Byrne, D. P., Imai, S., Jain, N., & Sarafidis, V. (2022). Instrument-free identification and estimation of differentiated products models using cost data. Journal of Econometrics, 228(2), 278-301.
https://doi.org/10.1016/j.jeconom.2021.12.006

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

David P. Byrne†

Susumu Imai‡

Neelam Jain§

Vasilis Sarafidis ¶

Abstract

We propose a new methodology for estimating demand and cost functions of differentiated products models when demand and cost data are available. The method deals with the endogeneity of prices to demand shocks and the endogeneity of outputs to cost shocks by using cost data. We establish identification, consistency and asymptotic normality of our two-step Sieve Nonlinear Least Squares (SNLLS) estimator for the commonly used logit and BLP demand function specification. Using Monte-Carlo experiments, we show that our method works well in contexts where commonly used instruments are correlated with demand and cost shocks and thus biased. We also apply our method to the estimation of deposit demand in the US banking industry.

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

JEL Codes: C13, C14, L13, L41

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

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

‡Corresponding author, Department of Economics and Business, Hokkaido University, susimaig@gmail.com

§Corresponding author, Department of Economics, City, University of London, Neelam.Jain.1@city.ac.uk

¶Department of Econometrics and Business Statistics, Monash University, vasilis.sarafidis@monash.edu

‖Department. of Economics, Setsunan University, hirukawa@econ.setsunan.ac.jp

1 Introduction

In this paper, we develop a new methodology for estimating models of differentiated products markets. Our approach requires commonly used demand-side data on products' prices, market shares, observed characteristics, and some firm-level cost data. The novelty of our method is that it does not use instrumental variables strategies to deal with the endogeneity of prices to demand shocks in estimating demand, nor the endogeneity of outputs to cost shocks in estimating cost functions. Instead, we use cost data for identification and estimation. Market-level demand and cost data tend to be available for large industries that are subject to regulatory oversight (which often requires firms to report cost data); examples include banking, telecommunications, and nursing home care. Such major sectors of the economy represent natural settings for the application of our estimator.

Recently, in order to incorporate the rich heterogeneities of agents, aggregate demand or market share equations are becoming more complex and nonlinear, and thus, more instruments and their interactions are needed for the identification of parameters. It is, in general, a challenge to convincingly argue that all these instruments are valid. We show in our Monte-Carlo exercises and an empirical example, how a small subset of invalid instruments can greatly bias parameter estimates in unanticipated directions in nonlinear demand models. In contrast, we find that the cost data based approach that we propose tends to deliver consistent parameter estimates in Monte-Carlo examples and reasonable parameter estimates in the empirical example. By comparing the IV-based parameter estimates and those obtained by the cost-based approach, in particular, finding instruments that provide parameter estimates similar to the cost-based ones, our method can provide additional support for the chosen instruments.

The frameworks of interest for this paper are the logit and random coefficient logit models of Berry (1994) and Berry et al. (1995) (hereafter, BLP). They have had a substantial impact on empirical research in IO and various other areas of economics.[1] These models incorporate unobserved heterogeneity in product quality and use instruments to deal with the endogeneity of prices to such heterogeneity. As Berry and Haile (2014) and others point out, as long as there are instruments available, demand functions can be identified using market-level data. Popular instruments include cost shifters such as market wages, product characteristics of other products in a market ("BLP instruments"), and the price of a given product in other markets ("Hausman

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

instruments"). The attractiveness of this approach is that even in the absence of cost data, firms' marginal cost functions can be recovered with the assumption that firms set prices to maximize profits given their rivals' prices.[2]

Nonetheless, BLP also propose using cost data, if available, for a variety of purposes, including improving their parameter estimates as well as understanding the relationship between prices and marginal costs. Recently, some researchers have started incorporating cost data as an additional source of identification. For instance, Houde (2012) combines wholesale gasoline prices with first order conditions that characterize stations' optimal pricing strategies to identify stations' marginal cost function. Crawford and Yurukoglu (2012) and Byrne (2015) similarly exploit first order conditions and firm-level cost data to identify the cost functions of cable companies. Some papers have also used demand and cost data to test assumptions regarding conduct in oligopoly models. See, for instance, Byrne (2015), McManus (2007), Clay and Troesken (2003), Kim and Knittel (2003), and Wolfram (1999). Kutlu and Sickles (2012) estimate market power while allowing for inefficiency in production. Like previous research, these researchers use instrumental variables (IVs) to identify demand in a first step.

We extend the existing research on BLP-type models by developing and formalizing new ways to obtain additional identification with cost data.[3] Our main theoretical finding is that by combining demand and cost data, and by using the restriction that marginal revenue equals marginal cost in equilibrium, one can jointly identify the price coefficients in the BLP demand model and a nonparametric cost function. The implicit exclusion restrictions that we exploit for identification are (1) price p, and market share s determine marginal revenue but do not directly enter the cost function, (2) output q and market share are related through $q = sQ$, where Q is market size and (3) output q enters the cost function but does not directly enter the demand function. The type of cost data we have in mind comes from firms' income statements and balance sheets, among other sources. Such data has been used extensively in a large parallel literature on cost function estimation in empirical IO.[4]

We do not need any variation in market size to identify and estimate the BLP-demand model.

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

[3]At a broader level, our paper shares a common theme with De Loecker (2011), where the usefulness of demand-side data in identifying production functions and measuring productivity is investigated.

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

For logit demand, we need some variation in market size, but such variation does not need to be exogenous. In our Monte-Carlo simulations, we show that our methodology yields consistent estimates even when market size is correlated with demand and cost shocks, as well as input prices.

We also show that we can identify the price coefficient without any instruments, even if the true market size is unobservable, as long as it is a function of observables. Here, Bresnahan and Reiss (1990) assume the exclusion restriction that the variables determining the market size do not enter in the demand function. We show that we can still identify the price coefficient with a different restriction that the variables determining the market size do not enter in the cost function. Such a restriction is more reasonable because typically, the determinants of market size are demographics, which do not affect cost directly.

In this paper, we propose to use cost data to control for the supply shock in ways that are similar to the control function approach. Assuming that cost is a function of output, input prices, observed characteristics and a cost shock, we show that cost (or expected cost conditional on observables) can be used to control for the cost shock. Our results then imply that one does not need to use instruments to identify the parameters for endogenous prices in differentiated goods demand, nor to identify the cost function where output is potentially endogenous. We also do not need to assume that researchers have some information on the correlation of the demand and cost shocks, which is used by MacKaye and Miller (2018) to identify the price coefficient of demand, nor do we need orthogonality between observed and unobserved product characteristics to identify the price coefficient.[5]

To address concerns in the empirical IO literature about the reliability of cost data for studying firm behavior[6], we minimize the use of cost data. In particular, we use it only to alleviate the endogeneity issue of product prices to demand shocks. We also show that our parametric identification results go through in model specifications that allow for cost data with measurement error as well as systematic over/under reporting by firms. Furthermore, we impose minimal assumptions on our nonparametric cost function. The main requirement we have is that the total cost function is strictly increasing in output and is strictly increasing and strictly convex in cost shock.

[5]Petrin and Seo (2016) propose an identification and estimation scheme that allows for observed and unobserved characteristics in the demand equation to be endogenously determined. They skilfully exploit the optimal choice of observed characteristics to create additional moments. However, under the BLP random coefficient model of demand, the number of first order conditions is less than the number of parameters. Therefore, additional moment restrictions are required.

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

Our identification analysis highlights a Curse of Dimensionality problem that likely makes an estimator based on the direct application of the identification results impractical. This motivates our two-step Sieve Non-Linear Least Squares (SNLLS) estimator, which does not suffer from the dimensionality problem. This estimator is semi-parametric in that it recovers a parametric logit or BLP demand structure and a nonparametric cost function. We also show how this estimator can be adapted to accommodate various data and specification issues that arise in practice. These include endogenous product characteristics, imposing restrictions on cost functions such as homogeneity of degree one in input prices, dealing with the difference between accounting cost and economic cost, missing cost data for some products or firms, and multi-product firms.

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

We then apply our methodology to the estimation of deposit demand in the US banking industry. We find that our method works well. The magnitude of the coefficient estimate on deposit interest rate is smaller than the ones obtained in the existing literature such as Dick (2008) and Ho and Ishii (2012). Further, we find that the IV-based method yields a negative coefficient on deposit interest rate whereas ours is positive which is what one would expect.

A prominent example of papers that exploit first order conditions to estimate demand parameters is Smith (2004). The author estimates a demand model using consumer-level choice data for supermarket products. The author does not, however, have product-level price data. To overcome this missing data problem, the author develops an identification strategy that uses data on national price-cost margins and identifies the price coefficient in the demand model as the one that rationalizes these national margins. Our study differs in that we focus on the more common situation where a researcher has data on prices, aggregate market shares, and total costs, but not marginal costs.

This paper is organized as follows. In Section 2, we specify the differentiated products model of interest and review the IV based estimation approach in the literature. In Section 3, we study identification when demand and cost data are available and develop our formal identification results. In Section 4, we propose the two-step SNLLS estimator and analyze its large sample properties. Section 5 contains a Monte-Carlo study that illustrates the effectiveness of our estimator in environments where standard approaches to demand estimation yield biased results. In Section 6 we apply our methodology to the estimation of deposit demand in the banking industry. In Section 7 we conclude. The Appendix contains several proofs and further details of the deposit demand estimation exercise.

2 Differentiated products models and IV estimation

2.1 Differentiated products models

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

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

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

We assume that there are $m = 1 \ldots M$ isolated markets that have respective market sizes Q_m.[7] Each market has $j = 0 \ldots J_m$ products whose aggregate demand across individuals is

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

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

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

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

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

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

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

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

In the case of BLP, one allows the price coefficient and coefficients on the observed characteristics to be different for different consumers. Specifically, α has a distribution function $F_\alpha(.; \boldsymbol{\theta}_\alpha)$, where $\boldsymbol{\theta}_\alpha$ is the parameter vector of the distribution, and similarly, $\boldsymbol{\beta}$ has a distribution function $F_{\boldsymbol{\beta}}(.; \boldsymbol{\theta}_\beta)$ with parameter vector $\boldsymbol{\theta}_\beta$. The probability a consumer with coefficients α and $\boldsymbol{\beta}$ purchases product j is identical to that provided by the market share formula in Equation (2). The aggregate market share is obtained by integrating over the distributions of α and $\boldsymbol{\beta}$,

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

The mean utility is then defined to be $\delta_{jm} = \mathbf{x}_{jm}\boldsymbol{\mu}_\beta + \mu_\alpha p_{jm} + \xi_{jm}$, with $\delta_{0m} = 0$ for the outside good.

2.1.1 Recovering demand shocks

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

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

and therefore,

$$\boldsymbol{\delta}_m\left(\boldsymbol{\theta}\right) = \mathbf{s}^{-1}\left(\mathbf{s}_m; \boldsymbol{\theta}\right). \tag{5}$$

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

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

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

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

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

$$\xi_{jm}(\boldsymbol{\theta}) \equiv \xi_j\left(\mathbf{p}_m, \mathbf{s}_m, \mathbf{X}_m; \boldsymbol{\theta}\right) = \delta_{jm}\left(\boldsymbol{\theta}\right) - \mathbf{x}_{jm}\boldsymbol{\mu}_\beta - \mu_\alpha p_{jm}. \tag{7}$$

2.1.2 IV estimation of demand

A simple regression of $\delta_{jm}\left(\boldsymbol{\theta}\right) = \mathbf{x}_{jm}\boldsymbol{\beta} + \alpha p_{jm} + \xi_{jm}$ with $\delta_{jm}\left(\boldsymbol{\theta}\right)$ being the dependent variable and $\mathbf{x}_{jm}$ and p_{jm} being the regressors would yield a biased estimate of the price coefficient. This is because firms likely set higher prices for products with higher unobserved product quality, which creates a correlation between p_{jm} and ξ_{jm}, which violates the OLS orthogonality condition $E[\xi_{jm}p_{jm}] = 0$.

Researchers use a variety of demand instruments denoted as $\mathbf{z}_{jm}$ to overcome this issue. In particular, researchers construct a GMM estimator for $\boldsymbol{\theta}$ by assuming the following population moment conditions are satisfied at the true value of the demand parameters $\boldsymbol{\theta}_0$:

$$E[\xi_{jm}(\boldsymbol{\theta}_0)\,\mathbf{z}_{jm}] = \mathbf{0}, \tag{8}$$

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

2.2 Supply

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

$$C_{jm} = C\left(q_{jm}, \mathbf{w}_{jm}, \mathbf{x}_{jm}, \upsilon_{jm}; \boldsymbol{\tau}\right), \tag{9}$$

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

For the sake of expositional simplicity, we assume for now there is one firm for each product. Then, we can write firm j's profit function as

$$\begin{aligned} \pi_{jm} &= p_{jm} \times q_{jm} - C\left(q_{jm}, \mathbf{w}_{jm}, \mathbf{x}_{jm}, \upsilon_{jm}; \boldsymbol{\tau}\right). \\ q_{jm} &= Q_m \times s_j\left(\mathbf{p}_m, \mathbf{X}_m, \boldsymbol{\xi}_m; \boldsymbol{\theta}\right) \end{aligned} \tag{10}$$

Let MR_{jm}, be the marginal revenue of firm j in market m, where $j = 1, \ldots, J_m$, $m = 1, \ldots, M$. Keeping with BLP, we assume that firms act as differentiated products Bertrand price competitors. Therefore, the optimal price and quantity of product j in market m are determined by the first order condition (F.O.C.) that equates marginal revenue and marginal cost

$$\underbrace{MR_{jm} = \frac{\partial p_{jm} q_{jm}}{\partial q_{jm}} = p_{jm} + s_{jm}\left[\frac{\partial s_j\left(\mathbf{p}_m, \mathbf{X}_m, \boldsymbol{\xi}_m; \boldsymbol{\theta}\right)}{\partial p_{jm}}\right]^{-1}}_{MR_{jm}} = MC_{jm} = \underbrace{\frac{\partial C\left(q_{jm}, \mathbf{w}_{jm}, \mathbf{x}_{jm}, \upsilon_{jm}; \boldsymbol{\tau}\right)}{\partial q_{jm}}}_{MC_{jm}}. \tag{11}$$

Note that given the market share inversion in Equation (4), and the specification of mean utility $\boldsymbol{\delta}_m$, $\boldsymbol{\xi}_m$ is a function of $\mathcal{D}_m \equiv \{\mathbf{p}_m, \mathbf{s}_m, \mathbf{X}_m\}$ and $\boldsymbol{\theta}$. Therefore, marginal revenue of firm j

in market m, MR_{jm} in Equation (11) can be written as a function of observables and parameters as follows:

$$MR_{jm} \equiv MR_j\left(\mathcal{D}_m, \theta\right) \equiv MR_j\left(\mathbf{p}_m, \mathbf{s}_m, \mathbf{X}_m; \boldsymbol{\theta}\right). \tag{12}$$

where $MR_j\left(\mathbf{p}_m, \mathbf{s}_m, \mathbf{X}_m; \boldsymbol{\theta}\right)$ is the j th element of the vector of marginal revenue functions in market m, denoted by $\mathbf{MR}\left(\mathbf{p}_m, \mathbf{s}_m, \mathbf{X}_m; \boldsymbol{\theta}\right)$. Equations (11) and (12) imply that demand parameters can potentially be identified if there is data on marginal cost[8] or even without such data, if the cost function can be estimated and its derivative with respect to output can be taken.[9]

2.2.1 Cost function estimation

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

$$C_{jm} = C\left(q_{jm}, \mathbf{w}_{jm}, \mathbf{x}_{jm}, \upsilon_{jm}; \boldsymbol{\tau}\right) \Rightarrow \upsilon_{jm} = \upsilon\left(q_{jm}, \mathbf{w}_{jm}, \mathbf{x}_{jm}, C_{jm}; \boldsymbol{\tau}\right). \tag{13}$$

Like demand estimation, there are important endogeneity concerns with standard approaches to estimating cost functions. Specifically, output q_{jm} is endogenously determined by profit-maximizing firms as in Equation (11), and is potentially negatively correlated with the cost shock υ_{jm}. That is, all else equal, less efficient firms tend to produce less. In dealing with this issue, researchers have traditionally focused on selected industries where endogeneity can be ignored, or used instruments for output.

The IV approach to cost function estimation typically uses excluded demand shifters as instruments. We denote this vector of cost instruments by $\widetilde{\mathbf{z}}_{jm}$. We can estimate $\boldsymbol{\tau}$ assuming that the following population moments are satisfied at the true value of the cost parameters $\boldsymbol{\tau}_0$: $E\left[\upsilon_{jm}\left(q_{jm}, \mathbf{w}_{jm}, \mathbf{x}_{jm}, C_{jm}; \boldsymbol{\tau}_0\right)\widetilde{\mathbf{z}}_{jm}\right] = \mathbf{0}$.[10]

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

[9] Note that the pricing equation that Berry et al. (1995) take to the data (their Equation 3.6) is similar to our Equation (11) except that they assume that marginal cost is log-linear in output and observed product characteristics, i.e., $MC_{jm} = exp\left(\mathbf{w}_{jm}\boldsymbol{\gamma}_w + q_{jm}\gamma_q + \upsilon_{jm}\right)$. They use instruments to deal with the endogeneity of output with cost shocks and of prices to demand shocks. As long as the parametric specification of the supply side is accurate and there are enough instruments for identification, demand side and F.O.C. based orthogonality conditions are sufficient for identification of the parameters. We in contrast assume cost to be a nonparametric function of output, input prices, observed product characteristics and the cost shock, and use observed cost to control for the cost shock. This point is further explained in the context of identification for the logit model. Unlike Berry et al. (1995) and the subsequent papers, we do not use the demand orthogonality equation (8).

[10] Also note that the joint estimation of the market share function (Equation (2) or (3)), profit maximization Equation (11) and cost function (13) in a standard manner does not solve the endogeneity issue without additional exclusion restrictions that are similar in spirit to the IV assumptions.

3 Identification and estimation of the price coefficient

In this section, we demonstrate that the endogeneity concerns in estimating the price coefficient in demand can be mitigated if we use both demand and cost data. We prove identification for the BLP random coefficients model of demand, and then illustrate the main idea in a simple monopoly model with Berry (1994) logit demand. We find, however, that the estimation strategy that directly follows the parametric identification argument is likely to be subject to a Curse of Dimensionality problem in practice. Therefore, we pursue a different parametric estimation strategy in Section 4 that avoids this problem.

3.1 Identification in the BLP model

Consider the static differentiated products oligopoly model defined in Section 2. In market m there are J_m firms selling one product each.[11]. Let $\mathcal{R}_{jm} = \{\mathbf{p}_m, \mathbf{s}_m, \mathbf{X}_m, Q_m, q_{jm}, \mathbf{w}_m, C_m, j\}$ contain all the relevant data for firm j in market m.

3.1.1 Main assumptions

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

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

Assumption 2 *Markets are isolated. That is, outputs, market shares, prices and costs in market m are functions of only the observed and unobserved variables in market m.*

Assumption 3 *Input price vector $\mathbf{w}_m$: $\mathbf{w}_{jm} = \mathbf{w}_m$, for all j, m.*[12]

Assumption 4 *Market share s_{jm} is specified as in Equation (3). The distributions of α and each element of $\boldsymbol{\beta}$ are assumed to be independently normal, i.e., $\alpha \sim N\left(\mu_\alpha, \sigma_\alpha^2\right), \beta_k \sim N\left(\mu_{\boldsymbol{\beta}k}, \sigma^2_{\boldsymbol{\beta}k}\right)$, $k = 1, \ldots, K$. Further, $\mu_{\boldsymbol{\beta}k} = 0$, $k = 1, \ldots, K$;$\mu_\alpha < 0$.*

Assumption 5 *Let C^*_{jm} denote true cost. Then,*

$$C^*_{jm} \equiv C^v\left(q_{jm}, \mathbf{w}_m, \mathbf{x}_{jm}, \upsilon_{jm}\right) + e_f\left(\mathbf{w}_m, \mathbf{x}_{jm}\right) + \varsigma_{jm}, \tag{14}$$

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

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

where $C^v()$ is the variable cost component, which is a continuous function of q, w, x andυ,[13] *strictly increasing, and continuously differentiable in q and υ, and is strictly convex in υ; $e_f()$ is the deterministic component of the fixed cost, a continuous function of w and x;*[14] *ς is a fixed cost shock, i.i.d., with mean zero and independent of $\{\mathbf{X}_m, \mathbf{w}_m, \upsilon_{jm}, \boldsymbol{\xi}_m, Q_m\}$. Further, for any $q > 0$, $\mathbf{w} > \mathbf{0}$ and $\mathbf{x} \in \mathcal{X}$,*

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

Assumption 6 *Bertrand-Nash equilibrium holds in each market. That is, for any $j = 1, \ldots, J_m \geq 3$, $m = 1, \ldots, M$, firm j in market m chooses its price p_{jm} to equalize marginal revenue and marginal cost, given market size Q_m and prices of other firms in the same market $\mathbf{p}_{-j,m}$. The profit for firm j in market m is specified as in Equation (10)* [15]*.*

Assumption 7 *Let $\mathbf{Q}_{-m} = (Q_1, Q_2, \ldots, Q_{m-1}, Q_{m+1}, \ldots Q_M)$; $\mathbf{W}_{-m} = (\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_{m-1}, \mathbf{w}_{m+1}, \ldots \mathbf{w}_M)$; $\boldsymbol{\Xi}_{-j,-m} = \left(\boldsymbol{\xi}_1^\dagger, \ldots, \boldsymbol{\xi}_{m-1}^\dagger, \boldsymbol{\xi}_{-jm}^\dagger, \boldsymbol{\xi}_{m+1}^\dagger, \ldots \boldsymbol{\xi}_M^\dagger\right)$, where $\boldsymbol{\xi}_{-jm}$ is the vector of demand shocks of all firms in market m other than firm j; $\boldsymbol{\Upsilon}_{-j,-m}$ and $\mathcal{X}_{-j,-m}$ are defined analogously for vectors of cost shocks and observed characteristics. Further, let $\mathbf{V}_{-j,-m} \equiv (\mathbf{Q}_{-m}, \mathbf{W}_{-m}, \boldsymbol{\Xi}_{-j,-m}, \boldsymbol{\Upsilon}_{-j,-m}, \mathcal{X}_{-j,-m})$.*[16] *For all $jm : j = 1, \ldots, J_m, m = 1, \ldots, M$, given J_m, $\upsilon_{jm} \in R_+$ given Q_m, ξ_{jm}, $\mathbf{w}_m$, $\mathbf{x}_{jm}$ and $\mathbf{V}_{-j,-m}$; $\xi_{jm} \in R$ given Q_m, υ_{jm}, $\mathbf{w}_m$, $\mathbf{x}_{jm}$ and $\mathbf{V}_{-j,-m}$; $\mathbf{w}_m \in R_+^L$, given Q_m, ξ_{jm}, υ_m, $\mathbf{x}_{jm}$ and $\mathbf{V}_{-j,-m}$ and $x_{kjm} \in R$ given the other variables.*[17]

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

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

[14] We can also include $\mathbf{x}_{f,jm}$, a vector of additional variables that determine the fixed cost, into the deterministic component of the fixed cost, so that its specification becomes $e_f(\mathbf{w}_m, \mathbf{x}_{jm}, \mathbf{x}_{f,jm})$. However, we omit $\mathbf{x}_{f,jm}$ for the sake of expositional simplicity.

[15] It is straightforward to show that in the case of the market share function specified as in Equation (3), firms choose prices such that marginal revenue equals marginal cost. Note that we have assumed this for expositional purposes only. It is not required for identification. As long as firms maintain a functional relationship between marginal revenue and marginal cost, i.e., firms require MR to be a function of MC in equilibrium, and not necessarily equal to MC, we can identify the price coefficient. For estimation, we also require this function to be strictly increasing. This makes our framework applicable to firms that are under government regulation and firms under organizational incentives or behavioral aspects that prevent them from setting $MR = MC$.

[16] Note that we do not impose any support conditions on the market size Q_m. Later, when we present a simple example where the market share function is specified to be a logit function, then we need the assumption that the conditional support of the market size shock Q_m is a non-singleton set.

[17] Its support can also be R_+

by the measurement error [18]. *That is,*

$$C_{jm} = C^*_{jm} + e_m(q_{jm}, \mathbf{w}_m, \mathbf{x}_{jm}) + \nu_{jm}. \tag{15}$$

where $e_m()$ *is a continuous function and* ν_{jm} *is i.i.d. with mean* 0 *and independent of* $\{\mathbf{X}_m, \mathbf{w}_m, \upsilon_{jm}, \boldsymbol{\xi}_m, Q_m, \varsigma_m\}$[19]. *All data except costs are observed without measurement error.*

Using Equations (14) and (15), we obtain $C_{jm} = C^v(q_{jm}, \mathbf{w}_m, \mathbf{x}_{jm}, \upsilon_{jm}) + e_f(\mathbf{w}_m, \mathbf{x}_{jm}) + \varsigma_{jm} + e_m(q_{jm}, \mathbf{w}_m, \mathbf{x}_{jm}) + \nu_{jm}$. Now, let

$$e(q, \mathbf{w}, \mathbf{x}) \equiv e_f(\mathbf{w}, \mathbf{x}) + e_m(q, \mathbf{w}, \mathbf{x}).$$

Then,

$$C_{jm} = C^v(q_{jm}, \mathbf{w}_m, \mathbf{x}_{jm}, \upsilon_{jm}) + e(q_{jm}, \mathbf{w}_m, \mathbf{x}_{jm}) + \nu_{jm} + \varsigma_{jm}. \tag{16}$$

We show that cost data identifies the parameters of the distribution of the random coefficients on price $(\mu_\alpha, \sigma_\alpha)$, as well as σ_{β_k}, $k = 1, \ldots, K$, the standard deviations of $\boldsymbol{\beta}$.[20] Note that we can only identify $(\mu_{\alpha 0}, \sigma_{\alpha 0}, \boldsymbol{\sigma}_{\boldsymbol{\beta}0})$ and $\xi_{0jm} + \mathbf{x}_{jm}\boldsymbol{\mu}_{\beta 0}$ in the absence of further restrictions imposed on the model. However, the identification of $(\mu_{\alpha 0}, \sigma_{\alpha 0}, \boldsymbol{\sigma}_{\boldsymbol{\beta}0})$ and $\xi_{0jm} + \mathbf{x}_{jm}\boldsymbol{\mu}_{\beta 0}$ is sufficient to identify the marginal revenue, thus, markup, which is the primary interest of most empirical exercises. The additional orthogonality assumption $E(\mathbf{x}_{jm}\xi_{jm}) = 0$ identifies $\boldsymbol{\mu}_\beta$. Therefore, from now on, except when notified otherwise, we will denote the vector of true parameters $\boldsymbol{\theta}_{c0}$ to be $(\alpha_0, \boldsymbol{\beta}'_0)$ for the logit demand specification and $\left(\mu_{\alpha 0}, \sigma_{\alpha 0}, \boldsymbol{\sigma}'_{\boldsymbol{\beta}0}\right)$ for the BLP specification.

Recall that marginal revenue in the BLP model is a function of prices, market shares, observed characteristics, and demand shocks. Also, marginal cost is a function of output, input prices, observed characteristics, and variable cost shocks. We use the exclusion restriction that given output q_{jm}, firm j's market share s_{jm} does not affect its marginal cost, and given the firm's market share, its output does not affect its marginal revenue. It is the market size that

[18] Note that we could also allow for systematic misreporting of true costs. So, for example, if firms report costs truthfully but with an error, then $\nu(C^*) = C^*$. Alternatively, if firms systematically under-report their true costs, then we could consider a specification like $\nu(C^*) = \nu C^*$ where $0 < \nu < 1$. Over-reporting could be captured by the same specification with $\nu > 1$.

[19] We can also include $\mathbf{x}_{me,jm}$, a vector of additional variables that determine the measurement error, into the deterministic component of the measurement error, so that its specification becomes $e_m(q_{jm}, \mathbf{w}_m, \mathbf{x}_{jm}, \mathbf{x}_{me,jm},)$. However, we omit $\mathbf{x}_{me,jm}$ for the sake of expositional simplicity.

[20] Fox et al. (2012) establishes identification of the random coefficient discrete choice models with exogenous RHS variables. For the identification of the random coefficient discrete choice models with endogenous RHS variables, Berry and Haile (2014) discuss the general orthogonality conditions by instruments that identify the BLP parameters. Here, we prove identification of the BLP random coefficient discrete choice models without instruments.

provides the link between market share and output.

Notice that we follow the convention that firms set prices so that marginal revenue equals marginal cost. However, in the BLP specification, firms equalizing marginal revenue to marginal cost does not imply that they are maximizing profits. This is because as long as the price coefficient is normally distributed, firms maximize profits by setting prices to plus infinity, i.e., no interior maximum exists. The logic is simple: consider, for simplicity, a monopolist firm. Even if the firm increases its price to infinity, it can still sell the product to consumers whose price coefficients are positive, and thus, obtain positive market share, which equals $s = Pr(\alpha > 0)$, and thus, output $q = Qs$. Therefore, the monopolist would earn an infinite revenue and only pay the finite total cost of producing the output $q > 0$.

Such discrepancy between profit maximization and the equality between the marginal revenue and marginal cost will vanish if we impose the restriction that $Pr(\alpha \geq 0; \boldsymbol{\theta}_{c0}) = 0$. This is essentially pre-imposing a restriction that the price coefficient is negative for all individuals. Instead, we follow the literature and allow for the support of the price coefficient to be positive as well in our identification/estimation, so that we can find the model specification and the estimation procedure that results in the price coefficient estimates of most individuals to be negative. Then, we can truncate away the positive component of the estimated price coefficient distribution so that MR=MC is consistent with profit maximization.

3.1.2 The Main Result

Proving identification for the BLP model of demand in oligopoly markets is complex because of the need to integrate with respect to the random coefficients α and $\boldsymbol{\beta}$ to form the market share. Thus, we present the formal proofs in the Appendix. In this subsection, we outline the logic of the proof. In the next subsection, we illustrate the idea with the logit demand structure.

Suppose the true variable cost is observable, i.e., $C_{jm} = C^v(q_{jm}, \mathbf{w}_m, \mathbf{x}_{jm}, \upsilon_{jm})$. Since, from Assumption 6, firms choose prices to equate marginal revenue to marginal cost, and since from Assumption 5, variable cost is an increasing and continuously differentiable function of output, and since the random coefficient BLP market share function is also differentiable, for any $q_{jm} > 0$ in the data, the following first order condition holds in equilibrium:

$$MR_j(\mathbf{p}_m, \mathbf{s}_m, \mathbf{X}_m; \boldsymbol{\theta}_{c0}) = MC(q_{jm}, \mathbf{w}_m, \mathbf{x}_{jm}, \upsilon_{jm}), \; q_{jm} = Q_m s_{jm}. \tag{17}$$

Since marginal cost is assumed to be positive, it follows that the marginal revenue for firms in the data must be positive as well. We exploit the inversion technique discussed in Equation (13) and use observed variable cost to control for the variable cost shock in a way that is similar to the control function approach. Substituting Equation (13) into Equation (17), we can eliminate the variable cost shock to derive

$$\begin{aligned} MR_j\left(\mathbf{p}_m, \mathbf{s}_m, \mathbf{X}_m; \boldsymbol{\theta}_{c0}\right) &= MC\left(q_{jm}, \mathbf{w}_m, \mathbf{x}_{jm}, \upsilon\left(q_{jm}, \mathbf{w}_m, \mathbf{x}_{jm}, C_{jm}\right)\right) \\ &= MC\left(q_{jm}, \mathbf{w}_m, \mathbf{x}_{jm}, C_{jm}\right). \end{aligned} \tag{18}$$

Thus, our first order condition, given by Equation (18) requires only the true demand coefficients $\boldsymbol{\theta}_{c0}$ and the observables. To identify the price coefficients, $(\mu_{\alpha 0}, \sigma_{\alpha 0})$, we pair up firms in two different markets m and $m^{\dagger}$ that have the same output, input prices, observed product characteristics, *and* variable cost ($q_{jm} = q_{jm^{\dagger}} = q$, $\mathbf{w}_m = \mathbf{w}_{m^{\dagger}} = \mathbf{w}$, $\mathbf{x}_m = \mathbf{x}_{m^{\dagger}} = \mathbf{x}$ *and* $C_{jm} = C_{j^{\dagger}m^{\dagger}} = C$). That is, $C_{jm} = C^v\left(q, \mathbf{w}, \mathbf{x}, \upsilon_{jm}\right) = C_{j^{\dagger}m^{\dagger}} = C^v\left(q, \mathbf{w}, \mathbf{x}, \upsilon_{j^{\dagger}m^{\dagger}}\right)$. Since the variable cost is a strictly increasing function of the variable cost shock, it follows that these firms must have the same variable cost shock, and thus, the same marginal cost and thus, the same marginal revenue. That is:

$$MR_j\left(p_m, s_m, \mathbf{X}_m, \boldsymbol{\theta}_{c0}\right) = MR_{j^{\dagger}}\left(p_{m^{\dagger}}, \mathbf{s}_{m^{\dagger}}, \mathbf{X}_{m^{\dagger}}, \boldsymbol{\theta}_{c0}\right). \tag{19}$$

This equality identifies $\boldsymbol{\theta}_{c0}$.

Since true variable cost is not observable and total cost includes fixed cost as well, we show next how the above logic can be applied to such cases:

Lemma 1 *Let Assumptions 1-3 and 5-8 be satisfied. Then, let* $\mathcal{R} = \{\mathbf{p}, \mathbf{s}, \mathbf{X}, q, \mathbf{w}, Q, j\}$ *and* $\mathcal{R}^{\dagger} = \left\{\mathbf{p}^{\dagger}, \mathbf{s}^{\dagger}, \mathbf{X}^{\dagger}, q^{\dagger}, \mathbf{w}^{\dagger}, Q^{\dagger}, j^{\dagger}\right\}$ *be two sets of observations in two different markets, such that* $q = q^{\dagger}$, $\mathbf{w} = \mathbf{w}^{\dagger}$ $\mathbf{x}_j = \mathbf{x}^{\dagger}_{j^{\dagger}} \equiv \mathbf{x}$ *for some* j, $j^{\dagger}$. *Let* $\mathcal{D} = \{\mathbf{p}, \mathbf{s}, \mathbf{X}\}$ *and* $\mathcal{D}^{\dagger} = \left\{\mathbf{p}^{\dagger}, \mathbf{s}^{\dagger}, \mathbf{X}^{\dagger}\right\}$, *and let* $\boldsymbol{\theta}_{c0}$ *be the true parameter value. Then,*

$$MR_j\left(\mathbf{p}, \mathbf{s}, \mathbf{X}; \boldsymbol{\theta}_{c0}\right) = MR_{j^{\dagger}}\left(\mathbf{p}^{\dagger}, \mathbf{s}^{\dagger}, \mathbf{X}^{\dagger}; \boldsymbol{\theta}_{c0}\right) \tag{20}$$

if and only if

$$E\left[C\middle|\left(\widetilde{q} = q, \widetilde{\mathbf{w}} = \mathbf{w}, \widetilde{\mathbf{p}} = \mathbf{p}, \widetilde{\mathbf{s}} = \mathbf{s}, \widetilde{\mathbf{X}} = \mathbf{X}, j\right)\right] = E\left[C\middle|\left(\widetilde{q} = q, \widetilde{\mathbf{w}} = \mathbf{w}, \widetilde{\mathbf{p}} = \mathbf{p}^{\dagger}, \widetilde{\mathbf{s}} = \mathbf{s}^{\dagger}, \widetilde{\mathbf{X}} = \mathbf{X}^{\dagger}, j^{\dagger}\right)\right] \tag{21}$$

Proof. First, given observation $\mathcal{R} = \{\mathbf{p}, \mathbf{s}, \mathbf{X}, q, \mathbf{w}, Q, j\}$, we can find $\mathcal{R}^\dagger = \left\{\mathbf{p}^\dagger, \mathbf{s}^\dagger, \mathbf{X}^\dagger, q, \mathbf{w}, Q^\dagger, j^\dagger\right\}$, from Assumption 7. Now, suppose Equation (20) holds. Then, from Assumptions 5 and 6, because marginal cost is strictly increasing in υ, for any q, $\mathbf{w}$, $\mathbf{x}$, there exists a unique υ such that

$$MR_j\left(\mathbf{p}, \mathbf{s}, \mathbf{X}; \boldsymbol{\theta}_{c0}\right) = MR_{j^\dagger}\left(\mathbf{p}^\dagger, \mathbf{s}^\dagger, \mathbf{X}^\dagger; \boldsymbol{\theta}_{c0}\right) = MC\left(q, \mathbf{w}, \mathbf{x}, \upsilon\right). \tag{22}$$

Since $(q, \mathbf{w}, \mathbf{x}) = \left(q^\dagger, \mathbf{w}^\dagger, \mathbf{x}^\dagger\right)$, from Equation (22), the two firms $\mathcal{R}$ and $\mathcal{R}^\dagger$ have the same variable cost shock, i.e., $\upsilon = \upsilon^\dagger$. Then,

$$C^v\left(q, \mathbf{w}, \mathbf{x}, \upsilon\right) + e\left(q, \mathbf{w}, \mathbf{x}\right) = C^v\left(q^\dagger, \mathbf{w}^\dagger, \mathbf{x}^\dagger, \upsilon^\dagger\right) + e\left(q^\dagger, \mathbf{w}^\dagger, \mathbf{x}^\dagger\right). \tag{23}$$

Then, since Equation (22) uniquely determines υ, υ is a function of $(q, \mathbf{w}, \mathbf{p}, \mathbf{s}, \mathbf{X})$. Thus, Assumptions 5 and 8 result in

$$\begin{aligned}
&E\left[C\middle|\left(\widetilde{q} = q, \widetilde{\mathbf{w}} = \mathbf{w}, \widetilde{\mathbf{p}} = \mathbf{p}, \widetilde{\mathbf{s}} = \mathbf{s}, \widetilde{\mathbf{X}} = \mathbf{X}, j\right)\right], \\
&= E\left[C^v\left(q, \mathbf{w}, \mathbf{x}, \upsilon\right) + e\left(q, \mathbf{w}, \mathbf{x}\right) + \varsigma_{jm} + \nu_{jm}\middle|\left(\widetilde{q} = q, \widetilde{\mathbf{w}} = \mathbf{w}, \widetilde{\mathbf{p}} = \mathbf{p}, \widetilde{\mathbf{s}} = \mathbf{s}, \widetilde{\mathbf{X}} = \mathbf{X}, j\right)\right] \\
&= C^v\left(q, \mathbf{w}, \mathbf{x}, \upsilon\right) + e\left(q, \mathbf{w}, \mathbf{x}\right)
\end{aligned} \tag{24}$$

and similarly,

$$E\left[C\middle|\left(\widetilde{q} = q^\dagger, \widetilde{\mathbf{w}} = \mathbf{w}^\dagger, \widetilde{\mathbf{p}} = \mathbf{p}^\dagger, \widetilde{\mathbf{s}} = \mathbf{s}^\dagger, \widetilde{\mathbf{X}} = \mathbf{X}^\dagger, j^\dagger\right)\right] = C^v\left(q^\dagger, \mathbf{w}^\dagger, \mathbf{x}^\dagger, \upsilon^\dagger\right) + e\left(q^\dagger, \mathbf{w}^\dagger, \mathbf{x}^\dagger\right). \tag{25}$$

Then, Equation (21) holds from Equations (23), (24) and (25).

Next suppose Equation (21) holds. From Assumptions 5 and 6, $MR_j\left(\mathbf{p}, \mathbf{s}, \mathbf{X}; \boldsymbol{\theta}_{c0}\right) = MC\left(q, \mathbf{w}, \mathbf{x}, \upsilon\right)$ holds for a unique υ. Thus, the cost shock υ can be expressed as $\upsilon\left(q, w, \mathbf{p}, \mathbf{s}, \mathbf{X}, j\right)$, Therefore, from Equations (24) and (25), we can derive

$$\begin{aligned}
&E\left[C\middle|\left(\widetilde{q} = q, \widetilde{\mathbf{w}} = \mathbf{w}, \widetilde{\mathbf{p}} = \mathbf{p}, \widetilde{\mathbf{s}} = \mathbf{s}, \widetilde{\mathbf{X}} = \mathbf{X}, j\right)\right] = C^v\left(q, \mathbf{w}, \mathbf{x}, \upsilon\left(q, w, \mathbf{p}, \mathbf{s}, \mathbf{X}, j\right)\right) + e\left(q, \mathbf{w}, \mathbf{x}\right) \\
= \ &E\left[C\middle|\left(\widetilde{q} = q, \widetilde{\mathbf{w}} = \mathbf{w}, \widetilde{\mathbf{p}} = \mathbf{p}^\dagger, \widetilde{\mathbf{s}} = \mathbf{s}^\dagger, \widetilde{\mathbf{X}} = \mathbf{X}^\dagger, j^\dagger\right)\right] = C^v\left(q, \mathbf{w}, \mathbf{x}, \upsilon\left(q, w, \mathbf{p}^\dagger, \mathbf{s}^\dagger, \mathbf{X}^\dagger, j^\dagger\right)\right) + e\left(q, \mathbf{w}, \mathbf{x}\right)
\end{aligned}$$

Therefore, $\upsilon \equiv \upsilon\left(q, \mathbf{w}, \mathbf{p}, \mathbf{s}, \mathbf{X}, j\right) = \upsilon\left(q, \mathbf{w}, \mathbf{p}^\dagger, \mathbf{s}^\dagger, \mathbf{X}^\dagger, j^\dagger\right) = \upsilon^\dagger$, and hence,

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

■

Next, we state the condition that enables us find the true parameter $\boldsymbol{\theta}_{c0}$ from those pairs of firms.

Condition 1 *Let* $\mathcal{D} = \{\mathbf{p}, \mathbf{s}, \mathbf{X}\}$ *and* $\mathcal{D}^\dagger = \{\mathbf{p}^\dagger, \mathbf{s}^\dagger, \mathbf{X}^\dagger\}$ *be two sets of observations in two different markets, and let* $\boldsymbol{\theta}_{c0}$ *be the true parameter value. Then, for any given* $\boldsymbol{\theta}_c \neq \boldsymbol{\theta}_{c0}$*, there exist* $\mathcal{D}$ *and* $\mathcal{D}^\dagger$ *,*$\mathcal{D} \neq \mathcal{D}^\dagger$ *for any reordering of the firms in market* $\dagger$*, and* j *and* $j^\dagger$ *that satisfy the following properties.*

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

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

3. *Either* $MR_j(\mathbf{p}, \mathbf{s}, \mathbf{X}, \boldsymbol{\theta}_c) = MR_{j^\dagger}\left(\mathbf{p}^\dagger, \mathbf{s}^\dagger, \mathbf{X}^\dagger, \boldsymbol{\theta}_c\right)$ *or* $MR_j(\mathbf{p}, \mathbf{s}, \mathbf{X}, \boldsymbol{\theta}_{c0}) = MR_{j^\dagger}\left(\mathbf{p}^\dagger, \mathbf{s}^\dagger, \mathbf{X}^\dagger, \boldsymbol{\theta}_{c0}\right)$ *but not both, and* $MR_j(\mathbf{p}, \mathbf{s}, \mathbf{X}, \boldsymbol{\theta}_{c0}) > 0$ *and* $MR_j\left(\mathbf{p}^\dagger, \mathbf{s}^\dagger, \mathbf{X}^\dagger, \boldsymbol{\theta}_{c0}\right) > 0$*.*

It is important to note that those two sets of vectors, $\mathcal{D}$ and $\mathcal{D}^\dagger$ need to come from different markets. Otherwise, for any $\boldsymbol{\theta}_c \neq \boldsymbol{\theta}_{c0}$, $MR_j(\mathbf{p}, \mathbf{s}, \mathbf{X}, \boldsymbol{\theta}_c) = MR_{j^\dagger}\left(\mathbf{p}^\dagger, \mathbf{s}^\dagger, \mathbf{X}^\dagger, \boldsymbol{\theta}_c\right)$ because trivial reordering of the firms in market $\dagger$ would result in $\mathbf{p} = \mathbf{p}^\dagger$, $\mathbf{s} = \mathbf{s}^\dagger$, $\mathbf{X} = \mathbf{X}^\dagger$ and $j = j^\dagger$.

Then, it is straightforward to show that given the assumptions, as long as the marginal revenue satisfies Condition 1, the true parameter of the demand function is identified. We state it below in a lemma.

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

Proof. Consider $\mathcal{A}(q, \mathbf{w}) \equiv \left\{\left(\mathcal{D}, j, \mathcal{D}^\dagger, j^\dagger\right) : \mathbf{x}_j = \mathbf{x}_{j^\dagger}, E[C|\mathcal{D}, q, \mathbf{w}, \mathbf{x}, j] = E\left[C|\mathcal{D}^\dagger, q, \mathbf{w}, \mathbf{x}, j^\dagger\right]\right\}$ to be the pairs of all firms that have the same output, same input price, same observed characteristics and the same expected cost but in different markets. Furthermore, let $\mathcal{B}(\boldsymbol{\theta}_c) \equiv \left\{\left(\mathcal{D}, j, \mathcal{D}^\dagger, j^\dagger\right) : \mathbf{x}_j = \mathbf{x}_{j^\dagger}, MR_j(\mathcal{D}; \boldsymbol{\theta}_c) = MR_{j^\dagger}\left(\mathcal{D}^\dagger; \boldsymbol{\theta}_c\right)\right\}$. Then, from Lemma 1, $\mathcal{A}(q, \mathbf{w}) = \mathcal{B}(\boldsymbol{\theta}_{c0})$. Furthermore, Condition 1 implies $\mathcal{B}(\boldsymbol{\theta}) \neq \mathcal{B}(\boldsymbol{\theta}_{c0})$ for $\boldsymbol{\theta}_c \neq \boldsymbol{\theta}_{c0}$. Together, we derive $\mathcal{B}(\boldsymbol{\theta}_{c0}) = \mathcal{A}(q, \mathbf{w})$ if and only if $\boldsymbol{\theta}_c = \theta_{c0}$. Therefore, $\boldsymbol{\theta}_{c0}$ is identified. ■

Note that in proving the Lemmas above or in the Condition, we did not need to make any assumptions about independence of any of the variables from each other. That is, the Lemmas go through and the Condition can be satisfied regardless of possible correlation across input prices, variable cost shock, observed characteristics, demand shock and market size, within markets or across markets. For example, a positive correlation between market size and demand/variable cost shock can arise as a larger market size may induce firms to invest in higher quality or more advertising, which improves unobserved product quality but increases cost. But this does not break our identification strategy.[21] Thus, these findings illustrate that given cost data, one does not need any IV- or orthogonality assumptions.[22]

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

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

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

Of course, in practice, (22) may not be satisfied exactly. However, we show later that the estimator that is based on the identification argument above relies only on the data that satisfies (22) approximately. See Section 4 and the appendix for more details, and proofs of consistency and asymptotic normality of the estimator.

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

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

3.1.3 Monopoly with Logit Demand: An Example

We provide the intuition for how demand and cost data identify the price coefficient using a single-product monopolist facing logit demand. For simplicity, we assume there is no fixed cost,

[21] Later, we show that for the identification of the parameters of the distribution of the BLP price coefficients, market size does not need to have any variation. In our Monte-Carlo analysis, we provide a scenario where market size is correlated with demand shock. Results demonstrate consistency of our estimator. Furthermore, notice that an important component of it is the F.O.C. Equation (17). Any violation of this F.O.C. may result in $\boldsymbol{\theta}_0$ not being identified. An example would be if higher prices and more advertising spending signal product quality, as in the model of Milgrom and Roberts (1986).

[22] It is important to note that each pair of observations satisfying Condition 1 can be generated from different equilibria. Since the observables $\left\{q, \mathbf{w}, \mathbf{X}, \mathbf{p}, \mathbf{s}, E\left\{C \middle| q, \mathbf{w}, \mathbf{X}, \mathbf{p}, \mathbf{s}\right\}\right\}$ uniquely determine the pair of firms that have the same cost shock υ, and the marginal cost, the above procedure identifies the true price coefficient even when multiple equilibria exist.

and true cost is observable. Given Assumption 6, the following first order condition holds in equilibrium:

$$MR(p_m, s_m, \mathbf{x}_m, \boldsymbol{\theta}_{c0}) = p_m + \frac{1}{(1-s_m)\,\alpha_0} = MC\left(q_m, \mathbf{w}_m, \mathbf{x}_m, \upsilon_m\right), \; q_m = Q_m s_m. \tag{26}$$

To identify the price coefficient, we pair up firms that have different market sizes and prices ($Q_m \neq Q_{m^\dagger}$ and $p_m \neq p_{m^\dagger}$) but have the same output, input prices, observed product characteristics, *and* cost ($q_m = q_{m^\dagger} = q$, $\mathbf{w}_m = \mathbf{w}_{m^\dagger} = \mathbf{w}$, $\mathbf{x}_m = \mathbf{x}_{m^\dagger} = \mathbf{x}$ *and* $C_m = C_{m^\dagger}$). That is, $C_m = C^v\left(q, \mathbf{w}, \mathbf{x}, \upsilon_m\right) = C_{m^\dagger} = C^v\left(q, \mathbf{w}, \mathbf{x}, \upsilon_{m^\dagger}\right)$. It follows that these firms must have the same variable cost shock ($\upsilon_m = \upsilon_{m^\dagger}$), and thus, the same marginal cost and marginal revenue so that we can identify the price coefficient from the equality of marginal revenues as below:

$$p_m + \frac{1}{(1-s_m)\,\alpha_0} = p_{m^\dagger} + \frac{1}{(1-s_{m^\dagger})\,\alpha_0}, \quad \alpha_0 = -\frac{1}{p_m - p_{m^\dagger}}\left[\frac{1}{1-s_m} - \frac{1}{1-s_{m^\dagger}}\right], \tag{27}$$

Since the right hand side contains only observables, the true price coefficient α_0 is identified. Note that $\alpha_0 \neq 0$, as assumed, must hold for marginal revenue to be bounded. It then follows that α_0 is identified from such a pair of data points.

