Identification and Estimation of Demand Models with Endogenous Product Entry*

Victor Aguirregabiria †
University of Toronto
CEPR

Alessandro Iaria ‡
University of Bristol
CEPR

Senay Sokullu §
University of Bristol

February 27, 2023

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 selection in demand estimation. The non-additivity of demand unobservables in firms' expected profit and the potential multiplicity of equilibria in the entry game, make this selection problem challenging. Standard two-step methods to account for selection provide inconsistent estimates of structural parameters. Recent proposed methods require parametetric assumptions on the distribution of unobservables as well as repeatedlly computing for equilibria in the game. In this paper, we show the identification of demand parameters in a structural model of demand, price competition, and market entry that allows for a nonparametric distribution of demand unobservables. We propose a two-step method that extends standard methods to control for selection bias. 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 for differentiated product; Endogenous product availability; Selection bias; Market entry; Multiple equilibria; Identification; Estimation; Demand for airlines.

JEL codes: C14, C34, C35, C57, D22, L13, L93.

*We are grateful for helpful comments from Yuanyuan Wan and from seminar participants at University of Mannheim.

†Department of Economics, University of Toronto. 150 St. George Street, Toronto, ON, M5S 3G7, Canada, victor.aguirregabiria@utoronto.ca.

‡Department of Economics, University of Bristol. The Priory Road Complex, Priory Road, BS8 1TU, Bristol, UK. alessandro.iaria@bristol.ac.uk.

§Department of Economics, University of Bristol. 8 Woodland Road, Bristol, BS81TN, UK. senay.sokullu@bristol.ac.uk.

1 Introduction

Since the seminal work by Tobin (1958), Amemiya (1973) and Heckman (1976), addressing endogenous selection has been a fundamental topic in microeconometrics. Dealing with censored observations (i.e., zeroes) in the estimation of consumer demand has motivated the development of important methods to account for sample selection. Most of the early literature studied the demand for a single product, but there are also early applications to demand systems using Amemiya's multivariate Tobit model (Amemiya, 1974; Yen, 2005; Yen and Lin, 2006). More recently, the selection problem in the estimation of demand systems has received substantial attention in the context of structural models of demand for differentiated products. An important dimension to distinguish recently proposed methods is the sources of zeroes. Gandhi, Lu, and Shi (2022) assume that zeroes in market shares result from purchases of finite number of consumers. Dubé, Hortaçsu, and Joo (2021) study zeroes that result from consumers not including some products in their *consideration sets*. Ciliberto, Murry, and Tamer (2021) and Li, Mazur, Park, Roberts, Sweeting, and Zhang (2022) study sample selection in demand estimation when zeroes are the result of firms' market entry decisions in the context of an oligopoly model of competition.

This paper deals with the estimation of demand for differentiated products using market level data when there is censoring/selection because of firms not offering some products in some markets or periods. Demand estimation usually relies on data from multiple geographic markets and/or time periods. Often, some products are not offered in some markets or periods. When making their market entry decisions, firms 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. This problem appears in many demand applications and industries such as demand for airlines (Berry, Carnall, and Spiller, 2006; Berry and Jia, 2010; Aguirregabiria and Ho, 2012), for supermarket chains (Smith, 2004), for radio stations (Sweeting, 2013), or for personal computers (Eizenberg, 2014). 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 assuming 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 practically convenient, it is based on restrictions on firms' information that may not be plausible in some empirical applications.

[1] For instance, this is the approach in Aguirregabiria and Ho (2012), or 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).

In fact, these restrictions are testable and can be rejected by the data.

Two important features of this demand-entry model make this selection problem non-standard. First, demand unobservables enter non-additively in the binary choice models for firms' entry decisions. Without further restrictions, the selection term is not only a function of the entry probabilities (i.e., propensity scores) and it depends also on the observable variables affecting market entry. This implies that standard identification results and two-step estimation methods in the literature do not apply (Newey, Powell, and Walker, 1990; Ahn and Powell, 1993; Powell, 2001; Das, Newey, and Vella, 2003; Aradillas-Lopez, Honoré, and Powell, 2007; Newey, 2009). Second, the model has multiple equilibria, both in the entry game and in the pricing game. Given the same observable characteristics, markets can select different equilibria. This introduces an additional source of unobserved heterogeneity that may affect sample selection.

In this paper, we study the identification of demand parameters in a structural model of demand, price competition, and market entry where the distribution of demand unobservables is nonparametrically specified. As in other selection models, a key feature is the specification of firms' information about demand unobservables at the moment of their product entry decisions. We assume that firms observe two signals about these demand variables: a discrete signal with finite support which is common knowledge to all the firms in the market; and a continuous (real-valued) signal which is private information of each firm and independent across firms. This structure is different to the ones in models in Ciliberto, Murry, and Tamer (2021) and Li, Mazur, Park, Roberts, Sweeting, and Zhang (2022). These previous papers assume that firms' have complete information both in the product entry game and in the price competition game. There is not uncertainty in the models of those papers. In our model, we maintain the standard assumption of complete information in the Bertrand pricing game, but we consider that firms have uncertainty about demand when they make their product entry decisions, and that firms have some private information about their own future demand. This feature of our model facilitates dealing with the selection problem. Furthermore, it allows us to consider a nonparametric distribution for all the unobservables in the model. In contrast, the previous methods mentioned above consider a fully parametric distribution (i.e., normal) for the unobservables.

The paper has three main contributions. First, we present new identification results in this model. We show that the probability of product entry conditional on firms' information about the unobserved components of demand is nonparametrically identified. 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 finite 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. Third, 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.

The motivation and purpose of our paper is closely related to the work of Ciliberto, Murry, and Tamer (2021) and Li, Mazur, Park, Roberts, Sweeting, and Zhang (2022). These papers develop methods for the estimation of structural models that bring together Berry, Levinsohn, and Pakes (1995)'s framework and games of market/product entry as in Bresnahan and Reiss (1990, 1991), and Berry (1992). 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 of 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 the supply side structure is fully consistent with an equilibrium model, but this supply side — i.e., marginal and entry costs and the distribution of all unobservables — is nonparametrically specified. Furthermore, our estimation method is computationally simple as it does not require the computation of equilibria.[2]

Our model of product entry relates to the literature of structural models of market entry with incomplete information such as Seim (2006), Aguirregabiria and Mira (2007), Pakes, Ostrovsky, and Berry (2007), Pakes, Ostrovsky, and Berry (2007), or Sweeting (2009). More specifically, the specification of the unobservables in the entry-decision part of our model — that combines common-knowledge unobsevables with finite support and private information unobservables with continuous support — is closely related to the models in Xiao (2018) and Aguirregabiria and Mira (2019).

Our estimation method builds on and extends the literature on semiparametric estimation of sample selection models (Newey, Powell, and Walker, 1990; Ahn and Powell, 1993; Powell, 2001; Das, Newey, and Vella (2003); Aradillas-Lopez, Honoré, and Powell, 2007; Newey, 2009).

[2] An interesting feature of the methods in Ciliberto, Murry, and Tamer (2021) and Li, Mazur, Park, Roberts, Sweeting, and Zhang (2022) 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, Mazur, Park, Roberts, Sweeting, and Zhang (2022). Our semiparametric framework is mainly designed for the robust and computationally convenient estimation of demand. However, given estimates of demand parameters and unobservables (residuals) from our method, one can 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.

A distinguishing feature of our model and method is that the selection term has a nonparametric finite mixture structure.

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. 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 on the US airline industry. We summarize and conclude in section 8.

2 Model

2.1 Demand

The demand system follows the BLP framework (Berry, Levinsohn, and Pakes, 1995). 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 T markets indexed by $t \in \mathcal{M} = \{1, 2, ..., M\}$, where a market can be a geographic location, a time period, or a combination of both. Consumers living in market t can buy only the products available in that market. Firms make entry decisions – independently across markets – and compete at the local market level after entry.

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

$$U_{hjt} \equiv \delta(p_{jt}, \boldsymbol{x}_{jt}) + v(p_{jt}, \boldsymbol{x}_{jt}, \upsilon_{ht}) + \varepsilon_{hjt}, \tag{1}$$

where p_{jt} and $\boldsymbol{x}_{jt}$ are the price and other characteristics, respectively, of product j in market t; $\delta_{jt} \equiv \delta(p_{jt}, \boldsymbol{x}_{jt})$ is the average (indirect) utility of product j in market t; and $v(p_{jt}, \boldsymbol{x}_{jt}, \upsilon_{ht}) + \varepsilon_{hjt}$ represents a household-specific deviation from the average utility. The term $v(p_{jt}, \boldsymbol{x}_{jt}, \upsilon_{ht})$ depends on the vector of random coefficients υ_{ht} that is unobserved to the researcher with distribution $F_{\upsilon}(\cdot|\boldsymbol{\sigma})$, where $\boldsymbol{\sigma}$ is a vector of parameters. The term ε_{hjt} is unobserved to the researcher and is i.i.d. over (h, j, t) with type 1 extreme value distribution. Following the standard specification, the average utility of product j is:

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

where α and $\boldsymbol{\beta}$ are parameters. Variable ξ_{jt} captures the characteristics of product j in market t which are unobserved to the researcher. The outside option is represented by $j = 0$ and its indirect utility is normalized to $U_{h0t} = \varepsilon_{h0t}$.

Let $a_{jt} \in \{0, 1\}$ be the indicator for product j being available in market t, and let $\boldsymbol{a}_t \equiv (a_{jt} :$

$j \in \mathcal{J}$) denote the vector with the indicators for the availability of every product in market t. The outside option $j = 0$ is always available in every market. Every household chooses a product to maximize utility. Let s_{jt} be the market share of product j in market t, i.e., the proportion of households choosing product j:

$$s_{jt} = d_j(\boldsymbol{\delta}_t, \boldsymbol{a}_t, \boldsymbol{p}_t, \boldsymbol{x}_t) \equiv \int \frac{a_{jt} \ \exp\left(\delta_{jt} + v(p_{jt}, \boldsymbol{x}_{jt}, \upsilon)\right)}{1 + \sum_{i=1}^{J} a_{it} \ \exp\left(\delta_{it} + v(p_{it}, \boldsymbol{x}_{it}, \upsilon)\right)} \, dF_\upsilon(\upsilon). \tag{3}$$

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

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:

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

where $\boldsymbol{s}_t^{(\boldsymbol{a})}$ is the subvector from $\boldsymbol{s}_t$ containing the market shares for only those products available in the market, and a similar definition applies to subvectors $\boldsymbol{d}^{(\boldsymbol{a})}$, $\boldsymbol{\delta}_t^{(\boldsymbol{a})}$, $\boldsymbol{p}_t^{(\boldsymbol{a})}$, and $\boldsymbol{x}_t^{(\boldsymbol{a})}$. Lemma 1 establishes that the invertibility property in Berry (1994) applies to the 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 $\boldsymbol{s}_t^{(\boldsymbol{a})} = \boldsymbol{d}^{(\boldsymbol{a})}(\boldsymbol{\delta}_t^{(\boldsymbol{a})}, \boldsymbol{p}_t^{(\boldsymbol{a})}, \boldsymbol{x}_t^{(\boldsymbol{a})})$ is invertible with respect to $\boldsymbol{\delta}_t^{(\boldsymbol{a})}$ such that for every product in this subsystem (i.e., for every product with $a_{jt} = 1$) the inverse function $\delta_{jt}^{(\boldsymbol{a})} = d_j^{(\boldsymbol{a})-1}(\boldsymbol{s}_t^{(\boldsymbol{a})}, \boldsymbol{p}_t^{(\boldsymbol{a})}, \boldsymbol{x}_t^{(\boldsymbol{a})})$ 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 a product available in market t, we have:

$$\delta_{jt} = \alpha \ p_{jt} + \boldsymbol{x}_{jt}' \ \boldsymbol{\beta} + \xi_{jt} \quad \text{if and only if} \quad a_{jt} = 1. \tag{5}$$

Importantly, this regression equation for product j only depends on the availability of product j and not on the availability of the other products. Therefore, the selection problem in the estimation of the demand of product j can be described in terms of the conditional expectation,

$$\mathbb{E}\left(\xi_{jt} \mid a_{jt} = 1\right). \tag{6}$$

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

To appreciate the value of this property, consider instead the case of the *Almost Ideal Demand System (AIDS)* (Deaton and Muellbauer, 1980). In that model, each value of the vector $\boldsymbol{a}_t$ implies a different expression for the regression equation that relates the demand of product j to the log-prices of all the available products. Therefore, in the AIDS model, the selection problem in the estimation of the demand of product j is not only related to the availability of that product but to the availability of all products in the system. In other words, the selection term cannot be represented in terms of $\mathbb{E}(\xi_{jt} \mid a_{jt} = 1)$ but in terms of $\mathbb{E}(\xi_{jt} \mid \boldsymbol{a}_t = \boldsymbol{a})$. 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 next Example illustrates Lemma 1 in the case of a 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 to which product j belongs. The indirect utility function is $U_{htj} \equiv \delta_{jt} + (1-\sigma)\, v_{ht,r(j)} + \varepsilon_{htj}$, where the variables v_{ht,r_j} and ε_{htj} are i.i.d. Type I extreme value and mutually independent, and $\sigma \in [0,1]$ is a parameter. This model implies $s_{jt} = d_j^{(\boldsymbol{a}_t)}(\boldsymbol{\delta}_t) = d_{r_j}^{(\boldsymbol{a}_t)}\ d_{j|r_j}^{(\boldsymbol{a}_t)}$ with

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

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

$$\delta_{jt} = \ln\left(\frac{s_{jt}}{s_{0t}}\right) - \sigma\ \ln\left(\frac{\sum_{i \in r_j} s_{it}}{s_{0t}}\right), \tag{8}$$

and it implies the regression equation:

$$\ln\left(\frac{s_{jt}}{s_{0t}}\right) = \sigma\ \ln\left(\frac{\sum_{i \in r_j} s_{it}}{s_{0t}}\right) + \alpha\ p_{jt} + \boldsymbol{x}'_{jt}\ \boldsymbol{\beta} + \xi_{jt}. \tag{9}$$

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

2.2 Price Competition

Firms' market entry decisions, prices, and quantities are determined as an equilibrium of a two-stage game. In the first stage, 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 Nash-Bertrand equilibrium of a pricing game. This two-stage game is played separately across markets.

