Identification and Estimation of Demand Models with Endogenous Product Entry and Exit*

Victor Aguirregabiria †
University of Toronto
CEPR

Alessandro Iaria ‡
University of Bristol
CEPR

Senay Sokullu §
University of Bristol

May 24, 2024

Abstract

This paper addresses the endogeneity of firms' entry and exit decisions in estimating demand for differentiated products. Under standard conditions, the selection propensity score is insufficient to control selection bias, leading to the inconsistency of conventional methods of handling selection. We introduce a novel, straightforward two-step approach to estimate demand while accounting for endogenous product entry and exit. In the first step, our method estimates a nonparametric finite mixture model of product entry and exit that accommodates latent market types. Assuming a finite-dimensional support set for the latent variable does not introduce misspecification bias in the product entry model or the corresponding selection term used in demand estimation. In the second step, our method estimates the demand parameters while controlling for selection by incorporating product entry probabilities for each latent market type. We apply this approach to data from the airline industry, revealing that conventional methods to address selection bias underestimate demand price elasticities.

Keywords: Demand for differentiated product; Endogenous product availability; Selection bias; Market entry and exit; Multiple equilibria; Identification; Estimation; Demand for airlines.

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

*We are grateful for helpful comments from Andreea Enache, Christos Genakos, Michele Fioretti, Philip Haile, Mathieu Marcoux, Anders Munk-Nielsen, David Pacini, Bertel Schjerning, Philipp Schmidt-Dengler, Yuanyuan Wan, Ao Wang, Christine Zulehner; from seminar participants at the Universities of Bolzano, Cambridge (Judge Business School), Copenhagen, Glasgow, Helsinki, Mannheim, Penn State, Rochester, Sciences Po, Vienna; and workshop participants at the IAAE (2023), MaCCI Summer Institute (2022), and Midwest Econometrics Group (2023).

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

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

§School of Economics, University of Bristol. The Priory Road Complex, Priory Road, BS8 1TU, Bristol, UK. senay.sokullu@bristol.ac.uk.

1 Introduction

Since the influential work by Tobin (1958), Amemiya (1973), and Heckman (1976), addressing endogenous selection has become a fundamental topic in microeconometrics. The need to handle censored observations, mainly zeros, in consumer demand estimation has made this economic application a significant driver in developing methods to account for sample selection. While much of the early literature focused on analyzing demand for a single product, there have also been early applications to demand systems (Amemiya, 1974, Yen, 2005, Yen and Lin, 2006).

More recently, the selection problem in estimating demand systems has attracted researchers' attention in the context of structural models of demand and supply for differentiated products. A key aspect distinguishing the recently proposed methods is the source of the zeros in market shares. The first group of studies considers that the choice probabilities in the model are strictly positive, but observed market shares may be zero because of sampling error due to a small number of consumers in the data (Gandhi, Lu, and Shi, 2023). A second approach assumes that some market shares are zero because consumers in certain markets exclude these products from their choice set or *consideration set* (Dubé, Hortaçsu, and Joo, 2021).[1] Finally, a third group of studies examines sample selection bias in demand estimation when zeros arise from firms' market entry decisions (Conlon and Mortimer, 2013, Ciliberto, Murry, and Tamer, 2021; Li, Mazur, Park, Roberts, Sweeting, and Zhang, 2022).

This paper studies the estimation of demand for differentiated products using market-level data while accounting for censoring or selection resulting from firms not offering certain products in specific markets or periods. Demand estimation typically relies on data collected from multiple geographic markets and periods, where it is common for certain products to be unavailable in specific markets or periods. When firms make their market entry decisions, they possess information about the demand for their products, particularly regarding demand components that are not observable to the researcher. Firms are more inclined to enter markets with higher expected demand. Failure to consider this selection process can introduce significant biases in estimating demand parameters. This issue arises across various demand applications and industries, such as the demand for airlines (Berry, Carnall, and Spiller, 2006; Berry and Jia, 2010; Aguirregabiria and Ho, 2012), supermarket chains (Smith, 2004), radio stations (Sweeting, 2013),

[1] There is a growing empirical literature on consideration sets, where typically all products are assumed to be available in the market, but consumers consider subsets of these due to inattention or costly search. These heterogeneous consideration sets are usually unobserved by the researcher. The estimators proposed to overcome this complication do not consider the endogenous selection problem addressed in this paper (Goeree, 2008; Abaluck and Adams-Prassl, 2021; Barseghyan, Coughlin, Molinari, and Teitelbaum, 2021; Lu, 2022; and Moraga-González, Sándor, and Wildenbeest, 2023). For more details on this literature, see Crawford, Griffith, and Iaria (2021).

personal computers (Eizenberg, 2014), or ice cream (Draganska, Mazzeo, and Seim, 2009).[2]

The selection problem in this demand-entry model exhibits a key characteristic that distinguishes it from more standard cases. The multi-dimensional nature of demand unobservables and their non-additive impact on firms' entry decisions prevent the model from meeting a crucial monotonicity condition. This condition is necessary for the selection propensity score (i.e., the product entry probabilities) to serve as a sufficient statistic to control for selection in the estimation of demand parameters (see Proposition 2 and 3 in Angrist, 1997). Moreover, the model exhibits multiple equilibria in the entry and pricing games. Different equilibria may be selected across markets, adding another potential source of non-monotonicity to the selection equation. Consequently, the conventional identification results and two-step estimation methods found in the literature are not applicable (Newey, Powell, and Walker, 1990; Ahn and Powell, 1993; Powell, 2001; Das, Newey, and Vella, 2003; Aradillas-Lopez, Honoré, and Powell, 2007; Newey, 2009).

This paper studies the identification of demand parameters in a structural model of demand, product entry, and price competition where the distribution of demand unobservables is nonparametric. On the supply side, we consider a two-stage game in which firms first decide whether to sell their products and then the prices to charge for the available products. The first stage game of product entry is general, allowing for different information structures, static or dynamic behavior, multiplicity of equilibria, and nonparametric distribution of all unobservables. The second stage price competition game is a standard Bertrand game of complete information as in Berry, Levinsohn, and Pakes (1995).

The main contribution of this paper is to establish the sequential (two-step) identification of demand parameters in the presence of endogenous selection and price endogeneity. A key element in our result is representing firms' product entry decisions using a nonparametric finite mixture model. Drawing on recent findings in tensor decomposition (Kargas, Sidiropoulos, and Fu, 2018), we demonstrate that this representation is general and applies broadly to our structural model of demand and product entry/exit decisions. This general representation holds under very general

[2] Recent research on nonparametric identification of demand systems accounts for product entry and exit as a source of identifying variation in consumer choice sets. However, these studies focus on the exogenous component of such variation and do not examine the endogenous selection problem tied to product entry and exit. See the work of Berry and Haile, (2014, 2022), as well as the recent survey papers by Berry and Haile, (2021) and Gandhi and Nevo, (2021). One common approach to tackle this selection problem is to incorporate fixed effects, such as product, market, and time fixed effects while assuming that the remaining portion of the error term in the demand equation is unknown to firms when they make their market/product entry decisions. This approach is employed in studies by Aguirregabiria and Ho (2012), Sweeting (2013), and Eizenberg (2014). Although the fixed effects approach is convenient in practice, it relies on assumptions regarding firms' information that may not be realistic in certain empirical applications. Furthermore, these assumptions can be subject to testing and potential rejection by the data. Notably, the model considered in this paper encompasses the fixed effects model as a specific case.

conditions, including nonparametric distributions of unobservables with continuous support and varying information structures in the entry/exit game, whether complete or incomplete.

The mixture model's unobservable component, which we refer to as the unobserved market type, encapsulates all the common knowledge that firms possess at the moment of entry regarding demand unobservables. Despite the potential multidimensionality and continuous support of this unobservable, Kargas, Sidiropoulos, and Fu (2018) show that using finite-dimensional support for this latent variable — specifically, with at most 2^{J-1} points of support where J is the number of products — does not misspecify the probabilities of product entry/exit. Building on this insight, we show that the selection term in the demand equation can be represented as a convolution of the conditional choice probabilities given the unobserved market type and the distribution of unobserved market types. Applying results from the nonparametric finite mixtures literature, we demonstrate that data on firms' product entry decisions nonparametrically identify these conditional choice probabilities and the distribution of unobserved market types. Given these identified objects, we establish the identification of demand parameters and the function that controls for endogenous selection. Our proof of sequential identification addresses two significant challenges in nonparametric finite mixture models: identification up to label swapping and establishing a lower bound on the number of unobserved market types.

Building on our constructive proof of identification, we propose a simple two-step estimator to address endogenous selection and price endogeneity in demand estimation. In the first step, we use a semiparametric finite mixture model to estimate the conditional choice probabilities of product entry and the distribution of unobserved market types. By directly targeting the nonparametric finite mixture representation by Kargas, Sidiropoulos, and Fu (2018), our first step estimator does not require the researcher to know or to assume the details of the specific model of product entry that generated the data, and it is still consistent with a large class of such games. In the second step, we estimate demand parameters using a Generalized Method of Moments (GMM) approach that controls for both endogenous product availability and price endogeneity.

The paper illustrates the proposed method with data from the airline industry. Our findings highlight the importance of accounting for endogenous product entry and a finite mixture of unobserved market types, as failing to do so can lead to significant biases. Specifically, neglecting latent classes or unobserved market types while accounting for endogenous selection results in meaningful attenuation biases in the estimates of demand elasticities.

Our paper is motivated by the work of Draganska, Mazzeo, and Seim (2009), Ciliberto, Murry, and Tamer (2021), and Li, Mazur, Park, Roberts, Sweeting, and Zhang (2022). These authors develop methods for estimating structural models that combine the demand for differentiated products in Berry, Levinsohn, and Pakes (1995) with games of market/product entry as in Bresnahan and Reiss (1990, 1991) and Berry (1992). They focus on the joint estimation of all structural parameters in the model, including demand, marginal costs, entry costs, and the probability distribution of unobservable factors. To jointly estimate the full model, these authors employ nested fixed point algorithms, which require solving multiple times for the equilibria of a two-step game. Consequently, they rely on strong parametric assumptions for all the structural functions and the distribution of unobservables. In contrast, we adopt a sequential approach to identify and estimate the structural parameters of the demand model. Our approach yields identification in a nonparametric specification of the model that guarantees a supply-side structure consistent with equilibrium behavior. We do not impose strong assumptions on marginal costs, entry costs, or the distribution of unobservables. Furthermore, our estimation method offers computational simplicity as it does not require the computation of equilibria. Finally, the method and its computational advantages apply both to static games of market entry and dynamic games of market entry and exit.[3]

We contribute to the literature on structural models of market entry and exit with the insight that the nonparametric finite mixture representation by Kargas, Sidiropoulos, and Fu (2018) extends the applicability of the econometric techniques developed for games with incomplete information (as the static games in Seim (2006); Draganska, Mazzeo, and Seim (2009); Xiao (2018); Aguirregabiria and Mira (2019) or the dynamic games in Aguirregabiria and Mira (2007); Pakes, Ostrovsky, and Berry (2007); Sweeting (2009)) to a larger class of entry and exit games.

Our estimation method contributes to the literature on the semiparametric estimation of sample selection models, with seminal contributions by Newey, Powell, and Walker (1990); Powell (2001); Newey (2009). A distinctive aspect of our method is the multi-dimensionality of the selection term in the second step of estimation. Specifically, the selection term involves the convolution of functions representing market entry probabilities for each latent market type. Our method builds on the semiparametric sieve method in Das, Newey, and Vella (2003); Newey (2009) and the pairwise-differencing approach in Powell (2001); Aradillas-Lopez, Honoré, and Powell (2007). We enhance the applicability of this method by broadening its scope in the context of our model.

[3]Given estimates of demand parameters and unobservables from our method, one can obtain estimates of marginal costs and entry costs with less stringent parametric assumptions than those required for the joint estimation of the full structural model. Similar to Ciliberto, Murry, and Tamer (2021) and Li, Mazur, Park, Roberts, Sweeting, and Zhang (2022), we can rely on the estimates of our model to implement a wide range of counterfactual experiments accounting for the endogeneity of product entry and exit. Accounting for endogenous product entry/exit is particularly valuable when simulating the effects of a merger, as demonstrated by Li, Mazur, Park, Roberts, Sweeting, and Zhang (2022). In section 6.4, we discuss the implementation of these counterfactuals.

Our method also relates to the literature on estimating Conditional Average Treatment Effects (CATE) using finite mixture models for the propensity score (Haviland and Nagin, 2005; Haviland, Nagin, Rosenbaum, and Tremblay, 2008; Lanza, Coffman, and Xu, 2013). In the first step of our method, we follow a similar approach with the only difference played by the conditional independence of firms' entry decisions (which is not assumed but implied by the representation in Kargas, Sidiropoulos, and Fu (2018)) to obtain nonparametric identification. However, the second step of our method differs substantially. Previous studies in the latent propensity score literature assign each individual in the sample to a latent class using techniques such as the highest posterior probability (modal assignment) and treating the assigned class as an observable variable. In contrast, our method recognizes that an individual's (firm's) latent class remains unobservable, even after estimating the nonparametric finite mixture model using an infinite sample. Consequently, controlling for selection bias in the second step requires the inclusion of propensity scores for all potential latent types.

The remainder of this paper is structured as follows. Section 2 introduces our model and underlying assumptions. Section 3 deals with the selection problem within this framework. Our identification results are outlined in Section 4. In Section 5, we detail our estimation methodology, followed by an empirical application to the US airline industry in Section 6. Finally, Section 7 provides a summary and concluding remarks.

