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

Victor Aguirregabiria † Alessandro Iaria ‡ Senay Sokullu §

October 29, 2025

Abstract

Firms are more likely to introduce products in markets where they anticipate stronger demand. They also possess information that is unobserved to researchers. This creates endogenous selection bias in the estimation of demand parameters. With differentiated products, the entry decision violates the monotonicity conditions required for standard selection-correction methods to yield consistent demand estimates. Existing studies address this issue either by imposing strong assumptions about firms' information on demand at the time of entry or by jointly estimating a full equilibrium model of demand, pricing, and entry. Both strategies make the estimation of demand heavily reliant on supply-side assumptions. We propose a new semiparametric estimation method that addresses these limitations. Our approach exploits the correlation across products in their market-entry decisions to identify entry probabilities conditional not only on observable characteristics but also on latent variables that capture unobserved interdependencies among firms' entry choices. We refer to these probabilities as latent propensity scores. We show that the selection bias term in the demand equation is a convolution of these latent propensity scores and is therefore identifiable. Building on this result, we develop a two-step semiparametric estimator in the spirit of standard sample-selection correction methods. Applying our method to data from the airline industry, we find that conventional approaches to correcting for selection bias substantially underestimate price elasticities of demand.

Keywords: Demand for differentiated product; Product entry; Selection bias; Airline markets.

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

*We are grateful for helpful comments from Roy Allen, Gaurab Aryal, Giovanni Compiani, Andreea Enache, Christos Genakos, Ying Fan, Michele Fioretti, Philip Haile, Nail Kashaev, Mathieu Marcoux, Mateusz Mysliwski, Anders Munk-Nielsen, David Pacini, Bertel Schjerning, Philipp Schmidt-Dengler, Jesse Shapiro, Yutec Sun, Yuanyuan Wan, Ao Wang, and Christine Zulehner; as well as from seminar participants at the Universities of Bolzano, Copenhagen, ENSAI-Rennes, Glasgow, Helsinki GSE, Cambridge (Judge), Mannheim, NHH-Bergen, Penn State, Rochester, Sciences Po, TSE, and Vienna; and from participants at the Advances in Demand Analysis Workshop (2024), BSE Summer Forum on Structural Microeconometrics (2024), CEPR IO Meeting (2024), Cowles Conference on Models & Measurement (2024), IAAE (2023), MaCCI Summer Institute (2022), and Midwest Econometrics Group Meeting (2023).

†University of Toronto and CEPR. victor.aguirregabiria@utoronto.ca.

‡University of Bristol and CEPR. alessandro.iaria@bristol.ac.uk.

§University of Bristol. senay.sokullu@bristol.ac.uk.

1 Introduction

Estimating demand systems for differentiated products typically relies on data spanning multiple geographic markets and time periods. In these settings, it is common for some products to be unavailable in certain markets or at particular times. Firms tend to introduce products in markets where they anticipate stronger demand, drawing on information about market conditions that is unobservable to researchers. As a result, the observed pattern of product availability is not random but reflects firms' private expectations about demand. This endogenous selection into markets can generate substantial bias in the estimation of demand parameters in regression-based models. This issue is prevalent across various industries, including airlines (Berry et al., 2006; Berry and Jia, 2010; Aguirregabiria and Ho, 2012), supermarket chains (Smith, 2004), radio stations (Sweeting, 2013), personal computers (Eizenberg, 2014), and ice cream (Draganska et al., 2009).

The selection problem in this structural model of demand and product entry exhibits a distinctive feature that sets it apart from more conventional cases. Specifically, the demand unobservables are multi-dimensional and have a non-additive effect on firms' expected profits. This breaks a key monotonicity condition typically required for the selection equation. Without this condition, the selection propensity score – the probability of product entry given exogenous observables – cannot serve as a sufficient statistic to control for selection bias in the estimation of demand parameters (Angrist, 1997). Furthermore, the model involves multiple equilibria in both the entry and pricing games. The possibility that different equilibria are selected across markets introduces additional non-monotonicity in the selection equation. As a result, standard identification results and two-step estimation methods that rely on the propensity score are not applicable in this context (e.g., Ahn and Powell, 1993; Das et al., 2003; Aradillas-Lopez et al., 2007; Newey, 2009).[1]

[1] Importantly, instrumental variable approaches cannot address this form of selection bias. Consistent estimation typically requires control-function methods that explicitly model the selection process. See Vella (1998), Heckman and Navarro-Lozano (2004), Wooldridge (2015).

The growing interest in estimating models of oligopoly competition that endogenize firms' product-entry decisions across geographic markets has made the associated selection problem increasingly salient. The standard approach in this literature begins with the estimation of a demand system. However, in the absence of instrumental-variable or control-function methods to address selection bias, these studies typically impose strong assumptions about firms' information sets at the time of entry. Such assumptions effectively rule out endogenous product selection based on unobserved demand shocks. Examples include Aguirregabiria and Ho (2012), Fan (2013), Sweeting (2013), Eizenberg (2014), and Fan and Yang (2020). Motivated by the importance of this issue, Ciliberto et al. (2021) and Li et al. (2022) develop methods that jointly estimate the full structural model of demand, price competition, and product entry. Although these approaches fully account for selection bias in demand estimation, they make demand identification heavily dependent on supply-side assumptions—such as the nature of competition, the functional form of cost functions, and the distributional assumptions on unobservables.

The main contribution of this paper is to establish new, more general conditions for the sequential (two-step) identification of demand parameters when product entry is endogenous. Our approach leverages the cross-product correlation in firms' market-entry decisions to recover entry probabilities that are conditioned not only on observable characteristics but also on latent variables capturing unobserved interdependencies among firms' choices. We refer to these as *latent propensity scores*. These probabilities are constructed by integrating over the distribution of unobservables that satisfy a monotonicity condition, while conditioning on those that violate it.

Our identification result proceeds in two steps. First, we establish the nonparametric identification of the latent propensity scores. This step exploits a key feature of the model: the unobservables that violate monotonicity in a product's entry decision are precisely the demand shocks of other products that could potentially enter the market. These unobservables generate the interdependence among firms' entry decisions. Consequently, the joint distribution of

entry decisions follows a mixture model structure, where the unobservables driving this interdependence act as the mixing variables. Second, we show that the selection-bias term in the demand equation can be expressed as a convolution of these latent propensity scores, and is therefore identifiable.

Building on our constructive proof of identification, we propose a transparent and computationally simple two-step estimator that jointly corrects for endogenous product selection and price endogeneity in demand estimation. In the first step, we estimate each product's latent propensity score using a semiparametric mixture model that captures unobserved interdependencies in firms' entry decisions. In the second step, we recover the demand parameters through a Control Function–GMM procedure that accounts for both endogenous product availability and price endogeneity. This approach yields consistent estimates under minimal assumptions about firms' information, the structure of competition, and the functional forms on the supply side.

We illustrate the proposed method using data from the airline industry. The results demonstrate the importance of accounting for endogenous product entry when estimating demand parameters and highlight the limitations of conventional selection-correction approaches. Specifically, standard methods that impose strong informational or structural restrictions substantially underestimate price elasticities of demand. We also uncover significant selection bias in the estimation of marginal costs derived from Bertrand pricing equations. Moreover, our reduced-form estimation of entry probabilities—capturing rich correlations in firms' entry decisions—provides economically meaningful insights. In particular, we find that models that ignore or restrict correlated unobservables in market-entry decisions tend to overstate the degree of market contestability, predicting a higher likelihood of new entry following mergers than what is supported by the data.

Our paper contributes to the literature on sample selection bias in demand estimation when zeros arise from firms' market entry decisions, including the seminal works of Draganska et al. (2009), Conlon and Mortimer (2013), Ciliberto et al. (2021), and Li et al. (2022). These

studies develop methods for estimating structural models that integrate differentiated-product demand systems à la Berry et al. (1995) with market or product entry games following Bresnahan and Reiss (1990, 1991) and Berry (1992). Their approach involves joint estimation of demand, marginal cost, and entry-cost parameters using nested fixed-point algorithms. While powerful, these methods rely on strong parametric assumptions about functional forms and the distribution of unobservables. In contrast, our paper proposes a sequential estimation strategy that identifies the demand parameters without imposing specific assumptions about the supply side. This approach ensures robustness to a wide range of supply-side structures and greatly simplifies computation by avoiding the need to solve for equilibrium outcomes. Moreover, the framework and its computational advantages extend naturally to both static and dynamic games of market entry and exit.[2]

Our approach contributes to the growing literature on structural models of oligopoly competition that endogenize firms' product entry decisions while explicitly incorporating demand systems for differentiated products. Contributions in this line of research include Aguirregabiria and Ho (2012), Fan (2013), Sweeting (2013), Eizenberg (2014), Schaumans and Verboven (2015), Fan and Yang (2020), Bontemps et al. (2023), Caoui and Steck (2023), and Liu and Luo (2025). These studies estimate structural parameters through a sequential approach that begins with the estimation of the demand system. To address potential selection bias from endogenous product entry, they impose restrictive assumptions about firms' information sets—specifically, that firms lack information about unobserved components of demand when making entry decisions. These assumptions effectively rule out selection on unobservables and simplify identification, but at the cost of misspecification biases. In contrast, we relax this restriction, allowing firms to possess information about demand shocks at the time of

[2]Given the estimated demand parameters and unobservables from our method, one can subsequently recover marginal and entry costs under weaker parametric assumptions than those required in joint structural estimation. As in Ciliberto et al. (2021) and Li et al. (2022), our estimates can be used to conduct a variety of counterfactual experiments that account for the endogeneity of product entry and exit—an essential feature when simulating merger effects, as demonstrated by Li et al. (2022). Section 6.4 provides details on the implementation of these counterfactuals.

entry. This not only addresses selection bias in demand estimation but also corrects the misspecification it induces in the entry game, where firms' entry choices are endogenously correlated through shared information about demand fundamentals.

Our estimation method contributes to the literature on semiparametric estimation of sample selection models (see, e.g., Das et al. (2003); Newey (2009); Powell (2001); Aradillas-Lopez et al. (2007)). We extend two-step propensity-score control function approaches to settings where the unobservables in the selection equation violate the standard monotonicity condition. Specifically, when the selection decision arises within a system of simultaneous selection equations, and the non-monotonic unobservables are those generating dependence across selection decisions, we show that it is still possible to identify a control function that corrects for selection bias. As fat as we know, this is a novel result in this literature. Our approach can be applied to other sample selection problems that share this structural feature, such as labor market models with two-sided matching (Choo and Siow (2006), Galichon and Salanié (2022)), models of joint household decisions (Browning et al. (2014)), or peer effects models with endogenous network formation (Graham (2017), De Paula et al. (2018)).

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 framework follows the canonical model of demand and oligopoly competition in differentiated product markets in industrial organization. We outline it here to define notation and highlight the main assumptions. Proposition 1 derives a simple property of this model that is fundamental to understanding the form of the selection bias examined in the paper.

The demand system follows the BLP framework (Berry et al., 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.[3] Demand and price 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:

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

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 $v(p_{jt}, \boldsymbol{x}_{jt}, v_{ht}) + \varepsilon_{hjt}$ represents a household-specific deviation from the average utility. The term $v(p_{jt}, \boldsymbol{x}_{jt}, v_{ht})$ depends on the vector of random coefficients v_{ht} with distribution $F_v(\cdot|\boldsymbol{\sigma})$, 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.

[3]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 et al. (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.

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. Similarly, the component of utility that depends on consumer-level random coefficients takes the multiplicative form:

$$v(p_{jt}, \boldsymbol{x}_{jt}, v_{ht}) \; = \; \left(p_{jt}, \; \boldsymbol{x}_{jt}\right)' \; \boldsymbol{\Omega}_{\sigma} \; \boldsymbol{v}_{ht} \tag{3}$$

where $\boldsymbol{\Omega}_{\sigma}$ is a $(K+1) \times (K+1)$ matrix that is a known, continuously differentiable function of the parameter vector $\boldsymbol{\sigma}$, and $\boldsymbol{v}_{ht}$ is a vector of random variables with a known distribution. The outside option is represented by $j = 0$ and its indirect utility is normalized to $U_{h0t} = \varepsilon_{h0t}$. We denote by $\boldsymbol{\theta}_d$ the column vector of demand parameters, $\boldsymbol{\theta}_d = (\alpha, \; \boldsymbol{\beta}', \; \boldsymbol{\sigma}')'$.