Firms that do not enter the market earn zero profit. Let Π_{jt} be the profit of firm j if active in market t. This equals revenues minus costs:

$$\Pi_{jt} = p_{jt}\ q_{jt} - c(q_{jt}; \boldsymbol{x}_{jt}, \omega_{jt}) - f(\boldsymbol{x}_{jt}, \eta_{jt}), \tag{10}$$

where q_{jt} is the quantity sold (i.e., market share s_{jt} times market size H_t), $c(q_{jt}; \boldsymbol{x}_{jt}, \omega_{jt})$ is the variable cost function, and $f(\boldsymbol{x}_{jt}, \eta_{jt})$ is the fixed entry cost. Variables ω_{jt} and η_{jt} are unobservable to the researcher. Given firms' entry decisions in $\boldsymbol{a}_t$, the best response function in the Bertrand pricing game implies the following system of pricing equations:

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

where mc_{jt} is the marginal cost $\partial c_{jt} / \partial q_{jt}$. A solution to this system of equations is a Nash-Bertrand equilibrium. The pricing game may have multiple equilibria. We do not impose restrictions on equilibrium selection and allow for each market to potentially select a different equilibrium. We use scalar variable τ_t^2 to index the equilibrium type selected in the Bertrand game, i.e., in step 2 of the two-stage game.

Let $\boldsymbol{x}_t \equiv (\boldsymbol{x}_{jt} : j \in \mathcal{J})$ be the vector with all the exogenous variables that are observable to the researcher, affecting demand or costs. Vectors $\boldsymbol{\xi}_t$ and $\boldsymbol{\omega}_t$ have similar definitions. We use $VP_j(\boldsymbol{a}_t, \boldsymbol{x}_t, \boldsymbol{\xi}_t, \boldsymbol{\omega}_t, \tau_t^2)$ to denote the indirect variable profit function for firm j that results from plugging into the expression $p_{jt}\ q_{jt} - c(q_{jt}; \boldsymbol{x}_{jt}, \omega_{jt})$ the value of (p_{jt}, q_{jt}) from the Nash-Bertrand equilibrium given $(\boldsymbol{a}_t, \boldsymbol{x}_t, \boldsymbol{\xi}_t, \boldsymbol{\omega}_t, \tau_t^2)$.

2.3 Market entry game and information structure

Firms' entry decisions are determined as an equilibrium of a game of market entry. The profit of being inactive is normalized to zero for all firms. Firms have uncertainty about their profits if active in the market. Their 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 1 and 2 summarize our conditions on the information structure and on the entry cost function, respectively.

ASSUMPTION 1. *Firm j's information at the moment of its entry decision in market t consists of $(\boldsymbol{x}_t, \kappa_t, \tau_t^1, \eta_{jt})$.*

A. κ_t is a signal for the demand-cost variables $(\boldsymbol{\xi}_t, \boldsymbol{\omega}_t, \tau_t^2)$. It is common knowledge for the firms, has discrete and finite support that we denote as $\mathcal{K}$, and its probability distribution conditional on $\boldsymbol{x}_t$ is $f_\kappa(\kappa_t|\boldsymbol{x}_t)$ and is nonparametrically specified.

B. Variable τ_t^1 represents the type of equilibrium selected in the entry game.

C. Variable η_{jt} is a component of the entry cost that is private information of firm j, independently distributed over firms, and independent of $(\kappa_t, \boldsymbol{x}_t)$ with CDF F_η, which is strictly increasing over the real line.

D. Vector $(\boldsymbol{\xi}_t, \boldsymbol{\omega}_t, \tau_t^2, \kappa_t, \tau_t^1, \eta_{jt})$ is unobserved to the researcher. Conditional on κ_t, variables $\boldsymbol{\xi}_t$, $\boldsymbol{\omega}_t$, and η_{jt} are independent of $\boldsymbol{x}_t$. ■

For our analysis, the payoff-relevant information in discrete variable κ_t plays the same role as the equilibrium-selection discrete variable τ_t^1. Therefore, for notational simplicity, we omit τ_t^1 and interpret κ_t as representing both equilibrium selection and payoff relevant variables. We represent this variable as an index with support $\{1, 2, ..., \mathcal{K}(\boldsymbol{x}_t)\}$. Similarly, with some abuse of notation, for the rest of the paper we represent vector of unobservables $(\boldsymbol{\xi}_t, \boldsymbol{\omega}_t, \tau_t^2)$ using the more compact notation $\boldsymbol{\xi}_t$.

Let $\pi_j(\boldsymbol{a}, \boldsymbol{x}_t, \kappa_t, \eta_{jt})$ be firm j's expected profit given its information about demand and costs and conditional on the hypothetical entry profile $\boldsymbol{a} \in \{0, 1\}^J$. Under Assumptions 1:

$$\pi_j(\boldsymbol{a}, \boldsymbol{x}_t, \kappa_t, \eta_{jt}) \;=\; \int VP_j(\boldsymbol{a}, \boldsymbol{x}_t, \boldsymbol{\xi}_t)\; dF_{j,\xi}\left(\boldsymbol{\xi}_t \mid \kappa_t, \eta_{jt}\right) - f(\boldsymbol{x}_{jt}, \eta_{jt}), \tag{12}$$

where $F_{j,\xi}\left(\boldsymbol{\xi}_t \mid \kappa_t, \eta_{jt}\right)$ is the CDF of $\boldsymbol{\xi}_t$ conditional on (κ_t, η_{jt}).

ASSUMPTION 2. *For any value $(\boldsymbol{a}, \boldsymbol{x}_t, \kappa_t)$, the function $\pi_j(\boldsymbol{a}, \boldsymbol{x}_t, \kappa_t, \eta_{jt})$ is strictly monotonic in η_{jt}. Without loss of generality, we consider that this function is decreasing in η_{jt}. Therefore, for any scalar value π^0 and any value $(\boldsymbol{a}, \boldsymbol{x}_t, \kappa_t)$, the equation $\pi^0 = \pi_j(\boldsymbol{a}, \boldsymbol{x}_t, \kappa_t, \eta_{jt})$ is invertible with respect to η_{jt}. That is, there is an inverse function π_j^{-1} such that $\eta_{jt} = \pi_j^{-1}(\boldsymbol{a}, \boldsymbol{x}_t, \kappa_t, \pi^0)$, and for any other scalar π^1 with $\pi^1 \leq \pi^0$, we have that $\pi_j^{-1}(\boldsymbol{a}_{-j}, \boldsymbol{x}_t, \kappa_t, \pi^1) \geq \eta_{jt}$.* ■

The presence of firms' private information implies that the entry game is of incomplete information. Given $(\boldsymbol{x}_t, \kappa_t)$, a *Bayesian Nash Equilibrium (BNE)* of this game can be represented as an J-tuple of entry probabilities, one for each firm, $(P_{jt} : j \in \mathcal{J})$. To describe this BNE, we first define a firm's expected profit function that accounts for its uncertainty about other firms' entry decisions.

$$\pi_j^P(\boldsymbol{x}_t, \kappa_t, \eta_{jt}) = \sum_{\boldsymbol{a}_{-j} \in \{0,1\}^{J-1}} \left(\prod_{i \neq j} [P_{it}]^{a_i} [1 - P_{it}]^{1-a_i} \right) \pi_j(\boldsymbol{a}_{-j}, \boldsymbol{x}_t, \kappa_t, \eta_{jt}). \tag{13}$$

Under Assumption 2, the expected profit function $\pi_j^P(\boldsymbol{x}_t, \kappa_t, \eta_{jt})$ is strictly monotonic and invertible in η_{jt}.[3] Let $\pi_j^{P(-1)}$ be this inverse function. Then, given the entry probabilities of its competitors, firm j's best response is to enter in the market if and only if:

$$\eta_{jt} \;\leq\; \pi_j^{P(-1)}(\boldsymbol{x}_t, \kappa_t, 0) \tag{14}$$

Taking this into account, we can define a BNE in this game as follows.

DEFINITION 1 (BNE). *Under Assumptions 1-2 and given* $(\boldsymbol{x}_t, \kappa_t)$*, a Bayesian Nash Equilibrium (BNE) can be represented as a* J*-tuple of probabilities,* $\{P_{jt} \equiv P_j(\boldsymbol{x}_t, \kappa_t) : j \in \mathcal{J}\}$ *that solve the following system of* J *best-response equations in the space of probabilities:*

$$P_{jt} \;=\; F_\eta\left(\pi_j^{P(-1)}(\boldsymbol{x}_t, \kappa_t, 0) \right), \tag{15}$$

where $\pi_j^{P(-1)}$ *is the inverse function with respect to* η_{jt} *of the expected profit in (13).* ■

This framework includes as particular cases different specifications of information structure about the unobservables which have been considered in structural models of market entry and product introduction, such as models with only complete information unobservables (e.g., Ciliberto and Tamer, 2009; Ciliberto, Murry, and Tamer, 2021), models with only private information unobservables (e.g., Seim, 2006; Sweeting, 2009; Bajari, Hong, Krainer, and Nekipelov, 2010), and models that include both types of unobservables (e.g., Grieco, 2014; Aguirregabiria and Mira, 2019).

[3] If π_j is strictly monotonic in η_{jt} for any possible entry profile $\boldsymbol{a}$, then a convex linear combination of π_j for different entry profiles is also strictly monotonic in η_{jt}.

3 Selection Problem

For simplicity and concreteness, we describe our sample selection problem using the nested logit demand model from Example 1. We use the starred variables s^*_{jt} and p^*_{jt} to represent *latent variables*. That is, s^*_{jt} and p^*_{jt} represent the latent market share and price, respectively, that we would observe if product j were offered in market t. Using these latent variables, we can write the following demand system:

$$\ln\left(\frac{s^*_{jt}}{s_{0t}}\right) \;=\; \sigma\;\ln\left(\frac{s^*_{jt}+S_{-jt}}{s_{0t}}\right)+\alpha\;p^*_{jt}+\boldsymbol{x}'_{jt}\;\boldsymbol{\beta}+\xi_{jt}, \tag{16}$$

where $S_{-jt} \equiv \sum_{i\neq j, i\in r_j} s_{it}$ is the aggregate market share of all products in group r_j other than product j. Latent variables (s^*_{jt}, p^*_{jt}) are equal to the observed variables (s_{jt}, p_{jt}) if and only if product j is offered in market t:

$$\{s^*_{jt} = s_{jt} \text{ and } p^*_{jt} = p_{jt}\} \text{ if and only if } \{a_{jt} = 1\} \tag{17}$$

The econometric model is completed with firm j's best response entry decision:[4]

$$a_{jt} \;=\; 1\left\{\eta_{jt} \;\leq\; \pi_j^{P(-1)}(\boldsymbol{x}_t, \kappa_t)\right\}. \tag{18}$$

Equations (16) to (18) imply the following regression equation for any product with $a_{jt} = 1$:

$$\ln\left(\frac{s_{jt}}{s_{0t}}\right) \;=\; \sigma\;\ln\left(\frac{s_{jt}+S_{-jt}}{s_{0t}}\right)+\alpha\;p_{jt}+\boldsymbol{x}'_{jt}\;\boldsymbol{\beta}+\lambda_j(\boldsymbol{x}_t)+\widetilde{\xi}_{jt}, \tag{19}$$

where $\lambda_j(\boldsymbol{x}_t)$ is the *selection bias function*, $\mathbb{E}\left[\xi_{jt} \mid \boldsymbol{x}_t, a_{jt} = 1\right]$. That is,

$$\lambda_j(\boldsymbol{x}_t) = \int \xi_{jt}\;1\left\{\eta_{jt} \;\leq\; \pi_j^{P(-1)}(\boldsymbol{x}_t, \kappa_t)\right\}\;\frac{f_{\xi,\eta,\kappa}\left(\xi_{jt}, \eta_{jt}, \kappa_t|\boldsymbol{x}_t\right)}{Pr\left(a_{jt}=1|\boldsymbol{x}_t\right)}\;d\left(\xi_{jt},\eta_{jt}, \kappa_t\right), \tag{20}$$

and:

$$Pr\left(a_{jt}=1|\boldsymbol{x}_t\right) \equiv \mathbb{E}\left[a_{jt} \mid \boldsymbol{x}_t\right] = \int 1\left\{\eta_{jt} \;\leq\; \pi_j^{P(-1)}(\boldsymbol{x}_t, \kappa_t)\right\}\;f_{\eta,\kappa}\left(\eta_{jt}, \kappa_t|\boldsymbol{x}_t\right)\;d\left(\eta_{jt}, \kappa_t\right), \tag{21}$$

where $f_{\eta,\kappa}$ and $f_{\xi,\eta,\kappa}$ are the joint density functions of (η_{jt}, κ_t) and $(\xi_{jt},\eta_{jt}, \kappa_t)$, conditional of $\boldsymbol{x}_t$, respectively.

Note that the selection bias function $\lambda_j(\boldsymbol{x}_t)$ is a nonparametric function of all its arguments

[4]With some abuse of notation, we use function $\pi_j^{P(-1)}(\boldsymbol{x}_t, \kappa_t)$ to represent $\pi_j^{P(-1)}(\boldsymbol{x}_t, \kappa_t, 0)$.