2 Model

The demand system follows the BLP framework (Berry, Levinsohn, and Pakes, 1995). For the sake of notational simplicity, we focus on single-product firms. In section 2.4, we discuss how to adapt our model and methodology to the case of multi-product firms. There are J firms indexed by $j \in \mathcal{J} = \{1, 2, ..., J\}$ and T markets indexed by $t \in \{1, 2, ..., T\}$, where a market can be a geographic location, a period, or a combination of both. Consumers living in a market t can buy only the products available in that market. 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 or not to be active 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.[4] Demand and price

[4]While this assumption is standard in the literature on empirical industrial organization, there are important exceptions of structural models of entry which allow potential entrants to internalize network externalities across markets, as Bontemps, Gualdani, and Remmy (2023); Jia (2008); Aguirregabiria and Ho (2012). However, these structural models of network formation do not consider the endogenous sample selection problem we study in this paper.

competition are static. Our model accommodates static and dynamic games of firms' product entry (and exit) decisions.

2.1 Demand

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 I 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 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 the product that maximizes its utility. Let s_{jt} be the market share of product j in market t, i.e., the proportion of households choosing product j:

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

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

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

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

where $\boldsymbol{s}_t^{(\boldsymbol{a})}$ is the subvector of $\boldsymbol{s}_t$ containing the market shares for only those products available. A similar definition applies to the subvector $\boldsymbol{\delta}_t^{(\boldsymbol{a})}$. Lemma 1 establishes that the invertibility property in Berry (1994) applies to the demand system (4) for any 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 $\boldsymbol{s}_t^{(\boldsymbol{a})} = \boldsymbol{d}_t^{(\boldsymbol{a})}(\boldsymbol{\delta}_t^{(\boldsymbol{a})})$ is invertible in $\boldsymbol{\delta}_t^{(\boldsymbol{a})}$ such that for every product in this subsystem (i.e., for every product with $a_{jt} = 1$) the inverse function $\delta_{jt}^{(\boldsymbol{a})} = d_{jt}^{-1}(\boldsymbol{s}_t^{(\boldsymbol{a})})$ 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:

$$d_{jt}^{-1}(\boldsymbol{s}_t^{(\boldsymbol{a}_t)}) \;=\; \delta_{jt} \;=\; \alpha\, p_{jt} + \boldsymbol{x}_{jt}'\, \boldsymbol{\beta} + \xi_{jt} \quad \text{if and only if} \quad a_{jt} = 1. \tag{5}$$

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 characterization of the selection term is an 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 the AIDS, each value of the vector $\boldsymbol{a}_t$ implies a different set of regressors and slope parameters in the regression equation that relates the demand of product j to the log-prices of the available products. Therefore, in the AIDS model, the selection bias within the demand equation for product j does not depend solely on the availability of that particular product but rather on the availability profile of all products within the system. In other words, the selection term cannot be represented in terms of $\mathbb{E}\left(\xi_{jt} \mid a_{jt} = 1\right)$ but must instead be expressed in terms of $\mathbb{E}\left(\xi_{jt} \mid a_{jt} = 1,\ \boldsymbol{a}_{-jt} = \boldsymbol{a}_{-j}\right)$. Consequently, in the AIDS model, we have a different selection term for each value of the vector $\boldsymbol{a}_{-j}$ representing the availability of products other than j. This structure makes the selection problem multi-dimensional and significantly complicates identification and estimation when the number of products J is large.

As discussed in section 2.4, Lemma 1 is unaffected in the case of multi-product firms and so is the structure of the resulting selection term, which can still be represented in terms of

$\mathbb{E}(\xi_{jt} \mid a_{jt} = 1)$ even if firm j owns other products. The following 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 denote by r_j 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 variables v and ε are independently distributed, ε and $(1-\sigma)\, v + \varepsilon$ are i.i.d. type I extreme value, and $\sigma \in [0,1]$ is a parameter (Cardell, 1997). 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

Let Π_{jt} be the profit of firm j if active in market t. This profit 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 unobserved to the researcher.

Given firms' entry decisions, 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 restrict equilibrium selection and allow each market to select its 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 exogenous variables observed by the researcher affecting demand or costs, with support $\mathcal{X}$ (in which each element can be continuous or discrete). Vectors $\boldsymbol{\xi}_t$ and $\boldsymbol{\omega}_t$ have similar definitions. Let $\boldsymbol{a}_{-jt}$ be the vector with the entry decisions of every firm other than j. We use $VP_j(\boldsymbol{a}_{-jt}, \boldsymbol{x}_t, \boldsymbol{\xi}_t, \boldsymbol{\omega}_t, \tau_t^2)$ 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 $(a_{jt} = 1, \boldsymbol{a}_{-jt}, \boldsymbol{x}_t, \boldsymbol{\xi}_t, \boldsymbol{\omega}_t, \tau_t^2)$.

2.3 Market entry game

This section presents a game of product entry consistent with a broad class of models in the literature. It includes complete information games, as in Ciliberto and Tamer (2009) and Ciliberto, Murry, and Tamer (2021), as well as incomplete information games with common knowledge unobservables, as in Grieco (2014) and Aguirregabiria and Mira (2019). Our model also accommodates a flexible structure regarding firms' information about demand unobservables at the time of entry, covering extreme cases where firms have no uncertainty or no signal about these variables or any intermediate scenario. Our identification results apply to this broad spectrum of market entry games.

Firms' entry decisions are determined as an equilibrium of a market entry game. The profit of being inactive is normalized to zero. Before entry, firms may have uncertainty about their profits when active in the market. Their information about demand and costs plays a key role in their entry decisions and, therefore, on the implied joint probabilities of entry.

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* $(\boldsymbol{\xi}_t, \boldsymbol{\omega}_t, \tau_t^2)$ *which is common knowledge for the firms and its probability distribution conditional on* $\boldsymbol{x}_t$ *is* $f_\kappa(\kappa_t \mid \boldsymbol{x}_t)$*. As a possible* *scenario, the signal κ_t might encompass the entire vector $(\boldsymbol{\xi}_t, \boldsymbol{\omega}_t, \tau_t^2)$, ensuring that firms face no uncertainty regarding demand and variable costs at the moment of entry.*

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

C. *Variable η_{jt} is a signal that is private information of firm j, independently distributed over firms, and independent of $(\kappa_t, \boldsymbol{x}_t)$ with CDF F_η. As a possible scenario, variable η_{jt} can have a degenerate probability distribution such that the entry game is of complete information.*

D. *Vector $(\boldsymbol{\xi}_t, \boldsymbol{\omega}_t, \tau_t^2, \kappa_t, \tau_t^1, \eta_{jt})$ is unobserved to the researcher. Variables $(\boldsymbol{\xi}_t, \boldsymbol{\omega}_t)$ are independent of $\boldsymbol{x}_t$.* ■

For notational simplicity, we omit τ_t^1 and interpret κ_t as representing both equilibrium selection and payoff relevant variables. Similarly, with some abuse of notation, for the rest of the paper we represent the vector of unobservables $(\boldsymbol{\xi}_t, \boldsymbol{\omega}_t, \tau_t^2)$ using the more compact notation $\boldsymbol{\xi}_t$.

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

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

where $F_{j,\xi}\left(\boldsymbol{\xi}_t \mid \kappa_t, \eta_{jt}\right)$ is a CDF and represents firm j's beliefs about the distribution of $\boldsymbol{\xi}_t$ conditional on (κ_t, η_{jt}). As $F_{j,\xi}\left(\boldsymbol{\xi}_t \mid \kappa_t, \eta_{jt}\right)$ is j-specific, the same market-level signal κ_t can affect the beliefs about $\boldsymbol{\xi}_t$ of different firms in different ways. Function $fc(\boldsymbol{x}_{jt}, \kappa_t, \eta_{jt})$ represents the fixed cost and entry cost of operating in the market.

Assumption 1(C) states that this entry game can accommodate complete information if the distribution of each η_{jt} is degenerate; otherwise, it is a game of incomplete information. Below, we describe an equilibrium of the game as a Bayesian Nash Equilibrium (BNE). However, it is essential to note that this solution concept encompasses a complete information Nash Equilibrium (NE) when each η_{jt} has a degenerate probability distribution.

Given $(\boldsymbol{x}_t, \kappa_t)$, a Bayesian Nash Equilibrium (BNE) of this game can be represented as a 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.

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

Firm j's best response is to enter the market if and only if this expected profit exceeds zero. Considering this, we can define a BNE in this game as follows.

DEFINITION 1 (BNE). *Under Assumption 1 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:*

$$P_{jt} = \int 1\{\pi_j^P(\boldsymbol{x}_t, \kappa_t, \eta_{jt}) \geq 0\} \, dF_\eta(\eta_{jt}) . \qquad \blacksquare \tag{14}$$

This framework accommodates various information structures corresponding to different scenarios considered by the literature on structural market entry models. When $\text{Var}(\kappa_t) = 0$, the entry game only features private information unobservables, as examined in studies such as Seim (2006), Sweeting (2009), and Bajari, Hong, Krainer, and Nekipelov (2010). As $\text{Var}(\eta_{jt}) = 0$, the entry game is of complete information, as in the work by Ciliberto and Tamer (2009) and Ciliberto, Murry, and Tamer (2021). In instances where $\text{Var}(\kappa_t) > 0$ and $\text{Var}(\eta_{jt}) > 0$, the model describes an entry game including both categories of unobservable factors, as in work by Grieco (2014) and Aguirregabiria and Mira (2019).

2.4 Multi-product firms

We briefly discuss how the proposed model, the results above, and the characterization of the selection problem in section 3 below can be extended to the case of multi-product firms. We still use $j \in \mathcal{J}$ to index products, but now we introduce the firm sub-index f and define $\mathcal{J}_f \subseteq \mathcal{J}$ as the set of products owned by firm f. The product entry decisions of firm f are described by vector $\boldsymbol{a}_{ft} \equiv (a_{jt} : j \in \mathcal{J}_f) \in \{0, 1\}^{|\mathcal{J}_f|}$.

First, note that Lemma 1's applicability remains unaffected by the product ownership structure. This Lemma only relies on the structure of the demand system. Therefore, regardless of the product ownership structure, the selection problem in the estimation of the demand of product j is still described in terms of the conditional expectation $\mathbb{E}(\xi_{jt} \mid a_{jt} = 1)$.

Second, Assumption 1, which describes a firm's information at the time of its entry decisions into market t, remains unchanged. The only difference is that we need to represent a firm's private information using a vector with as many elements as the products owned by this firm; that is, $\boldsymbol{\eta}_{ft} \equiv (\eta_{ft}(\boldsymbol{a}_f) : \boldsymbol{a}_f \in \{0, 1\}^{|\mathcal{J}_f|})$. For instance, in the case of a two-product firm, $\eta_{ft}(1, 0)$ is the latent component of entry cost when the firm offers product 1 while excluding product 2. Under Assumption 1, equation (12), describing the expected profit of a firm, readily extends to multi-product firms as follows:

$$\pi_f\left(\boldsymbol{a}_f, \boldsymbol{a}_{-f}, \boldsymbol{x}_t, \kappa_t, \boldsymbol{\eta}_{ft}\right) \;=\; \int VP_f\left(\boldsymbol{a}_f, \boldsymbol{a}_{-f}, \boldsymbol{x}_t, \boldsymbol{\xi}_t\right)\; dF_{f,\xi}\left(\boldsymbol{\xi}_t \mid \kappa_t, \boldsymbol{\eta}_{ft}\right) - f(\boldsymbol{x}_{ft}, \boldsymbol{\eta}_{ft}), \tag{15}$$

where $F_{f,\xi}\left(\boldsymbol{\xi}_t \mid \kappa_t, \boldsymbol{\eta}_{ft}\right)$ is a CDF and represents firm j's beliefs about the distribution of $\boldsymbol{\xi}_t$ conditional on $(\kappa_t, \boldsymbol{\eta}_{ft})$.

Given this expected profit, the definition of a Bayesian Nash Equilibrium (BNE) in the entry model for multi-product firms remains fundamentally the same as in the single-product case outlined earlier. The only distinction is that, in the multi-product scenario, an entry probability is associated with selecting a specific product portfolio.

While the preceding discussion illustrates the similar structure shared by the selection problem with single- and multi-product firms, it also highlights a significant practical difference. The dimension of the vector of choice probabilities we need to control for to deal with sample selection bias grows exponentially with the number of products per firm. In some applications, this can pose a substantial challenge in the practical implementation of our method.

2.5 Dynamic game of product entry and exit

Our framework and identification results can accommodate cases in which firms' decisions about product availability come from a Markov Perfect Equilibrium (MPE) of a dynamic game of product entry and exit, where firms are forward-looking. In this dynamic game, a firm's fixed cost is denoted as $f(a_{it}, a_{i,t-1}, \boldsymbol{x}_{jt}, \eta_{jt})$, where $f(1, 0, \boldsymbol{x}_{jt}, \eta_{jt})$ represents the cost of entry, $f(0, 1, \boldsymbol{x}_{jt}, \eta_{jt})$ is the cost of exit, $f(1, 1, \boldsymbol{x}_{jt}, \eta_{jt})$ is the fixed cost when a product stays in the market, and $f(0, 0, \boldsymbol{x}_{jt}, \eta_{jt})$ can be normalized to zero.