Let $a_{jt} \in 0, 1$ denote the indicator that product j is offered in market t, and define $\boldsymbol{a}_t \equiv (a_{jt} : j \in \mathcal{J})$ as the vector collecting the offer indicators for all products in market t. The outside option $j = 0$ is always offered 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:

$$s_{jt} = d_{jt}(\boldsymbol{\delta}_t, \; \boldsymbol{a}_t, \; \boldsymbol{\sigma}) \equiv \int \frac{a_{jt} \; \exp\left(\delta_{jt} + \left[p_{jt}, \; \boldsymbol{x}_{jt}\right]' \; \boldsymbol{\Omega}_{\sigma} \; \boldsymbol{v}\right)}{1 + \sum_{i=1}^{J} a_{it} \; \exp\left(\delta_{it} + \left[p_{it}, \; \boldsymbol{x}_{it}\right]' \; \boldsymbol{\Omega}_{\sigma} \; \boldsymbol{v}\right)} \, dF_v(v). \tag{4}$$

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

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 offered. Define $\mathcal{J}_t^{\boldsymbol{a}} \equiv \{j \in \mathcal{J} : a_{jt} = 1\}$, $\boldsymbol{s}_t^{\boldsymbol{a}} = (s_{jt} : j \in \mathcal{J}_t^{\boldsymbol{a}})$, and $\boldsymbol{\delta}_t^{\boldsymbol{a}} = (\delta_{jt} : j \in \mathcal{J}_t^{\boldsymbol{a}})$. Then, we

represent this system as:

$$\boldsymbol{s}_t^{\boldsymbol{a}} = \boldsymbol{d}_t^{\boldsymbol{a}}(\delta_t^{\boldsymbol{a}}, \boldsymbol{\sigma}), \tag{5}$$

Proposition 1 establishes that, for any configuration of $\boldsymbol{a}$, the demand system (5) satisfies the invertibility property of Berry (1994), and that the resulting inverse system depends on all demand parameters, rendering them identifiable.

PROPOSITION 1. *Suppose that the outside option $j = 0$ is always offered. Then, for any value of the vector $\boldsymbol{a} \in \{0,1\}^J$:*

a. The demand system in equation (5) is invertible in $\delta_t^{\boldsymbol{a}}$ such that for every product with $a_{jt} = 1$ the inverse function $\delta_{jt} = d_{jt}^{-1}(\boldsymbol{s}_t^{\boldsymbol{a}}, \boldsymbol{\sigma})$ exists.

b. The inverse function $d_{jt}^{-1}(\boldsymbol{s}_t^{\boldsymbol{a}}, \boldsymbol{\sigma})$ depends on all the parameters in the vector $\boldsymbol{\sigma}$, and matrix $\mathbb{E}\left(\frac{\partial d_{jt}^{-1}(\boldsymbol{s}_t^{\boldsymbol{a}}, \sigma)}{\partial \sigma} \frac{\partial d_{jt}^{-1}(\boldsymbol{s}_t^{\boldsymbol{a}}, \sigma)}{\partial \sigma'}\right)$ is full rank such that $\boldsymbol{\sigma}$ is identifiable. ■

Proof of Proposition 1: See Appendix A.1.

For a product offered in market t, we have:

$$d_{jt}^{-1}(\boldsymbol{s}_t^{\boldsymbol{a}}, \boldsymbol{\sigma}) \;=\; \alpha\, p_{jt} + \boldsymbol{x}_{jt}'\,\boldsymbol{\beta} + \xi_{jt} \quad \text{if and only if} \quad a_{jt} = 1. \tag{6}$$

Importantly, after applying Berry's inversion, the selection condition for the existence of the regression equation for a product–market observation (j,t) depends only on product j (and the outside option 0) being offered in market t, and not on which other products are offered in that market. Consequently, the selection bias in estimating the demand for product j can be expressed in terms of the following conditional expectation:

$$\mathbb{E}\left(\xi_{jt} \mid a_{jt} = 1\right). \tag{7}$$

This characterization of the selection term is an implication of working directly with the inverse demand system, as represented by equation (6).[4]

As discussed in section 2.4, Proposition 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 $\mathbb{E}\left(\xi_{jt} \mid a_{jt} = 1\right)$ even if the firm owns other products. The following Example illustrates Proposition 1 in the case of a nested logit model.

EXAMPLE 1 (Nested logit model). The J products are partitioned into $R+1$ mutually exclusive groups indexed by $r \in \{0, 1, ..., R\}$. We denote by r_j the group to which product j belongs. The outside good is the single element of group $r = 0$. 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:

$$s_{jt} \;=\; d_j\left(\boldsymbol{\delta}_t^{\boldsymbol{a}}, \sigma\right) \;=\; d_{r_j}\left(\boldsymbol{\delta}_t^{\boldsymbol{a}}, \sigma\right) \cdot d_{j|r_j}\left(\boldsymbol{\delta}_t^{\boldsymbol{a}}\right) \tag{8}$$

with:

$$d_{r_j}\left(\boldsymbol{\delta}_t^{\boldsymbol{a}}, \sigma\right) \;=\; \frac{a_{jt}\; e^{\delta_{jt}}}{\sum_{i \in r_j} a_{it}\; e^{\delta_{it}}} \quad \text{and} \quad d_{j|r_j}\left(\boldsymbol{\delta}_t^{\boldsymbol{a}}\right) \;=\; \frac{\left[\sum_{i \in r_j} a_{it}\; e^{\delta_{it}}\right]^{\frac{1}{1-\sigma}}}{\sum_{r=0}^{R}\left[\sum_{i \in r} a_{it}\; e^{\delta_{it}}\right]^{\frac{1}{1-\sigma}}}. \tag{9}$$

If $a_{0t} = 1$ and $a_{jt} = 1$, this model implies that $s_{0t} > 0$ and $s_{jt} > 0$, and the inverse function $d_{jt}^{-1}\left(\boldsymbol{s}_t^{\boldsymbol{a}}, \boldsymbol{\sigma}\right)$ exists regardless of the value of a_{it} for any product i different from j. It is

[4] 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 offered 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.

straightforward to show that this inverse function has the following form:

$$d_{jt}^{-1}\left(\boldsymbol{s}_t^{\boldsymbol{a}},\, \boldsymbol{\sigma}\right) \;=\; \ln\left(\frac{s_{jt}}{s_{0t}}\right) - \sigma\, \ln\left(\frac{\sum_{i\in r_j} s_{it}}{s_{0t}}\right), \tag{10}$$

and it implies the regression equation:

$$\ln\left(\frac{s_{jt}}{s_{0t}}\right) = \sigma\, \ln\left(\frac{\sum_{i\in r_j} s_{it}}{s_{0t}}\right) + \alpha\, p_{jt} + \boldsymbol{x}_{jt}'\, \boldsymbol{\beta} + \xi_{jt}. \tag{11}$$

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:

$$\Pi_{jt} \;=\; p_{jt}\, q_{jt} \;-\; c(q_{jt},\, \boldsymbol{x}_{jt},\, \omega_{jt}) \;-\; f_{jt}, \tag{12}$$

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_{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:

$$p_{jt} \;=\; mc_{jt} \;-\; d_{jt}\left(\boldsymbol{\delta}_t^{\boldsymbol{a}}, \sigma\right) \left[\frac{\partial d_{jt}\left(\boldsymbol{\delta}_t^{\boldsymbol{a}}, \sigma\right)}{\partial p_{jt}}\right]^{-1} \text{ for every } j \in \mathcal{J}_t^{\boldsymbol{a}}, \tag{13}$$

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.

Let $\boldsymbol{x}_t \equiv (\boldsymbol{x}_{jt} : j \in \mathcal{J})$ denote the vector of exogenous variables observed by the researcher that affect demand or costs, with support $\mathcal{X}$ (each element of which may be continuous or

discrete). The vectors $\boldsymbol{\xi}_t$ and $\boldsymbol{\omega}_t$ are defined analogously. Let $\boldsymbol{a}_{-jt}$ denote the vector of entry decisions for all firms other than j. We define the function

$$VP_{jt} \;=\; VP_j\left(\boldsymbol{a}_{-jt},\, \boldsymbol{x}_t,\, \boldsymbol{\xi}_t,\, \boldsymbol{\omega}_t\right) \tag{14}$$

as firm j's *indirect variable profit*, obtained by substituting into the expression $p_{jt}\, q_{jt} - c(q_{jt}; \boldsymbol{x}_{jt}, \omega_{jt})$ the equilibrium values of prices and quantities from the Nash-Bertrand equilibrium given $(a_{jt} = 1,\, \boldsymbol{a}_{-jt},\, \boldsymbol{x}_t,\, \boldsymbol{\xi}_t,\, \boldsymbol{\omega}_t)$.[5]

2.3 Market entry game

This section introduces a model of product entry that encompasses a broad class of games studied in the literature. It nests complete-information frameworks such as those in Ciliberto and Tamer (2009) and Ciliberto et al. (2021), as well as incomplete-information settings with common knowledge unobservables, as in Grieco (2014) and Aguirregabiria and Mira (2019). The model also allows for flexible information structures regarding firms' knowledge of demand shocks at the time of entry, ranging from cases with full information to those with complete uncertainty, and including intermediate scenarios with imperfect signals. This general formulation ensures that the identification results developed in this paper apply to a wide spectrum of market entry environments.

Firms' entry decisions arise as the equilibrium outcome of this game. The payoff from remaining inactive is normalized to zero. Prior to making their entry decisions, firms may face uncertainty about their potential profits if active in the market. Their information about demand and cost fundamentals is therefore central to the entry process, as it shapes both individual incentives and the joint distribution of firms' equilibrium entry decisions.

[5]The pricing game may admit multiple equilibria. We do not impose any restriction on equilibrium selection and allow each market to select its own equilibrium. For notational simplicity, we do not explicitly include an unobservable variable —say τ_t— to index the equilibrium selected in the Bertrand game, although it can be interpreted as part of the broader vector of unobservables.

Assumption 1 summarizes our conditions on the information structure and the unobservables to the researcher.[6]

ASSUMPTION 1. *At the time firm j makes its entry decision in market t, its information set consists of $(\boldsymbol{x}_t, \boldsymbol{\kappa}_t, \boldsymbol{\eta}_{jt})$.*

a. The vector $\boldsymbol{x}_t$ of variables observable to the researcher is also common knowledge among all firms.

b. The vector $\boldsymbol{\kappa}_t$ represents all information about demand and cost fundamentals $(\boldsymbol{\xi}_t, \boldsymbol{\omega}_t)$ and fixed costs that is common knowledge among firms but unobserved by the researcher. In one possible scenario, $\boldsymbol{\kappa}_t$ may include the entire vector $(\boldsymbol{\xi}_t, \boldsymbol{\omega}_t)$, implying that firms face no uncertainty about demand or variable costs at the time of entry.

c. The vector $\boldsymbol{\eta}_{jt}$ represents firm j's private information about its own demand or cost fundamentals. The vectors $\boldsymbol{\eta}_{jt}$ are assumed to be independently distributed across firms and independent of $(\boldsymbol{\kappa}_t, \boldsymbol{x}_t)$. As a special case, variable $\boldsymbol{\eta}_{jt}$ may have a degenerate distribution, in which case the entry game reduces to one of complete information.

d. All the unobservables for the researcher, $(\boldsymbol{\xi}_t, \boldsymbol{\omega}_t, \boldsymbol{\kappa}_t, \boldsymbol{\eta}_{jt})$, are assumed independent of the exogenous observables in $\boldsymbol{x}_t$. ■

For simplicity, and with some abuse of notation, for the rest of the paper we represent the vector of unobservables $(\boldsymbol{\xi}_t, \boldsymbol{\omega}_t)$ using the more compact notation $\boldsymbol{\xi}_t$.

