BLP use the sum of product characteristics over other firms as instruments for p_{jm}. It is also important to recall that even if $\boldsymbol{\beta}$ (or $\boldsymbol{\mu}_{\boldsymbol{\beta}}$) cannot be consistently estimated, in our procedure, $\boldsymbol{\theta}_{c0}$ is still estimated consistently, and so are marginal revenue and profit margin.

4. *Missing cost data*: Because the SNLLS part of our estimator does not involve any orthogonality conditions, and because the random components of the measurement error of cost and fixed cost are assumed to be i.i.d, choosing only those firms for which cost data is available does not result in selection bias in estimation. It is important to notice, however, that we still need demand-side data for all the firms in the same market to compute marginal revenue.

5. *Multi-product firms*: Even though firms produce multiple products, in most accounting statements, only the total cost, as opposed to product-level cost, is reported. In such cases, with logit or the BLP functional form, we can still estimate the parameters of the model by putting additional reasonable restrictions on market share functions and the cost function.

4.3 Large sample properties

In the appendix, we prove consistency and asymptotic normality of our estimator. These proofs are based on the asymptotic analysis of sieve estimators by Bierens (2014).

4.4 Bootstrap procedure for calculating the standard errors

In this subsection, we propose a bootstrap procedure for deriving the standard errors of $\widehat{\boldsymbol{\theta}}_{cM}$. In equilibrium models, bootstrapping by resampling the demand shocks ξ_{jm} and supply shocks υ_{jm} is computationally demanding because the equilibrium prices $\mathbf{p}_m$ and the market shares $\mathbf{s}_m$ need to be recomputed for each market m. Instead, Fu and Wolpin (2018), and others, propose conducting nonparametric bootstrap where one would resample market outcomes $(\mathbf{X}_m, \mathbf{p}_m, \mathbf{s}_m, \mathbf{w}_m, \mathbf{q}_m, \mathbf{C}_m)$, $m = 1, \ldots, M$ and estimate based on the resampled market data. However, the results of this procedure may be subject to small sample issues due to the relatively small number of markets. Furthermore, the bootstrapped parameter estimates would likely be affected by the additional variation from the resampled $\mathbf{X}_m$ and $\mathbf{w}_m$, which could overestimate the standard errors. In addition, if the demand and supply shocks, $\mathbf{X}_m$ and $\mathbf{w}_m$, are correlated across markets, then this correlation needs to be dealt with in resampling.

In our bootstrap, we instead resample $\nu_{jm} + \varsigma_{jm}$ to reconstruct the cost data and then, reestimate the parameters. The procedure is valid since we assume that $\nu_{jm} + \varsigma_{jm}$ is independent of other variables, which we leave unchanged. We describe the procedure below.

Step 1 Estimate the parameters $\widehat{\boldsymbol{\theta}}_{cM}^{(1)}$ and $\widehat{\gamma}_M^{(1)}$ using $C_{jm}, \mathbf{x}_{jm}, s_{jm}, p_{jm}, q_{jm}, \mathbf{w}_m$, $j = 1, \ldots, J_m$, $m = 1, \ldots, M$.

Step 2 Derive the residuals

$$\left(\widehat{\nu+\varsigma}\right)_{jm}=C_{jm}-\sum_{l=1}^{L_M}\widehat{\gamma}_{lM}^{(1)}\psi_l\left(q_{jm},\mathbf{w}_m,\mathbf{x}_{jm},MR_{jm}\left(\widehat{\boldsymbol{\theta}}_{cM}^{(1)}\right)\right).$$

Step 3 Resample with replacement from $\left\{\left(\widehat{\nu+\varsigma}\right)_{jm},j=1,\ldots,J_m,m=1,\ldots,M\right\}$ to generate $\left\{\left(\widetilde{\nu+\varsigma}\right)_{jm},j=1,\ldots,J_m,m=1,\ldots,M\right\}$.

Step 4 Generate the bootstrapped cost

$$\widehat{C}_{jm}=\sum_{l=1}^{L_M}\widehat{\gamma}_{lM}^{(1)}\psi_l\left(q_{jm},\mathbf{w}_m,\mathbf{x}_{jm},MR_{jm}\left(\widehat{\boldsymbol{\theta}}_{cM}^{(1)}\right)\right)+\left(\widetilde{\nu+\varsigma}\right)_{jm}.$$

Step 5 Go back to Step 1 with $\widehat{C}_{jm}$ instead of C_{jm}, and reestimate to derive $\widehat{\boldsymbol{\theta}}_{cM}^{(2)}$, $\widehat{\boldsymbol{\gamma}}_{M}^{(2)}$ using $\widehat{C}_{jm},\mathbf{x}_{jm},s_{jm},p_{jm},q_{jm},\mathbf{w}_m,\ j=1,\ldots,J_m,\ m=1,\ldots,M$.

Repeat the above steps M_B-1 times to derive $\boldsymbol{\theta}_{cM}^{(l_B)}$, $l_B=1,\ldots,M_B$ and report standard errors from the M_B bootstrapped parameter estimates.

5 Monte Carlo experiments

In this section, we present results from a series of Monte Carlo experiments that highlight the finite sample performance of our estimator. To generate samples, we use the following random coefficients logit demand model:

$$s_{jm}(\boldsymbol{\theta})=\int_\alpha\int_\beta\frac{\exp\left(\mathbf{x}_{jm}\beta+p_{jm}\alpha+\xi_{jm}\right)}{\sum_{j=0}^{J_m}\exp\left(\mathbf{x}_{jm}\beta+p_{jm}\alpha+\xi_{jm}\right)}\frac{1}{\sigma_\alpha}\phi\left(\frac{\alpha-\mu_\alpha}{\sigma_\alpha}\right)\frac{1}{\sigma_\beta}\phi\left(\frac{\beta-\mu_\beta}{\sigma_\beta}\right)d\alpha d\beta, \quad (34)$$

where we set the number of product characteristics K to be 1, and $\phi()$ to be the density of the standard normal distribution. We assume that each market has four firms, each producing one product (e.g., $J_m=J=4$). Hence consumers in each market have a choice of $j=1,\ldots,4$ differentiated products or not purchasing any of them ($j=0$).