ASSUMPTION 1-Dyn. *Suppose that t represents time. Conditions (A) to (D) in Assumption 1 hold, and we have the following additional conditions. (E) The vector of state variables at period t, $\boldsymbol{x}_t$, includes the entry decisions of all the firms at the previous period, $(a_{j,t-1} : j = 1, 2, ..., J)$. (F) The exogenous product characteristics in vector $\boldsymbol{x}_t$ and the latent market type κ_t are either time-invariant or follow a first-order Markov process. (G) The private information signal η_{jt} is independently and identically distributed over time and independent across firms.* ■

The conditions in Assumption 1-Dyn are standard in the literature of empirical dynamic games of oligopoly competition (see Aguirregabiria, Collard-Wexler, and Ryan, 2021). Under Assumption 1-Dyn, the value of being or not in the market depends on the state variables $(\boldsymbol{x}_t, \kappa_t)$ and on the private information shock η_{jt}. Let $v_j^P(\boldsymbol{x}_t, \kappa_t, \eta_{jt})$ be the difference between the value functions of being in the market and not being in the market at period t. This function can be represented as the sum of two functions: the difference between current profits and the difference between expected continuation values. Similar to a BNE in a static entry game, a MPE in a dynamic game can be characterized in terms of J conditional choice probabilities.

DEFINITION 2 (MPE). *Suppose that Assumptions 1-2 hold. Then, a Markov Perfect Equilibrium (MPE) can be represented as a J-tuple of probability functions $\{P_j(\boldsymbol{x}_t, \kappa_t) : j \in \mathcal{J})\}$ that solve the following system of best response equations in the space of probability functions:*

$$P_j\left(\boldsymbol{x}_t, \kappa_t\right) \;=\; \int 1\{v_j^P(\boldsymbol{x}_t, \kappa_t, \eta_{jt}) \geq 0\}\; dF_\eta\left(\eta_{jt}\right). \qquad \blacksquare \tag{16}$$

For the rest of the paper, we will not distinguish whether the choice probabilities $P_j\left(\boldsymbol{x}_t, \kappa_t\right)$ come from a BNE of a static entry game or from a MPE of a dynamic game of entry and exit. All our identification results apply to both cases.

3 Selection problem

For simplicity and concreteness, we describe our sample selection problem using the nested logit demand model from Example 1 (stressing that none of our results require such a restriction). 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:

$$\ln\left(\frac{s_{jt}^*}{s_{0t}}\right) \;=\; \sigma\;\ln\left(\frac{s_{jt}^* + S_{-jt}}{s_{0t}}\right) + \alpha\; p_{jt}^* + \boldsymbol{x}_{jt}'\;\boldsymbol{\beta} + \xi_{jt}, \tag{17}$$

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:

$$\{s_{jt}^* = s_{jt} \text{ and } p_{jt}^* = p_{jt}\} \text{ if and only if } a_{jt} = 1. \tag{18}$$

Firm j's best response entry decision completes the econometric model:

$$a_{jt} \;=\; 1\left\{\pi_j^P(\boldsymbol{x}_t, \kappa_t, \eta_{jt}) \geq 0\right\}. \tag{19}$$

Equations (17) to (19) imply the following regression equation for any product with $a_{jt} = 1$:

$$\ln\left(\frac{s_{jt}}{s_{0t}}\right) \;=\; \sigma\ \ln\left(\frac{s_{jt}+S_{-jt}}{s_{0t}}\right) + \alpha\ p_{jt} + \boldsymbol{x}'_{jt}\ \boldsymbol{\beta} + \lambda_j(\boldsymbol{x}_t) + \widetilde{\xi}_{jt}, \tag{20}$$

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,

$$\lambda_j(\boldsymbol{x}_t) = \int \xi_{jt}\ 1\left\{\pi_j^P(\boldsymbol{x}_t, \kappa_t, \eta_{jt}) \geq 0\right\}\ \frac{f_{\xi,\eta,\kappa}\left(\xi_{jt}, \eta_{jt}, \kappa_t \mid \boldsymbol{x}_t\right)}{\bar{P}_j\left(\boldsymbol{x}_t\right)}\ d\left(\xi_{jt},\eta_{jt},\kappa_t\right), \tag{21}$$

where $f_{\xi,\eta,\kappa}$ is the joint density function of $(\xi_{jt},\eta_{jt},\kappa_t)$ conditional of $\boldsymbol{x}_t$, and $\bar{P}_j\left(\boldsymbol{x}_t\right)$ is the *selection propensity score*,

$$\bar{P}_j\left(\boldsymbol{x}_t\right) \equiv \Pr\left(a_{jt}=1 \mid \boldsymbol{x}_t\right) = \int 1\left\{\pi_j^P(\boldsymbol{x}_t, \kappa_t, \eta_{jt}) \geq 0\right\}\ f_{\eta,\kappa}\left(\eta_{jt}, \kappa_t \mid \boldsymbol{x}_t\right)\ d\left(\eta_{jt}, \kappa_t\right). \tag{22}$$

In the econometrics literature on sample selection, it is well-known that estimating equation (20) using instrumental variables, where $\lambda_j(\boldsymbol{x}_t) + \widetilde{\xi}_{jt}$ is treated as the error term, is unfeasible. This is because $\lambda_j(\boldsymbol{x}_t)$ is an unknown function of all exogenous variables in the model, leaving no viable candidates as valid instruments (Wooldridge, 2010). To address sample selection, a control function approach can be employed to account for the selection term $\lambda_j(\boldsymbol{x}_t)$. However, we cannot identify demand parameters without additional structure on this selection term. The selection term is an unknown function of all exogenous variables, preventing us from disentangling the direct effect of $\boldsymbol{x}_{jt}$ on consumer demand (as represented by the vector of parameters $\boldsymbol{\beta}$) from its effect through the selection term.

In this context, the standard approach in the literature is to establish conditions under which this selection term depends solely on the *selection propensity score* $\bar{P}_j\left(\boldsymbol{x}_t\right)$, i.e., $\lambda_j(\boldsymbol{x}_t) = \rho_j(\bar{P}_j\left(\boldsymbol{x}_t\right))$. As such, identification and estimation follow a standard two-step procedure. In a first step, we nonparametrically estimate $\bar{P}_j\left(\boldsymbol{x}_t\right)$ from data on $(a_{jt}, \boldsymbol{x}_t)$. Then, in a second step, one can apply the semiparametric series estimator in Das, Newey, and Vella (2003) and Newey (2009), or the pairwise differencing method in Powell (2001) and Aradillas-Lopez (2012). Valid instruments in this regression are observed $\boldsymbol{x}_{-jt}$ characteristics of products other than j, i.e., the so-called BLP instruments.

Angrist (1997) establishes necessary and sufficient conditions for the selection propensity score to be a sufficient statistic to control for sample selection bias in a very general class of selection models that includes our demand/entry model as a particular case. The following Lemma 2 presents these conditions and is an adaptation of Propositions 2 and 3 in Angrist (1997).

LEMMA 2. *(A) [Necessary and sufficient condition] Conditioning on the propensity score* $\bar{P}_j(\boldsymbol{x}_t)$ *controls for selection bias in the estimation of demand parameters if and only if* $\Pr(\xi_{jt}, a_{jt} \mid \boldsymbol{x}_t, \bar{P}_j(\boldsymbol{x}_t)) = \Pr(\xi_{jt}, a_{jt} \mid \bar{P}_j(\boldsymbol{x}_t))$. *(B) [Weak sufficient condition] Suppose that for any two values in the support of* $\boldsymbol{x}_t$*, say* $\boldsymbol{x}^0$ *and* $\boldsymbol{x}^1$*, the sign of* $\Pr(a_{jt} = 1 \mid \boldsymbol{x}^1, \xi_{jt}) - \Pr(a_{jt} = 1 \mid \boldsymbol{x}^0, \xi_{jt})$ *is the same (almost surely) for every value* ξ_{jt} *in the support of this random variable. Then, conditioning on the propensity score* $\bar{P}_j(\boldsymbol{x}_t)$ *controls for selection bias in the estimation of demand parameters.* ■

Lemma 2(A) provides a clear, necessary, and sufficient condition for the propensity score to control for selection. This condition, consequently, implies the identification of demand parameters. Combining Lemma 2 with equation (19) characterizing the optimal entry decision, the condition stated in Lemma 2(A) can be equivalently expressed as follows:

$$\Pr(\xi_{jt},\ \pi_j^P(\boldsymbol{x}_t, \kappa_t, \eta_{jt}) \mid \boldsymbol{x}_t, \bar{P}_j(\boldsymbol{x}_t)) = \Pr(\xi_{jt},\ \pi_j^P(\boldsymbol{x}_t, \kappa_t, \eta_{jt}) \mid \bar{P}_j(\boldsymbol{x}_t)). \tag{23}$$

Unfortunately, this condition involves an endogenous equilibrium object rather than the model's primitives, making it challenging to verify in models where the expected profit function is a complex endogenous entity. Nevertheless, the representation of this condition in equation (23) demonstrates that assuming independence between the unobservables $(\boldsymbol{\xi}_t, \kappa_t, \eta_{jt})$ and $\boldsymbol{x}_t$ is insufficient for the condition to hold. The functional form of the profit function π_j^P, particularly how the unobservables enter into this function, is crucial. This is further illustrated in Examples 2 and 3.

EXAMPLE 2. Suppose that the expected profit function has the following structure:

$$\pi_j^P(\boldsymbol{x}_t, \kappa_t, \eta_{jt}) = \gamma_{1j}(\boldsymbol{x}_t) - \gamma_{2j}(\boldsymbol{x}_t)\ \gamma_{3j}(\kappa_t, \eta_{jt}), \tag{24}$$

where $\gamma_{1j}(.)$, $\gamma_{2j}(.)$, and $\gamma_{3j}(.)$ are scalar real-valued functions. Given this structure, the optimal entry decision becomes $a_{jt} = 1\{\gamma_{3j}(\kappa_t, \eta_{jt}) \leq \gamma_{1j}(\boldsymbol{x}_t)/\gamma_{2j}(\boldsymbol{x}_t)\}$. Assume that κ_t and η_{jt} are jointly independent of $\boldsymbol{x}_t$, which implies independence between $\gamma_{3jt} \equiv \gamma_{3j}(\kappa_t, \eta_{jt})$ and $\boldsymbol{x}_t$. Consequently, the selection propensity score is given by $\bar{P}_j(\boldsymbol{x}_t) = F_{\gamma_{3j}}(\gamma_{1j}(\boldsymbol{x}_t)/\gamma_{2j}(\boldsymbol{x}_t))$, where $F_{\gamma_{3j}}$ is the cumulative distribution function of γ_{3jt}. Under these conditions, it is evident that the necessary and sufficient condition in Lemma 2(A) holds, ensuring that $\bar{P}_j(\boldsymbol{x}_t)$ effectively controls for selection. ■

Example 2 illustrates a single-crossing structure in the the selection or entry decision function. This single-crossing condition is in terms of the scalar function $\gamma_{3j}(\kappa_t, \eta_t)$, which encapsulates the influence of all unobservables on the expected profit function. While this condition is sufficient but not necessary for the propensity score to account for selection, finding examples where the conditions in Lemma 2(A) or 2(B) are satisfied without this structure is challenging. Furthermore, for a general class of demand and entry models, this single-crossing structure never holds. We illustrate this in Example 3.

EXAMPLE 3. This example illustrates the crucial role that the multi-dimensional aspect of demand unobservables plays in preventing the propensity score from effectively controlling for selection bias. To make this concrete, consider a model of market entry with complete information (no η_{jt}) and no uncertainty, such that $\kappa_t = \boldsymbol{\xi}_t$. For simplicity, assume there are only two firms potentially competing in the market ($J = 2$). Suppose the expected profit function for product 1 is given by:

$$\pi_1^P(\boldsymbol{x}_t, \boldsymbol{\xi}_t) = \gamma_1(\boldsymbol{x}_t)\ (\boldsymbol{x}_{1t}\ \boldsymbol{\beta} + \xi_{1t}) + \gamma_2(\boldsymbol{x}_t)\ (\boldsymbol{x}_{1t}\ \boldsymbol{\beta} + \xi_{1t})\ (\boldsymbol{x}_{2t}\ \boldsymbol{\beta} + \xi_{2t}) \tag{25}$$

where $\gamma_1(\boldsymbol{x}_t)$ and $\gamma_2(\boldsymbol{x}_t)$ are scalar real-valued functions. The non-additive structure in the second additive term, involving the demands for products 1 and 2, implies that this profit function does not exhibit the single-crossing property described in Example 2. Moreover, it can be shown that the conditions outlined in Lemma 2 do not hold, indicating that the propensity score alone is insufficient to control for selection bias. ■

In the remainder of this section, we derive an expression that characterizes the selection term $\lambda_j(\boldsymbol{x}_t)$ as a function of the distribution of κ_t and the equilibrium CCPs $P_j(\boldsymbol{x}_t, \kappa_t)$. This characterization is crucial for our identification results. For this result, we make the following Assumption.

ASSUMPTION 2. *Let $P_j(\boldsymbol{x}_t, \kappa_t)$ be the probability $\Pr(a_{jt} = 1 \mid \boldsymbol{x}_t, \kappa_t)$. Conditional on κ_t and $P_j(\boldsymbol{x}_t, \kappa_t)$, variables (ξ_{jt}, a_{jt}) are jointly independent of $\boldsymbol{x}_t$. That is,* $\Pr(\xi_{jt}, a_{jt} \mid \boldsymbol{x}_t, \kappa_t, P_j(\boldsymbol{x}_t, \kappa_t)) = \Pr(\xi_{jt}, a_{jt} \mid \kappa_t, P_j(\boldsymbol{x}_t, \kappa_t))$. ■

A more structural condition that ensures Assumption 2 is the strict monotonicity of the profit function with respect to η_{jt}. For example, in the same spirit as Example 2, a sufficient condition for Assumption 2 is that the expected profit function takes the following form: $\pi_j^P(\boldsymbol{x}_t, \kappa_t, \eta_{jt}) = \gamma_{1j}(\boldsymbol{x}_t, \kappa_t) - \gamma_{2j}(\boldsymbol{x}_t, \kappa_t)\ \gamma_{3j}(\eta_{jt})$.