Let $\pi_j(\boldsymbol{a}_{-j}, \boldsymbol{x}_t, \boldsymbol{\kappa}_t, \eta_{jt})$ 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:

$$\pi_j(\boldsymbol{a}_{-j}, \boldsymbol{x}_t, \boldsymbol{\kappa}_t, \eta_{jt}) \;=\; \int VP_j(\boldsymbol{a}_{-j}, \boldsymbol{x}_t, \boldsymbol{\xi}_t)\, dF_{j,\xi}\left(\boldsymbol{\xi}_t \mid \boldsymbol{\kappa}_t, \eta_{jt}\right) - fc(\boldsymbol{x}_{jt}, \boldsymbol{\kappa}_t, \eta_{jt}), \tag{15}$$

[6] The entry game has multiple equilibria. We adopt the same general approach as for the pricing game. We do not impose any restrictions on equilibrium selection, but for notational simplicity, we do not explicitly include an unobservable variable to index the selected equilibrium. It can be interpreted that vector $\boldsymbol{\kappa}_t$ includes the equilibrium selection index.

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

Assumption 1 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 $(\boldsymbol{x}_t, \boldsymbol{\kappa}_t)$, 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.

$$\pi_j^P(\boldsymbol{x}_t, \boldsymbol{\kappa}_t, \eta_{jt}) = \sum_{\boldsymbol{a}_{-j} \in \{0,1\}^{J-1}} \left(\prod_{i \neq j} [P_{it}]^{a_i} \left[1 - P_{it}\right]^{1-a_i} \right) \pi_j(\boldsymbol{a}_{-j}, \boldsymbol{x}_t, \boldsymbol{\kappa}_t, \eta_{jt}). \tag{16}$$

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

$$P_{jt} \;=\; \int 1\{\pi_j^P(\boldsymbol{x}_t, \boldsymbol{\kappa}_t, \eta_{jt}) \geq 0\}\, dF_\eta\left(\eta_{jt}\right). \qquad \blacksquare \tag{17}$$

This framework accommodates various information structures corresponding to different scenarios considered by the literature on structural market entry models. When $\text{Var}(\boldsymbol{\kappa}_t) = 0$,

the entry game only features private information unobservables, as examined in studies such as Seim (2006), Sweeting (2009), and Bajari et al. (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 et al. (2021). In instances where $\text{Var}(\boldsymbol{\kappa}_t) > 0$ 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 Proposition 1's applicability remains unaffected by the product ownership structure. This Proposition 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 $\mathbb{E}\left(\xi_{jt} \mid a_{jt} = 1\right)$.

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 (15), describing the expected profit of a firm, readily extends to multi-product firms as follows:

$$\pi_f\left(\boldsymbol{a}_f, \boldsymbol{a}_{-f}, \boldsymbol{x}_t, \boldsymbol{\kappa}_t, \boldsymbol{\eta}_{ft}\right) = \int VP_f\left(\boldsymbol{a}_f, \boldsymbol{a}_{-f}, \boldsymbol{x}_t, \boldsymbol{\xi}_t\right) \, dF_{f,\xi}\left(\boldsymbol{\xi}_t \mid \boldsymbol{\kappa}_t, \boldsymbol{\eta}_{ft}\right) - f(\boldsymbol{x}_{ft}, \boldsymbol{\eta}_{ft}), \tag{18}$$

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

Given this expected profit, the definition of a 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.

The preceding discussion illustrates the similar structure shared by the selection problem with single- and multi-product firms. Our identification and estimation procedures are unchanged in the multi-product case.

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 $\boldsymbol{\kappa}_t$ follow a first-order Markov process or are time-invariant.

g. The private information signal $\boldsymbol{\eta}_{jt}$ 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 et al., 2021). Under Assumption 1-Dyn, the value of being or not in the market depends on the state variables $(\boldsymbol{x}_t, \boldsymbol{\kappa}_t)$ and on the private information shock η_{jt}. Let $v_j^P(\boldsymbol{x}_t, \boldsymbol{\kappa}_t, \eta_{jt})$ 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. ***Markov Perfect Equilibrium.*** *Suppose that Assumptions 1-Dyn hold. Then, a Markov Perfect Equilibrium (MPE) can be represented as a J-tuple of probability functions* $\{P_j(\boldsymbol{x}_t, \boldsymbol{\kappa}_t) : j \in \mathcal{J}\}$ *that solve the following system of best response equations in the space of probability functions:*

$$P_j(\boldsymbol{x}_t, \boldsymbol{\kappa}_t) = \int 1\{v_j^P(\boldsymbol{x}_t, \boldsymbol{\kappa}_t, \eta_{jt}) \geq 0\} \, dF_\eta(\eta_{jt}) . \qquad \blacksquare \tag{19}$$

For the rest of the paper, we will not distinguish whether the choice probabilities $P_j(\boldsymbol{x}_t, \boldsymbol{\kappa}_t)$ 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:

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

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:

$$\{s_{jt}^* = s_{jt} \text{ and } p_{jt}^* = p_{jt}\} \text{ if and only if } a_{jt} = 1. \tag{21}$$

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

$$a_{jt} \;=\; 1\left\{\pi_j^P(\boldsymbol{x}_t, \boldsymbol{\kappa}_t, \eta_{jt}) \geq 0\right\}. \tag{22}$$

Equations (20) to (22) imply the following regression equation for any product with $a_{jt} = 1$:

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

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,

$$\lambda_j(\boldsymbol{x}_t) = \int \xi_{jt}\; 1\left\{\pi_j^P(\boldsymbol{x}_t, \boldsymbol{\kappa}_t, \eta_{jt}) \geq 0\right\}\; \frac{f_{\xi,\eta,\kappa}\left(\xi_{jt}, \eta_{jt}, \boldsymbol{\kappa}_t \mid \boldsymbol{x}_t\right)}{\bar{P}_j\left(\boldsymbol{x}_t\right)}\; d\left(\xi_{jt}, \eta_{jt}, \boldsymbol{\kappa}_t\right), \tag{24}$$

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

$$\bar{P}_j\left(\boldsymbol{x}_t\right) \equiv \Pr\left(a_{jt} = 1 \mid \boldsymbol{x}_t\right) = \int 1\left\{\pi_j^P(\boldsymbol{x}_t, \boldsymbol{\kappa}_t, \eta_{jt}) \geq 0\right\}\; f_{\eta,\kappa}\left(\eta_{jt}, \boldsymbol{\kappa}_t \mid \boldsymbol{x}_t\right)\; d\left(\eta_{jt}, \boldsymbol{\kappa}_t\right). \tag{25}$$

In the econometrics literature on sample selection, it is well-known that estimating equation (23) 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* $\bar{P}_j(\boldsymbol{x}_t)$, i.e., $\lambda_j(\boldsymbol{x}_t) = \rho_j(\bar{P}_j(\boldsymbol{x}_t))$. As such, identification and estimation follow a standard two-step procedure. In a first step, one nonparametrically estimates $\bar{P}_j(\boldsymbol{x}_t)$ from data on $(a_{jt}, \boldsymbol{x}_t)$. Then, in a second step, one can apply the semiparametric series estimator in Das et al. (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-type 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 Proposition 3 presents these conditions and is an adaptation of Propositions 2 and 3 in Angrist (1997).

PROPOSITION 2. *Under Assumption 1:*

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. ■

Proposition 3(a) provides a necessary and sufficient condition for the propensity score to control for selection. This condition, consequently, implies the identification of demand parameters. Combining Proposition 3 with equation (22) characterizing the optimal entry decision, the condition stated in Proposition 3(a) can be equivalently expressed as:

$$\Pr(\xi_{jt},\ \pi_j^P(\boldsymbol{x}_t, \boldsymbol{\kappa}_t, \eta_{jt}) \mid \boldsymbol{x}_t, \bar{P}_j(\boldsymbol{x}_t)) = \Pr(\xi_{jt},\ \pi_j^P(\boldsymbol{x}_t, \boldsymbol{\kappa}_t, \eta_{jt}) \mid \bar{P}_j(\boldsymbol{x}_t)). \tag{26}$$

Unfortunately, this condition involves an 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 (26) demonstrates that assuming independence between the unobservables $(\boldsymbol{\xi}_t, \boldsymbol{\kappa}_t, \eta_{jt})$ and $\boldsymbol{x}_t$ is insufficient for the condition to hold. The crucial aspect is the functional form of the profit function π_j^P and the way the unobservables affect it. This is further illustrated in Examples 2 and 3.

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

$$\pi_j^P(\boldsymbol{x}_t, \boldsymbol{\kappa}_t, \eta_{jt}) = \gamma_{1j}(\boldsymbol{x}_t) - \gamma_{2j}(\boldsymbol{x}_t)\ \gamma_{3j}(\boldsymbol{\kappa}_t, \eta_{jt}), \tag{27}$$

where $\gamma_{1j}(\cdot)$, $\gamma_{2j}(\cdot)$, and $\gamma_{3j}(\cdot)$ are scalar real-valued functions. The optimal entry decision becomes $a_{jt} = 1\{\gamma_{3j}(\boldsymbol{\kappa}_t, \eta_{jt}) \leq \gamma_{1j}(\boldsymbol{x}_t)/\gamma_{2j}(\boldsymbol{x}_t)\}$. Assume that $\boldsymbol{\kappa}_t$ and η_{jt} are jointly independent of $\boldsymbol{x}_t$, which implies independence between $\gamma_{3jt} \equiv \gamma_{3j}(\boldsymbol{\kappa}_t, \eta_{jt})$ 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 Proposition 3(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 $\gamma_{3j}(\boldsymbol{\kappa}_t, \eta_{jt})$, 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 Proposition 3 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 $\boldsymbol{\kappa}_t = \boldsymbol{\xi}_t$. 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:

$$\pi_1^P(\boldsymbol{x}_t, \boldsymbol{\xi}_t) = \gamma_1(\boldsymbol{x}_t)\ (\boldsymbol{x}_{1t}'\ \boldsymbol{\beta} + \xi_{1t}) + \gamma_2(\boldsymbol{x}_t)\ (\boldsymbol{x}_{1t}'\ \boldsymbol{\beta} + \xi_{1t})\ (\boldsymbol{x}_{2t}'\ \boldsymbol{\beta} + \xi_{2t}) \tag{28}$$

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 Proposition 3 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 $\boldsymbol{\kappa}_t$ and the equilibrium CCPs $P_j(\boldsymbol{x}_t, \boldsymbol{\kappa}_t)$. This characterization is crucial for our identification results. For this result, we make the following

Assumption 2.

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

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: $\pi_j^P(\boldsymbol{x}_t, \boldsymbol{\kappa}_t, \eta_{jt}) = \gamma_{1j}(\boldsymbol{x}_t, \boldsymbol{\kappa}_t) - \gamma_{2j}(\boldsymbol{x}_t, \boldsymbol{\kappa}_t)\, \gamma_{3j}(\eta_{jt})$.

