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

Victor Aguirregabiria † Alessandro Iaria ‡ Senay Sokullu §

April 8, 2026

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, Georgia, Glasgow, Harvard, 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 Generalized Method of Moments (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 same economic logic points to extensions to richer environments, including dynamic games of market entry and exit, though such extensions require adapting the first-step identification argument to the corresponding state variables.[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 (2026), 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

[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 5.4 provides details on the implementation of these counterfactuals.

this restriction, allowing firms to possess information about demand shocks at the time of 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 far 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 the class of models and underlying assumptions. Section 3 describes the structure of the selection problem within this class of models. Section 4 outlines our identification results and estimation method. In Section 5, we illustrate our method with an empirical application to the US airline industry. Finally, Section 6 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 Berry-Levinsohn-Pakes (BLP) framework (Berry et al., 1995). For the sake of notational simplicity, we focus on single-product firms. In Section A, we discuss the intuition for adapting the model and the selection-correction argument 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. Appendix B discusses how the same logic can be extended to dynamic games of firms' product entry and exit.

2.1 Demand

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

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

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

[3] While this assumption is standard in the literature on empirical industrial organization, there are important exceptions, such as structural models of entry that allow potential entrants to internalize network externalities across markets, e.g., 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.

of parameters. The term ε_{hjt} is unobserved to the researcher and is i.i.d. over (h, j, t) with type I extreme value distribution.

Following the standard specification, the average utility of product j is:

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

where α and $\boldsymbol{\beta}$ are parameters. Variable ξ_{jt} captures the characteristics of product j in market t unobserved to the researcher. Throughout the paper, we normalize $\mathbb{E}(\xi_{jt}) = 0$. Similarly, the component of utility that depends on consumer-level random coefficients takes the multiplicative form:

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

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 $\boldsymbol{\theta} \equiv (\alpha,\, \boldsymbol{\beta}',\, \boldsymbol{\sigma}')'$ the column vector of demand parameters.

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:

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

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 subsystem of demand equations that includes the market shares, average utilities, and characteristics of the products that are 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 $\boldsymbol{\delta}_t^{\boldsymbol{a}} = (\delta_{jt} : j \in \mathcal{J}_t^{\boldsymbol{a}})$. We represent this system as:

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

Proposition 1 establishes that, for any configuration of $\boldsymbol{a}_t$, the demand system (5) satisfies the invertibility property with respect to $\boldsymbol{\delta}_t^{\boldsymbol{a}}$ (Berry, 1994).

PROPOSITION 1. *Suppose that the outside option $j = 0$ is always offered. Fix any value of the vector $\boldsymbol{a}_t \in \{0, 1\}^J$ and define the set of feasible interior market shares for the offered products as*

$$\mathcal{S}^{\boldsymbol{a}} \equiv \left\{ \boldsymbol{s}^{\boldsymbol{a}} \in (0,1)^{|\mathcal{J}_t^{\boldsymbol{a}}|} : \sum_{j \in \mathcal{J}_t^{\boldsymbol{a}}} s_j < 1 \right\}.$$

Then the demand system in equation (5) defines a one-to-one mapping from $\mathbb{R}^{|\mathcal{J}_t^{\boldsymbol{a}}|}$ into $\mathcal{S}^{\boldsymbol{a}}$. Therefore, for every $\boldsymbol{s}_t^{\boldsymbol{a}} \in \mathcal{S}^{\boldsymbol{a}}$, the inverse function $\boldsymbol{\delta}_t^{\boldsymbol{a}} = (\boldsymbol{d}_t^{\boldsymbol{a}})^{-1}(\boldsymbol{s}_t^{\boldsymbol{a}}, \boldsymbol{\sigma})$ exists and is unique. ■

Proof of Proposition 1: See Appendix C.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:

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

Whenever $a_{jt} = 1$, we write the inverse-demand term explicitly as $d_{jt}^{-1}\left(\boldsymbol{s}_t^{\boldsymbol{a}_t}, \boldsymbol{\sigma}\right)$, where $\boldsymbol{s}_t^{\boldsymbol{a}_t}$ is the vector of observed market shares of the products offered in market t. This characterization of the selection term follows from working directly with the inverse demand system, as represented by equation (6).[4]

As discussed in Appendix A, 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 as $\mathbb{E}\left(\xi_{jt} \mid a_{jt} = 1\right)$ 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 $U_{hjt} \equiv \delta_{jt} + v_{ht,r_j} + (1 - \sigma)\ \varepsilon_{hjt}$, where variables v and ε are independently distributed, ε and $v + (1-\sigma)\varepsilon$ are i.i.d. type I extreme value, $\sigma \in [0,1]$ is a parameter, and v has a $C(\sigma)$ distribution as defined in Cardell (1997). This model implies:*

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

with:

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

If $a_{0t} = 1$ and $a_{jt} = 1$, this model implies that $s_{0t} > 0$ and $s_{jt} > 0$, and the inverse function $d_{jt}^{-1}\left(\boldsymbol{s}_t^{\boldsymbol{a}}, \sigma\right)$

[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.

exists regardless of the value of a_{it} for any product i different from j. It is straightforward to show that this inverse function has the following form:

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

and it implies the regression equation:

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

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:

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

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 fc_{jt} is the fixed entry cost. Variable ω_{jt} is unobserved to the researcher.

Given firms' entry decisions, the best response function in the Bertrand pricing game implies the following system of pricing equations:

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

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

[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.

individual incentives and the joint distribution of firms' equilibrium entry decisions.

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 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 entry cost. The vectors* $\boldsymbol{\eta}_{jt}$ *are assumed to be independently distributed across firms and independent of* $(\boldsymbol{\xi}_t, \boldsymbol{\kappa}_t, \boldsymbol{x}_t)$. *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* $\boldsymbol{x}_t$. ■

Assumption 1 provides a flexible specification of firms' information sets at the time of entry. By allowing the common-knowledge component $\boldsymbol{\kappa}_t$ to range from a minimal set of market-level signals to the full vector of demand and cost fundamentals $(\boldsymbol{\xi}_t, \boldsymbol{\omega}_t)$, the assumption nests environments with substantial uncertainty as well as those in which firms face effectively complete information about market conditions. Likewise, by introducing firm-specific private

[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.

information $\boldsymbol{\eta}_{jt}$, the assumption accommodates a broad class of incomplete-information entry games, while also allowing the special case of complete information when $\boldsymbol{\eta}_{jt}$ is degenerate. Finally, part (d), which assumes independence between the unobserved shocks and the observable covariates $\boldsymbol{x}_t$, is entirely standard in empirical IO. This exogeneity condition underpins identification in both demand estimation and market entry models and aligns with conventional econometric practice in the literature.

To simplify notation, in expressions for expected profits we sometimes write $\boldsymbol{\xi}_t$ as shorthand for the full vector of demand and cost fundamentals $(\boldsymbol{\xi}_t, \boldsymbol{\omega}_t)$. When the distinction matters, we keep the two components separate.

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

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

where $F_{j,\xi}\left(\boldsymbol{\xi}_t \mid \boldsymbol{\kappa}_t\right)$ is a CDF and represents firm j's beliefs about the distribution of $\boldsymbol{\xi}_t$ conditional on $\boldsymbol{\kappa}_t$. Function $fc(\boldsymbol{x}_{jt}, \boldsymbol{\kappa}_t, \boldsymbol{\eta}_{jt})$ represents the fixed cost and entry cost of operating in the market. In the special case where firms face no uncertainty about market conditions at the time of entry, the common-knowledge vector is $\boldsymbol{\kappa}_t = \boldsymbol{\xi}_t$ and firm j's profit function simplifies to $\pi_j(\boldsymbol{a}_{-j}, \boldsymbol{x}_t, \boldsymbol{\xi}_t, \boldsymbol{\eta}_{jt}) \;=\; VP_j(\boldsymbol{a}_{-j}, \boldsymbol{x}_t, \boldsymbol{\xi}_t) - fc(\boldsymbol{x}_{jt}, \boldsymbol{\eta}_{jt})$.

Assumption 1 states that this entry game can accommodate complete information if the distribution of each $\boldsymbol{\eta}_{jt}$ 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, this solution concept encompasses a complete information Nash Equilibrium (NE) when each $\boldsymbol{\eta}_{jt}$ 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.

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

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:*

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

2.4 Special cases in the literature

This framework offers a general formulation that encompasses, as special cases, existing models of endogenous product entry with a structural demand system. To highlight this generality, Table 1 summarizes several recent influential contributions, focusing on features relevant to this paper: the presence and nature of endogenous selection on unobservables, and in particular whether selection occurs on demand-side unobservables.

As shown in Table 1, most empirical applications of endogenous product entry assume that firms do not know demand or marginal-cost unobservables at the time they make entry decisions. In other words, these models rule out—by assumption—the possibility of selection on demand (variable-profit) unobservables. Two important exceptions are the models in Ciliberto et al. (2021) and Li et al. (2022).

Table 1: Models of endogenous product entry with structural differentiated-product demand

Paper	Industry - Product Entry	Selection on demand (or MC) unobservables	What firms know about ξ_t at entry	Unobservables in entry cost	Entry game
Aguirregabiria and Ho (2012)	US airlines - Route (city-pair)	NO, once airline & route FEs are accounted for.	ξ_t is unknown. Firms know airline & route FEs but not demand shocks ξ_t.	Only η_{jt}. Private information shocks assumed independent of ξ_t.	Incomplete information dynamic game
Sweeting (2013)	US radio - Station genre	NO, once observable lagged variables are accounted for.	ξ_t is unknown. Follows AR(1). Firms know lagged but not current ξ.	Only η_{jt}. Private information shocks assumed independent of ξ_t.	Incomplete information dynamic game
Eizenberg (2014)	US home PC - PC models	NO. Selection on entry cost unobservables, but are assumed independent of ξ_t.	ξ_t is unknown. *"Firms only observe the realizations of ξ_t after committing to product choices".*	Includes common knowledge κ_t but are assumed independent of ξ_t.	Complete information static game
Fan and Yang (2020)	US smartphones - Phone models	NO. Selection on entry cost unobservables, but are assumed independent of ξ_t.	ξ_t is unknown. Firms know brand FEs but not demand shocks ξ_t.	Includes common knowledge κ_t but are assumed independent of ξ_t.	Complete information static game
Ciliberto et al. (2021)	US airlines - Route (city-pair)	YES	ξ_t is known. Demand unobservables are known to firms when making product choices.	Includes common knowledge κ_t that can be correlated with ξ_t.	Complete information static game
Li et al. (2022)	US airlines - Route (city-pair)	YES	ξ_t is known. Demand unobservables are known to firms when making product choices.	Includes common knowledge κ_t that can be correlated with ξ_t.	Complete information static game
Bontemps et al. (2023)	US airlines - Airline's network of non-stop routes	NO. Selection on entry cost unobservables, but are assumed independent of ξ_t.	ξ_t is unknown. Firms don't know demand unobservables when making network choices.	Includes common knowledge κ_t but are assumed independent of ξ_t.	Complete information static game

3 Structure of the selection problem

3.1 Selection-bias function

For the selection problem, it is convenient to work directly with the inverse-demand outcome from Proposition 1. Recall from equation (6) that, for any observation with $a_{jt} = 1$, the demand inversion yields $d_{jt}^{-1}(\boldsymbol{s}_t^{\boldsymbol{a}_t}, \boldsymbol{\sigma}) = \alpha\, p_{jt} + \boldsymbol{x}_{jt}'\boldsymbol{\beta} + \xi_{jt}$. Whether product j is offered is determined by firm j's equilibrium entry decision:

$$a_{jt} \;=\; \mathbb{1}\left\{\pi_j^P(\boldsymbol{x}_t, \boldsymbol{\kappa}_t, \boldsymbol{\eta}_{jt}) \geq 0\right\}. \tag{18}$$

Since the sample of observations for which the demand equation holds is selected, we decompose the unobservable ξ_{jt} into its conditional mean and a residual: $\xi_{jt} = \lambda_j(\boldsymbol{x}_t) + \widetilde{\xi}_{jt}$, where $\lambda_j(\boldsymbol{x}_t) \equiv \mathbb{E}(\xi_{jt} \mid \boldsymbol{x}_t, a_{jt} = 1)$ is the *selection-bias function* and $\widetilde{\xi}_{jt}$ is mean independent of $(\boldsymbol{x}_t, a_{jt} = 1)$ by construction. This gives the following regression equation for any observation with $a_{jt} = 1$:

$$d_{jt}^{-1}\left(\boldsymbol{s}_t^{\boldsymbol{a}_t}, \boldsymbol{\sigma}\right) \;=\; \alpha\; p_{jt} + \boldsymbol{x}_{jt}'\;\boldsymbol{\beta} + \lambda_j(\boldsymbol{x}_t) + \widetilde{\xi}_{jt}. \tag{19}$$

The structure of the selection-bias function plays a key role in the identification of demand parameters $(\alpha,\, \boldsymbol{\beta},\, \boldsymbol{\sigma})$. To characterize this structure, define the *propensity score*—the probability of product j being offered, conditional on observables—as:

$$\overline{P}_j\left(\boldsymbol{x}_t\right) \;\equiv\; \Pr\left(a_{jt} = 1 \mid \boldsymbol{x}_t\right) \;=\; \int \mathbb{1}\left\{\pi_j^P(\boldsymbol{x}_t, \boldsymbol{\kappa}_t, \boldsymbol{\eta}_{jt}) \geq 0\right\}\; f_\kappa(\boldsymbol{\kappa}_t)\, f_\eta(\boldsymbol{\eta}_{jt})\, d\boldsymbol{\kappa}_t\, d\boldsymbol{\eta}_{jt}. \tag{20}$$

For values of $\boldsymbol{x}_t$ such that $\overline{P}_j(\boldsymbol{x}_t) > 0$, the selection-bias function can be written as:

$$\lambda_j(\boldsymbol{x}_t) \;=\; \frac{\mathbb{E}\left(\xi_{jt}\, a_{jt} \;\middle|\; \boldsymbol{x}_t\right)}{\overline{P}_j\left(\boldsymbol{x}_t\right)}. \tag{21}$$

In the econometrics literature on sample selection, it is well known that estimating equation (19) by instrumental variables—treating $\lambda_j(\boldsymbol{x}_t) + \widetilde{\xi}_{jt}$ as the composite error—is generally infeasible. The reason is that $\lambda_j(\boldsymbol{x}_t)$ is an unknown function of all exogenous variables in the model, so any candidate instrument would also enter the error term through the selection function and therefore violate the exclusion restriction (Wooldridge, 2010). A natural alternative is a control-function approach that explicitly accounts for the selection component $\lambda_j(\boldsymbol{x}_t)$. However, without additional structure, this term remains an unrestricted function of the same exogenous variables that appear in demand. Consequently, the direct effect of $\boldsymbol{x}_{jt}$ on consumer demand (captured by $\boldsymbol{\beta}$) cannot be separated from its indirect effect operating through selection. In other words, absent further restrictions, the demand parameters are not identified.