Under Assumption 2, we have that:

$$
\begin{aligned}
\Pr(\xi_{jt} \mid a_{jt}=1, \boldsymbol{x}_t, \kappa_t, P_j(\boldsymbol{x}_t,\kappa_t)) &= \frac{\Pr(\xi_{jt}, a_{jt}=1 \mid \boldsymbol{x}_t, \kappa_t, P_j(\boldsymbol{x}_t,\kappa_t))}{\Pr(a_{jt}=1 \mid \boldsymbol{x}_t, \kappa_t, P_j(\boldsymbol{x}_t,\kappa_t))} \\
&= \frac{\Pr(\xi_{jt}, a_{jt}=1 \mid \kappa_t, P_j(\boldsymbol{x}_t,\kappa_t))}{P_j(\boldsymbol{x}_t,\kappa_t)}.
\end{aligned}
\tag{26}
$$

It then follows that,

$$
\begin{aligned}
\mathbb{E}(\xi_{jt} \mid \boldsymbol{x}_t, \kappa_t, a_{jt}=1) &= \int \xi_{jt} \, \frac{\Pr(\xi_{jt}, a_{jt}=1 \mid \kappa_t, P_j(\boldsymbol{x}_t,\kappa_t))}{P_j(\boldsymbol{x}_t,\kappa_t)} \, d\xi_{jt} \\
&\equiv \psi_j\left(P_j\left(\boldsymbol{x}_t,\kappa_t\right), \kappa_t\right).
\end{aligned}
\tag{27}
$$

To obtain the selection function $\lambda_j(\boldsymbol{x}_t) \equiv \mathbb{E}(\xi_{jt} \mid \boldsymbol{x}_t, a_{jt}=1)$, we must integrate (27) over the distribution of κ_t conditional on $(\boldsymbol{x}_t, a_{jt}=1)$. That is:

$$
\lambda_j(\boldsymbol{x}_t) = \int \psi_j\left(P_j\left(\boldsymbol{x}_t,\kappa_t\right),\kappa_t\right) \; f_{j,\kappa}\left(\kappa_t \mid \boldsymbol{x}_t\right) \; d\kappa_t, \tag{28}
$$

where $f_{j,\kappa}\left(\kappa_t \mid \boldsymbol{x}_t\right)$ denotes the distribution of κ_t conditional on $(\boldsymbol{x}_t, a_{jt}=1)$ and has the following structure:

$$
f_{j,\kappa}(\kappa_t \mid \boldsymbol{x}_t) \equiv \Pr(\kappa_t \mid \boldsymbol{x}_t, a_{jt}=1) = \frac{P_j\left(\boldsymbol{x}_t,\kappa_t\right)}{\bar{P}_j\left(\boldsymbol{x}_t\right)} f_\kappa(\kappa_t \mid \boldsymbol{x}_t). \tag{29}
$$

Equation (28) shows that, without additional constraints, the selection term depends not only on the propensity score $\bar{P}_j\left(\boldsymbol{x}_t\right)$ but also on the distribution of κ_t and the entry probabilities conditional on both $\boldsymbol{x}_t$ and κ_t. It is important to note that the functions ψ_j are unknown to the researcher and may vary with κ_t. Consequently, if κ_t has continuous support, there could be an infinite number of these functions. Given this structure of the selection term, it is evident that we cannot separately identify the demand parameters and the ψ_j functions. This holds true even in the hypothetical scenario in which the researcher is able to identify, in a preliminary step based on firms' entry decisions, the distribution of κ_t and the entry probabilities conditional on both $\boldsymbol{x}_t$ and κ_t.

To address this identification challenge, we strengthen Assumption 2 in the following way.

ASSUMPTION 2*. *Let κ_t^* be a proxy variable for κ_t with finite support $\mathcal{K}$. Define $P_j(\boldsymbol{x}_t,\kappa_t^*) \equiv \Pr(a_{jt}=1 \mid \boldsymbol{x}_t, \kappa_t^*)$. Conditional on κ_t^* and $P_j(\boldsymbol{x}_t,\kappa_t^*)$, variables (ξ_{jt}, a_{jt}) are jointly independent of $\boldsymbol{x}_t$. That is, $\Pr(\xi_{jt}, a_{jt} \mid \boldsymbol{x}_t, \kappa_t^*, P_j(\boldsymbol{x}_t,\kappa_t^*)) = \Pr(\xi_{jt}, a_{jt} \mid \kappa_t^*, P_j(\boldsymbol{x}_t,\kappa_t^*))$.* ■

Under Assumptions 1 and 2* and supposing that the proxy κ_t^* is unobserved to the researcher, the same derivations as above show that the selection term can be expressed as:

$$\lambda_j(\boldsymbol{x}_t) \;=\; \sum_{\kappa_t^* \in \mathcal{K}} \psi_j \left(P_j\left(\boldsymbol{x}_t, \kappa_t^*\right), \kappa_t^*\right) \; f_{j,\kappa^*}\left(\kappa_t^* \mid \boldsymbol{x}_t\right), \tag{30}$$

with $f_{j,\kappa^*}\left(\kappa_t^* \mid \boldsymbol{x}_t\right) \equiv \frac{P_j(\boldsymbol{x}_t,\kappa_t^*)}{P_j(\boldsymbol{x}_t)} \; f_{\kappa^*}(\kappa_t^* \mid \boldsymbol{x}_t)$. In the reminder of the paper, we refer to the unobserved κ_t^* interchangeably as latent class, unobserved market type, or unobserved proxy.

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 $\boldsymbol{x}_t \in \mathcal{X}$ and the vectors of firms' entry decisions $\boldsymbol{a}_t \in \{0,1\}^J$. The space $\mathcal{X}$ can be discrete or continuous. 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 proxies for the demand/cost variables, $\boldsymbol{f}_{\kappa^*} \equiv (f_{\kappa^*}(\kappa^* \mid \boldsymbol{x})$: for every $\kappa^*, \boldsymbol{x})$; the corresponding conditional entry 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 conditional distribution of the demand unobservables, $f_{\xi|\eta,\kappa^*}$:

$$\boldsymbol{\theta} \;\equiv\; \left(\, \boldsymbol{\theta}_\delta,\; \boldsymbol{P}_{\kappa^*},\; \boldsymbol{f}_{\kappa^*},\; f_{\xi|\eta,\kappa^*},\; F_\eta \,\right). \tag{31}$$

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 entry probabilities $\boldsymbol{P}_{\kappa^*}$ are nonparametrically specified.

We consider a two-step sequential procedure for the identification of $\boldsymbol{\theta}_\delta$. First, given the empirical distribution of firms' entry decisions, we establish the identification of the equilibrium probabilities $\boldsymbol{P}_{\kappa^*}$ and the distribution $\boldsymbol{f}_{\kappa^*}$. Then, given the structure of the selection bias function in (30), we show the identification of $\boldsymbol{\theta}_\delta$.

4.2 First step: Game of market entry

4.2.1 A general representation of the probability of entry profiles

In this section, we show that for a broad class of market entry games, and under minimal assumptions, the probabilities of firms' entry profiles, $\Pr(a_{1t}, a_{2t}, ..., a_{Jt}|\boldsymbol{x}_t)$, can be conveniently represented as a nonparametric finite mixture. Recent advances in tensor or multi-way linear algebra have proven useful in the representation of general high-dimensional arrays in terms of simpler lower-dimensional ones and are now ubiquitously applied in the fields of, e.g., signal processing, statistics, data mining, and machine learning (Sidiropoulos, De Lathauwer, Fu, Huang, Papalexakis, and Faloutsos, 2017; Kolda and Bader, 2009). By interpreting multivariate probability mass functions as multi-way arrays, these tensor decomposition techniques have helped researchers represent potentially complex multivariate probabilistic processes in terms of simpler univariate probabilities (Dunson and Xing, 2009; Yang and Dunson, 2016).

In the context of market entry games, the observed joint probability of the entry decisions $\boldsymbol{a}_t \in \{0,1\}^J$ of J firms conditional on a vector of exogenous variables $\boldsymbol{x}_t \in \mathcal{X}$ can be seen as a bounded (between 0 and 1) J-way tensor. Leveraging properties of the canonical polyadic decomposition (Harshman et al., 1970; Carroll and Chang, 1970), Kargas, Sidiropoulos, and Fu (2018) show that any J-dimensional probability mass function admits a very convenient nonparametric finite mixture or latent class representation. The following Lemma summarizes this result and is an adaptation to our context of Proposition 1 by Kargas, Sidiropoulos, and Fu (2018).

LEMMA 3. *For any* $(\boldsymbol{a}, \boldsymbol{x}) \in \{0,1\}^J \times \mathcal{X}$ *with* $J \geq 3$*, any arbitrary probability mass function* $\Pr(\boldsymbol{a}_t = \boldsymbol{a} \mid \boldsymbol{x}_t = \boldsymbol{x})$ *admits the nonparametric finite mixture representation:*

$$\Pr(\boldsymbol{a}_t = \boldsymbol{a} \mid \boldsymbol{x}_t = \boldsymbol{x}) = \sum_{\kappa^* \in \mathcal{K}(\boldsymbol{x})} f_{\kappa^*}(\kappa^* \mid \boldsymbol{x}) \left[\prod_{j=1}^{J} [P_j(\boldsymbol{x}, \kappa^*)]^{a_j} \;\; [1 - P_j(\boldsymbol{x}, \kappa^*)]^{1-a_j} \right], \tag{32}$$

with $\mathcal{K}(\boldsymbol{x})$ *a discrete and finite collection of latent classes with at most* $|\mathcal{K}(\boldsymbol{x})| \leq 2^{J-1}$ *components,* $f_{\kappa^*}(\kappa^* \mid \boldsymbol{x})$ *the probability of latent class* κ^* *conditional on* $\boldsymbol{x}$*, and* $P_j(\boldsymbol{x}, \kappa^*)$ *the probability of entry of firm* j *conditional on* $\boldsymbol{x}$ *and latent class* κ^**.* ■

This result states that *any* arbitrary probability mass function $\Pr(\boldsymbol{a}_t = \boldsymbol{a} \mid \boldsymbol{x}_t = \boldsymbol{x})$, which could arise from *any* game of product entry, can be represented as a convenient nonparametric finite mixture with: (i) a *finite* number of latent classes $\kappa^* \in \mathcal{K}(\boldsymbol{x})$ with at most $|\mathcal{K}(\boldsymbol{x})| \leq 2^{J-1}$ components and (ii) the entry probability $P_j(\boldsymbol{x}, \kappa^*)$ of each firm j *conditionally independent* from the others $P_i(\boldsymbol{x}, \kappa^*), i \neq j$. Finite mixture representation (32) is inherently *nonparametric* in that Lemma 3 does not pose any further restriction on $f_{\kappa^*}(\kappa^* \mid \boldsymbol{x})$ and $P_j(\boldsymbol{x}, \kappa^*)$ beyond the fact that these are probability mass functions. Moreover, it is *pointwise* with respect to $\boldsymbol{x}$, as for any different value of $\boldsymbol{x} \in \mathcal{X}$ the probability mass function $\Pr(\boldsymbol{a}_t = \boldsymbol{a} \mid \boldsymbol{x}_t = \boldsymbol{x})$ may admit a different nonparametric finite mixture representation (e.g., with a different number of latent classes, different probabilities $f_{\kappa^*}(\kappa^* \mid \boldsymbol{x})$ and $P_j(\boldsymbol{x}, \kappa^*)$). Finally, while Lemma 3 guarantees the existence of at least "a" representation as in (32), such a representation may not be unique. We return to the issue of uniqueness in section 4.2 when discussing about identification (see also related discussion in Kargas, Sidiropoulos, and Fu (2018)).

Remarkably, nonparametric finite mixture representation (32) resembles the joint probability of entry implied by the games of incomplete information studied in Aguirregabiria and Mira (2019); Xiao (2018). More formally, Lemma 3 shows that any arbitrary joint probability of entry $\Pr(\boldsymbol{a}_t = \boldsymbol{a} \mid \boldsymbol{x}_t = \boldsymbol{x})$ can be represented "as if" the J firms were playing an entry game of incomplete information along the lines of those proposed by Aguirregabiria and Mira (2019); Xiao (2018). Importantly though, and perhaps surprisingly, Lemma 3 highlights how both the discrete and finite number of latent classes (or unobserved market types) $\mathcal{K}(\boldsymbol{x})$ and the conditional independence of the firms' entry decisions are without loss of generality.

The identification of the entry probabilities $\boldsymbol{P}_{\kappa^*}$ and the distribution $\boldsymbol{f}_{\kappa^*}$ in the nonparametric finite mixture representation in (32) 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 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 $\boldsymbol{x}$. To simplify notation, for the rest of this subsection we then omit both $\boldsymbol{x}$ and the market subscript t .

