Identification and Estimation of Demand Models with Endogenous Product Entry and Multiple Equilibria

Victor Aguirregabiria
University of Toronto and CEPR

Alessandro Iaria
University of Bristol and CEPR

Senay Sokullu
University of Bristol

December 31, 2020

Abstract

The estimation of demand models for differentiated products often relies on data from multiple geographic markets and time periods. The fact that firms do not offer some products in some markets generates a problem of endogenous product selection in demand estimation. The high dimension and non-additivity of the demand unobservables that enter firms' expected profit, and the potential multiplicity of equilibria in the entry game, make this selection problem challenging. In particular, standard two-step estimation methods to account for selection are inconsistent. We show the identification of demand parameters in a structural model of demand, price competition, and market entry that allows for a flexible specification of firms' information on demand and for a nonparametric distribution of demand unobservables. We propose a simple two-step estimator in the spirit of traditional methods to control for endogenous selection. We illustrate our method using simulated data and real data from the airline industry. We show that not accounting for endogenous product entry generates substantial biases that can be even larger than those from ignoring price endogeneity.

Keywords: Demand of differentiated product; Endogenous product availability; Selection bias; Market entry; Multiple equilibria; Identification; Estimation.

JEL codes: C13, C35, C57.

Victor Aguirregabiria. *University of Toronto and CEPR*. Address: 150 St. George Street. Toronto, ON, M5S 3G7, Canada. E-mail: `victor.aguirregabiria@utoronto.ca`
Alessandro Iaria. *University of Bristol and CEPR.* The Priory Road Complex, Priory Road, BS8 1TU, Bristol, UK. E-mail: `alessandro.iaria@bristol.ac.uk`
Senay Sokullu. *University of Bristol.* 8 Woodland Road, Bristol, BS81TN, UK.
E-mail: `senay.sokullu@bristol.ac.uk`

1 Introduction

Since the seminal work by Amemiya (1973) and Heckman (1976), the selection problem has been a fundamental topic in microeconometrics. Dealing with censored observations (i.e., zeroes) in the estimation of consumer demand, has an even longer tradition, as illustrated by Tobin (1958)'s seminal work and the so called Tobit model. The estimation of consumer demand models has motivated the development of important methods to account for sample selection. Most of the early literature studies the demand of a single product, but there are applications to demand systems using Amemiya's multivariate Tobit model (Amemiya, 1974; Yen, 2005; Yen and Lin, 2006). These demand systems with censored demand (or zeroes) share three important features: (i) unobservables enter additively in the demand and selection equations; (ii) the system includes only a few products (broad product categories, such as alcohol); and (iii) endogenous selection comes from consumer demand decisions and not from firms' supply decisions. Recently, the selection problem has received substantial attention in the estimation of demand of differentiated products, where the number of products can be hundreds or more. See the work by Gandhi, Lu, and Shi (2017), Ciliberto, Murry, and Tamer (2018), Li et al. (2018), and Dubé, Hortaçsu, and Joo (2020). This recent literature tries to relax the restrictions (i), (ii), or/and (iii) mentioned above.

This paper deals with the estimation of demand of differentiated products using market level data when there is censoring/selection because of firms not offering some products in some markets. Demand estimation often relies on data from multiple geographic markets and/or time periods. Some products are not offered in some markets. This problem appears in many applications such as airline markets (Berry, Carnall, and Spiller, 2006; Berry and Jia, 2010; Aguirregabiria and Ho, 2012), consumer choice of supermarket (Smith, 2004), or choice of radio station (Sweeting, 2013), among others. When making their market entry decisions, firms may have information about the demand of their products, and more specifically about components of demand that are unobserved to the researcher. Firms are more likely to enter markets where expected demand is higher. Not accounting for this selection can generate substantial biases in the estimation of demand parameters. When panel data are available, a simple approach to control for this selection problem consists of including fixed effects — product, market, and time fixed effects — and assume that the remaining part of the error term in the demand equation is unknown to firms when they make their market/product entry decisions.[1] Though this approach is convenient because of its simplicity, it is based on restrictions on firms' information that may not be plausible in some empirical applications. However, this simple approach illustrates that the nature of this selection problem, and its solution, depends on firms' information about demand at the moment of market entry.

This selection problem is not standard. A firm supplies a product in a market if its expected profit is greater than zero. In an oligopoly market, this expected profit depends on demand unobservables in a non-additive and complicated form, as it results from a two-stage equilibrium game: a first stage entry game, and a second stage price competition game between active firms. The high dimension and non-additivity of the demand unobservables that enter firms' expected profit, and the potential multiplicity of equilibria in this game, make this selection problem challenging. In particular, standard two-step estimation methods to account for selection are inconsistent.

This paper presents several contributions. We study the identification of demand parameters in a structural model of demand, price competition, and market entry that allows for a flexible specification of firms' information on (unobserved) demand, and for a nonparametric distribution of demand unobservables. First, we present new identification results in this model. We show that the probability of product entry conditional on firms' information about (unobserved) demand is nonparametrically identified. This result exploits the additivity between observable and unobservable product characteristics in consumer utility. Given these entry probabilities, we show the identification of demand parameters. Second, we propose a simple two-step estimator in the spirit of traditional methods to control for endogenous selection. In the first step, we estimate a nonparametric continuous mixture model for the choice probabilities of product entry. In a second step, we estimate demand parameters using a GMM that accounts for both endogenous product availability and price endogeneity. We illustrate our method using simulated data and real data from the airline industry.

Our work is related to several recent papers that propose methods to deal with the selection problem in the estimation of demand of differentiated products. This selection problem is often referred as *the problem of zeroes in market shares*. Our paper also builds on

[1] For instance, this is the approach in Aguirregabiria and Ho (2012) and Eizenberg (2014). A similar but weaker restriction consists in assuming that the residual error term — after controlling for fixed effects — follows a first order autoregressive process, and the innovation shock of this process is not known to firms when they make their entry decisions. This is the approach in Sweeting (2013).

and brings together the literatures on identification and estimation of games, nonparametric finite mixtures, and sample selection corrections.

The motivation and purpose of our paper is closely related to the work of Ciliberto, Murry, and Tamer (2018) and to Li et al. (2018). These papers develop methods for the estimation of structural models that bring together the Berry, Levinsohn, and Pakes (1995, BLP hereafter) framework and a game of market/product entry. These papers are interested in the identification and estimation of all the structural parameters in the model, including demand, marginal costs, entry costs, and the probability distribution of the corresponding unobservables. The estimation of the full model requires the application of nested fixed point algorithms with the consequent repeated solution for the equilibria in the two-step game. For this purpose, these authors impose strong parametric restrictions on all the structural functions and on the distribution of the unobservables. In contrast, our approach focuses on the identification and estimation of demand parameters only, and derives identification results where all the primitives except consumer utility function – i.e., marginal and entry costs and the distribution of all the unobservables – are nonparametrically specified. Furthermore, our estimation method is computationally simple as it does not require solving for an equilibrium. Our approach is in the spirit of – and extends – the literature on semiparametric estimation of sample selection models (Newey, Powell, and Walker, 1990; Ahn and Powell, 1993; Newey, 2009; Aradillas-Lopez et al, 2007).[2]

The method proposed in Dubé, Hortaçsu, and Joo (2020) has similar motivation as our approach. However, their model is quite different to ours. In particular, it is reduced form model of market entry and competition, and the restrictions in their model are not compatible with a two-stage game of oligopoly competition in a differentiated product industry. Their model seems more consistent with some forms of monopolistic competition.

Ellickson and Misra (2012) propose a semiparametric extension of Dubin and McFadden (1984) to augment reduced form profit functions in entry games with data on rev-

[2] An interesting feature of the methods in Ciliberto, Murry, and Tamer (2018) and Li et al. (2018) is that the estimated model can be used for counterfactual experiments that account for the endogeneity of product entry. For instance, this is particularly useful when simulating the effects of a merger, as illustrated by Li et al. (2018). In contrast, our semi-parametric framework is mainly designed for the robust and computationally simple estimation of demand. Of course, given the estimated demand parameters and residual unobservables, it is possible to obtain estimates of marginal costs and entry costs under weaker parametric restrictions than the ones imposed for the joint estimation of the full structural model.

enues. Their approach assumes additive separability of the unobservable component in the revenue function, such that this framework cannot accommodate the type of demand unobservables in the Berry, Levinshon, and Pakes (1995) demand model. Moreover, their identification relies on equilibrium uniqueness in the entry game. In contrast, our method bridges the BLP demand models with the nonparametric identification of games with multiple equilibria as in Aguirregabiria and Mira (2019) and Xiao (2018).

The rest of the paper is organized as follows. Section 2 presents our model and assumptions. Section 3 describes the selection problem in this model, and why it is not standard . Section 4 presents our identification results. We describe our estimation method in section 5. Section 6 presents Monte Carlo experiments, and section 7 an empirical application using US airline markets. We summarize and conclude in section 8.

2 Model

2.1 A simple example

Suppose an economy with two types of geographic markets: rich and poor markets. We index markets by m and, for simplicity, suppose that all the markets have similar population size, H. A telecommunication company must decide whether to offer broadband internet in each market. Consumer average willingness to pay for broadband internet in market m is δ_m. The firm incurs a fixed cost F to offer broad band internet in a market. The telecommunication company has perfect information on the market-specific utilities δ_m and maximizes profits such that it provides broadband internet to those markets for which profit is positive. That is, market m gets broadband internet if $[p(\delta_m) - c]$ $q(\delta_m) - F \geq 0$, where $p(\delta_m)$ and $q(\delta_m)$ represent the monopolist's profit maximizing price and quantity – that depend on consumer willingness to pay – and c is the per unit cost. Variable profit $[p(\delta_m) - c]$ $q(\delta_m)$ is an increasing function of δ_m. Suppose that the value of δ_m in rich markets is large enough such that the company offers the product in these markets. In contrast, poor markets do not get broadband because δ_m in these markets is not large enough.

The government understands that it may not be feasible for the telecommunication company to offer the service to poor areas without some public financial support. The government decides to give the company a per-market lump-sum subsidy, S_m, defined as

$S_m = F - [p(\delta_m) - c]\, q(\delta_m)$, such that the profit maximizing company is willing to offer broadband in every market. To construct this amount of subsidy in a poor market, the government needs to know F, c, and the value of δ_m in that market. The government's engineers can get precise estimates of F and c but they do not have direct information on δ_m.

To estimate δ_m in poor markets, the government only has information on demand in rich markets. If demand follows a logit model, we have that $\ln(q_m/H) = \delta_m - p_m$, such that δ_m can be estimated as $\ln(q_m/H) + p_m$. Suppose that the government estimates the value of δ in a poor market by using the sample mean of $\delta's$ in the observed rich markets. The firm's profit maximization implies that the markets that were receiving broadband have larger willingness to pay than the markets that do not. Therefore, this sample mean over-estimates the value δ_m in poor markets. This estimation bias implies a level of subsidy that is too small to encourage the firm to offer broadband in poor markets. The government can use a more sophisticated OLS approach by taking into account that $\delta_m = x_m'\beta + \xi_m$, where x_m is a vector of observable socioeconomic characteristics in market m, and ξ_m is unobservable. Ignoring the selection problem, the government can estimate β by applying OLS in the linear regression $\ln(q_m/H) + p_m = x_m'\beta + \xi_m$ using the subsample of rich markets. This approach also implies a biased estimate of the true values of β and δ_m. Econometric methods that control for sample selection can generate consistent estimates of β and $x_m'\beta$.

In this simple example, the selection problem is standard because there is only one product and the firm is a monopolist. The rest of this section presents a model of demand, price competition, and product entry in an oligopoly differentiated product industry.

2.2 Demand

The demand system follows Berry, Levinsohn, and Pakes (1995), or BLP hereafter. Throughout the paper, we maintain the assumption of single-product firms. There are J firms indexed by $j \in \mathcal{J} = \{1, 2, ..., J\}$, and M geographic markets indexed by $m \in \mathcal{M} = \{1, 2, ..., M\}$. Consumers living in market m can buy only the products available in that market. Firms make entry decisions in each market and compete at the local market level after entry.

The indirect utility of household h in market m from buying product j is:

$$U_{hjm} \equiv \delta(p_{jm}, \boldsymbol{x}_{jm}) + v(p_{jm}, \boldsymbol{x}_{jm}, v_h) + \varepsilon_{hjm}, \tag{1}$$

where p_{jm} and $\boldsymbol{x}_{jm}$ are the price and other characteristics, respectively, of product j in market m; $\delta_{jm} \equiv \delta(p_{jm}, \boldsymbol{x}_{jm})$ is the average (indirect) utility of product j in market m; and $v(p_{jm}, \boldsymbol{x}_{jm}, v_h) + \varepsilon_{hjm}$ represents household-specific utility deviation for product j, with zero mean when averaged over households in market m. The term $v(p_{jm}, \boldsymbol{x}_{jm}, v_h)$ depends on the vector of random coefficients v_h that is unobserved to the researcher and distributed according to $F_v(\cdot|\boldsymbol{\sigma})$, with $\boldsymbol{\sigma}$ the parameters characterizing this distribution. The term ε_{hjm} is unobserved to the researcher and is assumed to be distributed Type I extreme value and i.i.d. over (h, j, m). As is standard, we specify the average utility of product j as:

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

with α and $\boldsymbol{\beta}$ being parameters. Variable ξ_{jm} captures the characteristics of product j in market m which are valuable to the households but unobserved to the researcher. The outside option is represented by $j = 0$ and its indirect utility is normalized to $U_{h0m} = \varepsilon_{h0m}$.

Let $a_{jm} \in \{0, 1\}$ be the indicator for product/firm j being available in market m, and vector $\boldsymbol{a}_m \equiv (a_{jm} : j \in \mathcal{J})$ denote the products available in market m. We assume that the outside option $j = 0$ is always available in every local market. Every household chooses a product to maximize their utility. Let s_{jm} be the market share of product j in market m, i.e., the proportion of households choosing product j:

$$s_{jm} = d_j(\boldsymbol{\delta}_m, \boldsymbol{a}_m, \boldsymbol{p}_m, \boldsymbol{x}_m) \equiv \int \frac{a_{jm} \ \exp\left(\delta_{jm} + v(p_{jm}, \boldsymbol{x}_{jm}, v)\right)}{1 + \sum_{i=1}^{J} a_{im} \ \exp\left(\delta_{im} + v(p_{im}, \boldsymbol{x}_{im}, v)\right)} dF_v(v). \tag{3}$$

This system of J equations represents the demand system in market m. We can represent this system in a vector form as: $\boldsymbol{s}_m = \boldsymbol{d}(\boldsymbol{\delta}_m, \boldsymbol{a}_m, \boldsymbol{p}_m, \boldsymbol{x}_m)$.

For our analysis, it is convenient to define the sub-system of demand equations that includes market shares, average utilities, and product characteristics of only those products available in the market. We represent this system as:

$$\boldsymbol{s}_m^{(\boldsymbol{a})} = \boldsymbol{d}^{(\boldsymbol{a})}(\boldsymbol{\delta}_m^{(\boldsymbol{a})}, \boldsymbol{p}_m^{(\boldsymbol{a})}, \boldsymbol{x}_m^{(\boldsymbol{a})}), \tag{4}$$

where $\boldsymbol{s}_m^{(\boldsymbol{a})} \equiv (s_{jm} : j \in \mathcal{J} \text{ and } a_{jm} = 1)$ and similarly for the sub-vectors $\boldsymbol{d}^{(\boldsymbol{a})}$, $\boldsymbol{\delta}_m^{(\boldsymbol{a})}$, $\boldsymbol{p}_m^{(\boldsymbol{a})}$, and $\boldsymbol{x}_m^{(\boldsymbol{a})}$. Lemma 1 establishes that the invertibility property by Berry (1994) applies to demand system (4) for any possible value of $\boldsymbol{a}$.

