Estimating Nonseparable Selection Models: A Functional Contraction Approach*

Fan Wu Yi Xin

May 30, 2026

Abstract

We propose a novel method for estimating nonseparable selection models. We show that, for a given selection function, the potential outcome distributions are nonparametrically identified from the selected outcome distributions and can be recovered using a simple iterative algorithm based on a contraction mapping. This result enables a full-information approach to estimating selection models without imposing parametric or separability assumptions on the outcome equation. We propose a two-step estimation strategy for the potential outcome distributions and the parameters of the selection function and establish the consistency and asymptotic normality of the resulting estimators. Monte Carlo simulations demonstrate that our approach performs well in finite samples. The method is applicable to a wide range of empirical settings, including consumer demand models with only transaction prices, auctions with incomplete bid data, and Roy models with data on accepted wages.

Keywords: Sample Selection, Nonseparable Models, Functional Contraction, Potential Outcome Distribution, Semiparametric Estimation, Demand Estimation, Auction, Roy Models.
JEL Codes: C14, C24, C51, L11, D44, J31.

*Wu: Peking University HSBC Business School, University Town of Shenzhen, China. Email: fanwu@phbs.pku.edu.cn. Xin: Division of the Humanities and Social Sciences, California Institute of Technology, 1200 East California Blvd., MC 228-77, Pasadena, CA 91125. Email: yixin@caltech.edu.
We appreciate valuable discussions with Yonghong An, Jeremy Fox, Michael Keane, Philip Haile, Yingyao Hu, Yao Luo, Luciano Pomatto, Yuya Sasaki, Robert Sherman, Xun Tang, Omer Tamuz, and Ao Wang. We thank seminar and conference participants at Caltech, CUFE, Peking University, Tsinghua University, UC Irvine, University of Michigan, USC, 2025 Asian Summer School in Econometrics and Statistics, Econometric Society World Congress 2025, and DSE Conference 2025. Financial support from the Ronald and Maxine Linde Institute of Economic and Management Sciences is gratefully acknowledged.

1 Introduction

Sample selection issues arise in many empirical settings. In consumer demand studies, researchers often observe only the transaction prices of chosen products (Goldberg, 1996; Cicala, 2015; Crawford et al., 2018; Allen et al., 2019; D'Haultfœuille et al., 2019; Sagl, 2023; Cosconati et al., 2025). In auctions, available data may consist solely of the winning bids (Athey and Haile, 2002; Komarova, 2013; Guerre and Luo, 2019; Allen et al., 2024). In labor economics, wage data are typically observed only for individuals who choose to work (Gronau, 1974; Heckman, 1974), and in the Roy model (Roy, 1951), earnings within an occupation are observed only for those who self-select into that sector.

Observing only a selected sample of outcomes—such as prices, bids, or wages—poses significant challenges for estimating two key elements. First is the selection function, which specifies how agents choose among alternatives, for example, through a consumer demand system, an auction's winning rule, or a labor force participation decision. Second is the distribution of outcomes *prior to selection*, often referred to as "potential outcomes" in the literature. Flexibly estimating potential outcome distributions is crucial in many empirical contexts, such as analyzing price distributions to understand firms' pricing strategies and wage distributions to examine inequality.

Our paper proposes a new approach to estimating nonseparable selection models by exploiting a previously unrecognized one-to-one mapping between the outcome distributions before and after selection. Our key contribution is a constructive identification result showing that, for a given selection function, the potential outcome distributions are nonparametrically identified from the selected outcome distributions and can be recovered using a simple iterative algorithm. Consequently, the only remaining object to be estimated is the selection function, which can be recovered from observed choice patterns. Our method enables a full-information approach to estimating selection models without imposing parametric or separability assumptions on the outcome equation. In addition, it does not rely on an excluded variable in the selection equation.

Formally, we consider a discrete choice problem in which each alternative is associated with a potential outcome distribution. A selection function maps a vector of realized potential outcomes to a probability distribution over the alternatives. For example, in the consumer demand setting, each alternative represents a product, and the potential outcome is the offered price, with the selection function micro-founded by the consumer's utility maximization problem. We allow the outcome equations to be fully nonparametric with nonseparable error terms and to vary flexibly across different alternatives. We assume that potential outcomes across alternatives are independent conditional on observable and unobservable characteristics, which allows for correlation across outcomes when conditioning only on observables. In addition, we allow the unobservable component in the outcome equation to enter directly into agents' preferences over alternatives, thereby capturing selection on unobservables.

We analyze how the selection model maps the potential outcome distributions to the distributions of selected outcomes and seek to *invert* the mapping. The key insight of our approach is that, given the selection model and potential outcome distributions across all alternatives, we can derive the likelihood of an outcome being selected at each price. Conversely, if this selection likelihood were known, we could recover the potential outcome distributions from the selected outcome distributions using Bayes' rule. This two-way relationship characterizes a fixed-point problem.

Building on this intuition, we construct an operator whose fixed point is the potential outcome distributions and establish sufficient conditions for it to be a functional contraction (Theorems 1 and 2). Our results imply that, given the selection function and the distributions of selected outcomes, we can nonparametrically identify the potential outcome distributions. Moreover, this identification result is constructive: starting with any initial guess for the potential outcome distributions, we iteratively apply the operator. This process converges to the potential outcome distributions associated with the selection function.

We then embed this identification result into a two-step estimation strategy for the unobserved potential outcome distributions and the parameters of the selection function. In the first step, we estimate the selected outcome distribution conditional on both observable and unobservable covariates. In the second step, we propose a nested fixed-point algorithm to estimate the parameters of the selection function: in the inner loop, for any candidate selection function, we recover the potential outcome distributions by iterating the operator, while in the outer loop we search for the parameter values that maximize the likelihood of the observed choice patterns. The potential outcome distributions are then obtained by reapplying the fixed-point algorithm at the estimated parameters of the selection function.

We establish the consistency and asymptotic normality of the proposed estimators in Theorems 3 and 4. To examine their finite sample properties, we conduct Monte Carlo simulations across various designs of the outcome equation. Our results show that the biases in our estimators are generally small, and the standard deviation decreases as the sample size increases across all simulation designs. Our nonparametric estimation of the potential outcome distributions outperforms the classic Heckman parametric two-step approach and the quantile selection model of Arellano and Bonhomme (2017) with linear quantile functions and a Gaussian copula, particularly when the outcome equation contains nonseparable error terms. In addition, we show that our approach does not require an excluded variable in the selection equation and remains robust even when the selection function is misspecified by the econometrician.

In a companion paper Cosconati et al. (2025), we apply our method to estimate consumer demand for auto insurance products when only transaction prices are observed. We nonparametrically estimate the offered price distribution for each insurance company and allow these distributions to vary fully flexibly across firms. The substantial heterogeneity in the recovered price distributions reflects differences in firms' information technologies and cost structures, which are key primitives we estimate through a supply-side competition model. We omit the details of this application here and refer readers to Cosconati et al. (2025) for the full empirical setup and results.

Related Literature Our paper contributes to the extensive theoretical literature on sample selection models. An early solution to sample selection bias is full information maximum likelihood (FIML) estimation based on parametric assumptions, as in Heckman (1974) and Lee (1982, 1983). More commonly employed methods for sample selection models are the two-step control function approach pioneered by Heckman (1976, 1979). A substantial body of theoretical work has been developed to relax the distributional assumptions in the two stages of the estimation procedure (e.g., Ahn and Powell, 1993; Andrews and Schafgans, 1998; Chen and Khan, 2003; Das et al., 2003; Newey, 2007, 2009; Fernández-Val et al., 2024; Chernozhukov et al., 2025). For a comprehensive survey of semiparametric two-step estimation methods for selection models, see Vella (1998).

Compared with the existing methods, our approach offers several key advantages. First, we allow the outcome equation to be nonparametric and nonseparable in the error terms, and we exploit the full information in the selected outcome distribution to recover the entire distribution of potential outcomes. Newey (2007) and Fernández-Val et al. (2024) use control function approaches to correct for sample selection in nonseparable models with binary and censored selection rules, respectively, and they focus on identifying certain global and local parameters of the outcome distribution.

Second, our method accommodates fully heterogeneous effects of covariates on outcomes, whereas most existing approaches that estimate conditional mean models restrict covariates to affecting only the location of the outcome distribution.[1] More recently, Arellano and Bonhomme (2017) propose a method to correct for sample selection in quantile regression models by modeling the copula of the error terms in the outcome and selection equations. Although identification under more general settings is discussed, their estimation strategy primarily considers linear quantile models and copulas characterized by a low-dimensional set of parameters.

Third, our approach does not require an instrument that shifts the choice probability without entering the outcome equation, which is central to identification in two-step methods. In practice, finding such an instrument can be challenging (see Vella (1998) for further discussion).[2] Furthermore, our method does not rely on identification-at-infinity arguments. Instead, when the conditioning set of variables includes an unobservable component, our method requires instruments for estimating the distribution of selected outcomes conditional on the unobservable in the first step. We adopt the measurement error framework in Hu (2008) and Hu and Schennach (2008), where the key requirement is to find instruments such that, conditional on the latent variable, the outcome and the instruments are independent.

Estimation of the selection function in our model is closely related to the demand estimation literature following the seminal work of Berry (1994) and Berry et al. (1995). In particular, observed choice patterns play the same role as market shares in recovering consumer preference parameters. Our method addresses the problem of missing full price menu that arises in many demand estimation contexts (e.g., Goldberg, 1996; Cicala, 2015; Crawford et al., 2018; Allen et al., 2019; D'Haultfœuille et al., 2019; Sagl, 2023; Cosconati et al., 2025), an issue that is especially relevant in

[1] An exception is the recent paper by Chernozhukov et al. (2025), which proposes a semiparametric generalization of the Heckman selection model that allows for rich forms of heterogeneity in the effects of covariates on both outcomes and selection.

[2] D'Haultfoeuille and Maurel (2013) and D'Haultfœuille et al. (2018) develop estimation methods for semiparametric sample selection models without an instrument or a large-support regressor, leveraging the independence-at-infinity assumption.

the presence of price discrimination or personalized pricing.[3]

At a broader conceptual level, our reliance on the structural restrictions implied by the selection model resonates with the nonparametric identification literature on auction models with missing bids. Athey and Haile (2002) show that the symmetric independent private values (IPV) models are identified with the transaction price by exploiting a one-to-one mapping between an order statistic and its parent distribution. Komarova (2013) analyzes asymmetric second-price auctions where only the winning bids and the winner's identity are observed. A related result for generalized competing risks models can be found in Meilijson (1981). More recently, Guerre and Luo (2019) examine nonparametric identification of symmetric IPV first-price auctions with only winning bids, accounting for unobserved competition. In these auction models, the selection rule is deterministic conditional on bids (the highest bidder wins), which allows order-statistic arguments to be applied. In contrast, our selection model assigns a probability distribution over alternatives and is therefore closer in spirit to multi-attribute auction environments (see e.g., Krasnokutskaya et al. (2020)). Moreover, our framework can flexibly accommodate asymmetries across alternatives, whereas bidder asymmetries are known to pose significant challenges in auction models (see the discussion in the handbook chapter by Athey and Haile (2007)).