Under Assumption 2, we have that:

$$\begin{aligned} \Pr(\xi_{jt} \mid a_{jt} = 1, \boldsymbol{x}_t, \boldsymbol{\kappa}_t, P_j(\boldsymbol{x}_t, \boldsymbol{\kappa}_t)) &= \frac{\Pr(\xi_{jt}, a_{jt} = 1 \mid \boldsymbol{x}_t, \boldsymbol{\kappa}_t, P_j(\boldsymbol{x}_t, \boldsymbol{\kappa}_t))}{\Pr(a_{jt} = 1 \mid \boldsymbol{x}_t, \boldsymbol{\kappa}_t, P_j(\boldsymbol{x}_t, \boldsymbol{\kappa}_t))} \\ &= \frac{\Pr(\xi_{jt}, a_{jt} = 1 \mid \boldsymbol{\kappa}_t, P_j(\boldsymbol{x}_t, \boldsymbol{\kappa}_t))}{P_j(\boldsymbol{x}_t, \boldsymbol{\kappa}_t)}. \end{aligned} \tag{29}$$

It then follows that,

$$\begin{aligned} \mathbb{E}(\xi_{jt} \mid \boldsymbol{x}_t, \boldsymbol{\kappa}_t, a_{jt} = 1) &= \int \xi_{jt} \frac{\Pr(\xi_{jt}, a_{jt} = 1 \mid \boldsymbol{\kappa}_t, P_j(\boldsymbol{x}_t, \boldsymbol{\kappa}_t))}{P_j(\boldsymbol{x}_t, \boldsymbol{\kappa}_t)}\, d\xi_{jt} \\ &\equiv \psi_j\left(P_j\left(\boldsymbol{x}_t, \boldsymbol{\kappa}_t\right), \boldsymbol{\kappa}_t\right). \end{aligned} \tag{30}$$

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 (30) over the distribution of $\boldsymbol{\kappa}_t$ conditional on $(\boldsymbol{x}_t, a_{jt} = 1)$. That is:

$$\lambda_j(\boldsymbol{x}_t) = \int \psi_j\left(P_j\left(\boldsymbol{x}_t, \boldsymbol{\kappa}_t\right), \boldsymbol{\kappa}_t\right)\, f_{j,\boldsymbol{\kappa}}\left(\boldsymbol{\kappa}_t \mid \boldsymbol{x}_t\right)\, d\boldsymbol{\kappa}_t, \tag{31}$$

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

following structure:

$$f_{j,\kappa}(\boldsymbol{\kappa}_t \mid \boldsymbol{x}_t) \equiv \Pr(\boldsymbol{\kappa}_t \mid \boldsymbol{x}_t, a_{jt} = 1) = \frac{P_j(\boldsymbol{x}_t, \boldsymbol{\kappa}_t)}{\bar{P}_j(\boldsymbol{x}_t)} f_\kappa(\boldsymbol{\kappa}_t \mid \boldsymbol{x}_t). \tag{32}$$

Equation (31) shows that, without additional constraints, the selection term depends not only on the propensity score $\bar{P}_j(\boldsymbol{x}_t)$ but also on the distribution of $\boldsymbol{\kappa}_t$ and the entry probabilities conditional on both $\boldsymbol{x}_t$ and $\boldsymbol{\kappa}_t$. It is important to note that the functions ψ_j are unknown to the researcher and may vary with $\boldsymbol{\kappa}_t$. Consequently, if $\boldsymbol{\kappa}_t$ 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 $\boldsymbol{\kappa}_t$ and the entry probabilities conditional on both $\boldsymbol{x}_t$ and $\boldsymbol{\kappa}_t$.

To address this identification challenge, we strengthen Assumption 2 as follows.

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 Assumption 2 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:

$$\lambda_j(\boldsymbol{x}_t) \;=\; \sum_{\kappa_t^* \in \mathcal{K}} \psi_j\left(P_j(\boldsymbol{x}_t, \kappa_t^*), \kappa_t^*\right) \; f_{j,\kappa^*}(\kappa_t^* \mid \boldsymbol{x}_t), \tag{33}$$

with $f_{j,\kappa^*}(\kappa_t^* \mid \boldsymbol{x}_t) \equiv \frac{P_j(\boldsymbol{x}_t, \kappa_t^*)}{\bar{P}_j(\boldsymbol{x}_t)} f_{\kappa^*}(\kappa_t^* \mid \boldsymbol{x}_t)$. 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 $\boldsymbol{\theta}_\delta \equiv (\alpha, \boldsymbol{\beta}, \sigma)$; probability distribution of proxies for the demand/cost variables, $\boldsymbol{f}_{\kappa^*} \equiv (f_{\kappa^*}(\kappa^* \mid \boldsymbol{x}) : \text{for every } \kappa^*, \boldsymbol{x})$; the corresponding conditional entry probabilities, $\boldsymbol{P}_{\kappa^*} \equiv (P_j(\boldsymbol{x}, \kappa^*) : \text{for every } j, \boldsymbol{x}, \kappa^*)$; the probability distribution of private information F_η, and the conditional distribution of the demand unobservables, $f_{\xi|\eta,\kappa^*}$:

$$\boldsymbol{\theta} \equiv \left(\boldsymbol{\theta}_\delta,\ \boldsymbol{P}_{\kappa^*},\ \boldsymbol{f}_{\kappa^*},\ f_{\xi|\eta,\kappa^*},\ F_\eta \right). \tag{34}$$

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 (33), 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 et al., 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 et al. (2018) show that any J-dimensional probability mass function admits a very convenient nonparametric finite mixture or latent class representation. The following Proposition **??** summarizes this result and is an adaptation to our context of Proposition 1 by Kargas et al. (2018).

PROPOSITION 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:*

$$\Pr(\boldsymbol{a}_t = \boldsymbol{a} \mid \boldsymbol{x}_t = \boldsymbol{x}) = \sum_{\kappa^* \in \mathcal{K}(\boldsymbol{x})} f_{\kappa^*}(\kappa^* \mid \boldsymbol{x}) \left[\prod_{j=1}^{J} \left[P_j(\boldsymbol{x}, \kappa^*)\right]^{a_j} \left[1 - P_j(\boldsymbol{x}, \kappa^*)\right]^{1-a_j}\right], \tag{35}$$

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 (35) is inherently *nonparametric* in that Proposition **??** 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 Proposition **??** guarantees the existence of at least "a" representation as in (35), 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 et al. (2018)).

Remarkably, nonparametric finite mixture representation (35) resembles the joint probability of entry implied by the games of incomplete information studied in Aguirregabiria and Mira (2019); Xiao (2018). More formally, Proposition **??** 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, and perhaps surprisingly, Proposition **??** 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 (35) has been studied by Hall and Zhou (2003), Hall et al. (2005), Allman et al. (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 (35) is typically unknown to the researcher. Following ideas similar to Bonhomme et al. (2016), Xiao (2018), and Aguirregabiria and Mira (2019), we start our first-step identification argument by providing sufficient conditions for the unique determination of $|\mathcal{K}|$ 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:

$$\Pr(Y_2 = y_2, Y_3 = y_3) \;=\; \sum_{\kappa^*=1}^{|\mathcal{K}|} \Pr(Y_2 = y_2 \mid \kappa^*)\; \Pr(Y_3 = y_3 \mid \kappa^*)\; f_{\kappa^*}(\kappa^*) \tag{36}$$

or, in matrix notation,

$$\boldsymbol{P}_{Y_2,Y_3} \;=\; \boldsymbol{P}_{Y_2|\kappa^*}\; diag(\boldsymbol{f}_{\kappa^*})\; \boldsymbol{P}'_{Y_3|\kappa^*}, \tag{37}$$

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 $f_{\kappa^*}(\kappa^*)$.

PROPOSITION 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 et al. (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 Proposition 5 is an adaptation of Theorem 4 and Corollary 5 in Allman et al. (2009).

PROPOSITION 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 Proposition 4 is in general more stringent than order condition (ii) in Proposition 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 Proposition 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 Proposition 1. For those markets with $a_{jt} = 1$, the demand equation can be expressed as:

$$\delta_j(\boldsymbol{s}_t, \boldsymbol{\sigma}) \;=\; \alpha\, p_{jt} + \boldsymbol{x}_{jt}'\,\boldsymbol{\beta} + \xi_{jt}, \qquad \text{for } a_{jt} = 1 \tag{38}$$

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 v_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 (30) and we would have a standard selection problem represented by the semiparametric partially linear model:

$$\delta_j(\boldsymbol{s}_t, \boldsymbol{\sigma}) = \alpha \; p_{jt} + \boldsymbol{x}'_{jt} \, \boldsymbol{\beta} + \psi_j \left(P_j(\boldsymbol{x}_t, \kappa_t^*), \kappa_t^* \right) + \widetilde{\xi}_{jt}. \tag{39}$$

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 (39). We instead need to deal with the more complex selection bias function:

$$\lambda_j(\boldsymbol{x}_t) \equiv \mathbb{E}\left(\xi_{jt} \mid \boldsymbol{x}_t, a_{jt} = 1 \right) = \sum_{\kappa^*=1}^{|\mathcal{K}(\boldsymbol{x}_t)|} f_{j,\kappa^*}(\kappa^* \mid \boldsymbol{x}_t) \; \psi_j \left(P_j(\boldsymbol{x}_t, \kappa^*), \kappa^* \right) \; = \; \boldsymbol{f}'_{j,\kappa^*,t} \; \boldsymbol{\psi}_j(\boldsymbol{P}_{j,t}), \tag{40}$$

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:

$$\delta_j(\boldsymbol{s}_t, \boldsymbol{\sigma}) = \alpha \; p_{jt} + \boldsymbol{x}'_{jt} \, \boldsymbol{\beta} + \boldsymbol{f}'_{j,\kappa^*,t} \; \boldsymbol{\psi}_j(\boldsymbol{P}_{j,t}) + \widetilde{\xi}_{jt}. \tag{41}$$

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 (41) then clarifies that $(\boldsymbol{P}_{j,t}, \boldsymbol{f}_{\kappa^*,t})$ is a sufficient statistic for the selection bias function.

Proposition 6 establishes a necessary and sufficient condition for the identification of $\boldsymbol{\theta}_\delta \equiv (\alpha, \boldsymbol{\beta}, \boldsymbol{\sigma})$ from equation (41). It is an application of Theorem 6 in Rothenberg (1971).

PROPOSITION 6. *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* (41) *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 $\boldsymbol{Z}_{jt} \equiv (\mathbb{E}(\partial\delta_{jt}/\partial\sigma \mid \boldsymbol{x}_t),\ \mathbb{E}(p_{jt} \mid \boldsymbol{x}_t),\ \boldsymbol{x}_{jt}')'$.

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 market structure. 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 probabili-

ties estimated in the first step.[7] Then, we apply GMM to jointly estimate the coefficients in the sieve approximation and the structural demand parameters. The standard errors of the second-step demand estimates can be computed using either the asymptotic approximations and formulas in Newey (2009) or the bootstrap procedure we detail in Appendix B. A key computational advantage of the proposed bootstrap procedure is that it does not require the repeated estimation of the first step, which, even for moderate $|\mathcal{K}|$, may take up to several hours for each individual execution.

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 et al., 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:

$$f_{\kappa^*}(\kappa^* \mid \boldsymbol{x}_t) \;=\; \frac{\exp\{\boldsymbol{r}_t^{f\prime}\,\gamma_\kappa^f\}}{\sum_{\kappa'=1}^{|\mathcal{K}|}\exp\{\boldsymbol{r}_t^{f\prime}\,\gamma_{\kappa'}^f\}}, \tag{42}$$

where, for $\kappa^* = 1, 2, ..., |\mathcal{K}|$, $\gamma_{\kappa^*}^f$ is a vector of parameters with dimension $L_f \times 1$ and normalization $\gamma^f(1) = 0$. 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:

$$P_j(\boldsymbol{x}_t, \kappa^*) \;=\; \Lambda\left(\boldsymbol{r}_t^{P\prime}\,\gamma_{j\kappa^*}^P\right), \tag{43}$$

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

[7] 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 et al. (2007).

finite mixture model is:

$$\ell(\gamma^{f,P}) \;=\; \sum_{t=1}^{T} \ln \left(\sum_{\kappa^*=1}^{|\mathcal{K}|} f_{\kappa^*}(\kappa^* \mid \boldsymbol{x}_t, \gamma^f) \prod_{j=1}^{J} \Lambda\left(\boldsymbol{r}_t^{P\prime}\, \gamma_{j\kappa^*}^{P}\right)^{a_{jt}} \left[1 - \Lambda\left(\boldsymbol{r}_t^{P\prime}\, \gamma_{j\kappa^*}^{P}\right)\right]^{1-a_{jt}} \right), \tag{44}$$

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

We estimate the vector of parameters $\gamma^{f,P}$ 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 et al. (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 et al. (2003) and Das et al. (2003).

5.2 Second step: Estimation of demand parameters

Following Das et al. (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 $P_j(\boldsymbol{x}_t, \kappa^*)$:

$$\begin{aligned} \psi_j\left(P_j(\boldsymbol{x}_t, \kappa^*), \kappa^*\right) \;&\approx\; \boldsymbol{r}^{\psi}\left(P_j(\boldsymbol{x}_t, \kappa^*)\right)' \gamma_{j\kappa^*}^{\psi} \\ &=\; \left[1,\; \ln P_j(\boldsymbol{x}_t, \kappa^*),\; \ln P_j(\boldsymbol{x}_t, \kappa^*)^2,\; ...,\; \ln P_j(\boldsymbol{x}_t, \kappa^*)^{L_\psi}\right] \gamma_{j\kappa^*}^{\psi}, \end{aligned} \tag{45}$$

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

$$\boldsymbol{f}'_{j,\kappa^*,t}\boldsymbol{\psi}_j(\boldsymbol{P}_{j,t}) \;\approx\; \boldsymbol{h}'_{j,t}\,\boldsymbol{\gamma}^{\psi}_j \;=\; \sum_{\kappa^*=1}^{|\mathcal{K}|}\sum_{\ell=0}^{L_\psi} \gamma^{\psi}_{\ell,j\kappa^*}\, f_{j,\kappa^*}(\kappa^* \mid \boldsymbol{x}_t)\;\left(\ln P_j(\boldsymbol{x}_t,\kappa^*)\right)^{\ell} \tag{46}$$

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

Plugging equation (46) into demand equation (41), we obtain the regression equation:

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

Equation (47) can be estimated by GMM. Following Das et al. (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 endogenous product entry in demand estimation in this industry has recently been explored by Ciliberto et al. (2021) and Li et al. (2022).