LEMMA 1. *Suppose that the outside option $j = 0$ is always available. Then, for any value of the vector $\boldsymbol{a} \in \{0,1\}^J$, the system $\boldsymbol{s}_m^{(\boldsymbol{a})} = \boldsymbol{d}^{(\boldsymbol{a})}(\boldsymbol{\delta}_m^{(\boldsymbol{a})}, \boldsymbol{p}_m^{(\boldsymbol{a})}, \boldsymbol{x}_m^{(\boldsymbol{a})}; \boldsymbol{\sigma})$ is invertible with respect to $\boldsymbol{\delta}_m^{(\boldsymbol{a})}$ such that for every product in this subsystem (i.e., for every product with $a_{jm} = 1$) the inverse function $\delta_{jm}^{(\boldsymbol{a})} = d_j^{(\boldsymbol{a})-1}(\boldsymbol{s}_m^{(\boldsymbol{a})}, \boldsymbol{p}_m^{(\boldsymbol{a})}, \boldsymbol{x}_m^{(\boldsymbol{a})}; \boldsymbol{\sigma})$ exists.* ■

Proof of Lemma 1. If the outside option $j = 0$ is available, then – for any value of the vector $\boldsymbol{a}$ – the system of equations (4) satisfies the conditions for invertibility in Berry (1994). ■

For the observed availability vector in market m, the average utility of product j, $\delta_{jm} \equiv d_j^{(\boldsymbol{a}_m)-1}(\boldsymbol{s}_m^{(\boldsymbol{a}_m)}, \boldsymbol{p}_m^{(\boldsymbol{a}_m)}, \boldsymbol{x}_m^{(\boldsymbol{a}_m)}; \boldsymbol{\sigma})$, can be expressed as:

$$\delta_{jm} = \alpha \, p_{jm} + \boldsymbol{x}'_{jm} \, \boldsymbol{\beta} + \xi_{jm} \quad \textbf{if and only if } a_{jm} = 1 \tag{5}$$

This condition is particularly important to deal with the selection problem we study. The regression equation corresponding to the demand of product j only depends on the availability of product j and not on the availability of the other products. More specifically, the selection problem in the estimation of the demand of product j can be described only in terms of the conditional expectation

$$\mathbb{E}\left[\xi_{jm} \mid a_{jm} = 1\right]. \tag{6}$$

This is an important implication of working directly with the inverse demand system, as clarified by equation (5).

To appreciate the value of this property, consider instead the case of an *Almost Ideal Demand System (AIDS)* (Deaton and Muellbauer, 1980). In that model, for every value of the vector $\boldsymbol{a}_m$, we have a different regression equation for the demand of product j as a function of the log-prices *of all available products*. In the AIDS model, the selection problem in the estimation of the demand of product j is not only related to the availability of product j — say $\mathbb{E}\left[\xi_{jm} \mid a_{jm} = 1\right]$ — but also to the availability of the other products — i.e., $\mathbb{E}\left[\xi_{jm} \mid \boldsymbol{a}_m = \boldsymbol{a}\right]$. That is, in the AIDS model we have a different selection term for each value of the vector $\boldsymbol{a}$. This structure makes the selection problem multi-dimensional and significantly complicates identification and estimation when the number of products J is large. The following Example illustrates Lemma 1 in the case of the nested logit model.

Example 1 (Nested logit model). The J products are partitioned into R mutually exclusive groups indexed by r. We use r_j to represent the group of product j. The indirect utility function is $U_{hmj} \equiv \delta_{jm} + \frac{1}{\sigma-1}\, v_{hmr(j)} + \varepsilon_{hmj}$, where the variables v_{hmr_j} and ε_{hmj} are i.i.d. Type I extreme value and mutually independent, and $\sigma > 0$ is a parameter. This model implies $s_{jm} = d_j^{(\boldsymbol{a}_m)}(\boldsymbol{\delta}_m) = d_{r_j}^{(\boldsymbol{a}_m)}\, d_{j|r_j}^{(\boldsymbol{a}_m)}$ with

$$d_{j|r_j}^{(\boldsymbol{a}_m)} = \frac{a_{jm}\, e^{\delta_{jm}}}{\sum_{i \in r_j} a_{im}\, e^{\delta_{im}}} \text{ and } d_{r_j}^{(\boldsymbol{a}_m)} = \frac{\left[\sum_{i \in r_j} a_{im}\, e^{\delta_{im}}\right]^{\frac{1}{\sigma-1}}}{1 + \sum_{r=1}^{R} \left[\sum_{i \in r} a_{im}\, e^{\delta_{im}}\right]^{\frac{1}{\sigma-1}}}. \tag{7}$$

If $a_{jm} = 1$ and $s_{0m} > 0$, the inverse function $d_j^{(\boldsymbol{a}_m)-1}(\cdot)$ exists — regardless of the value of a_{im} for any product i different from j. It is straightforward to show that this inverse function has the following form:

$$\delta_{jm} = \ln\left(\frac{s_{jm}}{s_{0m}}\right) - \sigma\, \ln\left(\frac{\sum_{i \in r_j} s_{im}}{s_{0m}}\right), \tag{8}$$

and it implies the regression equation

$$\ln\left(\frac{s_{jm}}{s_{0m}}\right) = \sigma\, \ln\left(\frac{\sum_{i \in r_j} s_{im}}{s_{0m}}\right) + \alpha\, p_{jm} + \boldsymbol{x}'_{jm}\, \boldsymbol{\beta} + \xi_{jm}. \tag{9}$$

Given $s_{0m} > 0$, this regression equation holds whenever $a_{jm} = 1$. ■

2.3 Price Competition

Firms' market entry decisions, prices, and quantities are determined as an equilibrium of a two-stage game. In the first stage of this game, firms maximize their expected profit by choosing whether to be active or not in the market. In the second stage, prices and quantities of the active firms are determined as a Bertrand equilibrium of price competition. This two-stage game is played independently across markets.

The profit of being inactive is normalized to zero for all firms. Let Π_{jm} be the profit of firm j if active in market m. This equals revenues minus costs:

$$\Pi_{jm} = p_{jm}\, q_{jm} - c_j(q_{jm}; \boldsymbol{x}_{jm}, \omega_{jm}) - f_j(\boldsymbol{x}_{jm}, \eta_{jm}), \tag{10}$$

where q_{jm} is the quantity sold (i.e., market share s_{jm} times market size H_m), $c_j(q_{jm}; \boldsymbol{x}_{jm}, \omega_{jm})$ is the variable cost function, and $f_j(\boldsymbol{x}_{jm}, \eta_{jm})$ is the fixed cost of entry. Vector $\boldsymbol{x}_{jm}$ includes variables affecting demand or/and costs and observable to the researcher. We use $\boldsymbol{x}_m \equiv (\boldsymbol{x}_{jm} : j \in \mathcal{J})$ to represent the vector of all the exogenous variables that are observable to the researcher, either in demand or in costs.[3] Variables ω_{jm} and η_{jm} are unobservable to the researcher.

Taking as given the vector of entry decisions, $\boldsymbol{a}_m$, the best response function in the Bertrand competition game implies the following system of pricing equations:

$$p_{jm} = mc_{jm} - d_{jm}^{(\boldsymbol{a}_m)} \left[\frac{\partial d_{jm}^{(\boldsymbol{a}_m)}}{\partial p_{jm}} \right]^{-1}, \text{ for every } j \in \mathcal{J}, \tag{11}$$

where mc_{jm} is the marginal cost $\partial c_{jm} / \partial q_{jm}$. A solution to this system of equations is a Bertrand equilibrium. Given $(\boldsymbol{a}_m, \boldsymbol{x}_m, \boldsymbol{\xi}_m, \boldsymbol{\omega}_m)$, the model may have multiple Bertrand equilibria. We do not impose restrictions on equilibrium selection and allow for each market to select a different equilibrium.

Let $V_j(\boldsymbol{a}_m, \boldsymbol{x}_m, \boldsymbol{\xi}_m, \boldsymbol{\omega}_m)$ be the indirect variable profit function for firm j that results from plugging into the expression $p_{jm}\, q_{jm} - c_j(q_{jm}; \boldsymbol{x}_{jm}, \omega_{jm})$ the value for (p_{jm}, q_{jm}) from the (selected) Bertrand equilibrium given $(\boldsymbol{a}_m, \boldsymbol{x}_m, \boldsymbol{\xi}_m, \boldsymbol{\omega}_m)$. Lemma 2 establishes a property of this profit function that we apply to obtain our identification results.

Assumption 1. *Suppose that $\boldsymbol{x}_{jm} = (z_{jm}, \widetilde{\boldsymbol{x}}_{jm})$ where z_{jm} is a scalar, and the model satisfies the following conditions. (i) Consumer utility does not have random coefficients in product characteristics z_{jm} and p_{jm}: i.e., $U_{hjm} = \beta_z z_{jm} + \alpha\, p_{jm} + \xi_{jm} + v(\widetilde{\boldsymbol{x}}_{jm}, v_h) + \varepsilon_{hjm}$. (ii) Marginal cost mc_{jm} is constant and it depends linearly on z_{jm} and ω_{jm}: i.e., $mc_{jm} = \gamma_z z_{jm} + \omega_{jm} + \widetilde{mc}_j(\widetilde{\boldsymbol{x}}_{jm})$. (iii) The equilibrium selection mechanism may depend on $\widetilde{\boldsymbol{x}}_m$ but it does not depend on $(z_{jm}, \xi_{jm}, \omega_{jm})$.* ■

LEMMA 2. *Under Assumption 1, the equilibrium variable profit function V_j has the following structure:*

$$\begin{array}{lcl} V_{jm} & = & V_j(\boldsymbol{a}_m, \widetilde{\boldsymbol{x}}_m, \boldsymbol{z}_m + \boldsymbol{\xi}_m^*) \ \text{ where } \boldsymbol{\xi}_m^* \equiv (\xi_{jm}^* : j \in \mathcal{J}) \\ \text{with} & & \xi_{jm}^* = \dfrac{\xi_{jm} + \alpha\, \omega_{jm}}{\beta_z + \alpha\, \gamma_z}. \quad \blacksquare \end{array} \tag{12}$$

Proof of Lemma 2. We omit here the market subindex m. Under condition (i), the demand

[3] Note that this notation is compatible with exclusion restrictions. That is, some variables in the vector $\boldsymbol{x}_{jm}$ may affect demand but not costs or viceversa, or may affect only one of the two types of costs. Our identification results do not require the presence of these exclusion restrictions. As in other BLP-type models, our model implies identifying exclusion restrictions even when the only exogenous observable variables are the product characteristics in demand, $\boldsymbol{x}_{jm}$.

system is:

$$s_j = \int \frac{a_j \ \exp\left\{\beta_z z_j + \alpha p_j + \xi_j + v(\widetilde{\boldsymbol{x}}_j, v)\right\}}{1 + \sum_{i=1}^{J} a_i \ \exp\left\{\beta_z z_i + \alpha p_i + \xi_i + v(\widetilde{\boldsymbol{x}}_i, v)\right\}} dF_v(v). \tag{13}$$

Define the price cost margin, $\tau_j \equiv p_j - mc_j$. Replacing p_j by $mc_j + \tau_j$ in equation (13), and taking into account condition (ii), we can rewrite the demand system as:

$$s_j = \int \frac{a_j \ \exp\left\{(\beta_z + \alpha\gamma_z) \ \left(z_j + \xi_j^*\right) + \alpha\tau_j + \widetilde{v}(\widetilde{\boldsymbol{x}}_j, v)\right\}}{1 + \sum_{i=1}^{J} a_i \ \exp\left\{(\beta_z + \alpha\gamma_z) \ \left(z_i + \xi_i^*\right) + \alpha\tau_i + \widetilde{v}(\widetilde{\boldsymbol{x}}_i, v)\right\}} dF_v(v). \tag{14}$$

where variable ξ_j^* is defined above, and $\widetilde{v}(\widetilde{\boldsymbol{x}}_j, v) \equiv \alpha \ \widetilde{mc}_j(\widetilde{\boldsymbol{x}}_j) + v(\widetilde{\boldsymbol{x}}_j, v)$. Now, consider the equilibrium equations in (11). Given these pricing equations and the representation of the system in (14), an equilibrium can be represented as a vector of price cost margins $\boldsymbol{\tau} \equiv (\tau_1, \tau_2, ..., \tau_J)$ that satisfies the following system of equations:

$$\tau_j = -d_j^{(\boldsymbol{a})}\left(\boldsymbol{\tau}, \boldsymbol{a}, \widetilde{\boldsymbol{x}}, \boldsymbol{z} + \boldsymbol{\xi}^*\right) \ \left[\frac{\partial d_j^{(\boldsymbol{a})}\left(\boldsymbol{\tau}, \boldsymbol{a}, \widetilde{\boldsymbol{x}}, \boldsymbol{z} + \boldsymbol{\xi}^*\right)}{\partial \tau_j}\right]^{-1} \quad \text{for every } j \in \mathcal{J}, \tag{15}$$

It is clear that any solution in $\boldsymbol{\tau}$ to this system of equations depends on the vectors $\boldsymbol{z}$, $\boldsymbol{\xi}$, and $\boldsymbol{\omega}$ only through the vector $\boldsymbol{z} + \boldsymbol{\xi}^*$. Finally, condition (iii) implies that, given $\boldsymbol{z} + \boldsymbol{\xi}^*$, the equilibrium variable profit does not depend also on $\boldsymbol{z}$, $\boldsymbol{\xi}$, or $\boldsymbol{\omega}$ through the equilibrium selection mechanism. ■

Lemma 2 has two main implications for identification. First, though the model includes two different unobservables per product (one for demand and one for marginal cost), the structure of the model – under conditions (i) and (ii) – implies that only a linear combination of them, ξ_j^*, affects a firm's profit. Second, the rate of substitution between observable z_j and unobservable ξ_j^* in the equilibrium profit function is equal to one.

2.4 Market entry game and information structure

Firms' entry decisions are determined as the result of a game of market entry. The market-specific profit of being inactive is normalized to zero for all firms. Firms may have uncertainty about their profits if active in the market. Firms' information about demand and costs plays a key role in their entry decisions and, therefore, on the selection problem in the estimation of demand. Assumptions 2 and 3 summarize our conditions on the information structure and on the entry cost function, respectively.