Outline The rest of the paper is organized as follows. Section 2 formally introduces our model and provides an illustrative example. Section 3 presents the main theoretical results. In Section 4, we describe our estimation strategy and establish the asymptotic properties of the estimators. Section 5 reports results from our Monte Carlo simulations, and Section 6 discusses the empirical application in Cosconati et al. (2025). Section 7 concludes. All proofs are collected in the appendix.

2 Model

In Sections 2–3, all analyses are conditional on a vector of characteristics (x, x^*), where x denotes observables and x^* denotes unobservables. The structure of the model and the main theoretical results *do not* depend on whether the conditioning set

[3] A recent paper by D'Haultfœuille et al. (2019) addresses a related challenge in demand estimation under unobserved price discrimination by imposing supply-side restrictions, such as assumptions about firm conduct (e.g., Bertrand competition), and assuming identical costs across consumers.

includes unobserved components, although the presence of unobservables introduces additional challenges for estimation, which we address in Section 4. Because all results in these two sections are stated conditional on (x, x^*), we omit these variables from the notation to simplify exposition.

Throughout the paper, we use a consumer demand example to illustrate the main results and clarify key ideas. In this context, potential outcomes are offered prices, while selected outcomes are selected (or transaction) prices. We use these terms interchangeably when discussing the demand example. Nevertheless, our approach is broadly applicable to a wide class of selection models.

Consider a discrete choice problem. There is a finite set of alternatives $\mathcal{J} = \{1, \cdots, J\}$. Each alternative is associated with a price distribution. Let $G_j \in \Delta([\underline{p}_j, \overline{p}_j])$ represent the price distribution associated with alternative j, where $\Delta(Y)$ denotes the set of all probability measures over a set $Y \subset \mathbb{R}$. We assume that $p_j \sim G_j$ are independently distributed across alternatives (conditional on x and x^*). The collection of G_j is denoted by $G = \prod_{j \in \mathcal{J}} G_j$. We refer to G as the *offered* price distribution.

A *selection function* is denoted by $f = (f_1, f_2, \cdots, f_J)$ where f_j maps prices $\boldsymbol{p} = (p_1, \cdots, p_J)$ to a strictly positive probability of selecting alternative $j \in \mathcal{J}$.[4] We assume that f_j is continuous on the compact price support,

$$f_j : \prod_j [\underline{p}_j, \overline{p}_j] \to (0, 1],$$

with $\sum_{j \in \mathcal{J}} f_j \leq 1$. The inequality allows for the case with an outside option. (Since the support is compact, strict positivity and continuity imply that each f_j is bounded away from zero.) The selection function is a primitive of the model. To provide a microfoundation, for example, f might be derived from a consumer's utility maximization problem as illustrated in Section 2.1.

Let $\boldsymbol{p}_{-j} = (p_1, \cdots, p_{j-1}, p_{j+1}, \cdots, p_J)$ denote the vector of prices excluding j's

[4]The assumption that the probability of selecting each alternative is strictly positive is analogous to the overlap assumption in the treatment effect literature, which requires each individual to have a positive probability of receiving each treatment level. This assumption is crucial for recovering the offered price distribution. To illustrate, consider a scenario where $f_j = 0$ whenever p_j falls within a certain subset of $[\underline{p}_j, \overline{p}_j]$. In this case, any p_j within that subset would not be observed in the data, making it impossible to identify G_j within that subset without introducing additional assumptions.

price. The probability of selecting j conditional on p_j is given by

$$Pr_j(p_j; G) = \int_{\boldsymbol{p}_{-j}} f_j(p_j, \boldsymbol{p}_{-j}) \prod_{k \neq j} dG_k(p_k), \tag{1}$$

where $Pr_j(\cdot; G)$ is a function defined on $[\underline{p}_j, \overline{p}_j]$. Independent prices across different alternatives (conditional on x and x^*) allow us to express the joint distribution of $\boldsymbol{p}_{-j}$ as the product of their individual marginal distribution functions.

Let $\tilde{G}_j \in \Delta([\underline{p}_j, \overline{p}_j])$ represent the price distribution conditional on selecting alternative j. We derive $\tilde{G}_j$ using Bayes' rule:

$$\tilde{G}_j(p) = \frac{\int_{\underline{p}_j}^{p} Pr_j(y; G) dG_j(y)}{\int_{\underline{p}_j}^{\overline{p}_j} Pr_j(y; G) dG_j(y)}. \tag{2}$$

Note that G_j and $\tilde{G}_j$ share the same support, as selection function f_j is strictly positive. Let $\tilde{G} = \prod_{j \in \mathcal{J}} \tilde{G}_j$ and we call $\tilde{G}$ the *selected* price distribution. Equations (1) and (2) define a mapping from G to $\tilde{G}$. Let $F\colon \prod_j \Delta([\underline{p}_j, \overline{p}_j]) \to \prod_j \Delta([\underline{p}_j, \overline{p}_j])$ denote this mapping, i.e., $\tilde{G} = F(G)$.

In many empirical settings, researchers have access only to the selected price distribution, such as the distribution of transaction prices, accepted wages, or winning bids. However, the key primitives of interest are often the offered price distribution, such as the distributions of posted prices, wage offers, or submitted bids. How to recover the offered price distribution G from the selected price distribution $\tilde{G}$? Note that both G and $\tilde{G}$ are collections of J probability measures. Therefore, the cardinality of unknowns and constraints are exactly the same in Equation (2) (assuming the selection function is known). Since a probability measure is an infinite-dimensional object, the key challenge is solving for a collection of infinite-dimensional objects entangled in a nonlinear system. We will explore this in detail in Section 3.

2.1 An Illustrative Example

We now present a simple example to illustrate the key assumptions of our model and compare them to the standard assumptions in the literature. Consider a consumer choosing between two products, $j = 1, 2$, to maximize her utility. The utilities from products 1 and 2 are:

$$u_1 = \gamma p_1 + x^* \kappa + \varepsilon_1, \tag{3}$$

$$u_2 = \gamma p_2 + \varepsilon_2, \tag{4}$$

where p_j represents the price of product j for this consumer, and ε_j is an idiosyncratic utility shock. We also allow an unobservable characteristic x^* to enter directly into the utility from product 1. This captures settings in which unobserved traits affect preferences across alternatives. For example, in insurance markets, high-risk consumers may prefer products offered by certain firms; similarly, in labor markets, workers with higher productivity may prefer certain types of jobs. Our model can flexibly allow utility to depend on observable consumer attributes as well, but we omit these terms here for simplicity of presentation.

In this model, the price sensitivity parameter γ, coefficient κ, and the distribution of ε_j determine the selection function f, for any fixed x^*. If $\varepsilon_1 - \varepsilon_2 \sim \mathcal{N}(0,1)$, the selection function for product 1 takes the standard binary probit form:

$$f_1(p_1, p_2; x^*) = 1 - \Phi_{\mathcal{N}} \left(\gamma(p_2 - p_1) - x^* \kappa \right),$$

where $\Phi_{\mathcal{N}}$ denotes the CDF of the standard normal distribution. For simplicity, we denote the difference in unobservables across the two utilities as $\tilde{\varepsilon} = x^* \kappa + (\varepsilon_1 - \varepsilon_2)$.

In this illustrative example, we consider a simple linear outcome equation with an additive error term. For each product $j = 1, 2$, the price is generated by the following equation:

$$p_j = x\beta_j + x^* \delta_j + \eta_j \equiv x\beta_j + \eta_j^*, \tag{5}$$

where x denotes observable characteristics, x^* denotes unobservable characteristics, and η_j is an idiosyncratic shock. We define the composite error in the pricing equation as $\eta_j^* \equiv x^* \delta_j + \eta_j$, which, for simplicity, is assumed to be independent of x. We assume that the true underlying price shocks η_1 and η_2 are independent. However, when x^* is unobserved by the econometrician, the composite errors η_1^* and η_2^* may be correlated through x^*.

Suppose the econometrician observes the price of product 1 only when it is chosen by the consumer. We derive the conditional mean of p_1 given that it is observed:

$$\begin{aligned}
E(p_1|x, u_1 > u_2) &= x\beta_1 + E(\eta_1^*|\gamma p_1 + x^*\kappa + \varepsilon_1 - (\gamma p_2 + \varepsilon_2) > 0) \\
&= x\beta_1 + E(\eta_1^*|x \underbrace{\gamma(\beta_1 - \beta_2)}_{\beta^*} + \underbrace{[\gamma(\eta_1^* - \eta_2^*) + \tilde{\varepsilon}]}_{\text{composite error: } \varepsilon^*} > 0) \\
&= x\beta_1 + E(\eta_1^*|x\beta^* + \varepsilon^* > 0). \qquad (6)
\end{aligned}$$

The conditioning term $x\beta^* + \varepsilon^* > 0$ in Equation (6) represents the reduced-form selection model typically seen in the literature. Sample selection issue arises when η_1^* and ε^* are correlated, so that $E(\eta_1^*|x\beta^* + \varepsilon^* > 0) \neq 0$. In the two-step estimation literature, researchers often impose assumptions on the joint distribution of $(\varepsilon^*, \eta_1^*, \eta_2^*)$. For example,

$$\begin{bmatrix} \varepsilon^* \\ \eta_1^* \\ \eta_2^* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 \end{bmatrix} \right).$$

We now take a closer look at the correlation between the error in the selection model (ε^*) and the error in the outcome equation (η_1^*). Specifically,

$$\begin{aligned}
cov(\varepsilon^*, \eta_1^*) &= cov(\gamma(\eta_1^* - \eta_2^*) + \tilde{\varepsilon}, \eta_1^*) \\
&= \gamma var(\eta_1^*) - \gamma cov(\eta_1^*, \eta_2^*) + cov(\eta_1^*, \tilde{\varepsilon}). \qquad (7)
\end{aligned}$$

Equation (7) shows that the error term η_1^* directly enters the composite error ε^*, generating the first term $\gamma var(\eta_1^*) \neq 0$ unless $\gamma = 0$. This correlation is *by construction* in selection models, as agents make decisions after observing the potential outcomes. The second term in Equation (7) reflects the correlation between the composite errors in the two outcome equations. It is important to emphasize that our independence assumption—that p_j are independently distributed across alternatives conditional on (x, x^*)—requires independence of the underlying shocks η_1 and η_2, not of the composite errors η_1^* and η_2^*. Thus, our framework accommodates settings in which prices across alternatives are correlated conditional on observables. For instance, if x^* captures a worker's unobserved productivity, then wage offers from two firms may appear correlated when x^* is not conditioned on.