On the supply-side, we assume firms compete on prices, use labor and capital inputs in production and have a Cobb-Douglas production function. Given output, input prices $\mathbf{w}=[w,r]'$ (w is the wage and r is the rental rate of capital), total cost and marginal cost functions are specified as[20]

$$C\left(q,w,r,x,\upsilon\right)=x\left(\alpha_c+\beta_c\right)\left[\frac{1}{B}\left(\frac{w}{\alpha_c}\right)^{\alpha_c}\left(\frac{r}{\beta_c}\right)^{\beta_c}\upsilon q\right]^{\frac{1}{\alpha_c+\beta_c}}$$

$$MC\left(q,w,r,x,\upsilon\right)=x\left[\frac{1}{B}\left(\frac{w}{\alpha_c}\right)^{\alpha_c}\left(\frac{r}{\beta_c}\right)^{\beta_c}\upsilon\right]^{\frac{1}{\alpha_c+\beta_c}}q^{\frac{1}{\alpha_c+\beta_c}-1}.$$

Note that the cost function is homogeneous of degree one in input prices.[21]

To create our Monte Carlo samples, we generate wage, rental rate, variable cost shock, market size Q_m, and observable product characteristics x_{jm} as follows:

$$w_m\sim i.i.d.TN\left(\mu_w,\sigma_w\right),\quad e.g.,w_m=\mu_w+\sigma_w\varrho_{wm},\quad \varrho_{wm}\sim i.i.d.TN\left(0,1\right).$$

$$r_m\sim i.i.d.TN\left(\mu_r,\sigma_r\right),\quad e.g.,r_m=\mu_r+\sigma_r\varrho_{rm},\quad \varrho_{rm}\sim i.i.d.TN\left(0,1\right).$$

$$Q_m\sim i.i.d.U\left(Q_L,Q_H\right).$$

$$x_{jm}\sim i.i.d.TN\left(\mu_x,\sigma_x\right),\quad e.g.,x_{jm}=\mu_x+\sigma_x\varrho_{xjm},\quad \varrho_{xjm}\sim i.i.d.TN\left(0,1\right).$$

$TN\left(0,1\right)$ is the truncated standard normal distribution, where we truncate both upper and lower 0.82 percentiles. $U\left(Q_L,Q_H\right)$ is the uniform distribution with lower bound of Q_L and upper bound of Q_H. Furthermore, we specify the variable cost shock as follows:

$$\upsilon_{jm}=\mu_\upsilon+\sigma_\upsilon\varrho_{\upsilon jm}+\zeta_Q\Phi^{-1}\left(\delta+\left(1.0-2\delta\right)\frac{Q_m-Q_L}{Q_H-Q_L}\right),\quad \varrho_{\upsilon jm}\sim i.i.d.TN\left(0,1\right).$$

For transforming the uniformly distributed market size shock to truncated normal distribution, we use small positive $\delta=0.025$ for truncation. We truncate the distribution of the shocks to ensure that the true cost function is positive and bounded given the parameter values of the cost function we set (which will be discussed later). We let the cost shock υ_{jm} be positively correlated with the market size shock, i.e., we set ζ_Q to be 0.2.

Importantly, we specify the unobserved characteristics so as to allow for correlation between ξ_{jm} and input prices, the cost shock, market size and the observed characteristics of products

[20] In our Monte Carlo, we assume away the deterministic components of the fixed cost and the measurement error.

[21] The cost function given the Cobb-Douglas production technology is defined as

$$C\left(q,w,r,x,\upsilon\right)=argmin_{L,K}wL+rK\quad subject\ to\ q=B\upsilon^{-1}L^{\alpha_c}K^{\beta_c}x^{-1/(\alpha_c+\beta_c)}.$$

other than j in market m denoted by $xo_{jm} \equiv (1/3) \sum_{l \neq j} \varrho_{xlm}$. Specifically, we set:

$$\xi_{jm} = \delta_0 + \delta_\xi \varrho_{\xi jm} + \delta_w \varrho_{wm} + \delta_r \varrho_{rm} + \delta_\upsilon \varrho_{\upsilon jm} + \delta_Q \Phi^{-1} \left(\delta + (1.0 - 2\delta) \frac{Q_m - Q_L}{Q_H - Q_L} \right) + \delta_{xo} xo_{jm},$$

where ϱ_ξ is the idiosyncratic component of the demand shock. We set $\delta_l = \frac{1}{2\sqrt{6}}$ for $l \in \{\xi, w, r, \upsilon, Q, xo\}$.

By construction, neither input prices nor observed characteristics of other products can be used as valid instruments for prices in demand estimation. Furthermore, since both demand and variable cost shocks are correlated with market size, one cannot use the variation of market size as an instrument for prices, or for output in the cost function estimation discussed in Subsection 2.2. We let the sum of the random terms $\nu_{jm} + \varsigma_{jm}$ be distributed $TN\left(0, \sqrt{Var\left(\nu + \varsigma\right)}\right)$ where $\sqrt{Var\left(\nu + \varsigma\right)} = \sqrt{\sigma_\nu^2 + \sigma_\varsigma^2} = 0.2$.