Assumption 2. *The information set of firm j at the moment of its entry decision in market m consists of the triple* $(\boldsymbol{x}_m, \boldsymbol{\kappa}_m, \eta_{jm})$. *(1)* $\boldsymbol{x}_m \equiv (\boldsymbol{x}_{jm} : j \in \mathcal{J})$ *is the vector of product characteristics affecting demand and costs that are observable to the researcher. (2)* $\boldsymbol{\kappa}_m \equiv (\kappa_{jm} : j \in \mathcal{J})$ *is a vector of noisy signals for the demand-cost variables* ξ_j^* *such that, for every product j:*

$$\xi_j^* = \kappa_{jm} + e_{jm} \tag{16}$$

where e_{jm} *represents the error or noise in signal* κ_{jm} *and it is independent of* $(\boldsymbol{x}_m, \boldsymbol{\kappa}_m)$. *Both* $\boldsymbol{x}_m$ *and* $\boldsymbol{\kappa}_m$ *are common knowledge for all the firms. (3) Variable* η_{jm} *in the fixed cost function* $f_j(\boldsymbol{x}_{jm}, \eta_{jm})$ *is private information of firm j and independently distributed over firms with CDF* F_η. ■

Assumption 3. *The fixed cost function does not depend on* z_{jm} *and is additive in* η_{jm}:

$$f_j(\boldsymbol{x}_{jm}, \eta_{jm}) = \overline{f}_j(\widetilde{\boldsymbol{x}}_{jm}) + \eta_{jm}. \quad ■ \tag{17}$$

Let $\pi_j(\boldsymbol{a}, \boldsymbol{x}_m, \boldsymbol{\kappa}_m)$ be firm j's expected variable profit given its information about demand and costs, $(\boldsymbol{x}_m, \boldsymbol{\kappa}_m)$, and conditional on the hypothetical entry profile $\boldsymbol{a}$. Under Assumptions 1 to 3, we have that:

$$\begin{aligned} \pi_j(\boldsymbol{a}, \boldsymbol{x}_m, \boldsymbol{\kappa}_m) &= \int V_j(\boldsymbol{a}, \widetilde{\boldsymbol{x}}_m, \boldsymbol{z}_m + \boldsymbol{\xi}_m^*) \; p\left(\boldsymbol{\xi}_m^* | \boldsymbol{x}_m, \boldsymbol{\kappa}_m\right) \; d\boldsymbol{\xi}_m^* - \overline{f}_j(\widetilde{\boldsymbol{x}}_{jm}) \\ &= \int V_j(\boldsymbol{a}, \widetilde{\boldsymbol{x}}_m, \boldsymbol{z}_m + \boldsymbol{\kappa}_m + \boldsymbol{e}_m) \; p\left(\boldsymbol{e}_m\right) \; d\boldsymbol{e}_m - \overline{f}_j(\widetilde{\boldsymbol{x}}_{jm}) \end{aligned} \tag{18}$$

This expression shows that the expected variable profit has structure $\pi_j(\boldsymbol{a}, \widetilde{\boldsymbol{x}}_m, \boldsymbol{z}_m + \boldsymbol{\kappa}_m)$. We establish this property in Lemma 3.

LEMMA 3. *Under Assumptions 1 to 3, given an entry profile* $\boldsymbol{a} = (a_j : j \in \mathcal{J})$, *a firm's expected profit (up to the private information* η_{jm}*) depends on* $(\boldsymbol{x}_m, \boldsymbol{\kappa}_m)$ *according to the structure,* $\pi_j(\boldsymbol{a}, \widetilde{\boldsymbol{x}}_m, \boldsymbol{w}_m)$ *with* $\boldsymbol{w}_m = \boldsymbol{z}_m + \boldsymbol{\kappa}_m$. ■

The presence of firms' private information implies that the entry game is one of incomplete information. Given $(\boldsymbol{x}_m, \boldsymbol{\kappa}_m)$, a *Bayesian Nash Equilibrium (BNE)* in this game can be represented as an J-tuple of entry probabilities, one for each firm, $(P_{jm} : j \in \mathcal{J})$ that satisfies the following system of equilibrium restrictions:

$$P_{jm} = F_\eta\left(\pi_{jm}^P\right) \tag{19}$$

where π^P_{jm} is firm j's expected profit – up to η_{jm} – taking into account the uncertainty about other firms' entry decisions, and given that the other firms make entry choices according to their entry probabilities $(P_{im} : i \neq j)$. That is,

$$\pi^P_{jm} = \sum_{\boldsymbol{a}_{-j} \in \{0,1\}^{J-1}} \left(\prod_{i \neq j} [P_{im}]^{a_i} [1 - P_{im}]^{1-a_i} \right) \pi_j(\boldsymbol{a}_{-j}, \widetilde{\boldsymbol{x}}_m, \boldsymbol{z}_m + \boldsymbol{\kappa}_m) \tag{20}$$

LEMMA 4. *Under Assumptions 1 to 3, the equilibrium choice probabilities, and the expected profit function depend on $(\boldsymbol{x}_m, \boldsymbol{\kappa}_m)$ according to the following structure:*

$$P_{jm} = P_j(\widetilde{\boldsymbol{x}}_m, \boldsymbol{w}_m) \quad \text{and} \quad \pi^P_{jm} = \pi^P_j(\widetilde{\boldsymbol{x}}_m, \boldsymbol{w}_m), \tag{21}$$

with $\boldsymbol{w}_m = \boldsymbol{z}_m + \boldsymbol{\kappa}_m$. ■

This general framework includes as particular cases the different specifications of information structure about the unobservables $(\boldsymbol{\xi}_m, \boldsymbol{\omega}_m, \boldsymbol{\eta}_m)$ appeared in the literature, such as: all the unobservables are common knowledge (Berry, Levinsohn, and Pakes, 1995; Ciliberto, Murry, and Tamer, 2018), firm-specific unobservables are each firm's private information (Seim, 2006; Bajari et al., 2010); and intermediate cases where some unobservables are common knowledge and others are private information (Aradillas-Lopez, 2010; Grieco, 2014; Aguirregabiria and Mira, 2019).

3 Sample Selection Problem

For simplicity and concreteness, we describe our sample selection problem using the nested logit demand model from Example 1. Our econometric model can be described in terms of three equations: (i) an equation for the *latent market share*, s^*_{jm}, and the latent price, p^*_{jm}, under the hypothetical condition that product j is offered in market m,

$$\ln\left(\frac{s^*_{jm}}{s_{0m}}\right) = \sigma \ \ln\left(\frac{s^*_{jm} + S_{-jm}}{s_{0m}}\right) + \alpha \ p^*_{jm} + \boldsymbol{x}'_{jm} \ \boldsymbol{\beta} + \xi_{jm}, \tag{22}$$

with $S_{-jm} \equiv \sum_{i \neq j, i \in r_j} s_{im}$; (ii) a selection condition establishing that market share and price are observable and equal to their latent counterparts only if the product is offered in the market:

$$s^*_{jm} = s_{jm} \text{ and } p^*_{jm} = p_{jm} \iff a_{jm} = 1; \tag{23}$$

and (iii) firms' best response equations in the entry model:

$$a_{jm} = 1\left\{\pi_j^P(\widetilde{\boldsymbol{x}}_m, \boldsymbol{z}_m + \boldsymbol{\kappa}_m) - \eta_{jm} \geq 0\right\}. \tag{24}$$

Given equations (22) to (24), we have the following regression equation for any observation with $a_{jm} = 1$:

$$\ln\left(\frac{s_{jm}}{s_{0m}}\right) = \sigma \ \ln\left(\frac{s_{jm} + S_{-jm}}{s_{0m}}\right) + \alpha \ p_{jm} + \boldsymbol{x}_{jm}' \ \boldsymbol{\beta} + \lambda_j(\boldsymbol{x}_m) + \widetilde{\xi}_{jm}, \tag{25}$$

where $\lambda_j(\boldsymbol{x}_m)$ is the selection term:

$$\begin{aligned} \lambda_j(\boldsymbol{x}_m) \ &\equiv \ \mathbb{E}\left[\xi_{jm} \mid \boldsymbol{x}_m, a_{jm} = 1\right] \\ &= \ \int \xi_{jm} \, 1\left\{\pi_j^P(\widetilde{\boldsymbol{x}}_m, \boldsymbol{z}_m + \boldsymbol{\kappa}_m) - \eta_{jm} \geq 0\right\} \ \frac{f_{\xi,\eta,\boldsymbol{\kappa}}\left(\xi_{jm}, \eta_{jm}, \boldsymbol{\kappa}_m\right)}{\mathbb{P}_j\left(\boldsymbol{x}_m\right)} \ d\left(\xi_{jm}, \eta_{jm}, \boldsymbol{\kappa}_m\right), \end{aligned} \tag{26}$$

and $\mathbb{P}_j\left(\boldsymbol{x}_m\right)$ is the *Conditional Choice Probability* (CCP):

$$\mathbb{P}_j\left(\boldsymbol{x}_m\right) \ \equiv \ \mathbb{E}\left[a_{jm} \mid \boldsymbol{x}_m\right] = \int 1\left\{\pi_j^P(\widetilde{\boldsymbol{x}}_m, \boldsymbol{z}_m + \boldsymbol{\kappa}_m) - \eta_{jm} \geq 0\right\} \ f_{\eta,\boldsymbol{\kappa}}\left(\eta_{jm}, \boldsymbol{\kappa}_m\right) \ d\left(\eta_{jm}, \boldsymbol{\kappa}_m\right), \tag{27}$$

where $f_{\eta,\boldsymbol{\kappa}}$ and $f_{\xi,\eta,\boldsymbol{\kappa}}$ are the joint density functions of $\left(\eta_{jm}, \boldsymbol{\kappa}_m\right)$ and $\left(\xi_{jm}, \eta_{jm}, \boldsymbol{\kappa}_m\right)$, respectively.

Without further restrictions, the selection term $\lambda_j(\boldsymbol{x}_m)$ is a nonparametric function of all its arguments in $\boldsymbol{x}_m$. This implies that the demand model — the parameters σ, α, and $\boldsymbol{\beta}$ — is not identified: we cannot disentangle the direct effect of $\boldsymbol{x}_{jm}$ on demand from its indirect effect through the selection term. We now present two examples that illustrate the relationship between firms' information sets and the identification of the demand model.

Example 2: No signals $\boldsymbol{\kappa}_m$. Suppose that: (1*) $\boldsymbol{\kappa}_m = 0$ such that, at the moment of entry, firms do not have any information about the demand/cost variables $\boldsymbol{\xi}_m^*$; and (2*) a unique equilibrium is played across all market entry games with the same observables $\boldsymbol{x}_m$. Under Assumptions 1 to 3 and conditions (1*) and (2*), the selection term $\lambda_j(\boldsymbol{x}_m)$ only depends on the CCP $\mathbb{P}_j\left(\boldsymbol{x}_m\right)$: that is, $\lambda_j(\boldsymbol{x}_m) = \rho_j(\mathbb{P}_j\left(\boldsymbol{x}_m\right))$ for some function $\rho_j(\cdot)$.

The proof is straightforward. Under conditions (1*)–(2*), the expected profit function π_{jm}^P and the equilibrium choice probabilities only depend on $\boldsymbol{x}_m$ but not on $\boldsymbol{\kappa}_m$. The equilibrium entry probability $P_j\left(\boldsymbol{x}_m\right)$ is equal to the "empirical" probability $\mathbb{P}_j\left(\boldsymbol{x}_m\right) \equiv$

$\mathbb{E}\left(a_{jm}|\boldsymbol{x}_m\right)$. This empirical probability satisfies the equilibrium condition $\mathbb{P}_j(\boldsymbol{x}_m) = F_\eta(\pi_j^P(\boldsymbol{x}_m))$, such that $\pi_j^P(\boldsymbol{z}_m) = F_\eta^{-1}(\mathbb{P}_j(\boldsymbol{x}_m))$, and the entry condition can be represented as $a_{jm} = 1\{F_\eta^{-1}(\mathbb{P}_j(\boldsymbol{x}_m)) - \eta_{jm} \geq 0\}$. Furthermore, the independence between η_{jm} and $\boldsymbol{x}_m$ implies that:

$$
\begin{aligned}
\lambda_j(\boldsymbol{x}_m) &= \int \xi_{jm}\, 1\left\{\eta_{jm} \leq F_\eta^{-1}(\mathbb{P}_j\left(\boldsymbol{x}_m\right))\right\} \frac{f_{\xi,\eta}\left(\xi_{jm},\eta_{jm}\right)}{\mathbb{P}_j\left(\boldsymbol{x}_m\right)}\, d\xi_{jm}\, d\eta_{jm} \\
&= \rho_j(\mathbb{P}_j\left(\boldsymbol{x}_m\right))
\end{aligned}
\tag{28}
$$

Therefore, the demand equation can be represented as:

$$
\ln\left(\frac{s_{jm}}{s_{0m}}\right) = \sigma\, \ln\left(\frac{s_{jm} + S_{-jm}}{s_{0m}}\right) + \alpha\, p_{jm} + \boldsymbol{x}_{jm}'\, \boldsymbol{\beta} + \rho_j(\mathbb{P}_j\left(\boldsymbol{x}_m\right)) + \widetilde{\xi}_{jm}, \tag{29}
$$

This result has important implications for identification and estimation. Regression equation (29) is a standard semiparametric partially linear model with two endogenous regressors, $\ln[(s_{jm} + S_{-jm})/s_{0m}]$ and p_{jm}, and with the nonparametric component $\rho_j(\mathbb{P}_j\left(\boldsymbol{x}_m\right))$ only depending on the CCP of firm j. As such, identification and estimation can follow the standard two-step procedure as in, for example, Powell (2001).

In a first step, one can nonparametrically estimate $\boldsymbol{P}(\boldsymbol{x}_m) = \left(\mathbb{P}_1\left(\boldsymbol{x}_m\right), ..., \mathbb{P}_J\left(\boldsymbol{x}_m\right)\right)$ from data on $(\boldsymbol{a}_m, \boldsymbol{x}_m)$. Then, in a second step, by relying on observations from markets m and m' with $\mathbb{P}_j\left(\boldsymbol{x}_m\right) = \mathbb{P}_j\left(\boldsymbol{x}_{m'}\right)$, but with $\ln[(s_{jm} + S_{-jm})/s_{0m}] \neq \ln[(s_{jm'} + S_{-jm'})/s_{0m'}]$, $p_{jm} \neq p_{jm'}$, and $\boldsymbol{x}_{jm} \neq \boldsymbol{x}_{jm'}$, one can identify $(\sigma, \alpha, \boldsymbol{\beta})$ by correcting for both sample selection and endogeneity following Powell (2001).

(i) Sample Selection. Compute the difference $\ln(s_{jm}/s_{0m}) - \ln(s_{jm'}/s_{0m'})$ from (29) to get rid of the nonparametric selection function $\rho_j(\mathbb{P}_j\left(\boldsymbol{x}_m\right))$.

(ii) Endogenity. Apply standard IV arguments using instruments derived from $\boldsymbol{x}_m$ to the linear regression $\ln(s_{jm}/s_{0m}) - \ln(s_{jm'}/s_{0m'})$ on $\ln[(s_{jm} + S_{-jm})/s_{0m}] - \ln[(s_{jm'} + S_{-jm'})/s_{0m'}]$, $p_{jm} - p_{jm'}$, and $\boldsymbol{x}_{jm} - \boldsymbol{x}_{jm'}$. Valid instruments in this IV regression can be the observed characteristics of products other than j (the so-called BLP instruments). ■

Under the conditions in Example 2, the market entry condition has only one unobservable variable – η_{jm} – that enters additively in the inequality that defines the selection/entry decision. The model becomes a standard sample selection model. However, though practically convenient, the conditions in Example 2 are not realistic and likely to be rejected in most empirical applications.[4] In particular, restriction (1*) — firms do not

[4]The model of Example 2 is over-identified, allowing for the testability of its over-identifying restrictions.

have any information about demand/cost variables ξ_{jm}^{*} when making entry decisions – seems very unrealistic. Firms typically have more information about their own business than the researcher, for example about their demand. This restriction is key for the potential misspecification of the selection control function in the estimation of demand.

Example 3. Distribution of $\boldsymbol{\kappa}_m$ has finite support. Consider the model under Assumptions 1 to 3, but with the additional additional restriction that $\boldsymbol{\kappa}_m$ has finite support. Let $P_j(\boldsymbol{x}_m, \boldsymbol{\kappa}_m)$ be the equilibrium probabilities in the entry game in market m. By definition, we have that $P_j(\boldsymbol{x}_m, \boldsymbol{\kappa}_m) \equiv \mathbb{E}[a_{jm}|\boldsymbol{x}_m, \boldsymbol{\kappa}_m]$, and:

on eqn 31and the line below, on eqn 32, and eqn 34.

$$P_j(\boldsymbol{x}_m, \boldsymbol{\kappa}_m) = F_\eta \left(\pi_j^P(\boldsymbol{x}_m, \boldsymbol{\kappa}_m) \right), \tag{30}$$

This expression, together with the invertibility of the CDF F_η, implies a one-to-one relationship between the equilibrium CCPs and equilibrium expected profit: $\pi_j^P(\boldsymbol{x}_m, \boldsymbol{\kappa}_m) = F_\eta^{-1}(P_j(\boldsymbol{x}_m, \boldsymbol{\kappa}_m))$.