$\boldsymbol{x}_t$. Therefore, based on equation (19), and without further restrictions, the demand parameters σ, α, and $\boldsymbol{\beta}$ are not identified. That is, we cannot disentangle the direct effect of $\boldsymbol{x}_{jt}$ on demand (as represented by the vector of parameters $\boldsymbol{\beta}$) from the indirect effect that comes from the selection bias function $\lambda_j(\boldsymbol{x}_t)$.

Examples 2 and 3 below present restrictions that imply the identification of demand parameters. Example 2 is simple but it imposes strong restrictions on the unobservables. Our main identification results in section 4 are closely related to Example 3.

EXAMPLE 2 (No signals κ_t). Suppose that: (1*) $\kappa_t = 0$ such that, at the moment of the entry decision, the only information that a firm has about the demand/cost variables $\boldsymbol{\xi}_t$ is its private information variable η_{jt}; and (2*) a unique equilibrium is played across all entry games with the same observables $\boldsymbol{x}_t$. Under Assumptions 1-2 and conditions (1*) and (2*), the selection term $\lambda_j(\boldsymbol{x}_t)$ only depends on the CCP $P_j(\boldsymbol{x}_t)$: that is, $\lambda_j(\boldsymbol{x}_t) = \rho_j(P_j(\boldsymbol{x}_t))$ for some function $\rho_j(\cdot)$.

The proof is straightforward. Under conditions (1*)–(2*), the inverse profit function $\pi_j^{P(-1)}$ and the equilibrium entry probability P_j only depend on $\boldsymbol{x}_t$ but not on κ_t. The equilibrium entry probability $P_j(\boldsymbol{x}_t)$ is equal to the conditional expectation $\mathbb{E}(a_{jt}|\boldsymbol{x}_t)$, which is nonparametrically identified. Furthermore, this probability satisfies the equilibrium condition $P_j(\boldsymbol{x}_t) = F_\eta(\pi_j^{P(-1)}(\boldsymbol{x}_t))$, such that $\pi_j^{P(-1)}(\boldsymbol{x}_t) = F_\eta^{-1}(P_j(\boldsymbol{x}_t))$, and the entry condition can be represented as $a_{jt} = 1\{\eta_{jt} \leq F_\eta^{-1}(P_j(\boldsymbol{x}_t))\}$. Independence between (ξ_{jt}, η_{jt}) and $\boldsymbol{x}_t$ then implies that:

$$\lambda_j(\boldsymbol{x}_t) \;=\; \int \xi_{jt}\; 1\left\{\eta_{jt} \leq F_\eta^{-1}(P_j(\boldsymbol{x}_t))\right\}\; \frac{f_{\xi,\eta}(\xi_{jt}, \eta_{jt})}{P_j(\boldsymbol{x}_t)}\; d\xi_{jt}\; d\eta_{jt} \;=\; \rho_j(P_j(\boldsymbol{x}_t)) \tag{22}$$

Therefore, the demand equation can be represented as:

$$\ln\left(\frac{s_{jt}}{s_{0t}}\right) = \sigma\; \ln\left(\frac{s_{jt} + S_{-jt}}{s_{0t}}\right) + \alpha\; p_{jt} + \boldsymbol{x}_{jt}'\; \boldsymbol{\beta} + \rho_j(P_j(\boldsymbol{x}_t)) + \widetilde{\xi}_{jt}. \tag{23}$$

The result in equation (23) has important implications for identification and estimation. Regression equation (23) is a standard semiparametric partially linear model with two endogenous regressors, $\ln[(s_{jt} + S_{-jt})/s_{0t}]$ and p_{jt}, and with the nonparametric component $\rho_j(P_j(\boldsymbol{x}_t))$ 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 $\boldsymbol{P}(\boldsymbol{x}_t) = (P_1(\boldsymbol{x}_t), ..., P_J(\boldsymbol{x}_t))$ from data on $(\boldsymbol{a}_t, \boldsymbol{x}_t)$. Then, in a second step, by relying on observations from markets t and t' with $P_j(\boldsymbol{x}_t) = P_j(\boldsymbol{x}_{t'})$, but with $\ln[(s_{jt} + S_{-jt})/s_{0t}] \neq \ln[(s_{jt'} + S_{-jt'})/s_{0t'}]$, $p_{jt} \neq p_{jt'}$, and $\boldsymbol{x}_{jt} \neq$

$\boldsymbol{x}_{jt'}$, one can identify $(\sigma, \alpha, \boldsymbol{\beta})$ by correcting for both sample selection and endogeneity. More specifically, the difference $\ln(s_{jt}/s_{0t}) - \ln(s_{jt'}/s_{0t'})$ from (23) gets rid of the selection bias term $\rho_j(P_j(\boldsymbol{x}_t))$. Then, in the linear regression in deviations for $\ln(s_{jt}/s_{0t}) - \ln(s_{jt'}/s_{0t'})$ on $\ln[(s_{jt} + S_{-jt})/s_{0t}] - \ln[(s_{jt'} + S_{-jt'})/s_{0m'}]$, $p_{jt} - p_{jt'}$, and $\boldsymbol{x}_{jt} - \boldsymbol{x}_{jt'}$, we can use a standard IV where instruments are derived from $\boldsymbol{x}_t$. For instance, valid instruments in this regression are observed $\boldsymbol{x}$ characteristics of products other than j, i.e., the so-called BLP instruments. ■

Under the assumptions of Example 2, the market entry condition has only one unobservable variable, η_{jt}, which, after inverting the profit function, enters additively in the inequality that defines the selection/entry decision. The model becomes a standard semiparametric sample selection model. However, though practically convenient, the conditions in Example 2 are likely to be rejected in many empirical applications.[5] In particular, restriction (1*) — i.e., η_{jt} is the only relevant information that firm j has about demand/cost variables ξ_{jt} when making its entry decision — seems unrealistic. If this restriction does not hold, the identification/estimation approach described above is inconsistent because the equation in differences based on matching the CCP $P_j(\boldsymbol{x}_t)$ does not fully control for the selection bias function.

EXAMPLE 3 (κ_t has finite support). Consider the model under Assumptions 1-2, including Assumption 1(A) on the finite support of κ_t. Let $P_j(\boldsymbol{x}_t, \kappa_t)$ be the equilibrium probabilities in the entry game in market t. By definition, we have that $P_j(\boldsymbol{x}_t, \kappa_t) \equiv \mathbb{E}[a_{jt}|\boldsymbol{x}_t, \kappa_t]$, and:

$$P_j(\boldsymbol{x}_t, \kappa_t) \;=\; F_\eta\left(\pi_j^{P(-1)}(\boldsymbol{x}_t, \kappa_t)\right), \tag{24}$$

Similar to Example 2, the model implies a one-to-one relationship between $P_j(\boldsymbol{x}_t, \kappa_t)$ and the inverse expected profit function: i.e., $\pi_j^{P(-1)}(\boldsymbol{x}_t, \kappa_t) = F_\eta^{-1}(P_j(\boldsymbol{x}_t, \kappa_t))$. Define $\widetilde{\lambda}_j(\boldsymbol{x}_t, \boldsymbol{\kappa}_t) \equiv \mathbb{E}[\xi_{jt}|\boldsymbol{x}_t, \kappa_t, a_{jt} = 1]$. Applying the one-to-one relationship between CCP and inverse profit function, we have that this selection function is a function only of the CCP and κ_t:

$$\begin{aligned}\widetilde{\lambda}_j(\boldsymbol{x}_t, \kappa_t) \;&=\; \int \xi_{jt}\, 1\left\{\eta_{jt} \leq F_\eta^{-1}(P_j(\boldsymbol{x}_t, \kappa_t))\right\} \frac{f_{\xi,\eta|\kappa}(\xi_{jt}, \eta_{jt}|\kappa_t)}{P_j(\boldsymbol{x}_t, \kappa_t)}\, d(\xi_{jt}, \eta_{jt}) \\ &\equiv\; \psi_j\left(P_j(\boldsymbol{x}_t, \kappa_t), \kappa_t\right).\end{aligned} \tag{25}$$

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

By definition, we have the following relationship between $\widetilde{\lambda}_j$ and the *selection bias function* $\lambda_j(\boldsymbol{x}_t) \equiv \mathbb{E}[\xi_{jt}|\boldsymbol{x}_t, a_{jt} = 1]$ that appears in the demand equation:

$$\lambda_j(\boldsymbol{x}_t) \;=\; \sum_{\kappa_t \in \mathcal{K}(\boldsymbol{x}_t)} \widetilde{\lambda}_j(\boldsymbol{x}_t, \kappa_t)\; f_\kappa\left(\kappa_t|\boldsymbol{x}_t\right). \tag{26}$$

Combining the demand equation with equations (25) and (26), we obtain the regression equation:

$$\begin{aligned}\ln\left(\frac{s_{jt}}{s_{0t}}\right) \;&=\; \sigma\;\ln\left(\frac{s_{jt}+S_{-jt}}{s_{0t}}\right) + \alpha\; p_{jt} + \boldsymbol{x}'_{jt}\;\boldsymbol{\beta} \\ &+ \sum_{\kappa_t \in \mathcal{K}(\boldsymbol{x}_t)} \psi_j\left(P_j\left(\boldsymbol{x}_t, \kappa_t\right), \kappa_t\right)\; f_\kappa\left(\kappa_t|\boldsymbol{x}_t\right) + \widetilde{\xi}_{jt}.\end{aligned} \tag{27}$$

The regression equation in (27), and more specifically, the structure of the selection bias function, has two properties that play a key role in our identification results. First, $P_j\left(\boldsymbol{x}_t, \kappa_t\right)$ and the distribution $f_\kappa\left(\kappa_t\right)$ are nonparametrically identified using data on firms' entry decisions. And second, conditional on $\{P_j\left(\boldsymbol{x}_t, \kappa\right), f_\kappa\left(\kappa|\boldsymbol{x}_t\right) : \kappa \in \mathcal{K}\}$, the vector of product characteristics $\boldsymbol{x}_{jt}$ has full-rank. We establish these results in section 4. ■

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 T markets. For every market t, the researcher observes the vector of exogenous variables $\boldsymbol{x}_t$ and the vectors of firms' entry decisions $\boldsymbol{a}_t$. For those firms active in market t, the researcher observes prices $\boldsymbol{p}_t$ and market shares $\boldsymbol{s}_t$.

Let $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ be the vector of all the parameters in 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 $\boldsymbol{\theta}_\delta \equiv (\alpha, \boldsymbol{\beta}, \boldsymbol{\sigma})$; probability distribution of demand/cost signals, $\boldsymbol{f}_\kappa \equiv (f_\kappa(\kappa|\boldsymbol{x}) :$ for every $\kappa, \boldsymbol{x})$; equilibrium choice probabilities, $\boldsymbol{P}_\kappa \equiv (P_j(\boldsymbol{x}, \kappa) :$ for every $j, \boldsymbol{x}, \kappa)$; the probability distribution of private information F_η, and the distribution of unobserved demand conditional on signals, $f_{\xi|\eta,\kappa}$.

$$\boldsymbol{\theta} \;\equiv\; \left(\; \boldsymbol{\theta}_\delta,\; \boldsymbol{P}_\kappa,\; \boldsymbol{f}_\kappa,\; f_{\xi|\eta,\kappa},\; F_\eta \;\right). \tag{28}$$

In this paper, we are interested in the identification of demand parameters $\boldsymbol{\theta}_\delta$ when the distributions $\boldsymbol{f}_{\kappa}$ and $f_{\xi|\eta,\kappa}$ and the equilibrium choice probabilities $\boldsymbol{P}_{\kappa}$ are non-parametrically 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 distribution of firms' entry decisions, we establish the identification of the equilibrium probabilities $\boldsymbol{P}_{\kappa}$ and the distribution f_{κ}. Then, given the structure of the selection function in (27), we show the identification of $\boldsymbol{\theta}_{\delta}$.

4.2 First Step: Game of Market Entry

Let $|\mathcal{K}(\boldsymbol{x})|$ be the number of points in the support of κ_t when $\boldsymbol{x}_t = \boldsymbol{x}$ such that the support set is $\{1, 2, ..., |\mathcal{K}(\boldsymbol{x})|\}$. The identification of the equilibrium probabilities $\boldsymbol{P}_{\kappa}$ and the distribution $\boldsymbol{f}_{\kappa}$ is based on the structure of the joint probability distribution of the entry decisions of the J firms. For any value $(\boldsymbol{a}, \boldsymbol{x}) \in \{0,1\}^J \times \mathcal{X}$:

$$\Pr\left(\boldsymbol{a}_t = \boldsymbol{a} | \boldsymbol{x}_t = \boldsymbol{x}\right) \quad = \quad \sum_{\kappa=1}^{|\mathcal{K}(\boldsymbol{x})|} f_{\kappa}(\kappa|\boldsymbol{x}) \left[\prod_{j=1}^{J} \left[P_j(\boldsymbol{x},\kappa)\right]^{a_j} \quad \left[1 - P_j(\boldsymbol{x},\kappa)\right]^{1-a_{jt}}\right] \tag{29}$$

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, Neeman, Pakyari, and Elmore (2005), Allman, Matias, and Rhodes (2009), and Kasahara and Shimotsu (2014), among others. Identification is based on the assumption of independence between firms' entry decisions once we condition on $\boldsymbol{x}_t$ and κ_t.

In this first step, the proof of identification is pointwise for each value of $\boldsymbol{x}$. For simplicity in notation, for the rest of this subsection, we omit any further reference to $\boldsymbol{x}$ and to the market subscript t .

