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

David P. Byrne† Masayuki Hirukawa‡

Susumu Imai§ Vasilis Sarafidis¶

June 11, 2014

Abstract

This paper proposes a methodology for jointly estimating the demand and cost functions of differentiated goods oligopoly models when demand and cost data are available. The method deals with the endogeneity of prices to unobserved product quality in the demand function, as well as the endogeneity of output to unobserved cost shocks in firms’ cost functions. Our method does not, however, require instruments to do so. We establish non-parametric identification, consistency and asymptotic normality of our estimator. Using Monte-Carlo experiments, we demonstrate that it works well in situations where instruments are correlated with the error term in the demand and cost functions, and where the standard IV approach results in bias.

*We are grateful to seminar participants at the University of Melbourne, Monash University, ANU, Latrobe University, University Technology Sydney, UBC and the University of New South Wales for comments.
†Dept. of Economics, University of Melbourne, byrned@unimelb.edu.au
‡Dept. of Economics, Setsunan University, hirukawa@econ.setsunan.ac.jp
§Business School, University of Technology Sydney, susumu.imai@uts.edu.au
¶Dept. of Econometrics and Business Statistics, Monash University, vasilis.sarafidis@monash.edu
1 Introduction

In this paper, we develop a new methodology for estimating models of differentiated products markets. Our approach relies on the availability of standard demand-side data on products’ prices, market shares, and product characteristics, some firm-level cost data, and the assumption that firms set prices to maximize profits. The key novelty of our method is it does not require instrumental variables to deal with the endogeneity of prices to demand shocks in estimating demand, nor to account for the endogeneity of output to cost shocks in estimating cost functions.

Our study is motivated by questions surrounding the validity of IV-based identification strategies for differentiated products models, as well as recent applications that have started to leverage cost data for model testing and identification. The econometric frameworks of interest are the logit and random coefficient logit models of Berry (1994) and Berry, Levinsohn, and Pakes (1995) (hereafter, BLP), methodologies that have had a substantial impact on empirical research in IO and various other areas of economics. These models incorporate unobserved heterogeneity in product quality, and use instruments to deal with the endogeneity of prices to these demand shocks. As Berry and Haile (2014) and others have pointed out, so long as there are instruments available, fairly flexible demand functions can be identified using market-level data. Further, in the absence of cost data, firms’ marginal cost functions can be recovered, given a consistently estimated demand system, assumptions on cost structure (such as constant marginal costs), and the assumption that firms set prices to maximize profits.

A central issue then as far as the asymptotic properties of these estimators and the predictions they deliver is concerned the validity of the instruments. Commonly used IVs include cost shifters such as market wages, product characteristics of other products in a market ("BLP

---

1 Leading examples from IO include measuring market power (Nevo 2001), quantifying the welfare gains from new products (Petrin 2002), or merger evaluation (Nevo 2000). Applications of these methods to fields include measuring media slant (Gentzkow and Shapiro 2010), evaluating trade policy (Berry, Levinsohn, and Pakes 1999), and identifying sorting across neighbourhoods (Bayer, Ferreira, and McMillan 2007).
instruments”), and the price of a given product in other markets (“Hausman instruments”). As with most IV-based identification strategies, there are potential issues with these instruments. Market-level cost shifters such as wages tend to exhibit little variation across firms or over time, which implies they generate little exogenous variation in prices conditional on market fixed effects. Recent work on endogenous product characteristics raises questions about validity of BLP instruments.\(^2\) The use of Hausman instruments is compromised if demand shocks are correlated across markets, perhaps due to spatial correlation or national advertising campaigns. These issues have been the subject of intense debate in the literature (Hausman 1997 and Bresnahan 1997).

Despite these concerns regarding instrument validity, there have been no studies to date with regard to the BLP specification and its reliance on instruments. Indeed, virtually all methodological innovations as well as applications based on the BLP model have, to date, relied on price instruments and the profit maximization assumption to identify the model’s demand and cost parameters.\(^3\) Recent applications have, however, started incorporating cost data to help with model of identification. For instance, Houde (2012) uses wholesale gasoline prices, combined with station-level pricing first order conditions, to identify gasoline stations’ marginal cost functions. to identify the model’s cost parameters. Crawford and Yurukoglu (2012) and Byrne (2014) similarly exploit first order conditions and firm-level cost data to identify cable companies’ cost functions.\(^4\) Like previous research, these papers also use IVs to identify demand in a first step. Given demand estimates, cost parameters are identified using moments based on first order conditions and cost data in a second step.

Motivated by these applications and methodological concerns, we study identification in

\(^2\)See Crawford (2012) for an excellent overview of this burgeoning literature.

\(^3\)There has been some research assessing numerical difficulties with the BLP algorithm (Dube, Fox, and Su 2012, Knittel and Metaxoglou 2012), and the use of optimal instruments to help alleviate these difficulties (Reynart and Verboven 2014). All of these studies also used IVs for identification.

\(^4\)A number of papers have also used demand and cost data to test assumptions regarding conduct in oligopoly models. See, for instance, Byrne (2014), McManus (2007), Clay and Troesken (2003), Kim and Knittel (2003), Wolfram (1999), Genesove and Mullin (1998).
BLP-type models when researchers have access to standard demand-side data on prices, market shares, and product characteristics, and some firm-level cost data. The type of cost data we have in mind is that which comes from firms’ income statements and balance sheets. Such data has been used extensively in a large parallel literature on cost function estimation in empirical IO, where the endogeneity of output to unobserved cost shocks is the central identification problem. Our main finding is that by combining these data, one can jointly identify non-parametric demand and cost functions by using variation in market size (which does not need to be exogenous) and the profit maximization assumption. Neither price nor quantity instruments are needed to correct for the endogeneity of prices and quantities to unobserved demand and cost shocks. The methodology is thus robust to fundamental specification concerns over instrument validity in empirical research on differentiated products and cost function estimation. Our approach is unifying in that it combines data typically used in these respective literatures.

We first propose an identification and estimation strategy that is based on the parametric model of demand, such as the one by Berry (1994) and Berry, Levinsohn, and Pakes (1995). Our identification and estimation strategy combines three ideas. First, we note that because unobserved demand shocks in the BLP model perfectly rationalize the data, the model’s predicted marginal revenue can be written as a function of data and demand parameters only. Second, assuming firms act as differentiated Bertrand price competitors, marginal revenue equals marginal costs in a Nash Equilibrium. Given a non-parametric cost function that is convex in cost shocks, we can exploit these equilibrium conditions to recover each product’s cost shocks as an unspec-

---

5This is not dissimilar to the work of De Loecker (2011) who investigates the usefulness of previously unused demand-side data in identifying production functions and recovering productivity measures.

6Numerous 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 mark-ups for a variety of industries. For identification, researchers often either use IVs for quantities, or need to argue that in the market they study quantities are effectively exogenous from firms’ point of view, and that given this they strictly focus on cost minimization. See Arocena, Saal, and Coelli (2012) for a recent application that makes such an industry-based argument.

7Importantly, we also prove that marginal revenue is uniquely identified in BLP-type models, which is required for our key results to go through.
ified function of data and the demand parameters. That is, assuming profit maximization we can use marginal revenue and cost-related observables to construct a “control function” for cost shocks.

Finally, we use the cost data to estimate a non-parametric cost function, where we control for the cost shock in estimation with our control function. This control function approach to cost estimation is very similar to the approach of Olley and Pakes (1996). They similarly derive a control function based on the first order conditions from underlying economic model and use it to control for unobserved productivity shocks in estimating production functions. It turns out that marginal revenue works as a proper control function only when the demand parameter is at its true value; in that case, the cost function has the best fit in terms of explaining the data. Therefore, we jointly estimate the non-parametric cost function, control function, and hence demand parameters, where the cost function is non-parametrically specified as a sieve function of output, input price and marginal revenue. The corresponding non-linear least squares estimation routine is essentially an application of the sieve estimator from Chen (2007) and Bierens (2013). As we show, the large sample properties of our estimator are similar to these sieve estimators.

We then demonstrate that with the cost data, instrument-free identification does not require any functional form assumptions on the demand side. We prove that the marginal revenue and marginal cost are jointly nonparametrically identified by the sample analog of the first order condition, which chooses the two close points in the data that equates the sample analog of the marginal revenue and the marginal cost.

We further illustrate how our estimator can be adapted to incorporate a number of additional features that will likely arise in practice. These include imposing restrictions to ensure

---

8 In their case, the first order conditions govern firms’ strategic investment decisions. Using investment, capital, and labor cost data, these equations can be inverted to construct control functions for firms’ productivity shocks. Their work spawned an enormous literature on the measurement of firm productivity that has been particularly influential in international trade. See Melitz and Redding (2014) for an overview of this research.
the cost function is properly defined (e.g., homogeneity in input prices)\textsuperscript{9}, allowing for differences between economic and accounting costs, dealing with missing cost data for certain products or firms, multi-product firms, and estimating fixed costs. Through a set of Monte-Carlo experiments we illustrate the ability of our estimator to deliver consistent demand and cost parameter estimators when prices and output are correlated with demand and cost shocks. We also consider environments where demand and cost shocks are correlated, and where standard IV-based approaches break down.\textsuperscript{10}

The only paper that we can find that exploits first order conditions to estimate demand parameters is Smith (2004), who studies multi-store competition and mergers in the UK supermarket industry. He estimates a demand model using consumer-level choice data for within-store products. He does not, however, have product-level price data. To overcome this missing data problem Smith uses data on national price-cost margins for firms in his sample to identify the price coefficient in the demand model that rationalizes the observed mark-ups. Our study differs considerably in that we focus on the more common situation where a researcher has price data and aggregate market shares. Indeed, we directly build on the general BLP framework, whereas Smith develops a clever identification strategy that is unique to his empirical and industrial setting.

This paper is organized as follows. In Section 2, we specify the differentiated products model of interest. We discuss identification of the model in Section 3, propose our semi-parametric sieve estimator, and analyze its asymptotic properties. Section 4 contains a Monte-Carlo study that illustrates the effectiveness of our estimator in environments where standard approaches to demand and cost estimation yield biased results. In Section 5 we conclude, where we discuss the potential for applications of our methodology by academic or policy economists.

\textsuperscript{9}Like the cost-function estimation literature, we check for convexity in input prices after estimation.

\textsuperscript{10}A further result from the experiments speaks to the relative numerical performance of ours and IV-based estimators. Whereas we easily obtain convergence in our estimation routines for most Monte-Carlo samples, like Dube, Fox, and Su (2012) and Knittel and Metaxoglou (2012) we find the BLP algorithm to be quite unstable.
2 Differentiated products models and IV estimation

2.1 Demand estimation

Consider the following standard differentiated products demand model. Consumer \( i \) in market \( m \) gets the following utility from consuming product \( j \):

\[
 u_{ijm} = x'_{jm}\beta + \alpha p_{jm} + \xi_{jm} + \epsilon_{ijm},
\]

(1)

where \( x_j \) is a \( K \times 1 \) vector of observed product characteristics, \( p_{jm} \) is price, \( \xi_{jm} \) is the unobserved product quality (or demand) shock that both firms and consumers account for in making their decisions, and \( \epsilon_{ijm} \) term is an idiosyncratic taste shock. Denote the demand parameter vector by \( \theta = [\beta', \alpha]' \).

Suppose there are \( m = 1 \ldots M \) isolated markets that have respective market sizes \( Q_m \). Each market has \( j = 0 \ldots J_m \) products whose aggregate demand across individuals is

\[
 q_{jm} = s_{jm} Q_m,
\]

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

\[
 s_{jm}(\theta) = s (X_m, p_m, \xi_m, j; \theta) = \frac{\exp \left( x'_{jm}\beta + \alpha p_{jm} + \xi_{jm} \right)}{\sum_{j=0}^{J_m} \exp \left( x'_{jm}\beta + \alpha p_{jm} + \xi_{jm} \right)} = \frac{\exp (\delta_{jm})}{\sum_{j=0}^{J_m} \exp (\delta_{jm})},
\]

(2)

where \( p_m = [p_{0m}, p_{1m}, \ldots, p_{J_m}]' \), a \( (J_m+1) \times 1 \) vector, \( X_m = [x_{0m}, x_{1m}, \ldots, x_{J_m}]' \), a \( (J_m+1) \times K \) matrix, \( \xi_m = [\xi_{0m}, \xi_{1m}, \ldots, \xi_{J_m}]' \), and \( \delta_{jm} = x'_{jm}\beta + \alpha p_{jm} + \xi_{jm} \) is the “mean utility” for product.

\(^{11}\)With panel data, the \( m \) index can correspond to a market-period.
Following standard practice, we label good \( j = 0 \) the “outside good” resulting from not buying one of the \( j = 1, \ldots, J_m \) inside goods. To identify the level and scale of utility in the model, we normalize the outside good’s product characteristics, price and demand shock to \( x_{0m} = 0, p_{0m} = 0, \xi_{0m} = 0 \) for all \( m \).

In the case of BLP, the demand structure is identical except that one allows the price coefficient to be a random variable. That is, consumer \( i \) has price coefficient \( \alpha_i \), where \( \alpha_i \) has a distribution function \( F(\cdot; \theta_\alpha) \), where \( \theta_\alpha \) is the parameter of the distribution.\(^{12}\) Conditional on \( \alpha_i \), the probability consumer \( i \) 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 \( \alpha_i \), i.e.

\[
\begin{align*}
s_{jm}(\theta) &= s(X_m, p_m, \xi_m, j; \theta) = \int_{\alpha_i} \frac{\exp(x'_{jm} \beta + \alpha_i p_{jm} + \xi_{jm})}{\sum_{j=0}^{J_m} \exp(x'_{jm} \beta + \alpha_i p_{jm} + \xi_{jm})} dF(\cdot; \theta_\alpha). \\
\end{align*}
\]

Often the distribution of \( \alpha_i \) is assumed to be normal, implying that \( \theta_\alpha \) consists of the first two moments of the distribution, the mean and the variance, \( \theta_\alpha = [\alpha, \sigma_\alpha]' \), and \( \theta = [\beta', \alpha, \sigma_\alpha]' \).

**Estimation**

Given \( \theta \) and data on market shares, prices and product characteristics, we can solve for the vector \( \delta_m = [\delta_{0m}, \delta_{1m}, \ldots, \delta_{J_m m}]' \) through market share inversion. This involves finding the value of \( \delta_m \) that solves \( s(\delta_m, \theta) - s_m = 0 \), where

\[
\begin{align*}
s(\delta_m, \theta) &= (s(\delta_m(\theta), 0; \theta), \ldots, s(\delta_m(\theta), J_m; \theta))'.
\end{align*}
\]

\(^{12}\)We can also allow for random product characteristics coefficients for each of the \( k = 1 \ldots K \) product characteristics (e.g., allow for \( \beta_{ik} \)'s). The standard parametric assumption on each of these distributions is also i.i.d normal.
i.e.

\[ s(\delta_m(\theta), j; \theta) - s_{jm} = 0, \text{ for } j = 0, \ldots, J_m, \]  

(4)

and therefore,

\[ \delta_m(\theta) = s^{-1}(s_m; \theta), \]  

(5)

where \( s_m = (s_{1m}, \ldots, s_{J_m}m)'. \) That is, we find the vector of mean utilities that perfectly align the model’s predicted market shares to those observed in the data. For example, under the simple logit model we can easily recover the demand shocks for product \( j \) using its market share and the share of the outside good,

\[ \delta_{jm} = \log(s_{jm}) - \log(s_{0m}). \]

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 \( \delta_m \) that solves (4) under some regularity conditions.

Using the inferred values of \( \delta_{jm} \) for all products and markets, we can construct a GMM estimator for \( \theta \) by assuming the following population moment conditions are satisfied at the true value of the demand parameters, \( \theta_0 \)

\[ E[\xi_{jm}(\theta_0)z_{jm}] = 0, \]

where \( \xi_{jm}(\theta) = \delta_{jm} - x'_{jm}\beta - \alpha p_{jm} \) and \( z_{jm} \) is an \( L \times 1 \) vector of instruments. These typically include the market characteristics \( x_{jm} \), market-level demographic variables, and so on. They also include excluded instruments from the demand model that are used to correct for the endogeneity of \( p_{jm} \) to \( \xi_{jm} \), which we discuss in a moment. For estimation, we can construct the
sample analogue to the population moment conditions,

$$\frac{1}{N_{jm}} \sum_{j,m} \xi_{jm}(\theta)z_{jm}$$

where $N_{jm}$ is the number of product-market observations. Using these sample moment conditions, it is straightforward to obtain the GMM estimator of $\theta, \hat{\theta}$.

**Identification**

An endogeneity problem naturally arises in differentiated product markets since firms tend to charge higher prices if their products have higher unobserved product quality. Researchers use a variety of excluded demand instruments to overcome this issue, though each type of instrument has its pitfalls.

Cost shifters are often used as price instruments. This is inline with traditional market equilibrium analysis which identifies the demand curve from shifts in the supply curve caused by cost shifters. Popular examples are input prices, $w_{jm}$. 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 also affect the demand of the product in the same local market through changes in consumer income. Changes in other input prices such as gasoline or electricity could reasonably be expected to affect both firms’ and consumers’ choices. Even further, higher input prices may induce firms to reduce product quality.

In instances where cost shifters are likely to satisfy exclusion restrictions, they are often weak instruments. For example, if one assumes input prices are exogenously determined in some external market (such as the labor market), then all firms will face the same input prices. Therefore, cost shifters may not have sufficient within-market variation across firms to identify the demand parameters.\(^{13}\)

\(^{13}\)These concerns are similar to those from the productivity literature in empirical IO. See, for example, Olley
In the absence of cost shifters, researchers often use product characteristics of rivals’ products or market structure characteristics like the number of firms as price instruments. One naturally would worry, however, just like prices, these variables are endogenous with respect to unobserved demand shocks. Indeed, Crawford and Yurukoglu (2012), Fan (2012), Byrne (2014), and others have documented that product characteristics, like prices, are strategic choices made by firms that depend on demand shocks.