This has important implications on the structure of the selection function $\lambda_j(\boldsymbol{x}_m)$. First, note that by definition:

$$\mathbb{P}_j(\boldsymbol{x}_m) = \sum_{\boldsymbol{\kappa}_m \in \mathcal{K}} P_j(\boldsymbol{x}_m, \boldsymbol{\kappa}_m) \, f_{\boldsymbol{\kappa}}(\boldsymbol{\kappa}_m) \tag{31}$$

where $\mathcal{K}$ represents the finite support set of $\boldsymbol{\kappa}_m$. Define $\widetilde{\lambda}_j(\boldsymbol{x}_m, \boldsymbol{\kappa}_m) \equiv \mathbb{E}[\xi_{jm}|\boldsymbol{x}_m, \boldsymbol{\kappa}_m, a_{jm} = 1]$. Similarly, by definition, we have that,

$$\lambda_j(\boldsymbol{x}_m) = \sum_{\boldsymbol{\kappa}_m \in \mathcal{K}} \widetilde{\lambda}_j(\boldsymbol{x}_m, \boldsymbol{\kappa}_m) \, f_{\boldsymbol{\kappa}}(\boldsymbol{\kappa}_m) \tag{32}$$

Also, applying the relationship $\pi_j^P(\boldsymbol{x}_m, \boldsymbol{\kappa}_m) = F_\eta^{-1}(P_j(\boldsymbol{x}_m, \boldsymbol{\kappa}_m))$, we have that:

$$\begin{aligned} \widetilde{\lambda}_j(\boldsymbol{x}_m, \boldsymbol{\kappa}_m) &= \int \xi_{jm} \, 1 \left\{ F_\eta^{-1}(P_j(\boldsymbol{x}_m, \boldsymbol{\kappa}_m)) - \eta_{jm} \geq 0 \right\} \frac{f_{\xi,\eta|\boldsymbol{\kappa}}(\xi_{jm}, \eta_{jm}|\boldsymbol{\kappa}_m)}{P_j(\boldsymbol{x}_m, \boldsymbol{\kappa}_m)} \, d(\xi_{jm}, \eta_{jm}), \\ &\equiv \psi_j \left(P_j(\boldsymbol{x}_m, \kappa_m), \kappa_m \right) \end{aligned} \tag{33}$$

Combining the demand equation with equations (32) and (33), we obtain the regression equation:

$$\ln\left(\frac{s_{jm}}{s_{0m}}\right) = \sigma \, \ln\left(\frac{s_{jm} + S_{-jm}}{s_{0m}}\right) + \alpha \, p_{jm} + \boldsymbol{x}_{jm}' \, \boldsymbol{\beta} + \left[\sum_{\boldsymbol{\kappa}_m \in \mathcal{K}} \psi_j \left(P_j(\boldsymbol{x}_m, \kappa_m), \kappa_m \right) \, f_{\boldsymbol{\kappa}}(\kappa_m) \right] + \widetilde{\xi}_{jm}, \tag{34}$$

An important implication is that the selection term can be represented as a function of an identifiable vector of dimension potentially lower than $\boldsymbol{x}_m$. ■

In the next section, we present identification results on the nonparametric identification of the equilibrium CCPs $P_j(\boldsymbol{x}_m, \kappa_m)$ and the distribution of $\boldsymbol{\kappa}_m$, and given this result, we establish the identification of demand parameters.

4 Identification

4.1 Data and Sequential Identification

Suppose that each of the J firms is a potential entrant in every local market. The researcher observes these firms in a random sample of M markets. For every market m, the researcher observes the vector of exogenous variables $\boldsymbol{x}_m$, and the vectors of firms' entry decisions ($\boldsymbol{a}_m$), prices ($\boldsymbol{p}_m$), and market shares ($\boldsymbol{s}_m$).

Let $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ be the vector that includes all the parameters of the model, where $\boldsymbol{\Theta}$ is the parameter space. This vector has infinite dimension because some of the structural parameters are real-valued functions. The vector $\boldsymbol{\theta}$ has the following components: demand parameters $\boldsymbol{\theta}_\delta \equiv (\alpha, \boldsymbol{\beta}, \sigma)$; probability distribution of demand/cost signals, F_κ; equilibrium choice probabilities, $\boldsymbol{P}_{j,x,\kappa} \equiv (P_j(\boldsymbol{x}, \boldsymbol{\kappa}) : \text{for every } j, \boldsymbol{x}, \boldsymbol{\kappa})$; the probability distribution of private information F_η, and the distribution of unobserved demand conditional on signals, $f_{\xi|\eta,\boldsymbol{\kappa}}$. In summary, $\boldsymbol{\theta} \equiv (\boldsymbol{\theta}_\delta,\ \boldsymbol{P}_{j,x,\kappa},\ f_\kappa,\ f_{\xi|\eta,\boldsymbol{\kappa}},\ F_\eta)$. In this paper, we are interested in the identification of demand parameters $\boldsymbol{\theta}_\delta$ when the distributions f_κ and $f_{\xi|\eta,\boldsymbol{\kappa}}$ and the equilibrium choice probabilities $\boldsymbol{P}_{j,x,\kappa}$ are nonparametrically specified.

We consider a two-step sequential procedure for the identification of $\boldsymbol{\theta}_\delta$, along the lines of the procedure in Example 2 above. First, given the empirical entry probabilities $\mathbb{P}_j(\boldsymbol{x})$, we establish the identification of the equilibrium probabilities $\boldsymbol{P}_{j,x,\kappa}$ and the distribution f_κ. Then, given the structure of the selection function in (34), we show that $\boldsymbol{\theta}_\delta$ is identified using instrumental variables in the context of a partially linear model as in Powell (2001).

4.2 First Step: Game of Market Entry

In this subsection, we present two different results on the identification of $\boldsymbol{P}_{j,x,\kappa}$ and f_κ. The first result restricts the distribution of $\boldsymbol{\kappa}_m$ to have finite support, but this distribution is fully nonparametric. The second result is based on the additivity between $\boldsymbol{z}_m$ and $\boldsymbol{\kappa}_m$

in the expected profit function, as established in Lemma 3. This result applies to either a continuos or discrete distribution of $\boldsymbol{\kappa}_m$ but it limits the number of unknown parameters for the researcher in this distribution.

4.2.1 Identification based on finite support of κ_m

Assumption 4. *The probability distribution of the vector of signals $\boldsymbol{\kappa}_m$ has finite support.* ■

The identification of the equilibrium probabilities $\boldsymbol{P}_{j,x,\kappa}$ and the distribution f_κ is based on the joint conditional probability of entry of the J firms in any market:

$$\mathbb{P}(\boldsymbol{a}_m|\boldsymbol{x}_m) \;=\; \sum_{\kappa=1}^{L} f_\kappa(\kappa|\boldsymbol{x}_m)\left[\prod_{j=1}^{J}\left[P_{j,\boldsymbol{x}_m,\kappa}\right]^{a_{jm}}\;\left[1-P_{j,\boldsymbol{x}_m,\kappa}\right]^{1-a_{jm}}\right]. \tag{35}$$

where L is the number of κ's with $f_\kappa(\kappa) > 0$. This system of equations describes a non-parametric finite mixture model. The identification of this class of models has been studied by Hall and Zhou (2003), Hall et al. (2005), Allman et al. (2009), and Kasahara and Shimotsu (2014). Identification is based on the assumption of independence between firms' entry decisions once we condition on $\boldsymbol{x}_m$ and κ_m. Note that this system of equations does not involve any restriction across equations for different values of $\boldsymbol{x}_m$. We can then study the point-wise identification of this system, for each value of $\boldsymbol{x}$. For notational simplicity, in the remaining part of this subsection we omit any further reference to $\boldsymbol{x}_m$ and to the market subscript m.

The number of components L in the finite mixture (35) is typically unknown to the researcher. Following ideas similar to Bonhomme et al. (2016), Xiao (2018), and Aguirregabiria and Mira (2019), we start our first step identification argument by providing sufficient conditions for the unique determination of L from observable items. In particular, we adapt to our context Proposition 2 in Aguirregabiria and Mira (2019) and Lemma 1 in Xiao (2018).

Suppose that $J \geq 3$ and let (Y_1, Y_2, Y_3) be three random variables that represent a partition of the vector of firms' actions $(a_1, a_2, ..., a_J)$ such that Y_1 is equal to the action of one firm (if J is odd) or two firms (if J is even), and variables Y_2 and Y_3 evenly divide the actions of the rest of the firms. Denote by $\tilde{J}$ the number of firms collected in Y_i, $i = 2, 3$, with $J = 2\tilde{J} + 1$ if J is odd or $J = 2\tilde{J} + 2$ if J is even. For $i = 1, 2, 3$, let $\boldsymbol{P}_{Y_i}(\kappa)$ be the vector of CCPs for each element of Y_i conditional on component κ. The main idea is then to identify L from the observed joint distribution of Y_2 and Y_3:

$$\mathbb{P}(Y_2 = y_2, Y_3 = y_3) \equiv \sum_{\kappa=1}^{L} \Pr(Y_2 = y_2|\kappa)\ \Pr(Y_3 = y_3|\kappa)\ f_\kappa(\kappa)$$

or, in matrix notation,

$$\boldsymbol{P}_{Y_2,Y_3} \ \equiv \boldsymbol{P}_{Y_2|\kappa}\ diag(\boldsymbol{f}_\kappa)\ \boldsymbol{P}'_{Y_3|\kappa}, \tag{36}$$

where: $\boldsymbol{P}_{Y_2,Y_3}$ is the $2^{\tilde{J}} \times 2^{\tilde{J}}$ matrix with elements $\mathbb{P}(y_2, y_3)$; $\boldsymbol{P}_{Y_i|\kappa}$ is the $2^{\tilde{J}} \times L$ matrix with elements $\Pr(Y_i = y|\kappa)$; and $diag(\boldsymbol{f}_\kappa)$ is the $L \times L$ diagonal matrix with the probabilities $f_\kappa(\kappa)$.[5]

LEMMA 5. *Without further restrictions, $Rank(\boldsymbol{P}_{Y_2,Y_3})$ is a lower bound to the true value of parameter L. Furthermore, if (i) $L < 2^{\tilde{J}}$ and (ii) for $i = 2, 3$ the L vectors $\boldsymbol{P}_{Y_i}(\kappa = 1)$, $\boldsymbol{P}_{Y_i}(\kappa = 2)$, ..., $\boldsymbol{P}_{Y_i}(\kappa = L)$ are linearly independent, then $L = Rank(\boldsymbol{P}_{Y_2,Y_3})$.* ■

The point identification of the number of components L from the observed matrix $\boldsymbol{P}_{Y_2,Y_3}$ hinges on a "large enough" number of firms $\tilde{J}$ and on the matrices $\boldsymbol{P}_{Y_2|\kappa}$ and $\boldsymbol{P}_{Y_3|\kappa}$ being of full column rank, so that the CCPs associated to each component κ cannot be obtained as linear combinations of the others.

Given L and a value of $\boldsymbol{x}$, the number of restrictions in (35) is equal to $2^J - 1$ (i.e., the number of non-redundant values of the vector $\boldsymbol{a} = (a_1, a_2, ..., a_J)$), and the number of free parameters is $JL + (L - 1)$, i.e., JL CCPs, and $L - 1$ mixing probabilities. Therefore, a necessary order condition for identification is $2^J \geq (J + 1)L$, or equivalently $L \leq 2^J/(J + 1)$. Even with only two market types, $L = 2$, sequential identification requires more than two firms/products, $J \geq 3$.[6]

Allman et al. (2009) study the identification of nonparametric multinomial finite mixture models that include our binary choice model as a particular case. They establish that a mixture with L components is identified if $J \geq 3$ and $L \leq 2^J/(J + 1)$. The following is an application to our model of Theorem 4 and Corollary 5 in Allman et al. (2009).

PROPOSITION 1. *Suppose that: (i) $J \geq 3$; (ii) $L \leq 2^J/(J + 1)$; and (iii) for $i = 1, 2, 3$, the L vectors $\boldsymbol{P}_{Y_i}(\kappa = 1)$, $\boldsymbol{P}_{Y_i}(\kappa = 2)$, ..., $\boldsymbol{P}_{Y_i}(\kappa = L)$ are linearly independent. Then, the probability*

[5] Remember that if J is odd, then $\tilde{J} = (J - 1)/2$; while if it is even, then $\tilde{J} = (J - 2)/2$.

[6] As shown in Aguirregabiria and Mira (2019), identification can be achieved with $J = 2$ if the sequential identification approach is replaced with a joint identification approach. See Proposition 6 in Aguirregabiria and Mira (2019), and section 5.4 in that paper, where the authors present an example of a two player ($J = 2$) binary choice game with $L = 2$ that is identified.

distribution of κ – $f_\kappa(\kappa)$ for $\kappa = 1, 2, ..., L$ – and the equilibrium CCPs – $P_j(\kappa)$ for $j = 1, 2, ..., J$ and $\kappa = 1, 2, ..., L$ – are uniquely identified up to label swapping ■

Note that the order condition (i) of Lemma 5 is in general more stringent than the order condition (ii) of Proposition 1: that is, for $J \geq 3$, we have that $2^{\tilde{J}} \leq 2^J/(J+1)$. In this sense, for any $J \geq 3$, when the conditions in Lemma 5 hold and the L vectors $\boldsymbol{P}_{Y_1}(\kappa = 1)$, $\boldsymbol{P}_{Y_1}(\kappa = 2)$, ..., $\boldsymbol{P}_{Y_1}(\kappa = L)$ are linearly independent, then $L = \text{Rank}(\boldsymbol{P}_{Y_2,Y_3})$, and the distribution of κ and the equilibrium CCPs are uniquely identified, up to label swapping.

4.2.2 Identification based on additivity between z_m and κ_m in expected profit

Assumption 4′. *(i) The expected profit function $\pi_j^P(\widetilde{\boldsymbol{x}}_m, \boldsymbol{w}_m)$ is strictly monotonic in each of the arguments w_{im} that form the vector $\boldsymbol{w}_m = \boldsymbol{z}_m + \boldsymbol{\kappa}_m$. (ii) $\boldsymbol{\kappa}_m$ is independent of $\boldsymbol{z}_m$ (though it may have dependence with respect to $\widetilde{\boldsymbol{x}}_m$) with a CDF F_κ that is known to the researcher up to a finite vector of parameters $\boldsymbol{\theta}_\kappa$. (iii) η_{jm} is independent of $(\boldsymbol{\kappa}_m, \boldsymbol{z}_m)$ and is i.i.d. with a CDF F_η that is strictly increasing over the real line and is known to the researcher. (iv) $\boldsymbol{z}_m$ is a vector of continuous random variables with support $\mathbb{R}^J$.* ■

Proposition 2 presents our first identification result in our sequential approach.