Another common concern regarding selection bias arises from potential correlation between errors in the outcome equation (e.g., η_1^*) and those in the structural selection model (e.g., $\tilde{\varepsilon}$), as represented by the third term in Equation (7). For example, unobserved productivity factors may create correlation between a worker's willingness to work and their wage. Our model also accommodates this type of correlation by allowing unobservable characteristics x^* to enter both the selection equation and the outcome equation.

This simple two-product example illustrates how our notation for the selection function and the offered price distribution maps to the conventions commonly used in the existing literature, and highlights how our setting accommodates all key sources of correlation that give rise to selection bias. The general framework introduced in Section 2 is substantially more flexible than this illustrative case. In particular, our model allows the outcome equation in (5) to be fully flexible and nonparametrically specified with a nonseparable error term. Moreover, we impose minimal assumptions on the selection function. It can accommodate nonparametric, nonseparable relationships between observable and unobserved errors, offering much greater flexibility than the utility specification in Equations (3) and (4); in fact, it *does not* even need to be derived from a utility maximization problem. Our framework also allows for alternative-specific unobserved heterogeneity (such as product quality or job amenities), which is a desirable feature in many empirical contexts.

3 Main Results

We now present our main theoretical results on how to recover the offered price distribution from the selected price distribution. As the selected price distribution is derived from the offered price distribution through Bayes' rule in Equation (2), we can first invert Equation (2):

$$G_j(p_j) = \frac{\int_{\underline{p}_j}^{p_j} d\tilde{G}_j(p)/Pr_j(p;G)}{\int_{\underline{p}_j}^{\overline{p}_j} d\tilde{G}_j(p)/Pr_j(p;G)}. \tag{8}$$

Note that if the selection probability $Pr_j(\cdot;G)$—that is, the probability of selecting product j conditional on its offered price—were known, then recovering the offered price distribution from Equation (8) would be straightforward.

We illustrate this inversion process using the simulated example in Figure 1. The red solid line plots the selected price density for an alternative. Dividing this density by the probability that the alternative is chosen at each price, $Pr_j(p; G)$, yields the unnormalized offered price density shown by the blue dashed line. The gap between these two densities captures the selection mechanism: when a lower price is offered, agents are more likely to accept it, whereas higher prices make them more likely to choose other alternatives. The offered price distribution shown by the blue solid line is then obtained after normalization, which corresponds to the denominator in Equation (8).

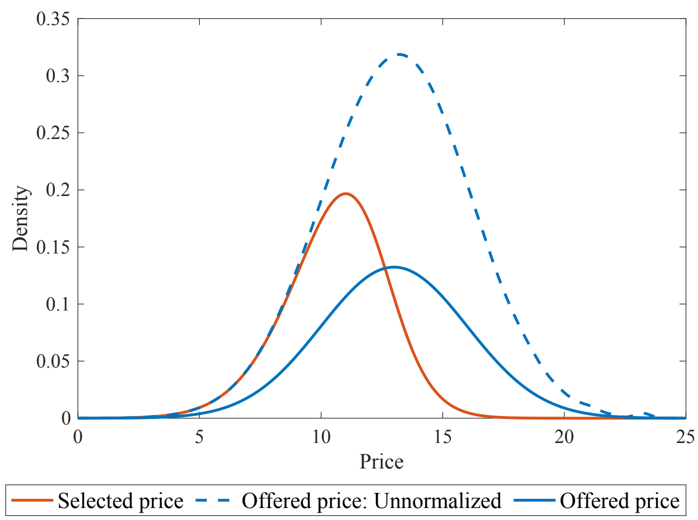

Figure 1: Densities of offered and selected prices. We draw offered prices from $\mathcal{N}(13, 9)$, and the probability that the agent given price p chooses this alternative is given by $\exp(10 - p)/(0.1 + \exp(10 - p))$.

However, $Pr_j(\cdot; G)$ is *not* known, because it depends on the offered price distribution G, which we seek to recover. A tentative solution is to start with a conjecture Ψ for the offered price distribution and use it to compute the implied selection probability $Pr_j(\cdot; \Psi)$. Equation (8) then delivers an *updated* conjecture of the offered price distribution. This procedure, which maps a conjectured offered price distribution into its updated version, defines an operator $T\colon \prod_j \Delta([\underline{p}_j, \overline{p}_j]) \to \prod_j \Delta([\underline{p}_j, \overline{p}_j])$ as follows.

$$(T\Psi)_j(p_j) = \frac{\int_{\underline{p}_j}^{p_j} d\tilde{G}_j(p)/Pr_j(p; \Psi)}{\int_{\underline{p}_j}^{\overline{p}_j} d\tilde{G}_j(p)/Pr_j(p; \Psi)}, \tag{9}$$

where $\Psi = (\Psi_1, \Psi_2, \cdots, \Psi_J) \in \prod_j \Delta([\underline{p}_j, \overline{p}_j])$. Importantly, if the conjecture Ψ is correct, i.e., $\Psi = G$, then the selection probability $Pr_j(\cdot;\Psi)$ is correctly specified, ensuring that the updated conjecture $T\Psi$ also equals G. Thus, the offered price distribution G is a fixed point of the operator T.

The operator T is a contraction if there exists some real number $0 \leq \rho < 1$ such that for all $\Psi, \Phi \in \prod_j \Delta([\underline{p}_j, \overline{p}_j])$,

$$D(T\Psi, T\Phi) \leq \rho D(\Psi, \Phi),$$

given some metric D.[5] In the remainder of this section, we first construct the metric D and then characterize the modulus ρ. We discuss several special cases of our model at the end.

3.1 Hilbert's Projective Metric

We first consider probability measures in the one dimensional space. Let Ψ_j and Φ_j denote two probability measures in $\Delta([\underline{p}_j, \overline{p}_j])$. Recall that two probability measures Ψ_j and Φ_j are equivalent, denoted $\Psi_j \sim \Phi_j$, if they are absolutely continuous with respect to each other. Since $f_j > 0$, $G_j \sim \tilde{G}_j$. When $\Psi_j \sim \Phi_j$, the Radon-Nikodym derivative,

$$\frac{d\Psi_j}{d\Phi_j} : [\underline{p}_j, \overline{p}_j] \to (0, \infty),$$

exists, as guaranteed by the Radon-Nikodym Theorem. If both Ψ_j and Φ_j have continuous densities, the Radon-Nikodym derivative simplifies to the ratio of densities. Note that

$$\Psi_j = \Phi_j \quad \Leftrightarrow \quad \frac{d\Psi_j}{d\Phi_j}(p) = 1 \qquad \Phi_j\text{-a.e.}$$

For two probability measures $\Psi_j, \Phi_j \in \Delta([\underline{p}_j, \overline{p}_j])$, we measure their distance using Hilbert's projective metric.[6] If $\Psi_j \sim \Phi_j$,

$$d_H(\Psi_j, \Phi_j) = \ln \frac{\operatorname{ess\,sup}_{p \in [\underline{p}_j, \overline{p}_j]} \frac{d\Phi_j}{d\Psi_j}(p)}{\operatorname{ess\,inf}_{p \in [\underline{p}_j, \overline{p}_j]} \frac{d\Phi_j}{d\Psi_j}(p)}.$$

[5] We adopt the convention that $+\infty$ and $+\infty$ are not comparable, but $c < +\infty$ for any $c \in \mathbb{R}_+$.

[6] Although Hilbert's metric is projective on the cone of positive finite measures, its restriction to probability measures is a genuine metric on each equivalence class.

If Ψ_j and Φ_j are not equivalent, set

$$d_H(\Psi_j, \Phi_j) = +\infty.$$

Given our operator T in Equation (9), for all $\Psi_j, \Phi_j \in \Delta([\underline{p}_j, \overline{p}_j])$,

$$(T\Psi)_j \sim \tilde{G}_j \sim (T\Phi)_j.$$

Thus,

$$d_H((T\Psi)_j, (T\Phi)_j) = \ln \operatorname*{ess\,sup}_{p_j} \frac{d(T\Psi)_j}{d(T\Phi)_j}(p_j) + \ln \operatorname*{ess\,sup}_{p_j} \frac{d(T\Phi)_j}{d(T\Psi)_j}(p_j).$$

The selected price distribution $\tilde{G}_j$ appears in both $(T\Psi)_j$ and $(T\Phi)_j$. As a result, $\tilde{G}_j$ cancels out in the distance above. Moreover, the denominator in our operator is a normalizing factor, which is also canceled out after we take the sum of log ratios. Consequently, the distance between $(T\Psi)_j$ and $(T\Phi)_j$ relies only on the ratio between selection probabilities:

$$d_H((T\Psi)_j, (T\Phi)_j) \leq \sup_{p_j} \ln \frac{Pr_j(p_j; \Psi)}{Pr_j(p_j; \Phi)} + \sup_{p_j} \ln \frac{Pr_j(p_j; \Phi)}{Pr_j(p_j; \Psi)}.$$

The equality holds when $\tilde{G}_j$ admits full support on $[\underline{p}_j, \overline{p}_j]$. Since $f_j > 0$ is continuous with compact support, Pr_j is bounded away from 0. Thus, $d((T\Psi)_j, G_j)$, $d((T\Psi)_j, \tilde{G}_j)$ and $d((T\Psi)_j, (T\Phi)_j)$ are all finite.

Next, we define a metric in the space $\prod_j \Delta([\underline{p}_j, \overline{p}_j])$ by taking the maximum distance among all alternatives:

$$D(\Psi, \Phi) = \max_{j \in \mathcal{J}} d_H(\Psi_j, \Phi_j)$$

for any $\Psi, \Phi \in \prod_j \Delta([\underline{p}_j, \overline{p}_j])$. From now on, we work with the metric space $(\prod_j \Delta([\underline{p}_j, \overline{p}_j]), D)$.

