Byrne, D. P., Imai, S., Jain, N., Sarafidis, V. & Hirukawa, M. (2015). Identification and Estimation of Differentiated Products Models using Market Size and Cost Data (Report No. 15/05). London, UK: Department of Economics, City University London.

School of Arts and Social Sciences
CITY UNIVERSITY LONDON

Department of Economics

Identification and Estimation of Differentiated Products Models using Market Size and Cost Data

David P. Byrne
University of Melbourne

Susumu Imai[1]
University of Technology Sydney and Queen's University

Neelam Jain
City University London

Vasilis Sarafidis
Monash University

Masayuki Hirukawa
Setsunan University

Department of Economics
Discussion Paper Series
No. 15/05

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

Identification and Estimation of Differentiated Products Models using Market Size and Cost Data[*]

David P. Byrne[†] Susumu Imai[‡]

Neelam Jain[§] Vasilis Sarafidis[¶]

and Masayuki Hirukawa[‖]

August 8, 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, by using variation in market size that does not need to be exogenous, and cost data. We establish nonparametric 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, Micheal Keane, Robin Sickles and the seminar participants at the University of Melbourne, Monash University, ANU, Latrobe University, UTS, UBC, UNSW, University of Sydney, University of Queensland, City University 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, e-mail: Susumu.Imai@uts.edu.au.

[§]Department of Economics, City University London, e-mail: Neelam.Jain.1@city.ac.uk.

[¶]Department of Econometrics and Business Statistics, Monash University, e-mail: vasilis.sarafidis@monash.edu.

[‖]Department of Economics, Setsunan University, e-mail: hirukawa@econ.setsunan.ac.jp

Keywords: Differentiated Goods Oligopoly, Instruments, Parametric Identification, Nonparametric Identification, Cost data.

JEL Codes: C13, C14, L13, L41

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 and some firm-level cost data. The novelty of our method is it does not use conventional instrumental variables strategies 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. Instead, we use variation in market size (which does not need to be exogenous) and cost data for identification.

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 such heterogeneity. As Berry and Haile (2014) and others point out, as long as there are instruments available, fairly flexible demand functions can be identified using market-level data. Popular 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"). The attractiveness of this approach is that even 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.[2]

Recently, some researchers have 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 (2015) similarly exploit first order conditions and firm-level cost data to identify the cost

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

[2] There has been some research assessing numerical difficulties with the BLP algorithm (Dube et al. (2012) and Knittel and Metaxoglou (2012)), and the use of optimal instruments to help alleviate these difficulties (Reynaert and Verboven (2014)).

functions of cable companies.[3] Kutlu and Sickles (2012) estimate market power while allowing for inefficiency in production by exploiting cost data. Like previous research, these researchers use instrumental variables (IVs) to identify demand in a first step.

Our study is motivated by these recent applications that combine cost data with standard demand data for model identification and testing.[4] The type of cost data we have in mind comes from firms' income statements and balance sheets, among other sources. Such data has been used extensively in a large parallel literature on cost function estimation in empirical IO.[5] We thus believe our study is promising since it aims to unify this literature with research on differentiated products models.

We extend the existing research on BLP-type models by developing and formalizing new ways to obtain additional identification with cost data. Our main theoretical finding is that by combining demand and cost data, one can jointly identify BLP demand and a nonparametric cost function using variation in market size Q and the restriction that marginal revenue is a function of marginal cost.[6] The implicit exclusion restrictions that we exploit for identification are: (1) price p and market share s determine marginal revenue but do not directly enter in the cost function; and (2) ouput $q = sQ$ enters the cost function but does not directly enter the demand function.

Our paper is related to Bresnahan and Reiss (1990), who use variation in market size to empirically analyze firms' markups. The challenge in this direction of research is how to use the exclusion restriction for estimation if we allow for the supply shock, which we need to control for. In this paper, we propose to use the cost data. Formally, we argue that variation in market share, (i.e., variation in market size) that keeps output, input prices and expected cost (conditional on observed demand and supply variables) the same, should come solely from variation in the demand shock, not from changes in

[3] A number of papers have also used demand and cost data to test assumptions regarding conduct in oligopoly models. See, for instance, Byrne (2015), McManus (2007), Clay and Troesken (2003), Kim and Knittel (2003), and Wolfram (1999).

[4] 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.

[5] 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.

[6] As in the existing research on BLP models, profit maximization is only required to identify the cost function. We show that we can consistently estimate demand even if firms are not profit maximizing.

the supply shock. Hence, our results imply that one does not need to use conventional instruments to identify the endogenous price parameters in differentiated goods demand, nor to identify the cost function parameters where output is potentially endogenous.

In the empirical IO literature, it is often argued that cost data is unreliable and should not be used for the purposes of studying firm behavior.[7] In light of these concerns, we try to use cost data in as limited manner as possible. In particular, we use it only to alleviate the endogeneity issue of product price to demand shocks. We also show that our identification results go through in model specifications that allow for cost data with measurement error as well as systematic over/under reporting by firms. Furthermore, we impose minimal assumptions on our nonparametric cost function: we require only that the true total cost be increasing in output, input price and the cost shock. We do not need to derive the marginal cost, analytically or numerically, in identifying and estimating logit or BLP demand and cost functions.

We also prove nonparametric identification to demonstrate that our identification strategy is not entirely based on functional form assumptions on the demand side. We prove that marginal revenue and marginal cost are jointly nonparametrically identified by the sample analog of the first order condition that equates marginal revenue and marginal cost corresponding to two close points in the data. We do so on a cross section of data, without any functional form restrictions on demand or costs, nor on the observed variables and unobserved demand and cost shocks, and without any use of orthogonality conditions between observed variables and demand/cost shocks. From marginal revenue, one can locally identify a nonparametric market share function.

Our nonparametric identification analysis also highlights a Curse of Dimensionality that likely makes an estimator based on the nonparametric identification argument and the direct application of the parametric identification argument impractical for applied research. This motivates our efficient Non-Linear Least Squares (NLLS)-sieve estimator, which does not suffer from the dimensionality problem. This estimator is semi-parametric in that it recovers a parametric logit or BLP demand structure and a non-parametric cost function. We also show how this estimator can be adapted to accommodate various data and specification issues that arise in practice. These include endogenous product

[7]See Nevo (2001) and Fisher and McGowan (1983).

characteristics, imposing restrictions to ensure well-behaved cost functions, dealing with the difference between accounting cost and economic cost, missing cost data for certain products or firms, and multi-product firms.

Through a set of Monte-Carlo experiments, we illustrate how our estimator delivers consistent parameter estimates when demand shock is not only correlated with equilibrium price and output, but also with cost shock, input prices and market size, and when cost shock is correlated with market size. In such a setting there are no valid instruments to account for price endogeneity, and market size alone cannot control for the supply side. The IV estimates, on the other hand, are shown to have bias.[8]

A prominent example of papers that exploit 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 the one that rationalizes these national margins.[9] Our study differs considerably in that we focus on the more common situation where a researcher has data on prices, aggregate market shares and total costs, but not marginal costs. Indeed, we directly build on the general BLP framework.

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 study identification when demand and cost data are available and develop our formal identification results. Section 4 proposes estimators that are based on our identification arguments and analyzes the large sample properties of our preferred NLLS-sieve estimator. 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 and discuss potential applications of our estimator.

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

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

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 $\boldsymbol{\theta} = [\boldsymbol{\beta}',\ \alpha]'$ 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.[10] 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 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 market share equation by $s\left(\boldsymbol{\delta}_m(\boldsymbol{\theta}), j\right) \equiv s\left(\mathbf{p}_m, \mathbf{X}_m, \boldsymbol{\xi}_m, j, \boldsymbol{\theta}\right)$ where $\boldsymbol{\delta}_m(\boldsymbol{\theta}) = [\delta_{0m}(\boldsymbol{\theta}), \delta_{1m}(\boldsymbol{\theta}), \ldots, \delta_{J_m m}(\boldsymbol{\theta})]'$ is a $J_m + 1 \times 1$ 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 (i.e., $\mathbf{x}_{0m} = \mathbf{0}$, $p_{0m} = 0$,

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

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 coefficients on the observed characteristics to be different for different consumers. Specifically, α has a distribution function $F_\alpha(.;\boldsymbol{\theta}_\alpha)$, where $\boldsymbol{\theta}_\alpha$ is the parameter vector of the distribution, and similarly, $\boldsymbol{\beta}$ has a distribution function $F_{\boldsymbol{\beta}}(.;\boldsymbol{\theta}_\beta)$ with parameter vector $\boldsymbol{\theta}_\beta$. The probability a consumer with 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 $\boldsymbol{\beta}$,

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

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.

2.1.1 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 $\boldsymbol{\delta}_m$ for market m that solves $\mathbf{s}(\boldsymbol{\delta}_m, \boldsymbol{\theta}) - \mathbf{s}_m = \mathbf{0}$, 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 predicted by the model. That is, market share inversion involves solving the following set of J_m equations,

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

and therefore,

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

The vector of mean utilities that solves these equations perfectly aligns 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 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 mean utilities in hand, recovering the structural demand shocks is straightforward,

$$\xi_{jm}(\boldsymbol{\theta}) \equiv \xi\left(\mathbf{p}_m, \mathbf{X}_m, \boldsymbol{\xi}_m, j, \boldsymbol{\theta}\right) = \delta_{jm}(\boldsymbol{\theta}) - \mathbf{x}'_{jm}\boldsymbol{\beta} - \alpha p_{jm} \tag{6}$$

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. That is,

$$\xi_{jm}(\boldsymbol{\theta}) \equiv \xi\left(\mathbf{p}_m, \mathbf{X}_m, \boldsymbol{\xi}_m, j, \boldsymbol{\theta}\right) = \delta_{jm}(\boldsymbol{\theta}) - \mathbf{x}'_{jm}\boldsymbol{\mu}_\beta - \boldsymbol{\mu}_\alpha p_{jm} \tag{7}$$

2.1.2 IV estimation of demand

A simple regression 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 regressors would yield a biased estimate of the price coefficient. This is because firms likely set higher prices for products with higher unobserved product quality, which creates correlation between p_{jm} and ξ_{jm}.

Researchers use a variety of 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, good instruments are not easy to find, as we discuss below.

Cost shifters are often used as instruments. This is in line with traditional market equilibrium analysis which identifies the demand curve from shifts in the supply curve. 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 products in the same local market through changes in consumer income. Changes in other input prices such as gasoline or electricity could affect both firms' and consumers' choices. Moreover, higher input prices may induce firms to reduce unobserved product quality, and hence could also undermine the exclusion restriction.

In instances where cost shifters are likely to satisfy the exclusion restriction, they are often weak instruments. For example, if one assumes that input prices are exogenously determined in external labor and capital markets, 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 the utility function in (1).

BLP originally proposed using product characteristics of rivals' products as price instruments. As we can see in equation (3), a firm's market share is a function of prices and observed product characteristics of all firms in its market. Therefore, the exclusion restriction that enables one to use rivals' product characteristics as price instruments heavily depends on functional form assumptions for the utility function and the distribution function of utility shocks.

A further potential issue with these instruments is that product characteristics, like prices, may be chosen strategically by firms. Indeed recently a literature on endogenous characteristics has emerged,[11] which raises concerns that product characteristics are also endogenous. In this case, IVs are needed for prices as well as product characteristics for identification.

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 products across markets that create cross-market correlation in product prices. However, researchers need to ensure they use such instruments in cases where there is no spatial correlation in demand shocks across markets as this would render these instruments invalid. Regional demand shocks or national marketing campaigns, for example, could generate such correlation.[12]

[11] See Crawford (2012) for an overview of this literature.

[12] 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 is correlated across markets.

2.2 Supply

The cost of producing q_{jm} units of product j in market m, C_{jm}, 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,

$$C_{jm} = C\left(q_{jm}, \mathbf{w}_{jm}, \upsilon_{jm}, \boldsymbol{\tau}\right), \tag{8}$$

where $\boldsymbol{\tau}$ is the cost parameter vector. In addition, $C()$ is assumed to be continuously differentiable with respect to output.

Given this cost function and the demand model above, we can write firm j's profit function as

$$\pi_{jm} = p_{jm} \times s\left(\mathbf{p}_m, \mathbf{X}_m, \boldsymbol{\xi}_m, j, \boldsymbol{\theta}\right) \times Q_m - C\left(s\left(\mathbf{p}_m, \mathbf{X}_m, \boldsymbol{\xi}_m, j, \boldsymbol{\theta}\right) \times Q_m, \mathbf{w}_{jm}, \upsilon_{jm}, \boldsymbol{\tau}\right), \tag{9}$$

where we assume there is one firm for each product. Keeping with BLP, we for now assume that firms act as differentiated products Bertrand price competitors. Therefore, the optimal price and quantity of product j in market m are determined by the first order condition (F.O.C.) that equates marginal revenue and marginal cost

$$\underbrace{p_{jm} + s_{jm}\left[\frac{\partial s\left(\mathbf{p}_m, \mathbf{X}_m, \boldsymbol{\xi}_m, j, \boldsymbol{\theta}\right)}{\partial p_{jm}}\right]^{-1}}_{MR_{jm}} = \underbrace{\frac{\partial C\left(q_{jm}, \mathbf{w}_{jm}, \upsilon_{jm}, \boldsymbol{\tau}\right)}{\partial q_{jm}}}_{MC_{jm}}. \tag{10}$$

Note that MR_{jm} in (10) can be written solely as a function of prices, market shares and parameters. This is because, given the market share inversion in (5), and the specification of mean utility δ_{jm}, demand shock $\boldsymbol{\xi}_m$ is a function of $\mathbf{X}_m$, $\mathbf{p}_m$, $\mathbf{s}_m$ and $\boldsymbol{\theta}$. Therefore, marginal revenue can be written as

$$MR_{jm} \equiv MR_{jm}(\boldsymbol{\theta}) \equiv MR(\mathbf{p}_m, \mathbf{s}_m, \mathbf{X}_m, j, \boldsymbol{\theta}). \tag{11}$$

This turns out to be quite useful in developing our identification and estimation approach below. Also, equations (10) and (11) imply that demand parameters can potentially be identified if there is data on marginal cost or even without such data, if the cost function can be estimated and its derivative with respect to output can be taken.

2.2.1 Cost function estimation

As with demand estimation, a similar inversion procedure can be used to recover unobserved cost shocks,

$$C_{jm} = C\left(q_{jm}, \mathbf{w}_{jm}, \upsilon_{jm}, \boldsymbol{\tau}\right) \Rightarrow \upsilon_{jm}(\boldsymbol{\tau}) = C^{-1}\left(q_{jm}, \mathbf{w}_{jm}, C_{jm}, \boldsymbol{\tau}\right). \tag{12}$$

Like demand estimation, there are important endogeneity concerns with standard approaches to estimating cost functions. Specifically, output q_{jm} is endogenously determined by profit maximizing firms equating marginal revenue to marginal cost as in equation (10), and is potentially negatively correlated with the cost shock υ_{jm}. That is, all else equal, less efficient firms tend to produce less. In dealing with this issue, researchers have traditionally focused on selected industries where endogeneity can be ignored, or used instruments for output.

The IV approach to cost function estimation typically uses excluded demand shifters as instruments. We denote this 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 has pitfalls that are similar to the ones we discussed with IV demand estimation. In particular, typical excluded demand shifters such as demographics affect all firms, and thus generate little within-market, across-firm variation in equilibrium output for identification. Moreover, one cannot completely rule out the possibility of correlation between demand shifters and cost shocks.

3 Identification and estimation of the price coefficient

In this section, we investigate the benefits of jointly using demand and cost data to identify the model. It turns out that the endogeneity concerns in estimating the demand and cost parameters can be mitigated if the parameters are jointly estimated using such data. Fundamental to this result is having variation in market size across markets, which is allowed to be arbitrarily correlated with unobserved demand and cost shocks.

The identification analysis is developed in three parts. For now we focus solely on identification of the price coefficient; later we extend the analysis to include other demand parameters such as the coefficients on product characteristics, and cost function. First we present the main idea for our general identification results in a simple monopoly model with Berry (1994) logit demand. We then maintain the simple logit demand structure, and prove identification under various extensions to the supply-side of the model: oligopoly, cost data with measurement error and systematic misreporting, and fixed costs. Second, we prove identification in a setting with a richer BLP/random coefficients model of demand. Here we emphasize the need to assume that marginal revenue identifies the price coefficient and show this assumption holds for logit and BLP demand. Third, we show the marginal revenue and market share functions are in fact non-parametrically identified. We find, however, that the estimation strategy that directly follows our parametric or nonparametric identification argument is likely to be subject to a Curse of Dimensionality in practice. This motivates us to pursue a different parametric estimation strategy in Section 4.

3.1 Identification in the Logit model

3.1.1 Main assumptions

We begin by considering identification in the simplest possible environment: a monopolist who sells one product and faces logit demand. The following six assumptions are the main ones needed for identification in this model and its extensions.

Assumption 1 *Researchers have data on outputs, product prices, market shares, input prices, and observed product characteristics of firms. In addition, data on firms' 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 in output, input prices and cost shock.*

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

We denote MC_{jm} to be the marginal cost of firm j in market m, i.e. $MC_{jm} = MC(q_{jm}, \mathbf{w}_{jm}, \upsilon_{jm})$.

Assumption 5 *Firms set prices such that marginal revenue is a function of marginal cost, taking as given their rivals' prices. That is,* $MR_{jm} = \zeta(MC_{jm})$.

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* $\mathbf{p}_m$, $\mathbf{s}_m$, $\mathbf{X}_m$, υ_{jm} *and* $\boldsymbol{\xi}_m$ *such that under the true parameter vector* $\boldsymbol{\theta}_0$, $\delta_0 - 1 \leq \left[\frac{\partial \ln s(\mathbf{p}_m, \mathbf{X}_m, \boldsymbol{\xi}_m, j, \boldsymbol{\theta}_0)}{\partial \ln p_{jm}}\right]^{-1} \leq -\delta_0$ *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* $\alpha_0 < 0$ *for the logit model and* $\mu_{\alpha 0} < 0$ *for the BLP random coefficient model.*[13]

For simplicity, throughout, we also assume that firms in the same market m share the common input price vector $\mathbf{w}_m$. This assumption can be weakened without changing any of the results below.

3.1.2 Monopoly

The intuition for how the price coefficient can be identified by using demand and cost data jointly can be illustrated with a single-product monopolist facing logit demand. Assuming for the moment that the monopolist maximizes profits, the following first order condition holds in equilibrium:

$$MR(\mathbf{p}_m, \mathbf{s}_m, \mathbf{X}_m, \boldsymbol{\theta}_0) = p_m + \frac{1}{(1 - s_m)\,\alpha_0} = MC(q_m, \mathbf{w}_m, \upsilon_m), \; q_m = Q_m s_m,$$

[13] Assumption 6 is not often discussed in the BLP setup. If we generate demand shocks that have reasonably large variance and are independent of 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 occur 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. The latter two strategies potentially result in bias of the price coefficient estimate. As we will see later, since our identification and estimation strategy of the price coefficient does not use any orthogonality conditions involving demand shocks, it is 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.

where α_0 is the true price coefficient. We use the exclusion restriction that the monopolist's market share s_m does not directly enter in the marginal cost function, and its output q_m does not directly enter in the marginal revenue equation. It is the market size Q_m that provides the link between market share and output, and variation between the two. Now, suppose there were no cost shocks. That is, the marginal cost function is $MC(q_m, \mathbf{w}_m)$. Then, if we find a pair of monopolists in different markets with different market sizes, prices and market shares but with the same output and input price vector, then we know they must have the same marginal cost. Thus, the equality of marginal revenue to marginal cost implies

$$p_m + \frac{1}{(1-s_m)\alpha_0} = p_{m'} + \frac{1}{(1-s_{m'})\alpha_0}, \quad \alpha_0 = -\frac{1}{p_m - p_{m'}}\left[\frac{1}{1-s_m} - \frac{1}{1-s_{m'}}\right] \tag{13}$$

and the price coefficient α_0 is identified.

The same argument cannot be made in the presence of a cost shock υ_m to MC_m, i.e. $MC_m = MC(q_m, \mathbf{w}_m, \upsilon_m)$. Then, marginal costs of two firms with the same output q_m and input prices $\mathbf{w}_m$ will not be equal, thus, for these two firms, equation (13) will not hold. In order to allow for a cost shock, we modify this identification argument and pair up firms that have different Q_m and s_m, but have the same q_m, $\mathbf{w}_m$ *and* C_m. It follows that these firms must have the same cost shock, and thus, we can identify the price coefficient with equation (13)

Note that we have assumed profit maximization for expositional purposes only. It is not required for identification. Assumption 5 ensures that if we pair firms such that they have the same output, input prices, total cost and (hence) cost shock, then $MC_m = MC_{m'} \Rightarrow MR_m = MR_{m'}$ and we can identify the price coefficient using equation (13).

3.1.3 Oligopoly

We now extend the identification argument to the oligopoly case, where in market m there are J_m firms selling one product each.[14]. Let C^d_{jm} be the observed cost, and let $P_{jm} = \{Q_m, q_{jm}, \mathbf{w}_m, \mathbf{p}_m, \mathbf{s}_m, \mathbf{X}_m, j\}$ contain the other relevant information (beyond cost) about firm j in market m.

[14]We extend our results to the multi-product case in Section 4.

In addition to Assumptions 1-6, identification in the oligopoly case relies on the following assumption,

Assumption 7 *There exists a pair of observations* $P_{jm} = \{Q_m, q_{jm}, \mathbf{w}_m, \mathbf{p}_m, \mathbf{s}_m, \mathbf{X}_m, j\}$ *and* $P_{j'm'} = \{Q_{m'}, q_{j'm'}, \mathbf{w}_{m'}, \mathbf{p}_{m'}, \mathbf{s}_{m'}, \mathbf{X}_{m'}, j'\}$ *that satisfy* $\mathbf{w}_m = \mathbf{w}_{m'}$*,* $q_{jm} = q_{j'm'}$*,* $C^d_{jm} = C^d_{j'm'}$ *and* $p_{jm} \neq p_{j'm'}$*.*

Going forward, we drop the cost parameter vector $\boldsymbol{\tau}$ as we will treat the cost function $C(\cdot)$ as nonparametric for the remainder of the paper.

Proposition 1 *Suppose Assumptions 1-7 are satisfied. Then,* $\mathbf{w}_m = \mathbf{w}_{m'}$*,* $q_{jm} = q_{jm'}$*,* $C^d_{jm} = C^d_{j'm'}$ *implies* $\upsilon_{jm} = \upsilon_{j'm'}$*, where* υ_{jm} *is the cost shock corresponding to the set of observations* P_{jm}*, and* $\upsilon_{j'm'}$ *for* $P_{j'm'}$*. Further, under the logit demand model the true price coefficient* α_0 *is identified by*

$$\alpha_0 = -\frac{1}{p_{jm} - p_{j'm'}}\left[\frac{1}{1 - s_{jm}} - \frac{1}{1 - s_{j'm'}}\right]. \tag{14}$$

Proof. Suppose $\upsilon_{jm} > \upsilon_{j'm'}$. Then, since the cost function is strictly increasing in υ,

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

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

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

Because $MR = \zeta(MC)$ by Assumption 5, if data points jm and $j'm'$ have the same marginal cost then they must have the same marginal revenue. In the case of logit model, this implies

$$p_{jm} + \frac{1}{\left(1 - s_{jm}\right)\alpha_0} = p_{j'm'} + \frac{1}{\left(1 - s_{j'm'}\right)\alpha_0}.$$

It then follows that α_0 is identified from such a pair of data points as follows

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

■

The above result highlights the importance of the variation of market size for identification. If all the data came from a single market, or from two markets of the same size, then $q_{jm} = q_{j'm'}$ implies $s_{jm} = s_{j'm'}$, thus $p_{jm} = p_{j'm'}$ if $\alpha_0 \neq 0$ and thus α_0 cannot be identified from (14), unless the true value is $\alpha_0 = 0$. The example also illustrates the role of cost data in identifying the price coefficient. As long as we can find firms with the same cost, output, and input price, they will have the same cost shock and marginal cost, thereby allowing us to difference away their supply side effects in a pairwise fashion.

What Assumption 7 states is that there exist two markets m, m' with different market sizes $Q_m \neq Q_{m'}$ with two firms jm, $j'm'$ that have the same input price vector $\mathbf{w}_m = \mathbf{w}_{m'}$ and the same output $q_{jm} = q_{j'm'}$, but potentially different vectors of demand shocks $\boldsymbol{\xi}_m \neq \boldsymbol{\xi}_{m'}$, thus, different market shares $\boldsymbol{s}_m \neq \boldsymbol{s}_{m'}$. Two such firms can exist even if we allow for correlation between the input price and the cost shocks. In the same manner, correlation between demand shocks and input price, or between demand shocks and cost shocks do not prevent us from finding pairs of firms satisfying Assumption 7, as long as there is sufficient variation in market size and demand shocks.

In addition to allowing for such correlation between observables and unobservables within markets, our approach also allows demand and cost shocks to be correlated across markets, as long as the correlation is not perfect. Further yet, we do not need to assume exogeneity of $\mathbf{X}_m$ to identify α as is typically assumed. In sum, these findings illustrate that given cost data, one does not need any conventional IV- or orthogonality assumptions. And, as long as marginal revenue is a function of, but not necessarily equal to, marginal cost, we can identify α without assuming profit maximization. This makes our framework applicable to firms that are under government regulation and firms under organizational incentives or behavioral aspects that prevent them from maximizing profit.

We note that, in practice, Assumption 7 is unrealistic. However, a similar argument can be made for pairs that satisfy the equalities in Assumption 7 approximately.

3.1.4 Measurement error and misreporting in costs

Two important issues are likely to arise in practice with the above identification strategy. First, suppose there exist two pairs that satisfy Assumption 7 and each pair provides 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. Second, it is widely accepted that cost data are measured with error.

In light of these issues, we weaken the model specification by allowing for both measurement error as well as systematic misreporting of true costs. The following assumption generalizes our cost function specification.

Assumption 8 *The observed cost of firm j in market m, C^d_{jm} differs from the true cost C_{jm} as follows.*

$$C^d_{jm} = \varphi\left(C_{jm}\right) + \eta_{jm}. \tag{15}$$

where φ is a strictly increasing function and measurement error η_{jm} is i.i.d. distributed with mean 0 and variance σ^2_η. In addition, we assume measurement error is independent of $\{q_{jm}, \mathbf{w}_{jm}, \mathbf{p}_m, \mathbf{s}_m, \mathbf{X}_m, j\}$, for all j, m.

So, for example, if firms report costs truthfully but with error then $\varphi(C) = C$. Alternatively, if firms systematically under-report their true costs, then we could consider a specification like $\varphi(C) = \varphi C$ where $0 < \varphi < 1$. Over-reporting could be captured by the same specification, but where $\varphi > 1$.

It turns out that in order to identify the price coefficient given cost data characterized by Assumption 8, the only modification we need to make to our identification argument is that we work with firms with the same mean cost conditional on $\{q_{jm}, \mathbf{w}_{jm}, \mathbf{p}_m, \mathbf{s}_m, \mathbf{X}_m, j\}$, rather than firms with the same cost data. Assumption 9 formalizes this requirement.

