Queen's Economics Department Working Paper No. 1336

Instrument-free Identification and Estimation of Differentiated Products Models

David Byrne
University of Melbourne

Susumu Imai
UTS and Queen's University

Vasilis Sarafidis
Monash University

Masayuki Hirukawa
Setsunan University

Department of Economics
Queen's University
94 University Avenue
Kingston, Ontario, Canada
K7L 3N6

1-2015

Instrument-free Identification and Estimation of Differentiated Products Models[*]

David P. Byrne[†] Susumu Imai[‡]

Vasilis Sarafidis[§] and Masayuki Hirukawa[¶]

January 15, 2015

Abstract

We propose a new methodology for estimating the demand and cost functions of 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, but does not require instruments for identification. We establish non-parametric identification, consistency and asymptotic normality of our estimator. Using Monte-Carlo experiments, we show our method works well in contexts where instruments are correlated with demand and cost shocks, and where commonly-used instrumental variables estimators are biased and numerically unstable.

[*]We are grateful to Herman Bierens, Neelam Jain, Micheal Keane, Robin Sickles and the seminar and participants at the University of Melbourne, Monash University, ANU, Latrobe University, UTS, UBC, UNSW, University of Sydney, University of Queensland, City University London and Queen's for comments.

[†]Dept. of Economics, University of Melbourne, `byrned@unimelb.edu.au`.

[‡]Corresponding author, Economics Discipline Group, University of Technology Sydney and Queen's University, `Susumu.Imai@uts.edu.au`.

[§]Dept. of Econometrics and Business Statistics, Monash University, `vasilis.sarafidis@monash.edu`.

[¶]Dept. of Economics, Setsunan University, `hirukawa@econ.setsunan.ac.jp`.

1 Introduction

In this paper, we develop a new methodology for estimating models of differentiated products markets. Our approach requires commonly used demand-side data on products' prices, market shares, observed characteristics, some firm-level cost data, and the common assumption that firms set prices to maximize profits. The key novelty of our method is it does not require instruments variables to deal with the endogeneity of prices to demand shocks in estimating demand, nor the endogeneity of outputs to cost shocks in estimating cost functions.

Our study is motivated by questions surrounding the validity of instrument-based identification strategies for differentiated products models, and by recent applications that use cost data for model identification and testing. The frameworks of interest are the logit and random coefficient logit models of Berry (1994) and Berry et al. (1995) (hereafter, BLP), methodologies that have had a substantial impact on empirical research in IO and various other areas of economics.[1] These models incorporate unobserved heterogeneity in product quality, and use instruments to deal with the endogeneity of prices to these demand shocks.[2] As Berry and Haile (2014) and others point, as long as there are instruments available, fairly flexible demand functions can be identified using market-level data. Further, in the absence of cost data, firms' marginal cost functions can be recovered with a consistently estimated demand system, and the assumption that firms set prices to maximize profits given their rivals' prices.

A central issue then for the asymptotic properties of these estimators and the predictions they deliver is instrument validity. Commonly used instruments include cost shifters such as market wages, product characteristics of other products in a market ("BLP instruments"), and the price of a given product in other markets ("Hausman instruments"). As with most IV-based

[1] Leading examples from IO include measuring market power (Nevo, 2001), quantifying welfare gains from new products (Petrin, 2002), or merger evaluation (Nevo, 2000). Applications of these methods to other fields include measuring media slant (Genzkow and Shapiro, 2010), evaluating trade policy (Berry et al., 1999), and identifying sorting across neighborhoods (Bayer et al., 2007).

[2] We will use the terms unobserved heterogeneity and demand shock interchangeably.

identification strategies, there are potential concerns with all of these instruments. Market-level cost shifters such as wages tend to exhibit little variation across firms or over time. This implies they generate little exogenous variation in prices conditional on market and firm fixed effects. Recent work on endogenous product characteristics raises questions about validity of BLP instruments.[3] As with prices, firms are potentially strategic in determining their product lines, which can create correlation between product characteristics and unobserved demand shocks. Hausman instruments are compromised if demand shocks are correlated across markets, perhaps due to spatial correlation in demand or national advertising campaigns.[4]

Despite these concerns, studies regarding instrument validity in the estimation of BLP models are scarce. Indeed, virtually all methodological innovations and applications based on the BLP model have relied on price instruments and profit maximization to identify demand and cost parameters.[5] Recent applications have, however, started incorporating cost data as an additional source of identification. For instance, Houde (2012) combines wholesale gasoline prices with first order conditions that characterize stations' optimal pricing strategies to identify stations' marginal cost function parameters. Crawford and Yurukoglu (2012) and Byrne (2014) similarly exploit first order conditions and firm-level cost data to identify the cost functions of cable companies.[6] Using some cost data, Kutlu and Sickles (2012) estimate market power while allowing for inefficiency in production. Like previous research, all of these applications use instrumental variables to identify demand in a first step.

Motivated by this recent rise in the use of cost data to empirically study oligopoly, we study

[3]See Crawford (2012) for an overview of this burgeoning literature. A commonly held view among empirical IO researchers is that product lines tend to be fixed in the short run and firms mainly compete on prices. The longer the time horizon of a given study, the more endogenous product characteristics becomes a concern.

[4]See Hausman (1997) and Bresnahan (1997) for a lively debate on the validity of Hausman instruments.

[5]There has been some research assessing numerical difficulties with the BLP algorithm (Dube, Fox, and Su, 2012), Knittel and Metaxoglou, 2014), and the use of optimal instruments to help alleviate these difficulties (Reynart and Verboven,2014). All of these studies use instruments for identification.

[6]A number of papers have also used demand and cost data to test assumptions regarding conduct in oligopoly models. See, for instance, Byrne (2014), McManus (2007), Clay and Trosken (2003), Kim and Knittel (2003), Wolfram (1999).

identification in BLP-type models when researchers have access to standard demand-side data (prices, market shares, product characteristics), and some firm-level cost data.[7] 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 cost function estimation in empirical IO.[8] Our approach is unifying in that it combines data typically used this and the differentiated products literature.

Our main theoretical finding is that by combining these data, one can jointly identify non-parametric demand and cost functions using variation in market size and the profit maximization assumption (market size does not have to be exogenous). Neither price nor quantity instruments are needed to correct for the endogeneity of prices and quantities to unobserved demand and cost shocks. The methodology is thus robust to fundamental specification concerns over instrument validity in empirical research on differentiated products and cost function estimation.[9]

Getting into specifics, we develop an econometric model and approach that is based on the parametric model of demand from Berry (1994) and BLP. Our identification and estimation strategy combines three ideas. First, we note that because unobserved demand shocks in the BLP model perfectly rationalize the data, the model's predicted marginal revenue can be written as a function of demand data and demand parameters only.[10] Second, assuming firms act as differentiated Bertrand price competitors, marginal revenue will equal marginal costs for

[7] At a broader level, our paper shares a common theme with De Loecker (2011). In particular, he investigates the usefulness of previously unused demand-side data in identifying production functions and measuring productivity.

[8] Numerous studies have used such data to estimate flexible cost functions (e.g., quadratic, translog, generalized leontief) to identify economies of scale or scope, measure marginal costs, and quantify markups for a variety of industries. For identification, researchers either use instruments for quantities, or argue that in the market they study quantities are effectively exogenous from firms' point of view.

[9] Recent skepticism over instrument validity has led many researchers in fields such as empirical labor and development to favor using randomized field experiments, random or quasi-random natural experiments, and discontinuities in the endogenous regressor of interest. Implementing randomized experiments with firms in differentiated products markets is particularly challenging because of profit maximization motives. In particular, firms are hesitant to publish new insights and data from academic collaborations that provide them with a competitive advantage over their rivals. Even if one were to implement a field experiment or have access to a handful of natural experiments or discontinuties, they likely would not provide enough variation to identify all the parameters of a differentiated products model, which importantly include firm and market fixed effects that account for unobserved demand and cost shocks.

[10] Importantly, we also prove that marginal revenue uniquely identifies the demand parameters.

firms in a Nash Equilibrium. Given a non-parametric cost function that is increasing in cost shocks, we can exploit these equilibrium conditions to recover each product's cost shock as an unspecified function of cost-related observables and marginal revenue. That is, assuming profit maximization, we can use marginal revenue to construct a "control function" for cost shocks. An important insight that we exploit for this step in our procedure is that marginal revenue is a function of market shares in BLP models, whereas marginal cost is a function of output.

Finally, we use cost data to estimate a nonparametric cost function, where we control for the cost shock with marginal revenue. We call such cost function pseudo-cost function.[11] It turns out that marginal revenue works as a proper control function only when the demand parameters are at their true values; in that case, the pseudo-cost function best fits the cost data. Therefore, we jointly estimate the pseudo-cost function, control function, and hence demand parameters, where the pseudo-cost function is nonparametrically specified as a sieve function of output, input price, and marginal revenue.

We then demonstrate that with cost data, instrument-free identification does not require any functional form assumptions on the demand side either. We prove that marginal revenue and marginal cost are jointly nonparametrically identified by the sample analog of the first order condition, which corresponds to two close points in the data that equates marginal revenue and cost. From the marginal revenue one can locally identify a nonparametric nonparametric market share function.

We believe our instrument-free, nonparametric identification results make a fundamental contribution to broader literature and persistent debates over structural econometrics. Specifically, they can be viewed as a counterexample to the commonly held view that structural estimation

[11]This control function approach to cost estimation is similar to the approach of Olley and Pakes (1996) and other control function approaches to production function estimation. They derive a control function based on the first order condition of the underlying economic model and use it to control for the unobserved productivity shock. In their case, the first order conditions govern firms' strategic investment decisions. Using investment, capital, and labor cost data, these equations can be inverted to construct a control function for firms' productivity shocks.

necessarily requires functional form assumptions or exclusion restrictions.[12] We demonstrate that neither are needed to identify the demand and cost functions of differentiated products models. Only the structure of the economic model is required for identification (e.g., Bertrand-Nash Equilibrium pricing among firms with regular cost functions).

We further show how our estimator can be adapted to incorporate a number of additional features that often arise in practice. These include endogenous product characteristics, imposing restrictions to ensure properly-defined cost functions (e.g., homogeneity in input prices)[13], allowing for differences between economic and accounting costs, missing cost data for certain products or firms, the presence of multi-product firms, and estimating fixed costs. Through a set of Monte-Carlo experiments, we illustrate the ability of our estimator to deliver consistent demand parameter estimates when prices and output are correlated with demand and cost shocks; and when cost shocks, input prices and market size are all correlated with the demand shocks (e.g. when there are no valid instruments to account for the endogeneity of prices).[14]

The only other paper that we could find that exploits first order conditions to estimate demand parameters is Smith (2004). He estimates a demand model using consumer-level choice data for supermarket products. He does not, however, have product-level price data. To overcome this missing data problem, he develops a clever identification strategy that uses data on national price-cost margins, and identifies the price coefficient in the demand model as that which rationalizes these national margins. Our study differs considerably in that we focus on the more common situation where a researcher has data on prices, aggregate market shares and

[12] See Angrist and Pischke (2010) for such a critical view of structural methods. See also Keane (2009) for an overview of these common criticisms of structural models and arguments in support of the notion that econometric research - both structural and atheoretic reduced form "experimentalist" approaches - cannot avoid making numerous assumptions.

[13] Keeping with the cost-function estimation literature, we check for convexity in output after estimation.

[14] A further result from the experiments speaks to the relative numerical performance of ours and instrumental variable estimators. Whereas we easily obtain convergence in our estimation routines, for most Monte-Carlo samples, like Dube et al. (2012) and Knittel and Metaxoglou (2014) we find the BLP algorithm to be quite unstable.

total costs, but not marginal costs. Indeed, we directly build on the general BLP framework.[15]

This paper is organized as follows. In Section 2, we specify the differentiated products model of interest and review the IV based estimation approach in the literature. In Section 3, we propose our semiparametric sieve based Nonlinear Least Squares (NLLS)-GMM estimator, and discuss parametric and nonparametric identification of the model. We then analyze the large sample properties of our sieve NLLS-GMM estimator in Section 4. Section 5 contains a Monte-Carlo study that illustrates the effectiveness of our estimator in environments where standard approaches to demand estimation yield biased results. In Section 6 we conclude.

2 Differentiated products models and IV estimation

2.1 Differentiated products models

Consider the following standard differentiated products discrete choice demand model. Consumer i in market m gets the following utility from consuming one unit of product j:

$$u_{ijm} = \mathbf{x}_{jm}'\boldsymbol{\beta} + \alpha p_{jm} + \xi_{jm} + \epsilon_{ijm}, \tag{1}$$

where $\mathbf{x}_{jm}$ is a $K\times 1$ 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. Denote the demand parameter vector by $\boldsymbol{\theta} = \left[\boldsymbol{\beta}',\ \alpha\right]'$ where $\boldsymbol{\beta}$ is a $K \times 1$ vector.

Suppose there are $m = 1 \ldots M$ isolated markets that have respective market sizes Q_m.[16]

[15] Genesove and Mullin (1998) also use data on marginal cost to estimate the conduct parameters of the homogenous goods oligopoly model.

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

Each market has $j = 0 \ldots J_m$ products whose aggregate demand across individuals is

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

where q_{jm} denotes output and s_{jm} denotes the market share. In the case of the Berry (1994) logit demand model which assumes ϵ_{ijm} has a logit distribution, the aggregate market share for product j in market m is

$$s_{jm}(\boldsymbol{\theta}) \equiv s\left(\mathbf{p}_m, \mathbf{X}_m, \boldsymbol{\xi}_m, j, \boldsymbol{\theta}\right) = \frac{\exp\left(\mathbf{x}'_{jm}\boldsymbol{\beta} + \alpha p_{jm} + \xi_{jm}\right)}{\sum_{k=0}^{J_m} \exp\left(\mathbf{x}'_{km}\boldsymbol{\beta} + \alpha p_{km} + \xi_{km}\right)} = \frac{\exp\left(\delta_{jm}\right)}{\sum_{k=0}^{J_m} \exp\left(\delta_{km}\right)}, \tag{2}$$

where $\mathbf{p}_m = [p_{0m}, p_{1m}, ..., p_{J_m m}]'$ is a $(J_m + 1) \times 1$ vector, $\mathbf{X}_m = [\mathbf{x}_{0m}, \mathbf{x}_{1m}, ..., \mathbf{x}_{J_m m}]'$ 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 $\delta_{jm} = \mathbf{x}'_{jm}\boldsymbol{\beta} + \alpha p_{jm} + \xi_{jm}$ is the "mean utility" of product j. Notice from the definition of mean utility that we can also denote the share equation by $s\left(\boldsymbol{\delta}(\boldsymbol{\theta}), j\right) \equiv s\left(\mathbf{p}_m, \mathbf{X}_m, \boldsymbol{\xi}_m, j, \boldsymbol{\theta}\right)$ where $\boldsymbol{\delta}(\boldsymbol{\theta}) = [\delta_{0m}(\boldsymbol{\theta}), \delta_{1m}(\boldsymbol{\theta}), \ldots, \delta_{J_m m}(\boldsymbol{\theta})]'$ is a $J_m + 1$ by 1 vector of mean utilities.

Following standard practice, we label good $j = 0$ as the "outside good" that corresponds to not buying any one of the $j = 1, \ldots, J_m$ goods. We normalize the outside good's product characteristics, price, and demand shock to zero (e.g., $\mathbf{x}_{0m} = \mathbf{0}$, $p_{0m} = 0$, and $\xi_{0m} = 0$ for all m), which implies $\delta_{0m}(\boldsymbol{\theta}) = 0$. This normalization, together with the logit assumption for the distribution of ϵ_{ijm}, identifies the level and scale of utility.

In the case of BLP, one allows the price coefficient and the coefficients on the observed characteristics to be different for different consumers. Specifically, α has a distribution function $F_\alpha\left(.; \boldsymbol{\theta}_\alpha\right)$, where $\boldsymbol{\theta}_\alpha$ is the parameter vector of the distribution, and similarly, $\boldsymbol{\beta}$ has a distribution function $F_{\boldsymbol{\beta}}\left(.; \boldsymbol{\theta}_\beta\right)$ with parameter vector $\boldsymbol{\theta}_\beta$. The probability a consumer with the coefficients α and $\boldsymbol{\beta}$ purchases product j is identical to that provided by the market share formula in equation

(2). The aggregate market share is obtained by integrating over the distribution of α and $\boldsymbol{\beta}$,

$$s\left(\mathbf{p}_m, \mathbf{X}_m, \boldsymbol{\xi}_m, j, \boldsymbol{\theta}\right) = \int_\alpha \int_{\boldsymbol{\beta}} \frac{\exp\left(\mathbf{x}'_{jm}\boldsymbol{\beta} + \alpha p_{jm} + \xi_{jm}\right)}{\sum_{j=0}^{J_m} \exp\left(\mathbf{x}'_{jm}\boldsymbol{\beta} + \alpha p_{jm} + \xi_{jm}\right)} dF_{\boldsymbol{\beta}}\left(\boldsymbol{\beta}; \boldsymbol{\theta}_\beta\right) dF_\alpha\left(\alpha; \boldsymbol{\theta}_\alpha\right). \tag{3}$$

Often the distributions of α and each element of $\boldsymbol{\beta}$ are assumed to be independently normal, implying that the parameters consist of mean and standard deviation, i.e., $\boldsymbol{\theta}_\alpha = [\mu_\alpha, \sigma_\alpha]'$, $\boldsymbol{\theta}_{\beta k} = [\mu_{\beta k}, \sigma_{\beta k}]'$, $k = 1, ..., K$. The mean utility is then defined to be $\delta_{jm} = \mathbf{x}'_{jm}\boldsymbol{\mu}_\beta + \mu_\alpha p_{jm} + \xi_{jm}$, with $\delta_{0m} = 0$ for the outside good.

Recovering demand shocks

Given $\boldsymbol{\theta}$ and data on market shares, prices and product characteristics, we can solve for the vector $\boldsymbol{\delta}_m$ through market share inversion. This involves finding the values of vector $\boldsymbol{\delta}_m$ for market m that solve $\mathbf{s}(\boldsymbol{\delta}_m, \boldsymbol{\theta}) - \mathbf{s}_m = \mathbf{0}$, where $\mathbf{s}_m = (s_{0m}, s_{1m}, ..., s_{J_m m})'$ is the observed market share and $s\left(\boldsymbol{\delta}_m(\boldsymbol{\theta}), j, \boldsymbol{\theta}\right)$ is the market share of firm j in the model, i.e.

$$s\left(\boldsymbol{\delta}_m(\boldsymbol{\theta}), j, \boldsymbol{\theta}\right) - s_{jm} = 0, \ \textit{for } j = 0, \ldots, J_m, \tag{4}$$

and therefore,

$$\boldsymbol{\delta}_m(\boldsymbol{\theta}) = \mathbf{s}^{-1}(\mathbf{s}_m, \boldsymbol{\theta}), \tag{5}$$

That is, we find the vector of mean utilities that perfectly align the model's predicted market shares to those observed in the data.

In the context of the logit model, Berry (1994) shows we can easily recover the mean utilities for product j using its market share and the share of the outside good as $\delta_{jm}(\boldsymbol{\theta}) = \log(s_{jm}) - \log(s_{0m})$, $j = 1, \ldots, J_m$ (with δ_{0m} normalized to 0). In the random coefficient case, there is no such closed form formula for market share inversion. Instead, BLP propose a contraction mapping algorithm that recovers the unique $\boldsymbol{\delta}_m(\boldsymbol{\theta})$ that solves (5) under some regularity conditions.

With the mean utilities in hand, recovering the structural demand shocks is straightforward,

$$\xi_{jm}(\boldsymbol{\theta}) \equiv \xi\left(\mathbf{p}_m, \mathbf{X}_m, \boldsymbol{\xi}_m, j, \boldsymbol{\theta}\right) = \delta_{jm}(\boldsymbol{\theta}) - \mathbf{x}'_{jm}\boldsymbol{\beta} - \alpha p_{jm}$$

for the logit model. For the BLP model, we use $\boldsymbol{\mu}_\beta$ instead of $\boldsymbol{\beta}$ and μ_α instead of α as coefficients.

IV estimation of demand

A simple regression analysis of $\delta_{jm}(\boldsymbol{\theta}) = \mathbf{x}'_{jm}\boldsymbol{\beta} + \alpha p_{jm} + \xi_{jm}$ with $\delta_{jm}(\boldsymbol{\theta})$ being the dependent variable and $\mathbf{x}'_{jm}$ and p_{jm} being the RHS variables won't give us unbiased estimates of the price coefficient because of the likely correlation between the product price p_{jm} and the unobserved product quality ξ_{jm}. Such endogeneity problem naturally arises in differentiated product markets since firms tend to charge higher prices if their products have higher unobserved product quality.

Researchers use a variety of excluded demand instruments to overcome this issue. That is, using the inferred values of ξ_{jm} for all products and markets, we can construct a GMM estimator for $\boldsymbol{\theta}$ by assuming the following population moment conditions are satisfied at the true value of the demand parameters $\boldsymbol{\theta}_0$: $E[\xi_{jm}(\boldsymbol{\theta}_0)\mathbf{z}_{jm}] = \mathbf{0}$, where $\mathbf{z}_{jm}$ is an $L \times 1$ vector of instruments.

However, as discussed below, each type of instrument has its pitfalls. Cost shifters are often used as price instruments. This is in line with traditional market equilibrium analysis which identifies the demand curve from shifts in the supply curve caused by cost shifters. Popular examples are input prices, $\mathbf{w}_{jm}$. However, one cannot rule out the possibility that the exclusion restriction of cost shifters in the demand function does not hold. Input prices, like wages, may affect demand of the products in the same local market through changes in consumer income.

Changes in other input prices such as gasoline or electricity could reasonably be expected to affect both firms' and consumers' choices. Even further, higher input prices may induce firms to reduce product quality.

In instances where cost shifters are likely to satisfy exclusion restrictions, they are often weak instruments. For example, if one assumes that input prices are exogenously determined in some external market (such as the labor market), then all firms will face the same input prices. Therefore, cost shifters may not have sufficient within-market variation across firms to identify the demand parameters, especially if market fixed effects are included in demand specification.

In the absence of cost shifters, researchers often use product characteristics of rivals' products or market structure characteristics like the number of firms as price instruments. One naturally would worry that these variables are endogenous with respect to unobserved demand shocks.[17]

A final commonly-used set of instruments is the set of prices of product j in markets other than m (Nevo (2001); Hausman (1997)). The strength of these instruments comes from common cost shocks for product j across markets that create cross-market correlation in product j's prices. These instruments are invalid, however, if there is spatial correlation in demand shocks across markets. Regional demand shocks, for example, could generate such correlation.[18]

2.2 Supply

The cost of producing q_{jm} units of product j is assumed to be a strictly increasing function of output, $L \times 1$ vector of input prices $\mathbf{w}_{jm}$, and a cost shock υ_{jm}. That is,

$$C_{jm} = C\left(q_{jm}, \mathbf{w}_{jm}, \upsilon_{jm}, \boldsymbol{\tau}\right), \tag{6}$$

[17] Indeed, Crawford and Yurukoglu (2012), Fan (2013), Byrne (2014), and others have documented that product characteristics are strategic choices made by firms that depend on demand shocks.

[18] Firm, market, and year fixed effects are typically included in the set of instruments when panel data are available. So the exclusion restriction fails if the innovation in the demand shock in period t for product j is correlated across markets.

where C_{jm} is the total cost of producing product j in market m, and $\boldsymbol{\tau}$ is a cost parameter vector. In addition, $C()$ is assumed to be continuously differentiable and convex with respect to output. Given this cost function and the demand model above, we can write firm j's profit function[19] as

$$\pi_{jm} = p_{jm} \times s\left(\mathbf{p}_m, \mathbf{X}_m, \boldsymbol{\xi}_m, j, \boldsymbol{\theta}\right) \times Q_m - C\left(s\left(\mathbf{p}_m, \mathbf{X}_m, \boldsymbol{\xi}_m, j, \boldsymbol{\theta}\right) \times Q_m, \mathbf{w}_{jm}, \upsilon_{jm}, \boldsymbol{\tau}\right), \quad (7)$$

We assume that firms act as differentiated products Bertrand price competitors. Therefore, the optimal price and quantity of product j in market m is determined by the first order condition (F.O.C.) that equates marginal revenue and marginal cost

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

It follows from the inversion in (4) and the specification of mean utility δ_{jm} that $\boldsymbol{\xi}_m$ is a function of $\mathbf{X}_m$, $\mathbf{p}_m$, $\mathbf{s}_m$ and $\boldsymbol{\theta}$. Therefore, marginal revenue is also a function of those variables. That is,

$$MR_{jm} \equiv MR_{jm}(\boldsymbol{\theta}) \equiv MR(\mathbf{X}_m, \mathbf{p}_m, \mathbf{s}_m, j, \boldsymbol{\theta}).$$

Cost function estimation

The above discussion implies that once the cost function is estimated, one can take the derivative to obtain the marginal cost, and thereby identify the marginal revenue, and thus, the demand parameters. Below, we briefly discuss the cost function estimation. Similar to the inversion procedure in demand, the unobserved cost shock satisfies:

$$C_{jm} = C\left(q_{jm}, \mathbf{w}_{jm}, \upsilon_{jm}, \boldsymbol{\tau}\right) \Rightarrow \upsilon_{jm}(\boldsymbol{\tau}) = C^{-1}\left(q_{jm}, \mathbf{w}_{jm}, C_{jm}, \boldsymbol{\tau}\right). \quad (9)$$

[19] Here, we are assuming there is one firm for each product. We will relax this later.

As in demand estimation, there are important endogeneity concerns with standard approaches to cost function estimation. That is, output q_{jm} would potentially be negatively correlated with the cost shock υ_{jm}. This would be especially troublesome for demand estimation since it is based on the endogenous choice of output, i.e. quantity demanded. Traditionally, researchers have either argued that the problem can be ignored[20] or tried to find instruments for output.

In principle, one can estimate the cost function parameters by using the excluded demand shifters as instruments for output. We denote the vector of cost instruments by $\tilde{\mathbf{z}}_{jm}$. We can estimate $\boldsymbol{\tau}$ assuming that the following population moments are satisfied at the true value of the cost parameters $\boldsymbol{\tau}_0$: $E\left[\upsilon_{jm}(\boldsymbol{\tau}_0)\tilde{\mathbf{z}}_{jm}\right] = \mathbf{0}$. This approach potentially involves issues that are similar to the ones we discussed in applying the IV strategy to demand estimation. That is, typical instruments such as demand shifters (e.g., market demographics) affect all firms, thus generating insufficient within-market across-firm variation in equilibrium output for identification. Furthermore, one cannot completely rule out the possibility of correlation between demand shifters and the cost shock.

Instead, we jointly estimate demand and cost functions consistently with endogenous price and output but without any instruments. Below, we discuss the basic idea behind the instrument-free identification and estimation strategy.[21]

[20] See, for example, Arocena et al. (2012) for an application that exploits the institutional environment to identify cost functions.

[21] There is literature that estimates cost or production functions under imperfect competition without instruments. Examples of works are: Roeger (1995) and Klette (1999). Most of them impose restrictions on the production function, or on the demand side, in addition to the demand model specification. For example, Roeger (1995) assumes a constant returns to scale production function. In Klette (1999), mark-up is based on the "conjectured" price elasticity of demand, which is not determined endogenously from the equilibrium of the model. Recent literature on production function estimation, such as Olley and Pakes (1996), Levinsohn and Petrin (2003), Ackerberg, et. al. (2006) and Gandhi et. al. (2014) estimate production function in a way that is robust to the product market structure and endogeneity of inputs, but imposes some functional form assumptions and "timing" and other assumptions on the input and productivity processes.

3 Instrument-free identification and estimation of the price coefficient

Instead of using instruments, we estimate the demand and cost parameters directly from the first order condition of profit maximization. The following six assumptions are the main ones needed for the identification. We begin by focusing on identification of the price coefficient and the cost function; other assumptions will be added and discussed later as we extend our analysis to include the other demand parameters.

Assumption 1 *Researchers have data on outputs, product prices, market shares, input prices, and observed product characteristics of firms. In addition, data on total cost are available.*

Assumption 2 *Marginal revenue is a function of observed product characteristics, product prices and market shares. Marginal cost is a function of output, input prices and cost shock.*

Assumption 3 *The cost function is strictly increasing, continuously differentiable and strictly convex in output, and strictly increasing and continuously differentiable in cost shock and input price. Furthermore, the marginal cost function is strictly increasing and continuous in cost shock.*

Assumption 4 *Markets are isolated. Market size is not a deterministic function of demand/supply shocks, and/or demand/supply shifters.*[22]

Assumption 5 *Firms are profit maximizing, and set prices such that marginal cost equals marginal revenue, taking as given their rivals' prices.*

Assumption 6 *The support of the supply shock υ_{jm} is in R^+ and the support of the demand shock $\boldsymbol{\xi}_m$ is in R^{J_m}. However, only firms that have υ_{jm}, $\boldsymbol{\xi}_m$, $\mathbf{X}_m$, $\mathbf{p}_m$ and $\mathbf{s}_m$ such that under*

[22]Notice that the residual variation of market size, which is independent to demand/supply shocks, and/or demand/supply shifters, cannot be used as instruments because demand/supply shocks are not observed.

the true parameter vector $\boldsymbol{\theta}_0$*,* $\delta_0 - 1 \leq \left[\frac{\partial lns(\mathbf{X}_m, \mathbf{p}_m, \boldsymbol{\xi}_m, j, \boldsymbol{\theta}_0)}{\partial lnp_{jm}}\right]^{-1} \leq -\delta_0$ *for a small* $\delta_0 > 0$*, are observed in the market. The rest of the firms are out of the market. Furthermore, for the sake of simplicity, we assume* $\alpha < 0$ *for the logit model and* $\mu_\alpha < 0$ *for the BLP random coefficient model.*[23]

Notice that none of these assumptions implies an instrumental variables restriction. In order to identify the coefficient on the observed characteristics ($\boldsymbol{\beta}$ or $\boldsymbol{\mu}_\beta$)we also need the following assumption.

Assumption 7 *Unobserved quality* $\boldsymbol{\xi}_m$ *is orthogonal to observed product characteristics* $\mathbf{X}_m$*.*[24]

It is worth noting that the orthogonality condition in Assumption 7 is not required for identification of either the price parameters, nor for σ_β.

3.1 Identification of the Logit model

Before developing our general identification results, we first present our main idea using a simple example based on the Berry (1994) logit model of demand without measurement error in the cost data. We then extend our analysis of identification to the BLP (1995) random coefficient model of demand with measurement error in the cost data. Finally, we prove identification for a nonparametric model of demand with measurement error in the cost data. Since the main issue is identification of the coefficient on endogenous prices, we initially abstract from the treatment of controls $\mathbf{X}_m$.

[23] We believe Assumption 6 is often overlooked in the BLP setup. If we generate demand shocks that have reasonably large variance and are independent to other exogenous variables and cost shocks, then even for many parameter values with negative μ_α some outcomes will have market shares with positive slope with respect to price. In effect, previous researchers may have either: (1) allowed positive slopes to happen in the data; (2) implicitly avoided parameters that generate these anomalies; or (3) implicitly assumed that only demand shocks that generate negative slope are selected in the data. It is clear that the latter two strategies results in bias of the price coefficient estimate. As we will see later, since our identification and estimation strategy of the price coefficient do not use any orthogonality conditions involving demand shocks, they not subject to this form of selection bias. However, our estimator for $\boldsymbol{\beta}$ or $\boldsymbol{\mu}_{\boldsymbol{\beta}}$ will be subject to some bias.

[24] One can use other instruments for the orthogonality conditions as well.

Consider a pair of firms $(Q_m, \mathbf{w}_{jm}, s_{jm}, p_{jm})$ and $\left(Q_{m'}, \mathbf{w}_{j'm'}, s_{j'm'}, p_{j'm'}\right)$ in different markets (e.g., $m \neq m'$) whose demand shocks are ξ_{jm} and $\xi_{j'm'}$, respectively. Under the logit specification, we their respective market shares are

$$s_{jm} = \frac{\exp\left(\alpha p_{jm} + \xi_{jm}\right)}{\sum_{k=0}^{J_m} \exp\left(\alpha p_{km} + \xi_{km}\right)}, \quad s_{j'm'} = \frac{\exp\left(\alpha p_{j'm'} + \xi_{j'm'}\right)}{\sum_{k'=0}^{J_m} \exp\left(\alpha p_{k'm'} + \xi_{k'm'}\right)},$$

where $0 < s_{jm} < 1$ and $0 < s_{j'm'} < 1$. Lemma 1 below formalizes the source of identification for this parametric demand model. To prove it, we need the following assumption in addition to Assumptions 1-6.

Assumption 8 *There exists a pair of observations that satisfies* $Q_m \neq Q_{m'}$, $\mathbf{w}_{jm} = \mathbf{w}_{j'm'}$, $q_{jm} = s_{jm} Q_m = q_{j'm'} = s_{j'm'} Q_{m'}$ *and* $C_{jm} = C_{j'm'}$.

Before stating and proving the lemma, we make one final change to our differentiated products model from Section 2. Specifically, we drop the cost parameter vector $\boldsymbol{\tau}$ as we will treat the cost function $C(\cdot)$ as being nonparametric for the remainder of the paper.

Lemma 1 *Suppose Assumptions 1-6 and Assumption 8 are satisfied. Then,* $\upsilon_{jm} = \upsilon_{j'm'}$ *and under the logit model of demand,* α *is identified by*

$$\alpha = -\frac{1}{p_{jm} - p_{j'm'}} \left[\frac{1}{1 - s_{jm}} - \frac{1}{1 - s_{j'm'}} \right]. \tag{10}$$

Proof. Suppose $\upsilon_{jm} > \upsilon_{j'm'}$. Then, from strict monotonicity of the cost function in terms of the cost shock υ

$$C\left(q_{jm}, \mathbf{w}_{jm}, \upsilon_{jm}\right) = C\left(q_{j'm'}, \mathbf{w}_{j'm'}, \upsilon_{jm}\right) > C\left(q_{j'm'}, \mathbf{w}_{j'm'}, \upsilon_{j'm'}\right),$$

contradicting $C_{jm} = C_{j'm'}$. A similar contradiction obtains for $\upsilon_{jm} < \upsilon_{j'm'}$. Therefore, $\upsilon_{jm} =$

$\upsilon_{j'm'}$. As a result, the marginal cost of the two observations is the same. That is,

$$MC\left(s_{jm}Q_m, \mathbf{w}_{jm}, \upsilon_{jm}\right) = MC\left(s_{j'm'}Q_{m'}, \mathbf{w}_{j'm'}, \upsilon_{j'm'}\right).$$

Because marginal revenue equals marginal cost, for these two data points, their marginal revenues must be the same. Then, in the case of logit model,

$$p_{jm} + \frac{1}{\left(1 - s_{jm}\right)\alpha} = p_{j'm'} + \frac{1}{\left(1 - s_{j'm'}\right)\alpha}.$$

Since $Q_m \neq Q_{m'}$, $s_{jm} \neq s_{j'm'}$ and thus, for bounded negative α, $p_{jm} \neq p_{j'm'}$. It then follows that α is identified from such a pair of data points as follows:

$$\alpha = -\frac{1}{p_{jm} - p_{j'm'}}\left[\frac{1}{1 - s_{jm}} - \frac{1}{1 - s_{j'm'}}\right].$$

■

It is important to note that at no point do we explicitly or implicitly use market size Q_m as an instrument in identifying α. Instead, we exploit residual variation in market size that is independent of the demand and supply shocks and/or demand and supply shifters. In other words, for the same level of output, differences in market size imply differences in market share. These differences can be used to separately identify the demand and cost function parameters.

Of course, in practice Assumption 8 is unrealistic. However, a similar argument can be made for pairs that satisfy the equalities in Assumption 8 approximately.