PROPOSITION 2. *Suppose that Assumptions 1 to 3 and 4′ hold. (A) The equilibrium choice probability functions $\boldsymbol{P}_{j,x,\kappa}$ are nonparametrically identified over the whole support of the variables $(\boldsymbol{x}_m, \boldsymbol{\kappa}_m)$. (B) The vector of parameters $\boldsymbol{\theta}_\kappa$ that defines the CDF of $\boldsymbol{\kappa}_m$ is identified as long as the dimension of $\boldsymbol{\theta}_\kappa$ is not larger than $J - 1$.* ■

Proof of Proposition 2. For notational simplicity, we omit the argument $\widetilde{\boldsymbol{x}}_m$ but all the results below can be interpreted as conditional to an arbitrary value of this vector of observable variables. Define $\mathbb{P}_j(\boldsymbol{z}) \equiv \Pr(a_{jm} = 1 | \boldsymbol{z}_m = \boldsymbol{z}) = \mathbb{E}[a_{jm} | \boldsymbol{z}_m = \boldsymbol{z}]$. It is clear that $\mathbb{P}_j(\boldsymbol{z})$ is nonparametrically identified for every $\boldsymbol{z} \in \mathbb{R}^J$. According to the model:

$$\mathbb{P}_j(\boldsymbol{z}) = \int P_j(z_1 + \kappa_1, ..., z_J + \kappa_J) \, dF_\kappa(\kappa_1, ..., \kappa_J) \tag{37}$$

We want to show the identification of the function $P_j(w_1, ..., w_J)$ for every $(w_1, ..., w_J) \in \mathbb{R}^J$. For the sake of illustration, we start showing identification for the simpler case with $J = 1$. Also, given that the results are for a given firm/product j, we also omit the subindex j.

Proof for $J = 1$. In this case $J - 1 = 0$ such that Proposition 1(B) implies that there is not any $\boldsymbol{\theta}_\kappa$ to identify and the CDF F_κ is known to the researcher. We need to show the identification of $P(w)$ for every $w \in \mathbb{R}^J$.

The entry equilibrium condition is $a_m = 1\{\ \pi^P(z_m + \kappa_m) - \eta_m \geq 0\}$. Strict monotonicity of the expected profit function implies that we can represent the equilibrium condition as $a_m = 1\{\ v_m \leq z_m\}$, where v_m is the random variable equal to $\pi^{-1}(\eta_m) - \kappa_m$, and $\pi^{-1}(.)$ is the inverse function of $\pi^P(.)$. Therefore, for any $z \in \mathbb{R}$, we have that $\mathbb{P}(z) = F_v(z)$, where F_v is the CDF of v_m. This means that F_v is identified over the whole real line. Next, given the CDFs F_v and F_κ we identify the CDF of the random variable $\widetilde{\eta}_m \equiv \pi^{-1}(\eta_m)$ applying deconvolution properties (see Horowitz, 1998, chapter 4). For $t \in \mathbb{R}$, let $\varphi_v(t)$, $\varphi_{\widetilde{\eta}}(t)$, and $\varphi_\kappa(t)$ be the characteristic function of the variables v, $\widetilde{\eta}$ and κ, respectively. By definition, for any random variable x with CDF F_x, we have that $\varphi_x(t) = \int_{-\infty}^{+\infty} \exp\{i\ t\ x\}$ $dF_x(x)$. Independence between $\widetilde{\eta}$ and κ, together with $v = \widetilde{\eta} - \kappa$, implies that $\varphi_v(t) = \varphi_{\widetilde{\eta}}(t) / \varphi_\kappa(t)$, such that $\varphi_{\widetilde{\eta}}(t) = \varphi_v(t)\ \varphi_\kappa(t)$. Taking into account the relationship between the characteristic function and the CDF of a continuous random variable, we have that for any value $u_0 \in \mathbb{R}$:

$$F_{\widetilde{\eta}}(u_0) = \int_{-\infty}^{u_0} \int_{-\infty}^{+\infty} \frac{1}{2\pi} \exp\{-i\ t\ u\}\ \ \varphi_v(t)\ \varphi_\kappa(t)\ dt\ du \tag{38}$$

with $\varphi_v(t) = \int_{-\infty}^{+\infty} \exp\{i\ t\ v\}\ dF_v(v)$ and $\varphi_\kappa(t) = \int_{-\infty}^{+\infty} \exp\{i\ t\ \kappa\}\ dF_\kappa(\kappa)$. Therefore, equation (38) provides a closed-form expression representation of the identification of $F_{\widetilde{\eta}}$ given F_v and F_κ. Finally, given the CDF $F_{\widetilde{\eta}}$, we have that for any value $w \in \mathbb{R}$, $F_{\widetilde{\eta}}(w) = \Pr(\pi^{-1}(\eta_m) \leq w) = \Pr(\eta_m \leq \pi^P(w)) = F_\eta(\pi^P(w)) = P(w)$ such that $P(w)$ is identified as $P(w) = F_{\widetilde{\eta}}(w)$. Or taking into account equation (38) and $F_v(z) = \mathbb{P}(z)$, for any value $w \in \mathbb{R}$:

$$P(w) = \int_{-\infty}^{u_0} \int_{-\infty}^{+\infty} \frac{1}{2\pi} \exp\{-i\ t\ u\}\ \ \varphi_\kappa(t)\ \left[\int_{-\infty}^{+\infty} \exp\{i\ t\ z\}\ d\mathbb{P}(z)\right]\ dt\ du \tag{39}$$

Note that in this proof we have not used any parametric restriction on the distribution of the unobserved private information η_m.

Proof for $J > 1$. For the moment, consider that the distribution of the vector $\boldsymbol{\kappa}_m$ is known to the researcher. The entry equilibrium condition is $a_m = 1\{\ \pi^P(z_{1m} + \kappa_{1m}, ..., z_{Jm} + \kappa_{Jm}) - \eta_m \geq 0\}$. Strict monotonicity of the expected profit function with respect to each of the J arguments, say argument i, implies that we can represent the equilibrium condition as $a_m = 1\{v_{im} \leq z_{im}\}$, where v_{im} is the random variable $\widetilde{\eta}_{im} - \kappa_{im}$, with $\widetilde{\eta}_{im} \equiv \pi^{-1(i)}(\eta_m, \boldsymbol{z}_{-im} + \kappa_{-im})$ and $\pi^{-1(i)}$ is the inverse function of $\pi^P(.)$ with respect to its $i-th$ argument. Note that this inverse function not only depends on η_m but also on the vector $\boldsymbol{z}_{-im} + \boldsymbol{\kappa}_{-im}$ with elements $z_{\ell m} + \kappa_{\ell m}$ for every $\ell \neq i$. For this reason, the random variable v_{im} depends on $\boldsymbol{z}_{-im}$, and we use $F_{v_i|\boldsymbol{z}_{-i}}$ to represent the CDF of v_{im} conditional on $\boldsymbol{z}_{-im}$. Given the representation $a_m = 1\{v_{im} \leq z_{im}\}$, we have that for any vector $\boldsymbol{z} \in \mathbb{R}^J$, $\mathbb{P}(\boldsymbol{z}) = \Pr(v_{im} \leq z_i | \boldsymbol{z}_m = \boldsymbol{z}) = F_{v_i|\boldsymbol{z}_{-i}}(z_i)$. This means that the conditional distribution $F_{v_i|\boldsymbol{z}_{-i}}(u)$ is identified everywhere, for any $u \in \mathbb{R}$, and $\boldsymbol{z}_{-i} \in \mathbb{R}^{J-1}$.

4.3 Second Step: Identification of Demand Parameters

Following the discussion in section 2.2, we represent the demand system using the inverse $d_j^{(\boldsymbol{a})-1}(\boldsymbol{s}_m^{(\boldsymbol{a})}, \boldsymbol{p}_m^{(\boldsymbol{a})}, \boldsymbol{x}_m^{(\boldsymbol{a})})$ from Lemma 1. Among those markets with $a_{jm} = 1$, this system can be expressed as:

$$\delta_{jm}(\boldsymbol{\sigma}) = \alpha\, p_{jm} + \boldsymbol{x}'_{jm}\, \boldsymbol{\beta} + \xi_{jm}, \tag{40}$$

where we use the notation $\delta_{jm}(\boldsymbol{\sigma})$ to emphasize that δ_{jm} is a function of the parameters $\boldsymbol{\sigma}$ characterizing the distribution of the random coefficients v_h. The selection problem appears because the unobservable ξ_{jm} is not mean independent of the market entry (or product availability) condition $a_{jm} = 1$. Therefore, moment conditions that are valid under exogenous product selection may no longer be valid when ξ_{jm} and a_{jm} are not independent.

Suppose for a moment that the market type κ_m were observable to the researcher after identification in the first step. In this case, the selection term would be $\psi_j\left(P_j(\boldsymbol{x}_m, \kappa_m), \kappa_m\right)$ from equation (33) and we would have — as in Example 2 — a relatively standard selection problem represented by the semiparametric partially linear model:

$$\delta_{jm}(\boldsymbol{\sigma}) = \alpha\, p_{jm} + \boldsymbol{x}'_{jm}\, \boldsymbol{\beta} + \psi_j\left(P_j(\boldsymbol{x}_m, \kappa_m), \kappa_m\right) + e_{jm}. \tag{41}$$

A key complication of the selection problem in our model is that the market type κ_m is unobserved to the researcher. After the first step identification, we do not know the unobserved type of a market but only its probability distribution conditional on $\boldsymbol{x}_m$. Therefore, in the second step we cannot condition on κ_m as in Example 2 or (41) and, as derived in equation (33) of Example 3, we instead need to deal with the more complex selection function:

$$\lambda_j(\boldsymbol{x}_m) \quad \equiv \mathbb{E}\left[\xi_{jm}|\boldsymbol{x}_m, a_{jm}=1\right] = \sum_{\kappa=1}^{L} f_\kappa(\kappa)\, \psi_j\left(P_j(\boldsymbol{x}_m,\kappa),\kappa\right) = \boldsymbol{f}'_\kappa \boldsymbol{\psi}_j(\boldsymbol{P}_{jm}), \tag{42}$$

where $\boldsymbol{f}_\kappa$, $\boldsymbol{P}_{jm}$, and $\boldsymbol{\psi}_j(\boldsymbol{P}_{jm})$ are all vectors of dimension $L \times 1$ such that $\boldsymbol{f}_\kappa \equiv (f_\kappa(\kappa) : \kappa = 1,2,...,L)$, $\boldsymbol{P}_{jm} \equiv (P_j(\boldsymbol{x}_m,\kappa) : \kappa = 1,2,...,L)$, and $\boldsymbol{\psi}_j(\boldsymbol{P}_{jm}) \equiv (\psi_j\left[P_j(\boldsymbol{x}_m,\kappa),\kappa\right] : \kappa = 1,2,...,L)$.

An important implication of Assumption 4 is that the selection term can be represented as a function of an identifiable vector of dimension potentially lower than $\boldsymbol{x}_m$. Let us denote $\boldsymbol{P}_{jm} \equiv (P_j(\boldsymbol{x}_m,\kappa) : \kappa = 1,2,...,L)$, the vector collecting the CCPs identified in the first step. Then, equation (42) implies:

$$\mathbb{E}\left[\xi_{jm}|\boldsymbol{P}_{jm}, a_{jm}=1\right] = \ \mathbb{E}\left[\mathbb{E}\left[\xi_{jm}|\boldsymbol{x}_m, a_{jm}=1\right]|\boldsymbol{P}_{jm}, a_{jm}=1\right] = \boldsymbol{f}'_\kappa \boldsymbol{\psi}_j(\boldsymbol{P}_{jm}) \tag{43}$$

We are interested in the identification of the true demand parameters $\boldsymbol{\theta}_\delta \equiv (\alpha, \boldsymbol{\beta}, \boldsymbol{\sigma})$ given a nonparametric specification of the functions $\psi_j(P_j(\boldsymbol{x}_m,\kappa),\kappa)$, $\kappa = 1,2,...,L$. As in Example 2, we follow the pairwise differencing strategy proposed by Ahn and Powell (1993), Powell (2001), and Aradillas-Lopez et al. (2007). For firm j active in markets m and n with $\boldsymbol{P}_{jm} = \boldsymbol{P}_{jn}$, one can eliminate the nonparametric selection term by differencing it out. Identification of $\boldsymbol{\theta}_\delta$ then hinges on the availability of valid instruments both to address the problem of price endogeneity and to pin down the parameters of the distribution of random coefficients, $\boldsymbol{\sigma}$. Define the vector of regressors as $\widetilde{\boldsymbol{x}}_{jm} = (p_{jm}, \boldsymbol{x}'_{jm})'$ and the associated demand parameters as $\widetilde{\boldsymbol{\beta}} \equiv (\alpha, \boldsymbol{\beta})$, then for any two markets m and n such that $\boldsymbol{P}_{jm} = \boldsymbol{P}_{jn}$:

$$\delta_{jm}(\boldsymbol{\sigma}) - \delta_{jn}(\boldsymbol{\sigma}) = (\widetilde{\boldsymbol{x}}_{jm} - \widetilde{\boldsymbol{x}}_{jn})'\widetilde{\boldsymbol{\beta}} + (e_{jm} - e_{jn}). \tag{44}$$

Define the vector of instruments as $\widetilde{\boldsymbol{\zeta}}_{jm} = (\boldsymbol{\zeta}'_{jm}, \boldsymbol{x}'_{jm})'$, where $\boldsymbol{\zeta}_{jm}$ is constructed from $\boldsymbol{x}_m$, and the moment function obtained from (44) by $\boldsymbol{g}(e_{jm}, e_{jn}, \boldsymbol{\theta}_\delta) = (\widetilde{\boldsymbol{\zeta}}_{jm} - \widetilde{\boldsymbol{\zeta}}_{jn})(e_{jm} - e_{jn})$. Our second step identification is then based on the moment conditions:

$$\mathbb{E}[\boldsymbol{g}(e_{jm}, e_{jn}, \boldsymbol{\theta}^0_\delta)|\boldsymbol{P}_{jm} = \boldsymbol{P}_{jn}, a_{jm} = a_{jn} = 1] = 0, \tag{45}$$

whose Jacobian matrix can be expressed as:

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}_\delta) = \frac{\partial \mathbb{E}[\boldsymbol{g}(e_{jm}, e_{jn}, \boldsymbol{\theta}_\delta)|\boldsymbol{P}_{jm} = \boldsymbol{P}_{jn}, a_{jm} = a_{jn} = 1]}{\partial \boldsymbol{\theta}'_\delta}. \tag{46}$$

The following is a sufficient condition commonly used for identification in semiparametric partially linear models, see for example, Chamberlain (1986), Robinson (1988), Coslett (1991), Ahn and Powell (1993), Powell (2001), Aradillas-Lopez et al. (2007), or Newey (2009).

Assumption 5. *The following conditions hold. (A) Exogeneity of the instruments:* $\mathbb{E}[e_{jm} - e_{jn} \mid \widetilde{\boldsymbol{\zeta}}_{jm} - \widetilde{\boldsymbol{\zeta}}_{jn}, \boldsymbol{P}_{jm} = \boldsymbol{P}_{jn}, a_{jm} = a_{jn} = 1] = 0$. *(B) The conditional distribution of* $\boldsymbol{P}_{jm}$ *given* $a_{jm} = 1$ *is absolutely continuous with respect to the Lebesgue measure and bounded from above. (C) The Jacobian matrix* $\boldsymbol{\Sigma}(\boldsymbol{\theta}_\delta)$ *defined in equation (46) has constant and full column rank in a neighborhood of the true demand parameters* $\boldsymbol{\theta}_\delta$. ■

PROPOSITION 3. *Given Assumptions 1 to 5, the true value of the demand parameters* $\boldsymbol{\theta}_\delta \equiv (\alpha, \boldsymbol{\beta}, \boldsymbol{\sigma})$ *is locally identified.* ■

5 Estimation and Inference

In this section, we describe an estimation method based on our identification result in Proposition 1. Similar to identification, estimation also unfolds in two sequential steps. In the first step, we initially determine the number of unobserved market types $L(\boldsymbol{x}_m)$ and then estimate $\boldsymbol{f}_\kappa$ and $\boldsymbol{P}_{jm}$ building on the non-parametric procedure proposed by Xiao (2018). The second step instead follows Aradillas-Lopez et al. (2007): we use our first step $\widehat{\boldsymbol{f}}_\kappa$ and $\widehat{\boldsymbol{P}}_{jm}$ to construct an estimate of $\boldsymbol{P}_{jm}$ necessary to difference out the selection term $\boldsymbol{f}_\kappa' \boldsymbol{\psi}_j(\boldsymbol{P}_{jm})$ as in (44) and to obtain estimates of the true demand parameters $\boldsymbol{\theta}_\delta$.