Assumption 9 *There exist two firms with $P = \{Q_m, q_{jm}, \mathbf{w}_m, \mathbf{p}_m, \mathbf{s}_m, \mathbf{X}_m, j\}$ and $P' = \{Q_{m'}, q_{j'm'}, \mathbf{w}_{m'}, \mathbf{p}_{m'}, \mathbf{s}_{m'}, \mathbf{X}_{m'}, j'\}$, where $q_{jm} = q_{j'm'}$, $\mathbf{w}_m = \mathbf{w}_{m'}$, $\mathbf{p}_m \neq \mathbf{p}_{m'}$ and*

$$E\left[C^d | q_{jm}, \mathbf{w}_m, \mathbf{p}_m, \mathbf{s}_m, \mathbf{X}_m, j\right] = E\left[C^d | q_{j'm'}, \mathbf{w}_{m'}, \mathbf{p}_{m'}, \mathbf{s}_{m'}, \mathbf{X}_{m'}, j'\right].$$

Proposition 2 *Suppose Assumptions 1-6, 8 and 9 are satisfied. Then, $\upsilon_{jm} = \upsilon_{j'm'}$ and under the logit model of demand, α_0 is identified.*

Proof. The proof is very similar to that of Proposition 1. Since the measurement errors are i.i.d. and independent of $\{q_{jm}, \mathbf{w}_m, \mathbf{p}_m, \mathbf{s}_m, \mathbf{X}_m, j\}$ and $\{q_{j'm'}, \mathbf{w}_{m'}, \mathbf{p}_{m'}, \mathbf{s}_{m'}, \mathbf{X}_{j'}, j'\}$,

$$\begin{aligned} E\left[C^d|q_{jm}, \mathbf{w}_m, \mathbf{p}_m, \mathbf{s}_m, \mathbf{X}_m, j\right] &= E\left[\left(\varphi\left(C\left(q_{jm}, \mathbf{w}_m, \upsilon_{jm}\right)\right) + \eta\right)|q_{jm}, \mathbf{w}_m, \mathbf{p}_m, \mathbf{s}_m, \mathbf{X}_m, j\right] \\ &= \varphi\left(C\left(q_{jm}, \mathbf{w}_m, \upsilon_{jm}\right)\right). \end{aligned}$$

Similarly,

$$E\left[C^d|q_{j'm'}, \mathbf{w}_{m'}, \mathbf{p}_{m'}, \mathbf{s}_{m'}, \mathbf{X}_{j'}, j'\right] = \varphi\left(C\left(q_{jm}, \mathbf{w}_m, \upsilon_{j'm'}\right)\right).$$

By Assumption 9,

$$E\left[C^d|q_{jm}, \mathbf{w}_m, \mathbf{p}_m, \mathbf{s}_m, \mathbf{X}_m, j\right] = E\left[C^d|q_{j'm'}, \mathbf{w}_{m'}, \mathbf{p}_{m'}, \mathbf{s}_{m'}, \mathbf{X}_{j'}, j'\right].$$

Thus, given $q_{jm} = q_{j'm'}$, $\mathbf{w}_m = \mathbf{w}_{m'}$, and $\varphi()$ being an increasing function, it follows that $\upsilon_{jm} = \upsilon_{j'm'}$. Therefore,

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

and,

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

and the price coefficient α_0 is identified. ■

The conditional mean function $E\left[C^d|q, \mathbf{w}, \mathbf{p}, \mathbf{s}, \mathbf{X}, j\right]$ can be recovered from the data by kernel or sieve based regression where the dependent variable is the cost data C^d and the independent variables are sieve polynomials of $(q, \mathbf{w}, \mathbf{p}, \mathbf{s}, \mathbf{X})$. In practice, this would likely be subject to a Curse of Dimensionality.

3.1.5 Fixed costs

We can further extend the above identification argument to include a fixed cost. To begin, we denote fixed costs as

$$F\left(\upsilon_{jm}\right) + \varsigma_{jm}^F,$$

where ς_{jm}^F is independent of υ_{jm}, and wlog, has mean zero. The modified cost function that includes variable and fixed costs is

$$\tilde{C}\left(q_{jm}, \mathbf{w}_m, \upsilon_{jm}\right) + \varsigma_{jm}^F \equiv C\left(q_{jm}, \mathbf{w}_m, \upsilon_{jm}\right) + F\left(\upsilon_{jm}\right) + \varsigma_{jm}^F.$$

where $\tilde{C}\left(q_{jm}, \mathbf{w}_m, \upsilon_{jm}\right) \equiv C\left(q_{jm}, \mathbf{w}_m, \upsilon_{jm}\right) + F\left(\upsilon_{jm}\right)$ and the relationship between the observed and true costs is updated to be

$$C_{jm}^d = \varphi\left(\tilde{C}\left(q_{jm}, \mathbf{w}_m, \upsilon_{jm}\right) + \varsigma_{jm}^F\right) + \eta_{jm}.$$

That is, we allow for the systematic misreporting of the true cost, which now includes the random term of the fixed cost, ς_{jm}^F.

In this set-up, as long as $\tilde{C}_{jm} \equiv \tilde{C}\left(q_{jm}, \mathbf{w}_m, \upsilon_{jm}\right)$ satisfies Assumption 3 then Proposition 2 holds. To see why, first note that

$$E_{(\varsigma^F,\eta)}\left[\varphi\left(\tilde{C}_{jm} + \varsigma^F\right) + \eta | \tilde{C}_{jm}\right] > E_{(\varsigma^F,\eta)}\left[\varphi\left(\tilde{C}_{j'm'} + \varsigma^F\right) + \eta | \tilde{C}_{j'm'}\right]$$

implies

$$E_{\varsigma^F}\left[\varphi\left(\tilde{C}_{jm} + \varsigma^F\right)\right] > E_{\varsigma^F}\left[\varphi\left(\tilde{C}_{j'm'} + \varsigma^F\right)\right].$$

Because $\varphi\left(\right)$ is an increasing function, this implies $\tilde{C}_{jm} > \tilde{C}_{j'm'}$. The opposite inequalities also hold. That is,

$$E_{(\varsigma^F,\eta)}\left[\varphi\left(\tilde{C}_{jm} + \varsigma^F\right) + \eta | \tilde{C}_{jm}\right] < E_{(\varsigma^F,\eta)}\left[\varphi\left(\tilde{C}_{j'm'} + \varsigma^F\right) + \eta | \tilde{C}_{j'm'}\right]$$

implies $\tilde{C}_{jm} < \tilde{C}_{j'm'}$. Therefore,

$$E_{(\varsigma^F,\eta)}\left[\varphi\left(\tilde{C}_{jm} + \varsigma^F\right) + \eta | \tilde{C}_{jm}\right] = E_{(\varsigma^F,\eta)}\left[\varphi\left(\tilde{C}_{j'm'} + \varsigma^F\right) + \eta | \tilde{C}_{j'm'}\right]$$

implies $\tilde{C}_{jm} = \tilde{C}_{j'm'}$. Therefore,

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

and the same identification proof from Proposition 2 goes through.

3.2 Identification of marginal revenue

We now enrich the demand-side of the model, and show that parameters of the distribution of price coefficients in the BLP model are identified as well. As with the logit model, the identification analysis has two parts. The first part, which is related to supply, concerns finding pairs of firms with the same output, input price vector and conditional mean cost. The second part, which is related to demand, concerns whether price coefficients can be identified from these firm pairs with the same marginal revenue. Since the cost function is nonparametric, the first part of the analysis remains unchanged for different parametric demand specifications. Only the second part changes, for which we present general conditions for identification. We then show that the logit and BLP demand model satisfy these conditions.

3.2.1 General conditions for identification

We make the following assumptions on the support of the market size, demand and cost shock, and marginal cost. These greatly simplify the identification proofs without imposing any orthogonality restrictions.

Assumption 10 *The following properties hold for all $jm : j = 1, \ldots, J_m, m = 1, \ldots, M$. Let $\mathcal{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_m)$. Let the vector of market size of all markets other than m be $\mathbf{Q}_{-m} = (Q_1, Q_2, \ldots, Q_{m-1}, Q_{m+1}, \ldots Q_M)$. Similarly let the vector of input prices of all markets other than m be denoted as $\mathbf{W}_{-m} = (\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_{m-1}, \mathbf{w}_{m+1}, \ldots \mathbf{w}_M)$. Let the vector of demand shock of all firms other than jm be $\boldsymbol{\Xi}_{-j,-m} = \left(\boldsymbol{\xi}_1', \ldots, \boldsymbol{\xi}_{m-1}', \boldsymbol{\xi}_{-jm}', \boldsymbol{\xi}_{m+1}', \ldots \boldsymbol{\xi}_M'\right)$, where $\boldsymbol{\xi}_{-jm}$ is the vector of demand shocks of all firms in market m other than the firm j. The cost shock of all firms other than jm is defined analogously to $\boldsymbol{\Xi}_{-j,-m}$ and is denoted by $\boldsymbol{\Upsilon}_{-j,-m}$. Further, define $\mathbf{V}_{-j,-m} \equiv (\mathbf{Q}_{-m}, \mathbf{W}_{-m}, \boldsymbol{\Xi}_{-j,-m}, \boldsymbol{\Upsilon}_{-j,-m}, \mathcal{X})$. Then, the support of the market size Q_m given ξ_{jm}, υ_{jm}, $\mathbf{w}_m$ and $\mathbf{V}_{-j,-m}$ is the positive real line, and so is the support of the cost shock υ_{jm} given Q_m, ξ_{jm}, $\mathbf{w}_m$ and $\mathbf{V}_{-j,-m}$. Similarly, the support of the demand shock ξ_{jm} given Q_m, υ_{jm}, $\mathbf{w}_m$ and $\mathbf{V}_{-j,-m}$ is R, the real line.*

Furthermore, the support of the input price $\mathbf{w}_m$ *given* Q_m*,* ξ_{jm}*,* υ_m *and* $\mathbf{V}_{-j,-m}$ *is* R^L_+*, where* L *is the number of inputs.*

Assumption 11 *For any* $q > 0$*,* $\mathbf{w} > \mathbf{0}$*, the marginal cost function satisfies the following properties:*

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

Furthermore,

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

The key identification assumption for a general parametric demand function is that marginal revenue identifies the price coefficient. That is, one can find two vectors of prices, market shares and matrices of observed characteristics that have the same marginal revenue under the true price parameters, but different marginal revenues under the wrong price parameters. We introduce some notation before stating this assumption formally. Let the parameter vector have two components, $\boldsymbol{\theta} = (\boldsymbol{\theta}_{-p}, \boldsymbol{\theta}_p) \in \boldsymbol{\Theta}_{-p} \times \boldsymbol{\Theta}_p$ where $\boldsymbol{\Theta}_{-p}$ is the parameter space of $\boldsymbol{\theta}_{-p}$ and $\boldsymbol{\Theta}_p$ is the parameter space of $\boldsymbol{\theta}_p$. Roughly, $\boldsymbol{\theta}_p$ corresponds to the price coefficient that we identify below.

Let $\boldsymbol{\nu}$ and $\boldsymbol{\nu}'$ be two sets of vectors of prices, market shares, and observed product characteristics that can be generated as an equilibrium of the oligopoly model under corresponding assumptions, with J rows for the former and J' rows for the latter, corresponding to two markets,

$$\boldsymbol{\nu} = \left\{\mathbf{p}, \mathbf{s}, \mathbf{X}\right\}, \ \boldsymbol{\nu}' = \left\{\mathbf{p}', \mathbf{s}', \mathbf{X}'\right\}, \ \boldsymbol{\nu} \neq \boldsymbol{\nu}'.$$

Assumption 12 *For any given* $\boldsymbol{\theta}_p \neq \boldsymbol{\theta}_{p0}$*, there exist* $\boldsymbol{\nu}$ *and* $\boldsymbol{\nu}', \boldsymbol{\nu} \neq \boldsymbol{\nu}'$*, and* j *that satisfy the following properties.*

1. $p_l > 0$*,* $0 < s_l < 1$ *for* $l = 1, ..., J$ *and* $p'_l > 0$*,* $0 < s'_l < 1$*, for* $l = 1, .., J'$*, and* $0 < \sum_{l=1}^{J} s_l < 1$*,* $0 < \sum_{l=1}^{J'} s'_l < 1$*.*

2. *For any* $\boldsymbol{\theta}_{-p} \in \boldsymbol{\Theta}_{-p}$*,*

$$MR(\mathbf{p}, \mathbf{s}, \mathbf{X}, j, \boldsymbol{\theta}_{-p}, \boldsymbol{\theta}_{p0}) = MR\left(\mathbf{p}', \mathbf{s}', \mathbf{X}', j, \boldsymbol{\theta}_{-p}, \boldsymbol{\theta}_{p0}\right)$$
$$\textit{and } MR(\mathbf{p}, \mathbf{s}, \mathbf{X}, j, \boldsymbol{\theta}_{-p}, \boldsymbol{\theta}_{p}) \neq MR\left(\mathbf{p}', \mathbf{s}', \mathbf{X}', j, \boldsymbol{\theta}_{-p}, \boldsymbol{\theta}_{p}\right).$$

where $\theta_0 = (\theta_{-p0}, \theta_{p0})$ is the true parameter vector.

Proposition 3 *Suppose cost data is generated as in equation (15), and Assumptions 1-6 and Assumption 10-12 are satisfied. Then, $\boldsymbol{\theta}_{p0}$ is identified.*

Proof. See Appendix ■

3.2.2 Identification in the logit and BLP model

It is important to note that Assumption 12 is a high level assumption; it is not necessarily satisfied in all demand models. For example, if marginal revenue is a multiplicative, separable function of $\boldsymbol{\theta}_p$, then $\boldsymbol{\theta}_p$ is not identifiable by MR. To see this, notice that for any $\boldsymbol{\theta}_p$, and for any $(\mathbf{p}, \mathbf{s}, \mathbf{X})$ and $(\mathbf{p}', \mathbf{s}', \mathbf{X}')$, $MR(\mathbf{p}, \mathbf{s}, \mathbf{X}, j, \boldsymbol{\theta}) = MR(\mathbf{p}', \mathbf{s}', \mathbf{X}', j, \boldsymbol{\theta})$ implies $MR(\mathbf{p}, \mathbf{s}, \mathbf{X}, j, \boldsymbol{\theta}_{-p})\, g(\boldsymbol{\theta}_p) = MR(\mathbf{p}', \mathbf{s}', \mathbf{X}', j, \boldsymbol{\theta}_{-p})\, g(\boldsymbol{\theta}_p)$ for some $g(\cdot)$. Hence, Assumption 12 is violated.

The marginal revenue function for logit and BLP does not have a multiplicative, separable form, nor do most functions commonly used by researchers. The important question, then, is whether the logit and BLP demand models satisfy Assumption 12. The proposition below answers this question for the monopoly case.

Proposition 4 *Suppose Assumptions 1-6 and Assumptions 10-11 are satisfied, and there exist two firms jm and jm' with demand variables $\boldsymbol{\nu}_m = \{\mathbf{p}_m, \mathbf{s}_m, \mathbf{X}_m\}$ and $\boldsymbol{\nu}_{m'} = \{\mathbf{p}_{m'}, \mathbf{s}_{m'}, \mathbf{X}_{m'}\}$ where $p_{jm} \neq p_{jm'}$. Then,*

a. *Assumption 12 is satisfied for the logit model for monopoly markets.*

b. *Assumption 12 is satisfied for the BLP model without observed product characteristics for monopoly markets if $s_m \neq s_{m'}$ and*

$$\eta_0 \equiv \frac{\mu_{\alpha 0}}{\sigma_{\alpha 0}} < -\frac{1}{2\phi(0)}. \tag{16}$$

Proof. See Appendix. ■

We can further include controls $\mathbf{X}_m$ into the demand model, and show that the cost data identifies parameters of the random coefficients on price $(\mu_\alpha, \sigma_\alpha)$, as well as $\boldsymbol{\sigma}_{\boldsymbol{\beta}}$, the standard deviation of the distribution of $\boldsymbol{\beta}$. The details are shown in the Appendix, where we formally prove identification when the observed product characteristic x_{jm} is a scalar for all firms jm. As we can see from equations (6) and (7), what is not identified from the cost data alone is $\boldsymbol{\mu}_\beta$. This is because given $(\mu_\alpha, \sigma_\alpha)$, and $\boldsymbol{\sigma_\beta}$, only $\xi_{jm} + \mathbf{x}'_{jm}\boldsymbol{\mu}_\beta$ is identified without further restrictions imposed from the model.

Proving Assumption 12 for the BLP model with oligopoly markets is a straightforward extension of Proposition 4 and is thus omitted. It requires data that contain firms with high and similar prices: $p_{1m} = p_{2m} = ...p_{J_m m} = p$ for sufficiently high p. Despite the need for these strong assumptions in the formal argument for parametric identification of the BLP model, we later show that the parameters are well-identified in our Monte-Carlo experiments.[15]

Next we prove nonparametric identification to illustrate that identification of the demand function does not rely on its functional form restrictions like logit or BLP. Readers who are more interested in our new estimation procedure can skip Subsection 3.3 and directly move to Section 4.

3.3 Nonparametric identification of marginal revenue function

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. To simplify our discussion, we again first focus on monopoly markets and then extend our identification arguments to oligopoly markets.

We begin by making the following auxiliary assumptions for the monopoly model:

Assumption 13 *Let $MR(p, \mathbf{x}, \xi)$ be the marginal revenue specified as a function of price p, vector of product characteristics $\mathbf{x}$ and the demand shock ξ.*

a. *$MR(p, \mathbf{x}, \xi)$ is strictly increasing in p.*

b. *For any $\mathbf{x}$, $\mathbf{x}'$ and two pairs of prices and market shares (p, s) and (p', s') such that*

[15] The proof for the BLP model relies on firms with very high prices for identification. This is unattractive, but necessary to deal with the complexity in separately identifying parameters of the distribution of the random coefficients. As we will see below, nonparametric identification of marginal revenue and the market share equation does not rely on having such firms. It does however require stronger assumptions on cost and conduct.

$s = s'$ *and* $p > p'$,

$$MR\left(p, \mathbf{x}, \xi\left(p, s, \mathbf{x}\right)\right) > MR\left(p', \mathbf{x}', \xi\left(p', s', \mathbf{x}'\right)\right).$$

where recall the demand shock ξ *is an unspecified function of* p, s *and* $\mathbf{x}$.

Assumption 14 *The market share function* $s\left(p, \mathbf{x}, \xi\right)$ *is strictly decreasing and continuous in* p *and strictly increasing and continuous in* ξ. *Furthermore,*

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

Assumption 15 *Firms maximize profits, setting prices to equate marginal revenue and marginal cost. Furthermore,*

$$C^d_{jm} = C\left(q_{jm}, \mathbf{w}_{jm}, \upsilon_{jm}\right) + \eta_{jm}$$

where $E\left(\eta_{jm} | q_{jm}, \mathbf{w}_{jm}, \upsilon_{jm}\right) = 0$

Assumption 3′ *For any* $\mathbf{w}$, *the marginal cost function is nondecreasing and continuous in* q *and increasing and continuous in* υ. *Furthermore, for any* $\mathbf{w} > 0$, *and* $q > 0$,

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

Given these additional assumptions, we prove that marginal cost and marginal revenue are nonparametrically identified. The following proposition formally states our claim.

Proposition 5 *Suppose Assumptions 1, 2, 3′, 4, 5, 6 and Assumptions 13, 14 and 15 are satisfied. Consider a monopolist in market* m *with the set of observables* $P_m = \{Q_m, \mathbf{w}_m, q_m, p_m, s_m, \mathbf{x}\}$, *and demand shock and cost shock* ξ_m *and* υ_m, *respectively.*

a. *Suppose the marginal cost is increasing in output. Then consider the monopolist firm in market* m' *with observables* $P_{m'} = \{Q_{m'}, \mathbf{w}_{m'}, q_{m'}, p_{m'}, s_{m'}, \mathbf{x}_{m'}\}$ *close to* P_m, *such that* $\mathbf{w}_m = \mathbf{w}_{m'}$ *and* $\mathbf{x}_m = \mathbf{x}_{m'}$ *and is generated by the same demand shock (*$\xi_{m'} =$

$\xi_m = \xi$) and cost shock ($\upsilon_{m'} = \upsilon_m = \upsilon$) but has a different market size $Q_{m'} > Q_m$. It follows that

$$s_m > s_{m'},\ \ p_m < p_{m'},\ \ q_m < q_{m'}, \tag{17}$$

and

$$p_m \left[1 + \frac{lnp_{m'} - lnp_m}{lns_{m'} - lns_m}\right] = \frac{E\left[C^d|q_{m'}, \mathbf{w}_{m'}, p_{m'}, s_{m'}, \mathbf{x}_{m'}\right] - E\left[C^d|q_m, \mathbf{w}_m, p_m, s_m, \mathbf{x}_m\right]}{q_{j'm'} - q_m} + O\left(|Q_{m'} - Q_m|\right). \tag{18}$$

b. *Suppose the marginal cost is increasing in output. Suppose we have a firm m' with $P_{m'}$ close to P_m, such that both (17) and (18) hold. Then, the true marginal cost at $\{Q_m, \mathbf{w}_m, q_m, p_m, s_m, \mathbf{x}_m\}$, MC_m satisfies*

$$MC_m = \frac{E\left[C^d|q_{m'}, \mathbf{w}_{m'}, p_{m'}, s_{m'}, \mathbf{x}_{m'}\right] - E\left[C^d|q_m, \mathbf{w}_m, p_m, s_m, \mathbf{x}_m\right]}{q_{m'} - q_m} + O\left(|Q_{m'} - Q_m|\right),$$

and the nonparametric estimate of MC_m is given by,

$$\widehat{MC}_m = \frac{E\left[C^d|q_{m'}, \mathbf{w}_{m'}, p_{m'}, s_{m'}, \mathbf{x}_{m'}\right] - E\left[C^d|q_m, \mathbf{w}_m, p_m, s_m, \mathbf{x}_m\right]}{q_{m'} - q_m} \tag{19}$$

c. *Suppose the marginal cost is constant in output. Then consider another firm m' with $P_{m'} = \{Q_{m'}, \mathbf{w}_{m'}, q_{m'}, p_{m'}, s_{m'}, \mathbf{x}_{m'}\}$ close to P_m, that is generated by the same demand shock ($\xi_{m'} = \xi_m = \xi$) and cost shock ($\upsilon_{m'} = \upsilon_m = \upsilon$) and that has a different market size $Q_{m'} \neq Q_m$. It follows that*

$$s_{m'} = s_m,\ \ p_{m'} = p_m,\ \ q_{m'} = Q_{m'} s_{m'}, \tag{20}$$

and

$$MC_m = \frac{E\left[C^d|q_{m'}, \mathbf{w}_{m'}, p_{m'}, s_{m'}, \mathbf{x}_{m'}\right] - E\left[C^d|q_m, \mathbf{w}_m, p_m, s_m, \mathbf{x}_m\right]}{q_{m'} - q_m}. \tag{21}$$

Proof. See Appendix. ■

Parts a b and c of Proposition 5 state that the level of marginal revenue can be identified.[16] That is, parts a and b say that for a point P_m in the population, if we can find a nearby point $P_{m'}$ with the same $\mathbf{x}$ and $\mathbf{w}$, 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 in (19).

Part c of Proposition 5 states the following: if one cannot find $P_{m'}$ close to P_m where the sample analog of marginal revenue equals marginal cost, and one finds a nearby point whose prices and market shares are the same as P_m, but output is different, then it is likely that the marginal cost is a constant. Thus one can derive the marginal cost as in equation (21).

In implementing the above identification approach, in practice $E\left[C^d|q_m, \mathbf{w}_m, p_m, s_m, \mathbf{x}_m\right]$ could be nonparametrically estimated in a first step. Given the profit maximization assumption, we could then obtain a nonparametric marginal revenue estimate $\widehat{MR}_m$ from the corresponding marginal cost estimate in a second step, $\widehat{MR}_m = \widehat{MC}_m$.[17]

It is fairly straightforward to see that the logit model with a negative price coefficient satisfies Assumptions 13 a and b. We conducted an extensive numerical analysis with the BLP demand model with negative μ_β in monopoly markets and found that in all cases we tried, Assumptions 13 a. and b. are satisfied as long as marginal revenue is positive. However, we have not provided a formal proof, and thus one cannot completely rule out the possibility of Assumption 13 being violated.

Fortunately, Assumption 13 b. can be tested. Consider two monopoly firms who have the same input prices and product characteristics, and whose output, market size, and market shares are close to each other. In particular, for the point $\{Q_m, q_m, \mathbf{w}_m, p_m, s_m, \mathbf{x}_m\}$, take another close point $\{Q_{m'}, q_{m'}, \mathbf{w}_{m'}, p_{m'}, s_{m'}, \mathbf{x}_{m'}\}$ that satisfies $Q_m = Q_{m'}$, $s_m = s_{m'} = s$, thus $q_m = q_{m'}$, but $p_m < p_{m'}$. Then, if $E\left(C^d|q_m, \mathbf{w}_m, p_m, s_m, \mathbf{x}_m\right) < E\left(C^d|q_{m'}, \mathbf{w}_{m'}, p_{m'}, s_{m'}, \mathbf{x}_{m'}\right)$,

[16] Notice that a marginal revenue function that is multiplicatively separable in price parameters and other parameters and variables, is now identifiable given the additional identification assumptions.

[17] Recall that with parametric identification, we only needed to assume that marginal revenue was a function of marginal cost. For Proposition 5 however, to prove nonparametric identification, we require marginal revenue to equal marginal cost. It is the parametric functional form restriction that helped weaken the profit maximization assumption previously.

it implies that $C(q_m, \mathbf{w}_m, \upsilon_m) < C(q_{m'}, \mathbf{w}_{m'}, \upsilon_{m'})$, thus, $\upsilon_m < \upsilon_{m'}$, and given $q_m = q_{m'}$, $MR_m = MC(q_m, \mathbf{w}_m, \upsilon_m) < MC(q_{m'}, \mathbf{w}_{m'}, \upsilon_{m'}] = MR_{m'}$, and Assumption 13 b holds. If, on the other hand, $E\left[C_m^d | q_m, \mathbf{w}_m, p_m, s_m, \mathbf{x}_m\right] \geq E\left[C_{m'}^d | q_{m'}, \mathbf{w}_{m'}, p_{m'}, s_{m'}, \mathbf{x}_{m'}\right]$, then $MR_m = MC(q_m, \mathbf{w}_m, \upsilon_m) \geq MC(q_{m'}, \mathbf{w}_{m'}, \upsilon_{m'}) = MR_{m'}$ and Assumption 13 b does not hold. Therefore, by testing the hypothesis $E\left[C_m^d | q_m, \mathbf{w}_m, p_m, s_m, \mathbf{x}_m\right] < E\left[C_{m'}^d | q_{m'}, \mathbf{w}_{m'}, p_{m'}, s_{m'}, \mathbf{x}_{m'}\right]$ given $q_m = q_{m'}$, one can test Assumption 13 b.

Furthermore, even if Assumption 13 does not hold, if it is reasonable to assume that marginal cost does not vary with output locally around the point P_m, then one can still nonparametrically identify marginal cost, and thus marginal revenue by using equation (21) in part c of Proposition 5.