Note also that the Logit model satisfies Condition 1. In particular, since there is a unique solution to the above equation, it is straightforward that for any $\alpha \neq \alpha_0$, the pair of observations used above do not satisfy the marginal revenue equality, and thus, point 3 of Condition 1 holds.

Note that our argument in the logit demand case relies on market size Q to be different across the two markets (and thus, market shares). If the above two firms have the same output ($q_m = q_{m^\dagger}$), and the same market size ($Q_m = Q_{m^\dagger}$), then they have the same market share ($s_m = q_m/Q_m = q_{m^\dagger}/Q_{m^\dagger} = s_{m^\dagger}$). If they have the same marginal revenue, then from Equation (26), $p_m = p_{m^\dagger}$ holds, and thus, the price coefficient cannot be identified.

In contrast, the identification of $\boldsymbol{\theta}_{c0}$ in the BLP demand model holds even without any variation in market size across markets. To see why, note that even though the same market size leads to the pair of firms having the same market share, under BLP, these firms can have different $[\partial s/\partial p]$ and thus, different prices in the relationship. That is,

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

Therefore,

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

Thus, the relationship

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

is the additional source of variation in BLP specification that helps identify the parameters. [23]

3.1.4 Identification of Market Size

We now consider the case where market size Q_m is not observed, and thus needs to be estimated.[24] We follow Bresnahan and Reiss (1991) and specify the market size as follows:

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

where $\mathbf{z}_m$ is a $1 \times K_{\mathbf{z}}$ vector of observables in market m. Then, the true market share of firm j in market m, denoted by s^*_{jm} is unobservable. Bresnahan and Reiss (1991) and other literature on this issue assume that variables that determine market size are not included in the market share equation. However, since many variables in the RHS of the market size equation are demographic variables, they are likely to affect consumer demand as well. Therefore, we do not impose this restriction. That is, the modified utility function for individual i in market m consuming product j is

$$u_{ijm} = \mathbf{x}_{jm}\boldsymbol{\beta}_{\mathbf{x}} + \mathbf{z}_m\boldsymbol{\beta}_{\mathbf{z}} + \alpha p_{jm} + \xi_{jm} + \epsilon_{ijm}, \tag{28}$$

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

[24] While BLP assumes that the market size of automobile is the entire population, it is reasonable to assume that there is likely to be a fraction of individuals who would not consider buying an automobile regardless of its price or product characteristics. They may be living too far away from the area where the firms are located, or they may not be in need of the product. For example, individuals without a driver's license won't be interested in buying cars.

On the other hand, following the literature, we assume that the variables determining market size are not included in the cost function. This assumption is reasonable as demographic variables usually do not enter the production function.

We prove identification for the logit demand model here, and for the BLP model in the Appendix. We first show that $\widetilde{\boldsymbol{\lambda}}_{\mathbf{z}0} \equiv \boldsymbol{\lambda}_{\mathbf{z}0}/\lambda_{\mathbf{z}01}$ is identified as long as we can find $k = 1, \ldots, K_{\mathbf{z}}-1$ pairs of firms in different markets $j^{(k)}m^{(k)}$ and $j^{\dagger(k)}m^{\dagger(k)}$ satisfying $q_{j^{(k)}m^{(k)}} = q_{j^{\dagger(k)}m^{\dagger(k)}} = q^{(k)}$, $\mathbf{w}_{m^{(k)}} = \mathbf{w}_{m^{\dagger(k)}} = \mathbf{w}^{(k)}$, $\mathbf{x}_{j^{(k)}m^{(k)}} = \mathbf{x}_{j^{\dagger(k)}m^{\dagger(k)}} = \mathbf{x}^{(k)}$,[25] and the following equivalence of expected cost:

$$\begin{aligned} & E\left[C\middle|\widetilde{q} = q_{j^{(k)}m^{(k)}}, \widetilde{\mathbf{w}} = \mathbf{w}_{m^{(k)}}, \widetilde{\mathbf{x}} = \mathbf{x}_{j^{(k)}m^{(k)}}, \widetilde{p} = p_{j^{(k)}m^{(k)}}, \widetilde{\mathbf{z}} = \mathbf{z}_{m^{(k)}}, \widetilde{\mathbf{X}} = \mathbf{X}_{m^{(k)}}\right] \\ = \ & E\left[C\middle|\widetilde{q} = q_{j^{\dagger(k)}m^{\dagger(k)}}, \widetilde{\mathbf{w}} = \mathbf{w}_{m^{\dagger(k)}}, \widetilde{\mathbf{x}} = \mathbf{x}_{j^{\dagger(k)}m^{\dagger(k)}}, \widetilde{p} = p_{j^{\dagger(k)}m^{\dagger(k)}}, \widetilde{\mathbf{z}} = \mathbf{z}_{m^{\dagger(k)}}, \widetilde{\mathbf{X}} = \mathbf{X}_{m^{\dagger(k)}}\right] \\ & , \end{aligned} \tag{29}$$

and $p_{j^{(k)}m^{(k)}} = p_{j^{\dagger(k)}m^{\dagger(k)}}$ (all these equalities are allowed to hold approximately as well), but $\mathbf{z}_{m^{(k)}} \neq \mathbf{z}_{m^{\dagger(k)}}$, and the matrix $K_{\mathbf{z}} - 1 \times K_{\mathbf{Z}}$ matrix $\mathbf{Z}$ whose columns are $\mathbf{z}_{m^{(k)}} - \mathbf{z}_{m^{\dagger(k)}}$ is of full rank. Then, we use the same argument as before to show that these two firms have the same variable cost shocks and marginal costs, and thus, they must have the same marginal revenue at true market shares, i.e.,

$$p_{j^{(k)}m^{(k)}} - \frac{1}{\left(1 - s^*_{j^{(k)}m^{(k)}}\right)\alpha_0} = p_{j^{\dagger(k)}m^{\dagger(k)}} - \frac{1}{\left(1 - s^*_{j^{\dagger(k)}m^{\dagger(k)}}\right)\alpha_0}. \tag{30}$$

Then, it follows that within each pair, the true market shares must be equal. Thus,

$$\begin{aligned} ln\left(s^*_{j^{(k)}m^{(k)}}\right) &= ln\left(q_{j^{(k)}m^{(k)}}\right) - ln\left(Q_{m^{(k)}}\right) = ln\left(q_{j^{(k)}m^{(k)}}\right) - \lambda_{c0} - \mathbf{z}_{m^{(k)}}\boldsymbol{\lambda}_{\mathbf{z}0} \\ &= ln\left(s^*_{j^{\dagger(k)}m^{\dagger(k)}}\right) = ln\left(q_{j^{\dagger(k)}m^{\dagger(k)}}\right) - \lambda_{c0} - \mathbf{z}_{m^{\dagger(k)}}\boldsymbol{\lambda}_{\mathbf{z}0}, \end{aligned} \tag{31}$$

which results in

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

which identifies $\widehat{\boldsymbol{\lambda}}_{\mathbf{z}0} \equiv \boldsymbol{\lambda}_{\mathbf{z}0}/\lambda_{\mathbf{z}01}$. Such $K_{\mathbf{z}} - 1$ pairs exist as long as we assume that conditional on $(\boldsymbol{\xi}_m, \boldsymbol{\upsilon}_m, \mathbf{w}_m, \mathbf{X}_m)$, the support of $\mathbf{z}_m$ is $R^{K_{\mathbf{z}}}$.

The identification of the parameter that is left is equivalent to the case when $\mathbf{z}_m$ is one dimensional, which we denote as z_m from now on. For that, we need firms with $q_{jm} = q_{j^\dagger m^\dagger} = q$,

[25] Note that we select a pair of markets and a pair of firms corresponding to each k.

$\mathbf{w}_m = \mathbf{w}_{m^\dagger} = \mathbf{w}$, $\mathbf{x}_m = \mathbf{x}_{m^\dagger} = \mathbf{x}$, Equations (29) and (30), and $p_{jm} \neq p_{j^\dagger m^\dagger}$. Then, we substitute $s^*_{jm} = q_{jm}/exp\left(\lambda_{c0} + \lambda_{z0} z_m\right)$, $s^*_{j^\dagger m^\dagger} = q_{j^\dagger m^\dagger}/exp\left(\lambda_{c0} + \lambda_{z0} z_{m^\dagger}\right)$ to obtain

$$p_{jm} - p_{j^\dagger m^\dagger} = -\frac{1}{\alpha_0}\left[\frac{q}{exp\left(\lambda_{c0} + \lambda_{z0} z_m\right) - q} - \frac{q}{exp\left(\lambda_{c0} + \lambda_{z0} z_{m^\dagger}\right) - q}\right] \equiv \Delta p\left(q, z_m, z_{m^\dagger}\right).$$

Then, pairs of firms with $q_{jm} = q_{j^\dagger m^\dagger} = q$, $\mathbf{w}_m = \mathbf{w}_{m^\dagger} = \mathbf{w}$, $\mathbf{x}_m = \mathbf{x}_{m^\dagger} = \mathbf{x}$ with the same $p_{jm} - p_{j^\dagger m^\dagger}$ but different q, z_m, $z_{m^\dagger}$ will identify λ_{c0}, λ_{z0} regardless of the value of α_0. More concretely, given Equation (29) and $p_{jm} \neq p_{j^\dagger m^\dagger}$, and $z_m = 0$, $z_{m^\dagger} = \Delta z > 0$ being small, we have

$$p_{jm} - p_{j^\dagger m^\dagger} = -\frac{1}{\alpha_0}\left[\frac{q}{exp\left(\lambda_{c0}\right) - q} - \frac{q}{exp\left(\lambda_{c0} + \lambda_{z0} z_{m^\dagger}\right) - q}\right], \tag{32}$$

and for small $z_{m^\dagger} = \Delta z > 0$,

$$\frac{q}{exp\left(\lambda_{c0} + \lambda_{z0} z_{m^\dagger}\right) - q} \approx \frac{q}{exp\left(\lambda_{c0}\right) - q} - \frac{q}{\left(exp\left(\lambda_{c0}\right) - q\right)^2}\left[exp\left(\lambda_{c0} + \lambda_{z0} z_{m^\dagger}\right) - exp\left(\lambda_{c0}\right)\right].$$

Therefore,

$$\Delta p\left(q, 0, \Delta z\right) \equiv p_{jm} - p_{j^\dagger m^\dagger} \approx -\frac{1}{\alpha_0}\frac{q}{\left(exp\left(\lambda_{c0}\right) - q\right)^2} exp\left(\lambda_{c0}\right)\lambda_{z0}\Delta z$$

holds. Next, we do the same with $q' \neq q$. Then, letting $B\left(q, q', z_m, z_{m^\dagger}\right) \equiv \Delta p\left(q, z_m, z_{m^\dagger}\right)/\Delta p\left(q', z_m, z_{m^\dagger}\right)$, we have

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

which identifies λ_{c0}. Then, we do the same with $z_m = z$, $z_{m^\dagger} = z + \Delta z$, and, given λ_{c0}, we do similar calculations to derive

$$B\left(q, q', z, \Delta z\right) \equiv \frac{\Delta p\left(q, z, \Delta z\right)}{\Delta p\left(q', z, \Delta z\right)} \approx \frac{q}{q'}\left[\frac{exp\left(\lambda_{c0} + \lambda_{z0} z\right) - q'}{exp\left(\lambda_{c0} + \lambda_{z0} z\right) - q}\right]^2 = \frac{q}{q'}\left[1 + \frac{q - q'}{exp\left(\lambda_{c0} + \lambda_{z0} z\right) - q}\right]^2.$$

Since λ_{c0} is already identified, the above equation identifies λ_{z0}.

4 Estimation

In practice, an estimator that directly applies the parametric identification results in Section 3.1.2 will likely suffer from a Curse of Dimensionality. To implement such an estimator, one would need to obtain a nonparametric estimate of the conditional mean cost.

$$E\left[C|\widetilde{q}=q_{jm},\widetilde{\mathbf{w}}=\mathbf{w}_m,\widetilde{\mathbf{p}}=\mathbf{p}_m,\widetilde{\mathbf{s}}=\mathbf{s}_m,\widetilde{\mathbf{X}}=\mathbf{X}_m,j\right].$$

Furthermore, we need to find pairs with $q_{jm}\approx q_{j^\dagger m^\dagger}$, $\mathbf{x}_{jm}\approx\mathbf{x}_{j^\dagger m^\dagger}$, $\mathbf{w}_m\approx\mathbf{w}_{m^\dagger}$. For most markets of interest, $\mathbf{X}_m$ will contain some product characteristics across a non-negligible number of firms. This makes the dimensionality problem potentially quite severe.

Because of this dimensionality issue, we construct an estimator that exploits the parametric marginal revenue in such a way that the calculation of conditional mean cost is no longer required. This estimator conditions on marginal revenue, which is a parametric function of the observables, rather than the conditional expected cost.

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

Lemma 3 *Suppose that Assumptions 5, 6 and 8 are satisfied. Then,* $C^v(q_{jm},\mathbf{w}_m,\mathbf{x}_{jm},\upsilon_{jm})+e(q_{jm},\mathbf{w}_m,\mathbf{x}_{jm})=\varphi(q_{jm},\mathbf{w}_m,\mathbf{x}_{jm},MR_{jm}(\boldsymbol{\theta}_{c0}))$*, for firm* j *in market* m *with observables* $\{\mathbf{p}_m,\mathbf{s}_m,\mathbf{X}_m,q_{jm},\mathbf{w}_m,j\}$*, where* φ *is a function that is strictly increasing and continuous in marginal revenue.*

Proof. Because the marginal cost function is strictly increasing and continuous in υ, there exists an inverse function that is increasing and continuous such that $\upsilon=\upsilon(q,\mathbf{w},\mathbf{x},MC)$. This implies that we can use (an unspecified function of) q, $\mathbf{w}$, $\mathbf{x}$ and MC: $\upsilon(q,\mathbf{w},\mathbf{x},MC)$, to control for υ. Therefore, we can let the deterministic component of the cost to be $C^d(q,\mathbf{w},\mathbf{x},\upsilon)\equiv C^v(q,\mathbf{w},\mathbf{x},\upsilon)+e(q,\mathbf{w},\mathbf{x})=\varphi(q,\mathbf{w},\mathbf{x},MC)$, where φ is an increasing and continuous function of MC. Because $MR=MC$, at the true parameter vector $\boldsymbol{\theta}_{c0}$,

$$C^d(q,\mathbf{w},\mathbf{x},\upsilon) \quad = \quad \varphi(q,\mathbf{w},\mathbf{x},MC)=\varphi(q,\mathbf{w},\mathbf{x},MR).$$

■

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

4.1 Two-step Sieve Non-linear least squares (SNLLS) estimator

Using the above lemma, we construct an estimator that is based on the control function approach. In the first step, it selects parameters $\boldsymbol{\theta}_c$ to fit the pseudo-cost function to the cost data using a nonparametric sieve regression (Chen (2007); Bierens (2014)) as follows.

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

$$\varphi\left(q_{jm}, \mathbf{w}_m, \mathbf{x}_{jm}, MR_{jm}\left(\boldsymbol{\theta}_{c0}\right)\right) \quad = \quad \sum_{l=1}^{\infty} \gamma_{l0} \psi_l\left(q_{jm}, \mathbf{w}_m, \mathbf{x}_{jm}, MR_{jm}\left(\boldsymbol{\theta}_{c0}\right)\right), \tag{33}$$

where $\psi_1(\cdot), \psi_2(\cdot), \ldots$ are the basis functions for the sieve and $\gamma_1, \gamma_2, \ldots$ is a sequence of their coefficients, satisfying $\sum_{l=1}^{\infty} |\gamma_{l0}| < \infty$.[26]

From Equation (33), we obtain

$$E\left[\left(C_{jm} - \sum_{l=1}^{\infty} \gamma_{l0} \psi_l\left(q_{jm}, \mathbf{w}_m, \mathbf{x}_{jm}, MR_{jm}\left(\boldsymbol{\theta}_{c0}\right)\right)\right)^2\right] \quad = \quad E\left[\left(\nu_{jm} + \varsigma_{jm}\right)^2\right] = \sigma_\nu^2 + \sigma_\varsigma^2 \tag{34}$$

Our estimator is based on Equation (33). It is useful to introduce some additional notation before formally defining it. Let M be the number of markets, and L_M an integer that increases with M. For some bounded but sufficiently large constant $T > 0$, let $\Gamma_k(T) = \{\pi_k \boldsymbol{\gamma} : \|\pi_k \boldsymbol{\gamma}\| \leq T\}$ where π_k is the operator that applies to an infinite sequence $\boldsymbol{\gamma} = \{\gamma_n\}_{n=1}^{\infty}$, replacing γ_n, $n > k$ with zeros. That is, for $n \leq k$, $\pi_k \gamma_n = \gamma_n$, and for $n > k$, $\pi_k \gamma_n = 0$. The norm $\|\mathbf{x}\|$ is defined as $\|\mathbf{x}\| = \sqrt{\sum_{k=1}^{\infty} x_k^2}$.

Then, we can prove the following (The proof is in the Appendix.):

Proposition 2 *Suppose Assumptions 1-8, 9 and Condition 1 are satisfied. Then*

$$\left[\boldsymbol{\theta}_{c0}, \boldsymbol{\gamma}_0\right] = argmin_{(\boldsymbol{\theta}, \gamma) \in \Theta_c \times \Gamma} E\left[C_{jm} - \sum_{l=1}^{\infty} \gamma_l \psi_l\left(q_{jm}, \mathbf{w}_m, \mathbf{x}_{jm}, MR_{jm}\left(\boldsymbol{\theta}_c\right)\right)\right]^2, \tag{35}$$

where $\Gamma = lim_{M \to \infty} \Gamma_{L_M}(T)$; and equation (35) identifies $\boldsymbol{\theta}_{c0}$.

[26] Suppose the vector $(q_{jm}, \mathbf{w}_{jm}, \mathbf{x}_{jm}, MR_{jm})$ belongs to a compact finite-dimensional Euclidean space, $\mathcal{W}$. Then, if $\varphi(q_{jm}, \mathbf{w}_{jm}, \mathbf{x}_{jm}, MR_{jm})$ is a continuous function on $\mathcal{W}$, from the Stone-Weierstrass Theorem, it follows that the function can be approximated arbitrarily well by a polynomial function of a sufficiently higher order. However, the polynomials may not converge absolutely, so we still need to assume absolute convergence.

The sample analog of Equation (35), given a sample of M markets is:

$$\left[\widehat{\boldsymbol{\theta}}_{cM}, \widehat{\boldsymbol{\gamma}}_M\right] = argmin_{(\boldsymbol{\theta}_c,\gamma)\in\Theta_c\times\Gamma_{L_M}(T)} \frac{1}{\sum_{m=1}^{M} J_m} \sum_{m=1}^{M}\sum_{j=1}^{J_m}\left[C_{jm} - \sum_{l=1}^{L_M}\gamma_l\psi_l\left(q_{jm}, \mathbf{w}_{jm}, \mathbf{x}_{jm}, MR_{jm}\left(\boldsymbol{\theta}_c\right)\right)\right]^2. \tag{36}$$

The set $\Gamma_{L_M}(T)$ makes clear the fact that the complexity of the sieve is increasing in the sample's number of markets.[27] Our SNLLS (Sieve-NLLS) approach deals with issues of endogeneity by adopting a control function approach for the unobserved cost shock υ_{jm}. With our estimator, the right-hand side of Equation (36) is minimized only when parameters are at their true value $\boldsymbol{\theta}_{c0}$ so that the computed marginal revenue equals the true marginal revenue, and thus works as a control function for the supply shock υ_{jm}. If $\boldsymbol{\theta}_c \neq \boldsymbol{\theta}_{c0}$, then using the false marginal revenue adds noise, which increases the right-hand side of the sum of squared residuals in Equation (36). As the above argument makes clear, the true demand parameter $\boldsymbol{\theta}_{c0}$ can be obtained as a by-product of this control function approach.[28]

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

$$\widehat{\boldsymbol{\beta}}_M = \left(\sum_{m=1}^{M}\mathbf{X}_m'\mathbf{X}_m\right)^{-1}\sum_{m=1}^{M}\mathbf{X}_m'\left(\widehat{\boldsymbol{\delta}}_m - \widehat{\alpha}_M\mathbf{p}_m\right)\ or\ \widehat{\boldsymbol{\mu}}_{\beta M} = \left(\sum_{m=1}^{M}\mathbf{X}_m'\mathbf{X}_m\right)^{-1}\sum_{m=1}^{M}\mathbf{X}_m'\left(\widehat{\boldsymbol{\delta}}_m - \widehat{\mu}_{\alpha M}\mathbf{p}_m\right) \tag{37}$$

Both Equations (36) and (37) constitute our two-step SNLLS estimator for parameters $\boldsymbol{\theta} = \left(\mu_\alpha, \sigma_\alpha, \boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\sigma}_{\boldsymbol{\beta}}\right)$.[29]

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

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

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

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

Note that practitioners need to be careful in the nonparametric estimation of the pseudo-cost function if $\mathbf{x}_{jm}$ includes discrete variables, just like in any cases where nonparametric estimation involves both discrete and continuous variables.

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

[29] Note that if the exclusion restriction is not met, we can still obtain consistent parameter estimate of $\boldsymbol{\beta}$ (or $\boldsymbol{\mu}_{\boldsymbol{\beta}}$) if we assume that the observed characteristics of other products $\mathbf{X}_{-jm}$ and ξ_{jm} are uncorrelated. Then, we

4.2 Further specification and data issues

We have thus far worked with the standard differentiated products model from Berry (1994) and BLP. Depending on the empirical context, however, a number of specification and data-related issues can potentially arise. In this section, we list some empirical settings in which our estimator can be adapted by modifying the SNLLS part of the objective function in Equation (36):[30] They are:

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

2. *Endogenous product characteristics*: We can deal with the case where firms also choose product characteristics, by including the additional first order conditions in our estimator.

3. *Cost function restrictions*: We can incorporate the restriction that the cost function satisfies homogeneity of degree one in input price. Incorporating such a restriction in the estimation procedure has the benefit of reducing the dimensionality of the nonparametric pseudo-cost function.

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

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

can use $\mathbf{X}_{-jm}$ as instruments for $\mathbf{x}_{jm}$. In contrast, the literature uses these variables as instruments for both p_{jm} and $\mathbf{x}_{jm}$. Berry et al. (1995) use the sum of product characteristics over other firms as instruments for p_{jm}. In that case, if no other instruments are available, only functional form restrictions identify the coefficients of p_{jm} and $\mathbf{x}_{jm}$. It is also important to recall that even if $\boldsymbol{\beta}$ (or $\boldsymbol{\mu}_{\boldsymbol{\beta}}$) cannot be consistently estimated, in our procedure $\boldsymbol{\theta}_c$ is still estimated consistently, and so are marginal revenue and profit margin.

[30]The detailed discussions are in the Appendix.

4.3 Large sample properties

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

4.4 Bootstrap procedure for calculating the standard errors

In this section, we propose a bootstrap procedure for deriving the standard errors of $\widehat{\boldsymbol{\theta}}_{cM}$. In equilibrium models, bootstrapping by resampling the demand shocks ξ_{jm} and supply shocks υ_{jm} is computationally demanding because the equilibrium prices $\mathbf{p}_m$ and the market shares $\mathbf{s}_m$ need to be recomputed for each market m. Researchers such as Fu and Wolpin (2018) conduct nonparametric bootstrap where they resample market outcomes $(\mathbf{X}_m, \mathbf{p}_m, \mathbf{s}_m, \mathbf{w}_m, \mathbf{q}_m)$ for markets $m = 1, \ldots, M$, and estimate based on the resampled market data. Then, one does not need to recompute the equilibrium prices and market shares. However, their results may be subject to small sample issues due to the relatively small number of markets. Furthermore, the bootstrapped parameter estimates would likely be affected by the additional variation from the resampled $\mathbf{X}_m$ and $\mathbf{w}_m$ as well, which could overestimate the standard error. In addition, if the demand and supply shocks, and $\mathbf{X}_m$, $\mathbf{w}_m$ are correlated across markets, then this correlation needs to be dealt with in resampling.

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

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

Step 2 Derive the residual

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

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

Step 4 Generate the bootstrapped cost

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

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

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

5 Monte-Carlo experiments

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

$$s_{jm}(\boldsymbol{\theta}) = \int_\alpha \int_\beta \frac{\exp\left(\mathbf{x}_{jm}\beta + p_{jm}\alpha + \xi_{jm}\right)}{\sum_{j=0}^{J_m} \exp\left(\mathbf{x}_{jm}\beta + p_{jm}\alpha + \xi_{jm}\right)} \frac{1}{\sigma_\alpha} \phi\left(\frac{\alpha - \mu_\alpha}{\sigma_\alpha}\right) \frac{1}{\sigma_\beta} \phi\left(\frac{\beta - \mu_\beta}{\sigma_\beta}\right) d\alpha d\beta, \quad (38)$$

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

On the supply-side, we assume firms compete on prices a la differentiated products Bertrand competition, use labor and capital inputs in production and have a Cobb-Douglas production function. Given output, input prices $\mathbf{w} = [w, r]'$ (w is the wage and r is the rental rate of capital), total cost and marginal cost functions are specified as[31]

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

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

[31] In our Monte-Carlo, we assume away the deterministic components of the fixed cost and the measurement error, $e(q, \mathbf{w}, \mathbf{x})$.

Notice that in the above specification, the cost function is homogeneous of degree one in input prices.[32]

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

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

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

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

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

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

$$\upsilon_{jm} = \mu_\upsilon + \sigma_\upsilon \varrho_{\upsilon jm} + \zeta_Q \Phi^{-1}\left(\delta + (1.0 - 2\delta)\frac{Q_m - Q_L}{Q_H - Q_L}\right), \quad \varrho_{\upsilon m} \sim i.i.d.TN\left(0,1\right). \tag{39}$$

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

Importantly, we specify the unobserved quality so as to allow for correlation between ξ_{jm} and input price, variable cost shock and market size. We also let ξ_{jm} be correlated with the observed characteristics of the products other than j. Specifically, we set:

$$\xi_{jm} = \delta_0 + \delta_\xi \varrho_{\xi jm} + \delta_w \varrho_{wm} + \delta_r \varrho_{rm} + \delta_\upsilon \varrho_{\upsilon jm} + \delta_Q \Phi^{-1}\left(\delta + (1.0 - 2\delta)\frac{Q_m - Q_L}{Q_H - Q_L}\right) + \delta_{xo}\frac{1}{3}\sum_{l \neq j} \varrho_{xlm}. \tag{40}$$

where ϱ_ξ is the idiosyncratic component of the demand shock. We set $\delta_l = \frac{1}{2\sqrt{6}}$ for $l \in \{\xi, w, r, \upsilon, Q, xo\}$. By construction, neither input prices, nor observed characteristics of other

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

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

Table 1: Monte-Carlo Parameter Values

μ_α	σ_α	μ_β	σ_β	μ_X	σ_X	α_c	β_c	μ_w	σ_w	μ_r	σ_r
2.0	0.5	1.0	0.2	1.7	0.5	0.4	0.4	1.0	0.2	1.0	0.2
μ_υ	σ_υ	δ_υ		Q_L	Q_H	δ_0	δ_ξ	B			
0.3	0.1	0.2		5.0	10.0	4.0	0.5	1.0			

products can be used as valid instruments for prices in demand estimation. Furthermore, since both demand and variable cost shocks are correlated with market size, one cannot use the variation of market size as an instrument for prices, or for output for cost estimation discussed in Subsection 2.2. We set the deterministic component of the measurement error to be zero and $\sqrt{Var\left(\nu+\varsigma\right)} = \sqrt{\sigma_\nu^2+\sigma_\varsigma^2} = 0.2$.

To solve for the equilibrium price, quantity, and market share for each oligopoly firm, we use golden section search on price.[33]

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

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

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

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

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

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

Table 2: Sample Statistics of Simulated Data.

variables	Mean	Std. Dev
Price (p_m)	4.1035	1.2388
Output (q_m)	1.3948	0.6615
Quality (ξ_m)	4.0079	0.4356
Market Share (s_m)	0.1914	0.0909
Wage (w_m)	1.0069	0.1831
Rent (r_m)	1.0005	0.1947
Cost (C_m)	2.8144	1.0045
x_m	1.7037	0.4652
$\sigma_{\nu+\varsigma} = 0.2$		

since the estimated parameter values are close to their true values, the standard deviations and RMSEs are close to each other as well. Overall, these Monte-Carlo results demonstrate the validity of our approach.[34] In Table 4, we present the results where we allow for the observed characteristics x to be correlated with the unobserved characteristics ξ. That is,

$$\xi_{jm} = \delta_0 + \delta_\xi \varrho_{\xi jm} + \delta_w \varrho_{wm} + \delta_r \varrho_{rm} + \delta_\upsilon \varrho_{\upsilon jm} + \delta_Q \Phi^{-1}\left(\delta + (1.0 - 2\delta)\frac{Q_m - Q_L}{Q_H - Q_L}\right) + \delta_{xo}\frac{1}{3}\sum_{l \neq j} \varrho_{xlm} + \delta_x \varrho_{xjm}. \tag{41}$$

where $\delta_0 = 4.0$, $\delta_\xi = \delta_w = \delta_\upsilon = \delta_Q = \delta_x = 1/\left(2\sqrt{7}\right)$. Now, by construction, no observed variable of the firm can be used as a valid instrument for its prices in demand estimation. What we can see in Table 4 is that the parameters $\boldsymbol{\theta}_c$ are consistently estimated. On the other hand, μ_β is estimated to be around 1.2, much higher than the true coefficient 1.0. The upward bias is due to the positive correlation between the demand shock ξ_{jm} and the random term of the observed characteristics ϱ_{xjm} as specified in Equation (41). However, since rest of the parameters are estimated consistently, the markup of the firms can still be recovered consistently.

In Table 5, we report the results where we set the variation in market size to be zero. We can see that overall, means of the parameter estimates become closer to the true values, and the standard deviations and RMSEs become smaller as sample size increases. In the Monte-Carlo results with the sample size of 200, one out of a hundred simulation/estimation exercises did not converge, so we removed it, and took the sample statistics over 99 parameter estimates. By comparing the results in Table 3, we can see that the standard deviations and the RMSEs are higher than the ones where we had variation in market size. We conclude that even though the variation in market size is not needed, it helps in improving the accuracy of the estimators.

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

Table 3: SNLLS Estimator of Random Coefficient Demand Parameters.

No. of	Sample		$\widehat{\mu}_\alpha$			$\widehat{\sigma}_\alpha$			CPU min.
Markets	Size	No. Poly	Mean	Std. Dev	RMSE	Mean	Std. Dev	RMSE	
50	200	144	-2.1775	0.8230	0.8379	0.4547	0.2146	0.2183	274.0
100	400	171	-2.1953	0.5389	0.5707	0.5472	0.1781	0.1834	788.5
200	800	204	-2.0044	0.1837	0.1828	0.5062	0.0937	0.0934	1795.4
400	1600	256	-2.0203	0.1235	0.1246	0.5037	0.0527	0.0525	2278.4
True			-2.0000			0.5			
No. of	Sample		$\widehat{\mu}_\beta$			$\widehat{\sigma}_\beta$			Obj. Fcn
Markets	Size	No. Poly	Mean	Std. Dev	RMSE	Mean	Std. Dev	RMSE	
50	200	144	1.1296	0.5865	0.5978	0.2583	0.2841	0.2887	1.3919D-2
100	400	171	1.0692	0.2722	0.2796	0.2646	0.2130	0.2216	2.2941D-2
200	800	204	0.9880	0.1178	0.1178	0.2036	0.0846	0.0842	2.9442D-2
400	1600	256	1.0066	0.0857	0.0855	0.2069	0.0563	0.0564	3.3647D-2
True			1.0000			0.2			

$\sigma_{\nu+\varsigma} = 0.2$

Table 4: SNLLS Estimator of Random Coefficient Demand Parameters with correlated x and ξ

No. of	Sample		$\widehat{\mu}_\alpha$			$\widehat{\sigma}_\alpha$			CPU
Markets	Size	No. Poly	Mean	Std. Dev	RMSE	Mean	Std. Dev	RMSE	min
50	200	144	-2.2042	0.9538	0.9708	0.5077	0.1957	0.1948	762.2
100	400	171	-2.0710	0.2716	0.2794	0.5172	0.1173	0.1179	1526.7
200	800	204	-2.0118	0.1505	0.1502	0.5001	0.0656	0.0653	3189.4
400	1600	256	-2.0058	0.0883	0.0880	0.5020	0.0337	0.0336	5770.0
True			-2.0000			0.5			
No. of	Sample		$\widehat{\mu}_\beta$			$\widehat{\sigma}_\beta$			Obj. Fct.
Markets	Size	No. Poly	Mean	Std. Dev	RMSE	Mean	Std. Dev	RMSE	
50	200	144	1.3146	0.6606	0.7287	0.2314	0.2558	0.2564	1.3923D-2
100	400	171	1.2221	0.1582	0.2723	0.2135	0.1107	0.1109	2.3205D-2
200	800	204	1.1968	0.0897	0.2160	0.2015	0.0785	0.0781	2.9529D-2
400	1600	256	1.1904	0.0560	0.1983	0.1978	0.0519	0.0517	3.3751D-2
True			1.0000			0.2			

$\sigma_{\nu+\varsigma} = 0.2$; x and ξ are correlated.

Next, we consider the case where market size is not observable, and needs to be estimated. We assume that the equation for market size is defined as follows.

$$Q_m^* = \lambda_0 + \lambda_1 z_m \tag{42}$$

where Q_m^* is the unobserved market size, and we set $z_m = Q_m$. We set $\lambda_0 = 0$, $\lambda_1 = 1$. Then, the true market share vector is $\mathbf{s}_m = \mathbf{q}_m / Q_m^*$. We keep the BLP market share equation as specified in Equation (38) except that market size is unobservable and therefore, parameters λ_0 and λ_1 need to be jointly estimated. Note that market shares $\mathbf{s}_m$ are unobservable as well. In Table

Table 5: SNLLS Estimator of Random Coefficient Demand Parameters: No Market Size Variation

No. of	Sample		$\widehat{\mu}_\alpha$			$\widehat{\sigma}_\alpha$			CPU
Markets	Size	No. Poly	Mean	Std. Dev	RMSE	Mean	Std. Dev	RMSE	min
50	200	144	-2.3739	1.2429	1.2919	0.4413	0.3349	0.3383	821.4
100	400	171	-2.2805	0.6776	0.7302	0.5229	0.2151	0.2152	2013.2
200	800	204	-2.0092	0.1966	0.1958	0.4989	0.0958	0.0954	3622.3
400	1600	256	-2.0400	0.1551	0.1595	0.5001	0.0592	0.0589	6249.1
True			-2.0000			0.5			
No. of	Sample		$\widehat{\mu}_\beta$			$\widehat{\sigma}_\beta$			Obj. Fct.
Markets	Size	No. Poly	Mean	Std. Dev	RMSE	Mean	Std. Dev	RMSE	
50	200	144	1.2317	0.7813	0.8111	0.2209	0.2096	0.2096	1.3857D-2
100	400	171	1.1369	0.3768	0.3991	0.2897	0.2089	0.2263	2.3007D-2
200	800	204	1.0026	0.1130	0.1125	0.1969	0.1095	0.1089	2.9524D-2
400	1600	256	1.0225	0.0884	0.0908	0.2050	0.0728	0.0726	3.3901D-2
True			1.0000			0.2			

$\sigma_{\nu+\varsigma} = 0.2$.

Table 6: SNLLS Estimator of BLP Demand Model with unobservable Market Size.

		A			B		
		z_m not in market share function			z_m in market share function		
Parameter	True	Mean	Std. Dev	RMSE	Mean	Std.Dev	RMSE
$\widehat{\mu}_\alpha$	-2.0	-2.0247	0.1252	0.1270	-1.9642	0.1300	0.1342
$\widehat{\sigma}_\alpha$	0.5	0.5098	0.0432	0.0406	0.5064	0.0499	0.0501
$\widehat{\mu}_\beta$ (β)	1.0	1.0047	0.1004	0.1000	0.9963	0.2540	0.2527
$\widehat{\sigma}_\beta$	0.2	0.2079	0.0569	0.0572	0.1991	0.0734	0.0730
λ_c	0.0	-0.0341	0.2491	0.2502	0.0019	0.0962	0.0957
λ_z	1.0	1.0203	0.0725	0.0750	1.0597	0.1364	0.1483
Market size: 500, sample size:2000, Number of polynomials:256.							

6, panel A, we present the statistics of the parameter estimates that were generated from 100 repeated simulation/estimation exercises, based on the model used in Table 3 with unobservable market size. In addition, in panel B, we report the results where we set $x_{jm} = \Phi^{-1}(z_m)$ in Equation (38). As we can see, in both cases, means of the parameter estimates are close to the true values.

In Table 7, we compare the estimated parameters using our two-step SNLLS method with the standard IV approach using instruments that are commonly used in the literature. These are: wage, rental rate and observed product characteristics of own and rival firms and their interactions. Results show that in a parameter setup where the instruments are invalid, while our two-step SNLLS estimates are consistent, the IV estimates of the demand parameters are biased.

In the first row (SNLLS 1) of Table 7, we show results of our estimator when the demand

Table 7: SNLLS and IV Estimators.

	No. of	Sample	$\widehat{\mu}_\alpha$			$\widehat{\sigma}_\alpha$			CPU
	Markets	Size	Mean	Std. Dev	RMSE	Mean	Std. Dev	RMSE	min
SNLLS 1	500	2000	-2.0247	0.0845	0.0877	0.5049	0.0362	0.0363	9400.8
SNLLS 2	500	2000	-2.0335	0.0865	0.0924	0.5072	0.0334	0.0340	7932.0
IV1	500	2000	-1.9779	0.0863	0.0887	0.4749	0.0692	0.0733	11571.3
IV2	500	2000	-1.6145	0.0877	0.3953	0.6087	0.0513	0.1200	13607.4
IV3	500	2000	-1.4533	0.0778	0.5522	0.1570	0.0555	0.3474	11793.1
IV4	500	2000	-1.2784	0.0861	0.7267	0.3935	0.0545	0.1195	12875.1
	True		-2.0000			0.5			
	No. of	Sample	$\widehat{\mu}_\beta$			$\widehat{\sigma}_\beta$			Obj. Fct.
	Markets	Size	Mean	Std. Dev	RMSE	Mean	Std. Dev	RMSE	
SNLLS 1	500	2000	1.0114	0.0481	0.0492	0.2005	0.0408	0.0406	3.4868D-2
SNLLS 2	500	2000	1.0157	0.0518	0.0539	0.2038	0.0425	0.0424	3.5042D-2
IV 1	500	2000	0.9824	0.0520	0.0546	0.1967	0.1466	0.1459	8.3437D-4
IV 2	500	2000	0.6197	0.0459	0.3830	0.2065	0.1483	0.1477	1.4978D-3
IV 3	500	2000	0.7751	0.0304	0.2269	0.2032	0.1113	0.1108	9.4921D-4
IV 4	500	2000	0.5334	0.0355	0.4679	0.1587	0.1197	0.1261	1.2029D-3
	True		1.0000			0.2			

$\sigma_{\nu+\varsigma} = 0.2$.
SNLLS 1: instruments are valid: $\delta_\xi = 0.5, \delta_w = \delta_r = \delta_\upsilon = \delta_Q = \delta_{xo} = 0$.
SNLLS 2: Input prices are correlated with demand shock: $\delta_\xi = \delta_w = \delta_r = \frac{1}{2\sqrt{3}}, \delta_\upsilon = \delta_Q = \delta_{xo} = 0$.
IV 1: instruments are valid: $\delta_\xi = 0.5, \delta_w = \delta_r = \delta_\upsilon = \delta_Q = \delta_{xo} = 0$.
IV 2: input prices are correlated with demand shock: $\delta_\xi = \frac{1}{2\sqrt{1.08}},\ \delta_w = \delta_r = \frac{1}{10\sqrt{1.08}}, \delta_\upsilon = \delta_Q = \delta_{xo} = 0$
IV 3: rival product observed characteristics are correlated with demand shock.
$\delta_\xi = \frac{1}{2\sqrt{1.04}}, \delta_{xo} = \frac{1}{10\sqrt{1.04}}, \delta_w = \delta_r = \delta_\upsilon = \delta_Q = 0$.
IV 4: Input price, variable cost shock, market size and rival product observed characteristics are correlated with demand shock.
$\delta_\xi = \frac{1}{2\sqrt{1.20}}, \delta_w = \delta_r = \delta_\upsilon = \delta_Q = \delta_{xo} = \frac{1}{10\sqrt{1.20}}$

shock is orthogonal to the other variables, (e.g., $\delta_\xi = 0.5,\ \delta_w = \delta_r = \delta_\upsilon = \delta_Q = \delta_{xo} = 0$), so that the instruments are valid. As we can see, the two-step SNLLS estimated coefficients are close to the true values, as are the IV estimates presented in the third row (IV1). However, the standard deviations of the IV estimates of σ_α and σ_β are higher than those of the two-step SNLLS estimates. That is, higher order interactions of the instruments may be needed to estimate σ_α and σ_β using instruments as accurately as the two-step SNLLS ones. Next, in the fourth row (IV2), we show the statistics of the IV estimates when the input prices are not valid instruments. That is, we set the demand shock and the correlation between the demand shock and other variables to be $\delta_\xi = 1/\left(2\sqrt{1.08}\right)$ $\delta_\upsilon = \delta_Q = \delta_{x0} = 0$,$\delta_w = \delta_r = 0.2\delta_\xi$[35]. We can see that while the two-step SNLLS estimates given in the second row (SNLLS2) are close to the

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

true values, the IV estimated price coefficient is smaller in value than the true parameter -2.0, i.e., we have an upward bias. The positive direction of bias is to be expected because the error term, which is the unobserved quality, is set up to be positively correlated with the instruments. Notice also that the coefficient estimate on the observed characteristics is downwardly biased, and the heterogeneity parameter of price effect, σ_α is upwardly biased.

Next, in row IV3, we present the IV results where the rival firms' observed product characteristics $\mathbf{X}_{-jm}$ are correlated with own unobserved characteristics ξ_{jm}. That is, we set $\delta_{xo} = \frac{1}{10\sqrt{1.04}}, \delta_\xi = \frac{1}{2\sqrt{1.04}}, \delta_w = \delta_r = \delta_\upsilon = \delta_Q = 0$. Hence, the observed characteristics of rival firms cannot be used as instruments for own price. Results show that the IV-estimated price coefficient again has a positive bias. The parameter μ_β is again estimated with a negative bias, and so is σ_α, unlike results in Table 3, where we show the two-step SNLLS estimator delivers consistent parameter estimates even if the demand shock is correlated with rival product characteristics.

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

Overall, we conclude that our two-step SNLLS estimator provides parameter estimates that have little bias even in situations where the commonly-used instruments are invalid and thus the IV estimates are biased. In addition, our two-step SNLLS estimator performs well even when market size is not observable, and the variable that determines market size is correlated with the demand shock or enters directly in the market share equation. Furthermore, with similar convergence criteria, the CPU minutes required for the two-step SNLLS estimator are less than the CPU minutes required for the IV estimates. It is worthwhile noting that this does not automatically imply that our two-step SNLLS estimator is computationally superior to the IV estimates, since the IV-GMM procedure has to estimate the additional parameter μ_β, whereas in the two-step SNLLS μ_β is obtained simply by OLS. Nonetheless, we tentatively conclude that our sieve-estimation procedure does not seem to impose excessive computational burden.

6 Empirical application to U.S. banking industry

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

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

where $\mathbf{x}_{jm}$ is a vector of observed characteristics of bank j in market m, which consists of log of number of branches of bank j, log of number of markets served by bank j and log of one plus bank age; r_{djm} is the deposit interest rate of bank j in market m net of the service charge, and ξ_{jm} is the unobserved characteristic of bank j in market m. Finally, ϵ_{ijm} is the random residual term in the utility function, which is assumed to be i.i.d. Extreme-Value distributed.

Then, the market share of deposits for bank j in market m is

$$s_{jm} = \int_{\alpha}\int_{\boldsymbol{\beta}} \frac{\exp\left(\mathbf{x}_{jm}\boldsymbol{\beta} + r_{djm}\alpha + \xi_{jm}\right)}{\left[1 + \sum_{k=1}^{J}\exp\left(\mathbf{x}_{km}\boldsymbol{\beta} + r_{dkm}\alpha + \xi_{km}\right)\right]} \left[\prod_{l=1}^{K}\frac{1}{\sigma_{\beta l}}\phi\left(\frac{\beta_l}{\sigma_{\beta l}}\right)d\beta_l\right]\frac{1}{\sigma_{\alpha}}\phi\left(\frac{\alpha-\mu_{\alpha}}{\sigma_{\alpha}}\right)d\alpha$$

where ϕ is the density function of the standard normal distribution. We can rewrite the above equation by applying the change of variables as follows:

$$s_{jm} = \int_{\alpha}\int_{\boldsymbol{\beta}} \frac{exp\left(\left(\mathbf{x}_{jm}\circ\boldsymbol{\sigma}_{\boldsymbol{\beta}}\right)\boldsymbol{\beta} + r_{djm}\left(\sigma_{\alpha}\alpha + \mu_{\alpha}\right) + \xi_{jm}\right)}{1 + \sum_{k} exp\left(\left(\mathbf{x}_{km}\circ\boldsymbol{\sigma}_{\boldsymbol{\beta}}\right)\boldsymbol{\beta} + r_{dkm}\left(\sigma_{\alpha}\alpha + \mu_{\alpha}\right) + \xi_{km}\right)}\phi\left(\boldsymbol{\beta}\right)d\boldsymbol{\beta}\phi\left(\alpha\right)d\alpha,$$

and the market share of the outside option is

$$s_{0m} = \int_{\alpha}\int_{\boldsymbol{\beta}} \frac{exp\left(\xi_{0m}\right)}{1 + \sum_{k} exp\left(\left(\mathbf{x}_{km}\circ\boldsymbol{\sigma}_{\boldsymbol{\beta}}\right)\boldsymbol{\beta} + r_{dkm}\left(\sigma_{\alpha}\alpha + \mu_{\alpha}\right) + \xi_{km}\right)}\phi\left(\boldsymbol{\beta}\right)d\boldsymbol{\beta}\phi\left(\alpha\right)d\alpha.$$

Note that we expect the sign of μ_{α} to be positive because consumers prefer higher deposit interest rate. We follow Dick (2008) and let the outside option be depositing in credit unions.

Notice that since individuals receive deposit interest rate on their deposits, banks need to loan out or invest the deposits to earn any revenues. Thus, we assume revenue to be $\left(r_{jm} - r_{djm}\right)q_{jm}$, where r_{jm} is the interest rate earned by bank j in market m and Q_m as earlier is the market size variable. In this empirical example, the market size is the total number of deposits (including credit unions). We set the interest rate to be $r_{jm} = r$, where r is the interest rate on the government treasury notes in January 2002.[36] Then, the marginal revenue of deposit is

$$MR_{jm}(\boldsymbol{\theta}) = (r - r_{djm}) - s_{jm}\left[\frac{\partial s_j(\mathbf{r}_{dm}, \mathbf{X}_m, \boldsymbol{\xi}_m; \boldsymbol{\theta})}{\partial \mathbf{r}_{dm}}\right]^{-1}$$

where

$$\begin{aligned}
&\frac{\partial s_j(\mathbf{r}_{dm}, \mathbf{X}_m, \boldsymbol{\xi}_m; \boldsymbol{\theta})}{\partial \mathbf{r}_{dm}} \\
= &\int_\alpha \int_{\boldsymbol{\beta}} \left[\frac{exp\left((\mathbf{x}_{jm} \circ \boldsymbol{\sigma}_{\boldsymbol{\beta}})\boldsymbol{\beta} + r_{djm}(\sigma_\alpha \alpha + \mu_\alpha) + \xi_{jm}\right)}{\left[1 + \sum_k exp\left((\mathbf{x}_{km} \circ \boldsymbol{\sigma}_{\boldsymbol{\beta}})\boldsymbol{\beta} + r_{dkm}(\sigma_\alpha \alpha + \mu_\alpha) + \xi_{km}\right)\right]^2}\right. \\
&\times \left.\left[1 + \sum_{k \neq j} exp\left((\mathbf{x}_{km} \circ \boldsymbol{\sigma}_{\boldsymbol{\beta}})\boldsymbol{\beta} + r_{dkm}(\sigma_\alpha \alpha + \mu_\alpha) + \xi_{km}\right)\right]\right](\sigma_\alpha \alpha + \mu_\alpha)\phi(\boldsymbol{\beta})\, d\boldsymbol{\beta}\phi(\alpha)\, d\alpha.
\end{aligned} \tag{43}$$

We define the variable cost of a bank to be the total annual labor cost. That is, we assume we can measure the variable cost, and thus do not need to specify the fixed cost. We also assume away the deterministic measurement error component of the cost data. We define the cost function as $C(q_{jm}, w_m, \mathbf{x}_{jm}, \upsilon_{jm})$, where q_{jm} is the total deposits of bank j in market m, w_m is the wage level in market m, and υ_{jm} is the variable cost shock of bank j in market m. Since the cost function is homogeneous of degree one in wage, we can rewrite the cost function as follows:

$$C(q_{jm}, w_m, \mathbf{x}_{jm}, \upsilon_{jm}) = w_m \varphi\left(q_{jm}, \mathbf{x}_{jm}, \frac{MR_{jm}}{w_m}\right) = w_m \sum_l \gamma_l \psi_l\left(q_{jm}, \mathbf{x}_{jm}, \frac{MR_{jm}}{w_m}\right),$$

where ψ_l, $l = 1, \ldots$ are the basis functions of the sieve and γ_l are the coefficients. Hence, we estimate $\boldsymbol{\theta}_{c0}$ by minimizing the following objective function:

$$\frac{1}{\sum J_m} \sum_{j,m}\left[\frac{C_{jm}}{w_m} - \sum_{l=1}^{\Gamma_{L_M}(T)} \gamma_l \psi_l\left(q_{jm}, \mathbf{x}_{jm}, \frac{MR_j(\mathbf{r}_{dm}, r, \mathbf{s}_m, \mathbf{X}_m, \boldsymbol{\theta}_c)}{w_m}\right)\right]^2. \tag{44}$$

[36] There are three possible choices of variables for the interest rate r_{jm}. One could use the interest rate on assets such as government bonds, loan interest rate, or a basket of the rate of returns on loans and other financial assets. Note that setting the interest rate to be the loan interest rate would raise an additional endogeneity issue of bank lending. Since one of the important goals of this empirical analysis is to demonstrate the validity of our estimator, we decided to focus on the deposit side so that our results are comparable to the literature on deposit demand estimation. We use the interest rate on the government treasury notes in January 2002, since one can reasonably assume it to be exogenous. We believe that an interesting future direction of research would be an empirical analysis of the banking industry where both deposits and loans are endogenous. To the best of our knowledge, structural empirical analysis of banking that includes both deposits and loans (or investments) is scarce in the literature.

with respect to both $\boldsymbol{\theta}_c$ and $\boldsymbol{\gamma}$. In the second step, we estimate $\boldsymbol{\mu}_{\boldsymbol{\beta}}$ by OLS where the dependent variable is $\delta_{jm}\left(\widehat{\boldsymbol{\theta}}_c\right) - r_{djm}\widehat{\mu}_\alpha$ and the vector of independent variables is $\mathbf{x}_{jm}$. We obtain $\delta_{jm}\left(\widehat{\boldsymbol{\theta}}_c\right)$ using the inversion algorithm explained earlier. That is, the equation we estimate is:

$$\delta_{jm}\left(\widehat{\boldsymbol{\theta}}_c\right) - r_{djm}\widehat{\mu}_\alpha = \mathbf{x}_{jm}\boldsymbol{\mu}_{\boldsymbol{\beta}} + \xi_{jm}. \tag{45}$$

We calculate the standard errors of the parameter estimates of $\widehat{\boldsymbol{\theta}}_c$ by bootstrap, i.e., resampling the residuals in Equation (44).