4.2.1 Identification of the number of components $|\mathcal{K}|$

The number of components $|\mathcal{K}|$ in the finite mixture (29) is typically unknown to the researcher. Following ideas similar to Bonhomme, Jochmans, and Robin (2016), Xiao (2018), and Aguirregabiria and Mira (2019), we start our first step identification argument by providing sufficient conditions for the unique determination of $|\mathcal{K}|$ from observables. 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, $i = 2, 3$, such that $\tilde{J} = (J-1)/2$ if J is odd, and $\tilde{J} = (J-2)/2$ 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 the number of components $|\mathcal{K}|$ from the observed joint distribution of Y_2 and Y_3:

$$\Pr(Y_2 = y_2, Y_3 = y_3) \;=\; \sum_{\kappa=1}^{|\mathcal{K}|} \Pr(Y_2 = y_2|\kappa) \;\Pr(Y_3 = y_3|\kappa)\; f_\kappa(\kappa) \tag{30}$$

or, in matrix notation,

$$\boldsymbol{P}_{Y_2,Y_3} \;=\; \boldsymbol{P}_{Y_2|\kappa}\; diag(\boldsymbol{f}_\kappa)\; \boldsymbol{P}'_{Y_3|\kappa}, \tag{31}$$

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

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

The point identification of the number of components $|\mathcal{K}|$ 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.

4.2.2 Identification of equilibrium CCPs and distribution of κ

Allman, Matias, and Rhodes (2009) study the identification of non-parametric multinomial finite mixtures that include our binary choice model as a particular case. They establish that a mixture with $|\mathcal{K}|$ components is identified if $J \geq 3$ and $|\mathcal{K}| \leq 2^J/(J + 1)$. The following Lemma 3 is an application to our model of Theorem 4 and Corollary 5 by Allman, Matias, and Rhodes (2009).

LEMMA 3. *Suppose that: (i) $J \geq 3$;* (ii) $|\mathcal{K}| \leq 2^J/(J + 1)$*; and (iii) for $i = 1, 2, 3$, the $|\mathcal{K}|$ vectors $\boldsymbol{P}_{Y_i}(\kappa = 1)$, $\boldsymbol{P}_{Y_i}(\kappa = 2)$, ..., $\boldsymbol{P}_{Y_i}(\kappa = |\mathcal{K}|)$ are linearly independent. Then, the probability distribution of κ — i.e., $f_\kappa(\kappa)$ for $\kappa = 1, 2, ..., |\mathcal{K}|$ — and the equilibrium CCPs — i.e., $P_j(\kappa)$ for $j = 1, 2, ..., J$ and $\kappa = 1, 2, ..., |\mathcal{K}|$ — are uniquely identified up to label swapping.* ■

Remark 1. Note that the order condition (i) in Lemma 2 is in general more stringent than the order condition (ii) of Lemma 3: 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 2 holds and the $|\mathcal{K}|$ vectors $\boldsymbol{P}_{Y_1}(\kappa = 1)$, $\boldsymbol{P}_{Y_1}(\kappa = 2)$, ..., $\boldsymbol{P}_{Y_1}(\kappa = |\mathcal{K}|)$ are linearly independent, then $|\mathcal{K}| = \text{Rank}(\boldsymbol{P}_{Y_2,Y_3})$ and the distribution of κ and the equilibrium CCPs are uniquely identified, up to label swapping.

Remark 2. Remember that Lemmas 2 and 3 apply to the model where the distribution of κ_t conditional on $\boldsymbol{x}_t$ is completely unrestricted. In our model, to achieve identification in the second step and also to improve the precision of the estimator in empirical applications, we assume that κ_t and $\boldsymbol{x}_t$ are independently distributed. This implies that the order condition of identification in our model is weaker than those in Lemmas 2 and 3.

4.3 Second Step: Identification of Demand Parameters

Following the discussion in section 2.1, we represent the demand system using the inverse $d_j^{(\boldsymbol{a})-1}(\boldsymbol{s}_t^{(\boldsymbol{a})}, \boldsymbol{p}_t^{(\boldsymbol{a})}, \boldsymbol{x}_t^{(\boldsymbol{a})})$ from Lemma 1. For those markets with $a_{jt} = 1$, the demand equation can be expressed as:

$$\delta_j(\boldsymbol{s}_t, \boldsymbol{\sigma}) \;=\; \alpha\; p_{jt} + \boldsymbol{x}'_{jt}\; \boldsymbol{\beta} + \xi_{jt}, \qquad \text{for } a_{jt} = 1 \tag{32}$$

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

Suppose for a moment that the market type κ_t were observable to the researcher after identification in the first step. In this case, the selection term would be $\psi_j\left(P_j(\boldsymbol{x}_t, \kappa_t), \kappa_t\right)$ from equation (25) and we would have — as in Example 2 — a relatively standard selection problem represented by the semi-parametric partially linear model:

$$\delta_j(\boldsymbol{s}_t, \boldsymbol{\sigma}) = \alpha\; p_{jt} + \boldsymbol{x}'_{jt}\; \boldsymbol{\beta} + \psi_j\left(P_j(\boldsymbol{x}_t, \kappa_t), \kappa_t\right) + \widetilde{\xi}_{jt}. \tag{33}$$

A key complication of the selection problem in our model is that the market type κ_t 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 $\boldsymbol{x}_t$. Therefore, in the second step we cannot condition on κ_t as in equation (33). We instead need to deal with the more complex selection bias function:

$$\lambda_j(\boldsymbol{x}_t) \equiv \mathbb{E}\left[\xi_{jt} | \boldsymbol{x}_t, a_{jt} = 1\right] = \sum_{\kappa=1}^{|\mathcal{K}(\boldsymbol{x}_t)|} f_\kappa(\kappa|\boldsymbol{x}_t)\; \psi_j\left(P_j(\boldsymbol{x}_t, \kappa), \kappa\right) \;=\; \boldsymbol{f}'_{\kappa,t}\; \boldsymbol{\psi}_j(\boldsymbol{P}_{j,t}), \tag{34}$$

where $\boldsymbol{f}_{\kappa,t}$, $\boldsymbol{P}_{j,t}$, and $\boldsymbol{\psi}_j(\boldsymbol{P}_{j,t})$ are all vectors of dimension $|\mathcal{K}(\boldsymbol{x}_t)| \times 1$ such that $\boldsymbol{f}_{\kappa,t} \equiv (f_\kappa(\kappa|\boldsymbol{x}_t) : \kappa = 1, 2, ..., |\mathcal{K}(\boldsymbol{x}_t)|)$, $\boldsymbol{P}_{j,t} \equiv (P_j(\boldsymbol{x}_t, \kappa) : \kappa = 1, 2, ..., |\mathcal{K}(\boldsymbol{x}_t)|)$, and $\boldsymbol{\psi}_j(\boldsymbol{P}_{j,t}) \equiv (\psi_j\left(P_j(\boldsymbol{x}_t, \kappa), \kappa\right) :$

$\kappa = 1, 2, ..., |\mathcal{K}(\boldsymbol{x}_t)|$). Therefore, the regression equation of our model is:

$$\delta_j(\boldsymbol{s}_t, \boldsymbol{\sigma}) = \alpha\ p_{jt} + \boldsymbol{x}'_{jt}\ \boldsymbol{\beta} + \boldsymbol{f}'_{\kappa,t}\ \boldsymbol{\psi}_j(\boldsymbol{P}_{j,t}) + \widetilde{\xi}_{jt}. \tag{35}$$

where $\mathbb{E}[\widetilde{\xi}_{jt}|\boldsymbol{x}_t] = \mathbb{E}[\widetilde{\xi}_{jt}|\boldsymbol{P}_{j,t}, \boldsymbol{f}_{\kappa,t}] = 0$. This last equality establishes that $(\boldsymbol{P}_{j,t}, \boldsymbol{f}_{\kappa,t})$ is a sufficient statistic for the selection function. This property is key for the identification result in this second step.

The following Proposition establishes a high-level necessary and sufficient condition for the identification of $\boldsymbol{\theta}_\delta \equiv (\alpha, \boldsymbol{\beta}, \boldsymbol{\sigma})$ from equation (35).

PROPOSITION 1. *Define the vector* $\boldsymbol{Z}_{jt} \equiv (\ \mathbb{E}[\partial\delta_{jt}/\partial\boldsymbol{\sigma} \mid \boldsymbol{x}_t], \mathbb{E}[p_{jt} \mid \boldsymbol{x}_t], \boldsymbol{x}'_{jt}\ )'$*, and let* $\widetilde{\boldsymbol{Z}}_{jt}$ *be the deviation (or residual)* $\boldsymbol{Z}_{jt} - \mathbb{E}[\boldsymbol{Z}_{jt} \mid \boldsymbol{P}_{j,t}, \boldsymbol{f}_{\kappa,t}]$*. Then, given that* $\mathbb{E}[\widetilde{\xi}_{jt}|\boldsymbol{x}_t] = \mathbb{E}[\widetilde{\xi}_{jt}|\boldsymbol{P}_{j,t}, \boldsymbol{f}_{\kappa,t}] = 0$*, a necessary and sufficient condition for the identification of* $\boldsymbol{\theta}_\delta \equiv (\alpha, \boldsymbol{\beta}, \boldsymbol{\sigma})$ *in equation* (35) *is that matrix* $\mathbb{E}[\widetilde{\boldsymbol{Z}}_{jt}\ \widetilde{\boldsymbol{Z}}'_{jt}]$ *is full-rank.* ■

Intuitively, Proposition 1 says that the identification of $\boldsymbol{\theta}_\delta$ requires that, after differencing out any dependence with respect to $(\boldsymbol{P}_{j,t}, \boldsymbol{f}_{\kappa,t})$, there should be no perfect collinearity in the vector of "explanatory variables" $\boldsymbol{Z}_{jt} \equiv (\ \mathbb{E}[\partial\delta_{jt}/\partial\boldsymbol{\sigma} \mid \boldsymbol{x}_t], \mathbb{E}[p_{jt} \mid \boldsymbol{x}_t], \boldsymbol{x}'_{jt}\ )'$.

Proposition 1 does not provide identification conditions that apply directly to primitives of the model. However, based on this Proposition, it is straightforward to establish necessary identification conditions that apply to primitives of the model, or to objects which are more closely related to primitives. First, we need $J \geq 2$, otherwise there are no exclusion restrictions to deal with the endogeneity of prices, i.e., $\mathbb{E}[p_{jt} \mid \boldsymbol{x}_t]$ would be a linear combination of $\boldsymbol{x}_{jt}$. Second, the vector of entry probabilities $\boldsymbol{P}_{j,t}$ should depend on $\boldsymbol{x}_{it}$ for $i \neq j$. Otherwise, keeping $\boldsymbol{P}_{j,t}$ fixed implies also fixing $\boldsymbol{x}_{jt}$ and the vector of parameters $\boldsymbol{\beta}$ would not be identified. Hence, there should be effective competition in firms' market entry decisions. For instance, in the absence of observable variables affecting entry but not demand, the model would not be identified for a monopolist or under monopolistic competition. Third, the number of points in the support of κ should be smaller than the number of variables in vector $\boldsymbol{x}_t$: i.e., $|\mathcal{K}(\boldsymbol{x}_t)| < dim(\boldsymbol{x}_t)$. Otherwise, controlling for $\boldsymbol{P}_{j,t}$ would be equivalent to controlling for the whole vector $\boldsymbol{x}_t$, and no parameter in $\boldsymbol{\theta}_\delta$ would be identified.

5 Estimation and Inference

In this section, we build on our previous constructive proof of identification and present a two step estimation method that mimics our sequential identification result. In the first step, we estimate the number of unobserved market types $|\mathcal{K}(\boldsymbol{x}_t)|$, the distribution $\boldsymbol{f}_{\kappa,t}$, and the vector of CCPs for each unobserved type. For the second step, we use the method of sieves. We specify each function $\psi_j\left(P_j(\boldsymbol{x}_t,\kappa),\kappa\right)$ as a polynomial function of the entry probability $P_j(\boldsymbol{x}_t,\kappa)$ such that the selection term $\boldsymbol{f}'_{\kappa,t}\boldsymbol{\psi}_j(\boldsymbol{P}_{j,t})$ is a linear-in-parameters function of products between powers of $P_j(\boldsymbol{x}_t,\kappa)$ and densities $f_\kappa(\kappa|\boldsymbol{x}_t)$. Then, we plug in this expression the estimates of $P_j(\boldsymbol{x}_t,\kappa)$ and $f_\kappa(\kappa|\boldsymbol{x}_t)$ from the first step, and jointly estimate the coefficients in the sieves approximation and the structural demand parameters $\boldsymbol{\theta}_\delta$.

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

In this section, we describe our approach to estimate $\boldsymbol{f}_{\kappa,t}$ and $\boldsymbol{P}_{j,t}$ from data on firms' entry decisions across markets. The method that we present here builds on the nonparametric finite mixture methods in Kasahara and Shimotsu (2014), Xiao (2018), and Aguirregabiria and Mira (2019) and extends these methods by allowing vector $\boldsymbol{x}_t$ to include continuous variables.