In this setting, the standard strategy in the literature is to impose conditions under which the selection term depends only on the propensity score $\overline{P}_j(\boldsymbol{x}_t)$, that is, $\lambda_j(\boldsymbol{x}_t) = \rho_j\left(\overline{P}_j(\boldsymbol{x}_t)\right)$. This restriction is powerful because it reduces the infinite-dimensional nuisance function $\lambda_j(\boldsymbol{x}_t)$ to a single-index object. Once this dimensionality reduction holds, the effect of observables on demand can be separated from their effect through selection, restoring identification. Estimation then follows a standard two-step procedure. In the first step, the propensity score $\overline{P}_j(\boldsymbol{x}_t)$ is estimated nonparametrically using data on $(a_{jt}, \boldsymbol{x}_t)$. In the second step, one recovers the structural parameters using semiparametric methods, such as the series estimators in Das et al. (2003) and Newey (2009), or the pairwise differencing approaches in Powell (2001) and Aradillas-Lopez (2012). The remaining endogeneity of price arising from the standard simultaneity problem in demand estimation can be handled in the usual way: valid instruments are characteristics of other products, $\boldsymbol{x}_{-jt}$—the familiar BLP-type instruments.

3.2 Failure of the ordinary propensity score

Unfortunately, the conditions under which the selection-bias correction depends solely on the propensity score can fail in models of endogenous product entry and oligopoly competition, even under simple specifications. Intuitively, entry decisions in these environments depend on equilibrium profitability, which is jointly determined by unobserved demand and cost variables of all the competing products. As a result, the selection rule need not satisfy the single-index structure required for the propensity score to eliminate selection bias.

Proposition 2 presents this benchmark result for the standard propensity-score logic. Under the exogeneity conditions maintained in the paper, LATE-style monotonicity of the entry indicator is equivalent to a scalar threshold representation of that indicator, as in Theorem 1 of Vytlacil (2002). These structural conditions are then sufficient for conditioning on the ordinary propensity score to eliminate selection bias, following Propositions 2 and 3 in Angrist (1997).

PROPOSITION 2. *Under independence between the unobservables* $(\boldsymbol{\xi}_t, \boldsymbol{\kappa}_t, \boldsymbol{\eta}_{jt})$ *and the observables* $\boldsymbol{x}_t$*, as in Assumption 1[d], consider the following conditions:*

a. Monotonicity: For any $\boldsymbol{x}$ *and* $\boldsymbol{x}'$*, either* $\mathbb{1}\{\pi_j^P(\boldsymbol{x}, \boldsymbol{\kappa}_t, \boldsymbol{\eta}_{jt}) \geq 0\} \geq \mathbb{1}\{\pi_j^P(\boldsymbol{x}', \boldsymbol{\kappa}_t, \boldsymbol{\eta}_{jt}) \geq 0\}$ *for all* $(\boldsymbol{\kappa}_t, \boldsymbol{\eta}_{jt})$*, or the inequality is reversed for all* $(\boldsymbol{\kappa}_t, \boldsymbol{\eta}_{jt})$*.*

b. Single-index representation: There are real-valued functions γ_{1j} *and* γ_{2j} *such that:* $\pi_j^P(\boldsymbol{x}_t, \boldsymbol{\kappa}_t, \boldsymbol{\eta}_{jt}) \geq 0 \iff \gamma_{1j}(\boldsymbol{x}_t) \;\geq\; \gamma_{2j}(\boldsymbol{\kappa}_t, \boldsymbol{\eta}_{jt})$*.*

c. Conditional independence: $\Pr(\xi_{jt}, a_{jt} \mid \boldsymbol{x}_t, \overline{P}_j(\boldsymbol{x}_t)) \;=\; \Pr(\xi_{jt}, a_{jt} \mid \overline{P}_j(\boldsymbol{x}_t))$*.*

Under the regularity conditions in Theorem 1 of Vytlacil (2002), conditions (a) and (b) are equivalent. Furthermore, if either condition (a) or (b) holds, then (c) is satisfied. Finally, condition (c) is necessary and sufficient for the ordinary propensity score $\overline{P}_j(\boldsymbol{x}_t)$ *to control for selection bias.* ■

Although the conditions in Proposition 2 involve endogenous objects and are not straightforward to verify, the proposition is useful as a benchmark. The following example uses a minimal two-product environment satisfying Assumption 1 to show that condition (c) easily fails in oligopoly models with endogenous entry, and that, as a consequence, the ordinary propensity score often cannot control for selection bias in demand estimation.

EXAMPLE 2 (***Failure of propensity-score sufficiency under oligopoly entry***)**.** *Consider a simple setting satisfying Assumption 1 and the following features: (i) a standard logit demand system (no random coefficients); (ii) constant marginal costs $c_i(\boldsymbol{x}_{it})$ and exogenous price–cost margins $PCM_i(\boldsymbol{x}_t)$, for $i \in \mathcal{J}$; (iii) the $J-1$ products $i \neq j$ are always offered, while product j's entry is endogenous; (iv) the demand shocks $\boldsymbol{\xi}_t = (\xi_{it} : i \in \mathcal{J})$ are common knowledge among firms at the entry stage, and there is no private information: in the notation of Assumption 1, we have $\boldsymbol{\kappa}_t = \boldsymbol{\xi}_t$, and $\boldsymbol{\eta}_{jt}$ is degenerate; (v) entry costs depend only on observables; and (vi) $PCM_j(\boldsymbol{x}_t)\, H_t > EC_j(\boldsymbol{x}_{jt})$.[7] For the remainder of this example, we omit the subscript t to simplify the notation.*

Entry condition. *Firm j's expected profit, upon entry, is $\pi_j^P = PCM_j(\boldsymbol{x})\, H\, s_j - EC_j(\boldsymbol{x}_j)$. Therefore, the entry condition $\pi_j^P \geq 0$ can be represented as:*

$$\frac{e^{v_j(\boldsymbol{x})+\xi_j}}{1 + e^{v_j(\boldsymbol{x}_j)+\xi_j} + \sum_{i\neq j} e^{v_i(\boldsymbol{x})+\xi_i}} \;\geq\; \frac{EC_j(\boldsymbol{x}_j)}{PCM_j(\boldsymbol{x})\ H} \tag{22}$$

where $v_i(\boldsymbol{x}) \;\equiv\; \boldsymbol{x}_i'\boldsymbol{\beta} - \alpha[PCM_i(\boldsymbol{x}) + c_i(\boldsymbol{x}_i)]$. After simple operations, we can represent this entry condition as a threshold rule for the demand unobservable ξ_j:[8]

$$a_j \;=\; \mathbb{1}\Big\{\xi_j \;\geq\; \ln\Big(1 + \sum_{i\neq j} e^{v_i(\boldsymbol{x})+\xi_i}\Big) - B_j(\boldsymbol{x})\Big\}, \tag{23}$$

with $B_j(\boldsymbol{x}) \;\equiv\; v_j(\boldsymbol{x}) + \ln\left[PCM_j(\boldsymbol{x})\ H - EC_j(\boldsymbol{x}_j)\right] - \ln\left[EC_j(\boldsymbol{x}_j)\right]$.

[7] Without this condition, the probability of entry for firm j, conditional on observables $\boldsymbol{x}$, would be zero: even if the firm's market share were arbitrarily close to one, its profit would still be negative. In that case, the selection problem would be irrelevant for those values of $\boldsymbol{x}$.

[8] Multiplying equation (22) by the denominators, we get:

$$PCM_j(\boldsymbol{x})\ H\ e^{v_j(\boldsymbol{x})+\xi_j} \geq EC_j(\boldsymbol{x}_j)\ \Big(1 + e^{v_j(\boldsymbol{x})+\xi_j} + \sum_{i\neq j} e^{v_i(\boldsymbol{x})+\xi_i}\Big).$$

Moving the term involving $e^{v_j(\boldsymbol{x})+\xi_j}$ to the left:

$$\big(PCM_j(\boldsymbol{x})\ H - EC_j(\boldsymbol{x}_j)\big)\ e^{v_j(\boldsymbol{x})+\xi_j} \geq EC_j(\boldsymbol{x}_j)\ \Big(1 + \sum_{i\neq j} e^{v_i(\boldsymbol{x})+\xi_i}\Big).$$

Taking logarithms and rearranging terms, we get equation (23)

Failure of monotonicity and single-index (conditions (a)–(b), Proposition 2). *In this example, failure of monotonicity can be verified directly. Consider* $\boldsymbol{x}$ *and* $\boldsymbol{x}'$ *such that* $[v_k(\boldsymbol{x}') - v_k(\boldsymbol{x})] > [B_j(\boldsymbol{x}') - B_j(\boldsymbol{x})] > 0$ *for some* $k \neq j$*, and* $v_i(\boldsymbol{x}') - v_i(\boldsymbol{x}) = 0$ *for all* $i \neq j, k$*. Define the function:*

$$\Delta_j(\xi_k) \equiv \ln\Big(1 + \sum_{i \neq j} e^{v_i(\boldsymbol{x}') + \xi_i}\Big) - \ln\Big(1 + \sum_{i \neq j} e^{v_i(\boldsymbol{x}) + \xi_i}\Big) - [B_j(\boldsymbol{x}') - B_j(\boldsymbol{x})] \tag{24}$$

This function is continuous and strictly increasing, with $\Delta_j(\xi_k) \to -[B_j(\boldsymbol{x}') - B_j(\boldsymbol{x})] < 0$ *as* $\xi_k \to -\infty$*, and* $\Delta_j(\xi_k) \to [v_k(\boldsymbol{x}') - v_k(\boldsymbol{x})] - [B_j(\boldsymbol{x}') - B_j(\boldsymbol{x})] > 0$ *as* $\xi_k \to +\infty$*. Hence, a crossing point exists: the ordering of the entry threshold reverses, so monotonicity fails. By equivalence, the single-index representation also fails.*

Failure of sufficiency of the propensity score (condition (c), Proposition 2). *Again, consider* $\boldsymbol{x}$ *and* $\boldsymbol{x}'$ *such that* $[v_k(\boldsymbol{x}') - v_k(\boldsymbol{x})] > [B_j(\boldsymbol{x}') - B_j(\boldsymbol{x})] > 0$ *for some* $k \neq j$*, and* $v_i(\boldsymbol{x}') - v_i(\boldsymbol{x}) = 0$ *for all* $i \neq j, k$*. By continuity of* $\overline{P}_j(\boldsymbol{x})$ *and* $\Pr(\xi_j, a_j \mid \boldsymbol{x})$ *with respect to* $v_k(\boldsymbol{x})$ *and* $B_j(\boldsymbol{x})$*, and given the structure of the entry condition in this example, one can adjust these values so that* $\overline{P}_j(\boldsymbol{x}') = \overline{P}_j(\boldsymbol{x})$*, while still having* $\Pr(\xi_j, a_j \mid \boldsymbol{x}') \neq \Pr(\xi_j, a_j \mid \boldsymbol{x})$*. Thus, the propensity score is not sufficient.* ■

This example shows that, even in simple environments, the ordinary propensity score need not be sufficient to control for selection bias. The reason is that the rival's component of the latent state, κ_{kt}, affects firm j's entry threshold through oligopoly competition. As a result, selection depends on the full latent state $\boldsymbol{\kappa}_t$, whereas the ordinary propensity score averages over $\boldsymbol{\kappa}_t$ and therefore loses relevant information. This motivates characterizing the selection-bias function in terms of entry probabilities conditional on $(\boldsymbol{x}_t, \boldsymbol{\kappa}_t)$.