We use data on banks in the year 2002 from similar sources as Dick (2008). That is, we obtained the branch level information, such as the number of branches, the total deposits for each branch and the branch location from the FDIC (Federal Deposit Insurance Corporation). We collected data on bank characteristics from the balance sheet and income statements of banks in the UBPR (Uniform Bank Performance Report) from the FFIEC (Federal Financial Institutions Examination Council). Information on county level weekly wage is obtained from the annualized data of Quarterly Census of Employment and Wages (QCEW) data files managed by the Bureau of Labor Statistics. The definition of a market is either metropolitan statistical areas (MSA's) or counties if the area is not included in any of the MSA's. In the Appendix, we provide the sample statistics and discuss some additional data and estimation issues.

By closely inspecting the data, we noticed that in some markets, credit unions seem to be effectively nonexistent as an outside option. Therefore, we removed the markets in which market share of credit unions is less than 1%. Furthermore, we use the cost data of only those banks that operate in a single market. We did so because banks who operate in multiple markets may not exercise third degree discrimination, which then violates Assumption 6. Finally, for the sake of reducing the computational burden, we restrict the sample to markets where there are 40 banks or less. In the final sample, the number of banks whose cost data we use is 2067, whereas the number of all banks is 3230. That is, about two thirds of the banks are single-market banks. As we can see in the sample statistics and the results in Table 8, the number of banks that operate only in a single market is high enough for identification of $\boldsymbol{\theta}_c$. It is important to remember that $\boldsymbol{\theta}_c$ is identified based on the assumption of independence of the measurement error to other variables. Therefore, using cost data of only those banks that serve one market does not result in any selection bias for estimation of $\boldsymbol{\theta}_c$, as long as the data of all banks are available for the variables in the marginal revenue function. For the estimation of $\boldsymbol{\mu}_{\boldsymbol{\beta}}$, selection matters, and thus we use all banks in the data.

Table 8: Parameter estimates of deposit demand

Coefficients		mean	std. dev	mean	std. dev
		A: Two-step SNLLS		B: IV-GMM	
deposit interest rate	μ_α	31.92***	(0.663)	-208.8***	(1.767)
	σ_α	9.829E-3	(0.0313)	2.697***	(0.0257)
$\boldsymbol{\beta}_1$ log number of branches	$\mu_{\beta 1}$	4.256***	(0.0534)	2.067	(12.31)
	$\sigma_{\beta 1}$	0.3357***	(0.0319)	9.646	(16.35)
$\boldsymbol{\beta}_2$ log number of markets	$\mu_{\beta 2}$	0.5539***	(0.0209)	0.0858	(1.694)
	$\sigma_{\beta 2}$	5.044E-4	(6.20E-3)	1.481E-3	(1.477E-3)
$\boldsymbol{\beta}_3$ log age of bank plus one	$\mu_{\beta 3}$	4.802***	(0.0450)	1.356	(8.027)
	$\sigma_{\beta 3}$	2.892E-3	(0.0109)	0.1129	(0.2637)

Standard errors are in parentheses. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Note that since we use only those banks that operate in a single market for cost estimation, we do not include the number of markets served in the cost function.

We present our results in Table 8. In panel A, we report the results where we use the two-step SNLLS procedure. In panel B, we report the results where we use the IV-GMM procedure. We used regional wage, regional housing price index, observed product characteristics (log number of branches, log number of markets, log age plus one) of rival banks in the market, and the interactions of those variables as instruments.

Our estimated price coefficient is around 32 and the average price elasticity is 1.64. The proportion of banks whose elasticity is less than one is 4 %. In contrast the IV-estimated price coefficient on deposit interest rate in Dick (2008) ranges from 54.19 to 100.23, depending on the inclusion of the bank/market/state fixed effects. One potential difference between our analysis and that by Dick (2008) is that they only use banks in MSAs, which are predominantly urban. Indeed, our estimated price effects are closer to the ones in Ho and Ishii (2012), who also include rural markets in their analysis and find that the own-price elasticity is smaller in rural markets. The price elasticity is expected to be lower in rural markets, where the distance to branches of other banks is likely to be greater, and our data includes rural markets as well.

We see in panel B of the table that the IV estimated price coefficient is negative and significant. It is unintuitive because it implies that a higher deposit interest rate reduces deposits. Furthermore, the IV estimated parameters $(\boldsymbol{\mu_\beta}, \boldsymbol{\sigma_\beta})$ are all insignificant.[37], whereas in panel A, the two-step SNLLS estimated coefficients on observed characteristics are all positive, as is

[37]We have tried various versions of setups and instruments, both logit and BLP demand. What we find is that among the cases we tried, only the logit specification with a relatively small number of instruments resulted in the price coefficients with the plausible sign. What we infer from those results is: Since BLP has more parameters that need to be estimated than the logit specification, it requires more instruments, and in this particular case, some of the instruments or some of the polynomials of the instruments are invalid.

intuitive, and significant.

7 Conclusion

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

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

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

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

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

References

Bayer, P., F. Ferreira, and R. McMillan (2007): "A Unified Framework for Measuring Preferences for Schools and Neighborhoods," *Journal of Political Economy*, 115, 588–638.

Berry, S. T. (1994): "Estimating Discrete-Choice Models of Product Differentiation," *RAND Journal of Economics*, 25, 242–262.

Berry, S. T. and P. Haile (2014): "Identification in Differentiated Products Markets Using Market Level Data," *Econometrica*, forthcoming.

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

——— (1999): "Voluntary Export Restraints on Automobiles: Evaluating a Strategic Trade Policy," *American Economic Review*, 89, 400–430.

BIERENS, H. (2014): "Consistent and Asymptotic Normality of Sieve ML Estimators Under Low-Level Conditions," *Econometric Theory*, 30, 1021–1076.

BRESNAHAN, T. F. AND P. C. REISS (1991): "Entry and Competition in Concentrated Markets," *Journal of Political Economy*, 99, 977–1009.

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

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

CHU, C. S. (2010): "The Effect of Satellite Entry on Cable Television Prices and Product Quality," *Rand Journal of Economics*, 41, 730–764.

CLAY, K. AND W. TROESKEN (2003): "Further Tests of Static Oligopoly Models: Whiskey, 1882-1898," *Journal of Industrial Economics*, 51, 151–166.

CRAWFORD, G. S. (2012): "Accommodating Endogenous Product Choices: A Progress Report," *International Journal of Industrial Organization*, 30, 315–320.

CRAWFORD, G. S. AND A. YURUKOGLU (2012): "The Welfare Effects of Bundling in Multichannel Television Markets," *American Economic Review*, 102, 301–317.

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

DICK, A. A. (2008): "Demand Estimation and Consumer Welfare in the Banking Industry," *Journal of Banking and Finance*, 32, 1661–1676.

DUBE, J.-P., J. T. FOX, AND C.-L. SU (2012): "Improving the Numerical Performance of Static and Dynamic Aggregate Discrete Choice Random Coefficients Demand Estimation," *Econometica*, 5, 2231–2267.

FAN, Y. (2013): "Ownership Consolidation and Product Characteristics: A Study of the US Daily Newspaper Market." *American Economic Review*, 103, 1598–1628.

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

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

Fu, C. and K. I. Wolpin (2018): “Structural Estimation of a Becker-Ehrlich Equilibrium Model of Crime: Allocating Police Across Cities to Reduce Crime,” *Review of Economic Studies*, 85, 2097–2138.

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

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

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

Ho, K. and J. Ishii (2012): “Location and Competition in Retail Banking,” *International Journal of Industrial Organization*, 29, 537–546.

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

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

Knittel, C. and K. Metaxoglou (2012): “Estimation of Random Coefficient Demand Models: Challenges, Difficulties, and Warnings,” *Review of Economics and Statistics*, forthcoming.

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

MacKaye, A. and N. H. Miller (2018): “Instrument-Free Demand Estimation,” *mimeo, Harvard University.*

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

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

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

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

Petrin, A. and B. Seo (2016): "Identification and Estimation of Discrete Choice Demand Models when Observed and Unobserved Characteristics are Correlated," *University of Minnesota Working Paper*.

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

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

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

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

A Appendix

A.1 Identification of the random coefficient BLP model with endogeneity. Proof of Proposition 1.

We prove that given Assumptions 1-8, the BLP demand function satisfies Condition 1. Then, from Lemma 2, identification follows.

Consider two firms j and $j^{\dagger}$ in different markets. The first market has demand side observables $(\mathbf{X}, \mathbf{p}, \mathbf{s})$ where $\mathbf{X}$ is a $J \times K$ matrix of observed product characteristics, $\mathbf{p}$ is the $J \times 1$ price vector and $\mathbf{s}$ is the $J \times 1$ market share vector. Let $\left(\mathbf{X}^{\dagger}, \mathbf{p}^{\dagger}, \mathbf{s}^{\dagger}\right)$ denote the corresponding observables for the second market where the number of products/firms is $J^{\dagger}$. We start by assuming that these two firms satisfy 1 and 2 of Condition 1. We then prove that among those pairs, there exists one that satisfies 3 of Condition 1. We denote $\Phi\left(\right)$ to be the standard normal distribution function and $\phi\left(\right)$ to be its density function. Then, the market shares of firms j and $j^{\dagger}$ are respectively:

$$
\begin{aligned}
s_j &= \int_{\alpha}\int_{\boldsymbol{\beta}} \frac{\exp\left(\mathbf{x}_j\boldsymbol{\beta} + p_j\alpha + \xi_{j,0}\right)}{\left[1+\sum_{l=1}^{J}\exp\left(\mathbf{x}_l\boldsymbol{\beta} + p_l\alpha + \xi_{l,0}\right)\right]} \left[\prod_{k=1}^{K}\frac{1}{\sigma_{\beta 0k}}\phi\left(\frac{\beta_k - \mu_{\beta 0k}}{\sigma_{\beta 0k}}\right)d\beta_k\right]\frac{1}{\sigma_{\alpha 0}}\phi\left(\frac{\alpha-\mu_{\alpha 0}}{\sigma_{\alpha 0}}\right)d\alpha \\
s_{j^{\dagger}}^{\dagger} &= \int_{\alpha}\int_{\boldsymbol{\beta}} \frac{\exp\left(\mathbf{x}_{j^{\dagger}}^{\dagger}\boldsymbol{\beta} + p_{j^{\dagger}}^{\dagger}\alpha + \xi_{j^{\dagger},0}^{\dagger}\right)}{\left[1+\sum_{l=1}^{J^{\dagger}}\exp\left(\mathbf{x}_l^{\dagger}\boldsymbol{\beta} + p_l^{\dagger}\sigma_{\alpha 0}\alpha + \xi_{l,0}^{\dagger}\right)\right]} \left[\prod_{k=1}^{K}\frac{1}{\sigma_{\beta 0k}}\phi\left(\frac{\beta_k - \mu_{\beta 0k}}{\sigma_{\beta 0k}}\right)d\beta_k\right]\frac{1}{\sigma_{\alpha 0}}\phi\left(\frac{\alpha-\mu_{\alpha 0}}{\sigma_{\alpha 0}}\right)d\alpha.
\end{aligned}
$$

Now, denote, $\eta_{\alpha 0} = \mu_{\alpha 0}/\sigma_{\alpha 0}$. Then, by a change of variables such that $\alpha^* = \alpha/\sigma_{\alpha 0} - \eta_{\alpha 0}$ and $\beta_j^* = \beta_j/\sigma_{\beta 0j} - \mu_{\beta 0k}/\sigma_{\beta 0j}$, we obtain

$$
\begin{aligned}
s_j &= \int_{\alpha^*}\int_{\boldsymbol{\beta}^*} \frac{\exp\left(\left(\mathbf{x}_j \circ \boldsymbol{\sigma}_{\boldsymbol{\beta}0}\right)\boldsymbol{\beta}^* + p_j\sigma_{\alpha 0}\left(\alpha^* + \eta_{\alpha 0}\right) + \mathbf{x}_j\boldsymbol{\mu}_{\boldsymbol{\beta}0} + \xi_{j,0}\right)}{\left[1+\sum_{l=1}^{J}\exp\left(\left(\mathbf{x}_l \circ \boldsymbol{\sigma}_{\boldsymbol{\beta}0}\right)\boldsymbol{\beta}^* + p_l\sigma_{\alpha 0}\left(\alpha^* + \eta_{\alpha 0}\right) + \mathbf{x}_l\boldsymbol{\mu}_{\boldsymbol{\beta}0} + \xi_{l,0}\right)\right]}\phi\left(\boldsymbol{\beta}^*\right)d\boldsymbol{\beta}^*\phi\left(\alpha^*\right)d\alpha^* \\
s_{j^{\dagger}}^{\dagger} &= \int_{\alpha^*}\int_{\boldsymbol{\beta}^*} \frac{\exp\left(\left(\mathbf{x}_{j^{\dagger}}^{\dagger} \circ \boldsymbol{\sigma}_{\boldsymbol{\beta}0}\right)\boldsymbol{\beta}^* + p_{j^{\dagger}}^{\dagger}\sigma_{\alpha 0}\left(\alpha^* + \eta_{\alpha 0}\right) + \mathbf{x}_{j^{\dagger}}^{\dagger}\boldsymbol{\mu}_{\boldsymbol{\beta}0} + \xi_{j^{\dagger},0}^{\dagger}\right)}{\left[1+\sum_{l=1}^{J^{\dagger}}\exp\left(\left(\mathbf{x}_l^{\dagger} \circ \boldsymbol{\sigma}_{\boldsymbol{\beta}0}\right)\boldsymbol{\beta}^* + p_l^{\dagger}\sigma_{\alpha 0}\left(\alpha^* + \eta_{\alpha 0}\right) + \mathbf{x}_{l^{\dagger}}^{\dagger}\boldsymbol{\mu}_{\boldsymbol{\beta}0} + \xi_{l,0}^{\dagger}\right)\right]}\phi\left(\boldsymbol{\beta}^*\right)d\boldsymbol{\beta}^*\phi\left(\alpha^*\right)d\alpha^*
\end{aligned}
$$

where $\mathbf{x}_j \circ \boldsymbol{\sigma}_{\boldsymbol{\beta}0} = \left(x_{j1}\sigma_{\boldsymbol{\beta}01}, \ldots, x_{jK}\sigma_{\boldsymbol{\beta}0K}\right)$, and $\phi\left(\boldsymbol{\beta}^*\right) \equiv \prod_{k=1}^{K}\phi\left(\beta_k\right)$ is the joint standard normal density function. We then denote α^* to be α and $\boldsymbol{\beta}^*$ to be $\boldsymbol{\beta}$, and $\mathbf{x}_j\boldsymbol{\mu}_{\boldsymbol{\beta}} + \xi_j$ to be ξ_j (similarly for $\mathbf{x}_{j^\dagger}^{\dagger}\boldsymbol{\mu}_{\boldsymbol{\beta}}+\xi_{j^\dagger}^{\dagger}$ to be $\xi_{j^\dagger}^{\dagger}$). Then,

$$
\begin{aligned}
s_j &= \int_\alpha\int_{\boldsymbol{\beta}} \frac{\exp\left(\left(\mathbf{x}_j\circ\boldsymbol{\sigma}_{\boldsymbol{\beta}0}\right)\boldsymbol{\beta}+p_j\sigma_{\alpha 0}\left(\alpha+\eta_{\alpha 0}\right)+\xi_{j,0}\right)}{\left[1+\sum_{l=1}^{J}\exp\left(\left(\mathbf{x}_l\circ\boldsymbol{\sigma}_{\boldsymbol{\beta}0}\right)\boldsymbol{\beta}+p_l\sigma_{\alpha 0}\left(\alpha+\eta_{\alpha 0}\right)+\xi_{l,0}\right)\right]}\phi\left(\boldsymbol{\beta}\right)d\boldsymbol{\beta}\phi\left(\alpha\right)d\alpha \\
s_{j^\dagger}^{\dagger} &= \int_\alpha\int_{\boldsymbol{\beta}} \frac{\exp\left(\left(\mathbf{x}_{j^\dagger}^{\dagger}\circ\boldsymbol{\sigma}_{\boldsymbol{\beta}0}\right)\boldsymbol{\beta}+p_{j^\dagger}^{\dagger}\sigma_{\alpha 0}\left(\alpha+\eta_{\alpha 0}\right)+\xi_{j^\dagger,0}^{\dagger}\right)}{\left[1+\sum_{l=1}^{J^\dagger}\exp\left(\left(\mathbf{x}_l^{\dagger}\circ\boldsymbol{\sigma}_{\boldsymbol{\beta}0}\right)\boldsymbol{\beta}+p_l^{\dagger}\sigma_{\alpha 0}\left(\alpha+\eta_{\alpha 0}\right)+\xi_{l,0}^{\dagger}\right)\right]}\phi\left(\boldsymbol{\beta}\right)d\boldsymbol{\beta}\phi\left(\alpha\right)d\alpha
\end{aligned}
$$

Notice now that the parameters that enter in the market share function are $(\eta_\alpha,\sigma_\alpha,\boldsymbol{\sigma}_{\boldsymbol{\beta}})$, which corresponds to $\boldsymbol{\theta}_c=(\mu_\alpha,\sigma_\alpha,\boldsymbol{\sigma}_{\boldsymbol{\beta}})$. Then,

$$
\begin{aligned}
&\frac{\partial s_j}{\partial p_j} = \\
&\int_\alpha\int_\beta \frac{\exp\left(\left(\mathbf{x}_j\circ\boldsymbol{\sigma}_{\boldsymbol{\beta}0}\right)\boldsymbol{\beta}+p_j\sigma_{\alpha 0}\left(\alpha+\eta_{\alpha 0}\right)+\xi_{j,0}\right)\left[1+\sum_{l\neq j}\exp\left(\left(\mathbf{x}_l\circ\boldsymbol{\sigma}_{\boldsymbol{\beta}0}\right)\boldsymbol{\beta}+p_l\sigma_{\alpha 0}\left(\alpha+\eta_{\alpha 0}\right)+\xi_{l,0}\right)\right]}{\left[1+\sum_{l=1}^{J}\exp\left(\left(\mathbf{x}_l\circ\boldsymbol{\sigma}_{\boldsymbol{\beta}0}\right)\boldsymbol{\beta}+p_l\sigma_{\alpha 0}\left(\alpha+\eta_{\alpha 0}\right)+\xi_{l,0}\right)\right]^2} \\
&\times\sigma_{\alpha 0}\left(\alpha+\eta_{\alpha 0}\right)\phi\left(\boldsymbol{\beta}\right)d\boldsymbol{\beta}\boldsymbol{\phi}\left(\boldsymbol{\alpha}\right)d\alpha
\end{aligned}
$$

and a similar expression holds for $\partial s_{j^\dagger}^{\dagger}/\partial p_{j^\dagger}$.

Note that if we let

$$
\boldsymbol{\Sigma}_{\boldsymbol{\beta}} = \begin{bmatrix} \sigma_{\boldsymbol{\beta}1} & 0 & \dots & 0 \\ 0 & \sigma_{\boldsymbol{\beta}2} & & 0 \\ & & \ddots & \\ 0 & \dots & 0 & \sigma_{\boldsymbol{\beta}K} \end{bmatrix},
$$

then $\left(\mathbf{x}_j\circ\boldsymbol{\sigma}_{\boldsymbol{\beta}0}\right)\boldsymbol{\beta}=\mathbf{x}_j\boldsymbol{\Sigma}_{\boldsymbol{\beta}0}\boldsymbol{\beta}$. For the sake of simplicity, we choose two markets in which prices of firms in each market are different from each other. Further, let $p_l<p_{l+1}$ for $l=1,\dots,J$, and similarly for the second market. Also, let $\mathbf{v}_{i,0}\equiv\boldsymbol{\xi}_{i,0}/\sigma_{\alpha 0}$.

Note that the market share function can be expressed as the aggregate of individuals with different α, $\boldsymbol{\beta}$ and the i.i.d. extreme value distributed preference shock $\boldsymbol{\epsilon}$ as below. That is, consider an individual with specific α, $\boldsymbol{\beta}$ and $\boldsymbol{\epsilon}$ values who chooses product $i=0,\dots,J$, (where 0 is no purchase) if

$$
\mathbf{x}_i\circ\boldsymbol{\sigma}_{\boldsymbol{\beta}0}\boldsymbol{\beta}+p_i\sigma_{\alpha 0}\left(\alpha+\eta_{\alpha 0}\right)+\sigma_{\alpha 0}v_{i,0}+\epsilon_i\geq\mathbf{x}_l\circ\boldsymbol{\sigma}_{\boldsymbol{\beta}0}\boldsymbol{\beta}+p_l\sigma_{\alpha 0}\left(\alpha+\eta_{\alpha 0}\right)+\sigma_{\alpha 0}v_{l,0}+\epsilon_l,\ l\neq i
$$

where $\mathbf{x}_0=0$, $p_0=0$ $v_{00}=0$ for normalization. From now on, we set the price vector to be $\gamma\mathbf{p}$, where $\gamma>0$ is a scalar. Since we only consider cases where γ is large for the identification of $\eta_{\alpha 0}$, we set $\gamma > \underline{\gamma}$ for some $\underline{\gamma} > 0$. Furthermore, we define $\mathbf{v}_0(\gamma, \mathbf{X}, \mathbf{p}, \mathbf{s}, \boldsymbol{\theta}_{c0}) \equiv \boldsymbol{\xi}(\mathbf{X}, \gamma\mathbf{p}, \mathbf{s}, \boldsymbol{\theta}_{c0}) / (\gamma\sigma_{\alpha 0})$. Then,

$$\begin{aligned} s_i &= \int_\alpha \int_{\boldsymbol{\beta}} \int_{\boldsymbol{\epsilon}} \prod_{l \neq i} I\left(\mathbf{x}_i \boldsymbol{\Sigma}_{\beta 0} \boldsymbol{\beta} + \gamma p_i \sigma_{\alpha 0}(\alpha + \eta_{\alpha 0}) + \gamma\sigma_{\alpha 0} v_{i,0}(\gamma, .) + \epsilon_i \right. \\ &\geq \left. \mathbf{x}_l \boldsymbol{\Sigma}_{\beta 0} \boldsymbol{\beta} + \gamma p_l \sigma_{\alpha 0}(\alpha + \eta_{\alpha 0}) + \gamma\sigma_{\alpha 0} v_{l,0}(\gamma, .) + \epsilon_l\right) f(\boldsymbol{\epsilon})\, d\boldsymbol{\epsilon}\, \phi(\boldsymbol{\beta})\, d\boldsymbol{\beta}\, \phi(\alpha)\, d\alpha \end{aligned} \tag{46}$$

Then, dividing the integrand by γ, and holding s_i fixed, as $\gamma \to \infty$[38], we obtain

$$\begin{aligned} s_i &= lim_{\gamma \to \infty} \int_\alpha \int_{\boldsymbol{\beta}} \int_{\boldsymbol{\epsilon}} \prod_{l \neq i} I\left(\frac{\mathbf{x}_i}{\gamma} \boldsymbol{\Sigma}_{\beta 0} \boldsymbol{\beta} + p_i \sigma_{\alpha 0}(\alpha + \eta_{\alpha 0}) + \sigma_{\alpha 0} v_{i,0}(\gamma, .) + \frac{\epsilon_i}{\gamma} \right. \\ &\geq \left. \frac{\mathbf{x}_l}{\gamma} \boldsymbol{\Sigma}_{\beta 0} \boldsymbol{\beta} + p_l \sigma_{\alpha 0}(\alpha + \eta_{\alpha 0}) + \sigma_{\alpha 0} v_{l,0}(\gamma, .) + \frac{\epsilon_l}{\gamma}\right) f(\boldsymbol{\epsilon})\, d\boldsymbol{\epsilon}\, \phi(\boldsymbol{\beta})\, d\boldsymbol{\beta}\, \phi(\alpha)\, d\alpha \end{aligned}$$

where $f(\boldsymbol{\epsilon}) \equiv \prod_{l=1}^{J} g(\epsilon_l)$ is the joint distribution of $\boldsymbol{\epsilon}$, with $g()$ being the extreme value density function. As $\gamma \to \infty$, $\mathbf{x}_l/\gamma \to \mathbf{0}$, $l = 1, \ldots, J$, $\boldsymbol{\epsilon}/\gamma \to \mathbf{0}$ for any $\boldsymbol{\epsilon} \in R^J$. Therefore, for a fixed vector $\widetilde{\boldsymbol{v}}_0$, the following holds:

$$\begin{aligned} & I\left(\frac{\mathbf{x}_i}{\gamma} \boldsymbol{\Sigma}_{\beta 0} \boldsymbol{\beta} + p_i \sigma_{\alpha 0}(\alpha + \eta_{\alpha 0}) + \sigma_{\alpha 0} \widetilde{v}_{i,0} + \frac{\epsilon_i}{\gamma} \geq \frac{\mathbf{x}_l}{\gamma} \boldsymbol{\Sigma}_{\beta 0} \boldsymbol{\beta} + p_l \sigma_{\alpha 0}(\alpha + \eta_{\alpha 0}) + \sigma_{\alpha 0} \widetilde{v}_{l,0} + \frac{\epsilon_l}{\gamma}\right) \\ \to \;& I\left(p_i(\alpha + \eta_{\alpha 0}) + \widetilde{v}_{i,0} \geq p_l(\alpha + \eta_{\alpha 0}) + \widetilde{v}_{l,0}\right) \;\; as\ \gamma \to \infty. \end{aligned}$$

Furthermore, $\left|\prod_{l \neq i} I()\right| \leq 1$ and $\int_\alpha \int_{\boldsymbol{\beta}} \int_{\boldsymbol{\epsilon}} \left[\prod_{l \neq i} 1\right] f(\boldsymbol{\epsilon})\, d\boldsymbol{\epsilon}\, \phi(\boldsymbol{\beta})\, d\boldsymbol{\beta}\, \phi(\alpha)\, d\alpha = 1$. Therefore, from the Dominated Convergence Theorem,

$$\begin{aligned} s_i &= \int_\alpha \int_{\boldsymbol{\beta}} \int_{\boldsymbol{\epsilon}} \prod_{l \neq i} I\left(p_i(\alpha + \eta_{\alpha 0}) + \widetilde{v}_{i,0} \geq p_l(\alpha + \eta_{\alpha 0}) + \widetilde{v}_{l,0}\right) f(\boldsymbol{\epsilon})\, d\boldsymbol{\epsilon}\, \phi(\boldsymbol{\beta})\, d\boldsymbol{\beta}\, \phi(\alpha)\, d\alpha \\ &= \int_\alpha \prod_{l \neq i} I\left(p_i(\alpha + \eta_{\alpha 0}) + \widetilde{v}_{i,0} \geq p_l(\alpha + \eta_{\alpha 0}) + \widetilde{v}_{l,0}\right) \phi(\alpha)\, d\alpha \\ &\equiv \widetilde{s}_i(\mathbf{p}, \widetilde{\mathbf{v}}_0, \boldsymbol{\theta}_{c0}), \end{aligned}$$

And,

$$s_0 = \int_\alpha \prod_{l=1}^{J} I\left(p_l(\alpha + \eta_{\alpha 0}) + \widetilde{v}_{l,0} \leq 0\right) \phi(\alpha)\, d\alpha.$$

where, for $i = 0, \ldots, J$, $\widetilde{\mathbf{v}}_0$ satisfies $\widetilde{s}(\mathbf{p}, \widetilde{\mathbf{v}}_0, \boldsymbol{\theta}_0) = \mathbf{s}$.

[38] Note that prices $\gamma\mathbf{p}$ and market shares $\mathbf{s}$ are observed, and the unobserved product characteristics $\boldsymbol{\nu}$ is implied through the functional relationship, given $\mathbf{X}$ and the true parameters. Given our assumptions, we can find demand shocks that generate $\gamma\mathbf{p}$ and $\mathbf{s}$, for arbitrarily large γ.

Next, we prove that such $\widetilde{\mathbf{v}}_0$ exists and is unique, as long as we normalize $\widetilde{v}_{0,0}$ to be zero. In particular, we prove that given $\widetilde{v}_{0,0}=0$, $p_0=0$, $\widetilde{v}_{i,0}$, $i>0$ can be derived recursively as follows.

$$\widetilde{v}_{i+1,0}=\widetilde{v}_{i,0}+(p_i-p_{i+1})\left[\eta_{\alpha 0}+\Phi^{-1}\left(\sum_{l\leq i}s_l\right)\right].$$

First, let $\widetilde{\mathbf{v}}_0$ to be an arbitrary $J\times 1$ vector. Denote

$$a_{i,l}=\frac{\widetilde{v}_{i,0}-\widetilde{v}_{l,0}}{p_i-p_l},\ l\neq i \tag{47}$$

Since we assumed $p_i<p_{i+1}$ for $i=0,\ldots,J-1$, i is chosen if

$$\begin{aligned}\alpha+\eta_{\alpha 0}+a_{i,l}\geq 0 \quad & if \quad i>l\\ \alpha+\eta_{\alpha 0}+a_{i,l}\leq 0 \quad & if \quad i<l.\end{aligned}$$

Then, for $i\geq 1$, let $\underline{a}_i=min_{l<i}\left\{a_{i,l}\right\}$ $\overline{a}_i=max_{l>i}\left\{a_{i,l}\right\}$. Then,

$$\begin{aligned}I\left(choose\ i\right) &= \prod_{l=0}^{i-1}I\left(\alpha\geq-\eta_{\alpha 0}-a_{i,l}\right)\prod_{l=i+1}^{J}I\left(\alpha\leq-\eta_{\alpha 0}-a_{i,l}\right)\\ &= I\left(\alpha\geq max_{l<i}\left\{-\eta_{\alpha 0}-a_{i,l}\right\}\right)I\left(\alpha\leq min_{l>i}\left\{-\eta_{\alpha 0}-a_{i,l}\right\}\right).\end{aligned}$$

Therefore,

$$\widetilde{s}_i\left(\mathbf{p},\widetilde{\mathbf{v}}_0,\boldsymbol{\theta}_0\right)=\int_\alpha I\left(-\eta_{\alpha 0}-\underline{a}_i\leq\alpha\leq-\eta_{\alpha 0}-\overline{a}_i\right)\phi\left(\alpha\right)d\alpha.$$

Furthermore,

$$I\left(choose\ 0\right)=I\left(\alpha\leq min_{l>0}\left\{-\eta_{\alpha 0}-a_{0l}\right\}\right).$$

Therefore, as before,

$$\widetilde{s}_0\left(\mathbf{p},\widetilde{\mathbf{v}}_0,\boldsymbol{\theta}_0\right)=\int_\alpha I\left(-\eta_{\alpha 0}-\underline{a}_0\leq\alpha\leq-\eta_{\alpha 0}-\overline{a}_0\right)\phi\left(\alpha\right)d\alpha.$$

where $\underline{a}_0=\infty$, $\overline{a}_0=max_{l>0}\left\{a_{0l}\right\}$. Similarly,

$$\widetilde{s}_J=\int_\alpha I\left(-\eta_{\alpha 0}-\underline{a}_J\leq\alpha\leq-\eta_{\alpha 0}-\overline{a}_J\right)\phi\left(\alpha\right)d\alpha.$$

$\underline{a}_J=min_{l<J}\left\{a_{Jl}\right\}$, $\overline{a}_J=-\infty$.

Given $s_i>0$ for any $i=0,\ldots,J$ (2 of Condition 1), $-\underline{a}_i<-\overline{a}_i$ holds. Furthermore, note that

$$\underline{a}_{i+1} = min_{l<i+1}\left\{a_{i+1,l}\right\} \leq a_{i+1,i} = a_{i,i+1} \leq max_{l>i}\left\{a_{i,l}\right\} = \overline{a}_i. \tag{48}$$

Therefore, $-\overline{a}_i \leq -\underline{a}_{i+1}$ $i = 0, \ldots, J-1$. Now, suppose $-\overline{a}_i < -\underline{a}_{i+1}$. Then, $\ldots - \overline{a}_{i-1} \leq -\underline{a}_i \leq -\overline{a}_i < -\underline{a}_{i+1} \leq -\overline{a}_{i+1} \leq -\underline{a}_{i+2} \ldots$. It implies that an individual with $\alpha \in \left(-\eta_{\alpha 0} - \overline{a}_i, -\eta_{\alpha 0} - \underline{a}_{i+1}\right)$ will not choose any of the available product choices $i \in \{0, 1, \ldots, J\}$ including the no-purchase option 0, which is a contradiction. Therefore, $-\overline{a}_i = -\underline{a}_{i+1}$ holds. Furthermore, Equation (48) and $\overline{a}_i = \underline{a}_{i+1}$ imply $\overline{a}_i = \underline{a}_{i+1} = a_{i,i+1}$. Therefore,

$$s_0 = \widetilde{s}_0\left(\mathbf{p}, \widetilde{\mathbf{v}}_0, \boldsymbol{\theta}_0\right) = \int_\alpha I\left(\alpha \leq -\eta_{\alpha 0} - \frac{\widetilde{v}_{10}}{p_1}\right)\phi\left(\alpha\right)d\alpha = \Phi\left(-\eta_{\alpha 0} - \frac{\widetilde{v}_{10}}{p_1}\right) \tag{49}$$

and for $i \geq 1$,

$$\begin{aligned} s_i &= \widetilde{s}_i\left(\mathbf{p}, \widetilde{\mathbf{v}}_0, \boldsymbol{\theta}_0\right) \equiv \int_\alpha I\left(-\eta_{\alpha 0} - \frac{\widetilde{v}_{i-1,0} - \widetilde{v}_{i,0}}{p_{i-1} - p_i} \leq \alpha \leq -\eta_{\alpha 0} - \frac{\widetilde{v}_{i,0} - \widetilde{v}_{i+1,0}}{p_i - p_{i+1}}\right)\phi\left(\alpha\right)d\alpha \\ &= \Phi\left(-\eta_{\alpha 0} - \frac{\widetilde{v}_{i,0} - \widetilde{v}_{i+1,0}}{p_i - p_{i+1}}\right) - \Phi\left(-\eta_{\alpha 0} - \frac{\widetilde{v}_{i-1,0} - \widetilde{v}_{i,0}}{p_{i-1} - p_i}\right) \end{aligned} \tag{50}$$

and

$$s_J = 1 - \Phi\left(-\eta_{\alpha 0} - \frac{\widetilde{v}_{J-10} - \widetilde{v}_{J,0}}{p_{J-1} - p_J}\right) \tag{51}$$

Therefore, from Equations (49) to (51), we can derive

$$\Phi\left(-\eta_{\alpha 0} - \frac{\widetilde{v}_{i,0} - \widetilde{v}_{i+1,0}}{p_i - p_{i+1}}\right) = \sum_{l \leq i} s_l.$$

Thus,

$$\widetilde{v}_{i+1,0} = \widetilde{v}_{i,0} + \left(p_i - p_{i+1}\right)\left[\eta_{\alpha 0} + \Phi^{-1}\left(\sum_{l \leq i} s_l\right)\right], \tag{52}$$

As $\widetilde{v}_{0,0}$ is normalized to be 0, $\widetilde{v}_{i+1,0}$, $i \geq 0$ can be recursively derived as follows:

$$\widetilde{v}_{i+1,0} = \sum_{k=0}^{i}\left(p_k - p_{k+1}\right)\left[\eta_{\alpha 0} + \Phi^{-1}\left(\sum_{l \leq k} s_l\right)\right]. \tag{53}$$

Therefore, we have proven that $\widetilde{\mathbf{v}}_0$ satisfying $\widetilde{s}\left(\mathbf{p}, \widetilde{\mathbf{v}}_0, \boldsymbol{\theta}_0\right) = \mathbf{s}$ exists and is unique.

Next, we show that $lim_{\gamma\to\infty}\mathbf{v}_0\left(\gamma,\mathbf{X},\mathbf{p},\mathbf{s},\boldsymbol{\theta}_{c0}\right)=\widetilde{\mathbf{v}}_0$. Now,

$$\begin{aligned}
&\int_\alpha\int_{\boldsymbol{\beta}}\int_{\boldsymbol{\epsilon}}\prod_{l\neq i}I\left[\left(p_i-p_l\right)\sigma_{\alpha0}\left(\alpha+\eta_{\alpha0}\right)\geq-\frac{\mathbf{x}_i-\mathbf{x}_l}{\gamma}\boldsymbol{\Sigma}_{\boldsymbol{\beta}0}\boldsymbol{\beta}-\frac{\epsilon_i-\epsilon_l}{\gamma}\right.\\
&\left.-\,\sigma_{\alpha0}\left(v_{i,0}\left(\gamma,\mathbf{X},\mathbf{p},\mathbf{s},\boldsymbol{\theta}_{c0}\right)-v_{l,0}\left(\gamma,\mathbf{X},\mathbf{p},\mathbf{s},\boldsymbol{\theta}_{c0}\right)\right)\right]f\left(\boldsymbol{\epsilon}\right)d\boldsymbol{\epsilon}\phi\left(\boldsymbol{\beta}\right)d\boldsymbol{\beta}\phi\left(\alpha\right)d\alpha\\
=\;&\int_\alpha\int_{\boldsymbol{\beta}\in\left[-\sqrt{\gamma},\sqrt{\gamma}\right]^K}\int_{\boldsymbol{\epsilon}\in\left[-\sqrt{\gamma},\sqrt{\gamma}\right]^J}\prod_{l\neq i}I\left[\left(p_i-p_l\right)\sigma_{\alpha0}\left(\alpha+\eta_{\alpha0}\right)\geq-\frac{\mathbf{x}_i-\mathbf{x}_l}{\gamma}\boldsymbol{\Sigma}_{\boldsymbol{\beta}0}\boldsymbol{\beta}-\frac{\epsilon_i-\epsilon_l}{\gamma}\right.\\
&\left.-\sigma_{\alpha0}\left(v_{i,0}\left(\gamma,\mathbf{X},\mathbf{p},\mathbf{s},\boldsymbol{\theta}_{c0}\right)-v_{l,0}\left(\gamma,\mathbf{X},\mathbf{p},\mathbf{s},\boldsymbol{\theta}_{c0}\right)\right)\right]f\left(\boldsymbol{\epsilon}\right)d\boldsymbol{\epsilon}\phi\left(\boldsymbol{\beta}\right)d\boldsymbol{\beta}\phi\left(\alpha\right)d\alpha\\
&+\int_\alpha\int_{\boldsymbol{\beta}\notin\left[-\sqrt{\gamma},\sqrt{\gamma}\right]^K}\int_{\boldsymbol{\epsilon}}\prod_{l\neq i}I\left[\left(p_i-p_l\right)\sigma_{\alpha0}\left(\alpha+\eta_{\alpha0}\right)\geq-\frac{\mathbf{x}_i-\mathbf{x}_l}{\gamma}\boldsymbol{\Sigma}_{\boldsymbol{\beta}0}\boldsymbol{\beta}-\frac{\epsilon_i-\epsilon_l}{\gamma}\right.\\
&\left.-\sigma_{\alpha0}\left(v_{i,0}\left(\gamma,\mathbf{X},\mathbf{p},\mathbf{s},\boldsymbol{\theta}_{c0}\right)-v_{l,0}\left(\gamma,\mathbf{X},\mathbf{p},\mathbf{s},\boldsymbol{\theta}_{c0}\right)\right)\right]f\left(\boldsymbol{\epsilon}\right)d\boldsymbol{\epsilon}\phi\left(\boldsymbol{\beta}\right)d\boldsymbol{\beta}\phi\left(\alpha\right)d\alpha\\
&+\int_\alpha\int_{\boldsymbol{\beta}}\int_{\boldsymbol{\epsilon}\notin\left[-\sqrt{\gamma},\sqrt{\gamma}\right]^J}\prod_{l\neq i}I\left[\left(p_i-p_l\right)\sigma_{\alpha0}\left(\alpha+\eta_{\alpha0}\right)\geq-\frac{\mathbf{x}_i-\mathbf{x}_l}{\gamma}\boldsymbol{\Sigma}_{\boldsymbol{\beta}0}\boldsymbol{\beta}-\frac{\epsilon_i-\epsilon_l}{\gamma}\right.\\
&\left.-\sigma_{\alpha0}\left(v_{i,0}\left(\gamma,\mathbf{X},\mathbf{p},\mathbf{s},\boldsymbol{\theta}_{c0}\right)-v_{l,0}\left(\gamma,\mathbf{X},\mathbf{p},\mathbf{s},\boldsymbol{\theta}_{c0}\right)\right)\right]f\left(\boldsymbol{\epsilon}\right)d\boldsymbol{\epsilon}\phi\left(\boldsymbol{\beta}\right)d\boldsymbol{\beta}\phi\left(\alpha\right)d\alpha. \qquad (54)
\end{aligned}$$

Then, for $\boldsymbol{\beta}\in\left[-\sqrt{\gamma},\sqrt{\gamma}\right]^K$ and $\boldsymbol{\epsilon}\in\left[-\sqrt{\gamma},\sqrt{\gamma}\right]^J$, and for $A\equiv\max_{k,i,l}\left|\left[-\left[\mathbf{x}_i-\mathbf{x}_l\right]\boldsymbol{\Sigma}_{\boldsymbol{\beta}0}\right]_k\right|K+2$

$$\begin{aligned}
\left|-\frac{\mathbf{x}_i-\mathbf{x}_l}{\gamma}\boldsymbol{\Sigma}_{\boldsymbol{\beta}0}\boldsymbol{\beta}-\frac{\epsilon_i-\epsilon_l}{\gamma}\right| &\leq \frac{1}{\gamma}\max_{k,i,l}\left|\left[-\left[\mathbf{x}_i-\mathbf{x}_l\right]\boldsymbol{\Sigma}_{\boldsymbol{\beta}0}\right]_k\right|K\sup_{\boldsymbol{\beta}\in\left[-\sqrt{\gamma},\sqrt{\gamma}\right]^K}\left|\beta_k\right|+\frac{1}{\gamma}\sup_{\boldsymbol{\epsilon}\in\left[-\sqrt{\gamma},\sqrt{\gamma}\right]^J}\left|\epsilon_i-\epsilon_l\right|\\
&= \frac{1}{\sqrt{\gamma}}\left[\max_{k,i,l}\left|\left[-\left[\mathbf{x}_i-\mathbf{x}_l\right]\boldsymbol{\Sigma}_{\boldsymbol{\beta}0}\right]_k\right|K+2\right]=\frac{A}{\sqrt{\gamma}}.
\end{aligned}$$

Furthermore, the 2nd and 3rd terms of Equation (54) can be made arbitrarily small by choosing γ to be arbitrarily large. Therefore, for any $\delta>0$, we can choose γ to be sufficiently large such that $\delta>A/\left(\sigma_{\alpha0}\sqrt{\gamma}\right)$ and

$$\begin{aligned}
&\int_\alpha\int_{\boldsymbol{\beta}\notin\left[-\sqrt{\gamma},\sqrt{\gamma}\right]^K}\int_{\boldsymbol{\epsilon}}\prod_{l\neq i}I\left[\left(p_i-p_l\right)\sigma_{\alpha0}\left(\alpha+\eta_{\alpha0}\right)\geq-\frac{\mathbf{x}_i-\mathbf{x}_l}{\gamma}\boldsymbol{\Sigma}_{\boldsymbol{\beta}0}\boldsymbol{\beta}-\frac{\epsilon_i-\epsilon_l}{\gamma}\right.\\
&\left.-\sigma_{\alpha0}\left(v_{i,0}\left(\gamma,\mathbf{X},\mathbf{p},\mathbf{s},\boldsymbol{\theta}_{c0}\right)-v_{l,0}\left(\gamma,\mathbf{X},\mathbf{p},\mathbf{s},\boldsymbol{\theta}_{c0}\right)\right)\right]f\left(\boldsymbol{\epsilon}\right)d\boldsymbol{\epsilon}\phi\left(\boldsymbol{\beta}\right)d\boldsymbol{\beta}\phi\left(\alpha\right)d\alpha\\
\leq\;&\int_\alpha\int_{\boldsymbol{\beta}\notin\left[-\sqrt{\gamma},\sqrt{\gamma}\right]^K}\int_{\boldsymbol{\epsilon}}f\left(\boldsymbol{\epsilon}\right)d\boldsymbol{\epsilon}\phi\left(\boldsymbol{\beta}\right)d\boldsymbol{\beta}\phi\left(\alpha\right)d\alpha\leq\frac{\delta}{2},\\
&\int_\alpha\int_{\boldsymbol{\beta}}\int_{\boldsymbol{\epsilon}\notin\left[-\sqrt{\gamma},\sqrt{\gamma}\right]^K}\prod_{l\neq i}I\left[\left(p_i-p_l\right)\sigma_{\alpha0}\left(\alpha+\eta_{\alpha0}\right)\geq-\frac{\mathbf{x}_i-\mathbf{x}_l}{\gamma}\boldsymbol{\Sigma}_{\boldsymbol{\beta}0}\boldsymbol{\beta}-\frac{\epsilon_i-\epsilon_l}{\gamma}\right.\\
&\left.-\sigma_{\alpha0}\left(v_{i,0}\left(\gamma,\mathbf{X},\mathbf{p},\mathbf{s},\boldsymbol{\theta}_{c0}\right)-v_{l,0}\left(\gamma,\mathbf{X},\mathbf{p},\mathbf{s},\boldsymbol{\theta}_{c0}\right)\right)\right]f\left(\boldsymbol{\epsilon}\right)d\boldsymbol{\epsilon}\phi\left(\boldsymbol{\beta}\right)d\boldsymbol{\beta}\phi\left(\alpha\right)d\alpha\\
\leq\;&\int_\alpha\int_{\boldsymbol{\beta}}\int_{\boldsymbol{\epsilon}\notin\left[-\sqrt{\gamma},\sqrt{\gamma}\right]^J}f\left(\boldsymbol{\epsilon}\right)d\boldsymbol{\epsilon}\phi\left(\boldsymbol{\beta}\right)d\boldsymbol{\beta}\phi\left(\alpha\right)d\alpha\leq\frac{\delta}{2}. \qquad (55)
\end{aligned}$$

Because $\delta > A/\left(\sigma_{\alpha 0}\sqrt{\gamma}\right)$, for $\boldsymbol{\beta} \in \left[-\sqrt{\gamma}, \sqrt{\gamma}\right]^K$ and $\boldsymbol{\epsilon} \in \left[-\sqrt{\gamma}, \sqrt{\gamma}\right]^J$,

$$(p_i - p_l)\,\sigma_{\alpha 0}\,(\alpha + \eta_{\alpha 0}) \geq -\frac{\mathbf{x}_i - \mathbf{x}_l}{\gamma}\boldsymbol{\Sigma}_{\boldsymbol{\beta}0}\boldsymbol{\beta} - \frac{\epsilon_i - \epsilon_l}{\gamma} - \sigma_{\alpha 0}\left(v_{i,0}\left(\gamma, \mathbf{X}, \mathbf{p}, \mathbf{s}, \boldsymbol{\theta}_{c0}\right) - v_{l,0}\left(\gamma, \mathbf{X}, \mathbf{p}, \mathbf{s}, \boldsymbol{\theta}_{c0}\right)\right)$$

implies

$$(p_i - p_l)\,\sigma_{\alpha 0}\,(\alpha + \eta_{\alpha 0}) \geq -\sigma_{\alpha 0}\delta - \sigma_{\alpha 0}\left(v_{i,0}\left(\gamma, \mathbf{X}, \mathbf{p}, \mathbf{s}, \boldsymbol{\theta}_{c0}\right) - v_{l,0}\left(\gamma, \mathbf{X}, \mathbf{p}, \mathbf{s}, \boldsymbol{\theta}_{c0}\right)\right).$$