A final commonly-used set of instruments are the prices of product $j$ in markets other than $m$ (Nevo (2001); Hausman (1997)). The identifying assumption is that the demand shock for product $j$ in market $m$ is uncorrelated with product $j$’s prices in other markets. The strength of these instruments come from common cost shocks for product $j$ across markets that creates cross-market correlation in product $j$’s prices. These instruments are invalid, however, if there is spatial correlation in demand shocks across markets. Regional demand shocks, for example, could generate such correlation.\textsuperscript{14} Unfortunately, there is little evidence of whether such cross-market correlation in demand shocks is problematic for identifying demand parameters in practice.

\section{2.2 Supply and cost function estimation}

\textbf{Supply}

The cost of producing $q_{jm}$ units of product $j$ is assumed to be a strictly convex function of output, input price $w_{jm}$, and a cost shock $\omega_{jm}$. That is,

\begin{equation}
C_{jm} = C(q_{jm}, w_{jm}, \omega_{jm}; \tau),
\end{equation}

\textsuperscript{14}Firm, market, and year fixed effects are typically included in the set of instruments when panel data are available. So the exclusion restriction fails if the innovation in the demand shock in period $t$ for product $j$ is correlated across markets.
where $C_{jm}$ is the total cost of producing product $j$ in market $m$, $C(\cdot)$ is a positive, strictly increasing, convex and continuously differentiable function, and $\tau$ is a cost parameter vector. Given this cost function and the demand model above, we can write firm $j$’s profit function as

$$
\pi_{jm} = p_{jm} \times s(X_m, p_m, \xi_m; j; \theta) \times Q_m - C(s(X_m, p_m, \xi_m; j; \theta) \times Q_m, w_{jm}, \omega_{jm}; \tau),
$$

(7)

Keeping with the standard differentiated products model, we assume that firms act as differentiated Bertrand price competitors. Therefore, the optimal price and quantity of product $j$ in market $m$ is determined by the first order condition (F.O.C.) that equates marginal revenue and marginal cost\(^{15}\)

$$
p_{jm} + s_{jm} \left[ \frac{\partial s(X_m, p_m, \xi_m; j; \theta)}{\partial p_{jm}} \right]^{-1} = \frac{\partial C(q_{jm}, w_{jm}, \omega_{jm}; \tau)}{\partial q_{jm}} \frac{\partial C(q_{jm}, w_{jm}, \omega_{jm}; \tau)}{\partial \omega_{jm}},
$$

(8)

where $p_{-jm}$ is a vector that contains the prices of all other firms in market $m$ other than the firm $j$. Notice that the marginal revenue function can be written as

$$
MR_{jm} = MR(X_m, p_m, \xi_m; j; \theta).
$$

It follows from the inversion in (4) and the definition of $\xi_{jm}(\theta) = \delta_{jm} - x'_j \beta - \alpha p_{jm}$, $\xi_m$ is fundamentally a function of $X_m, p_m, s_m$ and $\theta$. Therefore, marginal revenue is also a function of those variables. That is,

$$
MR_{jm} = MR(X_m, p_m, s_m, j; \theta) = MR_{jm}(\theta),
$$

where $MR_{jm}(\theta)$ is shorthand for $MR(X_m, p_m, s_m, j; \theta)$. Below, we will extensively use the first

---

\(^{15}\)Here, we are assuming there is one firm for each product. We will relax this later.
order condition (8), and two important aspects of $MR_{jm}$: it is only a function of data and parameters (and not unobservables), and it depends on market shares and not quantities.

Cost function estimation

Like with demand estimation, there are important endogeneity concerns with standard approaches to cost function estimation. Suppose, for example, that we were to estimate the following log-linearized version of the cost function in equation (6) by OLS:

$$\log(C_{jm}) = \tau_1 q_{jm} + \tau_2 w_{jm} + \omega_{jm}.$$  

In doing so, we would face an endogeneity issue that output $q_{jm}$ would potentially be negatively correlated with the cost shock $\omega_{jm}$. Traditionally, researchers have either ignored this problem or tried to find instruments for quantities. Similar to the inversion procedure in demand, the unobserved cost shock satisfies:

$$C_{jm} = C(q_{jm}, w_{jm}, \omega_{jm}; \tau) \Rightarrow \omega_{jm}(\tau) = C^{-1}(q_{jm}, w_{jm}, C_{jm}; \tau).$$  \hspace{1cm} (9)

So in principle, one can estimate the cost function parameters by using the excluded demand shifters as instruments for quantities. Denote the vector of cost instruments by $\tilde{z}_{jm}$. We can estimate $\tau$ assuming that the following population moments are satisfied at the true value of the cost parameters $\tau_0$:

$$E[\omega_{jm}(\tau_0)\tilde{z}_{jm}] = 0.$$  

The difficulty of this approach is that typically valid instruments such as demand shifters (e.g., market demographics) affect all firms. This implies they often do not generate sufficient within-market across-firm variation in equilibrium prices to identify the cost parameters. A further
difficulty with identification arises from the fact that one often wants to use a higher order polynomial to flexibly estimate the cost function. This implies, however, that potentially implausibly many excluded demand shifters are required to estimate all the parameters in the cost function.

3 Instrument-free identification and estimation

Instead of using instruments, we estimate the demand and cost parameters directly from the first order conditions. That is, we exploit restrictions implied by the oligopolistic firms’ first order condition in (8) to identify the demand and cost function parameters.

In the above Berry (1994) or BLP specification, marginal revenue can be written in the usual form

\[
M R_{jm}(\theta) = p_{jm} + s_{jm} \left[ \frac{\partial s \left( X_m, p_m, \xi_m, j; \theta \right)}{\partial p_{jm}} \right]^{-1}
\]

Notice that marginal revenue is a function of price and market share, which both do not appear in the marginal cost function from equation (8). We also use this for identification. In particular, we use observations that have different market shares in different markets (e.g. \( s_{jm} \neq s_{j'm'} \)), but the same levels of output \( q_{jm} = s_{jm} Q_m = q_{j'm'} = s_{j'm'} Q_{m'} \). It is important to note that at no point do we explicitly or implicit use market size \( Q_m \) as an instrument. We allow for the market size variable to be correlated with demand or cost shocks, but we do not allow it to be a deterministic function of such shocks. It must have variation that is independent to the variation in these shocks. In other words, for the same level of output, differences in market size imply differences in market share. These differences can be used to separately identify the demand and cost function parameters.

The idea that data on total costs can be used to help identify demand parameters is not immediately obvious, and perhaps even counterintuitive. Before developing our general identification results, we first make this idea more concrete through an example based on the simple Berry
(1994) logit model. Suppose that $-\infty < \alpha < 0$, and consider a pair of firms $(Q_m, w_{jm}, s_{jm}, p_{jm})$ and $(Q_{m'}, w_{j'm'}, s_{j'm'}, p_{j'm'})$ whose demand shocks are $\xi_{jm}$, $\xi_{j'm'}$. Under the logit specification we have

$$s_{jm} = \frac{\exp (\alpha p_{jm} + \xi_{jm})}{\sum_{j=0}^{J} \exp (\alpha p_{jm} + \xi_{jm})},$$

$$s_{j'm'} = \frac{\exp (\alpha p_{j'm'} + \xi_{j'm'})}{\sum_{j=0}^{J} \exp (\alpha p_{j'm'} + \xi_{j'm'})}.$$

We maintain that $C(\cdot)$ is positive, strictly increasing, convex and continuously differentiable.

Going forward, we drop $\tau$ since we will treat $C(\cdot)$ as non-parametric. We further assume that the solution of the F.O.C. for profit maximization is at the global optimum.

Suppose this pair of observations satisfies $Q_m \neq Q_{m'}$, $w_{jm} = w_{j'm'}$, $q_{jm} = s_{jm} Q_m = q_{j'm'} = s_{j'm'} Q_{m'}$ and $C_{jm} = C_{j'm'}$. For such a pair we claim that

$$\omega_{jm} = \omega_{j'm'}.$$

Suppose this is not the case, for instance $\omega_{jm} > \omega_{j'm'}$. Then, from the strict monotonicity of the cost function in terms of the cost shock $\omega$

$$C (q_{jm}, w_{jm}, \omega_{jm}) = C (q_{j'm'}, w_{j'm'}, \omega_{jm}) > C (q_{j'm'}, w_{j'm'}, \omega_{j'm'}) ,$$

contradicting $C_{jm} = C_{j'm'}$. Similarly for $\omega_{jm} < \omega_{j'm'}$. Therefore, $\omega_{jm} = \omega_{j'm'}$. As a result, the marginal cost of the two observations is the same. That is,

$$MC (s_{jm} Q_m, w_{jm}, \omega_{jm}) = MC (s_{j'm'} Q_{m'}, w_{j'm'}, \omega_{j'm'}) .$$

Therefore, for those two data points, because marginal revenue equals marginal cost, their

\footnote{We have not included the product characteristics here. Recall that if $x_{jm}$ is uncorrelated with $\xi_{jm}$, then we can use the $E[\xi_{jm}(\theta_0)x_{jm}] = 0$ moment conditions to identify $\beta$.}
marginal revenues must be the same. For instance, in the case of the logit model,

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

Notice that since \( Q_m \neq Q_{m'} \), \( s_{jm} \neq s_{j'm'} \) and thus, for bounded negative \( \alpha \), \( p_{jm} \neq p_{j'm'} \). It then follows that \( \alpha \) is identified from such pair of data points from

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

Of course, in practice one is unlikely to find a pair that exactly satisfies \( Q_m \neq Q_{m'} \), \( w_{jm} = w_{j'm'} \), \( q_{jm} = s_{jm}Q_m = q_{j'm'} = s_{j'm'}Q_{m'} \) and \( C_{jm} = C_{j'm'} \). However, a similar argument can be made for pairs that satisfy the above equality relationship approximately, and such pairs can be found even if market size is correlated with the demand and cost shocks.

The above example highlights the importance of the variation of market size \( Q_m \) for identification. If all the data came from a single market, then \( q_{jm} = q_{j'm'} \) implies \( s_{jm} = s_{j'm'} \), and thus \( \alpha \) could not be identified from (10).

Two issues are likely to arise in practice with this estimation strategy. Firstly, suppose there exists two pairs of firms where \( Q_m \neq Q_{m'} \), \( w_{jm} = w_{j'm'} \), \( q_{jm} = s_{jm}Q_m = q_{j'm'} = s_{j'm'}Q_{m'} \) and \( C_{jm} = C_{j'm'} \) for both pairs but yet provide two very different estimates of \( \alpha \). 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 quite different \( \alpha \) estimates. This issue arises because the specification of the model is too strong. It allows for no discrepancies in the estimate of \( \alpha \) across market pairs, except for that which is due to the within-pair distance of the variables.

Secondly, it is widely accepted that cost data are measured with error.\(^\text{17}\) If measurement

\(^{17}\text{For a discussion of this issue see, for example, Wang (2003)}\)
error is allowed, then one needs to have appropriate instruments to identify the model even if
the endogeneity of output is assumed away.

To handle both issues, we explicitly introduce an additive measurement error in the cost
function. To identify the demand parameters, we impose the restriction that the measurement
error is mean zero i.i.d. across products, markets, and time, and is independent from $q_{jm}$, $w_{jm}$,
$s_{jm}$, $p_{jm}$, the demand shock $\xi_{jm}$ and the cost shock $\omega_{jm}$. Under these assumptions, the observed
cost $C_{jm}$ includes measurement error, which is orthogonal to the true cost $C^d_{jm}$, and the observed
cost, i.e.

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

(11)

3.1 Pseudo cost function

**Definition 1** A pseudo-cost function is defined to be

$$PC(q_{jm}, w_{jm}, MC_{jm}),$$

where $MC_{jm}$ denotes marginal cost for product $j$ in market $m$.

Next, we state and prove a lemma that relates the cost function to the pseudo-cost function.
The lemma shows that given output and input prices, marginal cost, if observable, can be used
as a proxy for the cost shock. Because we assume profit maximization, marginal revenue equals
marginal cost, and thus, when the parameters of the demand functions are at their true values,
marginal revenue is observable from the demand data and can be used as a proxy for the cost
shock.

Before proving the lemma, we formally state two key assumptions of the supply-side model
and marginal cost function:

**Assumption 1** Within each market, each firm takes other firms’ strategies as given, and chooses
prices to equate marginal revenue and marginal cost. It is also assumed that the global optimum for profit maximization is attained at this point.

**Assumption 2** The marginal cost function is strictly increasing and continuous in \( \omega \).

**Lemma 1** Suppose that Assumptions 1, 2 are satisfied. Then,

\[
C(q_{jm}, w_{jm}, \omega_{jm}) = PC(q_{jm}, w_{jm}, MR_{jm}(\theta_0)),
\]

and the pseudo-cost function is increasing in marginal revenue.

**Proof.** First, we show that

\[
C(q_{jm}, w_{jm}, \omega_{jm}) = PC(q_{jm}, w_{jm}, MC_{jm}).
\]

Note that because \( MC \) is an increasing and continuous function of \( \omega_{jm} \) given \( q_{jm} \) and \( w_{jm} \), there exists an inverse function on the domain of \( MC(q_{jm}, w_{jm}, \omega_{jm}) \) such that

\[
\omega_{jm} = \omega(q_{jm}, w_{jm}, MC_{jm}).
\]

This implies that we can use (an unspecified function of) \( q_{jm}, w_{jm} \) and \( MC_{jm} \): \( \omega(q_{jm}, w_{jm}, MC_{jm}) \), to control for \( \omega_{jm} \). Substituting this “control function” for \( \omega_{jm} \) into the cost function, we obtain the pseudo-cost function from Definition 1

\[
C(q_{jm}, w_{jm}, \omega_{jm}) = PC(q_{jm}, w_{jm}, MC_{jm}).
\]

From the F.O.C. we know that marginal revenue must equal marginal cost when the demand parameters are at their true values, \( \theta_0 \). We can therefore substitute \( MR_{jm}(\theta_0) \) in for \( MC_{jm} \) in
the pseudo-cost function
\[
C(q_{jm}, w_{jm}, \omega_{jm}) = PC(q_{jm}, w_{jm}, MC_{jm}) = PC(q_{jm}, w_{jm}, MR_{jm}(\theta_0)).
\]

Finally, because \( \omega_{jm} \) is an increasing function of \( MC_{jm} \) given \( q_{jm}, w_{jm} \), and because \( MR_{jm}(\theta_0) = MC_{jm} \) under the profit maximization assumption at \( \theta_0 \), \( PC \) is also an increasing function of \( MR \).  

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

### 3.2 Proposed estimator

We now present our estimator and prove identification. The estimator selects demand and cost parameters to fit the pseudo-cost function to the cost data using a non-parametric sieve regression estimator (Chen, 2007; Bierens, 2013).

Suppose \( (q_{jm}, w_{jm}, MR_{jm}) \) comes from a compact finite dimensional Euclidean space, \( W \). Then if \( PC(q_{jm}, w_{jm}, MC_{jm}) \) is a continuous function over \( W \), from the Weierstrass Theorem it follows that the function can be approximated arbitrarily well by an infinite sequence of polynomials. That is,

\[
PC(q_{jm}, w_{jm}, MR_{jm}(\theta_0)) = \sum_{l=1}^{\infty} \gamma_l \psi_l(q_{jm}, w_{jm}, MR_{jm}(\theta_0)) \quad \forall \ (q_{jm}, w_{jm}, MR_{jm}(\theta_0)) \in W,
\]

(12)

where \( \psi_1(\cdot), \psi_2(\cdot), \ldots \) are the basis functions for the sieve and \( \gamma_1, \gamma_2, \ldots \) is a sequence of their corresponding coefficients.

Our estimator is derived from the approximation in (12). It is useful to introduce some
additional notation before formally defining it. Let $M$ be the number of markets in our sample. Define the sieve space for a set of $L_M$ basis functions as

$$
H_M = \left\{ h : h = \sum_{l \leq L_M} \gamma_l \psi_l (q_{jm}, w_{jm}, MR_{jm}(\theta)), \| h \|_H < B_M, (q_{jm}, w_{jm}, MR_{jm}) \in \mathcal{W} \right\},
$$

for some bounded constant $B_M > 0$ with $L_M \to \infty$ such that $H_1 \subseteq H_2 \subseteq \ldots \subseteq H_M \subseteq H_{M+1} \subseteq \ldots \subseteq H_{\infty}$, and where the pseudo-norm, $\| h \|_H$, is defined as

$$
\| h \|_H = E_{\mathcal{H}} |h (q_{jm}, w_{jm}, MR_{jm})|.
$$

The subscript in $L_M$ makes explicit the fact that the complexity of the sieve is increasing in the sample’s number of markets. The corresponding space of sieve coefficients is similarly defined as

$$
G_M = \left\{ \gamma : \sum_{l \leq L_M} \gamma_l \psi_l (q_{jm}, w_{jm}, MR_{jm}) \in H_M \right\}.
$$

At the true values of $\theta$ and $\gamma = [\gamma_1, \ldots]'$, the sieve approximation error is minimized. That is

$$
[\theta_0, \gamma_0] = \arg \min_{(\theta, \gamma) \in \Theta \times G} E \left[ C_{jm}^d - h (q_{jm}, w_{jm}, MR_{jm}(\theta)) \right]^2,
$$

where $\Theta$ is the demand parameter space. Our estimator solves the sample analogue to (13), given a sample of $M$ markets

$$
\left[ \hat{\theta}_M, \hat{\gamma}_M \right] = \arg \min_{(\theta, \gamma) \in \Theta \times G} \frac{1}{M} \sum_{m} \sum_{j} \left[ C_{jm}^d - h (q_{jm}, w_{jm}, MR_{jm}(\theta)) \right]^2.
$$

18This is equivalent to stating that the sieve approximation error is minimized at the true values of the demand parameters and sieve basis functions,

$$
\{[\theta_0, h_0] = \arg \min_{(\theta, h) \in \Theta \times H} \left[ |C_{jm} - h (q_{jm}, w_{jm}, MR_{jm}(\theta))| \right]^2.
$$
Given $\theta$, the choice of the compact set $(q_{jm}, w_{jm}, MR_{jm}(\theta))$ could be somewhat arbitrary. Let $\mathcal{W}$ be the compact set of variables $q_{jm}, w_{jm}, MR_{jm}$. We can make the lower bound of $q_{jm}$, $\bar{q}$ arbitrarily close to 0, and the upper bound $\bar{q}$ arbitrarily large. Similarly for the lower and upper bound of $w_{jm}$, $\bar{w}$ and $\bar{w}$. Then, given $(q_{jm}, w_{jm})$ in the compact set, for any $MR_{jm} \in [\underline{MR}, \overline{MR}]$, there exists a cost shock that $MR_{jm} = MC(q_{jm}, w_{jm}, \omega_{jm})$. Furthermore, for any $MR_{jm} \in [\underline{MR}, \overline{MR}]$, there exists $\xi_m$ such that $MR_{jm}(\theta) = MR(X_m, p_m, s_m, \xi_m; \theta)$.