3.2 Functional Contraction

Given a collection of conjectured offered price distributions

$$\Psi = (\Psi_1, \ldots, \Psi_J),$$

write

$$\Psi_{-j} = \bigotimes_{k \neq j} \Psi_k.$$

Then the selection probability conditional on p_j can be written as

$$Pr_j(p_j; \Psi) = \int_{\boldsymbol{p}_{-j}} f_j(p_j, p_{-j})\, d\Psi_{-j}(p_{-j}).$$

This expression reveals the key operator-theoretic structure of the model. The map $\Psi_{-j} \mapsto Pr_j(\cdot; \Psi)$ can be viewed as a positive linear integral operator, with f_j as its kernel. This allows us to use positive operator theory under Hilbert's projective metric.

The contraction property of T follows from the classical Birkhoff–Hopf contraction theorem for positive linear operators under Hilbert's projective metric. See Theorem A.4.1 in Lemmens and Nussbaum (2012). To state the result, for a positive integral operator with kernel f_j, define its projective diameter by

$$\Delta_j = \sup_{\substack{p_j, p_j' \in [\underline{p}_j, \overline{p}_j] \\ \boldsymbol{p}_{-j}, \boldsymbol{p}'_{-j} \in \prod_{k \neq j} [\underline{p}_k, \overline{p}_k]}} \left| \ln \frac{f_j(p_j, p_{-j}) f_j(p_j', p_{-j}')}{f_j(p_j, p_{-j}') f_j(p_j', p_{-j})} \right|.$$

Since f_j is continuous and strictly positive on compact support, Δ_j is finite. The quantity Δ_j measures how far the kernel f_j is from being multiplicatively separable in p_j and p_{-j}. If $f_j(p_j, p_{-j}) = a_j(p_j) b_j(p_{-j})$, then $\Delta_j = 0$. Let

$$\rho = (J - 1) \max_{j \in \mathcal{J}} \tanh\left(\frac{\Delta_j}{4}\right).$$

Theorem 1. *If $\rho < 1$, the operator T is a contraction with modulus at most ρ.*

Proof. See Appendix A.1. □

Theorem 1 establishes a key identification result for selection models. Whenever $\rho < 1$, the operator T admits a unique fixed point, the offered price distribution G. Theorem 1 implies that we can nonparametrically identify the potential outcome distributions G from any selected outcome distribution $\tilde{G}$, given the selection function f. Notably, the theorem imposes no assumptions on the functional form of the potential outcome distributions, allowing the outcome equation to be fully nonparametric and nonseparable in the error terms, and it applies to any selection function f satisfying the positivity and finite-projective-diameter condition.[7]

Moreover, the result in Theorem 1 provides a *constructive* method for solving for G. Take any $\Psi \in \prod_j \Delta([\underline{p}_j, \overline{p}_j])$, by Theorem 1,

$$D(T^n\Psi, G) = D(T^n\Psi, TG) \leq \rho D(T^{n-1}\Psi, G) \leq \rho^{n-1} D(T\Psi, G),$$

where $D(T\Psi, G)$ is finite. This implies

$$\lim_{n\to\infty} D(T^n\Psi, G) = 0,$$

$$\lim_{n\to\infty} T^n\Psi = G.$$

Thus, we can simply take an initial guess for the potential outcome distributions and iteratively apply the operator. As the number of iterations approaches infinity, this process converges to the potential outcome distributions associated with the selection function.

The condition in Theorem 1 is a joint restriction on the selection function and the price support. The primitive object in Theorem 1 is the projective diameter Δ_j of the selection function kernel. The quantity Δ_j measures the strength of the interaction between the price of alternative j and the prices of competing alternatives in the log selection rule. Equivalently, it measures how far the kernel f_j is from being multiplicatively separable in p_j and p_{-j}.

The condition is sufficient rather than necessary: the operator may remain a contraction even when the upper bound

$$(J-1) \max_{j\in\mathcal{J}} \tanh(\Delta_j/4)$$

exceeds one.[8] Nevertheless, the bound provides useful intuition. First, holding the selection function fixed, enlarging the price support weakly increases the projective diameter, because the supremum in Δ_j is taken over a larger set. Hence, larger price domains make it harder to control the ratio $D(T\Psi, T\Phi)/D(\Psi, \Phi)$ uniformly over all

[7] The continuity of f is not required.

[8] Although the Birkhoff–Hopf coefficient is sharp for each positive linear operator $\Psi_{-j} \mapsto Pr_j(\cdot;\Psi)$ under Hilbert's projective metric, the bound in Theorem 1 further uses product-measure and max-norm inequalities to control the full vector-valued operator T. These aggregation steps may be conservative.

conjectures Ψ and Φ.

Second, Δ_j is small when the effect of competing prices on the selection probability of alternative j is weak in a projective sense. In the extreme case where

$$f_j(p_j, p_{-j}) = a_j(p_j) b_j(p_{-j}),$$

we have $\Delta_j = 0$. In this case,

$$Pr_j(p_j; \Psi) = a_j(p_j) \int b_j(p_{-j})\, d\Psi_{-j}(p_{-j}),$$

so changing the conjectured distribution Ψ_{-j} only multiplies $Pr_j(\cdot; \Psi)$ by a positive constant. This constant cancels in the update step of T. Therefore, the distribution of competing prices has no projective effect on the recovery of G_j.

More generally, when Δ_j is small, misspecifying the conjectured distribution of competing prices changes the implied selection probability $Pr_j(\cdot; \Psi)$ only mildly in Hilbert's metric. The update $T\Psi$ is therefore close to the update that would be obtained under the true offered price distribution. This is the source of the contraction.

The factor $J - 1$ arises because the selection probability for alternative j depends on the product distribution of the $J - 1$ competing alternatives. When $J = 2$, the condition reduces to

$$\max_j \tanh(\Delta_j / 4) < 1.$$

Thus, as long as the projective diameter is finite, the binary-choice case satisfies the contraction condition automatically. With more than two alternatives, contraction requires the aggregate interaction between each alternative and its competitors to be sufficiently small.

Corollary 1. *In the binary-choice case ($J = 2$), the operator T is a contraction.*

The projective diameter Δ_j has a direct relationship with the semi-elasticity of the selection function. If $\ln f_j$ is continuously differentiable in p_j, then we can define

$$M_j = \sup_{p_j, p_{-j}, p'_{-j}} \left| \frac{\partial \ln f_j(p_j, p_{-j})}{\partial p_j} - \frac{\partial \ln f_j(p_j, p'_{-j})}{\partial p_j} \right|.$$

Theorem 1 implies the following corollary.

Corollary 2. *Suppose* $\ln f_j$ *is continuously differentiable in* p_j*. A sufficient condition for contraction is*

$$\frac{J-1}{4}\max_{j\in\mathcal{J}}(\bar{p}_j - \underline{p}_j)M_j < 1.$$

Thus, the semi-elasticity condition is a convenient primitive sufficient condition, while the projective-diameter condition in Theorem 1 provides a sharper contraction bound. The term M_j is small when the own-price responsiveness of alternative j remains stable as competitors' prices vary. This occurs, for example, when alternative j is highly differentiated or weakly substitutable with other alternatives. In that case, even if the conjectured distributions of competing prices are inaccurate, the resulting selection probability $Pr_j(\cdot;\Psi)$ remains close to the truth, and the Bayes update in T remains close to the correct update.

3.3 Special Cases

Thus far, we have not imposed any structure on the selection function. For a general selection function, we have to take the supremum over prices to compute the projective diameter. Now we impose an assumption on the selection function to determine where the supremum is attained.

Assumption 1 (Log Supermodularity)**.** Suppose $\ln f_j$ is continuously differentiable in p_j. For all $j \in \mathcal{J}$ and $p_j \in [\underline{p}_j, \bar{p}_j]$, $\frac{\partial \ln f_j(p_j, \boldsymbol{p}_{-j})}{\partial p_j}$ is weakly increasing in each p_k with $k \neq j$.

Under log supermodularity, the projective diameter in Theorem 1 is attained at the boundary. The result is as follows. Let

$$\rho^* = (J-1)\max_{j\in\mathcal{J}}\tanh\left\{\frac{1}{4}\left[\ln f_j(\overline{\boldsymbol{p}}) - \ln f_j(\underline{p}_j, \overline{\boldsymbol{p}}_{-j}) - \ln f_j(\bar{p}_j, \underline{\boldsymbol{p}}_{-j}) + \ln f_j(\underline{\boldsymbol{p}})\right]\right\}.$$

Theorem 2. *Suppose that Assumption 1 holds. If* $\rho^* < 1$*, the operator* T *is a contraction with modulus at most* ρ^**.*

Proof. See Appendix A.2. □

Under Assumption 1, the modulus ρ^* takes a much simpler form and is straightforward to compute. The log-supermodularity assumption holds in models widely adopted by empirical researchers. For example, the multinomial logit model satisfies

Assumption 1. The binary probit model described in Section 2.1 also satisfies the log-supermodularity condition in Assumption 1, so Theorem 2 applies.[9] However, Assumption 1 may not hold for probit models with three or more alternatives; in such cases, the more general results in Theorem 1 can be applied.

To summarize, our contraction results provide a novel method for identifying the potential outcome distribution from the selected outcome distribution, given any selection function f—whether parametric or nonparametric, and regardless of whether it is microfounded in a utility maximization problem. Moreover, the identification is constructive: starting with an initial guess, iterative application of the operator converges to the potential outcome distributions associated with the selection function. This powerful identification result exhausts all the information contained in the selected outcome distributions. Then the estimation of the selection model essentially reduces to recovering the selection function from observed choice patterns. We discuss the estimation strategy in the next section.

4 Estimation

We now turn to the estimation of the model's primitives, which include (1) the unobserved offered price distributions G and (2) the parameters in the selection function f. We propose a two-step estimation procedure. In the first step, we estimate the selected outcome distribution conditional on both observable and unobservable covariates using instruments. Once the selected outcome distribution has been recovered, for any given selection function f, the potential outcome distribution can be recovered iteratively using the contraction mapping results in Section 3. The second step nests this fixed-point problem within an estimation routine that recovers the parameters of the selection function from agents' observed choice patterns. Using the resulting

[9] To see this, we compute the log derivative for the binary probit model:

$$\frac{\partial \ln f_1(p_1, p_2)}{\partial p_1} = \frac{\gamma \phi_{\mathcal{N}}(\Delta)}{1 - \Phi_{\mathcal{N}}(\Delta)},$$

$$\frac{\partial^2 \ln f_1(p_1, p_2)}{\partial p_1 \partial p_2} = \gamma^2 \frac{d}{d\Delta}\left[\frac{\phi_{\mathcal{N}}(\Delta)}{1 - \Phi_{\mathcal{N}}(\Delta)}\right],$$

where $\Delta = \gamma(p_2 - p_1)$ and the term in the square bracket is known as the hazard rate or inverse Mills ratio. As Gaussian satisfies increasing hazard rate (Baricz, 2008), the log-supermodularity condition in Assumption 1 holds.

parameter estimates, we then re-run the fixed-point algorithm to recover the offered price distribution G.