5.1 First Step: Estimation of $\boldsymbol{f}_\kappa$ and $\boldsymbol{P}_{jm}$

In this section, we discuss the estimation of $\boldsymbol{f}_\kappa$ and $\boldsymbol{P}_{jm}$ from data on firms' entry decisions across markets. For this, we build on the nonparametric procedure proposed by Xiao Xiao (2018). We describe how to determine $L(\boldsymbol{x}_m) = \text{Rank}(\boldsymbol{P}_{Y_2,Y_3}(\boldsymbol{x}_m))$ and then — for given L — the nonparametric estimation of $\boldsymbol{f}_{\kappa m}$ and $\boldsymbol{P}_{jm}$. Note that, even though the two-step estimator proposed by Xiao (2018) is not robust to the general form of unobserved heterogeneity considered here, her first step is still valid for the estimation of $\boldsymbol{f}_{\kappa m}$ and $\boldsymbol{P}_{jm}$. As discussed by Aguirregabiria and Mira (2019), the identification up to label swapping in the first step would — in the context of our model — cause problems in the second step of Xiao (2018)'s procedure, i.e. the estimation of firms' profit functions. As illustrated in the previous section though, for the purpose of correcting demand estimates from sample selection, we only need estimates of $\boldsymbol{f}_{\kappa m}$ and $\boldsymbol{P}_{jm}$.

Employing the sequence of rank tests implemented by Xiao (2018) on the basis of Robin and Smith (2000), we illustrate how to determine $L(\boldsymbol{x}_m) = \text{Rank}(\boldsymbol{P}_{Y_2,Y_3}(\boldsymbol{x}_m))$. Xiao (2018) uses frequency estimators and assumes that $\boldsymbol{x}_m$ is discrete. We instead allow for $\boldsymbol{x}_m$ to contain both discrete and continuous variables. Denote the sub-vector of discrete variables in $\boldsymbol{x}_m$ by $\boldsymbol{x}_m^d$ and the continuous one by $\boldsymbol{x}_m^c$, where $\boldsymbol{x}_m^c \in \mathbb{R}^q$. For any realization $\boldsymbol{x}_m = \boldsymbol{x} \equiv (\boldsymbol{x}^c, \boldsymbol{x}^d)$, we estimate each element of the matrix of joint probabilities $\boldsymbol{P}_{Y_2,Y_3}(\boldsymbol{x})$ by:

$$\widehat{\Pr}(Y_{2m} = y_2, Y_{3m} = y_3, \boldsymbol{x}_m = \boldsymbol{x}) = \frac{\sum_{m=1}^{M} 1\left(Y_{2m} = y_2, Y_{3m} = y_3, \boldsymbol{x}_m^d = \boldsymbol{x}^d\right) K_{b_1}(\boldsymbol{x}_m^c, \boldsymbol{x}^c)}{\sum_{m=1}^{M} 1\left(\boldsymbol{x}_m^d = \boldsymbol{x}^d\right) K_{b_1}(\boldsymbol{x}_m^c, \boldsymbol{x}^c)}, \tag{47}$$

where $K_{b_1}(\boldsymbol{x}_m^c, \boldsymbol{x}^c) = \prod_{s=1}^{q} k\left(\frac{x_{ms}^c - x_s^c}{b_1}\right)$ and $k(\cdot)$ is a kernel function, b_1 a bandwidth, and $1(\cdot)$ the indicator function. In the remaining part of this section, we denote by $\widehat{\boldsymbol{P}}$ an estimator of matrix $\boldsymbol{P}$ along the lines of (47) and, when our arguments are conditional on a specific realization of $\boldsymbol{x}_m$, for notational simplicity we omit the reference to $\boldsymbol{x}_m$. Each rank test in the sequence $r = 1, ..., 2^{\tilde{J}} - 1$ has null hypothesis $H_0^r : \text{Rank}(\boldsymbol{P}_{Y_2,Y_3}) = r$ against the alternative $H_1^r : \text{Rank}(\boldsymbol{P}_{Y_2,Y_3}) > r$ and is based on a characteristic root statistic denoted by CRT_r. The sequence of tests starts with a null hypothesis of rank equal to $r = 1$. If the null is rejected, then $r = 2$ and the test is repeated, and so on. Along the sequence of tests, $\text{Rank}(\boldsymbol{P}_{Y_2,Y_3})$ is estimated to equal that value of r for which H_0^r obtains the first rejection.

In order to guarantee weak consistency of the rank estimator, Xiao (2018) adjusts the asymptotic size α_r of the test at each stage r to depend on the sample size M, α_{rM}. The critical region at stage r corresponding to the adjusted critical value $c_{1-\alpha_{rM}}^r$ is then denoted by $\{CRT_r \geq c_{1-\alpha_{rM}}^r\}$. Finally, the estimator of L is defined as:

$$\widehat{L} \equiv \min_{r \in \{1,...,2^{\tilde{J}}-1\}} \{r : CRT_r \geq c_{1-\alpha_{iM}}^i, i = 1, ..., r-1, CRT_r < c_{1-\alpha_{rM}}^r\}. \tag{48}$$

Xiao (2018) shows in her Appendix C the weak consistency of $\widehat{L}$. For more detail on rank tests, see also Robin and Smith (2000) and Kleibergen and Paap (2006).[7]

[7]Given the well known difficulties in characterizing the asymptotic distribution of rank estimators, in what follows we consider $\widehat{L}$ as a tool to facilitate model selection. In this sense, we interpret the weak consistency of $\widehat{L}$ as consistent model selection. See Xiao (2018) for a discussion, especially in relation to the next part of the estimation procedure.

Once the number of components L is determined, Xiao (2018)'s nonparametric estimator of $\boldsymbol{f}_{\kappa,m}$ and $\boldsymbol{P}_{jm}$ requires "collapsing" the $2^{\tilde{J}} \times 2^{\tilde{J}}$ matrix $\boldsymbol{P}_{Y_2,Y_3}$ of rank L into a smaller $L \times L$ non-singular matrix $\boldsymbol{P}_{\tilde{Y}_2,\tilde{Y}_3}$. The non-singular matrix $\boldsymbol{P}_{\tilde{Y}_2,\tilde{Y}_3}$ can be obtained from $\boldsymbol{P}_{Y_2,Y_3}$ by summing up some of its columns and rows, and Lemma 2 in Xiao (2018) proves that such a transformation is always possible given our assumptions.[8] Given $\boldsymbol{P}_{\tilde{Y}_2,\tilde{Y}_3}$, we define the full rank matrices $\boldsymbol{P}_{\tilde{Y}_2|\kappa}$ and $\boldsymbol{P}_{\tilde{Y}_3|\kappa}$ as in (36), the diagonal matrix $\boldsymbol{P}_{Y_1=a|\kappa} \equiv \text{diag}\left[\Pr(Y_1 = a|\kappa = 1), ..., \Pr(Y_1 = a|\kappa = L)\right]$, and the observed matrix $\boldsymbol{P}_{Y_1=a,\tilde{Y}_2,\tilde{Y}_3}$ as:

$$\boldsymbol{P}_{Y_1=a,\tilde{Y}_2,\tilde{Y}_3} \equiv \boldsymbol{P}_{\tilde{Y}_2|\kappa}\, \boldsymbol{P}_{Y_1=a|\kappa}\, diag\{\boldsymbol{f}_\kappa\}\, \boldsymbol{P}'_{\tilde{Y}_3|\kappa},$$

where all the matrices are of dimension $L \times L$. On the basis of the following eigen-decomposition:

$$\boldsymbol{P}_{Y_1=a,\tilde{Y}_2,\tilde{Y}_3}\, \boldsymbol{P}^{-1}_{\tilde{Y}_2,\tilde{Y}_3} = \boldsymbol{P}_{\tilde{Y}_2|\kappa}\, \boldsymbol{P}_{Y_1=a|\kappa}\, \boldsymbol{P}^{-1}_{\tilde{Y}_2|\kappa}, \tag{49}$$

we estimate $\boldsymbol{P}_{\tilde{Y}_2|\kappa}$ as the $L \times L$ eigenvector matrix:

$$\widehat{\boldsymbol{P}}_{\tilde{Y}_2|\kappa} = \rceil \left(\widehat{\boldsymbol{P}}_{Y_1=a,\tilde{Y}_2,\tilde{Y}_3}\, \widehat{\boldsymbol{P}}^{-1}_{\tilde{Y}_2,\tilde{Y}_3} \right), \tag{50}$$

where the operator $\rceil(\cdot)$ denotes the eigenvector function. Note that here the scale is determined by imposing that each column of $\boldsymbol{P}_{\tilde{Y}_2|\kappa}$ is a probability distribution that must sum up to 1. Given $\widehat{\boldsymbol{P}}_{\tilde{Y}_2|\kappa}$, we can estimate $\boldsymbol{f}_\kappa$ and $\boldsymbol{P}_{\tilde{Y}_3|\kappa}$, respectively, by:

$$\begin{aligned} \widehat{\boldsymbol{f}}_\kappa &= \widehat{\boldsymbol{P}}^{-1}_{\tilde{Y}_2|\kappa}\, \widehat{\boldsymbol{P}}_{\tilde{Y}_2} \\ \widehat{\boldsymbol{P}}_{\tilde{Y}_3|\kappa} &= \left[\left(\widehat{\boldsymbol{P}}_{\tilde{Y}_2|L}\, diag\{\widehat{\boldsymbol{f}}_\kappa\} \right)^{-1}\, \widehat{\boldsymbol{P}}_{\tilde{Y}_2,\tilde{Y}_3} \right]', \end{aligned} \tag{51}$$

where $\boldsymbol{P}_{\tilde{Y}_2}$ is the observed $L \times 1$ vector of marginal probabilities $\Pr(\tilde{Y}_2 = \tilde{y}_2)$, $\forall \tilde{y}_2$, $\mathbf{1}$ is a $L \times 1$ vector of 1's.

In the last step of the procedure, we estimate the CCPs for firm j on the basis of the two following systems of equations:

$$\begin{aligned} \boldsymbol{P}_{\tilde{Y}_2,j} &= \boldsymbol{P}_{\tilde{Y}_2|\kappa}\, diag\{\boldsymbol{f}_\kappa\}\, \boldsymbol{P}'_{j,\kappa} \quad \text{for any } a_j \text{ part of } Y_1 \text{ or } \tilde{Y}_3 \\ \boldsymbol{P}_{j,\tilde{Y}_3} &= \boldsymbol{P}_{j,\kappa}\, diag\{\boldsymbol{f}_\kappa\}\, \boldsymbol{P}'_{\tilde{Y}_3|\kappa} \quad \text{for any } a_j \text{ part of } \tilde{Y}_2, \end{aligned} \tag{52}$$

[8]Given that for any $\boldsymbol{P}_{Y_2,Y_3}$ there are many ways of constructing $L \times L$ matrices, Xiao (2018) suggests to create various candidates and then to pick the one associated to the smallest *condition number* (the smaller the condition number of a matrix, the more likely the matrix is to be non-singular).

where $\boldsymbol{P}_{\tilde{Y}_2,j}$ is the observed $L \times 2$ vector of joint probabilities $\Pr(\tilde{Y}_2 = \tilde{y}_2, a_j = a)$, $\forall \tilde{y}_2$, $a \in \{0,1\}$, $\boldsymbol{P}_{j,\tilde{Y}_3}$ is the analogous observed $2 \times L$ vector, and $\boldsymbol{P}_{j,\kappa}$ is the vector of firm j's CCPs. $\boldsymbol{P}_{j,\kappa}$ can finally be estimated for any firm j as:

$$\begin{aligned} \widehat{\boldsymbol{P}}_{j,\kappa} &= \left[\left(\widehat{\boldsymbol{P}}_{\tilde{Y}_2|\kappa}\, diag\{\widehat{\boldsymbol{f}}_\kappa\}\right)^{-1}\ \widehat{\boldsymbol{P}}_{\tilde{Y}_2,j}\right]' && \text{for any } a_j \text{ part of } Y_1 \text{ or } \tilde{Y}_3 \\ \widehat{\boldsymbol{P}}_{j,\kappa} &= \widehat{\boldsymbol{P}}_{j,\tilde{Y}_3}\left(diag\{\widehat{\boldsymbol{f}}_\kappa\}\ \widehat{\boldsymbol{P}}'_{\tilde{Y}_3|\kappa}\right)^{-1} && \text{for any } a_j \text{ part of } \tilde{Y}_2. \end{aligned} \tag{53}$$

Xiao (2018) characterizes the asymptotic properties of the procedures described here and shows that the proposed estimator is $\sqrt{M}$-consistent and asymptotically normal when $\boldsymbol{x}_m$ is discrete. To reduce estimation bias in the first step and guarantee a parametric rate of convergence of the second step estimator, we rely on higher order kernels in the first step estimator.

5.2 Second Step: Estimation of $\boldsymbol{\theta}_\delta$

Here we describe a GMM estimator $\widehat{\boldsymbol{\theta}}_\delta$ of the true demand parameters $\boldsymbol{\theta}_\delta^0 = (\widetilde{\boldsymbol{\beta}}^0, \boldsymbol{\sigma}^0)$ from demand model (44) that builds on Aradillas-Lopez et al. (2007). We refer the reader to Ahn and Powell (1993), Powell (2001), as well as Aradillas-Lopez (2012) for additional details on the original method. We illustrate that the second step estimator $\widehat{\boldsymbol{\theta}}_\delta$ achieves a parametric rate of convergence and is asymptotically normal despite the nonparametric first step estimator of $\boldsymbol{P}_{jm}$, $\widehat{\boldsymbol{P}}_{jm} \equiv \left(\widehat{P}_j(\boldsymbol{x}_m, \kappa) : \kappa = 1, 2, ..., L\right)$.

In demand model (44), the nonparametric selection term $\boldsymbol{f}'_{\kappa m}\boldsymbol{\psi}_j(\boldsymbol{P}_{jm})$ is differenced out relying on a pair of markets m and n where firm j is active and such that $\boldsymbol{f}'_{\kappa,m}\boldsymbol{\psi}_j(\boldsymbol{P}_{jm}) = \boldsymbol{f}'_{\kappa,n}\boldsymbol{\psi}_j(\boldsymbol{P}_{jn})$. In practice, there may be no such pair of markets for any firm j and this exact differencing may not be possible. To overcome this practical difficulty, Aradillas-Lopez et al. (2007) propose a GMM estimator on the basis of the differences between *all* possible pairs of markets (m, n) in which firm j is active, with each pair weighed according to the "distance" between the estimated $\boldsymbol{P}_{jm}$ and $\boldsymbol{P}_{jn}$:

$$\mathcal{D}_{jmn} = \frac{1}{(b_2)^L}\prod_{\kappa=1}^{L}\tilde{k}\left(\frac{\widehat{P}_{j,x_m,\kappa} - \widehat{P}_{j,x_n,\kappa}}{b_2}\right), \tag{54}$$

where $\tilde{k}(\cdot)$ is a kernel function, and b_2 is a bandwidth.