Therefore,

$$\begin{aligned}
& \int_\alpha \int_{\boldsymbol{\beta}} \int_{\boldsymbol{\epsilon}} \prod_{l \neq i} I\left[(p_i - p_l)\,\sigma_{\alpha 0}\,(\alpha + \eta_{\alpha 0}) \geq -\sigma_{\alpha 0}\delta\right. \\
& \left.-\sigma_{\alpha 0}\left(v_{i,0}\left(\gamma, \mathbf{X}, \mathbf{p}, \mathbf{s}, \boldsymbol{\theta}_{c0}\right) - v_{l,0}\left(\gamma, \mathbf{X}, \mathbf{p}, \mathbf{s}, \boldsymbol{\theta}_{c0}\right)\right)\right] f(\boldsymbol{\epsilon})\, d\boldsymbol{\epsilon}\, \phi(\boldsymbol{\beta})\, d\boldsymbol{\beta}\, \phi(\alpha)\, d\alpha \\
\geq\; & \int_\alpha \int_{\boldsymbol{\beta} \in \left[-\sqrt{\gamma}, \sqrt{\gamma}\right]^K} \int_{\boldsymbol{\epsilon} \in \left[-\sqrt{\gamma}, \sqrt{\gamma}\right]^K} \prod_{l \neq i} I\left[(p_i - p_l)\,\sigma_{\alpha 0}\,(\alpha + \eta_{\alpha 0}) \geq -\sigma_{\alpha 0}\delta\right. \\
& \left.-\sigma_{\alpha 0}\left(v_{i,0}\left(\gamma, \mathbf{X}, \mathbf{p}, \mathbf{s}, \boldsymbol{\theta}_{c0}\right) - v_{l,0}\left(\gamma, \mathbf{X}, \mathbf{p}, \mathbf{s}, \boldsymbol{\theta}_{c0}\right)\right)\right] f(\boldsymbol{\epsilon})\, d\boldsymbol{\epsilon}\, \phi(\boldsymbol{\beta})\, d\boldsymbol{\beta}\, \phi(\alpha)\, d\alpha \\
\geq\; & \int_\alpha \int_{\boldsymbol{\beta} \in \left[-\sqrt{\gamma}, \sqrt{\gamma}\right]^K} \int_{\boldsymbol{\epsilon} \in \left[-\sqrt{\gamma}, \sqrt{\gamma}\right]^K} \prod_{l \neq i} I\left[(p_i - p_l)\,\sigma_{\alpha 0}\,(\alpha + \eta_{\alpha 0}) \geq -\frac{\mathbf{x}_i - \mathbf{x}_l}{\gamma}\boldsymbol{\Sigma}_{\boldsymbol{\beta}0}\boldsymbol{\beta} - \frac{\epsilon_i - \epsilon_l}{\gamma}\right. \\
& \left.-\sigma_{\alpha 0}\left(v_{i,0}\left(\gamma, \mathbf{X}, \mathbf{p}, \mathbf{s}, \boldsymbol{\theta}_{c0}\right) - v_{l,0}\left(\gamma, \mathbf{X}, \mathbf{p}, \mathbf{s}, \boldsymbol{\theta}_{c0}\right)\right)\right] f(\boldsymbol{\epsilon})\, d\boldsymbol{\epsilon}\, \phi(\boldsymbol{\beta})\, d\boldsymbol{\beta}\, \phi(\alpha)\, d\alpha. \qquad (56)
\end{aligned}$$

Then, using Equations (55) and (56), we obtain

$$\begin{aligned}
& \int_\alpha \prod_{l \neq i} I\left[(p_i - p_l)\,\sigma_{\alpha 0}\,(\alpha + \eta_{\alpha 0}) \geq -\sigma_{\alpha 0}\delta\right. \\
& \left.-\sigma_{\alpha 0}\left(v_{i,0}\left(\gamma, \mathbf{X}, \mathbf{p}, \mathbf{s}, \boldsymbol{\theta}_{c0}\right) - v_{l,0}\left(\gamma, \mathbf{X}, \mathbf{p}, \mathbf{s}, \boldsymbol{\theta}_{c0}\right)\right)\right] \phi(\alpha)\, d\alpha + \delta \\
\geq\; & \int_\alpha \int_{\boldsymbol{\beta}} \int_{\boldsymbol{\epsilon}} \prod_{l \neq i} I\left[(p_i - p_l)\,\sigma_{\alpha 0}\,(\alpha + \eta_{\alpha 0}) \geq -\frac{\mathbf{x}_i - \mathbf{x}_l}{\gamma}\boldsymbol{\Sigma}_{\boldsymbol{\beta}0}\boldsymbol{\beta} - \frac{\epsilon_i - \epsilon_l}{\gamma}\right. \\
& \left.-\sigma_{\alpha 0}\left(v_{i,0}\left(\gamma, \mathbf{X}, \mathbf{p}, \mathbf{s}, \boldsymbol{\theta}_{c0}\right) - v_{l,0}\left(\gamma, \mathbf{X}, \mathbf{p}, \mathbf{s}, \boldsymbol{\theta}_{c0}\right)\right)\right] f(\boldsymbol{\epsilon})\, d\boldsymbol{\epsilon}\, \phi(\boldsymbol{\beta})\, d\boldsymbol{\beta}\, \phi(\alpha)\, d\alpha = s_i.
\end{aligned}$$

Similarly, for $\boldsymbol{\beta} \in \left[-\sqrt{\gamma}, \sqrt{\gamma}\right]^K$ and $\boldsymbol{\epsilon} \in \left[-\sqrt{\gamma}, \sqrt{\gamma}\right]^J$

$$(p_i - p_l)\,(\alpha + \eta_{\alpha 0}) \geq \delta - \left(v_{i,0}\left(\gamma, \mathbf{X}, \mathbf{p}, \mathbf{s}, \boldsymbol{\theta}_{c0}\right) - v_{l,0}\left(\gamma, \mathbf{X}, \mathbf{p}, \mathbf{s}, \boldsymbol{\theta}_{c0}\right)\right)$$

implies

$$(p_i - p_l)\,(\alpha + \eta_{\alpha 0}) \geq -\frac{\mathbf{x}_i - \mathbf{x}_l}{\sigma_{\alpha 0}\gamma}\boldsymbol{\Sigma}_{\boldsymbol{\beta}0}\boldsymbol{\beta} - \frac{\epsilon_i - \epsilon_l}{\sigma_{\alpha 0}\gamma} - \left(v_{i,0}\left(\gamma, \mathbf{X}, \mathbf{p}, \mathbf{s}, \boldsymbol{\theta}_{c0}\right) - v_{l,0}\left(\gamma, \mathbf{X}, \mathbf{p}, \mathbf{s}, \boldsymbol{\theta}_{c0}\right)\right).$$

Therefore, we can obtain

$$
\begin{aligned}
& \int_{\alpha}\prod_{l\neq i} I\left[\left(p_i-p_l\right)\left(\alpha+\eta_{\alpha 0}\right)\geq \delta\right. \\
& \left.-\left(v_{i,0}\left(\gamma,\mathbf{X},\mathbf{p},\mathbf{s},\boldsymbol{\theta}_{c0}\right)-v_{l,0}\left(\gamma,\mathbf{X},\mathbf{p},\mathbf{s},\boldsymbol{\theta}_{c0}\right)\right)\right]\phi\left(\alpha\right)d\alpha-\delta \\
\leq\ & \int_{\alpha}\int_{\boldsymbol{\beta}}\int_{\boldsymbol{\epsilon}}\prod_{l\neq i} I\left[\left(p_i-p_l\right)\left(\alpha+\eta_{\alpha 0}\right)\geq -\frac{\mathbf{x}_i-\mathbf{x}_l}{\sigma_{\alpha 0}\gamma}\boldsymbol{\Sigma}_{\boldsymbol{\beta}0}\boldsymbol{\beta}-\frac{\epsilon_i-\epsilon_l}{\sigma_{\alpha 0}\gamma}\right. \\
& \left.-\left(v_{i,0}\left(\gamma,\mathbf{X},\mathbf{p},\mathbf{s},\boldsymbol{\theta}_{c0}\right)-v_{l,0}\left(\gamma,\mathbf{X},\mathbf{p},\mathbf{s},\boldsymbol{\theta}_{c0}\right)\right)\right]f\left(\boldsymbol{\epsilon}\right)d\boldsymbol{\epsilon}\phi\left(\boldsymbol{\beta}\right)d\boldsymbol{\beta}\phi\left(\alpha\right)d\alpha=s_i.
\end{aligned}
$$

Then, if we let

$$
\begin{aligned}
\overline{s}_i\ \equiv\ & \int_{\alpha}\prod_{l\neq i} I\left[\left(p_i-p_l\right)\left(\alpha+\eta_{\alpha 0}\right)\geq -\delta\right. \\
& \left.-\left(v_{i,0}\left(\gamma,\mathbf{X},\mathbf{p},\mathbf{s},\boldsymbol{\theta}_{c0}\right)-v_{l,0}\left(\gamma,\mathbf{X},\mathbf{p},\mathbf{s},\boldsymbol{\theta}_{c0}\right)\right)\right]\phi\left(\alpha\right)d\alpha \qquad (57)\\
\underline{s}_i\ \equiv\ & \int_{\alpha}\prod_{l\neq i} I\left[\left(p_i-p_l\right)\left(\alpha+\eta_{\alpha 0}\right)\geq \delta\right. \\
& \left.-\left(v_{i,0}\left(\gamma,\mathbf{X},\mathbf{p},\mathbf{s},\boldsymbol{\theta}_{c0}\right)-v_{l,0}\left(\gamma,\mathbf{X},\mathbf{p},\mathbf{s},\boldsymbol{\theta}_{c0}\right)\right)\right]\phi\left(\alpha\right)d\alpha, \qquad (58)
\end{aligned}
$$

$$\underline{s}_i-\delta\leq s_i\leq \overline{s}_i+\delta. \qquad (59)$$

Also, from continuity of $\overline{s}_i$ and $\underline{s}_i$ in δ, we obtain

$$\left|\overline{s}_i-\underline{s}_i\right|\leq \varpi_i \qquad (60)$$

where ϖ_i can be made arbitrarily small by making δ sufficiently small, which can be done by making γ sufficiently large. Then, from the Triangle Inequality and Equations (59) and (60),

$$\left|\overline{s}_i-s_i\right|\leq\left|\overline{s}_i+\delta-s_i\right|+\delta\leq\left|\overline{s}_i+\delta-\left(\underline{s}_i-\delta\right)\right|+\delta\leq\left|\overline{s}_i-\underline{s}_i\right|+3\delta\leq\varpi_i+3\delta. \qquad (61)$$

Hence, $\left|\overline{s}_i-s_i\right|$ can be made arbitrarily small by making γ sufficiently large. Similarly, $\left|s_i-\underline{s}_i\right|\leq\left|s_i+\delta-\underline{s}_i\right|+\delta$, and since $\underline{s}_i-\delta\leq s_i\leq\overline{s}_i+\delta$,

$$\overline{s}_i+\delta-\left(\underline{s}_i-\delta\right)\geq s_i+\delta-\underline{s}_i\geq 0$$

. Thus,

$$\left|s_i-\underline{s}_i\right|\leq\left|s_i+\delta-\underline{s}_i\right|+\delta\leq\left|\overline{s}_i+\delta-\left(\underline{s}_i-\delta\right)\right|+\delta\leq\left|\overline{s}_i-\underline{s}_i\right|+3\delta\leq\varpi_i+3\delta. \qquad (62)$$

Furthermore, we can use similar arguments that resulted in Equations (52) and (53) to derive

$$\begin{aligned}(p_i - p_{i+1})\left[\eta_{\alpha 0} + \frac{\delta}{p_i - p_{i+1}} + \Phi^{-1}\left(\sum_{l\leq i}(\overline{s}_l + \delta)\right)\right] &\leq v_{i+10}(\gamma, \mathbf{X}, \mathbf{p}, \mathbf{s}, \boldsymbol{\theta}_{c0}) - v_{i,0}(\gamma, \mathbf{X}, \mathbf{p}, \mathbf{s}, \boldsymbol{\theta}_{c0}) \\ &\leq (p_i - p_{i+1})\left[\eta_{\alpha 0} - \frac{\delta}{p_i - p_{i+1}} + \Phi^{-1}\left(\sum_{l\leq i}(\underline{s}_l - \delta)\right)\right].\end{aligned}$$

where δ is chosen to be a small enough positive number to satisfy

$$\delta < min\left\{\frac{1 - \sum_{l\leq J-1}\overline{s}_l}{J}, min_l\left\{\underline{s}_l\right\}\right\}.$$

[39]Then, from Inequalities (61) and (62), for sufficiently small $\delta > 0$, both $0 < \sum_{l\leq i}(\overline{s}_l + \delta) < 1$ and $0 < \sum_{l\leq i}(\underline{s}_l - \delta) < 1$ hold for any $i = 0, \ldots, J-1$.

Therefore, for any $i = 0, \ldots, J-1$,

$$\begin{aligned}&\sum_{k=0}^{i}(p_k - p_{k+1})\left[\eta_{\alpha 0} + \frac{\delta}{p_i - p_{i+1}} + \Phi^{-1}\left(\sum_{l\leq k}(\overline{s}_l + \delta)\right)\right] \\ \leq\ & v_{i+10}(\gamma, \mathbf{X}, \mathbf{p}, \mathbf{s}, \boldsymbol{\theta}_{c0}) \leq \sum_{k=0}^{i}(p_k - p_{k+1})\left[\eta_{\alpha 0} - \frac{\delta}{p_i - p_{i+1}} + \Phi^{-1}\left(\sum_{l\leq k}(\underline{s}_l - \delta)\right)\right]\end{aligned}$$

and since

$$\begin{aligned}\sum_{k=0}^{i}(p_k - p_{k+1})\left[\eta_{\alpha 0} + \frac{\delta}{p_i - p_{i+1}} + \Phi^{-1}\left(\sum_{l\leq k}(\overline{s}_l + \delta)\right)\right] &\to \sum_{k=0}^{i}(p_k - p_{k+1})\left[\eta_{\alpha 0} + \Phi^{-1}\left(\sum_{l\leq k}s_l\right)\right] \\ \sum_{k=0}^{i}(p_k - p_{k+1})\left[\eta_{\alpha 0} - \frac{\delta}{p_i - p_{i+1}} + \Phi^{-1}\left(\sum_{l\leq k}(\underline{s}_l - \delta)\right)\right] &\to \sum_{k=0}^{i}(p_k - p_{k+1})\left[\eta_{\alpha 0} + \Phi^{-1}\left(\sum_{l\leq k}s_l\right)\right],\end{aligned}$$

as $\gamma \to \infty$, we have proven the claim.

By taking the derivative of the limit of the market share function of good j obtained above (see Equations (49) to (51)) with respect to its price, we obtain

$$\frac{\partial \widetilde{s}_j}{\partial p_j} = \phi\left(-\eta_{\alpha 0} - \frac{\widetilde{v}_{j,0} - \widetilde{v}_{j+1,0}}{p_j - p_{j+1}}\right)\frac{\widetilde{v}_{j,0} - \widetilde{v}_{j+1,0}}{(p_j - p_{j+1})^2} + \phi\left(-\eta_{\alpha 0} - \frac{\widetilde{v}_{j-1,0} - \widetilde{v}_{j,0}}{p_{j-1} - p_j}\right)\frac{\widetilde{v}_{j-1,0} - \widetilde{v}_{j,0}}{(p_j - p_{j-1})^2}.$$

Next, we show that limit of the derivative of the market share function is the same as derived

[39]Note that from Inequality (61), $lim_{\delta\searrow 0}\overline{\mathbf{s}} = \mathbf{s}$, therefore, $lim_{\delta\searrow 0}\left[1 - \sum_{l\leq J-1}\overline{s}_l\right] = 1 - \sum_{l\leq J-1}s_l > 0$ and the definition of $\underline{\mathbf{s}}$ imply $\delta > 0$.

above. Differentiating the market share function given $\mathbf{v}_0$, we obtain

$$\begin{aligned}
\frac{\partial s_j}{\partial p_j} &= \frac{\partial}{\partial p_j}\int_\alpha\int_{\boldsymbol{\beta}}\int_{\boldsymbol{\epsilon}}\prod_{l\neq j} I\left(\mathbf{x}_j\boldsymbol{\Sigma}_{\boldsymbol{\beta}0}\boldsymbol{\beta}+\gamma p_j\sigma_{\alpha0}\left(\alpha+\eta_{\alpha0}\right)+\gamma\sigma_{\alpha0}v_{j,0}\left(\gamma,\mathbf{X},\mathbf{p},\mathbf{s},\boldsymbol{\theta}_{c0}\right)+\epsilon_j\right.\\
&\quad\left.\geq \mathbf{x}_l\boldsymbol{\Sigma}_{\boldsymbol{\beta}0}\boldsymbol{\beta}+\gamma p_l\sigma_{\alpha0}\left(\alpha+\eta_{\alpha0}\right)+\gamma\sigma_{\alpha0}v_{l,0}\left(\gamma,\mathbf{X},\mathbf{p},\mathbf{s},\boldsymbol{\theta}_{c0}\right)+\epsilon_l\right) f\left(\boldsymbol{\epsilon}\right)d\boldsymbol{\epsilon}\phi\left(\boldsymbol{\beta}\right)d\boldsymbol{\beta}\phi\left(\alpha\right)d\alpha\\
&= \int_{\boldsymbol{\beta}}\int_{\boldsymbol{\epsilon}}\frac{\partial}{\partial p_j}\int_\alpha\prod_{l\neq j} I\left(\mathbf{x}_j\boldsymbol{\Sigma}_{\boldsymbol{\beta}0}\boldsymbol{\beta}+\gamma p_j\sigma_{\alpha0}\left(\alpha+\eta_{\alpha0}\right)+\gamma\sigma_{\alpha0}v_{j,0}\left(\gamma,\mathbf{X},\mathbf{p},\mathbf{s},\boldsymbol{\theta}_{c0}\right)+\epsilon_j\right.\\
&\quad\left.\geq \mathbf{x}_l\boldsymbol{\Sigma}_{\boldsymbol{\beta}0}\boldsymbol{\beta}+\gamma p_l\sigma_{\alpha0}\left(\alpha+\eta_{\alpha0}\right)+\gamma\sigma_{\alpha0}v_{l,0}\left(\gamma,\mathbf{X},\mathbf{p},\mathbf{s},\boldsymbol{\theta}_{c0}\right)+\epsilon_l\right)\phi\left(\alpha\right)d\alpha f\left(\boldsymbol{\epsilon}\right)d\boldsymbol{\epsilon}\phi\left(\boldsymbol{\beta}\right)d\boldsymbol{\beta}
\end{aligned} \tag{63}$$

First, since the function $\prod_{l\neq j} I\left(\right)$ is nonnegative and the measure space of $(\alpha,\boldsymbol{\beta},\boldsymbol{\epsilon})$ is σ-finite, from Tonelli's Theorem, we can interchange the order of integrals. Next, we prove that the integral and the derivative above can be interchanged. We first focus on the function that is inside the integral over $\boldsymbol{\beta}$ and $\boldsymbol{\epsilon}$, which is,

$$\begin{aligned}
G\left(\mathbf{X},\mathbf{p},\mathbf{v}_0,\gamma,\boldsymbol{\beta},\boldsymbol{\epsilon},\boldsymbol{\theta}_{c0}\right) &\equiv \int_\alpha\prod_{l\neq j} I\left(\mathbf{x}_j\boldsymbol{\Sigma}_{\boldsymbol{\beta}0}\boldsymbol{\beta}+\gamma p_j\sigma_{\alpha0}\left(\alpha+\eta_{\alpha0}\right)+\gamma\sigma_{\alpha0}v_{j,0}+\epsilon_j\right.\\
&\geq \left.\mathbf{x}_l\boldsymbol{\Sigma}_{\boldsymbol{\beta}0}\boldsymbol{\beta}+\gamma p_l\sigma_{\alpha0}\left(\alpha+\eta_{\alpha0}\right)+\gamma\sigma_{\alpha0}v_{l,0}+\epsilon_l\right)\phi\left(\alpha\right)d\alpha.
\end{aligned}$$

It suffices to show that there exists a function $H\left(\right)$ such that

$$\left|\frac{\partial G\left(\mathbf{X},\mathbf{p},\mathbf{v}_0,\gamma,\boldsymbol{\beta},\boldsymbol{\epsilon},\boldsymbol{\theta}_{c0}\right)}{\partial p_j}\right|\leq H\left(\mathbf{X},\mathbf{p},\mathbf{v}_0,\gamma,\boldsymbol{\beta},\boldsymbol{\epsilon},\boldsymbol{\theta}_{c0}\right)$$

where $\int_{\boldsymbol{\beta}}\int_{\boldsymbol{\epsilon}} H\left(\mathbf{X},\mathbf{p},\mathbf{v}_0,\gamma,\boldsymbol{\beta},\boldsymbol{\epsilon},\boldsymbol{\theta}_{c0}\right)f\left(\boldsymbol{\epsilon}\right)d\boldsymbol{\epsilon}\phi\left(\boldsymbol{\beta}\right)d\boldsymbol{\beta}<\infty$. We show it as follows: First,

$$\begin{aligned}
&G\left(\mathbf{X},\mathbf{p},\mathbf{v}_0,\gamma,\boldsymbol{\beta},\boldsymbol{\epsilon},\boldsymbol{\theta}_{c0}\right)\\
=&\int_\alpha\prod_{p_j-p_l>0} I\left(\alpha\geq-\eta_{\alpha0}-\frac{v_{j,0}-v_{l,0}}{p_j-p_l}-\frac{\left(\mathbf{x}_j-\mathbf{x}_l\right)}{\gamma\left(p_j-p_l\right)\sigma_{\alpha0}}\boldsymbol{\Sigma}_{\boldsymbol{\beta}0}\boldsymbol{\beta}-\frac{\epsilon_j-\epsilon_l}{\gamma\left(p_j-p_l\right)\sigma_{\alpha0}}\right)\\
&\times\prod_{p_j-p_l<0} I\left(\alpha\leq-\eta_{\alpha0}-\frac{v_{j,0}-v_{l,0}}{p_j-p_l}-\frac{\left(\mathbf{x}_j-\mathbf{x}_l\right)}{\gamma\left(p_j-p_l\right)\sigma_{\alpha0}}\boldsymbol{\Sigma}_{\boldsymbol{\beta}0}\boldsymbol{\beta}-\frac{\epsilon_j-\epsilon_l}{\gamma\left(p_j-p_l\right)\sigma_{\alpha0}}\right)\phi\left(\alpha\right)d\alpha\\
=&\int_\alpha I\left(Max_{\{l:p_j-p_l>0\}}\left\{-\eta_{\alpha0}-\frac{v_{j,0}-v_{l,0}}{p_j-p_l}-\frac{\left(\mathbf{x}_j-\mathbf{x}_l\right)}{\gamma\left(p_j-p_l\right)\sigma_{\alpha0}}\boldsymbol{\Sigma}_{\boldsymbol{\beta}0}\boldsymbol{\beta}-\frac{\epsilon_j-\epsilon_l}{\gamma\left(p_j-p_l\right)\sigma_{\alpha0}}\right\}\leq\alpha\right.\\
\leq&\left. Min_{\{l:p_j-p_l<0\}}\left\{-\eta_{\alpha0}-\frac{v_{j,0}-v_{l,0}}{p_j-p_l}-\frac{\left(\mathbf{x}_j-\mathbf{x}_l\right)}{\gamma\left(p_j-p_l\right)\sigma_{\alpha0}}\boldsymbol{\Sigma}_{\boldsymbol{\beta}0}\boldsymbol{\beta}-\frac{\epsilon_j-\epsilon_l}{\gamma\left(p_j-p_l\right)\sigma_{\alpha0}}\right\}\right)\phi\left(\alpha\right)d\alpha.
\end{aligned} \tag{64}$$

Now, if we let

$$\begin{aligned} & \underline{A}\left(\mathbf{X},\mathbf{p},\mathbf{v}_0,\boldsymbol{\beta},\boldsymbol{\epsilon},j,\gamma,\boldsymbol{\theta}_{c0}\right) \\ \equiv\ & Max_{\{l:p_j-p_l>0\}}\left\{-\eta_{\alpha 0}-\frac{v_{j,0}-v_{l,0}}{p_j-p_l}-\frac{(\mathbf{x}_j-\mathbf{x}_l)}{\gamma\left(p_j-p_l\right)\sigma_{\alpha 0}}\boldsymbol{\Sigma}_{\boldsymbol{\beta}0}\boldsymbol{\beta}-\frac{\epsilon_j-\epsilon_l}{\gamma\left(p_j-p_l\right)\sigma_{\alpha 0}}\right\} \end{aligned} \tag{65}$$

$$\begin{aligned} & \overline{A}\left(\mathbf{X},\mathbf{p},\mathbf{v}_0,\boldsymbol{\beta},\boldsymbol{\epsilon},j,\gamma,\boldsymbol{\theta}_{c0}\right) \\ \equiv\ & Min_{\{l:p_j-p_l<0\}}\left\{-\eta_{\alpha 0}-\frac{v_{j,0}-v_{l,0}}{p_j-p_l}-\frac{(\mathbf{x}_j-\mathbf{x}_l)}{\gamma\left(p_j-p_l\right)\sigma_{\alpha 0}}\boldsymbol{\Sigma}_{\boldsymbol{\beta}0}\boldsymbol{\beta}-\frac{\epsilon_j-\epsilon_l}{\gamma\left(p_j-p_l\right)\sigma_{\alpha 0}}\right\} \end{aligned} \tag{66}$$

, then

$$\begin{aligned} (64) &= \int_\alpha I\left(\underline{A}\left(\mathbf{X},\mathbf{p},\mathbf{v}_0,\boldsymbol{\beta},\boldsymbol{\epsilon},j,\gamma,\boldsymbol{\theta}_{c0}\right)\le\alpha\le\overline{A}\left(\mathbf{X},\mathbf{p},\mathbf{v}_0,\boldsymbol{\beta},\boldsymbol{\epsilon},j,\gamma,\boldsymbol{\theta}_{c0}\right)\right)\phi\left(\alpha\right)d\alpha \\ &= \left[\Phi\left(\overline{A}\left(\mathbf{X},\mathbf{p},\mathbf{v}_0,\boldsymbol{\beta},\boldsymbol{\epsilon},j,\gamma,\boldsymbol{\theta}_{c0}\right)\right)-\Phi\left(\underline{A}\left(\mathbf{X},\mathbf{p},\mathbf{v}_0,\boldsymbol{\beta},\boldsymbol{\epsilon},j,\gamma,\boldsymbol{\theta}_{c0}\right)\right)\right] \\ &\quad \times I\left[\underline{A}\left(\mathbf{X},\mathbf{p},\mathbf{v}_0,\boldsymbol{\beta},\boldsymbol{\epsilon},j,\gamma,\boldsymbol{\theta}_{c0}\right)<\overline{A}\left(\mathbf{X},\mathbf{p},\mathbf{v}_0,\boldsymbol{\beta},\boldsymbol{\epsilon},j,\gamma,\boldsymbol{\theta}_{c0}\right)\right] \end{aligned}$$

Thus, taking the derivative of (64) with respect to p_j given $v_{j,0}$ being constant, we obtain

$$\begin{aligned} & \frac{\partial}{\partial p_j}\int_\alpha I\left(\underline{A}\left(\mathbf{X},\mathbf{p},\mathbf{v}_0,\boldsymbol{\beta},\boldsymbol{\epsilon},j,\gamma,\boldsymbol{\theta}_{c0}\right)\le\alpha\le\overline{A}\left(\mathbf{X},\mathbf{p},\mathbf{v}_0,\boldsymbol{\beta},\boldsymbol{\epsilon},j,\gamma,\boldsymbol{\theta}_{c0}\right)\right)\phi\left(\alpha\right)d\alpha \\ =\ & \left[\phi\left(\overline{A}\right)\frac{\partial\overline{A}\left(\mathbf{X},\mathbf{p},\mathbf{v}_0,\boldsymbol{\beta},\boldsymbol{\epsilon},j,\gamma,\boldsymbol{\theta}_{c0}\right)}{\partial p_j}-\phi\left(\underline{A}\right)\frac{\partial\underline{A}\left(\mathbf{X},\mathbf{p},\mathbf{v}_0,\boldsymbol{\beta},\boldsymbol{\epsilon},j,\gamma,\boldsymbol{\theta}_{c0}\right)}{\partial p_j}\right]I\left[\underline{A}<\overline{A}\right] \end{aligned}$$

Now, let

$$\mathcal{L}^*\left(j\right)=argmax_{\{l:p_j-p_l>0\}}\left\{-\eta_{\alpha 0}-\frac{v_{j,0}-v_{l,0}}{p_j-p_l}-\frac{(\mathbf{x}_j-\mathbf{x}_l)}{\gamma\left(p_j-p_l\right)\sigma_{\alpha 0}}\boldsymbol{\Sigma}_{\boldsymbol{\beta}0}\boldsymbol{\beta}-\frac{\epsilon_j-\epsilon_l}{\gamma\left(p_j-p_l\right)\sigma_{\alpha 0}}\right\}. \tag{67}$$

Then, $\mathcal{L}^*\left(j\right)$ may have multiple elements. But the values of $\boldsymbol{\beta}$ and $\boldsymbol{\epsilon}$ that results in multiple elements have measure zero, and other than those cases, $\mathcal{L}^*\left(j\right)$ is a singleton. We denote the singleton element to be $l^*\left(j\right)$. Then,

$$\frac{\partial\underline{A}\left(\mathbf{X},\mathbf{p},\mathbf{v}_0,\boldsymbol{\beta},\boldsymbol{\epsilon},j,\gamma,\boldsymbol{\theta}_{c0}\right)}{\partial p_j}=\frac{v_{j,0}-v_{l^*(j),0}}{\left(p_j-p_{l^*(j)}\right)^2}+\frac{1}{\gamma}\frac{\left(\mathbf{x}_j-\mathbf{x}_{l^*(j)}\right)}{\left(p_j-p_{l^*(j)}\right)^2\sigma_{\alpha 0}}\boldsymbol{\Sigma}_{\boldsymbol{\beta}}\boldsymbol{\beta}+\frac{1}{\gamma}\frac{\epsilon_j-\epsilon_{l^*(j)}}{\left(p_j-p_{l^*(j)}\right)^2\sigma_{\alpha 0}}$$

except for the values of $\boldsymbol{\beta}$ and $\boldsymbol{\epsilon}$ when $\mathcal{L}^*\left(j\right)$ contains more than one element. Then, because

$\gamma > \underline{\gamma}$,

$$\left|\frac{\partial \underline{A}\left(\mathbf{X},\mathbf{p},\mathbf{v}_0,\boldsymbol{\beta},\boldsymbol{\epsilon},j,\gamma,\boldsymbol{\theta}_{c0}\right)}{\partial p_j}\right|$$

$$\leq \quad \frac{\max_l \left|v_{j,0}-v_{l,0}\right|}{\left(p_j-p_l\right)^2}+\max_{k,l}\left|\frac{\left[\left(\mathbf{x}_j-\mathbf{x}_l\right)\boldsymbol{\Sigma}_{\boldsymbol{\beta}0}\right]_k}{\underline{\gamma}\left(p_j-p_l\right)^2\sigma_{\alpha 0}}\right|\sum_k |\beta|_k+\max_l \frac{\left|\epsilon_j-\epsilon_l\right|}{\underline{\gamma}\left(p_j-p_l\right)^2\sigma_{\alpha 0}} \qquad (68)$$

Since $\boldsymbol{\beta}$, $\boldsymbol{\epsilon}$ are i.i.d. normally distributed, the RHS of the above inequality is integrable with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\epsilon}$ for any $\underline{\gamma} > 0$. Similar results hold for $\partial \overline{A}\left(\mathbf{X},\mathbf{p},\mathbf{v}_0,\boldsymbol{\beta},\boldsymbol{\epsilon},j,\gamma,\boldsymbol{\theta}_0\right)/\partial p_j$. Therefore, the derivative and the integral are interchangeable. Hence,

$$\frac{\partial}{\partial p_j}\int_{\boldsymbol{\beta}}\int_{\boldsymbol{\epsilon}}\int_{\alpha} I\left(\underline{A}\left(\mathbf{X},\mathbf{p},\mathbf{v}_0,\boldsymbol{\beta},\boldsymbol{\epsilon},j,\gamma,\boldsymbol{\theta}_{c0}\right)\leq\alpha\leq\overline{A}\left(\mathbf{X},\mathbf{p},\mathbf{v}_0,\boldsymbol{\beta},\boldsymbol{\epsilon},j,\gamma,\boldsymbol{\theta}_{c0}\right)\right)\phi\left(\alpha\right)d\alpha f\left(\boldsymbol{\epsilon}\right)d\boldsymbol{\epsilon}\phi\left(\boldsymbol{\beta}\right)d\boldsymbol{\beta}$$

$$=\int_{\boldsymbol{\beta}}\int_{\boldsymbol{\epsilon}}\left[\phi\left(\overline{A}\right)\frac{\partial \overline{A}\left(\mathbf{X},\mathbf{p},\mathbf{v}_0,\boldsymbol{\beta},\boldsymbol{\epsilon},j,\gamma,\boldsymbol{\theta}_{c0}\right)}{\partial p_j}-\phi\left(\underline{A}\right)\frac{\partial \underline{A}\left(\mathbf{X},\mathbf{p},\mathbf{v}_0,\boldsymbol{\beta},\boldsymbol{\epsilon},j,\gamma,\boldsymbol{\theta}_{c0}\right)}{\partial p_j}\right]$$

$$\times I\left[\underline{A}\left(\mathbf{X},\mathbf{p},\mathbf{v}_0,\boldsymbol{\beta},\boldsymbol{\epsilon},j,\gamma,\boldsymbol{\theta}_{c0}\right)<\overline{A}\left(\mathbf{X},\mathbf{p},\mathbf{v}_0,\boldsymbol{\beta},\boldsymbol{\epsilon},j,\gamma,\boldsymbol{\theta}_{c0}\right)\right]f\left(\boldsymbol{\epsilon}\right)d\boldsymbol{\epsilon}\phi\left(\boldsymbol{\beta}\right)d\boldsymbol{\beta}.$$

Now, from Equation (65), because $lim_{\gamma\to\infty}\mathbf{v}_0\left(\mathbf{X},\gamma\mathbf{p},\mathbf{s},\boldsymbol{\theta}_{c0}\right)=\widetilde{\mathbf{v}}_0$,

$$lim_{\gamma\to\infty}\underline{A}\left(\mathbf{X},\mathbf{p},\mathbf{v}_0,\boldsymbol{\beta},\boldsymbol{\epsilon},j,\gamma,\boldsymbol{\theta}_{c0}\right)=max_{l:p_j-p_l>0}\left\{-\eta_{\alpha 0}-\frac{\widetilde{v}_{j,0}-\widetilde{v}_{l,0}}{p_j-p_l}\right\}=-\eta_{\alpha 0}-\underline{a}_j$$

and similarly,

$$lim_{\gamma\to\infty}\overline{A}\left(\mathbf{X},\mathbf{p},\mathbf{v}_0,\boldsymbol{\beta},\boldsymbol{\epsilon},j,\gamma,\boldsymbol{\theta}_{c0}\right)=max_{l:p_j-p_l<0}\left\{-\eta_{\alpha 0}-\frac{\widetilde{v}_{j,0}-\widetilde{v}_{l,0}}{p_j-p_l}\right\}=-\eta_{\alpha 0}-\overline{a}_j$$

$$\mathcal{L}^*\left(j\right)\to argmax_{\{l:p_j-p_l\geq 0\}}\left\{-\eta_{\alpha 0}-\frac{\widetilde{v}_{j,0}-\widetilde{v}_{l,0}}{p_j-p_l}\right\}\ as\ \gamma\to\infty.$$

Hence,

$$I\left(\underline{A}\left(\mathbf{X},\mathbf{p},\mathbf{v}_0,\boldsymbol{\beta},\boldsymbol{\epsilon},j,\gamma,\boldsymbol{\theta}_{c0}\right)<\overline{A}\left(\mathbf{X},\mathbf{p},\mathbf{v}_0,\boldsymbol{\beta},\boldsymbol{\epsilon},j,\gamma,\boldsymbol{\theta}_{c0}\right)\right)$$

$$\to I(-\underline{a}_j<-\overline{a}_j)=1\ as\ \gamma\to\infty$$

Therefore,

$$\frac{\partial \underline{A}\left(\mathbf{X},\mathbf{p},\mathbf{v}_0,\boldsymbol{\beta},\boldsymbol{\epsilon},j,\gamma,\boldsymbol{\theta}_{c0}\right)}{\partial p_j}\to-\frac{\partial}{\partial p_j}\frac{\widetilde{v}_{j-1,0}-\widetilde{v}_{j,0}}{p_{j-1}-p_j}\ as\ \gamma\to\infty.$$

Similarly,

$$\frac{\partial \overline{A}\left(\mathbf{X},\mathbf{p},\mathbf{v}_0,\boldsymbol{\beta},\boldsymbol{\epsilon},j,\gamma,\boldsymbol{\theta}_{c0}\right)}{\partial p_j}\to-\frac{\partial}{\partial p_j}\frac{\widetilde{v}_{j,0}-\widetilde{v}_{j+1,0}}{p_j-p_{j+1}}\ as\ \gamma\to\infty.$$

Therefore,

$$
\begin{aligned}
& \frac{\partial}{\partial p_j} \int_{\alpha} \prod_{l=0, l\neq j}^{J} I\left(\mathbf{x}_j \boldsymbol{\Sigma}_{\boldsymbol{\beta}0} \boldsymbol{\beta} + \gamma p_j \sigma_{\alpha 0}\left(\alpha + \eta_{\alpha 0}\right) + \gamma \sigma_{\alpha 0} v_{j,0} + \epsilon_j \right. \\
\geq \quad & \left. \mathbf{x}_l \boldsymbol{\Sigma}_{\boldsymbol{\beta}0} \boldsymbol{\beta} + \gamma p_l \sigma_{\alpha 0}\left(\alpha + \eta_{\alpha 0}\right) + \gamma \sigma_{\alpha 0} v_{l,0} + \epsilon_l \right) \phi\left(\alpha\right) d\alpha \\
\to \quad & \phi\left(-\eta_{\alpha 0} - \frac{\widetilde{v}_{j,0} - \widetilde{v}_{j+1,0}}{p_j - p_{j+1}}\right) \frac{\widetilde{v}_{j,0} - \widetilde{v}_{j+1,0}}{\left(p_j - p_{j+1}\right)^2} + \phi\left(-\eta_{\alpha 0} - \frac{\widetilde{v}_{j-1,0} - \widetilde{v}_{j,0}}{p_{j-1} - p_j}\right) \frac{\widetilde{v}_{j-1,0} - \widetilde{v}_{j,0}}{\left(p_{j-1} - p_j\right)^2}
\end{aligned}
$$

Furthermore, Equation (68) shows that the conditions for the Dominated Convergence Theorem holds so that the limit and the integral can be interchanged to derive

$$
\begin{aligned}
& lim_{\gamma \to \infty} \int_{\boldsymbol{\beta}} \int_{\boldsymbol{\epsilon}} \left[\frac{\partial}{\partial p_j} \int_{\alpha} \prod_{l=0, l\neq j}^{J} I\left(\mathbf{x}_j \boldsymbol{\Sigma}_{\boldsymbol{\beta}0} \boldsymbol{\beta} + \gamma p_j \sigma_{\alpha 0}\left(\alpha + \eta_{\alpha 0}\right) + \gamma \sigma_{\alpha 0} v_{j,0}\left(\gamma\right) + \epsilon_j \right.\right. \\
\geq \quad & \left.\left. \mathbf{x}_l \boldsymbol{\Sigma}_{\boldsymbol{\beta}0} \boldsymbol{\beta} + \gamma p_l \sigma_{\alpha 0}\left(\alpha + \eta_{\alpha 0}\right) + \gamma \sigma_{\alpha 0} v_{l,0} + \epsilon_l \right) \phi\left(\alpha\right) d\alpha \right] f\left(\boldsymbol{\epsilon}\right) d\boldsymbol{\epsilon} \phi\left(\boldsymbol{\beta}\right) d\boldsymbol{\beta} \\
= \quad & \int_{\boldsymbol{\beta}} \int_{\boldsymbol{\epsilon}} lim_{\gamma \to \infty} \left[\phi\left(\overline{A}\right) \frac{\partial \overline{A}\left(\mathbf{X}, \mathbf{p}, \mathbf{v}_0, \boldsymbol{\beta}, \boldsymbol{\epsilon}, j, \gamma, \boldsymbol{\theta}_{c0}\right)}{\partial p_j} - \phi\left(\underline{A}\right) \frac{\partial \underline{A}\left(\mathbf{X}, \mathbf{p}, \mathbf{v}_0, \boldsymbol{\beta}, \boldsymbol{\epsilon}, j, \gamma, \boldsymbol{\theta}_{c0}\right)}{\partial p_j} \right] \\
\times \quad & I\left[\underline{A}\left(\mathbf{X}, \mathbf{p}, \mathbf{v}_0, \boldsymbol{\beta}, \boldsymbol{\epsilon}, j, \gamma, \boldsymbol{\theta}_{c0}\right) < \overline{A}\left(\mathbf{X}, \mathbf{p}, \mathbf{v}_0, \boldsymbol{\beta}, \boldsymbol{\epsilon}, j, \gamma, \boldsymbol{\theta}_{c0}\right)\right] f\left(\boldsymbol{\epsilon}\right) d\boldsymbol{\epsilon} \phi\left(\boldsymbol{\beta}\right) d\boldsymbol{\beta}. \\
= \quad & \phi\left(-\eta_{\alpha 0} - \frac{\widetilde{v}_{j,0} - \widetilde{v}_{j+1,0}}{p_j - p_{j+1}}\right) \frac{\widetilde{v}_{j,0} - \widetilde{v}_{j+1,0}}{\left(p_j - p_{j+1}\right)^2} + \phi\left(-\eta_{\alpha 0} - \frac{\widetilde{v}_{j-1,0} - \widetilde{v}_{j,0}}{p_{j-1} - p_j}\right) \frac{\widetilde{v}_{j-1,0} - \widetilde{v}_{j,0}}{\left(p_{j-1} - p_j\right)^2}
\end{aligned}
$$

Therefore, we have shown that the derivative of the limiting market share function is the same as the limit of its derivative.

We now argue that $\eta_{\alpha 0}$ is identified from the equality of marginal revenues at the limit. From what we have derived, the following holds

$$
\begin{aligned}
& lim_{\gamma \to \infty} \frac{MR_j\left(\mathbf{X}, \gamma \mathbf{p}, \mathbf{s}, \boldsymbol{\theta}_{c0}\right)}{\gamma} \\
& = lim_{\gamma \to \infty} \frac{1}{\gamma} \left[\gamma p_j + \left[\frac{\partial s_j\left(\mathbf{X}, \gamma \mathbf{p}, \mathbf{s}, \boldsymbol{\theta}_{c0}\right)}{\partial\left(\gamma p_j\right)} \right]^{-1} s_j \right] = lim_{\gamma \to \infty} \frac{1}{\gamma} \left[\gamma p_j + \gamma \left[\frac{\partial s_j\left(\mathbf{X}, \gamma \mathbf{p}, \mathbf{s}, \boldsymbol{\theta}_{c0}\right)}{\partial p_j} \right]^{-1} s_j \right] \\
& = p_j + \left[\phi\left(-\eta_{\alpha 0} - \frac{\widetilde{v}_{j,0} - \widetilde{v}_{j+1,0}}{p_j - p_{j+1}}\right) \frac{\widetilde{v}_{j,0} - \widetilde{v}_{j+1,0}}{\left(p_j - p_{j+1}\right)^2} + \phi\left(-\eta_{\alpha 0} - \frac{\widetilde{v}_{j-1,0} - \widetilde{v}_{j,0}}{p_{j-1} - p_j}\right) \frac{\widetilde{v}_{j-1,0} - \widetilde{v}_{j,0}}{\left(p_{j-1} - p_j\right)^2} \right]^{-1} s_j \qquad (69) \\
= \quad & p_j + \left[-\phi\left(\Phi^{-1}\left(\sum_{l \leq j} s_l\right)\right) \frac{\Phi^{-1}\left(\sum_{l \leq j} s_l\right) + \eta_{\alpha 0}}{p_j - p_{j+1}} - \phi\left(\Phi^{-1}\left(\sum_{l \leq j-1} s_l\right)\right) \frac{\Phi^{-1}\left(\sum_{l \leq j-1} s_l\right) + \eta_{\alpha 0}}{p_{j-1} - p_j} \right]^{-1} s_j \qquad (70)
\end{aligned}
$$

where $\widetilde{v}_{00} = 0$, $p_0 = 0$, and

$$\Phi\left(-\eta_{\alpha 0} - \frac{\widetilde{v}_{i,0} - \widetilde{v}_{i+1,0}}{p_i - p_{i+1}}\right) = \sum_{l \leq i} s_l.$$

We can similarly write down the limit of the marginal revenue function of the second market.

We now show that there exist $\mathbf{p}$, $\mathbf{p}^{\dagger}$, $\mathbf{s}$, $\mathbf{s}^{\dagger}$ and $j \in \{1, \ldots J-1\}$, $j^{\dagger} \in \left\{1, \ldots, J^{\dagger}-1\right\}$ such that

$$lim_{\gamma \to \infty} \frac{MR_j\left(\mathbf{X}, \gamma\mathbf{p}, \mathbf{s}, \boldsymbol{\theta}_{c0}\right)}{\gamma} = lim_{\gamma \to \infty} \frac{MR_{j^{\dagger}}\left(\mathbf{X}^{\dagger}, \gamma\mathbf{p}^{\dagger}, \mathbf{s}^{\dagger}, \boldsymbol{\theta}_{c0}\right)}{\gamma}. \tag{71}$$

Further, we can ensure that the derivative of the market share function with respect to price is negative by first choosing $\mathbf{s}$ and $\mathbf{s}^{\dagger}$so that $\sum_{l \leq j} s_l$ and $\sum_{l \leq j^{\dagger}} s_l^{\dagger}$ are small enough that $\Phi^{-1}\left(\sum_{l \leq j} s_l\right) + \eta_{\alpha 0} < 0$ as well as $\Phi^{-1}\left(\sum_{l \leq j^{\dagger}} s_l^{\dagger}\right) + \eta_{\alpha 0} < 0$. The equality above holds for some $(\mathbf{p}, \mathbf{s})$ and $\left(\mathbf{p}^{\dagger}, \mathbf{s}^{\dagger}\right)$ because if it does not hold, then we can adjust $p_j - p^{\dagger}_{j^{\dagger}}$ while keeping $p_j - p_{j+1}$, $j = 1, \ldots, J-1$, $p^{\dagger}_{j^{\dagger}} - p^{\dagger}_{j^{\dagger}+1}$, $j^{\dagger} = 1, \ldots, J^{\dagger}-1$ constant. Note that we can keep $\mathbf{s}$, $\mathbf{s}^{\dagger}$ the same as well by adjusting the demand shocks. Further, we can make p_j and $p^{\dagger}_{j^{\dagger}}$ large enough so that marginal revenues are positive. Then, Equation (71) identifies $\eta_{\alpha 0}$. To see why, consider the case where $p_j = p^{\dagger}_{j^{\dagger}}$. Then, from Equation (71), we derive

$$\eta_{\alpha 0} = \frac{\left(s^{\dagger}_{j^{\dagger}}/s_j\right)\left(C\left(\mathbf{p}, \mathbf{s}, j, j+1\right) + C\left(\mathbf{p}, \mathbf{s}, j-1, j\right)\right) - C\left(\mathbf{p}^{\dagger}, \mathbf{s}^{\dagger}, j^{\dagger}, j^{\dagger}+1\right) - C\left(\mathbf{p}^{\dagger}, \mathbf{s}^{\dagger}, j^{\dagger}-1, j^{\dagger}\right)}{B\left(\mathbf{p}, \mathbf{s}, j, j+1\right) + B\left(\mathbf{p}, \mathbf{s}, j-1, j\right) - \left(s^{\dagger}_{j^{\dagger}}/s_j\right)\left(B\left(\mathbf{p}^{\dagger}, \mathbf{s}^{\dagger}, j^{\dagger}, j^{\dagger}+1\right) + B\left(\mathbf{p}^{\dagger}, \mathbf{s}^{\dagger}, j^{\dagger}-1, j^{\dagger}\right)\right)}$$

where $B\left(\mathbf{p}, \mathbf{s}, j, j+1\right) \equiv \phi\left(\Phi^{-1}\left(\sum_{l \leq j} s_l\right)\right) / \left(p_j - p_{j+1}\right)$,

$C\left(\mathbf{p}, \mathbf{s}, j, j+1\right) \equiv \phi\left(\Phi^{-1}\left(\sum_{l \leq j} s_l\right)\right) \Phi^{-1}\left(\sum_{l \leq j} s_l\right) / \left(p_j - p_{j+1}\right)$ and these expressions for the second market are similarly defined. We can identify $\eta_{\alpha 0}$ as long as we can find $(\mathbf{p}, \mathbf{s})$ and $\left(\mathbf{p}^{\dagger}, \mathbf{s}^{\dagger}\right)$ such that the denominator is nonzero. Such price-market share combinations exist. For example, set $j = j^{\dagger}$, $\mathbf{s} = \mathbf{s}^{\dagger}$, $p_{j-1} - p_j = p^{\dagger}_{j^{\dagger}-1} - p^{\dagger}_{j^{\dagger}}$, $p_j - p_{j+1} \neq p^{\dagger}_{j^{\dagger}} - p^{\dagger}_{j^{\dagger}+1}$, and $\Phi^{-1}\left(\sum_{l \leq j} s_l\right) = -\eta_{\alpha 0}$. Note that since $\mathbf{s} = \mathbf{s}^{\dagger}$, market size variation is not needed for identification of $\eta_{\alpha 0}$.

Note also that while the above equality holds for some $(\mathbf{p}, \mathbf{s})$, $\left(\mathbf{p}^{\dagger}, \mathbf{s}^{\dagger}\right)$ and identifies $\eta_{\alpha 0}$, the price vectors used are not the actual prices which are infinite.