Step 1(a): Series Logit Estimator. To accommodate continuous explanatory variables in the estimation of the nonparametric finite mixture, we start our sequential procedure by introducing smoothness in the nonparametric estimates of the entry profile probabilities $\mathbb{P}(\boldsymbol{a}_t=\boldsymbol{a}|\boldsymbol{x}_t=\boldsymbol{x})$. Following Hirano, Imbens, and Ridder (2003), we consider a sieves approximation to $\mathbb{P}(\boldsymbol{a}_t=\boldsymbol{a}|\boldsymbol{x}_t=\boldsymbol{x})$ based on the following Series Logit model:

$$\mathbb{P}(\boldsymbol{a}_t=\boldsymbol{a}|\boldsymbol{x}_t=\boldsymbol{x}) \;=\; \frac{\exp\{\boldsymbol{r}(\boldsymbol{x})'\;\boldsymbol{\pi}(\boldsymbol{a})\}}{\sum_{\boldsymbol{a}'\in\{0,1\}^J}\exp\{\boldsymbol{r}(\boldsymbol{x})'\;\boldsymbol{\pi}(\boldsymbol{a}')\}} \tag{36}$$

where $\boldsymbol{r}(\boldsymbol{x})$ is a vector with L approximating functions, i.e., $(r_1(\boldsymbol{x}),r_2(\boldsymbol{x}),...,r_L(\boldsymbol{x}))'$, and $\boldsymbol{\pi}\equiv(\boldsymbol{\pi}(\boldsymbol{a}):\boldsymbol{a}\in\{0,1\}^J)$ is a vector of parameters with the normalization $\boldsymbol{\pi}(\boldsymbol{0})=0$. This vector of parameters is estimated using the maximum likelihood estimator:

$$\widehat{\boldsymbol{\pi}} \;=\; argmax_{\boldsymbol{\pi}}\sum_{t=1}^{T}\sum_{\boldsymbol{a}\in\{0,1\}^J}1\{\boldsymbol{a}_t=\boldsymbol{a}\}\;\ln\left(\frac{\exp\{\boldsymbol{r}(\boldsymbol{x_t})'\;\boldsymbol{\pi}(\boldsymbol{a})\}}{\sum_{\boldsymbol{a}'\in\{0,1\}^J}\exp\{\boldsymbol{r}(\boldsymbol{x}_t)'\;\boldsymbol{\pi}(\boldsymbol{a}')\}}\right) \tag{37}$$

Hence, for any value of $(\boldsymbol{a},\boldsymbol{x})$, we estimate the entry profile probability $\mathbb{P}(\boldsymbol{a}|\boldsymbol{x})$ using the Series Logit Estimator $\exp\{\boldsymbol{r}(\boldsymbol{x})'\widehat{\boldsymbol{\pi}}(\boldsymbol{a})\}\;/\;\sum_{\boldsymbol{a}'}\exp\{\boldsymbol{r}(\boldsymbol{x})'\widehat{\boldsymbol{\pi}}(\boldsymbol{a}')\}$.

The rest of the estimation in this first step applies pointwise to each value $\boldsymbol{x}_t$ in the sample. This first step estimation starts with the construction of the estimate $\widehat{\mathbb{P}}(\boldsymbol{a}|\boldsymbol{x}_t)$ using the Series Logit Estimator. For notational simplicity, for the rest of this subsection, we omit $\boldsymbol{x}_t$ as an argument.

Step 1(b): Estimating the number of market types $|\mathcal{K}|$. As shown is Lemma 2 above, the number of market types $|\mathcal{K}|$ is identified by $Rank(\boldsymbol{P}_{Y_2,Y_3})$. We estimate this rank using the sequence of rank tests proposed by Robin and Smith (2000) and Xiao (2018).

For a given value $\boldsymbol{x}_t$ and a partition of the entry profile $\boldsymbol{a}$ into the variables Y_1, Y_2, and Y_3, let $\boldsymbol{P}_{Y_2,Y_3}$ be the matrix defined in section 4.2.1 above, and let $\widehat{\boldsymbol{P}}_{Y_2,Y_3}$ be the estimator of this matrix based on the Series Logit estimation in Step 1(a). We describe here the method to estimate $|\mathcal{K}| = Rank(\widehat{\boldsymbol{P}}_{Y_2,Y_3})$.

Remember that matrix $\widehat{\boldsymbol{P}}_{Y_2,Y_3}$ has dimension $2^{\tilde{J}} \times 2^{\tilde{J}}$, with $\tilde{J} = (J-1)/2$ if J is odd, and $\tilde{J} = (J-2)/2$ if J is even. For any natural number $r \in \{1, ..., 2^{\tilde{J}} - 1\}$, consider the null hypothesis H_0^r : $\text{Rank}(\widehat{\boldsymbol{P}}_{Y_2,Y_3}) = r$ against the alternative H_1^r : $\text{Rank}(\widehat{\boldsymbol{P}}_{Y_2,Y_3}) > r$. This test is based on statistic CRT_r which represents the $r-th$ characteristic root of the matrix quadratic form, such that $CRT_1 \geq CRT_2 \geq ... \geq CRT_{2^{\tilde{J}}}$. Given a significance level α, there is a critical value $c_{1-\alpha}^r$ such that the acceptance region for H_0^r is $\{CRT_r \leq c_{1-\alpha}^r\}$.

The estimator of $\text{Rank}(\widehat{\boldsymbol{P}}_{Y_2,Y_3})$ is based on a sequence of this rank test. The sequence starts with a null hypothesis of rank equal to $r = 1$. If this null is rejected, then $r = 2$ and the test is repeated, and so on. Along this sequence of tests, $\text{Rank}(\widehat{\boldsymbol{P}}_{Y_2,Y_3})$ is estimated as the value of r for which H_0^r obtains the first rejection. That is:

$$\widehat{|\mathcal{K}|} \;=\; \widehat{\text{Rank}}(\widehat{\boldsymbol{P}}_{Y_2,Y_3}) \;=\; \min_{r\in\{1,...,2^{\tilde{J}}-1\}}\{r:\; CRT_r \leq c_{1-\alpha}^r\}. \tag{38}$$

In order to deal with multiple testing and to guarantee weak consistency of this rank estimator, Robin and Smith (2000) adjust the asymptotic size of the test at each stage r to depend on the sample size T, such that we have $\alpha_{r,M}$. See their Theorem 5.2 for a characterization of $\alpha_{r,M}$.[6]

Step 1(c): Estimating $\boldsymbol{f}_{\kappa,t}$ and $\boldsymbol{P}_{j,t}$. We follow the method in Xiao (2018). Once the number of components $\widehat{|\mathcal{K}|}$ is determined, we can collapse the $2^{\tilde{J}} \times 2^{\tilde{J}}$ matrix $\widehat{\boldsymbol{P}}_{Y_2,Y_3}$ of rank $\widehat{|\mathcal{K}|}$ into a smaller $\widehat{|\mathcal{K}|} \times \widehat{|\mathcal{K}|}$ non-singular matrix $\widehat{\boldsymbol{P}}_{\tilde{Y}_2,\tilde{Y}_3}$. This non-singular matrix can be obtained from $\widehat{\boldsymbol{P}}_{Y_2,Y_3}$ by summing up some of its columns and rows.[7]

[6]For more detail on rank tests, see also Kleibergen and Paap (2006). Given the well known difficulties in characterizing the asymptotic distribution of rank estimators, in what follows we consider this estimator of $|\mathcal{K}|$ as a tool to facilitate model selection. In this sense, we interpret the weak consistency of $|\mathcal{K}|$ as consistent model selection.

[7]Lemma 2 in Xiao (2018) proves that such a transformation is always possible. Given that for any $\boldsymbol{P}_{Y_2,Y_3}$

Given $\widehat{\boldsymbol{P}}_{\tilde{Y}_2,\tilde{Y}_3}$, we define the full rank matrices $\widehat{\boldsymbol{P}}_{\tilde{Y}_2|\kappa}$ and $\widehat{\boldsymbol{P}}_{\tilde{Y}_3|\kappa}$ as in equation (31), as well as the diagonal matrix $\widehat{\boldsymbol{P}}_{Y_1=a|\kappa} \equiv \text{diag}\left[\Pr(Y_1 = a|\kappa = 1), ..., \Pr(Y_1 = a|\kappa = L)\right]$. By construction, the observed matrix $\widehat{\boldsymbol{P}}_{Y_1,\tilde{Y}_2,\tilde{Y}_3}$ satisfies this equation:

$$\widehat{\boldsymbol{P}}_{Y_1,\tilde{Y}_2,\tilde{Y}_3} = \widehat{\boldsymbol{P}}_{\tilde{Y}_2|\kappa}\ \widehat{\boldsymbol{P}}_{Y_1|\kappa}\ diag\{\widehat{\boldsymbol{f}}_\kappa\}\ \widehat{\boldsymbol{P}}'_{\tilde{Y}_3|\kappa}, \tag{39}$$

where all the matrices are of dimension $|\widehat{\mathcal{K}}|\times|\widehat{\mathcal{K}}|$. On the basis of the following eigen-decomposition,

$$\widehat{\boldsymbol{P}}_{Y_1,\tilde{Y}_2,\tilde{Y}_3}\ \widehat{\boldsymbol{P}}^{-1}_{\tilde{Y}_2,\tilde{Y}_3} = \widehat{\boldsymbol{P}}_{\tilde{Y}_2|\kappa}\ \widehat{\boldsymbol{P}}_{Y_1|\kappa}\ \widehat{\boldsymbol{P}}^{-1}_{\tilde{Y}_2|\kappa}, \tag{40}$$

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

$$\widehat{\boldsymbol{P}}_{\tilde{Y}_2|\kappa} = \rceil\left(\widehat{\boldsymbol{P}}_{Y_1,\tilde{Y}_2,\tilde{Y}_3}\ \widehat{\boldsymbol{P}}^{-1}_{\tilde{Y}_2,\tilde{Y}_3}\right), \tag{41}$$

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:

$$\begin{aligned} \widehat{\boldsymbol{f}}_\kappa &= \widehat{\boldsymbol{P}}^{-1}_{\tilde{Y}_2|\kappa}\ \widehat{\boldsymbol{P}}_{\tilde{Y}_2} \\ \widehat{\boldsymbol{P}}_{\tilde{Y}_3|\kappa} &= \left[\left(\widehat{\boldsymbol{P}}_{\tilde{Y}_2|\kappa}\ diag\{\widehat{\boldsymbol{f}}_\kappa\}\right)^{-1}\ \widehat{\boldsymbol{P}}_{\tilde{Y}_2,\tilde{Y}_3}\right]', \end{aligned} \tag{42}$$

where $\boldsymbol{P}_{\tilde{Y}_2}$ is the observed $|\widehat{\mathcal{K}}| \times 1$ vector with the marginal probabilities $\Pr(\tilde{Y}_2)$.

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

$$\begin{aligned} \boldsymbol{P}_{\tilde{Y}_2,j} &= \boldsymbol{P}_{\tilde{Y}_2|\kappa}\ diag\{\boldsymbol{f}_\kappa\}\ \boldsymbol{P}'_{j,\kappa} \quad \text{for any } a_j = 1 \text{ part of } Y_1 \text{ or } \tilde{Y}_3 \\ \boldsymbol{P}_{j,\tilde{Y}_3} &= \boldsymbol{P}_{j,\kappa}\ diag\{\boldsymbol{f}_\kappa\}\ \boldsymbol{P}'_{\tilde{Y}_3|\kappa} \quad \text{for any } a_j = 1 \text{ part of } \tilde{Y}_2, \end{aligned} \tag{43}$$

where $\boldsymbol{P}_{\tilde{Y}_2,j}$ is the observed $|\widehat{\mathcal{K}}| \times 1$ vector with the joint probabilities $\Pr(\tilde{Y}_2 = \tilde{y}_2, a_j = 1)$, $\boldsymbol{P}_{j,\tilde{Y}_3}$ is the analogous observed $1 \times |\widehat{\mathcal{K}}|$ vector, and $\boldsymbol{P}_{j,\kappa}$ is the vector of firm j's CCPs. Finally, the

---

there are many ways of constructing $|\widehat{\mathcal{K}}| \times |\widehat{\mathcal{K}}|$ non-singular matrices, Xiao (2018) suggests to create various candidates and then to pick the one associated to the smallest *condition number*.The intuition is that the smaller the condition number of a matrix, the more likely the matrix is to be non-singular.

vector of CCPs $\boldsymbol{P}_{j,\kappa}$ can be estimated for any firm j as:

$$\begin{aligned}
\widehat{\boldsymbol{P}}_{j,\kappa} &= \left[\left(\widehat{\boldsymbol{P}}_{\tilde{Y}_2|\kappa}\, diag\{\widehat{\boldsymbol{f}}_{\kappa}\}\right)^{-1}\ \widehat{\boldsymbol{P}}_{\tilde{Y}_2,j}\right]' \quad \text{for any } a_j = 1 \text{ part of } Y_1 \text{ or } \tilde{Y}_3 \\
\widehat{\boldsymbol{P}}_{j,\kappa} &= \widehat{\boldsymbol{P}}_{j,\tilde{Y}_3}\left(diag\{\widehat{\boldsymbol{f}}_{\kappa}\}\ \widehat{\boldsymbol{P}}'_{\tilde{Y}_3|\kappa}\right)^{-1} \qquad \text{for any } a_j = 1 \text{ part of } \tilde{Y}_2.
\end{aligned} \tag{44}$$

Asymptotics. Xiao (2018) characterizes the asymptotic properties of the estimation procedure described here and shows that the proposed estimator is $\sqrt{T}$-consistent and asymptotically normal when $\boldsymbol{x}_t$ is discrete. In contrast, we allow for both continuous and discrete variables in $\boldsymbol{x}_t$, and we incorporate step 1(a) where we estimate the entry-profile probabilities $\mathbb{P}(\boldsymbol{a}|\boldsymbol{x})$ using a Series Logit Estimator (SLE). With continuous variables in $\boldsymbol{x}$, the SLE cannot achieve a $\sqrt{T}$ rate. However, under standard regularity conditions, this does not affect the $\sqrt{T}$-consistency of the estimators in steps 1(b) and 1(c). The proof of this result follows from Hirano, Imbens, and Ridder (2003). The fact that SLE is slower than $\sqrt{T}$ does affect the convergence rates of the estimators in steps 1b and 1c. However, this slower rate in the first step does not prevent the demand parameters $\boldsymbol{\theta}_{\delta}$ to be consistent and $\sqrt{T}$-asymptotically normal. This result follows from Das, Newey, and Vella (2003).