3.3.1 Oligopoly

We next consider oligopoly models with J_m firms in market m. We apply the same argument wlog to firm $j = 1$ in two different markets m and m' with variables $P_{1m} \equiv \{Q_m, q_{1m}, \mathbf{w}_m, p_{1m}, s_{1m}, \mathbf{s}_{-1m}, \mathbf{p}_{-1m}, \mathbf{X}_m\}$ and $P_{1m'} \equiv \{Q_{m'}, q_{1m'}, \mathbf{w}_{m'}, p_{1m'}, s_{1m'}, \mathbf{s}_{-1m'}, \mathbf{p}_{-1m'}, \mathbf{X}_{m'}\}$, where $\mathbf{s}_{-1m}$ and $\mathbf{p}_{-1m}$ are vectors of market shares and prices of firms other than firm 1 in market m and likewise for $\mathbf{s}_{-1m'}$ and $\mathbf{p}_{-1m'}$. As in Proposition 5, we need to find two close points in the data that have the same product characteristics ($\mathbf{X}_m = \mathbf{X}_{m'} = \mathbf{X}$) and input price vector ($\mathbf{w}_m = \mathbf{w}_{m'} = \mathbf{w}$). In addition, the two markets must satisfy the following properties

$$Q_m < Q_{m'}, \quad s_{1m} > s_{1m'}, \quad p_{1m} < p_{1m'}, \quad s_{1m}Q_m < s_{1m'}Q_{m'} \text{ and } \mathbf{p}_{-1m} = \mathbf{p}_{-1m'},$$

as well as

$$p_{1m}\left[1 + \frac{lnp_{1m'} - lnp_{1m}}{lns_{1m'} - lns_{1m}}\right] = \frac{E\left[C^d | q_{1m'}, \mathbf{w}_{m'}, \mathbf{p}_{m'}, \mathbf{s}_{m'}, \mathbf{X}_{m'}\right] - E\left[C^d | q_{1m}, \mathbf{w}_m, \mathbf{p}_m, \mathbf{s}_m, \mathbf{X}_m\right]}{q_{1m'} - q_{1m}} + O\left(|Q_{m'} - Q_m|\right).$$

Then, with only slight modifications to the proof of Proposition 5 for the monopoly case, we can prove nonparametric identification of the marginal revenue function for the oligopoly case. The relevant Proposition 6 and its proof are in Subsection C.2 of the Appendix.

3.3.2 Recovering the market share function

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

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

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

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

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

We can use this nonparametric estimate of the market share derivative to recover a nonparametric estimate of the demand function. Starting from the point $\bar{\mathbf{p}}, \bar{\mathbf{s}}, \bar{\mathbf{X}}$ (where $\bar{\mathbf{s}} = \mathbf{s}\left(\bar{\mathbf{p}}, \bar{\mathbf{X}}, \bar{\boldsymbol{\xi}}, 1\right)$ for some $\bar{\boldsymbol{\xi}}$), we derive the approximation of $\mathbf{s}\left(\bar{\mathbf{p}} + \Delta\mathbf{p}, \bar{\mathbf{X}}, \bar{\boldsymbol{\xi}}, 1\right)$, 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

$$\hat{s}_1 \equiv \hat{s}(\bar{\mathbf{p}} + \Delta\mathbf{p}, \bar{\mathbf{X}}, \bar{\boldsymbol{\xi}}, 1) = \bar{s} + \frac{\partial s(\widehat{\bar{\mathbf{p}}, \bar{\mathbf{X}}, \bar{\boldsymbol{\xi}}}, 1)}{\partial \mathbf{p}}' \Delta\mathbf{p},$$

where $\frac{\partial s(\widehat{\bar{\mathbf{p}}, \bar{\mathbf{X}}, \bar{\boldsymbol{\xi}}}, 1)}{\partial \mathbf{p}} = \left[\frac{\partial s(\widehat{\bar{p}, \bar{X}, \bar{\xi}}, 1)}{\partial p_1}, 0, \ldots, 0\right]'$. 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

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

Then,

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

Therefore, with some additional assumptions on the regularity of the marginal revenue function, one can show that

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

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

3.3.3 Curse of Dimensionality

In practice, a nonparametric estimator for the demand and cost parameters based on Propositions 5 and 6 will likely suffer from a Curse of Dimensionality. To implement such an estimator, one would need to obtain a nonparametric estimate of $E\left[C^d|q_{jm},\mathbf{w}_m,\mathbf{p}_m,\mathbf{s}_m,\mathbf{X}_m\right]$. 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 estimation we pursue the common practice where researchers use parametric restrictions to reduce the dimensionality of the estimation problem, essentially transforming the nonparametric estimation exercise into a semi-parametric one. In particular, we adopt the Berry (1994) logit or Berry et al. (1995) random coefficients demand model. This relaxes the need to condition on the individual variables $\mathbf{p}_m,\mathbf{s}_m,\mathbf{X}_m$ in developing our estimator; we only need to control for the marginal revenue MR_{jm}, which is a parametric function of these variables.

4 Estimation

An estimation strategy that reflects the parametric identification results the closest is to construct a pairwise differenced estimator that pairs up firms with similar outputs, input prices, and expected cost conditional on the vector of prices and market shares in the market. That is,

$$\begin{aligned}\boldsymbol{\theta}_{pJM} &= argmin_{\theta_p \in \Theta_p} \sum_{j,m} \sum_{j',m':(j',m')\neq(j,m)} \Big(MR_{jm}\left(\boldsymbol{\theta}_{-p}, \boldsymbol{\theta}_p\right) - MR_{j'm'}\left(\boldsymbol{\theta}_{-p}, \boldsymbol{\theta}_p\right) \Big)^2 \\ &\quad W_{\mathbf{h}} \Big(q_{jm} - q_{j'm'}, \mathbf{w}_{jm} - \mathbf{w}_{j'm'}, \widehat{E}\left[C^d | q_{jm}, \mathbf{w}_{jm}, \mathbf{p}_m, \mathbf{s}_m, \mathbf{X}_m, j \right] - \\ &\quad \widehat{E}\left[C^d | q_{j'm'}, \mathbf{w}_{j'm'}, \mathbf{p}_{m'}, \mathbf{s}_{m'}, \mathbf{X}_{m'}, j' \right], \mathbf{X}_{jm} - \mathbf{X}_{j'm'} \Big),\end{aligned}$$

where $W_{\mathbf{h}}$ is the kernel based weight function with the vector of bandwidth being $\mathbf{h}$, and $\widehat{E}\left[C^d | q, \mathbf{w}, \mathbf{p}, \mathbf{s}, \mathbf{X}, j\right]$ is the sample average of cost data conditional on $\{q, \mathbf{w}, \mathbf{p}, \mathbf{s}, \mathbf{X}, j\}$. The advantage of the above estimator is that it recovers demand parameters even if the firm is not profit maximizing, and thus, is potentially useful in industries where firms are under government regulation.

However, the need for $\widehat{E}\left[C^d | q, \mathbf{w}, \mathbf{p}, \mathbf{s}, \mathbf{X}, j\right]$ makes the Curse of Dimensionality of the estimator just as severe as the one for nonparametric identification discussed above. Thus the estimator may be impractical in most oligopoly markets where number of firms is sufficiently high. Therefore, from now on, we focus on developing an estimator that works well in such situations. To do so, we need to find a way to condition on marginal revenue, which is a parametric function of the observables, rather than the conditional expected cost.

As a first step towards constructing such an estimator, we define the MC-pseudo-cost function and the pseudo-cost function.

Definition 1 *An MC-pseudo-cost function is defined to be $\widetilde{PC}(q, \mathbf{w}, MC)$, where MC is the marginal cost. Similarly, a pseudo-cost function is defined to be $PC\left(q, \mathbf{w}, MR\right)$ where MR is the marginal revenue of the firm.*

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 under the assumption that marginal cost is an increasing function of the cost shock. Then, when we also assume that marginal cost is a function of mariginal revenue, one can use marginal revenue, if observable, as a proxy for the cost shock in order to relate the cost function to the pseudo-cost function. Now recall that marginal revenue can be expressed solely as a function of demand parameters and data, and thus is observable given the demand parameters. The lemma below formalizes this idea.

Lemma 1 *Suppose that Assumptions 2, 3, 5 and 6 are satisfied. Further, assume that marginal cost is strictly increasing and continuous in υ and marginal revenue is a strictly increasing and continuous function of marginal cost. Consider a firm $\{q, \mathbf{w}, \mathbf{p}, \mathbf{s}, \mathbf{X}, j\}$. Then, there exists a pseudo-cost function that satisfies $C(q, \mathbf{w}, \upsilon) = PC(q, \mathbf{w}, MR(\mathbf{p}, \mathbf{s}, \mathbf{X}, j, \boldsymbol{\theta}_0))$ and is increasing and continuous in marginal revenue.*

Proof. First, we show that there exists an MC-pseudo-cost function such that $C(q, \mathbf{w}, \upsilon) = \widetilde{PC}(q, \mathbf{w}, MC)$, where $\widetilde{PC}$ is a strictly increasing and continuous function of MC. Because the marginal cost function is strictly increasing and continuous in υ given q and $\mathbf{w}$, there exists an inverse function on the domain of $MC(q, \mathbf{w}, \upsilon)$ such that $\upsilon = \upsilon(q, \mathbf{w}, MC)$, where υ is an increasing and continuous function of MC. This implies that we can use (an unspecified function of) q, $\mathbf{w}$ and MC: $\upsilon(q, \mathbf{w}, MC)$, to control for υ. Substituting this "control function" for υ into the cost function, we obtain the MC-pseudo-cost function: $C(q, \mathbf{w}, \upsilon) = \widetilde{PC}(q, \mathbf{w}, MC)$. Because $MR(\mathbf{p}, \mathbf{s}, \mathbf{X}, j, \boldsymbol{\theta}_0) = \zeta(MC)$ and $\zeta()$ is assumed to be strictly increasing and continuous, the inverse function $\zeta^{-1}()$ is well defined, strictly increasing and continuous as well. Hence, $\widetilde{PC}$ can also be expressed as a strictly increasing and continuous function of MR.

$$\begin{aligned} C(q, \mathbf{w}, \upsilon) &= \widetilde{PC}(q, \mathbf{w}, MC) = \widetilde{PC}\left(q, \mathbf{w}, \zeta^{-1}(MR(\mathbf{p}, \mathbf{s}, \mathbf{X}, j, \boldsymbol{\theta}_0))\right) \\ &= PC(q, \mathbf{w}, MR(\mathbf{p}, \mathbf{s}, \mathbf{X}, j, \boldsymbol{\theta}_0)) \end{aligned}$$

■

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.

Using the above lemma, one could construct a new estimator that would pair up firms with similar outputs, input prices and marginal revenues given parameter $\boldsymbol{\theta}$. Specifically, the price parameters could be estimated as follows.

$$\begin{aligned}\boldsymbol{\theta}_{pJM} &= argmin_{\theta\in\Theta}\sum_{j,m}\sum_{j',m':(j',m')\neq(j,m)}\left(C^d_{jm}-C^d_{j'm'}\right)^2 \\ &\quad W_{\mathbf{h}}\left(q_{jm}-q_{j'm'},\mathbf{w}_{jm}-\mathbf{w}_{j'm'},MR_{jm}\left(\boldsymbol{\theta}_{-p},\boldsymbol{\theta}_p\right)-MR_{j'm'}\left(\boldsymbol{\theta}_{-p},\boldsymbol{\theta}_p\right)\right),\end{aligned}$$

where $MR_{jm}(\boldsymbol{\theta})$ is the shorthand of $MR(\mathbf{p}_m,\mathbf{s}_m,\mathbf{X}_m,j,\boldsymbol{\theta})$.The estimator is based on the argument that two firms jm and $j'm'$ that have similar ouput, input prices and marginal revenue should have similar pseudo-cost, and thus their costs in the data should be close except for the residual variation that is orthogonal to the output, input prices and the marginal revenue. This method avoids the Curse of Dimensionality because marginal revenue is a parametric function derived from the demand model.

The above pairwise difference based estimator is subject to some loss of efficiency because it does not impose the constraint that marginal cost is a funcion of marginal revenue exactly; it does so only approximately through the kernel weight function $W_{\mathbf{h}}$. Pairs of firms that violate the constraint are given low weight in the estimator, depending on the magnitude of violation, but not eliminated.

Next, we consider an estimator that imposes the restriction exactly. It selects demand parameters to fit the pseudo-cost function to the cost data using a nonparametric sieve regression (Chen (2007); Bierens (2014)).

4.1 Non-Linear Least Squares Estimator (NLLS)

We start with the following assumption.

Assumption 16 *The true pseudo-cost function can be expressed as a linear function of an infinite sequence of polynomials.*

$$PC\left(q_{jm},\mathbf{w}_{jm},MR_{jm}\left(\boldsymbol{\theta}_0\right)\right)=\sum_{l=1}^{\infty}\gamma_l\psi_l\left(q_{jm},\mathbf{w}_{jm},MR_{jm}\left(\boldsymbol{\theta}_0\right)\right), \tag{22}$$

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 $\sum_{l=1}^{\infty}\gamma_l^2<\infty$.[18]

[18]Suppose the vector $(q_{jm},\mathbf{w}_{jm},MR_{jm})$ comes from a compact finite dimensional Euclidean space, $\mathcal{W}$.

Our estimator is derived from the approximation of (22). 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 the true value of the parameter vector $\boldsymbol{\theta}_0$ and $\boldsymbol{\gamma}_0 = [\gamma_{10}, \ldots]'$, gives the smallest distance between the cost data and the sieve-approximated pseudo-cost function. That is

$$E\left[C_{jm}^d - \sum_l \gamma_l \psi_l\left(q_{jm}, \mathbf{w}_{jm}, MR_{jm}\left(\boldsymbol{\theta}_0\right)\right)\right]^2 \leq E\left[C_{jm}^d - \sum_l \gamma_l \psi_l\left(q_{jm}, \mathbf{w}_{jm}, MR_{jm}\left(\boldsymbol{\theta}\right)\right)\right]^2, \tag{23}$$

for any $\boldsymbol{\theta} \in \Theta$, where Θ is the demand parameter space, and $\Gamma = lim_{M\to\infty}\Gamma_{L_M}(T)$. As we have discussed above, if the demand function is logit or BLP, then the price parameters are identified and can be estimated as

$$[\boldsymbol{\theta}_{p0}, \boldsymbol{\gamma}_0] = \underset{(\boldsymbol{\theta}_p, \gamma)\in\Theta_p\times\Gamma}{\arg\min}\, E\left[C_{jm}^d - \sum_l \gamma_l \psi_l\left(q_{jm}, \mathbf{w}_{jm}, MR_{jm}\left(\boldsymbol{\theta}_{-p}, \boldsymbol{\theta}_p\right)\right)\right]^2, \tag{24}$$

and the sample analog of equation (24), given a sample of M markets is:

$$\left[\hat{\boldsymbol{\theta}}_{pM}, \hat{\boldsymbol{\gamma}}_M\right] = \underset{(\boldsymbol{\theta}_p, \gamma)\in\Theta_p\times\Gamma_{L_M}(T)}{\arg\min} \frac{1}{\sum_m J_m} \sum_{j,m} \left[C_{jm}^d - \sum_l \gamma_l \psi_l\left(q_{jm}, \mathbf{w}_{jm}, MR_{jm}\left(\boldsymbol{\theta}_{-p}, \boldsymbol{\theta}_p\right)\right)\right]^2. \tag{25}$$

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.[19]

Then, if $PC\left(q_{jm}, \mathbf{w}_{jm}, MR_{jm}\right)$ 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.

[19]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_{jm}^d and the RHS variables are $\psi_l\left(q_{jm}, \mathbf{w}_{jm}, MR_{jm}(\boldsymbol{\theta})\right)$, $l = 1, ..., L_M$. Then, construct the objective function, which is the average of squared residuals $Q_M(\boldsymbol{\theta}) = \frac{1}{\sum_m J_m} \sum_{j,m} \left[C_{jm}^d - \sum_{l=1}^{L_M} \hat{\gamma}_l \psi_l\left(q_{jm}, \mathbf{w}_{jm}, MR_{jm}(\boldsymbol{\theta})\right)\right]^2$.

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

Our sieve NLLS approach deals with issues of endogeneity by adopting a control function approach for the unobserved cost shock υ_{jm}. Our work borrows the idea from Ghandi et al. (2015) who investigate the control function approach in nonlinear BLP models where endogenous variables interact with error term, and standard instrumental variables strategy is insufficient. With our estimator, the right hand side of (25) is minimized only when the price parameters are at their true value $\boldsymbol{\theta}_{p0}$ so that the computed marginal revenue equals the true marginal revenue, and thus works as a control function for the supply shock υ_{jm}. If $\boldsymbol{\theta}_p \neq \boldsymbol{\theta}_{p0}$, then using the false marginal revenue adds noise, which increases the right hand size of the sum of squared residuals in (25). Thus, the true demand parameter $\boldsymbol{\theta}_{p0}$ can be obtained as a by-product of this control function approach.

The proposition that the estimator identifies the price parameter and its proof are provided in the Appendix.

4.2 Estimating taste parameters for product characteristics

Note that our estimator in equation (25) abstracts from product characteristics $\mathbf{x}$ in the demand model. When $\mathbf{x}$ is included, as we discussed in the identification section and as we will see in the Monte-Carlo results later, what is identified are α and $\boldsymbol{\xi} + \mathbf{x}'\boldsymbol{\beta}$ for the logit model and μ_α, σ_α, $\boldsymbol{\sigma_\beta}$ and $\boldsymbol{\xi} + \mathbf{x}'\boldsymbol{\mu_\beta}$ for the BLP model. In order to further identify $\boldsymbol{\beta}$ for the logit model and $\boldsymbol{\mu_\beta}$ for BLP, we include additional moment conditions in our estimator that leverage the (common) assumption that $E\left[\boldsymbol{\xi}_{jm}|\mathbf{X}_m\right] = 0$.[20] The modified

set of demand parameters.

[20] These orthogonality conditions also help in identifying the price coefficient in the logit or BLP models. That is, product characteristics of rival firms in the same market can be used as instruments, because of the restriction that they affect own price but do not enter directly in the deterministic component of the utility function of own product. Generally in a discrete choice model, identification likely requires some additional restrictions. For example, in a general discrete choice model with linear utility and random demand shock, individual i chooses product j, if

$$\alpha p_j + \mathbf{x}_j\boldsymbol{\beta} + \xi_j + \epsilon_{ij} \geq \alpha p_k + \mathbf{x}_k\boldsymbol{\beta} + \xi_k + \epsilon_{ik} \ \forall k.$$

That is,

$$I_j = 1 \ if \ \alpha\left(p_j - p_k\right) + \left(\mathbf{x}_j - \mathbf{x}_k\right)\boldsymbol{\beta} + \left(\xi_j - \xi_k\right) + \left(\epsilon_{ij} - \epsilon_{ik}\right) \geq 0 \ \forall k.$$

Therefore, characteristics of other products are included in the RHS of the choice equation and thus, they cannot be used as instruments for own demand, unless additional functional form restrictions are imposed. One example is to assume that the utility shock has a specific functional form such as logit, where market share equation can be expressed as

$$log\left(s_{jm}\right) - log\left(s_{0m}\right) = \alpha p_j + \mathbf{x}_j\boldsymbol{\beta} + \xi_j.$$

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,

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

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

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.

4.3 Cost function estimation

After estimating the marginal revenue function, we can recover the cost function. For that, we need to impose Assumption 15, i.e., there is no systematic misreporting of the cost. The steps in doing so are similar to those used in recovering of the market share function from marginal revenue.

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

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

where $\boldsymbol{\theta}_M$ is the estimated demand parameter, $\widehat{PC}(\cdot)$ is the estimated pseudo-cost function, and $W_{\mathbf{h}}$ is a kernel-based weight function.[21] Second, for a given input price $\mathbf{w}$,

The additional source of identification for the product characteristics instruments is the restriction that $\mathbf{p}_{0m} = 0$, $\mathbf{x}_{0m} = 0$, $\xi_{0m} = 0$. For example, the outside option of not buying an automobile implicitly does not include paying for public transportation.

[21]

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

starting at output $\bar{q}$ and total cost $\bar{C}$, there exists a cost shock $\overline{\upsilon}$ that corresponds to $MC(\bar{q}, \mathbf{w}, \bar{\upsilon}) = \overline{MR}$.[22] 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 $\bar{\upsilon}$,

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

where Δq represents a small change in quantity.[23] 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.

4.4 Further specification and data 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 (25) our estimator can be adapted to various empirical settings.

4.4.1 Economic versus accounting cost

The cost data we envision using comes from accounting statements of firms.[24] Such data do not necessarily reflect the economic cost that the firm considers in making input and output choices. More concretely, 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 production. For example, we may find details on firms' financial investments including their rate of return.[25] Suppose that return on a unit of financial investment is r_{jm}. Then, the

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

[23] In the Appendix, we provide detailed instructions on how to implement this iterative procedure.

[24] 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.

[25] 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 (2015).

opportunity cost of production is r_{jm} and the profit maximizing firm will produce and sell output until marginal revenue equals marginal cost that accounts for this cost, i.e.,

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

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

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

This NLLS is also consistent with the following specification, where firms do not equate marginal cost to marginal revenue: $MR_{jm}(\boldsymbol{\theta}) - r_{jm} = \varphi(MC(q_{jm}, \mathbf{w}_{jm}, \upsilon_{jm}))$.

Thus, 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.

4.4.2 Endogenous Product Characteristics

So far, we have followed the literature and assumed $\mathbf{x}_{jm}$ to be exogenous, i.e., orthogonal to $\boldsymbol{\xi}_{jm}$. However, if firms strategically choose prices *and* product characteristics, then elements of $\mathbf{x}_{jm}$ will be correlated with the demand shock $\boldsymbol{\xi}_{jm}$. 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.[26] To estimate demand parameters in this setting, one needs instruments to deal with the endogeneity of prices and product characteristics. In addition, the typi-

[26]See, for example, Chu (2010), Fan (2013), and Byrne (2015). 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, whose solution is computationally extremely burdensome.

cal set of “BLP instruments” for prices based on product characteristics become invalid. That is, researchers need more instruments and have fewer options for IVs.

In our framework, the cost function needs to be modified so that it also includes endogenous product characteristics $\mathbf{x}_e$. That is,

$$C\left(q, \mathbf{x}_e, \mathbf{w}, \upsilon, \boldsymbol{\upsilon}_e\right),$$

where the additional cost shocks $\boldsymbol{\upsilon}_e$ correspond to the shocks that affect the production of observed characteristics. The additinal F.O.C. for optimal product characteristics choice would then be

$$MR_{\mathbf{x}_e,jm}\left(\boldsymbol{\theta}_0\right) = MC_{\mathbf{X}_{e,jm}}\left(q_{jm}, \mathbf{x}_{e,jm}, \mathbf{w}_{jm}, \upsilon_{jm}, \boldsymbol{\upsilon}_{e,jm}\right),$$

where $MR_{\mathbf{x}_e,jm}$ is the vector of marginal revenue of firm j in market m with respect to the product characteristics choice. Then, the pseudo-cost function can be modified as follows,

$$PC\left(q_{jm}, \mathbf{x}_{ejm}, \mathbf{w}_{jm}, MR_{q,jm}\left(\boldsymbol{\theta}_0\right), MR_{\mathbf{x}_e,jm}\left(\boldsymbol{\theta}_0\right)\right),$$

where $MR_{q,jm}$ is the marginal revenue with respect to quantity choice. Then, the modified NLLS part would be

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

4.4.3 Cost Function Restrictions

So far, for the purpose of estimation, we have 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 restricted to have properties such as positive slopes, homogeneity of degree one in input prices, nor convexity in output or cost shock.

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 an input price w_1,

$$C(q, \mathbf{w}, \upsilon) = w_1 C(q, \frac{\mathbf{w}}{w_1}, \upsilon),$$

and

$$MC(q, \mathbf{w}, \upsilon) = w_1 \frac{\partial C(q, \mathbf{w}/w_1, \upsilon)}{\partial q}.$$

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

$$\frac{1}{\sum_m J_m} \sum_{j,m} \left[\frac{C^d_{jm}}{w_{1,jm}} - \sum_l \gamma_l \psi_l \left(q_{jm}, \frac{\mathbf{w}_{-1,jm}}{w_{1,jm}}, \frac{MR_{jm}(\boldsymbol{\theta})}{w_{1,jm}} \right) \right]^2, \tag{26}$$

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

We do not, however, impose positivity of slopes or convexity in output in estimation. Rather, we follow numerous papers in the cost function estimation literature and check these characteristics after estimating the cost function.

4.4.4 Missing cost data and multi product firms

Until now we have assumed cost data are available for all firms in the sample. However, it could very well be the case that we observe costs only 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 or orthogonality conditions, choosing only 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 marginal revenue and the GMM part of the objective function. Luckily, such demand-side data tends to be available to researchers for many industries. Allowing for missing cost data is important in alleviating the concern on the reliability of cost data. That is, in practice, researchers can go over the accounting cost data carefully and simply remove the problematic cost data.

A more difficult case of unobservable costs is when firms produce multiple products, but only the total cost across all products is observable in the data.[27] Suppose that each firm produces F outputs. Then, the cost function of the firm can be modified as

$$C\left(\mathbf{q}_{jm,1:F},\mathbf{w}_{jm},\boldsymbol{\upsilon}_{jm1:F}\right),$$

where $\mathbf{q}_{m1:F}=(q_{m1},...,q_{mF})$ is the vector of outputs of product 1 to product F, and the vector $\boldsymbol{\upsilon}$ is the F dimensional vector of cost shocks. Then, as long as the number of products is not too large (otherwise, we would face a Curse of Dimensionality issue in estimation), the NLLS component can be extended as follows,

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

Here, $\mathbf{X}_{m1:F}$, $\mathbf{P}_{m1:F}$ and $\mathbf{S}_{m1:F}$ are matrices of observed product characteristics, prices and market shares of all firms in the same market for all products they produce.

If the number of products is large, one should consider imposing more structure on the pseudo-cost function to avoid the Curse of Dimensionality.[28] Such a cost function could be specified as:

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

where MR_f is marginal revenue with respect to the operation of all products of the firm f and $I_{jm}(f)$ is an indicator function that equals 1 if product j in market m belongs to firm f and 0 otherwise; C_f^d is the total cost of the firm that includes the cost of all products, and η_f is the i.i.d. distributed measurement error of firm f's total cost. Denoting F to be the total number of firms in the data, the NLLS component of our estimator can be modified as follows,

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

[27] This is a 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.

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

5 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.

6 Monte Carlo Experiments

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

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

where we set the number of product characteristics K to be 1, and $\phi()$ to be 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 any of them ($j = 0$).