Then, However, we can show that the equality holds also for large γ, and identifies $\eta_{\alpha 0}$ as follows: Let

$$A\left(\mathbf{p},\mathbf{s},\eta_{\alpha 0},j\right)\equiv lim_{\gamma\to\infty}\frac{MR_j\left(\mathbf{X},\gamma\mathbf{p},\mathbf{s},\boldsymbol{\theta}_{c0}\right)}{\gamma}$$

$$= \quad p_j+\left[-\phi\left(\Phi^{-1}\left(\sum_{l\leq j}s_l\right)\right)\frac{\Phi^{-1}\left(\sum_{l\leq j}s_l\right)+\eta_{\alpha 0}}{p_j-p_{j+1}}-\phi\left(\Phi^{-1}\left(\sum_{l\leq j-1}s_l\right)\right)\frac{\Phi^{-1}\left(\sum_{l\leq j-1}s_l\right)+\eta_{\alpha 0}}{p_{j-1}-p_j}\right]^{-1}s_j$$

Then, for any $\eta_\alpha\neq\eta_{\alpha 0}$, $A\left(\mathbf{p},\mathbf{s},\eta_\alpha,j\right)\neq A\left(\mathbf{p}^\dagger,\mathbf{s}^\dagger,\eta_\alpha,j\right)$. Now, choose $\mathbf{p}'$ such that $p'_i-p'_{i+1}=p_i-p_{i+1}$ for $i=1,\ldots,J-1.$ and $p'_j=p_j+\delta$ for some small $\delta>0$. Then,

$$A\left(\mathbf{p}',\mathbf{s},\eta_{\alpha 0},j\right)-A\left(\mathbf{p},\mathbf{s},\eta_{\alpha 0},j\right)=\delta,\ A\left(\mathbf{p}',\mathbf{s},\eta_\alpha,j\right)-A\left(\mathbf{p},\mathbf{s},\eta_\alpha,j\right)=\delta$$

Similarly, we choose prices $\mathbf{p}''$ such that $p''_i-p''_{i+1}=p_i-p_{i+1}$ for $i=1,\ldots,J-1.$ and $p''_j=p_j-\delta$. Then,

$$A\left(\mathbf{p}'',\mathbf{s},\eta_{\alpha 0},j\right)-A\left(\mathbf{p},\mathbf{s},\eta_{\alpha 0},j\right)=-\delta,\ A\left(\mathbf{p}'',\mathbf{s},\eta_\alpha,j\right)-A\left(\mathbf{p},\mathbf{s},\eta_\alpha,j\right)=-\delta.$$

For any $\delta>0$, for a sufficiently large $\gamma>0$,

$$\left|\frac{MR_j\left(\mathbf{X},\gamma\mathbf{p}',\mathbf{s},\boldsymbol{\theta}_{c0}\right)}{\gamma}-A\left(\mathbf{p}',\mathbf{s},\eta_{\alpha 0},j\right)\right|<\frac{\delta}{3},\ \left|\frac{MR_j\left(\mathbf{X}^\dagger,\gamma\mathbf{p}^\dagger,\mathbf{s}^\dagger,\boldsymbol{\theta}_{c0}\right)}{\gamma}-A\left(\mathbf{p}^\dagger,\mathbf{s}^\dagger,\eta_{\alpha 0},j\right)\right|<\frac{\delta}{3}$$

$$\left|\frac{MR_j\left(\mathbf{X},\gamma\mathbf{p}'',\mathbf{s},\boldsymbol{\theta}_{c0}\right)}{\gamma}-A\left(\mathbf{p}'',\mathbf{s},\eta_{\alpha 0},j\right)\right|<\frac{\delta}{3},$$

$$\left|\frac{MR_j\left(\mathbf{X},\gamma\mathbf{p},\mathbf{s},\boldsymbol{\theta}_c\right)}{\gamma}-A\left(\mathbf{p},\mathbf{s},\eta_\alpha,j\right)\right|<\frac{\delta}{3},\ \left|\frac{MR_j\left(\mathbf{X}^\dagger,\gamma\mathbf{p}^\dagger,\mathbf{s}^\dagger,\boldsymbol{\theta}_c\right)}{\gamma}-A\left(\mathbf{p}^\dagger,\mathbf{s}^\dagger,\eta_\alpha,j\right)\right|<\frac{\delta}{3},$$

Then,

$$\frac{MR_i\left(\mathbf{X},\gamma\mathbf{p}',\mathbf{s},\boldsymbol{\theta}_{c0}\right)-MR_i\left(\mathbf{X}^\dagger,\gamma\mathbf{p}^\dagger,\mathbf{s}^\dagger,\boldsymbol{\theta}_{c0}\right)}{\gamma}>\frac{\delta}{3}>0$$

$$\frac{MR_i\left(\mathbf{X}^\dagger,\gamma\mathbf{p}'',\mathbf{s},\boldsymbol{\theta}_{c0}\right)-MR_i\left(\mathbf{X}^\dagger,\gamma\mathbf{p}^\dagger,\mathbf{s}^\dagger,\boldsymbol{\theta}_{c0}\right)}{\gamma}<-\frac{\delta}{3}<0$$

Therefore, from continuity of the marginal revenue function, it follows from the Intermediate Value Theorem that there exists $\widetilde{\mathbf{p}}$ such that $|\widetilde{\mathbf{p}}-\mathbf{p}|<\delta$, and

$$\frac{MR_i\left(\mathbf{X},\gamma\widetilde{\mathbf{p}},\mathbf{s},\boldsymbol{\theta}_{c0}\right)-MR_i\left(\mathbf{X}^\dagger,\gamma\mathbf{p}^\dagger,\mathbf{s}^\dagger,\boldsymbol{\theta}_{c0}\right)}{\gamma}=0.$$

Furthermore, from the Triangle Inequality,

$$\begin{aligned}
& \left|A(\mathbf{p},\mathbf{s},\eta_\alpha,j) - A\left(\mathbf{p}^\dagger,\mathbf{s}^\dagger,\eta_\alpha,j\right)\right| \\
\leq\ & \left|A(\mathbf{p},\mathbf{s},\eta_\alpha,j) - A(\widetilde{\mathbf{p}},\mathbf{s},\eta_\alpha,j)\right| + \left|A(\widetilde{\mathbf{p}},\mathbf{s},\eta_\alpha,j) - A\left(\mathbf{p}^\dagger,\mathbf{s}^\dagger,\eta_\alpha,j\right)\right| \\
\leq\ & \left|A(\mathbf{p},\mathbf{s},\eta_\alpha,j) - A(\widetilde{\mathbf{p}},\mathbf{s},\eta_\alpha,j)\right| + \left|A(\widetilde{\mathbf{p}},\mathbf{s},\eta_\alpha,j) - \frac{MR_j(\mathbf{X},\gamma\widetilde{\mathbf{p}},\mathbf{s},\boldsymbol{\theta}_c)}{\gamma}\right| \\
& + \left|\frac{MR_j\left(\mathbf{X},\gamma\mathbf{p}^\dagger,\mathbf{s}^\dagger,\boldsymbol{\theta}_c\right)}{\gamma} - A\left(\mathbf{p}^\dagger,\mathbf{s}^\dagger,\eta_\alpha,j\right)\right| + \frac{\left|MR_j(\mathbf{X},\gamma\widetilde{\mathbf{p}},\mathbf{s},\boldsymbol{\theta_c}) - MR_j\left(\mathbf{X},\gamma\mathbf{p}^\dagger,\mathbf{s}^\dagger,\boldsymbol{\theta_c}\right)\right|}{\gamma}
\end{aligned}$$

Therefore, for sufficiently small $\delta > 0$,

$$\begin{aligned}
& \frac{\left|MR_j(\mathbf{X},\gamma\widetilde{\mathbf{p}},\mathbf{s},\boldsymbol{\theta_c}) - MR_j\left(\mathbf{X},\gamma\mathbf{p}^\dagger,\mathbf{s}^\dagger,\boldsymbol{\theta_c}\right)\right|}{\gamma} \\
& \geq \left|A(\mathbf{p},\mathbf{s},\eta_\alpha,j) - A\left(\mathbf{p}^\dagger,\mathbf{s}^\dagger,\eta_\alpha,j\right)\right| - \left|A(\mathbf{p},\mathbf{s},\eta_\alpha,j) - A(\widetilde{\mathbf{p}},\mathbf{s},\eta_\alpha,j)\right| - \frac{2}{3}\delta \\
& > 0
\end{aligned}$$

This is because $\left|A(\mathbf{p},\mathbf{s},\eta_\alpha,j) - A\left(\mathbf{p}^\dagger,\mathbf{s}^\dagger,\eta_\alpha,j\right)\right| > 0$ from identification, and from continuity of $A(\mathbf{p},\mathbf{s},\eta_\alpha,j)$ with respect to $\mathbf{p}$, $\left|A(\mathbf{p},\mathbf{s},\eta_\alpha,j) - A(\widetilde{\mathbf{p}},\mathbf{s},\eta_\alpha,j)\right|$ can be made arbitrarily small for sufficiently small $\delta > 0$. Given the assumptions, we can find firms in the population in different markets m and m' satisfying $(\mathbf{X}_m,\mathbf{p}_m,\mathbf{s}_m) = (\mathbf{X},\gamma\widetilde{\mathbf{p}},\mathbf{s})$, $(\mathbf{X}_{m'},\mathbf{p}_{m'},\mathbf{s}_{m'}) = \left(\mathbf{X}^\dagger,\gamma\widetilde{\mathbf{p}}^\dagger,\mathbf{s}^\dagger\right)$. Therefore, $\eta_{\alpha 0}$ is identified.

We next show that the market size is identified. Recall that the equation for market size is $log(Q) = \lambda_0 + \mathbf{z}\boldsymbol{\lambda}_{z0}$. Let $\mathbf{s}$ be the true market share, i.e. $ln(s_{im}) = ln(q_{im}) - ln(Q_m) = ln(q_{im}) - \lambda_0 - \mathbf{z}\boldsymbol{\lambda}_{z0}$. Then, the recovered market share satisfies $\widehat{\mathbf{s}} = \chi(\mathbf{z})\,\mathbf{s}$, where $\chi(\mathbf{z}) \equiv exp\left[(\lambda_0 + \mathbf{z}\boldsymbol{\lambda}_{z0}) - \left(\widehat{\lambda} + \mathbf{z}\widehat{\boldsymbol{\lambda}}_z\right)\right]$. Then, if $\left(\widehat{\lambda},\widehat{\boldsymbol{\lambda}}_z\right) = (\lambda_0,\boldsymbol{\lambda}_{z0})$, $\chi(\mathbf{z}) = 1$ for all $\mathbf{z}$. On the other hand, if $\left(\widehat{\lambda},\widehat{\boldsymbol{\lambda}}_z\right) \neq (\lambda_0,\boldsymbol{\lambda}_{z0})$, then for some $\mathbf{z}$, we have $\chi(\mathbf{z}) \neq 1$. Hence, we show below that $\chi(\mathbf{z}) = 1$ for all $\mathbf{z}$.

To do so, we choose two firms with $\mathbf{z} = \mathbf{z}^\dagger$ so that $\widehat{\mathbf{s}} = \chi\mathbf{s}$, $\widehat{\mathbf{s}}^\dagger = \chi\mathbf{s}^\dagger$. Then, expected cost conditional on observables when true market share is observed is the same for these two firms if and only if it is also the same when the true market share is not observed. That is,

$$E\left[C\middle|\left(\widetilde{q} = q,\widetilde{\mathbf{w}} = \mathbf{w},\widetilde{\mathbf{p}} = \mathbf{p},\widetilde{\mathbf{s}} = \widehat{\mathbf{s}},\widetilde{\mathbf{X}} = \mathbf{X},j\right)\right] = E\left[C\middle|\left(\widetilde{q} = q,\widetilde{\mathbf{w}} = \mathbf{w},\widetilde{\mathbf{p}} = \mathbf{p}^\dagger,\widetilde{\mathbf{s}} = \widehat{\mathbf{s}}^\dagger,\widetilde{\mathbf{X}} = \mathbf{X}^\dagger,j^\dagger\right)\right] \tag{72}$$

if and only if

$$E\left[C|\left(\widetilde{q}=q,\widetilde{\mathbf{w}}=\mathbf{w},\widetilde{\mathbf{p}}=\mathbf{p},\widetilde{\mathbf{s}}=\mathbf{s},\widetilde{\mathbf{X}}=\mathbf{X},j\right)\right]=E\left[C|\left(\widetilde{q}=q,\widetilde{\mathbf{w}}=\mathbf{w},\widetilde{\mathbf{p}}=\mathbf{p}^{\dagger},\widetilde{\mathbf{s}}=\mathbf{s}^{\dagger},\widetilde{\mathbf{X}}=\mathbf{X}^{\dagger},j^{\dagger}\right)\right]$$

Therefore, from Lemma 1, for those two firms with $q=q^{\dagger}$, $\mathbf{w}=\mathbf{w}^{\dagger}$ and $\mathbf{x}=\mathbf{x}^{\dagger}$, we know that Equation (72) holds if and only if

$$MR_{j}\left(\mathbf{p},\mathbf{s},\mathbf{X};\boldsymbol{\theta}_{c0}\right)=MR_{j^{\dagger}}\left(\mathbf{p}^{\dagger},\mathbf{s}^{\dagger},\mathbf{X}^{\dagger};\boldsymbol{\theta}_{c0}\right) \tag{73}$$

Then, we prove that Condition (1) holds for $\boldsymbol{\theta}_{c0}'\equiv(\chi_{0},\eta_{\alpha0})$, i.e., we essentially prove joint identification of $\chi_{0}=1$ and $\eta_{\alpha0}$. That is, we show that for any $\eta_{\alpha0}<0$ and $\eta_{\alpha}<0$ such that $\eta_{\alpha}\neq\eta_{\alpha0}$, for $\chi\neq1$, there exist $(\mathbf{p},\mathbf{s})$ and $\left(\mathbf{p}^{\dagger},\mathbf{s}^{\dagger}\right)$ such that for $\mathbf{s}=\chi_{0}\mathbf{s}$ and $\mathbf{s}^{\dagger}=\chi_{0}\mathbf{s}^{\dagger}$, $\widehat{\mathbf{s}}=\chi\mathbf{s}$ and $\widehat{\mathbf{s}}^{\dagger}=\chi\mathbf{s}^{\dagger}$,

$$\begin{aligned}A\left(\mathbf{p},\mathbf{s},\eta_{\alpha0},j\right) &= A\left(\mathbf{p}^{\dagger},\mathbf{s}^{\dagger},\eta_{\alpha0},j^{\dagger}\right)\\ &\quad \textit{and either } A\left(\mathbf{p},\widehat{\mathbf{s}},\eta_{\alpha},j\right)\neq A\left(\mathbf{p}^{\dagger},\widehat{\mathbf{s}}^{\dagger},\eta_{\alpha},j^{\dagger}\right)\\ &\quad \textit{or max}\left\{\sum_{l\leq j}\widehat{s}_{l},\sum_{l\leq j^{\dagger}}\widehat{s}_{l}^{\dagger}\right\}>1 \textit{ or both.}\end{aligned} \tag{74}$$

On the other hand, if $\chi=1$, then $\widehat{\mathbf{s}}^{\dagger}=\mathbf{s}^{\dagger}$, and since in this case, $\eta_{\alpha0}$ is identified, $A\left(\mathbf{p},\mathbf{s},\eta_{\alpha0},j\right)=A\left(\mathbf{p}^{\dagger},\mathbf{s}^{\dagger},\eta_{\alpha0},j^{\dagger}\right)$ for any $(\mathbf{p},\mathbf{s})$ if and only if $A\left(\mathbf{p},\widehat{\mathbf{s}},\eta_{\alpha0},j\right)=A\left(\mathbf{p}^{\dagger},\widehat{\mathbf{s}}^{\dagger},\eta_{\alpha0},j^{\dagger}\right)$.

Now, consider the case $\chi\neq1$. Let p_{j+1}, $p_{j^{\dagger}+1}$, satisfy

$$p_{j}-p_{j+1}\to0,\ p_{j^{\dagger}}^{\dagger}-p_{j^{\dagger}+1}^{\dagger}\to0.$$

Now, choose $\mathbf{s}$ and $\mathbf{s}^{\dagger}$such that $\sum_{l\leq j}s_{l}=\sum_{l\leq j^{\dagger}}s_{l}^{\dagger}=\Phi\left(-\eta_{\alpha0}\right)$. Then

$$\begin{aligned}&-\phi\left(\Phi^{-1}\left(\sum_{l\leq j}s_{l}\right)\right)\frac{\Phi^{-1}\left(\sum_{l\leq j}s_{l}\right)+\eta_{\alpha0}}{p_{j}-p_{j+1}}-\phi\left(\Phi^{-1}\left(\sum_{l\leq j-1}s_{l}\right)\right)\frac{\Phi^{-1}\left(\sum_{l\leq j-1}s_{l}\right)+\eta_{\alpha0}}{p_{j-1}-p_{j}}\\ =\;&-\phi\left(\Phi^{-1}\left(\sum_{l\leq j-1}s_{l}\right)\right)\frac{\Phi^{-1}\left(\sum_{l\leq j-1}s_{l}\right)+\eta_{\alpha0}}{p_{j-1}-p_{j}}<0.\end{aligned}$$

The negative sign holds because $\sum_{l\leq j^{\dagger}-1}s_{l}^{\dagger}<\Phi\left(-\eta_{\alpha0}\right)$ and $\eta_{\alpha0}<0$. Hence, $p_{j}>A\left(\mathbf{p},\mathbf{s},\eta_{\alpha0},j\right)$. Then, choose $s_{j^{\dagger}}^{\dagger}=s_{j}$ such that $\sum_{l\leq j-1}s_{l}=\sum_{l\leq j^{\dagger}-1}s_{l}^{\dagger}$, but $p_{j-1}^{\dagger}-p_{j}^{\dagger}\neq p_{j^{\dagger}-1}^{\dagger}-p_{j^{\dagger}}^{\dagger}$. Thus,

$A(\mathbf{p},\mathbf{s},\eta_{\alpha 0},j) = A\left(\mathbf{p}^{\dagger},\mathbf{s}^{\dagger},\eta_{\alpha 0},j^{\dagger}\right)$, implies $p_j \neq p_j^{\dagger}$.

Now, let, $\widehat{\mathbf{s}} = \chi\mathbf{s}$, $\widehat{\mathbf{s}}^{\dagger} = \chi\mathbf{s}^{\dagger}$. Then, if $max\left\{\sum_{l\leq j}\widehat{s}_l, \sum_{l\leq j^{\dagger}}\widehat{s}_l^{\dagger}\right\} < 1$, and $\Phi^{-1}\left(\sum_{l\leq j}\widehat{s}_l\right) + \eta_\alpha \neq 0$, then,

$$\left[-\phi\left(\Phi^{-1}\left(\sum_{l\leq j}\widehat{s}_l\right)\right)\frac{\Phi^{-1}\left(\sum_{l\leq j}\widehat{s}_l\right)+\eta_\alpha}{p_j-p_{j+1}} - \phi\left(\Phi^{-1}\left(\sum_{l\leq j-1}\widehat{s}_l\right)\right)\frac{\Phi^{-1}\left(\sum_{l\leq j-1}\widehat{s}_l\right)+\eta_\alpha}{p_{j-1}-p_j}\right]^{-1}$$
$$\times\widehat{s}_j \to 0.$$

as $p_j - p_{j+1} \to 0$, and the same holds for firm $j^{\dagger}$. Therefore, $A(\mathbf{p},\widehat{\mathbf{s}},\eta_\alpha,j) \to p_j$, and $A\left(\mathbf{p}^{\dagger},\widehat{\mathbf{s}}^{\dagger},\eta_\alpha,j^{\dagger}\right) \to p_{j^{\dagger}}^{\dagger}$. Then, since $p_j \neq p_j^{\dagger}$, for sufficiently small $|p_j - p_{j+1}|$ and $\left|p_{j^{\dagger}}^{\dagger} - p_{j^{\dagger}+1}^{\dagger}\right|$, $A(\mathbf{p},\widehat{\mathbf{s}},\eta_\alpha,j) \neq A\left(\mathbf{p}^{\dagger},\widehat{\mathbf{s}}^{\dagger},\eta_\alpha,j^{\dagger}\right)$. If $max\left\{\sum_{l\leq j^{\dagger}}\widehat{s}_l, \sum_{l\leq j^{\dagger}}\widehat{s}_l^{\dagger}\right\} \geq 1$, then either $\Phi^{-1}\left(\sum_{l\leq j}\widehat{s}_l\right)$ or $\Phi^{-1}\left(\sum_{l\leq j^{\dagger}}\widehat{s}_l^{\dagger}\right)$ or both are not defined.

Next, consider the case where for $\sum_{l\leq j}s_l = \sum_{l\leq j^{\dagger}}s_l^{\dagger} = \Phi(-\eta_{\alpha 0})$, $\sum_{l\leq j}\widehat{s}_l = \sum_{l\leq j}\widehat{s}_l^{\dagger} = \Phi(-\eta_\alpha)$ holds. Then choose $\mathbf{p}$ and $\mathbf{p}^{\dagger}$ such that $p_j - p_{j+1} = p_{j^{\dagger}}^{\dagger} - p_{j^{\dagger}+1}^{\dagger}$ and $p_{j-1} - p_j = p_{j^{\dagger}-1}^{\dagger} - p_{j^{\dagger}}^{\dagger}$. Then,

$$-\phi\left(\Phi^{-1}\left(\sum_{l\leq j^{\dagger}}s_l^{\dagger}\right)\right)\frac{\Phi^{-1}\left(\sum_{l\leq j^{\dagger}}s_l^{\dagger}\right)+\eta_{\alpha 0}}{p_{j^{\dagger}}^{\dagger}-p_{j^{\dagger}+1}^{\dagger}} - \phi\left(\Phi^{-1}\left(\sum_{l\leq j^{\dagger}-1}s_l^{\dagger}\right)\right)\frac{\Phi^{-1}\left(\sum_{l\leq j^{\dagger}-1}s_l^{\dagger}\right)+\eta_{\alpha 0}}{p_{j^{\dagger}-1}^{\dagger}-p_{j^{\dagger}}^{\dagger}}$$

$$= -\phi\left(\Phi^{-1}\left(\sum_{l\leq j^{\dagger}-1}s_l^{\dagger}\right)\right)\frac{\Phi^{-1}\left(\sum_{l\leq j^{\dagger}-1}s_l^{\dagger}\right)+\eta_{\alpha 0}}{p_{j^{\dagger}-1}^{\dagger}-p_{j^{\dagger}}^{\dagger}}.$$

Next, choose $p_j^{\dagger}$ such that

$$p_j^{\dagger} = p_j - \left[\phi\left(\Phi^{-1}\left(\sum_{l\leq j-1}s_l\right)\right)\frac{\Phi^{-1}\left(\sum_{l\leq j-1}s_l\right)+\eta_{\alpha 0}}{p_{j-1}-p_j}\right]^{-1}s_j$$

$$+\left[\phi\left(\Phi^{-1}\left(\sum_{l\leq j^{+}-1}s_l^{\dagger}\right)\right)\frac{\Phi^{-1}\left(\sum_{l\leq j^{\dagger}-1}s_l^{\dagger}\right)+\eta_{\alpha 0}}{p_{j^{\dagger}-1}^{\dagger}-p_{j^{\dagger}}^{\dagger}}\right]^{-1}s_{j^{\dagger}}^{\dagger}$$

Then, for $\chi < 1$, since Φ is normally distributed,

$$\left[\phi\left(\Phi^{-1}\left(\chi\sum_{l\leq j-1}s_l\right)\right)\frac{\Phi^{-1}\left(\chi\sum_{l\leq j-1}s_l\right)+\eta_\alpha}{p_{j-1}-p_j}\right]^{-1}\chi s_j$$

$$-\left[\phi\left(\Phi^{-1}\left(\sum_{l\leq j-1} s_l\right)\right)\frac{\Phi^{-1}\left(\sum_{l\leq j-1} s_l\right)+\eta_\alpha}{p_{j-1}-p_j}\right]^{-1} s_j$$

can be made arbitrarily large by making $\sum_{l\leq j-1} s_l$ to be sufficiently small. Then, $A(\mathbf{p},\widehat{\mathbf{s}},\eta_\alpha,j)\neq A\left(\mathbf{p}^\dagger,\widehat{\mathbf{s}}^\dagger,\eta_\alpha,j^\dagger\right)$. Similar arguments can be made for $\chi>1$. Therefore, we have proven the identification of $(\chi_0,\eta_{\alpha 0})$ using the limiting marginal revenue functions. As we argued for identification of $\eta_{\alpha 0}$, while the argument above uses infinite prices, we can show that it works approximately for sufficiently large prices.

The above results hold because of the continuous differentiability of the market share function. Given the assumptions, we can find firms in the population in different markets m and m'satisfying $(\mathbf{X}_m,\mathbf{p}_m,\mathbf{s}_m)=(\mathbf{X},\widetilde{\mathbf{p}},\mathbf{s})$, $(\mathbf{X}_{m'},\mathbf{p}_{m'},\mathbf{s}_{m'})=\left(\mathbf{X}^\dagger,\widetilde{\mathbf{p}}^\dagger,\mathbf{s}^\dagger\right)$. Therefore, $\chi_0=1$ is identified. Given $\chi_0=1$, the true market shares can be recovered, and thus, from the earlier analysis, $\eta_{\alpha 0}$ is identified.

We next show that $\sigma_{\alpha 0}$ is identified. Since we have already shown that $\eta_{\alpha 0}$ and $\chi_0=1$ are identified, we assume we can recover the true market shares $\mathbf{s}$ and $\mathbf{s}^\dagger$, and only focus on identifying $(\sigma_\alpha,\boldsymbol{\sigma}_{\boldsymbol{\beta}})$ of $\boldsymbol{\theta}_c=(\eta_{\alpha 0},\sigma_\alpha,\boldsymbol{\sigma}_{\boldsymbol{\beta}})$.

Below, we start by assuming that $\boldsymbol{\theta}_c$ is not identified. Then, from Condition 1, there exists $\boldsymbol{\theta}_c=(\eta_{\alpha 0},\sigma_\alpha,\boldsymbol{\sigma}_{\boldsymbol{\beta}})\neq\boldsymbol{\theta}_{c0}$ such that for any $(\mathbf{X},\mathbf{p},\mathbf{s},j)$ and $\left(\mathbf{X}^\dagger,\mathbf{p}^\dagger,\mathbf{s}^\dagger,j^\dagger\right)$, that satisfies $\mathbf{x}_j=\mathbf{x}^\dagger_{j^\dagger}$, both

$$MR(\mathbf{X},\mathbf{p},\mathbf{s},j,\boldsymbol{\theta}_{c0})=MR\left(\mathbf{X}^\dagger,\mathbf{p}^\dagger,\mathbf{s}^\dagger,j^\dagger,\boldsymbol{\theta}_{c0}\right)>0,\ MR(\mathbf{X},\mathbf{p},\mathbf{s},j,\boldsymbol{\theta}_c)=MR\left(\mathbf{X}^\dagger,\mathbf{p}^\dagger,\mathbf{s}^\dagger,j^\dagger,\boldsymbol{\theta}_c\right)>0. \tag{75}$$

hold. Then, we demonstrate a contradiction.

Let $\gamma\equiv\sigma_{\alpha 0}/\sigma_\alpha$ and

$$\boldsymbol{\Gamma}_{\boldsymbol{\beta}}\equiv\begin{bmatrix}\gamma_{\boldsymbol{\beta}1} & 0 & \dots & 0\\ 0 & \gamma_{\boldsymbol{\beta}2} & & 0\\ & & \ddots & \\ 0 & \dots & 0 & \gamma_{\boldsymbol{\beta}K}\end{bmatrix},$$

where $\gamma_{\boldsymbol{\beta}l}\equiv\sigma_{\beta 0l}/\sigma_{\beta l}, l=1,\dots,K$. Notice that $\gamma_{\boldsymbol{\beta}l}>0$ for all l and $\gamma>0$. We suppose that either $\boldsymbol{\Gamma}_{\boldsymbol{\beta}}\neq I$ or $\gamma\neq 1$ or both, and show that for some $(\mathbf{X},\mathbf{p},\mathbf{s},j)$ and $\left(\mathbf{X}^\dagger,\mathbf{p}^\dagger,\mathbf{s}^\dagger,j^\dagger\right)$, Equation (75) is not satisfied. We consider the following cases in the order given: 1) $\gamma<1$, $max_l\{\gamma_{\boldsymbol{\beta}l}\}\leq 1$; 2) $\gamma>1$, $min_l\{\gamma_{\boldsymbol{\beta}l}\}\geq 1$; 3)$\gamma<1$, $max_l\{\gamma_{\boldsymbol{\beta}l}\}>1$; 4) $\gamma>1$, $min_l\{\gamma_{\boldsymbol{\beta}l}\}<1$; 5) $\gamma=1$, $max_l\{\gamma_{\boldsymbol{\beta}l}\}>1$; and 6) $\gamma=1$, $min_l\{\gamma_{\boldsymbol{\beta}l}\}<1$.

Case 1) : $\gamma < 1$, $max_l \{\gamma_{\boldsymbol{\beta} l}\} \leq 1$, Define $\mathbf{X}^{(n)} \equiv \mathbf{X}\boldsymbol{\Gamma}_{\boldsymbol{\beta}}^{n}$, $\mathbf{p}^{(n)} \equiv \gamma^n \mathbf{p}$. Then, $lim_{n\to\infty}\mathbf{p}^{(n)} = \mathbf{0}$. Let $\widetilde{\mathbf{X}} \equiv lim_{n\to\infty}\mathbf{X}^{(n)}$. Then, for firms $j = 1, \ldots J$ and characteristics $k = 1, \ldots, K$, $\widetilde{x}_{j,k} = x_{j,k}$ if $\gamma_{\boldsymbol{\beta} k} = 1$ and $\widetilde{x}_{j,k} = 0$ if $\gamma_{\boldsymbol{\beta} k} < 1$. Then, using $\sigma_{\alpha 0} = \gamma\sigma_\alpha$, we obtain

$$\begin{aligned} s_j\left(\mathbf{X}, \mathbf{p}, \boldsymbol{\xi}_0, \boldsymbol{\theta}_{c0}\right) &= \int_\alpha \int_{\boldsymbol{\beta}} \frac{\exp\left(\mathbf{x}_j \boldsymbol{\Sigma}_{\boldsymbol{\beta}0}\boldsymbol{\beta} + p_j\sigma_{\alpha 0}\left(\alpha + \eta_{\alpha 0}\right) + \xi_{j,0}\right)}{\left[1 + \sum_{l=1}^{J}\exp\left(\mathbf{x}_l \boldsymbol{\Sigma}_{\boldsymbol{\beta}0}\boldsymbol{\beta} + p_l\sigma_{\alpha 0}\left(\alpha + \eta_{\alpha 0}\right) + \xi_{l,0}\right)\right]}\phi\left(\boldsymbol{\beta}\right) d\boldsymbol{\beta}\phi\left(\alpha\right) d\alpha \\ &= \int_\alpha \int_{\boldsymbol{\beta}} \frac{\exp\left(\mathbf{x}_j \boldsymbol{\Gamma}_{\boldsymbol{\beta}}\boldsymbol{\Sigma}_{\boldsymbol{\beta}}\boldsymbol{\beta} + \gamma p_j\sigma_{\alpha}\left(\alpha + \eta_{\alpha 0}\right) + \xi_{j,0}\right)}{\left[1 + \sum_{l=1}^{J}\exp\left(\mathbf{x}_l \boldsymbol{\Gamma}_{\boldsymbol{\beta}}\boldsymbol{\Sigma}_{\boldsymbol{\beta}}\boldsymbol{\beta} + \gamma p_l\sigma_{\alpha}\left(\alpha + \eta_{\alpha 0}\right) + \xi_{l,0}\right)\right]}\phi\left(\boldsymbol{\beta}\right) d\boldsymbol{\beta}\phi\left(\alpha\right) d\alpha \\ &= s_j\left(\mathbf{X}^{(1)}, \mathbf{p}^{(1)}, \boldsymbol{\xi}_0, \boldsymbol{\theta}_c\right) \end{aligned} \tag{76}$$

$$\begin{aligned} &\frac{\partial s_j}{\partial p_j}\left(\mathbf{X}, \mathbf{p}, \boldsymbol{\xi}_0, \boldsymbol{\theta}_{c0}\right) \\ = &\int_\alpha \int_{\beta} \frac{\exp\left(\mathbf{x}_j \boldsymbol{\Sigma}_{\boldsymbol{\beta}0}\boldsymbol{\beta} + p_j\sigma_{\alpha 0}\left(\alpha + \eta_{\alpha 0}\right) + \xi_{j,0}\right)\left[1 + \sum_{l\neq j}\exp\left(\mathbf{x}_l \boldsymbol{\Sigma}_{\boldsymbol{\beta}0}\boldsymbol{\beta} + p_l\sigma_{\alpha 0}\left(\alpha + \eta_{\alpha 0}\right) + \xi_{l,0}\right)\right]}{\left[1 + \sum_{l=1}^{J}\exp\left(\mathbf{x}_l \boldsymbol{\Sigma}_{\boldsymbol{\beta}0}\boldsymbol{\beta} + p_l\sigma_{\alpha 0}\left(\alpha + \eta_{\alpha 0}\right) + \xi_{l,0}\right)\right]^2} \\ &\times\sigma_{\alpha 0}\left(\alpha + \eta_{\alpha 0}\right)\phi\left(\boldsymbol{\beta}\right) d\boldsymbol{\beta}\phi\left(\alpha\right) d\alpha \\ = &\gamma\int_\alpha \int_{\beta} \frac{\exp\left(\mathbf{x}_j \boldsymbol{\Gamma}_{\boldsymbol{\beta}}\boldsymbol{\Sigma}_{\boldsymbol{\beta}}\boldsymbol{\beta} + \gamma p_j\sigma_{\alpha}\left(\alpha + \eta_{\alpha 0}\right) + \xi_{j,0}\right)}{\left[1 + \sum_{l=1}^{J}\exp\left(\mathbf{x}_l \boldsymbol{\Gamma}_{\boldsymbol{\beta}}\boldsymbol{\Sigma}_{\boldsymbol{\beta}}\boldsymbol{\beta} + \gamma p_l\sigma_{\alpha}\left(\alpha + \eta_{\alpha 0}\right) + \xi_{l,0}\right)\right]^2} \\ &\times\left[1 + \sum_{l\neq j}\exp\left(\mathbf{x}_l \boldsymbol{\Gamma}_{\boldsymbol{\beta}}\boldsymbol{\Sigma}_{\boldsymbol{\beta}}\boldsymbol{\beta} + \gamma p_l\sigma_{\alpha}\left(\alpha + \eta_{\alpha}\right) + \xi_{l,0}\right)\right]\sigma_\alpha\left(\alpha + \eta_{\alpha 0}\right)\phi\left(\boldsymbol{\beta}\right) d\boldsymbol{\beta}\phi\left(\alpha\right) d\alpha \\ = &\gamma\frac{\partial s_j}{\partial p_j}\left(\mathbf{X}^{(1)}, \mathbf{p}^{(1)}, \boldsymbol{\xi}_0, \boldsymbol{\theta}_c\right) \end{aligned} \tag{77}$$

and similarly for $s^\dagger_{j^\dagger}$, $\frac{\partial s^\dagger_{j^\dagger}}{\partial p_{j^\dagger}}$. Thus,

$$s_j = s_j\left(\mathbf{X}^{(1)}, \mathbf{p}^{(1)}, \boldsymbol{\xi}_0, \boldsymbol{\theta}_c\right),\ s^\dagger_{j^\dagger} = s_{j^\dagger}\left(\mathbf{X}^{\dagger(1)}, \mathbf{p}^{\dagger(1)}, \boldsymbol{\xi}^\dagger_0, \boldsymbol{\theta}_c\right),\ s_j \neq s^\dagger_{j^\dagger}$$

And, if marginal revenues are equal for firms with $\left(\mathbf{X}, \mathbf{p}, \mathbf{s}, \boldsymbol{\theta}_{c0}\right)$ and $\left(\mathbf{X}, \mathbf{p}^\dagger, \mathbf{s}^\dagger, \boldsymbol{\theta}_{c0}\right)$, then they are also equal for $\left(\mathbf{X}^{(1)}, \mathbf{p}^{(1)}, \mathbf{s}, \boldsymbol{\theta}_c\right)$ and $\left(\mathbf{X}^{\dagger(1)}, \mathbf{p}^{\dagger(1)}, \mathbf{s}^{\dagger}, \boldsymbol{\theta}_c\right)$, because,

$$\begin{aligned}
& p_j + \left[\frac{\partial s_j}{\partial p_j}\right]^{-1} \left(\mathbf{X}, \mathbf{p}, \boldsymbol{\xi}_0, \boldsymbol{\theta}_{c0}\right) s_j \left(\mathbf{X}, \mathbf{p}, \boldsymbol{\xi}_0, \boldsymbol{\theta}_{c0}\right) \\
= \;& \gamma^{-1} p_j^{(1)} + \gamma^{-1} \left[\frac{\partial s_j}{\partial p_j}\right]^{-1} \left(\mathbf{X}^{(1)}, \mathbf{p}^{(1)}, \boldsymbol{\xi}_0, \boldsymbol{\theta}_c\right) s_j \left(\mathbf{X}^{(1)}, \mathbf{p}^{(1)}, \boldsymbol{\xi}_0, \boldsymbol{\theta}_c\right) \\
= \;& p_{j^\dagger}^{\dagger} + \left[\frac{\partial s_{j^\dagger}}{\partial p_{j^\dagger}}\right]^{-1} \left(\mathbf{X}^{\dagger}, \mathbf{p}^{\dagger}, \boldsymbol{\xi}_0^{\dagger}, \boldsymbol{\theta}_{c0}\right) s_{j^\dagger} \left(\mathbf{x}, \mathbf{p}^{\dagger}, \boldsymbol{\xi}_0^{\dagger}, \boldsymbol{\theta}_{c0}\right) \\
= \;& \gamma^{-1} p_{j^\dagger}^{\dagger(1)} + \gamma^{-1} \left[\frac{\partial s_{j^\dagger}}{\partial p_{j^\dagger}}\right]^{-1} \left(\mathbf{X}^{\dagger(1)}, \mathbf{p}^{\dagger(1)}, \boldsymbol{\xi}_0^{\dagger}, \boldsymbol{\theta}_c\right) s_j \left(\mathbf{X}^{\dagger(1)}, \mathbf{p}^{\dagger(1)}, \boldsymbol{\xi}_0^{\dagger}, \boldsymbol{\theta}_c\right) \qquad (78)
\end{aligned}$$

If $\boldsymbol{\theta}_c$ is not identified, then Equation (78) implies that there exist $\boldsymbol{\xi}_0^{(1)}$ and $\boldsymbol{\xi}_0^{\dagger(1)}$ that satisfy

$$s_j = s_j \left(\mathbf{X}^{(1)}, \mathbf{p}^{(1)}, \boldsymbol{\xi}_0^{(1)}, \boldsymbol{\theta}_{c0}\right), \; s_{j^\dagger}^{\dagger} = s_{j^\dagger} \left(\mathbf{X}^{\dagger(1)}, \mathbf{p}^{\dagger(1)}, \boldsymbol{\xi}_0^{\dagger(1)}, \boldsymbol{\theta}_{c0}\right),$$

$$\begin{aligned}
& p_j^{(1)} + \gamma^{-1} \left[\frac{\partial s_j \left(\mathbf{X}^{(1)}, \mathbf{p}^{(1)}, \boldsymbol{\xi}_0^{(1)}, \boldsymbol{\theta}_{c0}\right)}{\partial p_j^{(1)}}\right]^{-1} s_j \left(\mathbf{X}^{(1)}, \mathbf{p}^{(1)}, \boldsymbol{\xi}_0^{(1)}, \boldsymbol{\theta}_{c0}\right) \\
= \;& p_{j^\dagger}^{\dagger(1)} + \gamma^{-1} \left[\frac{\partial s_{j^\dagger} \left(\mathbf{X}^{\dagger(1)}, \mathbf{p}^{\dagger(1)}, \boldsymbol{\xi}_0^{\dagger(1)}, \boldsymbol{\theta}_{c0}\right)}{\partial p_{j^\dagger}^{(1)}}\right]^{-1} s_{j^\dagger}^{\dagger} \left(\mathbf{X}^{\dagger(1)}, \mathbf{p}^{\dagger(1)}, \boldsymbol{\xi}_0^{\dagger(1)}, \boldsymbol{\theta}_{c0}\right) \geq 0
\end{aligned}$$

That is, if $\boldsymbol{\theta}_c$ is not identified, then there exist unobserved product characteristics $\boldsymbol{\xi}_0^{(1)}$ and $\boldsymbol{\xi}_0^{(1)}$ such that market shares remain the same and marginal revenues are equal at the true parameter vector.

Using the same logic, we define $\boldsymbol{\xi}_0^{(n)}$ to satisfy $s_j = s_j \left(\mathbf{X}^{(n)}, \mathbf{p}^{(n)}, \boldsymbol{\xi}_0^{(n)}, \boldsymbol{\theta}_{c0}\right)$ and $\widetilde{\boldsymbol{\xi}}_0 = lim_{n \to \infty} \boldsymbol{\xi}_0^{(n)}$ which satisfies $\mathbf{s} = \mathbf{s}\left(\widetilde{\mathbf{X}}, \mathbf{0}, \widetilde{\boldsymbol{\xi}}_0, \boldsymbol{\theta}_{c0}\right)$ because of the Implicit Function Theorem. Then, as $n \to \infty$,

$$
\begin{aligned}
& \frac{\partial s_j\left(\mathbf{X}^{(n)},\mathbf{p}^{(n)},\boldsymbol{\xi}_0^{(n)},\boldsymbol{\theta}_{c0}\right)}{\partial p_j^{(n)}} \\
= \quad & \int_\alpha\int_\beta \frac{\exp\left(\mathbf{x}_j^{(n)}\boldsymbol{\Sigma}_{\boldsymbol{\beta}0}\boldsymbol{\beta}+p_j^{(n)}\sigma_{\alpha0}\left(\alpha+\eta_{\alpha0}\right)+\xi_{j,0}^{(n)}\right)}{\left[1+\sum_{l=1}^J\exp\left(\mathbf{x}_l^{(n)}\boldsymbol{\Sigma}_{\boldsymbol{\beta}0}\boldsymbol{\beta}+p_l^{(n)}\sigma_{\alpha0}\left(\alpha+\eta_{\alpha0}\right)+\xi_{l,0}^{(n)}\right)\right]^2} \\
& \times\left[1+\sum_{l\neq j}\exp\left(\mathbf{x}_l^{(n)}\boldsymbol{\Sigma}_{\boldsymbol{\beta}0}\boldsymbol{\beta}+p_l^{(n)}\sigma_{\alpha0}\left(\alpha+\eta_{\alpha0}\right)+\xi_{l,0}^{(n)}\right)\right] \\
& \times\sigma_{\alpha0}\left(\alpha+\eta_{\alpha0}\right)\phi\left(\alpha\right)\phi\left(\boldsymbol{\beta}\right)d\boldsymbol{\beta}d\alpha \\
\to \quad & \int_\alpha\int_\beta \frac{\exp\left(\widetilde{\mathbf{x}}_j\boldsymbol{\Sigma}_{\boldsymbol{\beta}0}\boldsymbol{\beta}+\widetilde{\xi}_{j,0}\right)\left[1+\sum_{l\neq j}\exp\left(\widetilde{\mathbf{x}}_l\boldsymbol{\Sigma}_{\boldsymbol{\beta}0}\boldsymbol{\beta}+\widetilde{\xi}_{l,0}\right)\right]}{\left[1+\sum_{l=1}^J\exp\left(\widetilde{\mathbf{x}}_l\boldsymbol{\Sigma}_{\boldsymbol{\beta}0}\boldsymbol{\beta}+\widetilde{\xi}_{l,0}\right)\right]^2} \\
& \times\sigma_{\alpha0}\left(\alpha+\eta_{\alpha0}\right)\phi\left(\alpha\right)\phi\left(\boldsymbol{\beta}\right)d\boldsymbol{\beta}d\alpha \\
= \sigma_{\alpha0}\eta_{\alpha0} \quad & \int_\beta \frac{\exp\left(\widetilde{\mathbf{x}}_j\boldsymbol{\Sigma}_{\boldsymbol{\beta}0}\boldsymbol{\beta}+\widetilde{\xi}_{j,0}\right)\left[1+\sum_{l\neq j}\exp\left(\widetilde{\mathbf{x}}_l\boldsymbol{\Sigma}_{\boldsymbol{\beta}0}\boldsymbol{\beta}+\widetilde{\xi}_{l,0}\right)\right]}{\left[1+\sum_{l=1}^J\exp\left(\widetilde{\mathbf{x}}_l\boldsymbol{\Sigma}_{\boldsymbol{\beta}0}\boldsymbol{\beta}+\widetilde{\xi}_{l,0}\right)\right]^2}\phi\left(\boldsymbol{\beta}\right)d\boldsymbol{\beta}<0
\end{aligned}
$$

because $\eta_{\alpha0}<0$. Therefore,

$$
p_j^{(n)}+\gamma^{-n}\left[\frac{\partial s_j\left(\mathbf{X}^{(n)},\mathbf{p}^{(n)},\boldsymbol{\xi}_0^{(n)},\boldsymbol{\theta}_{c0}\right)}{\partial p_j^{(n)}}\right]^{-1}s_j<0
$$

for sufficiently large n, which contradicts positivity of the marginal revenue.

Case 2) $\gamma>1$, $min_l\left\{\gamma_{\boldsymbol{\beta}l}\right\}\geq1$ Let $\widetilde{\gamma}\equiv1/\gamma$, $\widetilde{\gamma}_{\boldsymbol{\beta}l}\equiv1/\gamma_{\boldsymbol{\beta}l}$. Then $\sigma_\alpha=\widetilde{\gamma}\sigma_{\alpha0}$. Similarly, define $\mathbf{X}^{(n)}\equiv\mathbf{X}\widetilde{\boldsymbol{\Gamma}}_{\boldsymbol{\beta}}^n$, $\mathbf{p}^{(n)}\equiv\widetilde{\gamma}^n\mathbf{p}$. Then, $lim_{n\to\infty}\mathbf{p}^{(n)}=\mathbf{0}$. Let $\widetilde{\mathbf{X}}\equiv lim_{n\to\infty}\mathbf{X}^{(n)}$. Then, $\widetilde{x}_{j,l}=x_{j,l}$ if $\gamma_{\boldsymbol{\beta}l}=1$ and $\widetilde{x}_{j,l}=0$ if $\gamma_{\boldsymbol{\beta}l}>1$. Then, using the same steps as in Case 1), but starting with the parameter vector $\boldsymbol{\theta}_c$, we find that $s_j\left(\mathbf{X},\mathbf{p},\boldsymbol{\xi},\boldsymbol{\theta}_c\right)=s_j\left(\mathbf{X}^{(1)},\mathbf{p}^{(1)},\boldsymbol{\xi},\boldsymbol{\theta}_{c0}\right)$; To see this, note that:

$$
\begin{aligned}
s_j\left(\mathbf{X},\mathbf{p},\boldsymbol{\xi},\boldsymbol{\theta}_c\right) &= \int_\alpha\int_{\boldsymbol{\beta}}\frac{\exp\left(\mathbf{x}_j\boldsymbol{\Sigma}_{\boldsymbol{\beta}}\boldsymbol{\beta}+p_j\sigma_\alpha\left(\alpha+\eta_{\alpha0}\right)+\xi_j\right)}{\left[1+\sum_{l=1}^J\exp\left(\mathbf{x}_l\boldsymbol{\Sigma}_{\boldsymbol{\beta}}\boldsymbol{\beta}+p_l\sigma_\alpha\left(\alpha+\eta_{\alpha0}\right)+\xi_l\right)\right]}\phi\left(\boldsymbol{\beta}\right)d\boldsymbol{\beta}\phi\left(\alpha\right)d\alpha \\
&= \int_\alpha\int_{\boldsymbol{\beta}}\frac{\exp\left(\mathbf{x}_j\widetilde{\boldsymbol{\Gamma}}_{\boldsymbol{\beta}}\boldsymbol{\Sigma}_{\boldsymbol{\beta}0}\boldsymbol{\beta}+\widetilde{\gamma}p_j\sigma_{\alpha0}\left(\alpha+\eta_{\alpha0}\right)+\xi_j\right)}{\left[1+\sum_{l=1}^J\exp\left(\mathbf{x}_l\widetilde{\boldsymbol{\Gamma}}_{\boldsymbol{\beta}}\boldsymbol{\Sigma}_{\boldsymbol{\beta}0}\boldsymbol{\beta}+\widetilde{\gamma}p_l\sigma_{\alpha0}\left(\alpha+\eta_{\alpha0}\right)+\xi_l\right)\right]}\phi\left(\boldsymbol{\beta}\right)d\boldsymbol{\beta}\phi\left(\alpha\right)d\alpha \\
&= s_j\left(\mathbf{X}^{(1)},\mathbf{p}^{(1)},\boldsymbol{\xi},\boldsymbol{\theta}_{c0}\right) \qquad (79)
\end{aligned}
$$

It then follows that $\frac{\partial s_j}{\partial p_j}\left(\mathbf{X},\mathbf{p},\boldsymbol{\xi},\boldsymbol{\theta}_c\right)=\widetilde{\gamma}\frac{\partial s_j}{\partial p_j}\left(\mathbf{X}^{(1)},\mathbf{p}^{(1)},\boldsymbol{\xi},\boldsymbol{\theta}_{c0}\right)$and similarly for $s_{j^\dagger}^\dagger$, $\frac{\partial s_{j^\dagger}^\dagger}{\partial p_{j^\dagger}}$.Then by applying the non-identification condition as in the case above, and then iterating, we conclude that

$$p_j^{(n)} + \widetilde{\gamma}^{-n} \left[\frac{\partial s_j \left(\mathbf{X}^{(n)}, \mathbf{p}^{(n)}, \boldsymbol{\xi}^{(n)}, \boldsymbol{\theta}_c \right)}{\partial p_j^{(n)}} \right]^{-1} s_j < 0,$$

which is a contradiction to positivity of marginal revenue. Thus $\boldsymbol{\theta}_{c0}$ is identified.