To solve for the equilibrium price, quantity, and market share for each oligopoly firm, we use the golden section search on price.

Table 1 summarizes the parameter setup of the Monte Carlo experiments. Table 2 presents sample statistics from the simulated data of 1600 market-firm observations (there are 400 local markets). Note that $\sigma_{\nu+\varsigma}$ is about seven percent of the total cost. The parameter estimates of $\boldsymbol{\theta}_c = (\mu_\alpha, \sigma_\alpha, \sigma_\beta)$ are obtained by the following minimization algorithm:

$$\left[\widehat{\boldsymbol{\theta}}_M, \widehat{\boldsymbol{\gamma}}_M\right] = argmin_{(\boldsymbol{\theta}_c, \gamma) \in \Theta_c \times \Gamma_{k_M}(T)} \left[\frac{1}{\sum_{m=1}^M J_m} \sum_{jm} \left[\frac{C_{jm}}{r_m} - \sum_l \gamma_l \psi_l \left(q_{jm}, \frac{w_m}{r_m}, \mathbf{x}_{jm}, \frac{MR_{jm}\left(\boldsymbol{\theta}_c\right)}{r_m} \right) \right]^2 \right.$$

In this pseudo-cost function, we exploit the property that the cost function is homogeneous of degree one. For a detailed discussion, see the appendix. We then recover $\boldsymbol{\delta}$ by inversion and in the 2nd stage, we estimate the parameter μ_β as follows:

$$\widehat{\mu}_{\beta M} = \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}' \left(\widehat{\boldsymbol{\delta}}_M - \mathbf{p}\widehat{\mu}_{\alpha M} \right).$$

In Table 3, we present the Monte Carlo results for our two-step estimator. We report the mean, standard deviation, and square root of the mean squared errors (RMSE) of the parameter estimates from 100 Monte Carlo simulation/estimation replications. From the table, we see that as the sample size increases, the standard deviation and the RMSE of the parameter estimates decrease. The results highlight the consistency of our estimator. It is noteworthy that means of the estimates are quite close to their true values even with the small sample size of 200.

Table 1: Monte Carlo Parameter Values

Parameter	Description	Value
(a) Demand-side parameters		
μ_α	Price coef. mean	2.0
σ_α	Price coef. std. dev.	0.5
μ_β	Product characteristic coef. mean	1.0
σ_β	Product characteristic coef. std. dev.	0.2
μ_X	Product characteristic mean	3.0
σ_X	Product characteristic std. dev.	1.0
δ_0	Unobserved product quality mean	4.0
δ_ξ	Unobserved product quality std. dev.	0.5
Q_L	Lower bound on market size	5.0
Q_H	Upper bound on market size	10.0
(b) Supply-side parameters		
α_c	Labor coef. in Cobb-Douglas prod. fun.	0.4
β_c	Capital coef. in Cobb-Douglas prod. fun.	0.4
μ_w	Wage mean	1.0
σ_w	Wage std. dev.	0.2
μ_r	Rental rate mean	1.0
σ_r	Rental rate std. dev.	0.2
μ_υ	Cost shock mean	0.3
σ_υ	Cost shock std. dev.	0.1
J	Number of firms in each market	4
B	Scaling factor for output in the cost function	0.8326
(c) Cost measurement error		
$\sigma_{\nu+\varsigma}$	Measurement std. dev.	0.2
(d) Correlation parameters with unobservables ξ_{jm} and υ_{jm}		
δ_{xo}	ξ_{jm} and X_{-jm} correlation	$1/(2\sqrt{6})$
δ_w	ξ_{jm} and w_m correlation	$1/(2\sqrt{6})$
δ_r	ξ_{jm} and r_m correlation	$1/(2\sqrt{6})$
δ_υ	ξ_{jm} and υ_{jm} correlation	$1/(2\sqrt{6})$
δ_Q	ξ_{jm} and Q_m correlation	$1/(2\sqrt{6})$
ζ_Q	υ_{jm} and Q_m correlation	$1/(2\sqrt{6})$

Furthermore, since the estimated parameter values are close to their true values, the standard deviations and the RMSEs are close to each other as well. Overall, these results demonstrate the validity of our approach.[22]

In Table 4, we present the results where we allow for the observed characteristics x to be correlated with the unobserved characteristics ξ. That is,

$$\xi_{jm} = \delta_0 + \delta_\xi \varrho_{\xi jm} + \delta_w \varrho_{wm} + \delta_r \varrho_{rm} + \delta_\upsilon \varrho_{\upsilon jm} + \delta_Q \Phi^{-1}\left(\delta + (1.0 - 2\delta)\frac{Q_m - Q_L}{Q_H - Q_L}\right) + \delta_{xo} xo_{jm} + \delta_x \varrho_{xjm}, \tag{35}$$

where $\delta_0 = 4.0$, $\delta_\xi = \delta_w = \delta_r = \delta_\upsilon = \delta_Q = \delta_{xo} = \delta_x = 1/\left(2\sqrt{7}\right)$. Now, by construction, no observed variable of the firm can be used as a valid instrument for its price in demand estimation. We can see that $\boldsymbol{\theta}_c$ is consistently estimated. On the other hand, μ_β is estimated to be around

[22]Results with $\sigma_{\nu+\varsigma}$ larger than 0.2 are similar to the ones presented, but with larger standard deviations and RMSEs.