On the supply-side, we assume firms compete on prices a la differentiated products Bertrand competition 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 input price 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 cost shock υ, total cost and marginal cost functions are specified as,

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

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

Notice that in the above specification the cost function is homogenous of degree 1 in input prices.[29]

To create our Monte-Carlo samples, we generate the wage, rental rate, cost shock, market size Q_m, observable product characteristics $\mathbf{x}_{jm}$, and the idiosyncratic component of the demand shock ϱ_ξ, and as follows,

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

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

$$Q_m\sim i.i.d.U(Q_L,Q_H),\quad \mathbf{x}_{jm}\sim i.i.d.TN(\mu_{\mathbf{x}},\sigma_{\mathbf{x}}),$$

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

We assume market size to be uniformly distributed with lower bound Q_L and upper bound Q_H. We draw various random shocks from the truncated normal distribution $TN(\cdot)$, where we truncate both upper and lower 0.82 percentiles. For transforming the uniformly distributed market size shock to truncated normal distribution, we use a small positive $\delta=0.025$ for truncation. We truncate the distribution of shocks to ensure that the true cost function is positive and bounded, and the compactness of the set $\mathcal{W}$, which recall contains elements $(q_{jm},\mathbf{w}_m,MR_{jm})$, as is assumed in the asymptotics in the previous section. We let the cost shock υ_{jm} be positively correlated with the market size shock, i.e., $\delta_\upsilon>0$.

Importantly, we specify the unobserved quality so as to allow for correlation between

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

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

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	μ_r	σ_r
2.0	0.5	1.0	0.2	1.0	0.5	0.4	0.4	2.0	0.2	2.0	0.2
μ_υ	σ_υ	δ_υ		Q_L	Q_H	δ_0	σ_ξ	A	B		
0.5	0.2	0.2		5.0	10.0	4.0	0.5	0.01	1.0		

ξ_{jm} and input price, cost shock and market size. Specifically, we set:

$$\begin{aligned} \varrho_{\xi jm} &\sim i.i.d.TN(0,1), \\ \xi_{jm} &= \delta_0 + \delta_1 \varrho_{\xi jm} + \delta_2 \varrho_{wm} + \delta_3 \varrho_{rm} + \delta_4 \varrho_{\upsilon jm} + \delta_5 \Phi^{-1}\left(\delta + (1.0 - 2\delta)\frac{Q_m - Q_L}{Q_H - Q_L}\right). \end{aligned}$$

We set $\delta_l > 0$ for $l = 1, ..., 5$. Specifically, $\delta_1 = ... = \delta_5$, $\sigma_\xi = 0.5$. Hence, by construction, no variable can be used as a valid instrument for prices in demand estimation. Furthermore, since both demand and cost shocks are correlated with market size, one cannot use the variation of market size as an instrument for prices.

To solve for the equilibrium price, quantity, and market share for each oligopoly firm, we use golden section search on price.[30]

Table 1 summarizes the parameter setup of the Monte-Carlo experiments. Table 2 presents sample statistics from the 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 six percent of the total cost. The parameter estimates are obtained by the following minimization algorithm,

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

where we restrict the cost function to be homogenous of degree one in input price. Further,

[30]The algorithm for finding equilibria in oligopoly markets is available upon request.

Table 2: Sample Statistics of Simulated Data.

variables	Mean	Std. Dev
Price (p_m)	3.8239	0.8743
Output (q_m)	0.8931	0.6337
Quality (ξ_m)	3.9932	0.4385
Market Share (s_m)	0.1223	0.0858
Wage (w_m)	2.0011	0.1961
Rent (r_m)	1.9842	0.1781
Cost (C_m)	1.7088	0.7306
x_m	0.9791	0.4636

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

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 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 highlights the consistency of our estimator. It is noteworthy that 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.[31]

In Table 4, we report an 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, 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

[31] 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.

Market	Sample		$\hat{\mu}_\alpha$			$\hat{\sigma}_\alpha$			
Size	Size	No. Poly	Mean	Std. Dev	RMSE	Mean	Std. Dev	RMSE	
25	100	27	-2.0608	0.5754	0.5757	0.4604	0.1242	0.1298	
50	200	32	-2.0123	0.2925	0.2913	0.4944	0.0653	0.0652	
100	400	38	-1.9965	0.1854	0.1844	0.4993	0.0460	0.0457	
250	1000	48	-2.0128	0.1174	0.1175	0.4998	0.0283	0.0282	
True			-2.0000			0.5			
Market	Sample		$\hat{\mu}_\beta$			$\hat{\sigma}_\beta$			Obj. Fct.
Size	Size	No. Poly	Mean	Std. Dev	RMSE	Mean	Std. Dev	RMSE	
25	100	27	0.9900	0.1825	0.1819	0.1976	0.0796	0.0792	1.738D-3
50	200	32	0.9931	0.0893	0.0891	0.1955	0.0384	0.0385	2.013D-3
100	400	38	1.0059	0.0770	0.0769	0.2026	0.0170	0.0170	2.137D-3
250	1000	48	1.0037	0.0498	0.0497	0.2009	0.0076	0.0077	2.195D-3
True			1.0000			0.2			

Measurement error std. deviation: 0.1

be larger than those of the NLLS-GMM estimates from Table 3, except for those of $\hat{\mu}_\beta$ for the sample size of 1000. What we can see from these results is consistent with the identification results: the NLLS component of the estimator is sufficient for the estimation of $\mu_\alpha, \sigma_\alpha$, and σ_β. That is, to estimate price parameters, one only needs the cost data. The additional orthogonality conditions are only needed for the estimation of μ_β. However, the additional GMM component is effective in improving efficiency, in particular the efficiency of $\hat{\sigma}_\beta$, coefficients that determine the degree of heterogeneity and price elasticity in random coefficient models.

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 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 (NLS-GMM1), we show results of the NLLS-GMM estima-

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

Market	Sample		$\hat{\mu}_\alpha$			$\hat{\sigma}_\alpha$			
Size	Size	No. Poly	Mean	Std. Dev	RMSE	Mean	Std. Dev	RMSE	
25	100	27	-2.0088	0.8342	0.8347	0.4456	0.1735	0.1810	
50	200	32	-2.0200	0.3679	0.3667	0.5006	0.0952	0.0947	
100	400	38	-1.9951	0.2320	0.2309	0.5006	0.0502	0.0499	
250	1000	48	-1.9853	0.1319	0.1321	0.4960	0.0291	0.0293	
True			-2.0000			0.5			
Market	Sample		$\hat{\mu}_\beta$			$\hat{\sigma}_\beta$			Obj. Fct.
Size	Size	No. Poly	Mean	Std. Dev	RMSE	Mean	Std. Dev	RMSE	
25	100	27	1.0146	0.2950	0.2939	0.2165	0.1752	0.1751	1.552D-3
50	200	32	1.0017	0.1478	0.1471	0.2024	0.0744	0.0741	1.894D-3
100	400	38	1.0077	0.1004	0.1002	0.2026	0.0515	0.0513	2.068D-3
250	1000	48	0.9959	0.0487	0.0487	0.1989	0.0305	0.0303	2.152D-3
True			1.0000			0.2			

Measurement error std. deviation: 0.1

tor. We still obtain parameter estimates that are close to their true values. The results in the second row (IV1) 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. In the third row (IV2), 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 logit demand specification. We then reduced the degree of endogeneity to $\delta_2 = \delta_3 = \delta_4 = \delta_5 = 0.5\delta_1$ 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, i.e., 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.

In the fourth row (NLLS-GMM2), we report the results where we introduce correlation between demand and supply shocks across different markets. That is, we specify the

Table 5: IV Estimator for Logit Demand Parameters.

		$\hat{\alpha}$			$\hat{\beta}$		
Sample Size		Mean	Std. Dev	MSE	Mean	Std. Dev	MSE
1000	NLLS-GMM1[a]	-1.9993	0.1000	0.0995	0.9997	0.0375	0.0373
1000	IV1[b]	-2.0034	0.0549	0.0547	1.0036	0.0340	0.0334
1000	IV2[c]	-1.2388	0.0860	0.7660	0.7837	0.0397	0.2199
1000	NLLS-GMM2	-2.0168	0.1390	0.1393	0.9999	0.0373	0.0372
1000	IV3	-1.8255	0.2476	0.3019	0.9476	0.0789	0.0944
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:$\delta_2 = \delta_3 = \delta_4 = \delta_5 = 0.5\delta_1$

demand and supply shocks as

$$\begin{aligned}
\xi_{jm} &= 0.5\varrho_{\xi} + 0.5\zeta_m,\ \zeta_m = \rho\zeta_{m-1} + \sqrt{1-\rho^2}\varrho_{\zeta m}\ \textit{if}\ m > 1, \\
\zeta_1 &= \varrho_{\zeta 1}, \varrho_{\zeta k} \sim i.i.d.TN\left(0,1\right), k = 1, ..., M, \\
\upsilon_{jm} &= 0.5\varrho_{\upsilon} + 0.5\chi_m,\ \chi_m = \rho\chi_{m-1} + \sqrt{1-\rho^2}\varrho_{\chi m}\ \textit{if}\ m > 1, \\
\chi_1 &= \varrho_{\chi 1}, \varrho_{\chi k} \sim i.i.d.TN\left(0,1\right), k = 1, ..., M,
\end{aligned}$$

where ρ is set to be 0.5. The NLLS-GMM estimates are again close to the true values. In the fifth row (IV3), we present the IV estimates, where for firms in market m we use observed product characteristics of own and other firms, and theaverage price of market $m-1$ (for market 1, we use the average price of market M) as instruments. Results again indicate some bias in the IV estimates of parameters. In addition, variances of the estimated parameters are high.

7 Conclusion

We have developed a new methodology for estimating the demand and cost parameters of a BLP model. The method presumes that demand data on prices, market shares and product characteristics, and some information on firms' costs is available. With these data, in particular, including the cost data, we show that variation in market size and the exclusion restrictions implied by the model can be exploited to identify the model. These restrictions are: (1) price p and market shares s enter the marginal revenue function but do not directly enter the marginal cost function; and (2) total output q enters the marginal cost function but not the marginal revenue function.

Our approach identifies the demand parameters in the presence of price endogeneity, and a nonparametric cost function in the presence of output endogeneity, but does not use any conventional instruments. That is, it does not require demand and cost shocks to be uncorrelated within and across markets, with each other, nor with demand shifters, cost shifters or market size. Thus, our approach works in instances where standard IV-based estimation methods break down. Our Monte-Carlo experiments highlight this fact, as well as the numerical stability of our estimator. We have also shown that the marginal revenue function is nonparametrically identified. Moreover, our method can accommodate measurement error and misreporting with cost data, endogenous product characteristics, multi-product firms, difference between accounting and economic costs, and non-profit maximizing firms.

In which markets can our estimator be applied, and where would it be a potentially valuable addition to traditional IV-based estimation strategies? Market-level demand and cost data tend to be available for large industries that are subject to regulatory oversight (which often requires firms to report cost data); examples include banking and telecommunications.[32] Such major sectors of the economy represent natural settings for applying our estimator.

Our estimation strategy also presents an additional tool for policymakers who use BLP-type models. Prospective merger analysis is perhaps the most notable application where BLP models have had policy influence. This also represents a natural setting to apply our method since anti-trust authorities have the power to subpoena detailed cost data from firms for merger evaluation.[33] Fundamental to the predictions from merger simulations based on the standard IV approach (Nevo (2001)) is the estimated demand elasticity and inferred marginal costs from the supply-side first order conditions of the structural model. The demand elasticity and non-parametric cost estimates based on our

[32] Indeed, the parallel literatures on differentiated products markets and cost function estimation have relevant applications in these industries. See, for example, Ho and Ishii (2012) and Crawford and Yurukoglu (2012) for BLP applications to banking and cable television, and Wang (2003) and Kelly and Ying (2003) for corresponding cost function applications.

[33] For example, in the U.S. the Hart-Scott-Rodino Antitrust Improvements Act is the relevant law that gives authorities this power. Here, it is important to note our discussion from above of the fact that we can still obtain consistent estimates without needing cost data from all firms in an industry.

instrument-free approach can thus yield a complementary set of estimates and predictions regarding the welfare effects of prospective mergers.

More generally, by comparing results from our estimator and IV-based estimators, one could check the validity of various instruments. For example, such a comparison could be readily made in the context of cable TV industry where cost data is available, and where researchers have used Hausman-type instruments as price instruments (Crawford and Yurukoglu (2012)).

References

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*, forthcoming.

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. F. and P. C. Reiss (1990): "Entry in Monopoly Markets," *Review of Economic Studies*, 57, 531–553.

Byrne, D. P. (2015): "Testing Models of Differentiated Products Markets: Consolidation in the Cable TV Industry," *International Economic Review*, 56, 805–850.

Chen, X. (2007): "Large Sample Sieve Estimation of Semi-Nonparametric Models," in *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 Product Quality for Cable Television," *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.

FISHER, F. M. AND J. J. MCGOWAN (1983): "On the Misuse of Accounting Rates of Return to Infer Monopoly Power," *American Economic Review*, 73, 82–97.

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., K. IL KIM, AND A. PETRIN (2015): "Identification and Estimation in Discrete Choice Demand Models when Endogenous Variables Interact with the Error," *Unpublished Manuscript, Department of Economics, University of Wisconsin*.

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.

HO, K. AND J. ISHII (2012): "Location and Competition in Retail Banking," *International Journal of Industrial Organization*, 29, 537–546.

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

KELLY, M. T. AND J. S. YING (2003): "On Measuring Competitive Viability and Monopoly Power in Cable: An Empirical Cost Approach," *Review of Economics and Statistics*, 85, 963–970.

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.

KNITTEL, C. AND K. METAXOGLOU (2012): "Estimation of Random Coefficient Demand Models: Challenges, Difficulties, and Warnings," *Review of Economics and Statistics*, forthcoming.

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

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.

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.

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 3

We need to show that some firms in the population satisfying Assumption 12 have the same conditional expected cost. From Assumption 12, there exist two sets of variables $\boldsymbol{\nu}$ and $\boldsymbol{\nu}'$ and j such that, for all $\boldsymbol{\theta}_{-p} \in \boldsymbol{\Theta}_{-p}$,

$$MR\left(\mathbf{p}, \mathbf{s}, \mathbf{X}, j, \boldsymbol{\theta}_{-p}, \boldsymbol{\theta}_{p0}\right) = MR\left(\mathbf{p}', \mathbf{s}', \mathbf{X}', j, \boldsymbol{\theta}_{-p}, \boldsymbol{\theta}_{p0}\right),$$

Then, because of Assumptions 10 and 11, given q, $\mathbf{w}$, there exists υ such that

$$MR\left(\mathbf{p}, \mathbf{s}, \mathbf{X}, j, \boldsymbol{\theta}_{-p}, \boldsymbol{\theta}_{p0}\right) = MR\left(\mathbf{p}', \mathbf{s}', \mathbf{X}', j, \boldsymbol{\theta}_{-p}, \boldsymbol{\theta}_{p0}\right) = \zeta\left(MC\left(q, \mathbf{w}, \upsilon\right)\right). \tag{27}$$

where j is the row of the prices, market shraes and the observed characteristics of the firm in the vectors of prices, market shares and the matrix of observed characteristics. Hence, the supply and demand variables of the first firm are $\{q, \mathbf{w}, \mathbf{p}, \mathbf{s}, \mathbf{X}\}$, and the ones for the other firm are $\{q, \mathbf{w}, \mathbf{p}', \mathbf{s}', \mathbf{X}'\}$, and their cost shocks are the same: $\upsilon = \upsilon'$. It follows then that,

$$C\left(q, \mathbf{w}, \upsilon\right) = C\left(q, \mathbf{w}, \upsilon'\right).$$

Hence,

$$E\left[C^d \middle| \left(q, \mathbf{w}, \boldsymbol{\nu}\right)\right] = \varphi\left(C\left(q, \mathbf{w}, \upsilon'\right)\right) = \varphi\left(C\left(q, \mathbf{w}, \upsilon'\right)\right) = E\left[C^d \middle| \left(q, \mathbf{w}, \boldsymbol{\nu}'\right)\right].$$

Since, by Assumption 12, if $\boldsymbol{\theta}_p \neq \boldsymbol{\theta}_{p0}$,

$$MR\left(\mathbf{p}, \mathbf{s}, \mathbf{X}, j, \boldsymbol{\theta}_{-p}, \boldsymbol{\theta}_{p}\right) \neq MR\left(\mathbf{p}', \mathbf{s}', \mathbf{X}', j, \boldsymbol{\theta}_{-p}, \boldsymbol{\theta}_{p}\right) \tag{28}$$

we choose price parameters so that the two firms with the same output q, input price $\mathbf{w}$ and the conditional expected cost have the same marginal revenue.

B Parametric Identification of Marginal Revenue

B.1 Proof of Proposition 4: Logit Model

Proof. It is easy to show that the Berry (1994) logit demand model satisfies Assumption 12. For the parameter $\alpha \neq \alpha_0$, pick the two demand variables $\boldsymbol{\nu} = \{\mathbf{p}, \mathbf{s}, \mathbf{X}, j\}$ and $\boldsymbol{\nu}' = \{\mathbf{p}', \mathbf{s}', \mathbf{X}', j'\}$ with prices p_j, $p'_{j'}$ and market shares s_j, $s'_{j'}$ such that under α_0 their marginal revenues are equated, i.e.

$$p_j + \frac{1}{\left(1 - s_j\right)\alpha_0} = p'_{j'} + \frac{1}{\left(1 - s'_{j'}\right)\alpha_0} \Rightarrow \alpha_0 = -\frac{1}{p_j - p'_{j'}}\left[\frac{1}{1 - s'_j} - \frac{1}{1 - s'_{j'}}\right].$$

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

$$\alpha \neq -\frac{1}{p_j - p'_{j'}}\left[\frac{1}{1 - s_j} - \frac{1}{1 - s'_{j'}}\right],$$

thus,

$$p_j + \frac{1}{\left(1 - s_j\right)\alpha} \neq p'_{j'} + \frac{1}{\left(1 - s'_{j'}\right)\alpha}.$$

Therefore, the price coefficient satisfies Assumption 12. ■

B.2 Proof of Proposition 4: BLP Model

Proof. Next, we prove that the random coefficient BLP model also satisfies Assumption 12 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 $\alpha \sim N(\mu_{\alpha 0}, \sigma_{\alpha 0})$. We denote $\Phi()$ to be the distribution function of the standard normal distribution and $\phi()$ to be its density function. Consider the pair (s, p, ξ) and (s', p', ξ') that satisfy the share equation. Then,

$$\int_{\alpha} \frac{\exp(\xi + p\alpha)}{1 + \exp(\xi + p\alpha)} \frac{1}{\sigma_{\alpha 0}} \phi\left(\frac{\alpha - \mu_{\alpha 0}}{\sigma_{\alpha 0}}\right) d\alpha = \int_{\alpha} \frac{\exp(p(\alpha + \xi/p))}{1 + \exp(p(\alpha + \xi/p))} \frac{1}{\sigma_{\alpha 0}} \phi\left(\frac{\alpha - \mu_{\alpha 0}}{\sigma_{\alpha 0}}\right) d\alpha = s \tag{29}$$

and

$$\int_{\alpha} \frac{\exp(p'(\alpha + \xi'/p'))}{1 + \exp(p'(\alpha + \xi'/p'))} \frac{1}{\sigma_{\alpha 0}} \phi\left(\frac{\alpha - \mu_{\alpha 0}}{\sigma_{\alpha 0}}\right) d\alpha = s' \tag{30}$$

and we assume that they have the same marginal revenue:

$$\begin{aligned} MR(p, s, \mathbf{x}, \mu_\beta, \sigma_\beta, \mu_{\alpha 0}, \sigma_{\alpha 0}) &= p + p\left[\int_{\alpha} \frac{p\exp(p(\alpha + \xi/p))}{[1 + \exp(p(\alpha + \xi/p))]^2} \alpha \frac{1}{\sigma_{\alpha 0}} \phi\left(\frac{\alpha - \mu_{\alpha 0}}{\sigma_{\alpha 0}}\right) d\alpha\right]^{-1} s \\ = MR(p', s', \mathbf{x}, \mu_\beta, \sigma_\beta, \mu_{\alpha 0}, \sigma_{\alpha 0}) &= p' + p'\left[\int_{\alpha} \frac{p'\exp(p'(\alpha + \xi'/p'))}{[1 + \exp(p'(\alpha + \xi'/p'))]^2} \alpha \frac{1}{\sigma_{\alpha 0}} \phi\left(\frac{\alpha - \mu_{\alpha 0}}{\sigma_{\alpha 0}}\right) d\alpha\right]^{-1} s'. \end{aligned} \tag{31}$$

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

$$\int_{\alpha} \frac{\exp(p\sigma_{\alpha 0}(\tilde{\alpha} + a_0(p)))}{1 + \exp(p\sigma_{\alpha 0}(\tilde{\alpha} + a_0(p)))} \phi(\tilde{\alpha} - \eta_0)\, d\tilde{\alpha} = s, \quad \int_{\alpha} \frac{\exp(p'\sigma_{\alpha 0}(\tilde{\alpha} + a_0'(p)))}{1 + \exp(p'\sigma_{\alpha 0}(\tilde{\alpha} + a_0'(p)))} \phi(\tilde{\alpha} - \eta_0)\, d\tilde{\alpha} = s' \tag{32}$$

and the marginal revenue equation becomes

$$\begin{aligned} MR(p, s, \mathbf{x}, \mu_\beta, \sigma_\beta, \eta_0, \sigma_{\alpha 0}) &= p + p\left[\int_{\tilde{\alpha}} \frac{p\sigma_{\alpha 0}\exp(p\sigma_{\alpha 0}(\tilde{\alpha} + a_0(p)))}{[1 + \exp(p\sigma_{\alpha 0}(\tilde{\alpha} + a_0(p)))]^2} \tilde{\alpha}\phi(\tilde{\alpha} - \eta_0)\, d\tilde{\alpha}\right]^{-1} s \\ = MR(p', s', \mathbf{x}, \mu_\beta, \sigma_\beta, \eta_0, \sigma_{\alpha 0}) &= p' + p'\left[\int_{\tilde{\alpha}} \frac{p'\sigma_{\alpha 0}\exp(p'\sigma_{\alpha 0}(\tilde{\alpha} + a_0'(p)))}{[1 + \exp(p'\sigma_{\alpha 0}(\tilde{\alpha} + a_0'(p)))]^2} \tilde{\alpha}\phi(\tilde{\alpha} - \eta_0)\, d\tilde{\alpha}\right]^{-1} s'. \end{aligned} \tag{33}$$

Given (η, σ_α), we show that there exist (s, p) and (s', p') such that $(s, p) \neq (s', p')$ satisfying equations (32) and (33),

$$\int_{\tilde{\alpha}} \frac{\exp(p\sigma_\alpha(\tilde{\alpha} + a(p)))}{1 + \exp(p\sigma_\alpha(\tilde{\alpha} + a(p)))} \phi(\tilde{\alpha} - \eta)\, d\tilde{\alpha} = s, \quad \int_{\tilde{\alpha}} \frac{\exp(p'\sigma_\alpha(\tilde{\alpha} + a'(p)))}{1 + \exp(p'\sigma_\alpha(\tilde{\alpha} + a'(p)))} \phi(\tilde{\alpha} - \eta)\, d\tilde{\alpha} = s' \tag{34}$$

and

$$\begin{aligned} MR(p, s, \mathbf{x}, \mu_\beta, \sigma_\beta, \eta, \sigma_\alpha) = \quad & p + p\left[\int_{\tilde{\alpha}} \frac{p\sigma_\alpha\exp(p\sigma_\alpha(\tilde{\alpha} + a(p)))}{[1 + \exp(p\sigma_\alpha(\tilde{\alpha} + a(p)))]^2} \tilde{\alpha}\phi(\tilde{\alpha} - \eta)\, d\tilde{\alpha}\right]^{-1} s \\ \neq MR(p', s', \mathbf{x}, \mu_\beta, \sigma_\beta, \eta, \sigma_\alpha) = \quad & p' + p'\left[\int_{\tilde{\alpha}} \frac{p'\sigma_\alpha\exp(p'\sigma_\alpha(\tilde{\alpha} + a'(p)))}{[1 + \exp(p'\sigma_\alpha(\tilde{\alpha} + a'(p)))]^2} \tilde{\alpha}\phi(\tilde{\alpha} - \eta)\, d\tilde{\alpha}\right]^{-1} s'. \end{aligned} \tag{35}$$

Consider first the case $\eta_0 \neq \eta$. First, we abstract from observed heterogeneity and set $\mathbf{x} = 0$. We prove the following Lemma concerning the marginal revenue for large $p > 0$.

Lemma 2 *For any η, σ_α, for large $p>0$, $\mathbf{x}=0$,*

$$MR\left(p,s,\mathbf{x},\mu_\beta,\sigma_\beta,\eta,\sigma_\alpha\right)=p\left[1-\left[\left(\Phi^{-1}\left(s\right)-\eta\right)\phi\left(\Phi^{-1}\left(s\right)\right)+O\left(p^{-2}\right)\right]^{-1}s\right]. \tag{36}$$

Proof. Using integration by parts, we obtain

$$\int_{\tilde{\alpha}}\frac{\exp\left(p\sigma_\alpha\left(\tilde{\alpha}+a\left(p\right)\right)\right)}{\left[1+\exp\left(p\sigma_\alpha\left(\tilde{\alpha}+a\left(p\right)\right)\right)\right]}\phi\left(\tilde{\alpha}-\eta\right)d\tilde{\alpha}=1-\int_{\tilde{\alpha}}\frac{p\sigma_\alpha exp\left[p\sigma_\alpha\left(\tilde{\alpha}+a\left(p\right)\right)\right]}{\left[1+exp\left[p\sigma_\alpha\left(\tilde{\alpha}+a\left(p\right)\right)\right]\right]^2}\Phi\left(\tilde{\alpha}-\eta\right)d\tilde{\alpha} \tag{37}$$

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

$$\begin{aligned}(\ 37) \quad &= 1-\int_{\tilde{\alpha}}\frac{p\sigma_\alpha exp\left[p\sigma_\alpha\left(\tilde{\alpha}+a\left(p\right)\right)\right]}{\left[1+exp\left[p\sigma_\alpha\left(\tilde{\alpha}+a\left(p\right)\right)\right]\right]^2}\Big[\Phi\left(-a\left(p\right)-\eta\right)+\left(\tilde{\alpha}+a\left(p\right)\right)\phi\left(-a\left(p\right)-\eta\right)\\ &+\frac{1}{2}\left(\tilde{\alpha}+a\left(p\right)\right)^2\phi'\left(-a\left(p\right)-\eta\right)+\frac{1}{6}\left(\tilde{\alpha}+a\left(p\right)\right)^3\phi''\left(-a\left(p\right)-\eta\right)+\frac{1}{24}\left(\tilde{\alpha}+a\left(p\right)\right)^4\phi'''\left(a^*\left(\tilde{\alpha}\right)-\eta\right)\Big]\\ &d\tilde{\alpha}\end{aligned}$$

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

$$\int_{\tilde{\alpha}}\frac{p\sigma_\alpha exp\left[p\sigma_\alpha\left(\tilde{\alpha}+a\left(p\right)\right)\right]}{\left[1+exp\left[p\sigma_\alpha\left(\tilde{\alpha}+a\left(p\right)\right)\right]\right]^2}\left(\tilde{\alpha}+a_0\left(p\right)\right)d\tilde{\alpha}=\int_{\tilde{\alpha}}\frac{p\sigma_\alpha exp\left[p\sigma_\alpha\left(\tilde{\alpha}+a\left(p\right)\right)\right]}{\left[1+exp\left[p\sigma_\alpha\left(\tilde{\alpha}+a\left(p\right)\right)\right]\right]^2}\left(\tilde{\alpha}+a\left(p\right)\right)^3d\tilde{\alpha}=0.$$

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

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

and from the fourth central moment, we can derive that

$$\left|\frac{1}{24}\int_{\tilde{\alpha}}\frac{p\sigma_\alpha exp\left[p\sigma_\alpha\left(\tilde{\alpha}+a_0\left(p\right)\right)\right]}{\left[1+exp\left[p\sigma_\alpha\left(\tilde{\alpha}+a\left(p\right)\right)\right]\right]^2}\left(\tilde{\alpha}+a\left(p\right)\right)^4\phi'''\left(a^*\left(\tilde{\alpha}\right)-\eta\right)d\tilde{\alpha}\right|$$

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

where $C>0$ is a constant.