The above example highlights the importance of the variation of market size Q_m for identification. If all the data came from a single market, or from two markets with the same market size, then $q_{jm} = q_{j'm'}$ implies $s_{jm} = s_{j'm'}$, and thus α cannot not be identified from (10).

Two issues are likely to arise in practice with this estimation strategy. First, suppose there exist two pairs that satisfy Assumption 8 and each pair may provide a different estimate of α. This would immediately lead a practitioner to conclude that the model is misspecified since, if the model is correct, it is impossible to have two such pairs of markets that deliver different α estimates. This issue arises because the specification of the model is too strong. According to the model, given output and input price, cost data uniquely identify cost shocks. The second issue with the strategy is that it is widely accepted that cost data are measured with error.[25]

To handle both issues, in the next assumption we explicitly introduce an additive measurement error in the cost function.

Assumption 9 *The observed cost of firm j in market m, C^d_{jm} differs from the true cost C_{jm} by measurement error, i.e.*

$$C^d_{jm} = C_{jm} + \eta_{jm}. \tag{11}$$

Measurement error η_{jm} is i.i.d. distributed with mean 0 and variance σ^2_η. In addition, measurement error is independent of $(q_{jm}, \mathbf{w}_{jm}, \mathbf{p}_m, \mathbf{s}_m, \mathbf{X}_m)$, for all j, m.

3.2 Pseudo cost function

We now define the pseudo-cost function, which is the core component of our new estimator.

Definition 1 *A pseudo-cost function is defined to be $PC(q_{jm}, \mathbf{w}_{jm}, MC_{jm})$, where MC_{jm} denotes the marginal cost for product j in market m.*

Next, we state and prove a lemma that relates the cost function to the pseudo-cost function. The lemma shows that given output and input prices, marginal cost, if observable, can be used as a proxy for the cost shock. Because we assume profit maximization, we know that marginal revenue equals marginal cost at all points in the data. Now recall from the discussion in Section 2.2 that marginal revenue can be expressed strictly as a function of demand parameters and

[25]For a discussion of this issue see, for example, Wang (2003)

data. Together with the profit maximization assumption, this implies that when the demand parameters are at their true values, marginal cost is, in effect, observable. This allows us to relate the cost function to the pseudo-cost function in practice. The lemma formalizes this idea.

Lemma 2 *Suppose that Assumptions 2, 3, 5 and 6 are satisfied. Then,* $C(q_{jm}, \mathbf{w}_{jm}, \upsilon_{jm}) = PC(q_{jm}, \mathbf{w}_{jm}, MR_{jm}(\boldsymbol{\theta}_0))$ *, and the pseudo-cost function is increasing and continuous in marginal revenue.*

Proof. First, we show that $C(q_{jm}, \mathbf{w}_{jm}, \upsilon_{jm}) = PC(q_{jm}, \mathbf{w}_{jm}, MC_{jm})$. Note that because MC is an increasing and continuous function of υ_{jm} given q_{jm} and $\mathbf{w}_{jm}$, there exists an inverse function on the domain of $MC(q_{jm}, \mathbf{w}_{jm}, \upsilon_{jm})$ such that $\upsilon_{jm} = \upsilon(q_{jm}, \mathbf{w}_{jm}, MC_{jm})$, where υ is an increasing and continuous function of MC. This implies that we can use (an unspecified function of) q_{jm}, $\mathbf{w}_{jm}$ and MC_{jm}: $\upsilon(q_{jm}, \mathbf{w}_{jm}, MC_{jm})$, to control for υ_{jm}. Substituting this "control function" for υ_{jm} into the cost function, we obtain the pseudo-cost function from Definition 1: $C(q_{jm}, \mathbf{w}_{jm}, \upsilon_{jm}) = PC(q_{jm}, \mathbf{w}_{jm}, MC_{jm})$. From the F.O.C. we know that marginal revenue must equal marginal cost when the demand parameters are at their true values, $\boldsymbol{\theta}_0$. We can therefore substitute $MR_{jm}(\boldsymbol{\theta}_0)$ in for MC_{jm} in the pseudo-cost function:

$$C(q_{jm}, \mathbf{w}_{jm}, \upsilon_{jm}) = PC(q_{jm}, \mathbf{w}_{jm}, MC_{jm}) = PC(q_{jm}, \mathbf{w}_{jm}, MR_{jm}(\boldsymbol{\theta}_0)).$$

Finally, because υ_{jm} is an increasing and continuous function of MC_{jm} given q_{jm}, $\mathbf{w}_{jm}$, and because $MR_{jm}(\boldsymbol{\theta}_0) = MC_{jm}$ at $\boldsymbol{\theta}_0$, PC is also an increasing and continuous function of MR. ■

This lemma allows us to use the pseudo-cost function instead of the cost function in estimation. The advantage in doing so is that the former is only a function of data and parameters, whereas the latter depends on the unobservable cost shock υ.

3.3 Proposed estimator

We now present our estimator. It selects demand parameters to fit the pseudo-cost function to the cost data using a nonparametric sieve regression (Chen (2007); Bierens (2014)). We thus assume

Assumption 10 *The true pseudo-cost function can be expressed as a linear function of an infinite sequence of polynomials.*

$$PC\left(q_{jm}, \mathbf{w}_{jm}, MR_{jm}(\boldsymbol{\theta}_0)\right) = \sum_{l=1}^{\infty} \gamma_l \psi_l \left(q_{jm}, \mathbf{w}_{jm}, MR_{jm}(\boldsymbol{\theta}_0)\right), \tag{12}$$

where $\psi_1(\cdot), \psi_2(\cdot), \ldots$ are the basis functions for the sieve and $\gamma_1, \gamma_2, \ldots$ is a sequence of their coefficients, satisfying $\sum_{l=1}^{\infty} \gamma_l^2 < \infty$.[26]

Our estimator is derived from the approximation of (12). It is useful to introduce some additional notation before formally defining it. Let M be the number of markets, 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_k, $k > n$ with zeros. The norm $\|\mathbf{x}\|$ is defined as $\|\mathbf{x}\| = \sqrt{\sum_{k=1}^{\infty} x_k^2}$. We will prove later that at the true value of $\boldsymbol{\theta}$ and $\boldsymbol{\gamma} = [\gamma_1, \ldots]'$, the population distance between the cost data and the sieve-approximated pseudo-cost function is minimized. That is

$$[\boldsymbol{\theta}_0, \boldsymbol{\gamma}_0] = \underset{(\boldsymbol{\theta},\gamma)\in\Theta\times\Gamma}{\arg\min}\; E\left[C_{jm}^d - \sum_l \gamma_l \psi_l \left(q_{jm}, \mathbf{w}_{jm}, MR_{jm}(\boldsymbol{\theta})\right)\right]^2, \tag{13}$$

[26] Suppose the vector $(q_{jm}, \mathbf{w}_{jm}, MR_{jm})$ comes from a compact finite dimensional Euclidean space, $\mathcal{W}$. Then, if $PC\left(q_{jm}, \mathbf{w}_{jm}, MC_{jm}\right)$ is a continuous function over $\mathcal{W}$, from the Stone-Weierstrass Theorem it follows that the function can be approximated arbitrarily well by an infinite sequence of polynomials.

where Θ is the demand parameter space, and $\Gamma = lim_{M\to\infty}\Gamma_{L_M}(T)$. Our estimator solves the sample analogue of (13), given a sample of M markets:

$$\left[\hat{\boldsymbol{\theta}}_M, \hat{\boldsymbol{\gamma}}_M\right] = \underset{(\boldsymbol{\theta},\gamma)\in\Theta\times\Gamma_{L_M}(T)}{\arg\min} \frac{1}{\sum_m J_m} \sum_{j,m} \left[C^d_{jm} - \sum_l \gamma_l \psi_l \left(q_{jm}, \mathbf{w}_{jm}, MR_{jm}(\boldsymbol{\theta})\right)\right]^2. \quad (14)$$

The set $\Gamma_{L_M}(T)$ makes explicit the fact that the complexity of the sieve is increasing in the sample's number of markets.[27]

3.3.1 Identification

We now prove identification of the estimator. First, we state the assumption that marginal revenue identifies the demand function parameters.

Assumption 11 *Marginal revenue identifies the demand function parameters. More concretely, consider two firms jm and $j'm'$ in different markets $m \neq m'$ with the number of firms being J_m and $J_{m'}$, respectively. Denote $\tilde{\boldsymbol{\nu}}$ to be the vector of the first firm's output, input price, prices and market shares of all firms in the same market, and $\tilde{\tilde{\boldsymbol{\nu}}}$ being the same vector of the second firm. We assume the two firms have the same output and input price, i.e.,*

$$\tilde{\boldsymbol{\nu}} = (q, \mathbf{w}, \tilde{\mathbf{p}}, \tilde{\mathbf{s}}),\ \tilde{\tilde{\boldsymbol{\nu}}} = \left(q, \mathbf{w}, \tilde{\tilde{\mathbf{p}}}, \tilde{\tilde{\mathbf{s}}}\right),\ \tilde{\boldsymbol{\nu}} \neq \tilde{\tilde{\boldsymbol{\nu}}}.$$

If $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, then there exist two such firms with outcomes $\tilde{\boldsymbol{\nu}} \neq \tilde{\tilde{\boldsymbol{\nu}}}$, that satisfy and the following properties

[27] In the actual estimation exercise, the objective function can be constructed in the following 2 steps.

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

Step 2: Derive the estimates of $\hat{\gamma}_l$, $l = 1, ..., L_M$ by OLS, where the dependent variable is C^d_{jm} and the RHS variables are $\psi_l(q_{jm}, \mathbf{w}_{jm}, MR_{jm}(\boldsymbol{\theta}))$, $l = 1, ..., L_M$. Then, construct the objective function, which is the average of squared residuals $Q_M(\boldsymbol{\theta}) = \frac{1}{\sum_m J_m}\sum_{j,m}\left[C^d_{jm} - \sum_{l=1}^{L_M}\hat{\gamma}_l\psi_l(q_{jm}, \mathbf{w}_{jm}, MR_{jm}(\boldsymbol{\theta}))\right]^2$.

We choose $\boldsymbol{\theta}$ that minimizes the objective function $Q_M(\boldsymbol{\theta})$. In sum, we search for the demand parameters in an outer loop, and find the best fitting cost function on an inner loop for each candidate set of demand parameters.

1 $q > 0$, $\mathbf{w}_l > 0$, *for* $l = 1, ..., L$ *and* $\tilde{\mathbf{p}}_l > 0$, $0 < \tilde{\mathbf{s}}_l < 1$ *for* $l = 1, ..., J_m$ *and* $\tilde{\tilde{\mathbf{p}}}_l > 0$, $0 < \tilde{\tilde{\mathbf{s}}}_l < 1$, *for* $l = 1, .., J_{m'}$, *and* $0 < \sum_{l=1}^{J_m} \tilde{\mathbf{s}}_l < 1$, $0 < \sum_{l=1}^{J_{m'}} \tilde{\tilde{\mathbf{s}}}_l < 1$.

2 $MR_{jm}(\tilde{\mathbf{p}}, \tilde{\mathbf{s}}, \boldsymbol{\theta}) = MR_{j'm'}\left(\tilde{\tilde{\mathbf{p}}}, \tilde{\tilde{\mathbf{s}}}, \boldsymbol{\theta}\right)$ *and* $MR_{jm}(\tilde{\mathbf{p}}, \tilde{\mathbf{s}}, \boldsymbol{\theta_0}) \neq MR_{j'm'}\left(\tilde{\tilde{\mathbf{p}}}, \tilde{\tilde{\mathbf{s}}}, \boldsymbol{\theta}_0\right)$.[28]

3 *For any open sets* $\mathcal{A}$ *including* (q, w), $\tilde{\mathcal{B}}$ *including* $(\tilde{\mathbf{p}}, \tilde{\mathbf{s}})$ *and* $\tilde{\tilde{\mathcal{B}}}$ *including* $\left(\tilde{\tilde{\mathbf{p}}}, \tilde{\tilde{\mathbf{s}}}\right)$, $Prob\left(\mathcal{A} \times \tilde{\mathcal{B}}\right) > 0$, $Prob\left(\mathcal{A} \times \tilde{\tilde{\mathcal{B}}}\right) > 0$.

Proposition 1 *Suppose Assumptions 1-6, 9-11 hold. Then, equation (13) identifies* $\boldsymbol{\theta}_0$.

Proof. See Appendix.

We have shown above that in fitting the pseudo-cost function to the cost data, we identify the demand parameters. Notice that in the above estimation, the assumption we actually impose is somewhat weaker than profit maximization (i.e. marginal revenue equals marginal cost). We only require that marginal cost is an increasing function of marginal revenue.[29]

Our sieve NLLS approach deals with issues of endogeneity by adopting a control function approach for the unobserved cost shock υ_{jm}. With our estimator, the right hand side of (14) is minimized only when the demand parameters are at their true value $\boldsymbol{\theta}_0$ so that the computed

[28] As we will see later, this is essentially equivalent to assuming nonlinearity of the marginal revenue function.

[29] One may argue that a more straightforward estimation strategy is to construct a pairwise differenced estimator that pairs up firms in different markets with similar outputs, input prices, and marginal revenues. Specifically,

$$\theta^*_{JM} = argmin_{\theta \in \Theta} \sum_{j,m} \sum_{j',m':(j',m') \neq (j,m)} \left[\left(C^d_{jm} - C^d_{j'm'}\right)^2 W_h \left(q^d_{jm} - q^d_{j'm'}, \mathbf{w}_{jm} - \mathbf{w}_{j'm'}, MR_{jm}(\boldsymbol{\theta}) - MR_{j'm'}(\boldsymbol{\theta})\right)\right],$$

where pairs are a weighted as a function of the difference in their outputs, input prices, and marginal revenues,

$$\begin{aligned} & W_h \left(q_{jm} - q_{j'm'}, \mathbf{w}_{jm} - \mathbf{w}_{j'm'}, MR_{jm}(\boldsymbol{\theta}) - MR_{j'm'}(\boldsymbol{\theta})\right) \\ \equiv \; & \frac{K_{h_q}\left(q_{jm} - q_{j'm'}\right) K_{h_w}\left(\mathbf{w}_{jm} - \mathbf{w}_{j'm'}\right) K_{h_{MR}}\left(MR_{jm}(\boldsymbol{\theta}) - MR_{j'm'}(\boldsymbol{\theta})\right)}{\sum_{k,n} \sum_{k',n':(k',n') \neq (k,n)} K_{h_q}\left(q_{kn} - q_{k'n'}\right) K_{h_w}\left(\mathbf{w}_{kn} - \mathbf{w}_{k'n'}\right) K_{h_{MR}}\left(MR_{kn}(\boldsymbol{\theta}) - MR_{k'n'}(\boldsymbol{\theta})\right)} \end{aligned}$$

where $K_h(\cdot)$ is a kernel with bandwidth h. Since we are not pairing up firms with exactly the same marginal revenues, by construction, the monotone relationship between marginal revenue and marginal cost does not always hold within a pair, which results in a loss of efficiency for this estimator. In contrast, the sieve- based approach is built on the assumption that the monotone relationship between marginal revenue and marginal cost holds *exactly* for each firm. This additional constraint can be shown to increase efficiency. Furthermore, as we will see later, the sieve NLLS estimator is more flexible in dealing with some practical data issues such as multi-product firms than a pairwise differenced estimator.

marginal revenue equals the true marginal revenue, e.g., the marginal cost, and thus works as a control function for the supply shock υ_{jm}. If $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, then using the false marginal revenue adds noise, which increases the right hand size of the sum of squared residuals in (14). In this sense, we are adopting a pseudo-control function approach, but without any need for instruments. As the above argument makes clear, the true demand parameter $\boldsymbol{\theta}_0$ can be obtained as a by-product of this control function approach.

3.3.2 Identification of marginal revenue

It is important to note that Assumption 11 is a high level assumption; it is not necessarily satisfied in all demand models. For example, if marginal revenue is a linear function of $\boldsymbol{\theta}$, then for any positive constant $a > 0$, if we set $\boldsymbol{\theta} = a\boldsymbol{\theta}_0$, then, for any $(\tilde{\mathbf{p}}_m, \tilde{\mathbf{s}}_m)$ and $\left(\tilde{\tilde{\mathbf{p}}}_m, \tilde{\tilde{\mathbf{s}}}_m\right)$, $MR\left(\tilde{\mathbf{p}}_m, \tilde{\mathbf{s}}_m, j, \boldsymbol{\theta}\right) = MR\left(\tilde{\tilde{\mathbf{p}}}_m, \tilde{\tilde{\mathbf{s}}}_m, j, \boldsymbol{\theta}\right)$ implies $MR\left(\tilde{\mathbf{p}}_m, \tilde{\mathbf{s}}_m, j, \boldsymbol{\theta}\right) = aMR\left(\tilde{\mathbf{p}}_m, \tilde{\mathbf{s}}_m, j, \boldsymbol{\theta}_0\right)$, $MR\left(\tilde{\tilde{\mathbf{p}}}_m, \tilde{\tilde{\mathbf{s}}}_m, j, \boldsymbol{\theta}\right) = aMR\left(\tilde{\tilde{\mathbf{p}}}_m, \tilde{\tilde{\mathbf{s}}}_m, j, \boldsymbol{\theta}_0\right)$. Hence, $MR\left(\tilde{\mathbf{p}}_m, \tilde{\mathbf{s}}_m, j, \boldsymbol{\theta}_0\right) = MR\left(\tilde{\tilde{\mathbf{p}}}_m, \tilde{\tilde{\mathbf{s}}}_m, j, \boldsymbol{\theta}_0\right)$. Hence, Assumption 11 is violated. The important question, then, is whether standard differentiated products demand models satisfy Assumption 11. The lemma below answers this question.

Lemma 3 *Assumption 11 is satisfied for the logit model. It holds as well as for BLP model of demand without observed product characteristics for monopoly markets if*

$$\frac{\mu_{\alpha 0}}{\sigma_{\alpha 0}} < -\frac{1}{2\phi\left(0\right)}. \tag{15}$$

Proof. See Appendix.

Inequality (15) needs to be satisfied so that there exists (p, s) that generates negative slope of the market share with respect to price and the positive marginal revenue.[30]

[30] We have conducted extensive numerical studies, computing monopoly equilibria with different demand and supply shocks. We could not find any price and market share combination that has both negative market share-price slope and positive marginal revenue if the inequality is not satisfied. If it is satisfied, then, it can be shown that the point (p, s) where $s = 1/2$ and p sufficiently large satisfies both conditions.

Next, we include controls $\mathbf{X}_m$ into the demand model, and show that the cost data identifies the parameters of the random coefficients on price $(\mu_\alpha, \sigma_\alpha)$, and σ_β, the standard deviation of the distribution of β.

Lemma 4 *Suppose that the logit or BLP model now includes the exogenous demand controls* $\mathbf{X}_m$*. Assumption 11 is still satisfied for the logit model with respect to* α*. Assumption 11 is satisfied for the BLP model of demand for the parameters* $(\mu_\alpha, \sigma_\alpha)$ *and* σ_β *under monopoly as well.*

Proof. See Appendix.

In the proof for the BLP model, we had to rely on firms with very high prices for identification. That is unattractive, but necessary to deal with the complexity in separately identifying the parameters of the distribution of the random coefficients.[31] As we will see later, nonparametric identification of the marginal revenue and the market share equation does not rely on having such firms.

Proving Assumption 11 for the BLP model for oligopoly markets is a straightforward extension of Lemma 3 and is thus omitted. It requires data that contains firms with high and similar prices, e.g. $p_{1m} = p_{2m} = ...p_{J_m m} = p$ for sufficiently high p. Despite the need for these strong assumptions in the formal argument for parametric identification for the BLP model, we later show with Monte-Carlo experiments that our estimator identifies the coefficients of the logit model and the BLP model very well in datasets without such firms with high prices.

Estimating taste parameters for product characteristics

Recall that our estimator in equation (14) abstracted from having product characteristics $\boldsymbol{x}_{jm}$ in the demand model. Allowing for them now, we can identify β for the logit model and μ_β for

[31] We conducted simulation exercises to see the conditions that are required to observe data are satisfied with such high prices, i.e., the market share having negative slope with respect to price, and the marginal revenue to be positive. What we need is that the demand shock increases with price such that the resulting monopoly market share is not below 15 percent.

BLP if we include additional moment conditions in our estimator that leverage the (common) assumption that $E\left[\boldsymbol{\xi}_{jm}|\mathbf{X}_m\right] = 0$. Our modified estimator simply minimizes the weighted sum of the original NLLS objective function and the GMM objective function based on the sample analog of these orthogonality conditions between the observed and unobserved product characteristics. That is,

$$\left[\hat{\boldsymbol{\theta}}_M, \hat{\gamma}_M\right] = argmin_{(\boldsymbol{\theta},\gamma)\in\Theta\times\Gamma_{L_M}} \frac{1}{\sum_m J_m} \sum_{jm} \left[C^d_{jm} - \sum_{l=1}^{L_M} \gamma_l \psi_l \left(q_{jm}, \mathbf{w}_{jm}, MR_{jm}\left(\boldsymbol{\theta}\right)\right)\right]^2$$
$$+A\left[\frac{\sum_{m=1}^{M}\sum_{j=1}^{J_m}\hat{\boldsymbol{\xi}}_{jm}\left(\boldsymbol{\theta}\right)\mathbf{X}_m}{\sum_{m=1}^{M}\sum_{j=1}^{J_m}1}\right]' \mathbf{W}_M \left[\frac{\sum_{m=1}^{M}\sum_{j=1}^{J_m}\hat{\boldsymbol{\xi}}_{jm}\left(\boldsymbol{\theta}\right)\mathbf{X}_m}{\sum_{m=1}^{M}\sum_{j=1}^{J_m}1}\right]$$

where

$$\mathbf{W}_M = \left(\frac{1}{\sum_{m=1}^{M}\sum_{j=1}^{J_m}1}\sum_{m=1}^{M}\sum_{j=1}^{J_m}\hat{\boldsymbol{\xi}}_{jm}\left(\boldsymbol{\theta}_M\right)\mathbf{X}_m\mathbf{X}'_m\hat{\boldsymbol{\xi}}_{jm}\left(\boldsymbol{\theta}_M\right)\right)^{-1}$$

and A is a positive constant.

3.4 Nonparametric identification of marginal revenue function

We have so far assumed Berry (1994) logit or BLP random coefficient logit functional forms of demand. In this section, we show that marginal revenue is nonparametrically identified, and that the market share function can be recovered from nonparametric marginal revenue estimates. In practice, however, identification and estimation will be subject to a Curse of Dimensionality. This provides motivation for the use of the parametric methods described above in practice.

To simplify our discussion, we first focus on monopoly markets; this allows us to drop the j subscript. In addition, for the sake of brevity in notation below, we will not explicitly state the dependence of marginal revenue $MR(\cdot)$ and market shares $s(\cdot)$ on product characteristics $\mathbf{x}_{jm}$, nor the dependence of marginal costs $MC(\cdot)$ on input prices $\mathbf{w}_{jm}$. After establishing nonparametric identification in monopoly markets, we then discuss how the logic can be straightforwardly extended to oligopoly markets.

We begin by making the following auxiliary assumptions:

Assumption 12 *The marginal revenue function $MR(p,\xi)$ is strictly increasing in price. Furthermore, for any two pairs of prices and market shares (p_1, s_1) and (p_2, s_2) such that $s_1 = s_2$ and $p_1 > p_2$,*

$$MR_1 > MR_2,$$

where MR_i is the marginal revenue of firm i in the pair, and $MR\left(p,\xi\right)$ is the marginal revenue specified as a function of price p and the demand shock ξ.

Assumption 13 *The market share function $s\left(p,\xi\right)$ is strictly decreasing and continuous in p and strictly increasing and continuous in ξ. Furthermore,*

$$lim_{\xi\downarrow-\infty}s\left(p,\xi\right)=0, \quad lim_{\xi\uparrow\infty}s\left(p,\xi\right)=1 \quad and \quad lim_{p\uparrow\infty}s\left(p,\xi\right)=0.$$

Assumption 3′ *The marginal cost function is strictly increasing and continuous in υ. Furthermore, for any $q > 0$,*

$$lim_{\upsilon\downarrow 0}MC\left(q,\upsilon\right)=0, \quad and\ lim_{\upsilon\uparrow\infty}MC\left(q,\upsilon\right)=\infty.$$

Formally, we prove the following proposition:

Proposition 2 *Suppose Assumptions 1, 2, 3′, 4, 5, 6 and Assumptions 9, 12 and 13 are satisfied. Consider data on firms with the same product characteristics $\mathbf{x}$ and input prices $\mathbf{w}$.*

a. *Given q, the ordering of marginal revenue is nonparametrically identified from the cost data.*

b. *Suppose we have two points, (Q_1, q_1, p_1, s_1) and (Q_2, q_2, p_2, s_2), with the same demand shocks $(\xi_1 = \xi_2 = \xi)$ and cost shocks $(\upsilon_1 = \upsilon_2 = \upsilon)$ and different market sizes $Q_1 < Q_2$. It follows* *that*

$$s_1 > s_2,\ p_1 < p_2,\ q_1 < q_2, \tag{16}$$

and

$$p_1\left[1+\frac{lnp_2-lnp_1}{lns_2-lns_1}\right]=\frac{E\left[C^d|\left(q_2,p_2,s_2\right)\right]-E\left[C^d|\left(q_1,p_1,s_1\right)\right]}{q_2-q_1}+O\left(|Q_2-Q_1|\right). \tag{17}$$

c. *Suppose we have two close points,* (Q_1, q_1, p_1, s_1) *and* (Q_2, q_2, p_2, s_2)*, such that both (16) and (17) hold. Then, the true marginal cost at at* (Q_1, q_1, p_1, s_1)*,* MC_1 *satisfies*

$$MC_1=\frac{E\left[C^d|\left(q_2,p_2,s_2\right)\right]-E\left[C^d|\left(q_1,p_1,s_1\right)\right]}{q_2-q_1}+O\left(|Q_2-Q_1|\right)$$

Part a of the Proposition clarifies the source of identification in parametric demand models discussed earlier. The parameters of the logit and random coefficient logit models are identified from the ordering of the marginal revenues, which in turn is identified by the ordering of the nonparametrically derived average cost conditional on price, market share, output, observed product characteristics, and input prices.

Parts b and c of the Proposition go further in terms of identification; they state that the level of marginal revenue can be identified. Hence, linear marginal revenue models are now identifiable, which was not the case if we only had the ordering of the marginal revenue as the source of identification. Parts b and c say that if we find two nearby points with the same $\mathbf{x}_m$ and $\mathbf{w}_m$, satisfying some inequalities relating their market shares, prices and outputs, and if the first order condition using these points is approximately satisfied, then a nonparametric estimate of marginal cost can be computed from these points as the local slope of the average cost, where the average is taken over the total cost conditional on output, input price, observed product characteristics, prices, and market shares:

$$\widehat{MC}_1 = \frac{E\left[C^d | \left(q_2, \mathbf{w}, p_2, s_2, \mathbf{x}\right)\right] - E\left[C^d | \left(q_1, \mathbf{w}, p_1, s_1, \mathbf{x}\right)\right]}{q_2 - q_1}.$$

In practice $E\left[C^d | \left(q, \mathbf{w}, p, s, \mathbf{x}\right)\right]$ could be nonparametrically estimated in a first step. Assuming profit maximization, we can directly obtain a nonparametric marginal revenue estimate $\widehat{MR}_1$ from this marginal cost estimate, $\widehat{MR}_1 = \widehat{MC}_1$.[32]

It is fairly straightforward to see that Assumption 12 is satisfied for the logit model. For the random coefficient logit model, we conducted an extensive numerical analysis in monopoly markets and found that when market share is low (e.g., less than or equal to 15 percent of the market size), sometimes marginal revenue decreases with an increase in price. Even though this is an exceptional case of a monopolist having a very low market share, it shows that one cannot completely rule out the possibility of Assumption 12 not being satisfied. Fortunately, it can be tested. To do so, consider two monopoly firms whose output, market size, and market shares are close to each other. In particular, for the point (Q_1, q_1, p_1, s_1), take another close point (Q_2, q_2, p_2, s_2) that has the same $\mathbf{x}_1 = \mathbf{x}_2 = \mathbf{x}$ and $\mathbf{w}_1 = \mathbf{w}_2 = \mathbf{w}$, and that satisfies $Q_1 = Q_2$, $s_1 = s_2 = s$, hence $q_1 = q_2$, but $p_1 < p_2$. Then, if $E\left(C^d | q_2, \mathbf{w}, p_2, s_2, \mathbf{x}\right) > E\left(C^d | q_1, \mathbf{w}, p_1, s_1, \mathbf{x}\right)$, it implies that $C(q_2, \upsilon_2) > C(q_1, \upsilon_1)$, thus, $\upsilon_2 > \upsilon_1$, and given $q_1 = q_2$, $MR_1 = MC(q_1, \upsilon_1) < MC(q_2, \upsilon_2) = MR_2$, and Assumption 12 holds. If, on the other hand, $E\left(C^d | q_2, \mathbf{w}, p_2, s_2, \mathbf{x}\right) \leq E\left(C^d | q_1, \mathbf{w}, p_1, s_1, \mathbf{x}\right)$, then $MR_1 = MC(q_1, \upsilon_1) \geq MC(q_2, \upsilon_2) = MR_2$ and Assumption 12 does not hold. Therefore, by testing the hypothesis $E\left(C^d | q_2, \mathbf{w}, p_2, s_2, \mathbf{x}\right) > E\left(C^d | q_1, \mathbf{w}, p_1, s_1, \mathbf{x}\right)$, one can test Assumption 12.

We next consider oligopoly models with J firms. We apply the same argument to firm $j = 1$

[32] Recall that with parametric identification, we only needed to assume that marginal revenue was an increasing function of marginal cost. For Proposition 2 however, to prove nonparametric identification, we require that marginal revenue to equal marginal cost. It is the parametric functional form restriction that helped weaken the profit maximization assumption previously.

in two different markets. Define $\mathbf{s}_{-jm}$ and $\mathbf{p}_{-jm}$ to be the market share and price vectors of firms other than firm j in market m. As in Proposition 2, we need to find two close points in the data, e.g. two oligopoly outcomes in different markets that are similar. We denote the first outcome to be $(Q_1, q_{11}, p_{11}, s_{11}, \mathbf{s}_{-11}, \mathbf{p}_{-11})$[33] where q_{11}, p_{11}, and s_{11} are the quantity, price and market share of the firm 1 in market 1, and $\mathbf{s}_{-11}$, $\mathbf{p}_{-11}$ are market share and price vectors of firms other than 1 in market 1. We similarly define $(Q_2, q_{12}, p_{12}, s_{12}, \mathbf{s}_{-12}, \mathbf{p}_{-12})$ to be the vector of variables of market 2. These two points are chosen to satisfy the following properties

$$Q_1 < Q_2, \quad s_{11} > s_{12}, \quad p_{11} < p_{12}, \quad s_{11}Q_1 < s_{12}Q_2 \; and \; \mathbf{p}_{-11} = \mathbf{p}_{-12},$$

and

$$p_{11}\left[1 + \frac{lnp_{12} - lnp_{11}}{lns_{12} - lns_{11}}\right] = \frac{E\left[C^d | (q_{12}, \mathbf{p}_2, \mathbf{s}_2)\right] - E\left[C^d | (q_{11}, \mathbf{p}_1, \mathbf{s}_1)\right]}{q_{12} - q_{11}} + O\left(|Q_2 - Q_1|\right).$$

Then, with only slight modifications to the proof of Proposition 2 for the monopoly case, we can prove nonparametric identification of the marginal revenue function for the oligopoly case.[34]

3.4.1 Recovering the market share function

We can use this marginal revenue estimate to recover a non-parametric estimate of the market share function. Denote the non-parametric marginal revenue estimate of firm 1 evaluated at point $(\mathbf{X}_m, \mathbf{p}_m, \mathbf{s}_m)$ by $\widehat{MR}(\mathbf{X}_m, \mathbf{p}_m, \mathbf{s}_m, 1)$.[35] Using the definition of marginal revenue, we can

[33] We again suppress $\mathbf{X}_m$ and $\mathbf{w}_{jm}$ as we continue to condition on having firms that also have the same $\mathbf{X}_m$ and $\mathbf{w}_{jm}$ for these identification proofs.

[34] The relevant lemma, Lemma 5 and its proof are in Subsection C.2 of the Appendix.

[35] Notice that by conditioning on the point $\mathbf{X}_m$, $\mathbf{p}_m$, $\mathbf{s}_m$, we are effectively conditioning on the demand shock $\boldsymbol{\xi}_m$ as well since (recall) it perfectly rationalizes $\mathbf{s}_m$ given $\mathbf{X}_m$ and $\mathbf{p}_m$.

recover the derivative of the market share function at this point as

$$\frac{\partial s(\mathbf{X}_m, \mathbf{p}_m, \mathbf{s}_m, 1)}{\partial p_{1m}} = \left[\frac{MR(\mathbf{X}_m, \mathbf{p}_m, \mathbf{s}_m, 1) - p_{1m}}{s_{1m}}\right]^{-1}.$$

A non-parametric estimate of the market share derivative around the point $\mathbf{X}_m, \mathbf{p}_m, \mathbf{s}_m$ for firm 1 can then be calculated as

$$\frac{\widehat{\partial s(\mathbf{X}, \mathbf{p}, \mathbf{s}}, 1)}{\partial p_1} = \sum_m \left[\frac{\widehat{MR}(\mathbf{X}_m, \mathbf{p}_m, \mathbf{s}_m, 1) - p_{1m}}{s_{1m}}\right]^{-1} \frac{K_{\mathbf{h}}\left(\mathbf{X} - \mathbf{X}_m, \mathbf{p} - \mathbf{p}_m, \mathbf{s} - \mathbf{s}_m\right)}{\sum_n K_{\mathbf{h}}\left(\mathbf{X} - \mathbf{X}_n, \mathbf{p} - \mathbf{p}_n, \mathbf{s} - \mathbf{s}_n\right)},$$

where $K_{\mathbf{h}}(\cdot)$ is a kernel with bandwidth vector $\mathbf{h}$.