3.2.1 Identification

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

**Assumption 3** The marginal revenue identifies the demand function parameter in the compact set $\mathcal{W}$. That is, if $\theta \neq \theta_0$, then there exists two measurable sets $\tilde{C}, \hat{C}$

\[
\tilde{C} = \{ (q_{jm}, w_{jm}, \tilde{X}_m, \tilde{p}_m, \tilde{s}_m) : (\tilde{X}_m, \tilde{p}_m, \tilde{s}_m) \in \tilde{B}, \forall j, m \}
\]
\[
\hat{C} = \{ (q_{jm}, w_{jm}, \hat{X}_m, \hat{p}_m, \hat{s}_m) : (\hat{X}_m, \hat{p}_m, \hat{s}_m) \in \hat{B}, \forall j, m \},
\]

where $(q_{jm}, w_{jm}) \in \tilde{A}$ and $\tilde{A}$, $\tilde{B}$ and $\hat{B}$ are all measurable sets. Furthermore,

1. $\tilde{B} \cap \hat{B} = \emptyset$;

2. $\text{Prob}(\tilde{C}) > 0$, $\text{Prob}(\hat{C}) > 0$;

3. $(q_{jm}, w_{jm}, \overline{MR}_{jm}(\theta)) \in \mathcal{W}$, $(q_{jm}, w_{jm}, \overline{MR}_{jm}(\theta_0)) \in \mathcal{W}$,

\[
(q_{jm}, w_{jm}, \overline{MR}_{jm}(\theta)) \in \mathcal{W}, \quad (q_{jm}, w_{jm}, \overline{MR}_{jm}(\theta_0)) \in \mathcal{W}.
\]

4. For any $(q_{jm}, w_{jm}, \tilde{X}_m, \tilde{p}_m, \tilde{s}_m) \in \tilde{C}$, there exists $(q_{jm}, w_{jm}, \hat{X}_m, \hat{p}_m, \hat{s}_m) \in \hat{C}$ such that either

\[
\overline{MR}_{jm}(\theta) = \overline{MR}_{jm}(\theta_0) \quad \text{and} \quad \overline{MR}_{jm}(\theta_0) \neq \overline{MR}_{jm}(\theta) .
\]
or

\[ \tilde{MR}_{jm}(\theta) \neq \tilde{MR}_{jm}(\theta_0) \quad \text{and} \quad \tilde{MR}_{jm}(\theta_0) = \tilde{MR}_{jm}(\theta_0). \]

\footnote{As we will see later, this is essentially equivalent to assuming non-linearity of the marginal revenue function from the demand system.}

**Assumption 4** Measurement error \( \eta_{jm} \) is i.i.d distributed with mean 0 and variance \( \sigma^2_{\eta} \), where \( \sigma^2_{\eta} < \infty \). In particular, the measurement error is independent of \((q_{jm}, w_{jm}, X_m, p_m, s_m)\), for all \( j, m \).

**Proposition 1** Suppose Assumptions 1-4 hold. Then, the proposed sieve-pseudo-cost estimator in equation (14) identifies \( \theta_0 \).

**Proof.** Recall from above that for each firm the observed cost is

\[ C_{jm}^d = C_{jm} + \eta_{jm} \]

for firm/product \( j \) in market \( m \), and \( \eta_{jm} \) is the measurement error. Denote the sieve function of \( q_{jm}, w_{jm} \) and \( MR_{jm} \) as

\[ \psi(q_{jm}, w_{jm}, MR_{jm}(\theta), \gamma) = \sum_{l=1}^{\infty} \gamma_l \psi_l(q_{jm}, w_{jm}, MR_{jm}(\theta)). \]
Then,

\[
E[(C_{jm}^{d} - \psi(q_{jm}, w_{jm}, MR_{jm}^{(\theta)}, \gamma)) | q_{jm} = q, w_{jm} = w, MR_{jm}^{(\theta)} = MR] \\
= E[(C_{jm} - \psi(q_{jm}, w_{jm}, MR_{jm}^{(\theta)}, \gamma)) | q_{jm} = q, w_{jm} = w, MR_{jm}^{(\theta)} = MR] \\
+ 2E[(C_{jm} - \psi(q_{jm}, w_{jm}, MR_{jm}^{(\theta)}, \gamma)) \eta | q_{jm} = q, w_{jm} = w, MR_{jm}^{(\theta)} = MR] \\
+ E[\eta^2 | q_{jm} = q, w_{jm} = w, MR_{jm}^{(\theta)} = MR] \\
= E[(C_{jm} - \psi(q_{jm}, w_{jm}, MR_{jm}^{(\theta)}, \gamma)) | q_{jm} = q, w_{jm} = w, MR_{jm}^{(\theta)} = MR] + \sigma_{\eta}^2.
\]

Now, from Lemma 1 and the Weierstrass Theorem, there exists an infinite sequence \( \gamma_0 = \{\gamma_0^l\}_{l=1}^{\infty} \) such that

\[
C_{jm}^{*} = PC(q_{jm}, w_{jm}, MR_{jm}^{(\theta_0)}) = \psi(q_{jm}, w_{jm}, MR_{jm}^{(\theta_0)}, \gamma_0) = \sum_{l=1}^{\infty} \gamma_0^l \psi_l(q_{jm}, w_{jm}, MR_{jm}^{(\theta_0)})
\]

for the compact space \((q_{jm}, w_{jm}, MR(\theta_0)) \in W\). Therefore,

\[
E \left[ \left( C_{jm} - \sum_{l=1}^{\infty} \gamma_0^l \psi_l(q_{jm}, w_{jm}, MR_{jm}^{(\theta_0)}) \right)^2 | q_{jm} = q, w_{jm} = w, MR_{jm}^{(\theta_0)} = MR \right] \\
= 0 + \sigma_{\eta}^2.
\]

Let \( \tilde{\nu} = (q_{jm}, w_{jm}, vec(\tilde{X}_m'), \tilde{p}_m', \tilde{s}_m') \) and \( \tilde{\nu} = (q_{jm}, w_{jm}, vec(\tilde{X}_m'), \tilde{p}_m', \tilde{s}_m') \). Without loss of generality, we omit the market subscript \( m \). From the first case of the assumption on the identification of the marginal revenue, if \( \theta \neq \theta_0 \), \( \tilde{\nu} \in \tilde{C} \), and its corresponding \( \tilde{\nu} \in \tilde{C} \) such that

\[
\tilde{MR}_{jm}(\theta) = MR(\tilde{X}_m, \tilde{p}_m, \tilde{s}_m, j, \theta) = MR(\tilde{X}_m, \tilde{p}_m, \tilde{s}_m, j, \theta) = \tilde{MR}_{jm}(\theta) = MR
\]

\[
\tilde{MR}_{jm}(\theta_0) \neq \tilde{MR}_{jm}(\theta_0).
\]
Hence, for any $\gamma \in \mathcal{G}$,

$$
\sum_{l=1}^{\infty} \gamma_l \psi_l \left( q_{jm}, w_{jm}, \tilde{MR}_{jm}(\theta) \right) = \sum_{l=1}^{\infty} \gamma_l \psi_l \left( q_{jm}, w_{jm}, \tilde{\tilde{MR}}_{jm}(\theta) \right)
$$

but because the true pseudo-cost function is strictly increasing in $\tilde{MR}$ and $\tilde{\tilde{MR}}_{jm}(\theta_0) \neq \tilde{\tilde{MR}}_{jm}(\theta_0)$,

$$
\sum_{l=1}^{\infty} \gamma_0 \psi_l \left( q_{jm}, w_{jm}, \tilde{MR}_{jm}(\theta_0) \right) \neq \sum_{l=1}^{\infty} \gamma_0 \psi_l \left( q_{jm}, w_{jm}, \tilde{\tilde{MR}}(\theta_0) \right).
$$

Therefore, either

$$
\sum_{l=1}^{\infty} \gamma_0 \psi_l \left( q_{jm}, w_{jm}, \tilde{MR}_{jm}(\theta_0) \right) \neq \sum_{l=1}^{\infty} \gamma_l \psi_l \left( q_{jm}, w_{jm}, \tilde{\tilde{MR}}_{jm}(\theta) \right)
$$

or

$$
\sum_{l=1}^{\infty} \gamma_0 \psi_l \left( q_{jm}, w_{jm}, \tilde{MR}_{jm}(\theta_0) \right) \neq \sum_{l=1}^{\infty} \gamma_l \psi_l \left( q_{jm}, w_{jm}, \tilde{\tilde{MR}}_{jm}(\theta) \right)
$$

or both. Similarly, if the second case of the assumption holds, then if $\theta \neq \theta_0$, $\tilde{\nu} \in \tilde{\tilde{C}}$, and its corresponding $\tilde{\tilde{\nu}} \in \tilde{\tilde{C}}$ such that

$$
\tilde{\tilde{MR}}_{jm}(\theta) \equiv MR \left( \tilde{X}_m, \tilde{p}_m, \tilde{s}_m, j, \theta \right) \neq MR \left( \tilde{X}_m, \tilde{\tilde{p}}_m, \tilde{\tilde{s}}_m, j, \theta \right) \equiv \tilde{\tilde{MR}}_{jm}(\theta)
$$

, $\tilde{\tilde{MR}}_{jm}(\theta_0) = \tilde{\tilde{MR}}_{jm}(\theta_0)$.

Hence, for any $\gamma \in \mathcal{G}$,

$$
\sum_{l=1}^{\infty} \gamma_l \psi_l \left( q_{jm}, w_{jm}, \tilde{MR}_{jm}(\theta) \right) \neq \sum_{l=1}^{\infty} \gamma_l \psi_l \left( q_{jm}, w_{jm}, \tilde{\tilde{MR}}_{jm}(\theta) \right)
$$
but because the true pseudo-cost function is $\widetilde{MR}_{jm}(\theta_0) = \widetilde{MR}_{jm}(\theta_0)$,

$$\sum_{l=1}^{\infty} \gamma_l \psi_l(q_{jm}, w_{jm}, \overline{MR}_{jm}(\theta_0)) = \sum_{l=1}^{\infty} \gamma_l \psi_l(q_{jm}, w_{jm}, \overline{MR}(\theta_0)).$$

Therefore, again, either

$$\sum_{l=1}^{\infty} \gamma_l \psi_l(q_{jm}, w_{jm}, \overline{MR}_{jm}(\theta_0)) \neq \sum_{l=1}^{\infty} \gamma_l \psi_l(q_{jm}, w_{jm}, \overline{MR}_{jm}(\theta))$$

or

$$\sum_{l=1}^{\infty} \gamma_l \psi_l(q_{jm}, w_{jm}, \overline{MR}_{jm}(\theta_0)) \neq \sum_{l=1}^{\infty} \gamma_l \psi_l(q_{jm}, w_{jm}, \overline{MR}_{jm}(\theta))$$

or both.

Now, consider the former case. Then, because of continuity of $\sum_{l=1}^{\infty} \gamma_l \psi_l$ and $\sum_{l=1}^{\infty} \gamma_l \psi_l$, there exists a set $\tilde{B}(q_{jm}, w_{jm})$ measurable with respect to the conditional probability $P(\cdot | q_{jm}, w_{jm})$ and $P(\tilde{B}(q_{jm}, w_{jm}) | q_{jm}, w_{jm}) > 0$, where $\tilde{\nu} \in \tilde{B}(q_{jm}, w_{jm})$ and for any element $\tilde{b} \in \tilde{B}(q_{jm}, w_{jm})$

$$\sum_{l=1}^{\infty} \gamma_l \psi_l(q_{jm}, w_{jm}, MR_{jm}(\tilde{b}, \theta_0)) \neq \sum_{l=1}^{\infty} \gamma_l \psi_l(q_{jm}, w_{jm}, MR_{jm}(\tilde{b}, \theta))$$

Similarly for latter case. That is, there exists an open set $\tilde{\tilde{B}}(q_{jm}, w_{jm})$ measurable with respect to the conditional probability $P(\cdot | q_{jm}, w_{jm})$, and $P(\tilde{\tilde{B}}(q_{jm}, w_{jm}) | q_{jm}, w_{jm}) > 0$, where $\tilde{\tilde{\nu}} \in \tilde{\tilde{B}}(q_{jm}, w_{jm})$ and for any element $\tilde{\tilde{b}} \in \tilde{\tilde{B}}(q_{jm}, w_{jm})$

$$\sum_{l=1}^{\infty} \gamma_l \psi_l(q_{jm}, w_{jm}, MR_{jm}(\tilde{\tilde{b}}, \theta_0)) \neq \sum_{l=1}^{\infty} \gamma_l \psi_l(q_{jm}, w_{jm}, MR_{jm}(\tilde{\tilde{b}}, \theta))$$
This implies that

\[
E \left[ \left( \sum_{l=1}^{\infty} \gamma_0 \psi_l \left( q_{jm}, w_{jm}, \overline{MR}_{jm}(\theta_0) \right) \right) 
- \sum_{l=1}^{\infty} \gamma_l \psi_l \left( q_{jm}, w_{jm}, \overline{MR}_{jm}(\theta) \right) \right] ^2 \left| \tilde{B} \left( q_{jm}, w_{jm} \right) \right| > 0,
\]

and

\[
E \left[ \left( \sum_{l=1}^{\infty} \gamma_0 \psi_l \left( q_{jm}, w_{jm}, \overline{MR}_{jm}(\theta_0) \right) \right) 
- \sum_{l=1}^{\infty} \gamma_l \psi_l \left( q_{jm}, w_{jm}, \overline{MR}_{jm}(\theta) \right) \right] ^2 \left| \tilde{B} \left( q_{jm}, w_{jm} \right) \right| > 0,
\]

Therefore, integrating over \( q, w \) and \( MR \), we obtain that for any \( \gamma \) which makes \( PC(\cdot) \) to be strictly increasing or strictly decreasing in \( MR \), for \( \theta \neq \theta_0 \),

\[
E \left[ \left( C_{jm}^* - \sum_{l=1}^{\infty} \gamma_l \psi_l \left( q_{jm}, w_{jm}, MR_{jm}(\theta) \right) \right) \right] ^2
\]

\[
= E \left[ \left( \sum_{l=1}^{\infty} \gamma_0 \psi_l \left( q_{jm}, w_{jm}, MR_{jm}(\theta_0) \right) \right) 
- \sum_{l=1}^{\infty} \gamma_l \psi_l \left( q_{jm}, w_{jm}, MR_{jm}(\theta) \right) \right] ^2 \left| \tilde{B} \left( q_{jm}, w_{jm} \right) \right| dProb \left( \tilde{B} \left( q_{jm}, w_{jm} \right) \right) I \left( \tilde{B} \left( q_{jm}, w_{jm} \right) \neq \emptyset \right)
\]

\[
\geq E_{(q_{jm}, w_{jm})} \left[ E \left[ \left( \sum_{l=1}^{\infty} \gamma_0 \psi_l \left( q_{jm}, w_{jm}, MR_{jm}(\theta_0) \right) \right) 
- \sum_{l=1}^{\infty} \gamma_l \psi_l \left( q_{jm}, w_{jm}, MR_{jm}(\theta) \right) \right] ^2 \left| \tilde{B} \left( q_{jm}, w_{jm} \right) \right| dProb \left( \tilde{B} \left( q_{jm}, w_{jm} \right) \right) I \left( \tilde{B} \left( q_{jm}, w_{jm} \right) \neq \emptyset \right) \right]
\]

\[
+ E_{(q_{jm}, w_{jm})} \left[ E \left[ \left( \sum_{l=1}^{\infty} \gamma_0 \psi_l \left( q_{jm}, w_{jm}, MR_{jm}(\theta_0) \right) \right) 
- \sum_{l=1}^{\infty} \gamma_l \psi_l \left( q_{jm}, w_{jm}, MR_{jm}(\theta) \right) \right] ^2 \left| \tilde{B} \left( q_{jm}, w_{jm} \right) \right| dProb \left( \tilde{B} \left( q_{jm}, w_{jm} \right) \right) I \left( \tilde{B} \left( q_{jm}, w_{jm} \right) \neq \emptyset, \tilde{B} \left( q_{jm}, w_{jm} \right) = \emptyset \right) \right] > 0
\]
Therefore,

\[
E \left[ (C_{jm} - \psi(q_{jm}, w_{jm}, MR_{jm}(\theta), \gamma))^2 \right] \geq \sigma_{\eta}^2,
\]

with equality only holding for \( \theta = \theta_0 \), \( \psi(q_{jm}, w_{jm}, MR_{jm}(\theta_0)) = PC(q_{jm}, w_{jm}, MR_{jm}(\theta_0)) \).

Surprisingly, it turns out that in fitting the pseudo-cost function to the cost data, we identify the demand parameters. In Section 3.3, it is demonstrated that given a nonparametric cost function, the marginal revenue function can be identified nonparametrically. It would be natural to estimate the demand by exploiting the nonparametric identification directly. Actually, we advocate sieve nonlinear least squares ("NLLS") regression estimation, which relies on the functional form of the demand.

Why do we resort to this strategy? There are a few rationales. First, fully nonparametric estimation of the marginal revenue function is infeasible because of the number of regressors that must be included. In short, due to the curse of dimensionality, we are obliged to maintain parametric specification on the demand side.