4.2.2 Identification of the number of latent market types

The number of components $|\mathcal{K}|$ in finite mixture (32) 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' entry decisions $(a_1, a_2, ..., a_J)$ such that Y_1 is equal to the entry decision of one firm (if J is odd) or two firms (if J is even), and variables Y_2 and Y_3 evenly divide the entry decisions 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 matrix of probabilities for each possible value of Y_i — in the rows of the matrix — conditional on every possible value of κ^* — in the columns of the matrix. The main idea is then to identify the number of components $|\mathcal{K}|$ from the observed joint distribution of Y_2 and Y_3:

$$\Pr(Y_2 = y_2, Y_3 = y_3) \;=\; \sum_{\kappa^*=1}^{|\mathcal{K}|} \Pr(Y_2 = y_2 \mid \kappa^*) \; \Pr(Y_3 = y_3 \mid \kappa^*) \; f_{\kappa^*}(\kappa^*) \tag{33}$$

or, in matrix notation,

$$\boldsymbol{P}_{Y_2,Y_3} \;=\; \boldsymbol{P}_{Y_2|\kappa^*} \; diag(\boldsymbol{f}_{\kappa^*}) \; \boldsymbol{P}'_{Y_3|\kappa^*}, \tag{34}$$

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 $\Pr(Y_i = y \mid \kappa^*)$; and $diag(\boldsymbol{f}_{\kappa^*})$ is the $|\mathcal{K}| \times |\mathcal{K}|$ diagonal matrix with the probabilities $f_{\kappa^*}(\kappa^*)$.

LEMMA 4. *Without further restrictions, Rank($\boldsymbol{P}_{Y_2,Y_3}$) is a lower bound for 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 entry probabilities associated to each component κ^* cannot be obtained as linear combinations of the others.

4.2.3 Identification of equilibrium CCPs and distribution of latent types

Allman, Matias, and Rhodes (2009) study the identification of nonparametric 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 5 is an adaptation of Theorem 4 and Corollary 5 in Allman, Matias, and Rhodes (2009).

LEMMA 5. *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.* ■

Note that order condition (i) in Lemma 4 is in general more stringent than order condition (ii) in Lemma 5: 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 4 hold 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.

The identification of the distribution of κ_t^* and the equilibrium CCPs is up to label swapping, and pointwise or separately for each value of the observable $\boldsymbol{x}_t$. In the absence of additional assumptions, the combination of these two features leads to an identification issue in the second step of our method. In fact, in the estimation of the demand equation in the second step, we need to include $f_{\kappa^*}(\kappa^* \mid \boldsymbol{x}_t)$ and $P_j(\kappa^*, \boldsymbol{x}_t)$ for every value of κ^* as additional regressors, or more precisely as control variables. To construct these regressors, we need to be able to "match" the same latent type κ^* across different observed values of $\boldsymbol{x}_t$ in the sample. However, this task is not feasible without further assumptions.

Aguirregabiria and Mira (2019) discuss alternative assumptions that can solve this matching-latent-types problem. In our empirical application, we opt for the independence between κ_t and $\boldsymbol{x}_t$. This assumption addresses the challenge by altering the nature of identification in the first step: rather than being pointwise with respect to $\boldsymbol{x}_t$, identification holds uniformly across all values of $\boldsymbol{x}_t$. Therefore, though identification is still up to label swapping, the same label κ^* will apply to all values of $\boldsymbol{x}_t$, effectively removing the problem of matching-latent-types.

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:

$$\delta_j(\boldsymbol{s}_t, \boldsymbol{\sigma}) \;=\; \alpha\; p_{jt} + \boldsymbol{x}_{jt}'\;\boldsymbol{\beta} + \xi_{jt}, \qquad \text{for } a_{jt} = 1 \tag{35}$$

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 arises 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 or proxy 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 (27) and we would have a standard selection problem represented by the semiparametric partially linear model:

$$\delta_j(\boldsymbol{s}_t, \boldsymbol{\sigma}) = \alpha\; p_{jt} + \boldsymbol{x}_{jt}'\;\boldsymbol{\beta} + \psi_j\left(P_j(\boldsymbol{x}_t, \kappa_t^*), \kappa_t^*\right) + \widetilde{\xi}_{jt}. \tag{36}$$

A key complication of the selection problem in our model is that the market type or proxy κ_t^*

is unobserved to the researcher. After the first step of the identification procedure, 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 (36). We instead need to deal with the more complex selection bias function:

$$\lambda_j(\boldsymbol{x}_t) \equiv \mathbb{E}\left(\xi_{jt} \mid \boldsymbol{x}_t, a_{jt}=1\right) = \sum_{\kappa^*=1}^{|\mathcal{K}(\boldsymbol{x}_t)|} f_{j,\kappa^*}(\kappa^* \mid \boldsymbol{x}_t)\ \psi_j\left(P_j(\boldsymbol{x}_t,\kappa^*),\kappa^*\right) \ = \ \boldsymbol{f}'_{j,\kappa^*,t}\ \boldsymbol{\psi}_j(\boldsymbol{P}_{j,t}), \quad (37)$$

where $\boldsymbol{f}_{j,\kappa^*,t}$, $\boldsymbol{P}_{j,t}$, and $\boldsymbol{\psi}_j(\boldsymbol{P}_{j,t})$ are all vectors of dimension $|\mathcal{K}(\boldsymbol{x}_t)|\times 1$. Therefore, the regression equation of our model is:

$$\delta_j(\boldsymbol{s}_t,\boldsymbol{\sigma}) = \alpha\ p_{jt} + \boldsymbol{x}'_{jt}\ \boldsymbol{\beta} + \boldsymbol{f}'_{j,\kappa^*,t}\ \boldsymbol{\psi}_j(\boldsymbol{P}_{j,t}) + \widetilde{\xi}_{jt}. \quad (38)$$

Define $\boldsymbol{f}_{\kappa^*,t} \equiv (f_{\kappa^*}(\kappa^* \mid \boldsymbol{x}_t) : \kappa^* = 1,2,...,|\mathcal{K}(\boldsymbol{x}_t)|)$. Note that as $f_{j,\kappa^*}(\kappa^* \mid \boldsymbol{x}_t)$ is a known function of $(\boldsymbol{P}_{j,t},\boldsymbol{f}_{\kappa^*,t})$, equation (38) then clarifies that $(\boldsymbol{P}_{j,t},\boldsymbol{f}_{\kappa^*,t})$ is a sufficient statistic for the selection bias function.

Proposition 1 establishes a necessary and sufficient condition for the identification of $\boldsymbol{\theta}_\delta \equiv (\alpha,\boldsymbol{\beta},\boldsymbol{\sigma})$ from equation (38). It is an application of Theorem 6 in Rothenberg (1971).

PROPOSITION 1. *Define the vector* $\boldsymbol{Z}_{jt} \equiv \left(\mathbb{E}\left(\frac{\partial\delta_j(\boldsymbol{s}_t,\boldsymbol{\sigma})}{\partial\boldsymbol{\sigma}} \mid \boldsymbol{x}_t\right),\ \mathbb{E}\left(p_{jt} \mid \boldsymbol{x}_t\right),\ \boldsymbol{x}'_{jt}\right)'$*, 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* $\mathbb{E}\left(\widetilde{\xi}_{jt} \mid \boldsymbol{x}_t\right) = \mathbb{E}\left(\widetilde{\xi}_{jt} \mid \boldsymbol{P}_{j,t},\boldsymbol{f}_{\kappa^*,t}\right) = 0$*, a necessary and sufficient condition for the identification of* $\boldsymbol{\theta}_\delta \equiv (\alpha,\boldsymbol{\beta},\boldsymbol{\sigma})$ *in equation (38) is that matrix* $\mathbb{E}\left(\widetilde{\boldsymbol{Z}}_{jt}\ \widetilde{\boldsymbol{Z}}'_{jt}\right)$ *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 the primitives of the model. However, on the basis of 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 would not be exclusion restrictions to deal with price endogeneity, 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 would also imply 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, without observable variables affecting entry but not demand, the model would not be identified 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 present a two-step estimation method that mimics our sequential identification result. In the first step, we use a nonparametric sieve maximum likelihood method to estimate the distribution of unobserved market types, the vector of entry probabilities for each unobserved type, and the number of unobserved market types. In the second step, we use sieves to approximate the selection bias term as a function of the densities and entry probabilities estimated in the first step.[5] Then, we apply GMM to jointly estimate the coefficients in the sieve approximation and the structural demand parameters. We calculate asymptotic standard errors of the estimates in the second step using the method and formulas in Newey (2009).

5.1 First step: Estimation of CCPs and distribution of latent types

We use sieves to approximate the nonparametric functions $f_{\kappa^*}(\kappa_t^* \mid \boldsymbol{x}_t)$ and $P_j(\boldsymbol{x}_t, \kappa_t)$ (Hirano, Imbens, and Ridder, 2003, Chen, 2007). Let $\boldsymbol{r}_t^f \equiv \left(r_1^f(\boldsymbol{x}_t), r_2^f(\boldsymbol{x}_t), ..., r_{L_f}^f(\boldsymbol{x}_t) \right)'$ be a vector with a finite number L_f of basis functions. The density function $f_{\kappa^*}(\kappa_t^* \mid \boldsymbol{x}_t)$ has the following sieves multinomial logit structure:

$$f_{\kappa^*}(\kappa^* \mid \boldsymbol{x}_t) \;=\; \frac{\exp\{\boldsymbol{r}_t^{f\prime}\, \boldsymbol{\gamma}_\kappa^f\}}{\sum_{\kappa'=1}^{|\mathcal{K}|} \exp\{\boldsymbol{r}_t^{f\prime}\, \boldsymbol{\gamma}_{\kappa'}^f\}}, \tag{39}$$

where, for $\kappa^* = 1, 2, ..., |\mathcal{K}|$, $\boldsymbol{\gamma}_{\kappa^*}^f$ is a vector of parameters with dimension $L_f \times 1$ and normalization $\boldsymbol{\gamma}^f(1) = 0$. Similarly, let $\boldsymbol{r}_t^P \equiv \left(r_1^P(\boldsymbol{x}_t), r_2^P(\boldsymbol{x}_t), ..., r_{L_P}^P(\boldsymbol{x}_t) \right)'$ be a vector with a finite number L_P of basis functions. For any product j and any unobserved type or proxy κ^*, the entry probability function $P_j(\boldsymbol{x}_t, \kappa^*)$ has the following sieves binary logit structure:

$$P_j(\boldsymbol{x}_t, \kappa^*) \;=\; \Lambda\left(\boldsymbol{r}_t^{P\prime}\, \boldsymbol{\gamma}_{j\kappa^*}^P \right), \tag{40}$$

[5] The second step could alternatively be based on differencing out the selection bias term using a matching estimator as in Ahn and Powell (1993), Powell (2001), and Aradillas-Lopez, Honoré, and Powell (2007).

where $\Lambda(\cdot)$ is the logistic function. For $j = 1, 2, ..., J$ and $\kappa^* = 1, 2, ..., |\mathcal{K}|$, we have that $\boldsymbol{\gamma}^P_{j\kappa^*}$ is a vector of parameters of dimension $L_P \times 1$. The log-likelihood function of this nonparametric finite mixture model is:

$$\ell(\boldsymbol{\gamma}^{f,P}) \;=\; \sum_{t=1}^{T} \ln \left(\sum_{\kappa^*=1}^{|\mathcal{K}|} f_{\kappa^*}(\kappa^* \mid \boldsymbol{x}_t, \boldsymbol{\gamma}^f) \; \prod_{j=1}^{J} \Lambda\left(\boldsymbol{r}_t^{P\prime}\, \boldsymbol{\gamma}^P_{j\kappa^*}\right)^{a_{jt}} \left[1 - \Lambda\left(\boldsymbol{r}_t^{P\prime}\, \boldsymbol{\gamma}^P_{j\kappa^*}\right)\right]^{1-a_{jt}} \right), \tag{41}$$

where $\boldsymbol{\gamma}^{f,P}$ is a vector collecting the parameters $\{\boldsymbol{\gamma}^f_{\kappa^*}, \boldsymbol{\gamma}^P_{j\kappa^*} : \kappa^* = 1, 2, ..., |\mathcal{K}|; j = 1, 2, ..., J\}$, with a total of $L_f(|\mathcal{K}| - 1) + L_P|\mathcal{K}|J$ parameters.

We estimate the vector of parameters $\boldsymbol{\gamma}$ by Maximum Likelihood (MLE) using the EM algorithm (Pilla and Lindsay, 2001). Recent papers considering MLE and the EM algorithm to estimate nonparametric mixtures in discrete choice models include Bunting (2022), Bunting, Diegert, and Maurel (2022), Hu and Xin (2022), and Williams (2020). Following this statistical literature, we use Akaike and Bayesian Information Criteria (AIC and BIC, respectively) to determine the number of latent classes $|\mathcal{K}|$.

When $\boldsymbol{x}_t$ is discrete, the nonparametric MLE is $\sqrt{T}$-consistent and asymptotically normal. With continuous variables in $\boldsymbol{x}_t$, the nonparametric MLE cannot achieve a $\sqrt{T}$ rate. However, under standard regularity conditions, this does not affect the $\sqrt{T}$-consistency and asymptotic normality of the estimator of the demand parameters in the second step. The proof of this result follows from Hirano, Imbens, and Ridder (2003) and Das, Newey, and Vella (2003).