We use publicly available data from the US Department of Transportation for our analysis. Our working sample consists of the DB1B and T100 datasets. Specifically, we use quarterly data spanning 2012-Q1 to 2013-Q4 for routes between the airports at the 100 largest Metropolitan Statistical Areas (MSA) in the United States. These account 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 the same market as LGA to ORD. There are potentially 5,778 non-directional markets between the 108 airports, i.e., $108 \times 107/2$. However, many of these markets have not 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 only consider non-directional markets which were served in at least 50 quarters between 1994 and 2018. This results in 2,230 non-directional markets and 17,155 market-quarter observations.[8] We consider an airline a *potential entrant* in a non-directional airport pair in a given quarter if it 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).[9] 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 areas (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 averages of the market

[8] 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.

[9] Following Ciliberto et al. (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.

characteristics. Notably, in a significant portion of these markets (almost 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 probabilities, respectively. Furthermore, there is substantial heterogeneity in the correlations between entry, market size, and distance among airlines. 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 population (13.3 million people versus 8.4 million people). Different entry strategies are also evident on the basis of 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 of market entry

For the entry decisions, we consider the nonparametric sieve finite mixture Logit described in equations (42) and (43). The vector of explanatory variables $\boldsymbol{x}_t$ 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.[10]

In the analysis, we focus on specifications of the mixture Logit model in which the distribution of κ_t^* is independent of $\boldsymbol{x}_t$.[11] Our choice stems from the robustness of our demand estimates to incorporating dependence between κ_t^* and $\boldsymbol{x}_t$, and because this form of independence 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

[10] 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.

[11] Although we assume that $f_{\kappa^*}(\kappa_t^* \mid \boldsymbol{x}_t) = f_{\kappa^*}(\kappa_t^*)$, the distribution of κ_t^* conditional on $a_{jt} = 1$, $f_{j,\kappa^*}(\kappa_t^* \mid \boldsymbol{x}_t) \equiv \frac{P_j(\boldsymbol{x}_t, \kappa_t^*)}{\bar{P}_j(\boldsymbol{x}_t)} f_{\kappa^*}(\kappa_t^*)$, is still a function of $\boldsymbol{x}_t$. See equation (51) below.

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.

We explored various specifications of the mixture Logit model based on two key elements: the polynomial order in $\boldsymbol{x}_t$ used to construct the basis $\boldsymbol{r}_t^P$ and the number of elements in the support of κ_t^*. As our estimates of the demand parameters are robust to the selection of the basis $\boldsymbol{r}_t^P$ in the entry model, we only present results for the specification with $\boldsymbol{r}_t^P = \boldsymbol{x}_t$. Regarding the specification of the number of unobserved market types $|\mathcal{K}|$, Table 3 presents the goodness-of-fit statistics obtained from estimating four nested specifications of the mixture Logit model. The choice of the preferred specification is guided by the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC), along with the convergence performance of the EM algorithm, the accuracy of the parameter estimates of the entry model, and the robustness of the estimates of the demand model.

The introduction of unobserved market heterogeneity κ_t^* improves the entry model's goodness-of-fit. Comparing the specification without κ_t^* and the one with two unobserved market types in Table 3, we see a substantial increase in the log-likelihood and a decrease in both AIC and BIC. This form of unobserved market heterogeneity captures a strong correlation among airline entry decisions, a correlation not captured by the observable market and airline characteristics in $\boldsymbol{x}_t$. The inclusion of additional unobserved market types continues to positively impact the goodness-of-fit. However, this improvement has diminishing returns and it is very small when moving 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 issues in the specification with four unobserved market types. In this case, we obtain imprecise estimates for some of the parameters of the

entry model. These considerations, combined with the marginal improvement observed in the AIC and BIC criteria, lead us to favor the specification with $|\mathcal{K}| = 3$. Moreover, this choice is also motivated by the implied estimates of the demand model. As we illustrate below, the estimated own-price elasticities of demand with $|\mathcal{K}| = 3$ and $|\mathcal{K}| = 4$ are practically indistinguishable. In contrast, with $|\mathcal{K}| \leq 2$ we obtain estimates of the own-price elasticities which are substantially smaller.

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 et al. (2021) and estimate a nested logit model with two nests: a nest for all the airlines and another nest for the outside option.

$$\ln\left(\frac{s_{jt}}{s_{0t}}\right) = \alpha\, p_{jt} + \boldsymbol{x}'_{jt}\,\boldsymbol{\beta} + \sigma\, \ln\left(\frac{s_{jt}}{1 - s_{0t}}\right) + \boldsymbol{h}'_{jt}\,\boldsymbol{\gamma}^{\psi}_{j} + \widetilde{\xi}_{jt}. \tag{48}$$

We compute each directional route-specific market share in a given quarter s_{jt} as the total number of passengers who traveled that directional route with a non-stop flight of a specific airline in that given quarter (times 10, as the data are 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 term,

$\boldsymbol{h}'_{jt}\,\boldsymbol{\gamma}^{\psi}_{j}$, varies with the specification 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 $a_{jt} = 1\{\eta_{jt} \leq \boldsymbol{x}'_{jt}\boldsymbol{\gamma}^{P}_{j}\}$, with $\eta_{jt} \sim Logistic$, and $\xi_{jt} = \gamma^{\psi}_{j,1}\,\eta_{jt} + v_{jt}$, with v_{jt} independent of η_{jt} and $\boldsymbol{x}_t$. In this parametric specification, the selection term takes the following form:

$$\mathbb{E}\left(\xi_{jt} \mid a_{jt} = 1, \boldsymbol{x}_t\right) \;=\; \gamma^{\psi}_{j,1}\,\mathbb{E}\left(\eta_{jt} \mid \eta_{jt} \leq \boldsymbol{x}'_{jt}\boldsymbol{\gamma}^{P}_{j}\right) \;=\; \gamma^{\psi}_{j,1}\,\left[Euler - \ln\Lambda\left(\boldsymbol{x}'_{t}\,\boldsymbol{\gamma}^{P}_{j}\right)\right]. \quad (49)$$

   where $Euler \approx 0.5772$ represents Euler's constant. 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 $Euler - \ln\Lambda\left(\boldsymbol{x}'_t\boldsymbol{\gamma}^{P}_{j}\right)$.[12] Therefore, the vector of regressors controlling for endogenous selection is:

$$\boldsymbol{h}'_{jt} = \left[\left(Euler - \ln\Lambda\left(\boldsymbol{x}'_t\boldsymbol{\gamma}^{P}_{j}\right)\right)^{\ell} : \ell = 1, 2, 3\right]. \quad (50)$$

   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 $\kappa^*_t = \kappa^*$, and with conditional mixture distribution $f_{j,\kappa^*}(\kappa^* \mid \boldsymbol{x}_t) \;=\; \frac{\Lambda\left(\boldsymbol{x}'_t\boldsymbol{\gamma}^{P}_{j\kappa^*}\right) f_{\kappa^*}(\kappa^*)}{\sum_{\iota=1}^{|\mathcal{K}|}\Lambda\left(\boldsymbol{x}'_t\boldsymbol{\gamma}^{P}_{j\iota}\right) f_{\kappa^*}(\iota)}$. 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:

$$\boldsymbol{h}'_{jt} = \left[f_{j,\kappa^*}(\kappa^* \mid \boldsymbol{x}_t)\left(Euler - \ln\Lambda\left(\boldsymbol{x}'_t\boldsymbol{\gamma}^{P}_{j\kappa^*}\right)\right)^{\ell} : \ell = 1, 2, 3, \text{ and } \kappa^* = 1, 2, ..., |\mathcal{K}|\right]. \quad (51)$$

[12] Both in this and in the case of the semiparametric mixture Logit, estimates are very similar by approximating the selection function with polynomials of higher orders.

For all the 2SLS estimators, we use as instruments the number of competitors in the market and the average hub-size of the rest of the airlines, separately for origin and destination. We compute standard errors on the basis of the bootstrap procedure detailed in Appendix B.

Table 4 presents the estimates of the demand parameters, while Table 5 reports the average demand elasticities and Lerner indexes derived from these estimates. Comparing the estimates obtained using OLS with those from the standard 2SLS method — not accounting for potential selection bias — we observe a significant change in all parameter estimates when addressing the endogeneity of prices and within-nest market shares. Controlling for endogeneity meaningfully affects the average estimated own-price elasticity, which decreases from -1.60 to -5.55, and the corresponding average Lerner index, which decreases from 68.8% to 19.9%.

Turning to the consequences of controlling for endogenous market entry, we note the important role played by finite mixture unobserved heterogeneity. The estimates of parameters α and σ of a finite mixture model with $|\mathcal{K}| = 3$ are, compared to those of "Semiparametric" (assuming $|\mathcal{K}| = 1$), 15.9% and 28.8% higher (in absolute terms). These changes translate into an increase in the average estimated own-price elasticities of around 30%. Consequently, the corresponding average estimated Lerner index decreases from 18.9% to 15.1%. These effects are of substantial importance and lead to meaningful economic implications.

Parameter estimates and implied own-price elasticities of the standard 2SLS (not controlling for selection) and those of "Heckman" or "Semiparametric" (assuming $|\mathcal{K}| = 1$) are relatively similar. In contrast, parameter estimates and corresponding own-price elasticities remarkably change when we allow $|\mathcal{K}| > 1$. Although the estimated own-price elasticities of a model with $|\mathcal{K}| = 2$ are still meaningfully different from those of a model with $|\mathcal{K}| = 3$, the estimates implied by models with $|\mathcal{K}| = 3$ and $|\mathcal{K}| = 4$ are essentially indistinguishable. Collectively, these results stress the importance of allowing for "some" unobserved market heterogeneity to effectively control for endogenous selection, but also that as few unobserved market types as three may already be sufficient.

Figure 1 plots the empirical distributions of the estimated own-price elasticities. Each row

Table 4: Estimation of Demand Parameters

	Not control. for sel.		*Controlling for endogenous selection*				
	OLS	**2SLS**	**2SLS Heckman** $\lvert\mathcal{K}\rvert = 1$	**2SLS Semipar.** $\lvert\mathcal{K}\rvert = 1$	**2SLS Fin.-Mix.** $\lvert\mathcal{K}\rvert = 2$	**2SLS Fin.-Mix.** $\lvert\mathcal{K}\rvert = 3$	**2SLS Fin.-Mix.** $\lvert\mathcal{K}\rvert = 4$
Price (100$) ($\alpha$)	−0.643	−2.180	−2.193	−2.261	−2.392	−2.621	−2.697
	(0.0105)	(0.1378)	(0.2065)	(0.2077)	(0.2201)	(0.2448)	(0.2716)
Within Share (σ)	0.371	0.409	0.413	0.431	0.494	0.555	0.546
	(0.0058)	(0.0351)	(0.0529)	(0.0559)	(0.0622)	(0.0717)	(0.0821)
Distance (1000mi)	0.729	2.130	2.196	2.264	2.387	2.503	2.624
	(0.0306)	(0.1372)	(0.2074)	(0.2055)	(0.2133)	(0.2390)	(0.2648)
Distance^2	−0.216	−0.424	−0.453	−0.462	−0.493	−0.525	−0.502
	(0.0112)	(0.0244)	(0.0398)	(0.0392)	(0.0401)	(0.0440)	(0.0483)
hub-size orig. (100s)	1.637	2.272	1.999	1.320	1.709	1.677	1.444
	(0.0263)	(0.0382)	(0.0767)	(0.0919)	(0.1085)	(0.1206)	(0.1244)
hub-size dest. (100s)	1.613	2.242	1.995	1.310	1.703	1.674	1.436
	(0.0267)	(0.0385)	(0.0784)	(0.0933)	(0.1106)	(0.1228)	(0.1266)
Airline×Quarter FE	Y	Y	Y	Y	Y	Y	Y
# control var. entry	0	0	6	18	36	54	72
Observations	35,763	35,763	35,763	35,763	35,763	35,763	35,763

Bootstrap standard errors account for estimation error in the first step using the method described in Appendix B.