In the data, for each individual i, we observe their choice $y_i \in \mathcal{J}$ and the price of the selected product p_i. Let x_{ij} denote a vector of observable characteristics, and define $x_i = (x'_{i1}, \cdots, x'_{iJ})' \in X$. We let $x_i^* \in X^*$ denote an unobservable characteristic which may affect both the selection decision and the distribution of potential outcomes.

We assume that the selection function f is derived from a standard multinomial choice model with an indirect utility given by

$$u_{ij} = v_j(p_{ij}, x_{ij}, x_i^*, \varepsilon_{ij}; \theta),$$

where v_j is a known function indexed by a finite-dimensional parameter vector θ. Here, p_{ij} is the offered price of alternative j for individual i, and the vector of unobserved shocks $\varepsilon_i = (\varepsilon_{i1}, \cdots, \varepsilon_{iJ})$ follows a known joint distribution, such as Type 1 extreme value. Each individual chooses the alternative that maximizes utility, and the selection function f is captured by the parameter θ. Throughout the paper, we use θ_0 to denote the true parameter.

For example, a widely used specification takes the following form:

$$u_{ij} = \gamma p_{ij} + x'_{ij}\beta + \xi_j + x_i^* \kappa_j + \varepsilon_{ij}, \quad j = 1, 2, \cdots, J, \tag{10}$$

where ξ_j represents a scalar-valued unobserved characteristic of alternative j, such as product quality or brand loyalty. The term $x_i^* \kappa_j$ allows preferences for product j to vary with the unobservable characteristic x_i^*. In this example, $\theta = (\gamma, \beta, \boldsymbol{\xi}, \boldsymbol{\kappa})$, where $\boldsymbol{\xi} = (\xi_1, \cdots, \xi_J)$ and $\boldsymbol{\kappa} = (\kappa_1, \cdots, \kappa_J)$.

4.1 Two-Step Estimation Strategy

Step 1: Estimating Selected Outcome Distribution The key inputs for our contraction-mapping results are the selected outcome distributions $\tilde{G}$ conditional on (x, x^*). When all relevant covariates are observed, so that no unobserved component x^* is present, $\tilde{G}$ conditional on x can be easily estimated nonparametrically from the data, for example using kernel methods. We therefore do not elaborate on this case. The more challenging setting arises when an unobserved covariate x^* is present. In this case, $\tilde{G}$ conditional on (x, x^*) cannot be directly estimated from the observed data, and additional information about the unobserved covariates is required in order to recover this distribution.

We follow the instrumental variable approach of Hu (2008) to estimate the selected outcome distribution conditional on the unobservable x^* in the first step. We assume that the variables $\omega_i = \{x_i, y_i, p_i, z_{1i}, z_{2i}\}$ are observed in an i.i.d. sample and take values in a finite support.[10] The variables z_1 and z_2 serve as instrumental variables and are required to satisfy the following condition:

$$\begin{aligned} &h_{p,z_1|z_2,x,y}(p, z_1|z_2, x, y) \\ =&\sum_{x^*} h_{p|x^*,x,y}(p|x^*, x, y) h_{z_1|x^*,x,y}(z_1|x^*, x, y) h_{x^*|z_2,x,y}(x^*|z_2, x, y), \end{aligned} \tag{11}$$

where $h(\cdot)$ represents probability mass functions. Equation (11) shows that the joint distribution of (p, z_1) conditional on (z_2, x, y) can be expressed as a mixture over the latent variable x^*. This condition requires, first, that the two instrumental variables are informative about the latent variable x^*, and second, that once we condition on x^*, the price and the instruments are independent.[11] In practice, finding such instruments is often feasible. In insurance pricing, for example, the latent variable x^* may represent a consumer's unobserved risk type, which influences premiums. Realized claims can serve as proxy variables for this latent type. In labor applications, the latent variable might correspond to a worker's unobserved productivity, which affects wages. Measures such as work-performance evaluations or test scores can provide useful proxies in these settings.

Theorem 1 in Hu (2008) shows that, under additional rank and ordering assumptions, the unknown probability mass functions on the right-hand side of Equation

[10] The finite support assumption is not essential for identifying and estimating the selected outcome distributions in the first step. Hu and Schennach (2008) extends the results in Hu (2008) to settings with continuously distributed variables, so similar identification argument and estimation procedure remain valid without discreteness. This assumption is adopted here primarily to simplify the asymptotic normality results, which we discuss further in Section 4.2.

[11] Alternatively, suppose we have three instruments (z_1, z_2, z_3) such that,

$$\begin{aligned} &h_{z_3,z_1|z_2,x,y,p}(z_3, z_1|z_2, x, y, p) \\ =&\sum_{x^*} h_{z_3|x^*,x,y,p}(z_3|x^*, x, y, p) h_{z_1|x^*,x,y,p}(z_1|x^*, x, y, p) h_{x^*|z_2,x,y,p}(x^*|z_2, x, y, p). \end{aligned}$$

Under this condition, the instruments are allowed to depend arbitrarily on the price, while only requiring independence across instruments conditional on the latent variable.

(11), $h = (h_{p|x^*,x,y}, h_{z_1|x^*,x,y}, h_{x^*|z_2,x,y}) \in H$, are nonparametrically identified. We do not restate these additional assumptions here and instead refer readers to Hu (2008) for the technical details.

Given Equation (11), a maximum likelihood estimator of h can be obtained in a straightforward manner. We denote this estimator by $\hat{h} = (\hat{h}_{p|x^*,x,y}, \hat{h}_{z_1|x^*,x,y}, \hat{h}_{x^*|z_2,x,y})$. The term $\hat{h}_{p|x^*,x,y}$ represents the estimate of the selected price distribution conditional on (x, x^*), which corresponds to $\tilde{G}(x, x^*)$. We do not distinguish between these two objects in what follows. Finally, by taking the expectation of $\hat{h}_{x^*|z_2,x,y}$ with respect to the distribution of z_2, we obtain an estimate of the distribution of the latent variable x^* conditional on (x, y), which we denote by $\hat{h}_{x^*|x,y}$.

Step 2: Estimating Selection Function Parameters and Offered Price Distributions Given the first-step estimates $\hat{h}_{p|x^*,x,y}$ and $\hat{h}_{x^*|x,y}$, we propose a semi-parametric maximum likelihood estimator for parameter θ in the selection function:

$$\hat{\theta} = \arg\max_{\theta\in\Theta} \hat{Q}_n(\theta), \tag{12}$$

where

$$\hat{Q}_n(\theta) = \frac{1}{n}\sum_{i=1}^{n}\sum_{x^*} \hat{h}_{x^*|x,y}(x^*|x_i, y_i) \ln Prob_{y_i}(x_i, x^*, \theta, \hat{h}_{p|x^*,x,y}), \tag{13}$$

$$Prob_j(x, x^*; \theta, \tilde{G}) = \int_{\boldsymbol{p}} f_j(\boldsymbol{p}; x, x^*, \theta) d\big(F^{-1}(\tilde{G}(x, x^*); \theta, x, x^*)\big)(\boldsymbol{p}). \tag{14}$$

Equation (14) derives the probability that alternative j is chosen conditional on (x, x^*) for any utility parameters θ and any selected outcome distributions $\tilde{G}$. This probability is obtained by integrating the selection function $f_j(\boldsymbol{p}; x, x^*, \theta)$ with respect to the distribution of offered prices for all alternatives. Recall that F denotes the mapping from the offered price distribution G to the selected outcome distribution $\tilde{G}$ as defined in Equations (1) and (2). The inverse mapping F^{-1} in Equation (14) maps $\tilde{G}$ back to G.

To recover the offered price distribution, we rely on the contraction-mapping results in Theorem 1, which guarantees that we can replicate F^{-1} by iterating the operator T until convergence. Note that the operator T depends on two components: (1) the parameters of the selection function, θ, and (2) the selected outcome distributions $\tilde{G}$, for which a first-step estimate $\hat{h}_{p|x^*,x,y}$ is obtained. We use $\hat{T}_{\theta,\hat{h}}$ to denote the operator constructed given θ and $\hat{h}$. Let $\hat{T}^{\infty}_{\theta,\hat{h}}\Psi$ denote the limit of the iterates of $\hat{T}_{\theta,\hat{h}}$ starting from an initial distribution Ψ.[12]

Our second-step estimation follows a nested fixed-point algorithm. In the inner loop, for any candidate value of the parameter θ in the selection function, we obtain the fixed point of the operator T as $\hat{T}^{\infty}_{\theta,\hat{h}}\Psi$. Given the resulting offered price distribution, we compute agents' choice probabilities using Equation (14) and then construct the sample analogue of the likelihood function in Equation (13). In the outer loop, we then search over θ to maximize this likelihood.

Once $\hat{\theta}$ is obtained, a plug-in estimator of the offered price distribution G can be constructed by

$$\hat{G} = \hat{T}^{\infty}_{\hat{\theta},\hat{h}}\Psi.$$

This step essentially repeats the inner-loop procedure, except that we replace θ with its estimate $\hat{\theta}$.

4.2 Consistency and Asymptotic Normality

We now discuss the asymptotic properties of our proposed estimators $\hat{\theta}$ and $\hat{G}$. When constructing the model-implied choice probabilities in Equation (14), the inverse mapping F^{-1} appears, which maps the selected price distribution $\tilde{G}$ back to the offered price distribution G. We therefore begin by analyzing the properties of this inverse mapping F^{-1}.

Proposition 1. *Suppose $\rho < 1$. The mapping F is a homeomorphism. Moreover, both F and F^{-1} are Lipschitz continuous, with Lipschitz constants $1+\rho$ and $\frac{1}{1-\rho}$, respectively.*

Proof. See Appendix A.3. □

Proposition 1 has three important implications. First, because F is a homeomorphism, its inverse F^{-1} is well-defined, and we have $G = F^{-1}(\tilde{G})$. Second, the

[12]In practice, the algorithm used to solve the fixed point is terminated after a finite number of iterations. We show that the resulting approximation error is asymptotically negligible, provided that the number of iterations grows fast enough compared to the logarithm of the sample size. Further details are provided at the end of Section 4.2.

continuity of F^{-1} implies that if a consistent estimator $\tilde{G}_n$ of the selected outcome distribution is used in place of $\tilde{G}$, then

$$F^{-1}(\tilde{G}_n) \xrightarrow{p} F^{-1}(\tilde{G}) = G \quad \text{as} \quad \tilde{G}_n \xrightarrow{p} \tilde{G}.$$

Finally, since F^{-1} is Lipschitz continuous, $F^{-1}(\tilde{G}_n)$ converges to G at the same rate as $\tilde{G}_n$ converges to $\tilde{G}$.