Table 2: Sample Statistics from Simulated Data

Variable	Description	Mean	Std. Dev.
p_m	Price	4.104	1.239
x_{jm}	Product characteristic	1.704	0.465
ξ_m	Unobserved product quality	4.008	0.436
s_{jm}	Market share	0.191	0.091
q_{jm}	Output	1.395	0.662
C_{jm}	Total cost	2.814	1.005
w_m	Wage	1.007	0.183
r_m	Rental Rate	1.001	0.195

Notes: Sample statistics from simulated data from a Monte Carlo sample with 400 markets, $J = 4$ firms per market, and 1600 observations.

1.2, much higher than the true coefficient 1.0. The upward bias is due to the positive correlation between the demand shock ξ_{jm} and the random term of the observed characteristics ϱ_{xjm} as specified in Equation (35). However, since all the other parameters are estimated consistently, markups can still be recovered consistently.

In Table **??** in the appendix, we report the results when the variation in market size is set to be zero. We can see that overall, means of the parameter estimates become closer to the true values, and the standard deviations and the RMSEs become smaller as sample size increases. By comparing the results in Table 3, we find that the standard deviations and the RMSEs are higher than the ones where we had variation in market size. We conclude that even though the variation in market size is not needed, it helps in improving the accuracy of the estimators. We also report additional numerical examples in the appendix illustrating that market size variation is needed for estimating the logit demand parameters.

Next, we consider the case where market size is not observable, and needs to be estimated. We specify market size as follows:

$$Q_m^* = \lambda_0 + \lambda_1 z_m,$$

where Q_m^* is the unobserved market size, and we set $z_m = Q_m$, and $\lambda_0 = 0$, $\lambda_1 = 1$. Then, the true market share vector is $\mathbf{s}_m = \mathbf{q}_m / Q_m^*$. We keep the BLP market share equation as specified in Equation (34) except that market size is unobservable and therefore, parameters λ_0 and λ_1 need to be jointly estimated. Note that market shares $\mathbf{s}_m$ are unobservable as well. In Table 5, panel (a), we present the statistics of the parameter estimates that were generated from 100 repeated simulation/estimation exercises, based on the model used in Table 3 with unobservable market size. In addition, in panel (b), we report the results where we set $x_{jm} = \Phi^{-1}(z_m)$ in Equation (34). As we can see, in both cases, means of the parameter estimates are close to the true values.

Table 3: SNLLS Estimator of Random Coefficient Demand Parameters
(Product Characteristic x_{jm} and Unobserved Product Quality ξ_{jm} Uncorrelated)

(a) Price coefficients parameters

			$\hat{\mu}_\alpha$			$\hat{\sigma}_\alpha$			
Markets	Sample Size	Polynomials	Mean	Std. Dev.	RMSE	Mean	Std. Dev.	RMSE	CPU minutes
50	200	144	-2.178	0.823	0.838	0.455	0.215	0.218	274.0
100	400	171	-2.195	0.539	0.571	0.547	0.178	0.183	788.5
200	800	204	-2.004	0.184	0.183	0.506	0.094	0.093	1795.4
400	1600	256	-2.020	0.124	0.125	0.504	0.053	0.053	2278.4
True Value			-2.000			0.500			

(b) Product characteristic coefficients parameters

			$\hat{\mu}_\beta$			$\hat{\sigma}_\beta$			
Markets	Sample Size	Polynomials	Mean	Std. Dev.	RMSE	Mean	Std. Dev.	RMSE	Obj. Fun.
50	200	144	1.130	0.586	0.598	0.258	0.284	0.289	1.392D-2
100	400	171	1.069	0.272	0.280	0.265	0.213	0.222	2.294D-2
200	800	204	0.988	0.118	0.118	0.204	0.085	0.084	2.944D-2
400	1600	256	1.007	0.086	0.086	0.207	0.056	0.056	3.365D-2
True Value			1.000			0.200			

Notes: Monte Carlo experiment results are based on calibration described in panels (a)-(d) of Table 1. CPU minutes are the average estimation time in minutes across the simulation/estimation replications. Measurement error in cost data has a standard deviation of $\sigma_{\nu+\varsigma} = 0.2$ which is approximately seven percent of mean total cost.

Table 4: SNLLS Estimator of Random Coefficient Demand Parameters
(Product Characteristic x_{jm} and Unobserved Product Quality ξ_{jm} Correlated)

(a) Price coefficients parameters

			$\hat{\mu}_\alpha$			$\hat{\sigma}_\alpha$			
Markets	Sample Size	Polynomials	Mean	Std. Dev.	RMSE	Mean	Std. Dev.	RMSE	CPU minutes
50	200	144	-2.204	0.954	0.971	0.508	0.196	0.195	762.2
100	400	171	-2.071	0.272	0.279	0.517	0.117	0.118	1526.7
200	800	204	-2.012	0.150	0.150	0.500	0.066	0.065	3189.4
400	1600	256	-2.006	0.088	0.088	0.502	0.034	0.034	5770.0
True Value			-2.000			0.500			

(b) Product characteristic coefficients parameters

			$\hat{\mu}_\beta$			$\hat{\sigma}_\beta$			
Markets	Sample Size	Polynomials	Mean	Std. Dev.	RMSE	Mean	Std. Dev.	RMSE	Obj. Fun.
50	200	144	1.315	0.661	0.729	0.231	0.256	0.256	1.392D-2
100	400	171	1.222	0.158	0.272	0.213	0.111	0.111	2.320D-2
200	800	204	1.197	0.090	0.216	0.202	0.078	0.078	2.953D-2
400	1600	256	1.190	0.056	0.198	0.198	0.052	0.052	3.375D-2
True Value			1.000			0.200			

Notes: Monte Carlo experiment results are based on calibration described in panels (a)-(c) of Table 1 with ξ_{jm} distributed according to Equation (35) with $\delta_\xi = \delta_w = \delta_r = \delta_\upsilon = \delta_Q = \delta_{xo} = \delta_x = 1/\left(2\sqrt{7}\right)$. CPU minutes are the average estimation time in minutes across the simulation/estimation replications. Measurement error in cost data has a standard deviation of $\sigma_{\nu+\varsigma} = 0.2$ which is approximately seven percent of mean total cost.