5.2 Second step: Estimation of demand parameters

Following 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 L_ψ in the logarithm of the entry probability $P_j(\boldsymbol{x}_t, \kappa^*)$:

$$\begin{aligned} \psi_j\left(P_j(\boldsymbol{x}_t, \kappa^*), \kappa^*\right) \;&\approx\; \boldsymbol{r}^{\psi}\left(P_j(\boldsymbol{x}_t, \kappa^*)\right)' \boldsymbol{\gamma}^{\psi}_{j\kappa^*} \\ &=\; \left[1,\;\; \ln P_j(\boldsymbol{x}_t, \kappa^*),\;\; \ln P_j(\boldsymbol{x}_t, \kappa^*)^2,\;\; ...,\;\; \ln P_j(\boldsymbol{x}_t, \kappa^*)^{L_\psi}\right] \boldsymbol{\gamma}^{\psi}_{j\kappa^*}, \end{aligned} \tag{42}$$

where $\boldsymbol{\gamma}^{\psi}_{j\kappa^*} \equiv (\gamma^{\psi}_{0,j\kappa^*}, \gamma^{\psi}_{1,j\kappa^*}, ..., \gamma^{\psi}_{L_\psi,j\kappa^*})'$ is a vector of parameters. Given this approximation, the selection function is linear in $\boldsymbol{\gamma}^{\psi}_{j\kappa^*}$ and has the following expression:

$$\boldsymbol{f}'_{j,\kappa^*,t} \boldsymbol{\psi}_j(\boldsymbol{P}_{j,t}) \;\approx\; \boldsymbol{h}'_{j,t}\; \boldsymbol{\gamma}^{\psi}_j \;=\; \sum_{\kappa^*=1}^{|\mathcal{K}|} \sum_{\ell=0}^{L_\psi} \gamma^{\psi}_{j,\ell}(\kappa^*)\; f_{j,\kappa^*}(\kappa^* \mid \boldsymbol{x}_t)\; \left(\ln P_j(\boldsymbol{x}_t, \kappa^*)\right)^{\ell} \tag{43}$$

where $\boldsymbol{h}'_{j,t}$ is a vector with dimension $1 \times (L_\psi + 1)|\mathcal{K}|$ and elements $\{f_{j,\kappa^*}(\kappa^* \mid \boldsymbol{x}_t)\ (\ln P_j(\boldsymbol{x}_t, \kappa^*))^\ell :$ $\ell = 0, 1, ..., L_\psi; \kappa^* = 1, ..., |\mathcal{K}|\}$, and $\boldsymbol{\gamma}_j^\psi$ is a vector of parameters of the same dimension and with elements $\{\gamma_{j,\ell}^\psi(\kappa^*) : \ell = 0, 1, ..., L_\psi; \kappa^* = 1, ..., |\mathcal{K}|\}$.

Plugging equation (43) into demand equation (38), we obtain the regression equation:

$$\delta_j(\boldsymbol{s}_t, \boldsymbol{\sigma}) = \alpha\ p_{jt} + \boldsymbol{x}'_{jt}\ \boldsymbol{\beta} + \boldsymbol{h}'_{j,t}\ \boldsymbol{\gamma}_j^\psi + \widetilde{\xi}_{jt}. \tag{44}$$

Equation (44) can be estimated by GMM. Following Das, Newey, and Vella (2003), one can show that this two-step estimator of the vector of demand parameters $\boldsymbol{\theta}_\delta$ is $\sqrt{T}$-consistent and asymptotically normal.

6 Empirical application

6.1 Data and descriptive statistics

We apply our method to estimate demand in the US airline industry. The challenge of accounting for endogenous product entry in demand estimation in this industry has recently been explored by Ciliberto, Murry, and Tamer (2021) and Li, Mazur, Park, Roberts, Sweeting, and Zhang (2022).

We use publicly available data from the US Department of Transportation for our analysis. Our working sample consists of data from the DB1B and T100 databases. Specifically, we use quarterly data spanning from 2012-Q1 to 2013-Q4 for routes between the airports at the 100 largest Metropolitan Statistical Areas (MSA) in the United States. This accounts for 108 airports, as there are a few MSAs with more than one airport.

In terms of airlines' entry decisions, we define a market as a non-directional airport pair, where, for example, Chicago O'Hare (ORD) to New York La Guardia (LGA) is considered equivalent to LGA to ORD. There are potentially $5,778$ non-directional markets between the 108 airports, i.e., $108 * 107/2$. However, many of these markets have never had an incumbent airline with non-stop flights for several decades. These are typically airport pairs that are geographically too close or in smaller MSAs. In our sample, we consider only non-directional markets which were served in at least 50 quarters between the years 1994 to 2018. This accounts for $2,230$ non-directional markets in our sample, and $17,155$ market-quarter observations.[6] We consider an airline a *potential entrant* in a non-directional airport pair in a given quarter if it

[6]Given $2,230$ non-directional markets and eight quarters, the total number of market-quarter observations in our sample is $2,230 \times 8 = 17,840$. We however discard from the analysis 685 market-quarter observations for which we either do not observe some of the regressors or none of the six airlines included in the entry model is a potential entrant.

operates non-stop flights from either origin or destination airport (toward or from any airport), while an airline is an *entrant* in a non-directional airport pair in a given quarter if it operates non-stop flights between the origin and destination airports.

A product is defined as the combination of directional airport pair, airline, and an indicator for non-stop flight. For example, an American Airlines non-stop flight from LGA to ORD is a product. The airlines included in our analysis are American (AA), Delta (DL), United (UA), US Airways (US), Southwest (WN), a combined group of Low-Cost Carriers (LCC), and a combined group of the remaining carriers (Others).[7] Given the large number of carriers included in Others, we do not consider this combined group as a player in the entry game.

Following the empirical literature on the airline industry, we define market size as the geometric mean of the populations in the metropolitan regions (MSAs) of the two airports and market distance as the geodesic distance between the two airports.

Table 1 presents the distribution of the number of entrants and the average value of market characteristics. Notably, in a significant portion of these markets (30%), there are no airlines providing non-stop flights, and they are exclusively served with stop flights. Among the markets with non-stop flights, more than 90% are monopolies or duopolies. Furthermore, there is a strong positive correlation between the number of incumbents, market size, and distance.

Table 1: Distribution of Markets by Number of Entrants

Number of airlines	Frequency # markets-quarters (%)	Avg. market size in millions of people	Avg. market distance in miles
0 airlines	5,117 (29.83%)	7.09	734
1 airline	8,217 (47.90%)	8.82	913
2 airlines	2,637 (15.37%)	10.95	960
3 airlines	869 (5.07%)	13.00	1,117
4 airlines	233 (1.36%)	12.60	1,140
5 airlines	72 (0.42%)	20.16	1,255
$\geq$ 6 airlines	10 (0.06%)	17.54	320
Total	17,155 (100.00%)	8.95	882

Table 2 presents entry frequencies for each airline and the average market size and distance associated with their entry. We observe significant variation in airlines' entry probabilities, with WN and AA having the highest (27.5%) and the lowest (10.6%) entry propensities, respectively.

[7]Following Ciliberto, Murry, and Tamer (2021), the list of airlines included in the group LCC is: Alaska, JetBlue, Frontier, Allegiant, Spirit, Sun Country, and Virgin. The carriers in the group Others are small regional carriers, charters, and private jets.

Furthermore, substantial heterogeneity exists across airlines concerning the correlations between entry and market size and distance. For example, while WN enters markets that are not significantly different in size from the markets it does not enter (8.7 million people versus 9 million people), AA tends to enter markets with much larger average size (13.3 million people versus 8.4 million people). Moreover, different entry strategies are evident based on market distance. DL and US typically enter markets with an average distance of around 875-95 miles, whereas the markets served by LCC have an average distance of 1,171 miles.

Table 2: Entry Frequency by Airline

Airline	Frequency # markets-quarters (%)	Avg. market size in millions of people	Avg. market distance in miles
WN	4,714 (27.48%)	8.71	989
DL	3,285 (19.15%)	10.68	875
UA	3,244 (18.91%)	11.56	968
LCC	2,386 (13.91%)	11.42	1,171
US	2,001 (11.66%)	9.52	894
AA	1,820 (10.61%)	13.28	965

6.2 Estimation of the model for market entry

For the entry decisions, we consider the nonparametric sieve finite mixture Logit described in equations (39) and (40). The vector of explanatory variables in $\boldsymbol{x}_t$ includes: market size; market distance (see definitions above); the airline's own hub-size in the market, as measured by the sum of the airline's hub-sizes in the two airports; the average hub-sizes of the other airlines; and time indicators for each of the eight quarters in the sample.[8]

We have estimated various versions of the Logit entry model based on two key factors: the polynomial order in $\boldsymbol{x}_t$ used to construct the basis $\boldsymbol{r}_t^P$ and the number of points in the support of κ_t^*. Notably, the parameters of the entry model are specific to each airline and are unrestricted across airlines. Our estimation results for demand parameters are robust in relation to the selection of the basis $\boldsymbol{r}_t^P$ in the entry model. For brevity, we then present results only for the specification with $\boldsymbol{r}_t^P = \boldsymbol{x}_t$. Regarding the selection of the number of unobserved market types $|\mathcal{K}|$ and its influence on the estimation of demand parameters, as illustrated in Table 4 below, the primary and most substantial effect arises from allowing for *some* unobserved market

[8] We define the *hub-size* of an airline in an airport as the number of non-stop routes that the airline operates from that airport.

heterogeneity κ_t^*. Specifically, transitioning from a model without κ_t^* (i.e., one unobserved market type) to a model with two unobserved market types induces a noteworthy impact. Once we account for this type of unobserved market heterogeneity, demand estimates are however very similar among specifications that allow for two, three, or four unobserved market types.[9]

In this section, we present the estimation results for mixture Logit models, focusing on cases where the distribution of κ_t^* is independent of $\boldsymbol{x}_t$. Our choice stems from the robustness of our demand estimates to incorporating dependence between κ_t^* and $\boldsymbol{x}_t$, and because this independence assumption effectively addresses the identification challenge of matching latent market types due to label swapping (see discussion in Section 4.2.3). The robustness of our results to relaxing independence between κ_t^* and $\boldsymbol{x}_t$ is intuitive, as this assumption does not impose any exclusion restriction necessary to control for selection bias in the estimation of the demand parameters.

Furthermore, there is a practical computational rationale behind this decision. While estimating the mixture Logit model under the assumption of independence and with two or three unobserved market types requires only a few hours, and the Expectation-Maximization (EM) algorithm consistently converges, introducing dependence significantly complicates computations. The estimation process, even with just two unobserved market types, extends over several days of EM iterations, and the EM algorithm often fails to converge. While the model with dependence is theoretically identified, the practical implementation of the estimator considerably complicates in our empirical application.

Table 3 presents the goodness-of-fit statistics obtained from estimating four nested specifications of the market entry model. The selection of the preferred model is guided by the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), alongside other crucial considerations such as the convergence performance of the EM algorithm, the accuracy of parameter estimates of the entry model, and the robustness of the demand parameter estimates.

The introduction of unobserved market heterogeneity κ_t^* improves the model's goodness-of-fit. This is demonstrated by a substantial increase in the log-likelihood and a decrease in both AIC and BIC when comparing the model without κ_t^* and the model with two unobserved market types. This form of unobserved market heterogeneity captures a strong correlation among airlines' entry decisions, a correlation beyond the explanatory power of the observable market and airline characteristics in the vector $\boldsymbol{x}_t$.

The inclusion of additional unobserved market types continues to positively impact the goodness-of-fit. However, this impact shows diminishing returns, with small improvements when

[9]To simplify exposition, the estimation results for the model with four unobserved market types are not reported in Table 4 but are qualitatively similar to those for models with two or three unobserved market types.

we move from three to four unobserved market types. While the EM algorithm converges rapidly to the MLE in the specifications with two and three unobserved market types, we experience convergence problems in the model with four unobserved market types. In this case, we obtain imprecise estimates for some of the parameters in the entry model. These factors, combined with the marginal improvement observed in the AIC and BIC criteria as we transition from three to four unobserved market types, lead us to favor the model with $|\mathcal{K}| = 3$ unobserved market types.

Table 3: Estimation of Market Entry Model — Goodness-of-Fit Statistics

Statistics	**Logit** **# types = 1**	**Mixture Logit** **# types = 2**	**Mixture Logit** **# types = 3**	**Mixture Logit** **# types = 4**
Observations	$17,155$	$17,155$	$17,155$	$17,155$
Parameters	72	145	218	287
Log-likelihood	$-20,378$	$-18,985$	$-18,022$	$-17,621$
AIC	$40,900$	$38,261$	$36,481$	$35,817$
BIC	$41,458$	$39,385$	$38,170$	$38,041$

6.3 Estimation of demand parameters

For the demand system, we follow Ciliberto, Murry, and Tamer (2021) and estimate a nested logit demand with two nests: a nest for all the airlines and another nest for the outside alternative.

$$\ln\left(\frac{s_{jt}}{s_{0t}}\right) = \alpha\ p_{jt} + \boldsymbol{x}'_{jt}\boldsymbol{\beta} + \sigma\ \ln\left(\frac{s_{jt}}{1-s_{0t}}\right) + \boldsymbol{h}'_{jt}\ \boldsymbol{\gamma}_j^{\psi} + \widetilde{\xi}_{jt}. \tag{45}$$

We compute each directional route-specific market share in a given quarter s_{jt} as the total number of passengers who travelled that directional route by a specific airline in that given quarter (times 10, as the data is a survey of 10% of total traffic) divided by market size. The vector of product characteristics $\boldsymbol{x}_{jt}$ includes market distance and market distance squared, airline j's hub-size in the origin airport, airline j's hub-size in the destination airport, and airline $\times$ quarter fixed effects (indicators). The expression for the selection bias term, $\boldsymbol{h}'_{jt}\ \boldsymbol{\gamma}_j^{\psi}$, varies across the specifications of the market entry model, from the more restrictive parametric Logit model to the more general semiparametric finite mixture Logit model.