Given demand model (44), moment conditions (45), and weights as in (54), building on Aradillas-Lopez et al. (2007) we propose to estimate $\boldsymbol{\theta}_\delta^0$ by GMM as follows:

$$\widehat{\boldsymbol{\theta}}_\delta = \arg\min_{\boldsymbol{\theta}_\delta} \left(\sum_{j=1}^{J} \binom{N_j}{2} \right)^{-1} \sum_{j=1}^{J} \sum_{\substack{m,n \in \mathcal{N}_j \\ m<n}} \mathcal{D}_{jmn}\, \boldsymbol{g}(e_{jm}, e_{jn}, \boldsymbol{\theta}_\delta)'\, \boldsymbol{\Omega}\, \boldsymbol{g}(e_{jm}, e_{jn}, \boldsymbol{\theta}_\delta), \tag{55}$$

where $N_j = \sum_{m=1}^{M} a_{jm}$ is the number of markets where firm j is active, $\mathcal{N}_j$ the collection of these N_j markets, $\boldsymbol{g}(e_{jm}, e_{jn}, \boldsymbol{\theta}_\delta) = (\widetilde{\boldsymbol{\zeta}}_{jm} - \widetilde{\boldsymbol{\zeta}}_{jn})(e_{jm} - e_{jn})$ the moment function obtained from demand model (44), and $\boldsymbol{\Omega}$ a positive definite symmetric weighting matrix.

Because kernel methods as those used here are not accurate in the tails of distributions, following Robinson (1988) and Aradillas-Lopez et al. (2007), we trim the extreme realizations of $\boldsymbol{x}_m$ to reduce the bias in the nonparametric first step estimator $\widehat{\boldsymbol{P}}_{jm}$. Accordingly, our second step GMM estimator becomes:

$$\widehat{\boldsymbol{\theta}}_\delta = \arg\min_{\boldsymbol{\theta}_\delta} \frac{1}{\widetilde{M}} \sum_{j=1}^{J} \sum_{\substack{m,n \in \mathcal{N}_j \\ m<n}} \mathcal{D}_{jmn}\, \boldsymbol{g}(e_{jm}, e_{jn}, \boldsymbol{\theta}_\delta)'\, \boldsymbol{\Omega}\, \boldsymbol{g}(e_{jm}, e_{jn}, \boldsymbol{\theta}_\delta)\, \phi(\boldsymbol{x}_{jm})\, \phi(\boldsymbol{x}_{jn}), \tag{56}$$

where $\widetilde{M} = \left(\sum_{j=1}^{J} \binom{N_j}{2} \right)$ and $\phi(\cdot)$ is a trimming function.

Note that, for nested logit model (25) with $\ln(s_{jm}/s_{0m}) = \delta_{jm}$, in the just-identified case (56) simplifies to the linear IV estimator proposed by Powell (2001):

$$\widehat{\boldsymbol{\theta}}_\delta = \left[\left(\sum_{j=1}^{J} \binom{N_j}{2} \right)^{-1} \widehat{\boldsymbol{\Sigma}}_{\zeta x} \right]^{-1} \times \left[\left(\sum_{j=1}^{J} \binom{N_j}{2} \right)^{-1} \widehat{\boldsymbol{\Sigma}}_{\zeta \delta} \right] \tag{57}$$

where

$$\widehat{\boldsymbol{\Sigma}}_{\zeta x} = \sum_{j=1}^{J} \sum_{m=1}^{N_j-1} \sum_{n=m+1}^{N_j} \mathcal{D}_{jmn}\, (\boldsymbol{\zeta}_{jm} - \boldsymbol{\zeta}_{jn})\, (\tilde{\boldsymbol{x}}_{jm} - \tilde{\boldsymbol{x}}_{jn})'$$

$$\widehat{\boldsymbol{\Sigma}}_{\zeta \delta} = \sum_{j=1}^{J} \sum_{m=1}^{N_j-1} \sum_{n=m+1}^{N_j} \mathcal{D}_{jmn}\, (\boldsymbol{\zeta}_{jm} - \boldsymbol{\zeta}_{jn})\, (\delta_{jm} - \delta_{jn})'.$$

6 Monte Carlo Experiments

In this section, we present results from several Monte Carlo experiments. The purpose of these experiments is threefold. First, we want to evaluate the performance of the proposed estimation method to deal with sample selection. We are particularly interested in evaluating the loss of precision due to the nonparametric approach. Second, we are interested in measuring the magnitude of the biases associated to different forms of misspecification of the model. Finally, we compare our method with alternative approaches.

6.1 Data Generating Process

The industry consists of three firms ($J = 3$) and $M \in \{500,\ 1000,\ 5000,\ 10000\}$ geographic markets. Each firm sells one product.

6.1.1 Consumer demand.

Consumer demand is a nested logit with two nests. One nest includes only the outside alternative $j = 0$, and the other nest includes all the $J = 3$ products. Therefore, if $a_{jm} = 1$, the equation for the demand of product $j = 1, 2, 3$ is:

$$\ln\left(\frac{s_{jm}}{s_{0m}}\right) = \beta\, x_{jm} + \alpha\, p_{jm} + \sigma\, \ln\left(\frac{s_{jm}}{1 - s_{0m}}\right) + \xi_{jm}, \tag{58}$$

where $s_{jm}/(1 - s_{0m}) = s_{jm}/(s_{1m} + s_{2m} + s_{3m})$ is the within-nest market share of product j. Variable x_{jm} is a characteristic of product j that varies across markets.[9] We consider that $x_{jm} \sim i.i.d.\ |N(0, \sigma_x^2)|$. We use $\mathbf{x}_m$ to represent vector (x_{1m}, x_{2m}, x_{3m}).

6.1.2 Demand unobservables.

The demand unobservables $(\xi_{1m}, \xi_{2m}, \xi_{3m})$ are distributed according to a mixture of normals. More specifically, there are two 'types' of markets, indexed by $\kappa \in \{\ell, h\}$. The type of market m, κ_m, determines the mean of the normal distribution of the variables ξ_{jm}. This mean is equal to μ_ℓ if $\kappa_m = \ell$ and equal to μ_h if $\kappa_m = h$. Accordingly, we have that:

$$\xi_{jm} \sim 1\{\kappa_m = \ell\}\ N(\mu_\ell, \sigma_\xi^2) + 1\{\kappa_m = h\}\ N(\mu_h, \sigma_\xi^2) \tag{59}$$

[9] For instance, in the demand for air travel, consumers value an airline's degree of operation in the origin and destination airports of the market. Therefore, x_{jm} can be the number of other airports that the airline connects to from/to the airports in market m.

Market type κ_m is independent of $\mathbf{x}_m$ and *i.i.d.* across markets with $\Pr(\kappa_m = \ell) = f_\kappa(\ell)$. The realization of the normal random variables $N(\mu_\ell, \sigma_\xi^2)$ and $N(\mu_h, \sigma_\xi^2)$ is independent over markets and over firms. Note that the market type κ_m is the same for the three firms, and this introduces positive correlation between the variables ξ_{1m}, ξ_{2m}, and ξ_{3m}.

Note that variable ξ_{jm} is correlated with the unobservable component of the entry cost η_{jm}. See below in our description of the market entry game.

6.1.3 Price competition and marginal costs.

Given an hypothetical entry profile $\boldsymbol{a} = (a_1, a_2, a_3) \in \{0,1\}^3$, firms compete in prices a la Bertrand. In this nested logit model, equilibrium prices given entry profile $\boldsymbol{a}$ – that we represent as $\boldsymbol{p}_m(\boldsymbol{a}) = (p_{1m}(\boldsymbol{a}), p_{2m}(\boldsymbol{a}), p_{3m}(\boldsymbol{a}))$ – are the solution to the following system of best response equations:

$$p_{jm}(\boldsymbol{a}) = mc_{jm} - \frac{1-\sigma}{\alpha\left(1 - \sigma\,\frac{s_{jm}(\boldsymbol{p}_m(\boldsymbol{a}),\boldsymbol{a})}{s_{1m}(\boldsymbol{p}_m(\boldsymbol{a}),\boldsymbol{a})+s_{2m}(\boldsymbol{p}_m(\boldsymbol{a}),\boldsymbol{a})+s_{3m}(\boldsymbol{p}_m(\boldsymbol{a}),\boldsymbol{a})} - (1-\sigma)\ s_{jm}(\boldsymbol{p}_m(\boldsymbol{a}), \boldsymbol{a})\right)}. \tag{60}$$

To avoid the computation of this Bertrand equilibrium (for every market and Monte Carlo simulation), we consider that prices come from the following approximation to an equilibrium:

$$p_{jm}(\boldsymbol{a}) = mc_{jm} - \frac{1-\sigma}{\alpha\left(1 - \sigma\,\frac{s^*_{jm}(\mathbf{mc}_m,\boldsymbol{a})}{s^*_{1m}(\mathbf{mc}_m,\boldsymbol{a})+s^*_{2m}(\mathbf{mc}_m,\boldsymbol{a})+s^*_{3m}(\mathbf{mc}_m,\boldsymbol{a})} - (1-\sigma)\ s^*_{jm}(\mathbf{mc}_m, \boldsymbol{a})\right)}, \tag{61}$$

where $s^*_{jm}(\mathbf{mc}_m, \boldsymbol{a})$ represents the market share of product j in market m under entry profile $\boldsymbol{a}$ and under the hypothetical scenario that the price of each firm were equal to its marginal cost. That is, a firm's price is equal to its best response to the belief that other firms are charging their marginal costs.

A firm's marginal cost in market m is a deterministic function of x_{jm}: $mc_{jm} = \omega_0 + \omega_1\, x_{jm}$, where ω_0 and ω_1 are parameters.

6.1.4 Market entry game.

When making their entry decisions, firms know $\boldsymbol{x}_m$ and κ_m. Let $\boldsymbol{a}_{-j}$ represent the vector with the (hypothetical) entry decisions of all the firms except j. Let $\pi_j\left(\boldsymbol{a}_{-j}, \boldsymbol{x}_m, \kappa_m\right)$ be the expected variable profit for firm j given $\left(\boldsymbol{a}_{-j}, \boldsymbol{x}_m, \kappa_m\right)$. This variable profit is obtained integrating $\left[p_j\left(\boldsymbol{a}_{-j}, \boldsymbol{x}_m, \boldsymbol{\xi}_m\right) - mc_{jm}\right] s_j\left(\boldsymbol{a}_{-j}, \boldsymbol{x}_m, \boldsymbol{\xi}_m\right)$ over the distribution of $\boldsymbol{\xi}_m = (\xi_{1m}, \xi_{2m}, \xi_{3m})$ conditional on κ_m.

The entry cost of firm j in market m is $\gamma_j\, z_m + \eta_{jm}$, where γ_j is a parameter, z_m is a variable that is observable to the researcher, and η_{jm} is unobservable and *i.i.d.* over (j, m) with standard normal distribution. Variable z_m is *i.i.d.* over markets with a Uniform distribution over the interval $[z_{\min}, z_{\max}]$. The M sample realizations of z_m are generated by dividing interval $[z_{\min}, z_{\max}]$ into a grid of M equally spaced points. Let $\Delta = (z_{\max} - z_{\min})/(M-1)$. Then, for any market $m = 1, 2, ..., M$, we have that $z_m = z_{\min} + (m-1)\Delta$. We keep this grid fixed across all the Monte Carlo simulations. Firm j knows z_m and at its own η_{jm} when making its entry decision.

Given $(\boldsymbol{x}_m, z_m, \kappa_m)$, we solve for a Bayesian Nash equilibrium of the entry game by solving the following system of 3 equations and three unknown probabilities (P_{1m}, P_{2m}, P_{3m}). For $j = 1, 2, 3$:

$$P_{jm} = \Phi\left(\pi_j^P\left(\boldsymbol{x}_m, \kappa_m\right) - \gamma_j\, z_m\right) \tag{62}$$

where $\pi_j^P\left(\boldsymbol{x}_m, \kappa_m\right) = \sum_{\boldsymbol{a}_{-j} \in \{0,1\}^2} \left[\prod_{i \neq j} \left(P_{im}\right)^{a_i} \left(1 - P_{im}\right)^{1-a_i}\right] \pi_j\left(\boldsymbol{a}_{-j}, \boldsymbol{x}_m, \kappa_m\right)$, and $\Phi(.)$ is the CDF of the standard normal.

The unobserved entry variable η_{jm} is correlated with the demand unobservable ξ_{jm}. Note that this introduces a source of endogenous selection in addition to the one that comes from κ_m. Parameter $\sigma_{\eta,\xi}$ measures the correlation between η_{jm} and ξ_{jm}. Therefore, we have that:

$$\mathbb{E}\left(\xi_{jm} \mid \boldsymbol{x}_m, z_m, \kappa_m, a_{jm} = 1\right) = \mu_{\kappa_m} + \sigma_{\eta,\xi}\, \frac{\phi\left(\left(\pi_j^P\left(\boldsymbol{x}_m, \kappa_m\right) - \gamma_j\, z_m\right)\right)}{P_j\left(\boldsymbol{x}_m, z_m, \kappa_m\right)} \tag{63}$$

where $\phi(\cdot)$ is the standard normal density function. If the researcher knew the true distribution of the unobservables, then she would use equation (63) as the control function for selection in the second step of the method. We consider that the researcher does not have this information. However, in our experiments, we will evaluate what is the improvement in the precision of the estimates if the researcher had this information.

6.1.5 Solving for an equilibrium of entry game.

For a given value of $(\boldsymbol{x}_m, z_m, \kappa_m)$, we need solve for a Bayesian Nash equilibrium of the entry game. There are two main computational tasks involved in this solution.

First, we need to compute the expected variable profit $\pi_j\left(\boldsymbol{a}_{-j}, \boldsymbol{x}_m, \kappa_m\right)$ for every hypothetical value of $\boldsymbol{a}_{-j}$ by integrating $[p_j(\boldsymbol{a}_{-j}, \boldsymbol{x}_m, \boldsymbol{\xi}_m) - mc_{jm}]\, s_j(\boldsymbol{a}_{-j}, \boldsymbol{x}_m, \boldsymbol{\xi}_m)$ over the distribution of $(\xi_{1m}, \xi_{2m}, \xi_{3m})$ conditional on κ_m. We approximate this expectation by using Monte Carlo simulation. That is, for each market m, we simulate 500 random draws of ξ_{jm} from the normal distribution $N\left(\mu_{\kappa_m}, \sigma_{\xi}^2\right)$ and then obtain the average of $[p_j(\boldsymbol{a}_{-j}, \boldsymbol{x}_m, \boldsymbol{\xi}_m) - mc_{jm}]\, s_j(\boldsymbol{a}_{-j}, \boldsymbol{x}_m, \boldsymbol{\xi}_m)$ over these simulations.

Second, we need to solve numerically for a fixed point of the system of equations (62). We use fixed point iterations.

6.1.6 Generating simulations

Our Monte Carlo experiments are based on 100 Monte Carlo simulations or samples. We describe here the different steps to generate one single sample in our Monte Carlo simulations.

1. Generate the M values of z_m using the grid points described above.

2. Generate M independent random draws of $(x_m, \kappa_m : m = 1, 2, ...M)$ from the distribution of these variables.

3. For each market m, given $(\boldsymbol{x}_m,\ z_m,\ \kappa_m)$, compute the BNE CCPs $(P_j(\boldsymbol{x}_m, z_m, \kappa_m) : j = 1, 2, 3)$. Then, generate a_{jm} as a random from the Bernoulli with probability $P_j(\boldsymbol{x}_m, z_m, \kappa_m)$.

4. For each market m, given κ_m, generate a random draw of the variables $(\xi_{1m}, \xi_{2m}, \xi_{3m})$.

5. For each market m, given $(\boldsymbol{x}_m,\ \boldsymbol{a}_m,\ \boldsymbol{\xi}_m)$, compute equilibrium prices and market shares.

To better explore the performance of our estimator, we simulate 16 "types" of dataset by varying the number of markets $M \in \{500, 1.000, 5.000, 10.000\}$ and the number of times we observe entry decisions being made for each market m, $n_{obs} \in \{1, 10, 100, 1.000\}$. While M affects the properties of both the first and the second step estimators, n_{obs} only affects the precision of the first step estimator. That is, demand is always estimated on M observations irrespective of n_{obs}.

Table 1 summarizes the values of all the parameters in the DGP.