3.3 Latent propensity scores and the selection-bias function

In this section, we derive a representation of the selection-bias function $\lambda_j(\boldsymbol{x}_t)$ in terms of equilibrium entry probabilities conditional on $(\boldsymbol{x}_t, \boldsymbol{\kappa}_t)$. Define the equilibrium entry probability of product j, conditional on both observables and common-knowledge unobservables, as

$$P_j(\boldsymbol{x}_t, \boldsymbol{\kappa}_t) \equiv \Pr\left(a_{jt} = 1 \mid \boldsymbol{x}_t, \boldsymbol{\kappa}_t\right).$$

We refer to $P_j(\boldsymbol{x}_t, \boldsymbol{\kappa}_t)$ as the *latent propensity score*, to highlight both its connection to the ordinary propensity score and its dependence on the latent state $\boldsymbol{\kappa}_t$. The ordinary propensity score is simply the average of the latent propensity scores over the distribution of $\boldsymbol{\kappa}$:

$$\overline{P}_j(\boldsymbol{x}_t) = \int P_j(\boldsymbol{x}_t, \boldsymbol{\kappa})\, f_\kappa(\boldsymbol{\kappa})\, d\boldsymbol{\kappa}. \tag{25}$$

Proposition 3 shows that, under only the information-structure conditions in Assumption 1, the selection-bias function admits a mixture representation in terms of these latent propensity scores. Unlike the ordinary propensity-score approach, this representation does not require any monotonicity or single-index assumption.

PROPOSITION 3. *Under Assumption 1, and for values of $\boldsymbol{x}_t$ such that $\overline{P}_j(\boldsymbol{x}_t) > 0$, the selection-bias function $\lambda_j(\boldsymbol{x}_t)$ admits the following mixture representation:*

$$\lambda_j(\boldsymbol{x}_t) = \int \left[\frac{P_j(\boldsymbol{x}_t, \boldsymbol{\kappa})}{\overline{P}_j(\boldsymbol{x}_t)}\right] \mu_j(\boldsymbol{\kappa})\, f_\kappa(\boldsymbol{\kappa})\, d\boldsymbol{\kappa}. \tag{26}$$

with $\mu_j(\boldsymbol{\kappa}) \equiv \mathbb{E}(\xi_{jt} \mid \boldsymbol{\kappa}_t = \boldsymbol{\kappa})$. ■

Proof: In Appendix C.2.

Proposition 3 provides the first building block for our identification strategy and estimation procedure to control for endogenous product selection in the estimation of demand systems. In Section 4, we establish semiparametric identification of the latent propensity score function using the empirical joint distribution of entry decisions across all products, $\Pr(a_{1t}, a_{2t}, \ldots, a_{Jt} \mid \boldsymbol{x}_t)$. The key intuition is that cross-product dependence in entry decisions reveals how these decisions are jointly affected by the common unobserved component $\boldsymbol{\kappa}_t$. Variation in this dependence structure allows us to recover the latent propensity score without imposing restrictive parametric assumptions. Combined with Proposition 3, this implies that once the selection-bias function is approximated by sieve, it can be represented as a linear index in constructed regressors. As a consequence, the demand parameters can be identified.

4 Identification and estimation

4.1 Setting

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 $\boldsymbol{s}_t$.

Recall the vector of demand parameters $\boldsymbol{\theta} \equiv (\alpha, \boldsymbol{\beta}', \boldsymbol{\sigma}')'$. Let $\boldsymbol{P} \equiv \{P_j(\boldsymbol{x}, \boldsymbol{\kappa}) : \forall (j, \boldsymbol{x}, \boldsymbol{\kappa})\}$ be the collection of equilibrium latent propensity scores, and let $\boldsymbol{f}_\kappa \equiv \{f_\kappa(\boldsymbol{\kappa}) : \forall \boldsymbol{\kappa}\}$ denote the distribution of the unobserved heterogeneity $\boldsymbol{\kappa}$. Finally, let $\boldsymbol{\mu} \equiv \{\mu_j(\boldsymbol{\kappa}) : \forall (j, \boldsymbol{\kappa})\}$ denote the collection of conditional expectations $\mu_j(\boldsymbol{\kappa}) = \mathbb{E}(\xi_{jt} \mid \boldsymbol{\kappa}_t = \boldsymbol{\kappa})$.

We adopt a two-step sequential identification strategy for $\boldsymbol{\theta}$. In the first step, we use the empirical distribution of firms' entry decisions to identify the equilibrium probabilities $\boldsymbol{P}$ and the distribution f_κ. In the second step, we exploit the structure of the selection-bias function in equation (28) to identify the demand parameters $\boldsymbol{\theta}$ and the incidental parameters $\boldsymbol{\mu}$.

The econometric model consists of three sets of equations: (i) the demand regression equations,

$$d_{jt}^{-1}\left(\boldsymbol{s}_t^{\boldsymbol{a}_t}, \boldsymbol{\sigma}\right) = \alpha\, p_{jt} + \boldsymbol{x}_{jt}'\, \boldsymbol{\beta} + \lambda_j(\boldsymbol{x}_t) + \widetilde{\xi}_{jt}, \tag{27}$$

(ii) the equation for the selection-bias function,

$$\lambda_j(\boldsymbol{x}_t) \;=\; \int \left[\frac{P_j\,(\boldsymbol{x}_t, \boldsymbol{\kappa})}{\overline{P}_j\,(\boldsymbol{x}_t)}\right] \mu_j(\boldsymbol{\kappa})\, f_\kappa(\boldsymbol{\kappa})\, d\boldsymbol{\kappa}, \tag{28}$$

and (iii) the joint distribution of firms' entry decisions conditional on the observed state variables implied by the equilibrium of the entry game. Under Assumption 1, this distribution has a nonparametric mixture representation:

$$\Pr\left(a_{1t}, a_{2t}, \ldots, a_{Jt} \mid \boldsymbol{x}_t\right) \;=\; \int \left(\prod_{j=1}^{J} P_j(\boldsymbol{x}_t, \boldsymbol{\kappa})^{a_{jt}} \;\left[1 - P_j(\boldsymbol{x}_t, \boldsymbol{\kappa})\right]^{1-a_{jt}}\right) f_\kappa(\boldsymbol{\kappa})\, d\boldsymbol{\kappa}. \tag{29}$$

Conditional on $(\boldsymbol{x}_t, \boldsymbol{\kappa}_t)$, firms' entry decisions depend only on their private shocks, which are assumed to be independent across firms. As a result, entry decisions are conditionally independent given $(\boldsymbol{x}_t, \boldsymbol{\kappa}_t)$. All residual dependence across firms' entry decisions—beyond what is explained by observables—is therefore driven by the unobservables in $\boldsymbol{\kappa}_t$. This structure provides the key source of identification of $(\boldsymbol{P}, f_\kappa)$. The joint distribution of $(a_{1t}, a_{2t}, \ldots, a_{Jt})$ conditional on $\boldsymbol{x}_t$, and in particular the cross-sectional correlation in entry decisions, reveals how the common unobservable $\boldsymbol{\kappa}_t$ shifts firms' entry probabilities. In other words, the dependence across firms' decisions encodes information about both the distribution f_κ and the probability functions $P_j(\boldsymbol{x}_t, \boldsymbol{\kappa}_t)$.

We establish identification of $(\boldsymbol{\theta}, \boldsymbol{\mu}, \boldsymbol{P}, f_\kappa)$ using a two-step sieve-based approach. Specifically, we construct a sequence of sieve spaces over the support of the continuous unobserved heterogeneity $\boldsymbol{\kappa}_t$ and use these approximating spaces to represent the unknown objects $\boldsymbol{\mu}$, $\boldsymbol{P}$, and f_κ as functions of $\boldsymbol{\kappa}$.

4.2 Two-Step Sieve-Based Approach

Equation (28) shows that the selection-bias function depends on the latent heterogeneity only through three primitive objects: the latent propensity score $P_j(\boldsymbol{x}, \boldsymbol{\kappa})$, the density $f_\kappa(\boldsymbol{\kappa})$, and the conditional mean function $\mu_j(\boldsymbol{\kappa})$. This decomposition is key, as it reduces the infinite-dimensional dependence on $\boldsymbol{\kappa}$ to a structured combination of these objects and provides the foundation for the sieve-approximation approach that follows.

The latent heterogeneity vector $\boldsymbol{\kappa}_t$ may be continuously distributed with support $\mathcal{K}$ on $\mathbb{R}^{d_\kappa}$. For example, in our differentiated-products setting, $\boldsymbol{\kappa}_t$ may collect the product-specific demand and cost unobservables, $\boldsymbol{\kappa}_t = (\xi_{jt}, \omega_{jt} : j \in \mathcal{J}) \in \mathbb{R}^{2J}$. Accordingly, $P_j(\boldsymbol{x}, \boldsymbol{\kappa})$, $f_\kappa(\boldsymbol{\kappa})$, and $\mu_j(\boldsymbol{\kappa})$ are all functions defined on the space $\mathbb{R}^{d_\kappa}$.

To approximate these objects, we employ a discrete sieve. For each integer $L \geq 1$, let $\{\mathcal{K}_1, \mathcal{K}_2, \ldots, \mathcal{K}_L\}$ denote a partition of the support $\mathcal{K} \subseteq \mathbb{R}^{d_\kappa}$ of $\boldsymbol{\kappa}_t$. Each element of the partition can be interpreted as a *latent market type* (or latent sieve class), thereby providing a finite-dimensional approximation to the underlying continuous heterogeneity.

For each latent market type $\ell = 1, 2, \ldots, L$, define its probability

$$\widetilde{f}_\ell \;\equiv\; \Pr\left(\boldsymbol{\kappa}_t \in \mathcal{K}_\ell\right) \;=\; \int_{\boldsymbol{\kappa} \in \mathcal{K}_\ell} f_\kappa(\boldsymbol{\kappa})\, d\boldsymbol{\kappa}. \tag{30}$$

Define the type-specific average propensity score

$$\widetilde{P}_{j,\ell}(\boldsymbol{x}) \;\equiv\; \frac{1}{\widetilde{f}_\ell} \int_{\boldsymbol{\kappa} \in \mathcal{K}_\ell} P_j(\boldsymbol{x}, \boldsymbol{\kappa})\, f_\kappa(\boldsymbol{\kappa})\, d\boldsymbol{\kappa}, \tag{31}$$

as well as the type-specific average of the function $\mu_j(\boldsymbol{\kappa})$:

$$\widetilde{\mu}_{j,\ell} \;\equiv\; \frac{1}{\widetilde{f}_\ell} \int_{\boldsymbol{\kappa} \in \mathcal{K}_\ell} \mu_j(\boldsymbol{\kappa})\, f_\kappa(\boldsymbol{\kappa})\, d\boldsymbol{\kappa}. \tag{32}$$

These definitions induce a finite-mixture approximation to the distribution of entry decisions:

$$\Pr\left(a_{1t}, a_{2t}, \ldots, a_{Jt} \mid \boldsymbol{x}_t\right) \;\approx\; \widetilde{\Pr}^L(\boldsymbol{a}_t \mid \boldsymbol{x}) \;\equiv\; \sum_{\ell=1}^{L} \widetilde{f}_\ell \left(\prod_{j=1}^{J} \widetilde{P}_{j,\ell}(\boldsymbol{x})^{a_{jt}} \left[1 - \widetilde{P}_{j,\ell}(\boldsymbol{x})\right]^{1-a_{jt}} \right). \tag{33}$$

Likewise, the finite partition induces the sieve approximation

$$\lambda_j(\boldsymbol{x}) \;\approx\; \widetilde{\lambda}_j^L(\boldsymbol{x}) \;\equiv\; \sum_{\ell=1}^{L} \left[\frac{\widetilde{P}_{j,\ell}(\boldsymbol{x})}{\overline{P}_j(\boldsymbol{x})}\right] \widetilde{\mu}_{j,\ell}\, \widetilde{f}_\ell \tag{34}$$

Expression (34) is the discrete-support analog of the mixture representation in Proposition 3. Once the latent space has been partitioned into L classes, the selection-bias function is summarized by the finite-dimensional objects $\{\widetilde{f}_\ell\}_{\ell=1}^L$, $\{\widetilde{P}_{j,\ell}(\boldsymbol{x})\}_{\ell=1}^L$, and $\{\widetilde{\mu}_{j,\ell}\}_{\ell=1}^L$. Therefore, if the first step delivers sufficiently accurate approximations to the class probabilities and the class-specific propensity scores, and if the second step provides a sufficiently accurate approximation to the function $\mu_j(\boldsymbol{\kappa})$, then the induced approximation to $\lambda_j(\boldsymbol{x})$ will also be accurate.

The finite-mixture model should therefore be interpreted as a sieve, not as a literal restriction that the true latent heterogeneity has finite support. As L increases and the partition becomes finer, the type probabilities $\widetilde{f}_\ell$ approximate the true distribution f_κ, the type-specific choice probabilities $\widetilde{P}_{j,\ell}(\boldsymbol{x})$ approximate the true latent propensity score function $P_j(\boldsymbol{x}, \boldsymbol{\kappa})$, and the type-specific averages $\widetilde{\mu}_{j,\ell}$ approximate the true conditional expectation function $\mu_j(\boldsymbol{\kappa})$. Under standard regularity conditions—specifically, boundedness and continuity of the functions in $\boldsymbol{\kappa}$, uniformly in $\boldsymbol{x}$—there exists a sequence of partitions indexed by L such that the approximation $\widetilde{\Pr}^L(\boldsymbol{a} \mid \boldsymbol{x})$ converges to the true $\Pr(\boldsymbol{a} \mid \boldsymbol{x})$ for every $\boldsymbol{a} \in \{0,1\}^J$ and every $\boldsymbol{x}$, and the approximation $\widetilde{\lambda}_j^L(\boldsymbol{x})$ converges to the true $\lambda_j(\boldsymbol{x})$ for every j and every $\boldsymbol{x}$. Hence, the finite-support model can be viewed as a sieve approximation to the continuous-support structural model.