Second, sieve NLLS estimation using the pseudo-cost function is preferred, because it is built on the assumption that the first order condition \( MR = MC \) holds exactly for each firm. A more straightforward usage of the pairwise identification argument is to construct the following pairwise differenced estimator.

\[
\theta^*_{JM} = \arg\min_{\theta \in \Theta} \sum_{j,m,j',m'(j',m') \neq (j,m)} \left[ \left( C_{jm}^d - C_{j'm'}^d \right)^2 \right] W_h \left( q_{jm}^d - q_{j'm'}^d, w_{jm}^d - w_{j'm'}^d, MR_{jm}(\theta) - MR_{j'm'}(\theta) \right)
\]
where

$$W_h \left( q^d_{jm} - q^d_{j'm'}, w^d_{jm} - w^d_{j'm'}, MR_{jm}(\theta) - MR_{j'm'}(\theta) \right)$$

\[\equiv \frac{K_h q \left( q^d_{jm} - q^d_{j'm'} \right) K_h w \left( w^d_{jm} - w^d_{j'm'} \right) K_{hMR} \left( MR_{jm}(\theta) - MR_{j'm'}(\theta) \right)}{\sum_{k,n} \sum_{k',n':(k',n')\neq(k,n)} K_h q \left( q^d_{kn} - q^d_{k'n'} \right) K_h w \left( w^d_{kn} - w^d_{k'n'} \right) K_{hMR} \left( MR_{kn}(\theta) - MR_{k'n'}(\theta) \right)}\]

By construction, the first order condition is not assumed to hold exactly for any firm. To put it differently, each estimator loses efficiency because the equilibrium condition is not imposed explicitly. The sieve NLLS estimator can be viewed as a constrained estimator with the constraint $MR_{jm} = MC_{jm}$ for each $(j,m)$, which is thought to attain efficiency gain. As we will see later, the sieve NLLS estimator is more flexible in dealing with the data issue such as multi-branch firm, etc. than the pairwise differenced estimator.

In the above approach, we are using a control function approach, where we estimate the pseudo-cost function, using the marginal revenue as a control function for cost shock, only when the demand parameters are at the true value. Then, it turns out that the pseudo-cost function has the best fit to the data.

In the above approach, we can consider the issues of endogeneity in a similar way to the issue of endogeneity in cost function estimation. Recall that in estimating cost functions, an endogeneity issue arises because output $q_{jm}$ is correlated with the cost shock $\omega_{jm}$. There are two potential ways to deal with this endogeneity problem: find an instrument that is orthogonal to $\omega_{jm}$, or a control function approach that finds a variable that is a function of $\omega_{jm}$. With our estimator, the right hand side of (14) is minimized only when the demand parameters are at their true value $\theta_0$ so that the computed marginal revenue equals the true marginal revenue, i.e., the marginal cost, and thus works as a control function for the supply shock $\omega$. If $\theta \neq \theta_0$, then using the false marginal revenue adds noise, which increases the right hand size of the sum of squared residual in (14). In that sense, we are adopting a pseudo-control function approach.
to avoid using instruments. As the above makes clear, a by-product of this approach is we can also obtain the true demand parameter $\theta$.

### 3.2.2 Identification of marginal revenue

It is important to note that Assumption 3 is a high level assumption; it is not necessarily satisfied in all demand models. For example, if the marginal revenue is a linear function of $\theta$, then for any positive constant $a > 0$ if we set $\theta = a\theta_0$, then $MR\left(\tilde{X}_m, \tilde{p}_m, \tilde{s}_m, \theta\right) = MR\left(\tilde{X}_m, \tilde{p}_m, \tilde{s}_m, \theta\right)$ implies

$$MR\left(\tilde{X}_m, \tilde{p}_m, \tilde{s}_m, \theta\right) = aMR\left(\tilde{X}_m, \tilde{p}_m, \tilde{s}_m, \theta_0\right) = \tilde{MR}_{jm}(\theta) = a\tilde{MR}_{jm}(\theta_0).$$

Hence, if the marginal revenue is a linear function of $\theta$ then Assumption 3 is violated. The remaining issue is whether standard differentiated products demand models satisfy Assumption 3. Next, we prove that both logit model and the BLP model of demand satisfy Assumption 3. In proving Assumption 3 for BLP, we need to use specific pairs of prices and market shares to induce a contradiction. Then, the results would look odd if the parameters are such that in those points, the demand curves are upward sloping or the marginal revenue is nonpositive. To avoid this, we will impose an assumption on the parameter space, such that those anomalies do not arise even for large equilibrium prices. That is, we assume that $\eta = \mu/\sigma$ satisfies the below constraint, and will prove identification in the below parameter space.

$$\eta < \frac{1}{2\phi(0)}.$$
Then, for a monopolist with the market share being $s > 1/2$ that is close to $s = 1/2$, we can show that

$$0 < \left[ (\Phi^{-1}(s) - \eta) \phi(\Phi^{-1}(s)) \right]^{-1} s < 1$$

which then results in the demand curve having negative slope and the marginal revenue being positive even for arbitrarily large price.

**Assumption 5** $\eta$ satisfies

$$\eta < -\frac{1}{2\phi(0)}.$$

**Lemma 2** Assumption 3 is satisfied for the logit model. Suppose Assumption 5 is satisfied. Then, Assumption 3 is satisfied for the BLP model of demand.

**Proof.** See Appendix.

### 3.3 Non-parametric identification of marginal revenue function

We thus far have assumed Berry (1994) or BLP (1995) functional forms of demand, from which marginal revenue can be directly recovered. In this section, establish that marginal revenue is in fact non-parametrically identifiable if cost data are available, and that demand and cost parameters can be recovered from non-parametric marginal revenue estimates. In practice, however, identification will be subject to a Curse of Dimensionality. This motivates our use of flexible demand functional forms in deriving an instrument-free identification and estimation strategy that is not subject to such dimensionality problems.

To simplify our discussion, we focus on monopoly markets for now, which allows us to drop the $j$ subscript. Before discussing the nonparametric identification, we add the following Assumption.

**Assumption 6** Marginal revenue function $MR(p, \xi)$ is strictly increasing in price. Further-
more, Suppose that we have two pairs of prices and market shares \((p_1, s_1)\) and \((p_2, s_2)\) such that \(s_1 = s_2\) and \(p_1 > p_2\). Then,

\[ MR_1 > MR_2 \]

where \(MR_i\) is the marginal revenue of the pair \(i\).

**Assumption 7** The market share function \(s(p, \xi)\) is strictly decreasing and continuous in \(p\) and strictly increasing and continuous in \(\xi\). Furthermore,

\[ \lim_{\xi \to -\infty} s(p, \xi) = 0, \quad \lim_{\xi \to \infty} s(p, \xi) = 1 \quad \text{and} \quad \lim_{p \to \infty} s(p, \xi) = 0 \]

**Assumption 2’** The marginal cost function is strictly increasing and continuous in \(\omega\). Furthermore, for any \(q > 0\),

\[ \lim_{\omega \to -\infty} MC(q, w, \omega) = 0, \quad \text{and} \quad \lim_{\omega \to \infty} MC(q, w, \omega) = \infty. \]

Formally, we will prove the following proposition:

**Proposition 2** Suppose Assumptions 1 and 2’ and Assumption 5 are satisfied.

a. Given \((q, w)\), the ordering of the marginal revenue is nonparametrically identified from the cost data.

b. Assume that the share function \(s(x, p, \xi)\) is decreasing in price \(p\), the marginal revenue function \(MR(x, p, \xi)\) to be nondecreasing in price, and the cost function \(C^*(q, w, \omega)\) to be increasing and strictly convex in \(q\). Consider the point \((Q_1, q_1, w, x, p_1, s_1)\) where the demand shock is \(\xi_1\) and the cost shock \(\omega_1\). Suppose we have another point \((Q_2, q_2, w, x, p_2, s_2)\) that has the same demand shock \(\xi_2 = \xi_1\) and cost shock \(\omega_2 = \omega_1\), but has a larger market.
size $Q_2 > Q_1$. It follows that $s_2 < s_1$, $p_2 > p_1$ and $q_1 < q_2$, and

$$p_1 \left[ 1 + \frac{\ln p_2 - \ln p_1}{\ln s_2 - \ln s_1} \right] = \frac{E[C| (q_2, w, x, p_2, s_2)] - E[C| (q_1, w, x, p_1, s_1)]}{q_2 - q_1} + O(|q_2 - q_1|)$$

c. Assume that if we have two close points $(x, s_1, p_1, Q_1)$ and $(x, s_2, p_2, Q_2)$ such that

$s_1 > s_2$, $p_1 < p_2$, $Q_1 < Q_2$ and $s_1 Q_1 < s_2 Q_2$

that satisfies,

$$p_1 \left[ 1 + \frac{\ln p_2 - \ln p_1}{\ln s_2 - \ln s_1} \right] = \frac{E[C| (q_2, w, x, p_2, s_2)] - E[C| (q_1, w, x, p_1, s_1)]}{q_2 - q_1} + O(|q_2 - q_1|)$$

Then, the true marginal cost at point 1 $MC_1^*$ satisfies

$$MC_1^* = \frac{E[C| (q_2, w, x, p_2, s_2)] - E[C| (q_1, w, x, p_1, s_1)]}{q_2 - q_1} + O(|q_2 - q_1|)$$

Proof

Part a. For $(q_m, w_m) = (q_m', w_{m'}) = (q, w)$ it immediately follows from Lemma 1 and the above conditional expectations,

$$MR_m > MR_{m'} \iff E[C| (q, w, x_m, p_m, s_m)] > E[C| (q, w, x_{m'}, p_{m'}, s_{m'})]$$

$$MR_m < MR_{m'} \iff E[C| (q, w, x_m, p_m, s_m)] < E[C| (q, w, x_{m'}, p_{m'}, s_{m'})]$$

and

$$MR_m = MR_{m'} \iff E[C| (q, w, x_m, p_m, s_m)] = E[C| (q, w, x_{m'}, p_{m'}, s_{m'})].$$
Therefore,

\[ MR(x_m, p_m, s_m) = \zeta(q, w, E[C| (q, w, x_m, p_m, s_m)]) , \]

where \( \zeta \) is an increasing and continuous function of the third element. That is, \( E[C| (q, w, x_m, p_m, s_m)] \), is the nonparametric estimator of the relative ranking of the marginal revenue, given \( (q, w) \).

**Part b.** We choose the point \((Q_2, q_2, w, x, p_2, s_2)\) and such that it is close to the first point \((Q_1, q_1, w, x, p_1, s_1)\), and where the demand and cost shocks are the same \((\xi_1 = \xi_2, \omega_1 = \omega_2)\), but \(Q\) is larger for point 2 \(Q_2 > Q_1\). From now on, for notational simplicity, we omit \(x\) and \(w\).

It follows from the strict convexity of the cost function that

\[ MR(p_1, \xi_1) < \frac{\partial C^* (Q_2s_1, \omega_1)}{\partial q} \]  \hspace{1cm} (15)

Furthermore, consider \( \tilde{s} \) such that \( Q_2 \tilde{s} = Q_1 s_1 \). Then, \( \tilde{s} < s_1 \). From Assumption 6, there exists \( \tilde{p} > p_1 \) such that \( \tilde{s} = s(\tilde{p}, \xi_1) \). In that case, since \( MR(, \xi) \) is strictly increasing in \( p \),

\[ MR(\tilde{p}, \xi_1) > \frac{\partial C^* (Q_2 \tilde{s}, \omega_1)}{\partial q} = \frac{\partial C^* (Q_1 s_1, \omega_1)}{\partial q} = MR(p_1, \xi_1) \]  \hspace{1cm} (16)

Then, from equation 15 and 16, and from Intermediate Value Theorem, there exists \( p_2 > p_1 \) and \( s_2 = s(p_2, \xi_1) < s_1 \) such that

\[ MR(p_2, \xi_1) = \frac{\partial C^* (Q_2 s_2, \omega_1)}{\partial q} \]

is satisfied.

It must further follow that \( q_2 \geq q_1 \). To see this, suppose by way of contradiction that \( q_2 < q_1 \).
From the convexity of the cost function we have,

\[ MR(p_2, \xi_1) = \frac{\partial C^*(q_2, \omega_1)}{\partial q} < \frac{\partial C^*(q_1, \omega_1)}{\partial q} = MR(p_1, \xi_1) \]

which leads to a contradiction of the assumption that the marginal revenue function to be nondecreasing with respect to price.

Finally, from the F.O.C. of the profit maximization, at point 1, we have

\[ p_1 \left[ 1 + \left( \frac{\partial \ln s(p_1, \xi_1)}{\partial \ln p} \right)^{-1} \right] = MC_1^* = \frac{\partial C^*(Q_1s_1, \omega_1)}{\partial q} \]

Furthermore,

\[ \frac{\partial C^*(Q_1s_1, \omega_1)}{\partial q} = \frac{C^*(Q_2s_2, \omega_1) - C^*(Q_1s_1, \omega_1)}{Q_2s_2 - Q_1s_1} + O(|Q_2s_2 - Q_1s_1|) \]

and

\[ \left( \frac{\partial \ln s(p_1, \xi_1)}{\partial \ln p} \right)^{-1} = \frac{\ln(p(Q_2, \xi_1, \omega_1)) - \ln(p(Q_1, \xi_1, \omega_1))}{\ln(s(Q_2, \xi_1, \omega_1)) - \ln(s(Q_1, \xi_1, \omega_1))} + O(|Q_2s_2 - Q_1s_1|) \]

Therefore, the claim holds.

**Part c.** Finally, we prove c. Again, for notational simplicity, we omit \( x \) and \( w \). Denote

\[ MC_1^* = \frac{\partial C^*(q_1, \omega_1)}{\partial q} \]

and \( \hat{MC}_1 \) is the estimate of the marginal cost at \((q_1, \omega_1)\). Let \( \xi_1 \) be the unobserved demand quality at point 1 and \( \xi_2 \) be the unobserved demand quality at point 2. Since marginal cost
equals marginal revenue, i.e.

\[ MC^*_1 = MR(p_1, \xi_1) = p_1 \left[ 1 + \frac{\partial \ln s(p_1, \xi_1)}{\partial \ln p} \right]^{-1}. \]

Therefore,

\[ \left( \frac{\partial \ln s(p_1, \xi_1)}{\partial \ln p} \right)^{-1} = \frac{MC^*_1}{p_1} - 1. \]

Notice that \( p \) and \( s \) can also be written as functions of the underlying exogenous variables \( Q, \xi \) and \( \omega \). That is, \( p = p(Q, \xi, \omega) \) and \( s = s(Q, \xi, \omega) \). Then, as we have seen in b, for sufficiently small \( \Delta Q > 0 \), the points \( (p(Q_1 + \Delta Q, \xi_1, \omega_1), s(Q_1 + \Delta Q, \xi_1, \omega_1)) \) satisfy the following equation.

\[
p(Q_1, \xi_1, \omega_1) \left[ 1 + \frac{\ln(p(Q_1 + \Delta Q, \xi_1, \omega_1)) - \ln(p(Q_1, \xi_1, \omega_1))}{\ln(s(Q_1 + \Delta Q, \xi_1, \omega_1)) - \ln(s(Q_1, \xi_1, \omega_1))} \right] = MC^*_1 + O((\Delta Q)).
\]

Hence,

\[
\frac{\ln(p(Q_1 + \Delta Q, \xi_1, \omega_1)) - \ln(p(Q_1, \xi_1, \omega_1))}{\ln(s(Q_1 + \Delta Q, \xi_1, \omega_1)) - \ln(s(Q_1, \xi_1, \omega_1))} = \frac{MC^*_1}{p(Q_1, \xi_1, \omega_1)} - 1 + O((\Delta Q))
\]

Now, assume that the estimated marginal cost, denoted to be \( \hat{MC}_1 \) is less than the true marginal cost, i.e., \( \hat{MC}_1 < MC^*_1 \). Then, consider a vector of price and market share \((\hat{p}, \hat{s})\) such that \( \hat{s} = s(Q_1 + \Delta Q, \xi_1, \omega_1) \) and \( \hat{p} \) satisfies

\[
p(Q_1, \xi_1, \omega_1) \left[ 1 + \frac{\ln(\hat{p}) - \ln(p(Q_1, \xi_1, \omega_1))}{\ln(\hat{s}) - \ln(s(Q_1, \xi_1, \omega_1))} \right] = \hat{MC}_1.
\]

\[
\frac{\ln(\hat{p}) - \ln(p(Q_1, \xi_1, \omega_1))}{\ln(\hat{s}) - \ln(s(Q_1, \xi_1, \omega_1))} = \frac{\hat{MC}_1}{p(Q_1, \xi_1, \omega_1)} - 1 < \frac{MC^*_1}{p(Q_1, \xi_1, \omega_1)} - 1 + O(\Delta Q).
\]
Hence, for sufficiently small $\Delta Q > 0$, we have

\[
\frac{\ln (\hat{p}) - \ln (p(Q_1, \xi_1, \omega_1))}{\ln (\hat{s}) - \ln (s(Q_1, \xi_1, \omega_1))} < \frac{\ln (p(Q_1 + \Delta Q, \xi_1, \omega_1)) - \ln (p(Q_1, \xi_1, \omega_1))}{\ln (s(Q_1 + \Delta Q, \xi_1, \omega_1)) - \ln (s(Q_1, \xi_1, \omega_1))} < 0.
\]

Notice that $\hat{s} = s(Q_1 + \Delta Q, \xi_1, \omega_1)$ and the above equation implies $\hat{p} > p(Q_1 + \Delta Q, \xi_1, \omega_1)$.