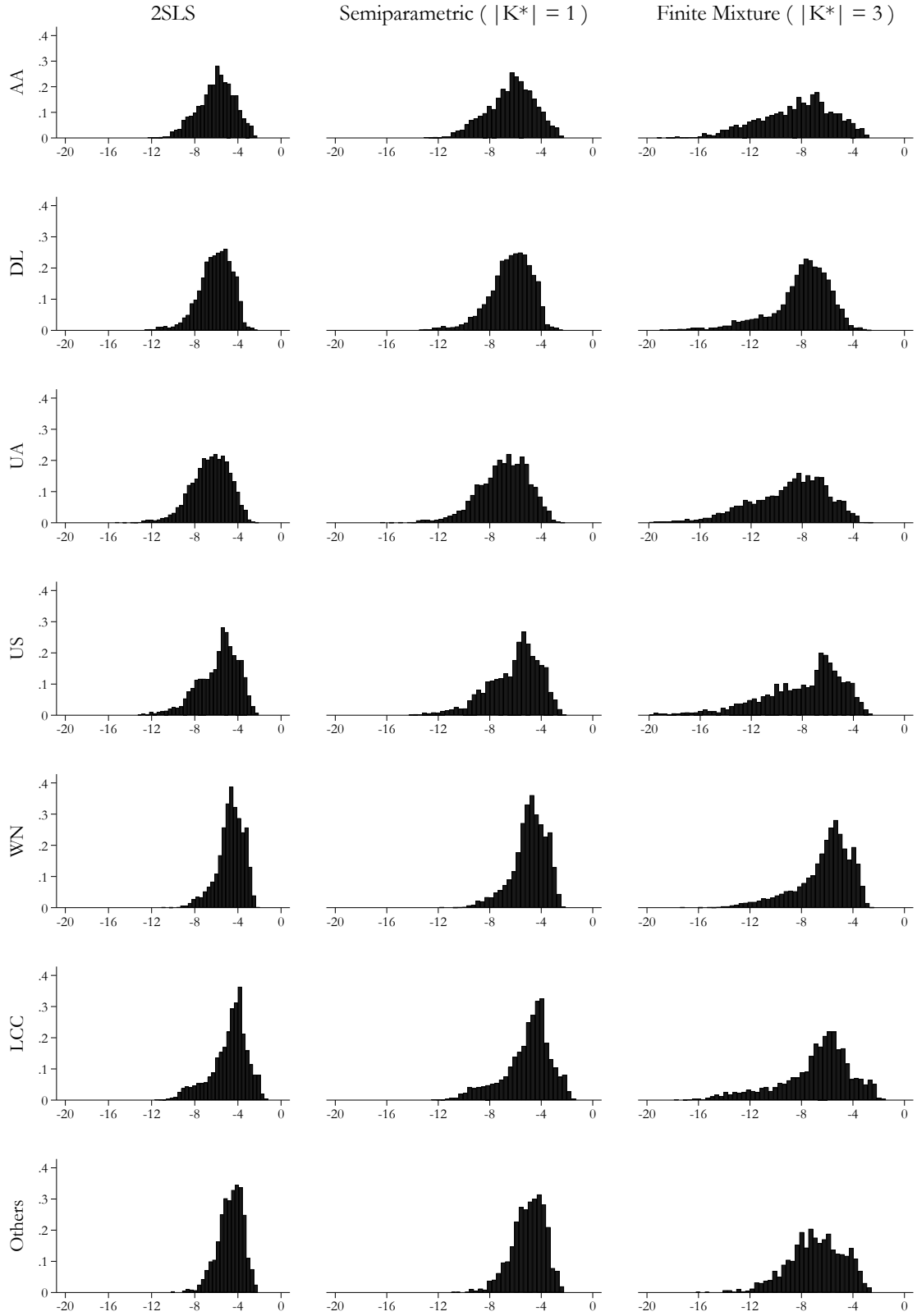
corresponds to an airline, while each column to a different 2SLS estimator: the first column plots results for the estimator that does not control for selection, the second column plots results for the estimator that controls for selection using a sieve method but no mixture, and the third column plots results for the estimator with three unobserved market types.

The histograms in this figure are constructed based on estimates of own-price elasticities at the airline-market-quarter level. The equation describing each own-price elasticity only 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 change in empirical distributions can only be attributed to changes in the values of the estimates $\widehat{\alpha}$ and $\widehat{\sigma}$ across the different estimators.

Table 5: Average Own-Price Elasticities and Lerner Indexes

	Not control. for sel.		*Controlling for endogenous selection*				
	OLS	**2SLS**	**2SLS Heckman** $\lvert\mathcal{K}\rvert = 1$	**2SLS Semipar.** $\lvert\mathcal{K}\rvert = 1$	**2SLS Fin.-Mix.** $\lvert\mathcal{K}\rvert = 2$	**2SLS Fin.-Mix.** $\lvert\mathcal{K}\rvert = 3$	**2SLS Fin.-Mix.** $\lvert\mathcal{K}\rvert = 4$
Own-Price Elasticity	−1.596	−5.549	−5.601	−5.849	−6.524	−7.605	−7.746
AA	−1.722	−6.013	−6.071	−6.363	−7.143	−8.399	−8.543
DL	−1.761	−6.082	−6.133	−6.382	−7.024	−8.067	−8.236
UA	−1.887	−6.573	−6.636	−6.936	−7.766	−9.090	−9.253
US	−1.665	−5.801	−5.856	−6.122	−6.854	−8.023	−8.167
WN	−1.354	−4.680	−4.719	−4.913	−5.411	−6.220	−6.350
LCC	−1.370	−4.808	−4.857	−5.095	−5.784	−6.870	−6.977
Others	−1.332	−4.705	−4.757	−5.006	−5.750	−6.915	−7.009
Lerner Index	68.8%	19.9%	19.7%	18.9%	17.2%	15.1%	14.7%
AA	62.7%	18.0%	17.9%	17.1%	15.5%	13.5%	13.2%
DL	60.4%	17.5%	17.3%	16.7%	15.3%	13.4%	13.1%
UA	56.9%	16.4%	16.2%	15.6%	14.1%	12.3%	12.1%
US	65.9%	19.0%	18.9%	18.1%	16.5%	14.5%	14.2%
WN	78.4%	22.8%	22.6%	21.8%	20.1%	17.8%	17.4%
LCC	82.1%	23.5%	23.3%	22.2%	19.9%	17.1%	16.8%
Others	79.2%	22.5%	22.3%	21.3%	18.9%	16.0%	15.8%
Observations	35,763	35,763	35,763	35,763	35,763	35,763	35,763

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



The empirical distributions in the first two columns of Figure 1 are very similar. In contrast, the empirical distributions based on the finite mixture estimates show substantially different locations and dispersions. Across all airlines, the larger estimates of $\widehat{\alpha}$ and $\widehat{\sigma}$ using the mixture method lead to a leftward shift and an amplification in the spread of the empirical distributions. These changes in the empirical distributions' location and dispersion may have important economic implications in any application that requires demand estimates as input for further analyses — irrespective of whether endogenous product entry and/or exit is in itself of any economic interest.

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 discuss these additional estimation procedures in the context of our empirical application.

6.4.1 Marginal costs

Based on an assumption about the nature of competition, such as Bertrand-Nash competition, the researcher would be 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 are observed to be active in the market.

For some empirical questions, given the marginal costs, the researcher may need to further estimate a marginal cost function: that is, a function that represents the effect of product characteristics and output on marginal costs. For this purpose, the researcher needs to estimate the parameters of a regression in which the dependent variable is the marginal cost estimate

and the explanatory variables are the exogenous product characteristics $\boldsymbol{x}_{jt}$ and, in the case of non-linear returns to scale, the output q_{jt}. As in this case of demand, 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.

We now illustrate these points in the context of our application. Following Ciliberto et al. (2021), we assume that the airlines engage in Bertrand-Nash competition and that each airline has marginal cost function that does not depend on output. Then, given demand equation (48), the marginal cost function of airline j in market-quarter t can be estimated from the following pricing equation:

$$p_{jt} + \frac{1-\sigma}{\alpha(1-\sigma s_{jt|g} - (1-\sigma)s_{jt})} = mc_{jt}, \tag{52}$$

where g denotes the nest that contains all the airlines and the marginal cost mc_{jt} is specified as:

$$mc_{jt} = \boldsymbol{x}'_{jt}\,\boldsymbol{\varphi} + \boldsymbol{h}'_{jt}\,\boldsymbol{\gamma}_j^{\psi,\text{mc}} + \widetilde{\omega}_{jt}, \tag{53}$$

with both $\boldsymbol{x}_{jt}$ and $\boldsymbol{h}_{jt}$ defined as in the case of demand equation (48), while the unobserved component of marginal cost is $\omega_{jt} \equiv \mathbb{E}\left(\omega_{jt} \mid a_{jt} = 1, \boldsymbol{x}_t\right) + \widetilde{\omega}_{jt}$. Similarly to the case of demand, the selection term is $\mathbb{E}\left(\omega_{jt} \mid a_{jt} = 1, \boldsymbol{x}_t\right) = \boldsymbol{h}'_{jt}\,\boldsymbol{\gamma}_j^{\psi,\text{mc}}$ and we consider the same specifications for $\boldsymbol{h}_{jt}$ as those described in section 6.3.

Table 6 reports the average marginal costs obtained from equation (52) and the demand estimates in Table 4 (see Appendix Figure 2 for the corresponding empirical distributions), while Table 7 presents our estimates of $\boldsymbol{\varphi}$ from equation (53). The estimates of $\boldsymbol{\varphi}$ in each column of Table 7 rely on the corresponding demand estimates of Table 4, so that, for example, the first column of Table 7 reports estimates of $\boldsymbol{\varphi}$ obtained by using the estimates of α and σ (i.e., plugging them in the left-hand side of (52)) from the first column of Table 4. Collectively, these results illustrate that although endogeneity of prices and of within-nest market shares play

Table 6: Average Marginal Costs

	Not control. for sel.		*Controlling for endogenous selection*				
	OLS	**2SLS**	**2SLS Heckman** $\lvert\mathcal{K}\rvert = 1$	**2SLS Semipar.** $\lvert\mathcal{K}\rvert = 1$	**2SLS Fin.-Mix.** $\lvert\mathcal{K}\rvert = 2$	**2SLS Fin.-Mix.** $\lvert\mathcal{K}\rvert = 3$	**2SLS Fin.-Mix.** $\lvert\mathcal{K}\rvert = 4$
Marginal Cost (100$)	0.766	1.718	1.721	1.736	1.769	1.810	1.817
AA	0.901	1.829	1.832	1.847	1.881	1.924	1.930
DL	1.049	2.032	2.036	2.050	2.082	2.123	2.130
UA	1.134	2.072	2.075	2.090	2.123	2.165	2.171
US	0.830	1.779	1.782	1.797	1.830	1.871	1.878
WN	0.464	1.461	1.464	1.478	1.510	1.549	1.557
LCC	0.434	1.330	1.333	1.349	1.384	1.427	1.432
Others	0.362	1.220	1.224	1.239	1.276	1.319	1.323
Observations	35,763	35,763	35,763	35,763	35,763	35,763	35,763

an important role on the implied marginal cost estimates from equation (52), endogenous selection seems to have less of an impact. Moreover, the parameter estimates of equation (53) (which uses the estimated mc_{jt} as a dependent variable) look remarkably similar across *all* columns of Table 7, including in the case of the OLS. From these findings, we can conclude that — at least in our sample — the unobserved component of entry η_{jt} appears to be strongly correlated with the unobserved component of demand ξ_{jt} but not with that of marginal cost ω_{jt}. In other words, heterogeneity in airlines' entry decisions appears to be primarily explained by demand-side rather than by marginal cost-side unobserved heterogeneity.