We now turn to the consistency and asymptotic normality of our estimators. To establish consistency, we rely on the fundamental consistency theorem for extremum estimators (Theorem 2.1 in Newey and McFadden (1994)). We construct the true population objective function as follows:

$$Q_0(\theta) = \mathbb{E}_{x,x^*} \sum_{j=1}^{J} \left(\int_{\boldsymbol{p}} f_j(\boldsymbol{p}; x, x^*, \theta_0) dG(x, x^*)(\boldsymbol{p}) \right) \ln \left(Prob_j(x, x^*, \theta, \tilde{G}) \right),$$

where $\int_{\boldsymbol{p}} f_j(\boldsymbol{p}; x, x^*, \theta_0) dG(x, x^*)(\boldsymbol{p})$ represents the true probability of selecting alternative j conditional on x and x^*.

We maintain the standing assumptions that f_j is strictly positive and continuous. The following additional technical conditions are required for consistency.

Assumption 2. (i) The space Θ of parameter θ is compact; (ii) for each x, x^*, the selection function $f(\boldsymbol{p}; x, x^*, \theta)$ is jointly continuous in θ and $\boldsymbol{p}$; (iii) the condition in Theorem 1 holds for all $\theta \in \Theta$, that is, $\sup_{\theta \in \Theta} \rho(\theta) \leq \bar{\rho} < 1$ for some $\bar{\rho}$.

Assumption 3 (Identification)**.** There does not exist $\theta' \in \Theta$, $\theta' \neq \theta_0$, offered price distributions G, $G' \in \left(\prod_j \Delta([\underline{p}_j, \overline{p}_j]) \right)^{X \times X^*}$ such that for all $j \in \mathcal{J}$ and x, x^*,

$$F(G(x, x^*); \theta_0, x, x^*) = F(G'(x, x^*); \theta', x, x^*),$$

$$\int_{\boldsymbol{p}} f_j(\boldsymbol{p}; x, x^*, \theta_0) dG(x, x^*)(\boldsymbol{p}) = \int_{\boldsymbol{p}} f_j(\boldsymbol{p}; x, x^*, \theta') dG'(x, x^*)(\boldsymbol{p}).$$

Assumption 2 (i) and (ii) are standard regularity conditions. Assumption 2 (iii) ensures that for all $\theta \in \Theta$, the operator T is a contraction. Assumption 3 imposes the identification condition, which requires that there does not exist another parameter that can yield the same selected price distribution and choice probabilities.

The identification condition merits additional discussion. The unknown objects in our model are the parameter vector θ in the selection function f and the offered price distribution G. A key insight from our contraction mapping result (Theorem 1) is that, for any given selection function f, the operator T admits a *unique* fixed point, and this fixed point corresponds to the offered price distribution associated with f. In other words, given f, the offered price distribution G is fully nonparametrically identified from the accepted price distribution.This is *not* an additional identifying assumption; it follows from the contraction mapping result. As a result, identification of the full model reduces to identification of the parameter vector θ in the selection function.

Assumption 3 essentially requires that variation in the observed choice probabilities conditional on (x, x^*)—which we have already identified in the first-step estimation—is sufficient to uniquely pin down the selection-function parameters. For example, under the commonly used demand specification in Equation (10), the unknown parameters include the price sensitivity parameter γ, the coefficients (β, κ_j) on the covariates (x, x^*), and the unobserved product characteristics ξ_j (with one of them normalized to zero without loss). The number of unknowns is $dim(x_{ij}) + 2J$, whereas the number of moments (i.e., conditional choice probabilities) available for identification is $|X||X^*|(J-1)$. As the dimensions of x and x^* increase, the variation in choice probabilities expands, generating an overidentified system for the utility parameters.

Moreover, if additional instrumental variables are available, such as exogenous cost shocks that shift the offered price distribution, these provide extra moment conditions for identifying the price sensitivity parameter as in the classical demand estimation literature. In the paper, we provide a high-level version of the identification condition for simplicity, but our framework can readily incorporate any additional instrumental variables when available. These extra moments can be included in the outer loop of the Step 2 estimation procedure described in Section 4.1. We summarize the consistency result in the following theorem.

Theorem 3 (Consistency)**.** *Under Assumptions 2 and 3,* $\hat{\theta} \xrightarrow{p} \theta_0$, $\hat{T}^{\infty}_{\hat{\theta},\hat{h}}\Psi \xrightarrow{p} G$.

Proof. See Appendix A.3. □

Next, we show that the estimator defined in Equation (12) is asymptotically normal. Let

$$\mathfrak{g}(\omega; \theta, h) = \nabla_\theta \left(\sum_{x^*} h_{x^*|x,y}(x^*|x, y) \ln Prob_y(x, x^*, \theta, h_{p|x^*,x,y}) \right),$$

where ∇_θ denotes the gradient operator with respect to θ. The estimator $\hat{\theta}$ solves the first-order condition

$$\frac{1}{n}\sum_{i=1}^{n}\mathfrak{g}(\omega_i;\theta,\hat{h}) = 0.$$

Moreover, we define

$$\mathfrak{m}(\omega_i,h) = \nabla_h \ln\left(\sum_{x^*} h_{p|x^*,x,y}(p_i|x^*,x_i,y_i)h_{z_1|x^*,x,y}(z_{1i}|x^*,x_i,y_i)h_{x^*|z_2,x,y}(x^*|z_{2i},x_i,y_i)\right).$$

We stack $\mathfrak{g}$ and $\mathfrak{m}$ to form

$$\tilde{\mathfrak{g}}(\omega,\theta,h) = [\mathfrak{g}(\omega,\theta,h)', \mathfrak{m}(\omega,h)']',$$

then the estimators in the first two steps can be viewed as a GMM estimator. We impose the following standard regularity conditions.

Assumption 4. (i) (θ_0,h_0) is in the interior of $\Theta\times H$. (ii) f is twice continuously differentiable in θ. (iii) $\mathbb{E}\nabla_{\theta,h}\tilde{\mathfrak{g}}(\omega;\theta_0,h_0)$ is nonsingular. (iv) After representing G and $\tilde{G}$ by free coordinates, $\nabla_G F(G,\theta_0)$ is nonsingular.

Theorem 4 (Asymptotic Normality)**.** *Suppose that Assumption 2, 3, and 4 hold. Then* $\hat{\theta}$, $\hat{h}$, $\hat{T}^{\infty}_{\hat{\theta},\hat{h}}\Psi$ *are* $\sqrt{n}$*-asymptotically normal and* $\sqrt{n}(\hat{\theta}-\theta_0)\xrightarrow{d}\mathcal{N}(0,V)$.[13]

Proof. See Appendix A.4. □

So far, the asymptotic results have been stated under the assumption that the operator is iterated infinitely many times. In practice, however, the iteration used to obtain the offered price distribution is stopped after a finite number of steps. The resulting approximation error is asymptotically negligible as long as the number of iterations grows fast enough relative to the logarithm of the sample size. Formally, let $m(n)$ denote the number of iterations given the sample size n. Consistency of our estimator (Theorem 3) can be achieved as long as $\lim_{n\to+\infty} m(n)\to\infty$. Asymptotic normality (Theorem 4) continues to hold if in addition, $\liminf_{n\to+\infty}\frac{m(n)}{\ln n} > \frac{1}{2}(\ln(1/\bar{\rho}))^{-1}$.

Finally, we discuss the finite support assumption imposed on the outcome p_i. This assumption is not essential for the consistency result in Theorem 3. As long as the estimator of h is consistent, our proposed estimator remains consistent even

[13]See the analytical form of V in the proof of Theorem 4.

when p_i is continuous. Assuming that p_i has finite support mainly keeps the proof of asymptotic normality in Theorem 4 tractable. If p_i is instead continuous, establishing asymptotic normality for a semiparametric two-step estimator typically requires a first-order expansion around the nonparametric estimator (see Theorem 8.1 in Newey and McFadden (1994)). In our setting, this would require expanding the function $\mathfrak{g}$ around $\hat{h}_{p|x^*,x,y}$. A standard argument would apply if $\hat{h}_{p|x^*,x,y}$ entered Equation (14) directly. However, in our case it enters only through F^{-1}, for which no analytic expression is available. As a result, working with the infinite-dimensional distribution $\tilde{G}$ is extremely challenging.

In practice, when the selected outcome distribution is estimated nonparametrically, even if p_i is conceptually continuous, the estimator necessarily evaluates its CDF on a finite grid of points. For this reason, this assumption does not impose a substantive restriction in applied work.

5 Monte Carlo Simulations

To examine how our estimators for θ and the offered price distribution G perform in finite samples, we conduct a Monte Carlo simulation experiment with $J = 2$. The utilities of individual i from the two alternatives are specified as follows:

$$\begin{aligned} u_{i1} &= -\gamma \ln(p_{i1}) + \xi_1 + \beta x_{i1} + \kappa x_i^* + \varepsilon_i, \\ u_{i2} &= -\gamma \ln(p_{i2}) + \xi_2, \end{aligned}$$

where p_{ij} and ξ_j are, respectively, the offered price and unobserved heterogeneity for alternative j; $x_{i1} \in \{0,1\}$ is a binary observable with $Pr(x_{i1} = 1) = 0.5$ that shifts individual i's choice probabilities; $x_i^* \in \{-1,1\}$ is a binary unobservable with $Pr(x_i^* = 1) = 0.5$; and $\varepsilon_i \sim N(0,1)$ is the error term. Throughout the main simulation exercises, we set the utility parameters as follows: $\gamma = 1$, $\xi_1 = 0$, $\xi_2 = 0.5$, $\beta = 0.5$ and $\kappa = 0.1$.[14] Let $y_i \in \{1,2\}$ denote the choice of individual i.

We consider four data generating processes for the offered prices. Let x_{i2} denote the observable characteristic of individual i that enters the pricing equation. We

[14]Note that x_{i1} is included in the utility specification to facilitate the implementation of the two-step method for sample selection. Our method does not require this type of excluded variable. We therefore also consider a simulation exercise in which x_{i1} is omitted, that is, $\beta = 0$. The results are reported in Tables 4 and 5 in Appendix B.

assume that x_{i2} takes values in $\{0, 0.25, 0.5, 0.75, 1\}$ with equal probability.