In Table 6, we compare the estimated parameters using our two-step SNLLS method with the standard IV approach using instruments that are commonly used in the literature. These are: wage, rental rate and observed product characteristics of own and rival firms and their

Table 5: SNLLS Estimator of Random Coefficient Demand Parameters (Unobservable Market Size)

		(a) z_m not in market share function			*(b) z_m in market share function*		
Parameter	True Value	Mean	Std. Dev.	RMSE	Mean	Std. Dev.	RMSE
$\hat{\mu}_\alpha$	2.000	-2.025	0.125	0.127	-1.964	0.130	0.134
$\hat{\sigma}_\alpha$	0.500	0.510	0.043	0.041	0.506	0.050	0.050
$\hat{\mu}_\beta$	1.000	1.005	0.100	0.100	0.996	0.254	0.253
$\hat{\sigma}_\beta$	0.200	0.208	0.057	0.057	0.199	0.073	0.073
$\hat{\lambda}_0$	0.000	-0.034	0.249	0.250	0.002	0.096	0.096
$\hat{\lambda}_1$	1.000	1.020	0.073	0.075	1.060	0.136	0.148

Notes: Monte Carlo experiment results are based on calibration described in panels (a)-(d) of Table 1 except we assume market size is unobserved and distributed according to equation (42) in the paper, where we set $\lambda_0 = 0$ and $\lambda_1 = 1$ in generating samples. All results are based on samples with 500 markets, sample size of 2000 market-firm observations, with 256 polynomials used in estimation. Measurement error in cost data has a standard deviation of $\sigma_{\nu+\varsigma} = 0.2$ which is approximately seven percent of mean total cost.

interactions. Results show that our two-step SNLLS estimates are consistent throughout whereas the IV estimates of the demand parameters are biased when the instruments are invalid.

In the first row (SNLLS 1) of Table 6, the demand shock is set to be orthogonal to the other variables, (i.e., $\delta_\xi = 0.5,\ \delta_w = \delta_r = \delta_\upsilon = \delta_Q = \delta_{xo} = 0$), so that the instruments are valid. As we can see, the two-step SNLLS estimated coefficients are close to the true values, as are the IV estimates presented in the third row (IV1). However, the standard deviations of the IV estimates of σ_α and σ_β are higher than those of the two-step SNLLS estimates. This implies that higher order interactions of the instruments may be needed in the IV method to estimate σ_α and σ_β as accurately as the two-step SNLLS ones. Next, in the fourth row (IV2), we set $\delta_\xi = 1/\left(2\sqrt{1.08}\right)$, $\delta_\upsilon = \delta_Q = \delta_{x0} = 0$, and $\delta_w = \delta_r = 0.2\delta_\xi$.[23] Thus, the input prices are not valid instruments. We can see that while the two-step SNLLS estimates in the second row (SNLLS2) are close to the true values, the IV estimated μ_α has an upward bias. The positive direction of bias is to be expected because ξ in Equation (5) is set up to be positively correlated with the instruments. Notice also that the coefficient estimate on the observed characteristic is biased downwards, and the heterogeneity parameter of price effect, σ_α is biased upwards.

Next, in row IV3, we present the IV results where the rival firms' observed product characteristics $\mathbf{X}_{-jm}$ are correlated with own unobserved characteristics ξ_{jm}. That is, we set $\delta_{xo} = \frac{1}{2\sqrt{1.04}}, \delta_\xi = \frac{1}{2\sqrt{1.04}}, \delta_w = \delta_r = \delta_\upsilon = \delta_Q = 0$. Hence, the observed characteristics of rival firms cannot be used as instruments for own price. Results show that the IV-estimated μ_α again has a

[23] In all the subsequent analysis where we allow correlation between the demand shocks and the other variables, these correlations are set to be smaller than the ones used for the SNLLS estimates. We also conducted the Monte Carlo experiments with larger correlations, but faced numerical difficulties during the IV estimation exercise.

Table 6: SNLLS and IV Estimators of Random Coefficient Demand Parameters (Variation in Market Size)

(a) Price coefficients parameters

	$\hat{\mu}_\alpha$			$\hat{\sigma}_\alpha$			
Experiment	Mean	Std. Dev.	RMSE	Mean	Std. Dev.	RMSE	CPU minutes
SNLLS1	-2.025	0.085	0.088	0.505	0.036	0.036	9400.8
SNLLS2	-2.034	0.087	0.092	0.507	0.033	0.034	7932.0
IV1	-1.978	0.086	0.089	0.475	0.069	0.073	11571.3
IV2	-1.614	0.088	0.395	0.609	0.051	0.120	13607.4
IV3	-1.453	0.078	0.552	0.157	0.055	0.347	11793.1
IV4	-1.278	0.086	0.727	0.394	0.054	0.119	12875.1
True Value	-2.000			0.500			

(b) Product characteristic coefficients parameters

	$\hat{\mu}_\beta$			$\hat{\sigma}_\beta$			
Experiment	Mean	Std. Dev.	RMSE	Mean	Std. Dev.	RMSE	Obj. Fun.
SNLLS1	1.011	0.048	0.048	0.200	0.041	0.041	3.487D-2
SNLLS2	1.016	0.052	0.054	0.204	0.043	0.042	3.504D-2
IV1	0.982	0.052	0.055	0.197	0.147	0.146	8.344D-4
IV2	0.620	0.046	0.383	0.206	0.148	0.148	1.498D-3
IV3	0.775	0.030	0.227	0.203	0.111	0.111	9.492D-4
IV4	0.533	0.035	0.468	0.159	0.120	0.126	1.203D-3
True Value	1.000			0.200			

Notes: Monte Carlo experiment results are based on calibration described in panels (a)-(d) of Table 1 with variations in experimental designs described at the end of this table's note. The results are based on randomly generated samples with 500 markets, sample size of 2000 market-firm observations, with 256 polynomials used in estimation. CPU minutes are the average estimation time in minutes across the simulation/estimation replications. Measurement error in cost data has a standard deviation of $\sigma_{\nu+\varsigma} = 0.2$ which is approximately seven percent of mean total cost.