6.4.2 Demand and marginal cost unobservables

The consistent estimation of demand and marginal cost parameters yields consistent estimates of the corresponding unobservable variables, ξ_{jmt} and ω_{jmt}, which can be obtained as residuals from the estimated equations. While the estimation of these equations is subject to selection bias, controlling for selection enables the researcher to achieve consistent estimation of the unobservable variables ξ_{jmt} and ω_{jmt} for the airlines that are observed to be active in the

Table 7: Estimation of Marginal Cost Parameters

	Not control. for sel.		*Controlling for endogenous selection*				
	OLS	**2SLS**	**2SLS Heckman** $\lvert\mathcal{K}\rvert = 1$	**2SLS Semipar.** $\lvert\mathcal{K}\rvert = 1$	**2SLS Fin.-Mix.** $\lvert\mathcal{K}\rvert = 2$	**2SLS Fin.-Mix.** $\lvert\mathcal{K}\rvert = 3$	**2SLS Fin.-Mix.** $\lvert\mathcal{K}\rvert = 4$
Distance (1000mi)	0.971	0.927	0.938	0.935	0.934	0.937	0.965
	(0.014)	(0.014)	(0.023)	(0.022)	(0.024)	(0.023)	(0.025)
Distance^2	−0.150	−0.139	−0.146	−0.144	−0.147	−0.149	−0.149
	(0.006)	(0.006)	(0.008)	(0.008)	(0.009)	(0.009)	(0.010)
hub-size orig. (100s)	0.247	0.382	0.237	0.103	0.326	0.348	0.288
	(0.013)	(0.013)	(0.024)	(0.031)	(0.034)	(0.034)	(0.034)
hub-size dest. (100s)	0.243	0.377	0.241	0.105	0.330	0.353	0.290
	(0.013)	(0.013)	(0.024)	(0.031)	(0.035)	(0.034)	(0.034)
Airline×Quarter FE	Y	Y	Y	Y	Y	Y	Y
# control var. entry	0	0	6	18	36	54	72
Observations	35,763	35,763	35,763	35,763	35,763	35,763	35,763

Bootstrap standard errors account for estimation error in the first step using the method described in Appendix B.

market.[13]

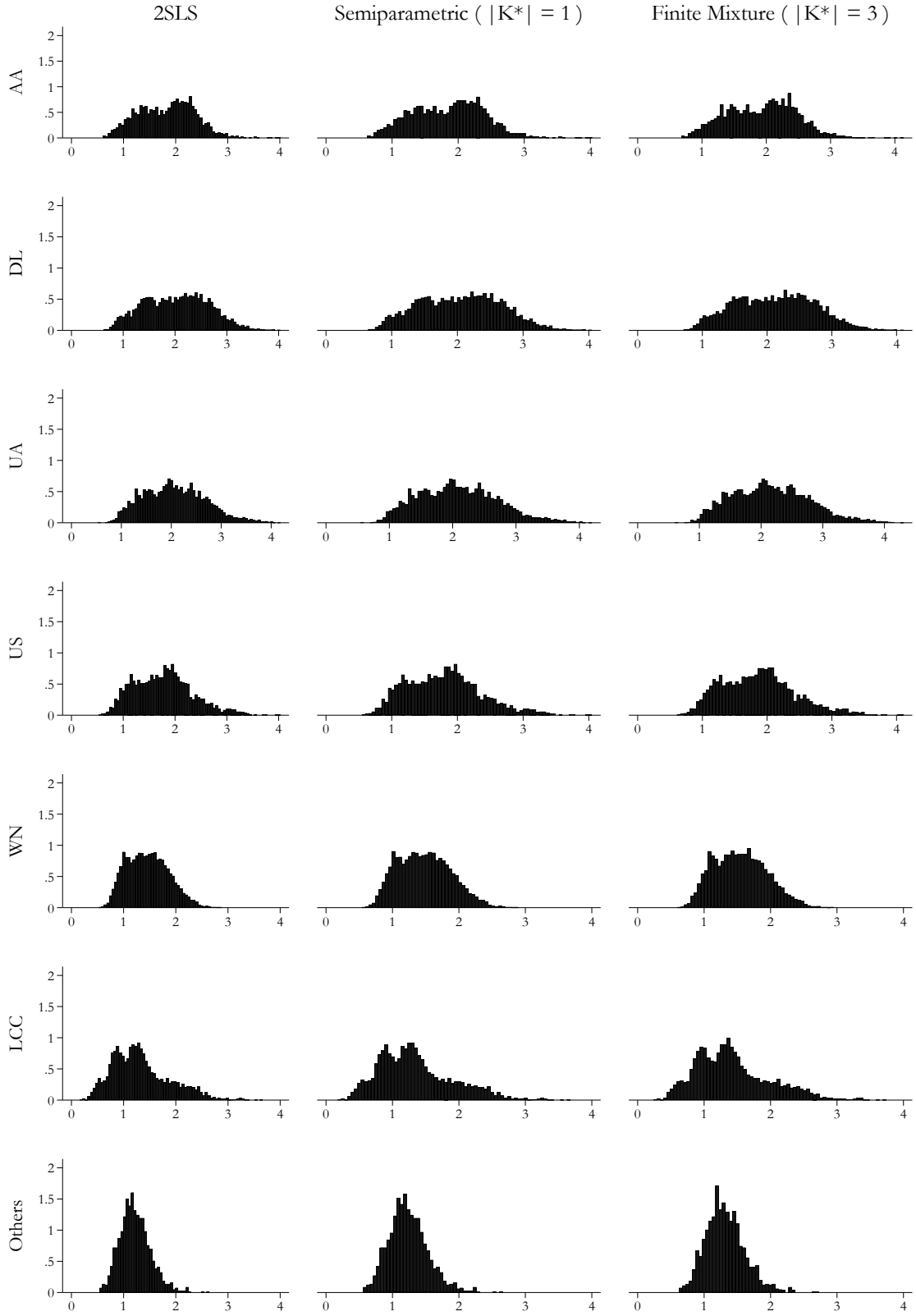
Naturally, the more complex estimation of the probability distribution governing these unobservables for *all* products requires one to address an additional extrapolation problem for the products observed to be *inactive* in the market. See related discussion in subsection 6.4.4 below.

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.

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

Figure 2: Distribution of Estimated Marginal Costs (Airline-Market-Quarter level)



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.[14]

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 the values of ξ_{jmt} and ω_{jmt} for 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 this particular market. To find values of ξ_{jmt} and ω_{jmt} that are consistent 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 $\mathbb{E}(\xi_{jmt}|\boldsymbol{x}_{mt}, a_{jmt}=$

[14] 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.

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 choose 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 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

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

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

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

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

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

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., Honoré, B., and Powell, J. (2007). Pairwise difference estimation with nonparametric control variables. *International Economic Review*, 48(4):1119–1158.

Armstrong, T. B., Bertanha, M., and Hong, H. (2014). A fast resample method for parametric and semiparametric models. *Journal of Econometrics*, 179(2):128–133.

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

Berry, S. (1992). Estimation of a model of entry in the airline industry. *Econometrica*, 60(4):889–917.

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

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

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

Berry, S., Levinsohn, J., and Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica*, 63(4):841–890.

Bonhomme, S., Jochmans, K., and Robin, J.-M. (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., Gualdani, C., and Remmy, K. (2023). Price competition and endogenous product choice in networks: Evidence from the us airline industry. Technical report, Working Paper.

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

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

Browning, M., Chiappori, P.-A., and Weiss, Y. (2014). Economics of the family. *Cambridge Books*.

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

Bunting, J., Diegert, P., and Maurel, A. (2022). Heterogeneity, uncertainty and learning: Semiparametric identification and estimation. *Working Paper*.

Caoui, E. H. and Steck, A. (2023). Diversification, market entry, and the global internet backbone. *Available at SSRN: http://dx.doi.org/10.2139/ssrn.4478868*.

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 Chang, J.-J. (1970). Analysis of individual differences in multidimensional scaling via an n-way generalization of "eckart-young" decomposition. *Psychometrika*, 35(3):283–319.

Cattaneo, M. D., Jansson, M., and Ma, X. (2019). Two-step estimation and inference with possibly many included covariates. *The Review of Economic Studies*, 86(3):1095–1122.

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

Choo, E. and Siow, A. (2006). Who marries whom and why. *Journal of Political Economy*, 114(1):175–201.

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

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

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

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

De Paula, Á., Richards-Shubik, S., and Tamer, E. (2018). Identifying preferences in networks with bounded degree. *Econometrica*, 86(1):263–288.

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

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

Dunson, D. B. and Xing, C. (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.

Fan, Y. (2013). Ownership consolidation and product characteristics: A study of the us daily newspaper market. *American Economic Review*, 103(5):1598–1628.

Fan, Y. and Yang, C. (2020). Competition, product proliferation, and welfare: A study of the us smartphone market. *American Economic Journal: Microeconomics*, 12(2):99–134.

Galichon, A. and Salanié, B. (2022). Cupid's invisible hand: Social surplus and identification in matching models. *The Review of Economic Studies*, 89(5):2600–2629.

Gonçalves, S., Hounyo, U., Patton, A. J., and Sheppard, K. (2023). Bootstrapping two-stage quasi-maximum likelihood estimators of time series models. *Journal of Business & Economic Statistics*, 41(3):683–694.

Graham, B. S. (2017). An econometric model of network formation with degree heterogeneity. *Econometrica*, 85(4):1033–1063.

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

Hall, P. and Zhou, X.-H. (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.

Heckman, J. and Navarro-Lozano, S. (2004). Using matching, instrumental variables, and control functions to estimate economic choice models. *Review of Economics and Statistics*, 86(1):30–57.

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

Hu, Y. and Xin, Y. (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., Sidiropoulos, N. D., and Fu, X. (2018). Tensors, learning, and "kolmogorov extension" for finite-alphabet random vectors. *IEEE Transactions on Signal Processing*, 66(18):4854–4868.

Kasahara, H. and Shimotsu, K. (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 Bader, B. W. (2009). Tensor decompositions and applications. *SIAM review*, 51(3):455–500.

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

Liu, H. and Luo, Y. (2025). Demand analysis under price rigidity and endogenous assortment: An application to china's tobacco industry. *arXiv preprint arXiv:2501.17251*.

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

Pilla, R. S. and Lindsay, B. G. (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*, pages 165–96.

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

Schaumans, C. and Verboven, F. (2015). Entry and competition in differentiated products markets. *Review of Economics and Statistics*, 97(1):195–209.

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., De Lathauwer, L., Fu, X., Huang, K., Papalexakis, E. E., and Faloutsos, C. (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.

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

Vella, F. (1998). Estimating models with sample selection bias: a survey. *Journal of Human Resources*, pages 127–169.

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.

Wooldridge, J. M. (2015). Control function methods in applied econometrics. *Journal of Human Resources*, 50(2):420–445.

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

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

Appendices

A Proofs of Propositions

A.1 Proof of Proposition 1

For $j \in \{0\} \cup \mathcal{J}^a$, let $d_j(\boldsymbol{\delta}^a, \boldsymbol{\sigma})$, as defined in equation (4), denote the model's prediction for the market share s_j. Define $P_j(\boldsymbol{\delta}^a, \boldsymbol{\sigma}, \boldsymbol{v})$ as the multinomial logit choice probability, whose integral over the distribution of $\boldsymbol{v}$ yields the demand function $d_j(\boldsymbol{\delta}^a, \boldsymbol{\sigma})$. Based on the model assumptions as described in section 2, this demand function satisfies the following properties:

P1. *Continuously differentiable* in $\boldsymbol{\delta}^a$.

P2. *Non-degenerate*: For any argument $(\boldsymbol{\delta}^a, \boldsymbol{\sigma})$ with $a_{jt} = 1$, we have, $0 < d_j(\boldsymbol{\delta}^a, \boldsymbol{\sigma}) < 1$.