Therefore, from 34 and 37, we obtain

$$(\ 37)=1-\Phi\left(-a\left(p\right)-\eta\right)-\frac{\pi^2}{6\sigma_\alpha^2p^2}\phi'\left(-a\left(p\right)-\eta\right)+O\left(p^{-4}\right)=s.$$

and if we let $a=\Phi^{-1}\left(s\right)-\eta=lim_{p\rightarrow\infty}a\left(p\right)$, then

$$1-\Phi\left(-a\left(p\right)-\eta\right)-\frac{\pi^2}{6\sigma_\alpha^2p^2}\phi'\left(-a\left(p\right)-\eta\right)+O\left(p^{-4}\right)=1-\Phi\left(-a-\eta\right)=\Phi\left(a+\eta\right). \tag{38}$$

Therefore,

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

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

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

and

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

Therefore,

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

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

$$\int_{\tilde{\alpha}} \frac{p\sigma_\alpha \exp[p\sigma_\alpha(\tilde{\alpha}+a(p))]}{[1+\exp[p\sigma_\alpha(\tilde{\alpha}+a(p))]]^2}\tilde{\alpha}\phi(\tilde{\alpha}-\eta)\,d\tilde{\alpha} = \int_{\tilde{\alpha}} \frac{p\sigma_\alpha \exp[p\sigma_\alpha(\tilde{\alpha}+a(p))]}{[1+\exp[p\sigma_\alpha(\tilde{\alpha}+a(p))]]^2}(\tilde{\alpha}+a(p))\phi(\tilde{\alpha}-\eta)\,d\tilde{\alpha}$$
$$-a(p)\int_{\tilde{\alpha}} \frac{p\sigma_\alpha \exp[p\sigma_\alpha(\tilde{\alpha}+a(p))]}{[1+\exp[p\sigma_\alpha(\tilde{\alpha}+a(p))]]^2}\phi(\tilde{\alpha}-\eta)\,d\tilde{\alpha}$$
$$= \int_{\tilde{\alpha}} \frac{p\sigma_\alpha \exp[p\sigma_\alpha(\tilde{\alpha}+a(p))]}{[1+\exp[p\sigma_\alpha(\tilde{\alpha}+a(p))]]^2}\Big[(\tilde{\alpha}+a(p))\phi(-a(p)-\eta) + (\tilde{\alpha}+a(p))^2\phi'(-a(p)-\eta)$$
$$+\frac{1}{2}(\tilde{\alpha}+a(p))^3\phi''(-a(p)-\eta) + \frac{1}{6}(\tilde{\alpha}+a(p))^4\phi'''(-a^*(\tilde{\alpha})-\eta)\Big]\,d\tilde{\alpha}$$
$$-a_0(p)\int_{\tilde{\alpha}} \frac{p\sigma_\alpha \exp[p\sigma_\alpha(\tilde{\alpha}+a(p))]}{[1+\exp[p\sigma_{\alpha 0}(\tilde{\alpha}+a(p))]]^2}\Big[\phi(-a(p)-\eta) + (\tilde{\alpha}+a(p))\phi'(-a(p)-\eta)$$
$$+\frac{1}{2}(\tilde{\alpha}+a(p))^2\phi''(-a(p)-\eta) + \frac{1}{6}(\tilde{\alpha}+a(p))^3\phi'''(-a(p)-\eta) + \frac{1}{24}(\tilde{\alpha}+a(p))^4\phi''''(-a^*(\tilde{\alpha})-\eta)\Big]$$
$$d\tilde{\alpha}.$$

Therefore,

$$\int_{\tilde{\alpha}} \frac{p\sigma_\alpha \exp[p\sigma_\alpha(\tilde{\alpha}+a(p))]}{[1+\exp[p\sigma_\alpha(\tilde{\alpha}+a(p))]]^2}\tilde{\alpha}\phi(\tilde{\alpha}-\eta)\,d\tilde{\alpha}$$
$$= \quad -a(p)\phi(-a(p)-\eta) + \left[\phi'(-a(p)-\eta) - \frac{a(p)}{2}\phi''(-a(p)-\eta)\right]\frac{\pi^2}{3\sigma_\alpha^2 p^2} + O\left(p^{-4}\right)$$
$$= \quad -a\phi(-a-\eta) + \left[(\phi(-a-\eta) - a\phi'(-a-\eta))\frac{\phi'(-a-\eta)}{2\phi(-a-\eta)} + \phi'(-a-\eta) - \frac{a}{2}\phi''(-a-\eta)\right]\frac{\pi^2}{3\sigma_\alpha^2 p^2}$$
$$+O\left(p^{-4}\right) \tag{39}$$
$$= \quad -a\phi(-a-\eta) + O\left(p^{-2}\right).$$

Therefore, using 38 and symmetry of ϕ around zero, we obtain

$$MR = p\left[1-\left(a\phi(-a-\eta)+O\left(p^{-2}\right)\right)^{-1}s\right] = p\left[1-\left[\left(\Phi^{-1}(s)-\eta\right)\phi\left(\Phi^{-1}(s)\right)+O\left(p^{-2}\right)\right]^{-1}s\right]. \tag{40}$$

Next, we show that given the assumptions, given Q and Q', there exist (ξ,υ) and (ξ',υ') that generate such (p,s) and (p',s'). That is, set $\xi = a(p)(p\sigma_\alpha) \approx \left(\Phi^{-1}(s)-\eta\right)p\sigma_\alpha$, $\xi' = a(p')(p'\sigma_\alpha) \approx \left(\Phi^{-1}(s')-\eta\right)p'\sigma_\alpha$, and set υ such that

$$p + p\left[\int_{\tilde{\alpha}} \frac{p\sigma_\alpha \exp[p\sigma_\alpha(\tilde{\alpha}+a(p))]}{[1+\exp[p\sigma_\alpha(\tilde{\alpha}+a(p))]]^2}\tilde{\alpha}\phi(\tilde{\alpha}-\eta)\,d\tilde{\alpha}\right]^{-1}s = MC(Qs,\upsilon)$$

which exists given the assumptions. Similarly for (p',s').

■

Now, consider η_0. Because of the Assumption b, $\eta_0 < -1/\left(2\phi\left(0\right)\right)$. We pick $s = 1/2$ and $s' \neq s$ that satisfies

$$-\frac{1}{2\phi\left(0\right)\eta_0} < \frac{s'}{\phi\left(\Phi^{-1}\left(s'\right)\right)\left(\Phi^{-1}\left(s'\right)-\eta_0\right)} < 1. \tag{41}$$

We next show that such s' exists. Let $g\left(\right)$ be the function as follows.

$$g\left(t\right) = \frac{t}{\phi\left(\Phi^{-1}\left(t\right)\right)\left(\Phi^{-1}\left(t\right)-\eta_0\right)}.$$

Because Then, g is continuous and , and

$$g\left(\frac{1}{2}\right) = -\frac{1}{2\phi\left(0\right)\eta_0},\ lim_{t\downarrow\Phi(\eta_0)}g\left(t\right) = \infty$$

Therefore, from intermediate value theorem, there exists such s', and because $\eta_0 < -1/\left(2\phi\left(0\right)\right) < 0$, $s' \in \left(\Phi\left(\eta_0\right), 1/2\right)$, and thus, $s' < s$ can be satisfied. Furtheremore, one can choose the prices p, p' such that given large $\overline{P}$, p, $p' > \overline{P}$ and the relative price $P = p'/p$, they satisfy

$$\begin{aligned} MR\left(p, s, \mathbf{x}, \mu_\beta, \sigma_\beta, \eta_0, \sigma_{\alpha 0}\right) &= p\left[1-\frac{s}{\left(\Phi^{-1}\left(s\right)-\eta_0\right)\phi\left(\Phi^{-1}\left(s\right)\right)+O\left(p^{-2}\right)}\right] \\ = MR\left(p', s', \mathbf{x}, \mu_\beta, \sigma_\beta, \eta_0, \sigma_{\alpha 0}\right) &= p'\left[1-\frac{s'}{\left(\Phi^{-1}\left(s'\right)-\eta_0\right)\phi\left(\Phi^{-1}\left(s'\right)\right)+O\left(p'^{-2}\right)}\right]. \end{aligned} \tag{42}$$

Such two prices can be chosen because the equation can be rewritten as

$$P = \left[1-\frac{s'}{\left(\Phi^{-1}\left(s'\right)-\eta_0\right)\phi\left(\Phi^{-1}\left(s'\right)\right)+O\left(p'^{-2}\right)}\right]^{-1}\left[1-\frac{s}{\left(\Phi^{-1}\left(s\right)-\eta_0\right)\phi\left(\Phi^{-1}\left(s\right)\right)+O\left(p^{-2}\right)}\right]$$

and given p and p'being large, RHS is roughly constant. Equation (42) can be rewritten as,

$$p - \frac{ps\phi^{-1}\left(\Phi^{-1}\left(s\right)\right)}{\left(\Phi^{-1}\left(s\right)-\eta_0\right)+O\left(p^{-2}\right)} = p' - \frac{p's'\phi^{-1}\left(\Phi^{-1}\left(s'\right)\right)}{\left(\Phi^{-1}\left(s'\right)-\eta_0\right)+O\left(p'^{-2}\right)}. \tag{43}$$

Then, what we prove next is that the equation (43) is only satisfied by η_0, but not by another negative η.

Lemma 3 *Given $s = 1/2$, and s' satisfying equation (41), and large $\tilde{p}$ and $\tilde{p}'$ such that*

$$\tilde{p} - \frac{\tilde{p}s\phi^{-1}\left(\Phi^{-1}\left(s\right)\right)}{\left(\Phi^{-1}\left(s\right)-\eta\right)} = \tilde{p}' - \frac{\tilde{p}'s'\phi^{-1}\left(\Phi^{-1}\left(s'\right)\right)}{\left(\Phi^{-1}\left(s'\right)-\eta\right)} > 0, \tag{44}$$

equation (44) has two solutions for η: η_1 and η_2 and the only one of them is negative.

Proof. 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,

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

Since $s = 1/2$, $B = 0$,

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

$$\left[1 + \frac{C}{\eta}\right] = \tilde{P}\left[1 - \frac{C'}{B' - \eta}\right]. \tag{45}$$

Then, from the assumptions, it is easy to see that the slope of market share with respect to price C/η is negative and $1 + C/\eta$ is positive. Furthermore, from equation (41), $-C/\eta < C'/\left(B' - \eta\right) < 1$. Therefore, equation (45) implies $\tilde{P} > 1$. Because $s' < s = 1/2$, $B' < 0$, and thus, $CB'/\left(1 - \tilde{P}\right) > 0$. Then, if we denote

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

$\eta_1 > 0$, and $\eta_2 < 0$. Therefore, claim holds. ■

Hence, only η_2 satisfied the assumption. Furthermore $\eta = \eta_2$ satisfying equation (44) can be made arbitrarily close to η_0 satisfying equations (29), (30) and (31) by choosing p, p' sufficiently large. Therefore, for s, s', p, and p', there only exists one η that satisfies (29), (30) and (31). Hence, claim holds. Next, consider the case where $\eta_0 = \eta$, $\sigma_{\alpha 0} \neq \sigma_\alpha$. First, consider σ_α such that $\sigma_\alpha > \sigma_{\alpha 0}$. Suppose Assumption 12 is not satisfied. Then, consider $s \neq s'$, and p, p' in the data mentioned in Assumption 10. For those two data, the following holds: Given $a_0(p)$, $a_0'(p)$ satisfying

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

, then, if we denote $MR\left(p, s, \mathbf{x}, \mu_\beta, \sigma_\beta, \eta_0, \sigma_{\alpha 0}\right)$ to be the marginal revenue given demand variables $(p, s, \mathbf{x})$, $\mathbf{x} = 0$ and parameters $\left(\mu_\beta, \sigma_\beta, \eta_0, \sigma_{\alpha 0}\right)$,

$$\begin{aligned} & MR\left(p, s, \mathbf{x}, \mu_\beta, \sigma_\beta, \eta_0, \sigma_{\alpha 0}\right) = p + p\left[\int_{\tilde{\alpha}} \frac{p\sigma_{\alpha 0}\exp\left[p\sigma_{\alpha 0}\left(\tilde{\alpha} + a_0(p)\right)\right]}{\left[1 + \exp\left[p\sigma_{\alpha 0}\left(\tilde{\alpha} + a_0(p)\right)\right]\right]^2}\tilde{\alpha}\phi\left(\tilde{\alpha} - \eta_0\right)d\tilde{\alpha}\right]^{-1} s \\ = \; & MR'\left(p', s', \mathbf{x}, \mu_\beta, \sigma_\beta, \eta_0, \sigma_{\alpha 0}\right) = p' + p'\left[\int_{\tilde{\alpha}} \frac{p'\sigma_{\alpha 0}\exp\left[p'\sigma_{\alpha 0}\left(\tilde{\alpha} + a_0'(p)\right)\right]}{\left[1 + \exp\left[p'\sigma_{\alpha 0}\left(\tilde{\alpha} + a_0'(p)\right)\right]\right]^2}\tilde{\alpha}\phi\left(\tilde{\alpha} - \eta_0\right)d\tilde{\alpha}\right]^{-1} s' \end{aligned} \tag{46}$$

Now, since we consider the case where Assumption 12 is not satisfied., the same relationship holds for σ_α instead of $\sigma_{\alpha 0}$. Then, for μ_α, σ_α such that $\mu_\alpha/\sigma_\alpha = \eta_0$, given $a(p)$, $a'(p)$ satisfying

$$\begin{aligned}\int_{\tilde{\alpha}} \frac{\exp\left(p\sigma_{\alpha}\left(\tilde{\alpha} + a(p)\right)\right)}{\left[1 + \exp\left(p\sigma_{\alpha}\left(\tilde{\alpha} + a(p)\right)\right)\right]}\phi\left(\tilde{\alpha} - \eta_0\right)d\tilde{\alpha} &= s \\ \int_{\tilde{\alpha}} \frac{\exp\left(p'\sigma_{\alpha}\left(\tilde{\alpha} + a'(p)\right)\right)}{\left[1 + \exp\left(p'\sigma_{\alpha}\left(\tilde{\alpha} + a'(p)\right)\right)\right]}\phi\left(\tilde{\alpha} - \eta_0\right)d\tilde{\alpha} &= s',\end{aligned}$$

$$\begin{aligned} & MR\left(p, s, \mathbf{x}, \mu_\beta, \sigma_\beta, \eta_0, \sigma_{\alpha}\right) = p + p\left[\int_{\tilde{\alpha}} \frac{p\sigma_{\alpha}\exp\left[p\sigma_{\alpha}\left(\tilde{\alpha} + a(p)\right)\right]}{\left[1 + \exp\left[p\sigma_{\alpha}\left(\tilde{\alpha} + a(p)\right)\right]\right]^2}\tilde{\alpha}\phi\left(\tilde{\alpha} - \eta_0\right)d\tilde{\alpha}\right]^{-1} s \\ = \; & MR\left(p', s', \mathbf{x}, \mu_\beta, \sigma_\beta, \eta_0, \sigma_{\alpha}\right) = p' + p'\left[\int_{\alpha} \frac{p'\sigma_{\alpha}\exp\left[p'\sigma_{\alpha}\left(\tilde{\alpha} + a'\{p\}\right)\right]}{\left[1 + \exp\left[p'\sigma_{\alpha}\left(\tilde{\alpha} + a'(p)\right)\right]\right]^2}\tilde{\alpha}\phi\left(\tilde{\alpha} - \eta_0\right)d\tilde{\alpha}\right]^{-1} s' \end{aligned} \tag{47}$$

To show that (p, s) and p', s' satisfying equation (46) actually exist in the population given the assumption, consider p, p' to be sufficiently large so that

$$\begin{aligned} MR\left(p, s, \mathbf{x}, \mu_\beta, \sigma_\beta, \eta_0, \sigma_{\alpha 0}\right) &= p\left[1 - \frac{s}{\left(\Phi^{-1}\left(s\right) - \eta_0\right)\phi\left(\Phi^{-1}\left(s\right)\right) + O\left(p^{-2}\right)}\right] \\ MR\left(p', s', \mathbf{x}, \mu_\beta, \sigma_\beta, \eta_0, \sigma_{\alpha 0}\right) &= p'\left[1 - \frac{s'}{\left(\Phi^{-1}\left(s'\right) - \eta_0\right)\phi\left(\Phi^{-1}\left(s'\right)\right) + O\left(p'^{-2}\right)}\right]. \end{aligned}$$

Then, one can set the relative price $P = p'/p$ such that

$$P = \left[1 - \frac{s'}{\left(\Phi^{-1}\left(s'\right) - \eta_0\right)\phi\left(\Phi^{-1}\left(s'\right)\right) + O\left(p'^{-2}\right)}\right]^{-1}\left[1 - \frac{s}{\left(\Phi^{-1}\left(s\right) - \eta_0\right)\phi\left(\Phi^{-1}\left(s\right)\right) + O\left(p^{-2}\right)}\right]$$

is satisfied because the RHS is approximately a constant function of p, p' for large p, p'. Hence, equation (46) is satisfied. Next, we show that given the assumptions, given Q and Q', there exist (ξ, υ) and (ξ', υ') that generate such (p, s) and (p', s'). That is, set $\xi = a_0\left(p\right)\left(p\sigma_{\alpha 0}\right) \approx \left(\Phi^{-1}\left(s\right) - \eta_0\right)p\sigma_{\alpha 0}$, and set υ such that

$$p + p\left[\int_{\tilde{\alpha}} \frac{p\sigma_{\alpha 0}\exp\left[p\sigma_{\alpha 0}\left(\tilde{\alpha} + a_0\left(p\right)\right)\right]}{\left[1 + \exp\left[p\sigma_{\alpha 0}\left(\tilde{\alpha} + a_0\left(p\right)\right)\right]\right]^2}\tilde{\alpha}\phi\left(\tilde{\alpha} - \eta\right)d\tilde{\alpha}\right]^{-1} s = MC\left(Qs, \upsilon\right)$$

which exists given the assumptions. Similarly for (p', s').

Now, 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 $\left(p^{(0)}, p^{(0)\prime}\right) = (p, p')$, $\left(p^{(k)}, p^{(k)\prime}\right) = \left(\left(\sigma_\alpha/\sigma_{\alpha 0}\right)^k p, \left(\sigma_\alpha/\sigma_{\alpha 0}\right)^k p'\right)$. Then, one can show that the firms $\left(p^{(k)}, s\right)$, $\left(p^{(k)\prime}, s'\right)$ exist in the population. That is, $\xi^{(k)} = a\left(p^{(k)}\right)\left(p^{(k)}\sigma_{\alpha 0}\right) \approx \left(\Phi^{-1}\left(s\right) - \eta_0\right)p^{(k)}\sigma_{\alpha 0}$, $\xi^{(k)\prime} = a_0\left(p'\right)\left(p'\sigma_{\alpha 0}\right) \approx \left(\Phi^{-1}\left(s'\right) - \eta_0\right)p^{(k)\prime}\sigma_{\alpha 0}$, and there exists $\upsilon^{(k)}$, $\upsilon^{(k)\prime}$ such that marginal revenue equals marginal cost.

Now, $p^{(1)} = \left(\sigma_\alpha/\sigma_{\alpha 0}\right)p > p$, $p^{(1)\prime} = \left(\sigma_\alpha/\sigma_{\alpha 0}\right)p' > p'$, and furthermore, $p^{(1)}\sigma_{\alpha 0} = p\sigma_\alpha$, $p^{(1)\prime}\sigma_{\alpha 0} = p'\sigma_\alpha$. Therefore,

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

therefore, $a_0\left(p^{(1)}\right) = a\left(p\right)$, $a_0\left(p^{(1)\prime}\right) = a\left(p'\right)$ and

$$\begin{aligned} (47) \times \sigma_\alpha &= p^{(1)}\sigma_{\alpha 0} + p^{(1)}\sigma_{\alpha 0}\left[\int_{\tilde{\alpha}} \frac{p^{(1)}\sigma_{\alpha 0}\exp\left[p^{(1)}\sigma_{\alpha 0}\left(\tilde{\alpha} + a_0\left(p^{(1)}\right)\right)\right]}{\left[1 + \exp\left[p^{(1)}\sigma_{\alpha 0}\left(\tilde{\alpha} + a_0\left(p^{(1)}\right)\right)\right]\right]^2}\tilde{\alpha}\phi\left(\tilde{\alpha} - \eta_0\right)d\tilde{\alpha}\right]^{-1} s \\ &= p^{(1)\prime}\sigma_{\alpha 0} + p^{(1)\prime}\sigma_{\alpha 0}\left[\int \frac{p^{(1)\prime}\sigma_{\alpha 0}\exp\left[p^{(1)\prime}\sigma_{\alpha 0}\left(\tilde{\alpha} + a_0\left(p\right)\right)\right]}{\left[1 + \exp\left[p^{(1)\prime}\sigma_{\alpha 0}\left(\tilde{\alpha} + a_0\left(p\right)\right)\right]\right]^2}\tilde{\alpha}\phi\left(\tilde{\alpha} - \eta_0\right)d\tilde{\alpha}\right]^{-1} s'. \end{aligned}$$

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 $\left(p^{(0)},p^{(0)\prime}\right)=(p,p')$, $\left(p^{(k)},p^{(k)\prime}\right)=\left(\left(\sigma_\alpha/\sigma_{\alpha 0}\right)^k p,\left(\sigma_\alpha/\sigma_{\alpha 0}\right)^k p'\right)$, and for any integer $k\geq 1$ such that.

$$MR\left(p^{(k)},s,\mathbf{x},\mu_\beta,\sigma_\beta,\eta_0,\sigma_{\alpha 0}\right)=p^{(k)}+p^{(k)}\left[\int_{\tilde{\alpha}}\frac{p^{(k)}\sigma_{\alpha 0}\exp\left[p^{(k)}\sigma_{\alpha 0}\left(\tilde{\alpha}+a_0^{(k)}\right)\right]}{\left[1+\exp\left[p^{(k)}\sigma_{\alpha 0}\left(\tilde{\alpha}+a_0^{(k)}\right)\right]\right]^2}\tilde{\alpha}\phi\left(\tilde{\alpha}-\eta_0\right)d\tilde{\alpha}\right]^{-1}s$$

$$=\quad p^{(k)\prime}+p^{(k)\prime}\left[\int_{\tilde{\alpha}}\frac{p^{(k)\prime}\sigma_{\alpha 0}\exp\left[p^{(k)\prime}\sigma_{\alpha 0}\left(\tilde{\alpha}+a_0^{(k)\prime}\right)\right]}{\left[1+\exp\left[p^{(k)\prime}\sigma_{\alpha 0}\left(\tilde{\alpha}+a_0^{(k)\prime}\right)\right]\right]^2}\tilde{\alpha}\phi\left(\tilde{\alpha}-\eta_0\right)d\tilde{\alpha}\right]^{-1}s'=MR\left(p^{(k)\prime},s',\mathbf{x},\mu_\beta,\sigma_\beta,\mu_{\alpha 0},\sigma_{\alpha 0}\right)$$

$$MR\left(p^{(k)},s,\mathbf{x},\mu_\beta,\sigma_\beta,\eta_0,\sigma_\alpha\right)=p^{(k)}+p^{(k)}\left[\int_{\tilde{\alpha}}\frac{p^{(k)}\sigma_\alpha\exp\left[p^{(k)}\sigma_\alpha\left(\tilde{\alpha}+a^{(k)}\right)\right]}{\left[1+\exp\left[p^{(k)}\sigma_\alpha\left(\tilde{\alpha}+a^{(k)}\right)\right]\right]^2}\tilde{\alpha}\phi\left(\tilde{\alpha}-\eta_0\right)d\tilde{\alpha}\right]^{-1}s$$

$$=\quad p^{(k)\prime}+p^{(k)\prime}\left[\int_{\tilde{\alpha}}\frac{p^{(k)\prime}\sigma_\alpha\exp\left[p^{(k)\prime}\sigma_\alpha\left(\tilde{\alpha}+a^{(k)\prime}\right)\right]}{\left[1+\exp\left[p^{(k)\prime}\sigma_\alpha\left(\tilde{\alpha}+a^{(k)\prime}\right)\right]\right]^2}\tilde{\alpha}\phi\left(\tilde{\alpha}-\eta_0\right)d\tilde{\alpha}\right]^{-1}s'=MR\left(p^{(k)\prime},s',\mathbf{x},\mu_\beta,\sigma_\beta,\mu_\alpha,\sigma_\alpha\right)$$

hold, where $a_0^{(k)}$ satisfies

$$\int_{\tilde{\alpha}}\frac{\exp\left(p^{(k)}\sigma_{\alpha 0}\left(\tilde{\alpha}+a_0^{(k)}\right)\right)}{\left[1+\exp\left(p^{(k)}\sigma_{\alpha 0}\left(\tilde{\alpha}+a_0^{(k)}\right)\right)\right]}\phi\left(\tilde{\alpha}-\eta_0\right)d\tilde{\alpha}=s$$

and $a^{(k)}$ satisfies

$$\int_{\tilde{\alpha}}\frac{\exp\left(p^{(k)}\sigma_\alpha\left(\tilde{\alpha}+a^{(k)}\right)\right)}{\left[1+\exp\left(p^{(k)}\sigma_\alpha\left(\tilde{\alpha}+a^{(k)}\right)\right)\right]}\phi\left(\tilde{\alpha}-\eta_0\right)d\tilde{\alpha}=s$$

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

$$\int_{\tilde{\alpha}}\frac{\exp\left(p^{(k)\prime}\sigma_{\alpha 0}\left(\tilde{\alpha}+a_0^{(k)\prime}\right)\right)}{\left[1+\exp\left(p^{(k)\prime}\sigma_{\alpha 0}\left(\tilde{\alpha}+a_0^{(k)\prime}\right)\right)\right]}\phi\left(\tilde{\alpha}-\eta_0\right)d\tilde{\alpha}=s'$$

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

$$\int_{\tilde{\alpha}}\frac{\exp\left(p^{(k)\prime}\sigma_\alpha\left(\tilde{\alpha}+a^{(k)\prime}\right)\right)}{\left[1+\exp\left(p^{(k)\prime}\sigma_\alpha\left(\tilde{\alpha}+a^{(k)\prime}\right)\right)\right]}\phi\left(\tilde{\alpha}-\eta_0\right)d\tilde{\alpha}=s'.$$