We can use this non-parametric estimate of the market share derivative to recover a non-parametric estimate of the demand function. Starting from the point $\bar{\mathbf{X}}, \bar{\mathbf{p}}, \bar{\mathbf{s}}$ (where $\bar{\mathbf{s}} = \mathbf{s}\left(\overline{\mathbf{X}}, \bar{\mathbf{p}}, \boldsymbol{\xi}, 1\right)$ for some $\boldsymbol{\xi}$), we derive the approximation of $\mathbf{s}\left(\overline{\mathbf{X}}, \bar{\mathbf{p}} + \Delta\mathbf{p}, \boldsymbol{\xi}, 1\right)$, that is, the market share of firm 1 with price vector $\bar{\mathbf{p}} + \Delta\mathbf{p}$ where $\Delta\mathbf{p} = [\Delta p_{1m}, 0, \ldots, 0]'$ and where Δp_{1m} is small. The approximation is computed as

$$\hat{s}(\overline{\mathbf{X}}, \bar{\mathbf{p}} + \Delta\mathbf{p}, \boldsymbol{\xi}, 1) = \bar{s} + \frac{\widehat{\partial s(\overline{\mathbf{X}}, \bar{\mathbf{p}}, \bar{\mathbf{s}}}, 1)}{\partial \mathbf{p}}' \Delta\mathbf{p},$$

where $\frac{\widehat{\partial s(\overline{\mathbf{X}}, \bar{\mathbf{p}}, \bar{\mathbf{s}}}, 1)}{\partial \mathbf{p}} = \left[\frac{\widehat{\partial s(\mathbf{X}, \mathbf{p}, \mathbf{s}}, 1)}{\partial p_1}, 0, \ldots, 0\right]'$. The market share function can be iteratively recovered in a similar fashion, where at iteration k the share estimate at price $\bar{\mathbf{p}} + k\Delta\mathbf{p}$ is

$$\hat{s}\left(\overline{\mathbf{X}}, \bar{\mathbf{p}} + k\Delta\mathbf{p}, \boldsymbol{\xi}, 1\right) = \hat{s}\left(\overline{\mathbf{X}}, \bar{\mathbf{p}} + (k-1)\Delta\mathbf{p}, \boldsymbol{\xi}, 1\right) + \frac{\partial s\left(\overline{\mathbf{X}}, \bar{\mathbf{p}} + \widehat{(k-1)}\Delta\mathbf{p}, \bar{\mathbf{s}}, 1\right)}{\partial \mathbf{p}}' \Delta\mathbf{p}.$$

Then,

$$\hat{s}\left(\overline{\mathbf{X}}, \bar{\mathbf{p}} + k\Delta\mathbf{p}, \boldsymbol{\xi}_m, 1\right) = s\left(\overline{\mathbf{X}}, \bar{\mathbf{p}} + k\Delta\mathbf{p}, \boldsymbol{\xi}, 1\right)$$
$$+ \sum_{l=1}^{k} \left[\left(\frac{\partial s\left(\overline{\mathbf{X}}, \widehat{\bar{\mathbf{p}} + l\Delta}\mathbf{p}, \boldsymbol{\xi}, 1\right)}{\partial \mathbf{p}} - \frac{\partial s\left(\overline{\mathbf{X}}, \mathbf{p} + l\Delta\mathbf{p}, \boldsymbol{\xi}, 1\right)}{\partial \mathbf{p}} \right)' \Delta\mathbf{p} + O\left(\|\Delta\mathbf{p}\|^2\right) \right].$$

Therefore,

$$\hat{s}\left(\overline{\mathbf{X}}, \bar{\mathbf{p}} + k\Delta\mathbf{p}, \boldsymbol{\xi}, 1\right) = s\left(\overline{\mathbf{X}}, \bar{\mathbf{p}} + k\Delta\mathbf{p}, \boldsymbol{\xi}, 1\right) + O\left(k\|\Delta\mathbf{p}\|^2\right) + k o_p\left(1\right)\|\Delta\mathbf{p}\|.$$

Hence, we can obtain a non-parametric market share function estimate given $\overline{\mathbf{X}}$ and $\mathbf{p}$.

3.4.2 Curse of Dimensionality

In practice, a non-parametric estimator for the demand and cost parameters based on parts b and c of Proposition 2 will likely suffer from a Curse of Dimensionality. To implement such an estimator, one would need to obtain a non-parametric estimate of $E[C^d_{jm}|(q_{jm}, \mathbf{w}_{jm}, \mathbf{X}_m, \mathbf{p}_m, \mathbf{s}_m)]$. For most markets of interest, $\mathbf{X}_m$ will contain a number of product characteristics across a non-negligible number of firms. This makes the dimensionality problem potentially quite severe.

Because of this dimensionality issue, in what follows we pursue the common practice where researchers use parametric restrictions to reduce the dimensionality of the estimation problem, essentially transforming the non-parametric estimation exercise into a semi-parametric one. In particular, we adopt the Berry (1994) logit or BLP (1995) random coefficients demand model. This relaxes the need to condition on the individual variables $\mathbf{X}_m, \mathbf{p}_m, \mathbf{s}_m$ in our pseudo-cost function estimator: we only need to control for a single MR_{jm} index, which is a parametric function of these variables.

3.5 Cost function estimation

After estimating the model's parameters with our pseudo-cost estimator in (14), we can recover the cost function from the marginal revenue estimates; the steps in doing so are similar to those used in recovering of the market share function from marginal revenue.[36]

The cost function can be recovered from the pseudo-cost function estimates in two steps. First, we non-parametrically estimate marginal cost for a given point $(q, \mathbf{w}, C)$ as follows,

$$\widehat{MC}\left(q, \mathbf{w}, C\right) = \sum_{jm} MR_{jm}\left(\boldsymbol{\theta}_M\right) W_h\left(q - q_{jm}, \mathbf{w} - \mathbf{w}_{jm}, C - \widehat{PC}\left(q_{jm}, \mathbf{w}_{jm}, MR_{jm}\left(\boldsymbol{\theta}_M\right), \boldsymbol{\gamma}_M\right)\right)$$

where $\boldsymbol{\theta}_M$ is the estimated demand parameter obtained from our pseudo-cost estimator, $\widehat{PC}(\cdot)$ is the predicted value of the estimated pseudo-cost function, and W_h is a kernel-based weight function.[37] Second, for a given input price $\mathbf{w}$, starting at output $\bar{q}$ and total cost $\bar{C}$, there exists a cost shock $\overline{v}$ that corresponds to $MC\left(\bar{q}, \mathbf{w}, \bar{v}\right) = \overline{MR}$.[38] Knowing this, we can use the following iteration for $k = 1, \ldots$ to recover the total cost for different levels of output given the cost shock $\bar{v}$,

$$\widehat{C}\left(\bar{q} + k\Delta q, \mathbf{w}, \bar{v}\right) = \widehat{C}\left(\bar{q} + (k-1)\Delta q, \mathbf{w}, \bar{v}\right) + \widehat{MC}\left(\bar{q} + (k-1)\Delta q, \mathbf{w}, \widehat{C}\left(\bar{q} + (k-1)\Delta q, \mathbf{w}, \bar{v}\right)\right)\Delta q.$$

where Δq represents a small change in quantity.[39] It is important to note that this procedure does not impose any constraints on the cost function. The additional source of information for recovering the cost function comes from the demand side of the model.

[36] Marginal revenue function can be parametric as well as nonparametric.

[37]
$$W_h\left(q - q_{jm}, \mathbf{w} - \mathbf{w}_{jm}, C - \widehat{PC}_{jm}\right) = \frac{K_{h_q}\left(q - q_{jm}\right) K_{h_W}\left(\mathbf{w} - \mathbf{w}_{jm}\right) K_{h_{MR}}\left(C - \widehat{PC}_{jm}\right)}{\sum_{kl} K_{h_q}\left(q - q_{kl}\right) K_{h_W}\left(\mathbf{w} - \mathbf{w}_{kl}\right) K_{h_{MR}}\left(C - \widehat{PC}_{kl}\right)}.$$

[38] We do not need to derive the value of $\overline{v}$, only the corresponding $\overline{MR}$.

[39] In Appendix D we provide detailed instructions on how to implement this iterative procedure and recover the cost function for a given $\mathbf{w}$, starting from a given set of $\bar{q}$, $\bar{C}$ and $\overline{MR}$ values.

3.6 Further issues

We have thus far worked with the standard differentiated products model from Berry (1994) and BLP. Depending on the empirical context, however, a number a specification and data-related issues can potentially arise. In this section, we demonstrate that with some modifications of the NLLS part of the objective function in (14) our estimator can be adapted to various empirical settings.

Endogenous Product Characteristics.

So far, we have followed the literature and assumed $\mathbf{X}_m$ to be exogenous. However, if firms strategically choose prices *and* product characteristics, then elements of $\mathbf{X}_m$ will be correlated with the demand shock $\boldsymbol{\xi}_m$. Researchers often abstract from this possibility by assuming they are studying a sufficiently short time horizon such that firms effectively take their product lines as fixed and compete strictly on prices. Over longer time horizons, this assumption likely breaks down in most markets. To accommodate endogenous product characteristics, researchers have recently started estimating BLP models that include first order conditions for optimal prices and product characteristics.[40] To estimate the demand parameters in this setting, one needs instruments to deal with the endogeneity of prices and product characteristics. In addition, the typical set of "BLP instruments" for prices based on product characteristics become invalid. That is, researchers need more instruments and have fewer options for IVs when trying to estimate differentiated products models with endogenous product characteristics

In our framework, we can explicitly incorporate endogeneity of product characteristics by adding the marginal revenue with respect to product characteristics choice in the pseudo-cost

[40]See, for example, Chu (2010), Fan (2013), and Byrne (2014). For an excellent overview of the empirical literature on endogenous product characteristics see Crawford (2012). It is worth noting that these applications all maintain the static decision-making assumption of BLP; firms are allowed to adjust their product characteristics period-by-period but are not forward-looking in doing so. A recent paper by Gowrisankaran and Rysman (2012) develops and estimates a dynamic version of a differentiated products oligopoly model, though the solution and estimation methods are computationally burdensome.

function as follows,

$$PC\left(q_{jm}, \mathbf{w}_{jm}, MR_{q_{jm}}\left(\boldsymbol{\theta}_0\right), MR_{\mathbf{X}_{jm}}\left(\boldsymbol{\theta}_0\right)\right),$$

where $MR_{q_{jm}}$ is the marginal revenue with respect to quantity choice and $MR_{\mathbf{X}_{jm}}$ is the marginal revenue with respect to the product characteristics choice. Then, the modified NLLS part would be

$$\frac{1}{\sum_m J_m}\sum_{j,m}\left[C^d_{jm} - \sum_l \gamma_l \psi_l\left(q_{jm}, \mathbf{w}_{jm}, MR_{q_{jm}}\left(\boldsymbol{\theta}\right), MR_{\mathbf{X}_{jm}}\left(\boldsymbol{\theta}\right)\right)\right]^2.$$

Cost Function Restrictions.

We have thus far not imposed any assumptions about the shape of the pseudo-cost function except that it is a smooth function of output, input price, and marginal revenue. Hence, the cost function that is recovered is not guaranteed to have properties such as homogeneity of degree one in input prices, nor convexity in output, which are required for the cost function to be well-defined.

Imposing the restriction of homogeneity in input prices in estimation is straightforward. If the cost function is homogenous of degree one with respect to input price, so is the marginal cost function. Hence for some input price $w_{1,jm}$,

$$C(q_{jm}, \mathbf{w}_{jm}, \upsilon_{jm}) = w_{1,jm}C(q_{jm}, \frac{\mathbf{w}_{jm}}{w_{1,jm}}, \upsilon_{jm})$$

and

$$MC(q_{jm}, \mathbf{w}_{jm}, \upsilon_{jm}) = w_{1,jm}\frac{\partial C(q_{jm}, \mathbf{w}_{jm}/w_{1,jm}, \upsilon_{jm})}{\partial q}.$$

We can thus modify the NLLS component of our pseudo-cost estimator to impose the homogeneity restriction as follows,

$$\frac{1}{\sum_m J_m} \sum_{j,m} \left[\frac{C^d_{jm}}{w_{1,jm}} - \sum_l \gamma_l \psi_l \left(q_{jm}, \frac{\mathbf{w}_{-1,jm}}{w_{1,jm}}, \frac{MR_{jm}(\boldsymbol{\theta})}{w_{1,jm}} \right) \right]^2, \tag{18}$$

where $\mathbf{w}_{-1,jm} = (w_{2,jm}, \ldots, w_{L,jm})$ are all the other input prices except for input price $w_{1,jm}$.

We do not, however, impose convexity in output in estimation. Rather, we follow numerous papers in the cost function estimation literature and check that convexity is satisfied after estimating the cost function.

Economic versus Accounting Cost

The cost data we envision using comes from the accounting statements of the firm.[41] Such data do not necessarily reflect the economic cost that the firm considers in making input and output choices. More concretely, by imposing profit maximization, we may not be appropriately taking into account the opportunity cost of the resources that are used in purchasing the necessary input to produce output. Fortunately, from accounting statements we may be able to obtain information on other activities that the firm may be pursuing in addition to the production of an output. For example, we may find details on firms' financial investments including their rate of return.[42] Suppose that the return on a unit of a financial investment is r_{jm}. Then, the opportunity cost of production is r_{jm} and the firm will produce and sell output until marginal revenue equals marginal cost that accounts for this cost,

$$MR_{jm}(\boldsymbol{\theta}) = MC(q_{jm}, \mathbf{w}_{jm}, \upsilon_{jm}) + r_{jm}.$$

[41] Indeed, accounting data are typically used in previous applications that estimate cost functions to evaluate market power, measure economies of scale or scope, and so on.

[42] In the application of our estimator to U.S. banking such information is readily available. Most large industries like banking that are subject to some form of regulatory oversight are likely to report such data. The cable TV industry is another good example; see, for example, the data described in Kelly and Ying (2003) or Byrne (2014).

Substituting this into our estimator, we obtain the modified NLLS part as follows:

$$\frac{1}{\sum_m J_m}\sum_{j,m}\left[C^d_{jm}-\sum_l \gamma_l\psi_l\left(q_{jm},\mathbf{w}_{jm},MR_{jm}\left(\boldsymbol{\theta}\right)-r_{jm}\right)\right]^2.$$

That is, as long as we can obtain information on the financial opportunities that the firm has other than production, we can incorporate them into our estimator. In such cases, the estimator will not be subject to bias even if the cost data we use corresponds to accounting costs.

Fixed costs

So far we have implicitly assumed that the cost data corresponds to variable costs. For the more general case where only total cost is given, our method can still be applied if we impose some additional assumptions on the firm's cost structure. For example, suppose there are fixed costs that correspond to rental payments, licensee fees, and so on that vary with variables in $\mathbf{x}_{jmf}$, but that do not vary with q_{jm}, $\mathbf{w}_{jm}$ nor MR_{jm}. We can modify the NLLS part of our estimator to account for such fixed costs as follows,

$$\frac{1}{M}\sum_{jm}\left[C^d_{jm}-\sum_l \gamma_l\psi_l\left(q_{jm},\mathbf{w}_{jm},MR_{jm}(\boldsymbol{\theta})\right)-\mathbf{x}_{jmf}\boldsymbol{\zeta}\right]^2,$$

where $\boldsymbol{\zeta}$ is an additional fixed cost parameter to be estimated.

Missing cost data and multi product firms

Until now we have assumed cost data are available for all firms in the sample. In the data, it could very well be the case that we only observe costs for some firms and not others. In that case, we can estimate the structural parameters consistently by constructing the NLLS part using only firms for which we have cost data. Because the NLLS part of our estimator does not involve any instruments, only choosing firms with available cost data in estimation will not result in selection bias. It is important to notice, however, that we still need demand-side data for all firms in the same market to compute the marginal revenue and the GMM part of the objective function. Luckily, such demand-side data tends to be available to researchers for many industries.[43]

A more difficult case of unobservable costs would be when firms produce multiple products, but only the total cost across all products is observable in the data.[44] Suppose that each firm produces F outputs. Then, as long as the numbers of products is not too large (otherwise, we would face a Curse of Dimensionality issues in estimation), the NLLS component can be extended as follows,

$$\frac{1}{M}\sum_{jm}\left[C_{jm}^{d}-\sum_{l}\gamma_{l}\psi_{l}\left(\mathbf{q}_{jm,1:F},\mathbf{w}_{jm},\mathbf{MR}_{j1:F}\left(\mathbf{X}_{m1:F},\mathbf{p}_{m1:F},\mathbf{s}_{m1:F},\boldsymbol{\theta}\right)\right)\right]^{2},$$

where $\mathbf{q}_{m1:F}=(q_{m1},...,q_{mF})$ is the vector of output of product 1 to product F. $\mathbf{X}_{m1:F}$, $\mathbf{p}_{m1:F}$ and $\mathbf{s}_{m1:F}$ are similarly defined.

If the number of products is large, one should consider imposing more structure on the pseudo-cost function to avoid the Curse of Dimensionality.[45] Such a cost function could be specified as:

$$C_{f}^{d}=\sum_{jm}C\left(q_{jm},\mathbf{w}_{jm},\upsilon_{jm},\boldsymbol{\tau}\right)I_{jm}\left(f\right)+\eta_{f}=\sum_{jm}PC\left(q_{jm},\mathbf{w}_{jm},MR\left(\mathbf{X}_{m},\mathbf{p}_{m},\mathbf{s}_{m},\boldsymbol{\theta}_{0}\right)\right)I_{jm}\left(f\right)+\eta_{f}$$

where $I_{jm}(f)$ is an indicator function that equals 1 if branch j in market m belongs to firm f and 0 otherwise. C_f is the total cost of the firm that includes the cost of all branches, and η_f is the i.i.d. distributed measurement error of firm f's total cost. Denoting F to be the total

[43] This is evidenced by the large number of applications across many industries in empirical IO the followed Berry (1994) and BLP.

[44] This is practical issue for our U.S. banking application. We have total costs for a given bank in a local market, however we do not know the individual branch-level costs for the bank within a market.

[45] This will likely be the case in our U.S. banking application where banks have potentially many branches in some Metropolitan Statistical Areas.

number of firms in the data, the NLLS component of our estimator can be modified as follows,

$$\frac{1}{F}\sum_{f}\left[C_f^d-\sum_{jm}\sum_{l}\gamma_l\psi_l\left(q_{jm},\mathbf{w}_{jm},MR_{jm}(\boldsymbol{\theta})\right)I_{jm}\left(f\right)\right]^2.$$

4 Large Sample Properties

Our estimator is derived from minimizing the objective function that is the sum of two components. The first NLLS component is sieve based, and the second component is the GMM objective function. In the Appendix, we prove consistency and asymptotic normality of the estimator. These proofs are based on the asymptotic analysis of sieve estimators by Chen (2007) and Bierens (2014), and the GMM asymptotics by Newey and McFadden (1994) and others.

5 Monte Carlo Experiments

This section presents results from some Monte-Carlo experiments that highlight the finite sample performance of our estimator. To generate samples, we use the following random coefficients logit demand model:

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

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

On the supply-side, we assume firms compete on prices ala differentiated Bertrand taking product characteristics as exogenously given, use labor and capital inputs in production, and have Cobb-Douglas production functions. We further assume input prices to be the same for all firms in a given market; the assumption is motivated by the practical reality that researchers typically only have access to market-level aggregate cost data. Then, given output, input prices $\mathbf{w} = [w\ r]'$ (where w is the wage and r is the rental rate of capital) and a productivity shock, total cost and marginal cost functions are specified as,

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

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

Notice that in the above specification the cost function is homogenous of degree 1 in input prices.[46]

To create our Monte-Carlo samples, we generate wages, rental rates, cost shocks υ, market size Q, the idiosyncratic component of the demand shock ϱ_ξ, and observable product characteristics $\mathbf{x}$ as follows,

$$w \sim TN(\mu_w, \sigma_w), \quad e.g., w = \mu_w + \varrho_w, \quad \varrho_w \sim TN(0, \sigma_w)$$

$$r \sim TN(\mu_r, \sigma_r), \quad e.g., r = \mu_r + \varrho_r, \quad \varrho_r \sim TN(0, \sigma_r)$$

$$\upsilon \sim TN(\mu_\upsilon, \sigma_\upsilon), \quad e.g., \upsilon = \mu_\upsilon + \varrho_\upsilon, \quad \varrho_\upsilon \sim TN(0, \sigma_\upsilon)$$

$$Q \sim U(Q_L, Q_H), \quad \varrho_\xi \sim TN(0, 1), \quad \mathbf{x} \sim TN(\mu_{\mathbf{x}}, \sigma_{\mathbf{x}}).$$

We draw variables from the truncated normal distribution $TN(\cdot)$ to ensure that the true cost

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

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

function is positive and bounded, and the compactness of the set $\mathcal{W}$ (which recall contains $(q, \mathbf{w}, MR)$). We truncate both upper and lower 0.82 percentiles. We further assume market size to be uniformly distributed with lower bound Q_L and upper bound Q_H.

Importantly, we specify the unobserved quality so as to allow for correlation between ξ and input price, cost shock and market size. Specifically, we set:

$$\xi = \delta_0 + \delta_1 \varrho_\xi + \delta_2 \varrho_w + \delta_3 \varrho_r + \delta_4 \varrho_\upsilon + \delta_5 \Phi^{-1}\left(\frac{Q - Q_L}{Q_H - Q_L}\right)$$

where Φ is the cumulative distribution function of the standard normal distribution. We set $\delta_l > 0$ for $l = 1, ..., 5$. Hence, by construction, no variable can be used as a valid instrument for prices in demand estimation.

To solve for the equilibrium price, quantity, and market share for each monopolist, we use golden section search on price.[47] In our Monte-Carlo experiments, we explicitly solve for the equilibrium price, market share, and quantity. Therefore, in our case, the impact of instruments

[47]For the oligopoly market, we compute the equilibrium for each market m as follows.

Step 1 : We generate Q_m, $\mathbf{X}_m$, w_m, r_m, $\boldsymbol{\upsilon}_m$ and $\boldsymbol{\xi}_m$ based on the above specification.

Step 2 : For firm j in market m, given other firms' price, market share and output $\mathbf{p}_{-jm}$, $\mathbf{s}_{-jm}$, we solve for the optimal price p_{jm}, market share s_{jm}, and output q_{jm} by using the F.O.C. of profit maximization.

$$MR_{jm} = p_{mj} + \left[\frac{\partial s\left(\mathbf{p}_m, \mathbf{X}_m, \boldsymbol{\xi}_m, j, \boldsymbol{\theta}\right)}{\partial p_{jm}}\right]^{-1} s_{jm} = MC_{mj} = \frac{\partial C\left(q_{jm}, w_m, r_m, \upsilon_{jm}\right)}{\partial q_{jm}}$$

where $q_{jm} = Q_m s_{jm}$. We do so by first bracketing p_{jm}, e.g., finding the interval $p_{jm} \in \left(\underline{p}, \overline{p}\right)$ so that $\underline{p}$, and the corresponding $\underline{s}$, $\underline{q}$ satisfies

$$\underline{p} + \left[\frac{\partial s\left(\underline{p}, \mathbf{p}_{-jm}, \mathbf{X}_m, \boldsymbol{\xi}_m, j, \boldsymbol{\theta}\right)}{\partial p}\right]^{-1} \underline{s} < \frac{\partial C\left(\underline{q}, w_m, r_m, \upsilon_{jm}\right)}{\partial q}$$

A similar procedure for $\overline{p}$, $\overline{s}$, and $\overline{q}$ yields

$$\overline{p} + \left[\frac{\partial s\left(\overline{p}, \mathbf{p}_{-jm}, \mathbf{X}_m, \boldsymbol{\xi}_m, j, \boldsymbol{\theta}\right)}{\partial p}\right]^{-1} \overline{s} > \frac{\partial C\left(\overline{q}, w_m, r_m, \upsilon_{jm}\right)}{\partial q}.$$

Then, we use the bisection method to compute the equilibrium price $p_{jm} \in \left(\underline{p}, \overline{p}\right)$ that satisfies

$$p_{jm} + \left[\frac{\partial s\left(\mathbf{p}_m, \mathbf{X}_m, \boldsymbol{\xi}_m, j, \boldsymbol{\theta}\right)}{\partial p_{jm}}\right]^{-1} s_{jm} = \frac{\partial C\left(q_{mj}, w_m, r_m, \upsilon_{jm}\right)}{\partial q_{jm}}.$$

We repeat the above algorithm for each $j = 1, \ldots, J_m$ until convergence, e.g. the F.O.C. condition for profit maximization being satisfied for each firm in market m.

Table 1: Monte-Carlo Parameter Values

μ_α	σ_α	μ_β	σ_β	μ_X	σ_X	α_c	β_c	μ_w	σ_w
2.0	0.5	1.0	0.2	1.0	0.5	0.4	0.4	2.0	0.2
μ_r	σ_r	μ_υ	σ_υ	Q_L	Q_H	δ_0	σ_ξ	A	B
2.0	0.2	0.5	0.2	5.0	10.0	4.0	0.5	0.01	1.0

on the endogenous variable is highly nonlinear and heterogeneous.

Table 1 summarizes the parameter setup of the Monte-Carlo experiments. Table 2 presents sample statistics from the corresponding simulated data where the sample size is set to 1000 market-firm observations (e.g., there are 250 local markets). We set the standard deviation of measurement error to be 0.1, about five percent of the total cost. The parameter estimates are obtained by the following minimization algorithm,

$$\begin{aligned}\left[\hat{\boldsymbol{\theta}}_M, \hat{\boldsymbol{\gamma}}_M\right] &= argmin_{(\boldsymbol{\theta},\gamma)\in\Theta\times\Gamma_{k_M}(T)}\Big[\frac{1}{\sum_{m=1}^{M} J_m}\sum_{jm}\left[\frac{C_{jm}^d}{r_m} - \sum_l \gamma_l\psi_l\left(q_{jm}, \frac{w_m}{r_m}, \frac{MR_{jm}(\boldsymbol{\theta})}{r_m}\right)\right]^2 \\ &+A\left[\frac{\sum_{m=1}^{M}\sum_{j=1}^{J_m}\hat{\boldsymbol{\xi}}_{jm}(\boldsymbol{\theta})\mathbf{X}_m}{\sum_{m=1}^{M}\sum_{j=1}^{J_m}1}\right]'\mathbf{W}_M\left[\frac{\sum_{m=1}^{M}\sum_{j=1}^{J_m}\hat{\boldsymbol{\xi}}_{jm}(\boldsymbol{\theta})\mathbf{X}_m}{\sum_{m=1}^{M}\sum_{j=1}^{J_m}1}\right],\end{aligned}$$

where note that in estimation we restrict the cost function to be homogenous of degree one in input price. Further, we set the weighting matrix $\mathbf{W}_M$ to be

$$\mathbf{W}_M = \left[\frac{\sum_{m=1}^{M}\sum_{j=1}^{J_m}\left(\hat{\boldsymbol{\xi}}_{jm}(\boldsymbol{\theta})\mathbf{X}_m\right)\left(\hat{\boldsymbol{\xi}}_{jm}(\boldsymbol{\theta})\mathbf{X}_m\right)'}{\sum_{m=1}^{M}\sum_{i=1}^{J_m}1}\right]^{-1}$$

We adopt the continuously updating GMM approach and estimate the weighting matrix $\mathbf{W}_M$ simultaneously with the estimation of the parameters.

In Table 3, we present the Monte-Carlo results for our NLLS-GMM estimator. We report the mean, standard deviation, and square root of the mean squared errors (RMSE) from 100 Monte-Carlo simulation/estimation replications. From the table, we see that as sample size increases, the standard deviation and the RMSE of the parameter estimates decrease. This

Table 2: Sample Statistics of Simulated Data.

variables	Mean	Std. Dev
Price (p_m)	3.3532	0.8352
Output (q_m)	1.1865	0.9068
Quality (ξ_m)	3.9932	0.4385
Market Share (s_m)	0.1578	0.1104
Wage (w_m)	2.0011	0.1961
Rent (r_m)	1.9842	0.1781
Cost (C_m)	1.7935	0.7600
x_m	0.9791	0.4636

Measurement error std. dev.: $\sigma_\eta = 0.1$

highlights the consistency of our estimator. It is noteworthy that the means of the estimates are quite close to their true values even with a small sample size of 100. Furthermore, since the estimated parameter values are very close to their true values, the standard deviations and RMSEs are very close to each other as well. Overall, these Monte-Carlo results demonstrate the validity of our approach.[48]

In Table 4, we report and additional set of Monte-Carlo results where we estimate $\mu_\alpha, \sigma_\alpha$, and σ_β by minimizing the NLLS objective function, whereas μ_β is estimated by minimizing the GMM objective function. Overall, the means of the parameter estimates are again close to their true values, and the standard deviations and RMSEs continue to decrease with sample size. Moreover, the standard deviations and RMSEs tend to be larger than those of the NLLS-GMM estimates from Table 3, with the following exceptions: standard deviation and RMSE of $\hat{\mu}_\alpha$ are lower on average than those of the NLLS-GMM estimator for a sample size of 100, and standard deviation and RMSE of $\hat{\mu}_\beta$ are lower on average than those of the NLLS-GMM estimator for a sample size of 1,000. We conclude that the NLLS component of the estimator is sufficient for the estimation of $\mu_\alpha, \sigma_\alpha$, and σ_β. However, the additional GMM component in the NLLS-GMM is effective in improving efficiency, in particular the efficiency of $\hat{\sigma}_\alpha$ and $\hat{\sigma}_\beta$, coefficients that determine the degree of heterogeneity and price elasticity in random coefficient models.

[48]Results with measurement error standard deviations larger than 0.1 are similar to the one presented, but with larger standard deviations and RMSEs.

Table 3: NLLS-GMM Estimator of Random Coefficient Demand Parameters.

			$\hat{\mu}_\alpha$			$\hat{\sigma}_\alpha$			
Market Size	Sample Size	No. Poly	Mean	Std. Dev	RMSE	Mean	Std. Dev	RMSE	
25	100	27	-2.0264	0.5345	0.5325	0.4702	0.1271	0.1299	
50	200	32	-2.0104	0.2628	0.2617	0.4978	0.0626	0.0623	
100	400	38	-1.9972	0.1271	0.1265	0.4987	0.0369	0.0367	
250	1000	48	-1.9986	0.0672	0.0668	0.4979	0.0194	0.0194	
True			-2.0000			0.5			
			$\hat{\mu}_\beta$			$\hat{\sigma}_\beta$			Obj. Fct.
Market Size	Sample Size	No. Poly	Mean	Std. Dev	RMSE	Mean	Std. Dev	RMSE	
25	100	27	0.9892	0.1633	0.1628	0.1939	0.0716	0.0715	1.737D-3
50	200	32	0.9936	0.0864	0.0862	0.1985	0.0529	0.0526	2.029D-3
100	400	38	1.0074	0.0650	0.0653	0.2023	0.0197	0.0198	2.138D-3
250	1000	48	0.9993	0.0374	0.0372	0.2017	0.0108	0.0108	2.194D-3
True			1.0000			0.2			

Measurement error std. deviation: 0.1

Table 4: Two-Step Estimator of Random Coefficient Demand Parameters.

			$\hat{\mu}_\alpha$			$\hat{\sigma}_\alpha$			
Market Size	Sample Size	No. Poly	Mean	Std. Dev	RMSE	Mean	Std. Dev	RMSE	
25	100	27	-2.0077	0.4924	0.4900	0.4572	0.1449	0.1504	
50	200	32	-2.0289	0.2718	0.2719	0.5011	0.0774	0.0770	
100	400	38	-2.0002	0.1492	0.1485	0.5025	0.0407	0.0406	
250	1000	48	-1.9878	0.0874	0.0878	0.4996	0.0237	0.0236	
True			-2.0000			0.5			
			$\hat{\mu}_\beta$			$\hat{\sigma}_\beta$			Obj. Fct.
Market Size	Sample Size	No. Poly	Mean	Std. Dev	RMSE	Mean	Std. Dev	RMSE	
25	100	27	0.9921	0.1695	0.1688	0.1957	0.1178	0.1173	1.546D-3
50	200	32	1.0042	0.1177	0.1172	0.2031	0.0547	0.0545	1.896D-3
100	400	38	1.0103	0.0768	0.0771	0.2044	0.0404	0.0404	2.069D-3
250	1000	48	0.9974	0.0360	0.0359	0.1991	0.0203	0.0203	2.150D-3
True			1.0000			0.2			

Measurement error std. deviation: 0.1

In Table 5, we present Monte-Carlo results where we estimate parameters using the standard IV approach. We use wage, rental rate and market size as instruments. We experienced numerical instability when we used the GMM estimator from BLP for the random coefficient model for this exercise. Since our main focus is on potential bias of the IV estimator, and not numerical issues, we decided to instead use the simpler and numerically more stable logit model. All parameter settings are the same as those from the BLP Monte-Carlo exercise, except for the restriction that $\sigma_\alpha = 0$ and $\sigma_\beta = 0$ and different values for δ_i, $i = 2, \ldots, 5$, which we will discuss

Table 5: IV Estimator for Logit Demand Parameters.

		$\hat{\alpha}$			$\hat{\beta}$		
Sample Size		Mean	Std. Dev	MSE	Mean	Std. Dev	MSE
1000	NLLS-GMM[a]	-1.9963	0.0623	0.0621	1.0015	0.0320	0.0319
1000	IV1[b]	-2.0163	0.1263	0.1267	1.0057	0.0451	0.0452
1000	IV2[c]	-0.8365	0.2276	1.1853	0.6906	0.0789	0.3192
True		-2.0000			1.0		

a: $\delta_2 = \delta_3 = \delta_4 = \delta_5 = \delta_1$, b: $\delta_2 = \delta_3 = \delta_4 = \delta_5 = 0$, c: $\delta_2 = \delta_3 = \delta_4 = \delta_5 = 0.3\delta_1$

in detail later. We also change the notation and use α instead of μ_α and β instead of μ_β to be consistent with Berry (1994).

In the first row of the table, we show results of the NLLS-GMM estimator. We still obtain parameter estimates that are close to their true values. The results in the second row are the ones for the IV estimator where instruments are not correlated with the demand shock, and thus, valid (e.g., where we set $\delta_2 = \delta_3 = \delta_4 = \delta_5 = 0$). We can see that means of the estimated parameters are close to their true values, although the standard deviations are relatively large. This is because of the inefficiency of the simple IV. In the third row, we show results where the instruments are invalid. We first tried the specification of $\delta_2 = \delta_3 = \delta_4 = \delta_5 = \delta_1$, as in the NLLS-GMM case. However, we faced numerical instability during estimation even for the simple IV estimator. We then reduced the degree of endogeneity to $\delta_2 = \delta_3 = \delta_4 = \delta_5 = 0.3\delta_1$ and reported the results from this exercise. We can see that in this case the estimated price coefficient is much higher than the true value of -2.0, e.g., we have an upward bias. The positive direction of bias is reasonable because the error term, which is the unobserved quality, is set up to be positively correlated with the instruments.

6 Conclusion

In this paper, we developed a new methodology for estimating demand and cost parameters of the differentiated goods models when cost data and input prices are available, in addition to the standard data on aggregate market shares and prices of products. Our approach, which exploits cost data and profit maximization of firms can, without instruments, identify demand parameters in the presence of price endogeneity, and the cost function nonparametrically in the presence of output endogeneity. Moreover, we have shown that the marginal revenue function is nonparametrically identified.

As the Monte-Carlo experiments show, our method works well in situations where instruments are correlated with structural unobservables in the model, and thus standard IV based estimation methods break down.

This is the reason why the estimator proposed in this paper could be useful in policy analysis, such as Anti-Trust cases, where governments may prefer to use the detailed cost data, which in some cases it may have the authority to obtain via subpoena, rather than instruments, whose choice could be a source of disagreement among various parties.

We believe that our results have implications for estimation of the models that assume profit maximization by firms. In most empirical literature on firm behavior, researchers have essentially used methods that are similar to the ones used in estimation of optimal behavior of individuals. The assumption behind it is that the researcher cannot observe any data on the objective function of individuals, i.e. their utility. On the other hand, for firms, we can actually observe a measure of the objective function, i.e. their revenue and cost, hence, their profit. Our results show that with data on the objective function, the exogenous variation that is conventionally used in nonparametric identification and estimation of structural parameters is no longer necessary.

Finally, when cost data is available, by comparing the results with cost data and without cost data but with instruments, one could check the validity of various instruments, which would be a useful guide on instrument choice even for industries whose cost data is not available.

References

Ackerberg, D. A., C. Kevin, and F. Garth (2006): "Structural Identification of Production Functions," *Unpublished Manuscript, UCLA Economics Department.*

Angrist, J. D. and J. S. Pischke (2010): "The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics," *Journal of Economic Perspectives*, 24, 3–30.

Arocena, P., S. D. S., and T. Coelli (2012): "Vertical and Horizontal Scope Economies in the Regulated U.S. Electricity Power Supply Industry," *Journal of Industrial Economics*, 60, 434–467.

Bayer, P., F. Ferreira, and R. McMillan (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. and P. Haile (2014): "Identification in Differentiated Products Markets Using Market Level Data," *Econometrica*, 82, 1749–1797.

Berry, S. T., J. Levinsohn, and A. Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63, 841–890.