Note that the parameters $\widetilde{\mu}_{j\ell}$ have a clear structural interpretation as $\mathbb{E}(\xi_{jt} \mid \boldsymbol{\kappa}_t \in \mathcal{K}_\ell)$. Thus, although the mixture model cannot be linked a priori with a specific partition of the continuous unobserved space, the estimation delivers an economically meaningful partition ex post. In particular, the estimates $\widetilde{\mu}_{j\ell}$ map each latent market type to the average unobserved demand for each product. For example, one type may feature high unobserved demand for products 1 and 2 but low for others—and this type may be associated with a low entry probability for product 1—while another type may exhibit high demand for products 1 and 3 and a high entry probability for product 1. This provides direct insight into how latent market types relate product selection to underlying demand heterogeneity.

4.3 First Step Identification

For a given number of support points L, equation (33) defines a nonparametric finite mixture model. This representation is closely related to discrete choice games with incomplete information and finite-support unobserved heterogeneity, as studied in Aguirregabiria and Mira (2019) and Xiao (2018). These papers draw on results from the nonparametric finite mixture literature—such as Hall and Zhou (2003), Allman et al. (2009), and Kasahara and Shimotsu (2014)—to establish identification of the mixture components $\{\widetilde{f}_\ell\}_{\ell=1}^L$, the component-specific choice probabilities $\{\widetilde{P}_{j,\ell}(\boldsymbol{x})\}_{\ell=1}^L$, and the number of latent types L. In particular, Theorem 4 and Corollary 5 in Allman et al. (2009) provide primitive conditions under which these objects are nonparametrically identified.

In nonparametric finite-mixture models, identification is obtained only up to a permutation of the latent types (*label swapping*). A key implication of our framework is that the latent heterogeneity $\boldsymbol{\kappa}$ is independent of the observed covariates $\boldsymbol{x}$. This independence ensures that any admissible relabelling of the latent types must be global, i.e., it applies uniformly across all values of $\boldsymbol{x}$. As a result, label swapping has no substantive consequences for subsequent analysis. In particular, it does not affect the construction of the control function in the second step of our identification and estimation procedure, since all objects entering that step are invariant to a common permutation of the latent types.

4.4 Second Step Identification

The second-step regression equation is:

$$d_{jt}^{-1}\left(\boldsymbol{s}_t^{a_t}, \boldsymbol{\sigma}\right) \;=\; \alpha\, p_{jt} + \boldsymbol{x}_{jt}'\boldsymbol{\beta} + \sum_{\ell=1}^{L} \widetilde{\mu}_{j\ell} \left[\frac{\widetilde{P}_{j,\ell}(\boldsymbol{x}_t)}{\overline{P}_j(\boldsymbol{x}_t)}\, \widetilde{f}_\ell \right] + \widetilde{\xi}_{jt}, \tag{35}$$

Hence, given the first-step objects, the selection-bias function is linear in the unknown coefficients $\{\widetilde{\mu}_{j\ell}\}_{\ell=1}^{L}$. This is an important implication of Proposition 3: the second-step unknown function $\mathbb{E}(\xi_{jt} \mid \boldsymbol{\kappa}_t = \boldsymbol{\kappa})$ enters the regression only through its type-averages, and those type-averages appear linearly once the first-step latent propensity scores and type probabilities are known.

Before establishing identification of the parameters in this regression equation, note a potential perfect-collinearity issue and the linear restriction on the $\widetilde{\mu}_{j\ell}$ that resolves it. First, by definition, the constructed regressors for the control function sum to one:

$$\sum_{\ell=1}^{L} \frac{\widetilde{P}_{j,\ell}(\boldsymbol{x}_t)\, \widetilde{f}_\ell}{\overline{P}_j(\boldsymbol{x}_t)} \;=\; \frac{\overline{P}_j(\boldsymbol{x}_t)}{\overline{P}_j(\boldsymbol{x}_t)} \;=\; 1 \tag{36}$$

Second, the definition of the $\widetilde{\mu}_{j\ell}$ parameters implies that

$$\sum_{\ell=1}^{L} \widetilde{\mu}_{j\ell}\, \widetilde{f}_\ell \;=\; \sum_{\ell=1}^{L} \mathbb{E}(\xi_{jt} \mid \boldsymbol{\kappa} \in \mathcal{K}_\ell)\ \Pr(\boldsymbol{\kappa} \in \mathcal{K}_\ell) \;=\; \mathbb{E}(\xi_{jt}) \;=\; 0 \tag{37}$$

Therefore, only $L-1$ of the L parameters $\widetilde{\mu}_{j\ell}$ are free. Taking type L as the reference type:

$$\widetilde{\mu}_{jL} \;=\; -\sum_{\ell=1}^{L-1} \frac{\widetilde{f}_\ell}{\widetilde{f}_L}\, \widetilde{\mu}_{j\ell}. \tag{38}$$

Substituting (38) into the selection-bias term,

$$\begin{aligned}\sum_{\ell=1}^{L} \widetilde{\mu}_{j\ell} \left[\frac{\widetilde{P}_{j,\ell}(\boldsymbol{x}_t)}{\overline{P}_j(\boldsymbol{x}_t)} \widetilde{f}_\ell \right] &= \sum_{\ell=1}^{L-1} \widetilde{\mu}_{j\ell} \left[\frac{\widetilde{P}_{j,\ell}(\boldsymbol{x}_t)}{\overline{P}_j(\boldsymbol{x}_t)} \widetilde{f}_\ell \right] - \left(\sum_{\ell=1}^{L-1} \frac{\widetilde{f}_\ell}{\widetilde{f}_L} \widetilde{\mu}_{j\ell} \right) \left[\frac{\widetilde{P}_{j,L}(\boldsymbol{x}_t)}{\overline{P}_j(\boldsymbol{x}_t)} \widetilde{f}_L \right] \\ &= \sum_{\ell=1}^{L-1} \widetilde{\mu}_{j\ell} \left[\frac{\widetilde{P}_{j,\ell}(\boldsymbol{x}_t) - \widetilde{P}_{j,L}(\boldsymbol{x}_t)}{\overline{P}_j(\boldsymbol{x}_t)} \widetilde{f}_\ell \right]\end{aligned} \tag{39}$$

Then the second-step regression can be written as

$$d_{jt}^{-1}\left(\boldsymbol{s}_t^{\boldsymbol{a}_t}, \boldsymbol{\sigma}\right) = \alpha\, p_{jt} + \boldsymbol{x}'_{jt}\boldsymbol{\beta} + \sum_{\ell=1}^{L-1} \widetilde{\mu}_{j\ell}\, r_{j\ell t} + \widetilde{\xi}_{jt}, \tag{40}$$

where the coefficient vector $\widetilde{\boldsymbol{\mu}}_j \equiv (\widetilde{\mu}_{j1}, \ldots, \widetilde{\mu}_{j,L-1})'$ contains the $L-1$ free parameters, and the constructed regressors $(r_{j1t}, r_{j2t}, ..., r_{j,L-1,t})$ have the following definition:

$$r_{j\ell t} = \frac{\widetilde{P}_{j,\ell}(\boldsymbol{x}_t) - \widetilde{P}_{j,L}(\boldsymbol{x}_t)}{\overline{P}_j(\boldsymbol{x}_t)} \widetilde{f}_\ell \tag{41}$$

Conditional on $\boldsymbol{\sigma}$, equation (40) is linear in the remaining parameters $(\alpha, \boldsymbol{\beta}, \widetilde{\boldsymbol{\mu}}_j)$. However, $\boldsymbol{\sigma}$ enters nonlinearly through the demand inverse $d_{jt}^{-1}(\boldsymbol{s}_t^{\boldsymbol{a}_t}, \boldsymbol{\sigma})$, so identification of the full parameter vector requires additional conditions. Proposition 4 provides sufficient conditions for local identification of all parameters through a standard rank condition.

PROPOSITION 4. *Let $\widetilde{\boldsymbol{\mu}}_j \equiv (\widetilde{\mu}_{j1}, \ldots, \widetilde{\mu}_{j,L-1})'$ and define the parameter vector $\boldsymbol{\vartheta}_j \equiv (\alpha, \boldsymbol{\beta}', \boldsymbol{\sigma}', \widetilde{\boldsymbol{\mu}}'_j)'$. Let $\boldsymbol{z}_{jt}$ denote a $[\dim(\boldsymbol{\sigma}) + 1] \times 1$ vector of instruments consisting of functions of the characteristics of products other than j, and define $\boldsymbol{r}_{jt} \equiv \left(r_{j1t}, r_{j2t}, \ldots, r_{j,L-1,t}\right)'$ and $\boldsymbol{w}_{jt} \equiv \left(\boldsymbol{z}'_{jt}, \boldsymbol{x}'_{jt}, \boldsymbol{r}'_{jt}\right)'$. Consider the vector-valued moment function evaluated over the selected sample:*

$$\boldsymbol{m}_j(\boldsymbol{\vartheta}_j) \equiv \mathbb{E}\left(\boldsymbol{w}_{jt} \left[d_{jt}^{-1}\left(\boldsymbol{s}_t^{\boldsymbol{a}_t}, \boldsymbol{\sigma}\right) - \alpha\, p_{jt} - \boldsymbol{x}'_{jt}\boldsymbol{\beta} - \boldsymbol{r}'_{jt}\widetilde{\boldsymbol{\mu}}_j \right] \,\middle|\, a_{jt} = 1 \right).$$

Suppose that the following conditions hold.

a. The objects $\left\{ \widetilde{f}_\ell, \widetilde{P}_{j,\ell}(\boldsymbol{x}_t) \right\}_{\ell=1}^{L}$ *are identified from the first step.*

b. The instruments are valid, so that $\mathbb{E}\left(\boldsymbol{w}_{jt}\, \widetilde{\xi}_{jt} \mid a_{jt} = 1 \right) = \mathbf{0}$.

c. The function $d_{jt}^{-1}\left(\boldsymbol{s}_t^{\boldsymbol{a}_t}, \boldsymbol{\sigma}\right)$ *is continuously differentiable in* $\boldsymbol{\sigma}$*, and the Jacobian matrix* $\boldsymbol{M}_j(\boldsymbol{\vartheta}_{j0}) \equiv \left. \frac{\partial \boldsymbol{m}_j(\boldsymbol{\vartheta}_j)}{\partial \boldsymbol{\vartheta}_j'} \right|_{\boldsymbol{\vartheta}_j = \boldsymbol{\vartheta}_{j0}}$ *has full column rank at the true parameter value* $\boldsymbol{\vartheta}_{j0}$.

Then, $\boldsymbol{\vartheta}_{j0}$ *is locally identified.* ■

4.4.1 Identification of marginal costs and fixed entry costs

As is standard in models of competition, once demand parameters are identified and a form of competition is specified (e.g., Bertrand competition), equilibrium price-cost margins PCM_{jt} can be recovered from firms' profit maximization conditions. Combining these with observed prices yields realized marginal costs mc_{jt}.

Estimating a marginal cost function—from the regression of mc_{jt} on observable product characteristics $\boldsymbol{x}_{jt}$ with an additive unobservable ω_{jt}—raises the same selection bias issues as in demand estimation. In particular, the regression involves the selection-bias function $\lambda_j^{mc}(\boldsymbol{x}_t) = \mathbb{E}(\omega_{jt} \mid \boldsymbol{x}_t, a_{jt} = 1)$. Fortunately, this selection-bias function has the same structure as in Proposition 3, though with different parameters:

$$\lambda_j^{mc}(\boldsymbol{x}_t) \;=\; \int \left[\frac{P_j\left(\boldsymbol{x}_t, \boldsymbol{\kappa}\right)}{\overline{P}_j\left(\boldsymbol{x}_t\right)} \right] \mu_j^{mc}(\boldsymbol{\kappa})\, f_\kappa(\boldsymbol{\kappa})\, d\boldsymbol{\kappa}. \tag{42}$$

with $\mu_j^{mc}(\boldsymbol{\kappa}) \equiv \mathbb{E}(\omega_{jt} \mid \boldsymbol{\kappa}_t = \boldsymbol{\kappa})$. Therefore, the same identification strategy can be applied to recover both the parameters in the marginal cost function and in the selection-bias function.

Given estimates of the demand and marginal cost parameters, together with the selection-bias parameters $\widetilde{\boldsymbol{\mu}}$ and $\widetilde{\boldsymbol{\mu}}^{mc}$, we can compute equilibrium variable profits for any given $\boldsymbol{x}_t$, latent market type $\ell = 1, \ldots, L$, and counterfactual entry configuration $\boldsymbol{a} \in \{0,1\}^J$. A key feature of our approach is that we estimate the expected unobserved demand and marginal costs, $\widetilde{\mu}_{j\ell}$ and $\widetilde{\mu}^{mc}_{j\ell}$, for every product j and latent type ℓ, regardless of whether the product is observed in the market. This allows us to compute counterfactual equilibrium outcomes for all products. In particular, for a given $\boldsymbol{x}_t$ and latent type ℓ, we can obtain equilibrium price-cost margins, market shares, and variable profits by setting $\xi_j = \widetilde{\mu}_{j\ell}$ and $\omega_j = \widetilde{\mu}^{mc}_{j\ell}$ for all $j \in \mathcal{J}$.