5.2 Second Step: Estimation of demand parameters

Here we describe a GMM estimator of demand parameters $\boldsymbol{\theta}_{\delta}$. In the same spirit as Das, Newey, and Vella (2003), we use the method of sieves and approximate each function $\psi_j\left(P_j(\boldsymbol{x}_t, \kappa), \kappa\right)$ using a polynomial of order Q of the entry probability $P_j(\boldsymbol{x}_t, \kappa)$[8]. That is:

$$\psi_j\left(\widehat{P}_{j,m,\kappa}, \kappa\right) \approx \gamma_{0,j,\kappa} + \gamma_{1,j,\kappa}\ \widehat{P}_{j,m,\kappa} + ... + \gamma_{Q,j,\kappa}\ (\widehat{P}_{j,m,\kappa})^Q \tag{45}$$

where $(\gamma_{0,j,\kappa}, \gamma_{1,j,\kappa}..., \gamma_{Q,j,\kappa})$ are parameters. Given this approximation, the selection function is linear in these γ parameters and has the following expression:

$$\widehat{\boldsymbol{f}}'_{\kappa,t}\boldsymbol{\psi}_j(\widehat{\boldsymbol{P}}_{j,t}) \approx \boldsymbol{h}'_{j,t}\ \boldsymbol{\gamma}_j = \sum_{\kappa=1}^{|\mathcal{K}(\boldsymbol{x}_t)|}\sum_{q=0}^{Q}\gamma_{q,j,\kappa}\ \widehat{f}_{\kappa}(\kappa|\boldsymbol{x}_t)\ (\widehat{P}_{j,m,\kappa})^q \tag{46}$$

where $\boldsymbol{h}'_{j,t}$ is the $1 \times (Q+1)|\mathcal{K}(\boldsymbol{x}_t)|$ vector with elements $\{\widehat{f}_{\kappa}(\kappa|\boldsymbol{x}_t)\ (\widehat{P}_{j,m,\kappa})^q : q = 0, 1, ..., Q; \kappa = 1, ..., |\mathcal{K}(\boldsymbol{x}_t)|\}$.

[8]The demand parameters $\boldsymbol{\theta}_{\delta}$ can also be estimated by differencing out the selection term following Ahn and Powel (1993) and Powell (2001).

Plugging equation (46) into the demand equation (35), we have the regression equation:

$$\delta_j(\boldsymbol{s}_t, \boldsymbol{\sigma}) = \alpha\ p_{jt} + \boldsymbol{x}'_{jt}\ \boldsymbol{\beta} + \boldsymbol{h}'_{j,t}\ \boldsymbol{\gamma}_j + \widetilde{\xi}_{jt}. \tag{47}$$

Once the order of polynomial K is chosen, equation 47 can be estimated by methods such as 2SLS or GMM. The estimation of our partially linear model is not different from estimation of parametric models in practice. However, by specifying ψ_j as a nonparametric function, we allow the order of the polynomial K grow with the sample size, T. This in turn effects the convergence rate of our estimator. Following Das, Newey, and Vella (2003), one can show that $\hat{\psi}_j$ is consistent and that the finite dimensional demand parameters $\boldsymbol{\theta}_\delta$ can be estimated consistently and shown to be $\sqrt{T}$-asymptotically normal.

As in Xiao(2018), our estimation is not affected by the problem of "identification up to label swapping". This is due to the specific form of the selection function given in equation (35), which does not need labelling the equilibrium selection probabilities and rather includes all values of selection probabilities and the corresponding CCP's for each firm j.

6 Monte Carlo Experiments

In this section, we present results from 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. 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 $M \in \{200, 500, 2000, 4000\}$ 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 $a_{jt} = 1$, the equation for the demand of product $j = 1, 2, 3$ is:

$$\ln\left(\frac{s_{jt}}{s_{0t}}\right) = \beta\ x_{jt} + \alpha\ p_{jt} + \sigma\ \ln\left(\frac{s_{jt}}{1 - s_{0t}}\right) + \xi_{jt}, \tag{48}$$

where $s_{jt}/(1-s_{0t}) = s_{jt}/(s_{1t}+s_{2t}+s_{3t})$ is the within-nest market share of product j. Variable x_{jt} is a characteristic of product j that varies across markets.[9] We consider $x_{jt} \sim i.i.d.$ log-normal$(0, \sigma_x^2)$. We denote vector (x_{1t}, x_{2t}, x_{3t}) by $\boldsymbol{x}_t$.

6.1.2 Demand unobservables.

The demand unobservables $(\xi_{1t}, \xi_{2t}, \xi_{3t})$ 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 t, κ_t, determines the mean of the normal distribution of the variables ξ_{jt}. This mean is equal to μ_ℓ if $\kappa_t = \ell$ and equal to μ_h if $\kappa_t = h$. Accordingly, we have:

$$\xi_{jt} \sim 1\{\kappa_t = \ell\}\ N(\mu_\ell, \sigma_\xi^2) + 1\{\kappa_t = h\}\ N(\mu_h, \sigma_\xi^2). \tag{49}$$

Market type κ_t is independent of $\boldsymbol{x}_t$ and $i.i.d.$ across markets with $\Pr(\kappa_t = \ell) = f_\kappa(\ell)$. 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_t is the same for the three firms, and this introduces positive correlation among ξ_{1t}, ξ_{2t}, and ξ_{3t}.

Note also that ξ_{jt} is correlated with the unobservable component of the entry cost η_{jt}. See below the 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 à la Bertrand. In this nested logit model, equilibrium prices given entry profile $\boldsymbol{a}$ — which we represent as $\boldsymbol{p}_t(\boldsymbol{a}) = (p_{1t}(\boldsymbol{a}),\, p_{2t}(\boldsymbol{a}),\, p_{3t}(\boldsymbol{a}))$ — are the solution to the following system of best response equations:

$$p_{jt}(\boldsymbol{a}) = mc_{jt} - \frac{1-\sigma}{\alpha\left(1 - \sigma\ \frac{s_{jt}(\boldsymbol{p}_t(\boldsymbol{a}),\boldsymbol{a})}{s_{1t}(\boldsymbol{p}_t(\boldsymbol{a}),\boldsymbol{a}) + s_{2t}(\boldsymbol{p}_t(\boldsymbol{a}),\boldsymbol{a}) + s_{3t}(\boldsymbol{p}_t(\boldsymbol{a}),\boldsymbol{a})} - (1-\sigma)\ s_{jt}(\boldsymbol{p}_t(\boldsymbol{a}), \boldsymbol{a})\right)}. \tag{50}$$

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

$$p_{jt}(\boldsymbol{a}) = mc_{jt} - \frac{1-\sigma}{\alpha\left(1 - \sigma\ \frac{s^*_{jt}(\boldsymbol{mc}_t,\boldsymbol{a})}{s^*_{1t}(\boldsymbol{mc}_t,\boldsymbol{a}) + s^*_{2t}(\boldsymbol{mc}_t,\boldsymbol{a}) + s^*_{3t}(\boldsymbol{mc}_t,\boldsymbol{a})} - (1-\sigma)\ s^*_{jt}(\boldsymbol{mc}_t, \boldsymbol{a})\right)}, \tag{51}$$

[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_{jt} can be the number of other airports that the airline connects to from/to the airports in market t.

where $s^*_{jt}(\boldsymbol{mc}_t, \boldsymbol{a})$ represents the market share of product j in market t 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 pricing to their marginal costs.

A firm's marginal cost in market t is a deterministic function of x_{jt}: $mc_{jt} = \omega_0 + \omega_1\ x_{jt}$, where ω_0 and ω_1 are parameters.

6.1.4 Market entry game.

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

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

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

$$P_{jt} = \Phi\left(\pi^P_j(\boldsymbol{x}_t, \kappa_t) - \gamma_j\ z_t\right), \tag{52}$$

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

The unobserved variable η_{jt} is correlated with the demand unobservable ξ_{jt}. Note that this introduces a source of endogenous selection in addition to the one that comes from κ_t. The parameter ρ measures the correlation between η_{jt} and ξ_{jt}. Therefore, we have that:

$$\mathbb{E}\left(\xi_{jt} \mid \boldsymbol{x}_t, z_t, \kappa_t, a_{jt} = 1\right) = \mu_{\kappa_t} + \rho\ \frac{\phi\left(\left(\pi^P_j(\boldsymbol{x}_t, \kappa_t) - \gamma_j\ z_t\right)\right)}{P_j(\boldsymbol{x}_t, z_t, \kappa_t)}, \tag{53}$$

where $\phi(\cdot)$ is the standard normal density function. If the researcher knew the true distribution of the unobservables, then she could use (53) to construct an appropriate control function and directly account for selection in the second step of the estimator. We suppose that the researcher does not have such information. However, in our experiments, we evaluate the improvement in the precision of the estimator if the researcher possessed such knowledge.

6.1.5 Solving for an equilibrium of entry game.

For a given value of $(\boldsymbol{x}_t,\ z_t,\ \kappa_t)$, we need solve for a Bayesian Nash equilibrium of the entry game. There are two main computational tasks involved.

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

Second, we need to solve numerically for a fixed point of the system of equations (52). We do that on the basis of fixed point iterations.

6.1.6 Generating simulations

Our Monte Carlo experiments are based on 100 Monte Carlo repetitions or samples. We describe here the different steps to generate a single sample.

1. Generate the T values of z_t using the grid points described above.

2. Generate T independent random draws of $(x_t, \kappa_t : t = 1, 2, ...M)$ from the distribution of these variables.

3. For each market t, given $(\boldsymbol{x}_t,\ z_t,\ \kappa_t)$, compute the BNE CCPs $(P_j(\boldsymbol{x}_t, z_t, \kappa_t) : j = 1, 2, 3)$. Then, generate a_{jt} as a random draw from a Bernoulli with probability $P_j(\boldsymbol{x}_t, z_t, \kappa_t)$.

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

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

As is common in real datasets, also in our simulated data we generate only one realization of the entry decisions in each market. This implies that the estimated probabilities of entry $\Pr(\boldsymbol{a}_m|\boldsymbol{x}_m, z_m)$ are measured too imprecisely by frequency counters within each market t (with 3 firms, there are 8 such probabilities in each t). We however found the use of some form of smoothing across markets essential to increase the precision of the first step estimator: e.g., kernels as in (**??**) or flexible multinomial logits with alternative-specific coefficients and polynomial expansions of $(\boldsymbol{x}_m, z_m)$.[10] For the results presented below, we estimate the probabilities of entry by a multinomial logit with 8 alternatives (as the possible realizations of $\boldsymbol{a}_m$), alternative-specific coefficients, and standard polynomials of z_m of degree up to 1 for $M \leq 500$ and up to 2 for $M > 500$. Table 1 summarizes the values of all the parameters in the d.g.p.

Table 1. Values of Parameters in d.g.p.

Parameter	True Value	Parameter	True Value
$\beta =$	2.0	$\sigma_\xi =$	1.0
$\alpha =$	-2.0	$\rho \in$	$[0,\ 0.8]$
$\sigma =$	0.6		
$\mu_\ell =$	-0.5	$z_{\min} =$	0.1
$\mu_h =$	0.5	$z_{\max} =$	3.0
$f_\kappa(\ell) =$	0.5	$\omega_0 = \omega_1 =$	0.3
$\sigma_x =$	0.7	$\gamma_1 = \gamma_2 = \gamma_3 =$	0.4

6.2 Estimators

For each of the four market sizes $M \in \{200, 500, 2000, 4000\}$ configurations, we generate 100 repetitions of the data and implement four estimators.

(i) OLS. This 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. This is the classic BLP estimator of a nested logit demand system. It accounts for the endogenity of price and the within-nest share, but it ignores endogenous sample selection. We construct instruments on the basis of x_t.

(iii) Our estimator (differencing). It estimates $P_j(\boldsymbol{x}_t, z_t, \kappa_t)$ and $f_\kappa(\ell)$ non-parametrically on the basis of n_{obs} entry observations per market t, and then feeds them to a second step semi-parametric estimator which controls for both endogeneity and selection by differencing out the

[10] Note that both forms of smoothing are consistent with our asymptotic theory and do not introduce any further complication. In fact, our theoretical arguments explicitly account for smoothing by kernels at this initial stage, while smoothing by multinomial logits estimated by MLE would actually imply an even faster rate of convergence (Powell, 2001; Newey, 2009).

selection term $\boldsymbol{f}'_\kappa \boldsymbol{\psi}_j(\boldsymbol{P}_{jt})$. In the second step estimator, we use the same instruments as in the 2SLS. This is the "differencing" version of our proposed estimator.[11]

(iv) Our estimator (sieve). This is the "sieve" version of our proposed estimator, where the semi-parametric second step approximates each $\psi_j(P_j(\boldsymbol{x}_t, z_t, \kappa_t), \kappa_m)$ by polynomial expansions of $P_j(\boldsymbol{x}_t, z_t, \kappa_t)$ — thereby estimating the selection term $\boldsymbol{f}'_\kappa \boldsymbol{\psi}_j(\boldsymbol{P}_{jt})$ instead of differencing it out.[12] In the second step, we use again the same instruments as in the 2SLS.

In general, the nested logit demand model used to generate the data does not include alternative-specific intercepts (i.e., they are equal to zero) as they are not separately identified by both versions of our estimator *(iii)* and *(iv)*. To simplify comparisons, we do not include alternative-specific intercepts also in the specifications of the OLS and the 2SLS.[13]

6.3 Results

Table 2 shows descriptive statistics for a typical repetition of the d.g.p. with 4000 markets.

Table 2. Summary Statistics from DGP

	Perc. zeros	Av. mkt share	Av. $\frac{p-c}{p}$
Firm 1	69.1%	13.3%	56.0%
Firm 2	70.4%	13.2%	56.3%
Firm 3	69.6%	13.3%	55.5%

Table 3 reports the average point estimates and their standard deviations computed over 100 repetitions for the d.g.p. with $M = 200$ and correlation $\rho = 0.6$.

[11] We implement this as the pairwise differencing estimator (**??**) with multivariate gaussian kernel to specify the weight function $\mathcal{D}_{jmn}$ and bandwidth chosen according to Silverman (2018)'s rule of thumb.