3) $\gamma < 1$, $max_l \{\gamma_{\boldsymbol{\beta} l}\} > 1$. Assume for some $k \in \{1, \ldots, K\}$, $\gamma_{\boldsymbol{\beta} k} > \gamma_{\boldsymbol{\beta} l}$ for any $l \neq k$.

Then, as $n \to \infty$, $x_{i,l}^{(n)}/\gamma_{\boldsymbol{\beta} k}^n \to 0$ for $l \neq k$ and $x_{i,k}^{(n)}/\gamma_{\boldsymbol{\beta} k}^n = x_{ik}$ for all $i = 1, \ldots, J$, $\mathbf{p}^{(n)}/\gamma_{\boldsymbol{\beta} k}^n \to 0$. Now, let $u_{i,0}^{(0)} = \boldsymbol{\xi}_{i,0}/\sigma_{\beta 0k}$, and $u_{i,0}^{(n)}$ be such that $s_i = s_i \left(\mathbf{X}^{(n)}, \mathbf{p}^{(n)}, \gamma_{\boldsymbol{\beta} k}^n \sigma_{\boldsymbol{\beta} 0k} \boldsymbol{u}_0^{(n)}, \boldsymbol{\theta}_{c0} \right)$. For the sake of simplicity, we assume that the k^{th} observed characteristics of firms in a market are different from each other. Now, wlog. order the firms such that $x_{1,k} < x_{2,k} \ldots < x_{J,k}$ and $x_{1,k}^\dagger < x_{2,k}^\dagger \ldots < x_{J^\dagger,k}^\dagger$. Then, denoting $\boldsymbol{\Sigma}_{\boldsymbol{\beta} 0,-k,-k}$ to be the submatrix of $\boldsymbol{\Sigma}_{\boldsymbol{\beta} 0}$ that does not include the kth row and the kth column, we get,

$$\begin{aligned}
& \int_{\beta_k} \prod_{l \neq j}^{J} I \left(\left(x_{l,k}^{(n)} - x_{j,k}^{(n)} \right) \sigma_{\boldsymbol{\beta} 0k} \beta_k + \left(\mathbf{x}_{l,-k}^{(n)} - \mathbf{x}_{j,-k}^{(n)} \right) \boldsymbol{\Sigma}_{\boldsymbol{\beta} 0-k,-k} \boldsymbol{\beta}_{-k} \right. \\
& \left. + \left(p_l^{(n)} - p_j^{(n)} \right) \sigma_{\alpha 0} \left(\alpha + \eta_{\alpha 0} \right) + \gamma_{\boldsymbol{\beta} k}^n \sigma_{\boldsymbol{\beta} 0k} \left(u_{l,0}^{(n)} - u_{j,0}^{(n)} \right) + \epsilon_l - \epsilon_j \leq 0 \right) \phi\left(\beta_k\right) d\beta_k \\
= & \int_{\beta_k} \prod_{l \neq j}^{J} I \left(\left(x_{l,k} - x_{j,k} \right) \sigma_{\boldsymbol{\beta} 0k} \beta_k + \frac{\mathbf{x}_{l,-k}^{(n)} - \mathbf{x}_{j,-k}^{(n)}}{\gamma_{\boldsymbol{\beta} k}^n} \boldsymbol{\Sigma}_{\boldsymbol{\beta} 0-k,-k} \boldsymbol{\beta}_{-k} \right. \\
& \left. + \left(\frac{p_l^{(n)} - p_j^{(n)}}{\gamma_{\boldsymbol{\beta} k}^n} \right) \sigma_{\alpha 0} \left(\alpha + \eta_{\alpha 0} \right) + \sigma_{\boldsymbol{\beta} 0k} (u_{l,0}^{(n)} - u_{j,0}^{(n)}) + \frac{\epsilon_l - \epsilon_j}{\gamma_{\boldsymbol{\beta} k}^n} \leq 0 \right) \phi\left(\beta_k\right) d\beta_k \\
= & \int_{\beta_k} \prod_{x_{l,k} - x_{j,k} > 0}^{J} I \left(\beta_k \leq - \frac{\mathbf{x}_{l,-k}^{(n)} - \mathbf{x}_{j,-k}^{(n)}}{\gamma_{\boldsymbol{\beta} k}^n \sigma_{\boldsymbol{\beta} 0k} \left(x_{l,k} - x_{j,k} \right)} \boldsymbol{\Sigma}_{\boldsymbol{\beta} 0-k,-k} \boldsymbol{\beta}_{-k} \right. \\
& \left. - \left(\frac{p_l^{(n)} - p_j^{(n)}}{\gamma_{\boldsymbol{\beta} k}^n \sigma_{\boldsymbol{\beta} 0k} \left(x_{l,k} - x_{j,k} \right)} \right) \sigma_{\alpha 0} \left(\alpha + \eta_{\alpha 0} \right) - \frac{u_{l,0}^{(n)} - u_{j,0}^{(n)}}{\left(x_{l,k} - x_{j,k} \right)} - \frac{\epsilon_l - \epsilon_j}{\gamma_{\boldsymbol{\beta} k}^n \sigma_{\boldsymbol{\beta} 0k} \left(x_{l,k} - x_{j,k} \right)} \right\} \Bigg) \\
& \prod_{x_{l,k} - x_{j,k} < 0}^{J} I \left(\beta_k \geq - \frac{\mathbf{x}_{l,-k}^{(n)} - \mathbf{x}_{j,-k}^{(n)}}{\gamma_{\boldsymbol{\beta} k}^n \sigma_{\boldsymbol{\beta} 0k} \left(x_{l,k} - x_{j,k} \right)} \boldsymbol{\Sigma}_{\boldsymbol{\beta} 0-k,-k} \boldsymbol{\beta}_{-k} \right. \\
& \left. - \left(\frac{p_l^{(n)} - p_j^{(n)}}{\gamma_{\boldsymbol{\beta} k}^n \sigma_{\boldsymbol{\beta} 0k} \left(x_{l,k} - x_{j,k} \right)} \right) \sigma_{\alpha 0} \left(\alpha + \eta_{\alpha 0} \right) - \frac{u_{l,0}^{(n)} - u_{j,0}^{(n)}}{\left(x_{l,k} - x_{j,k} \right)} - \frac{\epsilon_l - \epsilon_j}{\gamma_{\boldsymbol{\beta} k}^n \sigma_{\boldsymbol{\beta} 0k} \left(x_{l,k} - x_{j,k} \right)} \right\} \Bigg)
\end{aligned}$$

$$= \int_{\beta_k} I\left(\beta_k \leq Min_{x_{l,k}-x_{j,k}>0}\left\{-\frac{\mathbf{x}_{l,-k}^{(n)}-\mathbf{x}_{j,-k}^{(n)}}{\gamma_{\boldsymbol{\beta}k}^n \sigma_{\boldsymbol{\beta}0k}\left(x_{l,k}-x_{j,k}\right)}\boldsymbol{\Sigma}_{\boldsymbol{\beta}0-k,-k}\boldsymbol{\beta}_{-k}\right.\right.$$
$$\left.\left.-\left(\frac{p_l^{(n)}-p_j^{(n)}}{\gamma_{\boldsymbol{\beta}k}^n \sigma_{\boldsymbol{\beta}0k}\left(x_{l,k}-x_{j,k}\right)}\right)\sigma_{\alpha 0}\left(\alpha+\eta_{\alpha 0}\right)-\frac{u_{l,0}^{(n)}-u_{j,0}^{(n)}}{\left(x_{l,k}-x_{j,k}\right)}-\frac{\epsilon_l-\epsilon_j}{\gamma_{\boldsymbol{\beta}k}^n \sigma_{\boldsymbol{\beta}0k}\left(x_{l,k}-x_{j,k}\right)}\right\}\right)$$
$$\times I\left(\beta_k \geq Max_{x_{l,k}-x_{j,k}<0}\left\{-\frac{\mathbf{x}_{l,-k}^{(n)}-\mathbf{x}_{j,-k}^{(n)}}{\gamma_{\boldsymbol{\beta}k}^n \sigma_{\boldsymbol{\beta}0k}\left(x_{l,k}-x_{j,k}\right)}\boldsymbol{\Sigma}_{\boldsymbol{\beta}0-k}\boldsymbol{\beta}_{-k}\right.\right.$$
$$\left.\left.-\left(\frac{p_l^{(n)}-p_j^{(n)}}{\gamma_{\boldsymbol{\beta}k}^n \sigma_{\boldsymbol{\beta}0k}\left(x_{l,k}-x_{j,k}\right)}\right)\sigma_{\alpha 0}\left(\alpha+\eta_{\alpha 0}\right)-\frac{u_{l,0}^{(n)}-u_{j,0}^{(n)}}{\left(x_{l,k}-x_{j,k}\right)}-\frac{\epsilon_l-\epsilon_j}{\gamma_{\boldsymbol{\beta}k}^n \sigma_{\boldsymbol{\beta}0k}\left(x_{l,k}-x_{j,k}\right)}\right\}\right)\phi\left(\beta_k\right)d\beta_k$$

Note that $\frac{\mathbf{x}_{l,-k}^{(n)}-\mathbf{x}_{j,-k}^{(n)}}{\gamma_{\boldsymbol{\beta}k}^n \sigma_{\boldsymbol{\beta}0k}\left(x_{l,k}-x_{j,k}\right)}\boldsymbol{\Sigma}_{\boldsymbol{\beta}0-k,-k}\boldsymbol{\beta}_{-k} \to 0$, $\left(\frac{p_l^{(n)}-p_j^{(n)}}{\gamma_{\boldsymbol{\beta}k}^n \sigma_{\boldsymbol{\beta}0k}\left(x_{l,k}-x_{j,k}\right)}\right)\sigma_{\alpha 0}\left(\alpha+\eta_{\alpha 0}\right) \to 0$ and $\frac{\epsilon_l-\epsilon_j}{\gamma_{\boldsymbol{\beta}k}^n \sigma_{\boldsymbol{\beta}0k}\left(x_{l,k}-x_{j,k}\right)} \to 0$ as $n \to \infty$. Hence, using earlier arguments, from continuity, there exists $\widetilde{\mathbf{u}}_0$ that satisfies

$$s_0=\Phi\left(-\frac{\widetilde{u}_{1,0}}{x_{1,k}}\right)\ , s_i=\Phi\left(-\frac{\widetilde{u}_{i,0}-\widetilde{u}_{i+1,0}}{\left(x_{i,k}-x_{i+1,k}\right)}\right)-\Phi\left(-\frac{\widetilde{u}_{i,0}-\widetilde{u}_{i-1,0}}{\left(x_{i,k}-x_{i-1,k}\right)}\right), i=1,\ldots,J$$

It then follows that,

$$-\frac{\widetilde{u}_{i,0}-\widetilde{u}_{i+1,0}}{\left(x_{i,k}-x_{i+1,k}\right)}=\Phi^{-1}\left(\sum_{l\leq i} s_l\right)$$

Thus, we can derive $\widetilde{u}_{i,0}$ recursively, as earlier. Furthermore, following the similar arguments as before, we can show that $lim_{n\to\infty} u_{i,0}^{(n)}=\widetilde{u}_{i,0}$ for $i=1,\ldots,J$.

Next, we consider the derivative of the market share with respect to the price. Note that similarly as before, we define $a_{i,l} \equiv \frac{\widetilde{u}_{i,0}-\widetilde{u}_{l,0}}{\left(x_{i,k}-x_{l,k}\right)}$ and $\overline{a}_i$ and $\underline{a}_i$ similarly as well. Then,

$$\ldots<\underline{a}_{i+1}=\overline{a}_i=\frac{\widetilde{u}_{i,0}-\widetilde{u}_{i+1,0}}{\left(x_{i,k}-x_{i+1,k}\right)}<\underline{a}_i=\overline{a}_{i-1}=\frac{\widetilde{u}_{i,0}-\widetilde{u}_{i-1,0}}{\left(x_{i,k}-x_{i-1,k}\right)}<\ldots<\frac{\widetilde{u}_{1,0}}{x_{1,k}}$$

, and as before, we derive

$$\frac{\partial}{\partial p_j^{(n)}}\int_{\beta_k}\prod_{,l\neq j}^{J} I\left(\left(x_{l,k}^{(n)}-x_{j,k}^{(n)}\right)\sigma_{\boldsymbol{\beta}0k}\beta_k+\left(\mathbf{x}_{l,-k}^{(n)}-\mathbf{x}_{j,-k}^{(n)}\right)\boldsymbol{\Sigma}_{\boldsymbol{\beta}0-k,-k}\boldsymbol{\beta}_{-k}\right.$$
$$\left.+\left(p_l^{(n)}-p_j^{(n)}\right)\sigma_{\alpha 0}\left(\alpha+\eta_{\alpha 0}\right)+\gamma_{\boldsymbol{\beta}k}^n\sigma_{\boldsymbol{\beta}0k}\left(u_{l,0}^{(n)}-u_{j,0}^{(n)}\right)+\epsilon_l-\epsilon_j\leq 0\right)\phi\left(\beta_k\right)d\beta_k$$
$$=\frac{\partial}{\partial p_j^{(n)}}\int_{\beta_k}\prod_{l\neq j}^{J} I\left(\left(x_{l,k}-x_{j,k}\right)\sigma_{\boldsymbol{\beta}0k}\beta_k+\frac{\mathbf{x}_{l,-k}^{(n)}-\mathbf{x}_{j,-k}^{(n)}}{\gamma_{\boldsymbol{\beta}k}^n}\boldsymbol{\Sigma}_{\boldsymbol{\beta}0-k,-k}\boldsymbol{\beta}_{-k}\right.$$
$$\left.+\left(\frac{p_l^{(n)}-p_j^{(n)}}{\gamma_{\boldsymbol{\beta}k}^n}\right)\sigma_{\alpha 0}\left(\alpha+\eta_{\alpha 0}\right)+\sigma_{\boldsymbol{\beta}0k}(u_{l,0}^{(n)}-u_{j,0}^{(n)})+\frac{\epsilon_l-\epsilon_j}{\gamma_{\boldsymbol{\beta}k}^n}\leq 0\right)\phi\left(\beta_k\right)d\beta_k \qquad (80)$$

Now, let

$$\begin{aligned} & B\left(\mathbf{p}^{(n)}, \mathbf{u}_0^{(n)}, n, \mathbf{X}, \Gamma_{\boldsymbol{\beta}}, \boldsymbol{\beta}, \boldsymbol{\epsilon}, j, \boldsymbol{\theta}_{c0}\right) \\ \equiv\ & -\frac{\mathbf{x}_{l,-k}^{(n)} - \mathbf{x}_{j,-k}^{(n)}}{\gamma_{\boldsymbol{\beta}k}^{n} \sigma_{\boldsymbol{\beta}0k} \left(x_{l,k} - x_{j,k}\right)} \mathbf{\Sigma}_{\boldsymbol{\beta}0-k,-k} \boldsymbol{\beta}_{-k} \\ & -\left(\frac{p_l^{(n)} - p_j^{(n)}}{\gamma_{\boldsymbol{\beta}k}^{n} \sigma_{\boldsymbol{\beta}0k} \left(x_{l,k} - x_{j,k}\right)}\right) \sigma_{\alpha 0} \left(\alpha + \eta_{\alpha 0}\right) - \frac{u_{l,0}^{(n)} - u_{j,0}^{(n)}}{\left(x_{l,k} - x_{j,k}\right)} - \frac{\epsilon_l - \epsilon_j}{\gamma_{\boldsymbol{\beta}k}^{n} \sigma_{\boldsymbol{\beta}0k} \left(x_{l,k} - x_{j,k}\right)} \end{aligned}$$

Then let

$$\begin{aligned} & \underline{B}\left(\mathbf{p}^{(n)}, \mathbf{u}_0^{(n)}, n, \mathbf{X}, \Gamma_{\boldsymbol{\beta}}, \boldsymbol{\beta}, \boldsymbol{\epsilon}, j, \boldsymbol{\theta}_{c0}\right) \\ \equiv\ & Min_{x_{l,k} - x_{j,k} > 0} B\left(\mathbf{p}^{(n)}, \mathbf{u}_0^{(n)}, n, \mathbf{X}, \Gamma_{\boldsymbol{\beta}}, \boldsymbol{\beta}, \boldsymbol{\epsilon}, j, \boldsymbol{\theta}_{c0}\right) \end{aligned}$$

and

$$\begin{aligned} & \overline{B}\left(\mathbf{p}^{(n)}, \mathbf{u}_0^{(n)}, n, \mathbf{X}, \Gamma_{\boldsymbol{\beta}}, \boldsymbol{\beta}, \boldsymbol{\epsilon}, j, \boldsymbol{\theta}_{c0}\right) \\ \equiv\ & Max_{x_{lk} - x_{jk} < 0} B\left(\mathbf{p}^{(n)}, \mathbf{u}_0^{(n)}, n, \mathbf{X}, \Gamma_{\boldsymbol{\beta}}, \boldsymbol{\beta}, \boldsymbol{\epsilon}, j, \boldsymbol{\theta}_{c0}\right) \end{aligned}$$

Then,

$$\begin{aligned} \frac{\partial \underline{B}\left(\mathbf{p}^{(n)}, \mathbf{u}_0^{(n)}, , n, \mathbf{X}, \Gamma_{\boldsymbol{\beta}}, \boldsymbol{\beta}, \boldsymbol{\epsilon}, j, \boldsymbol{\theta}_{c0}\right)}{\partial p_j^{(n)}} &= \frac{\sigma_{\alpha 0} \left(\alpha + \eta_{\alpha 0}\right)}{\gamma_{\boldsymbol{\beta}k}^{n} \sigma_{\boldsymbol{\beta}0k} \left(x_{l_1,k} - x_{j,k}\right)}, \\ \frac{\partial \overline{B}\left(\mathbf{p}^{(n)}, \mathbf{u}_0^{(n)}, n, \mathbf{X}, \Gamma_{\boldsymbol{\beta}}, \boldsymbol{\beta}, \boldsymbol{\epsilon}, j, \boldsymbol{\theta}_{c0}\right)}{\partial p_j^{(n)}} &= \frac{\sigma_{\alpha 0} \left(\alpha + \eta_{\alpha 0}\right)}{\gamma_{\boldsymbol{\beta}k}^{n} \sigma_{\boldsymbol{\beta}0k} \left(x_{l_2,k} - x_{j,k}\right)} \end{aligned}$$

where

$$\begin{aligned} l_1 &= argmin_{l \neq j, x_{l,k} - x_{j,k} > 0} B\left(\mathbf{p}^{(n)}, \mathbf{u}_0^{(n)}, n, \mathbf{X}, \Gamma_{\boldsymbol{\beta}}, \boldsymbol{\beta}, \boldsymbol{\epsilon}, j, \boldsymbol{\theta}_{c0}\right) \\ l_2 &= argmax_{l \neq j, x_{l,k} - x_{j,k} \leq 0} B\left(\mathbf{p}^{(n)}, \mathbf{u}_0^{(n)}, n, \mathbf{X}, \Gamma_{\boldsymbol{\beta}}, \boldsymbol{\beta}, \boldsymbol{\epsilon}, j, \boldsymbol{\theta}_{c0}\right) \end{aligned}$$

except for the case where l_1 and l_2 have multiple values, which we ignore because those situations occur with probability zero. Therefore,

$$
\begin{aligned}
(80) &= \frac{\partial}{\partial p_j^{(n)}} \int_{\beta_k} I\left(\beta_k \leq \underline{B}\left(\mathbf{p}^{(n)}, \mathbf{u}_0^{(n)}, n, \mathbf{X}, \Gamma_{\boldsymbol{\beta}}, \boldsymbol{\beta}, \boldsymbol{\epsilon}, j, \boldsymbol{\theta}_{c0}\right)\right) \\
&\quad \times I\left(\beta_k \geq \overline{B}\left(\mathbf{p}^{(n)}, \mathbf{u}_0^{(n)}, n, \mathbf{X}, \Gamma_{\boldsymbol{\beta}}, \boldsymbol{\beta}, \boldsymbol{\epsilon}, j, \boldsymbol{\theta}_{c0}\right)\right) \phi\left(\beta_k\right) d\beta_k \\
&= \left[\phi\left(\underline{B}\left(\mathbf{p}^{(n)}, \mathbf{u}_0^{(n)}, n, \mathbf{X}, \Gamma_{\boldsymbol{\beta}}, \boldsymbol{\beta}, \boldsymbol{\epsilon}, j, \boldsymbol{\theta}_{c0}\right)\right) \frac{\sigma_{\alpha 0}\left(\alpha + \eta_{\alpha 0}\right)}{\gamma_{\boldsymbol{\beta} k}^n \sigma_{\boldsymbol{\beta} 0 k}\left(x_{l_1,k} - x_{j,k}\right)}\right. \\
&\quad \left. -\phi\left(\overline{B}\left(\mathbf{p}^{(n)}, \mathbf{u}_0^{(n)}, n, \mathbf{X}, \Gamma_{\boldsymbol{\beta}}, \boldsymbol{\beta}, \boldsymbol{\epsilon}, j, \boldsymbol{\theta}_{c0}\right)\right) \frac{\sigma_{\alpha 0}\left(\alpha + \eta_{\alpha 0}\right)}{\gamma_{\boldsymbol{\beta} k}^n \sigma_{\boldsymbol{\beta} 0 k}\left(x_{l_2,k} - x_{j,k}\right)}\right] \\
&\quad \times I\left(\underline{B}\left(\mathbf{p}^{(n)}, \mathbf{u}_0^{(n)}, n, \mathbf{X}, \Gamma_{\boldsymbol{\beta}}, \boldsymbol{\beta}, \boldsymbol{\epsilon}, j, \boldsymbol{\theta}_{c0}\right) > \overline{B}\left(\mathbf{p}^{(n)}, \mathbf{u}_0^{(n)}, n, \mathbf{X}, \Gamma_{\boldsymbol{\beta}}, \boldsymbol{\beta}, \boldsymbol{\epsilon}, j, \boldsymbol{\theta}_{c0}\right)\right) (81)
\end{aligned}
$$

Then,

$$
\begin{aligned}
&\gamma_{\boldsymbol{\beta} k}^n \times (81) \\
\rightarrow &\phi\left(lim_{n\rightarrow\infty}\left\{-\frac{u_{l_1,0}^{(n)} - u_{j,0}^{(n)}}{\left(x_{l_1,k} - x_{j,k}\right)}\right\}\right) \frac{\sigma_{\alpha 0}\left(\alpha + \eta_{\alpha 0}\right)}{\sigma_{\boldsymbol{\beta} 0 k}\left(x_{l_1,k} - x_{j,k}\right)} - \phi\left(lim_{n\rightarrow\infty}\left\{-\frac{u_{l_2,0}^{(n)} - u_{j,0}^{(n)}}{\left(x_{l_2 k} - x_{j,k}\right)}\right\}\right) \frac{\sigma_{\alpha 0}\left(\alpha + \eta_{\alpha 0}\right)}{\sigma_{\boldsymbol{\beta} 0 k}\left(x_{l_2,k} - x_{j,k}\right)} \\
&\times I\left(lim_{n\rightarrow\infty}\left\{-\frac{u_{l_1,0}^{(n)} - u_{j,0}^{(n)}}{\left(x_{l_1,k} - x_{j,k}\right)}\right\} > lim_{n\rightarrow\infty}\left\{-\frac{u_{l_2,0}^{(n)} - u_{j,0}^{(n)}}{\left(x_{l_2,k} - x_{j,k}\right)}\right\}\right) \\
= &\phi\left(-\frac{\widetilde{u}_{j+1,0} - \widetilde{u}_{j,0}}{\left(x_{j+1,k} - x_{j,k}\right)}\right) \frac{\sigma_{\alpha 0}\left(\alpha + \eta_{\alpha 0}\right)}{\sigma_{\boldsymbol{\beta} 0 k}\left(x_{j+1,k} - x_{j,k}\right)} - \phi\left(-\frac{\widetilde{u}_{j-1,0} - \widetilde{u}_{j,0}}{\left(x_{j-1,k} - x_{j,k}\right)}\right) \frac{\sigma_{\alpha 0}\left(\alpha + \eta_{\alpha 0}\right)}{\sigma_{\boldsymbol{\beta} 0 k}\left(x_{j-1,k} - x_{j,k}\right)}
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\gamma_{\boldsymbol{\beta} k}^n \frac{\partial s_j}{\partial p_j^{(n)}} &\rightarrow \int_{\boldsymbol{\beta}_{-k}} \int_{\alpha} \left[\phi\left(-\frac{\widetilde{u}_{j+1,0} - \widetilde{u}_{j,0}}{\left(x_{j+1,k} - x_{jk}\right)}\right) \frac{\sigma_{\alpha 0}\left(\alpha + \eta_{\alpha 0}\right)}{\sigma_{\boldsymbol{\beta} 0 k}\left(x_{j+1,k} - x_{j,k}\right)}\right. \\
&\quad \left. - \phi\left(-\frac{\widetilde{u}_{j-1,0} - \widetilde{u}_{j,0}}{\left(x_{j-1,k} - x_{j,k}\right)}\right) \frac{\sigma_{\alpha 0}\left(\alpha + \eta_{\alpha 0}\right)}{\sigma_{\boldsymbol{\beta} 0 k}\left(x_{j-1,k} - x_{j,k}\right)}\right] \phi\left(\alpha\right) d\alpha \phi\left(\boldsymbol{\beta}_{-k}\right) d\boldsymbol{\beta}_{-k} \\
&= \phi\left(-\frac{\widetilde{u}_{j+1,0} - \widetilde{u}_{j,0}}{\left(x_{j+1,k} - x_{j,k}\right)}\right) \frac{\sigma_{\alpha 0} \eta_{\alpha 0}}{\sigma_{\boldsymbol{\beta} 0 k}\left(x_{j+1,k} - x_{j,k}\right)} - \phi\left(-\frac{\widetilde{u}_{j-1,0} - \widetilde{u}_{j,0}}{\left(x_{j-1,k} - x_{j,k}\right)}\right) \frac{\sigma_{\alpha 0} \eta_{\alpha 0}}{\sigma_{\boldsymbol{\beta} 0 k}\left(x_{j-1,k} - x_{j,k}\right)} < 0
\end{aligned}
$$

because $x_{j+1,k} - x_{j,k} > 0$, $x_{j-1,k} - x_{j,k} < 0$, $\eta_{\alpha 0} < 0$. Therefore, for sufficiently large n,

$$
\begin{aligned}
& p_j^{(n)} + \left[\frac{\partial s_j}{\partial p_j^{(n)}}\right]^{-1} \left(\mathbf{X}^{(n)}, \mathbf{p}^{(n)}, \mathbf{u}_0^{(n)}, \boldsymbol{\theta}_0\right) s_j \\
= \; & \gamma^n p_j + \gamma_{\boldsymbol{\beta} k}^n \left[\gamma_{\boldsymbol{\beta} k}^n \frac{\partial s_j}{\partial p_j^{(n)}} \left(\mathbf{X}^{(n)}, \mathbf{p}^{(n)}, \mathbf{u}_0^{(n)}, \boldsymbol{\theta}_0\right)\right]^{-1} s_j < 0
\end{aligned}
$$

because $\gamma < 1 < \gamma_{\boldsymbol{\beta} k}$, which contradicts positivity of marginal revenue.

Next, we modify the above case to allow $\gamma_{\boldsymbol{\beta}l}$ to be the same for $l \in \mathcal{K} \subset \{1, 2, \ldots, K\}$. Denote $\gamma_{\boldsymbol{\beta}l} \equiv \gamma_{\boldsymbol{\beta}\mathcal{K}}$ for $l \in \mathcal{K}$. Let $\mathbf{x}_{j,\mathcal{K}}$ and $\mathbf{x}_{j,\mathcal{K}^c}$ be defined accordingly. Then, letting $k \in \mathcal{K}$,

$$
\begin{aligned}
s_j &= \int_\alpha \int_{\boldsymbol{\beta}} \int_\epsilon s_j \left(\mathbf{X}^{(n)}, \mathbf{p}^{(n)}, \mathbf{u}_0^{(n)}, \alpha, \boldsymbol{\beta}, \boldsymbol{\theta}_0 \right) f(\epsilon) d\epsilon \phi\left(\boldsymbol{\beta}\right) d\boldsymbol{\beta} \phi(\alpha) d\alpha \\
&= \int_\alpha \int_{\boldsymbol{\beta}} \int_\epsilon \prod_{l \neq j}^{J} I \left(\left(\mathbf{x}_{l,\mathcal{K}}^{(n)} - \mathbf{x}_{j,\mathcal{K}}^{(n)} \right) \circ \boldsymbol{\sigma}_{\boldsymbol{\beta}0\mathcal{K}} \boldsymbol{\beta}_{\mathcal{K}} + \left(\mathbf{x}_{l,\mathcal{K}^c}^{(n)} - \mathbf{x}_{j,\mathcal{K}^c}^{(n)} \right) \circ \boldsymbol{\sigma}_{\boldsymbol{\beta}0\mathcal{K}^c} \boldsymbol{\beta}_{\mathcal{K}^c} \right. \\
&\quad \left. + \left(p_l^{(n)} - p_j^{(n)} \right) \sigma_{\alpha 0} \left(\alpha + \eta_{\alpha 0} \right) + \gamma_{\boldsymbol{\beta}\mathcal{K}}^n \sigma_{\boldsymbol{\beta}0k} \left(u_{l,0}^{(n)} - u_{j,0}^{(n)} \right) + \epsilon_l - \epsilon_j \leq 0 \right) \phi\left(\beta\right) d\beta d\alpha \\
&= \int_\alpha \int_{\boldsymbol{\beta}} \int_\epsilon \prod_{l \neq j}^{J} I \left(\left(\mathbf{x}_{l,\mathcal{K}} - \mathbf{x}_{j,\mathcal{K}} \right) \circ \boldsymbol{\sigma}_{\boldsymbol{\beta}0\mathcal{K}} \boldsymbol{\beta}_{\mathcal{K}} + \frac{\mathbf{x}_{l,\mathcal{K}^c}^{(n)} - \mathbf{x}_{j,\mathcal{K}^c}^{(n)}}{\gamma_{\boldsymbol{\beta}\mathcal{K}}^n} \circ \boldsymbol{\sigma}_{\boldsymbol{\beta}0\mathcal{K}^c} \boldsymbol{\beta}_{\mathcal{K}^c} \right. \\
&\quad \left. + \left(\frac{p_l^{(n)} - p_j^{(n)}}{\gamma_{\boldsymbol{\beta}\mathcal{K}}^n} \right) \sigma_{\alpha 0} \left(\alpha + \eta_{\alpha 0} \right) + \sigma_{\boldsymbol{\beta}0k} (u_{l,0}^{(n)} - u_{j,0}^{(n)}) + \frac{\epsilon_l - \epsilon_j}{\gamma_{\boldsymbol{\beta}\mathcal{K}}^n} \leq 0 \right) f(\epsilon) d\epsilon \phi\left(\boldsymbol{\beta}\right) d\boldsymbol{\beta} \phi(\alpha) d\alpha
\end{aligned}
$$

Using earlier arguments, it is straightforward to show that,

$$
s_j \to \int_{\boldsymbol{\beta}_{\mathcal{K}}} \prod_{l \neq j} I \left(\left(\mathbf{x}_{l,\mathcal{K}} - \mathbf{x}_{j,\mathcal{K}} \right) \circ \boldsymbol{\sigma}_{\boldsymbol{\beta}0\mathcal{K}} \boldsymbol{\beta}_{\mathcal{K}} + \sigma_{\boldsymbol{\beta}0k} \left(\widetilde{u}_{l,0} - \widetilde{u}_{j,0} \right) \leq 0 \right) \phi\left(\boldsymbol{\beta}_{\mathcal{K}}\right) d\boldsymbol{\beta}_{\mathcal{K}}
$$

Similarly, we can show the following:

$$
\begin{aligned}
\gamma_{\boldsymbol{\beta}\mathcal{K}}^n \frac{\partial s_j}{\partial p_j^{(n)}} \quad \to \quad & \eta_{\alpha 0} \sigma_{\alpha 0} \int_{\boldsymbol{\beta}_{\mathcal{K}-k}} \left[\phi \left(-\frac{\mathbf{x}_{l_1,\mathcal{K}-k} - \mathbf{x}_{j,\mathcal{K}-k}}{\sigma_{\boldsymbol{\beta}0k} \left(x_{l_1,k} - x_{j,k} \right)} \circ \boldsymbol{\sigma}_{\boldsymbol{\beta}0\mathcal{K}-k} \boldsymbol{\beta}_{\mathcal{K}-k} - \frac{\widetilde{u}_{l_1,0} - \widetilde{u}_{j,0}}{\left(x_{l_1,k} - x_{j,k} \right)} \right) \frac{1}{\sigma_{\boldsymbol{\beta}0k} \left(x_{l_1,k} - x_{j,k} \right)} \right. \\
& \left. - \phi \left(-\frac{\mathbf{x}_{l_2,\mathcal{K}-k} - \mathbf{x}_{j,\mathcal{K}-k}}{\sigma_{\boldsymbol{\beta}0k} \left(x_{l_2,k} - x_{j,k} \right)} \circ \boldsymbol{\sigma}_{\boldsymbol{\beta}0\mathcal{K}-k} \boldsymbol{\beta}_{\mathcal{K}-k} - \frac{\widetilde{u}_{l_2,0} - \widetilde{u}_{j,0}}{\left(x_{l_2,k} - x_{j,k} \right)} \right) \frac{1}{\sigma_{\boldsymbol{\beta}0k} \left(x_{l_2,k} - x_{j,k} \right)} \right] \\
& \times I \left(-\frac{\mathbf{x}_{l_1,\mathcal{K}-k} - \mathbf{x}_{j,\mathcal{K}-k}}{\sigma_{\boldsymbol{\beta}0k} \left(x_{l_1,k} - x_{j,k} \right)} \circ \boldsymbol{\sigma}_{\boldsymbol{\beta}0\mathcal{K}-k} \boldsymbol{\beta}_{\mathcal{K}-k} - \frac{\widetilde{u}_{l_1,0} - \widetilde{u}_{j,0}}{\left(x_{l_1,k} - x_{j,k} \right)} \right. \\
& \left. > -\frac{\mathbf{x}_{l_2,\mathcal{K}-k} - \mathbf{x}_{j,\mathcal{K}-k}}{\sigma_{\boldsymbol{\beta}0k} \left(x_{l_2,k} - x_{j,k} \right)} \circ \boldsymbol{\sigma}_{\boldsymbol{\beta}0\mathcal{K}-k} \boldsymbol{\beta}_{\mathcal{K}-k} - \frac{\widetilde{u}_{l_2,0} - \widetilde{u}_{j,0}}{\left(x_{l_2,k} - x_{j,k} \right)} \right) \phi\left(\boldsymbol{\beta}_{\mathcal{K}-k}\right) d\boldsymbol{\beta}_{\mathcal{K}-k} < 0
\end{aligned}
$$

where l_1 and l_2 are defined as earlier. Therefore, for sufficiently large n,

$$
\begin{aligned}
p_j^{(n)} + \left[\frac{\partial s_j}{\partial p_j^{(n)}} \left(\mathbf{X}^{(n)}, \mathbf{p}^{(n)}, \mathbf{u}_0^{(n)}, \boldsymbol{\theta}_0 \right) \right]^{-1} s_j \quad &= \quad \gamma^n p_j + \gamma_{\boldsymbol{\beta}\mathcal{K}}^n \left[\gamma_{\boldsymbol{\beta}\mathcal{K}}^n \frac{\partial s_j}{\partial p_j^{(n)}} \left(\mathbf{X}^{(n)}, \mathbf{p}^{(n)}, \mathbf{u}_0^{(n)}, \boldsymbol{\theta}_0 \right) \right]^{-1} s_j \\
&= \quad \gamma^n p_j + \gamma_{\boldsymbol{\beta}\mathcal{K}}^n \left[\gamma_{\boldsymbol{\beta}\mathcal{K}}^n \frac{\partial s_j}{\partial p_j^{(n)}} \left(\mathbf{X}^{(n)}, \mathbf{p}^{(n)}, \mathbf{u}_0^{(n)}, \boldsymbol{\theta}_0 \right) \right]^{-1} s_j < 0
\end{aligned}
$$

which contradicts positivity of marginal revenue.

Case 4) $\gamma > 1$, $min_l \left\{ \gamma_{\boldsymbol{\beta}l} \right\} < 1$, As before, let $\widetilde{\gamma}_{\boldsymbol{\beta}l} \equiv 1/\gamma_{\boldsymbol{\beta}l}$, $\widetilde{\gamma} \equiv 1/\gamma$. Then $\widetilde{\gamma} < 1$, $max_l \left\{ \widetilde{\gamma}_{\boldsymbol{\beta}l} \right\} > 1$, and we can proceed as in Case 3).

Case 5) $\gamma = 1$, $max_l \{\gamma_{\boldsymbol{\beta} l}\} > 1$. We can show that this leads to contradiction as in Case 3).

Case 6) $\gamma = 1$, $min_l \{\gamma_{\boldsymbol{\beta} l}\} < 1$. Let $\widetilde{\gamma}_{\boldsymbol{\beta} l} \equiv 1/\gamma_{\boldsymbol{\beta} l}$, $\widetilde{\gamma} \equiv 1/\gamma$. Then, $\widetilde{\gamma} = 1$, $max_l \{\widetilde{\gamma}_{\boldsymbol{\beta} l}\} > 1$, and from Case 5), this leads to contradiction.

The results of Case 1) and Case 3) imply that $\gamma < 1$ leads to a contradiction. Similarly, the results of Case 2) and Case 4) imply that $\gamma > 1$ leads to a contradiction as well. Therefore, it follows that $\gamma = 1$ is the only possibility. Then, from Cases 5) and 6) it follows that $\gamma = 1$ and both $max_l \{\gamma_{\boldsymbol{\beta} l}\} \leq 1$ and $min_l \{\gamma_{\boldsymbol{\beta} l}\} \geq 1$ need to hold. Therefore, $\gamma_{\boldsymbol{\beta} l} = 1$ for $l = 1, \ldots K$ and therefore, $\boldsymbol{\sigma}_{\boldsymbol{\beta} 0}$ and $\sigma_{\alpha 0}$ are identified.

A.2 Identification of the SNLLS Estimator

Proof of Proposition 2:

For notational convenience, we denote

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

Recall that, from Equation (34),

$$E \left[\left(C_{jm} - \psi \left(q_{jm}, \mathbf{w}_{jm}, \mathbf{x}_{jm}, MR_{jm} \left(\boldsymbol{\theta}_{c0} \right), \boldsymbol{\gamma}_0 \right) \right)^2 \right] = \sigma_\nu^2 + \sigma_\varsigma^2,$$

because

$$\begin{aligned} C_{jm} &= C^v \left(q_{jm}, \mathbf{w}_m, \mathbf{x}_{jm}, \upsilon_{jm} \right) + e(q_{jm}, \mathbf{w}_m, \mathbf{x}_{jm}) + \varsigma_{jm} + \nu_{jm} \\ &= \psi \left(q_{jm}, \mathbf{w}_{jm}, \mathbf{x}_{jm}, MR_{jm} \left(\boldsymbol{\theta}_{c0} \right), \boldsymbol{\gamma}_0 \right) + \varsigma_{jm} + \nu_{jm} \end{aligned}$$

and ς_{jm}, ν_{jm} are assumed to be i.i.d. distributed and independent from $(q_{jm}, \mathbf{w}_{jm}, \mathbf{X}_m, \mathbf{p}_m, \mathbf{s}_m)$. Then, for any $(\boldsymbol{\theta}_c, \boldsymbol{\gamma})$

$$\begin{aligned} & E \left[\left(C_{jm} - \psi \left(q_{jm}, \mathbf{w}_{jm}, \mathbf{x}_{jm}, MR_{jm} \left(\boldsymbol{\theta}_c \right), \boldsymbol{\gamma} \right) \right)^2 \right] \\ = \; & E \left[\left(\psi \left(q_{jm}, \mathbf{w}_{jm}, \mathbf{x}_{jm}, MR_{jm} \left(\boldsymbol{\theta}_{c0} \right), \boldsymbol{\gamma}_0 \right) - \psi \left(q_{jm}, \mathbf{w}_{jm}, \mathbf{x}_{jm}, MR_{jm} \left(\boldsymbol{\theta}_c \right), \boldsymbol{\gamma} \right) \right)^2 \right] \\ & + 2E \left[\left(\psi \left(q_{jm}, \mathbf{w}_{jm}, \mathbf{x}_{jm}, MR_{jm} \left(\boldsymbol{\theta}_{c0} \right), \boldsymbol{\gamma}_0 \right) - \psi \left(q_{jm}, \mathbf{w}_{jm}, \mathbf{x}_{jm}, MR_{jm} \left(\boldsymbol{\theta}_c \right), \boldsymbol{\gamma} \right) \right) \left(\nu_{jm} + \varsigma_{jm} \right) \right] + E \left(\nu_{jm} + \varsigma_{jm} \right)^2 \\ = \; & E \left[\left(\psi \left(q_{jm}, \mathbf{w}_{jm}, \mathbf{x}_{jm}, MR_{jm} \left(\boldsymbol{\theta}_{c0} \right), \boldsymbol{\gamma}_0 \right) - \psi \left(q_{jm}, \mathbf{w}_{jm}, \mathbf{x}_{jm}, MR_{jm} \left(\boldsymbol{\theta}_c \right), \boldsymbol{\gamma} \right) \right)^2 \right] + \sigma_\nu^2 + \sigma_\varsigma^2 \\ \geq \; & \sigma_\nu^2 + \sigma_\varsigma^2 \end{aligned}$$

Therefore, if $(\boldsymbol{\theta}_{c*}, \boldsymbol{\gamma}_*)$ satisfies

$$[\boldsymbol{\theta}_{c*}, \boldsymbol{\gamma}_*] = argmin_{(\boldsymbol{\theta},\gamma)\in\Theta_c\times\Gamma} E\left[C_{jm} - \sum_{l=1}^{\infty} \gamma_l \psi_l \left(q_{jm}, \mathbf{w}_m, \mathbf{x}_{jm}, MR_{jm}\left(\boldsymbol{\theta}_c\right)\right)\right]^2, \tag{82}$$

Then,

$$\psi\left(q_j, \mathbf{w}_j, \mathbf{x}_{jm}, MR_j\left(\boldsymbol{\theta}_{c*}\right), \boldsymbol{\gamma}_*\right) = \psi\left(q_j, \mathbf{w}_j, \mathbf{x}_{jm}, MR_j\left(\boldsymbol{\theta}_{c0}\right), \boldsymbol{\gamma}_0\right) \tag{83}$$

needs to be satisfied for all $(q_j, \mathbf{w}_j, \mathbf{X}_m, \mathbf{p}_m, \mathbf{s}_m)$ in the population. Now, by assumption, given $(q_j, \mathbf{w}_j, \mathbf{x}_{jm})$, $\psi\left(, \boldsymbol{\gamma}_0\right)$ is continuous and strictly increasing in $MR_j\left(\boldsymbol{\theta}_{c0}\right)$. Then, we show that given $(q_j, \mathbf{w}_j, \mathbf{x}_{jm})$, $\psi\left(, \boldsymbol{\gamma}_*\right)$, is also continuous and strictly monotone function of $MR_j(\theta_{c*})$. Furthermore, let the numbering of firms be ordered by price in a market. That is, the lowest price firm is $j = 1$ and the highest price firm is $j = J_m$. Now, consider the firm j such that $2 \leq j \leq J_m - 1$. Consider a version of Equation (70) where we substitute $\mathbf{p}$ with $\gamma\widetilde{\mathbf{p}}$. That is,

$$\begin{aligned} \overline{MR}_j\left(\mathbf{X}, \widetilde{\mathbf{p}}, \mathbf{s}\boldsymbol{\theta}_{c0}\right) &\equiv lim_{\gamma\to\infty} \frac{MR_j\left(\mathbf{X}, \gamma\widetilde{\mathbf{p}}, \mathbf{s}, \boldsymbol{\theta}_{c0}\right)}{\gamma} \\ = \ \widetilde{p}_j &+ \left[-\phi\left(\Phi^{-1}\left(\sum_{l\leq j} s_l\right)\right)\frac{\Phi^{-1}\left(\sum_{l\leq j} s_l\right) + \eta_{\alpha 0}}{\widetilde{p}_j - \widetilde{p}_{j+1}} - \phi\left(\Phi^{-1}\left(\sum_{l\leq j-1} s_l\right)\right)\frac{\Phi^{-1}\left(\sum_{l\leq j-1} s_l\right) + \eta_{\alpha 0}}{\widetilde{p}_{j-1} - \widetilde{p}_j}\right]^{-1} s_j \end{aligned} \tag{84}$$

There, given $\mathbf{X}$ and $\widetilde{\mathbf{p}}$ where $\widetilde{p}_1 < \widetilde{p}_2 < \ldots < \widetilde{p}_J$, by choosing $\widetilde{p}_j$ to be positive but sufficiently small and $\mathbf{s}$ such that $\sum_{l\leq j} s_l$ is sufficiently small, we can make $\overline{MR}_j\left(\mathbf{X}, \widetilde{\mathbf{p}}, \mathbf{s}\boldsymbol{\theta}_0\right)$ to be negative. Denote such vector $\widetilde{\mathbf{p}}$ to be $\underline{\widetilde{\mathbf{p}}}$. Then, by increasing $\widetilde{p}_j$ and changing $\widetilde{\mathbf{p}}_{-j}$ such that the difference $\widetilde{p}_i - \widetilde{p}_{i+1} \equiv A_{i,i+1}$, $i = 1, \ldots, J-1$ remain the same, we can continuously increase $\overline{MR}_j\left(\mathbf{X}, \widetilde{\mathbf{p}}, \mathbf{s}\boldsymbol{\theta}_{c0}\right)$ from negative to an arbitrarily large number, say $B > 0$. Denote such vector $\widetilde{\mathbf{p}}$ to be $\overline{\widetilde{\mathbf{p}}}$. Hence, for sufficiently large $\gamma > 0$, for $\underline{\mathbf{p}} \equiv \gamma\underline{\widetilde{\mathbf{p}}}$ $MR_j\left(\mathbf{X}, \gamma\underline{\widetilde{\mathbf{p}}}, \mathbf{s}, \boldsymbol{\theta}_{c0}\right) < 0$, and for $\overline{\mathbf{p}} \equiv \gamma\overline{\widetilde{\mathbf{p}}}$, $MR_j\left(\mathbf{X}, \overline{\mathbf{p}}, \mathbf{s}, \boldsymbol{\theta}_{c0}\right) \equiv \overline{B} > \gamma B/2$. Furthermore, for sufficiently large $\gamma > 0$, the 2nd term of the Equation (84) is bounded. Because $MR_j\left(\right)$ is a continuous function of $\mathbf{p} \geq \mathbf{0}$, from the Intermediate Value Theorem, there exists $\mathbf{p}_0$ such that $p_{j,0} \in \left(\underline{p}_j, \overline{p}_j\right)$ and $p_{i,0} - p_{i+1,0} = \gamma A_{i,i+1}$, $i = 1, \ldots, J-1$, such that $MR_j\left(\mathbf{X}, \mathbf{p}_0, \mathbf{s}, \boldsymbol{\theta}_{c0}\right) = 0$. Similarly, there exists $\mathbf{p}_1$ such that $p_{j,1} \in \left(p_{0j}, \overline{p}_j\right)$ and $p_{i,1} - p_{i+1,1} = \gamma A_{i,i+1}$, $i = 1, \ldots, J-1$, such that $MR_j\left(\mathbf{X}, \mathbf{p}_1, \mathbf{s}, \boldsymbol{\theta}_{c0}\right) = \overline{B}/2$. If we continue applying the Intermediate Value Theorem, for any $y \in \left[0, \overline{B}\right]$, we have $\mathbf{p}\left(y\right)$ such that $MR_j\left(\mathbf{X}, \mathbf{p}\left(y\right), \mathbf{s}, \boldsymbol{\theta}_{c0}\right) = y$. Then, by construction,

$MR_j(\mathbf{X},\mathbf{p}(y),\mathbf{s},\boldsymbol{\theta}_{c0})$ is increasing in y. Therefore, $\psi(q_j,\mathbf{w}_j,\mathbf{x}_{jm},MR_j(\mathbf{X},\mathbf{p}(y),\mathbf{s},\boldsymbol{\theta}_{c0}),\boldsymbol{\gamma}_0)$ is also strictly increasing and continuous in y. Because $\psi(q_j,\mathbf{w}_j,\mathbf{x}_{jm},MR,\boldsymbol{\gamma}_*)$ is continuous in MR, in order the Equation (83) to hold, $MR_j(\mathbf{X},\mathbf{p}(y),\mathbf{s},\boldsymbol{\theta}_{c*})$ needs to be continuous in y as well.