Monte Carlo experiment designs:

SNLLS1, IV1: instruments are valid, $\delta_w = \delta_r = \delta_v = \delta_Q = \delta_{xo} = 0$

SNLLS2, IV2: input prices are correlated with demand shock: $\delta_\xi = \delta_w = \delta_r = \frac{1}{2\sqrt{3}}$, $\delta_v = \delta_Q = \delta_{xo} = 0$

IV3: rival product observed characteristics are correlated with demand shock, $\delta_\xi = \frac{1}{2\sqrt{1.04}}$, $\delta_{xo} = \frac{1}{10\sqrt{1.04}}$, $\delta_w = \delta_r = \delta_v = \delta_Q = 0$

IV4: input prices, variable cost shock, market size, and rival product observed characteristics are correlated with the demand shock, $\delta_\xi = \frac{1}{2\sqrt{1.20}}$, $\delta_w = \delta_r = \delta_v = \delta_Q = \delta_{xo} = \frac{1}{10\sqrt{1.20}}$

positive bias. The parameter μ_β is again estimated with a negative bias, and so is σ_α, unlike the results in Table 3, where we show the two-step SNLLS estimator delivers consistent parameter estimates even if the demand shock is correlated with rival product characteristics.

Finally, in row IV4, we report the results where all the instruments considered here are positively correlated with the demand shock. Again, we have an upward bias in the IV-estimated μ_α and downward bias in the estimates of σ_α, μ_β and σ_β.

Overall, we conclude that our two-step SNLLS estimator provides unbiased parameter estimates even in situations where the commonly-used instruments are invalid and thus the IV estimates are biased. In addition, our two-step SNLLS estimator performs well even when market size is not observable, and the variable that determines market size is correlated with the demand shock or enters directly in the market share equation. Furthermore, with similar convergence criteria, the CPU minutes required for the two-step SNLLS estimator are less than for the IV estimates. Thus, we tentatively conclude that our sieve-estimation procedure does not impose excessive computational burden.

6 Empirical application to U.S. banking industry

We next apply our method to the actual data on banks and depository institutions to estimate the demand for deposits. We estimate a slightly different version of the demand model estimated by Dick (2008). In particular, we assume that each consumer has one unit to deposit. The indirect utility function of individual i putting his/her deposits in bank j in market m is specified as:

$$u_{ijm} = \mathbf{x}_{jm}\boldsymbol{\beta} + r_{djm}\alpha + \xi_{jm} + \epsilon_{ijm},$$

where $\mathbf{x}_{jm}$ is a vector of observed characteristics of bank j in market m, which consists of log of number of its branches, log of number of markets served and log of one plus bank age; r_{djm} is the deposit interest rate of bank j in market m net of the service charge, and ξ_{jm} is its unobserved characteristic. Finally, ϵ_{ijm} is the random residual term in the utility function, which is assumed to be i.i.d. Extreme-Value distributed. We follow Dick (2008) and let the outside option be depositing in credit unions. Notice that since individuals receive interest rate on their deposits, banks need to loan out or invest the deposits to earn any revenues. Thus, we assume revenue to be $(r_{jm} - r_{djm})\, q_{jm}$, where r_{jm} is the interest rate earned by bank j in market m. We set the interest rate to be $r_{jm} = r$, where r is the interest rate on the government treasury notes in January 2002.[24] Market size Q_m is the total number of deposits (including credit unions).

Then, we can write down the market share function of this model and the marginal revenue function in a straightforward way. For details, see the appendix.

Our data is for year 2002 and comes from similar sources as Dick (2008). In the appendix, we provide information on the data sources, the sample statistics and discuss some data and estimation issues.

[24] There are three possible choices of variables for the interest rate r_{jm}. One could use the interest rate on assets such as government bonds, loan interest rate, or a basket of rates of returns on loans and other financial assets. Choosing the loan interest rate would raise the additional endogeneity issue of bank lending. To the best of our knowledge, the existing literature on banking focuses primarily on (endogenous) deposits. Since one of the important goals of this empirical analysis is to demonstrate the validity of our estimator by comparing our results with those in the existing literature, we use the interest rate on the government treasury notes in January 2002, since one can reasonably assume it to be exogenous. An interesting future direction of research would be to allow for both deposits and loans to be endogenous.

We use the cost data of only those banks that operate in a single market. We do so because banks that operate in multiple markets may not exercise third degree discrimination, which then violates Assumption 5. Also, for the sake of reducing the computational burden, we restrict the sample to markets with no more than 40 banks. In the final sample, the number of banks whose cost data we use is 2067, whereas the total number of banks is 3230. As we can see in the sample statistics in Table **??**, and the results in Table **??**, even after imposing the single-market restriction, we have enough sample size and variation in the data for identification. It is important to remember that $\boldsymbol{\theta}_c$ is identified based on the assumption of independence of the measurement error to the other variables. Therefore, using the cost data of only those banks that serve only one market does not result in any selection bias for the estimation of $\boldsymbol{\theta}_c$, as long as we have data on all the variables entering the marginal revenue function, for every bank in the sample. For estimating $\boldsymbol{\mu_\beta}$, selection matters, and thus we use the full sample.