——— (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. (1997): "Comment on Valuation of New Goods Under Perfect and Imperfect Competition," in *The Economics of New Goods*, ed. by T. Bresnahan and R. Gordon, Chicago: University of Chicago Press, 209–248.

Byrne, D. P. (2014): "Testing Models of Differentiated Products Markets: Consolidation in the Cable TV Industry," *International Economic Review*, forthcoming.

Chen, X. (2007): "Large Sample Sieve Estimation of Semi-Nonparametric Models," in *Handbook of Econometrics*, ed. by J. Heckman and E. Leamer, Oxford: North-Holland, 5549–5632.

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. and W. Troesken (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. and A. Yurukoglu (2012): "The Welfare Effects of Bundling in Multichannel Television Markets," *American Economic Review*, 102, 301–317.

De Loecker, J. (2011): "Product Differentiation, Multiproduct Firms, and Estimating the Impact of Trade Liberalization on Productivity," *Econometrica*, 79, 1407–1451.

Dube, J.-P., J. T. Fox, and C.-L. Su (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.

Genesove, D. and W. P. Mullin (1998): "Testing Static Oligopoly Models: Conduct and Cost in the Sugar Industry, 1890-1914," *RAND Journal of Economics*, 29, 355–377.

Gentzkow, M. and J. M. Shapiro (2010): "What Drives Media Slant? Evidence from U.S. Daily Newspapers," *Econometrica*, 78, 35–71.

Ghandi, A., S. Navarro, and D. Rivers (2014): "On the Identification of Production Functions: How Heterogenous is Productivity?" *Journal of Political Economy*, forthcoming.

Gowrisankaran, G. and M. Rysman (2012): "Dynamics of Consumer Demand for New Durable Goods," *Journal of Political Economy*, 120, 1173–1219.

Hausman, J. (1997): "Valuation of New Goods Under Perfect and Imperfect Competition," in *The Economics of New Goods*, ed. by T. Bresnahan and R. Gordon, Chicago: University of Chicago Press, 209–248.

Houde, J. F. (2012): "Spatial Differentiation and Vertical Mergers in Retail Markets for Gasoline," *American Economic Review*, 102, 2147–2182.

Keane, M. P. (2009): "Structural vs. Atheoretic Approaches to Econometrics," *Journal of Econometrics*, 156, 3–20.

Kelly, M. T. and J. S. Ying (2003): "On Measuring Competitive Viability and Monopoly Power in Cable: An Empirical Cost Approach," *The Review of Economics and Statistics*, 85, 962–970.

Kim, D.-W. and C. R. Knittel (2003): "Biases in Static Oligopoly Models? Evidence from the California Electricity Market," *Journal of Industrial Economics*, 54, 451–470.

Klette, J. T. (1999): "Market Power, Scale Economies and Productivity: Estimates from a Panel of Establishment Data," *Journal of Industrial Economics*, 47, 451–476.

Knittel, C. and K. Metaxoglou (2014): "Estimation of Random-Coefficient Demand Models: Two Empiricists' Perspective," *Review of Economics and Statistics*, 96, 34–59.

Kutlu, L. and R. Sickles (2012): "Estimation of Market Power in the Presence of Firm Leven Inefficiencies," *Journal of Econometrics*, 168, 141–155.

Levinsohn, J. and A. Petrin (2003): "Estimation of Production Functions Using Inputs to Control for Unobservables," *Review of Economic Studies*, 70, 317–341.

McManus, B. (2007): "Nonlinear Pricing in an Oligopoly Market: The Case of Specialty Coffee," *Rand Journal of Economics*, 38, 512–532.

Nevo, A. (2000): "Mergers with Differentiated Products: The Case of the Ready-to-Eat Cereal Industry," *RAND Journal of Economics*, 31, 395–421.

——— (2001): "Measuring Market Power in the Ready-to-Eat Cereal Industry," *Econometrica*, 69, 307–342.

Newey, W. K. and D. L. McFadden (1994): “Large Sample Estimation and Hypothesis Testing,” in *Handbood of Econometrics, Volume 4*, ed. by R. F. Engle and D. L. McFadden, Amsterdam: Elsevier B.V., 2111–2245.

Olley, S. G. and A. Pakes (1996): “The Dynamics of Productivity in the Telecommunications Equipment Industry,” *Econometrica*, 64, 1263–1297.

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

Reynaert, M. and F. Verboven (2014): “Improving the Performance of Random Coefficients Demand Models: The Role of Optimal Instruments,” *Journal of Econometrics*, 179, 83–98.

Roeger, W. (1995): “Can Imperfect Competition Explain the Difference between Primal and Dual Productivity Measures? Estimates for U.S. Manufacturing,” *Journal of Political Economy*, 103, 316–330.

Smith, H. (2004): “Supermarket Choice and Supermarket Competition in Market Equilibrium,” *Review of Economic Studies*, 71, 235–263.

Wang, C. J. (2003): “Productivity and Economies of Scale in the Production of Bank Service Value Added,” *FRB Boston Working Papers Series*, 03-7.

W.K.A, A. D. (1988): “Laws of large Numbers for Dependent Non-identically Distributed Random Variables,” *Econometric Theory*, 4, 458–467.

Wolfram, C. (1999): “Measuring Duopoly Power in the British Electricity Spot Market,” *American Economic Review*, 89, 805–826.

Appendix

A Proof of Proposition 1

Proof. For each firm the observed cost is

$$C^d_{jm} = C_{jm} + \eta_{jm}$$

for firm/product j in market m, and η_{jm} is the measurement error. Denote the sieve function of q_{jm}, $\mathbf{w}_{jm}$ and MR_{jm} as

$$\psi\left(q_{jm}, \mathbf{w}_{jm}, MR_{jm}(\boldsymbol{\theta}), \boldsymbol{\gamma}\right) \equiv \sum_{l=1}^{\infty} \gamma_l \psi_l\left(q_{jm}, \mathbf{w}_{jm}, MR_{jm}(\boldsymbol{\theta})\right).$$

Then, because of Assumption 9,

$$\begin{aligned} & E\left[\left(C^d_{jm} - \psi\left(q_{jm}, \mathbf{w}_{jm}, MR_{jm}(\boldsymbol{\theta}), \boldsymbol{\gamma}\right)\right)^2\right] \\ = \;& E\left[\left(C_{jm} - \psi\left(q_{jm}, \mathbf{w}_{jm}, MR_{jm}(\boldsymbol{\theta}), \boldsymbol{\gamma}\right)\right)^2\right] + 2E\left[\left(C_{jm} - \psi\left(q_{jm}, \mathbf{w}_{jm}, MR_{jm}(\boldsymbol{\theta}), \boldsymbol{\gamma}\right)\right)\eta_{jm}\right] + E\left(\eta^2_{jm}\right) \\ = \;& E\left[\left(C_{jm} - \psi\left(q_{jm}, \mathbf{w}_{jm}, MR_{jm}(\boldsymbol{\theta}), \boldsymbol{\gamma}\right)\right)^2\right] + \sigma^2_\eta. \end{aligned}$$

From Assumption 10, there exists an infinite sequence $\gamma_0 = \{\gamma_{0l}\}_{l=1}^{\infty}$ such that

$$C_{jm} = PC\left(q_{jm}, \mathbf{w}_{jm}, MR_{jm}\left(\boldsymbol{\theta}_0\right)\right) = \psi\left(q_{jm}, \mathbf{w}_{jm}, MR_{jm}(\boldsymbol{\theta}_0), \boldsymbol{\gamma}_0\right) = \sum_{l=1}^{\infty} \gamma_{0l} \psi_l\left(q_{jm}, \mathbf{w}_{jm}, MR_{jm}\left(\boldsymbol{\theta}_0\right)\right).$$

Therefore,

$$\begin{aligned} & E\left[\left(C^d_{jm} - \sum_{l=1}^{\infty} \gamma_{0l} \psi_l\left(q_{jm}, \mathbf{w}_{jm}, MR_{jm}\left(\boldsymbol{\theta}_0\right)\right)\right)^2 | q_{jm} = q, \mathbf{w}_{jm} = \mathbf{w}, MR_{jm}\left(\boldsymbol{\theta}_0\right)\right] \\ = \;& 0 + \sigma^2_\eta. \end{aligned}$$

From Assumption 11. 2, if $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$,

$$MR\left(\tilde{\mathbf{p}}_m, \tilde{\mathbf{s}}_m, j, \boldsymbol{\theta}\right) = MR\left(\tilde{\tilde{\mathbf{p}}}_m, \tilde{\tilde{\mathbf{s}}}_m, j, \boldsymbol{\theta}\right)$$

$$MR\left(\tilde{\mathbf{p}}_m, \tilde{\mathbf{s}}_m, j, \boldsymbol{\theta}_0\right) \neq MR\left(\tilde{\tilde{\mathbf{p}}}_m, \tilde{\tilde{\mathbf{s}}}_m, j, \boldsymbol{\theta}_0\right).$$

Hence,

$$\sum_{l=1}^{\infty} \gamma_{0l} \psi_l\left(q, \mathbf{w}, MR\left(\tilde{\mathbf{p}}_m, \tilde{\mathbf{s}}_m, j, \boldsymbol{\theta}\right)\right) = \sum_{l=1}^{\infty} \gamma_{0l} \psi_l\left(q, \mathbf{w}, MR\left(\tilde{\tilde{\mathbf{p}}}_m, \tilde{\tilde{\mathbf{s}}}_m, j, \boldsymbol{\theta}\right)\right).$$

Because the true pseudo-cost function is strictly increasing in MR, $MR\left(\tilde{\mathbf{p}}_m, \tilde{\mathbf{s}}_m, j, \boldsymbol{\theta}_0\right) \neq MR\left(\tilde{\tilde{\mathbf{p}}}_m, \tilde{\tilde{\mathbf{s}}}_m, j, \boldsymbol{\theta_0}\right)$ implies

$$\sum_{l=1}^{\infty} \gamma_{0l} \psi_l\left(q, \mathbf{w}, MR\left(\tilde{\mathbf{p}}_m, \tilde{\mathbf{s}}_m, j, \boldsymbol{\theta}_0\right)\right) \neq \sum_{l=1}^{\infty} \gamma_{0l} \psi_l\left(q, \mathbf{w}, MR\left(\tilde{\tilde{\mathbf{p}}}_m, \tilde{\tilde{\mathbf{s}}}_m, j, \boldsymbol{\theta_0}\right)\right).$$

Then, given $\boldsymbol{\gamma}$ and $\boldsymbol{\gamma}_0$ from continuity of pseudo-cost function with respect to marginal cost, and continuity of marginal revenue function with respect to $\mathbf{p}$ and $\mathbf{s}$, for sufficiently small open ball $\tilde{\mathcal{B}}$ that contains $\left(\tilde{\mathbf{p}}_m, \tilde{\mathbf{s}}_m\right)$ and $\tilde{\tilde{\mathcal{B}}}$ that contains $\left(\tilde{\tilde{\mathbf{p}}}_m, \tilde{\tilde{\mathbf{s}}}_m\right)$,

$$\begin{aligned} & sup_{\tilde{\mathbf{b}} \in \tilde{\mathcal{B}}} \left|\psi\left(q, \mathbf{w}, MR\left(\tilde{\mathbf{b}}, j, \boldsymbol{\theta}\right), \tilde{\boldsymbol{\gamma}}\right) - \psi\left(q, \mathbf{w}, MR\left(\tilde{\mathbf{p}}_m, \tilde{\mathbf{s}}_m, j, \boldsymbol{\theta}\right), \tilde{\boldsymbol{\gamma}}\right)\right| < \frac{\epsilon}{16},\; \tilde{\boldsymbol{\gamma}} = \boldsymbol{\gamma}, \boldsymbol{\gamma}_0, \\ & sup_{\tilde{\tilde{\mathbf{b}}} \in \tilde{\tilde{\mathcal{B}}}} \left|\psi\left(q, \mathbf{w}, MR\left(\tilde{\tilde{\mathbf{b}}}, j, \boldsymbol{\theta}\right), \tilde{\boldsymbol{\gamma}}\right) - \psi(q, \mathbf{w}, MR\left(\tilde{\tilde{\mathbf{p}}}_m, \tilde{\tilde{\mathbf{s}}}_m, j, \boldsymbol{\theta}\right), \tilde{\boldsymbol{\gamma}}\right)\right| < \frac{\epsilon}{16},\; \tilde{\boldsymbol{\gamma}} = \boldsymbol{\gamma}, \boldsymbol{\gamma}_0, \\ & sup_{\tilde{\mathbf{b}} \in \tilde{\mathcal{B}}} \left|\psi\left(q, \mathbf{w}, MR\left(\tilde{\mathbf{b}}, j, \boldsymbol{\theta}_0\right), \tilde{\boldsymbol{\gamma}}\right) - \psi(q, \mathbf{w}, MR\left(\tilde{\mathbf{p}}_m, \tilde{\mathbf{s}}_m, j, \boldsymbol{\theta}_0\right), \tilde{\boldsymbol{\gamma}}\right)\right| < \frac{\epsilon}{16},\; \tilde{\boldsymbol{\gamma}} = \boldsymbol{\gamma}, \boldsymbol{\gamma}_0, \\ & sup_{\tilde{\tilde{\mathbf{b}}} \in \tilde{\tilde{\mathcal{B}}}} \left|\psi\left(q, \mathbf{w}, MR\left(\tilde{\tilde{\mathbf{b}}}, j, \boldsymbol{\theta}_0\right), \tilde{\boldsymbol{\gamma}}\right) - \psi(q, \mathbf{w}, MR\left(\tilde{\tilde{\mathbf{p}}}_m, \tilde{\tilde{\mathbf{s}}}_m, j, \boldsymbol{\theta}_0\right), \tilde{\boldsymbol{\gamma}}\right)\right| < \frac{\epsilon}{16},\; \tilde{\boldsymbol{\gamma}} = \boldsymbol{\gamma}, \boldsymbol{\gamma}_0 \end{aligned} \tag{19}$$

are satisfied for ϵ such that

$$0<\epsilon<\left|\sum_{l=1}^{\infty}\gamma_{0l}\psi_l\left(q,\mathbf{w},MR\left(\tilde{\mathbf{p}}_m,\tilde{\mathbf{s}}_m,j,\boldsymbol{\theta}_0\right)\right)-\sum_{l=1}^{\infty}\gamma_{0l}\psi_l\left(q,\mathbf{w},MR\left(\tilde{\tilde{\mathbf{p}}}_m,\tilde{\tilde{\mathbf{s}}}_m,j,\boldsymbol{\theta}_\mathbf{0}\right)\right)\right|.$$

Then, from Assumption 11, for any $\tilde{\mathbf{b}}\in\tilde{\mathcal{B}}$, $\tilde{\tilde{\mathbf{b}}}\in\tilde{\tilde{\mathcal{B}}}$,

$$\left|\sum_{l=1}^{\infty}\gamma_l\psi_l\left(q,\mathbf{w},MR\left(\tilde{\mathbf{b}},j,\boldsymbol{\theta}\right)\right)-\sum_{l=1}^{\infty}\gamma_l\psi_l\left(q,\mathbf{w},MR\left(\tilde{\tilde{\mathbf{b}}},j,\boldsymbol{\theta}\right)\right)\right|<\frac{\epsilon}{8},$$

and

$$\left|\sum_{l=1}^{\infty}\gamma_{0l}\psi_l\left(q,\mathbf{w},MR\left(\tilde{\mathbf{b}},j,\boldsymbol{\theta}_0\right)\right)-\sum_{l=1}^{\infty}\gamma_{0l}\psi_l\left(q,\mathbf{w},MR\left(\tilde{\tilde{\mathbf{b}}},j,\boldsymbol{\theta}_0\right)\right)\right|>\frac{7}{8}\epsilon.$$

Then,

$$\begin{aligned}\frac{7}{8}\epsilon&<\left|\sum_{l=1}^{\infty}\gamma_{0l}\psi_l\left(q,\mathbf{w},MR\left(\tilde{\mathbf{b}},j,\boldsymbol{\theta}_0\right)\right)-\sum_{l=1}^{\infty}\gamma_{0l}\psi_l\left(q,\mathbf{w},MR\left(\tilde{\tilde{\mathbf{b}}},j,\boldsymbol{\theta}_\mathbf{0}\right)\right)\right|\\
&\leq\left|\sum_{l=1}^{\infty}\gamma_{0l}\psi_l\left(q,\mathbf{w},MR\left(\tilde{\mathbf{b}},j,\boldsymbol{\theta}_0\right)\right)-\sum_{l=1}^{\infty}\gamma_l\psi_l\left(q,\mathbf{w},MR\left(\tilde{\mathbf{b}},j,\boldsymbol{\theta}\right)\right)\right|\\
&\quad+\left|\sum_{l=1}^{\infty}\gamma_l\psi_l\left(q,\mathbf{w},MR\left(\tilde{\mathbf{b}},j,\boldsymbol{\theta}\right)\right)-\sum_{l=1}^{\infty}\gamma_l\psi_l\left(q,\mathbf{w},MR\left(\tilde{\tilde{\mathbf{b}}},j,\boldsymbol{\theta}\right)\right)\right|\\
&\quad+\left|\sum_{l=1}^{\infty}\gamma_l\psi_l\left(q,\mathbf{w},MR\left(\tilde{\tilde{\mathbf{b}}},j,\boldsymbol{\theta}\right)\right)-\sum_{l=1}^{\infty}\gamma_{0l}\psi_l\left(q,\mathbf{w},MR\left(\tilde{\tilde{\mathbf{b}}},j,,\boldsymbol{\theta}_0\right)\right)\right|,\end{aligned}$$

implying

$$\begin{aligned}\frac{3}{4}\epsilon&<\left|\sum_{l=1}^{\infty}\gamma_{0l}\psi_l\left(q,\mathbf{w},MR\left(\tilde{\mathbf{b}},j,\boldsymbol{\theta}_0\right)\right)-\sum_{l=1}^{\infty}\gamma_l\psi_l\left(q,\mathbf{w},MR\left(\tilde{\mathbf{b}},j,\boldsymbol{\theta}\right)\right)\right|\\
&\quad+\left|\sum_{l=1}^{\infty}\gamma_l\psi_l\left(q,\mathbf{w},MR\left(\tilde{\tilde{\mathbf{b}}},j,\boldsymbol{\theta}\right)\right)-\sum_{l=1}^{\infty}\gamma_{0l}\psi_l\left(q,\mathbf{w},MR\left(\tilde{\tilde{\mathbf{b}}},j,\boldsymbol{\theta}_0\right)\right)\right|.\end{aligned}$$

Hence, given $\tilde{\mathbf{b}}\in\tilde{\mathcal{B}}$ and $\tilde{\tilde{\mathbf{b}}}\in\tilde{\tilde{\mathcal{B}}}$, either

$$\frac{3}{8}\epsilon<\left|\sum_{l=1}^{\infty}\gamma_{0l}\psi_l\left(q,\mathbf{w},MR\left(\tilde{\mathbf{b}},j,\boldsymbol{\theta}_0\right)\right)-\sum_{l=1}^{\infty}\gamma_l\psi_l\left(q,\mathbf{w},MR\left(\tilde{\mathbf{b}},j,\boldsymbol{\theta}\right)\right)\right|$$

or

$$\frac{3}{8}\epsilon<\left|\sum_{l=1}^{\infty}\gamma_{0l}\psi_l\left(q,\mathbf{w},MR\left(\tilde{\tilde{\mathbf{b}}},j,\boldsymbol{\theta}_0\right)\right)-\sum_{l=1}^{\infty}\gamma_l\psi_l\left(q,\mathbf{w},MR\left(\tilde{\tilde{\mathbf{b}}},j,\boldsymbol{\theta}\right)\right)\right|$$

or both. Now, because of equation (19), this implies that either

$$\frac{1}{8}\epsilon<sup_{\tilde{\mathbf{b}}\in\tilde{\mathcal{B}}}\left|\sum_{l=1}^{\infty}\gamma_{0l}\psi_l\left(q,\mathbf{w},MR\left(\tilde{\mathbf{b}},j,\boldsymbol{\theta}_0\right)\right)-\sum_{l=1}^{\infty}\gamma_l\psi_l\left(q,\mathbf{w},MR\left(\tilde{\mathbf{b}},j,\boldsymbol{\theta}\right)\right)\right|$$

or

$$\frac{1}{8}\epsilon<sup_{\tilde{\tilde{\mathbf{b}}}\in\tilde{\tilde{\mathcal{B}}}}\left|\sum_{l=1}^{\infty}\gamma_{0l}\psi_l\left(q,\mathbf{w},MR\left(\tilde{\tilde{\mathbf{b}}},j,\boldsymbol{\theta}_0\right)\right)-\sum_{l=1}^{\infty}\gamma_l\psi_l\left(q,\mathbf{w},MR\left(\tilde{\tilde{\mathbf{b}}},j,\boldsymbol{\theta}\right)\right)\right|$$

or both.

This implies that either

$$\begin{aligned}&E\left[\left(\sum_{l=1}^{\infty}\gamma_{0l}\psi_l\left(q,\mathbf{w},MR\left(\tilde{\mathbf{b}},j,\boldsymbol{\theta}_\mathbf{0}\right)\right)\right.\right.\\
&\left.\left.-\sum_{l=1}^{\infty}\gamma_l\psi_l\left(q_{jm},\mathbf{w}_{jm},MR\left(\tilde{\mathbf{b}},j,\boldsymbol{\theta}\right)\right)\right)^2\mid\left(q,\mathbf{w},\tilde{\mathbf{b}}\right)\in\mathcal{A}\times\tilde{\mathcal{B}}\right]>0,\end{aligned}$$

or

$$E\left[\left(\sum_{l=1}^{\infty}\gamma_{0l}\psi_l\left(q,\mathbf{w},MR\left(\tilde{\tilde{\mathbf{b}}},j,\boldsymbol{\theta}_0\right)\right)\right.\right.$$
$$\left.\left.-\sum_{l=1}^{\infty}\gamma_l\psi_l\left(q,\mathbf{w},MR\left(\tilde{\tilde{\mathbf{b}}},j,\boldsymbol{\theta}\right)\right)\right)^2\mid\left(q,\mathbf{w},\tilde{\tilde{\mathbf{b}}}\right)\in\mathcal{A}\times\tilde{\tilde{\mathcal{B}}}\right]>0,$$

or both. Therefore, integrating over q, $\mathbf{w}$ and MR, we obtain that for $\boldsymbol{\theta}\neq\boldsymbol{\theta}_0$,

$$\begin{aligned}
&E\left[\left(C_{jm}-\sum_{l=1}^{\infty}\gamma_l\psi_l\left(q_{jm},\mathbf{w}_{jm},MR_{jm}(\boldsymbol{\theta})\right)\right)^2\right]\\
&=E\left[\left(\sum_{l=1}^{\infty}\gamma_{0l}\psi_l\left(q_{jm},\mathbf{w}_{jm},MR_{jm}(\boldsymbol{\theta}_0)\right)-\sum_{l=1}^{\infty}\gamma_l\psi_l\left(q_{jm},\mathbf{w}_{jm},MR_{jm}(\boldsymbol{\theta})\right)\right)^2\right]\\
&\geq E\left[\left(\sum_{l=1}^{\infty}\gamma_{0l}\psi_l\left(q,\mathbf{w},MR\left(\tilde{\mathbf{b}},j,\boldsymbol{\theta}_0\right)\right)\right.\right.\\
&\left.\left.-\sum_{l=1}^{\infty}\gamma_l\psi_l\left(q_{jm},\mathbf{w}_{jm},MR\left(\tilde{\mathbf{b}},j,\boldsymbol{\theta}\right)\right)\right)^2\mid\left(q,\mathbf{w},\tilde{\mathbf{b}}\right)\in\mathcal{A}\times\tilde{\mathcal{B}}\right]Prob\left(\mathcal{A}\times\tilde{\mathcal{B}}\right)\\
&+E\left[\left(\sum_{l=1}^{\infty}\gamma_{0l}\psi_l\left(q,\mathbf{w},MR\left(\tilde{\tilde{\mathbf{b}}},j,\boldsymbol{\theta}_0\right)\right)\right.\right.\\
&\left.\left.-\sum_{l=1}^{\infty}\gamma_l\psi_l\left(q,\mathbf{w},MR\left(\tilde{\tilde{\mathbf{b}}},j,\boldsymbol{\theta}\right)\right)\right)^2\mid\left(q,\mathbf{w},\tilde{\tilde{\mathbf{b}}}\right)\in\mathcal{A}\times\tilde{\tilde{\mathcal{B}}}\right]Prob\left(\mathcal{A}\times\tilde{\tilde{\mathcal{B}}}\right)>0
\end{aligned}$$

Therefore,

$$E\left[\left(C_{jm}^d-\sum_{l=1}^{\infty}\gamma_l\psi_l\left(q,\mathbf{w},MR_{jm}\left(\boldsymbol{\theta}\right)\right)\right)^2\right]\geq\sigma_\eta^2,$$

with equality only holding for $\boldsymbol{\theta}=\boldsymbol{\theta}_0$. ■

B Parametric Identification of marginal revenue

B.1 Proof of Lemma 3: Logit Model

Proof. It is easy to show that the Berry (1994) logit demand model satisfies Assumption 11. For the parameter $\alpha\neq\alpha_0$, pick the two firms j and j' in two different markets m, m' with prices p_{jm}, $p_{j'm'}$ and market shares s_{jm}, $s_{j'm'}$ such that under α their marginal revenues are equated, i.e.

$$p_{jm}+\frac{1}{(1-s_{jm})\,\alpha}=p_{j'm'}+\frac{1}{(1-s_{j'm'})\,\alpha}\Rightarrow\alpha=-\frac{1}{p_{jm}-p_{j'm'}}\left[\frac{1}{1-s_{jm}}-\frac{1}{1-s_{j'm'}}\right].$$

Then, for $\alpha\neq\alpha_0$,

$$\alpha_0\neq-\frac{1}{p_{jm}-p_{j'm'}}\left[\frac{1}{1-s_{jm}}-\frac{1}{1-s_{j'm'}}\right],$$

thus,

$$p_{jm}+\frac{1}{(1-s_{jm})\,\alpha_0}\neq p_{j'm'}+\frac{1}{(1-s_{j'm'})\,\alpha_0}.$$

Therefore, the price coefficient satisfies Assumption 11. ■

B.2 Proof of Lemma 3: BLP Model.

Proof. Next, we prove that the random coefficient BLP model also satisfies Assumption 11 in monopoly markets. We consider the data with $\mathbf{x} = 0$. Then, per period log utility component of a purchase is $u = p\alpha + \xi$, where $\alpha \sim N(\mu_\alpha, \sigma_\alpha)$. Consider the pair (s, p, ξ) and (s', p', ξ') that satisfy the share equation. Then,

$$\int_\alpha \frac{\exp(\xi + p\alpha)}{1 + \exp(\xi + p\alpha)} \frac{1}{\sigma_\alpha} \phi\left(\frac{\alpha - \mu_\alpha}{\sigma_\alpha}\right) d\alpha = \int_\alpha \frac{\exp(p(\alpha + \xi/p))}{1 + \exp(p(\alpha + \xi/p))} \frac{1}{\sigma_\alpha} \phi\left(\frac{\alpha - \mu_\alpha}{\sigma_\alpha}\right) d\alpha = s$$

and

$$\int_\alpha \frac{\exp(p'(\alpha + \xi'/p'))}{1 + \exp(p'(\alpha + \xi'/p'))} \frac{1}{\sigma_\alpha} \phi\left(\frac{\alpha - \mu_\alpha}{\sigma_\alpha}\right) d\alpha = s'$$

and we assume that they have the same marginal revenue:

$$\begin{aligned} MR = p + p\left[\int_\alpha \frac{p\exp(p(\alpha + \xi/p))}{[1 + \exp(p(\alpha + \xi/p))]^2} \alpha \frac{1}{\sigma_\alpha} \phi\left(\frac{\alpha - \mu_\alpha}{\sigma_\alpha}\right) d\alpha\right]^{-1} s \\ = p' + p'\left[\int_\alpha \frac{p'\exp(p'(\alpha + \xi'/p'))}{[1 + \exp(p'(\alpha + \xi'/p'))]^2} \alpha \frac{1}{\sigma_\alpha} \phi\left(\frac{\alpha - \mu_\alpha}{\sigma_\alpha}\right) d\alpha\right]^{-1} s'. \end{aligned}$$

Now, denote $\eta = \mu_\alpha/\sigma_\alpha$, $\eta_0 = \mu_{\alpha 0}/\sigma_{\alpha 0}$, $a(p) = \xi/(p\sigma_\alpha)$, $a'(p) = \xi'/(p'\sigma_\alpha)$, and $a_0(p) = \xi/(p\sigma_{\alpha 0})$, $a_0'(p) = a'/(p'\sigma_{\alpha 0})$. Furthermore, denote $\tilde{\alpha} = \alpha/\sigma_\alpha$ and $\tilde{\alpha}' = \alpha'/\sigma_\alpha$. Then, by change of variables,

$$\int_\alpha \frac{\exp(p\sigma_\alpha(\tilde{\alpha} + a(p)))}{1 + \exp(p\sigma_\alpha(\tilde{\alpha} + a(p)))} \phi(\tilde{\alpha} - \eta)\, d\tilde{\alpha} = s, \quad \int_\alpha \frac{\exp(p'\sigma_\alpha(\tilde{\alpha} + a'(p)))}{1 + \exp(p'\sigma_\alpha(\tilde{\alpha} + a'(p)))} \phi(\tilde{\alpha} - \eta)\, d\tilde{\alpha} = s' \tag{20}$$

and the marginal revenue equation becomes

$$\begin{aligned} MR &= p + p\left[\int_{\tilde{\alpha}} \frac{p\sigma_\alpha \exp(p\sigma_\alpha(\tilde{\alpha} + a(p)))}{[1 + \exp(p\sigma_\alpha(\tilde{\alpha} + a(p)))]^2} \tilde{\alpha}\phi(\tilde{\alpha} - \eta)\, d\tilde{\alpha}\right]^{-1} s \\ &= p' + p'\left[\int_{\tilde{\alpha}} \frac{p'\sigma_\alpha \exp(p'\sigma_\alpha(\tilde{\alpha} + a'(p)))}{[1 + \exp(p'\sigma_\alpha(\tilde{\alpha} + a'(p)))]^2} \tilde{\alpha}\phi(\tilde{\alpha} - \eta)\, d\tilde{\alpha}\right]^{-1} s'. \end{aligned} \tag{21}$$

Suppose Assumption 11 does not hold, then given (η, σ_α) and $(\eta_0, \sigma_{\alpha 0})$ such that $(\eta, \sigma_\alpha) \neq (\eta_0, \sigma_{\alpha 0})$, for any (s, p) and (s', p') such that $(s, p) \neq (s', p')$ satisfying equations (20) and (21),

$$\int_{\tilde{\alpha}} \frac{\exp(p\sigma_{\alpha 0}(\tilde{\alpha} + a_0(p)))}{1 + \exp(p\sigma_{\alpha 0}(\tilde{\alpha} + a_0(p)))} \phi(\tilde{\alpha} - \eta_0)\, d\tilde{\alpha} = s, \quad \int_{\tilde{\alpha}} \frac{\exp(p'\sigma_{\alpha 0}(\tilde{\alpha} + a_0'(p)))}{1 + \exp(p'\sigma_{\alpha 0}(\tilde{\alpha} + a_0'(p)))} \phi(\tilde{\alpha} - \eta_0)\, d\tilde{\alpha} = s' \tag{22}$$

and

$$\begin{aligned} MR_0 &= p + p\left[\int_{\tilde{\alpha}} \frac{p\sigma_{\alpha 0} \exp(p\sigma_{\alpha 0}(\tilde{\alpha} + a_0(p)))}{[1 + \exp(p\sigma_{\alpha 0}(\tilde{\alpha} + a_0(p)))]^2} \tilde{\alpha}\phi(\tilde{\alpha} - \eta_0)\, d\tilde{\alpha}\right]^{-1} s \\ &= p' + p'\left[\int_{\tilde{\alpha}} \frac{p'\sigma_{\alpha 0} \exp(p'\sigma_{\alpha 0}(\tilde{\alpha} + a_0'(p)))}{[1 + \exp(p'\sigma_{\alpha 0}(\tilde{\alpha} + a_0'(p)))]^2} \tilde{\alpha}\phi(\tilde{\alpha} - \eta_0)\, d\tilde{\alpha}\right]^{-1} s'. \end{aligned} \tag{23}$$

Consider first the case $\eta_0 \neq \eta$. Using integration by parts, we obtain

$$\int_{\tilde{\alpha}} \frac{\exp(p\sigma_\alpha(\tilde{\alpha} + a(p)))}{[1 + \exp(p\sigma_\alpha(\tilde{\alpha} + a(p)))]} \phi(\tilde{\alpha} - \eta)\, d\tilde{\alpha} = 1 - \int_{\tilde{\alpha}} \frac{p\sigma_\alpha exp[p\sigma_\alpha(\tilde{\alpha} + a(p))]}{[1 + exp[p\sigma_\alpha(\tilde{\alpha} + a(p))]]^2} \Phi(\tilde{\alpha} - \eta)\, d\tilde{\alpha} \tag{24}$$

Then, applying Taylor series expansion of $\Phi(\tilde{\alpha} - \eta)$ around $-a(p)$, we obtain

$$\begin{aligned} (24) \quad &= 1 - \int_{\tilde{\alpha}} \frac{p\sigma_\alpha exp[p\sigma_\alpha(\tilde{\alpha} + a(p))]}{[1 + exp[p\sigma_\alpha(\tilde{\alpha} + a(p))]]^2} \left[\Phi(-a(p) - \eta) + (\tilde{\alpha} + a(p))\phi(-a(p) - \eta)\right. \\ &\left.+ \frac{1}{2}(\tilde{\alpha} + a(p))^2 \phi'(-a(p) - \eta) + \frac{1}{6}(\tilde{\alpha} + a(p))^3 \phi''(-\alpha(p) - \eta) + \frac{1}{24}(\tilde{\alpha} + a(p))^4 \phi'''(a^*(\tilde{\alpha}) - \eta)\right] d\tilde{\alpha} \end{aligned}$$

where $a^*(\tilde{\alpha})$ is a continuous function of $\tilde{\alpha}$, and $sup_{\tilde{\alpha}} |\phi'''(a^*(\tilde{\alpha}) - \eta)| < B$ for some bounded constant $B > 0$. Notice that $\frac{p\sigma_\alpha exp[p\sigma_\alpha(\tilde{\alpha}+a(p))]}{[1+exp[p\sigma_\alpha(\tilde{\alpha}+a(p))]]^2}$ is symmetric around $-a(p)$. Hence,

$$\int_{\tilde{\alpha}} \frac{p\sigma_\alpha exp[p\sigma_\alpha(\tilde{\alpha} + a(p))]}{[1 + exp[p\sigma_\alpha(\tilde{\alpha} + a(p))]]^2} (\tilde{\alpha} + a(p))\, d\tilde{\alpha} = \int_{\tilde{\alpha}} \frac{p\sigma_\alpha exp[p\sigma_\alpha(\tilde{\alpha} + a(p))]}{[1 + exp[p\sigma_\alpha(\tilde{\alpha} + a(p))]]^2} (\tilde{\alpha} + a(p))^3\, d\tilde{\alpha} = 0.$$

Furthermore, from the formula for the variance of the logistic function,

$$\int_{\tilde{\alpha}} \frac{p\sigma_\alpha exp\left[p\sigma_\alpha\left(\tilde{\alpha}+a\left(p\right)\right)\right]}{\left[1+exp\left[p\sigma_\alpha\left(\tilde{\alpha}+a\left(p\right)\right)\right]\right]^2}\left(\tilde{\alpha}+a\left(p\right)\right)^2 d\tilde{\alpha}=\frac{\pi^2}{3p^2}$$

and from the fourth central moment, we can derive that

$$\left|\frac{1}{24}\int_{\tilde{\alpha}} \frac{p\sigma_\alpha exp\left[p\sigma_\alpha\left(\tilde{\alpha}+a\left(p\right)\right)\right]}{\left[1+exp\left[p\sigma_\alpha\left(\tilde{\alpha}+a\left(p\right)\right)\right]\right]^2}\left(\tilde{\alpha}+a\left(p\right)\right)^4 \phi'''\left(a^*\left(\tilde{\alpha}\right)-\eta\right) d\tilde{\alpha}\right|$$

$$\leq\left|\frac{1}{24}\int_{\tilde{\alpha}} \frac{p\sigma_\alpha exp\left[p\sigma_\alpha\left(\tilde{\alpha}+a\left(p\right)\right)\right]}{\left[1+exp\left[p\sigma_\alpha\left(\tilde{\alpha}+a\left(p\right)\right)\right]\right]^2}\left(\tilde{\alpha}+a\left(p\right)\right)^4 B d\tilde{\alpha}\right| \leq BC\frac{\pi^4}{p^4}\left(2^4-2\right)=O\left(p^{-4}\right)$$

where $C>0$ is a constant.