[12] After extensive experimentation, we concluded that this estimator is robust to the choice of the degree and the type of polynomials used. Operationally, for the results presented below we use standard polynomials of degree up to 4 for $M = 200$ and up to 5 for $M \in \{500, 2000, 4000\}$, but alternative specifications with different polynomials (e.g., Hermite or Bernstein) and/or different degrees lead to extremely similar conclusions. For example, in the case of standard polynomials of degree up to 5, this amounts to estimating 6 functions $\psi_{j\kappa}(P_{jt}(\kappa)) = \sum_{d=0}^{5} \varphi_{jkd} \times (P_{jt}(\kappa))^d$, i.e. one for each (j, κ) combination, and hence $6 \times 5 = 30$ additional $\boldsymbol{\varphi}$ parameters plus 3 alternative-specific intercepts (which will absorb the terms $f_\kappa(\ell) \times \varphi_{j\ell 0} + (1 - f_\kappa(\ell)) \times \varphi_{jh0}$) with respect to differencing the selection term out.

[13] Controlling for alternative-specific intercepts attenuates the estimation bias induced by endogenous sample selection on the structural parameters, but introduces huge estimation bias on these intercepts (whose true value is zero).

Table 3. Monte Carlo Experiments with M = 200 and $\rho = 0.6$

		True Value	OLS	2SLS	Our_{diff}	$\text{Our}_{\text{sieve}}$
β	Mean	2	1.3736	1.5619	1.9676	1.9722
	Std. Dev.		(0.1150)	(1.0228)	(0.2215)	(0.2132)
α	Mean	-2	-1.9314	-2.0456	-2.0016	-2.0079
	Std. Dev.		(0.0519)	(0.5502)	(0.0739)	(0.0670)
σ	Mean	0.6	0.5973	0.5740	0.5985	0.5976
	Std. Dev.		(0.0186)	(0.0968)	(0.0278)	(0.0280)

Table 4 reports the relative root mean square error (RMSE) of three estimators across four d.g.p.'s with correlation $\rho = 0.6$ and different values of T. The RMSE of each estimator in each d.g.p. is computed over 100 repetitions.

Table 4. Monte Carlo Experiments with different values of M and $\rho = 0.6$

	$M = 200$			$M = 500$		
	$\text{2SLS}/\text{Our}_{\text{diff}}$	$\text{2SLS}/\text{Our}_{\text{sieve}}$	$\text{Our}_{\text{diff}}/\text{Our}_{\text{sieve}}$	$\text{2SLS}/\text{Our}_{\text{diff}}$	$\text{2SLS}/\text{Our}_{\text{sieve}}$	$\text{Our}_{\text{diff}}/\text{Our}_{\text{sieve}}$
β	4.9694	5.1751	1.0414	6.2719	8.3841	1.3368
α	7.4727	8.1773	1.0943	9.5783	14.8747	1.5530
σ	3.5961	3.5727	0.9935	7.6558	11.4229	1.4920
	$M = 2000$			$M = 4000$		
	$\text{2SLS}/\text{Our}_{\text{diff}}$	$\text{2SLS}/\text{Our}_{\text{sieve}}$	$\text{Our}_{\text{diff}}/\text{Our}_{\text{sieve}}$	$\text{2SLS}/\text{Our}_{\text{diff}}$	$\text{2SLS}/\text{Our}_{\text{sieve}}$	$\text{Our}_{\text{diff}}/\text{Our}_{\text{sieve}}$
β	14.7125	16.9704	1.1535	42.3295	92.3250	2.1811
α	13.7524	22.2134	1.6152	76.5325	167.2902	2.1859
σ	12.5900	21.5862	1.7146	78.9348	163.9341	2.0768

Table 5 also reports the relative RMSE of three estimators across four d.g.p.'s with $M = 200$ and different values of the correlation ρ. The RMSE of each estimator in each d.g.p. is computed over 100 repetitions. The case of $\rho = 0$ represents the standard situation with only endogenous prices but no endogenous sample selection.

Table 5. Monte Carlo Experiments with M = 200 and different values of ρ

	$\rho = 0$			$\rho = 0.3$		
	2SLS/Our$_{\text{diff}}$	2SLS/Our$_{\text{sieve}}$	Our$_{\text{diff}}$/Our$_{\text{sieve}}$	2SLS/Our$_{\text{diff}}$	2SLS/Our$_{\text{sieve}}$	Our$_{\text{diff}}$/Our$_{\text{sieve}}$
β	0.5697	0.6053	1.0624	1.1002	1.5700	1.4271
α	1.4367	1.4651	1.0198	1.4943	2.4275	1.6245
σ	2.0455	1.5982	0.7813	0.9448	1.1013	1.1657
	$\rho = 0.6$			$\rho = 0.9$		
	2SLS/Our$_{\text{diff}}$	2SLS/Our$_{\text{sieve}}$	Our$_{\text{diff}}$/Our$_{\text{sieve}}$	2SLS/Our$_{\text{diff}}$	2SLS/Our$_{\text{sieve}}$	Our$_{\text{diff}}$/Our$_{\text{sieve}}$
β	4.9694	5.1751	1.0414	15.2161	21.6817	1.4249
α	7.4727	8.1773	1.0943	17.8260	34.1255	1.9144
σ	3.5961	3.5727	0.9935	12.8979	18.7658	1.4550

Figure 1 plots the relative RMSE of the 2SLS with respect to both version of our estimator for different values of the correlation ρ and $M = 200$. For any given value of ρ, we compute the RMSE of each estimator across 100 repetitions and its Euclidean norm across the three demand parameters: that is, $norm(RMSE) = (MSE(\widehat{\alpha}) + MSE(\widehat{\beta}) + MSE(\widehat{\sigma}))^{1/2}$. Then, we plot the ratios between $norm(RMSE)$ of the 2SLS and, respectively, that of Our$_{\text{diff}}$ and Our$_{\text{sieve}}$.

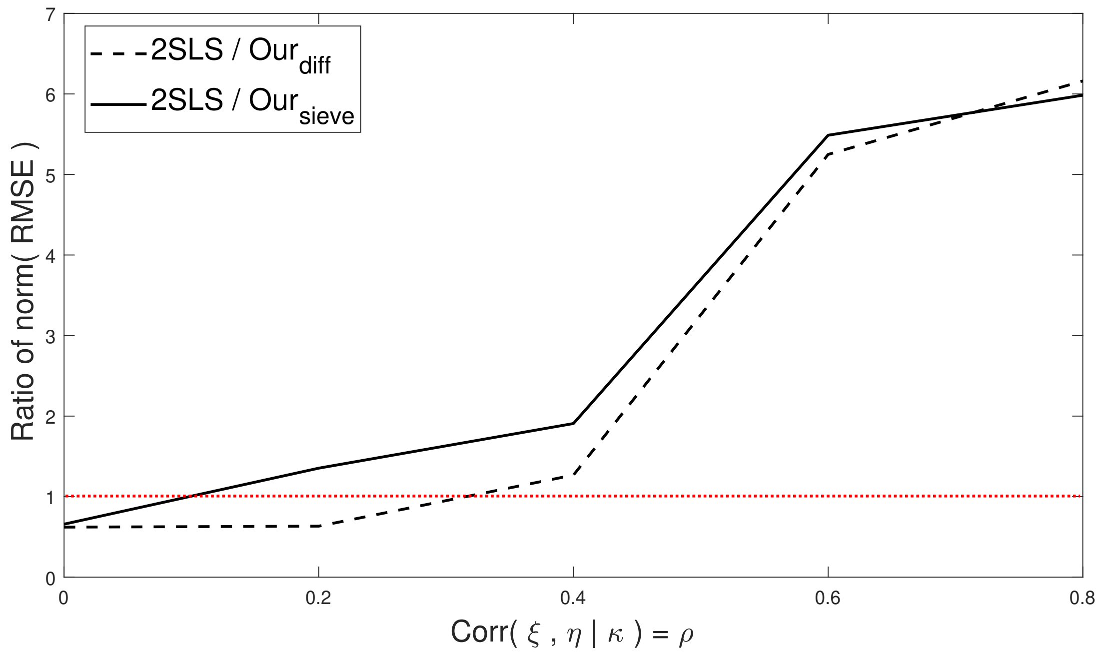
7 Empirical Application

TBW

8 Conclusions

TBW

Figure 1: Relative RMSE of 2SLS w.r.t. Our_{diff} and $\text{Our}_{\text{sieve}}$ for $M = 200$ as a function of ρ



Appendix: Monte Carlo Simulations – DGP

1. We consider $L = 2$ and $J = 3$.

2. We consider a continuous market-level z_t. We take some interval $Z = (z_{min}, z_{max})$ and divide it into a grid of T (i.e., total number of markets) equally spaced points. Call the distance between any two such points $\Delta \equiv \frac{z_{max} - z_{min}}{M-1}$ and then assign each point in this grid to a market t, so that $(z_1, z_2, ..., z_T) = (z_{min}, z_{min} + \Delta, ..., z_{min} + (M-1)\Delta)$. We keep this grid fixed across repetitions of the simulations.

3. The product sold by each firm j has an observable characteristic X_{jt} that varies across markets and is i.i.d. $X_{jt} \sim |N(0, \sigma_X^2)|$, for $j = 1, 2, 3$. When making entry decisions, each firm observes the realization of $X_t = (X_{1t}, X_{2t}, X_{3t})'$ and z_t, such that firms condition on four variables which are observable to the researcher.

4. The demand system is a nested logit with the $J = 3$ products belonging to the same nest g and the outside product $j = 0$ in its own nest:

$$\ln\left(\frac{s_{jt}}{s_{0t}}\right) \;=\; \beta\, X_{jt} + \alpha\, p_{jt} + \sigma\, \ln\left(s_{j|g}\right) + \xi_{jt},$$

where $s_{j|g}$ and s_{jt} are, respectively, the within-nest and the unconditional market shares of product j in market t; s_{0t} the market share of the outside product; p_{jt} the price; ξ_{jt} is an unobservable demand shock, and α, β, and σ are parameters.

5. The unobservable κ_t can take two possible values: $\kappa_t \in \{h, \ell\}$. Its probability distribution does not depend on (X_t, z_t) such that $h(\kappa|X_t, z_t) = h(\kappa)$. Assuming this – or more generally, that $h(\kappa|X_t, z_t)$ is a smooth function of X_t and z_t – allows us using smooth nonparametric methods instead of frequency counts in the estimation of $P_j(X_t, \boldsymbol{z}_t)$.

6. The value of κ_t determines the mean of the distribution of the unobservable demand shocks ξ_{jt}. More specifically, $\{\xi_{jt}|\kappa_t = \ell\} \sim N\left(\overline{\xi}_\ell, \sigma^2_\xi\right)$, and $\{\xi_{jt}|\kappa_t = h\} \sim N\left(\overline{\xi}_h, \sigma^2_\xi\right)$.

7. Marginal costs are i.i.d. across markets and firms with distribution $mc_{jt} \sim |N(0, \sigma^2_{mc})|$. Marginal costs are are determined before entry decisions are made, and they are common knowledge among all firms in the market.

8. Denote the set of products available in market t by $A_t \equiv \{a_{1t}, a_{2t}, a_{3t}\}$, and the set of competing products faced by firm j in market t by $A_{jt} = A_t \setminus \{a_{jt}\}$. For computational tractability, we assume that for given A_{jt}, X_t, and ξ_t, each firm j with $a_{jt} = 1$ chooses prices according to a simplified version of the single-product FOCs:

$$p_j(A_{jt}, X_t, \xi_t) = mc_{jt} - \frac{1-\sigma}{\alpha\left(1 - \sigma\ s^*_{j|g}(A_{jt}, X_t, \xi_t) - (1-\sigma)\ \ s^*_j(A_{jt}, X_t, \xi_t)\right)} \tag{54}$$

where $s^*_{j|g}(A_{jt}, X_t, \xi_t)$ and $s^*_j(A_{jt}, X_t, \xi_t)$ the within-nest and unconditional market shares of j given market structure $A_{jt} \cup \{a_{jt} = 1\}$ and where each firm sets its price equal to its own marginal cost.

9. Let $\Pi_j(A_{jt}, X_t, \xi_t)$ be the profit of firm j if it enters in market t given market structure A_{jt}, X_t, and ξ_t). By definition:

$$\Pi_j(A_{jt}, X_t, \xi_t) = (p_j(A_{jt}, X_t, \xi_t) - mc_{jt})\ \ s_j(A_{jt}, X_t, \xi_t) \tag{55}$$

where $p_j(A_{jt}, X_t, \xi_t)$ is the pricing function described in equation (54) above, and $s_j(A_{jt}, X_t, \xi_t)$ is product j's market share from the nested logit when all firms set their prices according to the pricing equation in (54).

10. Firm j's fixed cost from entering in market t is $\gamma_j z_t + \eta_{jt}$, where η_{jt} is firm's private information and it has a standard normal distribution. More precisely, we assume that

conditional on κ_t, variables (η_{jt}, ξ_{jt}) are i.i.d. across firms and markets jointly normal with correlation ρ:

$$\left(\begin{array}{c} \eta_{jt} \\ \xi_{jt} \end{array} \middle| \kappa_t = \kappa \right) \sim N\left(\left(\begin{array}{c} 0 \\ \overline{\xi}_\kappa \end{array} \right), \left(\begin{array}{cc} 1 & \rho\ \sigma_\xi \\ \rho\ \sigma_\xi & \sigma_\xi^2 \end{array} \right) \right), \tag{56}$$

It then follows that,

$$\xi_{jt} | \left(\kappa_t, \eta_{jt} \right) \sim N\left(\overline{\xi}_{\kappa_t} + \rho \sigma_\xi\ \eta_{jt}, \sigma_\xi^2 \left(1 - \rho^2 \right) \right), \tag{57}$$

from which we obtain,

$$\xi_{jt} | \left(\kappa_t, \eta_{jt} \right) = \overline{\xi}_{\kappa_t} + \rho \sigma_\xi\ \eta_{jt} + e_{jt}, \tag{58}$$

with $e_{jt} \sim N\left(0, \sigma_\xi^2 \left(1 - \rho^2\right)\right)$ and independent of (κ_t, η_{jt}).