We present our results in Table **??** in the appendix. Our estimated price coefficient is around 32 and the average price elasticity is 1.64. The proportion of banks whose elasticity is less than 1 is 4 %. In contrast the IV-estimated price coefficient on the deposit interest rate in Dick (2008) ranges from 54.19 to 100.23, depending on the inclusion of the bank/market/state fixed effects. A possible reason for this difference could be that Dick (2008) uses banks in areas which are predominantly urban whereas we also include rural areas. Indeed, our estimated price effects are closer to the ones in Ho and Ishii (2012), who include rural markets in their analysis. The price elasticity is expected to be lower in rural markets because the distance to branches of other banks is likely to be greater.

We see in Panel (b) of the table that the IV-estimated price coefficient is negative and significant. The negative sign is unintuitive because it implies that a higher deposit interest rate reduces deposits. Furthermore, the IV-estimated parameters $(\boldsymbol{\mu_\beta}, \boldsymbol{\sigma_\beta})$ are all insignificant,[25] whereas in Panel (a), the two-step SNLLS estimated coefficients are all positive, as is intuitive, and significant.

[25] We tried different instruments in both logit and BLP demand models and found that only the logit specification with a relatively small number of instruments resulted in positive price coefficients. This is not surprising because the number of parameters to be estimated is higher in the BLP set-up, requiring more instruments. In this application, it is very likely that some of the instruments or some of the polynomials of the instruments are invalid.

7 Conclusion

We have developed a new methodology for estimating demand and cost parameters of a differentiated products oligopoly model. The method uses data on prices, market shares, and product characteristics, and some data on firms' costs. Using this data, our approach identifies demand parameters in the presence of price endogeneity, and a nonparametric pseudo-cost function in the presence of output endogeneity without any instruments. That is, demand and variable cost shocks do not need to be uncorrelated with demand shifters, cost shifters or market size, and demand shocks can be correlated with the observed characteristics of other products. Also, demand and variable cost shocks are allowed to be uncorrelated with each other. Moreover, our method can accommodate measurement error and fixed cost in cost data, endogenous product characteristics, multi-product firms, difference between accounting and economic costs, and some non-profit maximizing firms. In addition, we allow market size to be unobservable, and show that even without conventional exclusion restrictions on the variables determining demand and market size, we are able to identify and recover the unobserved market size, and consistently estimate the demand parameters.

In our empirical application, we use data on the banking industry to compare our estimated price coefficient of deposit interest rate to the one in the literature estimated using IVs. Our results indicate that cost data identifies the demand parameters well. In contrast, studies such as Dick (2008), Ho and Ishii (2012) and others use a large number of instruments (often 20 or more) for estimating the demand parameters. The validity of all these instruments is often quite difficult to assess.

The small bootstrapped standard errors, especially for the $\boldsymbol{\theta}_c$ estimate in our banking application imply that the cost data and the nonparametric pseudo-cost function provide strong identification restrictions to control for endogeneity. This is also consistent with the favorable small sample Monte Carlo results provided earlier. In many situations in empirical work, researchers do not have enough identification power from instruments to have their estimated coefficients be significant. Even in such cases, the cost-based estimation method could provide significant parameter estimates. Then, our method has the potential to work well as a complement to the IV-based approach. As we have seen, both methodologies use similar variation in the data. The input price, which is used as an instrument also appears as one of the variables in the cost function in the cost-based approach. The main differences between the IV approach and our cost-based approach are: 1) in the cost-based approach, such variation is more explicitly modeled, which may improve efficiency; 2) unlike the IV approach, such variation does not need to be exogenous; 3) in our approach, the unobservable market size can be identified and estimated without strong exclusion restrictions. Thus, the results using our approach could provide some guidance on the specification of markets in the IV approach, and 4) the cost-based approach requires cost data.

Our estimation strategy also presents an alternative tool for anti-trust authorities since they have the power to subpoena detailed cost data from firms for merger evaluation. Fundamental to the predictions from merger simulations based on the standard IV approach (Nevo (2001)) is the estimated demand elasticity and inferred marginal costs from the supply-side first order conditions of the structural model. The demand elasticity and the nonparametric pseudo-cost estimates based on our instrument-free approach can yield a complementary set of estimates and predictions regarding the welfare effects of proposed mergers when reliable instruments are scarce, or there are differences in opinions among the parties on the validity of the instruments.

Our estimation procedure requires marginal revenue to equal marginal cost. We believe that a fruitful direction of future research would be to make the method applicable to situations where marginal revenue fails to be equal to marginal cost. Examples include firms facing capacity constraints, or when firms' decisions include dynamic considerations.

Acknowledgements

We are grateful to the editor, the associate editor and the anonymous referees for their valuable comments. We would also like to thank Masa Hirukawa, Herman Bierens, Micheal Keane, Steve Lehrer, Weili Ding, John Rust, Robin Sickles and the seminar participants at the University of Melbourne, Monash University, ANU, Latrobe University, UTS, UBC, UNSW, University of Sydney, University of Queensland, City, University of London, IGR-IAE (Rennes), University of Tokyo and Queen's University (Canada) for many useful comments. The project was supported by Grants-in-Aid for Scientific Research by the Ministry of Education, Culture, Sports, Science and Technology, Japan, Project Number 20K01585.

References

Aigner, D., Lovell, K.C.A., Schmidt, P., 1977. Formulation and estimation of stochastic frontier production function models. Journal of Econometrics 6, 21–37.