P3. *Strict monotonicity*: For any product $j \in \mathcal{J}^a$:

$$\frac{\partial d_j(\boldsymbol{\delta}^a, \boldsymbol{\sigma})}{\partial \delta_j} = \int P_j(\boldsymbol{\delta}^a, \boldsymbol{\sigma}, \boldsymbol{v}) \Big(1 - P_j(\boldsymbol{\delta}^a, \boldsymbol{\sigma}, \boldsymbol{v})\Big)\, dF_{\boldsymbol{v}}(\boldsymbol{v}) > 0.$$

Similarly, for $i \neq j \in \mathcal{J}^a$:

$$\frac{\partial d_j(\boldsymbol{\delta}^a, \boldsymbol{\sigma})}{\partial \delta_i} = -\int P_j(\boldsymbol{\delta}^a, \boldsymbol{\sigma}, \boldsymbol{v})\, P_i(\boldsymbol{\delta}^a, \boldsymbol{\sigma}, \boldsymbol{v})\, dF_{\boldsymbol{v}}(\boldsymbol{v}) < 0.$$

P4. *Jacobian matrix with strict diagonal dominance*: Let $\boldsymbol{D}(\boldsymbol{\delta}^a)$ denote the $|J^a| \times |J^a|$ Jacobian matrix with entries $D_{ij}(\boldsymbol{\delta}^a) = \frac{\partial d_j(\boldsymbol{\delta}^a, \boldsymbol{\sigma})}{\partial \delta_i}$. Taking into account the structure of demand as an integral of a multinomial logit probability, it is possible to show that for any $j \in \mathcal{J}^a$:

$$|D_{jj}(\boldsymbol{\delta}^a)| \; > \; \sum_{i \neq j} |D_{ij}(\boldsymbol{\delta}^a)|.$$

P5. *Globally invertible Jacobian matrix*: By Levy–Desplanques, the strict diagonal dominance of the Jacobian matrix implies that this matrix is non-singular.

P6. *Boundary bevavior*: Using monotonicity of the logit probabilities together with the Dominated Convergence Theorem, we have that

$$\lim_{\delta_j \to -\infty} d_j(\boldsymbol{\delta}^a, \boldsymbol{\sigma}) = \int \lim_{\delta_j \to -\infty} P_j(\boldsymbol{\delta}^a, \boldsymbol{\sigma}, \boldsymbol{v})\, dF_{\boldsymbol{v}}(\boldsymbol{v}) = \int 0\, dF_{\boldsymbol{v}}(\boldsymbol{v}) = 0$$

$$\lim_{\delta_j \to +\infty} d_j(\boldsymbol{\delta}^{\boldsymbol{a}}, \boldsymbol{\sigma}) = \int \lim_{\delta_j \to +\infty} P_j(\boldsymbol{\delta}^{\boldsymbol{a}}, \boldsymbol{\sigma}, \boldsymbol{v})\, dF_{\boldsymbol{v}}(\boldsymbol{v}) = \int 1\, dF_{\boldsymbol{v}}(\boldsymbol{v}) = 1$$

Based on properties P1 to P6, we prove the global invertibility result in Proposition 1.

Given a target vector of market shares $\boldsymbol{s}^* \in \Delta^{|\mathcal{J}^{\boldsymbol{a}}|}$ with each component satisfying $0 < s_j^* < 1$ for $j \in \mathcal{J}^{\boldsymbol{a}}$, we can establish existence of mean utilities that generate these shares. To formally establish existence using a fixed-point approach, we construct a rectangle:

$$\left[\underline{\boldsymbol{\delta}}, \overline{\boldsymbol{\delta}}\right] \;=\; \{\boldsymbol{\delta} \in \mathbb{R}^{|\mathcal{J}^{\boldsymbol{a}}|} : \underline{\delta}_j \leq \delta_j \leq \overline{\delta}_j \text{ for } j \in \mathcal{J}^{\boldsymbol{a}}\},$$

where we choose $\underline{\delta}_j$ sufficiently negative and $\overline{\delta}_j$ sufficiently positive such that:

$$d_j(\underline{\delta}_j, \boldsymbol{\delta}^{\boldsymbol{a}}_{-j}) \;<\; s_j^* \;<\; d_j(\overline{\delta}_j, \boldsymbol{\delta}^{\boldsymbol{a}}_{-j})$$

for any fixed $\boldsymbol{\delta}^{\boldsymbol{a}}_{-j}$ in the rectangle. The boundary behavior established in P6 above guarantees the existence of such values.

We define a continuous function $H : \left[\underline{\boldsymbol{\delta}}, \overline{\boldsymbol{\delta}}\right] \to \mathbb{R}^{|\mathcal{J}^{\boldsymbol{a}}|}$ by:

$$H_j(\boldsymbol{\delta}) = \delta_j + \ln(s_j^*) - \ln\left[d_j(\boldsymbol{\delta}^{\boldsymbol{a}})\right]$$

Next, define a mapping $G : \left[\underline{\boldsymbol{\delta}}, \overline{\boldsymbol{\delta}}\right] \to \left[\underline{\boldsymbol{\delta}}, \overline{\boldsymbol{\delta}}\right]$ by:

$$G_j(\boldsymbol{\delta}^{\boldsymbol{a}}) = \begin{cases} \underline{\delta}_j & \text{if } H_j(\boldsymbol{\delta}^{\boldsymbol{a}}) < \underline{\delta}_j, \\ H_j(\boldsymbol{\delta}^{\boldsymbol{a}}) & \text{if } \underline{\delta}_j \leq H_j(\boldsymbol{\delta}^{\boldsymbol{a}}) \leq \overline{\delta}_j, \\ \overline{\delta}_j & \text{if } H_j(\boldsymbol{\delta}^{\boldsymbol{a}}) > \overline{\delta}_j, \end{cases}$$

By Property P1, function G is continuous. Since Δ is compact and convex, Brouwer's Fixed Point Theorem guarantees the existence of $\boldsymbol{\delta}^* \in \Delta$ such that $G(\boldsymbol{\delta}^*) = \boldsymbol{\delta}^*$.

B Bootstrap Procedure for Second-Step Standard Errors

Our estimation procedure involves two steps (see section 5 for details). In the first step, we use a nonparametric sieve MLE to estimate the vector of $L_f(|\mathcal{K}| - 1) + L_P|\mathcal{K}|J$ parameters $\boldsymbol{\gamma}^{f,P}$ with elements $\{\gamma^f_{\kappa^*}, \gamma^P_{j\kappa^*} : \kappa^* = 1, 2, ..., |\mathcal{K}|; j = 1, 2, ..., J\}$, which govern the distribution of unobserved market types and the vector of entry probabilities for each unobserved type.

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. Then, we use a GMM to jointly

estimate the coefficients of the sieve approximation and the structural demand parameters $\boldsymbol{\theta}_\delta \equiv (\alpha, \boldsymbol{\beta}, \boldsymbol{\sigma})$. In particular, we approximate the selection bias term by $\boldsymbol{h}'_{j,t}\,\boldsymbol{\gamma}^{\psi}_{j}$, where $\boldsymbol{h}'_{j,t}$ is a vector of dimension $1 \times (L_\psi + 1)|\mathcal{K}|$ with elements $\{f_{j,\kappa^*}(\kappa^* \mid \boldsymbol{x}_t)\ \left(\ln P_j(\boldsymbol{x}_t, \kappa^*)\right)^{\ell} : \ell = 0, 1, ..., L_\psi; \kappa^* = 1, ..., |\mathcal{K}|\}$, and where $\boldsymbol{\gamma}^{\psi}_{j}$ is a vector of parameters of the same dimension with elements $\{\gamma^{\psi}_{\ell,j\kappa^*} : \ell = 0, 1, ..., L_\psi; \kappa^* = 1, ..., |\mathcal{K}|\}$.

Following Das et al. (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. However, given the sequential nature of the estimator, the standard errors of the estimates in the second step need to be corrected. One way to do this is to use the asymptotic approximations and formulas in Newey (2009). Given the complexity of our model, these however result in laborious calculations whose practical implementation must be adapted to the number of unobserved market types $|\mathcal{K}|$. For any given $|\mathcal{K}|$, the asymptotic formulas differ and the computer code must be modified accordingly.[15] To avoid this practical hurdle, we propose a convenient two-step bootstrap procedure. Importantly, this procedure does not require to repeatedly estimate the first step, which, even for moderate $|\mathcal{K}|$, may take up to several hours for each individual execution.

To avoid repeatedly estimating the first step (as in a nonparametric bootstrap), we use a two-step bootstrap procedure consisting of a first-step parametric bootstrap, based on the asymptotic normality of the first-step nonparametric MLE (which holds under standard regularity conditions),[16] followed by a second-step nonparametric bootstrap. In practice, the proposed two-step bootstrap procedure consists of the following steps:

1. Using the first-step estimates $\widehat{\boldsymbol{\gamma}}^{f,P}$ and their estimated variance-covariance matrix $\widehat{\boldsymbol{V}}_{f,P}$, draw a vector of parameters $\widehat{\boldsymbol{\gamma}}^{f,P}_{b}$ from the multivariate normal distribution $\mathcal{N}\left(\widehat{\boldsymbol{\gamma}}^{f,P}, \widehat{\boldsymbol{V}}_{f,P}\right)$.

2. For each draw $\widehat{\boldsymbol{\gamma}}^{f,P}_{b}$, calculate a corresponding $\boldsymbol{h}^{b}_{j,t}$ to be used in regression equation (47).

3. Generate S bootstrap samples. For given $\boldsymbol{h}^{b}_{j,t}$ and each of these bootstrap samples $s = 1, ..., S$, obtain an estimate $\widehat{\boldsymbol{\theta}}^{b,s}_{\delta}$ from regression equation (47).

4. Repeat B times steps 1-3, with $b = 1, ..., B$.

[15] Prior research has established the value of two-step bootstrap procedures, with notable applications in Armstrong et al. (2014), Gonçalves et al. (2023), and Cattaneo et al. (2019), among others. In particular, Gonçalves et al. (2023) demonstrate that in two-stage MLE procedures, bootstrap methods can effectively bypass coding errors that often arise when calculating asymptotic standard errors that involve complex first- and second-order derivatives.

[16] When $\boldsymbol{x}_t$ is discrete, the first-step nonparametric MLE estimator is $\sqrt{T}$-consistent and asymptotically normal, while with continuous $\boldsymbol{x}_t$ the rate of convergence will be slower than $\sqrt{T}$. Importantly, as discussed in section 5, this slower rate of convergence of the first-step estimator does not, however, affect the $\sqrt{T}$-consistency and asymptotic normality of the second-step estimator of the demand parameters $\boldsymbol{\theta}_\delta$.

This procedure generates $B \times S$ bootstrap estimates $\left(\widehat{\boldsymbol{\theta}}_{\delta}^{b,s}\right)_{b=1,\ldots,B;s=1,\ldots,S}$, which can be finally used to compute the standard errors of $\widehat{\boldsymbol{\theta}}_{\delta}$.

The consistency of two-step bootstrap procedures, such as the one we propose, is studied by Gonçalves et al. (2023). The validity of the procedure requires two main types of assumptions. First, one needs the two-step estimator to be consistent, including consistency of the first-step estimator and its asymptotic linear representation, along with regularity conditions on the second-step objective function. These conditions are inherently satisfied by our model's two-step estimation framework. Second, one needs additional regularity conditions such as *Assumption BG** in Gonçalves et al. (2023), which can be easily verified in our context: (i) The bootstrap estimator in the first step must be consistent and the first-step estimator must have a linear asymptotic representation. (ii) The second-step bootstrap objective function must be consistent. (iii) First-order derivatives of the second-step moment condition must satisfy a bootstrap uniform law of large numbers (ULLN). (iv) Bootstrap standardized moment conditions must converge in distribution to the original sample standardized moment conditions.

Our methodology also satisfies these additional assumptions. Our first-step estimator has a linear asymptotic representation and the first step of our bootstrap procedure involves parameter draws from the first-step estimator's distribution, satisfying the first requirement of *BG**.[17] Our second-step estimator is then a standard GMM, making the verification of the remaining conditions in *Assumption BG** straightforward.

[17] We implement the first-step estimator using Stata's gsem and incorporating restrictions sufficient to prevent label swapping and corner solutions. Given this, our first-step estimator has a linear asymptotic representation.