Together, we obtain,

$$(\ 24)=1-\Phi\left(-a\left(p\right)-\eta\right)-\frac{\pi^2}{6p^2}\phi'\left(-a\left(p\right)-\eta\right)+O\left(p^{-4}\right)=s=1-\Phi\left(-a-\eta\right)=\Phi\left(a+\eta\right).$$

where $a=\Phi^{-1}\left(s\right)-\eta=lim_{p\rightarrow\infty}a\left(p\right)$. Therefore,

$$-\left(a-a\left(p\right)\right)\phi\left(-a^*\left(p\right)-\eta\right)-\frac{\pi^2}{6p^2}\phi'\left(-a\left(p\right)-\eta\right)+O\left(p^{-4}\right)=0$$

where $a^*\left(p\right)$ is in between a and $a\left(p\right)$. Hence,

$$\left(a-a\left(p\right)\right)=-\frac{\phi'\left(-a\left(p\right)-\eta\right)\pi^2}{6\phi\left(-a^*\left(p\right)-\eta\right)p^2}+O\left(p^{-4}\right)=O\left(p^{-2}\right)$$

and

$$\begin{aligned}\frac{\phi'\left(-a\left(p\right)-\eta\right)}{6\phi\left(-a^*\left(p\right)-\eta\right)} &= \frac{\phi'\left(-a-\eta\right)}{6\phi\left(-a-\eta\right)}-\left(a\left(p\right)-a\right)\frac{\phi''\left(-a\left(p\right)-\eta\right)}{6\phi\left(-a^*\left(p\right)-\eta\right)}+\left(a^*\left(p\right)-a\right)\frac{\phi'\left(-a-\eta\right)\phi'\left(-a^*\left(p\right)-\eta\right)}{6\phi^2\left(-a^*\left(p\right)-\eta\right)}\\ + &\quad O\left(\left(a\left(p\right)-a\right)^2\right)+O\left(\left(a^*\left(p\right)-a\right)^2\right)+O\left(\left(a\left(p\right)-a\right)\left(a^*\left(p\right)-a\right)\right)=\frac{\phi'\left(-a-\eta\right)}{6\phi\left(-a-\eta\right)}+O\left(p^{-2}\right).\end{aligned}$$

Therefore,

$$\left(a-a\left(p\right)\right)=-\frac{\phi'\left(-a-\eta\right)\pi^2}{6\phi\left(-a-\eta\right)p^2}+O\left(p^{-4}\right)$$

Similarly, by applying Taylor series approximation of $\phi\left(\tilde{\alpha}-\eta\right)$ with respect to $\tilde{\alpha}$ around$-a\left(p\right)$, we obtain

$$\begin{aligned}&\int_{\tilde{\alpha}} \frac{p\sigma_\alpha \exp\left[p\sigma_\alpha\left(\tilde{\alpha}+a\left(p\right)\right)\right]}{\left[1+\exp\left[p\sigma_\alpha\left(\tilde{\alpha}+a\left(p\right)\right)\right]\right]^2}\tilde{\alpha}\phi\left(\tilde{\alpha}-\eta\right)d\tilde{\alpha}=\int_{\tilde{\alpha}} \frac{p\sigma_\alpha \exp\left[p\sigma_\alpha\left(\tilde{\alpha}+a\left(p\right)\right)\right]}{\left[1+\exp\left[p\sigma_\alpha\left(\tilde{\alpha}+a\left(p\right)\right)\right]\right]^2}\left(\tilde{\alpha}+a\left(p\right)\right)\phi\left(\tilde{\alpha}-\eta\right)d\tilde{\alpha}\\
&-a\left(p\right)\int_{\tilde{\alpha}} \frac{p\sigma_\alpha \exp\left[p\sigma_\alpha\left(\tilde{\alpha}+a\left(p\right)\right)\right]}{\left[1+\exp\left[p\sigma_\alpha\left(\tilde{\alpha}+a\left(p\right)\right)\right]\right]^2}\phi\left(\tilde{\alpha}-\eta\right)d\tilde{\alpha}\\
&=\int_{\tilde{\alpha}} \frac{p\sigma_\alpha \exp\left[p\sigma_\alpha\left(\tilde{\alpha}+a\left(p\right)\right)\right]}{\left[1+\exp\left[p\sigma_\alpha\left(\tilde{\alpha}+a\left(p\right)\right)\right]\right]^2}\left[\left(\tilde{\alpha}+a\left(p\right)\right)\phi\left(-a\left(p\right)-\eta\right)+\left(\tilde{\alpha}+a\left(p\right)\right)^2\phi'\left(-a\left(p\right)-\eta\right)\right.\\
&\left.+\frac{1}{2}\left(\tilde{\alpha}+a\left(p\right)\right)^3\phi''\left(-a\left(p\right)-\eta\right)+\frac{1}{6}\left(\tilde{\alpha}+a\left(p\right)\right)^4\phi'''\left(-a^*\left(\tilde{\alpha}\right)-\eta\right)\right]d\tilde{\alpha}\\
&-a\left(p\right)\int_{\tilde{\alpha}} \frac{p\sigma_\alpha \exp\left[p\sigma_\alpha\left(\tilde{\alpha}+a\left(p\right)\right)\right]}{\left[1+\exp\left[p\sigma_\alpha\left(\tilde{\alpha}+a\left(p\right)\right)\right]\right]^2}\left[\phi\left(-a\left(p\right)-\eta\right)+\left(\tilde{\alpha}+a\left(p\right)\right)\phi'\left(-a\left(p\right)-\eta\right)\right.\\
&\left.+\frac{1}{2}\left(\tilde{\alpha}+a\left(p\right)\right)^2\phi''\left(-a\left(p\right)-\eta\right)+\frac{1}{6}\left(\tilde{\alpha}+a\left(p\right)\right)^3\phi'''\left(-a\left(p\right)-\eta\right)+\frac{1}{24}\left(\tilde{\alpha}+a\left(p\right)\right)^4\phi''''\left(-a^*\left(\tilde{\alpha}\right)-\eta\right)\right]d\tilde{\alpha}\end{aligned}$$

Therefore,

$$
\begin{aligned}
& \int_{\tilde{\alpha}} \frac{p\sigma_\alpha \exp\left[p\sigma_\alpha\left(\tilde{\alpha}+a\left(p\right)\right)\right]}{\left[1+\exp\left[p\sigma_\alpha\left(\tilde{\alpha}+a\left(p\right)\right)\right]\right]^2}\tilde{\alpha}\phi\left(\tilde{\alpha}-\eta\right)d\tilde{\alpha} \\
= \quad & -a\left(p\right)\phi\left(-a\left(p\right)-\eta\right)+\left[\phi'\left(-a\left(p\right)-\eta\right)-\frac{a\left(p\right)}{2}\phi''\left(-a\left(p\right)-\eta\right)\right]\frac{\pi^2}{6p^2}+O\left(p^{-4}\right) \\
= \quad & -a\phi\left(-a-\eta\right)+\left[\left(-\phi\left(-a-\eta\right)+a\phi'\left(-a-\eta\right)\right)\frac{\phi'\left(-a-\eta\right)}{\phi\left(-a-\eta\right)}+\phi'\left(-a-\eta\right)-\frac{a}{2}\phi''\left(-a-\eta\right)\right]\frac{\pi^2}{6p^2} \\
& +O\left(p^{-4}\right) \qquad (25) \\
= \quad & -a\phi\left(-a-\eta\right)+O\left(p^{-2}\right)
\end{aligned}
$$

Therefore,

$$
MR = p\left[1-\left[\left(\Phi^{-1}\left(s\right)-\eta\right)\phi\left(\Phi^{-1}\left(s\right)\right)+O\left(p^{-2}\right)\right]^{-1}s\right].
$$

Then, obtain from the data all the market shares s whose corresponding prices p satisfy $p>\overline{p}$ for a large $\overline{p}<\overline{P}$, and let the set of market share with such a price to be $\mathcal{V}$. Then, because of Assumption 6, we need to make sure that price derivative of market share is negative and marginal revenue is positive. Therefore, we only focus on the parameters that satisfy $\eta < inf_{s\in\mathcal{V}}\left[\Phi^{-1}\left(s\right)-1/\left(\phi\left(\Phi^{-1}\left(s\right)\right)s\right)\right]$. Then, there exist (s,p), and (s',p'); $s\neq s'$ with p and p' large enough satisfying those two conditions.

We pick different values of $s,\ s'\in\mathcal{S}$, and the relative prices $P=p/p'$ so that they satisfy the following equation.

$$
MR = p\left[1-\frac{s}{\left(\Phi^{-1}\left(s\right)-\eta\right)\phi\left(\Phi^{-1}\left(s\right)\right)+O\left(p^{-2}\right)}\right] = p'\left[1-\frac{s'}{\left(\Phi^{-1}\left(s'\right)-\eta\right)\phi\left(\Phi^{-1}\left(s'\right)\right)+O\left(p'^{-2}\right)}\right]. \qquad (26)
$$

Such two points can be chosen because given the relative price P, in the above equation, both sides are roughly constant function of p and p' for large p and p'. Equation (26) can be rewritten as,

$$
MR = p-\frac{ps\phi^{-1}\left(\Phi^{-1}\left(s\right)\right)}{\left(\Phi^{-1}\left(s\right)-\eta\right)+O\left(p^{-2}\right)} = p'-\frac{p's'\phi^{-1}\left(\Phi^{-1}\left(s'\right)\right)}{\left(\Phi^{-1}\left(s'\right)-\eta\right)+O\left(p'^{-2}\right)}.
$$

Now, consider $\tilde{p}$ and $\tilde{p}'$ such that

$$
\tilde{p}-\frac{\tilde{p}s\phi^{-1}\left(\Phi^{-1}\left(s\right)\right)}{\left(\Phi^{-1}\left(s\right)-\eta\right)} = \tilde{p}'-\frac{\tilde{p}'s'\phi^{-1}\left(\Phi^{-1}\left(s'\right)\right)}{\left(\Phi^{-1}\left(s'\right)-\eta\right)}. \qquad (27)
$$

Denote $B\equiv\Phi^{-1}\left(s\right)$, $B'\equiv\Phi^{-1}\left(s'\right)$, $C\equiv s/\phi\left(\Phi^{-1}\left(s\right)\right)$, $C'\equiv s'/\phi\left(\Phi^{-1}\left(s'\right)\right)$, and $\tilde{P}=\tilde{p}'/\tilde{p}$. Then,

$$
\left[1-\frac{C}{B-\eta}\right] = \tilde{P}\left[1-\frac{C'}{B'-\eta}\right]
$$

$$
\left(B-\eta\right)\left(B'-\eta\right)\left(1-\tilde{P}\right)-C\left(B'-\eta\right)+\tilde{P}C'\left(B-\eta\right)=0
$$

$$
\eta^2-\left[B+B'-\frac{C-\tilde{P}C'}{1-\tilde{P}}\right]\eta+BB'-\frac{CB'-\tilde{P}C'B}{1-\tilde{P}}=0
$$

Then,

$$
\eta=\frac{1}{2}\left[B+B'-\frac{C-\tilde{P}C'}{1-\tilde{P}}\right]\pm\frac{1}{2}A,\quad A=\sqrt{\left[B+B'-\frac{C-\tilde{P}C'}{1-\tilde{P}}\right]^2+4\frac{CB'-\tilde{P}C'B}{1-\tilde{P}}-4BB'}.
$$

Now, set $s=1/2$, hence, $B=0$ and $C=1/\left(2\phi\left(0\right)\right)$. Then, from the assumption, it is easy to see that the slope of market share with respect to price C/η is negative and the marginal revenue $1+C/\eta$ is positive. In that case,

$$
\eta=\frac{1}{2}\left[B'-\frac{C-\tilde{P}C'}{1-\tilde{P}}\right]\pm\frac{1}{2}A,\quad A=\sqrt{\left[B'-\frac{C-\tilde{P}C'}{1-\tilde{P}}\right]^2+4\frac{CB'}{1-\tilde{P}}}.
$$

$$
\left[1+\frac{C}{\eta}\right]=\tilde{P}\left[1-\frac{C'}{B'-\eta}\right]
$$

Then, for $B'<0$ such that $B'>\eta$ close to η, $C'/\left(B'-\eta\right)$ can be made arbitrarily large. Hence, from Intermediate Value Theorem, one can choose $s'<s$, $B'<0$ such that $-C/\eta<C'/\left(B'-\eta\right)<1$. Then, $\tilde{P}>1$, hence

$CB'/\left(1-\tilde{P}\right)>0$. Then, if we denote

$$\eta_1=\frac{1}{2}\left[B+B'-\frac{C-\tilde{P}C'}{1-\tilde{P}}\right]+\frac{1}{2}A\quad \eta_2=\frac{1}{2}\left[B+B'-\frac{C-\tilde{P}C'}{1-\tilde{P}}\right]-\frac{1}{2}A,$$

$\eta_1>0$, and $\eta_2<0$. Hence, only η_2 is consistent with the negative demand curve slope. Furthermore $\eta=\eta_2$ satisfying equation (27) can be made arbitrarily close to η satisfying equation (26). Therefore, claim holds.

Next, consider the case where $\eta_0=\eta$, $\sigma_{\alpha 0}\neq\sigma_\alpha$. First, consider σ_α such that $\sigma_{\alpha 0}>\sigma_\alpha$. Suppose Assumption 11 is not satisfied. Then, consider $s\neq s'$, and p, p'in the data such that the following holds: Given $a\left(p\right)$, $a'\left(p\right)$ satisfying

$$\int_{\tilde{\alpha}}\frac{\exp\left(p\sigma_\alpha\left(\tilde{\alpha}+a\left(p\right)\right)\right)}{\left[1+\exp\left(p\sigma_\alpha\left(\tilde{\alpha}+a\left(p\right)\right)\right)\right]}\phi\left(\tilde{\alpha}-\eta\right)d\tilde{\alpha}=s$$
$$\int_{\tilde{\alpha}}\frac{\exp\left(p'\sigma_\alpha\left(\tilde{\alpha}+a'\left(p\right)\right)\right)}{\left[1+\exp\left(p'\sigma_\alpha\left(\tilde{\alpha}+a'\left(p\right)\right)\right)\right]}\phi\left(\tilde{\alpha}-\eta\right)d\tilde{\alpha}=s',$$

then,

$$\begin{aligned}&p+p\left[\int_{\tilde{\alpha}}\frac{p\sigma_\alpha\exp\left[p\sigma_\alpha\left(\tilde{\alpha}+a\left(p\right)\right)\right]}{\left[1+\exp\left[p\sigma_\alpha\left(\tilde{\alpha}+a\left(p\right)\right)\right]\right]^2}\tilde{\alpha}\phi\left(\tilde{\alpha}-\eta\right)d\tilde{\alpha}\right]^{-1}s\\ =\;&p'+p'\left[\int_{\tilde{\alpha}}\frac{p'\sigma_\alpha\exp\left[p'\sigma_\alpha\left(\tilde{\alpha}+a'\left(p\right)\right)\right]}{\left[1+\exp\left[p'\sigma_\alpha\left(\tilde{\alpha}+a'\left(p\right)\right)\right]\right]^2}\tilde{\alpha}\phi\left(\tilde{\alpha}-\eta\right)d\tilde{\alpha}\right]^{-1}s'.\end{aligned}$$

Now, because of the assumption of nonidentification, the same relationship holds for $\sigma_{\alpha 0}$ instead of σ_α, that is, given $a_0\left(p\right)$, $a_0'\left(p\right)$ satisfying

$$\begin{aligned}\int_{\tilde{\alpha}}\frac{\exp\left(p\sigma_{\alpha 0}\left(\tilde{\alpha}+a_0\left(p\right)\right)\right)}{\left[1+\exp\left(p\sigma_{\alpha 0}\left(\tilde{\alpha}+a_0\left(p\right)\right)\right)\right]}\phi\left(\tilde{\alpha}-\eta\right)d\tilde{\alpha}&=s\\ \int_{\tilde{\alpha}}\frac{\exp\left(p'\sigma_{\alpha 0}\left(\tilde{\alpha}+a_0'\left(p\right)\right)\right)}{\left[1+\exp\left(p'\sigma_{\alpha 0}\left(\tilde{\alpha}+a_0'\left(p\right)\right)\right)\right]}\phi\left(\tilde{\alpha}-\eta\right)d\tilde{\alpha}&=s',\end{aligned}$$

then,

$$\begin{aligned}&p+p\left[\int_{\tilde{\alpha}}\frac{p\sigma_{\alpha 0}\exp\left[p\sigma_{\alpha 0}\left(\tilde{\alpha}+a_0\left(p\right)\right)\right]}{\left[1+\exp\left[p\sigma_{\alpha 0}\left(\tilde{\alpha}+a_0\left(p\right)\right)\right]\right]^2}\tilde{\alpha}\phi\left(\tilde{\alpha}-\eta\right)d\tilde{\alpha}\right]^{-1}s\\ =\;&p'+p'\left[\int_{\alpha}\frac{p'\sigma_{\alpha 0}\exp\left[p'\sigma_{\alpha 0}\left(\tilde{\alpha}+a_0'\left\{p\right\}\right)\right]}{\left[1+\exp\left[p'\sigma_{\alpha 0}\left(\tilde{\alpha}+a_0'\left(p\right)\right)\right]\right]^2}\tilde{\alpha}\phi\left(\tilde{\alpha}-\eta\right)d\tilde{\alpha}\right]^{-1}s'\end{aligned}\tag{28}$$

In that case, if we define $p^{(1)}$ and $p^{(1)\prime}$ such that $p^{(1)}\sigma_\alpha=p\sigma_{\alpha 0}$, $p^{(1)\prime}\sigma_\alpha=p'\sigma_{\alpha 0}$ then, $p^{(1)}=\left(\sigma_{\alpha 0}/\sigma_\alpha\right)p>p$, $p^{(1)\prime}=\left(\sigma_{\alpha 0}/\sigma_\alpha\right)p'>p'$and

$$\begin{aligned}(28)\times\sigma_{\alpha 0}\;&=\;p^{(1)}\sigma_\alpha+p^{(1)}\sigma_\alpha\left[\int_{\tilde{\alpha}}\frac{p^{(1)}\sigma_\alpha\exp\left[p^{(1)}\sigma_\alpha\left(\tilde{\alpha}+a\left(p\right)\right)\right]}{\left[1+\exp\left[p^{(1)}\sigma_\alpha\left(\tilde{\alpha}+a\left(p\right)\right)\right]\right]^2}\tilde{\alpha}\phi\left(\tilde{\alpha}-\eta\right)d\tilde{\alpha}\right]^{-1}s\\ &=\;p^{(1)\prime}\sigma_\alpha+p^{(1)\prime}\sigma_\alpha\left[\int\frac{p^{(1)\prime}\sigma_\alpha\exp\left[p^{(1)\prime}\sigma_\alpha\left(\tilde{\alpha}+a_0\left(p\right)\right)\right]}{\left[1+\exp\left[p^{(1)\prime}\sigma_\alpha\left(\tilde{\alpha}+a_0\left(p\right)\right)\right]\right]^2}\tilde{\alpha}\phi\left(\alpha-\eta\right)d\tilde{\alpha}\right]^{-1}s'.\end{aligned}$$

This way, we can generate an increasing sequence of prices $\left(p^{(0)},p^{(0)\prime}\right)$, $\left(p^{(1)},p^{(1)\prime}\right)$,...,$\left(p^{(k)},p^{(k)\prime}\right)$,... such that $\left(p^{(0)},p^{(0)\prime}\right)=(p,p')$, $\left(p^{(k)},p^{(k)\prime}\right)=\left((\sigma_{\alpha 0}/\sigma_\alpha)^k p,(\sigma_{\alpha 0}/\sigma_\alpha)^k p'\right)$, and for any integer $k\geq 1$ such that.

$$\begin{aligned}
& p^{(k)}+p^{(k)}\left[\int_{\tilde{\alpha}}\frac{p^{(k)}\sigma_\alpha\exp\left[p^{(k)}\sigma_\alpha\left(\tilde{\alpha}+a^{(k)}\right)\right]}{\left[1+\exp\left[p^{(k)}\sigma_\alpha\left(\tilde{\alpha}-\eta\right)\right]\right]^2}\tilde{\alpha}\phi\left(\tilde{\alpha}+a^{(k)}\right)d\tilde{\alpha}\right]^{-1}s \\
= \; & p^{(k)\prime}+p^{(k)\prime}\left[\int_{\tilde{\alpha}}\frac{p^{(k)\prime}\sigma_\alpha\exp\left[p^{(k)\prime}\sigma_\alpha\left(\tilde{\alpha}+a^{(k)\prime}\right)\right]}{\left[1+\exp\left[p^{(k)\prime}\sigma_\alpha\left(\tilde{\alpha}+a^{(k)\prime}\right)\right]\right]^2}\tilde{\alpha}\phi\left(\tilde{\alpha}+a^{(k)\prime}\right)d\tilde{\alpha}\right]^{-1}s' \\
& p^{(k)}+p^{(k)}\left[\int_{\tilde{\alpha}}\frac{p^{(k)}\sigma_{\alpha 0}\exp\left[p^{(k)}\sigma_{\alpha 0}\left(\tilde{\alpha}+a_0^{(k)}\right)\right]}{\left[1+\exp\left[p^{(k)}\sigma_{\alpha 0}\left(\tilde{\alpha}+a_0^{(k)}\right)\right]\right]^2}\tilde{\alpha}\phi\left(\tilde{\alpha}+a_0^{(k)}\right)d\tilde{\alpha}\right]^{-1}s \\
= \; & p^{(k)\prime}+p^{(k)\prime}\left[\int_{\tilde{\alpha}}\frac{p^{(k)\prime}\sigma_{\alpha 0}\exp\left[p^{(k)\prime}\sigma_{\alpha 0}\left(\tilde{\alpha}+a_0^{(k)\prime}\right)\right]}{\left[1+\exp\left[p^{(k)\prime}\sigma_{\alpha 0}\left(\tilde{\alpha}+a_0^{(k)\prime}\right)\right]\right]^2}\tilde{\alpha}\phi\left(\tilde{\alpha}-\eta\right)d\tilde{\alpha}\right]^{-1}s'
\end{aligned}$$

where $a^{(k)}$ satisfies

$$\int_{\tilde{\alpha}}\frac{\exp\left(p^{(k)}\sigma_\alpha\left(\tilde{\alpha}+a^{(k)}\right)\right)}{\left[1+\exp\left(p^{(k)}\sigma_\alpha\left(\tilde{\alpha}+a^{(k)}\right)\right)\right]}\phi\left(\tilde{\alpha}-\eta\right)d\tilde{\alpha}=s$$

and$a_0^{(k)}$ satisfies

$$\int_{\tilde{\alpha}}\frac{\exp\left(p^{(k)}\sigma_{\alpha 0}\left(\tilde{\alpha}+a_0^{(k)}\right)\right)}{\left[1+\exp\left(p^{(k)}\sigma_{\alpha 0}\left(\tilde{\alpha}+a_0^{(k)}\right)\right)\right]}\phi\left(\tilde{\alpha}-\eta\right)d\tilde{\alpha}=s$$

and $a^{(k)\prime}$ satisfying

$$\int_{\tilde{\alpha}}\frac{\exp\left(p^{(k)\prime}\sigma_\alpha\left(\tilde{\alpha}+a^{(k)\prime}\right)\right)}{\left[1+\exp\left(p^{(k)\prime}\sigma_\alpha\left(\tilde{\alpha}+a^{(k)\prime}\right)\right)\right]}\phi\left(\tilde{\alpha}-\eta\right)d\tilde{\alpha}=s'$$

and$a_0^{(k)\prime}$ satisfies

$$\int_{\tilde{\alpha}}\frac{\exp\left(p^{(k)\prime}\sigma_{\alpha 0}\left(\tilde{\alpha}+a_0^{(k)\prime}\right)\right)}{\left[1+\exp\left(p^{(k)\prime}\sigma_{\alpha 0}\left(\tilde{\alpha}+a_0^{(k)\prime}\right)\right)\right]}\phi\left(\tilde{\alpha}-\eta\right)d\tilde{\alpha}=s'$$

Therefore,

$$\begin{aligned}
\frac{p^{(k+1)\prime}}{p^{(k+1)}} &= \frac{1+\left[\int_{\tilde{\alpha}}\frac{p^{(k+1)}\sigma_\alpha\exp\left[p^{(k+1)}\sigma_\alpha\left(\tilde{\alpha}+a^{(k+1)}\right)\right]}{\left[1+\exp\left[p^{(k+1)}\sigma_\alpha\left(\tilde{\alpha}+a^{(k+1)}\right)\right]\right]^2}\tilde{\alpha}\phi\left(\tilde{\alpha}-\eta\right)d\tilde{\alpha}\right]^{-1}s}{1+\left[\int_{\tilde{\alpha}}\frac{p^{(k+1)\prime}\sigma_\alpha\exp\left[p^{(k+1)\prime}\sigma_\alpha\left(\tilde{\alpha}+a^{(k+1)\prime}\right)\right]}{\left[1+\exp\left[p^{(k+1)\prime}\sigma_\alpha\left(\tilde{\alpha}+a^{(k+1)\prime}\right)\right]\right]^2}\tilde{\alpha}\phi\left(\tilde{\alpha}-\eta\right)d\tilde{\alpha}\right]^{-1}s'} \\
&= \frac{1+\left[\int_{\tilde{\alpha}}\frac{p^{(k)}\sigma_\alpha\exp\left[p^{(k)}\sigma_\alpha\left(\tilde{\alpha}+a^{(k)}\right)\right]}{\left[1+\exp\left[p^{(k)}\sigma_\alpha\left(\tilde{\alpha}+a^{(k)}\right)\right]\right]^2}\tilde{\alpha}\phi\left(\tilde{\alpha}-\eta\right)d\tilde{\alpha}\right]^{-1}s}{1+\left[\int_{\tilde{\alpha}}\frac{p^{(k)\prime}\sigma_\alpha\exp\left[p^{(k)\prime}\sigma_\alpha\left(\tilde{\alpha}+a^{(k)\prime}\right)\right]}{\left[1+\exp\left[p^{(k)\prime}\sigma_\alpha\left(\tilde{\alpha}+a^{(k)\prime}\right)\right]\right]^2}\tilde{\alpha}\phi\left(\tilde{\alpha}-\eta\right)d\tilde{\alpha}\right]^{-1}s'}=\frac{p^{(k)\prime}}{p^{(k)}}=...=\frac{p'}{p}
\end{aligned} \tag{29}$$

and by taking the limit,

$$lim_{k\rightarrow\infty}\frac{1+\left[\int_{\tilde{\alpha}}\frac{p^{(k)}\sigma_\alpha\exp\left[p^{(k)}\sigma_\alpha\left(\tilde{\alpha}+a^{(k)}\right)\right]}{\left[1+\exp\left[p^{(k)}\sigma_\alpha\left(\tilde{\alpha}+a^{(k)}\right)\right]\right]^2}\tilde{\alpha}\phi\left(\tilde{\alpha}-\eta\right)d\tilde{\alpha}\right]^{-1}s}{1+\left[\int_{\tilde{\alpha}}\frac{p^{(k)\prime}\sigma_\alpha\exp\left[p^{(k)\prime}\sigma_\alpha\left(\tilde{\alpha}+a^{(k)\prime}\right)\right]}{\left[1+\exp\left[p^{(k)\prime}\sigma_\alpha\left(\tilde{\alpha}+a^{(k)\prime}\right)\right]\right]^2}\tilde{\alpha}\phi\left(\tilde{\alpha}-\eta\right)d\tilde{\alpha}\right]^{-1}s'}=\frac{1+\left[\left(\eta-\Phi^{-1}\left(s\right)\right)\phi\left(\Phi^{-1}\left(s\right)\right)\right]^{-1}s}{1+\left[\left(\eta-\Phi^{-1}\left(s'\right)\right)\phi\left(\Phi^{-1}\left(s'\right)\right)\right]^{-1}s'}=\frac{p'}{p}.$$

Thus, for any k

$$G\equiv\frac{p'}{p}=\frac{1+\left[\left(\eta-\Phi^{-1}\left(s\right)\right)\phi\left(\Phi^{-1}\left(s\right)\right)\right]^{-1}s}{1+\left[\left(\eta-\Phi^{-1}\left(s'\right)\right)\phi\left(\Phi^{-1}\left(s'\right)\right)\right]^{-1}s'}=\frac{1+\left[\int_{\tilde{\alpha}}\frac{p^{(k)}\sigma_\alpha\exp\left[p^{(k)}\sigma_\alpha\left(\tilde{\alpha}+a^{(k)}\right)\right]}{\left[1+\exp\left[p^{(k)}\sigma_\alpha\left(\tilde{\alpha}+a^{(k)}\right)\right]\right]^2}\tilde{\alpha}\phi\left(\tilde{\alpha}-\eta\right)d\tilde{\alpha}\right]^{-1}s}{1+\left[\int_{\tilde{\alpha}}\frac{p^{(k)\prime}\sigma_\alpha\exp\left[p^{(k)\prime}\sigma_\alpha\left(\tilde{\alpha}+a^{(k)\prime}\right)\right]}{\left[1+\exp\left[p^{(k)\prime}\sigma_\alpha\left(\tilde{\alpha}+a^{(k)\prime}\right)\right]\right]^2}\tilde{\alpha}\phi\left(\tilde{\alpha}-\eta\right)d\tilde{\alpha}\right]^{-1}s'}.$$

Hence, if we denote

$$\begin{aligned}
B^{(k)} &= \int_\alpha \frac{p^{(k)}\sigma_\alpha \exp\left[p^{(k)}\sigma_\alpha\left(\tilde{\alpha}+a^{(k)}\right)\right]}{\left[1+\exp\left[p^{(k)}\sigma_\alpha\left(\tilde{\alpha}+a^{(k)}\right)\right]\right]^2}\tilde{\alpha}\phi\left(\tilde{\alpha}-\eta\right)d\tilde{\alpha}, \\
B'^{(k)} &= \int_\alpha \frac{p'^{(k)}\sigma_\alpha \exp\left[p'^{(k)}\sigma_\alpha\left(\tilde{\alpha}+a^{(k)\prime}\right)\right]}{\left[1+\exp\left[p'^{(k)}\sigma_\alpha\left(\tilde{\alpha}+a^{(k)\prime}\right)\right]\right]^2}\tilde{\alpha}\phi\left(\tilde{\alpha}-\eta\right)d\tilde{\alpha} \\
\frac{1+B^{(k)-1}s}{1+B'^{(k)-1}s'} &= \frac{p'}{p} \equiv G \\
B'^{(k)}s - GB^{(k)}s' &= B^{(k)}B'^{(k)}\left(G-1\right)
\end{aligned}$$

Now, denote $B^{(k)} = B + B\left(p^{(k)}\right)$, $B'^{(k)} = B' + B'\left(Gp^{(k)}\right)$, where $B = lim_{p\to\infty}B\left(p\right) = \left(\eta - \Phi^{-1}\left(s\right)\right)\phi\left(\Phi^{-1}\left(s\right)\right)$, $B' = lim_{p\to\infty}B'\left(p\right) = \left(\eta - \Phi^{-1}\left(s'\right)\right)\phi\left(\Phi^{-1}\left(s'\right)\right)$.

$$\left(B' + B'\left(Gp^{(k)}\right)\right)s - G\left(B + B\left(p^{(k)}\right)\right)s' = \left(B + B\left(p^{(k)}\right)\right)\left(B' + B'\left(Gp^{(k)}\right)\right)\left(G-1\right)$$

Because

$$\begin{aligned}
B's - GBs' &= BB'\left(G-1\right), \\
B'\left(Gp^{(k)}\right)s - GB\left(p^{(k)}\right)s' &= \left[BB'\left(Gp^{(k)}\right) + B'B\left(p^{(k)}\right) + B'\left(Gp^{(k)}\right)B\left(p^{(k)}\right)\right]\left(G-1\right) \\
\frac{s - B\left(G-1\right)}{B\left(p^{(k)}\right)} &= \frac{B'\left(G-1\right) + Gs'}{B'\left(Gp^{(k)}\right)} + \left(G-1\right)
\end{aligned}$$

Now, from equation (25), we know that

$$\begin{aligned}
B\left(p\right) &= \left[\left(-\phi\left(-a-\eta\right) + a\phi'\left(-a-\eta\right)\right)\frac{\phi'\left(-a-\eta\right)}{\phi\left(-a-\eta\right)} + \phi'\left(-a-\eta\right) - \frac{a}{2}\phi''\left(-a-\eta\right)\right]\frac{\pi^2}{6p^2} + O\left(p^{-4}\right) \\
B'\left(p\right) &= \left[\left(-\phi\left(-a'-\eta\right) + a'\phi'\left(-a'-\eta\right)\right)\frac{\phi'\left(-a'-\eta\right)}{\phi\left(-a'-\eta\right)} + \phi'\left(-a'-\eta\right) - \frac{a'}{2}\phi''\left(-a'-\eta\right)\right]\frac{\pi^2}{6G^2p^2} + O\left(p^{-4}\right)
\end{aligned}$$

Therefore, we can write the above equation as

$$\frac{s - B\left(G-1\right)}{p^{(k)2}b\left(p^{(k)}\right)} = \frac{B'\left(G-1\right) + Gs'}{\left(G^2p^{(k)2}\right)b'\left(Gp^{(k)}\right)} + \left(G-1\right)$$

where

$$\begin{aligned}
b\left(p\right) &= \left[\left(-\phi\left(-a-\eta\right) + a\phi'\left(-a-\eta\right)\right)\frac{\phi'\left(-a-\eta\right)}{\phi\left(-a-\eta\right)} + \phi'\left(-a-\eta\right) - \frac{a}{2}\phi''\left(-a-\eta\right)\right]\frac{\pi^2}{6} + O\left(p^{-2}\right) \\
b'\left(p\right) &= \left[\left(-\phi\left(-a'-\eta\right) + a'\phi'\left(-a'-\eta\right)\right)\frac{\phi'\left(-a'-\eta\right)}{\phi\left(-a'-\eta\right)} + \phi'\left(-a'-\eta\right) - \frac{a'}{2}\phi''\left(-a'-\eta\right)\right]\frac{\pi^2}{6G^2} + O\left(p^{-2}\right).
\end{aligned}$$

Hence,

$$\left[s - B\left(G-1\right)\right]b'\left(Gp^{(k)}\right) - \frac{\left[B'\left(G-1\right) + Gs'\right]b\left(p^{(k)}\right)}{G^2} = \left(G-1\right)p^{(k)2}b'\left(Gp^{(k)}\right)b\left(p^{(k)}\right)$$

Notice that $B'\left(G-1\right) + Gs' = B's/B$, $\left(s - B\left(G-1\right)\right) = GBs'/B'$. Hence,

$$\frac{GBs'}{B'}b'\left(Gp^{(k)}\right) - \frac{B'sb\left(p^{(k)}\right)}{BG^2} = \left(G-1\right)p^{(k)2}b'\left(Gp^{(k)}\right)b\left(p^{(k)}\right)$$

The RHS is a linear function of $p^{(k)2}$ and $p^{(k)2}O\left(p^{-(k)2}\right)$, whereas the LHS is a linear function of constant and $O\left(p^{-(k)2}\right)$. Therefore, in order for the equation to hold, either $G = 1$, or $b'\left(Gp^{(k)}\right) = O\left(p^{-(k)2}\right)$ or $b\left(p^{(k)}\right) = O\left(p^{-(k)2}\right)$ has to hold for any large k. That is, either $G = 1$ (implies $p^{(k)} = p'^{(k)}$), or $b'\left(Gp^{(k)}\right) = O\left(p^{-(k)2}\right)$ or $b\left(p^{(k)}\right) = O\left(p^{-(k)2}\right)$. Now, consider $b\left(p^{(k)}\right)$, whose constant term is

$$\left(-\phi\left(-a-\eta\right) + a\phi'\left(-a-\eta\right)\right)\frac{\phi'\left(-a-\eta\right)}{\phi\left(-a-\eta\right)} + \phi'\left(-a-\eta\right) - \frac{a}{2}\phi''\left(-a-\eta\right) = \left[\left(a+\eta\right)^2 + 1\right]\frac{a}{2}\phi\left(-a-\eta\right),$$

which is not 0 unless $\phi(-a-\eta)=0$, i.e. $a=\pm\infty$ or $a=0$. If $a=\pm\infty$, then $s=\Phi(a+\eta)$ either equals 1 or 0. If $a=0$, then $s=\Phi(\eta)$. That is, in order for $b\left(p^{(k)}\right)=O\left(p^{-(k)2}\right)$ to hold at high k, s has to be either close to 0 or 1, or close to $\Phi(\eta)$. Similarly for s'. Therefore, as long as both s, s' take on values that are not close to 0, or 1, or $\Phi(\eta)$, equation (29) cannot hold for large k. Therefore, for those s, s', and $p\neq p'$, claim holds.