DGP 1: $\ln(p_{ij}) = \delta_{0j} + \delta_{1j}x_{i2} + \delta_{2j}x_i^* + \eta_{ij}$, where $\eta_{ij} \sim N(0, \sigma_j^2)$. For alternative 1, we set $\delta_{01} = 0.2, \delta_{11} = 0.5, \delta_{21} = 0.1, \sigma_1 = 0.1$. For alternative 2, we set $\delta_{02} = 0.1, \delta_{12} = 1, \delta_{22} = 0.1, \sigma_2 = 0.2$.

DGP 2: $\ln(p_{ij}) = \delta_{0j} + \delta_{1j}x_{i2}^2 + \delta_{2j}x_i^* + \eta_{ij}$, where $\eta_{ij} \sim N(0, \sigma_j^2)$. For alternative 1, we set $\delta_{01} = 0.2, \delta_{11} = 0.5, \delta_{21} = 0.1, \sigma_1 = 0.1$. For alternative 2, we set $\delta_{02} = 0.1, \delta_{12} = 1, \delta_{22} = 0.1, \sigma_2 = 0.2$.

DGP 3: $\ln(p_{ij}) = \exp\left((\delta_{0j} + \delta_{1j}x_{i2})(\delta_{2j}x_i^* + \eta_{ij})\right)$, where $\eta_{ij} \sim N(1, \sigma_j^2)$. For alternative 1, we set $\delta_{01} = 0.2, \delta_{11} = 0.3, \delta_{21} = 0.1, \sigma_1 = 0.1$. For alternative 2, we set $\delta_{02} = 0.1, \delta_{12} = 0.5, \delta_{22} = 0.1, \sigma_2 = 0.2$.

DGP 4: $\ln(p_{ij}) = (\delta_{0j} + \delta_{1j}x_{i2}^2)(\delta_{2j}x_i^* + \eta_{ij})^{-1}$, where $\eta_{ij} \sim N(-2, \sigma_j^2)$. For alternative 1, we set $\delta_{01} = 0.2, \delta_{11} = 0.1, \delta_{21} = 0.1, \sigma_1 = 0.1$. For alternative 2, we set $\delta_{02} = 0.1, \delta_{12} = 0.3, \delta_{22} = 0.1, \sigma_2 = 0.2$.

Across all data generating processes, the unobserved characteristic x_i^* enters the pricing equations for both alternatives, which induces correlation in prices conditional on observables. In addition, x_i^* also enters the utility specification, allowing the unobserved type to jointly affect the prices individuals face and their preferences over alternatives. DGP 1 specifies an additively separable linear pricing equation, which is commonly assumed in empirical applications. DGP 2 introduces a nonlinear term. DGPs 3 and 4 consider scenarios where the pricing function takes a nonseparable form.[15]

For each DGP, we simulate offered prices and individual choices. To implement our estimator, we require an instrument z_i to recover the selected price distribution conditional on (x_{i1}, x_{i2}, x_i^*) in the first step, since x_i^* is unobserved. We construct such an instrument by assuming $z_i \sim Poisson(x_i^*)$ when $x_i^* = 1$, and $z_i = 0$ otherwise. This choice is motivated by settings where x_i^* can be interpreted as an individual's unobserved risk type, and such risk types may be reflected in the ex post realization of accidents, which are often modeled using a Poisson distribution. Because we impose

[15] Although the offered price distributions have unbounded support, in the implementation we approximate the support by the realized price range. Given the large sample size, this range contains almost all simulated probability mass. Later we show that the estimation of the offered price distribution performs well.

a parametric relationship between the instrument and the unobservable, only one instrument is needed.

We assume that the econometrician observes $(y_i, x_{i1}, x_{i2}, p_i, z_i)$, where p_i denotes the price of the chosen alternative. Using these data, we apply the procedure described in Section 4.1 to estimate the parameters of the selection function, $\theta = (\gamma, \xi_2, \beta, \kappa)$ with ξ_1 normalized to 0, along with the offered price distribution for each alternative.[16] For comparison, we first implement the classic Heckman parametric two-step method, assuming that the pricing equations are linearly separable and that the error terms in the selection and pricing equations follow a bivariate normal distribution. We also compare our estimator with the quantile selection model of Arellano and Bonhomme (2017). To implement their approach, we follow the standard practice of assuming that the quantile functions are linear in x_{i2} and that the dependence structure is governed by a Gaussian copula.[17] For each design, we run 500 simulations with sample sizes of 2,000 and 5,000 observations.

Table 1 reports the Monte Carlo biases, standard deviations, and root mean squared errors for the estimates of θ obtained using our method. Overall, the estimator performs well in finite samples across all DGPs, including those with nonseparable pricing equations. The biases are small, and the root mean squared errors remain modest for all parameters in the selection function. The standard deviation decreases as the sample size increases in all simulation designs.

For the cumulative distribution functions of $log(price)$, Tables 2 and 3 report the integrated squared biases and integrated mean squared errors for our proposed estimator, the Heckman two-step estimator, and the copula-based sample-selection correction estimator for quantile regression, separately for the two alternatives. Each row of the tables corresponds to the price distribution conditional on a specific value of x_{i2}. We also plot the true CDFs for alternatives 1 and 2 alongside the estimates produced by these models conditional on $x_{i2} = 0.25$ and $x_{i2} = 0.75$ in Figures 2 and 3, respectively. To save space, we report the CDF results only for the sample size of 2,000 observations, and we omit the figures for other values of the observable covariates.

Our method allows for nonparametric estimation of the offered price distributions,

[16] We estimate the cumulative distribution function of prices at 300 grid points.

[17] Although Arellano and Bonhomme (2017) discuss identification under more general settings, their empirical implementation focuses on cases in which the copula depends on a low-dimensional vector of parameters, which is the specification we adopt here.

Table 1: Simulation Results for Utility Parameters

	DGP 1					
	$N = 2000$			$N = 5000$		
	Bias	Std. Dev.	RMSE	Bias	Std. Dev.	RMSE
γ	-0.1017	0.1762	0.2033	-0.0697	0.1087	0.1290
β	-0.0058	0.0619	0.0621	-0.0013	0.0372	0.0371
κ	-0.0224	0.0491	0.0540	-0.0095	0.0350	0.0362
ξ_2	-0.0267	0.0570	0.0629	-0.0156	0.0349	0.0382
	DGP 2					
	$N = 2000$			$N = 5000$		
	Bias	Std. Dev.	RMSE	Bias	Std. Dev.	RMSE
γ	-0.1003	0.1672	0.1949	-0.0705	0.1067	0.1278
β	-0.0049	0.0626	0.0627	-0.0013	0.0379	0.0379
κ	-0.0185	0.0487	0.0521	-0.0072	0.0344	0.0351
ξ_2	-0.0197	0.0508	0.0544	-0.0114	0.0319	0.0338
	DGP 3					
	$N = 2000$			$N = 5000$		
	Bias	Std. Dev.	RMSE	Bias	Std. Dev.	RMSE
γ	-0.1092	0.2557	0.2778	-0.0685	0.1535	0.1680
β	-0.0011	0.0639	0.0639	0.0015	0.0367	0.0367
κ	-0.0194	0.0485	0.0522	-0.0076	0.0330	0.0339
ξ_2	-0.0074	0.0458	0.0463	-0.0032	0.0278	0.0279
	DGP 4					
	$N = 2000$			$N = 5000$		
	Bias	Std. Dev.	RMSE	Bias	Std. Dev.	RMSE
γ	0.0934	0.8041	0.8087	0.0666	0.4806	0.4847
β	0.0005	0.0613	0.0613	0.0038	0.0361	0.0363
κ	-0.0194	0.0474	0.0512	-0.0103	0.0335	0.0350
ξ_2	-0.0003	0.0451	0.0451	0.0016	0.0280	0.0280

whereas the alternative approaches impose parametric restrictions on either the conditional mean or quantiles of the pricing distributions, or on the dependence structure through the copula. Tables 2 and 3 show that our estimator achieves very low integrated squared bias and integrated mean squared error for the CDFs of $log(price)$ across all simulation designs and for all values of x_{i2}. In contrast, while the classic Heckman two-step method and the quantile selection model perform well in DGP 1, their biases and mean squared errors increase substantially as the pricing equation becomes more complex in DGPs 2–4. These results are expected, since the parametric assumptions underlying these methods, such as linear conditional mean or quantile functions and a Gaussian copula, are severely violated in these designs.

Figures 2 and 3 provide a visual illustration of these results. We can see that across all simulation designs, the estimated CDFs of $log(price)$ for both alternatives produced by our functional contraction approach closely track the true CDFs, as indicated by the black curves with "+" markers and the red solid curve in Figures 2 and 3. By comparison, the biases of the Heckman two-step method (blue dashed curves) and the quantile selection model (purple dash–dotted curves) can be substantial, particularly in DGPs 3 and 4. The direction and magnitude of these biases also vary with the values of the observable covariates.

Another key advantage of our approach is that it does not require an instrument to exogenously shift the selection probability. It is well known in the literature that the two-step method is nearly unidentified when the same regressors are used in both the selection function and the outcome equation even under strong parametric restrictions. This occurs because the inverse Mills ratio is approximately linear over a wide range of its argument. In practice, it is also difficult to find variables that affect selection but can be excluded from the outcome equation.

In contrast, our approach does not require such an excluded variable. To illustrate this, we conduct a set of Monte Carlo simulations where the excluded variable x_{i1} is removed from the indirect utility, using the same four DGPs for $log(price)$. The results for this specification are reported in Tables 4–5 in Appendix B. As shown, our estimator performs well in finite samples, even without an additional excluded variable to exogenously shift the selection probability. Our estimator consistently shows low biases across different DGPs and exhibits a decreasing standard deviation as the sample size increases.