Then,

$$\begin{aligned}\frac{p^{(k+1)\prime}}{p^{(k+1)}}&=\frac{1+\left[\int_{\tilde{\alpha}}\frac{p^{(k+1)}\sigma_{\alpha 0}\exp\left[p^{(k+1)}\sigma_{\alpha 0}\left(\tilde{\alpha}+a_0^{(k+1)}\right)\right]}{\left[1+\exp\left[p^{(k+1)}\sigma_{\alpha 0}\left(\tilde{\alpha}+a_0^{(k+1)}\right)\right]\right]^2}\tilde{\alpha}\phi\left(\tilde{\alpha}-\eta_0\right)d\tilde{\alpha}\right]^{-1}s}{1+\left[\int_{\tilde{\alpha}}\frac{p^{(k+1)\prime}\sigma_{\alpha 0}\exp\left[p^{(k+1)\prime}\sigma_{\alpha 0}\left(\tilde{\alpha}+a_0^{(k+1)\prime}\right)\right]}{\left[1+\exp\left[p^{(k+1)\prime}\sigma_{\alpha 0}\left(\tilde{\alpha}+a_0^{(k+1)\prime}\right)\right]\right]^2}\tilde{\alpha}\phi\left(\tilde{\alpha}-\eta_0\right)d\tilde{\alpha}\right]^{-1}s'}\\&=\frac{1+\left[\int_{\tilde{\alpha}}\frac{p^{(k)}\sigma_{\alpha 0}\exp\left[p^{(k)}\sigma_{\alpha 0}\left(\tilde{\alpha}+a_0^{(k)}\right)\right]}{\left[1+\exp\left[p^{(k)}\sigma_{\alpha 0}\left(\tilde{\alpha}+a_0^{(k)}\right)\right]\right]^2}\tilde{\alpha}\phi\left(\tilde{\alpha}-\eta_0\right)d\tilde{\alpha}\right]^{-1}s}{1+\left[\int_{\tilde{\alpha}}\frac{p^{(k)\prime}\sigma_{\alpha 0}\exp\left[p^{(k)\prime}\sigma_{\alpha 0}\left(\tilde{\alpha}+a_0^{(k)\prime}\right)\right]}{\left[1+\exp\left[p^{(k)\prime}\sigma_{\alpha 0}\left(\tilde{\alpha}+a_0^{(k)\prime}\right)\right]\right]^2}\tilde{\alpha}\phi\left(\tilde{\alpha}-\eta_0\right)d\tilde{\alpha}\right]^{-1}s'}=\frac{p^{(k)\prime}}{p^{(k)}}=...=\frac{p'}{p}\end{aligned}\qquad(48)$$

and by taking the limit,

$$lim_{k\to\infty}\frac{1+\left[\int_{\tilde{\alpha}}\frac{p^{(k)}\sigma_{\alpha 0}\exp\left[p^{(k)}\sigma_{\alpha 0}\left(\tilde{\alpha}+a_0^{(k)}\right)\right]}{\left[1+\exp\left[p^{(k)}\sigma_{\alpha 0}\left(\tilde{\alpha}+a_0^{(k)}\right)\right]\right]^2}\tilde{\alpha}\phi\left(\tilde{\alpha}-\eta_0\right)d\tilde{\alpha}\right]^{-1}s}{1+\left[\int_{\tilde{\alpha}}\frac{p^{(k)\prime}\sigma_{\alpha 0}\exp\left[p^{(k)\prime}\sigma_{\alpha 0}\left(\tilde{\alpha}+a_0^{(k)\prime}\right)\right]}{\left[1+\exp\left[p^{(k)\prime}\sigma_{\alpha 0}\left(\tilde{\alpha}+a_0^{(k)\prime}\right)\right]\right]^2}\tilde{\alpha}\phi\left(\tilde{\alpha}-\eta_0\right)d\tilde{\alpha}\right]^{-1}s'}=\frac{1+\left[\left(\eta_0-\Phi^{-1}\left(s\right)\right)\phi\left(\Phi^{-1}\left(s\right)\right)\right]^{-1}s}{1+\left[\left(\eta_0-\Phi^{-1}\left(s'\right)\right)\phi\left(\Phi^{-1}\left(s'\right)\right)\right]^{-1}s'}=\frac{p'}{p}.$$

Thus, for any k

$$G\equiv\frac{p'}{p}=\frac{1+\left[\left(\eta_0-\Phi^{-1}\left(s\right)\right)\phi\left(\Phi^{-1}\left(s\right)\right)\right]^{-1}s}{1+\left[\left(\eta_0-\Phi^{-1}\left(s'\right)\right)\phi\left(\Phi^{-1}\left(s'\right)\right)\right]^{-1}s'}=\frac{1+\left[\int_{\tilde{\alpha}}\frac{p^{(k)}\sigma_{\alpha 0}\exp\left[p^{(k)}\sigma_{\alpha 0}\left(\tilde{\alpha}+a_0^{(k)}\right)\right]}{\left[1+\exp\left[p^{(k)}\sigma_{\alpha 0}\left(\tilde{\alpha}+a_0^{(k)}\right)\right]\right]^2}\tilde{\alpha}\phi\left(\tilde{\alpha}-\eta_0\right)d\tilde{\alpha}\right]^{-1}s}{1+\left[\int_{\tilde{\alpha}}\frac{p^{(k)\prime}\sigma_{\alpha 0}\exp\left[p^{(k)\prime}\sigma_{\alpha 0}\left(\tilde{\alpha}+a_0^{(k)\prime}\right)\right]}{\left[1+\exp\left[p^{(k)\prime}\sigma_{\alpha 0}\left(\tilde{\alpha}+a_0^{(k)\prime}\right)\right]\right]^2}\tilde{\alpha}\phi\left(\tilde{\alpha}-\eta_0\right)d\tilde{\alpha}\right]^{-1}s'}.$$

Hence, if we denote

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

Now, denote $B=lim_{p\to\infty}B\left(p\right)=\left(\Phi^{-1}\left(s\right)-\eta_0\right)\phi\left(\Phi^{-1}\left(s\right)\right)$, $B'=lim_{p\to\infty}B'\left(p\right)=\left(\Phi^{-1}\left(s'\right)-\eta_0\right)\phi\left(\Phi^{-1}\left(s'\right)\right)$, and $B\left(p^{(k)}\right)=B^{(k)}-B$, $B'\left(Gp^{(k)}\right)=B'^{(k)}-B'$, hence

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

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 (39), we know that

$$\begin{aligned}
B\left(p\right) &= \left[\left(\phi\left(-a_0-\eta_0\right)-a_0\phi'\left(-a_0-\eta_0\right)\right)\frac{\phi'\left(-a_0-\eta_0\right)}{2\phi\left(-a_0-\eta_0\right)}+\phi'\left(-a_0-\eta_0\right)-\frac{a_0}{2}\phi''\left(-a_0-\eta_0\right)\right]\frac{\pi^2}{3\sigma_{\alpha 0}^2p^2}+O\left(p^{-4}\right)\\
B'\left(p\right) &= \left[\left(\phi\left(-a_0'-\eta_0\right)-a_0'\phi'\left(-a_0'-\eta_0\right)\right)\frac{\phi'\left(-a_0'-\eta_0\right)}{2\phi\left(-a_0'-\eta_0\right)}+\phi'\left(-a_0'-\eta_0\right)-\frac{a_0'}{2}\phi''\left(-a_0'-\eta_0\right)\right]\frac{\pi^2}{3\sigma_{\alpha 0}^2G^2p^2}+O\left(p^{-4}\right)
\end{aligned}$$

Therefore, we can write the above equation as

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

where

$$
\begin{aligned}
b(p) &= \left[(\phi(-a_0-\eta_0)-a_0\phi'(-a_0-\eta_0))\frac{\phi'(-a_0-\eta_0)}{2\phi(-a_0-\eta_0)}+\phi'(-a_0-\eta_0)-\frac{a_0}{2}\phi''(-a_0-\eta_0)\right]\frac{\pi^2}{3\sigma_{\alpha 0}^2}+O\left(p^{-2}\right)\\
b'(p) &= \left[(\phi(-a_0'-\eta_0)-a_0'\phi'(-a_0'-\eta_0))\frac{\phi'(-a_0'-\eta_0)}{2\phi(-a_0'-\eta_0)}+\phi'(-a_0'-\eta_0)-\frac{a_0'}{2}\phi''(-a_0'-\eta_0)\right]\frac{\pi^2}{3\sigma_{\alpha 0}^2}+O\left(p^{-2}\right),
\end{aligned}
$$

and

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

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

$$
\frac{GBs'}{B'}b'\left(Gp^{(k)}\right)-\frac{B'sb\left(p^{(k)}\right)G^2}{B}=(G-1)\,p^{(k)2}b'\left(Gp^{(k)}\right)b\left(p^{(k)}\right) \tag{49}
$$

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

$$
(\phi(-a_0-\eta_0)-a_0\phi'(-a_0-\eta_0))\frac{\phi'(-a_0-\eta_0)}{2\phi(-a_0-\eta_0)}+\phi'(-a_0-\eta_0)-\frac{a_0}{2}\phi''(-a_0-\eta_0)=-\frac{3}{2}(a_0+\eta_0)-a_0(a_0+\eta_0)^2+\frac{a_0}{2},
$$

which positive if $a_0<0$. Similarly for $a_0'<0$. Therefore, as long as both s, s' take on values that are less than $\Phi(\eta_0)$, $a_0<0$, $a_0'<0$ and $b\left(p^{(k)}\right)=C+O\left(p^{-(k)2}\right)$, $C>0$ and $b'\left(p^{(k)}\right)=C'+O\left(p^{-(k)2}\right)$, $C'>0$. Therefore, 49 cannot hold for large k, and thus, neither does (48) for large k. Therefore, for those s, s', and $p\neq p'$, for some $k\geq 1$,

$$
MR\left(p^{(k)},s,\mathbf{x},\mu_\beta,\sigma_\beta,\eta_0,\sigma_{\alpha 0}\right)=MR\left(p^{(k)\prime},s',\mathbf{x},\mu_\beta,\sigma_\beta,\mu_{\alpha 0},\sigma_{\alpha 0}\right)
$$

but

$$
MR\left(p^{(k)},s,\mathbf{x},\mu_\beta,\sigma_\beta,\eta_0,\sigma_\alpha\right)\neq MR\left(p^{(k)\prime},s',\mathbf{x},\mu_\beta,\sigma_\beta,\mu_\alpha,\sigma_\alpha\right).
$$

Therefore, Assumption 12 holds.

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

B.3 Identification of the BLP model with X.

Lemma 4 *Suppose the Assumptions use in Proposition 4 hold, except that the for BLP model* $\mathbf{X}$ *is a scalar and its support conditional on* $Q,\boldsymbol{\xi},\upsilon$ *is nonnegative.*

Assumption 12 is still satisfied for the logit model with respect to α*. Assumption 12 is satisfied for the BLP model of demand for the parameters* $(\mu_\alpha,\sigma_\alpha)$ *and* σ_β *under monopoly as well.*

Proof. The identification of the price coefficient α of the logit model of demand with covariates $\mathbf{X}$ is the same as in the proof of Proposition 4. Next, consider including the observed product characteristics into the BLP random coefficient model. Since we have shown in Proposition 4 that Assumption 12 is satisfied for μ_α,σ_α for the data with $\mathbf{X}=0$, we assume that μ_α,σ_α are identified. Because, for the sake of simplicity, we assumed its dimension to be one, and denote it as X. Furthermore, 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 $X = A_X p$ for some positive constant A_X. Then,

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

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 $\sigma_{A_X\beta} = A_X\sigma_\beta$, $\mu_{A_X\beta} = A_X\mu_\beta$, the corresponding marginal revenue equations are:

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

Now, we have

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

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

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

For large p,

$$\int_{\tilde{\gamma}} \frac{\exp\left(p\sigma_\gamma\left(\tilde{\gamma}+g\left(p\right)\right)\right)}{1+\exp\left(p\sigma_\gamma\left(\tilde{\gamma}+g\left(p\right)\right)\right)}\phi\left(\tilde{\gamma}-\eta_\gamma\right)d\tilde{\gamma} = \Phi\left(g\left(p\right)+\eta_\gamma\right)+O\left(p^{-2}\right)=s,$$

where

$$g = lim_{p\to\infty} g\left(p\right), \quad g=\Phi^{-1}\left(s\right)-\eta_\gamma,$$

$$\eta_\gamma = \frac{\mu_\gamma}{\sigma_\gamma} = \frac{\mu_\alpha + A_X\mu_\beta}{\sqrt{\sigma_\alpha^2+A_X^2\sigma_\beta^2}}.$$

Hence, for large p,

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

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.

$$\begin{aligned}
MR &= p+p\left[\left[\frac{\mu_\alpha}{\sigma_\alpha}-\Phi^{-1}\left(s\right)\right]exp\left(-\frac{1}{2}\Phi^{-1}\left(s\right)^2\right)+O\left(p^{-1}\right)\right]^{-1}s \\
&= p'+p'\left[\left[\frac{\mu_\alpha}{\sigma_\gamma'}-\left(\frac{\sigma_\alpha}{\sigma_\gamma'}\right)^2\Phi^{-1}\left(s\right)\right]exp\left(-\frac{1}{2}\Phi^{-1}\left(s\right)^2\right)+O\left(p'^{-1}\right)\right]^{-1}s. \qquad (50)
\end{aligned}$$

Because $\frac{\mu_\alpha}{\sigma_\alpha}$ is identified, given p, s MR can be derived. 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.

$$\frac{MR-p'}{p'}\left[\nu'\mu_\alpha-\sigma_\alpha^2\Phi^{-1}\left(s\right)\nu'^2\right]exp\left(-\frac{1}{2}\Phi^{-1}\left(s\right)^2\right)=s \qquad (51)$$

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

, 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 ν' satisfying equation (51) can be made arbitrarily close to $1/\sigma_\gamma'$ satisfying equation (50) by making p'arbitrarily large, $\sigma_\gamma'>0$ is identified. Furthermore, if $s=1/2$, $\Phi^{-1}\left(s\right)=0$, then equation (50) implies

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

holds approximately for large p, p', where $G=p'/p$ 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 5

Proof. Because in the pair of firms in the proof we consider, both firms have the same **w** and the same **x**, for the sake of notational simplicity, we will omit **w** and **x** from all the equations.
a. Under the profit maximization assumption, $MR_i = MC_i$ at both points m, m'. Given $Q_m < Q_{m'}$, it follows from the strict convexity of the cost function that

$$MR\left(p_m, \xi\right) < \frac{\partial C\left(Q_{m'} s_m, \upsilon\right)}{\partial q} \tag{52}$$

$C\left(Q_{m'} s_{m'}, \upsilon\right)$ is the cost function specification, where $Q_{m'} s_{m'}$ is the output and υ the cost shock. Furthermore, consider $\tilde{s}$ such that $Q_{m'}\tilde{s} = Q_m s_m$, which implies $\tilde{s} < s_m$. From Assumption 14, there exists $\tilde{p} > p_m$ such that $\tilde{s} = s\left(\tilde{p}, \xi\right)$. Since, from Assumption 13, $MR\left(p, \xi\right)$ is strictly increasing in p,

$$MR\left(\tilde{p}, \xi\right) > \frac{\partial C\left(Q_{m'}\tilde{s}, \upsilon\right)}{\partial q} = \frac{\partial C\left(Q_m s_m, \upsilon\right)}{\partial q} = MR\left(p_m, \xi\right). \tag{53}$$

It follows from equations (52) and (53), and the Intermediate Value Theorem that there exists $p_{m'} > p_m$ and $s_{m'}$ such that $\tilde{s} < s_{m'} = s\left(p_{m'}, \xi\right) < s_m$,

$$MR\left(p_{m'}, \xi\right) = \frac{\partial C\left(Q_{m'} s_{m'}, \upsilon\right)}{\partial q}$$

are satisfied. Furthermore, $q_m = Q_{m'}\tilde{s} < Q_{m'} s_{m'} = q_{m'}$. It is also straightforward to show that $s_{m'} - s_m$ is a continuous function of $Q_{m'} - Q_m$ given ξ and υ remaining unchanged.

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

$$p_m\left[1 + \frac{lnp_{m'} - lnp_m}{lns_{m'} - lns_m}\right] = \frac{E\left[C^d|q_{m'}, p_{m'}, s_{m'}\right] - E\left[C^d|q_m, p_m, s_m\right]}{q_{m'} - q_m} + O\left(|Q_{m'} - Q_m|\right).$$

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

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

where $\left(\frac{\partial lns(p,\xi)}{\partial lnp}\right)$ is the elasticity of demand.

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

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

To see this, consider the case $p' > p = p\left(Q, \xi, \upsilon\right)$. By assumption, $MR\left(p', \xi\right) > MR\left(p, \xi\right)$, and $s\left(p', \xi\right) < s\left(p, \xi\right)$, thus, $q' < q$, $MC\left(q', \upsilon\right) < MC\left(q, \upsilon\right)$ and thus the F.O.C. does not hold. A similar logic applies for $p < p'$. Then, given $p\left(Q, \xi, \upsilon\right)$, $s = s\left(p, \xi\right) = s\left(Q, \xi, \upsilon\right)$.

Now, marginal cost can be approximated using finite differences in total costs and quantities between points m and m' as follows:

$$\frac{\partial C\left(Q_m s_m, \upsilon\right)}{\partial q} = \frac{C\left(Q_{m'} s_{m'}, \upsilon\right) - C\left(Q_m s_m, \upsilon\right)}{Q_{m'} s_{m'} - Q_m s_m} + O\left(|Q_{m'} s_{m'} - Q_m s_m|\right) = \frac{C\left(Q_{m'} s_{m'}, \upsilon\right) - C\left(Q_m s_m, \upsilon\right)}{Q_{m'} s_{m'} - Q_m s_m} + O\left(|Q_{m'} - Q_m|\right).$$

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

$$\left(\frac{\partial lns\left(p_m,\xi\right)}{\partial lnp}\right)^{-1}=\frac{ln\left(p\left(Q_{m'},\xi,\upsilon\right)\right)-ln\left(p\left(Q_m,\xi,\upsilon\right)\right)}{ln\left(s\left(Q_{m'},\xi,\upsilon\right)\right)-ln\left(s\left(Q_m,\xi,\upsilon\right)\right)}+O\left(\left|Q_{m'}-Q_m\right|\right).$$

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

b. We denote the true marginal cost as

$$MC_m=\frac{\partial C\left(q_m,\upsilon_m\right)}{\partial q},$$

and $\widehat{MC}_m$ as the marginal cost estimate at (q_m,υ_m):

$$\widehat{MC}_m=\frac{E\left[C^d|q_{m'},p_{m'},s_{m'}\right]-E\left[C^d|q_m,p_m,s_m\right]}{q_{m'}-q_m}+O\left(\left|Q_{m'}-Q_m\right|\right).$$

From the first order condition, we know that the true marginal cost and marginal revenue are equal to each other. That is,

$$MC_m=MR\left(p_m,\xi_m\right)=p_m\left[1+\left[\frac{\partial lns\left(p_m,\xi_m\right)}{\partial lnp}\right]^{-1}\right]\Longrightarrow\left(\frac{\partial lns\left(p_m,\xi_m\right)}{\partial lnp}\right)^{-1}=\frac{MC_m}{p_m}-1.$$

Recall from our proof of part a that for sufficiently small $\Delta Q\equiv Q_{m'}-Q_m>0$, the points $\left(p\left(Q_m+\Delta Q,\xi_m,\upsilon_m\right),s\left(Q_m+\Delta Q,\xi_m,\upsilon_m\right)\right)$, $\left(p\left(Q_m,\xi_m,\upsilon_m\right),s\left(Q_m,\xi_m,\upsilon_m\right)\right)$ satisfy the following equation:

$$\frac{ln\left(p\left(Q_m+\Delta Q,\xi_m,\upsilon_m\right)\right)-ln\left(p\left(Q_m,\xi_m,\upsilon_m\right)\right)}{ln\left(s\left(Q_m+\Delta Q,\xi_m,\upsilon_m\right)\right)-ln\left(s\left(Q_m,\xi_m,\upsilon_m\right)\right)}=\frac{MC_m}{p\left(Q_m,\xi_m,\upsilon_m\right)}-1+O\left(\left(\Delta Q\right)\right)$$

Now, suppose that the estimated marginal cost $\widehat{MC}_m$ satisfies $limsup_{\Delta Q\downarrow 0}\widehat{MC}_m<MC_m$. We show that with such $\widehat{MC}_m$, one cannot find the two points that satisfy equation (18).

Consider a vector of price and market share $(\hat{p},\hat{s})$ close to (p_m,s_m) and $\hat{s}<s_m$ such that

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

Then, from continuity of market share with respect to market size, there exists sufficiently small $\Delta Q'>0$ such that $\hat{s}=s\left(Q_m+\Delta Q',\xi_m,\upsilon_m\right)$.

That is,

$$\frac{ln\left(\hat{p}\right)-ln\left(p\left(Q_m,\xi_m,\upsilon_m\right)\right)}{ln\left(\hat{s}\right)-ln\left(s\left(Q_m,\xi_m,\upsilon_m\right)\right)}=\frac{\widehat{MC}_m}{p\left(Q_m,\xi_m,\upsilon_m\right)}-1<\frac{MC_m}{p\left(Q_m,\xi_m,\upsilon_m\right)}-1+O\left(\Delta Q'\right).$$

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

$$\frac{ln\left(\hat{p}\right)-ln\left(p\left(Q_m,\xi_m,\upsilon_m\right)\right)}{ln\left(\hat{s}\right)-ln\left(s\left(Q_m,\xi_m,\upsilon_m\right)\right)}<\frac{ln\left(p\left(Q_m+\Delta Q',\xi_m,\upsilon_m\right)\right)-ln\left(p\left(Q_m,\xi_m,\upsilon_m\right)\right)}{ln\left(s\left(Q_m+\Delta Q',\xi_m,\upsilon_m\right)\right)-ln\left(s\left(Q_m,\xi_m,\upsilon_m\right)\right)}<0.$$

Given $\hat{s}=s\left(Q_m+\Delta Q',\xi_m,\upsilon_m\right)<s\left(Q_m,\xi_m,\upsilon_m\right)$ and Assumption 14, for the above inequality to hold, it must follow that $\hat{p}>p\left(Q_m+\Delta Q',\xi_m,\upsilon_m\right)$.

We now show that there exists such a pair $(\hat{s},\hat{p})$: specifically that there exists $(\xi_{m'},\upsilon_{m'})$ such that $\hat{s}=s\left(Q_m+\Delta Q',\xi_{m'},\upsilon_{m'}\right)$ and $\hat{p}=p\left(Q_m+\Delta Q',\xi_{m'},\upsilon_{m'}\right)$. For that, we need to show that $\xi_{m'}$ satisfying $\hat{s}=s\left(\hat{p},\xi_{m'}\right)$ and $\upsilon_{m'}$ satisfying $MR\left(\hat{p},\xi_{m'}\right)=MC\left(\hat{s}\left(Q_m+\Delta Q'\right),\upsilon_{m'}\right)$ exist. Since $\hat{p}>p\left(Q_m+\Delta Q',\xi_m,\upsilon_m\right)$, it follows by Assumption 14 that $s\left(\hat{p},\xi_m\right)<s\left(p\left(Q_m+\Delta Q',\xi_m,\upsilon_m\right),\xi_m\right)<1$. Furthermore, $s\left(\hat{p},\xi_m\right)>0$. Then, it follows from Assumption 14 and the Intermediate Value Theorem that there exists $\xi_{m'}>\xi_m$ such that $\hat{s}=s\left(\hat{p},\xi_{m'}\right)$.

Similarly, by Assumption 3' and by the Intermediate Value Theorem, we can find a $\upsilon_{m'}$ that satisfies $MR(\hat{p}, \xi_{m'}) = MC(\hat{s}(Q_m + \Delta Q'), \upsilon_{m'})$.