Next, consider the case for $\sigma_\alpha>\sigma_{\alpha 0}$. Similarly, we generate a decreasing sequence of prices $\left(p^{(0)},p^{(0)\prime}\right)$, $\left(p^{(1)},p^{(1)\prime}\right)$,...,$\left(p^{(k)},p^{(k)\prime}\right)$,... such that $\left(p^{(0)},p^{(0)\prime}\right)=(p,p')$, $p^{(k)}=(\sigma_{\alpha 0}/\sigma_\alpha)^k p<p^{(k-1)}$. Then, consider an arbitrarily large $\left(p^{(0)},p^{(0)\prime}\right)=(p,p')$. Then, as before, we can show identification. Therefore, claim holds. ■

B.3 Proof of Lemma 4

Proof. The identification of the price coefficient α of the logit model of demand with covariates $\mathbf{X}_m$ is the same as in the proof of Lemma 3. Next, consider including the observed product characteristics into the BLP random coefficient model. Since we have shown in Lemma 3 that Assumption 11 is satisfied for μ_α,σ_α for the data with $\mathbf{X}_m=0$, we assume that μ_α,σ_α are identified. Here, for simplicity, we assume its dimension to be one, and denote it as X. Then,

$$\int_\alpha\frac{\exp(\xi+p\alpha+X\beta)}{1+\exp(\xi+p\alpha+X\beta)}\frac{1}{\sigma_\alpha}\phi\left(\frac{\alpha-\mu_\alpha}{\sigma_\alpha}\right)\frac{1}{\sigma_\beta}\phi\left(\frac{\beta-\mu_\beta}{\sigma_\beta}\right)d\alpha d\beta=\int_\alpha\frac{\exp(\xi+p(\alpha+A_X\beta))}{1+\exp(\xi+p(\alpha+A_X\beta))}\frac{1}{\sigma_\alpha}\phi\left(\frac{\alpha-\mu_\alpha}{\sigma_\alpha}\right)\frac{1}{\sigma_\beta}\phi\left(\frac{\beta-\mu_\beta}{\sigma_\beta}\right)d\alpha d\beta.$$

Now, for identification, we choose in the data only firms whose observed control X is highly correlated with the observed price p. That is, we choose data with X such that $X=A_Xp$ for some positive constant A_X. Since α and β are assumed to be normally distributed and independent, $\gamma=\alpha+A_X\beta\sim N(\mu_\gamma,\sigma_\gamma)$, where $\mu_\gamma=\mu_\alpha+A_X\mu_\beta$, and $\sigma_\gamma=\sqrt{\sigma_\alpha^2+A_X^2\sigma_\beta^2}$. Similarly, we choose the other constant $A_X'\neq A_X$ such that $X'=A_X'p'$. Without loss of generality, we assume $A_X'>A_X\geq 0$. Then, as before, find s, s', p and p' such that

$$\int_{\alpha,\beta}\frac{\exp(\xi+p\alpha+X\beta)}{1+\exp(\xi+p\alpha+X\beta)}\frac{1}{\sigma_\alpha}\phi\left(\frac{\alpha-\mu_\alpha}{\sigma_\alpha}\right)\frac{1}{\sigma_\beta}\phi\left(\frac{\beta-\mu_\beta}{\sigma_\beta}\right)d\alpha d\beta=\int_\gamma\frac{p\exp(p(\gamma+\xi/p))}{1+\exp(p(\gamma+\xi/p))}\frac{1}{\sigma_\gamma}\phi\left(\frac{\gamma-\mu_\gamma}{\sigma_\gamma}\right)d\gamma=s.$$

and

$$\int_\gamma\frac{\exp(p'(\gamma+\xi'/p'))}{1+\exp(p'(\gamma+\xi'/p'))}\frac{1}{\sigma_\gamma}\phi\left(\frac{\gamma-\mu_\gamma}{\sigma_\gamma}\right)d\gamma=s'$$

and if we denote $\sigma_{X\beta}=A_X\sigma_\beta$, $\mu_{X\beta}=A_X\mu_\beta$, the corresponding marginal revenue equations are:

$$\begin{aligned}MR &= p+p\left[\int_\gamma\int_\alpha\frac{p\exp(p(\gamma+\xi/p))}{[1+\exp(p(\gamma+\xi/p))]^2}\alpha\frac{1}{\sigma_\alpha}\phi\left(\frac{\alpha-\mu_\alpha}{\sigma_\alpha}\right)\frac{1}{\sigma_{X\beta}}\phi\left(\frac{(\gamma-\alpha)-\mu_{X\beta}}{\sigma_{X\beta}}\right)d\alpha d\gamma\right]^{-1}s\\ &= p'+p'\left[\int_\gamma\int_\alpha\frac{p'\exp(p'(\gamma+\xi'/p'))}{[1+\exp(p'(\gamma+\xi'/p'))]^2}\alpha\frac{1}{\sigma_\alpha}\phi\left(\frac{\alpha-\mu_\alpha}{\sigma_\alpha}\right)\frac{1}{\sigma_{X'\beta}}\phi\left(\frac{(\gamma-\alpha)-\mu_{X'\beta}}{\sigma_{X'\beta}}\right)d\alpha d\gamma\right]^{-1}s'\end{aligned}$$

Now, we have

$$\begin{aligned}&\left(\frac{\alpha-\mu_\alpha}{\sigma_\alpha}\right)^2+\left(\frac{\alpha+\mu_{X\beta}-\gamma}{\sigma_{X\beta}}\right)^2=\left(\frac{\alpha-\mu_\alpha}{\sigma_\alpha}\right)^2+\left(\frac{\alpha-\mu_\alpha+\mu_\gamma-\gamma}{\sigma_{X\beta}}\right)^2\\ =\;&\frac{1}{\sigma_{X\beta}^2}\left[\left(\frac{\sigma_\gamma}{\sigma_\alpha}(\alpha-\mu_\alpha)\right)^2-2(\alpha-\mu_\alpha)(\gamma-\mu_\gamma)+(\gamma-\mu_\gamma)^2\right]\\ =\;&\frac{1}{\sigma_{X\beta}^2}\left[\frac{\sigma_\gamma}{\sigma_\alpha}(\alpha-\mu_\alpha)-\left(\frac{\sigma_\alpha}{\sigma_\gamma}\right)(\gamma-\mu_\gamma)\right]^2-\left[\frac{\sigma_\alpha^2-\sigma_\gamma^2}{\sigma_{X\beta}^2\sigma_\gamma^2}\right](\gamma-\mu_\gamma)^2=\frac{\sigma_\gamma^2}{\sigma_{X\beta}^2\sigma_\alpha^2}\left[\alpha-\mu_\alpha-\left(\frac{\sigma_\alpha}{\sigma_\gamma}\right)^2(\gamma-\mu_\gamma)\right]^2+\left[\frac{1}{\sigma_\gamma^2}\right](\gamma-\mu_\gamma)^2\end{aligned}$$

where, if we set $g(p)=\xi/(\sigma_\gamma p)$,

$$
\begin{aligned}
&\int_\alpha \frac{p\exp\left(p\left(\gamma+\xi/p\right)\right)}{\left[1+\exp\left(p\left(\gamma+\xi/p\right)\right)\right]^2}\alpha\frac{1}{\sigma_\alpha}\phi\left(\frac{\alpha-\mu_\alpha}{\sigma_\alpha}\right)\frac{1}{\sigma_{X\beta}}\phi\left(\frac{\left(\gamma-\alpha\right)-\mu_{X\beta}}{\sigma_{X\beta}}\right)d\alpha \\
= \;& exp\left(-\frac{1}{2}\left(\frac{\gamma-\mu_\gamma}{\sigma_\gamma}\right)^2\right)\frac{1}{\sigma_\alpha\sigma_{X\beta}}\frac{p\exp\left(p\left(\gamma+\sigma_\gamma g\left(p\right)\right)\right)}{\left[1+\exp\left(p\left(\gamma+\sigma_\gamma g\left(p\right)\right)\right)\right]^2}\int_\alpha \alpha exp\left(-\frac{1}{2}\left(\frac{\alpha-\mu_\alpha-\left(\frac{\sigma_\alpha}{\sigma_\gamma}\right)^2\left(\gamma-\mu_\gamma\right)}{\left(\sigma_{X\beta}\sigma_\alpha\right)/\sigma_\gamma}\right)^2\right)d\alpha \\
= \;& \frac{p\exp\left(p\left(\gamma+\sigma_\gamma g\left(p\right)\right)\right)}{\left[1+\exp\left(p\left(\gamma+\sigma_\gamma g\left(p\right)\right)\right)\right]^2}\frac{1}{\sigma_\gamma}\left[\mu_\alpha+\left(\frac{\sigma_\alpha}{\sigma_\gamma}\right)^2\left(\gamma-\mu_\gamma\right)\right]exp\left(-\frac{1}{2}\left(\frac{\gamma-\mu_\gamma}{\sigma_\gamma}\right)^2\right)
\end{aligned}
$$

Hence, for large p,

$$
\begin{aligned}
&\int_\gamma\int_\alpha \frac{p\exp\left(p\left(\gamma+\xi/p\right)\right)}{\left[1+\exp\left(p\left(\gamma+\xi/p\right)\right)\right]^2}\alpha\frac{1}{\sigma_\alpha}\phi\left(\frac{\alpha-\mu_\alpha}{\sigma_\alpha}\right)\frac{1}{\sigma_{X\beta}}\phi\left(\frac{\left(\gamma-\alpha\right)-\mu_{X\beta}}{\sigma_{X\beta}}\right)d\alpha d\gamma \\
= \;& \int_\gamma \frac{p\exp\left(p\left(\gamma+\sigma_\gamma g\left(p\right)\right)\right)}{\left[1+\exp\left(p\left(\gamma+\sigma_\gamma g\left(p\right)\right)\right)\right]^2}\frac{1}{\sigma_\gamma}\left[\mu_\alpha+\left(\frac{\sigma_\alpha}{\sigma_\gamma}\right)^2\left(\gamma-\mu_\gamma\right)\right]exp\left(-\frac{1}{2}\left(\frac{\gamma-\mu_\gamma}{\sigma_\gamma}\right)^2\right)d\gamma \\
= \;& \left[\frac{\mu_\alpha}{\sigma_\gamma}+\left(\frac{\sigma_\alpha}{\sigma_\gamma}\right)^2\left(-g\left(p\right)-\eta_\gamma\right)\right]exp\left(-\frac{1}{2}\left(-g\left(p\right)-\eta_\gamma\right)^2\right)+O\left(p^{-1}\right) \\
= \;& \left[\frac{\mu_\alpha}{\sigma_\gamma}-\left(\frac{\sigma_\alpha}{\sigma_\gamma}\right)^2\Phi^{-1}\left(s\right)\right]exp\left(-\frac{1}{2}\Phi^{-1}\left(s\right)^2\right)+O\left(p^{-1}\right).
\end{aligned}
$$

Furthermore,

$$
\int_\gamma \frac{\exp\left(p\sigma_\gamma\left(\tilde{\gamma}+g\left(p\right)\right)\right)}{1+\exp\left(p\sigma_\gamma\left(\tilde{\gamma}+g\left(p\right)\right)\right)}\phi\left(\tilde{\gamma}-\eta_\gamma\right)d\tilde{\gamma}=\Phi\left(g\left(p\right)+\eta_\gamma\right)+O\left(p^{-1}\right)=s,
$$

where

$$
\begin{gathered}
g=lim_{p\to\infty}g\left(p\right),\quad g=\Phi^{-1}\left(s\right)-\eta_\gamma, \\
\eta_\gamma=\frac{\mu_\gamma}{\sigma_\gamma}=\frac{\mu_\alpha+A_X\mu_\beta}{\sqrt{\sigma_\alpha^2+A_X^2\sigma_\beta^2}}.
\end{gathered}
$$

Now, take $A_X'>0$,$A_X=0$ and $s'=s$. Then, $\sigma_\gamma=\sigma_\alpha$. Choose large p, p' such that the two points have the same marginal revenue, i.e.

$$
\begin{aligned}
MR \;&=\; p+p\left[\left[\frac{\mu_\alpha}{\sigma_\alpha}-\Phi^{-1}\left(s\right)\right]exp\left(-\frac{1}{2}\Phi^{-1}\left(s\right)^2\right)+O\left(p^{-1}\right)\right]^{-1}s \\
&=\; p'+p'\left[\left[\frac{\mu_\alpha}{\sigma_\gamma'}-\left(\frac{\sigma_\alpha}{\sigma_\gamma'}\right)^2\Phi^{-1}\left(s\right)\right]exp\left(-\frac{1}{2}\Phi^{-1}\left(s\right)^2\right)+O\left(p'^{-1}\right)\right]^{-1}s. \qquad (30)
\end{aligned}
$$

Let $\nu'=1/\sigma_\gamma'$ be the precision of γ'. If we define $G=p'/p$, then for large p, p', the below equation is approximately satisfied.

$$
\frac{MR-p'}{p'}\left[\nu'\mu_\alpha-\sigma_\alpha^2\Phi^{-1}\left(s\right)\nu'^2\right]exp\left(-\frac{1}{2}\Phi^{-1}\left(s\right)^2\right)=s \qquad (31)
$$

$$
\sigma_\alpha^2\Phi^{-1}\left(s\right)\nu'^2-\mu_\alpha\nu'-sexp\left(\frac{1}{2}\Phi^{-1}\left(s\right)^2\right)\frac{p'}{p'-MR}=0
$$

, whose RHS is a function of v'. Because the constant term is negative, LHS is negative if $\nu'=0$. Therefore, if we choose $s>1/2$, $\Phi^{-1}\left(s\right)>0$, then one solution of ν' is positive and the other negative. Because v' has to be positive, there is only one value that satisfies the above equation. Since $\nu'=1/\sigma_\gamma'$ satisfying equation (31) can be made arbitrarily close to $\nu'=1/\sigma_\gamma'$ satisfying equation (30) by making p'arbitrarily large, $\sigma_\gamma'=\nu'^{-1}>0$ is identified. Furthermore, if $s=1/2$, $\Phi^{-1}\left(s\right)=0$, then

$$
1+1/\left[\frac{2\mu_\alpha}{\sigma_\alpha}\right]=G+G/\left[\frac{2\mu_\alpha}{\sigma_\gamma'}\right]
$$

holds approximately for large p, p' and thus, σ_γ' is identified in the same manner. Therefore, using data on market share satisfying $s\geq 1/2$, we can identify σ_β. ■

C Nonparametric Identification of Marginal Revenue.

C.1 Proof of Proposition 2

Proof.

a. Given the above model set-up, we can write the conditional expectation of the firm's total cost as:

$$\begin{aligned} E\left[C_m^d \middle| (q_m, p_m, s_m)\right] &= E\left[C_m + \eta_m \middle| (q_m, p_m, s_m)\right] \\ &= PC(q_m, MR(p_m, s_m)) \\ &= C_m \end{aligned}$$

where the first equality follows from Assumption 9, while the remaining equalities use Lemma 2 as well. Notice also that the pseudo-cost function can be expressed as a function of q, p and s because it is a function of p, $\mathbf{w}$ and MR, where $\mathbf{w}$ is suppressed and MR can be expressed as a function of p and s. For $q_m = q_{m'} = q$ it immediately follows from Lemma 2,

$$MR_m > MR_{m'} \Leftrightarrow E\left[C^d \middle| (q, p_m, s_m)\right] > E\left[C^d \middle| (q, p_{m'}, s_{m'})\right]$$

$$MR_m < MR_{m'} \Leftrightarrow E\left[C^d \middle| (q, p_m, s_m)\right] < E\left[C^d \middle| (q, p_{m'}, s_{m'})\right]$$

and

$$MR_m = MR_{m'} \Leftrightarrow E\left[C^d \middle| (q, p_m, s_m)\right] = E\left[C^d \middle| (q, p_{m'}, s_{m'})\right].$$

Therefore,

$$MR\left(p_m, s_m\right) = \zeta\left(q, E\left[C^d \middle| (q, p_m, s_m)\right]\right),$$

where ζ is an increasing and continuous function of the second element. That is, $E\left[C^d | (q, p_m, s_m)\right]$, is the nonparametric estimator of the relative ranking of the marginal revenue, given q.

b. Under the profit maximization assumption, $MR_i = MC_i$ at both points $i = 1, 2$. Given $Q_1 < Q_2$, it follows from the strict convexity of the cost function that

$$MR\left(p_1, \xi\right) < \frac{\partial C\left(Q_2 s_1, \upsilon\right)}{\partial q} \tag{32}$$

$C\left(Q_2 s_2, \upsilon\right)$ is the cost function specification, where $Q_2 s_2$ is the output and υ the cost shock. Furthermore, consider $\tilde{s}$ such that $Q_2 \tilde{s} = Q_1 s_1$, which implies $\tilde{s} < s_1$. From Assumption 12, there exists $\tilde{p} > p_1$ such that $\tilde{s} = s\left(\tilde{p}, \xi\right)$. Since, from Assumption 12, $MR\left(p, \xi\right)$ is strictly increasing in p,

$$MR\left(\tilde{p}, \xi\right) > \frac{\partial C\left(Q_2 \tilde{s}, \upsilon\right)}{\partial q} = \frac{\partial C\left(Q_1 s_1, \upsilon\right)}{\partial q} = MR\left(p_1, \xi\right). \tag{33}$$

It follows from equations (32) and (33), and the Intermediate Value Theorem that there exists $p_2 > p_1$ and s_2 such that $\tilde{s} < s_2 = s\left(p_2, \xi\right) < s_1$,

$$MR\left(p_2, \xi\right) = \frac{\partial C\left(Q_2 s_2, \upsilon\right)}{\partial q}$$

are satisfied. Furthermore, $q_1 = Q_2 \tilde{s} < Q_2 s_2 = q_2$. It is also straightforward to show that $s_2 - s_1$ is a continuous function of $Q_2 - Q_1$ given ξ and υ remaining unchanged.

To complete the proof of part b. it remains to show that,

$$p_1\left[1 + \frac{lnp_2 - lnp_1}{lns_2 - lns_1}\right] = \frac{E\left[C^d | (q_2, p_2, s_2)\right] - E\left[C^d | (q_1, p_1, s_1)\right]}{q_2 - q_1} + O\left(\left|Q_2 - Q_1\right|\right).$$

It is easy to show that the first order condition for profit maximization can be re-written as,

$$MR_1 = p_1\left[1 + \left(\frac{\partial lns\left(p_1, \xi\right)}{\partial lnp}\right)^{-1}\right] = MC_1 = \frac{\partial C\left(Q_1 s_1, \upsilon\right)}{\partial q},$$

where $\left(\frac{\partial lns(p_1,\xi)}{\partial lnp}\right)$ is the elasticity of demand. Marginal cost can be approximated using finite differences in total costs and quantities between points 1 and 2:

$$\frac{\partial C\left(Q_1 s_1, \upsilon\right)}{\partial q} = \frac{C\left(Q_2 s_2, \upsilon\right) - C\left(Q_1 s_1, \upsilon\right)}{Q_2 s_2 - Q_1 s_1} + O\left(\left|Q_2 s_2 - Q_1 s_1\right|\right) = \frac{C\left(Q_2 s_2, \upsilon\right) - C\left(Q_1 s_1, \upsilon\right)}{Q_2 s_2 - Q_1 s_1} + O\left(\left|Q_2 - Q_1\right|\right).$$

Now, notice that both price p and market share s can be expressed as a function of exogenous variables (Q, ξ, υ), i.e., $p = p(Q, \xi, \upsilon)$ and $s = s(Q, \xi, \upsilon)$, where we continue to simplify notation and suppress the dependence on observed product characteristics $\mathbf{x}$ and input prices $\mathbf{w}$. This is because, given p, ξ uniquely determins s, i.e., $s = s(p, \xi)$. Then, given Q, $q = Qs$ and ξ, υ uniquely determins p by,

$$MR = p\left[1 + \left(\frac{\partial lns\,(p, \xi)}{\partial lnp}\right)^{-1}\right] = MC = \frac{\partial C\,(q, \upsilon)}{\partial q},$$

Then, similarly as before, the elasticity of demand can be approximated using finite differences in prices and market shares between points 1 and 2:

$$\left(\frac{\partial lns\,(p_1, \xi)}{\partial lnp}\right)^{-1} = \frac{ln\,(p\,(Q_2, \xi, \upsilon)) - ln\,(p\,(Q_1, \xi, \upsilon))}{ln\,(s\,(Q_2, \xi, \upsilon)) - ln\,(s\,(Q_1, \xi, \upsilon))} + O\,(|Q_2 - Q_1|)\,.$$

The last part of the proposition immediately follows from the above re-written first order condition and these two approximations.

c. In proving the final part of the proposition, it is useful to distinguish between the true marginal cost and its estimate. We denote the true marginal cost as

$$MC_1 = \frac{\partial C\,(q_1, \upsilon_1)}{\partial q},$$

and $\widehat{MC}_1$ as the marginal cost estimate at (q_1, υ_1). From the first order condition, we know that the true marginal cost and marginal revenue are equal to each other. That is,

$$MC_1 = MR\,(p_1, \xi_1) = p_1\left[1 + \left[\frac{\partial lns\,(p_1, \xi_1)}{\partial lnp}\right]^{-1}\right],$$

which can be re-arranged to obtain the following equation,

$$\left(\frac{\partial lns\,(p_1, \xi_1)}{\partial lnp}\right)^{-1} = \frac{MC_1}{p_1} - 1.$$

Recall from our proof of part b that for sufficiently small $\Delta Q > 0$, the points $(p\,(Q_1 + \Delta Q, \xi_1, \upsilon_1)\,, s\,(Q_1 + \Delta Q, \xi_1, \upsilon_1))$ satisfy the following equation,

$$p\,(Q_1, \xi_1, \upsilon_1)\left[1 + \frac{ln\,(p\,(Q_1 + \Delta Q, \xi_1, \upsilon_1)) - ln\,(p\,(Q_1, \xi_1, \upsilon_1))}{ln\,(s\,(Q_1 + \Delta Q, \xi_1, \upsilon_1)) - ln\,(s\,(Q_1, \xi_1, \upsilon_1))}\right] = MC_1 + O\,((\Delta Q))\,.$$

Hence,

$$\frac{ln\,(p\,(Q_1 + \Delta Q, \xi_1, \upsilon_1)) - ln\,(p\,(Q_1, \xi_1, \upsilon_1))}{ln\,(s\,(Q_1 + \Delta Q, \xi_1, \upsilon_1)) - ln\,(s\,(Q_1, \xi_1, \upsilon_1))} = \frac{MC_1}{p\,(Q_1, \xi_1, \upsilon_1)} - 1 + O\,((\Delta Q))$$

Now, suppose that the estimated marginal cost is less than the true marginal cost, i.e., $\widehat{MC}_1 < MC_1$. Then, consider a vector of price and market share $(\hat{p}, \hat{s})$ such that $\hat{s} = s\,(Q_1 + \Delta Q, \xi_1, \upsilon_1)$[49] and $\hat{p}$ satisfy

$$p\,(Q_1, \xi_1, \upsilon_1)\left[1 + \frac{ln\,(\hat{p}) - ln\,(p\,(Q_1, \xi_1, \upsilon_1))}{ln\,(\hat{s}) - ln\,(s\,(Q_1, \xi_1, \upsilon_1))}\right] = \widehat{MC}_1.$$

That is,

$$\frac{ln\,(\hat{p}) - ln\,(p\,(Q_1, \xi_1, \upsilon_1))}{ln\,(\hat{s}) - ln\,(s\,(Q_1, \xi_1, \upsilon_1))} = \frac{\widehat{MC}_1}{p\,(Q_1, \xi_1, \upsilon_1)} - 1 < \frac{MC_1}{p\,(Q_1, \xi_1, \upsilon_1)} - 1 + O\,(\Delta Q)\,.$$

for sufficiently small $\Delta Q > 0$. Hence, for sufficiently small $\Delta Q > 0$, we have

$$\frac{ln\,(\hat{p}) - ln\,(p\,(Q_1, \xi_1, \upsilon_1))}{ln\,(\hat{s}) - ln\,(s\,(Q_1, \xi_1, \upsilon_1))} < \frac{ln\,(p\,(Q_1 + \Delta Q, \xi_1, \upsilon_1)) - ln\,(p\,(Q_1, \xi_1, \upsilon_1))}{ln\,(s\,(Q_1 + \Delta Q, \xi_1, \upsilon_1)) - ln\,(s\,(Q_1, \xi_1, \upsilon_1))} < 0.$$

Given $\hat{s} = s\,(Q_1 + \Delta Q, \xi_1, \upsilon_1) < s\,(Q_1, \xi_1, \upsilon_1)$ and Assumption 13, for the above inequality to hold it must follow that $\hat{p} > p\,(Q_1 + \Delta Q, \xi_1, \upsilon_1)$.

We now show that there exists such pair $(\hat{s}, \hat{p})$: specifically that there exists (ξ_2, υ_2) such that $\hat{s} = s\,(Q_1 + \Delta Q, \xi_2, \upsilon_2)$ and $\hat{p} = p\,(Q_1 + \Delta Q, \xi_2, \upsilon_2)$. For that, we need to show that ξ_2 satisfying $\hat{s} = s\,(\hat{p}, \xi_2)$ and υ_2 satisfying $MR\,(\hat{p}, \xi_2) = MC\,(\hat{s}\,(Q_1 + \Delta Q)\,, \upsilon_2)$ exist. $s\,(p, \xi)$ being a continuous and decreasing function of price, and

[49] Since $Q_1 + \Delta Q > Q_1$, $s\,(Q_1 + \Delta Q, \xi_1, \upsilon_1) < s\,(Q_1, \xi_1, \upsilon_1)$

$\hat{p} > p(Q_1 + \Delta Q, \xi_1, \upsilon_1)$ imply $s(\hat{p}, \xi_1) < s(p(Q_1 + \Delta Q, \xi_1, \upsilon_1), \xi_1)$. Since from Assumption 13, $lim_{\xi \uparrow \infty} s(\hat{p}, \xi) = 1 > s(p(Q_1 + \Delta Q, \xi_1, \upsilon_1), \xi_1)$, it follows from the Intermediate Value Theorem that there exists such $\xi_2 > \xi_1$.

Next, we show that there exists υ_2 that equates marginal revenue to marginal cost at $(\hat{p}, \hat{s})$. The marginal revenue of the point $(\hat{p}, s(\hat{p}, \xi_2))$ is

$$MR(\hat{p}, \xi_2) = \hat{p}\left[1 + \left(\frac{\partial lns(\hat{p}, \xi_2)}{\partial lnp}\right)^{-1}\right].$$

Since in Assumption 3', we assue that MC is an increasing and continuous function of υ, $lim_{\upsilon \downarrow 0} MC(\hat{s}(Q_1 + \Delta Q), \upsilon) = 0$ and $lim_{\upsilon \uparrow \infty} MC(\hat{s}(Q_1 + \Delta Q), \upsilon) = \infty$, again, by Intermediate Value Theorem, we can find such υ_2 that satisfies $MR(\hat{p}, \xi_2) = MC(\hat{s}(Q_1 + \Delta Q), \upsilon_2)$.

Figure 1 provides an illustrative exposition of the above argument, where $(p(Q_1 + \Delta Q, \xi_2, \upsilon_2), s(Q_1 + \Delta Q, \xi_2, \upsilon_2))$ is point F on the (incorrect) red demand curve one would infer based on the incorrect marginal cost estimate. Since the inferred demand curve (in red) has a steeper slope $\widehat{\left[\frac{\partial lnp}{\partial lns}\right]}$) than the true demand curve (in blue) in the figure, point F necessarily lies above the true demand curve going through point E. Because $\hat{s} = s(p(Q_1 + \Delta Q, \xi_2, \upsilon_2), \xi_2) = s(p(Q_1 + \Delta Q, \xi_1, \upsilon_1), \xi_1)$ and $p(Q_1 + \Delta Q, \xi_2, \upsilon_2) > p(Q_1 + \Delta Q, \xi_1, \upsilon_1)$, from Assumption 12 we know that the marginal revenue is higher at such a point (e.g., point F in the example):

$$MR(p(Q_1 + \Delta Q, \xi_2, \upsilon_2), \xi_2) > MR(p(Q_1 + \Delta Q, \xi_1, \upsilon_1), \xi_1)$$

The two red downward sloping lines in Figure 2 are the (true) demand curve going through point F, and its marginal revenue curve. Furthermore,

$$s(Q_1 + \Delta Q, \xi_2, \upsilon_2)(Q_1 + \Delta Q) = s(Q_1 + \Delta Q, \xi_1, \upsilon_1)(Q_1 + \Delta Q) \equiv q_1 + \Delta q.$$

Therefore,

$$MR(p(Q_1 + \Delta Q, \xi_2, \upsilon_2), \xi_2) = \frac{\partial C(q_1 + \Delta q, \upsilon_2)}{\partial q} > \frac{\partial C(q_1 + \Delta q, \upsilon_1)}{\partial q} = MR(p(Q_1 + \Delta Q, \xi_1, \upsilon_1), \xi_1),$$

which implies that $\upsilon_2 > \upsilon_1$. The upward sloping red line in Figure 2 is the marginal cost curve with υ_2. Therefore,

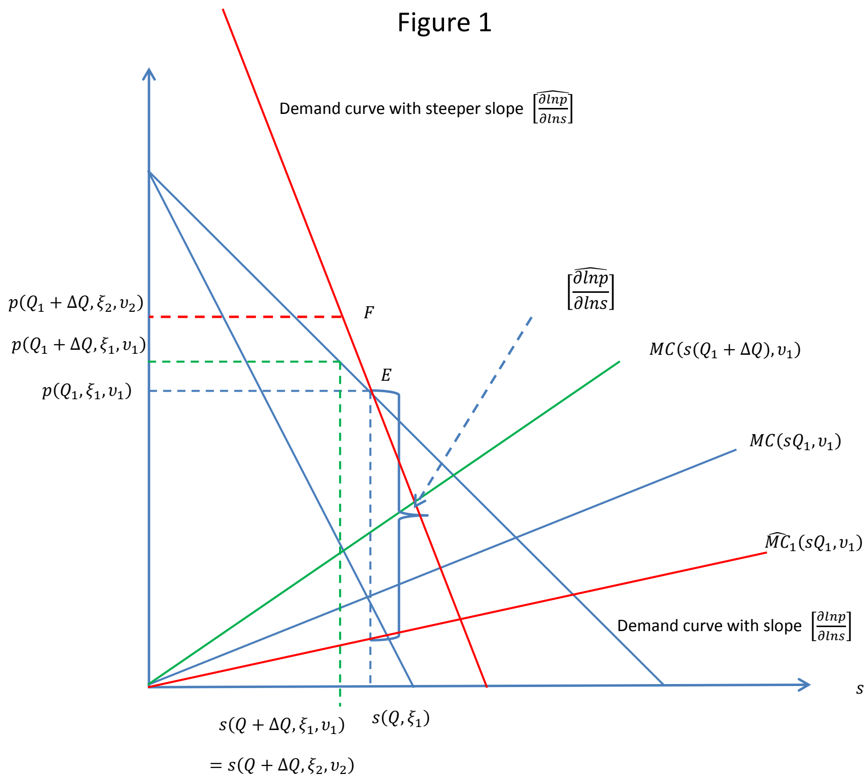
$$MC(q_1, \upsilon_2) > MC(q_1, \upsilon_1) > \widehat{MC}_1$$

and thus, for sufficiently small $\Delta q > 0$,

$$\frac{C(q_1 + \Delta q, \upsilon_2) - C(q_1, \upsilon_1)}{\Delta q} > \frac{C(q_1 + \Delta q, \upsilon_2) - C(q_1, \upsilon_2)}{\Delta q} > \widehat{MC}_1$$

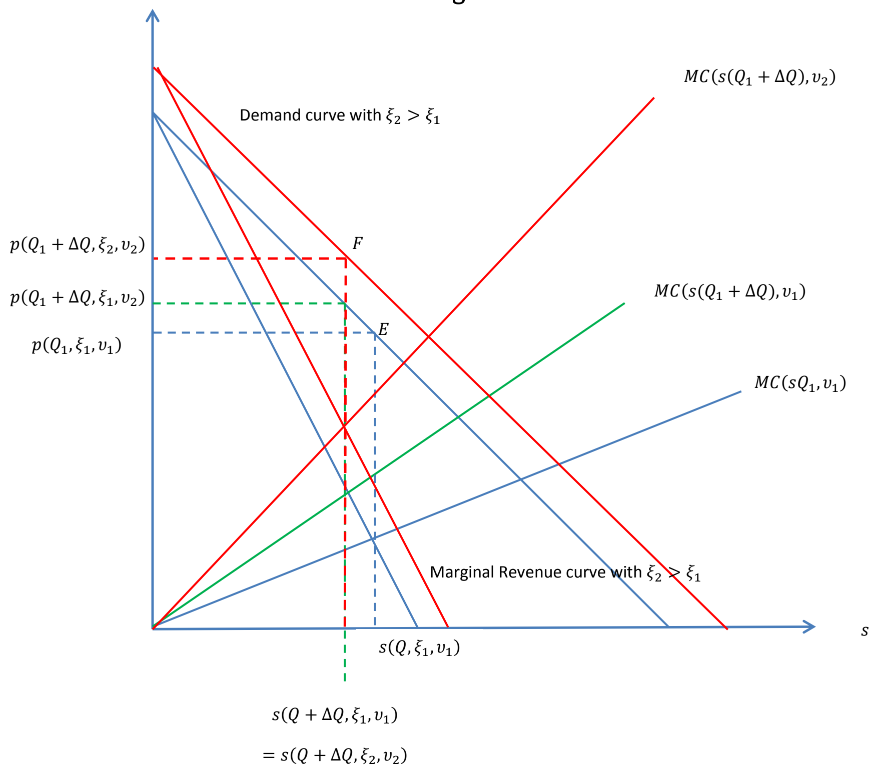
and $(C(q_1 + \Delta q, \upsilon_2) - C(q_1, \upsilon_1))/\Delta q - \widehat{MC}_1$ won't converge to zero as Δq goes to zero. Therefore, equation (17) does not hold. The proof for the case with the estimated marginal cost is greater than the true marginal cost (e.g., $\widehat{MC}_1 > MC_1$) follows similarly. ■

Figure 1



Demand curve with steeper slope $\left[\widehat{\frac{\partial lnp}{\partial lns}}\right]$
$\left[\widehat{\frac{\partial lnp}{\partial lns}}\right]$
$p(Q_1+\Delta Q,\xi_2,v_2)$
F
$p(Q_1+\Delta Q,\xi_1,v_1)$
$MC(s(Q_1+\Delta Q),v_1)$
E
$p(Q_1,\xi_1,v_1)$
$MC(sQ_1,v_1)$
$\widehat{MC}_1(sQ_1,v_1)$
Demand curve with slope $\left[\frac{\partial lnp}{\partial lns}\right]$
s
$s(Q+\Delta Q,\xi_1,v_1)$
$s(Q,\xi_1)$
$= s(Q+\Delta Q,\xi_2,v_2)$

Figure 2



C.2 Nonparametric Identification of Oligopoly Marginal Revenue.