Table 1. True Values of Parameters in DGP

Parameter	True Value	Parameter	True Value
$\beta =$	2.0	$\sigma_\xi^2 =$	2.0
$\alpha =$	-2.0	$\sigma_{\eta,\xi} =$	1.9999
$\sigma =$	0.6		
		$\sigma_x^2 =$	0.3
$\mu_\ell =$	-4.0	$z_{\min} =$	0.1
$\mu_h =$	4.0	$z_{\max} =$	0.2
$f_\kappa(\ell) =$	0.6	$\gamma_1 = \gamma_2 = \gamma_3 =$	1.0

6.2 Estimators

For each of the 16 possible (M, n_{obs}) configurations, we generate 100 repetitions of the data, and implement several estimators. The estimators we consider are the following.

(i) OLS. It is he most naive approach to the estimation of a nested logit demand system. It ignores not only endogenous sample selection but also the endogeneity of price and of the within-nest share.

(ii) 2SLS. It accounts for the endogenity of prices and within-nest share, but it ignores endogenous sample selection. We construct instruments on the basis of x_m. This is the classic BLP (1995) GMM estimator.

(iii) Our estimator. It estimates CCPs $P_j(\boldsymbol{x}_m, z_m, \kappa_m)$ and $f_\kappa(\ell)$ nonparametrically on the basis of n_{obs} entry observations per market m, and then feeds them in a second step semi-parametric estimator which controls for both endogeneity and selection. In the second step estimator, we use the same instruments as in the 2SLS. This is our proposed estimator.

(iv) Our estimator with true CCPs. It implements the second step of our semi-parametric estimator using as inputs the true values of the CCPs $P_j(\boldsymbol{x}_m, z_m, \kappa_m)$ and $f_\kappa(\ell)$, rather than their nonparametric estimates. The purpose of looking at this estimator is to obtain a measure of the loss of precision demand estimates due to the nonparametric estimation of the incidental parameters in the first step.

(v) Oracle estimator. Controls for endogeneity as in BLP (1995), but it also exactly controls for selection by subtracting the true value of the selection term $\boldsymbol{f}_\kappa' \boldsymbol{\psi}_j(\boldsymbol{P}_{jm})$ from the dependent variable of the 2SLS estimator. We call it Oracle because it relies on perfect

knowledge of the selection function. Of course, this estimator is unfeasible in an actual application. We use this estimator as a benchmark of comparison as it provides the most precise BLP-type estimator that controls for sample selection.

The OLS, 2SLS, and Oracle always control for product-specific intercepts even though the nested logit demand model used to generate the data does not include them (i.e., they are equal to zero).[10] Controlling for these intercepts attenuates the estimation bias induced by endogenous sample selection on the structural parameters. Importantly, we do not consider the (huge) estimation bias on these intercepts (i.e., difference from zero) in our comparisons with *Our estimator* and *Our estimator with true CCPs.*

6.3 Results

Table 2 shows some relevant statistics of the generated data for 5000 markets averaged over 100 replications.

Table 2. Summary Statistics from DGP

	Percentage of Zeros	Avr market Share	Average p-c/p
Firm 1	81.3%	0.16	79.9 %
Firm 2	81.4%	0.16	79.9%
Firm 3	81.4%	0.16	80%

Table 3 reports the average point estimates and their standard deviations computed over 100 Monte Carlo repetitions for the d.g.p. with $M = 1000$ and $n_{obs} = 1$.

[10] We set these product-specific intercepts to zero because *Our estimator* and *Our estimator with true CCPs* difference them out.

Table 3. Monte Experiment with $M = 1000$, $n_{obs} = 1$
Mean and Standard Deviation of Parameters Estimates

Parameter		True Value	OLS	2SLS	Our	Our(True P)	Oracle
β	Mean	2	**1.2161**	**1.6352**	**1.7754**	**1.9835**	**2.0011**
	Std. Dev.		(0.1483)	(0.1974)	(0.2573)	(0.0203)	(0.0071)
α	Mean	-2	**-1.9046**	**-1.9254**	**-1.9657**	**-1.9967**	**-2.0001**
	Std. Dev.		(0.0134)	(0.0136)	(0.0115)	(0.0017)	(0.0004)
σ	Mean	0.6	**0.5892**	**0.6677**	**0.6277**	**0.6042**	**0.6001**
	Std. Dev.		(0.0113)	(0.0287)	(0.0146)	(0.0022)	(0.0008)

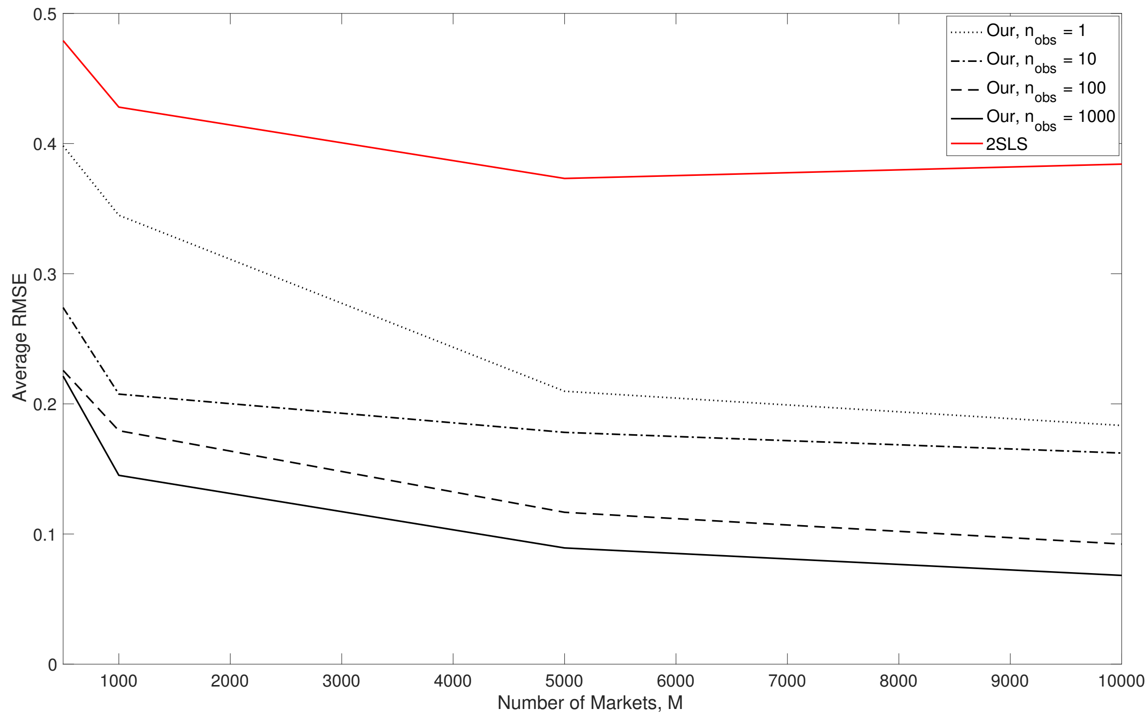
Table 4 reports the relative root mean square error (RMSE) of three pairs of estimators across the four d.g.p.'s with $n_{obs} = 1$. The RMSE of each estimator in each d.g.p. is computed over 100 repetitions.

Table 4. Monte Experiments with for different values of M
Ratios Between RMSEs of Different Estimators

	$M = 500$			$M = 1,000$		
	Our/2SLS	Our/Our-True	Our-True/Oracle	Our/2SLS	Our/Our-True	Our-True/Oracle
β	0.8435	10.9144	4.4753	0.8236	13.0788	3.6581
α	0.5138	7.9534	9.4354	0.4772	9.8504	9.327
σ	0.5743	7.0440	4.3611	0.4262	6.6343	5.5341
	$M = 5,000$			$M = 10,000$		
	Our/2SLS	Our/Our-True	Our-True/Oracle	Our/2SLS	Our/Our-True	Our-True/Oracle
β	0.5835	17.7773	3.9737	0.4947	18.0781	4.8414
α	0.1838	7.4314	12.8114	0.1232	6.4403	13.1430
σ	0.1509	4.3426	7.1205	0.0958	3.2940	9.2273

Figure 1 plots the average root mean square error (RMSE) of "Our" estimator and the 2SLS across the 16 possible configurations of (M, n_{obs}). Note that we report only one RMSE that aggregates the three demand parameters: that is, $RMSE = (MSE(\widehat{\alpha}) + MSE(\widehat{\beta}) + MSE(\widehat{\sigma}))^{1/2}$.

Figure 1: Average RMSE of "Our" and 2SLS estimators



7 Empirical Application

TBW

8 Conclusions

TBW

References

[1] Aguirregabiria, V., and C-Y. Ho (2012): "A dynamic oligopoly game of the US airline industry: Estimation and policy experiments," *Journal of Econometrics*, 168(1), 156-173.

[2] Aguirregabiria, V., and P. Mira (2019): “Identification of games of incomplete information with multiple equilibria and unobserved heterogeneity,” *Quantitative Economics*, 10(4), 1659–1701.

[3] Ahn, H., and J. L. Powell (1993): “Semiparametric estimation of censored selection models with a nonparametric selection mechanism,” *Journal of Econometrics*, 58(1-2), 3–29.

[4] Allman, E. S., C. Matias, and J. A. Rhodes (2009): “Identifiability of parameters in latent structure models with many observed variables,” The Annals of Statistics, 37(6A), 3099–3132.

[5] Amemiya, T. (1973): "Regression Analysis when the Dependent Variable is Truncated Normal," *Econometrica*, 41, 997-1017.

[6] Amemiya, T. (1974): “Multivariate Regression and Simultaneous Equation Models When the Dependent Variables Are Truncated Normal,” *Econometrica*, 42, 999–1012

[7] Aradillas-Lopez, A. (2010): “Semiparametric estimation of a simultaneous game with incomplete information,” Journal of Econometrics, 157(2), 409–431.

[8] Aradillas-Lopez, A. (2012): “Pairwise-difference estimation of incomplete information games,” Journal of Econometrics, 168(1), 120–140.

[9] Aradillas-Lopez, A., B. E. Honoré, and J. L. Powell (2007): “Pairwise difference estimation with nonparametric control variables,” International Economic Review, 48(4), 1119–1158.

[10] Bajari, P., H. Hong, J. Krainer, and D. Nekipelov (2010): “Estimating static models of strategic interactions,” Journal of Business & Economic Statistics, 28(4), 469–482.

[11] Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile prices in market equilibrium," Econometrica: Journal of the Econometric Society, pp. 841–890.

[12] Berry, S. T. (1994): "Estimating discrete-choice models of product differentiation," The RAND Journal of Economics, pp. 242–262.

[13] Berry, S., M. Carnall, and P. Spiller (2006): "Airline Hubs: Costs, Markups and the Implications of Customer Heterogeneity," in *Advances in Airline Economics*, Vol. 1: Competition Policy and Antitrust, D. Lee, ed. Elsevier Press. pp 183-214.

[14] Berry, S., and P. Jia (2010): "Tracing the woes: An empirical analysis of the airline industry," *American Economic Journal: Microeconomics*, 2(3), 1-43.

[15] Bonhomme, S., K. Jochmans, and J.-M. Robin (2016): "Non-parametric estimation of finite mixtures from repeated measurements," Journal of the Royal Statistical Society: Series B (Statistical Methodology), 78(1), 211–229.

[16] Chamberlain, G. (1986): "Asymptotic efficiency in semi-parametric models with censoring," Journal of Econometrics, 32(2), 189–218.

[17] Ciliberto, F., C. Murry, and E. T. Tamer (2018): "Market structure and competition in airline markets," Available at SSRN 2777820.

[18] Cosslett, S. (1991): "Semiparametric estimation of a regression model with sample selectivity," Nonparametric and semiparametric methods in econometrics and statistics, 175–197.

[19] Deaton, A., and J. Muellbauer (1980): "An almost ideal demand system," The American Economic Review, 70(3), 312–326.

[20] Dubé, J.-P. H., A. Hortaçsu, and J. Joo (2020): "Random-Coefficients Logit Demand Estimation with Zero-Valued Market Shares," National Bureau of Economic Research.

[21] Dubin, J. A., and D. L. McFadden (1984): "An econometric analysis of residential electric appliance holdings and consumption," Econometrica: Journal of the Econometric Society, 345–362.

[22] Ellickson, P. B., and S. Misra (2012): "Enriching interactions: Incorporating outcome data into static discrete games," Quantitative Marketing and Economics, 10(1), 1–26.

[23] Eizenberg, A. (2014): "Upstream innovation and product variety in the us home pc market," *Review of Economic Studies*, 81(3), 1003-1045.

[24] Gandhi, A., Z. Lu, and X. Shi (2017): "Estimating demand for differentiated products with zeroes in market share data," manuscript.

[25] Grieco, P. L. (2014): "Discrete games with flexible information structures: An application to local grocery markets," The RAND Journal of Economics, 45(2), 303–340.

[26] Hall, P., A. Neeman, R. Pakyari, and R. Elmore (2005): "Nonparametric inference in multivariate mixtures," Biometrika, 92(3), 667–678.

[27] Hall, P., and X.-H. Zhou (2003): "Nonparametric estimation of component distributions in a multivariate mixture," The Annals of Statistics, 31(1), 201–224.

[28] Heckman, J. J. (1976): "The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models," *Annals of Economic and Social Measurement*, 5(4), 475-492.

[29] Kasahara, H., and K. Shimotsu (2014): "Non-parametric identification and estimation of the number of components in multivariate mixtures," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(1), 97–111.

[30] Kleibergen, F., and R. Paap (2006): "Generalized reduced rank tests using the singular value decomposition," *Journal of Econometrics*, 133(1), 97–126.

[31] Li, S., J. Mazur, Y. Park, J. Roberts, A. Sweeting, and J. Zhang (2018): "Repositioning and market power after airline mergers," NBER working paper, No. 24214.

[32] Newey, W., and D. McFadden (1994): "Large sample estimation and hypothesis," *Handbook of Econometrics*, IV, Edited by RF Engle and DL McFadden, pp. 2112–2245.

[33] Newey, W., J. Powell, and J. Walker (1990): "Semiparametric estimation of selection models: some empirical results," *American Economic Review*, 80(2), 324-328.

[34] Newey, W. K. (2009): "Two-step series estimation of sample selection models," *The Econometrics Journal*, 12, S217–S229.

[35] Powell, J. L. (2001): "Semiparametric estimation of censored selection models," *Nonlinear Statistical Modeling*, pp. 165–96.

[36] Robin, J.-M., and R. J. Smith (2000): "Tests of rank," *Econometric Theory*, 16(2), 151–175.

[37] Robinson, P. M. (1988): "Root-N-consistent semiparametric regression," *Econometrica: Journal of the Econometric Society*, pp. 931–954.

[38] Seim, K. (2006): "An empirical model of firm entry with endogenous product-type choices," *The RAND Journal of Economics*, 37(3), 619–640.

[39] Smith, H. (2004): "Supermarket choice and supermarket competition in market equilibrium," *The Review of Economic Studies*, 71(1), 235-263.

[40] Sweeting, A. (2013): "Dynamic product positioning in differentiated product markets: The effect of fees for musical performance rights on the commercial radio industry," *Econometrica*, 81(5), 1763-1803.

[41] Tobin, J. (1958): "Estimation of relationships for limited dependent variables," *Econometrica*, 26(1), 24-36.

[42] Xiao, R. (2018): "Identification and estimation of incomplete information games with multiple equilibria," *Journal of Econometrics*, 203(2), 328–343.

[43] Yen, S. T. (2005): "A multivariate sample-selection model: Estimating cigarette and alcohol demands with zero observations," American Journal of Agricultural Economics, 87(2), 453-466.

[44] Yen, S. T., and B. H. Lin (2006): "A sample selection approach to censored demand systems," American Journal of Agricultural Economics, 88(3), 742-749.