Bayer, P., Ferreira, F., McMillan, R., 2007. A unified framework for measuring preferences for schools and neighborhoods. Journal of Political Economy 115, 588–638.

Berry, S.T., 1994. Estimating discrete-choice models of product differentiation. RAND Journal of Economics 25, 242–262.

Berry, S.T., Haile, P., 2014. Identification in differentiated products markets using market level data. Econometrica 82, 1749–1797.

Berry, S.T., Levinsohn, J., Pakes, A., 1995. Automobile prices in market equilibrium. Econometrica 63, 841–890.

Berry, S.T., Levinsohn, J., Pakes, A., 1999. Voluntary export restraints on automobiles: Evaluating a strategic trade policy. American Economic Review 89, 400–430.

Bierens, H., 2014. Consistent and asymptotic normality of sieve ml estimators under low-level conditions. Econometric Theory 30, 1021–1076.

Bresnahan, T.F., Reiss, P.C., 1991. Entry and competition in concentrated markets. Journal of Political Economy 99, 977–1009.

Byrne, D.P., 2015. Testing models of differentiated products markets: Consolidation in the cable tv industry. International Economic Review 56, 805–850.

Chen, X., 2007. Large sample sieve estimation of semi-nonparametric models, in: Heckman, J., Leamer, E. (Eds.), Handbook of Econometrics. North-Holland, Oxford, pp. 5549–5632.

Christensen, L.R., Greene, W.H., 1976. Economies of scale in u.s. electric power generation. Journal of Political Economy 84, 655–676.

Chu, C.S., 2010. The effect of satellite entry on cable television prices and product quality. Rand Journal of Economics 41, 730–764.

Clay, K., Troesken, W., 2003. Further tests of static oligopoly models: Whiskey, 1882-1898. Journal of Industrial Economics 51, 151–166.

Crawford, G.S., 2012. Accommodating endogenous product choices: A progress report. International Journal of Industrial Organization 30, 315–320.

Crawford, G.S., Yurukoglu, A., 2012. The welfare effects of bundling in multichannel television markets. American Economic Review 102, 301–317.

Dick, A.A., 2008. Demand estimation and consumer welfare in the banking industry. Journal of Banking and Finance 32, 1661–1676.

Dube, J.P., Fox, J.T., Su, C.L., 2012. Improving the numerical performance of static and dynamic aggregate discrete choice random coefficients demand estimation. Econometica 5, 2231–2267.

Fan, Y., 2013. Ownership consolidation and product characteristics: A study of the us daily newspaper market. American Economic Review 103, 1598–1628.

Fox, J.T., il Kim, K., Ryan, S.R., Bajari, P., 2012. The random coefficients logit model is identified. Journal of Econometrics 166, 204–212.

Fu, C., Wolpin, K.I., 2018. Structural estimation of a Becker-Ehrlich equilibrium model of crime: Allocating police across cities to reduce crime. Review of Economic Studies 85, 2097–2138.

Genesove, D., Mullin, W.P., 1998. Testing static oligopoly models: Conduct and cost in the sugar industry, 1890-1914. RAND Journal of Economics 29, 355–377.

Gentzkow, M., Shapiro, J.M., 2010. What drives media slant? evidence from u.s. daily newspapers. Econometrica 78, 35–71.

Gollop, F.M., Roberts, M.J., 1983. Environmental regulations and productivity growth: The case of fossil-fueled electric power generation. Journal of Political Economy 91, 654–674.

Gowrisankaran, G., Rysman, M., 2012. Dynamics of consumer demand for new durable goods. Journal of Political Economy 120, 1173–1219.

Ho, K., Ishii, J., 2012. Location and competition in retail banking. International Journal of Industrial Organization 29, 537–546.

Houde, J.F., 2012. Spatial differentiation and vertical mergers in retail markets for gasoline. American Economic Review 102, 2147–2182.

Khan, S., Tamer, E., 2010. Irregular identification, support conditions and inverse weight estimation. Econometrica 78, 2021–2042.

Kim, D.W., Knittel, C.R., 2003. Biases in static oligopoly models? evidence from the California electricity market. Journal of Industrial Economics 54, 451–470.

Knittel, C., Metaxoglou, K., 2014. Estimation of random-coefficient demand models: Two empiricists' perspective. Review of Economics and Statistics 96, 34–59.

Kutlu, L., Sickles, R., 2012. Estimation of market power in the presence of firm level inefficiencies. Journal of Econometrics 168, 141–155.

MacKaye, A., Miller, N.H., 2018. Instrument-free demand estimation. mimeograph, Harvard University .

Milgrom, P., 1986. Price and advertising signals of product quality. Journal of Political Economy 94, 796–821.

Nevo, A., 2000. Mergers with differentiated products: The case of the ready-to-eat cereal industry. RAND Journal of Economics 31, 395–421.

Nevo, A., 2001. Measuring market power in the ready-to-eat cereal industry. Econometrica 69, 307–342.

Petrin, A., 2002. Quantifying the benefits of new products: The case of the minivan. Journal of Political Economy 110, 705–729.

Petrin, A., Seo, B., 2019. Identification and estimation of discrete choice demand models when observed and unobserved characteristics are correlated. Working Paper, University of Minnesota .

Reynaert, M., Verboven, F., 2014. Improving the performance of random coefficients demand models: The role of optimal instruments. Journal of Econometrics 179, 83–98.

Smith, H., 2004. Supermarket choice and supermarket competition in market equilibrium. Review of Economic Studies 71, 235–263.

Wolfram, C., 1999. Measuring duopoly power in the british electricity spot market. American Economic Review 89, 805–826.