Let $VP_{j\ell}(\boldsymbol{a}, \boldsymbol{x}_t)$ denote the equilibrium variable profit of firm j evaluated at latent market type ℓ under entry configuration $\boldsymbol{a} \in \{0,1\}^J$. Using firms' entry probabilities, expected variable profits at the time of entry decisions can be written as:

$$VP^P_{j\ell}(\boldsymbol{x}_t) \;=\; \sum_{\boldsymbol{a}_{-j}\in\{0,1\}^{J-1}} VP_{j\ell}(a_j = 1, \boldsymbol{a}_{-j}, \boldsymbol{x}_t) \prod_{i\neq j} \widetilde{P}_{i,\ell}(\boldsymbol{x}_t)^{a_i} \left[1 - \widetilde{P}_{i,\ell}(\boldsymbol{x}_t)\right]^{1-a_i}. \tag{43}$$

Up to this point, the private information entering fixed costs has been allowed to be the general vector $\boldsymbol{\eta}_{jt}$. To recover fixed costs from equilibrium entry probabilities, we now impose the additional restriction that this private information can be summarized by a scalar shock η_{jt} entering additively in fixed costs. Specifically, suppose fixed costs take the form $fc_{j\ell}(\boldsymbol{x}_{jt}) + \sigma_{\eta_j}\eta_{jt}$, where η_{jt} has mean zero and known strictly increasing CDF F_η. Then, the equilibrium entry probabilities satisfy $\widetilde{P}_{j,\ell}(\boldsymbol{x}_t) = F_\eta\left(\left[VP^P_{j\ell}(\boldsymbol{x}_t) - fc_{j\ell}(\boldsymbol{x}_{jt})\right]/\sigma_{\eta_j}\right)$, and can therefore be inverted to obtain the regression-like equation:

$$F^{-1}_\eta\left(\widetilde{P}_{j,\ell}(\boldsymbol{x}_t)\right) \;=\; \frac{1}{\sigma_{\eta_j}} VP^P_{j\ell}(\boldsymbol{x}_t) \;-\; \frac{1}{\sigma_{\eta_j}} fc_{j\ell}(\boldsymbol{x}_{jt}) \tag{44}$$

Under the exclusion restriction that firm j's fixed cost depends only on its own characteristics $\boldsymbol{x}_{jt}$ (and not on those of other firms), this equation identifies both the scale parameter σ_{η_j} and the fixed cost function $fc_{j\ell}(\boldsymbol{x}_{jt})$ for each product and latent market type.

4.5 Estimation

In this section, we present a two-step estimation method that mimics our two-step sieve identification approach. In the first step, for a given number of market types L, we use a nonparametric sieve maximum likelihood method to estimate the distribution of unobserved market types, and the vector of entry probabilities for each unobserved type. In the second step, we construct the control variables r and apply the Generalized Method of Moments (GMM) to estimate demand parameters and control function parameters. Estimating the number of latent market types L warrants special attention, and we discuss it in Section 4.5.3.

4.5.1 First step: Estimation of conditional choice probabilities (CCPs) and distribution of latent types

For a given number of latent types L, we approximate the nonparametric functions $\{\widetilde{P}_{j\ell}(\boldsymbol{x}_t) : \forall(j,\ell)\}$ using sieves as functions of $\boldsymbol{x}_t$ (Hirano et al., 2003, Chen, 2007). For estimation, we specialize the general CDF F_η above to the Logistic case, so that $F_\eta = \Lambda$. Let $\boldsymbol{b}_t \equiv (b_1(\boldsymbol{x}_t), \ldots, b_{N_X}(\boldsymbol{x}_t))'$ be a vector with a finite number N_X of basis functions. For any product j and any latent type ℓ, the entry probability function $\widetilde{P}_{j\ell}(\boldsymbol{x}_t)$ has the following sieves binary logit structure:

$$\widetilde{P}_{j\ell}(\boldsymbol{x}_t) \;=\; \Lambda\left(\boldsymbol{b}_t'\,\gamma_{j\ell}\right), \tag{45}$$

where $\Lambda(\cdot)$ is the logistic function, and $\gamma_{j\ell}$ is a vector of parameters of dimension $N_X \times 1$.

Let $\gamma \equiv \{\gamma_{j\ell} : \forall(j,\ell)\}$ be the vector of $J\,L\,N_X$ parameters in the sieve approximation to the entry probabilities. And let $\widetilde{\boldsymbol{f}} \equiv \{\widetilde{f}_\ell : \forall\ell\}$ be the vector of probabilities for the latent types. The log-likelihood function of the finite mixture model is:

$$\ln \mathcal{L}_{1st}(\widetilde{\boldsymbol{f}}, \gamma) \;=\; \sum_{t=1}^{T} \ln\left(\sum_{\ell=1}^{L} \widetilde{f}_\ell \prod_{j=1}^{J} \Lambda\left(\boldsymbol{b}_t'\,\gamma_{j\ell}\right)^{a_{jt}} \left[1-\Lambda\left(\boldsymbol{b}_t'\,\gamma_{j\ell}\right)\right]^{1-a_{jt}}\right). \tag{46}$$

We estimate $(\widetilde{\boldsymbol{f}}, \gamma)$ by maximum likelihood estimation (MLE) using the Expectation-Maximization (EM) algorithm (Pilla and Lindsay, 2001).[9] The sieve is constructed using polynomial basis functions in $\boldsymbol{x}_t$, $\{b_n(\boldsymbol{x}_t) : n = 1, 2, \ldots N_X\}$. We select the dimension of this basis using the

[9]Recent applications of MLE–EM methods to nonparametric mixture models in discrete choice settings include Bunting (2022), Bunting et al. (2022), Hu and Xin (2022), and Williams (2020).

Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC), thereby balancing approximation flexibility against overfitting. Section 4.5.3 describes in detail the separate selection of the number of latent types L.

When $\boldsymbol{x}_t$ has discrete support, the nonparametric MLE achieves $\sqrt{T}$-consistency and asymptotic normality. With continuous covariates, the first-step nonparametric estimator converges at a slower rate. Crucially, however, this slower convergence does not compromise inference on the demand parameters in the second step. Under standard regularity conditions, the second-step estimator remains $\sqrt{JT}$-consistent and asymptotically normal. This result follows from the general semiparametric efficiency arguments in Hirano et al. (2003) and Das et al. (2003), among others.

4.5.2 Second step: Estimation of demand parameters

Given estimates from the first step, $(\widehat{\widehat{\boldsymbol{f}}}, \widehat{\gamma})$, the estimation of demand parameters in the second step is based on applying standard GMM to the regression demand equations augmented by the linear in parameters control function, as presented in equation (40). By definition, the constructed regressors for the control function are:

$$\widehat{r}_{j\ell t} \;=\; \frac{\widehat{\widehat{P}}_{j,\ell}(\boldsymbol{x}_t) - \widehat{\widehat{P}}_{j,L}(\boldsymbol{x}_t)}{\widehat{\widehat{P}}_j(\boldsymbol{x}_t)} \, \widehat{\widehat{f}}_\ell \tag{47}$$

Following Das et al. (2003), this two-step estimator of demand parameters is $\sqrt{JT}$-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. The correct standard errors can be computed using either the asymptotic approximations and formulas in Newey (2009) or the linearized bootstrap procedure we detail in Appendix D. Conditional on the selected first-step sieve specification, a key computational advantage of this bootstrap procedure is that it does not require repeated estimation of the first step.

4.5.3 Estimating the number of latent types

L is a tuning parameter for approximating the selection-bias term well enough to identify and estimate demand parameters consistently. For this reason, the main criterion to select L should be based on the demand equation and not so much on the goodness-of-fit of the joint entry distribution in the first step of the method. A value of L that is nearly optimal for approximating the entry distribution need not be optimal for approximating the control function. Specifically, a criterion that gives substantial weight to the goodness-of-fit of the entry distribution might select a value of L that is too small for the optimal approximation of the selection-bias function. This can easily happen when increasing L produces only a modest improvement in the likelihood of the entry model, but generates new directions of variation in the control variables r that matter a lot for the second-step regression and for the estimates of demand parameters. Therefore, selecting L using only the first-step fit may be misaligned with our ultimate objective.

That said, the goodness-of-fit in the first step should not be completely ignored when selecting L. The first step imposes a feasibility constraint: L must be small enough that the finite-mixture model is identified and can be estimated reliably.

For these reasons, we combine two BIC criteria to select L. The first-step criterion is a likelihood-based BIC from the estimation of the entry distribution mixture model:

$$\text{BIC}_{1\text{st}}(L) \;=\; -2 \, \ln \mathcal{L}_{1\text{st}}(\widehat{\widetilde{\boldsymbol{f}}}, \widehat{\gamma}, L) + (J \, L \, N_X + L - 1) \, \ln(T) \tag{48}$$

where $\mathcal{L}_{1\text{st}}(\widehat{\widetilde{\boldsymbol{f}}}, \widehat{\gamma}, L)$ is the likelihood function in the first-step estimation, and $(J \, L \, N_X + L - 1)$ is the number of parameters. The second-step criterion is a BIC based on the demand equation residuals:

$$\text{BIC}_{2\text{nd}}(L) \;=\; (TJ) \, \ln \widehat{\sigma}^2_{\xi}(L) + (L-1)J \, \ln(TJ) \tag{49}$$

where $\widehat{\sigma}^2_{\xi}(L)$ is the sample variance of the residuals in the estimation of the system of demand equations. Note that $(L-1)J$ represents the number of parameters in the control function.

We use the minimization of $\text{BIC}_{\text{2nd}}(L)$ as our selection criterion, and use the first-step criterion $\text{BIC}_{\text{1st}}(L)$ only as a constraint and diagnostic.

5 Empirical application

5.1 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).

Data sources. 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.

Airlines. 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).[10] Given the large number of carriers included in Others, we do not consider this combined group as a player in the entry game. In the notation of our model, j always indexes an airline.

Markets in the demand model. In the demand model, a market t is defined as a *directional airport pair* in a given quarter. For example, LGA→ORD in Q1 2012 and ORD→LGA in Q1 2012 are two distinct demand markets. In each demand market t, consumers choose among

[10] 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.

the airlines offering non-stop flights on that directional route (up to seven airlines, plus the outside option). Thus, s_{jt} denotes the market share of airline j on directional route t.

Markets in the entry model. In the entry model, a market t is defined as a *non-directional airport pair* in a given quarter, where, for example, Chicago O'Hare (ORD) to New York La Guardia (LGA) is the same market as LGA to ORD. Each non-directional entry market thus corresponds to two directional demand markets. 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.[11] The estimated control-function variables $\widehat{\boldsymbol{r}}_{jt}$ entering the demand equation for a given directional market are constructed from the entry probabilities estimated at the corresponding non-directional market level.

Potential entrants. 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.

Market size. 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.

Observable variables in $\boldsymbol{x}_{jt}$. The vector of exogenous observable variables at the airline-market level includes: market size, market distance, squared market distance, the airline's own hub-size in the origin airport, its hub-size in the destination airport, airline indicators, and quarter

[11] 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.

indicators. 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.

Table 2 presents the distribution of the number of entrants and averages of the market 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 2: 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 3 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

875–950 miles, whereas the markets served by LCC have an average distance of 1,171 miles.

Table 3: 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

5.2 First step: Estimation of the model of market entry

For the entry decisions, we consider the nonparametric sieve finite mixture Logit described in equation (46). We explore various specifications of the mixture Logit model based on the polynomial order in $\boldsymbol{x}_t$ used to construct the basis $\boldsymbol{b}_t$ and the number of latent market types L. As our estimates of the demand parameters are robust to the selection of the basis $\boldsymbol{b}_t$ in the entry model, we only present results here for the specification with $\boldsymbol{b}_t = \boldsymbol{x}_t$. Regarding the number of latent types L, Table 4 presents the goodness-of-fit statistics obtained from estimating four nested specifications of the mixture Logit model. The goodness-of-fit 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 latent market types improves the entry model's goodness-of-fit. Comparing the specification without $\boldsymbol{\kappa}_t$ and the one with two unobserved market types in Table 4, 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 $L = 3$. 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 $L = 3$ and $L = 4$ are practically indistinguishable. In contrast, with $L \leq 2$ we obtain estimates of the own-price elasticities which are substantially smaller.

Table 4: 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	$41,458$	$39,385$	$38,170$	$38,041$

5.3 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.

$$\ln\left(\frac{s_{jt}}{s_{0t}}\right) = \alpha\, p_{jt} + \boldsymbol{x}'_{jt}\,\boldsymbol{\beta} + \sigma\, \ln\left(\frac{s_{jt}}{1 - s_{0t}}\right) + \widehat{\boldsymbol{r}}'_{jt}\,\widetilde{\boldsymbol{\mu}}_j + \widetilde{\xi}_{jt}. \tag{50}$$

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 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, $\widehat{\boldsymbol{r}}'_{jt}\ \widetilde{\boldsymbol{\mu}}_j$, 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 without latent types.* We consider the entry model $a_{jt} = \mathbb{1}\{\eta_{jt} \leq \boldsymbol{x}'_{jt}\gamma_j^P\}$, with $\eta_{jt} \sim Logistic$, and $\xi_{jt} = \widetilde{\mu}_j\ \eta_{jt} + v_{jt}$, with v_{jt} independent of η_{jt} and $\boldsymbol{x}_t$. In this parametric specification, the selection term is given by the expected value of a truncated Logistic variable, which can be interpreted as the Logit analogue of the Heckman selection correction term:

$$\mathbb{E}\left(\xi_{jt} \mid a_{jt} = 1, \boldsymbol{x}_{jt}\right) \ = \ \widetilde{\mu}_j\ \mathbb{E}\left(\eta_{jt} \mid \eta_{jt} \leq \boldsymbol{x}'_{jt}\gamma_j^P\right) \ = \ \widetilde{\mu}_j\ m_j(\boldsymbol{x}_{jt}), \tag{51}$$

   where $m_j(\boldsymbol{x}_{jt})$ is the expectation of a truncated Logistic in terms of the truncation probability:

$$m_j(\boldsymbol{x}_{jt}) \equiv \ln \overline{P}_j(\boldsymbol{x}_{jt}) + \frac{1 - \overline{P}_j(\boldsymbol{x}_{jt})}{\overline{P}_j(\boldsymbol{x}_{jt})} \ln\left(1 - \overline{P}_j(\boldsymbol{x}_{jt})\right), \tag{52}$$

   and $\overline{P}_j(\boldsymbol{x}_{jt}) \equiv \Lambda\left(\boldsymbol{x}'_{jt}\gamma_j^P\right)$ is the propensity score implied by the Logit entry model.