We show that there exists such pair $(\hat{s}, \hat{p})$. In other words, there exists $(\xi_2, \omega_2)$ such that $\hat{s} = s(Q_1 + \Delta Q, \xi_2, \omega_2)$ and $\hat{p} = p(Q_1 + \Delta Q, \xi_2, \omega_2)$. First, we show that there exists a taste shock $\xi_2$ that satisfies $\hat{s} = s(\hat{p}, \xi_2)$. $s(p, \xi)$ being a continuous and decreasing function of price and $\hat{p} > p(Q_1 + \Delta Q, \xi_1, \omega_1)$ implies $s(\hat{p}, \xi_1) < s(p(Q_1 + \Delta Q, \xi_1, \omega_1), \xi_1)$. Furthermore, $\lim_{\xi \to \infty} s(\hat{p}, \xi) = 1 > s(p(Q_1 + \Delta Q, \xi_1, \omega_1), \xi_1)$. By Intermediate Value Theorem, there exists such $\xi_2 > \xi_1$. Next, we show that there exists $\omega_2$ that equates marginal revenue to marginal cost. The marginal revenue of the point $(\hat{p}, s(\hat{p}, \xi_2))$ is

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

Then, there exists $\omega_2$ that satisfies

\[
MR(\hat{p}, \xi_2) = MC(\hat{s}(Q_1 + \Delta Q), \omega_2),
\]

because $MC$ is an increasing and continuous function of $\omega$ and $\lim_{\omega \to -\infty} MC(\hat{s}(Q_1 + \Delta Q), \omega) = 0$ and $\lim_{\omega \to \infty} MC(\hat{s}(Q_1 + \Delta Q), \omega) = \infty$. Therefore, again, by Intermediate Value Theorem, we can find such $\omega_2$.

In Figure 1, $(p(Q_1 + \Delta Q, \xi_2, \omega_2), s(Q_1 + \Delta Q, \xi_2, \omega_2))$ is a point on the red line (demand curve with steeper slope $\left[ \frac{\partial \ln s}{\partial \ln p} \right]$ on the upper left side of E (this corresponds to point F in Figure 1). Now, because $s(p(Q_1 + \Delta Q, \xi_2, \omega_2), \xi_2) = s(p(Q_1 + \Delta Q, \xi_1, \omega_1), \xi_1)$ and $p(Q_1 + \Delta Q, \xi_2, \omega_2) > p(Q_1 + \Delta Q, \xi_1, \omega_1)$, from the Assumption 5, we know that the marginal revenue is higher at
such a point (i.e., \( F \)):

\[
MR(p(Q_1 + \Delta Q, \xi_2, \omega_2), \xi_2) > MR(p(Q_1 + \Delta Q, \xi_1, \omega_1), \xi_1)
\]

Furthermore,

\[
s(Q_1 + \Delta Q, \xi_2, \omega_2)(Q_1 + \Delta Q) = s(Q_1 + \Delta Q, \xi_1, \omega_1)(Q_1 + \Delta Q) \equiv q_1 + \Delta q.
\]

Therefore,

\[
MR(p(Q_1 + \Delta Q, \xi_2, \omega_2), \xi_2) = \frac{\partial C(q_1 + \Delta q, \omega_2)}{\partial q} > \frac{\partial C(q_1 + \Delta q, \omega_1)}{\partial q} = MR(p(Q_1 + \Delta Q, \xi_1, \omega_1), \xi_1)
\]

and thus, \( \omega_2 > \omega_1 \).

However, since \( \hat{MC}_1 < MC^*_1 \), if we set

\[
C^*(q_1 + \Delta q, \omega_2) = C^*(q_1, \omega_1) + \hat{MC}_1 \Delta q,
\]

then,

\[
C^*(q_1, \omega_1) + \hat{MC}_1 \Delta q < C^*(q_1, \omega_1) + MC^*_1 \Delta q = C^*(q_1 + \Delta q, \omega_1) + O(\Delta q^2).
\]

Then, for sufficiently small \( \Delta q > 0 \), i.e., for sufficiently small \( \Delta Q > 0 \),

\[
C^*(q_1 + \Delta q, \omega_2) < C^*(q_1 + \Delta q, \omega_1),
\]

therefore, \( \omega_2 < \omega_1 \), which is a contradiction. The proof for the case with the estimated marginal cost, denoted to be \( \hat{MC}_1 \) is greater than the true marginal cost, i.e., \( \hat{MC}_1 > MC_1 \) is similar as well.
What the above Theorem says is that if we find the two nearby points with the same \( x \) and \( w \), and where an approximation to the first order condition using these points is satisfied, then those two points will have similar unobserved demand and cost shocks, and the expected cost difference divided by the difference in output would be a good approximation of the marginal cost. Assuming profit maximization, such a non-parametric marginal cost estimate would also provide a non-parametric estimate of marginal revenue.

Next, we show that the market share function is also identified. WLOG, we start with the vector of market share \( s_{(0)} = s(x, p, \xi) \). Then, consider the new price \( p + \Delta t \). Then,

\[
\hat{s}(x, p + \Delta t, \xi) = s_{(0)} + B(x, p, \xi) \Delta t.
\]

Similarly, for any positive integer \( k > 0 \), denote \( \hat{s}_{(k)} = \hat{s}(x, p + k\Delta t, \xi) \). Then,

\[
\hat{s}(x, p + (k + 1) \Delta t, \xi) = \hat{s}_{(k)} + B(x, p + k\Delta t, \xi) \Delta t.
\]

The same procedure can be applied for any negative integer \( k < 0 \).

In practice, a non-parametric estimator based on parts b. and c. of the identification proof would likely suffer from a Curse of Dimensionality. One would need to obtain a non-parametric estimate of \( E[C|(q, w, x, p, s)] \) using a sieve which involves higher order polynomials in \((q, w, x, p, s)\). Recall it is necessary to control for \( x, p, s \) in fitting the non-parametric cost function to account for the cost shock through our control-function approach to cost function estimation. For many markets of interest, \( x \) will contain a number of product characteristics, further exacerbating the problem. In an oligopoly setting, one would further need to condition on the entire vector of product characteristics, prices, and market shares across firms to control for the cost shock, making the dimensionality problem even worse.
For these reasons, we follow the common practice where researchers use parametric restrictions to reduce the dimensionality of the estimation problem, essentially transforming the non-parametric estimation exercise into a semi-parametric one. In our specific case, we adopt the flexible BLP (1995) random coefficients demand model which essentially allows to construct a single marginal revenue estimate which is a function of $x, p, s$ that we condition on in our pseudo-cost estimation approach to control for cost shock, thereby relaxing the need to condition on the individual variables in $x, p, s$ in cost function estimation. That is, we overcome the dimensionality problem with such a parametric demand model, implying that our instrument-free estimation approach is semi-parametric where the demand function is parametric, but the cost function is non-parametrically specified.

### 3.4 Semi-parametric cost function estimation

After estimating the model’s parameters, we can recover the cost function. We can 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 have the pseudo-cost function from the above estimation procedure, $\hat{PC}(q_{jm}, w_{jm}, MR, \hat{\gamma}_M)$. Then, we can derive the non-parametric pseudo-marginal cost function as follows:

$$\hat{MC}(q_{jm}, w_{jm}, C) = \sum_{jm} MR \left( X_m, p_m, s_m, \hat{\theta}_M \right) K_h \left( q - q_{jm}, w - w_{jm}, C - \hat{PC}(q_{jm}, w_{jm}, MR, \hat{\gamma}_M) \right)$$
Step 2

Start with the data \( \bar{w}, \bar{q}, \) and \( MR. \) Then, define the proxy for the cost shock as \( \omega = \frac{MC(\bar{q}, \bar{w}, MR)}{MR}. \)

The cost for the output \( \bar{q} + \Delta q \) for small \( \Delta q, \bar{w}_m \) that has the same cost shock \( \omega \) is approximately

\[
\hat{C}(\bar{q} + \Delta q, \bar{w}, \omega) = \hat{PC}(\bar{q}, \bar{w}, MR, \gamma) + MR\Delta q.
\]

At iteration \( k, \) given \( \hat{C}_{k-1} = \hat{C}(\bar{q} + (k - 1) \Delta q, \bar{w}, \omega) \)

\[
\hat{C}(\bar{q} + k\Delta q, \bar{w}, \omega) = \hat{C}_{k-1} + MC(\bar{q} + (k - 1) \Delta q, \bar{w}, \hat{C}_{k-1}) \Delta q
\]

for any \( k. \) Then, from Taylor expansion, we know that for any \( k, \)

\[
\hat{C}(\bar{q} + k\Delta q, \bar{w}, \omega) = C(\bar{q} + k\Delta q, \bar{w}, \omega) + O((k\Delta q)^2)
\]

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

Step 3

Next derive the nonparametric estimate of the input demand:

\[
\hat{l}(q_{jm}, w_{jm}, C) = \sum_{jm} l_{jm} K_h \left(q - q_{jm}, w - w_{jm}, C - \hat{PC}(q_{jm}, w_{jm}, MR, \gamma) \right)
\]

Notice that from Shepard’s Lemma,

\[
l = \frac{\partial C(q_{jm}, w_{jm}, \omega)}{\partial w}
\]
Start, as before, with $\bar{q}$, $\bar{w}$, and $\hat{C}_0 = PC(\bar{q}, \bar{w}, \bar{MR})$. Next, we derive the cost for the output $\bar{q}$, $\bar{w} + \Delta w$ for small $\Delta w$ that has the same cost shock $\omega$. It is approximately:

$$\hat{C}_1 = \hat{C}(\bar{q}, \bar{w} + \Delta w, \omega) = \hat{C}_0 + \hat{l}(\bar{q}, \bar{w}, \hat{C}_0) \Delta w.$$ 

At iteration $k$, given $\hat{C}_{k-1} = \hat{C}(\bar{q}, \bar{w} + (k - 1) \Delta w, \omega)$

$$\hat{C}(\bar{q}, \bar{w} + k \Delta w, \omega) = \hat{C}_{k-1} + \hat{l}(\bar{q}, \bar{w} + (k - 1) \Delta w, \hat{C}_{k-1}) \Delta w$$

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

$$\hat{C}(\bar{q}, \bar{w} + k \Delta w, \omega) = C(\bar{q}, \bar{w} + k \Delta w, \omega) + O((k \Delta q)^2)$$

for any $k$.

It is important to notice that the above semiparametric procedure does not impose any constraints on the cost function, and that the demand model we use is the conventional BLP demand. Past literature that estimates cost function, such as Hall (1988), Roeger (1995), and Klette (1999) imposes restrictions on the production function, or on demand that is outside of the model equilibrium. For example, Roeger (1995) assumes a constant returns to scale production function. In Klette (1999), the mark-up is baed on the “conjectured” price elasticity of demand, which is not determined endogenously from the equilibrium of the model.

### 3.5 Further issues

**Cost Function Restrictions.**

In the above analysis, we imposed no assumptions on the nonparametric estimation of the Pseudo-Cost function, except that it is a smooth function of output, input price and the marginal
Hence, the cost function that is recovered will not have the properties, such as homogeneity of degree one in input prices, or convexity, that is commonly assumed in the literature of cost function estimation. Notice that restrictions need to be imposed on the pseudo-cost function that corresponds to the restrictions we want to impose on the cost function. Below, we show how to do this with the restriction that cost functions are homogenous of degree one with respect to input price. Let $w_{jm} = (w_{1,jm}, w_{2,jm}, \ldots, w_{K,jm})$ be the vector of input prices. If the cost function and the marginal cost functions are homegenous of degree 1 with respect to input price, then

$$C(q_{jm}, w_{jm}, \omega_{jm}) = w_{1,jm}c(q_{jm}, \frac{w_{jm}}{w_{1,jm}}, \omega_{jm})$$

and

$$MC(q_{jm}, w_{jm}, \omega_{jm}) = w_{1,jm} \frac{\partial c(q_{jm}, \frac{w_{jm}}{w_{1,jm}}, \omega_{jm})}{\partial q} \equiv w_{1,jm}mc(q_{jm}, \frac{w_{jm}}{w_{1,jm}}, \omega_{jm})$$

where

$$mc(q_{jm}, \frac{w_{jm}}{w_{1,jm}}, \omega_{jm}) = \frac{MR_{jm}(\theta)}{w_{1,jm}}$$

Therefore, the modified estimator that reflects the homogeneity of degree one restriction is the finite sample analog of the following NLLS problem.

$$\left[\begin{array}{c} \theta_0 \\ \gamma_0 \end{array}\right] = \arg\min_{(\theta, \gamma) \in \Theta \times \mathcal{G}} E \left[ \frac{C_{jm} - \sum_l \gamma_l \phi_l \left( q_{jm}, \frac{w_{-1,jm}}{w_{1,jm}}, \frac{MR_{jm}}{w_{1,jm}}(\theta) \right) }{w_{1,jm}} \right]^2.$$  \hspace{1cm} (17)

where $w_{-1,jm} = (w_{2,jm}, \ldots, w_{K,jm})$. That is,

$$\left[\begin{array}{c} \hat{\theta}_M \\ \hat{\gamma}_M \end{array}\right] = \arg\min_{(\theta, \gamma) \in \Theta \times \mathcal{G}} \frac{1}{\sum_m \sum_j m} \left[ \frac{C_{jm} - \sum_l \gamma_l \phi_l \left( q_{jm}, \frac{w_{-1,jm}}{w_{1,jm}}, \frac{MR_{jm}}{w_{1,jm}}(\theta) \right) }{w_{1,jm}} \right]^2.$$  \hspace{1cm} (18)
Economic versus Accounting Cost

The cost data we use is from the accounting statements of the firm. Therefore, it is an accounting variable, which may not necessarily reflect the economic cost that the firm considers in making input and output choices. More concretely, by imposing profit maximization, we may not be appropriately taking into account the opportunity cost of the resources that are used in purchasing the necessary input to produce output. Fortunately, in the accounting statements, we may be able to obtain information on other activities that the firm may be pursuing in addition the production of an output. If the firm is engaged in some financial investments. Then, in their accounting statements, we may find some details on them including the rate of return on their investments. Suppose that the return on a unit of a financial investment is \( r_{jm} \). Then, the opportunity cost of an additional cost of production is \( r_{jm} \), which can be obtained from the accounting data. Then, the firm will produce and sell output until

\[
MR_{jm} (\theta) = MC (q_{jm}, w_{jm}, \omega_{jm}) + r_{jm}
\]

Then, if we substitute this into our estimator, we obtain

\[
\left[ \hat{\theta}_M, \hat{\gamma}_M \right] = \arg \min_{(\theta, \gamma) \in \Theta \times G} \frac{1}{M} \sum_{j,m} \left[ C_{jm} - h (q_{jm}, w_{jm}, MR_{jm} (\theta) - r_{jm}) \right]^2.
\] (19)

That is, as long as we can obtain information about the opportunities that the firm has other than production, then we can incorporate them into our estimator as well. Then, the estimator will not be subject to bias even if the cost we use is the accounting cost that ignores the opportunity cost.
Fixed costs

The above estimator implicitly assumes that the cost data corresponds to variable costs. For the more general case where only total cost is given, our method can still be applied if we impose some additional assumptions. For example, suppose that fixed costs correspond to rental payments, licensee fees, etc., that do not vary with $q$, $w$ and $MR$, but varies with variables in $w_f$ that affect fixed costs. If we let $TC$ be the observed total cost and $C^*_jm$ be the true variable cost, then

$$TC_{jm} = C^*_jm + w_{jm} \beta + \eta$$

The modified estimator would then be:

$$[\theta_M, \gamma_M, \beta_M] = \argmin_{(\theta, \gamma, \beta) \in \Theta \times G \times \mathbb{R}} \sum_{jm} [TC_{jm} - PC(q_{jm}, w_{jm}, MR_{jm}(\theta, \gamma)) - w_{jm} \zeta]^2$$

Missing cost data and multi product firms

So far we have assumed that cost data is available for each firm. In the data, it could very well be the case that we only observe costs for some firms. The solution to this problem turns out to be straightforward. In this case, we estimate the structural parameters using only the F.O.C.’s for the firms that have cost data. Because we do not use any orthogonality conditions with instruments, only choosing firms with cost data will not result in selection bias for the demand parameter estimates. It is important to notice, however, that we still need demand-side data for all firms in the same market, as well as an assumption on market size (which affects the outside good market share), to form the oligopoly equilibrium marginal revenue. Luckily, such demand-side data tend to be available to researchers for many industries.

A more difficult case of unobservable cost would be when firms produce multiple products, but only the total cost across all products is observable in the data. Suppose that each firm
produces $F$ outputs. Then, as long as the numbers of products are not too large (which otherwise introduces a Curse of Dimensionality in estimation), the estimator can be straightforwardly extended as follows.

$$[\theta_M, \gamma_M] = \arg\min_{(\theta, \gamma) \in \Theta \times G_M} \sum_{j m} [C_{j m} - PC(q_{j m, 1:F}, w_{j m}, MR_{j 1:F}(x_{m 1:F}, p_{m 1:F}, s_{m 1:F}, \theta), \gamma)]^2$$

where $q_{m 1:F} = (q_{m 1}, ..., q_{m F})$ is the vector of output of product 1 to product $F$.

To avoid the Curse of Dimensionality, it is preferable that the number of products are small; if the number of products are large, one should consider a different specification of the cost. That would be the case, for example, in the banking industry where a firm potentially has multiple branches and we only observe the total cost for the bank across all of its branches. Such a bank’s cost function could be specified as:

$$C_f = \sum_{j m} C(q_{j m}, w_{j m}, o_{j m}; \tau) I_{jm}(f) + \eta_f = \sum_{j m} PC(q_{j m}, w_{j m}, MR(X_m, p_m, s_m, \theta_0)) I_{jm}(f) + \eta_f$$

where $I_{jm}(f)$ is an indicator function that equals 1 if branch $i$ in market $m$ belongs to firm $f$ and 0 otherwise. $C_f$ is the total cost of the firm that is recorded in the data including all the production operations, and $\eta_f$ is the measurement error for firm $f$’s total cost. Here we assume that only the total cost of the firm is observed with measurement error. We also assume, as before, that the measurement error is i.i.d. distributed. Then, the model’s parameters can be estimated by minimizing the following objective function:

$$[\hat{\theta}_M, \hat{\gamma}_M] = \arg\min_{(\theta, \gamma) \in \Theta \times G_M} \sum_f \left[ C_f - \sum_{j m} \hat{PC}(q_{j m}, w_{j m}, MR_{jm}(\theta), \gamma_M) I_{jm}(f) \right]^2$$
4 Large Sample Properties

As we discussed above, our estimator is a variant of the Sieve estimator. In this section we show the estimator is consistent. Notice that in our sample we have oligopolistic firms in the same market. They therefore may face the same input price, and their unobserved product quality and cost shocks may be correlated: $\text{corr}(\xi_{jm}, \omega_{jm}) \neq 0$. Furthermore, even if the firms in the same market faces different wages which are independent to each other, and even if the demand and supply shocks are independent, because of the strategic interaction, equilibrium prices and output of the firms in the same market are 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. Without loss of generality, we assume that in each market, the number of firms are $J$. Then, we first prove consistency of the estimator where in each market only the $j$th firm is chosen as a sample. Then, we prove consistency of the estimator that involves all firms in the market. Here, in our proof, we follow Bierens (2013) closely. All the assumptions below are the slight modifications of the ones by Bierens (2013) where we changed the signs to use it for minimization of the expected sum of squares of the errors rather than the maximization of the likelihood function.