1. *Parametric Logit* specification. We consider the entry model $a_{jt} = 1\{\eta_{jt} \leq \boldsymbol{x}'_{jt}\boldsymbol{\gamma}_j^P\}$, with $\eta_{jt} \sim$ *Logistic*, and $\xi_{jt} = \gamma_{j,1}^{\psi}\ \eta_{jt} + v_{jt}$, with v_{jt} independent of η_{jt} and $\boldsymbol{x}_t$. Under this

parametric Logit selection model, the selection term has the following form:

$$\mathbb{E}\left(\xi_{jt} \mid a_{jt}=1, \boldsymbol{x}_t\right) \;=\; \gamma^{\psi}_{j,1}\ \mathbb{E}\left(\eta_{jt} \mid \eta_{jt} \leq \boldsymbol{x}'_{jt}\boldsymbol{\gamma}^{P}_{j}\right) \;=\; \gamma^{\psi}_{j,1}\ \left[Euler - \ln\Lambda\left(\boldsymbol{x}'_t\ \boldsymbol{\gamma}^{P}_{j}\right)\right] \tag{46}$$

where $Euler$ represents *Euler's constant* ≈ 0.5772. For the parametric Logit model, the term $Euler - \ln\Lambda\left(\boldsymbol{x}'_t\boldsymbol{\gamma}^{P}_{j}\right)$ is analogous to the inverse Mills ratio in the context of the parametric Probit model.

2. *Semiparametric Logit without* κ^*_t. The entry model is still the Logit $a_{jt} = 1\{\eta_{jt} \leq \boldsymbol{x}'_{jt}\boldsymbol{\gamma}^{P}_{j}\}$, with $\eta_{jt} \sim Logistic$, but now $\mathbb{E}\left(\xi_{jt} \mid a_{jt}=1, \boldsymbol{x}_t\right)$ is a third order polynomial in $Euler - \ln\Lambda\left(\boldsymbol{x}'_t\boldsymbol{\gamma}^{P}_{j}\right)$. Therefore, the vector of regressors controlling for endogenous selection is:

$$\boldsymbol{h}'_{jt} = \left[\left(Euler - \ln\Lambda\left(\boldsymbol{x}'_t\boldsymbol{\gamma}^{P}_{j}\right)\right)^{\ell} : \ell = 1,2,3\right] \tag{47}$$

This semiparametric approach to control for selection follows Newey (2009).

3. *Semiparametric mixture Logit.* The entry model is the mixture Logit with entry decision $a_{jt} = 1\{\eta_{jt} \leq \boldsymbol{x}'_{jt}\boldsymbol{\gamma}^{P}_{j\kappa^*}\}$ for unobserved market type κ^*, and with mixture distribution $\Pr(\kappa^*_t = \kappa^*) = f_{\kappa^*}(\kappa^*)$. Conditional on $\kappa^*_t = \kappa^*$, the selection term $\mathbb{E}\left(\xi_{jt} \mid a_{jt}=1, \boldsymbol{x}_t, \kappa^*_t = \kappa^*\right)$ is a third order polynomial in $Euler - \ln\Lambda\left(\boldsymbol{x}'_t\boldsymbol{\gamma}^{P}_{j\kappa^*}\right)$. Accordingly, the vector of regressors controlling for endogenous selection is:

$$\boldsymbol{h}'_{jt} = \left[f_{\kappa^*}(\kappa^*)\left(Euler - \ln\Lambda\left(\boldsymbol{x}'_t\boldsymbol{\gamma}^{P}_{j\kappa^*}\right)\right)^{\ell} : \ell = 1,2,3, \text{ and } \kappa^* = 1,2,...,|\mathcal{K}|\right] \tag{48}$$

For all the 2SLS estimators, we use as instrumental variables the number of competitors in the market and the average *hub-size* of the rest of the airlines (separately for origin and destination airports).

Table 4 presents the estimates of the demand parameters, while Table 5 provides the average demand elasticities and Lerner indexes derived from these estimates. Comparing the estimates obtained using OLS with those from various 2SLS methods — whether accounting for selection effects or not — we observe a significant adjustment in all parameter estimates when addressing the endogeneity of price and within-nest market share. This correction notably impacts the average own-price elasticity, shifting it from -1.59 to values below -5.54, and the corresponding Lerner index, which transitions from 69% to less than 20%.

In the context of this paper, the most significant findings arise from our investigation into the effects of controlling for the endogeneity of market entry. Remarkably, the most pronounced impacts materialize when we introduce finite mixture unobserved heterogeneity to address selection bias. Upon incorporating a finite mixture, parameters α and σ experience absolute value increases of more than 18% and 34%, respectively. This change translates to an absolute value increase exceeding 50% in the average own-price elasticities. Consequently, the corresponding average Lerner index shifts from approximately 19% to 15%. These effects hold substantial importance, carrying meaningful economic implications.

Notably, the estimates of demand parameters and their corresponding elasticities exhibit considerable robustness concerning the selection of the number of mixtures. We observe a modest increase in price sensitivity of demand as we transition from two to three unobserved market types. The main impact stems from the introduction of a finite mixture to address selection bias, with the number of mixtures contributing overall less significantly.

Figure 1 presents the empirical distributions of estimated own-price elasticities. Each row corresponds to an airline, while each column pertains to a different 2SLS estimator: the first column presents the estimator without controlling for selection, the second column illustrates the estimator that controls for selection using a sieve method but no mixture, and the third column presents the estimator with a three-type mixture.

The histograms in this figure are constructed based on estimates of elasticities at the airline-market-quarter level. The equation describing each elasticity solely depends on data on price p_{jmt}, market shares s_{jmt} and s_{0mt}, and parameter estimates $\widehat{\alpha}$ and $\widehat{\sigma}$. It is important to note that the data regarding prices and market shares remain constant across the various columns in the figure. Therefore, any shift in the distribution can be attributed solely to changes in the values of estimates $\widehat{\alpha}$ and $\widehat{\sigma}$.

Table 4: Estimation of Demand Parameters

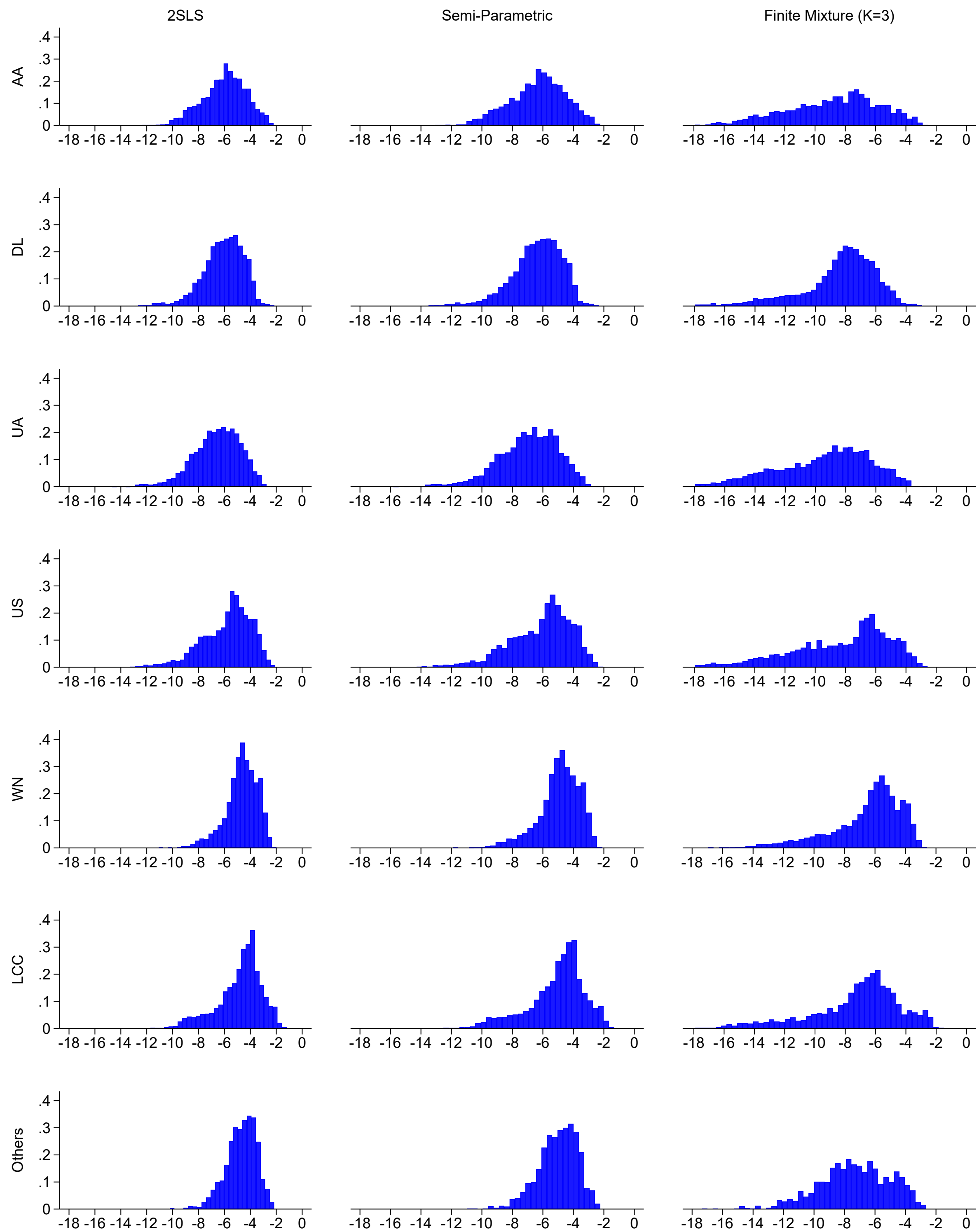
	Not control. for sel.		*Controlling for endogenous selection*			
	OLS	**2SLS**	**2SLS Heckman**	**2SLS Semi-P.**	**2SLS Fin-Mix** $\lvert\mathcal{K}\rvert = 2$	**2SLS Fin-Mix** $\lvert\mathcal{K}\rvert = 3$
Price (100$) ($\alpha$)	-0.643	-2.180	-2.193	-2.261	-2.574	-2.708
	(0.0105)	(0.1378)	(0.1348)	(0.1298)	(0.1549)	(0.1662)
Within Share (σ)	0.371	0.409	0.413	0.431	0.547	0.570
	(0.0058)	(0.0351)	(0.0389)	(0.0372)	(0.0509)	(0.0565)
Distance (1000mi)	0.729	2.130	2.196	2.264	2.497	2.472
	(0.0306)	(0.1372)	(0.1365)	(0.1310)	(0.1524)	(0.1572)
Distance^2	-0.216	-0.424	-0.453	-0.462	-0.496	-0.511
	(0.0112)	(0.0244)	(0.0252)	(0.0250)	(0.0276)	(0.0289)
hub-size orig. (100s)	1.637	2.272	1.999	1.320	1.593	1.383
	(0.0263)	(0.0382)	(0.0593)	(0.0625)	(0.0869)	(0.0989)
hub-size dest. (100s)	1.613	2.242	1.995	1.310	1.587	1.377
	(0.0267)	(0.0385)	(0.0595)	(0.0633)	(0.0872)	(0.0994)
Airline×Quarter FE	Y	Y	Y	Y	Y	Y
# control var. entry	0	0	6	18	36	54
Observations	35,763	35,763	35,763	35,763	35,763	35,763

Asymptotic standard errors account for estimation error in the first step using the method in Newey (2009).

Table 5: Average Own-Price Elasticities and Lerner Indexes

	Not control. for sel.		*Controllin for endogenous selection*			
	OLS	**2SLS**	**2SLS Heckman**	**2SLS Semi-P.**	**2SLS Fin-Mix** $\lvert\mathcal{K}\rvert = 2$	**2SLS Fin-Mix** $\lvert\mathcal{K}\rvert = 3$
Own-Price Elasticity	−1.596	−5.549	−5.601	−5.849	−7.406	−8.000
AA	−1.722	−6.013	−6.071	−6.363	−8.169	−8.857
DL	−1.761	−6.082	−6.133	−6.382	−7.871	−8.450
UA	−1.887	−6.573	−6.636	−6.936	−8.847	−9.573
US	−1.665	−5.801	−5.856	−6.122	−7.809	−8.450
WN	−1.354	−4.680	−4.719	−4.913	−6.068	−6.517
LCC	−1.370	−4.808	−4.857	−5.095	−6.674	−7.265
Others	−1.332	−4.705	−4.757	−5.006	−6.706	−7.337
Lerner Index	68.8%	19.9%	19.7%	18.9%	15.4%	14.4%
AA	62.7%	18.0%	17.9%	17.1%	13.8%	12.8%
DL	60.4%	17.5%	17.3%	16.7%	13.7%	12.8%
UA	56.9%	16.4%	16.2%	15.6%	12.6%	11.7%
US	65.9%	19.0%	18.9%	18.1%	14.8%	13.8%
WN	78.4%	22.8%	22.6%	21.8%	18.2%	17.1%
LCC	82.1%	23.5%	23.3%	22.2%	17.5%	16.3%
Others	79.2%	22.5%	22.3%	21.3%	16.4%	15.2%
Observations	35,763	35,763	35,763	35,763	35,763	35,763

Figure 1: Empirical Distribution of Estimated Elasticities (Airline-Market-Quarter level)



The histograms depicted in the first two columns are very similar. In contrast, the histograms based on the finite mixture estimates showcase significant alterations in both the location and the dispersion of the distribution of elasticities. Across all airlines, the incorporation of larger estimates for $\widehat{\alpha}$ and $\widehat{\sigma}$ using the mixture method leads to a noticeable leftward shift and an amplification in the spread of the histograms. These changes in the distributions' location and dispersion have important economic implications.

6.4 Estimation of costs and counterfactual experiments

In this paper, we focus on the consistent estimation of demand parameters in the presence of endogenous product entry. However, relying on the structure of our model, it is straightforward for researchers to estimate marginal costs, entry costs, and the joint distribution of unobservable variables. Given these estimated primitives, a variety of counterfactual experiments can be performed. In this subsection, we overview these supplementary estimation procedures within the framework of our empirical application.