11. The expected profit of firm j if it enters in market t given $(\kappa_t, A_{jt}, X_t, z_t, \eta_{jt})$ is (expectation over the distribution of unknown ξ_t):

$$\pi_j \left(\kappa_t, A_{jt}, X_t, \eta_{jt} \right) - \gamma_j\ z_t - \eta_{jt} \tag{59}$$

where $\pi_j \left(\kappa_t, A_{jt}, X_t, \eta_{jt} \right)$ is the expected variable profit function:

$$\pi_j \left(\kappa_t, A_{jt}, X_t, \eta_{jt} \right) \equiv \int_{\xi_t} \Pi_j \left(A_{jt}, X_t, \xi_t \right)\ dF_\xi \left(\xi_t | \kappa_t, \eta_{jt} \right) \tag{60}$$

where function $F_\xi(\xi_t | \kappa_t, \eta_{jt})$ is the normal CDF of $(\xi_t | \kappa_t, \eta_{jt})$. The expected profit associated to non-entry is zero.

12. Having η_{jt} as an argument in the expected profit function introduces a substantial complication in the computation of best response probabilities in the entry game. More specifically, computing best response probabilities requires the numerical inversion of the expected profit with respect to η_{jt}. This best response probability does not have a closed form expression. Note that this is not an issue for our estimation method, which can easily deal with η_{jt} as an argument in the expected profit. The computational difficulty is in the solution for an equilibrium of the model and thus the generation of the simulated data.

    To avoid this computational difficulty, in our Monte Carlo experiment we consider a DGP in which firms ignore the correlation between between ξ_{jt} and η_{jt} when calculating expected profit. That is, the expected profit function is $\pi_j \left(\kappa_t, A_{jt}, X_t \right)$, which does not depend on η_{jt} and has very similar definition as in equation (60), but where the integral is over the distribution $F_\xi\left(\xi_t|\,\kappa_t\right)$ instead of $F_\xi\left(\xi_t|\,\kappa_t, \eta_{jt}\right)$. That is:

$$\pi_j\left(\kappa_t, A_{jt}, X_t\right) \equiv \int_{\xi_t} \Pi_j\left(A_{jt}, X_t, \xi_t\right)\ dF_\xi\left(\xi_t|\,\kappa_t\right) \tag{61}$$

13. Note that the expected variable profit function $\pi_j\left(\kappa_t, A_{jt}, X_t\right)$ is conditional on the entry decisions of the other firms, as represented by A_{jt}. These competitors' decisions are unknown to firm j at the moment of its own entry decision. Therefore, this "conditional on market structure" expected profit function is not the expected profit that firm j uses to make its entry decision. To calculate the expected profit we need to integrate over the distribution of competitors' probabilities of entry.

   Let $\pi_j^e\left(\kappa_t, X_t, z_t\right)$ represents firm j's expected variable profit function. That is:

$$\pi_j^e\left(\kappa_t, X_t, z_t\right) = \sum_{A_{-jt}} \pi_j\left(\kappa_t, A_{jt}, X_t\right)\ \ \Pr\left(A_{jt}|\kappa_t, X_t, z_t\right) \tag{62}$$

   where $\Pr\left(A_{jt}|\kappa_t, X_t, z_t\right) = \Pr\left(a_{kt}|\kappa_t, X_t, z_t\right) \times \Pr\left[a_{rt}|\kappa_t, X_t, z_t\right)$ is the conditional probability that the other two firms, k and r, take entry decision $(a_{kt}, a_{rt}) \in \{(0,0),(1,0),(0,1),(1,1)\}$.

13. For a given value of (κ_t, X_t, z_t), let $P_{1t} \equiv P_1\left(\kappa_t, X_t, z_t\right)$, $P_{2t} \equiv P_2\left(\kappa_t, X_t, z_t\right)$, and $P_{3t} \equiv P_3\left(\kappa_t, X_t, z_t\right)$ be the equilibrium probabilities in the market entry game. These three probabilities are a solution to the following system of equations:

$$\begin{aligned}
P_{1t} &= \Phi\left(-\gamma_1\ z_t + \sum_{(a_2,a_3)\in\{0,1\}^2} \pi_1\left(\kappa_t, a_2, a_3, X_t\right) \prod_{v=2,3} \left[P_{vt}\right]^{a_v}\left[1-P_{vt}\right]^{(1-a_v)}\right) \\
P_{2t} &= \Phi\left(-\gamma_2\ z_t + \sum_{(a_1,a_3)\in\{0,1\}^2} \pi_2\left(\kappa_t, a_1, a_3, X_t\right) \prod_{v=1,3} \left[P_{vt}\right]^{a_v}\left[1-P_{vt}\right]^{(1-a_v)}\right) \\
P_{3t} &= \Phi\left(-\gamma_3\ z_t + \sum_{(a_1,a_2)\in\{0,1\}^2} \pi_3\left(\kappa_t, a_1, a_2, X_t\right) \prod_{v=1,2} \left[P_{vt}\right]^{a_v}\left[1-P_{vt}\right]^{(1-a_v)}\right)
\end{aligned} \tag{63}$$

   where $\Phi\left(\cdot\right)$ denotes the standard normal CDF.

14. **Computation of equilibrium probabilities of market entry**. For each value of (κ_t, X_t, z_t), we proceed as follows.

(i) For each market t and conditional on a value for κ_t, we simulate 500 random draws of the three random variables $\xi_{jt}|\kappa_t$, $j = 1, 2, 3$, from the normal distribution $N\left(\overline{\xi}_{\kappa_t}, \sigma_\xi^2\right)$.

(ii) For given draw ξ_t, firm j, and market structure A_{jt}, we compute price $p_j\left(A_{jt}, X_t, \xi_t\right)$ as if firm j were to enter in t and according to equation (54). We compute this price for every possible market structure $A_{jt} \in \{0,1\}^2$. We do the same for every firm j. Then, we use these prices and the corresponding market shares in the nested logit demand system to obtain the profit of firm j, $\Pi_j\left(A_{jt}, X_t, \xi_t\right)$, using equation (55). We repeat the same procedure for all the the 500 draws of ξ_t from step (i).

(iii) For each given firm j and for a given value of (κ_t, X_t, z_t) and a given value of market structure A_{jt}, we compute expected profit $\pi_j\left(\kappa_t, A_{jt}, X_t\right)$ by averaging the profit $\Pi_j\left(A_{jt}, X_t, \xi_t\right)$ (obtained in step (ii)) over the 500 random draws of ξ_t generated in step (i).

(iv) We repeat steps (ii) and (iii) for each firm $j = 1, 2, 3$ and each of the four values of market structure $A_{jt} \in \{(0,0), (1,0), (0,1), (1,1)\}$.

(v) Given the expected profits calculated in step (iv), we construct the system of equations in (63). We numerically solve this system of three equations with three unknowns to obtain the equilibrium CCPs $P_1\left(\kappa_t, X_t, z_t\right)$, $P_2\left(\kappa_t, X_t, z_t\right)$, and $P_3\left(\kappa_t, X_t, z_t\right)$.

We repeat previous steps (i) to (v) for every value of (κ_t, X_t, z_t).

15. **Simulating data for the endogenous variables in the model where ξ_{jt} and η_{jt} are independently distributed**. For each market t, we need to simulate data for the endogenous variables $\{a_{jt}, p_{jt}, s_{jt} : j = 1, 2, 3\}$.

For the simulation of variables ξ_{jt} for the firms who have decided to enter in the market, it is important taking into account that correlation between ξ_{jt} and η_{jt} and endogenous entry introduce censoring in the distribution of ξ_{jt}. Conditional on κ_t, variables ξ_{jt} and η_{jt} are jointly normal, such that $\mathbb{E}\left[\xi_{jt}|\kappa_t, \eta_{jt}\right] = \overline{\xi}_{\kappa_t} + \rho\sigma_\xi\ \eta_{jt}$. This joint normality assumption also implies:

$$\xi_{jt}|\left(\kappa_t, X_t, z_t, a_{jt} = 1\right) = -\rho\ \sigma_\xi\ \frac{\phi\left(\Phi^{-1}\left(P_j\left(\kappa_t, X_t, z_t\right)\right)\right)}{P_j\left(\kappa_t, X_t, z_t\right)} + \overline{\xi}_{\kappa_t} + e_{jt} \tag{64}$$

where $\phi(\cdot)$ is the density function and $\Phi^{-1}(\cdot)$ is the inverse CDF (or Quantile function) for the standard normal, and $e_{jt} \sim N\left(0, \sigma_\xi^2\left(1 - \rho^2\right)\right)$.

For a given value of the exogenous variables (X_t, z_t) we proceed as follows.

(i) For each market t, we use the probability $h(\kappa)$ to draw a value of κ_t.

(ii) Given the equilibrium probability of market entry $P_j(\kappa_t, X_t, z_t)$, we draw a value of a_{jt}. We do this for every firm $j = 1, 2, 3$. This generates the realized market structure $A_t = (a_{1t}, a_{2t}, a_{3t})$.

(iii) We take a random draw of variables ξ_{1t}, ξ_{2t}, and ξ_{3t} from the censored normal distribution described in equation (64).

(iv) Given X_t, the realized market structure A_t from step (ii), and the vector ξ_t from step (iii), we use the pricing equation in (54) to generate equilibrium prices (p_{1t}, p_{2t}, p_{3t}).

(v) Finally, given X_t, the realized market structure A_t from step (ii), the vector ξ_t from step (iii), and the vector of equilibrium prices from step (iv), we use the nested logit demand system to calculate the equilibrium market shares (s_{1t}, s_{2t}, s_{3t}).

Table 1. Values of Parameters in d.g.p.

Parameter	True Value	Parameter	True Value
$\beta =$	2.0	$\sigma_\xi =$	2.0
$\alpha =$	-2.0	$\rho =$	1.999
$\sigma =$	0.6		
		$\sigma_X = \sigma_{mc} =$	0.3
$\overline{\xi}_\ell =$	-4.0	$z_{\min} =$	0.1
$\overline{\xi}_h =$	4.0	$z_{\max} =$	2.0
$h(\ell) =$	0.6	$\gamma_1 = \gamma_2 = \gamma_3 =$	1.0

References

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.

Aguirregabiria, V., and P. Mira (2007): "Sequential estimation of dynamic discrete games," *Econometrica*, 75(1), 1–53.

——— (2019): "Identification of games of incomplete information with multiple equilibria and unobserved heterogeneity," *Quantitative Economics*, 10(4), 1659–1701.

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.

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.

Amemiya, T. (1973): "Regression analysis when the dependent variable is truncated normal," *Econometrica: Journal of the Econometric Society*, pp. 997–1016.

——— (1974): "Multivariate regression and simultaneous equation models when the dependent variables are truncated normal," *Econometrica*, pp. 999–1012.

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.

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.

Berry, S., M. Carnall, and P. T. Spiller (2006): "Airline hubs: costs, markups and the implications of customer heterogeneity," *Competition Policy and Antitrust*, pp. 183–213.

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

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

Berry, S. T. (1992): "Estimation of a Model of Entry in the Airline Industry," *Econometrica: Journal of the Econometric Society*, pp. 889–917.

——— (1994): "Estimating discrete-choice models of product differentiation," *The RAND Journal of Economics*, pp. 242–262.

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.

Bresnahan, T. F., and P. C. Reiss (1990): "Entry in monopoly market," *The Review of Economic Studies*, 57(4), 531–553.

——— (1991): "Entry and competition in concentrated markets," *Journal of Political Economy*, 99(5), 977–1009.

Ciliberto, F., C. Murry, and E. Tamer (2021): "Market structure and competition in airline markets," *Journal of Political Economy*, 129(11), 2995–3038.

Ciliberto, F., and E. Tamer (2009): "Market structure and multiple equilibria in airline markets," *Econometrica*, 77(6), 1791–1828.

Das, M., W. K. Newey, and F. Vella (2003): "Nonparametric estimation of sample selection models," *The Review of Economic Studies*, 70(1), 33–58.

Deaton, A., and J. Muellbauer (1980): "An almost ideal demand system," *The American economic review*, 70(3), 312–326.

Dubé, J.-P., A. Hortaçsu, and J. Joo (2021): "Random-coefficients logit demand estimation with zero-valued market shares," *Marketing Science*, 40(4), 637–660.

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

Gandhi, A., Z. Lu, and X. Shi (2022): "Estimating demand for differentiated products with zeroes in market share data," *Quantitative Economics*.

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.

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

Hall, P., and X.-H. Zhou (2003): "Nonparametric estimation of component distributions in a multivariate mixture," *The annals of statistics*, 31(1), 201–224.

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," 5(4), 475–492.

Hirano, K., G. W. Imbens, and G. Ridder (2003): "Efficient estimation of average treatment effects using the estimated propensity score," *Econometrica*, 71(4), 1161–1189.

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.

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

Li, S., J. Mazur, Y. Park, J. Roberts, A. Sweeting, and J. Zhang (2022): "Repositioning and market power after airline mergers," *The RAND Journal of Economics*, 53(1), 166–199.

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

Newey, W. K., J. L. Powell, and J. R. Walker (1990): "Semiparametric estimation of selection models: some empirical results," *The American Economic Review*, 80(2), 324–328.

Pakes, A., M. Ostrovsky, and S. Berry (2007): "Simple estimators for the parameters of discrete dynamic games (with entry/exit examples)," *The RAND Journal of Economics*, 38(2), 373–399.

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

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

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

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

Sweeting, A. (2009): "The strategic timing incentives of commercial radio stations: An empirical analysis using multiple equilibria," *The RAND Journal of Economics*, 40(4), 710–742.

——— (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.

Tobin, J. (1958): "Estimation of relationships for limited dependent variables," *Econometrica*, pp. 24–36.

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

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.

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.