Let $\mathbf{p}_m = (p_{1m}, \mathbf{p}_{-1m})$, where p_{1m} is the price of firm 1 in market m and $\mathbf{p}_{-1m}$ is the vector of prices of firms other than 1. Market share and other variables of market m are denoted similarly. Market share can be expressed as a function of price $\mathbf{p}_m$ and the unobserved product characteristics $\boldsymbol{\xi}_m$, i.e. $s(\mathbf{p}_m, \boldsymbol{\xi}_m)$. Let $s(\mathbf{p}_m, \boldsymbol{\xi}_m, j)$ be the corresponding market share of firm j in market m. Then, marginal revenue of firm 1 can be expressed as a function of own price p_{1m}, price of other firms $\mathbf{p}_{-1m}$, and the vector of unobserved product (firm) characteristics $\boldsymbol{\xi}_m$ of all firms in the market, i.e., $MR(p_{1m}, \mathbf{p}_{-1m}, \boldsymbol{\xi}_m, 1)$. Here, we again suppress observed product characteristics $\mathbf{X}$ and input price $\mathbf{w}$. They are assumed to be the same for all markets under consideration in the following Lemma. .

Next, we impose the following two assumptions, which are similar to Assumptions 11 and 12 for the monopoly case.

Assumption 14 *Marginal revenue is strictly increasing in own price p_{1m} given $\mathbf{p}_{-1m}$ and $\boldsymbol{\xi}_m$. Furthermore, suppose that we have two markets with $(\mathbf{p}_1, \mathbf{s}_1, \boldsymbol{\xi}_1)$ and $(\mathbf{p}_2, \mathbf{s}_2, \boldsymbol{\xi}_2)$, such that $s_{11} = s_{12} \equiv s$, $p_{11} > p_{12}$, and $\mathbf{p}_{-11} = \mathbf{p}_{-12} \equiv \mathbf{p}_{-1}$. Then,*

$$MR_1 (p_{11}, \mathbf{p}_{-1}, s, \mathbf{s}_{-11}, \boldsymbol{\xi}_1, 1) > MR(p_{12}, \mathbf{p}_{-1}, s, \mathbf{s}_{-12}, \boldsymbol{\xi}_2, 1).$$

Assumption 15 *Given the price $\mathbf{p}$ and the unobserved quality of other firms $\boldsymbol{\xi}_{-1}$, market share of firm 1, $s\left(\mathbf{p}, \xi, \boldsymbol{\xi}_{-1}, 1\right)$ is strictly increasing and continuous in ξ. Furthermore, given the other firms' prices $\mathbf{p}_{-1}$ and the unobserved characteristics $\boldsymbol{\xi}$, market share of firm 1, $s\left(p, \mathbf{p}_{-1}, \boldsymbol{\xi}, 1\right)$ is strictly decreasing and continuous in p. Furthermore,*

$$lim_{\xi \downarrow -\infty} s\left(\mathbf{p}, \xi, \boldsymbol{\xi}_{-1}, 1\right) = 0, \quad lim_{\xi \uparrow \infty} s\left(\mathbf{p}, \xi, \boldsymbol{\xi}_{-1}, 1\right) = 1 \ \ and \ \ lim_{p \uparrow \infty} s\left(p, \mathbf{p}_{-1}, \boldsymbol{\xi}, 1\right) = 0.$$

Lemma 5 *Suppose Assumptions 1,2,3',4,5 and Assumptions 13,14 are satisfied. Then,*

a. *Given q, the ordering of marginal revenue is nonparametrically identified from the cost data.*

b. *Suppose we have two points, $(Q_1, \mathbf{q}_1, \mathbf{p}_1, \mathbf{s}_1)$ and $(Q_2, \mathbf{q}_2, \mathbf{p}_2, \mathbf{s}_2)$, with the same demand shocks ($\boldsymbol{\xi}_1 = \boldsymbol{\xi}_2 = \boldsymbol{\xi}$), cost shocks that satisfy $\upsilon_{11} = \upsilon_{12} = \upsilon$, and different market sizes: $Q_1 < Q_2$. Then, there exist cost shocks $\boldsymbol{\upsilon}_{-11}$ and $\boldsymbol{\upsilon}_{-12}$ that are consistent with $\mathbf{p}_{-11} = \mathbf{p}_{-12} = \mathbf{p}_{-1}$. Furthermore, it follows that*

$$s_{11} > s_{12}, \quad p_{11} < p_{12}, \ \ q_{11} < q_{12}, \tag{34}$$

and

$$p_{11}\left[1 + \frac{lnp_{12} - lnp_{11}}{lns_{12} - lns_{11}}\right] = \frac{E\left[C^d \middle| (q_{12}, \mathbf{p}_2, \mathbf{s}_2)\right] - E\left[C^d \middle| (q_{11}, \mathbf{p}_1, \mathbf{s}_1)\right]}{q_{12} - q_{11}} + O\left(|Q_2 - Q_1|\right). \tag{35}$$

[50]

c. *Consider two close points, $(Q_1, \mathbf{q}_1, \mathbf{p}_1, \mathbf{s}_1)$ and $(Q_2, \mathbf{q}_2, \mathbf{p}_2, \mathbf{s}_2)$, such that*

$$Q_1 < Q_2, \ \ s_{11} > s_{12}, \ \ p_{11} < p_{12}, \ \ q_{11} < q_{12}, \ \ and \ \ \mathbf{p}_{-11} = \mathbf{p}_{-12} = \mathbf{p}_{-1},$$

and

$$p_{11}\left[1 + \frac{lnp_{12} - lnp_{11}}{lns_{12} - lns_{11}}\right] = \frac{E\left[C^d \middle| (q_{12}, \mathbf{p}_2, \mathbf{s}_2)\right] - E\left[C^d \middle| (q_{11}, \mathbf{p}_1, \mathbf{s}_1)\right]}{q_{12} - q_{11}} + O\left(|Q_2 - Q_1|\right).$$

Then, the true marginal cost at point 1, MC_1 satisfies

$$MC_1 = \frac{E\left[C^d \middle| (q_{12}, \mathbf{p}_2, \mathbf{s}_2)\right] - E\left[C^d \middle| (q_{11}, \mathbf{p}_1, \mathbf{s}_1)\right]}{q_{12} - q_{11}} + O\left(|Q_2 - Q_1|\right)$$

Proof.

a. The proof is the same as in Proposition 2, except for the price and market share being the vector $\mathbf{p}_m$, $\mathbf{s}_m$ instead of the scalar p_m, s_m.

b. Under the profit maximization assumption, $MR_{1m} = MC_{1m}$ at both markets $m = 1, 2$. Given $Q_2 > Q_1$, it follows from the strict convexity of the cost function,

$$MR\left(p_{11}, \mathbf{p}_{-1}, \boldsymbol{\xi}; 1\right) < \frac{\partial C\left(Q_2 s_{11}, \upsilon\right)}{\partial q} \tag{36}$$

Furthermore, consider $\tilde{s}$ such that $Q_2 \tilde{s} = Q_1 s_{11}$ which implies $\tilde{s} < s_{11}$. From Assumption 14, there exists $\tilde{p} > p_{11}$ such that $\tilde{s} = s\left(\tilde{p}, \mathbf{p}_{-1}, \boldsymbol{\xi}, 1\right)$. Since, from Assumption 13, $MR\left(p, \mathbf{p}_{-1}, \boldsymbol{\xi}, 1\right)$ is strictly increasing in p,

$$MR\left(\tilde{p}, \mathbf{p}_{-1}, \boldsymbol{\xi}, 1\right) > \frac{\partial C\left(Q_2 \tilde{s}, \upsilon\right)}{\partial q} = \frac{\partial C\left(Q_1 s_{11}, \upsilon\right)}{\partial q} = MR\left(p_{11}, \mathbf{p}_{-1}, \boldsymbol{\xi}, 1\right). \tag{37}$$

Because both marginal revenue and marginal cost functions are continuous, it follows from equations (36) and (37), and the Intermediate Value Theorem that there exists $p_{12} > p_{11}$ and s_{12} such that $\tilde{s} < s_{12} = s\left(p_{12}, \mathbf{p}_{-1}, \boldsymbol{\xi}, 1\right) < s_{11}$ and

$$MR\left(p_{12}, \mathbf{p}_{-1}, \boldsymbol{\xi}, 1\right) = \frac{\partial C\left(Q_2 s_{12}, \upsilon\right)}{\partial q}.$$

[50] The notation $C^d\left(q, \mathbf{p}, \mathbf{s}\right)$, where q is the own firm's (or, wlog. firm j's) output and $\mathbf{p}$, $\mathbf{s}$ are the vectors of prices and market shares of all firms in the market, is valid because, suppressing $\mathbf{w}$, the pseudo-cost function is defined to be $PC\left(q, MR_j\left(\mathbf{p}, \mathbf{s}\right)\right)$.

Then, $q_{12} = Q_2 s_{12} > Q_1 \tilde{s} = q_{11}$.
We also need to show that cost shocks $\boldsymbol{\upsilon}_{-11}$ and $\boldsymbol{\upsilon}_{-12}$ can be chosen at such level such that $\mathbf{p}_{-11} = \mathbf{p}_{-12} = \mathbf{p}_{-1}$ is satisfied. But this is obvious from Assumption 3', i.e. for any $j \neq 1$, one can find υ_{j1} such that

$$MR\left(p_{11}, \mathbf{p}_{-1}, \boldsymbol{\xi}, j\right) = \frac{\partial C\left(Q_1 s\left(p_{11}, \mathbf{p}_{-1}, \boldsymbol{\xi}, j\right), \upsilon_{j1}\right)}{\partial q}$$

and similarly, one can find υ_{j2} such that

$$MR\left(p_{12}, \mathbf{p}_{-1}, \boldsymbol{\xi}, j\right) = \frac{\partial C\left(Q_1 s\left(p_{12}, \mathbf{p}_{-1}, \boldsymbol{\xi}, j\right), \upsilon_{j2}\right)}{\partial q}.$$

Finally, it remains to show that,

$$p_{11}\left[1 + \frac{lnp_{12} - lnp_{11}}{lns_{12} - lns_{11}}\right] = \frac{E\left[C^d | \left(q_{12}, \mathbf{p}_2, \mathbf{s}_2\right)\right] - E\left[C^d | \left(q_{11}, \mathbf{p}_1, \mathbf{s}_1\right)\right]}{q_{12} - q_{11}} + O\left(|Q_2 - Q_1|\right).$$

It is easy to show that the first order condition for profit maximization for firm 1 can be re-written as,

$$p_1\left[1 + \left(\frac{\partial lns\left(p_1, \mathbf{p}_{-1}, \boldsymbol{\xi}, 1\right)}{\partial lnp}\right)^{-1}\right] = MC_1 = \frac{\partial C\left(Q_1 s_1, \upsilon\right)}{\partial q},$$

where $\left(\frac{\partial lns\left(p_1, \mathbf{p}_{-1} \boldsymbol{\xi}, 1\right)}{\partial lnp}\right)$ is the elasticity of demand. Further, marginal cost can be approximated using finite differences in total costs and quantities of firm 1 in markets 1 and 2,

$$\frac{\partial C\left(Q_1 s_{11}, \upsilon\right)}{\partial q} = \frac{C\left(Q_2 s_{12}, \upsilon\right) - C\left(Q_1 s_{11}, \upsilon\right)}{Q_2 s_{12} - Q_1 s_{11}} + O\left(|Q_2 s_{12} - Q_1 s_{11}|\right) = \frac{C\left(Q_2 s_{12}, \upsilon\right) - C\left(Q_1 s_{11}, \upsilon\right)}{Q_2 s_{12} - Q_1 s_{11}} + O\left(|Q_2 - Q_1|\right).$$

because both s_{11} and s_{12}, are continuous functions of Q_1, Q_2, respectively. Now, we can also express price and market share of firm 1 as functions of relevant exogenous variables (market size Q, demand shock $\boldsymbol{\xi}$, own cost shock υ) and the price of other firms $\mathbf{p}_{-1}$. The argument for this is similar as the one for the monopoly case. That is, given $\mathbf{p}_{-1}$ and p, from market share equation, $\boldsymbol{\xi}$ uniquely determines $\mathbf{s}_1$ as a function of p by $\mathbf{s}_1 = s\left(p, \mathbf{p}_{-1}, \boldsymbol{\xi}, 1\right)$. Then, given Q, $q = Q\mathbf{s}_1$, and υ uniquely determines q, p and $\mathbf{s}_1$ from the F.O.C:

$$p\left[1 + \left(\frac{\partial lns\left(p, \mathbf{p}_{-1}, \boldsymbol{\xi}, 1\right)}{\partial lnp}\right)^{-1}\right] = MC_1 = \frac{\partial C\left(q, \upsilon\right)}{\partial q}.$$

To see this, suppose that for the same Q,$\boldsymbol{\xi}$ and υ, we have two values of own price p^*, p' such that $p^* > p'$ that satify the F.O.C. Then, because marginal revenue is increasing in p, $MR^* > MR'$. On the other hand, since $\mathbf{s}_1^* < \mathbf{s}_1'$, $q^* < q'$, and thus, $MC^* < MC'$, contradicting the F.O.C. Therefore, the elasticity of demand can be approximated using finite differences in prices and market shares of firm 1 in markets 1 and 2,

$$\left(\frac{\partial lns\left(p_1, \mathbf{p}_{-1}, \boldsymbol{\xi}, 1\right)}{\partial lnp}\right)^{-1} = \frac{ln\left(p\left(Q_2, \boldsymbol{\xi}, \upsilon, \mathbf{p}_{-1}, 1\right)\right) - ln\left(p\left(Q_1, \boldsymbol{\xi}, \upsilon, \mathbf{p}_{-1}, 1\right)\right)}{ln\left(s\left(Q_2, \boldsymbol{\xi}, \upsilon, \mathbf{p}_{-1}, 1\right)\right) - ln\left(s\left(Q_1, \boldsymbol{\xi}, \upsilon, \mathbf{p}_{-1}, 1\right)\right)} + O\left(|Q_2 - Q_1|\right).$$

The last part of the proposition immediately follows from the above re-written first order condition and these two approximations.
c. In proving the final part of the proposition it is useful to distinguish between the true marginal cost and its estimate. Denote the true marginal cost as

$$MC_1 = \frac{\partial C\left(q_{11}, \upsilon_1\right)}{\partial q_{11}},$$

and let $\widehat{MC}_1$ be the marginal cost estimate at $\left(q_{11}, \upsilon_1\right)$. From the first order condition we know that true marginal cost and marginal revenue must be equal

$$MC_1 = MR\left(p_{11}, \mathbf{p}_{-1}, \boldsymbol{\xi}_1, 1\right) = p_{11}\left[1 + \frac{\partial lns\left(p_{11}, \mathbf{p}_{-1}, \boldsymbol{\xi}_1, 1\right)}{\partial lnp_{11}}\right]^{-1},$$

which can be re-arranged to obtain the following equation,

$$\left(\frac{\partial lns\left(p_{11}, \mathbf{p}_{-1}, \boldsymbol{\xi}_1, 1\right)}{\partial lnp_{11}}\right)^{-1} = \frac{MC_1}{p_{11}} - 1.$$

Recall from our proof of part b that for sufficiently small $\Delta Q > 0$, the points $(p(Q_1 + \Delta Q, \boldsymbol{\xi}_1, \upsilon_{11}, \mathbf{p}_{-1}, 1), s(Q_1 + \Delta Q, \boldsymbol{\xi}_1, \upsilon_{11}, \mathbf{p}_{-1}, 1))$ satisfy the following equation,

$$p(Q_1, \boldsymbol{\xi}_1, \upsilon_{11}, \mathbf{p}_{-1}, 1)\left[1 + \frac{ln(p(Q_1 + \Delta Q, \boldsymbol{\xi}_1, \upsilon_{11}, \mathbf{p}_{-1}, 1)) - ln(p(Q_1, \boldsymbol{\xi}_1, \upsilon_{11}, \mathbf{p}_{-1}, 1))}{ln(s(Q_1 + \Delta Q, \boldsymbol{\xi}_1, \upsilon_{11}, \mathbf{p}_{-1}, 1)) - ln(s(Q_1, \boldsymbol{\xi}_1, \upsilon_{11}, \mathbf{p}_{-1}, 1))}\right] = MC_1 + O((\Delta Q)).$$

Hence,

$$\frac{ln(p(Q_1 + \Delta Q, \boldsymbol{\xi}_1, \upsilon_{11}, \mathbf{p}_{-1}, 1)) - ln(p(Q_1, \boldsymbol{\xi}_1, \upsilon_{11}, \mathbf{p}_{-1}, 1))}{ln(s(Q_1 + \Delta Q, \boldsymbol{\xi}_1, \upsilon_{11}, \mathbf{p}_{-1}, 1)) - ln(s(Q_1, \boldsymbol{\xi}_1, \upsilon_{11}, \mathbf{p}_{-1}, 1))} = \frac{MC_1}{p(Q_1, \boldsymbol{\xi}_1, \upsilon_{11}, \mathbf{p}_{-1}, 1)} - 1 + O((\Delta Q))$$

Now, suppose that the estimated marginal cost for firm 1 is less than the true marginal cost, i.e., $\widehat{MC}_1 < MC_1$. Then, consider a vector of price and market share $(\hat{p}, \hat{s})$ such that $\hat{s} = s(Q_1 + \Delta Q, \boldsymbol{\xi}_1, \upsilon_{11}, \mathbf{p}_{-1}, 1)$ and $\hat{p}$ satisfies

$$p(Q_1, \boldsymbol{\xi}_1, \upsilon_{11}, \mathbf{p}_{-1}, 1)\left[1 + \frac{ln(\hat{p}) - ln(p(Q_1, \boldsymbol{\xi}_1, \upsilon_{11}, \mathbf{p}_{-1}, 1))}{ln(\hat{s}) - ln(s(Q_1, \boldsymbol{\xi}_1, \upsilon_{11}, \mathbf{p}_{-1}, 1))}\right] = \widehat{MC}_1.$$

$$\frac{ln(\hat{p}) - ln(p(Q_1, \boldsymbol{\xi}_1, \upsilon_{11}, \mathbf{p}_{-1}, 1))}{ln(\hat{s}) - ln(s(Q_1, \boldsymbol{\xi}_1, \upsilon_{11}, \mathbf{p}_{-1}, 1))} = \frac{\widehat{MC}_1}{p(Q_1, \boldsymbol{\xi}_1, \upsilon_{11}, \mathbf{p}_{-1}, 1)} - 1 < \frac{MC_1}{p(Q_1, \boldsymbol{\xi}_1, \upsilon_{11}, \mathbf{p}_{-1}, 1)} - 1 + O(\Delta Q).$$

Hence, for sufficiently small $\Delta Q > 0$, we have

$$\frac{ln(\hat{p}) - ln(p(Q_1, \boldsymbol{\xi}_1, \upsilon_{11}, \mathbf{p}_{-1}, 1))}{ln(\hat{s}) - ln(s(Q_1, \boldsymbol{\xi}_1, \upsilon_{11}, \mathbf{p}_{-1}, 1))} < \frac{ln(p(Q_1 + \Delta Q, \boldsymbol{\xi}_1, \upsilon_{11}, \mathbf{p}_{-1}, 1)) - ln(p(Q_1, \boldsymbol{\xi}_1, \upsilon_{11}, \mathbf{p}_{-1}, 1))}{ln(s(Q_1 + \Delta Q, \boldsymbol{\xi}_1, \upsilon_{11}, \mathbf{p}_{-1}, 1)) - ln(s(Q_1, \boldsymbol{\xi}_1, \upsilon_{11}, \mathbf{p}_{-1}, 1))} < 0.$$

Given $\hat{s} = s(Q_1 + \Delta Q, \boldsymbol{\xi}_1, \upsilon_{11}, \mathbf{p}_{-1}, 1) < s(Q_1, \boldsymbol{\xi}_1, \upsilon_{11}, \mathbf{p}_{-1}, 1)$, for the above inequality to hold, it must follow that $\hat{p} > p(Q_1 + \Delta Q, \boldsymbol{\xi}_1, \upsilon_{11}, \mathbf{p}_{-1}, 1)$.

We now show that there exists such pair $(\hat{s}, \hat{p})$, specifically that there exists $(\boldsymbol{\xi}_2, \boldsymbol{\upsilon}_2)$ such that $\xi_{12} > \xi_{11}$, $\boldsymbol{\xi}_{-12} = \boldsymbol{\xi}_{-11}$, and $\hat{s} = s(Q_1 + \Delta Q, \boldsymbol{\xi}_2, \upsilon_{12}, \mathbf{p}_{-1}, 1)$, $\hat{p} = p(Q_1 + \Delta Q, \boldsymbol{\xi}_2, \upsilon_{12}, \mathbf{p}_{-1}, 1)$. $s(p, \boldsymbol{\xi}, \mathbf{p}_{-1}, 1)$ being a continuous and decreasing function of price and $\hat{p} > p(Q_1 + \Delta Q, \boldsymbol{\xi}_1, \upsilon_{11}, \mathbf{p}_{-1}, 1)$ implies $s(\hat{p}, \boldsymbol{\xi}_1, \mathbf{p}_{-1}, 1) < s(p(Q_1 + \Delta Q, \boldsymbol{\xi}_1, \upsilon_{11}, \mathbf{p}_{-1}, 1), \boldsymbol{\xi}_1, \mathbf{p}_{-1}, 1)$. Since $s(p, \xi, \boldsymbol{\xi}_{-1}, \mathbf{p}_{-1}, 1)$ is continuous in ξ and $lim_{\xi\uparrow\infty} s(\hat{p}, \xi, \boldsymbol{\xi}_{-12}, \mathbf{p}_{-1}, 1) = 1 > s(p(Q_1 + \Delta Q, \boldsymbol{\xi}_1, \upsilon_{11}, \mathbf{p}_{-1}, 1), \boldsymbol{\xi}_1, \mathbf{p}_{-1}, 1)$, it follows from the Intermediate Value Theorem that there exists such $\boldsymbol{\xi}_2$.

Next, we show that there exists $\boldsymbol{\upsilon}_2$ that equates marginal revenue to marginal cost. The marginal revenue of the point $(\hat{p}, s(\hat{p}, \boldsymbol{\xi}_2), \mathbf{p}_{-1}, 1)$ is

$$MR(\hat{p}, \boldsymbol{\xi}_2, \mathbf{p}_{-1}, 1) = \hat{p}\left[1 + \left(\frac{\partial lns(\hat{p}, \boldsymbol{\xi}_2, \mathbf{p}_{-1}, 1)}{\partial lnp}\right)^{-1}\right].$$

Since MC is an increasing and continuous function of υ and $lim_{\upsilon\downarrow 0} MC(\hat{s}(Q_1 + \Delta Q), \upsilon) = 0$ and $lim_{\upsilon\uparrow\infty} MC(\hat{s}(Q_1 + \Delta Q), \upsilon) = \infty$, from Intermediate Value Theorem, there exists υ_{12} that satisfies

$$MR(\hat{p}, \boldsymbol{\xi}_2, \mathbf{p}_{-1}, 1) = MC(\hat{s}(Q_1 + \Delta Q), \upsilon_{12}),$$

Similarly, we can show that there exists υ_{j2}, $j \neq 1$ such that

$$MR(\hat{p}, \mathbf{p}_{-1}, \boldsymbol{\xi}_2, j) = \frac{\partial C(Q_1 s(\hat{p}, \mathbf{p}_{-1}, \boldsymbol{\xi}_2, j), \upsilon_{j2})}{\partial q}.$$

Because $\hat{s} = s(Q_1 + \Delta Q, \boldsymbol{\xi}_2, \upsilon_{12}, \mathbf{p}_{-1}, 1) = s(Q_1 + \Delta Q, \boldsymbol{\xi}_1, \upsilon_{11}, \mathbf{p}_{-1}, 1)$ and $\hat{p} = p(Q_1 + \Delta Q, \boldsymbol{\xi}_2, \upsilon_{12}, \mathbf{p}_{-1}, 1) > p(Q_1 + \Delta Q, \boldsymbol{\xi}_1, \upsilon_{11}, \mathbf{p}_{-1}, 1)$, from Assumption 13 we know that the marginal revenue is higher for firm 1 in market 2:

$$MR(p(Q_1 + \Delta Q, \boldsymbol{\xi}_2, \upsilon_{12}, \mathbf{p}_{-1}, 1), \boldsymbol{\xi}_2, \mathbf{p}_{-1}, 1) > MR(p(Q_1 + \Delta Q, \boldsymbol{\xi}_1, \upsilon_{11}, \mathbf{p}_{-1}, 1), \boldsymbol{\xi}_1, \mathbf{p}_{-1}, 1)$$

Furthermore,

$$s((Q_1 + \Delta Q, \boldsymbol{\xi}_2, \upsilon_{12}, \mathbf{p}_{-1}, 1), \boldsymbol{\xi}_2, \mathbf{p}_{-1}, 1)(Q_1 + \Delta Q) = s(Q_1 + \Delta Q, \boldsymbol{\xi}_1, \upsilon_{11}, \mathbf{p}_{-1}, 1)(Q_1 + \Delta Q) \equiv q_{11} + \Delta q_1.$$

where $q_{11} \equiv \mathbf{s}_{11} Q_1$. Therefore,

$$\begin{aligned} \tfrac{\partial C(q_{11} + \Delta q_1, \upsilon_{12})}{\partial q} &= MR(p(Q_1 + \Delta Q, \boldsymbol{\xi}_2, \upsilon_{12}, \mathbf{p}_{-1}, 1), \boldsymbol{\xi}_2, \mathbf{p}_{-1}, 1) \\ &> MR(p(Q_1 + \Delta Q, \boldsymbol{\xi}_1, \upsilon_{11}, \mathbf{p}_{-1}, 1), \boldsymbol{\xi}_1, \mathbf{p}_{-1}, 1) = \tfrac{\partial C(q_{11} + \Delta q_1, \upsilon_{11})}{\partial q} \end{aligned},$$

which implies that $\upsilon_{12} > \upsilon_{11}$. Therefore,

$$MC(q_{11}, \upsilon_{12}) > MC(q_{11}, \upsilon_{11}) > \widehat{MC}_1$$

and thus, for sufficiently small $\Delta q_1 > 0$,

$$\frac{C(q_{11} + \Delta q_1, \upsilon_{12}) - C(q_{11}, \upsilon_{11})}{\Delta q_1} > \frac{C(q_{11} + \Delta q_1, \upsilon_{12}) - C(q_{11}, \upsilon_{12})}{\Delta q_1} > \widehat{MC}_1$$

and $\left(C(q_{11} + \Delta q_1, \upsilon_{12}) - C(q_{11}, \upsilon_{11})\right)/\Delta q_1 - \widehat{MC}_1$ won't converge to zero as Δq_1 goes to zero. Therefore, equation (35) does not hold. The proof for the case with the estimated marginal cost is greater than the true marginal cost (e.g., $\widehat{MC}_1 > MC_1$) follows similarly. ∎

D Semi-Parametric Cost Function Estimation.

Once we have estimated the market share parameters, we can use the recovered marginal revenue and the pseudo-cost function to nonparametrically reconstruct the cost function. We do so in 3 steps, where we extensively use the supply-side F.O.C.'s and estimated marginal revenue.

Step 1

Suppose that we already estimated the pseudo-cost function $\widehat{PC}(q, \mathbf{w}, MR, \hat{\boldsymbol{\gamma}}_M)$. Then, we can derive the non-parametric pseudo-marginal cost function as follows:

$$\widehat{MC}(q, \mathbf{w}, C) = \sum_{jm} MR\left(\mathbf{X}_m, \mathbf{p}_m, \mathbf{s}_m, j, \hat{\boldsymbol{\theta}}_M\right) W_h\left(q - q_{jm}, \mathbf{w} - \mathbf{w}_{jm}, C - \widehat{PC}\left(q_{jm}, \mathbf{w}_{jm}, MR_{jm}\left(\hat{\boldsymbol{\theta}}_M\right), \hat{\boldsymbol{\gamma}}_M\right)\right)$$

where the weight function is

$$W_h\left(q - q_{jm}, \mathbf{w} - \mathbf{w}_{jm}, C - \widehat{PC}_{jm}\right) = \frac{K_{h_q}(q - q_{jm}) K_{h_W}(\mathbf{w} - \mathbf{w}_{jm}) K_{h_{MR}}\left(C - \widehat{PC}_{jm}\right)}{\sum_{kl} K_{h_q}(q - q_{kl}) K_{h_W}(\mathbf{w} - \mathbf{w}_{kl}) K_{h_{MR}}\left(C - \widehat{PC}_{kl}\right)}.$$

$MR\left(\mathbf{X}_m, \mathbf{p}_m, \mathbf{s}_m, j, \hat{\boldsymbol{\theta}}_M\right)$ can be both parametric or nonparametric.

Step 2

Start with an input price, output and (true) cost triple $\mathbf{w}$, $\bar{q}$, and $\bar{C}$. Then, there exists a cost shock $\bar{\upsilon}$ that corresponds to $\bar{MR} = \widehat{MC}\left(\bar{q}, \mathbf{w}, \bar{C}\right) = MC(\bar{q}, \mathbf{w}, \bar{\upsilon})$. Notice that we cannot derive the value of $\bar{\upsilon}$ because we have not constructed the cost function yet. For small Δq, the cost estimate for output $\bar{q} + \Delta q$, input price $\mathbf{w}$ and the same cost shock $\bar{\upsilon}$ is

$$\widehat{C}(\bar{q} + \Delta q, \mathbf{w}, \bar{\upsilon}) = \bar{C} + \bar{MR}\Delta q.$$

Then, from the consistency of the marginal revenue estimator (which we will prove later) and the Taylor series expansion,

$$\widehat{C}(\bar{q} + \Delta q, \mathbf{w}, \bar{\upsilon}) = C(\bar{q} + \Delta q, \mathbf{w}, \bar{\upsilon}) + \bar{MR}\Delta q + O\left((\Delta q)^2\right) + o_p(1)\Delta q.$$

At iteration k¿1, given $\widehat{C}_{k-1} = \widehat{C}(\bar{q} + (k-1)\Delta q, \mathbf{w}, \bar{\upsilon})$

$$\widehat{C}(\bar{q} + k\Delta q, \mathbf{w}, \bar{\upsilon}) = \widehat{C}_{k-1} + \widehat{MC}\left(\bar{q} + (k-1)\Delta q, \mathbf{w}, \widehat{C}_{k-1}\right)\Delta q.$$

Thus, from Taylor expansion, we know that for any $k > 0$,

$$\widehat{C}(\bar{q} + k\Delta q, \mathbf{w}, \bar{\upsilon}) = C(\bar{q} + k\Delta q, \mathbf{w}, \bar{\upsilon}) + O\left(k(\Delta q)^2\right) + k o_p(1)\Delta q$$

Thus, we can derive the approximate cost function for given input price $\bar{\mathbf{w}}$ and quantity q

Step 3

Next we derive the nonparametric estimate of the input demand. Denote $\mathbf{l}(q, \mathbf{w}, C)$ to be the vector of input demand given output q, input price $\mathbf{w}$ and cost C. Then, its nonparametric estimate is:

$$\hat{\mathbf{l}}(q, \mathbf{w}, C) = \sum_{jm} \mathbf{l}_{jm} W_h\left(q - q_{jm}, \mathbf{w} - \mathbf{w}_{jm}, C - \widehat{PC}\left(q_{jm}, \mathbf{w}_{jm}, MR_{jm}\left(\hat{\boldsymbol{\theta}}_M\right), \hat{\boldsymbol{\gamma}}_M\right)\right).$$

where $\mathbf{l}_{jm}$ is the vector of inputs of firm j in market m. Notice that from Shepard's Lemma,

$$\mathbf{l} = \frac{\partial C(q, \mathbf{w}, v)}{\partial \mathbf{w}}$$

Start, as before, with $\bar{q}$, $\mathbf{w}$, and $\bar{C}$. Next, we derive the cost for the output $\bar{q}$, $\mathbf{w} + \Delta\mathbf{w}$ for small $\Delta\mathbf{w}$ that has the same cost shock $\bar{v}$. It is approximately:

$$\widehat{C}_1 = \widehat{C}(\bar{q}, \mathbf{w} + \Delta\mathbf{w}, \bar{v}) = \bar{C} + \hat{\mathbf{l}}(\bar{q}, \mathbf{w}, \bar{C})\,\Delta\mathbf{w} + O\left(\left(\|\Delta\mathbf{w}\|^2\right)\right) + o_p(1)\,\|\Delta\mathbf{w}\|.$$

At iteration k¿1, given $\widehat{C}_{k-1} = \widehat{C}(\bar{q}, \mathbf{w} + (k-1)\Delta\mathbf{w}, \bar{v})$

$$\widehat{C}(\bar{q}, \mathbf{w} + k\Delta\mathbf{w}, \bar{v}) = \widehat{C}_{k-1} + \hat{\mathbf{l}}\left(\bar{q}, \mathbf{w} + (k-1)\Delta\mathbf{w}, \widehat{C}_{k-1}\right)\Delta\mathbf{w}$$

By iterating this, we can derive the approximated cost function, which satisfies

$$\widehat{C}(\bar{q}, \mathbf{w} + k\Delta\mathbf{w}, \bar{v}) = C(\bar{q}, \mathbf{w} + k\Delta\mathbf{w}, \bar{v}) + O\left(\left(k\|\Delta\mathbf{w}\|^2\right)\right) + k o_p(1)\,\|\Delta\mathbf{w}\|$$

for any $k > 0$.

E Large Sample Properties of the NLLS-GMM Estimator.