Because $\hat{s} = s(p(Q_m + \Delta Q', \xi_{m'}, \upsilon_{m'}), \xi_{m'}) = s(p(Q_m + \Delta Q', \xi_m, \upsilon_m), \xi_m)$ and $p(Q_m + \Delta Q', \xi_{m'}, \upsilon_{m'}) > p(Q_m + \Delta Q', \xi_m, \upsilon_m)$, from Assumption 13,

$$MR(p(Q_m + \Delta Q', \xi_{m'}, \upsilon_{m'}), \xi_{m'}) > MR(p(Q_m + \Delta Q', \xi_m, \upsilon_m), \xi_m).$$

Furthermore,

$$s(Q_m + \Delta Q', \xi_{m'}, \upsilon_{m'})(Q_m + \Delta Q') = s(Q_m + \Delta Q', \xi_m, \upsilon_m)(Q_m + \Delta Q') \equiv q_m + \Delta q'.$$

Therefore,

$$\frac{\partial C(q_m + \Delta q', \upsilon_{m'})}{\partial q} = MR(p(Q_m + \Delta Q', \xi_{m'}, \upsilon_{m'}), \xi_{m'}) > MR(p(Q_m + \Delta Q', \xi_m, \upsilon_m), \xi_m) = \frac{\partial C(q_m + \Delta q', \upsilon_m)}{\partial q},$$

which, from Assumption 3' implies $\upsilon_{m'} > \upsilon_m$. cost function is increasing in υ,

$$\frac{C(q_m + \Delta q', \upsilon_{m'}) - C(q_m, \upsilon_m)}{\Delta q'} > \frac{C(q_m + \Delta q', \upsilon_m) - C(q_m, \upsilon_m)}{\Delta q'} > \widehat{MC}_m$$

and because

$$lim_{\Delta q' \to 0} \frac{(C(q_m + \Delta q', \upsilon_{m'}) - C(q_m, \upsilon_m))}{\Delta q'} > MC(s_m Q_m, \upsilon_m) > \widehat{MC}_m = p(Q_m, \xi_m, \upsilon_m)\left[1 + \frac{ln(\hat{p}) - ln(p(Q_m, \xi_m, \upsilon_m))}{ln(\hat{s}) - ln(s(Q_m, \xi_m, \upsilon_m))}\right]$$

Equation (18) does not hold. The proof for the case where the estimated marginal cost is greater than the true marginal cost (e.g., $\widehat{MC}_m > MC_m$) follows similarly. We illustrate the basic logic of the proof of part (b), using Figures 1 and 2. We set the x-axis to be the market share s and y-axis to be the price p. Let AB to be the true demand curve with demand shock being ξ_m and let AC be the corresponding marginal revenue curve. Also, let the line $MC(sQ_m, \upsilon_m)$ be the true supply curve with cost shock υ_m and market size Q_m. Then, the equilibrium price and market share for (ξ_m, υ_m) is at point E. Furthermore, at $s(Q_m + \Delta Q', \xi_m, \upsilon_m) < s(Q_m, \xi_m, \upsilon_m)$, the equilibrium marginal revenue and marginal cost curve are given by H. Note that the marginal cost is given by a higher curve than $MC(sQ_m, \upsilon_m)$ because the x-axis is s not q and by part (a), we know that $q_{m'} > q_m$. Let $\widehat{MC}_m(sQ_m, \upsilon_m)$, the estimated marginal cost at the original point be lower than the true supply curve, $MC(sQ_m, \upsilon_m)$. First, because $\widehat{MC}_1$ is lower, at point E, the estimated markup, which is the inverse of the price elasticity of market share, is larger than the true markup. This implies that the estimated slope of the demand curve has to be steeper than the true demand curve. Therefore, the estimated demand curve going through point E is the red line FE instead of AB. Then, wlog, if we set the candidate point $P_{m'}$ to have the market share $s(Q + \Delta Q', \xi_m, \upsilon_m)$, the corresponding price is at F, above the original true demand curve AB. Next, we take a look at Figure 2, where the points E, F and H are also shown. In this Figure, we draw the true demand curve with $\xi_{m'} > \xi_m$, through F . The corresponding marginal revenue curve with $\xi_{m'}$ is IJ, so that for F to be observed, the true marginal cost curve must go through J (the red positively sloped line), but the estimated marginal cost curve $\widehat{MC}_m(s(Q_m + \Delta Q'), \upsilon_m)$ is too low for that. Hence, point F and E together do not satisfy Equation (17). A similar argument can be made for $\widehat{MC}_m(sQ_m, \upsilon_m) > MC_m(sQ_m, \upsilon_m)$ as well.

c. Suppose that around (Q_m, q_m, p_m, s_m), $MC(q, \upsilon) = MC(\upsilon)$. Then, for $(Q_{m'}, q_{m'}, p_{m'}, s_{m'})$, $Q_{m'} = Q_m + \Delta Q$, $\Delta Q > 0$ having the same υ implies the same MC, and also the same MR. Thus, given the demand shock ξ being the same, $p_{m'} = p_m$, $s_{m'} = s_m$, $q_{m'} = q_m + \Delta Q s_m$. Next, we prove the 2nd part of c. Choose the point $(Q_{m'}, q_{m'}, p_{m'}, s_{m'})$ such that $p_{m'} = p_m$, $s_{m'} = s_m$. Then,

$$MC(\upsilon_m) = \frac{E\left[C^d | q_{m'}, p_{m'}, s_{m'}\right] - E\left[C^d | q_m, p_m, s_m\right]}{q_{m'} - q_m}.$$

To see this, since $p_m = p_{m'}$, $s_m = s_{m'}$, $MR_m = MR_{m'} = MC(\upsilon_m) = MC(\upsilon_{m'})$, which implies

$\upsilon_m = \upsilon_{m'}$. Therefore,

$$\frac{E\left[C^d|q_{m'}, p_{m'}, s_{m'}\right] - E\left[C^d|q_m, p_m, s_m\right]}{q_{m'} - q_m} = \frac{C\left(q_m, \upsilon_m\right) + MC\left(\upsilon_m\right)\Delta Q s_m - C\left(q_m, \upsilon_m\right)}{\Delta Q s_m} = MC\left(\upsilon_m\right)$$

■

Figure 1

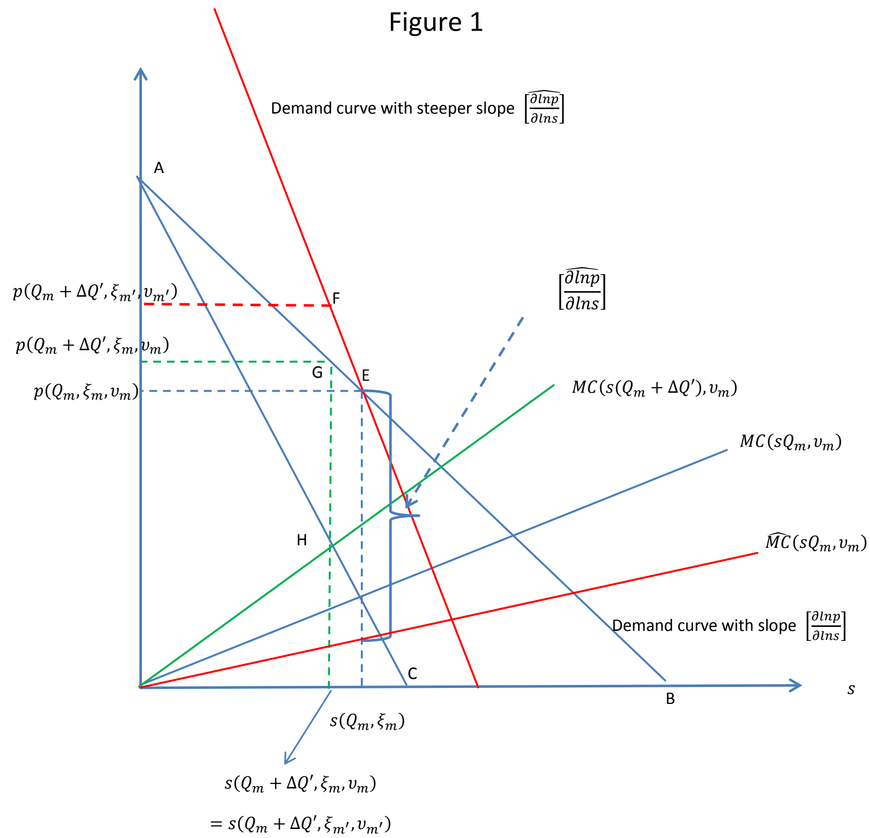
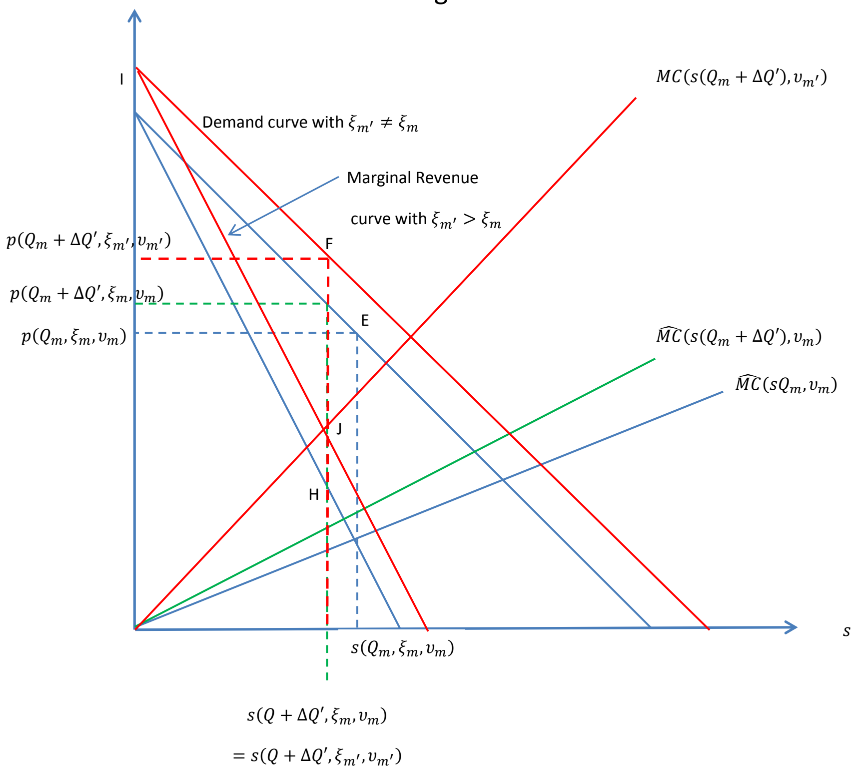


Figure 2



C.2 Nonparametric Identification of Oligopoly Marginal Revenue.

The marginal revenue of firm 1 in market m can be expressed as a function of own price p_{1m}, price of other firms $\mathbf{p}_{-1m}$, and the vector of observed and unobserved product (firm) characteristics $\mathbf{X}_m$ $\boldsymbol{\xi}_m$ of all firms in the market, i.e., $MR(p_{1m}, \mathbf{p}_{-1m}, \mathbf{X}_m, \boldsymbol{\xi}_m, 1)$. Next, we impose the following two assumptions, which are similar to Assumptions 11 and 12 for the monopoly case.

Assumption 17 *The marginal revenue of firm 1 (wlog, we set the firm under consideration to be the firm 1) in that market, denoted by* $MR(p_m, \mathbf{p}_{-1m}, \mathbf{X}_m, \boldsymbol{\xi}_m, 1)$ *is strictly increasing in own price* p_m. *Furthermore, suppose that we have the second market with demand variables* $\{\mathbf{p}_{m'}, \mathbf{s}_{m'}, \mathbf{X}_{m'}, \boldsymbol{\xi}_{m'}\}$ *, such* *that* $\mathbf{X}_m = \mathbf{X}_{m'} = \mathbf{X}$, $s_{1m} = s_{1m'} \equiv s$, $p_{1m} > p_{1m'}$, *and* $\mathbf{p}_{-1m} = \mathbf{p}_{-1m'} \equiv \mathbf{p}_{-1}$. *Then,*

$$MR\left(p_{1m}, \mathbf{p}_{-1m}, \mathbf{X}_m, \boldsymbol{\xi}\left(\mathbf{p}_m, \mathbf{s}_m, \mathbf{X}\right), 1\right) > MR(p_{1m'}, \mathbf{p}_{-1m'}, \mathbf{X}_{m'}, \boldsymbol{\xi}\left(\mathbf{p}_{m'}, \mathbf{s}_{m'}, \mathbf{X}\right), 1).$$

Assumption 18 *Given* $\left\{\mathbf{p}, \mathbf{X}, \boldsymbol{\xi}_{-1}\right\}$, *market share of firm 1,* $s\left(\mathbf{p}, \mathbf{X}, \xi, \boldsymbol{\xi}_{-1}, 1\right)$ *is strictly increasing and continuous in* ξ. *Furthermore, given* $\{\mathbf{p}_{-1}, \mathbf{X}, \boldsymbol{\xi}\}$, *market share of firm 1,* $s\left(p, \mathbf{p}_{-1}, \mathbf{X}, \boldsymbol{\xi}, 1\right)$ *is strictly decreasing and continuous in* p. *Furthermore,*

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

Proposition 6 *Suppose Assumptions 1,2,3',4, 5, 6, 10 and Assumptions 17, 18 are satisfied. Then,*

a. *Suppose marginal cost is increasing in output* q. *Suppose we have a firm with* $P_m = \{Q_m, \mathbf{w}_m, \mathbf{q}_m, \mathbf{p}_m, \mathbf{s}_m, \mathbf{X}_m, 1\}$ *and another firm with* $P_{m'} = \{Q_{m'}, \mathbf{w}_{m'}, \mathbf{q}_{m'}, \mathbf{p}_{m'}, \mathbf{s}_{m'}, \mathbf{X}_{m'}, 1\}$ *that is close to* P_m, *and* $\mathbf{w}_m = \mathbf{w}_{m'} = \mathbf{w}$, $\mathbf{X}_m = \mathbf{X}_{m'} = \mathbf{X}$ *and the same demand shocks* ($\boldsymbol{\xi}_m = \boldsymbol{\xi}_{m'} = \boldsymbol{\xi}$), *cost shocks that satisfy* $\upsilon_{1m} = \upsilon_{1m'} = \upsilon$, *and different market size:* $Q_m < Q_{m'}$. *Then, there exist cost shocks* $\boldsymbol{\upsilon}_{-1m}$ *and* $\boldsymbol{\upsilon}_{-1m'}$ *that are consistent with* $\mathbf{p}_{-1m} = \mathbf{p}_{-1m'} = \mathbf{p}_{-1}$. *Furthermore, it follows that*

$$s_{1m} > s_{1m'}, \quad p_{1m} < p_{1m'}, \; q_{1m} < q_{1m'}, \tag{54}$$

and

$$p_{1m}\left[1 + \frac{lnp_{1m'} - lnp_{1m}}{lns_{1m'} - lns_{1m}}\right] = \frac{E\left[C^d|q_{1m'}, \mathbf{w}_{m'}, \mathbf{p}_{m'}, \mathbf{s}_{m'}, \mathbf{X}_{m'}, 1\right] - E\left[C^d|q_{1m}, \mathbf{w}_m, \mathbf{p}_m, \mathbf{s}_m, \mathbf{X}_m, 1\right]}{q_{1m'} - q_{1m}} + O\left(|Q_{m'} - Q_m|\right). \tag{55}$$

b. *Suppose marginal cost is increasing in output* q. *Consider a firm with* $P_m = \{Q_m, \mathbf{w}_m, \mathbf{q}_m, \mathbf{p}_m, \mathbf{s}_m, \mathbf{X}_m, 1\}$ *and another firm with* $P_{m'} = \{Q_{m'}, \mathbf{w}_{m'}, \mathbf{q}_{m'}, \mathbf{p}_{m'}, \mathbf{s}_{m'}, \mathbf{X}_{m'}, 1\}$ *close to* P_m, *such that*

$$Q_m < Q_{m'}, \; s_{1m} > s_{1m'}, \; p_{1m} < p_{1m'}, \; q_{1m} < q_{1m'}, \; \text{and } \mathbf{p}_{-1m} = \mathbf{p}_{-1m'} = \mathbf{p}_{-1},$$

and equation (55) are satisfied. Then, the true marginal cost of firm 1, MC_{1m} *satisfies*

$$MC_{1m} = \frac{E\left[C^d|q_{1m'}, \mathbf{w}_{m'}, \mathbf{p}_{m'}, \mathbf{s}_{m'}, \mathbf{X}_{m'}, 1\right] - E\left[C^d|q_{1m}, \mathbf{w}_m, \mathbf{p}_m, \mathbf{s}_m, \mathbf{X}_m, 1\right]}{q_{1m'} - q_{1m}} + O\left(|Q_{m'} - Q_m|\right).$$

c. *Suppose marginal cost is constant in output. Then, consder a firm with* $P_m = \{Q_m, \mathbf{w}_m, \mathbf{q}_m, \mathbf{p}_m, \mathbf{s}_m, \mathbf{X}_m, 1\}$ *and another firm with* $P_{m'} = \{Q_{m'}, \mathbf{w}_{m'}, \mathbf{q}_{m'}, \mathbf{p}_{m'}, \mathbf{s}_{m'}, \mathbf{X}_{m'}, 1\}$ *close to* P_m, *that has the same demand shock vector* ($\boldsymbol{\xi}_{m'} = \boldsymbol{\xi}_m = \boldsymbol{\xi}$) *and cost shock* ($\upsilon_{1m'} = \upsilon_{1m} = \upsilon$) *and different market size* $Q_{m'} \neq Q_m$. *Then, it follows that* $\mathbf{s}_{m'} = \mathbf{s}_m$, $\mathbf{p}_{m'} = \mathbf{p}_m$ *and* $q_{m'} = Q_{m'} s_{1m}$ *and*

$$MC_{1m} = \frac{E\left[C^d|q_{1m'}, \mathbf{w}_{m'}, \mathbf{p}_{m'}, \mathbf{s}_{m'}, \mathbf{X}_{m'}, 1\right] - E\left[C^d|q_{1m}, \mathbf{w}_m, \mathbf{p}_m, \mathbf{s}_m, \mathbf{X}_m, 1\right]}{q_{1m'} - q_{1m}}. \tag{56}$$

Furthermore, suppose the marginal cost is constant in output. Then, consider a firm with $P_{m'} = \{Q_{m'}, \mathbf{w}_{m'}, \mathbf{q}_{m'}, \mathbf{p}_{m'}, \mathbf{s}_{m'}, \mathbf{X}_{m'}, 1\}$ *close to* P_m *that satisfies* $\mathbf{s}_{m'} = \mathbf{s}_m$, $\mathbf{p}_{m'} = \mathbf{p}_m$, $q_{m'} = Q_{m'} s_{1m}$ *and* $Q_{m'} \neq Q_m$. *Then,* $\boldsymbol{\xi}_{m'} = \boldsymbol{\xi}_m$, $\upsilon_{1m'} = \upsilon_{1m}$ *and and equation (56) is satisfied.*

Proof.
As before, we suppress **w** and **X** from the notation because in this proof, all firms under consideration have the same **w** and **X**.
a. Under the profit maximization assumption, $MR\left(\mathbf{p}_k, \mathbf{s}_k, \boldsymbol{\xi}_k, 1\right) = MC\left(q_{1k}, \upsilon_{1k}\right)$ at both markets $k = m$, m'. Given $Q_{m'} > Q_m$, it follows from the assumption that the marginal cost is strictly increasing in output,

$$MR\left(p_{1m}, \mathbf{p}_{-1m}, \boldsymbol{\xi}; 1\right) < \frac{\partial C\left(Q_{m'} s_{1m}, \upsilon\right)}{\partial q} \tag{57}$$

Furthermore, consider $\tilde{s}$ such that $Q_{m'}\tilde{s} = Q_m s_{1m} = q_{1m}$ which implies $\tilde{s} < s_{1m}$. From Assumption 18, there exists $\tilde{p} > p_{1m}$ such that $\tilde{s} = s\left(\tilde{p}, \mathbf{p}_{-1}, \boldsymbol{\xi}, 1\right)$. Since, from Assumption 17, $MR\left(p, \mathbf{p}_{-1}, \boldsymbol{\xi}, 1\right)$ is strictly increasing in p,

$$MR\left(\tilde{p},\mathbf{p}_{-1},\boldsymbol{\xi},1\right) > \frac{\partial C\left(Q_{m'}\tilde{s},\upsilon\right)}{\partial q} = \frac{\partial C\left(Q_{m}s_{1m},\upsilon\right)}{\partial q} = MR\left(p_{1m},\mathbf{p}_{-1},\boldsymbol{\xi},1\right). \tag{58}$$

Because both marginal revenue and marginal cost functions are continuous, it follows from equations (57) and (58), and the Intermediate Value Theorem that there exists $p_{1m'}$ such that $p_{1m} < p_{1m'} < \tilde{p}$ and $s_{1m'}$ such that $\tilde{s} < s_{1m'} = s\left(p_{1m'},\mathbf{p}_{-1},\boldsymbol{\xi},1\right) < s_{1m}$ and

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

Then, $q_{1m'} = Q_{m'}s_{1m'} > Q_{m'}\tilde{s} = q_{1m}$.
We also need to show that cost shocks $\boldsymbol{\upsilon}_{-1m}$ and $\boldsymbol{\upsilon}_{-1m'}$ can be chosen at such level such that $\mathbf{p}_{-1m} = \mathbf{p}_{-1m'} = \mathbf{p}_{-1}$ is satisfied. This is straightforward from Assumption 3', i.e. for any $j \neq 1$, one can find υ_{jm} such that

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

and similarly, one can find $\upsilon_{jm'}$ such that

$$MR\left(p_{1m'},\mathbf{p}_{-1},\boldsymbol{\xi},j\right) = \frac{\partial C\left(Q_{m}s\left(p_{1m'},\mathbf{p}_{-1},\boldsymbol{\xi},j\right),\upsilon_{jm'}\right)}{\partial q}.$$

Finally, it remains to show that,

$$p_{1m}\left[1+\frac{lnp_{1m'}-lnp_{1m}}{lns_{1m'}-lns_{1m}}\right] = \frac{E\left[C^{d}|q_{1m'},\mathbf{p}_{m'},\mathbf{s}_{m'}\right]-E\left[C^{d}|q_{1m},\mathbf{p}_{m},\mathbf{s}_{m}\right]}{q_{1m'}-q_{1m}} + O\left(|Q_{m'}-Q_{m}|\right).$$

The first order condition for profit maximization for firm 1 in market m can be re-written as,

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

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

$$\frac{\partial C\left(Q_{m}s_{1m},\upsilon\right)}{\partial q} = \frac{C\left(Q_{m'}s_{1m'},\upsilon\right)-C\left(Q_{m}s_{1m},\upsilon\right)}{Q_{m'}s_{1m'}-Q_{m}s_{1m}}+O\left(|Q_{m'}s_{1m'}-Q_{m}s_{1m}|\right) = \frac{C\left(Q_{m'}s_{1m'},\upsilon\right)-C\left(Q_{m}s_{1m},\upsilon\right)}{Q_{m'}s_{1m'}-Q_{m}s_{1m}}+O\left(|Q_{m'}-Q\right.$$

because both s_{1m} and $s_{1m'}$, are continuous functions of Q_m, $Q_{m'}$, respectively. Similarly, from the continuity of s_{1m}, $s_{1m'}$, p_{1m} and $p_{1m'}$, and the existence of the partial derivative, the marginal revenue can be approximated as,

$$\left(\frac{\partial lns\left(p_{1m},\mathbf{p}_{-1},\boldsymbol{\xi},1\right)}{\partial lnp}\right)^{-1} = \frac{lnp_{1m'}-lnp_{1m}}{ln\left(s\left(p_{1m'},\mathbf{p}_{-1},\boldsymbol{\xi},1\right)\right)-ln\left(s\left(p_{1m},\mathbf{p}_{-1},\boldsymbol{\xi},1\right)\right)} + O\left(|Q_{m'}-Q_{m}|\right).$$

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

$$MC_{1m} = \frac{\partial C\left(q_{1m},\upsilon_{1m}\right)}{\partial q_{1m}},$$

and let $\widehat{MC}_1$ be the marginal cost estimate at (q_{1m}, υ_{1m}). From the first order condition we know that true marginal cost and marginal revenue must be equal

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

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

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

Next, we show that 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, $\boldsymbol{\xi}$ uniquely determines s_1 by $s_1 = s\left(p, \mathbf{p}_{-1}, \boldsymbol{\xi}, 1\right)$. From the assumption, we know that the marginal revenue strictly increases with p given $\mathbf{p}_{-1}$ and $\boldsymbol{\xi}$.

Then, given Q, $q = Qs_1$, and υ satisfies the F.O.C:

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

Now, consider $p' > p$. Then, from the Assumption 17, $MR\left(p', \mathbf{p}_{-1}, \boldsymbol{\xi}, 1\right) > MR\left(p, \mathbf{p}_{-1}, \boldsymbol{\xi}, 1\right)$. Furthermore, from Assumption 18, $s_1' = s\left(p', \mathbf{p}_{-1}, \boldsymbol{\xi}, 1\right) < s_1 = s\left(p, \mathbf{p}_{-1}, \boldsymbol{\xi}, 1\right)$, $q_1' = Q_1\mathbf{s}_1' < Q_1\mathbf{s}_1 = q_1$. Because marginal cost is increasing in output, $MC\left(q_1', \upsilon\right) < MC\left(q_1, \upsilon\right)$, and thus, marginal revenue won't equal marginal cost. Similar argument can be made for $p' < p$. Therefore, $\{Q, \boldsymbol{\xi}, \upsilon, \mathbf{p}_{-1}\}$ uniquely determines p and s_1, and thus, the first firm's price and market share is a function of $\{Q, \boldsymbol{\xi}, \upsilon, \mathbf{p}_{-1}\}$. Therefore, following a similar arguement as the one in a, for sufficiently small $\Delta Q > 0$, the firm $\left(p\left(Q_m + \Delta Q, \boldsymbol{\xi}_m, \upsilon_{1m}, \mathbf{p}_{-1}, 1\right), s\left(Q_m + \Delta Q, \boldsymbol{\xi}_m, \upsilon_{1m}, \mathbf{p}_{-1}, 1\right)\right)$ satisfy the following equation,

$$p\left(Q_m, \boldsymbol{\xi}_m, \upsilon_{1m}, \mathbf{p}_{-1}, 1\right)\left[1 + \frac{ln\left(p\left(Q_m + \Delta Q, \boldsymbol{\xi}_m, \upsilon_{1m}, \mathbf{p}_{-1}, 1\right)\right) - ln\left(p\left(Q_m, \boldsymbol{\xi}_m, \upsilon_{1m}, \mathbf{p}_{-1}, 1\right)\right)}{ln\left(s\left(Q_m + \Delta Q, \boldsymbol{\xi}_m, \upsilon_{1m}, \mathbf{p}_{-1}, 1\right)\right) - ln\left(s\left(Q_m, \boldsymbol{\xi}_m, \upsilon_{1m}, \mathbf{p}_{-1}, 1\right)\right)}\right] = MC_{1m} + O\left(\left(\Delta Q\right)\right).$$

Hence,

$$\begin{aligned} & \frac{ln\left(p\left(Q_m + \Delta Q, \boldsymbol{\xi}_m, \upsilon_{1m}, \mathbf{p}_{-1}, 1\right)\right) - ln\left(p\left(Q_m, \boldsymbol{\xi}_m, \upsilon_{1m}, \mathbf{p}_{-1}, 1\right)\right)}{ln\left(s\left(Q_m + \Delta Q, \boldsymbol{\xi}_m, \upsilon_{1m}, \mathbf{p}_{-1}, 1\right)\right) - ln\left(s\left(Q_m, \boldsymbol{\xi}_m, \upsilon_{1m}, \mathbf{p}_{-1}, 1\right)\right)} \\ = & \frac{MC_{1m}}{p\left(Q_m, \boldsymbol{\xi}_m, \upsilon_{1m}, \mathbf{p}_{-1}, 1\right)} - 1 + O\left(\left(\Delta Q\right)\right) \end{aligned} \tag{59}$$

Now, suppose that the estimated marginal cost for firm 1 in market m is less than the true marginal cost, i.e., $limsup_{\Delta Q \downarrow 0}\widehat{MC}_{1m} < MC_{1m}$. Then, consider a vector of price and market share $(\hat{p}, \hat{s})$ that is close to (p_{1m}, s_{1m}), $\hat{s} < s_{1m}$ and satisfies

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

Then, from continuity of market share with respect to market size, there exists sufficiently small $\Delta Q' > 0$ such that $\hat{s} = s\left(Q_m + \Delta Q', \boldsymbol{\xi}_m, \upsilon_{1m}, \mathbf{p}_{-1}, 1\right)$ Thus,

$$\frac{ln\left(\hat{p}\right) - ln\left(p\left(Q_m, \boldsymbol{\xi}_m, \upsilon_{1m}, \mathbf{p}_{-1}, 1\right)\right)}{ln\left(\hat{s}\right) - ln\left(s\left(Q_m, \boldsymbol{\xi}_m, \upsilon_{1m}, \mathbf{p}_{-1}, 1\right)\right)} = \frac{\widehat{MC}_{1m}}{p\left(Q_m, \boldsymbol{\xi}_m, \upsilon_{1m}, \mathbf{p}_{-1}, 1\right)} - 1 < \frac{MC_{1m}}{p\left(Q_m, \boldsymbol{\xi}_m, \upsilon_{1m}, \mathbf{p}_{-1}, 1\right)} - 1 + O\left(\Delta Q'\right).$$

Hence, by using equation (59), for sufficiently small $\Delta Q' > 0$, we can derive:

$$\frac{ln\left(\hat{p}\right) - ln\left(p\left(Q_m, \boldsymbol{\xi}_m, \upsilon_{1m}, \mathbf{p}_{-1}, 1\right)\right)}{ln\left(\hat{s}\right) - ln\left(s\left(Q_m, \boldsymbol{\xi}_m, \upsilon_{1m}, \mathbf{p}_{-1}, 1\right)\right)} < \frac{ln\left(p\left(Q_m + \Delta Q', \boldsymbol{\xi}_m, \upsilon_{1m}, \mathbf{p}_{-1}, 1\right)\right) - ln\left(p\left(Q_m, \boldsymbol{\xi}_m, \upsilon_{1m}, \mathbf{p}_{-1}, 1\right)\right)}{ln\left(s\left(Q_m + \Delta Q', \boldsymbol{\xi}_m, \upsilon_{1m}, \mathbf{p}_{-1}, 1\right)\right) - ln\left(s\left(Q_m, \boldsymbol{\xi}_m, \upsilon_{1m}, \mathbf{p}_{-1}, 1\right)\right)} < 0.$$

The last inequality follows from the Assumption 18, where market share is assumed to be a strictly decreasing function of own price. Given $\hat{s} = s\left(Q_m + \Delta Q', \boldsymbol{\xi}_m, \upsilon_{1m}, \mathbf{p}_{-1}, 1\right) < s\left(Q_m, \boldsymbol{\xi}_m, \upsilon_{1m}, \mathbf{p}_{-1}, 1\right)$, for the above inequality to hold, it must follow that $\hat{p} > p\left(Q_m + \Delta Q', \boldsymbol{\xi}_m, \upsilon_{1m}, \mathbf{p}_{-1}, 1\right)$.
We now show that there exists such pair $(\hat{s}, \hat{p})$, specifically that there exists $(\boldsymbol{\xi}_{m'}, \boldsymbol{\upsilon}_{m'})$ such that $\xi_{1m'} > \xi_{1m}$, $\boldsymbol{\xi}_{-1m'} = \boldsymbol{\xi}_{-1m}$, and $\hat{s} = s\left(Q_m + \Delta Q', \boldsymbol{\xi}_{m'}, \upsilon_{1m'}, \mathbf{p}_{-1}, 1\right)$, $\hat{p} = p\left(Q_m + \Delta Q', \boldsymbol{\xi}_{m'}, \upsilon_{1m'}, \mathbf{p}_{-1}, 1\right)$. $s\left(p, \mathbf{p}_{-1}, \boldsymbol{\xi}_m, 1\right)$ being a continuous and decreasing function of own price and $\hat{p} > p\left(Q_m + \Delta Q', \boldsymbol{\xi}_m, \upsilon_{1m}, \mathbf{p}_{-1}, 1\right)$ implies $s\left(\hat{p}, \boldsymbol{\xi}_m, \mathbf{p}_{-1}, 1\right) < s\left(p\left(Q_m + \Delta Q', \boldsymbol{\xi}_m, \upsilon_{1m}, \mathbf{p}_{-1}, 1\right), \boldsymbol{\xi}_m, \mathbf{p}_{-1}, 1\right)$. Since $s\left(p, \mathbf{p}_{-1}, \xi, \boldsymbol{\xi}_{-1}, 1\right)$ is continuous and strictly increasing in ξ and $lim_{\xi\uparrow\infty} s\left(\hat{p}, \xi, \boldsymbol{\xi}_{-1m}, \mathbf{p}_{-1}, 1\right) = 1 > s\left(p\left(Q_m + \Delta Q', \boldsymbol{\xi}_m, \upsilon_{1m}, \mathbf{p}_{-1}, 1\right), \boldsymbol{\xi}_m, \mathbf{p}_{-1}, 1\right) > s\left(\hat{p}, \boldsymbol{\xi}_m, \mathbf{p}_{-1}, 1\right)$, it follows from the Intermediate Value Theorem with respect to ξ that there exists such $\boldsymbol{\xi}_{m'}$.