2. *Semiparametric without latent types.* The entry model is still the Logit $a_{jt} = \mathbb{1}\{\eta_{jt} \leq \boldsymbol{x}'_{jt}\gamma_j^P\}$, with $\eta_{jt} \sim Logistic$, but now $\mathbb{E}\left(\xi_{jt} \mid a_{jt} = 1, \boldsymbol{x}_{jt}\right)$ is approximated by a third order polynomial in the scalar control variable $m_j(\boldsymbol{x}_{jt})$. This is a one-dimensional control variable because, under the single-index restriction without latent types, the selection term depends on observables only through the propensity score.[12] This semiparametric

[12]Both in this and in the case of the semiparametric mixture Logit, estimates are very similar by approximating

approach to control for selection follows the standard series-based control-function strategy in Newey (2009).

3. *Finite mixture Logit with latent types.* This is our entry model described above, with $\widehat{\boldsymbol{r}}'_{jt} = (\widehat{r}_{j1t}, \widehat{r}_{j2t}, \ldots, \widehat{r}_{j,L-1,t})$ constructed from the first-step estimates.

For all the two-stage least squares (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 using the linearized bootstrap procedure detailed in Appendix D.

Table 5 presents the estimates of the demand parameters, while Table 6 reports the average demand elasticities and Lerner indexes derived from these estimates. Comparing the estimates obtained using ordinary least squares (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 $L = 3$ are, compared to those of "Semiparametric" (assuming $L = 1$), 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 $L = 1$) are relatively

the selection function with polynomials of higher orders.

similar. In contrast, parameter estimates and corresponding own-price elasticities remarkably change when we allow $L > 1$. Although the estimated own-price elasticities of a model with $L = 2$ are still meaningfully different from those of a model with $L = 3$, the estimates implied by models with $L = 3$ and $L = 4$ 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.

Table 5: Estimation of Demand Parameters

	Not control. for sel.		*Controlling for endogenous selection*				
	OLS	2SLS	2SLS Heckman $L = 1$	2SLS Semipar. $L = 1$	2SLS Fin.-Mix. $L = 2$	2SLS Fin.-Mix. $L = 3$	2SLS Fin.-Mix. $L = 4$
Price (100\$) (α)	−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 are based on the linearized bootstrap procedure in Appendix D and account for first-step estimation error conditional on the selected first-step specification.

Figure 1 plots the empirical distributions of the estimated own-price elasticities. Each row corresponds to an airline, while each column to a different 2SLS estimator: the first column

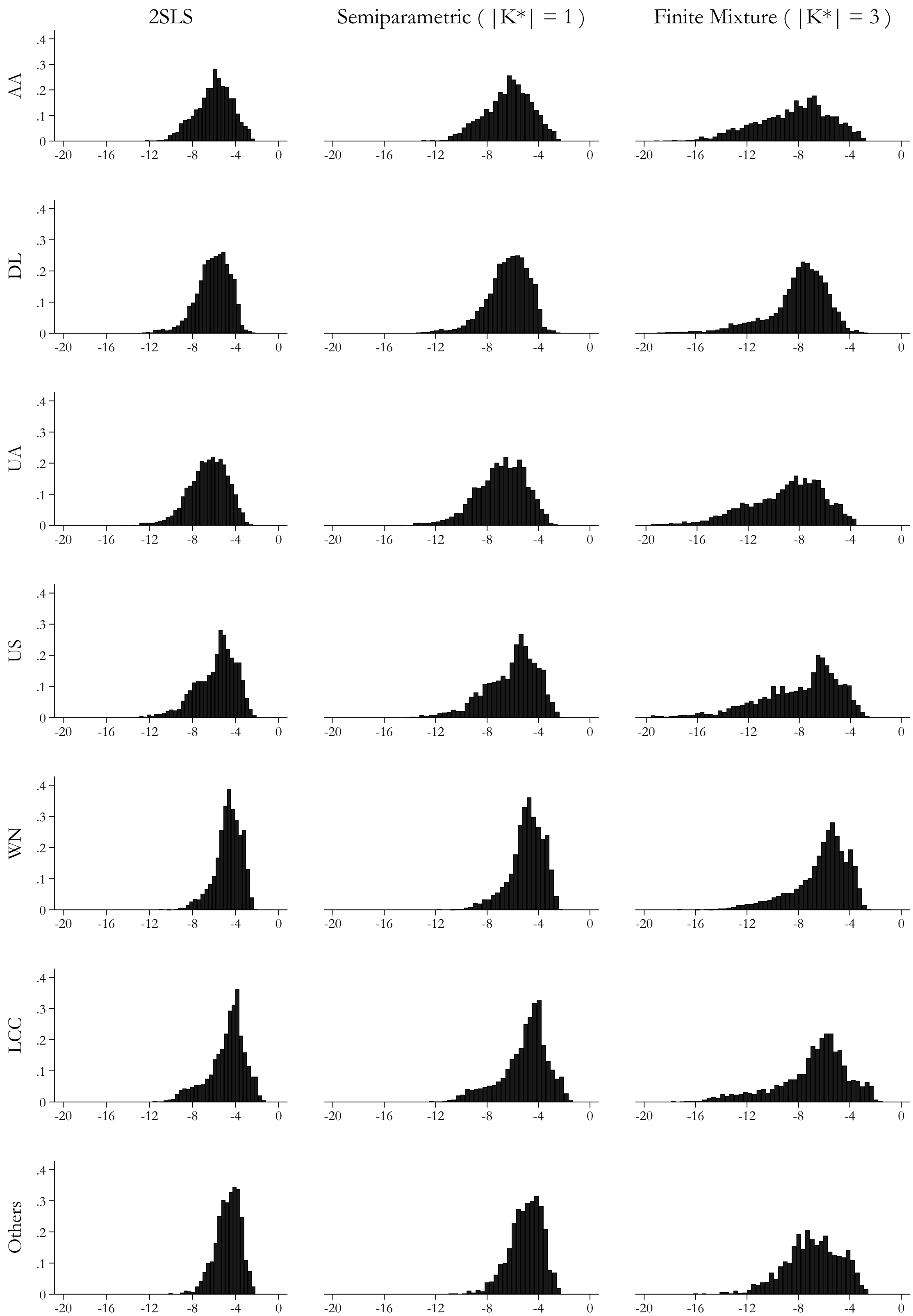
Table 6: Average Own-Price Elasticities and Lerner Indexes

	Not control. for sel.		*Controlling for endogenous selection*				
	OLS	2SLS	2SLS Heckman $L = 1$	2SLS Semipar. $L = 1$	2SLS Fin.-Mix. $L = 2$	2SLS Fin.-Mix. $L = 3$	2SLS Fin.-Mix. $L = 4$
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

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_{jt}, market shares s_{jt} and s_{0t}, 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.

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.

5.4 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.

5.4.1 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 characteristics $\boldsymbol{x}_{jt}$ and, in the case of non-linear returns to scale, the output q_{jt}. As in the 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 (50), the marginal cost function of airline j in market-quarter t can be estimated from the following pricing equation:

$$p_{jt} + \frac{1-\sigma}{\alpha(1-\sigma s_{jt|g} - (1-\sigma)s_{jt})} = mc_{jt}, \tag{53}$$

where g denotes the nest that contains all the airlines, $s_{jt|g} \equiv s_{jt}/(1-s_{0t})$ is the within-nest market share, and the marginal cost mc_{jt} is specified as:

$$mc_{jt} \;=\; \boldsymbol{x}'_{jt}\,\boldsymbol{\varphi} + \widehat{\boldsymbol{r}}'_{jt}\,\widetilde{\boldsymbol{\mu}}^{\text{mc}}_{j} + \widetilde{\omega}_{jt}, \tag{54}$$

with both $\boldsymbol{x}_{jt}$ and $\widehat{\boldsymbol{r}}_{jt}$ defined as in the case of demand equation (50), while $\widetilde{\boldsymbol{\mu}}^{\text{mc}}_{j}$ is a vector of $L-1$ parameters $\widetilde{\mu}^{\text{mc}}_{j\ell}$ with $\widetilde{\mu}^{\text{mc}}_{j\ell} = \mathbb{E}(\omega_{jt} \mid \boldsymbol{\kappa}_t \in \mathcal{K}_\ell)$.

Table 7 reports the average marginal costs obtained from equation (53) and the demand estimates in Table 5 (see Appendix Figure 2 for the corresponding empirical distributions), while Table 8 presents our estimates of $\boldsymbol{\varphi}$ from equation (54). The estimates of $\boldsymbol{\varphi}$ in each column of Table 8 rely on the corresponding demand estimates of Table 5, so that, for example, the first column of Table 8 reports estimates of $\boldsymbol{\varphi}$ obtained by using the estimates of α and σ (i.e., plugging them in the left-hand side of (53)) from the first column of Table 5. Collectively, these results illustrate that although endogeneity of prices and of within-nest market shares play an important role in the implied marginal cost estimates from equation (53), endogenous selection

Table 7: Average Marginal Costs

	Not control. for sel.		*Controlling for endogenous selection*				
	OLS	2SLS	2SLS Heckman $L=1$	2SLS Semipar. $L=1$	2SLS Fin.-Mix. $L=2$	2SLS Fin.-Mix. $L=3$	2SLS Fin.-Mix. $L=4$
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

seems to have less of an impact. Moreover, the parameter estimates of equation (54) (which uses the estimated mc_{jt} as a dependent variable) look remarkably similar across *all* columns of Table 8, 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 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

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

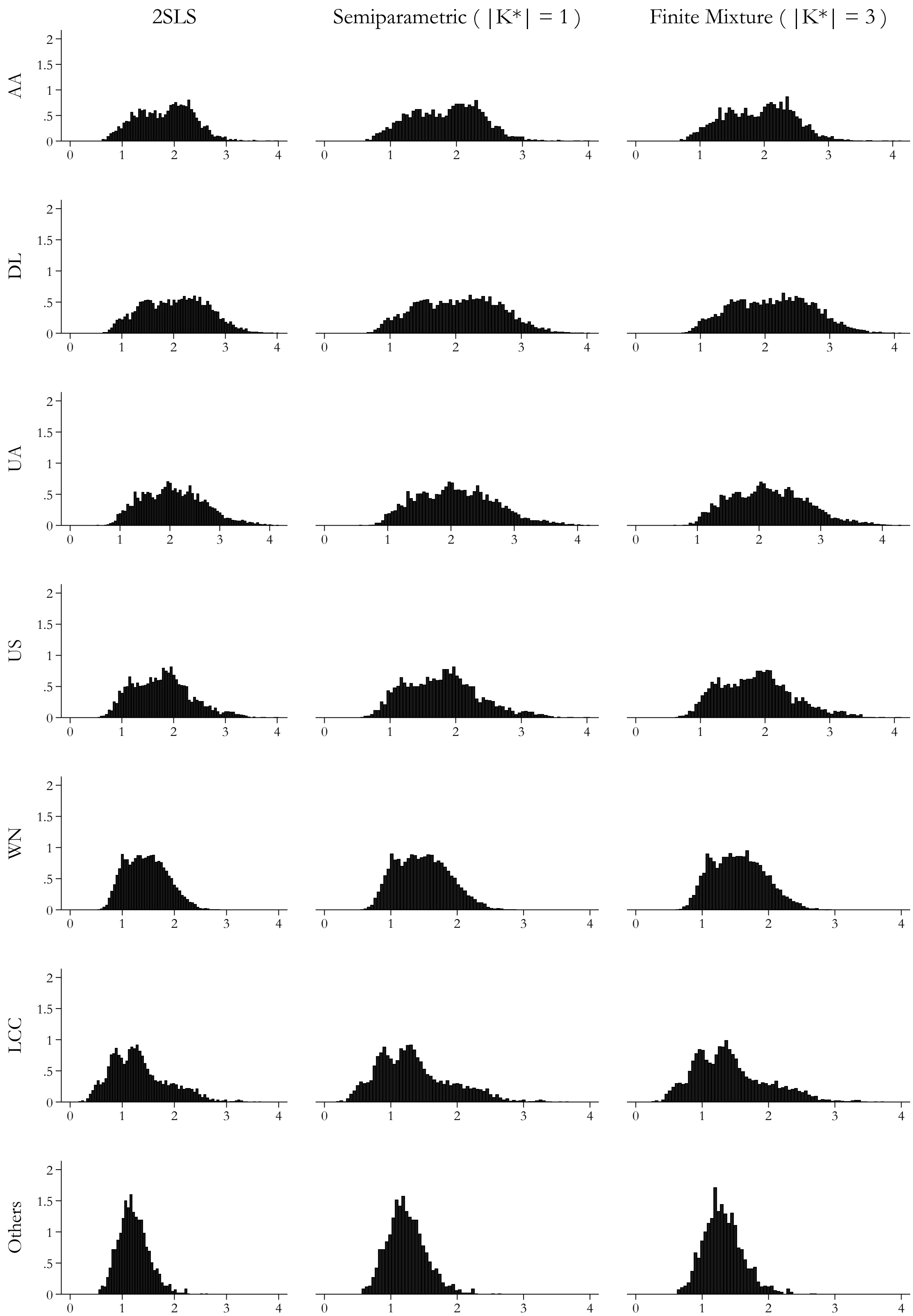


Table 8: Estimation of Marginal Cost Parameters

	Not control. for sel.		*Controlling for endogenous selection*				
	OLS	2SLS	2SLS Heckman $L=1$	2SLS Semipar. $L=1$	2SLS Fin.-Mix. $L=2$	2SLS Fin.-Mix. $L=3$	2SLS Fin.-Mix. $L=4$
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$\times$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 are based on the linearized bootstrap procedure in Appendix D and account for first-step estimation error conditional on the selected first-step specification.