Let $z_m = (q_{jm}, w_{jm}, C_{jm}, vec(X_m)', vec(p_m)', vec(s_m)')'$ and define

$$f(z_m, \chi) = \left[C_{jm} - \sum_l \gamma_l \psi_l(q_{jm}, w_{jm}, MR(X_m, p_m, s_m, j, \theta))\right]^2,$$

(20)

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

$$\chi_n = \begin{cases} \theta_n & \text{for } n = 1, ..., p, \\ \gamma_{n-p} & \text{for } n \geq p + 1. \end{cases}$$
The parameter space is assumed to be $\Xi \equiv \Theta \times \Gamma (T)$, where $\Theta$ is compact and

$$\Gamma (T) = \left\{ \gamma = \{ \gamma_n \}_{n=1}^{\infty} : \| \gamma \|^2 \leq T^2 \right\},$$

and is endowed with the metric $d(\chi_1, \chi_2) \equiv \| \chi_1 - \chi_2 \|$, where $\| \chi \| = \sqrt{\sum_{k=1}^{\infty} \chi_k^2}$. Define also

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

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

The following assumptions are made:

**Assumption 4.1**

(a) $z_1, z_2, \ldots, z_M$ are i.i.d. with support contained in a bounded open set $V$ of a Euclidean space.

(b) For each $\chi \in \Xi$, $f(z_m, \chi)$ is a Borel measurable real function of $z_m$.

(c) $f(z_m, \chi)$ is a.s. continuous in $\chi \in \Xi$.

(d) There exists a non-negative borel measurable real function $\overline{f(z)}$ such that $E\left[ \overline{f(z_m)} \right] < \infty$ and $f(z_m, \chi) < \overline{f(z)}$ for all $\chi \in \Xi$.

(e) There exists an element $\chi_0 \in \Xi$ such that $Q(\chi) > Q(\chi_0)$ for all $\chi \in \Xi \setminus \{ \chi_0 \}$.

(f) There exists an increasing sequence of compact subspaces $\Xi_k$ in $\Xi$ such that $\chi_0 \in \bigcup_{k=1}^{\infty} \Xi_k = \Xi \subset \Xi$. Furthermore, each sieve space $\Xi_k$ is isomorph to a compact subset of a Euclidean space.

(g) Each sieve space $\Xi_k$ contains an element $\chi_k$ such that $\lim_{M \to \infty} E[f(z_m, \chi)] = E[f(z_m, \chi_0)]$.

(h) The set $\Xi_\infty = \{ \chi \in \Xi : E[f(z_m, \chi)] = \infty \}$ does not contain an open set.

(i) There exists a compact set $\Xi_c$ containing $\chi_0$ such that $Q(\chi_0) < E\left[ \inf_{\chi \in \Xi_c, \chi_0} f(z_m, \chi) \right] < \infty$. 

47
Assumptions (a)-(e) are well established in the literature (see e.g. Bierens, 2013). For example, (d) is satisfied if we set \( f(z) = b < \infty \). (e) follows from the identification of \( \chi_0 \) in Proposition 2. (f) is required in order to make estimation feasible. In particular, since minimising \( \hat{Q}_N = M^{-1} \sum_{m=1}^{M} f(z_m, \chi) \) over \( \Xi \) is not possible given that \( \Xi \) is not even compact, (f) ensures that the minimization problem can be solved in terms of \( \Xi_{k_M} \), i.e.

\[
\hat{\chi}_M = \arg \min_{\chi \in \Xi_{k_M}} \hat{Q}_N (\chi),
\]

where \( k_M \) is an arbitrary sequence of \( M \) that satisfies \( k_M < M \), \( \lim_{M \to \infty} k_M = \infty \). While (g) is easy to verify for (20), this is not the case for (h), which has to be assumed. (i) is satisfied provided that

\[
\lim_{\tau \to \infty} E \left[ \inf_{\chi \in \Xi_{\tau}} f(z_m, \chi) \right] < E \left[ f(z_m, \chi_0) \right],
\]

where \( \Xi_{\tau} = X_{n=1}^{\infty} [-\chi_n \tau, -\chi_n \tau] \), and \( \{\chi_n\}_{n=1}^{\infty} \) satisfies \( \sum_{n=1}^{\infty} \chi_n < \infty \); \( \sup_{n \geq 1} |\chi_{0,n}| / \chi_n \leq 1 \) so long as \( \Xi_c = \Xi_{\tau} \) for \( \tau \) large enough. This seems a plausible condition.

Thus, following Bierens (2013), it is straightforward to show that \( \text{plim}_{M \to \infty} \| \chi_M - \chi_0 \| = 0 \).\(^{19}\)

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

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

for some \( T \) large enough such that \( \gamma_0 \in \Gamma_r (T) \) and associated metric \( \| \gamma_1 - \gamma_2 \|_r = \sum_{n=1}^{\infty} n^r |\gamma_{1,n} - \gamma_{2,n}| \),

\(^{19}\)Notice that this result holds true for any \( j \).
\( \gamma_i = \{ \gamma_{i,n} \}_{n=1}^{\infty} \). Furthermore, the sieve space is replaced by

\[
\Xi_r = \{ \chi = \{ \chi_n \}_{n=1}^{\infty} : \| \chi \|_r < T, T > \| \chi_0 \|_r \};
\]

\[
\Xi_{r,k} = \{ \pi_k \chi : \| \pi_k \chi \|_r < T \}.
\]

The following assumptions are employed:

**Assumption 4.2**

(a) The parameter space \( \Xi \) is defined with a norm \( \| \chi \|_r = \sum_{n=1}^{\infty} n^r |\chi_m| \) and associated metric \( d(\chi_1 - \chi_2) = \| \chi_1 - \chi_2 \|_r \).

(b) The true parameter \( \chi_0 = \{ \chi_{0,n} \}_{n=1}^{\infty} \) satisfies \( \| \chi_0 \|_r < \infty \).

(c) There exists \( k \in \mathbb{N} \) such that for \( k \) large enough \( \chi_{0,k} = \pi_k \chi_0 \in \Xi_{k,\text{Int}} \), where \( \Xi_{k,\text{Int}} \) is the interior of the sieve space \( \Xi_k \).

(d) \( f(Z, \chi) \) is a.s. twice continuously differentiable in an open neighborhood of \( \chi_0 \).

(e) For any subsequence \( k = k_M \) of the sample size \( M \) satisfying \( k_M \to \infty \) as \( M \to \infty \),

\[
\lim_{M \to \infty} \| \hat{\chi}_{k_M} - \chi_0 \|_r = 0.
\]

(b) imposes a boundedness condition on the true parameter values. (c) employs stronger requirements on the parameters being in the interior of the parameter space compared to 4.1.

The differentiability of the objective function in (d) is necessary for the asymptotic distribution of the estimator. (e) is straightforward to show given (a)-(d) and Assumption 4.1. Furthermore, we also assume:

**Assumption 4.3**

(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

$$E \left\| \nabla_j f (Z, \chi_0) - \nabla_j f (Z, \chi_{0,k}) \right\| \leq M_j \left\| \chi_0 - \chi_{0,k} \right\|_{r_0}$$

where $\nabla_j f (Z, \chi_0) = \partial f (Z, \chi_0) / \partial \chi_{0,j}$, $\sum_{k=1}^{\infty} 2^{-j} M_j < \infty$ and the sieve order $k = k_M$ is chosen such that

$$\lim_{M \to \infty} \sqrt{M} \sum_{n=k_M+1}^{\infty} n^{\tau_0} |\chi_{0,n}| = 0.$$

(b) For all $j \in \mathbb{N}$, $E \left[ \nabla_j f (Z, \chi_0) \right] = 0$.

(c) $\sum_{j=1}^{\infty} j 2^{-j} E \left[ (\nabla_j f (Z, \chi_0))^2 \right] < \infty$.

(d) $\sum_{j=1}^{\infty} j 2^{-j} E \left[ (\nabla_j f (Z, \chi_0))^2 \right] < \infty$.

(e) $\sum_{j=1}^{\infty} \sum_{n=1}^{\infty} (jn)^{-2-\omega} E [\left\| \nabla_j f (Z, \chi_0) \right\|] < \infty$.

(f) $\lim_{\epsilon \to 0} \sum_{j=1}^{\infty} \sum_{n=1}^{\infty} (jn)^{-2-\omega} E \left[ \sup_{\|\chi - \chi_0\|_r \leq \epsilon} |\nabla_j f (Z, \chi) - \nabla_j f (Z, \chi_0)| \right] = 0$.

(g) For at least one pair of positive integers $j, n$, $E \left[ \nabla_{j,p+l} f (Z, \chi_0) \right] \neq 0$.

(h) $\text{rank} (B_{k,k}) = k$ for each $k \geq p$, letting

$$B_{k,l} = \begin{bmatrix} E \left[ \nabla_{1,1} f (Z, \chi_0) \right] & \cdots & E \left[ \nabla_{1,n} f (Z, \chi_0) \right] \\
\vdots & \ddots & \vdots \\
E \left[ \nabla_{j,1} f (Z, \chi_0) \right] & \cdots & E \left[ \nabla_{j,n} f (Z, \chi_0) \right] \end{bmatrix}.$$ 

(b) postulates that the F.O.C. holds for the true parameter value, which we know from equation 15 that is satisfied. (c) imposes boundedness for the first-order derivatives. (d)-(f) are necessary in order to extract the parameters of interest via projection residuals. (g)-(h) impose necessary regularity conditions on the second-order derivatives, in fact (g) is already implied by
Then, given the Assumptions 4.1-4.3 we have

$$\sqrt{M} \left( \hat{\theta}_k - \theta_0 \right) \overset{d}{\rightarrow} N_p(0, \Sigma),$$

where $$\Sigma = \lim_{m \to \infty} \Sigma_k$$ with

$$\Sigma_k = \begin{pmatrix} I_p & O_{p,k-p} \end{pmatrix} B_{k,k}^{-1} V_k B_{k,k}^{-1} \begin{pmatrix} I_p \\ O_{k-p,p} \end{pmatrix},$$

and

$$V_k = \text{Var} \left[ (\nabla_1 f(Z, \chi_0), \nabla_2 f(Z, \chi_0), \cdots \nabla_k f(Z, \chi_0) )' \right].$$

### 5 Monte Carlo Experiments

This section presents evidence of the usefulness of our estimator from Monte Carlo experiments. For simplicity, and to demonstrate that the estimation method does not require any instruments, we estimate the following random coefficients logit demand equation:

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

where $$\phi(\mu_{\alpha_i}, \sigma_{\alpha})$$ is the density for a normal distribution with mean $$\mu_{\alpha_i}$$ and standard deviation $$\sigma_{\alpha}$$. In the Monte-Carlo experiment, we assume that each firm is a monopolist, i.e. $$J = 1$$. Consumers have two choices: either purchase the monopoly product $$j = 1$$ or the choose the outside option $$j = 0$$. Hence, without loss of generality, we do not need a market subscript and denote each firm and the market corresponding to it with a subscript $$i$$.  

51
The cost function is specified as:

\[ C(q_{jm}, w_{jm}, \omega_{jm}, \tau) = \left[ \tau_1 q_{jm} + \frac{1}{2} \tau_2 q_{jm}^2 \omega_{jm} \right] w_{jm} \]

Notice that in the above specification, cost function is homogenous of degree 1 in input price \( w_{jm} \). Furthermore, we draw both the demand shock \( \delta \) and the supply shock \( \omega \) from the normal distribution. For the supply shock, \( \omega \sim TN(\mu_\omega, \sigma_\omega) \). In order to guarantee that the marginal cost is positive, we trim the simulated \( \omega \) draws below and above symmetrically at the 2.5 percentiles. We then assume the market size to be distributed as \( Q_m \sim U(Q_L, Q_H) \). To solve for the equilibrium price, quantity, and market share for each monopolist, we use golden section search on price. Finally, we specify the demand shock \( \delta \) so as to allow for correlation between the demand shock with the cost shock and market size. Specifically, we set:

\[ \xi = \delta_1 \hat{\xi} + \delta_2 \hat{\xi} + \delta_3 \hat{\xi} + \delta_4 \Phi \left( \frac{Q - Q_L}{Q_H - Q_L} \right) \]

where \( \Phi \) is the cumulative distribution function of the standard normal distribution, \( \omega_\xi \sim TN(0, 1) \) and \( \delta_1 > 0, \delta_2 > 0, \delta_3 > 0 \). That is, the demand shock is specified as the linear function of cost shock and market size shock, transformed from the uniform to normal distribution, and the independent normally distributed demand shock component.

In the example below, we set \( (\mu_\alpha, \sigma_\alpha) = (2.0, 0.5), (\mu_\xi, \sigma_\xi) = (0.2, 4.0), (r_L, r_H) = (5.0, 20.0) \), and the measurement error of the cost function to be normally distributed \( \eta \sim TN(\mu_\eta, \sigma_\eta^2) \) where \( \mu_\eta = 0.0 \). Finally, we set \( \delta_1 = \delta_2 = \delta_3 = \delta_4 = 0.5 \).
Table 1: Sample Statistics of Simulated Data.

<table>
<thead>
<tr>
<th>variables</th>
<th>Mean</th>
<th>Std. Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price ($p_m$)</td>
<td>1.1064</td>
<td>0.3025</td>
</tr>
<tr>
<td>Output ($q_m$)</td>
<td>8.1728</td>
<td>4.2233</td>
</tr>
<tr>
<td>Quality ($\xi_m$)</td>
<td>1.0056</td>
<td>0.8940</td>
</tr>
<tr>
<td>Market Share ($s_m$)</td>
<td>0.2662</td>
<td>0.1153</td>
</tr>
<tr>
<td>Wage ($w_m$)</td>
<td>1.0012</td>
<td>0.1786</td>
</tr>
<tr>
<td>Cost ($C_m$)</td>
<td>1.9870</td>
<td>1.7151</td>
</tr>
</tbody>
</table>

Measurement error std. deviation: $\sigma_n = 0.0$

The parameter estimates are obtained by the following minimization algorithm

$$\left[ \hat{\theta}_M, \hat{\gamma}_M \right] = \arg \min_{(\theta, \gamma) \in \Theta \times G} \left[ \frac{1}{\sum_m J_m} \sum_{j,m} \left[ C_{jm}^d - h(q_{jm}, w_{jm}, MR_{jm}(\theta)) \right]^2 \
+ \left( \frac{1}{\sum_m J_m} \text{vec}(X)' \text{vec}(\hat{\xi}) \right)' \left( \frac{1}{\sum_m J_m} \text{vec}(X)' \text{vec}(\hat{\xi}) \right) \right].$$

We report the mean, standard error, and the mean squared errors of 100 simulation/estimation replications.

Table 1 presents the sample statistics of the simulated data. The low standard deviation of the quality variable relative to its model standard error is due to the fact that we draw from the truncated normal distribution instead of normal distribution. In Table 2, we present the Monte-Carlo simulation/estimation results for the nonlinear least squares estimation procedure for the random coefficient logit model. We report the average of 100 simulation/estimation replications. First, we show the results where the standard error of the measurement error is 0.0.

From the table, we see that as the sample size increases, the standard deviation and the mean squared error of the parameter estimates decrease. That implies consistency of our estimator, i.e., our estimators converge to the true value in probability as the sample size increases. In fact, it is noteworthy that the mean of the parameter estimates are quite close to the true value, even with sample size as small as 100. Furthermore, since the estimated parameter values are very
Table 2: Sieve Estimator of Random Coefficient Demand Parameters.

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>No. Poly</th>
<th>( \hat{\mu}_\alpha ) Mean</th>
<th>( \hat{\sigma}_\alpha ) Mean</th>
<th>( \hat{\sigma}_\alpha ) Std. Dev</th>
<th>( \hat{\sigma}_\alpha ) MSE</th>
<th>( \hat{\mu}_X ) Mean</th>
<th>( \hat{\sigma}_X ) Mean</th>
<th>( \hat{\mu}_X ) Std. Dev</th>
<th>( \hat{\sigma}_X ) MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>27</td>
<td>-1.9981</td>
<td>0.0268</td>
<td>0.0268</td>
<td>0.4970</td>
<td>0.0394</td>
<td>0.0393</td>
<td>0.4970</td>
<td>0.0393</td>
</tr>
<tr>
<td>200</td>
<td>27</td>
<td>-2.0003</td>
<td>0.0122</td>
<td>0.0122</td>
<td>0.5008</td>
<td>0.0182</td>
<td>0.0181</td>
<td>0.5008</td>
<td>0.0181</td>
</tr>
<tr>
<td>500</td>
<td>64</td>
<td>-1.9888</td>
<td>0.0114</td>
<td>0.0114</td>
<td>0.4986</td>
<td>0.0158</td>
<td>0.0158</td>
<td>0.4986</td>
<td>0.0158</td>
</tr>
<tr>
<td>1000</td>
<td>64</td>
<td>-1.9989</td>
<td>0.0073</td>
<td>0.0073</td>
<td>0.4988</td>
<td>0.0106</td>
<td>0.0106</td>
<td>0.4988</td>
<td>0.0106</td>
</tr>
<tr>
<td>True</td>
<td></td>
<td>-2.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Measurement error std. deviation: \( \sigma_\eta = 0.0 \)

close to the true parameters, the standard errors and the MSE’s are very close to each other.