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

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

Since MC is an increasing and continuous function of υ and $lim_{\upsilon\downarrow 0} MC\left(\hat{s}\left(Q_m + \Delta Q'\right), \upsilon\right) = 0$ and $lim_{\upsilon\uparrow\infty} MC\left(\hat{s}\left(Q_m + \Delta Q'\right), \upsilon\right) = \infty$, from Intermediate Value Theorem, there exists $\upsilon_{1m'}$ that satisfies

$$MR\left(\hat{p}, \boldsymbol{\xi}_{m'}, \mathbf{p}_{-1}, 1\right) = MC\left(\hat{s}\left(Q_m + \Delta Q'\right), \upsilon_{1m'}\right),$$

Similarly, we can show that there exists $\upsilon_{jm'}$, $j \neq 1$ such that

$$MR\left(\hat{p}, \mathbf{p}_{-1}, \boldsymbol{\xi}_{m'}, j\right) = \frac{\partial C\left(Q_m s\left(\hat{p}, \mathbf{p}_{-1}, \boldsymbol{\xi}_{m'}, j\right), \upsilon_{jm'}\right)}{\partial q}.$$

Because $\hat{s} = s\left(Q_m + \Delta Q', \boldsymbol{\xi}_{m'}, \upsilon_{1m'}, \mathbf{p}_{-1}, 1\right) = s\left(Q_m + \Delta Q', \boldsymbol{\xi}_m, \upsilon_{1m'}, \mathbf{p}_{-1}, 1\right)$ and $\hat{p} = p\left(Q_m + \Delta Q', \boldsymbol{\xi}_{m'}, \upsilon_{1m'}, \mathbf{p}_{-1}, 1\right) > p\left(Q_m + \Delta Q', \boldsymbol{\xi}_m, \upsilon_{1m}, \mathbf{p}_{-1}, 1\right)$, if we denote $\tilde{p} \equiv p\left(Q_m + \Delta Q', \boldsymbol{\xi}_m, \upsilon_{1m}, \mathbf{p}_{-1}, 1\right)$, $\tilde{\mathbf{s}} \equiv \mathbf{s}\left(\tilde{p}, \mathbf{p}_{-1}, \boldsymbol{\xi}_m\right)$, then $\hat{p} > \tilde{p}$, $\tilde{s}_m = \hat{s}$ and from Assumption 17 we know that:

$$MR\left(\hat{p}, \mathbf{p}_{-1}, \boldsymbol{\xi}_{m'}\left(\hat{p}, \mathbf{p}_{-1}, \hat{\mathbf{s}}\right), 1\right) > MR\left(\tilde{p}, \boldsymbol{\xi}_m\left(\tilde{p}, \mathbf{p}_{-1}, \tilde{\mathbf{s}}\right), \mathbf{p}_{-1}, 1\right)$$

Furthermore,

$$\left(Q_m + \Delta Q'\right)\hat{s} = \left(Q_m + \Delta Q'\right)\tilde{s} \equiv q_{1m} + \Delta q'.$$

where $q_{1m} \equiv Q_m s_{1m}$. Therefore,

$$\begin{aligned}\frac{\partial C\left(q_{1m} + \Delta q', \upsilon_{1m'}\right)}{\partial q} &= MR\left(\hat{p}, \mathbf{p}_{-1}, \boldsymbol{\xi}_{m'}, 1\right)\\ > MR\left(\tilde{p}, \mathbf{p}_{-1}, \boldsymbol{\xi}_m, 1\right) &= \frac{\partial C\left(q_{1m} + \Delta q', \upsilon_{1m}\right)}{\partial q},\end{aligned}$$

which implies that $\upsilon_{1m'} > \upsilon_{1m}$. Therefore,

$$MC\left(q_{1m}, \upsilon_{1m'}\right) > MC\left(q_{1m}, \upsilon_{1m}\right) > \widehat{MC}_{1m}$$

and because marginal cost is increasing in output and υ, and so is the cost function, for sufficiently small $\Delta q'$,

$$\frac{C\left(q_{1m} + \Delta q', \upsilon_{1m'}\right) - C\left(q_{1m}, \upsilon_{1m}\right)}{\Delta q} > \frac{C\left(q_{1m} + \Delta q', \upsilon_{1m'}\right) - C\left(q_{1m}, \upsilon_{1m'}\right)}{\Delta q} > \widehat{MC}_{1m}$$

and because

$$\begin{aligned}& lim_{\Delta q'\to 0}\frac{C\left(q_{1m} + \Delta q', \upsilon_{1m'}\right) - C\left(q_{1m}, \upsilon_{1m'}\right)}{\Delta q} > MC\left(q_{1m}, \upsilon_{1m}\right) > \widehat{MC}_{1m}\\ = \quad & p\left(Q_m, \boldsymbol{\xi}_m, \upsilon_{1m}, \mathbf{p}_{-1}, 1\right)\left[1 + \frac{ln\left(\hat{p}\right) - ln\left(p\left(Q_m, \boldsymbol{\xi}_m, \upsilon_{1m}, \mathbf{p}_{-1}, 1\right)\right)}{ln\left(\hat{s}\right) - ln\left(s\left(Q_m, \boldsymbol{\xi}_m, \upsilon_{1m}, \mathbf{p}_{-1}, 1\right)\right)}\right]\end{aligned}$$

,equation (55) does not hold. The proof for the case with the estimated marginal cost is greater than the true marginal cost (e.g., $\widehat{MC}_{1m} > MC_{1m}$) follows similarly.

c. Suppose for any (q, υ), $MC(q,\upsilon) = MC(\upsilon)$. Then, for $P_{m'} = \{Q_{m'}, \mathbf{w}_{m'}, \mathbf{q}_{m'}, \mathbf{p}_{m'}, \mathbf{s}_{m'}, \mathbf{X}_{m'}, 1\}$ having the same $\boldsymbol{\upsilon}_{m'} = \boldsymbol{\upsilon}_m \equiv \boldsymbol{\upsilon}$, $MC(q_{jm}, \upsilon_j) = MC(q_{jm'}, \upsilon_j)$. Threfore, $MR(p_{1m}, \mathbf{p}_{-1}, \boldsymbol{\xi}, 1) = MR(p_{1m'}, \mathbf{p}_{-1}, \boldsymbol{\xi}, 1)$, which implies $p_{1m} = p_{1m'}$. Therefore, $\mathbf{p}_m = \mathbf{p}_{m'} \equiv \mathbf{p}$ and thus, $\mathbf{s}_m = \mathbf{s}(\mathbf{p}, \boldsymbol{\xi}) = \mathbf{s}_{m'}$. Finally, consider the case where the marginal cost is constant. Then, one can find a close point $(Q_{m'}, \mathbf{p}_{m'}, \mathbf{s}_{m'})$ such that $Q_{m'} \neq Q_m$, $\mathbf{p}_{m'} = \mathbf{p}_m$, $\mathbf{s}_{m'} = \mathbf{s}_m$, and thus, $\boldsymbol{\xi}_m = \boldsymbol{\xi}_{m'}$, and therefore, those two firms have the same marginal revenue, i.e. the same marginal cost, i.e. $MC(q_{1m}, \upsilon_{1m}) = MC(\upsilon_{1m}) = MC(q_{1m'}, \upsilon_{1m'}) = MC(\upsilon_{1m'})$ Because marginal cost is increasing in cost shock, $\upsilon_{1m} = \upsilon_{1m'} = \upsilon$, and thus,

$$\frac{E\left[C^d | q_{1m'}, \mathbf{p}_{m'}, \mathbf{s}_{m'}, 1\right] - E\left[C^d | q_{1m}, \mathbf{p}_m, \mathbf{s}_m, 1\right]}{q_{1m'} - q_{1m}} = \frac{C(q_{1m'}, \upsilon) - C(q_{1m}, \upsilon)}{q_{1m'} - q_{1m}} = \frac{MC(\upsilon)(q_{1m'} - q_{1m})}{q_{1m'} - q_{1m}} = MC(\upsilon)$$

and claim holds. ■

D Identification of the NLLS Sieve Estimator

We now prove identification of the NLLS Sieve Estimator defind in equations (23) and (25).

We also assume invertibility of unobserved product characteristics $\boldsymbol{\xi}$ from the vector of prices, market shares and observed product characteristics.

Assumption 19 *For any $J = 1, \ldots,$ $\mathbf{s}$ such that $0 < s_j < 1$, $j = 0, \ldots, J$, $\sum_{j=0}^{J} = 1$, and $\mathbf{p} \in R_+^J$, $\mathbf{X} \in R^J$, there exists a vector of unobserved characteristics $\boldsymbol{\xi}$ such that $\mathbf{s} = \mathbf{s}(\mathbf{p}, \mathbf{X}, \boldsymbol{\xi})$. Furthermore, we assume that given $\mathbf{p}$, $\mathbf{X}$, $\boldsymbol{\xi} = \boldsymbol{\xi}(\mathbf{s}, \mathbf{p}, \mathbf{X})$ is a continuous function.*

Proposition 7 *Suppose Assumptions 1-6, 8, 16-19 hold. Then, for any $\boldsymbol{\theta} \in \Theta$,*

$$E\left[\left(C_{jm}^d - \psi(q_{jm}, \mathbf{w}_{jm}, MR_{jm}(\boldsymbol{\theta}_0), \boldsymbol{\gamma}_0)\right)^2\right] = \sigma_\eta^2 \leq E\left[\left(C_{jm}^d - \psi(q_{jm}, \mathbf{w}_{jm}, MR_{jm}(\boldsymbol{\theta}), \boldsymbol{\gamma})\right)^2\right]$$

and equation (23) identifies $\boldsymbol{\theta}_{0p}$. That is, we prove the following. For any $\boldsymbol{\theta}_*$ that satisfies

$$E\left[\left(C_{jm}^d - \psi(q_{jm}, \mathbf{w}_{jm}, MR_{jm}(\boldsymbol{\theta}_*), \boldsymbol{\gamma}_*)\right)^2\right] = E\left[\left(C_{jm}^d - \psi(q_{jm}, \mathbf{w}_{jm}, MR_{jm}(\boldsymbol{\theta}_0), \boldsymbol{\gamma}_0)\right)^2\right] = \sigma_\eta^2$$

, $\boldsymbol{\theta}_{*p} = \boldsymbol{\theta}_{0p}$.

Proof. For each firm the observed cost is

$$C_{jm}^d = \varphi(C_{jm}) + \eta_{jm} = \varphi(PC(q_{jm}, \mathbf{w}_{jm}, MR(\mathbf{p}_m, \mathbf{s}_m, \mathbf{X}_m, j, \boldsymbol{\theta}_0))) + \eta_{jm}$$

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

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

Then, because of Assumption 8,

$$\begin{aligned}
& E\left[\left(C_{jm}^d - \psi(q_{jm}, \mathbf{w}_{jm}, MR_{jm}(\boldsymbol{\theta}), \boldsymbol{\gamma})\right)^2\right] \\
= \ & E\left[\left(\varphi(C_{jm}) - \psi(q_{jm}, \mathbf{w}_{jm}, MR_{jm}(\boldsymbol{\theta}), \boldsymbol{\gamma})\right)^2\right] + 2E\left[\left(\varphi(C_{jm}) - \psi(q_{jm}, \mathbf{w}_{jm}, MR_{jm}(\boldsymbol{\theta}), \boldsymbol{\gamma})\right)\eta_{jm}\right] + E\left(\eta_{jm}^2\right) \\
= \ & E\left[\left(\varphi(C_{jm}) - \psi(q_{jm}, \mathbf{w}_{jm}, MR_{jm}(\boldsymbol{\theta}_0), \boldsymbol{\gamma})\right)^2\right] + \sigma_\eta^2.
\end{aligned}$$

From Assumption 16, there exists an infinite sequence $\gamma_0 = \{\gamma_{0l}\}_{l=1}^{\infty}$ such that

$$\varphi\left(C_{jm}\right) = \varphi\left(PC\left(q_{jm}, \mathbf{w}_{jm}, MR_{jm}\left(\boldsymbol{\theta}_0\right)\right)\right) = \psi\left(q_{jm}, \mathbf{w}_{jm}, MR_{jm}\left(\boldsymbol{\theta}_0\right), \boldsymbol{\gamma}_0\right).$$

Therefore,

$$E\left[\left(C_{jm}^d - \psi\left(q_{jm}, \mathbf{w}_{jm}, MR_{jm}\left(\boldsymbol{\theta}\right), \boldsymbol{\gamma}\right)\right)^2\right] \geq E\left[\left(C_{jm}^d - \psi\left(q_{jm}, \mathbf{w}_{jm}, MR_{jm}\left(\boldsymbol{\theta}_0\right), \boldsymbol{\gamma}_0\right)\right)^2\right] + \sigma_\eta^2 = \sigma_\eta^2$$

and

$$E\left[\left(C_{jm}^d - \psi\left(q_{jm}, \mathbf{w}_{jm}, MR_{jm}\left(\boldsymbol{\theta}_*\right), \boldsymbol{\gamma}_*\right)\right)^2\right] = E\left[\left(\psi\left(q_{jm}, \mathbf{w}_{jm}, MR_{jm}\left(\boldsymbol{\theta}_0\right), \boldsymbol{\gamma}_0\right) - \psi\left(q_{jm}, \mathbf{w}_{jm}, MR_{jm}\left(\boldsymbol{\theta}_*\right), \boldsymbol{\gamma}_*\right)\right)^2\right] + \sigma_\eta^2.$$

First, assume $\boldsymbol{\theta}_{*p} \neq \boldsymbol{\theta}_{0p}$. Because $\psi\left(., ., ., \boldsymbol{\gamma}_0\right)$ is a continuous function on the support of $\left(q, \mathbf{w}, MR\left(\boldsymbol{\theta}_0\right)\right)$, in order for the equality

$$E\left[\left(C_{jm}^d - \psi\left(q_{jm}, \mathbf{w}_{jm}, MR_{jm}\left(\boldsymbol{\theta}_*\right), \boldsymbol{\gamma}_*\right)\right)^2\right] = \sigma_\eta^2$$

to hold,

$$\psi\left(q_j, \mathbf{w}_j, MR_j\left(\boldsymbol{\theta}_*\right), \boldsymbol{\gamma}_*\right) = \psi\left(q_j, \mathbf{w}_j, MR_j\left(\boldsymbol{\theta}_0\right), \boldsymbol{\gamma}_0\right)$$

needs to be satisfied for all $\left(q_j, \mathbf{w}_j, MR_j\left(\boldsymbol{\theta}_0\right)\right)$ belonging to the support of the joint distribution. Since, by assumption, the true pseudo-cost function is increasing in $MR_j(\theta_0)$, there exists a function d such that $d(MR\left(\mathbf{p}, \mathbf{s}, \mathbf{X}, j, \boldsymbol{\theta}_{*-p}, \boldsymbol{\theta}_{*p}\right)) = MR\left(\mathbf{p}, \mathbf{s}, \mathbf{X}, j, \boldsymbol{\theta}_{0-p}, \boldsymbol{\theta}_{0p}\right)$ has to hold for the variables in the support. Therefore,

$$\varphi\left(q_{jm}, \mathbf{w}_{jm}, MR_{jm}\left(\boldsymbol{\theta}_0\right), \boldsymbol{\gamma}_0\right) = \varphi\left(q_{jm}, \mathbf{w}_{jm}, MR_{jm}\left(\boldsymbol{\theta}_*\right), \boldsymbol{\gamma}_*\right),$$

and, $MR_{jm}\left(\boldsymbol{\theta}_0\right) = d\left(MR_{jm}\left(\boldsymbol{\theta}_*\right)\right)$, where d is a continuous function.

From Assumption 12, if $\boldsymbol{\theta}_{*p} \neq \boldsymbol{\theta}_{0p}$, there exists two firms with $\tilde{\boldsymbol{\nu}}$ and $\tilde{\tilde{\boldsymbol{\nu}}}$ such that for any $\boldsymbol{\theta}_{-p} \in \times_{-p}$

$$MR\left(\tilde{\mathbf{p}}, \tilde{\mathbf{s}}, \tilde{\mathbf{X}}, j, \boldsymbol{\theta}_{-p}, \boldsymbol{\theta}_{0p}\right) = MR\left(\tilde{\tilde{\mathbf{p}}}, \tilde{\tilde{\mathbf{s}}}, \tilde{\tilde{\mathbf{X}}}, j, \boldsymbol{\theta}_{-p}, \boldsymbol{\theta}_{0p}\right). \tag{60}$$

but

$$MR\left(\tilde{\mathbf{p}}, \tilde{\mathbf{s}}, \tilde{\mathbf{X}}, j, \boldsymbol{\theta}_{-p}, \boldsymbol{\theta}_{*p}\right) \neq MR\left(\tilde{\tilde{\mathbf{p}}}, \tilde{\tilde{\mathbf{s}}}, \tilde{\tilde{\mathbf{X}}}, j, \boldsymbol{\theta}_{-p}, \boldsymbol{\theta}_{*p}\right)$$

Now, set $\boldsymbol{\theta}_{-p} = \boldsymbol{\theta}_{*-p}$. Then,

$$\begin{aligned} MR\left(\tilde{\mathbf{p}}, \tilde{\mathbf{s}}, \tilde{\mathbf{X}}, j, \boldsymbol{\theta}_{*-p}, \boldsymbol{\theta}_{*p}\right) &\neq MR\left(\tilde{\tilde{\mathbf{p}}}, \tilde{\tilde{\mathbf{s}}}, \tilde{\tilde{\mathbf{X}}}, j, \boldsymbol{\theta}_{*-p}, \boldsymbol{\theta}_{*p}\right) \qquad (61) \\ &\Rightarrow d\left(MR\left(\tilde{\mathbf{p}}, \tilde{\mathbf{s}}, \tilde{\mathbf{X}}, j, \boldsymbol{\theta}_{*-p}, \boldsymbol{\theta}_{*p}\right)\right) \neq d\left(MR\left(\tilde{\tilde{\mathbf{p}}}, \tilde{\tilde{\mathbf{s}}}, \tilde{\tilde{\mathbf{X}}}, j, \boldsymbol{\theta}_{*-p}, \boldsymbol{\theta}_{*p}\right)\right), \\ &\Rightarrow MR\left(\tilde{\mathbf{p}}, \tilde{\mathbf{s}}, \tilde{\mathbf{X}}, j, \boldsymbol{\theta}_{0-p}, \boldsymbol{\theta}_{0p}\right) \neq \left(MR\left(\tilde{\tilde{\mathbf{p}}}, \tilde{\tilde{\mathbf{s}}}, \tilde{\tilde{\mathbf{X}}}, j, \boldsymbol{\theta}_{0-p}, \boldsymbol{\theta}_{0p}\right)\right) \end{aligned}$$

which contradicts equation 60. ■

E 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 $\widehat{PC}\left(q,\mathbf{w},MR,\hat{\boldsymbol{\gamma}}_M\right)$. Then, we can derive the nonparametric pseudo-marginal cost function as follows:

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

where the weight function is

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

$MR\left(\mathbf{X}_m,\mathbf{p}_m,\mathbf{s}_m,j,\hat{\boldsymbol{\theta}}_M\right)$ 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 $\bar{MR}=\widehat{MC}\left(\bar{q},\mathbf{w},\bar{C}\right)=MC\left(\bar{q},\mathbf{w},\bar{\upsilon}\right)$. 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

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

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

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

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

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

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

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

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

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

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

$$\mathbf{l}=\frac{\partial C\left(q,\mathbf{w},\upsilon\right)}{\partial\mathbf{w}}$$

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 $\bar{\upsilon}$. It is approximately:

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

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

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

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

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

for any $k>0$.

F 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[34]. 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 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 $\mathbf{y}_m=\left(\mathbf{q}_m,vec\left(\mathbf{W}_m\right)',\mathbf{C}_m^d,vec\left(\mathbf{X}_m\right)',vec\left(\mathbf{p}_m\right)',vec\left(\mathbf{s}_m\right)'\right)'$, where $\mathbf{C}_m^d=\left(C_{1m}^d,C_{2m}^d,...,C_{Jm}^d\right)'$, $\mathbf{W}_m=\left(\mathbf{w}_{1m},\mathbf{w}_{2m},...,\mathbf{w}_{jm}\right)'$and define

$$f\left(\mathbf{y}_m,\boldsymbol{\chi}\right)=\sum_{j=1}^{J}\left[C_{jm}^d-\sum_{l}\gamma_l\psi_l\left(q_{jm},\mathbf{w}_{jm},MR\left(\mathbf{X}_m,\mathbf{p}_m,\mathbf{s}_m,j,\boldsymbol{\theta}_c\right)\right)\right]^2, \tag{62}$$

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

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

[34]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 W.K.A (1988) and the related literature. As we have discussed earlier, those assumptions are not required for identification.

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

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

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

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 $\underline{Q(\boldsymbol{\chi}_0)} < \infty$.
(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 62. (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 $\sum_{n=1}^{\infty} \overline{\chi}_n < \infty$; $\sup_{n\geq 1} |\chi_{0,n}| / \overline{\chi}_n \leq 1$. 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

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

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

$$\begin{aligned} & inf_{\boldsymbol{\chi}\in\Xi_\eta} plim_{M\to\infty} \frac{1}{M}\sum_{m=1}^{M} f(\mathbf{y}_m, \boldsymbol{\chi}) \geq lim_{\epsilon\downarrow 0}\left[min_{k=1,...,K_\epsilon} E\left[inf_{\boldsymbol{\chi}_*\in B_\epsilon(\boldsymbol{\chi}_k)\cap\Xi_\eta} f(\mathbf{y}_m, \boldsymbol{\chi}_*)\right]\right] \\ = \; & inf_{\boldsymbol{\chi}\in\Xi_\eta} E[f(\mathbf{y}_m, \boldsymbol{\chi})] > E[f(\mathbf{y}_m, \boldsymbol{\chi}_0)]. \end{aligned} \tag{63}$$

Furthermore, from SLLN, we obtain

$$inf_{\boldsymbol{\chi}\in\bar{\Xi}\setminus\Xi_\tau} \frac{1}{M}\sum_{m=1}^{M} f(\mathbf{y}_m, \boldsymbol{\chi}) \geq \frac{1}{M}\sum_{m=1}^{M} inf_{\boldsymbol{\chi}\in\bar{\Xi}\setminus\Xi_\tau} f(\mathbf{y}_m, \boldsymbol{\chi}) \xrightarrow{a.s.} E\left[inf_{\boldsymbol{\chi}\in\bar{\Xi}\setminus\Xi_\tau} f(\mathbf{y}_m, \boldsymbol{\chi})\right] > E[f(\mathbf{y}_m, \boldsymbol{\chi}_0)]. \tag{64}$$

From 63 and 64, we derive

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

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

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

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

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

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,

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

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

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

and similarly, for $\boldsymbol{\theta}_\beta$ such that $\|\boldsymbol{\theta}_\beta - \boldsymbol{\theta}_{\beta 0}\| \geq \eta$,

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

and

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

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

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

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

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

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

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

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

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

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

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

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

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

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

(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

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

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

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

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

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),

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

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

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

Now, we impose an addititional assumption that

Assumption E.5

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

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

$$\boldsymbol{\Upsilon} = \left[\begin{array}{cc} A^2\left[\mathbf{G}'\mathbf{W}\boldsymbol{\Sigma}_{\boldsymbol{g}}\mathbf{W}\mathbf{G}\right]_{1,1} & A^2\left[\mathbf{G}'\mathbf{W}\boldsymbol{\Sigma}_{\boldsymbol{g}}\mathbf{W}\mathbf{G}\right]_{1,2:p} \\ A^2\left[\mathbf{G}'\mathbf{W}\boldsymbol{\Sigma}_{\boldsymbol{g}}\mathbf{W}\mathbf{G}\right]_{2:p,1} & A^2\left[\mathbf{G}'\mathbf{W}\boldsymbol{\Sigma}_{\boldsymbol{g}}\mathbf{W}\mathbf{G}\right]_{2:p,2:p} + \int_0^1 \int_0^1 a(u_1)\, \boldsymbol{\Gamma}(u_1, u_2)\, a(u_2)\, du_1 du_2 \end{array}\right]$$

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