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.

Appendices

A Multi-product firms

We briefly discuss how the proposed model and the intuition behind the selection problem in Section 3 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 the applicability of Proposition 1 is unaffected by the product ownership structure. This proposition relies only on the structure of the demand system. Therefore, regardless of ownership, the selection problem in estimating demand for product j is still described by the conditional expectation $\mathbb{E}\left(\xi_{jt} \mid a_{jt} = 1\right)$.

Second, the same basic information structure can be maintained at the time of entry, except that a firm's private information must now be indexed by the product portfolios it may choose. In particular, we represent this private information as $\boldsymbol{\eta}_{ft} \equiv (\eta_{ft}(\boldsymbol{a}_f) : \boldsymbol{a}_f \in \{0,1\}^{|\mathcal{J}_f|})$. We also define $\boldsymbol{x}_{ft} \equiv (\boldsymbol{x}_{jt} : j \in \mathcal{J}_f)$. For instance, in the case of a two-product firm, $\boldsymbol{\eta}_{ft} \equiv (\eta_{ft}(0,0),\, \eta_{ft}(0,1), \eta_{ft}(1,0),\, \eta_{ft}(1,1))$ where, say, $\eta_{ft}(1,0)$ is the latent component of entry cost when the firm offers product 1 while excluding product 2. Under the same information structure as in Assumption 1, equation (15), which describes the expected profit of a firm, extends to multi-product firms as follows:

$$\pi_f\left(\boldsymbol{a}_f, \boldsymbol{a}_{-f}, \boldsymbol{x}_t, \boldsymbol{\kappa}_t, \boldsymbol{\eta}_{ft}\right) = \int VP_f\left(\boldsymbol{a}_f, \boldsymbol{a}_{-f}, \boldsymbol{x}_t, \boldsymbol{\xi}_t\right)\, dF_{f,\xi}\left(\boldsymbol{\xi}_t \mid \boldsymbol{\kappa}_t\right) - fc(\boldsymbol{x}_{ft}, \boldsymbol{\kappa}_t, \boldsymbol{\eta}_{ft}), \tag{55}$$

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

Given this expected-profit representation, the relevant equilibrium objects in the entry model for multi-product firms are firm-level probabilities of choosing product portfolios. Accordingly, the first-step finite-mixture representation in the main text would need to be reformulated at the portfolio level rather than at the level of individual product entries. The underlying intuition is nevertheless the same: common latent heterogeneity generates dependence across firms' portfolio choices, and the product-level selection term for any given product is obtained by aggregating over the portfolios that include that product.

B Dynamic game of product entry and exit

Our framework can also be adapted to 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 $fc(a_{jt}, a_{j,t-1}, \boldsymbol{x}_{jt}, \boldsymbol{\kappa}_t, \boldsymbol{\eta}_{jt})$, where $fc(1, 0, \boldsymbol{x}_{jt}, \boldsymbol{\kappa}_t, \boldsymbol{\eta}_{jt})$ represents the cost of entry, $fc(0, 1, \boldsymbol{x}_{jt}, \boldsymbol{\kappa}_t, \boldsymbol{\eta}_{jt})$ is the cost of exit, $fc(1, 1, \boldsymbol{x}_{jt}, \boldsymbol{\kappa}_t, \boldsymbol{\eta}_{jt})$ is the fixed cost when a product remains in the market, and $fc(0, 0, \boldsymbol{x}_{jt}, \boldsymbol{\kappa}_t, \boldsymbol{\eta}_{jt})$ can be normalized to zero.

ASSUMPTION 1-Dyn. *Suppose that t represents time. Conditions (a) to (c) in Assumption 1 hold, and we have the following additional conditions.*

e. Let $\boldsymbol{x}_t \equiv \left(\boldsymbol{x}_t^{ex}, (a_{j,t-1} : j = 1, 2, ..., J)\right)$ denote the observed state vector at period t, where $\boldsymbol{x}_t^{ex}$ collects the exogenous product characteristics and $(a_{j,t-1} : j = 1, 2, ..., J)$ are the lagged entry decisions of all firms.

f. The exogenous product characteristics $\boldsymbol{x}_t^{ex}$ and the latent market type $\boldsymbol{\kappa}_t$ follow a first-order Markov process or are time-invariant.

g. The private information shock $\boldsymbol{\eta}_{jt}$ is independently and identically distributed over time and independent across firms. ■

The conditions in Assumption 1-Dyn are standard in the literature on 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 observed state vector $\boldsymbol{x}_t = \left(\boldsymbol{x}_t^{ex}, (a_{j,t-1} : j = 1, 2, ..., J)\right)$, on the latent market type $\boldsymbol{\kappa}_t$, and on the private information shock $\boldsymbol{\eta}_{jt}$. Let $v_j^P(\boldsymbol{x}_t, \boldsymbol{\kappa}_t, \boldsymbol{\eta}_{jt})$ 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 the static entry game, an MPE in a dynamic game can be characterized in terms of J conditional choice probabilities (CCPs). These dynamic CCPs play the same role as the latent propensity scores in the static model, so the logic of our selection-correction argument extends naturally to the dynamic case. In many applications, one can therefore proceed by treating the relevant dynamic CCPs as the key first-step objects. If one also wishes to model persistent latent heterogeneity together with endogenous state dependence more formally, the first-step finite-mixture representation can be adapted in the standard way to condition on the relevant dynamic state or initial conditions.

DEFINITION 2. ***Markov Perfect Equilibrium.*** *Suppose that Assumption 1-Dyn holds. 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:*

$$P_j\left(\boldsymbol{x}_t, \boldsymbol{\kappa}_t\right) \;=\; \int \mathbb{1}\left\{v_j^P(\boldsymbol{x}_t, \boldsymbol{\kappa}_t, \boldsymbol{\eta}_{jt}) \geq 0\right\} \, dF_\eta\left(\boldsymbol{\eta}_{jt}\right). \qquad \blacksquare \tag{56}$$

For exposition, we use the same notation $P_j\left(\boldsymbol{x}_t, \boldsymbol{\kappa}_t\right)$ for these dynamic choice probabilities. Thus, at the level of economic intuition, the dynamic case is a direct extension of the static one: dynamic CCPs replace static entry probabilities, and the same selection-correction logic applies once these objects are available.

C Proofs of Propositions

C.1 Proof of Proposition 1

Fix a configuration of offered products $\mathcal{J}_t^{\boldsymbol{a}}$ and let $n \equiv |\mathcal{J}_t^{\boldsymbol{a}}|$. For $j \in \{0\} \cup \mathcal{J}_t^{\boldsymbol{a}}$, let $P_j(\boldsymbol{\delta^a}, \boldsymbol{\sigma}, \boldsymbol{v})$ denote the multinomial logit choice probability conditional on $\boldsymbol{v}$, and let $d_j(\boldsymbol{\delta^a}, \boldsymbol{\sigma})$ be its integral with respect to F_v. The subsystem in equation (5) is therefore just a random-coefficients logit demand system with product set $\mathcal{J}_t^{\boldsymbol{a}}$ and outside option 0.

Let $\boldsymbol{J}(\boldsymbol{\delta^a})$ denote the $n \times n$ Jacobian matrix of the mapping $\boldsymbol{d}_t^{\boldsymbol{a}}(\boldsymbol{\delta^a}, \boldsymbol{\sigma})$, with typical entry

$$J_{ji}(\boldsymbol{\delta^a}) \equiv \frac{\partial d_j(\boldsymbol{\delta^a}, \boldsymbol{\sigma})}{\partial \delta_i}.$$

Standard derivatives of the multinomial logit probabilities imply that, for every $j \in \mathcal{J}_t^{\boldsymbol{a}}$,

$$J_{jj}(\boldsymbol{\delta^a}) = \int P_j(\boldsymbol{\delta^a}, \boldsymbol{\sigma}, \boldsymbol{v}) \Big[1 - P_j(\boldsymbol{\delta^a}, \boldsymbol{\sigma}, \boldsymbol{v}) \Big] \, dF_v(\boldsymbol{v}) > 0,$$

and, for every $i \neq j$,

$$J_{ji}(\boldsymbol{\delta^a}) = -\int P_j(\boldsymbol{\delta^a}, \boldsymbol{\sigma}, \boldsymbol{v}) P_i(\boldsymbol{\delta^a}, \boldsymbol{\sigma}, \boldsymbol{v}) \, dF_v(\boldsymbol{v}) < 0.$$

Hence, for every row j,

$$J_{jj}(\boldsymbol{\delta^a}) - \sum_{i \neq j} |J_{ji}(\boldsymbol{\delta^a})| = \int P_j(\boldsymbol{\delta^a}, \boldsymbol{\sigma}, \boldsymbol{v}) P_0(\boldsymbol{\delta^a}, \boldsymbol{\sigma}, \boldsymbol{v}) \, dF_v(\boldsymbol{v}) > 0,$$

because the outside option is always available and therefore $P_0 > 0$. Thus $\boldsymbol{J}(\boldsymbol{\delta^a})$ is a strictly diagonally dominant Z-matrix with positive diagonal for every $\boldsymbol{\delta^a}$. Every principal submatrix has the same property, so $\boldsymbol{J}(\boldsymbol{\delta^a})$ is a P-matrix. By the Gale–Nikaido theorem, the mapping $\boldsymbol{d}_t^{\boldsymbol{a}}(\boldsymbol{\delta^a}, \boldsymbol{\sigma})$ is globally injective on $\mathbb{R}^n$.

To establish existence for every feasible share vector in $\mathcal{S}^{\boldsymbol{a}}$, note that the subsystem is exactly the setting covered by the standard BLP inversion argument in Berry (1994): the outside option is always present, each inside share lies in $(0, 1)$, and the outside share equals $1 - \sum_{j \in \mathcal{J}_t^a} s_j > 0$. Therefore, for every $\boldsymbol{s}_t^a \in \mathcal{S}^a$, there exists a vector $\boldsymbol{\delta}_t^a$ such that

$$\boldsymbol{d}_t^a(\boldsymbol{\delta}_t^a, \boldsymbol{\sigma}) = \boldsymbol{s}_t^a.$$

Global injectivity then implies that this vector is unique. Hence the inverse mapping $(\boldsymbol{d}_t^a)^{-1}$ exists and is unique on $\mathcal{S}^a$. ■

C.2 Proof of Proposition 3

By definition, the selection-bias function is:

$$\lambda_j(\boldsymbol{x}_t) \;=\; \mathbb{E}\left(\xi_{jt} \mid \boldsymbol{x}_t, a_{jt} = 1\right) \;=\; \frac{\mathbb{E}\left(\xi_{jt}\, a_{jt} \mid \boldsymbol{x}_t\right)}{\Pr\left(a_{jt} = 1 \mid \boldsymbol{x}_t\right)} \tag{57}$$

By Assumption 1, the entry decision a_{jt} is a structural function of the common-knowledge variables $(\boldsymbol{x}_t, \boldsymbol{\kappa}_t)$ and the private information $\boldsymbol{\eta}_{jt}$. Therefore, independence between ξ_{jt} and $\boldsymbol{\eta}_{jt}$ implies:

$$\begin{aligned}
\mathbb{E}\left(\xi_{jt}\, a_{jt} \mid \boldsymbol{x}_t\right) &= \mathbb{E}\left(\mathbb{E}\left(\xi_{jt}\, a_{jt} \mid \boldsymbol{x}_t, \boldsymbol{\kappa}_t\right) \mid \boldsymbol{x}_t\right) \\
&= \mathbb{E}\left(\mathbb{E}\left(\xi_{jt} \mid \boldsymbol{x}_t, \boldsymbol{\kappa}_t\right) \mathbb{E}\left(a_{jt} \mid \boldsymbol{x}_t, \boldsymbol{\kappa}_t\right) \mid \boldsymbol{x}_t\right) \\
&= \mathbb{E}\left(\mu_j(\boldsymbol{\kappa}_t)\, P_j(\boldsymbol{x}_t, \boldsymbol{\kappa}_t) \mid \boldsymbol{x}_t\right)
\end{aligned} \tag{58}$$

The second equality follows because, conditional on $(\boldsymbol{x}_t, \boldsymbol{\kappa}_t)$, the entry decision depends only on the private information shock $\boldsymbol{\eta}_{jt}$, which is independent of ξ_{jt} by Assumption 1[c]. The third equality uses the definition of the latent propensity score,

$$\mathbb{E}\left(a_{jt} \mid \boldsymbol{x}_t, \boldsymbol{\kappa}_t\right) = P_j(\boldsymbol{x}_t, \boldsymbol{\kappa}_t),$$

and Assumption 1[d], which implies that the distribution of $(\xi_{jt}, \boldsymbol{\kappa}_t)$ does not depend on $\boldsymbol{x}_t$ and therefore

$$\mathbb{E}\left(\xi_{jt} \mid \boldsymbol{x}_t, \boldsymbol{\kappa}_t\right) = \mathbb{E}\left(\xi_{jt} \mid \boldsymbol{\kappa}_t\right) = \mu_j(\boldsymbol{\kappa}_t).$$