In Table 3, we report the simulation/estimation results where the measurement error variance is 0.4, which is about 25 percent of the sample mean of the simulated cost. The sample mean of the estimated parameters are still fairly close to the true values, even with the sample size being 100. Furthermore, overall, both the standard deviation and the MSE of the parameter estimates decrease with the sample size. The exception is the case where the sample size is 100. There, the standard deviation and the MSE of the parameter estimates are smaller than those of sample size 200. But then, we excluded one simulation/estimation replication where the inversion algorithm did not converge. Overall, these Monte-Carlo results demonstrate the validity of our approach.

In Table 4, we report simulation/estimation results of the static oligopoly model where in each market there are four firms. It is numerically challenging to compute the Nash Equilibrium of an oligopoly model. In order to avoid any numerical inaccuracies in the simulated sample, instead of solving for the oligopoly equilibrium given the demand and cost shock, we generated the artificial data in reverse. That is, we randomly generated equilibrium prices and unobserved qualities, and then solve for the market share and the cost shock that are consistent with the
Table 3: Sieve Estimator of Random Coefficient Demand Parameters.

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>No. Poly</th>
<th>$\hat{\mu}_c$ Mean</th>
<th>$\hat{\mu}_c$ Std. Dev</th>
<th>$\hat{\mu}_c$ MSE</th>
<th>$\hat{\sigma}_c$ Mean</th>
<th>$\hat{\sigma}_c$ Std. Dev</th>
<th>$\hat{\sigma}_c$ MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>27</td>
<td>-2.3487</td>
<td>0.8642</td>
<td>0.9278</td>
<td>0.6064</td>
<td>0.5035</td>
<td>0.5121</td>
</tr>
<tr>
<td>200</td>
<td>27</td>
<td>-2.2490</td>
<td>1.6247</td>
<td>1.6356</td>
<td>0.5234</td>
<td>0.7646</td>
<td>0.7611</td>
</tr>
<tr>
<td>500</td>
<td>64</td>
<td>-2.0164</td>
<td>0.3320</td>
<td>0.3308</td>
<td>0.4511</td>
<td>0.3922</td>
<td>0.3933</td>
</tr>
<tr>
<td>1000</td>
<td>64</td>
<td>-2.0581</td>
<td>0.2584</td>
<td>0.2636</td>
<td>0.5118</td>
<td>0.3721</td>
<td>0.3704</td>
</tr>
<tr>
<td>True</td>
<td></td>
<td>-2.0000</td>
<td></td>
<td>0.5</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>No. Poly</th>
<th>$\hat{\mu}_X$ Mean</th>
<th>$\hat{\mu}_X$ Std. Dev</th>
<th>$\hat{\mu}_X$ MSE</th>
<th>$\hat{\sigma}_X$ Mean</th>
<th>$\hat{\sigma}_X$ Std. Dev</th>
<th>$\hat{\sigma}_X$ MSE</th>
<th>Obj. Fct.</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>27</td>
<td>1.0840</td>
<td>0.2646</td>
<td>0.2764</td>
<td>0.2656</td>
<td>0.2356</td>
<td>0.2434</td>
<td>8.7998</td>
</tr>
<tr>
<td>200</td>
<td>27</td>
<td>1.0676</td>
<td>0.4798</td>
<td>0.4821</td>
<td>0.2210</td>
<td>0.2975</td>
<td>0.2967</td>
<td>21.0449</td>
</tr>
<tr>
<td>500</td>
<td>64</td>
<td>1.0081</td>
<td>0.1152</td>
<td>0.1149</td>
<td>0.1778</td>
<td>0.1196</td>
<td>0.1210</td>
<td>53.2872</td>
</tr>
<tr>
<td>1000</td>
<td>64</td>
<td>1.0264</td>
<td>0.1210</td>
<td>0.1232</td>
<td>0.1908</td>
<td>0.0750</td>
<td>0.0752</td>
<td>114.8280</td>
</tr>
<tr>
<td>True</td>
<td></td>
<td>1.0000</td>
<td></td>
<td>0.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Measurement error std. deviation: $\sigma_\eta = 0.4$

As before, the demand shock is set to be correlated with the wage. That is, $p_{jm} \sim TN(4.0, 1.0)$, $\xi_{jm} = \xi_{1im} + \delta_1 \omega_{wm}$. where $\xi_{1im} \sim TN(2.0, 1.0)$, $\omega_{wm} \sim TN(0.0, 1.0)$, $\delta_1 = 0.1$, $w_m = 0.2(2.0 + \omega_{wm})$. What we assume here is that the firms in the same market have the same wage. The other cost function parameters, except for those of the cost shock, are the same as before.

In the third row of Table 4, we report the Monte-Carlo results of the simulation where we know the cost of each firm. We generated artificial data for 200 markets, where each market has four firms. As before, the estimated parameters are close to the true values, with standard deviations and mean square errors being relatively small. In the fourth row, we report results where we assume that we only have data on the total sum of the four firms’ costs. Even though we still find the mean of the estimates to be reasonably close to the true values, the standard deviations and the mean square errors are quite large. This is, of course, due to the lack of data for each firm.

Table 5 presents Monte-Carlo results where we estimate the parameters using the standard IV approach. For instruments we use wage and market size. Notice that this is somewhat
Table 4: Sieve Estimator of Random Coefficient Demand Parameters. Oligopoly

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>No. Poly</th>
<th>Mean</th>
<th>Std. Dev</th>
<th>MSE</th>
<th>Mean</th>
<th>Std. Dev</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>200 × 4</td>
<td>11</td>
<td>-1.8949</td>
<td>0.0849</td>
<td>0.1349</td>
<td>0.5321</td>
<td>0.0842</td>
<td>0.0898</td>
</tr>
<tr>
<td>400 × 4</td>
<td>11</td>
<td>-2.2059</td>
<td>1.4209</td>
<td>1.4287</td>
<td>0.5825</td>
<td>0.3624</td>
<td>0.3699</td>
</tr>
<tr>
<td>True</td>
<td>11</td>
<td>-2.0000</td>
<td>0.5</td>
<td></td>
<td></td>
<td>0.5</td>
<td></td>
</tr>
</tbody>
</table>

Measurement error std. deviation: $\sigma_\eta = 0.2$

different from the way Dube et. al. (2009) and others have done to Monte-Carlos to assess BLP models. In their specifications, the instrument is orthogonal to the residual demand $\xi$, and that the market price is a linear function of $\xi$ and the instruments $z$. They correspondingly report IV BLP estimates that are close to the true values.

In our Monte-Carlo experiments, our instruments are the cost shifters, i.e. input price, and the market size. Given those and demand and cost shocks, we explicitly solve for the equilibrium price, market share, and quantity. Therefore, in our case, the impact of the instruments to the endogenous variable is both nonlinear and heterogeneous. This is in stark contrast to the conventional Monte-Carlo design where the impact is linear and homogeneous. As we will see below, the Monte-Carlo performance of the IV estimator is far worse than those in the recent literature, even if we ignore computational performance.

The reasons for the poor performance of the IV estimator could be due to the nonlinear and heterogenous impact of the instruments on the endogenous variable. We have dramatically reduced the variance of the unobserved quality term to $\delta_1 = 0.4$ for the case where there is no source of endogeneity bias, i.e. no correlation between the unobserved quality term and the instruments. The reason for doing so is if we used the original specification, the parameter estimates of the IV would become very unstable and take on extremely large values. In Table 5 we can see that the mean parameter estimate $\hat{\mu}_\alpha$ is below the true value, and the standard deviation parameter estimate $\hat{\sigma}_\alpha$ is above the true value. Those biases decrease as sample size increases from 2000 to 4000, but the change is small. Even with a sample size of 4000, and the
Table 5: IV Estimator for Random Coefficient Demand Parameters.

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>(δ₁, δ₂)</th>
<th>Mean</th>
<th>Std. Dev</th>
<th>MSE</th>
<th>Mean</th>
<th>Std. Dev</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>(0.4,0.0)</td>
<td>-3.1499</td>
<td>2.5415</td>
<td>2.7779</td>
<td>1.1381</td>
<td>1.4936</td>
<td>1.6173</td>
</tr>
<tr>
<td>4000</td>
<td>(0.4,0.0)</td>
<td>-2.8502</td>
<td>1.7806</td>
<td>1.9651</td>
<td>0.9697</td>
<td>1.0993</td>
<td>1.1904</td>
</tr>
<tr>
<td>2000</td>
<td>(0.2,0.2)</td>
<td>0.0753</td>
<td>0.6203</td>
<td>2.1651</td>
<td>4.0419</td>
<td>8.5773</td>
<td>9.2401</td>
</tr>
<tr>
<td>4000</td>
<td>(0.2,0.2)</td>
<td>-0.7591</td>
<td>0.0211</td>
<td>1.2411</td>
<td>0.0562</td>
<td>0.0051</td>
<td>0.4944</td>
</tr>
<tr>
<td>True</td>
<td></td>
<td>-2.0000</td>
<td></td>
<td></td>
<td></td>
<td>0.5</td>
<td></td>
</tr>
</tbody>
</table>

unobserved quality term having small variation, the estimation results are far away from the large sample values.

Next, we report Monte-Carlo results where we allow for positive correlation between the unobserved quality term and the wage. We do so by setting δ₁ = δ₂ = 0.2. From Table 5 we can see that in this case ˆμα is much higher than the true value of −2.0, and that the upward bias is still large at the sample size of 4000.

The positive direction of bias is a reasonable result because the error term is correlated with the instrument. To clarify the issue, consider the following linear model:

\[ y = x\beta + u \]

where \( x \) is the endogenous variable with \( X_{i1} = 1 \) and \( Cov(X_{i2}, u_i) > 0 \), and \( z \) the instrument where \( Z_{i1} = 1 \), where, for simplicity, we assume that both vectors have the same number of columns. Then, the simple IV estimator is

\[ \beta_{IV} = (z'x)^{-1}z'y = \beta + (z'x)^{-1}z'u > \beta \]

If the instruments are positively correlated with the right-hand side variable, because of the positive correlation between the instruments and \( u, \beta_{IV} > \beta \). Our Monte-Carlo setup is similar in that we have positive correlation between price and wage, and positive correlation between wage and the unobserved quality term.
6 Conclusion

This paper has developed a new methodology for estimating demand and cost parameters of differentiated goods oligopoly models when cost data is available. Our integrated approach, which exploits equilibrium predictions from the model is able, without instruments, to identify the demand parameters in the presence of price endogeneity, and the cost parameters in the presence of output endogeneity. Moreover, firms’ cost functions can be non-parametrically estimated with our framework, and we have shown that the market share function is nonparametrically identified as well. To implement the estimator, one needs standard data on aggregate market shares and products’ prices, in addition to data on total costs and input prices. Importantly, as a result of our instrument-free approach to identification, cost data is only needed for some products and firms to consistently estimate the structural parameters.

As our Monte-Carlo experiments show, our method works well in situations where instruments are correlated with structural unobservables in the model, and thus where standard estimation methods break down. Going forward, we will further illustrate the usefulness of our estimator with an application to retail banking, a sector where logit and BLP models have delivered new insights into demand (Dick, 2007; Ho and Ishii 2011), and where the differentiated Bertrand assumption has been used as a competitive benchmark (Jaumandreu and Lorences, 2002).

We believe that our results have implications on the estimation of the models that are related to the optimal behavior of firms. In most literature that uses firms’ profit maximization, researchers have essentially use methods that are similar to the ones used in the estimation of optimal behavior of individuals, such as consumption or labour supply decisions, which essentially are, based on the idea of revealed preference. The assumption behind it is that the researcher cannot observe any data on the objective function of individuals, i.e. their utility.
On the other hand, for firms, we can actually observe a measure of the objective function of the firms, i.e. their revenue and cost, thus, their profit. Our results show that in such case, the exogenous variation that is conventionally used in the nonparametric identification and estimation of structural parameters are not necessary any more, and, as we have seen in the nonparametric identification proof. Hence, a promising direction of future research would be to investigate how far we can go with the non-revealed preference approach in trying to empirically understand further on the optimal behavior of firms.

References


A Identification of marginal revenue

Proof.
It is easy to show that the Berry (1994) logit demand model from (2) satisfies Assumption 3. Suppose that for \( p_{jm}, p_{j'm'} \) and \( s_{jm}, s_{j'm'} \), \( \alpha \neq \alpha_0 \)

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

Then, for \( \alpha \neq \alpha_0 \),

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

and

\[
p_{jm} + \frac{1}{1 - s_{jm}} \alpha_0 \neq p_{j'm'} + \frac{1}{1 - s_{j'm'}} \alpha_0.
\]

Therefore, the price coefficient satisfies Assumption 3. Thus, the price coefficient is identified.

Recall that the above analysis abstracted from having product characteristics, \( x_{jm} \). Without further restrictions \( \beta \) is not identified: once \( \alpha = \alpha_0 \) is identified, then \( A_{jm} = x'_{jm} \beta + \xi_{jm} \) is identified, as we can see from equation (4) and (5), but we cannot separately identify \( x'_{jm} \beta \) and \( \xi_{jm} \). The additional restriction that identifies \( \beta_0 \) could be the following orthogonality condition,

\[
\xi_{jm} = \log(s_{jm}) - \log(s_{0m}) - x'_{jm} \beta - \alpha_0 p_{jm}; \quad E[\xi_{jm}|x_{jm}] = 0
\]

By assuming this orthogonality condition we are following standard practice in the literature (Berry (1994), BLP, and most applications assume the exogeneity of product characteristics). Like with all differentiated products markets studies, the validity of this assumption will depend on the industry context and time horizon of a given study.

Next, we prove that the random coefficient BLP model also satisfies Assumption 3. First, consider the monopoly case. Here, we are primarily interested in the identification of the parameters of the distribution of the price coefficient. To do so, we consider the data where \( x = 0 \). Then, per period log utility component of a purchase is \( u = p_0 + \xi, \) where \( \alpha \sim N(\mu_{\alpha}, \sigma_{\alpha}) \). Consider the pair \((s, p, \xi)\) and \((s', p', \xi')\) that satisfy the share equation. Then,

\[
\int_{\alpha} \frac{\exp(\xi + p_0)}{1 + \exp(\xi + p_0)} \frac{1}{\sigma_{\alpha}} \phi \left( \frac{\alpha - \mu_{\alpha}}{\sigma_{\alpha}} \right) d\alpha = \int_{\alpha} \frac{\exp(p(\alpha + \xi/p))}{1 + \exp(p(\alpha + \xi/p))} \frac{1}{\sigma_{\alpha}} \phi \left( \frac{\alpha - \mu_{\alpha}}{\sigma_{\alpha}} \right) d\alpha = s
\]

and

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

and the marginal revenue equation:

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

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

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

and the marginal revenue equation:

\[
MR = p + p \int_{\tilde{\alpha}} \frac{p\sigma_{\alpha} \exp(p\sigma_{\alpha}(\tilde{\alpha} + a))}{1 + \exp(p\sigma_{\alpha}(\tilde{\alpha} + a))} \frac{1}{\sigma_{\alpha}} \phi(\tilde{\alpha} - \eta) d\tilde{\alpha} = s
\]

with the parameter \( \eta_{0} = \mu_{\alpha}/\sigma_{\alpha}, a_0 = \eta_0/p, a'_0 = \eta_0/p' \) needs to satisfy

\[
\int_{\tilde{\alpha}} \frac{\exp(p\sigma_{\alpha}(\tilde{\alpha} + a_0))}{1 + \exp(p\sigma_{\alpha}(\tilde{\alpha} + a_0))} \phi(\tilde{\alpha} - \eta_0) d\tilde{\alpha} = s, \quad \int_{\tilde{\alpha}} \frac{\exp(p'(\sigma_{\alpha}(\tilde{\alpha} + a'_0))}{1 + \exp(p'(\sigma_{\alpha}(\tilde{\alpha} + a'_0))} \phi(\tilde{\alpha} - \eta_0) d\tilde{\alpha} = s'
\]
Consider first the case \( \eta_0 \neq \eta \). Now, let \( p \) and \( p' \) to be arbitrarily large. Then,
\[
\int_{\tilde{\alpha}} \frac{\exp (p \sigma \alpha (\tilde{\alpha} + a))}{1 + \exp (p \sigma \alpha (\tilde{\alpha} + a))} \tilde{\alpha} \phi (\tilde{\alpha} - \eta) \, d\tilde{\alpha} = \int_{\tilde{\alpha}} I (\tilde{\alpha} \geq -a) \phi (\tilde{\alpha} - \eta) \, d\tilde{\alpha} + O \left( p^{-1} \right)
\]
\[
= \int_{\tilde{\alpha}} I (\tilde{\alpha} \geq -a - \eta) \phi (\tilde{\alpha}) \, d\tilde{\alpha} + O \left( p^{-1} \right) = 1 - \Phi (-a - \eta) + O \left( p^{-1} \right) = \Phi (a + \eta) + O \left( p^{-1} \right) = s.
\]
Similarly,
\[
\Phi (a' + \eta) + O \left( p'^{-1} \right) = s'
\]
and,
\[
\int_{\tilde{\alpha}} \frac{p \sigma \alpha \exp (p \sigma \alpha (\tilde{\alpha} + a))}{1 + \exp (p \sigma \alpha (\tilde{\alpha} + a))} \tilde{\alpha} \phi (\tilde{\alpha} - \eta) \, d\tilde{\alpha} = \int_{\tilde{\alpha}} \Phi (a + \eta) + O \left( p^{-1} \right).
\]
Therefore,
\[
MR = p \left[ 1 - \left( \Phi^{-1} (s) - \eta \right) \phi (\Phi^{-1} (s)) + O \left( p^{-1} \right) \right]^{-1} s
\].
. Then, for \( \eta < 0 \), there exists \( s > 1/2, \Phi^{-1} (s) > 0 \) such that
\[
0 < \left( \Phi^{-1} (s) - \eta \right) \phi (\Phi^{-1} (s)) + O \left( p^{-1} \right) < 1
\].
In that case,
\[
MR (p) = p \left[ 1 - \left( \Phi^{-1} (s) - \eta \right) \phi (\Phi^{-1} (s)) + O \left( p^{-1} \right) \right]^{-1} s > 0.
\]
Therefore, if we pick \( s, s' \) such that the above inequality holds, then by choosing \( P = p / p' \) for large \( p, p' \) that satisfies
\[
MR = p \left[ 1 - \frac{\Phi^{-1} (s) - \eta \phi (\Phi^{-1} (s)) + O \left( p^{-1} \right)}{p \Phi^{-1} (s) - \eta \phi (\Phi^{-1} (s)) + O \left( p^{-1} \right)} \right] = p' \left[ 1 - \frac{\Phi^{-1} (s') - \eta \phi (\Phi^{-1} (s')) + O \left( p'^{-1} \right)}{p \Phi^{-1} (s') - \eta \phi (\Phi^{-1} (s')) + O \left( p'^{-1} \right)} \right].
\]
. i.e.,
\[
MR = p - \frac{ps \phi^{-1} (\Phi^{-1} (s))}{(\Phi^{-1} (s) - \eta \phi (\Phi^{-1} (s)) + O \left( p^{-1} \right))} = p' - \frac{p's' \phi^{-1} (\Phi^{-1} (s'))}{(\Phi^{-1} (s') - \eta \phi (\Phi^{-1} (s')) + O \left( p'^{-1} \right))},
\]
where \( B \equiv \Phi^{-1} (s), B' \equiv \Phi^{-1} (s'), C \equiv s \phi^{-1} (\Phi^{-1} (s)), C' \equiv s' \phi^{-1} (\Phi^{-1} (s')) \).
\[
(B - \eta) (B' - \eta) (p - p') = pCB' - p'C' B - \eta (pC - p'C')
\]
\[
(p - p') \left[ \eta^2 - (B + B') \eta + BB' \right] + \eta (pC - p'C') - pCB' + p'C'B = 0
\]
\[
= (1 - P) \left[ \eta^2 - (B + B') \eta + BB' \right] + \eta (C - PC') - [CB' + PC'B] = 0
\]
which has at most two solutions. Hence, one can choose \( s, s' \) so that by setting \( P = p' / p \) such that one solution is \( \eta \) but the other solution is not \( \eta_0 \).
holds: Suppose that given $a$, $a'$ satisfying

$$
\int_0^\infty \frac{\exp(p\sigma_0 (\alpha + a))}{[1 + \exp(p\sigma_0 (\alpha + a))]^2} \alpha \phi (\alpha - \eta) d\alpha = s
$$

$$
\int_0^\infty \frac{\exp(p'\sigma_0 (\alpha + a'))}{[1 + \exp(p'\sigma_0 (\alpha + a'))]^2} \alpha \phi (\alpha - \eta) d\alpha = s'
$$