6.4.1 Marginal costs

Based on an assumption about the nature of competition, such as Bertrand-Nash competition, we are able to estimate marginal costs at the airline-market-quarter level as the residuals from the pricing equation. It is important to note that these marginal costs can be computed only for those airlines that we observe being active in the market.

For some empirical questions, the researcher may need to estimate the marginal cost function: that is, the function that represents the causal effect of product characteristics and output on marginal costs. For this purpose, the researcher needs to estimate the parameters of a regression wherein the dependent variable is the marginal cost estimate and the the explanatory variables are the exogenous product characteristics $\boldsymbol{x}_{jt}$ and the output q_{jt}. A crucial consideration is that this regression is subject to selection bias due to endogenous product entry. Remarkably, the structure of the selection term in this equation mirrors that in the demand equation. We can then control for selection bias in the estimation of the marginal cost function using exactly the same control variables that we have used for the estimation of the demand parameters.

6.4.2 Demand and marginal cost unobservables

The consistent estimation of demand and marginal cost parameters inherently yields consistent estimates for the corresponding unobservable variables: ξ_{jmt} and ω_{jmt}. These unobservables are estimated as residuals from the estimated equations. While the estimation of these equations is subject to selection bias, the introduction of controls for selection enables us to achieve consistent estimation of the structural parameters and of the variables ξ_{jmt} and ω_{jmt} for the airlines active in the market.[10]

Naturally, the more complex estimation of the probability distribution governing these unobservables for all products, both those observed being active and inactive in the market, requires one to address the issue of endogenous selection.

6.4.3 Counterfactuals at the intensive margin

Once the challenge of endogenous selection has been addressed in the estimation of demand and marginal cost parameters, counterfactual experiments that hold constant the set of airlines and market structure can be performed without further complications.

6.4.4 Counterfactuals at the extensive margin

Another class of counterfactual experiments involves changes to the set of active firms and/or products within the market. In this category, the most straightforward experiment is the exogenous removal of certain products from the market. Given the availability of data on the exogenous demand and marginal cost attributes of all products, performing this type of counterfactual does not significantly differ from the *counterfactuals at the intensive margin* discussed above. This type of counterfactual includes as a particular case the evaluation of a merger which ignores firms' endogenous responses at the extensive margin.[11]

Counterfactual experiments that involve the introduction of new products require data on the exogenous attributes of the new or hypothetical products. In our empirical analysis of the airline industry, we observe $\boldsymbol{x}_{jmt}$ for every airline-market-quarter product, irrespective of whether the airline is active in the market. Specifically, data on the airline's hub-size at both the origin and destination airports, as well as the airline-quarter fixed effects, are available for both active airlines and potential entrants. However, the unobservable factors ξ_{jmt} and ω_{jmt} are unknown to the researcher for potential entrants. To perform this type of counterfactual, the researcher needs to determine the values of these unobservables also for the potential entrants.

In principle, the researcher could set values for ξ_{jmt} and ω_{jmt} for the potential entrants at the unconditional mean of these variables, which is zero. However, this approach raises a significant concern: it contradicts the fact that these airlines opted not to enter in this particular market.

[10] Importantly, in calculating $\widehat{\xi}_{jmt}$ and $\widehat{\omega}_{jmt}$, one should not remove the estimated selection term from these residuals.

[11] In this class of models, the evaluation of the effects of a counterfactual merger requires making an assumption about the values of exogenous product characteristics for the new merging entity/firm. However, this complication is present regardless of the endogenous product selection issue that we address in this paper.

To establish values of ξ_{jmt} and ω_{jmt} that align with observed endogenous entry decisions, one must consider $\mathbb{E}\left(\xi_{jmt}|\boldsymbol{x}_{mt}, a_{jmt}=0\right)$ and $\mathbb{E}\left(\omega_{jmt}|\boldsymbol{x}_{mt}, a_{jmt}=0\right)$ respectively.

While our estimation method yields consistent semiparametric estimates of the expected values $\mathbb{E}(\xi_{jmt}|\boldsymbol{x}_{mt}, a_{jmt}=1)$ and $\mathbb{E}(\omega_{jmt}|\boldsymbol{x}_{mt}, a_{jmt}=1)$, it is silent with respect to $\mathbb{E}(\xi_{jmt}|\boldsymbol{x}_{mt}, a_{jmt}=0)$ and $\mathbb{E}(\omega_{jmt}|\boldsymbol{x}_{mt}, a_{jmt}=0)$. Achieving point identification for the latter requires supplementary constraints, such as parametric assumptions or symmetry restrictions. An alternative approach instead involves estimating semiparametric bounds for these expected values. This information can then be used to select appropriate values for ξ_{jmt} and ω_{jmt}.

7 Conclusions

In local geographic markets, we typically find only a subset of all the differentiated products available in an industry. Firms strategically select specific products that better match the preferences of local consumers. When making market entry decisions, firms possess information about the demand for their products, particularly regarding unobservable demand components. Firms tend to enter markets with higher expected demand. Neglecting this selection process can introduce significant biases in the estimation of demand parameters. This issue is common across various demand applications and industries. Existing methods to address this issue typically rely on strong parametric assumptions about demand unobservables and firms' information.

In this paper, we investigate the identification of demand parameters within a structural model that encompasses demand, price competition, and market entry (static or dynamic), while specifying the distribution of demand unobservables in a nonparametric finite mixture manner. The paper makes three main contributions. First, it establishes sequential identification of the demand parameters in this model. We demonstrate that the selection term in the demand equation results from a convolution of the probabilities of product entry for each discrete unobserved market type and the densities associated with these market types. We show that data on firms' product entry decisions nonparametrically identify the probabilities of product entry conditional on the market type and the density of unobserved market types. Under mild conditions on the observable variables, demand parameters are identified after controlling for the nonparametric entry probabilities and densities for each market type.

Second, we propose a simple two-step estimator to address endogenous selection. In the first step, we estimate a nonparametric finite mixture model to determine the choice probabilities of product entry. In the second step, demand parameters are estimated using a Generalized Method of Moments (GMM) approach that accounts for both endogenous product availability and price endogeneity.

Third, we illustrate the proposed method by applying it to data from the airline industry. The findings highlight the importance of allowing for a finite mixture of unobserved market types when controlling for endogenous product entry, as failure to do so can lead to significant biases.

References

Abaluck, J., and A. Adams-Prassl (2021): "What do consumers consider before they choose? Identification from asymmetric demand responses," *The Quarterly Journal of Economics*, 136(3), 1611–1663.

Aguirregabiria, V., A. Collard-Wexler, and S. Ryan (2021): "Dynamic games in empirical industrial organization," in *Handbook of Industrial Organization, Volume 4*, ed. by K. Ho, A. Hortaçsu, and A. Lizzeri, pp. 225–343. Elsevier.

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. Powell (1993): "Semiparametric estimation of censored selection models with a nonparametric selection mechanism," *Journal of Econometrics*, 58(1-2), 3–29.

Allman, E., C. Matias, and J. 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*, 41(6), 997–1016.

——— (1974): "Multivariate regression and simultaneous equation models when the dependent variables are truncated normal," *Econometrica*, 42(6), 999–1012.

Angrist, J. D. (1997): "Conditional independence in sample selection models," *Economics Letters*, 54(2), 103–112.

Aradillas-Lopez, A. (2012): "Pairwise-difference estimation of incomplete information games," *Journal of Econometrics*, 168(1), 120–140.

Aradillas-Lopez, A., B. Honoré, and J. 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.

Barseghyan, L., M. Coughlin, F. Molinari, and J. Teitelbaum (2021): “Heterogeneous choice sets and preferences,” *Econometrica*, 89(5), 2015–2048.

Berry, S. (1992): “Estimation of a Model of Entry in the Airline Industry,” *Econometrica*, 60(4), 889–917.

——— (1994): “Estimating discrete-choice models of product differentiation,” *The RAND Journal of Economics*, 25(2), 242–262.

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. Haile (2014): “Identification in differentiated products markets using market level data,” *Econometrica*, 82(5), 1749–1797.

——— (2021): “Foundations of demand estimation,” in *Handbook of Industrial Organization, Volume 4*, ed. by K. Ho, A. Hortaçsu, and A. Lizzeri, pp. 1–62. Elsevier.

——— (2022): “Nonparametric Identification of Differentiated Products Demand Using Micro Data,” *arXiv working paper*, 2204.06637.

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*, 63(4), 841–890.

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.

Bontemps, C., C. Gualdani, and K. Remmy (2023): “Price competition and endogenous product choice in networks: Evidence from the us airline industry,” Discussion paper, Working Paper.

Bresnahan, T., and P. Reiss (1991): “Entry and competition in concentrated markets,” *Journal of Political Economy*, 99(5), 977–1009.

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

Bunting, J. (2022): “Continuous permanent unobserved heterogeneity in dynamic discrete choice models,” *arXiv preprint arXiv:2202.03960*.

Bunting, J., P. Diegert, and A. Maurel (2022): “Heterogeneity, Uncertainty and Learning: Semiparametric Identification and Estimation,” *Working Paper*.

Cardell, S. (1997): "Variance Components Structures for the Extreme-Value and Logistic Distributions with Application to Models of Heterogeneity," *Econometric Theory*, 13(2), 185–213.

Carroll, J. D., and J.-J. Chang (1970): "Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition," *Psychometrika*, 35(3), 283–319.

Chen, X. (2007): "Large sample sieve estimation of semi-nonparametric models," *Handbook of Econometrics*, 6, 5549–5632.

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.

Conlon, C., and J. Mortimer (2013): "Demand estimation under incomplete product availability," *American Economic Journal: Microeconomics*, 5(4), 1–30.

Crawford, G., R. Griffith, and A. Iaria (2021): "A survey of preference estimation with unobserved choice set heterogeneity," *Journal of Econometrics*, 222(1), 4–43.

Das, M., W. 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.

Draganska, M., M. Mazzeo, and K. Seim (2009): "Beyond plain vanilla: Modeling joint product assortment and pricing decisions," *Quantitative Marketing and Economics*, 7, 105–146.

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.

Dunson, D. B., and C. Xing (2009): "Nonparametric Bayes modeling of multivariate categorical data," *Journal of the American Statistical Association*, 104(487), 1042–1051.

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

Gandhi, A., Z. Lu, and X. Shi (2023): "Estimating demand for differentiated products with zeroes in market share data," *Quantitative Economics*, 14(2), 381–418.

Gandhi, A., and A. Nevo (2021): "Empirical models of demand and supply in differentiated products industries," in *Handbook of Industrial Organization, Volume 4*, ed. by K. Ho, A. Hortaçsu, and A. Lizzeri, pp. 63–139. Elsevier.

GOEREE, M. S. (2008): "Limited information and advertising in the US personal computer industry," *Econometrica*, 76(5), 1017–1074.

GRIECO, P. (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.

HARSHMAN, R. A., ET AL. (1970): "Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multi-modal factor analysis," *UCLA working papers in phonetics*, 16(1), 84.

HAVILAND, A., AND D. NAGIN (2005): "Causal inferences with group based trajectory models," *Psychometrika*, 70(3), 557–578.

HAVILAND, A., D. NAGIN, P. ROSENBAUM, AND R. TREMBLAY (2008): "Combining group-based trajectory modeling and propensity score matching for causal inferences in nonexperimental longitudinal data.," *Developmental Psychology*, 44(2), 422.

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

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.

HU, Y., AND Y. XIN (2022): "Identification and estimation of dynamic structural models with unobserved choices," *Available at SSRN 3634910.*

JIA, P. (2008): "What happens when Wal-Mart comes to town: An empirical analysis of the discount retailing industry," *Econometrica*, 76(6), 1263–1316.

KARGAS, N., N. D. SIDIROPOULOS, AND X. FU (2018): "Tensors, learning, and "Kolmogorov extension" for finite-alphabet random vectors," *IEEE Transactions on Signal Processing*, 66(18), 4854–4868.

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.

KOLDA, T. G., AND B. W. BADER (2009): "Tensor decompositions and applications," *SIAM review*, 51(3), 455–500.

Lanza, S., D. Coffman, and S. Xu (2013): "Causal inference in latent class analysis," *Structural Equation Modeling: A Multidisciplinary Journal*, 20(3), 361–383.

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.

Lu, Z. (2022): "Estimating multinomial choice models with unobserved choice sets," *Journal of Econometrics*, 226(2), 368–398.

Moraga-González, J., Z. Sándor, and M. Wildenbeest (2023): "Consumer search and prices in the automobile market," *The Review of Economic Studies*, 90(3), 1394–1440.

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

Newey, W., J. Powell, and J. 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.

Pilla, R. S., and B. G. Lindsay (2001): "Alternative EM methods for nonparametric finite mixture models," *Biometrika*, 88(2), 535–550.

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

Rothenberg, T. (1971): "Identification in parametric models," *Econometrica*, 39(3), 577–591.

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

Sidiropoulos, N. D., L. De Lathauwer, X. Fu, K. Huang, E. E. Papalexakis, and C. Faloutsos (2017): "Tensor decomposition for signal processing and machine learning," *IEEE Transactions on signal processing*, 65(13), 3551–3582.

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*, 26(1), 24–36.

Williams, B. (2020): "Nonparametric identification of discrete choice models with lagged dependent variables," *Journal of Econometrics*, 215(1), 286–304.

Wooldridge, J. M. (2010): *Econometric analysis of cross section and panel data*. MIT press.

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

Yang, Y., and D. B. Dunson (2016): "Bayesian conditional tensor factorizations for high-dimensional classification," *Journal of the American Statistical Association*, 111(514), 656–669.

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