Using (58) in (57), and noting that $\Pr\left(a_{jt} = 1 \mid \boldsymbol{x}_t\right) = \overline{P}_j(\boldsymbol{x}_t)$, we obtain:

$$\lambda_j(\boldsymbol{x}_t) \;=\; \frac{\mathbb{E}\left(\mu_j(\boldsymbol{\kappa}_t)\, P_j(\boldsymbol{x}_t, \boldsymbol{\kappa}_t) \mid \boldsymbol{x}_t\right)}{\overline{P}_j(\boldsymbol{x}_t)}. \tag{59}$$

Finally, Assumption 1[d] implies that the conditional distribution of $\boldsymbol{\kappa}_t$ given $\boldsymbol{x}_t$ is simply $f_\kappa(\boldsymbol{\kappa})$. Therefore,

$$\lambda_j(\boldsymbol{x}_t) \;=\; \int \left[\frac{P_j(\boldsymbol{x}_t, \boldsymbol{\kappa})}{\overline{P}_j(\boldsymbol{x}_t)}\right] \mu_j(\boldsymbol{\kappa})\, f_\kappa(\boldsymbol{\kappa})\, d\boldsymbol{\kappa}. \tag{60}$$

This proves Proposition 3. ■

D Bootstrap Procedure for Second-Step Standard Errors

Our estimation procedure involves two steps. In the first step, we use a nonparametric sieve MLE to estimate the vector of parameters $(\widetilde{f}, \gamma)$ which govern the distribution of latent market types and the entry probabilities for each latent type. In the second step, we use first-step estimates to construct the control variables in the linear-in-parameters control function, and then apply GMM to jointly estimate demand parameters $\boldsymbol{\theta}$ and the parameters in the control function, $\widetilde{\boldsymbol{\mu}} \equiv \{\widetilde{\boldsymbol{\mu}}_j\}_{j=1}^{J}$. Following Das et al. (2003) and Newey (2009), this two-step estimator is asymptotically linear under standard regularity conditions. Chen et al. (2003) provide general bootstrap results for semiparametric estimators with preliminary nonparametric components, while Armstrong et al. (2014) and Gonçalves et al. (2023) motivate fast linearized resampling methods for multi-step estimators.

Re-estimating the first-step sieve MLE in every bootstrap replication would be computationally expensive in our setting. For this reason, and conditional on the selected value of L and the selected sieve basis, we use a one-step linearized common-weight bootstrap. The idea is to approximate the infeasible bootstrap that would re-estimate the first-step likelihood in each replication, while preserving the joint effect of the common resampling shock on both stages. The key requirement is that the same bootstrap perturbation be used in both steps, so that the bootstrap reproduces the covariance between first-step estimation error and second-step sampling error.

Let $u = 1, 2, \ldots, U$ index the resampling units in the first-step sample, which in our application correspond to non-directional market-quarter observations. [13] Let $\boldsymbol{\psi}$ denote a finite-dimensional unconstrained parameterization of the first-step parameters, combining the sieve coefficients in γ and an unconstrained reparameterization of the latent-type probabilities. Specifically, the type probabilities are mapped to the real line via a softmax transformation,

[13] The same resampling units are used in both steps. In the second step, $\widehat{m}_u(\boldsymbol{\theta}, \widetilde{\boldsymbol{\mu}}; \boldsymbol{\psi})$ aggregates all directional-market observations belonging to non-directional market-quarter unit u, so that the bootstrap weights operate at the market level throughout.

$\widetilde{f}_\ell = \exp(c_\ell) / \sum_{k=1}^{L} \exp(c_k)$, and $\boldsymbol{\psi}$ collects the $L-1$ free log-ratio parameters $(c_1, \ldots, c_{L-1})$ together with γ. This ensures that all components of $\boldsymbol{\psi}$ are unbounded, so that the linear Newton update in Step 2 below cannot violate the simplex constraint on $\widetilde{\boldsymbol{f}}$.

Let $\ell_u(\boldsymbol{\psi})$ denote the contribution of unit u to the first-step log-likelihood in (46), and let $\dot{\ell}_u(\boldsymbol{\psi}) \equiv \partial \ell_u / \partial \boldsymbol{\psi}$ denote the corresponding score contribution.[14] Also let $\widehat{\boldsymbol{m}}_u(\boldsymbol{\theta}, \widetilde{\boldsymbol{\mu}}; \boldsymbol{\psi})$ denote the stacked second-step GMM moment contribution associated with unit u, when the control variables are constructed from $\boldsymbol{\psi}$. More generally, $\widehat{\boldsymbol{m}}_u$ may aggregate all second-step observations associated with the same resampling unit. Define the outer-product-of-gradients (OPG) estimator of the information matrix:

$$\widehat{\boldsymbol{H}} \equiv \frac{1}{U} \sum_{u=1}^{U} \dot{\ell}_u(\widehat{\boldsymbol{\psi}})\, \dot{\ell}_u(\widehat{\boldsymbol{\psi}})', \tag{61}$$

where $\widehat{\boldsymbol{\psi}}$ is the estimated first-step parameter vector. The OPG matrix is positive semi-definite by construction and avoids the computation of second derivatives of the mixture log-likelihood, which can be numerically unstable near the boundary of the parameter space. Under correct specification, it is a consistent estimator of the Fisher information matrix and is therefore asymptotically equivalent to $-\mathbb{E}[\partial^2 \ell_u / \partial \boldsymbol{\psi}, \partial \boldsymbol{\psi}']$, the negative expected Hessian of the log-likelihood.[15]

1. **Step 1: Draw common bootstrap weights.** For bootstrap replication b, draw multinomial weights $\{w_u^{*(b)} : u = 1, \ldots, U\}$, equivalently by sampling the resampling units with replacement. The same weights are used in both steps of the bootstrap replication.

2. **Step 2: Update the first step by a one-step correction.** Starting from $\widehat{\boldsymbol{\psi}}$, construct the

[14] The procedure requires that individual unit-level scores $\dot{\ell}_u(\widehat{\boldsymbol{\psi}})$ be available at the MLE. In our GSEM/EM framework these are computed analytically as a byproduct of the E-step.

[15] While the OPG estimator is positive semi-definite by construction, positive semi-definiteness does not guarantee invertibility. In finite-mixture models, near-boundary solutions or near-collinear mixture components can cause $\widehat{H}$ to be close to singular. In practice, we find $\hat{H}$ to be invertible in our application; if near-singularity were encountered, a Moore–Penrose pseudoinverse or small ridge correction could be applied.

bootstrap first-step update

$$\widehat{\boldsymbol{\psi}}^{*(b)} = \widehat{\boldsymbol{\psi}} + \widehat{\boldsymbol{H}}^{-1} \left[\frac{1}{U} \sum_{u=1}^{U} \left(w_u^{*(b)} - 1 \right) \dot{\ell}_u(\widehat{\boldsymbol{\psi}}) \right]. \tag{62}$$

This is the first-order approximation to the first-step estimator that would be obtained by re-estimating the sieve likelihood on the bootstrap sample. Because all components of $\boldsymbol{\psi}$ are unconstrained, the linear update cannot the linear update cannot violate the simplex constraint on $\widetilde{\boldsymbol{f}}$.[16] The updated vector $\widehat{\boldsymbol{\psi}}^{*(b)}$ is mapped back into $(\widehat{\widetilde{\boldsymbol{f}}}^{*(b)}, \widehat{\gamma}^{*(b)})$ via the softmax transformation for the type probabilities, while keeping the local ordering of latent types fixed. Keeping the local ordering fixed ensures that bootstrap draws remain in the labeling region of the full-sample MLE, avoiding label switching across replications. As discussed in Section 4.3, a global relabeling of latent types has no substantive consequences for the second-step objects. The Newton step is sufficiently small ($O_p(U^{-1/2})$) that label switching does not arise in practice.

3. **Step 3: Reconstruct the control variables.** Using $(\widehat{\widetilde{\boldsymbol{f}}}^{*(b)}, \widehat{\gamma}^{*(b)})$, compute the bootstrap analogs of the first-step choice probabilities and of the control-function regressors:

$$\widehat{r}_{j\ell u}^{*(b)} = \frac{\widehat{\widetilde{P}}_{j,\ell}^{*(b)}(\boldsymbol{x}_u) - \widehat{\widetilde{P}}_{j,L}^{*(b)}(\boldsymbol{x}_u)}{\widehat{\bar{P}}_j^{*(b)}(\boldsymbol{x}_u)} \widehat{\widetilde{f}}_\ell^{*(b)}. \tag{63}$$

The denominator $\widehat{\bar{P}}_j^{*(b)}(x_u)$ is the bootstrap analog of the estimated ordinary propensity score for product j. This step is well-defined provided $\widehat{\bar{P}}_j(x_u)$ is bounded away from zero in a neighborhood of $\widehat{\boldsymbol{\psi}}$.[17]

---

[16] More precisely, the unconstrained reparameterization ensures the type probabilities $\widetilde{f}\ell$ remain on the simplex after the Newton update. It does not preclude numerical issues elsewhere in the parameter space—for instance, estimated entry probabilities $\widetilde{P}j, \ell(\cdot)$ remain in $(0, 1)$ by the logistic functional form, but near-zero values could amplify the control-variable construction in Step 3. We verify the absence of such issues in our application.

[17] The logistic functional form in equation 45 ensures $\widetilde{P}j, \ell(x_u) \in (0, 1)$ for all ℓ, and therefore $\bar{P}j(x_u) = \sum \ell \widetilde{P}j, \ell(x_u)\widetilde{f}_\ell \in (0, 1)$. In our application the minimum estimated propensity score across all units is strictly positive, so the denominator is bounded away from zero throughout.

4. **Step 4: Re-estimate the second step using the same weights.** Define the weighted bootstrap moment vector

$$\widehat{g}^{*(b)}(\boldsymbol{\theta},\widetilde{\boldsymbol{\mu}}) \;=\; \frac{1}{U}\sum_{u=1}^{U} w_u^{*(b)}\,\widehat{\boldsymbol{m}}_u(\boldsymbol{\theta},\widetilde{\boldsymbol{\mu}};\widehat{\boldsymbol{\psi}}^{*(b)}). \tag{64}$$

Obtain the bootstrap second-step estimate $(\widehat{\boldsymbol{\theta}}^{*(b)},\widehat{\widetilde{\boldsymbol{\mu}}}^{*(b)})$ by minimizing the corresponding weighted GMM criterion, holding fixed the GMM weighting matrix estimated in the original sample.[18]

Repeating these steps for $b = 1,\ldots,B$ yields bootstrap draws $\{\widehat{\boldsymbol{\theta}}^{*(b)},\widehat{\widetilde{\boldsymbol{\mu}}}^{*(b)}\}_{b=1}^{B}$, which we use to estimate the variance-covariance matrix of the second-step estimator.

This bootstrap avoids repeated estimation of the first-step EM algorithm while preserving the first-order effect of first-step estimation error on the second step. Under the standard regularity conditions that justify the asymptotic linear representation of the two-step estimator, together with local smoothness and nonsingularity of the first-step likelihood so that the one-step update is first-order equivalent to re-estimating the first-step likelihood on the resampled data, and smoothness of the mapping from first-step parameters into the control variables — which requires in particular that the ordinary propensity scores $\bar{P}_j(x_t)$ entering the denominator of the control variables in (39) be bounded away from zero uniformly over the support of x_t — the procedure provides a first-order valid approximation to the infeasible common-weight bootstrap that re-estimates both stages. This validity is conditional on the selected sieve specification; see Chen et al. (2003) for general bootstrap validity results for semiparametric two-step estimators, Armstrong et al. (2014) for conditions under which the linearized one-step update is first-order equivalent to the infeasible bootstrap, and

[18] In our application, equation 50 is linear in the parameters $(\alpha,\boldsymbol{\beta},\sigma,\widetilde{\boldsymbol{\mu}})$ conditional on $\widehat{\boldsymbol{\psi}}^{(b)}$, so the second step reduces to a standard 2SLS estimator. We refer to the step generically as GMM to maintain consistency with Section 4.5, but in practice it is solved as a linear IV problem. Holding the weighting matrix fixed across bootstrap replications is a standard first-order valid simplification: differences between the bootstrap-reoptimized and the full-sample weighting matrices are $O_p(U^{-1/2})$ and affect variance estimates only at second order.

Gonçalves et al. (2023) for the case of quasi-maximum likelihood estimators with preliminary nonparametric components. The bootstrap does not account for uncertainty from selecting L or the dimension of the sieve basis. As is standard in semiparametric two-step estimation, we treat these tuning parameters as fixed, conditioning inference on the chosen model complexity.

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.

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.

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. (2026). Diversification, market entry, and the global internet backbone. *Review of Economic Studies*.

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.

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

Chen, X., Linton, O., and van Keilegom, I. (2003). Estimation of semiparametric models when the criterion function is not smooth. *Econometrica*, 71(5):1591–1608.

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.

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

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.

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.

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.

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

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

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.

Vytlacil, E. (2002). Independence, monotonicity, and latent index models: An equivalence result. *Econometrica*, 70(1):331–341.

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.