, we have

$$
p + p \left[ \int_0^\infty \frac{p\sigma_0 \exp[p\sigma_0 (\alpha + a)]}{[1 + \exp[p\sigma_0 (\alpha + a)]]^2} \alpha \phi (\alpha - \eta) d\alpha \right]^{-1} s
\]

$$
= p' + p' \left[ \int_0^\infty \frac{p'\sigma_0 \exp[p'\sigma_0 (\alpha + a')]}{[1 + \exp[p'\sigma_0 (\alpha + a')]]^2} \alpha \phi (\alpha - \eta) d\alpha \right]^{-1} s'.
$$

Then, given $a_0$, $a_0'$ satisfying

$$
\int_0^\infty \frac{\exp(p\sigma_{a0} (\alpha + a_0))}{[1 + \exp(p\sigma_{a0} (\alpha + a_0))]^2} \alpha \phi (\alpha - \eta) d\alpha = s
$$

$$
\int_0^\infty \frac{\exp(p'\sigma_{a0} (\alpha + a_0'))}{[1 + \exp(p'\sigma_{a0} (\alpha + a_0'))]^2} \alpha \phi (\alpha - \eta) d\alpha = s'
$$

, the following holds as well.

$$
p + p \left[ \int_0^\infty \frac{p\sigma_{a0} \exp[p\sigma_{a0} (\alpha + a_0)]}{[1 + \exp[p\sigma_{a0} (\alpha + a_0)]]^2} \alpha \phi (\alpha - \eta) d\alpha \right]^{-1} s
\]

$$
= p' + p' \left[ \int_0^\infty \frac{p'\sigma_{a0} \exp[p'\sigma_{a0} (\alpha + a_0')]}{[1 + \exp[p'\sigma_{a0} (\alpha + a_0')]]^2} \alpha \phi (\alpha - \eta) d\alpha \right]^{-1} s'.
$$

In that case, if we define $p^{(1)}$ such that $p^{(1)}\sigma_0 = p\sigma_{a0}$, $p^{(1)'}\sigma_0 = p'\sigma_{a0}$ then, $p^{(1)} = (\sigma_{a0}/\sigma_0)p > p$, $p^{(1)'} = (\sigma_{a0}/\sigma_0)p' > p'$ and

$$
\int_0^\infty \frac{\exp(p^{(1)}\sigma_0 (\alpha + a_0))}{[1 + \exp(p^{(1)}\sigma_0 (\alpha + a_0))]^2} \alpha \phi (\alpha - \eta) d\alpha = \int_0^\infty \frac{\exp(p\sigma_{a0} (\alpha + a_0))}{[1 + \exp(p\sigma_{a0} (\alpha + a_0))]^2} \alpha \phi (\alpha - \eta) d\alpha = s
$$

$$
\int_0^\infty \frac{\exp(p^{(1)'}\sigma_0 (\alpha + a_0'))}{[1 + \exp(p^{(1)'}\sigma_0 (\alpha + a_0'))]^2} \alpha \phi (\alpha - \eta) d\alpha = \int_0^\infty \frac{\exp(p'\sigma_{a0} (\alpha + a_0'))}{[1 + \exp(p'\sigma_{a0} (\alpha + a_0'))]^2} \alpha \phi (\alpha - \eta) d\alpha = s'
$$

and

$$
p^{(1)}\sigma_0 + p^{(1)'}\sigma_0 \left[ \int_0^\infty \frac{p^{(1)}\sigma_0 \exp[p^{(1)}\sigma_0 (\alpha + a_0)]}{[1 + \exp[p^{(1)}\sigma_0 (\alpha + a_0)]]^2} \alpha \phi (\alpha - \eta) d\alpha \right]^{-1} s
\]

$$
= p\sigma_{a0} + p\sigma_{a0} \left[ \int_0^\infty \frac{p\sigma_{a0} \exp[p\sigma_{a0} (\alpha + a_0)]}{[1 + \exp[p\sigma_{a0} (\alpha + a_0)]]^2} \alpha \phi (\alpha - \eta) d\alpha \right]^{-1} s
\]

$$
= p'\sigma_{a0} + p'\sigma_{a0} \left[ \int_0^\infty \frac{p'\sigma_{a0} \exp[p'\sigma_{a0} (\alpha + a_0)]}{[1 + \exp[p'\sigma_{a0} (\alpha + a_0)]]^2} \alpha \phi (\alpha - \eta) d\alpha \right]^{-1} s'
\]

$$
= p^{(1)}\sigma_0 + p^{(1)'}\sigma_0 \left[ \int_0^\infty \frac{p^{(1)}\sigma_0 \exp[p^{(1)}\sigma_0 (\alpha + a_0)]}{[1 + \exp[p^{(1)}\sigma_0 (\alpha + a_0)]]^2} \alpha \phi (\alpha - \eta) d\alpha \right]^{-1} s'.
$$

Thus,

$$
p^{(1)} + p^{(1)} \left[ \int_0^\infty \frac{p^{(1)}\sigma_0 \exp[p^{(1)}\sigma_0 (\alpha + a_0)]}{[1 + \exp[p^{(1)}\sigma_0 (\alpha + a_0)]]^2} \alpha \phi (\alpha - \eta) d\alpha \right]^{-1} s
\]

$$
= p^{(1)'} + p^{(1)'} \left[ \int_0^\infty \frac{p^{(1)'}\sigma_0 \exp[p^{(1)'}\sigma_0 (\alpha + a_0)]}{[1 + \exp[p^{(1)'}\sigma_0 (\alpha + a_0)]]^2} \alpha \phi (\alpha - \eta) d\alpha \right]^{-1} s'.
$$
Therefore, the following equation also holds:

\[
p^{(1)} + p^{(1)} = p^{(1)} + p^{(1)} \left[ \frac{\int p^{(1)} \sigma_a \exp \left[ \frac{p^{(1)} \sigma_a (\alpha + a)}{1 + \exp [p^{(1)} \sigma_a (\alpha + a)]]} \right]}{1 + \exp [p^{(1)} \sigma_a (\alpha + a)]]} \right]^{-1} s
\]

\[
= p^{(1)} + p^{(1)} \left[ \frac{\int p^{(1)} \sigma_a \exp \left[ \frac{p^{(1)} \sigma_a (\alpha + a)}{1 + \exp [p^{(1)} \sigma_a (\alpha + a)]]} \right]}{1 + \exp [p^{(1)} \sigma_a (\alpha + a)]]} \right]^{-1} s'
\]

where \(a_0, a'_0\) satisfy

\[
\int \frac{\exp \left( p^{(1)} \sigma_a (\alpha + a) \right)}{1 + \exp \left[ p^{(1)} \sigma_a (\alpha + a) \right]} \phi (\alpha - \eta) d\alpha = s
\]

\[
\int \frac{\exp \left( p^{(1)} \sigma_a (\alpha + a'_0) \right)}{1 + \exp \left[ p^{(1)} \sigma_a (\alpha + a'_0) \right]} \phi (\alpha - \eta) d\alpha = s'
\]

Similarly, we generate an increasing sequence of prices \((p^{(0)}, p^{(0')}, (p^{(1)}, p^{(1')}, \ldots, (p^{(k)}, p^{(k')}, \ldots)\) such that \((p^{(0)}, p^{(0')}) = (p, p'), (p^{(k)}) = (\sigma_{a_0}/\sigma_a)^k p, and for any integer \(k \geq 1\) such that

\[
p^{(k)} + p^{(k)} = p^{(k)} + p^{(k)} \left[ \frac{\int p^{(k)} \sigma_a \exp \left[ \frac{p^{(k)} \sigma_a (\alpha + a^{(k)})}{1 + \exp [p^{(k)} \sigma_a (\alpha + a^{(k)})]]} \right]}{1 + \exp [p^{(k)} \sigma_a (\alpha + a^{(k)})]]} \right]^{-1} s
\]

\[
= p^{(k')} + p^{(k')} \left[ \frac{\int p^{(k')} \sigma_a \exp \left[ \frac{p^{(k')} \sigma_a (\alpha + a^{(k)})}{1 + \exp [p^{(k')} \sigma_a (\alpha + a^{(k)})]]} \right]}{1 + \exp [p^{(k')} \sigma_a (\alpha + a^{(k)})]]} \right]^{-1} s'
\]

where \(a^{(k)}, a^{(k')}\) satisfy

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

\[
\int \frac{\exp \left( p^{(k')} \sigma_a (\alpha + a^{(k)}) \right)}{1 + \exp \left[ p^{(k')} \sigma_a (\alpha + a^{(k)}) \right]} \phi (\alpha - \eta) d\alpha = s'
\]

and \(a_0^{(k)}, a_0^{(k')}\) satisfy

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

\[
\int \frac{\exp \left( p^{(k')} \sigma_{a_0} (\alpha + a_0^{(k)}) \right)}{1 + \exp \left[ p^{(k')} \sigma_{a_0} (\alpha + a_0^{(k)}) \right]} \phi (\alpha - \eta) d\alpha = s'
\]
and
\[
p^{(k+1)}\sigma_\alpha + p^{(k+1)}\sigma_\alpha \left[ \int_\alpha p^{(k+1)}\sigma_\alpha \exp \left[ \frac{p^{(k+1)}\sigma_\alpha (\alpha + a^{(k+1)})}{1 + \exp \left[ p^{(k+1)}\sigma_\alpha (\alpha + a^{(k+1)}) \right]^2} \right] \alpha \phi (\alpha - \eta) \, d\alpha \right]^{-1} s
\]
\[
= p^{(k)}\sigma_\alpha 0 + p^{(k)}\sigma_\alpha 0 \left[ \int_\alpha p^{(k)}\sigma_\alpha 0 \exp \left[ \frac{p^{(k)}\sigma_\alpha 0 (\alpha + a^{(k)})}{1 + \exp \left[ p^{(k)}\sigma_\alpha 0 (\alpha + a^{(k)}) \right]^2} \right] \alpha \phi (\alpha - \eta) \, d\alpha \right]^{-1} s'
\]
\[
= p^{(k+1)}\prime_\alpha + p^{(k+1)}\prime_\alpha \left[ \int_\alpha p^{(k+1)}\prime_\alpha \exp \left[ \frac{p^{(k+1)}\prime_\alpha (\alpha + a^{(k+1)\prime})}{1 + \exp \left[ p^{(k+1)}\prime_\alpha (\alpha + a^{(k+1)\prime}) \right]^2} \right] \alpha \phi (\alpha - \eta) \, d\alpha \right]^{-1} s'
\]
Then,
\[
\frac{p^\prime}{p} = \frac{1 + \left[ (\eta - \Phi^{-1}(s')) \Phi^{-1}(s') \right]^{-1} s}{1 + \left[ (\eta - \Phi^{-1}(s')) \Phi^{-1}(s') \right]^{-1} s'} = \frac{1 + \left[ \int_\alpha p^{(k)}\sigma_\alpha \exp \left[ \frac{p^{(k)}\sigma_\alpha (\alpha + a)}{1 + \exp \left[ p^{(k)}\sigma_\alpha (\alpha + a) \right]^2} \right] \alpha \phi (\alpha - \eta) \, d\alpha \right]^{-1} s}{1 + \left[ \int_\alpha p^{(k)}\sigma_\alpha \exp \left[ \frac{p^{(k)}\sigma_\alpha (\alpha + a')}{1 + \exp \left[ p^{(k)}\sigma_\alpha (\alpha + a') \right]^2} \right] \alpha \phi (\alpha - \eta) \, d\alpha \right]^{-1} s'}
\]

Assumption 3 also implies that if \(\sigma_{\alpha_0}\) as for \(\sigma_\alpha\). That is,
\[
p + p \left[ \int_\alpha p\sigma_\alpha \exp \left[ \frac{p\sigma_\alpha (\alpha + a_0)}{1 + \exp \left[ p\sigma_\alpha (\alpha + a_0) \right]^2} \right] \alpha \phi (\alpha - \eta) \, d\alpha \right]^{-1} s
\]
\[
= p^\prime + p^\prime \left[ \int_\alpha p^\prime \sigma_\alpha 0 \exp \left[ \frac{p^\prime \sigma_\alpha 0 (\alpha + a'_0)}{1 + \exp \left[ p^\prime \sigma_\alpha 0 (\alpha + a'_0) \right]^2} \right] \alpha \phi (\alpha - \eta) \, d\alpha \right]^{-1} s'
\]

Then,
\[
p + p \left[ \int_\alpha p\sigma_\alpha \exp \left[ \frac{p\sigma_\alpha (\alpha + a)}{1 + \exp \left[ p\sigma_\alpha (\alpha + a) \right]^2} \right] \alpha \phi (\alpha - \eta) \, d\alpha \right]^{-1} s
\]
\[
= p^\prime + p^\prime \left[ \int_\alpha p^\prime \sigma_\alpha \exp \left[ \frac{p^\prime \sigma_\alpha (\alpha + a')}{1 + \exp \left[ p^\prime \sigma_\alpha (\alpha + a') \right]^2} \right] \alpha \phi (\alpha - \eta) \, d\alpha \right]^{-1} s'
\]

If we let \(p^{(0)}_1 = p\), \(p^{(0)\prime}_1 = p'\), we can derive the sequence \(p^{(1)}_1\sigma_\alpha = p^{(0)}_1\sigma_\alpha\), \(p^{(1)\prime}_1\sigma_\alpha = p^{(0)\prime}_1\sigma_\alpha\), then \(p^{(1)}_1 < p^{(0)}_1\) and
\( p_1^{(1)} < p_1^{(0)} \), thus, similarly, we can obtain the sequence \( p_1^{(k+1)} = (\sigma_\alpha/\sigma_{\alpha 0}) p_1^{(k)} < p_1^{(k)} \) and

\[
p_1^{(k)} + p_1^{(k)} \left[ \int_\alpha \frac{p_1^{(k)} \sigma_{\alpha 0} \exp \left[ p_1^{(k)} \sigma_{\alpha 0} (\alpha + a^{(k)}_{a 01}) \right]}{1 + \exp \left[ p_1^{(k)} \sigma_{\alpha 0} (\alpha + a^{(k)}_{a 01}) \right]^2} \alpha \phi (\alpha - \eta) \, d\alpha \right]^{-1} s
\]

implies,

\[
p_1^{(k)} + p_1^{(k)} \left[ \int_\alpha \frac{p_1^{(k)} \sigma_{\alpha 0} \exp \left[ p_1^{(k)} \sigma_{\alpha 0} (\alpha + a^{(k)}_{a 11}) \right]}{1 + \exp \left[ p_1^{(k)} \sigma_{\alpha 0} (\alpha + a^{(k)}_{a 11}) \right]^2} \alpha \phi (\alpha - \eta) \, d\alpha \right]^{-1} s'
\]

Together, we obtain

\[
G \equiv \frac{p'}{p} = \frac{1 + (\eta \phi (0))^{-1} s}{1 + [(\eta - \Phi^{-1} (s')) \phi (\Phi^{-1} (s'))]^{-1} s'} = \lim_{k \to \infty} \frac{p_1^{(k)} \alpha \phi (\alpha - \eta) \, d\alpha}{p_1^{(k)} + \left[ \int_\alpha \frac{\sigma_{\alpha 0} \exp \left[ p_1^{(k)} \sigma_{\alpha 0} (\alpha + a^{(k)}_{a 01}) \right]}{1 + \exp \left[ p_1^{(k)} \sigma_{\alpha 0} (\alpha + a^{(k)}_{a 01}) \right]^2} \alpha \phi (\alpha - \eta) \, d\alpha \right]^{-1} s'}
\]

Because \( s \neq s' \), the equality does not hold. Thus, the claim holds. Now, consider the case where the market share of other firms, \( s_{-i} = (s_1, s_2, ..., s_{i-1}, s_{i+1}, ..., s_J) \) are all given. Then, the proof is just a minor modification of the earlier proof of the monopolist.