Now, suppose there exists a segment $\left[\underline{y},\overline{y}\right]\in\left[0,\overline{B}\right]$ where $MR_j(\mathbf{X},\mathbf{p}(y),\mathbf{s},\boldsymbol{\theta}_{c*})$ is constant. Then, for any two values y, y', satisfying $y>y'$, $y\in\left[\underline{y},\overline{y}\right]$ and $y'\in\left[\underline{y},\overline{y}\right]$, $MR_j(\mathbf{X},\mathbf{p}(y),\mathbf{s},\boldsymbol{\theta}_{c*})=MR_j(\mathbf{X},\mathbf{p}(y'),\mathbf{s},\boldsymbol{\theta}_{c*})$ implies

$$\psi(q_j,\mathbf{w}_j,\mathbf{x}_{jm},MR_j(\mathbf{X},\mathbf{p}(y),\mathbf{s},\boldsymbol{\theta}_{c*}),\boldsymbol{\gamma}_*)=\psi\left(q_j,\mathbf{w}_j,\mathbf{x}_{jm},MR_j\left(\mathbf{X},\mathbf{p}\left(y'\right),\mathbf{s},\boldsymbol{\theta}_{c*}\right),\boldsymbol{\gamma}_*\right)$$

whereas

$$\psi(q_j,\mathbf{w}_j,\mathbf{x}_{jm},MR_j(\mathbf{X},\mathbf{p}(y),\mathbf{s},\boldsymbol{\theta}_{c0}),\boldsymbol{\gamma}_0)>\psi\left(q_j,\mathbf{w}_j,\mathbf{x}_{jm},MR_j\left(\mathbf{X},\mathbf{p}\left(y'\right),\mathbf{s},\boldsymbol{\theta}_{c0}\right),\boldsymbol{\gamma}_0\right)$$

which contradicts Equation (83). Next, consider the case where $MR_j(\mathbf{X},\mathbf{p}(y),\mathbf{s},\boldsymbol{\theta}_{c*})$ is strictly increasing in $y\in\left[\underline{y},\widehat{y}\right]$ and strictly decreasing in $y\in[\widehat{y},\overline{y}]$. Then, from continuity, for $y<\widehat{y}$ sufficiently close to $\widehat{y}$, $MR_j(\mathbf{X},\mathbf{p}(\overline{y}),\mathbf{s},\boldsymbol{\theta}_{c*})<MR_j(\mathbf{X},\mathbf{p}(y),\mathbf{s},\boldsymbol{\theta}_{c*})<MR_j(\mathbf{X},\mathbf{p}(\widehat{y}),\mathbf{s},\boldsymbol{\theta}_{c*})$ holds Then, from the intermediate value function, there exist $y'\in[\widehat{y},\overline{y}]$ such that $MR_j(\mathbf{X},\mathbf{p}(y'),\mathbf{s},,\boldsymbol{\theta}_{c*})=MR_j(\mathbf{X},\mathbf{p}(y),\mathbf{s},,\boldsymbol{\theta}_{c*})$. Again, this contradicts $\psi(,\boldsymbol{\gamma}_0)$ being strictly increasing in MR. Similar argument can be made for the case where $MR_j(\mathbf{X},\mathbf{p}(y),\mathbf{s},\boldsymbol{\theta}_{c*})$ is strictly decreasing in $y\in\left[\underline{y},\widehat{y}\right]$ and strictly increasing in $y\in[\widehat{y},\overline{y}]$. Thus, we have shown that $MR_j(\mathbf{X},\mathbf{p}(y),\mathbf{s},\boldsymbol{\theta}_{c*})$ is strictly monotone in y, and therefore, from Equation (83), $\psi(,\boldsymbol{\gamma}_*)$, is a strictly monotone function of $MR_j(\boldsymbol{\theta}_{c*})$.

Recall that in Proposition 1 we have shown that the BLP demand function satisfies Condition 1. From Condition 1, there exist two firms with $(\mathbf{p},\mathbf{s},\mathbf{X},q,\mathbf{w},j)$ and $\left(\mathbf{p}^\dagger,\mathbf{s}^\dagger,\mathbf{X}^\dagger,q,\mathbf{w},j^\dagger\right)$ and

$$MR(\mathbf{p},\mathbf{s},\mathbf{X},j,\boldsymbol{\theta}_{c0})=MR\left(\mathbf{p}^\dagger,\mathbf{s}^\dagger,\mathbf{X}^\dagger,j^\dagger,\boldsymbol{\theta}_{c0}\right).$$

Then,

$$\psi(q,\mathbf{w},\mathbf{x},MR(\mathbf{p},\mathbf{s},\mathbf{X},j,\boldsymbol{\theta}_{c0}),\boldsymbol{\gamma}_0)=\psi\left(q,\mathbf{w},\mathbf{x},MR\left(\mathbf{p}^\dagger,\mathbf{s}^\dagger,\mathbf{X}^\dagger,j^\dagger,\boldsymbol{\theta}_{c0}\right),\boldsymbol{\gamma}_0\right)$$

which implies, from Equation (83)

$$\psi\left(q,\mathbf{w},\mathbf{x},MR\left(\mathbf{p},\mathbf{s},\mathbf{X},j,\boldsymbol{\theta}_{c*}\right),\boldsymbol{\gamma}_*\right)=\psi\left(q,\mathbf{w},\mathbf{x},MR\left(\mathbf{p}^{\dagger},\mathbf{s}^{\dagger},\mathbf{X}^{\dagger},j^{\dagger},\boldsymbol{\theta}_{c*}\right),\boldsymbol{\gamma}_*\right)$$

Then, from strict monotonicity of the function $\psi\left(,\boldsymbol{\gamma}_*\right)$ with respect to MR,

$$MR\left(\mathbf{p},\mathbf{s},\mathbf{X},j,\boldsymbol{\theta}_{c*}\right)=MR\left(\mathbf{p}^{\dagger},\mathbf{s}^{\dagger},\mathbf{X}^{\dagger},j,\boldsymbol{\theta}_{c*}\right)$$

Because BLP satisfies Condition 1, $\boldsymbol{\theta}_{c*}=\boldsymbol{\theta}_{c0}$.

A.3 Semi-Parametric Cost Function Estimation

The cost function can be recovered from the pseudo-cost function estimate in three steps.

Step 1

Suppose that we already estimated $\widehat{\psi}\left(q,\mathbf{w},\mathbf{x},MR,\widehat{\boldsymbol{\gamma}}_M\right)$. We then nonparametrically estimate marginal cost for a given point $(q,\mathbf{w},\mathbf{x},C)$ as follows,

$$\begin{aligned}\widehat{MC}\left(q,\mathbf{w},\mathbf{x},C\right) &= \sum_{jm} MR\left(\mathbf{X}_m,\mathbf{p}_m,\mathbf{s}_m,j,\widehat{\boldsymbol{\theta}}_{cM}\right)\\ &\quad \times W_{\mathbf{h}}\left(q-q_{jm},\mathbf{w}-\mathbf{w}_{jm},\mathbf{x}-\mathbf{x}_{jm},C-\widehat{\psi}\left(q_{jm},\mathbf{w}_{jm},\mathbf{x}_{jm},MR_{jm}\left(\widehat{\boldsymbol{\theta}}_{cM}\right),\widehat{\boldsymbol{\gamma}}_M\right)\right)\end{aligned}$$

where $\boldsymbol{\theta}_{cM}$ is the estimated demand parameter, and the weight function is

$$\begin{aligned}& W_{\mathbf{h}}\left(q-q_{jm},\mathbf{w}-\mathbf{w}_{jm},\mathbf{x}-\mathbf{x}_{jm},C-\widehat{\psi}_{jm}\right)\\ =\ & \frac{K_{h_q}\left(q-q_{jm}\right)K_{h_w}\left(\mathbf{w}-\mathbf{w}_{jm}\right)K_{h_x}\left(\mathbf{x}-\mathbf{x}_{jm}\right)K_{h_{MR}}\left(C-\widehat{\psi}_{jm}\right)}{\sum_{kl}K_{h_q}\left(q-q_{kl}\right)K_{h_w}\left(\mathbf{w}-\mathbf{w}_{kl}\right)K_{h_x}\left(\mathbf{x}-\mathbf{x}_{jm}\right)K_{h_{MR}}\left(C-\widehat{\psi}_{kl}\right)}.\end{aligned}$$

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

Step 2

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

$$\widehat{C}\left(\overline{q}+\Delta q,\mathbf{w},\mathbf{x},\overline{\upsilon}\right)=\overline{C}+\widehat{MC}\left(\overline{q},\mathbf{w},\mathbf{x},\overline{C}\right)\Delta q.$$

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

$$\widehat{C}\left(\overline{q}+\Delta q,\mathbf{w},\mathbf{x},\overline{\upsilon}\right)=C\left(\overline{q}+\Delta q,\mathbf{w},\mathbf{x},\overline{\upsilon}\right)+O\left(\left(\Delta q\right)^{2}\right)+o_{p}\left(1\right)\Delta q.$$

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

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

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

$$\widehat{C}\left(\overline{q}+k\Delta q,\mathbf{w},\mathbf{x},\overline{\upsilon}\right)=C\left(\overline{q}+k\Delta q,\mathbf{w},\mathbf{x},\overline{\upsilon}\right)+O\left(k\left(\Delta q\right)^{2}\right)+ko_{p}\left(1\right)\Delta q$$

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

Step 3

Next, we derive the cost function for different input price. First, we derive the nonparametric estimate of the input demand. Denote $\mathbf{l}\left(q,\mathbf{w},\mathbf{x},C\right)$ to be the vector of input demand given output q, input price $\mathbf{w}$, observed characteristics $\mathbf{x}$ and cost C. Then, its nonparametric estimate is:

$$\widehat{\mathbf{l}}\left(q,\mathbf{w},\mathbf{x},C\right)=\sum_{jm}\mathbf{l}_{jm}W_{h}\left(q-q_{jm},\mathbf{w}-\mathbf{w}_{jm},\mathbf{x}-\mathbf{x}_{jm},C-\widehat{\psi}\left(q_{jm},\mathbf{w}_{jm},\mathbf{x}_{jm},MR_{jm}\left(\widehat{\boldsymbol{\theta}}_{M}\right),\widehat{\boldsymbol{\gamma}}_{M}\right)\right).$$

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

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

Start, as before, with $\overline{q}$, $\mathbf{w}$, and $\overline{C}$. Next, we derive the cost for the output $\overline{q}$, $\mathbf{w}+\Delta\mathbf{w}$ for small $\Delta\mathbf{w}$ that has the same variable cost shock $\overline{\upsilon}$. Then:

$$\widehat{C}_{1}=\widehat{C}\left(\overline{q},\mathbf{w}+\Delta\mathbf{w},\mathbf{x},\bar{\upsilon}\right)=\overline{C}+\widehat{\mathbf{l}}\left(\overline{q},\mathbf{w},\mathbf{x},\overline{C}\right)\Delta\mathbf{w}+O\left(\left(\left\Vert \Delta\mathbf{w}\right\Vert ^{2}\right)\right)+o_{p}\left(1\right)\left\Vert \Delta\mathbf{w}\right\Vert .$$

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

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

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

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

for any $k > 0$.

A.4 Further specification and data issues

A.4.1 Economic versus accounting cost

The cost data we envision using comes from accounting statements of firms.[40] Such data do not necessarily reflect the economic cost that the firm considers in making input and output choices. More concretely, we may not be appropriately taking into account the opportunity cost of the resources that are used in purchasing the necessary input to produce output. Fortunately, from accounting statements, we may be able to obtain information on other activities that the firm may be pursuing in addition to the production of output. For example, we may find details on firms' financial investments including their rate of return.[41] Suppose that the return on a unit of a financial investment is r_{jm}. Then, the opportunity cost of production is r_{jm}, and the firm will produce and sell output until marginal revenue equals marginal cost that incorporates this cost, i.e.,

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

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

$$\frac{1}{\sum_m J_m} \sum_{j,m} \left[C_{jm} - \psi_l\left(q_{jm}, \mathbf{w}_m, \mathbf{x}_{jm}, MR_{jm}\left(\boldsymbol{\theta}_c\right) - r_{jm}, \boldsymbol{\gamma}\right)\right]^2.$$

SNLLS is also consistent with the following specification, where firms do not equate marginal cost to marginal revenue: $MR_{jm}(\boldsymbol{\theta}) - r_{jm} = \nu\left(MC\left(q_{jm}, \mathbf{w}_m, \mathbf{x}_{jm}, \upsilon_{jm}\right)\right)$. That is, as long as we can obtain information on the financial opportunities that the firm has other than production,

[40] Indeed, accounting data are typically used in the literature that estimates cost functions to evaluate market power, measure economies of scale or scope, and so on.

[41] In the application of our estimator to the U.S. banking industry, such information is readily available. Most large industries like banking that are subject to some form of regulatory oversight are likely to report such data.

we can incorporate them into our estimator. Then, the estimator will not be subject to bias even if the cost data we use corresponds to accounting costs.

A.4.2 Endogenous product characteristics.

Delete the entire subsubsection: If firms strategically choose prices *and* product characteristics, then elements of $\mathbf{X}_m$ will be correlated with the demand shock ξ_{jm}. To accommodate endogenous product characteristics, researchers have recently started estimating BLP models that include first order conditions for optimal prices and product characteristics.[42] In particular, Petrin and Seo (2016) use the first order conditions of product choice and additional panel data moment restrictions to identify the price coefficient. The endogeneity of product characteristics does not prevent us from identifying the price coefficient, and thus, markups, as discussed before. It is an issue only if researchers want to estimate the coefficients of the observed characteristics.

To deal with endogenous product characteristics, we follow a strategy that is similar to Petrin and Seo (2016). We differ from them by using cost data instead of instruments. We modify the cost function as below:

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

where $\boldsymbol{\upsilon}_{\mathbf{x}}$ corresponds to the shock that affects the production of observed characteristics. The additional F.O.C. for optimal product characteristics would then be

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

where $MR_{\mathbf{x},jm}$ is the vector of marginal revenue of firm j in market m of the product characteristics choice. Then, the SNLLS part would be modified as follows:

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

where MR_{jm} is, as before, marginal revenue with respect to output.

[42] See, for example, Chu (2010), Fan (2013), and Byrne (2015). For an excellent overview of the empirical literature on endogenous product characteristics see Crawford (2012). It is worth noting that these applications all maintain the static decision-making assumption of BLP; firms are allowed to adjust their product characteristics period-by-period but are not forward-looking in doing so. A recent paper by Gowrisankaran and Rysman (2012) develops and estimates a dynamic version of a differentiated products oligopoly model, the solution of which is computationally extremely burdensome.

A.4.3 Cost function restrictions.

So far, in the estimation exercise, we have not imposed any assumptions on the shape of the pseudo-cost function φ except that it is a smooth function of output, input price, observed product characteristics and marginal revenue. The cost function that is recovered is not restricted to have properties such as positive slopes, homogeneity of degree one in input prices, nor convexity in output or cost shock.

Imposing the restriction of homogeneity in input prices in estimation is straightforward. If the cost function is homogeneous of degree one in input price, so is the marginal cost function. Then, for an input price w_1, which is the first element of the vector $\mathbf{w}$ of input prices,

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

and

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

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

$$\frac{1}{\sum_m J_m} \sum_{j,m} \left[\frac{C_{jm}}{w_{1,m}} - \sum_l \gamma_l \psi_l \left(q_{jm}, \frac{\mathbf{w}_{-1,m}}{w_{1,m}}, \mathbf{x}_{jm}, \frac{MR_{jm}(\boldsymbol{\theta}_c)}{w_{1,m}} \right) \right]^2, \tag{85}$$

where $\mathbf{w}_{-1,m} = (w_{2,m}, \ldots, w_{L,m})$ is the vector of input prices except $w_{1,m}$.

A.4.4 Missing cost data and multi product firms

Until now we have assumed that cost data are available for all firms in the sample. However, it could very well be the case that we observe costs only for some firms and not others. Even in this case, we can estimate the structural parameters consistently by constructing the SNLLS part using only those firms for which we have cost data. Because the SNLLS part of our estimator does not involve any orthogonality conditions, and because the random components of the fixed cost and the measurement error of cost are assumed to be i.i.d, choosing firms in this way for estimation will not result in selection bias. It is important to notice, however, that we still need demand-side data for all firms in the same market to compute marginal revenue. Luckily, such demand-side data tends to be available to researchers for many industries. Allowing for missing cost data is important in reducing the concern about reliability of cost data. That is, in practice, researchers can go over the accounting cost data carefully and simply remove the cost data that have anomalies.

A more difficult situation would be when firms produce multiple products, but only the total cost of all products is observable in the data. To address this issue, we follow the setup of BLP closely. Let $\mathcal{F}_f$ be the set of product-market combinations that is produced by firm f . That is, $\mathcal{F}_f = \{(j, m) : I_f(j, m) = 1\}$ where $I_f(j, m) = 1$ if the j^{th} product in market m is produced by firm f and 0 if otherwise. Let $\mathbf{q}_f$ be the vector of outputs of products produced by firm f. That is, $\mathbf{q}_f = \left(q_{j1,m_1}, q_{j2,m_2}, \ldots, q_{j_{n_f},m_{n_f}}\right)$ and n_f is the number of product-market combinations produced by firm f. In our vector notation, we follow the convention of ordering of the products in the same market and the products in different markets. That is, if the firm f has two different products j and j' with $j < j'$ in the same market m, then q_{jm} comes before $q_{j'm}$ in the vector $\mathbf{q}_f$. If the firm f has two different products j in market m and j' in different market m' with $m < m'$, then q_{jm} comes before $q_{j'm'}$ in the vector $\mathbf{q}_f$. Similarly, let $\mathbf{X}_f$ be the $(K \times n_f)$ matrix of observable product characteristics and $\boldsymbol{\upsilon}_f$ be the vector of cost shocks of products belonging to $\mathcal{F}_f$. Similarly, denote $\mathbf{mc}_f$, $\mathbf{p}_f$ and $\mathbf{s}_f$ to be the vector of marginal costs, prices and market shares of products belonging to $\mathcal{F}_f$. Input prices are denoted by the matrix $\mathbf{w}_f$ with L input prices for each market the firm operates in. Then, we can modify the cost function of firm f as

$$C\left(\mathbf{q}_f, \mathbf{w}_f, \mathbf{X}_f, \boldsymbol{\upsilon}_f\right).$$

Then, the F.O.C. for profit maximization for product and market $(j, m) \in \mathcal{F}_f$ is:

$$s_{jm} + \sum_{(r',m') \in \mathcal{F}_f} \left(p_{r',m'} - mc_{r',m'}\right) \frac{\partial s_{r',m'}}{\partial p_{jm}} = 0.$$

Hence, we can write

$$\mathbf{mc}_f = \mathbf{p}_f + \boldsymbol{\Delta}_f^{-1} \mathbf{s}_f,$$

where row h and column i of the matrix $\boldsymbol{\Delta}_f$ is $[\boldsymbol{\Delta}_f]_{hi} = \partial s_{j_h m_h} / \partial p_{j_i,m_i}$. Since markets are isolated, $\partial s_{j_h,m_h} / \partial p_{j_i,m_i} = 0$ for $m_h \neq m_i$.

Then, the SNLLS component can be modified as follows:

$$\frac{1}{F} \sum_f \left[C_f - \sum_l \gamma_{lf} \psi_{lf} \left(\mathbf{q}_f, \mathbf{w}_f, \mathbf{X}_f, \mathbf{p}_f + \boldsymbol{\Delta}_f^{-1} \left(\mathbf{X}_f, \mathbf{p}_f, \mathbf{s}_f, \boldsymbol{\theta}_0\right) \mathbf{s}_f \left(\mathbf{X}_f, \mathbf{p}_f, \mathbf{s}_f, \boldsymbol{\theta}_0\right) \right) \right]^2,$$

where F is the number of firms;ψ_{lf} is the lth polynomial for firm f and γ_{lf} is the corresponding coefficient. These basis polynomials vary with firms because they depend on the number of products produced by each firm. Estimation can proceed as long as the number of product-market combination for a firm is not too large. Otherwise, we would face a Curse of Dimensionality issue in estimation.

If the number of products n_f is large, one could consider imposing more structure on the cost function. For example, we could specify the total cost of firm f as the sum of the costs of all the products that it produces. That is:

$$\begin{aligned} C_f &= \sum_i C\left(q_{j_i,m_i}, \mathbf{w}_{m_i}, \mathbf{X}_{j_i,m_i}, \upsilon_{j_i,m_i}\right) + \nu_f \\ &= \sum_i \sum_l \gamma_l \psi_l \left(q_{j_i,m_i}, \mathbf{w}_{m_i}, \mathbf{X}_{j_i,m_i}, \mathbf{p}_{j_i,m_i} + \left[\boldsymbol{\Delta}_f^{-1}\left(\mathbf{X}_f, \mathbf{p}_f, \mathbf{s}_f, \boldsymbol{\theta}_0\right) \mathbf{s}_f\left(\mathbf{X}_f, \mathbf{p}_f, \mathbf{s}_f, \boldsymbol{\theta}_0\right)\right]_i \right) + \nu_f, \end{aligned}$$

Hence, the SNLLS component of our estimator can be modified as follows,

$$\frac{1}{F}\sum_f \left[C_f - \sum_i \sum_l \gamma_l \psi_l \left(q_{j_i,m_i}, \mathbf{w}_{m_i}, \mathbf{X}_{j_i,m_i}, \mathbf{p}_{j_i,m_i} + \left[\boldsymbol{\Delta}_f^{-1}\left(\mathbf{X}_f, \mathbf{p}_f, \mathbf{s}_f, \boldsymbol{\theta}_0\right) \mathbf{s}_f\left(\mathbf{X}_f, \mathbf{p}_f, \mathbf{s}_f, \boldsymbol{\theta}_0\right)\right]_i \right) \right]^2.$$

The identification argument works as follows: Consider two firms f and $f^\dagger$ that operate in two markets. That is, firm f sells product j in market m and product j' in market m'. Similarly, firm $f^\dagger$ sells product $j^\dagger$ in market $m^\dagger$ and product $j^{\dagger\prime}$ in market $m^{\dagger\prime}$. Suppose the two firms also satisfy $q_{jm} = q_{j^\dagger m^\dagger}$, $\mathbf{w}_m = \mathbf{w}_{m^\dagger}$, $\mathbf{x}_{jm} = \mathbf{x}_{j^\dagger m^\dagger}$, and the same holds for product/market $j'm'$ and $j^{\dagger\prime}m^{\dagger\prime}$. Suppose further that the two firms have the same marginal revenue for products jm and $j^\dagger m^\dagger$, and $j'm'$ and $j^{\dagger\prime}m^{\dagger\prime}$. Then, the two true costs minus the fixed cost shocks for firm f in market m, and firm $f^\dagger$ in market $m^\dagger$ are the same, and similar result holds for firm f's operations in market m' and the firm $f^\dagger$'s operation in market $m^{\dagger\prime}$. Therefore, the true parameter will equalize the expected cost of the two firms. Then, arguments similar to Proposition 2 shows that the true demand parameters can be obtained by the SNLLS estimator.

Note if we were to use the data on the total costs of the firms that operate in multiple markets, they need to practice third-degree price discrimination. Then, the question arises as to what to do if they don't. In that case, we don't use their cost data and only use the cost data from firms that either operate in a single market or exercise third-degree price discrimination if they operate in multiple markets. Such choice of firms whose cost data we use does not result in any sample selection issues because the selection does not alter the conditional distribution of the measurement error.

A.5 Large Sample Properties

In this section we show that the estimator is consistent and asymptotically normal. Notice that in our sample, we have oligopolistic firms in the same market. Because of strategic interaction, equilibrium prices and outputs of the firms in the same market are likely to be correlated. To avoid the difficulty arising from such within-market correlation, our consistency proof will primarily exploit the large number of isolated markets, with the assumption that wages, unobserved product quality and variable cost shocks are independent across markets

The assumption of independence of variables across markets is employed for simplicity. We leave the asymptotic analysis with some across market dependence for future research. For Strong Law of Large Numbers under weaker assumptions, see W.K.A (1988) and the related literature. As we have discussed earlier, those assumptions are not required for identification. Without loss of generality, we assume that in each market, the number of firms is J. Notice that in our estimation procedure, we have two steps: one that involves the difference between the cost in the data and the nonparametrically approximated pseudo-cost function to identify $\boldsymbol{\theta}_c$ (which is α for the Berry logit model and $(\mu_\alpha, \sigma_\alpha)$ and σ_β for the BLP random coefficient logit model) and the second step is the OLS estimation based on the orthogonality condition $\boldsymbol{\xi}_m \perp \mathbf{X}_m$ which identifies$\boldsymbol{\beta}$ for the Berry logit model and $\boldsymbol{\mu}_\beta$ for the BLP model. We start with denoting $\boldsymbol{\theta} = (\boldsymbol{\theta}_\beta, \boldsymbol{\theta}_c)$, where $\boldsymbol{\theta}_\beta$ is the vector of parameters estimated in the second step. In our proof, for the pseudo-cost function part, we follow Bierens (2014) closely.

For the description of our objective function, let $\mathbf{y}_m = \left(\mathbf{q}_m, vec\left(\mathbf{W}_m\right)', \mathbf{C}_m, vec\left(\mathbf{X}_m\right)', vec\left(\mathbf{p}_m\right)', vec\left(\mathbf{s}_m\right)'\right)'$, $\mathbf{z}_m = \left(\mathbf{q}_m, vec\left(\mathbf{W}_m\right)', vec\left(\mathbf{X}_m\right)', vec\left(\mathbf{p}_m\right)', vec\left(\mathbf{s}_m\right)'\right)'$, where $\mathbf{C}_m = (C_{1m}, C_{2m}, ..., C_{Jm})'$, $\mathbf{W}_m = (\mathbf{w}_{1m}, \mathbf{w}_{2m}, ..., \mathbf{w}_{jm})'$.

For the sake of analytical simplicity, we derive the asymptotic results for a slightly modified component of the objective function for firm j in market m, which is

$$f_j\left(\mathbf{y}_m, \boldsymbol{\theta}_c, t, b\right) \equiv \left[C_{jm} - \varphi\left(q_{jm}, \mathbf{w}_{jm}, \mathbf{x}_{jm}, MR_{jm}\left(\boldsymbol{\theta}_c\right)\right)\right]^2 \frac{h\left(\frac{MR_{jm}(\boldsymbol{\theta}_c)-t}{b}\right)}{\frac{1}{MJ}\sum_{m=1}^{M}\sum_{j=1}^{J} h\left(\frac{MR_{jm}(\boldsymbol{\theta}_c)-t}{b}\right)}.$$

We set the function $h() \geq 0$ to be twice differentiable; equal to 1 for $x \leq 0$; decreasing in x for $x > 0$ and equal to zero for $x \geq 4$. Since the objective function is zero for $MR_{jm}(\boldsymbol{\theta}_c) \geq t + 4b$, and the rest of the variables in the objective function are assumed to be bounded, it follows that the objective function is bounded, which simplifies the proofs. For large $t > 0$, the modified objective function is practically equivalent to the original objective function

$[C_{jm} - \varphi(q_{jm}, \mathbf{w}_{jm}, \mathbf{x}_{jm}, MR_{jm}(\boldsymbol{\theta}_c))]^2$.

We next define $h()$. Let

$$\widetilde{g}(u) \equiv \left[\frac{1}{2} - \frac{1}{12}u^3\right] I(0 \leq u \leq 1) + \left[1 - \frac{1}{2}u - \frac{1}{12}(2-u)^3\right] I(1 < u \leq 2)$$

$$g(u) = \widetilde{g}(0) + \widetilde{g}(u) I(0 \leq u \leq 2) - \widetilde{g}(4-u) I(2 < u \leq 4)$$

and let

$$h(u) \equiv I(u < 0) + g(u) I(0 \leq u \leq 4).$$

Then, h is a continuous function and

$$h(u) = \begin{cases} 1 & if\, u \leq 0 \\ \frac{11}{12} & if\, u = 1 \\ \frac{1}{2} & if\ u = 2 \\ \frac{1}{12} & if\ u = 3 \\ 0 & if\ u \geq 4 \end{cases}$$

Also,

$$\widetilde{g}'(u) \equiv -\frac{1}{4}u^2 I(0 \leq u \leq 1) + \left[-\frac{1}{2} + \frac{1}{4}(2-u)^2\right] I(1 < u \leq 2)$$

and h' is a continuous function with

$$h'(u) = \begin{cases} 0 & if\, u \leq 0 \\ -\frac{1}{4} & if\, u = 1 \\ -\frac{1}{2} & if\ u = 2 \\ -\frac{1}{4} & if\ u = 3 \\ 0 & if\ u \geq 4 \end{cases}$$

Similarly,

$$\widetilde{g}''(u) \equiv -\frac{1}{2}uI(0 \leq u \leq 1) - \left[\frac{1}{2}(2-u)\right] I(1 < u \leq 2)$$

and h'' is a continuous function with

$$h''(u) = \begin{cases} 0 & if\ u \leq 0 \\ -\frac{1}{2} & if\ u = 1 \\ 0 & if\ u = 2 \\ -\frac{1}{2} & if\ u = 3 \\ 0 & if\ u \geq 4 \end{cases}$$

Now, define,

$$\widetilde{f}_j(\mathbf{y}_m, \boldsymbol{\theta}_c, t, b) \equiv f_j(\mathbf{y}_m, \boldsymbol{\theta}_c)\, \widetilde{w}_{jm}(MR_{jm}(\boldsymbol{\theta}_c), t, b)$$

where

$$f_j(\mathbf{y}_m, \boldsymbol{\theta}_c) \equiv [C_{jm} - \varphi(q_{jm}, \mathbf{w}_{jm}, \mathbf{x}_{jm}, MR_{jm}(\boldsymbol{\theta}_c))]^2$$

where

$$\widetilde{w}_{jm}(MR_{jm}(\boldsymbol{\theta}_c), t, b) \equiv \frac{h\left(\frac{MR_{jm}(\boldsymbol{\theta}_c)-t}{b}\right)}{E\left[h\left(\frac{MR_{jm}(\boldsymbol{\theta}_c)-t}{b}\right)|\mathbf{z}_m\right]}$$

for $t > 0$, and $b > 0$. We also define

$$A \equiv \frac{E\left[\frac{1}{b}h'\left(\frac{MR_{jm}(\boldsymbol{\theta}_c)-t}{b}\right)\right]}{E\left[h\left(\frac{MR_{jm}(\boldsymbol{\theta}_c)-t}{b}\right)|\mathbf{z}_m\right]^2}$$

Then,

$$\frac{\partial \widetilde{w}_{jm}(MR_{jm}, t, b)}{\partial MR} = \frac{\frac{1}{b}h'\left(\frac{MR_{jm}-t}{b}\right)}{E\left[h\left(\frac{MR_{jm}(\boldsymbol{\theta}_c)-t}{b}\right)|\mathbf{z}_m\right]} - Ah\left(\frac{MR_{jm}-t}{b}\right)$$

Let

$$\begin{aligned} B &= E\left[h\left(\frac{MR_{jm}(\boldsymbol{\theta}_c)-t}{b}\right)|\mathbf{z}_m\right] \\ \frac{\partial \widetilde{w}_{jm}(MR_{jm}, t, b)}{\partial MR \partial MR} &= \frac{1}{B}\left[\frac{1}{b^2}h''\left(\frac{MR_{jm}-t}{b}\right)\right] - \frac{\frac{1}{b}h'\left(\frac{MR_{jm}-t}{b}\right)}{B^2}\frac{\partial B}{\partial MR} - \frac{\partial}{\partial MR}Ah\left(\frac{MR_{jm}-t}{b}\right) \end{aligned}$$

Then, let

$$\widetilde{f}(\mathbf{y}_m, \boldsymbol{\chi}, t, b) \equiv \frac{1}{J}\sum_{j=1}^{J} f_j(\mathbf{y}_m, \boldsymbol{\chi})\, \widetilde{w}_{jm}(MR_{jm}(\boldsymbol{\theta}_c), t, b), \tag{86}$$

and $Q(\boldsymbol{\chi}) = E\left[\widetilde{f}(\mathbf{y}_m, \boldsymbol{\chi}, t, b)\,|\mathbf{z}_m\right]$, where $\boldsymbol{\chi} = (\boldsymbol{\theta}_c', \boldsymbol{\gamma}')' = \{\chi_n\}_{n=1}^{\infty}$, with

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

where r is the number of parameters in $\boldsymbol{\theta}_c$. Denote, as before, $\boldsymbol{\chi}_0 = (\boldsymbol{\theta}_{c0}', \boldsymbol{\gamma}_0')'$ to be the true parameter vector. In both the SNP logit estimation exercise of Bierens (2014) and ours, the polynomials are functions of parametric functions. In our model, the parameters $\boldsymbol{\theta}_c$ are identified. Let $\boldsymbol{\theta}_c \in \Theta_c$ be compact and let the parameter space $\Xi = \Theta_c \times \Gamma$, $(\boldsymbol{\theta}_c, \boldsymbol{\gamma}) \in \Xi$ satisfy

$$\Xi = \{\times_{n=1}^{\infty}[-\overline{\chi}_n, \overline{\chi}_n]\}$$

where $\overline{\chi}_n$, $n = 1, \ldots$ are a priori chosen positive sequence satisfying $\sum_{n=1}^{\infty} \overline{\chi}_n^2 < \infty$, $sup_{n \geq r} |\chi_n| / \overline{\chi}_n \leq 1$. We also assume that the parameter space is endowed with the metric $d(\boldsymbol{\chi}_1, \boldsymbol{\chi}_2) \equiv \|\boldsymbol{\chi}_1 - \boldsymbol{\chi}_2\|$, where $\|\boldsymbol{\chi}\| = \sqrt{\sum_{k=1}^{\infty} \chi_k^2}$. Let $\boldsymbol{\chi}_0$ be the vector of true parameters. Define also

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

where $k \in \mathbb{N}$, $\Gamma_k(T) = \{\pi_k \boldsymbol{\gamma} : \|\pi_k \boldsymbol{\gamma}\| \leq T\}$, and π_k is the operator that applies to an infinite sequence $\boldsymbol{\gamma} = \{\gamma_n\}_{n=1}^{\infty}$, replacing all the γ_n's for $n > k$ with zeros. Suppose further that $\boldsymbol{\chi}_0 \in \Xi$. Furthermore, since we assume that firms choose prices so that the marginal revenue equals marginal cost, we restrict the parameter space Θ_c so that any $\boldsymbol{\theta}_c \in \Theta_c$ satisfies $MR(\mathbf{X}, \mathbf{p}, \mathbf{s}, j, \boldsymbol{\theta}_c) \geq 0$ for all $(\mathbf{X}, \mathbf{p}, \mathbf{s}, j)$ in the population.

In proving consistency, for the sake of simplicity, we add the following assumptions:

Assumption 10 $(\mathbf{w}_m, \mathbf{X}_m, Q_m, \boldsymbol{\xi}_m, \boldsymbol{\upsilon}_m, \boldsymbol{\varsigma}_m, \boldsymbol{\nu}_m)$, $m = 1, \ldots, M$ *are i.i.d. across markets.*

Define $\widetilde{\mathbf{d}}_m \equiv (Q_m, vec(\mathbf{w}_m)', vec(\boldsymbol{\xi}_m)', vec(\boldsymbol{\upsilon}_m)', vec(\mathbf{X}_m)')$.

Assumption 11 *The support of* $\widetilde{\mathbf{d}}_m$*, denoted as* $\mathcal{D}_d$*, is compact, i.e.,* $\mathcal{D}_d \equiv [\underline{Q}, \overline{Q}] \times [\underline{w}, \overline{w}]^L \times [-\overline{\xi}, \overline{\xi}]^J \times [\underline{\upsilon}, \overline{\upsilon}]^J \times [-\overline{x}, \overline{x}]$.

Assumption 12 *The parameter space* Ξ *is compact. The true parameter* $\boldsymbol{\chi}_0$ *is in the interior of* Ξ.

$\widetilde{f}(\mathbf{y}_m, \boldsymbol{\chi}, \tau, b)$, $\mathbf{y}_m$, $m = 1, \ldots, M$, $\boldsymbol{\chi}_0$ and Ξ satisfy the assumptions that are similar to the Assumption 4.1 of Bierens (2014). Some inequalities are reversed because in our case, our estimator is derived by minimizing the objective function, whereas in Bierens (2014), it is based on maximization.

a $\mathbf{y}_1$, $\mathbf{y}_2$, ..., $\mathbf{y}_M$ are i.i.d. The support of $\mathbf{y}_m$ is contained in an open set $\mathcal{Y}$ of the Euclidean space.

b Ξ is a metric space with metric $d(\boldsymbol{\chi}_1, \boldsymbol{\chi}_2)$.

c For each $\boldsymbol{\chi} \in \Xi$, $\widetilde{f}(\mathbf{y}_m, \boldsymbol{\chi}, \tau, b)$ is a Borel measurable real function of $\mathbf{y}_m$.

d $\widetilde{f}(\mathbf{y}_m, \boldsymbol{\chi}, \tau, b)$ is a.s. continuous in $\boldsymbol{\chi} \in \Xi$.

e There exists a non-negative Borel measurable real function $\underline{f}(\mathbf{y})$ such that $E\left[\underline{f}(\mathbf{y}_m)|\mathbf{z}_m\right] > -\infty$ and $\widetilde{f}(\mathbf{y}_m, \boldsymbol{\chi}, \tau, b) > \underline{f}(\mathbf{y}_m)$ for all $\boldsymbol{\chi} \in \Xi$.

f There exists an element $\boldsymbol{\chi}_0 \in \Xi$ such that $Q(\boldsymbol{\chi}) > Q(\boldsymbol{\chi}_0)$ for all $\boldsymbol{\chi} \in \Xi \backslash \{\boldsymbol{\chi}_0\}$, and $Q(\boldsymbol{\chi}_0) < \infty$.

g There exists an increasing sequence of compact subspaces Ξ_k in Ξ such that $\boldsymbol{\chi}_0 \in \overline{\bigcup_{k=1}^{\infty} \Xi_k} = \overline{\Xi} \subset \Xi$.

h Each sieve space Ξ_k is isomorph to a compact subset of a Euclidean space.

i Each sieve space Ξ_k contains an element $\boldsymbol{\chi}_k$ such that, $lim_{k\to\infty} E\left[\widetilde{f}(\mathbf{y}_m, \boldsymbol{\chi}_k, \tau, b)|\mathbf{z}_m\right] = E\left[\widetilde{f}(\mathbf{y}_m, \boldsymbol{\chi}_0, \tau, b)|\mathbf{z}_m\right]$.

j The set $\Xi_\infty = \left\{\boldsymbol{\chi} \in \Xi : E\left[\widetilde{f}(\mathbf{y}_m, \boldsymbol{\chi}, \tau, b)\right] = \infty|\mathbf{z}_m\right\}$ does not contain an open ball.

Lemma 4 *Assumptions (a)∼(j) hold.*

Proof. First, we show that assumption (a) holds. First, since market size is bounded above by $\overline{Q}$, $q_{jm} \leq \overline{Q}$ for any j, m by definition of market share $s_{jm} = q_{jm}/Q_m < 1$, and thus, $q_{jm} < Q_m \leq \overline{Q}$. From Assumption 5, the variable cost function is continuously differentiable in output. Therefore, $MC(q_{jm}, \mathbf{w}_m, \mathbf{X}_m, \upsilon_{jm})$ is also bounded. Suppose $p_{jm} \to \infty$. Then, if all other prices are bounded, since $\boldsymbol{\xi}_m$ is bounded, $s_{jm} \to Prob(\alpha : \alpha + \eta_{\alpha 0} > 0) = \Phi(\eta_{\alpha 0}) < 1/2$. Therefore, $lim_{p_{jm}\to\infty} p_{jm} s_{jm} \to \infty$. Now, consider for a small $\delta > 0$, $s'_{jm} = \Phi(\eta_{\alpha 0}) + \delta$, and p'_{jm} corresponding to it. Then, because $lim_{p_{jm}\to\infty} s_{jm} = \Phi(\eta_{\alpha 0}) > 0$ and $lim_{p_{jm}\to\infty} p_{jm} s_{jm} \to \infty$, there exists $(\widetilde{s}_{jm}, \widetilde{p}_{jm})$ such that $\widetilde{s}_{jm} < s'_{jm}$ and $\widetilde{p}_{jm} > p'_{jm}$, and $p'_{jm} s'_{jm} - \widetilde{p}_{jm}\widetilde{s}_{jm} < 0$. Therefore, from the mean value theorem, there exists $s^*_{jm} \in \left(\widetilde{s}_{jm}, s'_{jm}\right)$ and the corresponding p^*_{jm} such that

$$\frac{p'_{jm}s'_{jm} - \widetilde{p}_{jm}\widetilde{s}_{jm}}{s'_{jm} - \widetilde{s}_{jm}} = \frac{\partial p^*_{jm}s^*_{jm}}{\partial s_{jm}} < 0.$$

On the other hand, consider $s'_{jm} = \Phi\left(\eta_{\alpha 0}\right) - 2\delta$ for a small $\delta > 0$, and p'_{jm} corresponding to it. Then, because $lim_{p_{jm}\to\infty} s_{jm} = \Phi\left(\eta_{\alpha 0}\right) > 0$ and $lim_{p_{jm}\to\infty} p_{jm}s_{jm} \to \infty$, there exists $(\widetilde{s}_{jm}, \widetilde{p}_{jm})$ such that $\widetilde{s}_{jm} > s'_{jm}$ and $\widetilde{p}_{jm} > p'_{jm}$, and $\widetilde{p}_{jm}\widetilde{s}_{jm} - p'_{jm}s'_{jm} > 0$ and can be made arbitrarily large by choosing arbitrarily high $\widetilde{p}_{jm}$. Therefore,

$$lim_{\widetilde{p}\to\infty}\frac{\widetilde{p}_{jm}\widetilde{s}_{jm} - p'_{jm}s'_{jm}}{\widetilde{s}_{jm} - s'_{jm}} > lim_{\widetilde{p}\to\infty}\frac{\widetilde{p}_{jm}\widetilde{s}_{jm} - p'_{jm}s'_{jm}}{3\delta} = \infty.$$

Therefore, from the mean value theorem, there exists $s^*_{jm} \in \left(s'_{jm}, \widetilde{s}_{jm}\right)$ and the corresponding p^*_{jm} such that

$$lim_{\widetilde{p}\to\infty}\frac{\widetilde{p}_{jm}\widetilde{s}_{jm} - p'_{jm}s'_{jm}}{\widetilde{s}_{jm} - s'_{jm}} = \frac{\partial p^*_{jm}s^*_{jm}}{\partial s_{jm}} = \infty.$$

From the boundedness of the marginal cost function, we conclude that for sufficiently large p, marginal revenue is higher than the marginal cost. We can show that similar results occur if the price goes to infinity for multiple firms as well. Therefore, we conclude that prices are bounded. Furthermore, $s_{jm} \in (0, 1)$. Lastly, $C_{jm} \in (-\infty, \infty)$. Therefore, (a) holds.

Assumption (b) is also satisfied with $d\,()$ being the Euclidean metric. (c), (d) are also satisfied given the Borel measurability and the continuity of the market share function and the deterministic component of the cost $C^v\left(q_{jm}, \mathbf{w}_m, \mathbf{x}_{jm}, \upsilon_{jm}\right) + e\left(q_{jm}, \mathbf{w}_m, \mathbf{x}_{jm}\right)$, and the Borel measurability of the random component of the fixed cost and the measurement error, $\nu_{jm} + \varsigma_{jm}$.

Assumption (e) is satisfied because $\widetilde{f}_j\left(\mathbf{y}, \chi\right) \geq 0$ for any $(\mathbf{y}, \boldsymbol{\chi})$ if we set $\underline{f}\left(\boldsymbol{y}\right) \equiv -1$, from equation 86. Assumption (f) follows from the identification of $\boldsymbol{\chi}_0$ in Proposition 2 as follows:

$$\begin{aligned} & E\left[\left[C_{jm} - \psi\left(q_{jm}, \mathbf{w}_{jm}, \mathbf{x}_{jm}, MR_{jm}\left(\boldsymbol{\theta}_c\right), \boldsymbol{\gamma}\right)\right]^2 \widetilde{w}_{jm}\left(MR_{jm}\left(\boldsymbol{\theta}_c\right), t, b\right)\left|\mathbf{z}_m\right.\right] \\ \geq\ & E\left[\left(\nu_{jm} + \eta_{jm}\right)^2\right] \end{aligned}$$

where strict inequality holds if $(\boldsymbol{\theta}_c, \boldsymbol{\gamma}) \neq (\boldsymbol{\theta}_{c0}, \boldsymbol{\gamma}_0)$ for sufficiently large t, $\overline{\xi}$ and $\overline{x}$, so that our identification proof goes through.

Assumptions (g) and (h) hold from Assumption 9(Equation (33)). Next, we consider Assumption (i). We set $\boldsymbol{\chi}_k$ to be $\boldsymbol{\chi}_k = \pi_k\boldsymbol{\chi}$ and $\chi_l = \theta_{c0l}$ for $l = 1, \ldots, p$. Furthermore, for

$k \geq p$,

$$
\begin{aligned}
& E\left[\widetilde{f}_j\left(\mathbf{y}_m, \boldsymbol{\chi}_{0k}, t, b\right)|\mathbf{z}_m\right] \\
\leq\ & E\left[v_{jm}^2+\eta_{jm}^2\right]+E\left[\left(\sum_{l=k+1}^{\infty} \psi_l\left(q_{jm}, \mathbf{w}_m, \mathbf{x}_{jm}, MR_{jm}\left(\boldsymbol{\theta}_{c0}\right)\right) \gamma_{0l}\right)^2 w_{jm}\left(MR_{jm}\left(\boldsymbol{\theta}_{c0}\right), t, b\right)|\mathbf{z}_m\right] \\
\leq\ & E\left[v_{jm}^2+\eta_{jm}^2\right]+E\left[\sup_{l \geq k+1} \psi_l\left(q_{jm}, \mathbf{w}_m, \mathbf{x}_{jm}, MR_{jm}\left(\boldsymbol{\theta}_0\right)\right)^2 w_{jm}\left(MR_{jm}\left(\boldsymbol{\theta}_{c0}\right), t, b\right)|\mathbf{z}_m\right]\left(\sum_{l=k+1}^{\infty}\left|\gamma_{0l}\right|\right)^2 \\
\leq\ & E\left[v_{jm}^2+\eta_{jm}^2\right]+E\left[\sup_{l \geq p+1} \psi_l\left(q_{jm}, \mathbf{w}_m, \mathbf{x}_{jm}, MR_{jm}\left(\boldsymbol{\theta}_0\right)\right)^2 w_{jm}\left(MR_{jm}\left(\boldsymbol{\theta}_{c0}\right), t, b\right)|\mathbf{z}_m\right]\left(\sum_{l=p+1}^{\infty}\left|\gamma_{0l}\right|\right)^2
\end{aligned}
$$

and the RHS is uniformly bounded. Therefore, from the Dominated Convergence theorem, $E\left[\widetilde{f}_j\left(\mathbf{y}_m, \boldsymbol{\chi}_{0k}, t, b\right)|\mathbf{z}_m\right] \rightarrow E\left[\widetilde{f}_j\left(\mathbf{y}_m, \boldsymbol{\chi}_0, t, b\right)|\mathbf{z}_m\right]$ as $k \rightarrow \infty$. Hence, (i) is satisfied.

Furthermore, by construction, $\widetilde{f}\left(\mathbf{y}_m, \boldsymbol{\chi}_k, t, b\right)$ is uniformly bounded, and hence,

$$
\Xi_{\infty}=\left\{\boldsymbol{\chi} \in \Xi: E\left[\widetilde{f}\left(\mathbf{y}_m, \boldsymbol{\chi}, t, b\right)|\mathbf{z}_m\right]=\infty\right\}
$$

is an empty set, and assumption (j) is satisfied. ■

Therefore, the assumptions that are equivalent of Assumption 4.1 of Bierens (2014) are satisfied. Then, consider the ϵ-neighborhood of a parameter vector $\boldsymbol{\chi}_* \in \Xi$. Then, because $\mathbf{y}_m$ is i.i.d, for $\boldsymbol{\chi}$ in the ϵ neighborhood of $\boldsymbol{\chi}_*$, the random variable $inf_{\boldsymbol{\chi} \in \Xi, d(\boldsymbol{\chi}_*, \boldsymbol{\chi})<\epsilon} \widetilde{f}_j\left(\mathbf{y}_m, \boldsymbol{\chi}, t, b\right)$ is also i.i.d. Thus, we can use the Kolmogorov's LLN. To do so, we need to show that one can take the expectation of $inf_{\boldsymbol{\chi}_* \in \Xi, d(\boldsymbol{\chi}, \boldsymbol{\chi}_*)<\epsilon} \widetilde{f}_j\left(\mathbf{y}_m, \boldsymbol{\chi}_*, t, b\right)$ over $\mathbf{y}_m$, i.e., that it is integrable over $\mathbf{y}_m$. Since, by construction, $\widetilde{f}\left(\mathbf{y}_m, \boldsymbol{\chi}_k, t, b\right)$ is nonnegative and uniformly bounded, given any $\boldsymbol{\chi} \in \Xi$, for any $\epsilon>0$,

$$
E\left[\left|inf_{\boldsymbol{\chi}_* \in \Xi, d(\boldsymbol{\chi}, \boldsymbol{\chi}_*)<\epsilon} \widetilde{f}_j\left(\mathbf{y}_m, \boldsymbol{\chi}_*, t, b\right)\right||\mathbf{z}_m\right]<\infty
$$

Hence, from Kolmogorov's LLN,

$$
\frac{1}{M} \sum_{m=1}^M inf_{\boldsymbol{\chi}_* \in \Xi, d(\boldsymbol{\chi}, \boldsymbol{\chi}_*)<\epsilon} \widetilde{f}_j\left(\mathbf{y}_m, \boldsymbol{\chi}_*, t, b\right) \rightarrow_{a.s.} E\left[inf_{\boldsymbol{\chi}_* \in \Xi, d(\boldsymbol{\chi}, \boldsymbol{\chi}_*)<\epsilon} \widetilde{f}_j\left(\mathbf{y}_m, \boldsymbol{\chi}_*, t, b\right)|\mathbf{z}_m\right] \quad (87)
$$

as $M \rightarrow \infty$. Now, let

$$
f\left(\mathbf{y}_m, \boldsymbol{\chi}, t, b\right) \equiv \frac{1}{J} \sum_{j=1}^J f_j\left(\mathbf{y}_m, \boldsymbol{\chi}\right) w_{jm}\left(MR_{jm}\left(\boldsymbol{\theta}_c\right), t, b\right)
$$

where

$$\begin{aligned} w_{jm}\left(MR_{jm}\left(\boldsymbol{\theta}_c\right), t, b\right) &\equiv \frac{h\left(\frac{MR_{jm}(\boldsymbol{\theta}_c)-t}{b}\right)}{[MJ]^{-1}\sum_{m=1}^{M}\sum_{j=1}^{J} h\left(\frac{MR_{jm}(\boldsymbol{\theta}_c)-t}{b}\right)} \\ &= \frac{E\left[h\left(\frac{MR_{jm}(\boldsymbol{\theta}_c)-t}{b}\right)|\mathbf{z}_m\right]}{[MJ]^{-1}\sum_{m=1}^{M}\sum_{j=1}^{J} h\left(\frac{MR_{jm}(\boldsymbol{\theta}_c)-t}{b}\right)} \times \widetilde{w}_{jm}\left(MR_{jm}\left(\boldsymbol{\theta}_c\right), t, b\right) \end{aligned}$$

Now, let

$$\widehat{E}_M\left[h\left(\frac{MR_{jm}\left(\boldsymbol{\theta}_c\right)-t}{b}\right)\right] \equiv \frac{1}{MJ}\sum_{m=1}^{M}\sum_{j=1}^{J} h\left(\frac{MR_{jm}\left(\boldsymbol{\theta}_c\right)-t}{b}\right)$$

and let

$$A_M \equiv \frac{E\left[h\left(\frac{MR_{jm}(\boldsymbol{\theta}_c)-t}{b}\right)|\mathbf{z}_m\right]}{\widehat{E}_M\left[h\left(\frac{MR_{jm}(\boldsymbol{\theta}_c)-t}{b}\right)\right]}.$$

By applying the Kolmogorov Law of Large Numbers, as $M \to \infty$,

$$\widehat{E}_M\left[h\left(\frac{MR_{jm}\left(\boldsymbol{\theta}_c\right)-t}{b}\right)\right] \overset{a.s.}{\to} E\left[h\left(\frac{MR_{jm}\left(\boldsymbol{\theta}_c\right)-t}{b}\right)|\mathbf{z}_m\right]$$

For sufficiently large $t > 0$, $E\left[h\left(\frac{MR_{jm}(\boldsymbol{\theta}_c)-t}{b}\right)|\mathbf{z}_m\right] > 0$, thus,

$$A_M \overset{a.s.}{\to} 1 \tag{88}$$

Then, let

$$f\left(\mathbf{y}_m, \boldsymbol{\chi}, t, b\right) \equiv A_M \times \widetilde{f}\left(\mathbf{y}_m, \boldsymbol{\chi}, t, b\right)$$

Therefore, from Equation (87) and (88), for any $\epsilon > 0$

$$\frac{1}{M}\sum_{m=1}^{M} \inf_{\boldsymbol{\chi}_* \in \boldsymbol{\Theta}_c \times \boldsymbol{\Gamma}, d(\boldsymbol{\chi}, \boldsymbol{\chi}_*) < \epsilon} f\left(\mathbf{y}_m, \boldsymbol{\chi}_*, t, b\right) \overset{a.s.}{\to} E\left[\inf_{\boldsymbol{\chi}_* \in \boldsymbol{\Theta}_c \times \boldsymbol{\Gamma}, d(\boldsymbol{\chi}, \boldsymbol{\chi}_*) < \epsilon} f\left(\mathbf{y}_m, \boldsymbol{\chi}_*, t, b\right)|\mathbf{z}_m\right].$$

Therefore, Theorem 4.1 of Bierens holds. That is, $plim_{M\to\infty} d\left(\boldsymbol{\chi}_M, \boldsymbol{\chi}_0\right) = 0$ if and only if for any compact subset Ξ_c of Ξ with $\boldsymbol{\chi}_0$ in its interior, $lim_{M\to\infty} Pr\left(\boldsymbol{\chi}_{bM} \in \Xi_c\right) = 1$.