In this section we show that the estimator is consistent and asymptotically normal. Notice that in our sample, we have oligopolistic firms in the same market. Because of strategic interaction, equilibrium prices and outputs of the firms in the same market are likely to be correlated. To avoid the difficulty arising from such within-market correlation, our consistency proof will primarily exploit the large number of isolated markets, with the assumption that wages, unobserved product quality and cost shocks are independent across markets[51]. Without loss of generality, we assume that in each market, the number of firms is J. Notice that in our objective function, we have two separate components: one that involves the difference between the cost in the data and the nonparametrically approximated pseudo-cost function, which identifies α for the Berry logit model and $(\mu_\alpha, \sigma_\alpha)$ and σ_β for the BLP random coefficient logit model. The second component is the objective function that is based on the orthogonality condition $\boldsymbol{\xi}_m \perp \mathbf{X}_m$, which identifies $\boldsymbol{\beta}$ for the logit model and μ_β for the BLP. We denote $\boldsymbol{\theta} = (\theta_\beta, \boldsymbol{\theta}_c)$, where $\boldsymbol{\theta}_c$ is the vector of the parameters identified from the difference between the cost data and the pseudo-cost function. That is, $\boldsymbol{\theta}_c = \alpha$ for the Berry logit model and $\boldsymbol{\theta}_c = (\mu_\alpha, \sigma_\alpha, \sigma_\beta)$ for the BLP model. We denote θ_β to be the vector of parameters that are identified by the orthogonality condition $\boldsymbol{\xi}_m \perp \mathbf{X}_m$, which is$\boldsymbol{\beta}$ for the Berry logit model and $\boldsymbol{\mu}_\beta$ for the BLP model.

In our proof, for the pseudo-cost function part, we follow Bierens (2014) closely. Most of the assumptions below are slight modifications of the ones by Bierens 2014, where we changed the signs to use them for minimization of the joint objective function rather than maximization of the likelihood function.

Let $\mathbf{y}_m = \left(\mathbf{q}_m, vec(\mathbf{W}_m)', \mathbf{C}_m^d, vec(\mathbf{X}_m)', vec(\mathbf{p}_m)', vec(\mathbf{s}_m)'\right)'$, where $\mathbf{C}_m^d = \left(C_{1m}^d, C_{2m}^d, ..., C_{Jm}^d\right)'$, $\mathbf{W}_m = (\mathbf{w}_{1m}, \mathbf{w}_{2m}, ..., \mathbf{w}_{jm})'$and define

$$f(\mathbf{y}_m, \boldsymbol{\chi}) = \sum_{j=1}^{J}\left[C_{jm}^d - \sum_l \gamma_l \psi_l\left(q_{jm}, \mathbf{w}_{jm}, MR(\mathbf{X}_m, \mathbf{p}_m, \mathbf{s}_m, j, \boldsymbol{\theta}_c)\right)\right]^2, \tag{38}$$

and $Q(\boldsymbol{\chi}) = E[f(\mathbf{y}_m, \boldsymbol{\chi})]$, where $\boldsymbol{\chi} = (\boldsymbol{\theta}_c', \boldsymbol{\gamma}')' = \{\chi_n\}_{n=1}^\infty$, with

$$\chi_n = \begin{cases} \theta_{cn} & \text{for } n = 1, ..., p, \\ \gamma_{n-p} & \text{for } n \geq p+1. \end{cases}$$

where p is the number of parameters in $\boldsymbol{\theta}_c$. Parameter space is $\Xi \equiv \Theta_c \times \Gamma(T)$, where $\boldsymbol{\theta}_c \in \Theta_c$ is compact and

$$\Gamma(T) = \left\{\boldsymbol{\gamma} = \{\gamma_n\}_{n=1}^\infty : \|\boldsymbol{\gamma}\| \leq T\right\},$$

and is endowed with the metric $d(\boldsymbol{\chi}_1, \boldsymbol{\chi}_2) \equiv \|\boldsymbol{\chi}_1 - \boldsymbol{\chi}_2\|$, where $\|\boldsymbol{\chi}\| = \sqrt{\sum_{k=1}^\infty \chi_k^2}$.

[51]The assumption of independence of variables across markets are employed for simplicity. We leave the asymptotic analysis with some across market dependence for future research. For Strong Law of Large Numbers under weaker assumptions, see Andrews (1988) and the related literature. As we have discussed earlier, those assumptions are not required for identification.

Let $\boldsymbol{\chi}_0$ be the vector of true parameters. Define also

$$\Xi_k = \begin{cases} \Theta & \text{for } k \leq p, \\ \Theta \times \Gamma_{k-p}(T) & \text{for } k \geq p+1, \end{cases}$$

where $k \in \mathbb{N}$, $\Gamma_k(T) = \{\pi_k \boldsymbol{\gamma} : \|\pi_k \boldsymbol{\gamma}\| \leq T\}$, and π_k is the operator that applies to an infinite sequence $\boldsymbol{\gamma} = \{\gamma_n\}_{n=1}^{\infty}$, replacing all the γ_n's for $n > k$ with zeros.

The following assumptions are made:

Assumption E.1

(a) $\mathbf{y}_1$, $\mathbf{y}_2$, ..., $\mathbf{y}_M$ are i.i.d. with support contained in a bounded open set $\mathcal{V}$ of a Euclidean space.
(b) For each $\boldsymbol{\chi} \in \Xi$, $f(\mathbf{y}_m, \boldsymbol{\chi})$ is a Borel measurable real function of $\mathbf{y}_m$.
(c) $f(\mathbf{y}_m, \boldsymbol{\chi})$ is a.s. continuous in $\boldsymbol{\chi} \in \Xi$.
(d) There exists a non-negative borel measurable real function $\underline{f(\mathbf{y})}$ such that $E\left[\underline{f(\mathbf{y}_m)}\right] > -\infty$ and $f(\mathbf{y}_m, \boldsymbol{\chi}) > \underline{f(\mathbf{y})}$ for all $\boldsymbol{\chi} \in \Xi$.
(e) There exists an element $\boldsymbol{\chi}_0 \in \Xi$ such that $Q(\boldsymbol{\chi}) > Q(\boldsymbol{\chi}_0)$ for all $\boldsymbol{\chi} \in \Xi \backslash \{\boldsymbol{\chi}_0\}$, and $Q(\boldsymbol{\chi}_0) < \infty$.
(f) There exists an increasing sequence of compact subspaces Ξ_k in Ξ such that $\boldsymbol{\chi}_0 \in \overline{\bigcup_{k=1}^{\infty} \Xi_k} = \overline{\Xi} \subset \Xi$. Furthermore, each sieve space Ξ_k is isomorph to a compact subset of a Euclidean space.
(g) Each sieve space Ξ_k contains an element $\boldsymbol{\chi}_k$ such that, $lim_{k\to\infty} E[f(\mathbf{y}_m, \boldsymbol{\chi}_k)] = E[f(\mathbf{y}_m, \boldsymbol{\chi}_0)]$.
(h) The set $\Xi_\infty = \{\boldsymbol{\chi} \in \Xi : E[f(\mathbf{y}_m, \boldsymbol{\chi})] = \infty\}$ does not contain an open ball.
(i) There exists a compact set Ξ_c containing $\boldsymbol{\chi}_0$ such that $Q(\boldsymbol{\chi}_0) < E\left[inf_{\boldsymbol{\chi} \in \overline{\Xi} \backslash \Xi_c} f(\mathbf{y}_m, \boldsymbol{\chi})\right] < \infty$.

Assumptions (a)-(f) are well established in the literature (see e.g. Bierens 2014). For example, (d) is satisfied because of the definition of $f() \geq 0$ from equation 38. (e) follows from the identification of $\boldsymbol{\chi}_0$ in Proposition 2. (f) is required in order to make estimation feasible. In particular, since minimising $\widehat{Q}_M = M^{-1} \sum_{m=1}^{M} f(\mathbf{y}_m, \boldsymbol{\chi})$ over Ξ is not possible given that Ξ is not even compact, (f) ensures that the minimization problem can be solved in terms of Ξ_{k_M}, i.e.

$$\widehat{\boldsymbol{\chi}}_M = \arg \min_{\boldsymbol{\chi} \in \Xi_{k_M}} \widehat{Q}_M(\boldsymbol{\chi}),$$

where k_M is an arbitrary sequence of M that satisfies $k_M < M$, $\lim_{M\to\infty} k_M = \infty$. We will assume

$$E[f(\mathbf{y}_m, \boldsymbol{\chi}_0)] < \lim_{\tau\to\infty} E\left[inf_{\boldsymbol{\chi} \in \overline{\Xi} \backslash \Xi_\tau} f(\mathbf{y}_m, \boldsymbol{\chi})\right],$$

where $\Xi_\tau = X_{n=1}^{\infty} [-\overline{\chi}_n \tau, -\overline{\chi}_n \tau]$, and $\{\overline{\chi}_n\}_{n=1}^{\infty}$ satisfies $\sum\limits_{n=1}^{\infty} \overline{\chi}_n < \infty$; $\sup_{n\geq 1} |\chi_{0,n}| / \overline{\chi}_n \leq 1$. Then, there exists $\tau < \infty$ such that, if we set $\Xi_\tau = \Xi_c$, (i) holds. Then, from Kolmogorov's Strong Law of Large Numbers, for a given $\boldsymbol{\chi} \in \Xi_\tau$

$$\frac{1}{M} \sum_{m=1}^{M} inf_{\boldsymbol{\chi}_* \in \Xi, \|\boldsymbol{\chi} - \boldsymbol{\chi}_*\| < \epsilon} f(\mathbf{y}_m, \boldsymbol{\chi}_*) \xrightarrow{a.s.} E\left[inf_{\boldsymbol{\chi}_* \in \Xi, \|\boldsymbol{\chi} - \boldsymbol{\chi}_*\| < \epsilon} f(\mathbf{y}_m, \boldsymbol{\chi}_*)\right] \;\; as\ M \to \infty.$$

Furthermore, Now, for an arbitrarily small $\eta > 0$, let $\Xi_\eta = \{\boldsymbol{\chi} : \|\boldsymbol{\chi} - \boldsymbol{\chi}_0\| \geq \eta\} \cap \Xi_c$. Then,

$$lim_{\epsilon\downarrow 0} inf_{\boldsymbol{\chi}_* \in \Xi_\eta, \|\boldsymbol{\chi} - \boldsymbol{\chi}_*\| < \epsilon} f(\mathbf{y}_m, \boldsymbol{\chi}_*) \geq f(\mathbf{y}_m, \boldsymbol{\chi})$$

And from Monotone Convergence Theorem,

$$lim_{\epsilon\downarrow 0} E\left[inf_{\boldsymbol{\chi}_* \in \Xi_\eta, \|\boldsymbol{\chi} - \boldsymbol{\chi}_*\| < \epsilon} f(\mathbf{y}_m, \boldsymbol{\chi}_*)\right] \geq E[f(\mathbf{y}_m, \boldsymbol{\chi})]$$

Let $\{B_\epsilon(\boldsymbol{\chi})\}_{\boldsymbol{\chi} \in \Xi_\eta}$ be the open cover of the compact set Ξ_η, i.e. $B_\epsilon(\boldsymbol{\chi}) = \{\tilde{\boldsymbol{\chi}} : \|\tilde{\boldsymbol{\chi}} - \boldsymbol{\chi}\| < \epsilon\}$ Then, it has a finite subcover of $\{B_\epsilon(\boldsymbol{\chi}_k)\}_{k=1}^{K_\epsilon}$satisfying

$$min_{k=1,\ldots,K_\epsilon} \frac{1}{M} \sum_{m=1}^{M} inf_{\boldsymbol{\chi}_* \in B_\epsilon(\boldsymbol{\chi}_k) \cap \Xi_\eta} f(\mathbf{y}_m, \boldsymbol{\chi}_*) \xrightarrow{a.s.} min_{k=1,\ldots,K_\epsilon} E\left[inf_{\boldsymbol{\chi}_* \in B_\epsilon(\boldsymbol{\chi}_k) \cap \Xi_\eta} f(\mathbf{y}_m, \boldsymbol{\chi}_*)\right].$$

as $M \to \infty$. Therefore, from Assumption E.1, (e)

$$\begin{aligned} & inf_{\boldsymbol{\chi} \in \Xi_\eta} plim_{M\to\infty} \frac{1}{M} \sum_{m=1}^{M} f(\mathbf{y}_m, \boldsymbol{\chi}) \geq lim_{\epsilon\downarrow 0} \left[min_{k=1,\ldots,K_\epsilon} E\left[inf_{\boldsymbol{\chi}_* \in B_\epsilon(\boldsymbol{\chi}_k) \cap \Xi_\eta} f(\mathbf{y}_m, \boldsymbol{\chi}_*)\right]\right] \\ = \;& inf_{\boldsymbol{\chi} \in \Xi_\eta} E[f(\mathbf{y}_m, \boldsymbol{\chi})] > E[f(\mathbf{y}_m, \boldsymbol{\chi}_0)]. \end{aligned} \tag{39}$$

Furthermore, from SLLN, we obtain

$$inf_{\boldsymbol{\chi}\in\overline{\Xi}\backslash\Xi_\tau}\frac{1}{M}\sum_{m=1}^{M}f\left(\mathbf{y}_m,\boldsymbol{\chi}\right)\geq\frac{1}{M}\sum_{m=1}^{M}inf_{\boldsymbol{\chi}\in\overline{\Xi}\backslash\Xi_\tau}f\left(\mathbf{y}_m,\boldsymbol{\chi}\right)\underrightarrow{a.s.}E\left[inf_{\boldsymbol{\chi}\in\overline{\Xi}\backslash\Xi_\tau}f\left(\mathbf{y}_m,\boldsymbol{\chi}\right)\right]>E\left[f\left(\mathbf{y}_m,\boldsymbol{\chi}_0\right)\right]. \quad (40)$$

From 39 and 40, we derive

$$inf_{\boldsymbol{\chi}\in\Xi,\|\boldsymbol{\chi}-\boldsymbol{\chi}_0\|\geq\eta}plim_{M\to\infty}\frac{1}{M}\sum_{m=1}^{M}f\left(\mathbf{y}_m,\boldsymbol{\chi}\right)>E\left[f\left(\mathbf{y}_m,\boldsymbol{\chi}_0\right)\right]$$

Next, we consider the moment-based objective function. Denote $\mathbf{v}_m=\left(\mathbf{y}_m,vec\left(\mathbf{Z}_m\right)\right)$ where $\mathbf{Z}_{jm}$ is the vector of instruments for firm j. Furthermore, let $\mathbf{g}\left(\mathbf{v}_m,j,\boldsymbol{\theta}\right)=\boldsymbol{\xi}_j\left(\mathbf{p}_m,\mathbf{X}_m,\mathbf{s}_m,\boldsymbol{\theta}\right)\mathbf{Z}_{jm}$, $\mathbf{g}_M\left(j,\boldsymbol{\theta}\right)=\frac{1}{M}\sum_{m=1}^{M}\boldsymbol{\xi}_j\left(\mathbf{p}_m,\mathbf{X}_m,\mathbf{s}_m,\boldsymbol{\theta}\right)\mathbf{Z}_{jm}$, i.e.,

$$\mathbf{g}\left(\mathbf{v}_m,\boldsymbol{\theta}\right)=\begin{bmatrix}\mathbf{g}\left(\mathbf{v}_m,1,\boldsymbol{\theta}\right)\\ \vdots\\ \mathbf{g}\left(\mathbf{v}_m,J,\boldsymbol{\theta}\right)\end{bmatrix},\mathbf{g}_M\left(\boldsymbol{\theta}\right)=\begin{bmatrix}\mathbf{g}_M\left(\mathbf{v}_m,1,\boldsymbol{\theta}\right)\\ \vdots\\ \mathbf{g}_M\left(\mathbf{v}_m,J,\boldsymbol{\theta}\right)\end{bmatrix}$$

, and $\mathbf{G}_{jM}\left(\boldsymbol{\theta}\right)=\partial\mathbf{g}_{jM}\left(\boldsymbol{\theta}\right)/\partial\boldsymbol{\theta}$. Then, we assume the following.

Assumption E.2

a) We assume that $\mathbf{v}_1,...,\mathbf{v}_M$ are i.i.d. distributed, and therefore, for any parameter $\boldsymbol{\theta}\in\Theta$, $\mathbf{g}\left(\mathbf{v}_m,\boldsymbol{\theta}\right)$, $m=1,...,M$ are also i.i.d. distributed.
b) $\mathbf{W}$ is symmetric and positive definite, and $\mathbf{W}E\left[\mathbf{g}\left(\mathbf{v}_m,\boldsymbol{\theta}\right)\right]=0$ only if $\boldsymbol{\theta}_\beta=\boldsymbol{\theta}_{\beta0}$.
c) $\mathbf{g}\left(\mathbf{v}_m,\boldsymbol{\theta}\right)$ is a continuously differentiable function of $\boldsymbol{\theta}$.
d) $E\left[sup_{\boldsymbol{\theta}\in\Theta,j}\left\|\mathbf{g}\left(\mathbf{v}_m,j,\boldsymbol{\theta}\right)\right\|\right]<\infty$.
e) $E\left[\mathbf{g}\left(\mathbf{v}_m,\boldsymbol{\theta}_0\right)\mathbf{g}\left(\mathbf{v}_m,\boldsymbol{\theta}_0\right)'\right]$ is positive definite.
f) $sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\leq\delta_M}\left\|\partial\mathbf{g}_M\left(\boldsymbol{\theta}\right)/\partial\boldsymbol{\theta}\right\|=O_p\left(1\right)$ for $\delta_M\to0$ as $M\to\infty$.

Assumption (c) and (f) implies stochastic equicontinuity, which implies Assumption (v) of Theorem 7.2, Newey and McFadden (1994). This result is used late for asymptotic normality proof.
Following the proof by Newey and McFadden (1994), Theorem 2.6, we can show that

$$sup_{\boldsymbol{\theta}\in\Theta}\left\|\mathbf{g}_M\left(\boldsymbol{\theta}\right)'\mathbf{W}\mathbf{g}_M\left(\boldsymbol{\theta}\right)-E\left[\mathbf{g}\left(\mathbf{v}_m,\boldsymbol{\theta}\right)\right]'\mathbf{W}E\left[\mathbf{g}\left(\mathbf{v}_m,\boldsymbol{\theta}\right)\right]\right\|\underrightarrow{P}0.$$

For any $\boldsymbol{\theta}\neq\boldsymbol{\theta}_0$, suppose first that $\boldsymbol{\theta}_c\neq\boldsymbol{\theta}_{c0}$, i.e. $\|\boldsymbol{\theta}_c-\boldsymbol{\theta}_{c0}\|\geq\eta$ for some $\eta>0$. Then,

$$plim_{M\to\infty}\mathbf{g}_M\left(\boldsymbol{\theta}\right)'\mathbf{W}\mathbf{g}_M\left(\boldsymbol{\theta}\right)\geq0=E\left[g\left(\mathbf{v}_m,\boldsymbol{\theta}_0\right)\right]'\mathbf{W}E\left[g\left(\mathbf{v}_m,\boldsymbol{\theta}_0\right)\right]=plim_{M\to\infty}\mathbf{g}_M\left(\boldsymbol{\theta}_0\right)'\mathbf{W}\mathbf{g}_M\left(\boldsymbol{\theta}_0\right).$$

Furthermore, since $\|\boldsymbol{\chi}-\boldsymbol{\chi}_0\|\geq\|\boldsymbol{\theta}_c-\boldsymbol{\theta}_{c0}\|\geq\eta$

$$plim_{M\to\infty}inf_{\boldsymbol{\chi}\in\Xi,\|\boldsymbol{\chi}-\boldsymbol{\chi}_0\|\geq\eta}\frac{1}{M}\sum_{m=1}^{M}f\left(\mathbf{y}_m,\boldsymbol{\chi}\right)>E\left[f\left(\mathbf{y}_m,\boldsymbol{\chi}_0\right)\right]$$

and similarly, for $\boldsymbol{\theta}_\beta$ such that $\|\boldsymbol{\theta}_\beta-\boldsymbol{\theta}_{\beta0}\|\geq\eta$,

$$plim_{M\to\infty}\mathbf{g}_M\left(\boldsymbol{\theta}\right)'\mathbf{W}\mathbf{g}_M\left(\boldsymbol{\theta}\right)>0=E\left[g\left(\mathbf{v}_m,\boldsymbol{\theta}_0\right)\right]'\mathbf{W}E\left[g\left(\mathbf{v}_m,\boldsymbol{\theta}_0\right)\right]$$

and

$$plim_{M\to\infty}inf_{\boldsymbol{\chi}\in\Xi,\|\boldsymbol{\chi}-\boldsymbol{\chi}_0\|\geq\eta}\frac{1}{M}\sum_{m=1}^{M}f\left(\mathbf{y}_m,\boldsymbol{\chi}\right)\geq E\left[f\left(\mathbf{y}_m,\boldsymbol{\chi}_0\right)\right].$$

Therefore, $lim_{M\to\infty}P\left(\left(\|\mu_{\beta M}-\mu_{\beta0}\|\geq\eta\right)\cup\left(\|\boldsymbol{\theta}_{cM}-\boldsymbol{\theta}_{c0}\|\geq\eta\right)\right)=0$, and we have shown that $plim\left[\boldsymbol{\theta}_M,\boldsymbol{\gamma}_M\right]=\left[\boldsymbol{\theta}_0,\boldsymbol{\gamma}_0\right]$. If we were to use the two-step GMM, then the weighting matrix is $\mathbf{W}=E\left[\mathbf{g}\left(\mathbf{v}_m,\boldsymbol{\theta}_0\right)\mathbf{g}\left(\mathbf{v}_m,\boldsymbol{\theta}_0\right)'\right]^{-1}$ and its sample analog, $\mathbf{W}_M=\left[\mathbf{g}_M\left(\boldsymbol{\theta}_M\right)\mathbf{g}_M\left(\boldsymbol{\theta}_M\right)'\right]^{-1}$. Then, if $\boldsymbol{\theta}_M$ is the estimator with the initial positive definite weight matrix $\mathbf{W}_0$, then, we have shown that $plim_{M\to\infty}\boldsymbol{\theta}_{0M}=\boldsymbol{\theta}_0$. Hence, from continuity of $\mathbf{g}\left(\mathbf{v}_m,\boldsymbol{\theta}\right)$ with respect to $\boldsymbol{\theta}$. and intertibility of $E\left[\mathbf{g}\left(\mathbf{v}_m,\boldsymbol{\theta}\right)\mathbf{g}\left(\mathbf{v}_m,\boldsymbol{\theta}\right)'\right]$,

$$plim_{M\to\infty}\mathbf{W}_M=\mathbf{W}.$$

Then, since the assumptions of theorem 2.6, Newey and McFadden (1994) are satisfied, $\boldsymbol{\theta}_M\underrightarrow{P}\boldsymbol{\theta}$as $M\to\infty$.

Next, we prove asymptotic normality. To do so, let

$$\Gamma_r\left(T\right)=\left\{\boldsymbol{\gamma}=\left\{\gamma_n\right\}_{n=1}^{\infty}:\sum_{n=1}^{\infty}n^r\left|\gamma_n\right|\leq T\right\},$$

for some T large enough such that $\boldsymbol{\gamma}_0 \in \Gamma_r(T)$ and associated metric $\|\boldsymbol{\gamma}_1-\boldsymbol{\gamma}_2\|_r = \sum\limits_{n=1}^{\infty} n^r |\gamma_{1,n}-\gamma_{2,n}|$, $\boldsymbol{\gamma}_i = \{\gamma_{i,n}\}_{n=1}^{\infty}$. Furthermore, the sieve space is replaced by

$$\begin{aligned}\Xi_r &= \left\{\boldsymbol{\chi}=\{\chi_n\}_{n=1}^{\infty} : \|\boldsymbol{\chi}\|_r < T,\, T > \|\boldsymbol{\chi}_0\|_r\right\};\\ \Xi_{r,k} &= \left\{\pi_k\boldsymbol{\chi} : \|\pi_k\boldsymbol{\chi}\|_r < T\right\}.\end{aligned}$$

The following assumptions are employeed:

Assumption E.3

(a) Parameter space Ξ is defined with a norm $\|\boldsymbol{\chi}\|_r = \sum_{n=1}^{\infty} n^r |\chi_m|$ and the associated metric $d(\boldsymbol{\chi}_1,\boldsymbol{\chi}_2) = \|\boldsymbol{\chi}_1-\boldsymbol{\chi}_2\|_r$.
(b) True parameter $\boldsymbol{\chi}_0 = \{\chi_{0,n}\}_{n=1}^{\infty}$ satisfies $\|\boldsymbol{\chi}_0\|_r < \infty$.
(c) There exists $k \in \mathbb{N}$ such that for k large enough $\boldsymbol{\chi}_{0,k} = \pi_k\boldsymbol{\chi}_0 \in \Xi_k^{Int}$, where Ξ_k^{Int} is the interior of the sieve space Ξ_k.
(d) $f(\mathbf{y}_m,\boldsymbol{\chi})$ is a.s. twice continuously differentiable in an open neighborhood of $\boldsymbol{\chi}_0$.
(e) For any subsequence $k=k_M$ of the sample size M satisfying $k_M \to \infty$ as $M \to \infty$, $plim_{M\to\infty}||\hat{\boldsymbol{\chi}}_{k_M} - \boldsymbol{\chi}_0||_r = 0$.

(b) imposes a boundedness condition on the true parameter values. (c) employs stronger requirements on the parameters than Assumption E.1. That is, the true parameters need to be in the interior of the parameter space. The differentiability of the objective function in (d) is necessary for the derivation of the asymptotic distribution of the estimator. (e) is straightforward to show given (a)-(d) and Assumption E.1. Furthermore, we also assume:

Assumption E.4

(a) There exists a nonnegative integer $r_0 < r$ such that the following local Lipschitz conditions hold for all positive integer $j \in \mathbb{N}$ we have

$$E\left\|\nabla_j f(\mathbf{y},\boldsymbol{\chi}_0) - \nabla_j f(\mathbf{y},\boldsymbol{\chi}_{0,k})\right\| \le C_j \left\|\boldsymbol{\chi}_0 - \boldsymbol{\chi}_{0,k}\right\|_{r_0}$$

where $\nabla_j f(\mathbf{y}_m,\boldsymbol{\chi}_0) = \partial f(\mathbf{y}_m,\boldsymbol{\chi}_0)/\partial\boldsymbol{\chi}_{0,j}$, $\sum_{j=1}^{\infty} 2^{-j} C_j < \infty$ and the sieve order $k=k_M$ is chosen such that

$$\lim_{M\to\infty} \sqrt{M} \sum_{n=k_M+1}^{\infty} n^{r_0} |\boldsymbol{\chi}_{0,n}| = 0.$$

(b) For all $j \in \mathbb{N}$, $E[\nabla_j f(\mathbf{y},\boldsymbol{\chi}_0)] = 0$.
(c) $\sum_{j=1}^{\infty} j2^{-j} E\left[(\nabla_j f(\mathbf{y},\boldsymbol{\chi}_0))^2\right] < \infty$.
For some $\tau \ge 0$,
(d) $\sum_{j=1}^{\infty}\sum_{n=1}^{\infty} (jn)^{-2-\tau} E[|\nabla_{j,n} f(\mathbf{y},\boldsymbol{\chi}_0)|] < \infty$, where $\nabla_{j,k} f(\mathbf{y}_m,\boldsymbol{\chi}_0) = \partial^2 f(\mathbf{y}_m,\boldsymbol{\chi}_0)/\left(\partial\boldsymbol{\chi}_{0,j}\partial\boldsymbol{\chi}_{0,k}\right)$.
(e) $\lim_{\epsilon\downarrow 0}\sum_{j=1}^{\infty}\sum_{n=1}^{\infty} (jn)^{-2-\tau} E\left[sup_{\|\boldsymbol{\chi}-\boldsymbol{\chi}^0\|_r \le \epsilon} |\nabla_{j,n} f(\mathbf{y},\boldsymbol{\chi}) - \nabla_{j,n} f(\mathbf{y},\boldsymbol{\chi}_0)|\right] = 0$.
(f) For at least one pair of positive integers j,n, $E[\nabla_{j,p+n} f(\mathbf{y},\boldsymbol{\chi}_0)] \ne 0$.
(g) $rank(B_{k,k}) = k$ for each $k \ge p$, where

$$B_{k,l} = \begin{bmatrix} E[\nabla_{1,1} f(\mathbf{y},\boldsymbol{\chi}_0)] & \dots & E[\nabla_{1,n} f(\mathbf{y},\boldsymbol{\chi}_0)] \\ \vdots & \ddots & \vdots \\ E[\nabla_{j,1} f(\mathbf{y},\boldsymbol{\chi}_0)] & \dots & E[\nabla_{j,n} f(\mathbf{y},\boldsymbol{\chi}_0)] \end{bmatrix}.$$

(b) postulates that the F.O.C. holds for the true parameter value, which we know is satisfied. (c) imposes boundedness for the first-order derivatives. (d),(e) are necessary in order to extract the parameters of interest via projection residuals. (f), (g) impose necessary regularity conditions on the second-order derivatives, in fact (f) is already implied by identification of $\boldsymbol{\chi}_0$.

Let

$$\hat{W}_n(u) = \sum_{k=1}^{K_n}\left[\frac{1}{\sqrt{M}}\sum_{j=1}^{M} \nabla_k f_j(\hat{\boldsymbol{\chi}}_n)\right]\eta_k(u)$$

$$\hat{V}_n(u) = \sum_{k=1}^{K_n}\left[\frac{1}{\sqrt{M}}\sum_{j=1}^{M} \left(\nabla_k f_j(\boldsymbol{\chi}^0) - \nabla_k f_j(\boldsymbol{\chi}_n^0)\right)\right]\eta_k(u)$$

$$\hat{Z}_n(u) = \sum_{k=1}^{K_n}\left[\frac{1}{\sqrt{M}}\sum_{j=1}^{M} \nabla_k f_j(\hat{\boldsymbol{\chi}}_n)\right]\eta_k(u)$$

$$\hat{b}_{l,n}(u) = -\sum_{k=1}^{K_n}\left[\frac{1}{\sqrt{M}}\sum_{j=1}^{M}\nabla_{k,l}f_j\left(\boldsymbol{\chi}_n^0 + \lambda_k\left(\hat{\boldsymbol{\chi}}_n - \boldsymbol{\chi}_n^0\right)\right)\right]\eta_k(u)$$

where $\eta_k(u) = 2^{-k}\sqrt{2}cos(k\pi u)$. Recall that in this case, n denotes the number of parameters, including sieve polynomials. Now, as in Bierens 2014, let

$$\hat{\boldsymbol{a}}_n(u) = \left(\hat{a}_{1,n}(u), \hat{a}_{2,n}(u), \ldots, \hat{a}_{p,n}(u)\right)$$

be the residual of the following projection

$$\hat{b}_{l,n}(u) = A\left[\hat{b}_{p+1,n}(u), ..., \hat{b}_{n,n}(u)\right] + \hat{a}_{l,n}(u)$$

Then, given the Assumptions E.1-E.4 we have

$$\int_0^1 \hat{a}_n(u)\hat{a}_n(u)' du\sqrt{M}\left(\hat{\boldsymbol{\theta}}_{cM} - \boldsymbol{\theta}_{c0}\right) = \int_0^1 \hat{a}_n(u)\left(\hat{Z}_n(u) - \hat{W}_n(u) - \hat{V}_n(u)\right)du$$

where $\hat{a}_n(u)\hat{a}_n(u)'$ is a p by p matrix, and $\hat{\boldsymbol{\theta}}_{cM} - \boldsymbol{\theta}_{c0}$ a p by 1 vector. Now, from the arguments similar to the Theorem 7.2 of Newey and McFadden (1994),

$$\begin{aligned} & \mathbf{G}_M(\boldsymbol{\theta}_M)'\mathbf{W}_M\mathbf{g}_M(\boldsymbol{\theta}_M) \\ = \quad & \mathbf{G}_M(\boldsymbol{\theta}_M)'\mathbf{W}_M\mathbf{g}_M(\boldsymbol{\theta}_0) + \mathbf{G}_M(\boldsymbol{\theta}_M)'\mathbf{W}_M\mathbf{G}_M\left(\hat{\boldsymbol{\theta}}\right)(\boldsymbol{\theta}_M - \boldsymbol{\theta}_0). \end{aligned}$$

where $\mathbf{G}_M(\boldsymbol{\theta}_M) = \partial\mathbf{g}_M(\boldsymbol{\theta}_M)/\partial\boldsymbol{\theta}$, and $\hat{\boldsymbol{\theta}}$is the intermediate value between $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_M$. Hence, together,

$$\begin{aligned} & \begin{bmatrix} A\left[\mathbf{G}_M(\boldsymbol{\theta}_M)'\mathbf{W}_M\mathbf{G}_M\left(\hat{\boldsymbol{\theta}}\right)\right]_{1,1} & A\left[\mathbf{G}_M(\boldsymbol{\theta}_M)'\mathbf{W}_M\mathbf{G}_M\left(\hat{\boldsymbol{\theta}}\right)\right]_{1,2:p} \\ A\left[\mathbf{G}_M(\boldsymbol{\theta}_M)'\mathbf{W}_M\mathbf{G}_M\left(\hat{\boldsymbol{\theta}}\right)\right]_{2:p,1} & A\left[\mathbf{G}_M(\boldsymbol{\theta}_M)'\mathbf{W}_M\mathbf{G}_M\left(\hat{\boldsymbol{\theta}}\right)\right]_{2:p,2:p} + \int_0^1 \hat{a}_{n_M}(u)\hat{a}_{n_M}(u)' du \end{bmatrix}\sqrt{M}\begin{bmatrix} \hat{\theta}_{\beta M} - \theta_\beta \\ \hat{\boldsymbol{\theta}}_{cM} - \boldsymbol{\theta}_c \end{bmatrix} \\ = \quad & \sqrt{M}\begin{bmatrix} -A\left[\mathbf{G}_M(\boldsymbol{\theta}_M)'\mathbf{W}_M\mathbf{g}_M(\boldsymbol{\theta}_0)\right]_1 \\ -A\left[\mathbf{G}_M(\boldsymbol{\theta}_M)'\mathbf{W}_M\mathbf{g}_M(\boldsymbol{\theta}_0)\right]_{2:p} + \int_0^1 \hat{a}_{n_M}(u)\left(\hat{Z}_{n_M}(u) - \hat{W}_{n_M}(u) - \hat{V}_{n_M}(u)\right)du \end{bmatrix} \end{aligned}$$

Now, we impose an addititional assumption that

Assumption E.5

$$\mathbf{F} = \begin{bmatrix} A\left[\mathbf{G}'\mathbf{W}\mathbf{G}\right]_{1,1} & A\left[\mathbf{G}'\mathbf{W}\mathbf{G}\right]_{1,2:p} \\ A\left[\mathbf{G}'\mathbf{W}\mathbf{G}\right]_{2:p,1} & A\left[\mathbf{G}'\mathbf{W}\mathbf{G}\right]_{2:p,2:p} + \int_0^1 a(u)a(u)' du \end{bmatrix}$$

is a full rank matrix, thus, invertible.

Then,

$$\sqrt{M}\left(\hat{\boldsymbol{\theta}}_M - \boldsymbol{\theta}_0\right) \xrightarrow{d} N_p\left(\mathbf{0}, \mathbf{F}^{-1}\boldsymbol{\Upsilon}\mathbf{F}'^{-1}\right),$$

where

$$\boldsymbol{\Upsilon} = \begin{bmatrix} A^2\left[\mathbf{G}'\mathbf{W}\boldsymbol{\Sigma}_{\boldsymbol{g}}\mathbf{W}\mathbf{G}\right]_{1,1} & A^2\left[\mathbf{G}'\mathbf{W}\boldsymbol{\Sigma}_{\boldsymbol{g}}\mathbf{W}\mathbf{G}\right]_{1,2:p} \\ A^2\left[\mathbf{G}'\mathbf{W}\boldsymbol{\Sigma}_{\boldsymbol{g}}\mathbf{W}\mathbf{G}\right]_{2:p,1} & A^2\left[\mathbf{G}'\mathbf{W}\boldsymbol{\Sigma}_{\boldsymbol{g}}\mathbf{W}\mathbf{G}\right]_{2:p,2:p} + \int_0^1\int_0^1 a(u_1)\boldsymbol{\Gamma}(u_1,u_2)a(u_2)du_1du_2 \end{bmatrix}$$

and $\Gamma(u_1, u_2) = E\left[Z(u_1)Z(u_2)\right]$.