Next, we explain how we set up the parameter space so that Assumption 4.2 of Bierens (2014) is satisfied. That is,

Assumption 13 *(Assumption 4.2, Bierens (2014)) Either*

(a) $\overline{\Xi} = \overline{\cup_{n=1}^{\infty}\Xi_n}$ *is compact itself, or*

(b) *There exists a compact set* Ξ_c *containing* $\boldsymbol{\chi}_0$ *such that* $Q(\boldsymbol{\chi}_0) < E\left[\inf_{\boldsymbol{\chi}\in\overline{\Xi}\backslash\Xi_c} \widetilde{f}(\mathbf{y}_m, \boldsymbol{\chi}, \tau, b)\,|\mathbf{z}_m\right] < \infty$.

Assumption 12 and (g) guarantee (a). Then, we can prove consistency using an equivalent of Theorem 4.2 of Bierens (2014). That is,

Theorem 1 *(Theorem 4.2, Bierens (2014)) Under assumptions that are equivalent to the Assumptions 4.1 and 4.2 of Bierens (2014),* $plim_{M\to\infty} d(\boldsymbol{\chi}_M, \boldsymbol{\chi}_0) = 0$.

Therefore, we proved consistency of our estimator.

Next, we prove the asymptotic normality of our estimator. First, we prove that $\partial A_M/\partial\theta_{cl} \overset{a.s.}{\to} 0$ and $\partial^2 A_M/\partial\theta_{ck}\partial\theta_{cl} \overset{a.s.}{\to} 0$. Now, let

$$a_M(\boldsymbol{\theta}_c) \equiv log\left(E\left[h\left(\frac{MR_{jm}(\boldsymbol{\theta}_c) - t}{b}\right)|\mathbf{z}_m\right]\right) - log\left(\widehat{E}_M\left[h\left(\frac{MR_{jm}(\boldsymbol{\theta}_c) - t}{b}\right)\right]\right).$$

Furthermore, since $h\left(\frac{MR_{jm}(\boldsymbol{\theta}_c)-t}{b}\right)$ satisfies the conditions required for interchanging the integration and derivatives, we do so and apply the Kolmogorov's Law of Large Numbers to derive

$$\widehat{E}_M\left[\frac{\partial}{\partial\theta_{cl}}h\left(\frac{MR_{jm}(\boldsymbol{\theta}_c) - t}{b}\right)\right] \overset{a.s.}{\to} E\left[\frac{\partial}{\partial\theta_{cl}}h\left(\frac{MR_{jm}(\boldsymbol{\theta}_c) - t}{b}\right)|\mathbf{z}_m\right],\ l = 1,\ldots,p$$

$$\frac{\partial}{\partial\theta_{cl}}a_M(\boldsymbol{\theta}_c) = \frac{E\left[\frac{\partial}{\partial\theta_{cl}}h\left(\frac{MR_{jm}(\boldsymbol{\theta}_c)-t}{b}\right)|\mathbf{z}_m\right]}{E\left[h\left(\frac{MR_{jm}(\boldsymbol{\theta}_c)-t}{b}\right)|\mathbf{z}_m\right]} - \frac{\widehat{E}_M\left[\frac{\partial}{\partial\theta_{cl}}h\left(\frac{MR_{jm}(\boldsymbol{\theta}_c)-t}{b}\right)\right]}{\widehat{E}_M\left[h\left(\frac{MR_{jm}(\boldsymbol{\theta}_c)-t}{b}\right)\right]} \overset{a.s.}{\to} 0$$

Since,

$$\frac{\partial}{\partial\theta_{cl}}a_M(\boldsymbol{\theta}_c) = \frac{1}{A_M}\frac{\partial A_M}{\partial\theta_{cl}},\ \textit{and}\ A_M \overset{a.s.}{\to} 1,$$

$$\frac{\partial A_M}{\partial\theta_{cl}} \overset{a.s.}{\to} 0$$

By the same logic, it is straightforward to show

$$\frac{\partial}{\partial\theta_{ck}\theta_{cl}}a_M(\boldsymbol{\theta}_c) \quad \overset{a.s.}{\to} \quad 0,$$

and thus,

$$\frac{\partial^2 A_M}{\partial\theta_{ck}\partial\theta_{cl}} \overset{a.s.}{\to} 0$$

We next discuss how we either impose or derive Assumption 6.1 of Bierens (2014).

Assumption 14 *(Assumption 6.1, Bierens (2014))*

(a) *Parameter space Ξ is endowed with the norm*

$$\|\boldsymbol{\chi}\|_r = \sum_{n=1}^{\infty} n^r |\chi_n|$$

and the associated metric $d(\boldsymbol{\chi}_1, \boldsymbol{\chi}_2) = \|\boldsymbol{\chi}_1 - \boldsymbol{\chi}_2\|_r$.

(b) *The true parameter $\boldsymbol{\chi}_0 = \{\chi_{0,n}\}_{n=1}^{\infty}$ satisfies $\|\boldsymbol{\chi}_0\|_r < \infty$.*

(c) *There exists $k \in \mathbb{N}$ such that for any $n \geq k$ $\boldsymbol{\chi}_{0,n} = \pi_n \boldsymbol{\chi}_0 \in \Xi_n^{Int}$, where Ξ_n^{Int} is the interior of the sieve space Ξ_n.*

(d) *$f(\mathbf{y}_m, \boldsymbol{\chi}, \tau, b)$ is a.s. twice continuously differentiable in an open neighborhood of $\boldsymbol{\chi}_0$.*

We assume (a) and (b). For (c), as explained earlier, we follow Bierens (2014) and construct the compact parameter space so that (c) is satisfied. The logit and BLP marginal revenues are twice differentiable and the cost function is also assumed to be twice continuously differentiable. Furthermore, $\widetilde{w}_{jm}(MR_{jm}, t, b)$ is also constructed to be twice continuously differentiable. Therefore, (d) holds.

It is straightforward to show that Assumption 6.2 of Bierens (2014) is satisfied given Assumption 14 and the earlier consistency proof. That is,

Assumption 15 *For any subsequence $k = k_M$ of the sample size M satisfying $k_M \to \infty$ as $M \to \infty$, $plim_{M\to\infty} ||\boldsymbol{\chi}_{k_M} - \boldsymbol{\chi}_0||_r = 0$.*

Then, Bierens (2014) proves in Lemma 6.1 that there exists a subsequence $K_n \leq n$, $\lim_{n\to\infty} K_n = \infty$ such that

$$\lim_{M\to\infty} Pr\left[\sum_{m=1}^{M} \nabla_k \widetilde{f}\left(\mathbf{y}_m, \boldsymbol{\chi}_{n_M}, t, b\right) = 0 \; for \; k = 1, \ldots, K_{n_M}\right] = 1$$

Furthermore, we also assume:

Assumption 16 *(a) :* $E\left[(\nu + \zeta)^4\right] < \infty$

First, we set $r=2$. Then, we prove that Assumption 6.3 of Bierens (2014) is satisfied. That is,

Lemma 5 *(Assumption 6.3, Bierens (2014))There exists a nonnegative integer $r_0 < r$ such that the following local Lipschitz conditions hold for all positive integer $l \in \mathbb{N}$ we have*

$$E\left\|\nabla_l \widetilde{f}\left(\mathbf{y},\boldsymbol{\chi}_0,t,b\right)-\nabla_l \widetilde{f}\left(\mathbf{y},\boldsymbol{\chi}_{0,k},t,b\right)|\mathbf{z}_m\right\| \leq C_l \left\|\boldsymbol{\chi}_0-\boldsymbol{\chi}_{0,k}\right\|_{r_0}$$

where $\nabla_l \widetilde{f}\left(\mathbf{y}_m,\boldsymbol{\chi}_0,t,b\right)=\partial \widetilde{f}\left(\mathbf{y}_m,\boldsymbol{\chi}_0,t,b\right)/\partial \boldsymbol{\chi}_{0,l}$, $\sum_{l=1}^{\infty} 2^{-l} C_l < \infty$ and the sieve order $k=k_M$ is chosen such that

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

Proof. We choose $k_M > p$ for any $M>0$.

$$\begin{aligned}
\widetilde{f}_j\left(\mathbf{y}_m,\boldsymbol{\chi},t,b\right) &= f_j\left(\mathbf{y}_m,\boldsymbol{\theta}_c\right)\widetilde{w}_{jm}\left(MR_{jm}\left(\boldsymbol{\theta}_c\right),t,b\right)\\
&= \left[C_{jm}-\psi\left(q_{jm},\mathbf{w}_{jm},\mathbf{x}_{jm},MR_{jm}\left(\boldsymbol{\theta}_c\right),\boldsymbol{\gamma}\right)\right]^2\widetilde{w}_{jm}\left(MR_{jm}\left(\boldsymbol{\theta}_c\right),t,b\right)
\end{aligned}$$

For $l>p$, $k>p$,

$$E\left[\left|\nabla_{l,k}\widetilde{f}\left(\mathbf{y}_m,\boldsymbol{\chi},t,b\right)\right||\mathbf{z}_m\right]=E\left[2\left|\psi_{l-p}\psi_{k-p}\widetilde{w}_{jm}\left(MR_{jm}\left(\boldsymbol{\theta}_c\right),t,b\right)\right||\mathbf{z}_m\right]<B.$$

Therefore, from the mean value theorem,

$$E\left\|\nabla_l \widetilde{f}\left(\mathbf{y},\boldsymbol{\chi}_0,t,b\right)-\nabla_l \widetilde{f}\left(\mathbf{y},\boldsymbol{\chi}_{0,k},t,b\right)|\mathbf{z}_m\right\| \leq B\sum_{n=k+1}^{\infty}\left|\chi_{0n}\right|=B\left\|\boldsymbol{\chi}_0-\boldsymbol{\chi}_{0,k}\right\|_{r_0}$$

and claim holds.

For $k\leq p$, $l>p$,

$$\begin{aligned}
\nabla_k \widetilde{f}_j\left(\mathbf{y}_m,\boldsymbol{\chi},t,b\right) &= -2\left[\nu_{jm}+\eta_{jm}+\psi\left(q_{jm},\mathbf{w}_{jm},\mathbf{x}_{jm},MR_{jm}\left(\boldsymbol{\theta}_{c0}\right)\gamma_0\right)-\psi\left(q_{jm},\mathbf{w}_{jm},\mathbf{x}_{jm},MR_{jm}\left(\boldsymbol{\theta}_c\right),\boldsymbol{\gamma}\right)\right]\\
&\quad\times\left[\psi_{MR}\left(q_{jm},\mathbf{w}_{jm},\mathbf{x}_{jm},MR_{jm}\left(\boldsymbol{\theta}_c\right),\boldsymbol{\gamma}\right)\widetilde{w}_{jm}\left(MR_{jm}\left(\boldsymbol{\theta}_c\right),t,b\right)\right.\\
&\quad\left.+\psi\left(q_{jm},\mathbf{w}_{jm},\mathbf{x}_{jm},MR_{jm}\left(\boldsymbol{\theta}_c\right),\boldsymbol{\gamma}\right)\frac{\partial}{\partial MR}\widetilde{w}_{jm}\left(MR_{jm}\left(\boldsymbol{\theta}_c\right),t,b\right)\right]\times\frac{\partial MR_{jm}\left(\boldsymbol{\theta}_c\right)}{\partial\theta_{ck}}\\
&\equiv -2A_{k1}\times A_{k2}\times\frac{\partial MR_{jm}\left(\boldsymbol{\theta}_c\right)}{\partial\theta_{ck}}
\end{aligned}$$

$$\nabla_{k,l}\widetilde{f}_j\left(\mathbf{y}_m,\boldsymbol{\chi},t,b\right)=2\psi_{l-p}\times A_{k2}\times\frac{\partial MR_{jm}\left(\boldsymbol{\theta}_c\right)}{\partial\theta_{ck}}-2A_{k1}\times\frac{\partial A_2}{\partial\gamma_{l-p}}\times\frac{\partial MR_{jm}\left(\boldsymbol{\theta}_c\right)}{\partial\theta_{ck}}$$

$$\frac{\partial A_2}{\partial \gamma_{l-p}} = \psi_{MR,l-p}\widetilde{w}_{jm} + \psi_{l-p}\frac{\partial}{\partial MR}\widetilde{w}_{jm}\left(MR_{jm}\left(\boldsymbol{\theta}_c\right), t, b\right)$$

Since $\psi\left(, MR_{jm}\left(\boldsymbol{\theta}_{c0}\right)\right)$, $\psi\left(, MR_{jm}\left(\boldsymbol{\theta}_c\right), \boldsymbol{\gamma}\right)$, $\psi_{MR}\left(MR_{jm}\left(\boldsymbol{\theta}_c\right), \boldsymbol{\gamma}\right)$, $\psi_{MR,l-p}\left(, MR_{jm}\left(\boldsymbol{\theta}_c\right), \boldsymbol{\gamma}\right)$ and $\partial\widetilde{w}_{jm}\left(MR_{jm}\left(\boldsymbol{\theta}_c\right), t, b\right)/\partial MR$, $\partial MR_{jm}\left(\boldsymbol{\theta}_c\right)/\partial\theta_{cl}$ are uniformly bounded over the compact domain,

$$E\left[\nabla_k f\left(\mathbf{y}_m, \boldsymbol{\chi}, t, b\right)|\mathbf{z}_m\right] < B$$

for sufficiently large $B > 0$.

■

Next, Assumption 6.4 of Bierens (2014) requires that for all $k \in \mathbb{N}$, $E\left[\nabla_k \widetilde{f}\left(\mathbf{y}, \boldsymbol{\chi}_0, t, b\right)|\mathbf{z}_m\right] = 0$. This holds because of the F.O.C. for the parameter value $\boldsymbol{\chi}_0$.

We show next that Assumption 6.5 of Bierens (2014): $\sum_{j=1}^{\infty} j2^{-j}E\left[\left(\nabla_j \widetilde{f}\left(\mathbf{y}, \boldsymbol{\chi}_0, t, b\right)\right)^2|\mathbf{z}_m\right] < \infty$, holds. From the twice continuous differentiability of the pseudo-cost function $\psi\left(\right)$ and the marginal revenue function, we have for $l \leq p$,

$$\begin{aligned}\nabla_l \widetilde{f}_j\left(\mathbf{y}_m, \boldsymbol{\chi}, t, b\right) &= -2\left[\nu_{jm} + \eta_{jm} + \psi\left(q_{jm}, \mathbf{w}_{jm}, \mathbf{x}_{jm}, MR_{jm}\left(\boldsymbol{\theta}_{c0}\right)\boldsymbol{\gamma}_0\right) - \psi\left(q_{jm}, \mathbf{w}_{jm}, \mathbf{x}_{jm}, MR_{jm}\left(\boldsymbol{\theta}_c\right), \boldsymbol{\gamma}\right)\right] \\ &\quad \times\left[\psi_{MR}\left(q_{jm}, \mathbf{w}_{jm}, \mathbf{x}_{jm}, MR_{jm}\left(\boldsymbol{\theta}_c\right), \boldsymbol{\gamma}\right)\widetilde{w}_{jm}\left(MR_{jm}\left(\boldsymbol{\theta}_c\right), t, b\right)\right. \\ &\quad \left.+\psi\left(q_{jm}, \mathbf{w}_{jm}, \mathbf{x}_{jm}, MR_{jm}\left(\boldsymbol{\theta}_c\right), \boldsymbol{\gamma}\right)\frac{\partial}{\partial MR}\widetilde{w}_{jm}\left(MR_{jm}\left(\boldsymbol{\theta}_c\right), t, b\right)\right]\frac{\partial MR_{jm}\left(\boldsymbol{\theta}_c\right)}{\partial\theta_{cl}}.\end{aligned}$$

Since $\varphi\left(, MR_{jm}\left(\boldsymbol{\theta}_{c0}\right)\right)$, $\psi\left(, MR_{jm}\left(\boldsymbol{\theta}_c\right), \boldsymbol{\gamma}\right)$, $\psi_{MR}\left(q_{jm}, \mathbf{w}_{jm}, \mathbf{x}_{jm}, MR_{jm}\left(\boldsymbol{\theta}_c\right), \boldsymbol{\gamma}\right)$ and $\partial\widetilde{w}_{jm}\left(MR_{jm}\left(\boldsymbol{\theta}_c\right), t, b\right)/\partial MR$, $\partial MR_{jm}\left(\boldsymbol{\theta}_c\right)/\partial\theta_{cl}$ are uniformly bounded over the compact domain,

$$E\left[\left[\nabla_l f\left(\mathbf{y}_m, \boldsymbol{\chi}, t, b\right)\right]^2|\mathbf{z}_m\right] < B$$

for sufficiently large $B > 0$. Furthermore, for $l > p$,

$$\begin{aligned}\nabla_l \widetilde{f}\left(\mathbf{y}_m, \boldsymbol{\chi}, t, b\right) &= -2\left[\nu_{jm} + \eta_{jm} + \varphi\left(, MR_{jm}\left(\boldsymbol{\theta}_{c0}\right)\right) - \varphi\left(q_{jm}, \mathbf{w}_{jm}, \mathbf{x}_{jm}, MR_{jm}\left(\boldsymbol{\theta}_c\right), \boldsymbol{\gamma}\right)\right] \\ &\quad \times\psi_{l-p}\left(q_{jm}, \mathbf{w}_{jm}, \mathbf{x}_{jm}, MR_{jm}\left(\boldsymbol{\theta}_{cb}\right), \boldsymbol{\gamma}\right)\widetilde{w}_{jm}\left(MR_{jm}\left(\boldsymbol{\theta}_c\right), t, b\right)\end{aligned}$$

Since $\varphi\left(, MR_{jm}\left(\boldsymbol{\theta}_{c0}\right)\right)$, $\varphi\left(q_{jm}, \mathbf{w}_{jm}, \mathbf{x}_{jm}, MR_{jm}\left(\boldsymbol{\theta}_c\right), \boldsymbol{\gamma}\right)$, $\psi_{l-p}\left(q_{jm}, \mathbf{w}_{jm}, \mathbf{x}_{jm}, MR_{jm}\left(\boldsymbol{\theta}_{cb}\right), \boldsymbol{\gamma}\right)$ and $\widetilde{w}_{jm}\left(MR_{jm}\left(\boldsymbol{\theta}_c\right), t, b\right)$ are uniformly bounded over the compact domain,

$$E\left[\left(\nabla_l \widetilde{f}\left(\mathbf{y}_m, \boldsymbol{\chi}, t, b\right)\right)^2|\mathbf{z}_m\right] < B$$

for sufficiently large $B > 0$. Therefore, $E\left[\left(\nabla_l \widetilde{f}\left(\mathbf{y}_m, \boldsymbol{\chi}, t, b\right)\right)^2 |\mathbf{z}_m\right] < B$. Integrating over $\mathbf{z}_m$, we can show that the claim is satisfied.

Next, we prove that (a)-(c) of Assumption 6.6 of Bierens (2014) is satisfied.

Lemma 6 *(Assumption 6.6, Bierens (2014)) For some $\tau > 0$,*

(a) $\sum_{j=1}^{\infty}\sum_{k=1}^{\infty}(jk)^{-2-\tau} E\left[\left|\nabla_{j,k}\widetilde{f}\left(\mathbf{y}, \boldsymbol{\chi}_0, t, b\right)\right| |\mathbf{z}_m\right] < \infty$, *where* $\nabla_{j,k}\widetilde{f}\left(\mathbf{y}_m, \boldsymbol{\chi}_0, t, b\right) = \partial^2 \widetilde{f}\left(\mathbf{y}_m, \boldsymbol{\chi}_0, t, b\right) / \left(\partial \boldsymbol{\chi}_{0,j} \partial \boldsymbol{\chi}_{0,k}\right)$.

(b) $\lim_{\epsilon\downarrow 0}\sum_{j=1}^{\infty}\sum_{k=1}^{\infty}(jk)^{-2-\tau} E\left[sup_{\|\boldsymbol{\chi}-\boldsymbol{\chi}_0\|_r\leq\epsilon}\left|\nabla_{j,k}\widetilde{f}\left(\mathbf{y}, \boldsymbol{\chi}, t, b\right) - \nabla_{j,k}\widetilde{f}\left(\mathbf{y}, \boldsymbol{\chi}_0, t, b\right)\right| |\mathbf{z}_m\right] = 0$.

(c) *For at least one pair of positive integers l, n,* $E\left[\nabla_{l,p+n}\widetilde{f}\left(\mathbf{y}, \boldsymbol{\chi}_0, t, b\right)|\mathbf{z}_m\right] \neq 0$.

(a): For $j > p$, $k > p$,

$$E\left[\left|\nabla_{j,k}\widetilde{f}\left(\mathbf{y}_m, \boldsymbol{\chi}, t, b\right)\right| |\mathbf{z}_m\right] = E\left[2\left|\psi_{j-p}\psi_{k-p}\widetilde{w}_{jm}\left(MR_{jm}\left(\boldsymbol{\theta}_c\right), t, b\right)\right| |\mathbf{z}_m\right] < B$$

for sufficiently large $B > 0$.

For $k \leq p$, $l > p$,

$$E\left[\left|\nabla_{k,l}\widetilde{f}_j\left(\mathbf{y}_m, \boldsymbol{\chi}, t, b\right)\right| |\mathbf{z}_m\right] = 2E\left[\left|A_{kl,j}\right| |\mathbf{z}_m\right]$$

where

$$A_{kl,j} = \psi_{l-p} \times A_{k2} \times \frac{\partial MR\left(\boldsymbol{\theta}_c\right)}{\partial \theta_{ck}} - A_{k1} \times \frac{\partial A_{k2}}{\partial \gamma_{l-p}} \times \frac{\partial MR\left(\boldsymbol{\theta}_{cb}\right)}{\partial \theta_{ck}}$$

From earlier arguments, it follows that,

$$E\left[\left|\nabla_{k,l}\widetilde{f}_j\left(\mathbf{y}_m, \boldsymbol{\chi}, t, b\right)\right| |\mathbf{z}_m\right] = 2E\left[\left|A_{kl,j}\right| |\mathbf{z}_m\right] < B$$

Finally, for $k \leq p$, $l \leq p$,

$$A_{kl} = \left[\frac{\partial A_{k1}}{\partial MR} A_{k2} A_{k3} + A_{k1}\frac{\partial A_{k2}}{\partial MR} A_{k3}\right]\frac{\partial MR\left(\boldsymbol{\theta}_c\right)}{\partial \theta_{cl}} + A_{k1}A_{k2}\frac{\partial A_{k3}}{\partial \theta_{cl}}$$

where, given $\boldsymbol{\theta}_{c0}$, and thus, $MR_{jm}\left(\boldsymbol{\theta}_{c0}\right)$ being fixed,

$$\frac{\partial A_{k1}}{\partial MR} = -\psi_{MR}\left(q_{jm}, \mathbf{w}_{jm}, \mathbf{x}_{jm}, MR_{jm}\left(\boldsymbol{\theta}_c\right), \boldsymbol{\gamma}\right)$$

$$\frac{\partial A_{k2}}{\partial MR} = \psi_{MR,MR}\left(, MR_{jm}\left(\boldsymbol{\theta}_c\right), \boldsymbol{\gamma}\right)\widetilde{w}_{jm}\left(\right) + 2\psi_{MR}\left(, MR_{jm}\left(\boldsymbol{\theta}_c\right), \boldsymbol{\gamma}\right)\frac{\partial}{\partial MR}\widetilde{w}_{jm}\left(MR_{jm}\left(\boldsymbol{\theta}_c\right), t, b\right)$$
$$+\psi\left(, MR_{jm}\left(\boldsymbol{\theta}_c\right), \boldsymbol{\gamma}\right)\frac{\partial^2}{\partial MR\partial MR}\widetilde{w}_{jm}\left(MR_{jm}\left(\boldsymbol{\theta}_c\right), t, b\right)$$

$$A_{k3} = \frac{\partial MR\left(\boldsymbol{\theta}_c\right)}{\partial \theta_{cj}}, \ \frac{\partial A_{k3}}{\partial \theta_{cl}} = \frac{\partial^2 MR\left(\boldsymbol{\theta}_c\right)}{\partial \theta_{ck}\partial \theta_{cl}}$$

Once again, by earlier arguments, it follows that,

$$E\left[\left|\nabla_{k,l}\widetilde{f}\left(\mathbf{y}_m, \boldsymbol{\chi}, t, b\right)\right| |\mathbf{z}_m\right] = 2E\left[|B|\right] < \infty.$$

Therefore, (a) holds.

Note that all the terms of $\nabla_{k,l}f\left(\mathbf{y}_m, \tau, b, \boldsymbol{\chi}\right)$ are continuous for $\|\boldsymbol{\chi} - \boldsymbol{\chi}_0\|_r \leq \epsilon$ for sufficiently small $\epsilon > 0$. Furthermore, it is uniformly continuous because the parameters belong to a compact set. Therefore, (b) holds.

(c) For $k > p$,

$$\nabla_k\widetilde{f}_j\left(\mathbf{y}_m, t, b, \boldsymbol{\chi}_c\right) = -2\left[C_{jm} - \psi\left(q_{jm}, \mathbf{w}_{jm}, \mathbf{x}_{jm}, MR_{jm}\left(\boldsymbol{\theta}_c\right), \boldsymbol{\gamma}\right)\right]$$

$$\times\psi_{k-p}\left(q_{jm}, \mathbf{w}_{jm}, \mathbf{x}_{jm}, MR_{jm}\left(\boldsymbol{\theta}_c\right)\right)\widetilde{w}_{jm}\left(MR_{jm}\left(\boldsymbol{\theta}_c\right), t, b\right) \tag{89}$$

Suppose that for $k > p$ and $k = l$, from Equation (89),

$$E\left[\nabla_{k,l}\widetilde{f}_j\left(\mathbf{y}_m, \boldsymbol{\chi}, t, b\right)|\mathbf{z}_m\right] = 2\psi^2_{k-p}\widetilde{w}_{jm}\left(MR_{jm}\left(\boldsymbol{\theta}_c\right), t, b\right) \neq 0$$

Therefore, (c) holds.

Now, let $\eta_k\left(u\right)$s be orthogonal weight functions on $[0, 1]$ such that $\sigma_k = \int_0^1 \eta_k\left(u\right)^2 du$ converges fast enough to zero as k goes to infinity. Further, n in $\boldsymbol{\chi}_n$ denotes the number of parameters, including the coefficients on sieve polynomials, so that $\chi_k = 0$ for all $k > n$. K_n used below is the subsequence defined in Lemma 6.1. of Bierens (2014).

$$\begin{aligned}
\widetilde{U}(u) &\equiv \left[\frac{1}{\sqrt{M}}\sum_{m=1}^{M}\widetilde{f}(\mathbf{y}_m,\boldsymbol{\chi}_0,t,b)\right]\eta_k(u), \quad \widehat{U}(u) \equiv \left[\frac{1}{\sqrt{M}}\sum_{m=1}^{M}f(\mathbf{y}_m,\boldsymbol{\chi}_0,t,b)\right]\eta_k(u) = A_M\widetilde{U}(u) \\
\widetilde{W}_n(u) &\equiv \sum_{k=1}^{K_n}\left[\frac{1}{\sqrt{M}}\sum_{m=1}^{M}\nabla_k\widetilde{f}(\mathbf{y}_m,\boldsymbol{\chi}_n,t,b)\right]\eta_k(u), \\
\widehat{W}_n(u) &\equiv \sum_{k=1}^{K_n}\left[\frac{1}{\sqrt{M}}\sum_{m=1}^{M}\nabla_k f(\mathbf{y}_m,\boldsymbol{\chi}_n,t,b)\right]\eta_k(u) \\
\widetilde{V}_n(u) &\equiv \sum_{k=1}^{K_n}\left[\frac{1}{\sqrt{M}}\sum_{m=1}^{M}\left(\nabla_k\widetilde{f}(\mathbf{y}_m,\boldsymbol{\chi}_0,t,b)-\nabla_k\widetilde{f}(\mathbf{y}_m,\boldsymbol{\chi}_{0n},t,b)\right)\right]\eta_k(u), \\
\widehat{V}_n(u) &\equiv \sum_{k=1}^{K_n}\left[\frac{1}{\sqrt{M}}\sum_{m=1}^{M}\left(\nabla_k f(\mathbf{y}_m,\boldsymbol{\chi}_0,t,b)-\nabla_k f(\mathbf{y}_m,\boldsymbol{\chi}_{0n},t,b)\right)\right]\eta_k(u) \\
\widetilde{Z}_n(u) &\equiv \sum_{k=1}^{K_n}\left[\frac{1}{\sqrt{M}}\sum_{m=1}^{M}\nabla_k\widetilde{f}(\mathbf{y}_m,\boldsymbol{\chi}_0,t,b)\right]\eta_k(u), \\
\widehat{Z}_n(u) &\equiv \sum_{k=1}^{K_n}\left[\frac{1}{\sqrt{M}}\sum_{m=1}^{M}\nabla_k f(\mathbf{y}_m,\boldsymbol{\chi}_0,t,b)\right]\eta_k(u) \\
\widetilde{b}_{l,n}(u) &\equiv -\sum_{k=1}^{K_n}\left[\frac{1}{M}\sum_{m=1}^{M}\nabla_{k,l}\widetilde{f}(\mathbf{y}_m,\boldsymbol{\chi}_{0n}+\lambda_k(\boldsymbol{\chi}_n-\boldsymbol{\chi}_{0n}),t,b)\right]\eta_k(u) \\
\widehat{b}_{l,n}(u) &\equiv -\sum_{k=1}^{K_n}\left[\frac{1}{M}\sum_{m=1}^{M}\nabla_{k,l}f(\mathbf{y}_m,\boldsymbol{\chi}_{0n}+\lambda_k(\boldsymbol{\chi}_n-\boldsymbol{\chi}_{0n}),t,b)\right]\eta_k(u)
\end{aligned}$$

where $\eta_k(u)$s are orthogonal weight functions on $[0,1]$ such that $\sigma_k = \int_0^1 \eta_k(u)^2\,du$ converges fast enough to zero as k goes to infinity. Note that in this case, n denotes the number of parameters, including sieve polynomials. Then,

$$\sum_{l=1}^{n}\widetilde{b}_{l,n}(u)\sqrt{M}(\chi_{n,l}-\chi_{0,l}) = \widetilde{Z}_n(u)-\widetilde{W}_n(u)-\widetilde{V}_n(u)$$

where, from Lemma 1, Bierens (2014), $sup_{0\le u\le 1}\left|\widetilde{W}_{n_N}(u)\right| = o_p(1)$ and $sup_{0\le u\le 1}\left|\widetilde{V}_{n_N}(u)\right| = o_p(1)$ because of Assumption 6.3 of Bierens (2014).

Next, we state Lemma 6.2 of Bierens (2014).

Lemma 7 *Under Assumption 6.4 and 6.5, $\widetilde{Z}_n \Rightarrow Z$ on $[0,1]$ where Z is a mean zero Gaussian* *process with covariance function*

$$\begin{aligned}\Gamma(u_1,u_2) &= E\left[Z(u_1)Z(u_2)\,|\mathbf{z}_m\right]\\ &= \sum_{k=1}^{\infty}\sum_{l=1}^{\infty}E\left[\nabla_k\widetilde{f}(\mathbf{y}_m,\boldsymbol{\chi}_0,t,b)\nabla_l\widetilde{f}(\mathbf{y}_m,\boldsymbol{\chi}_0,t,b)\eta_k(u_1)\eta_l(u_2)\,|\mathbf{z}_m\right].\end{aligned}$$

Moreover,

$$sup_{0\le u_1\le 1,0\le u_2\le 1}\left|\Gamma(u_1,u_2)\right|<\infty$$

and thus, Assumptions 6.1-6.5 imply that

$$\sum_{l=1}^{n}\widetilde{b}_{l,n}(u)\sqrt{M}\left(\chi_{n,l}-\chi_{0,l}\right)\Rightarrow Z(u).$$

Now, $\widetilde{f}(\mathbf{y}_m,\boldsymbol{\chi}_0,t,b)$ is i.i.d. distributed with mean $E\left[\widetilde{f}(\mathbf{y}_m,\boldsymbol{\chi}_0,t,b)\right]$ and variance

$$E\left[\widetilde{f}(\mathbf{y}_m,\boldsymbol{\chi}_0,t,b)-E\left[\widetilde{f}(\mathbf{y}_m,\boldsymbol{\chi}_0,t,b)\right]^2|\mathbf{z}_m\right]=E\left[(\nu+\zeta)^4+\left(\sigma_\nu^2+\sigma_\zeta^2\right)^2\right]<\infty.$$

Thus, from the Central Limit Theorem, we obtain

$$\frac{1}{\sqrt{M}}\sum_{m=1}^{M}\widetilde{f}(\mathbf{y}_m,\boldsymbol{\chi}_0,t,b)\Rightarrow N\left(E\left[\widetilde{f}(\mathbf{y}_m,\boldsymbol{\chi}_0,t,b)\right],Var\left(\widetilde{f}(\mathbf{y}_m,\boldsymbol{\chi}_0,t,b)\right)\right).$$

Now,

$$\widehat{Z}_n(u)=A_M\widetilde{Z}_n+\left[\sum_{k=1}^{p}\frac{\partial A_M}{\partial\theta_{ck}}\right]\left[\frac{1}{\sqrt{M}}\sum_{m=1}^{M}\widetilde{f}(\mathbf{y}_m,\boldsymbol{\chi}_0,t,b)\right]\eta_k(u)\Rightarrow Z$$

because $A_M\overset{a.s.}{\to}1$, $\partial A_M/\partial\theta_{ck}\overset{a.s.}{\to}0$, $A_M\widetilde{Z}_n\Rightarrow Z$,$\left[\sum_{k=1}^{p}\frac{\partial A_M}{\partial\theta_{ck}}\right]\left[\frac{1}{\sqrt{M}}\sum_{m=1}^{M}\widetilde{f}(\mathbf{y}_m,\boldsymbol{\chi}_0,t,b)\right]\eta_k(u)\overset{P}{\to}0$. Similarly, we derive $sup_{0\le u\le 1}\left|\widehat{W}_{n_N}(u)\right|=o_p(1)$ and $sup_{0\le u\le 1}\left|\widehat{V}_{n_N}(u)\right|=o_p(1)$

Therefore, $\widehat{Z}_n$ also satisfies Lemma 6.2. Furthermore, for $l>p$,

$$\widehat{b}_{l,n}(u)=A_M\widetilde{b}_{l,n}(u)\Rightarrow Z(u)$$

and for $l\le p$, let

$$\widetilde{b}_n(u)\equiv-\sum_{k=1}^{K_n}\left[\frac{1}{M}\sum_{m=1}^{M}\nabla_k\widetilde{f}\left(\mathbf{y}_m,\boldsymbol{\chi}_{0n}+\lambda_k\left(\boldsymbol{\chi}_n-\boldsymbol{\chi}_{0n}\right),t,b\right)\right]\eta_k(u).$$

Then,

$$
\begin{aligned}
& \sum_{l=1}^{n} \widehat{b}_{l,n}(u)\sqrt{M}\left(\chi_{n,l}-\chi_{0,l}\right) \\
= \; & A_M \sum_{l=1}^{n} \widetilde{b}_{l,n}(u)\sqrt{M}\left(\chi_{n,l}-\chi_{0,l}\right) - \sum_{l=1}^{p} \frac{\partial A_M}{\partial \theta_{cl}} \widetilde{b}_n(u)\sqrt{M}\left(\chi_{n,l}-\chi_{0,l}\right) \\
& - \sum_{l=1}^{p}\sum_{k=1}^{p} \frac{\partial A_M}{\partial \theta_{cl}\partial \theta_{ck}} \left[\frac{1}{\sqrt{M}} \sum_{m=1}^{M} \widetilde{f}\left(\mathbf{y}_m, \boldsymbol{\chi}_{0n} + \lambda_k\left(\boldsymbol{\chi}_n - \boldsymbol{\chi}_{0n}\right), t, b\right)\right] \eta_k(u)\sqrt{M}\left(\chi_{n,k}-\chi_{0,k}\right) \\
& \Rightarrow Z(u)
\end{aligned}
$$

because the 2nd and the 3rd terms converge both to zero in distribution.

Then, given Assumptions 6.1-6.5, we have

$$
\sum_{l=1}^{n} \widehat{b}_{l,n}(u)\sqrt{M}\left(\chi_{n,l}-\chi_{0,l}\right) \Rightarrow Z(u)
$$

Next, we decompose this equation as follows.

$$
\begin{aligned}
& \left[\widehat{b}_{1,n}(u), \ldots, \widehat{b}_{p,n}(u)\right]\sqrt{M}\left(\boldsymbol{\theta}_{cn} - \boldsymbol{\theta}_{c0}\right) + \sum_{l=1}^{n-p} \widehat{b}_{l+p,n}(u)\sqrt{M}\left(\gamma_{n,l}-\gamma_{0,l}\right) \\
= \; & \widehat{Z}_n(u) - \widehat{W}_n(u) - \widehat{V}_n(u) \Rightarrow Z(u)
\end{aligned}
$$

Now, as in Bierens (2014), let

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

be the residual of the following projection

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

Then, one can show that

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

where $\widehat{\boldsymbol{a}}_n(u)\widehat{\boldsymbol{a}}_n(u)'$ is a p by p matrix, and $\boldsymbol{\theta}_{cn} - \boldsymbol{\theta}_{c0}$ a p by 1 vector.

Then, from Lemma 6.3, Bierens (2014),

$$plim_{M\to\infty}\int_0^1 \widehat{\boldsymbol{a}}_n(u)\,\widehat{\boldsymbol{a}}_n(u)'\,du=\int_0^1 \boldsymbol{a}(u)\,\boldsymbol{a}(u)'\,du$$

where $a(u)$ is the residual of the following projection exercise

$$b(u)=(b_1(u),...,b_p(u))=A\left[b_{p+1}(u),...,b_\infty(u)\right]+a(u)\,,\;\; l=1\dots p$$

where $\boldsymbol{b}(u)\in L_0^2(0,1)$ satisfies

$$\left\|\widehat{b}_{l,n}-b_l\right\|=\sqrt{\int_0^1\left(\widehat{b}_{l,n}(u)-b_l(u)\right)^2du}=o_p(1)$$

for $l=1,\dots,p$ and

$$\sum_{l=p+1}^{n}\rho_l\left\|\widehat{b}_{l,n}-b_l\right\|=o_p(1)$$

and

$$\liminf_{n\to\infty}\left\|\sum_{l=p+1}^{n}\rho_l b_l\right\|>0$$

We impose the Assumptions 6.7 of Bierens (2014), which is:

Assumption 17 *Assumption 6.7, Bierens (2014): Let*

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

$rank(B_{k,k})=k$ for each $k\geq p$.

Next, we impose Assumption 6.8:

$$\lim_{n\to\infty}\inf_{k\to\infty}det\left(L_p^{(n,k)}\right)>0$$

where $L^{(n,k)}$ satisfies $\Phi_k^{1/2}B_{n,k}=Q_{k,n}L^{(n,k)}$, and $\Phi_k=diag\left(2^{-2},2^{-4},\dots,2^{-2k}\right)$, $Q_{k,n}$ is a $k\times n$ orthogonal matrix and $L_p^{(n,k)}$ is the upper-left $p\times p$ block of the triangular matrix $L^{(n,k)}$. Then, Lemma 6.4 in Bierens (2014) holds, and

$$\mathbf{F} = \int_0^1 a(u)\, a(u)'\, du$$

is a full rank matrix, thus, invertible. Then,

$$\sqrt{M}\left(\boldsymbol{\theta}_{cM} - \boldsymbol{\theta}_{c0}\right) \overset{d}{\to} N_p\left(\mathbf{0}, \mathbf{F}^{-1}\boldsymbol{\Upsilon}\mathbf{F}'^{-1}\right),$$

where

$$\boldsymbol{\Upsilon} = \int_0^1 \int_0^1 a(u_1)\, \boldsymbol{\Gamma}(u_1, u_2)\, a(u_2)\, du_1 du_2.$$

Then, for the logit model, $\boldsymbol{\theta}_{c0} = \{\alpha\}$ and for the BLP model, $\boldsymbol{\theta}_{c0} = \{\mu_\alpha, \sigma_\alpha, \sigma_{\beta k}, k = 1, \ldots, K\}$. Then, because of consistency, $plim_{M\to\infty}\widehat{\boldsymbol{\theta}}_{cM} = \boldsymbol{\theta}_0$. Hence, $\widehat{\boldsymbol{\delta}}_{mM} = s^{-1}\left(\mathbf{s}_m, \widehat{\boldsymbol{\theta}}_{cM}\right) \to_p \boldsymbol{\delta}_{m0}$ for each $m = 1, \ldots, M$ and therefore, given the orthogonality assumption $E\left[\mathbf{X}_m\boldsymbol{\xi}_m\right] = 0$, given conditions for the consistency of the OLS estimator for the logit model

$$\widehat{\boldsymbol{\beta}}_M = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\left[\widehat{\boldsymbol{\delta}}_M - \widehat{\alpha}\mathbf{p}\right] \to_p \boldsymbol{\beta}_0,$$

or

$$\widehat{\boldsymbol{\mu}}_{\boldsymbol{\beta}M} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\left[\widehat{\boldsymbol{\delta}}_M - \widehat{\mu}_\alpha\mathbf{p}\right] \to_p \boldsymbol{\mu}_{\boldsymbol{\beta}0},$$

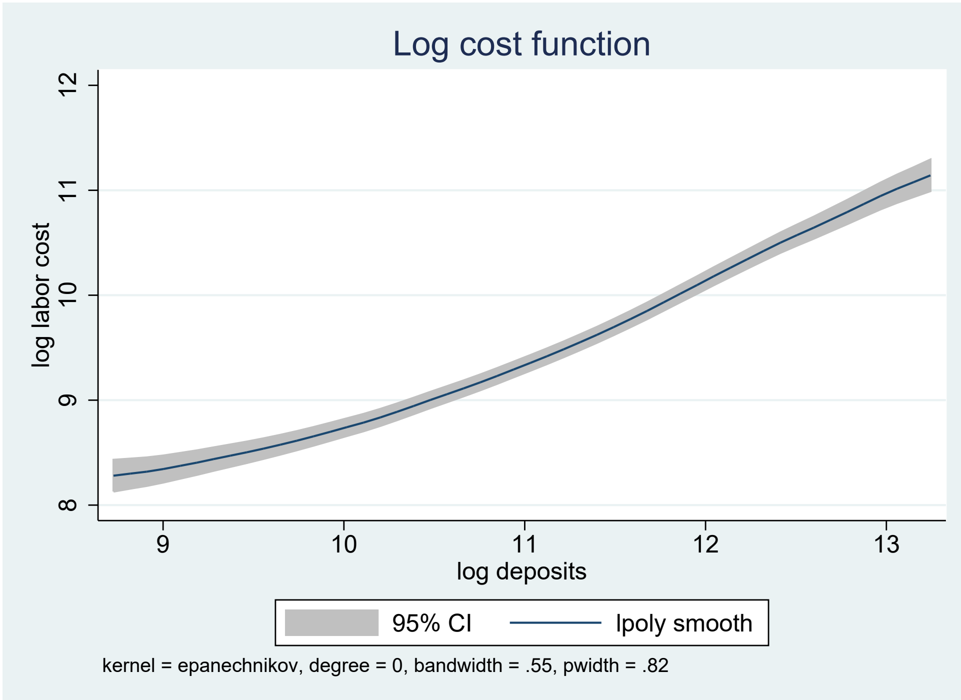
as $M \to \infty$. Therefore, $\widehat{\boldsymbol{\xi}}_M \to_p \boldsymbol{\xi}_0$. Therefore, if we impose standard assumptions to ensure $plim_{M\to\infty}\frac{1}{M}\mathbf{X}'\mathbf{X} = E\left(\mathbf{X}_m'\mathbf{X}_m\right)$, and $E\left(\mathbf{X}_m'\mathbf{X}_m\right)$ being nonsingular, $plim_{M\to\infty}\frac{1}{M}\sum_{m=1}^M \mathbf{X}_m'\left(\widehat{\boldsymbol{\xi}_m'\boldsymbol{\xi}_m}\right)\mathbf{X}_m = E\left[\mathbf{X}_m'\boldsymbol{\Sigma}_0\mathbf{X}_m\right]$, then

$$\sqrt{M}\left(\widehat{\boldsymbol{\mu}}_{\boldsymbol{\beta}} - \boldsymbol{\mu}_{\boldsymbol{\beta}0}\right) \to_d N\left(0, E\left(\mathbf{X}_m'\mathbf{X}_m\right)^{-1} E\left[\mathbf{X}_m'\boldsymbol{\Sigma}_0\mathbf{X}_m\right] E\left(\mathbf{X}_m'\mathbf{X}_m\right)^{-1}\right)$$

as $M \to \infty$.

G: Details on the empirical application on the banking sector

We use the bank level total deposits as the output and the bank level employee salaries as the variable cost. We derive the deposit interest rate by calculating the deposit interest payment per bank level total deposit minus per deposit service fee. Market level weekly wage is used as the input price. The controls we use are the log of population per branch, log of number of markets served and the log of bank age plus one. We construct the bank level dataset of these variables by



merging the data from various sources. We obtain the bank level total deposits and the number of branches for each market from Federal Deposit Insurance Corporation (FDIC), employee salaries, deposit and loan interest rates at the bank level from the balance sheet information reported to the Federal Financial Institutions Examination Council (FFIEC). Information on county level weekly wage and county level population is obtained from the Census and the Bureau of Labor Statistics. We define markets as Metropolitan Statistical Areas (MSA's) for urban areas and counties for rural areas.

By closely inspecting the data, we noticed that in many markets, credit unions seem to be effectively nonexistent as an outside option. We made a judicious choice of removing the markets whose market share of credit unions is less than 1 %. More concretely, we start with 16417 banks in 2325 markets. After removing the markets that have missing data on banks and credit unions, we have 11647 banks in 1117 markets. Then, removing the markets with small share of credit unions, we are left with 10513 banks in 954 markets. Finally, for the sake of reducing the computational burden, we only selected markets where the total number of banks are no more than 40. Then, we are left with 8155 banks in 914 markets.

In Figure 3, we show the kernel fitted relationship between log deposits and log of salaries divided by weekly wage, where deposits and total salaries in millions of dollars. As we can see, overall, the relationship between log deposits and log variable cost is smoothly increasing.

Table 9: Sample Statistics

	mean	std. dev
Bank deposits (in million $)	182.18	735.42
Bank deposit market share	0.09411	0.1137
Bank total salaries (in million $)	455.88	1396.0
Outside option deposit market share	0.1528	0.1310
Deposit interest rate	0.02514	7.484E-3
No. of banks per market	14.08	9.534
No. of markets per bank	2.428	6.751
Number of branches per market	4.137	7.835
Bank age	75.68	47.08
2002 Jan. treasury note interest rate	0.05536	0
Housing price index	139.4	14.02
Sample size	8155	
No. of banks	3230	
No. of markets	914	
No. of single market banks	2067	