Our method requires the econometrician to correctly specify the functional form of

Table 2: Simulation Results for CDF of $log(p_1)$

	DGP 1					
	Functional Contraction		Heckman Two-Step		Quantile Selection	
	IBias^2	IMSE	IBias^2	IMSE	IBias^2	IMSE
$x_{i2} = 0$	0.0005	0.0017	0.0001	0.0039	0.0001	0.0046
$x_{i2} = 0.25$	0.0004	0.0015	0.0001	0.0033	0.0001	0.0038
$x_{i2} = 0.5$	0.0002	0.0012	0.0001	0.0027	0.0001	0.0031
$x_{i2} = 0.75$	0.0002	0.0010	0.0001	0.0023	0.0001	0.0026
$x_{i2} = 1$	0.0001	0.0010	0.0001	0.0021	0.0001	0.0023
	DGP 2					
	Functional Contraction		Heckman Two-Step		Quantile Selection	
	IBias^2	IMSE	IBias^2	IMSE	IBias^2	IMSE
$x_{i2} = 0$	0.0005	0.0017	0.0235	0.0275	0.0185	0.0231
$x_{i2} = 0.25$	0.0005	0.0017	0.0017	0.0053	0.0031	0.0072
$x_{i2} = 0.5$	0.0003	0.0014	0.0124	0.0153	0.0145	0.0177
$x_{i2} = 0.75$	0.0002	0.0011	0.0037	0.0062	0.0052	0.0080
$x_{i2} = 1$	0.0001	0.0010	0.0133	0.0154	0.0106	0.0129
	DGP 3					
	Functional Contraction		Heckman Two-Step		Quantile Selection	
	IBias^2	IMSE	IBias^2	IMSE	IBias^2	IMSE
$x_{i2} = 0$	0.0007	0.0019	0.0207	0.0259	0.0017	0.0071
$x_{i2} = 0.25$	0.0002	0.0013	0.0061	0.0102	0.0011	0.0052
$x_{i2} = 0.5$	0.0002	0.0013	0.0017	0.0049	0.0013	0.0049
$x_{i2} = 0.75$	0.0002	0.0012	0.0016	0.0041	0.0002	0.0037
$x_{i2} = 1$	0.0002	0.0013	0.0053	0.0074	0.0007	0.0040
	DGP 4					
	Functional Contraction		Heckman Two-Step		Quantile Selection	
	IBias^2	IMSE	IBias^2	IMSE	IBias^2	IMSE
$x_{i2} = 0$	0.0014	0.0023	0.0432	0.0463	0.0390	0.0425
$x_{i2} = 0.25$	0.0014	0.0024	0.0147	0.0182	0.0153	0.0190
$x_{i2} = 0.5$	0.0011	0.0021	0.0412	0.0443	0.0358	0.0392
$x_{i2} = 0.75$	0.0012	0.0021	0.0106	0.0141	0.0059	0.0100
$x_{i2} = 1$	0.0005	0.0018	0.0270	0.0304	0.0346	0.0384

Note: The IBias^2 of a function h is calculated as follows. Let $\hat{h}_r$ be the estimate of h from the r-th simulated dataset, and $\bar{h}(p) = \frac{1}{R}\sum_{r=1}^{R}\hat{h}_r(p)$ be the point-wise average over R simulations. The integrated squared bias is calculated by numerically integrating the point-wise squared bias $(\bar{h}(p) - h(p))^2$ over the distribution of p. The integrated MSE is computed in a similar way. The values reported in each row correspond to the price distributions conditional on a given value of x_{i2}. The results shown in this table are based on 500 Monte Carlo replications with a sample size of 2,000. Corresponding results for a sample size of 5,000 are available upon request.

Table 3: Simulation Results for CDF of $log(p_2)$

	DGP 1					
	Functional Contraction		Heckman Two-Step		Quantile Selection	
	IBias^2	IMSE	IBias^2	IMSE	IBias^2	IMSE
$x_{i2} = 0$	0.0002	0.0008	0.0000	0.0016	0.0001	0.0018
$x_{i2} = 0.25$	0.0002	0.0009	0.0000	0.0019	0.0001	0.0020
$x_{i2} = 0.5$	0.0002	0.0010	0.0000	0.0023	0.0001	0.0024
$x_{i2} = 0.75$	0.0002	0.0011	0.0000	0.0028	0.0001	0.0030
$x_{i2} = 1$	0.0002	0.0012	0.0000	0.0034	0.0001	0.0038
	DGP 2					
	Functional Contraction		Heckman Two-Step		Quantile Selection	
	IBias^2	IMSE	IBias^2	IMSE	IBias^2	IMSE
$x_{i2} = 0$	0.0002	0.0008	0.0231	0.0246	0.0211	0.0228
$x_{i2} = 0.25$	0.0002	0.0008	0.0058	0.0075	0.0067	0.0086
$x_{i2} = 0.5$	0.0002	0.0009	0.0200	0.0220	0.0213	0.0234
$x_{i2} = 0.75$	0.0002	0.0010	0.0030	0.0058	0.0040	0.0069
$x_{i2} = 1$	0.0003	0.0011	0.0368	0.0401	0.0329	0.0365
	DGP 3					
	Functional Contraction		Heckman Two-Step		Quantile Selection	
	IBias^2	IMSE	IBias^2	IMSE	IBias^2	IMSE
$x_{i2} = 0$	0.0031	0.0036	0.0492	0.0515	0.0030	0.0047
$x_{i2} = 0.25$	0.0005	0.0011	0.0190	0.0212	0.0052	0.0072
$x_{i2} = 0.5$	0.0003	0.0010	0.0047	0.0068	0.0019	0.0041
$x_{i2} = 0.75$	0.0002	0.0009	0.0014	0.0035	0.0002	0.0026
$x_{i2} = 1$	0.0002	0.0011	0.0112	0.0131	0.0049	0.0072
	DGP 4					
	Functional Contraction		Heckman Two-Step		Quantile Selection	
	IBias^2	IMSE	IBias^2	IMSE	IBias^2	IMSE
$x_{i2} = 0$	0.0011	0.0016	0.1443	0.1459	0.1123	0.1141
$x_{i2} = 0.25$	0.0009	0.0015	0.0568	0.0592	0.0455	0.0483
$x_{i2} = 0.5$	0.0005	0.0011	0.1095	0.1109	0.0952	0.0966
$x_{i2} = 0.75$	0.0003	0.0009	0.0290	0.0308	0.0133	0.0155
$x_{i2} = 1$	0.0002	0.0007	0.0530	0.0546	0.0722	0.0741

Note: The IBias^2 of a function h is calculated as follows. Let $\hat{h}_r$ be the estimate of h from the r-th simulated dataset, and $\bar{h}(p) = \frac{1}{R}\sum_{r=1}^{R} \hat{h}_r(p)$ be the point-wise average over R simulations. The integrated squared bias is calculated by numerically integrating the point-wise squared bias $(\bar{h}(p) - h(p))^2$ over the distribution of p. The integrated MSE is computed in a similar way. The values reported in each row correspond to the price distributions conditional on a given value of x_{i2}. The results shown in this table are based on a 500 Monte Carlo replications with a sample size of 2,000. Corresponding results for a sample size of 5,000 are available upon request.

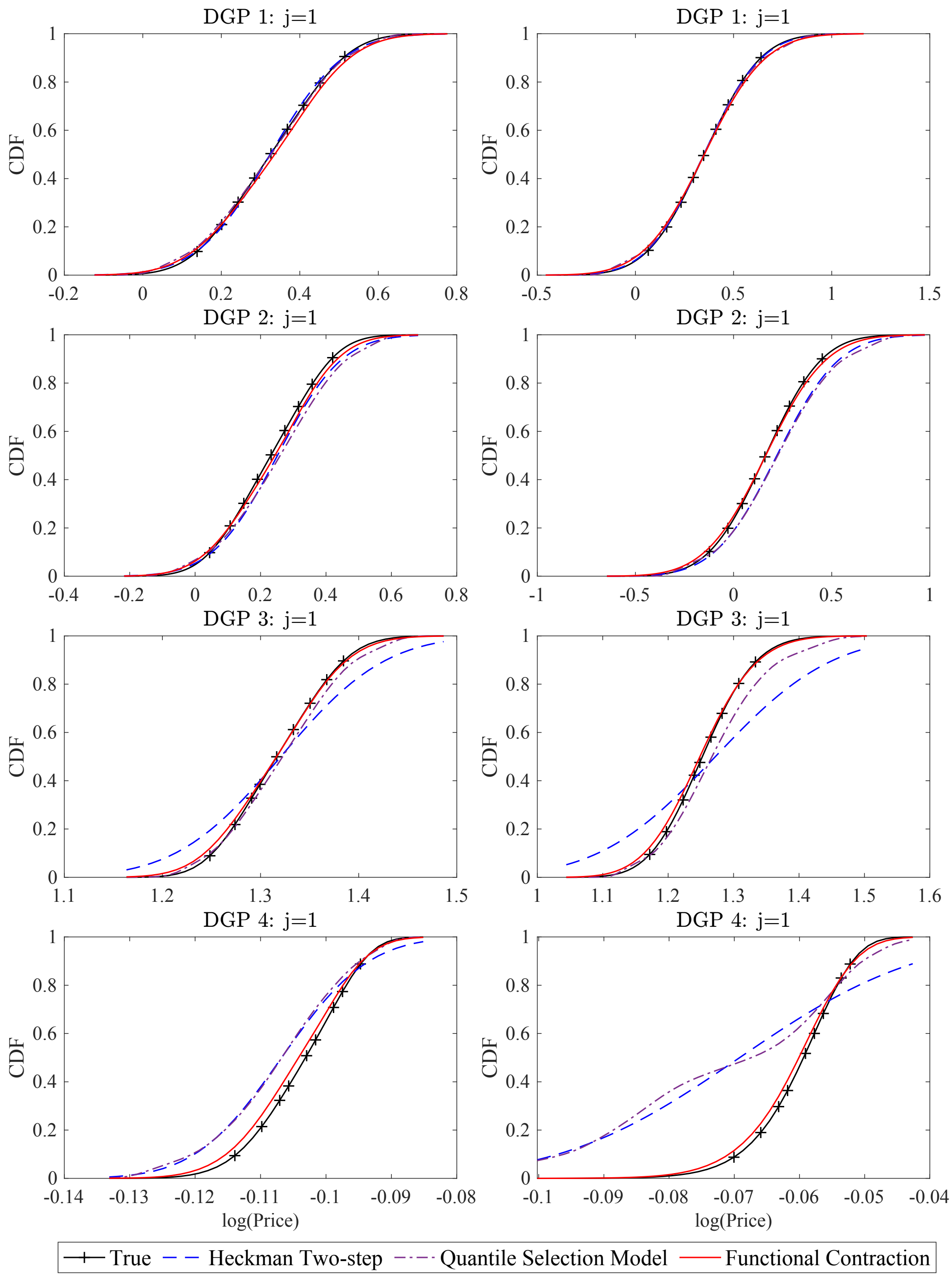

Figure 2: CDFs of $log(price)$ for alternatives 1 and 2, conditional on $x_{i2} = 0.25$. The black curve with "+" markers, the blue dashed curve, the purple dash–dotted curve, and the red solid curve correspond to the true CDF, the Heckman two-step estimate, the quantile selection estimate, and the functional contraction estimate, respectively, based on 500 Monte Carlo replications with a sample size of 2,000.

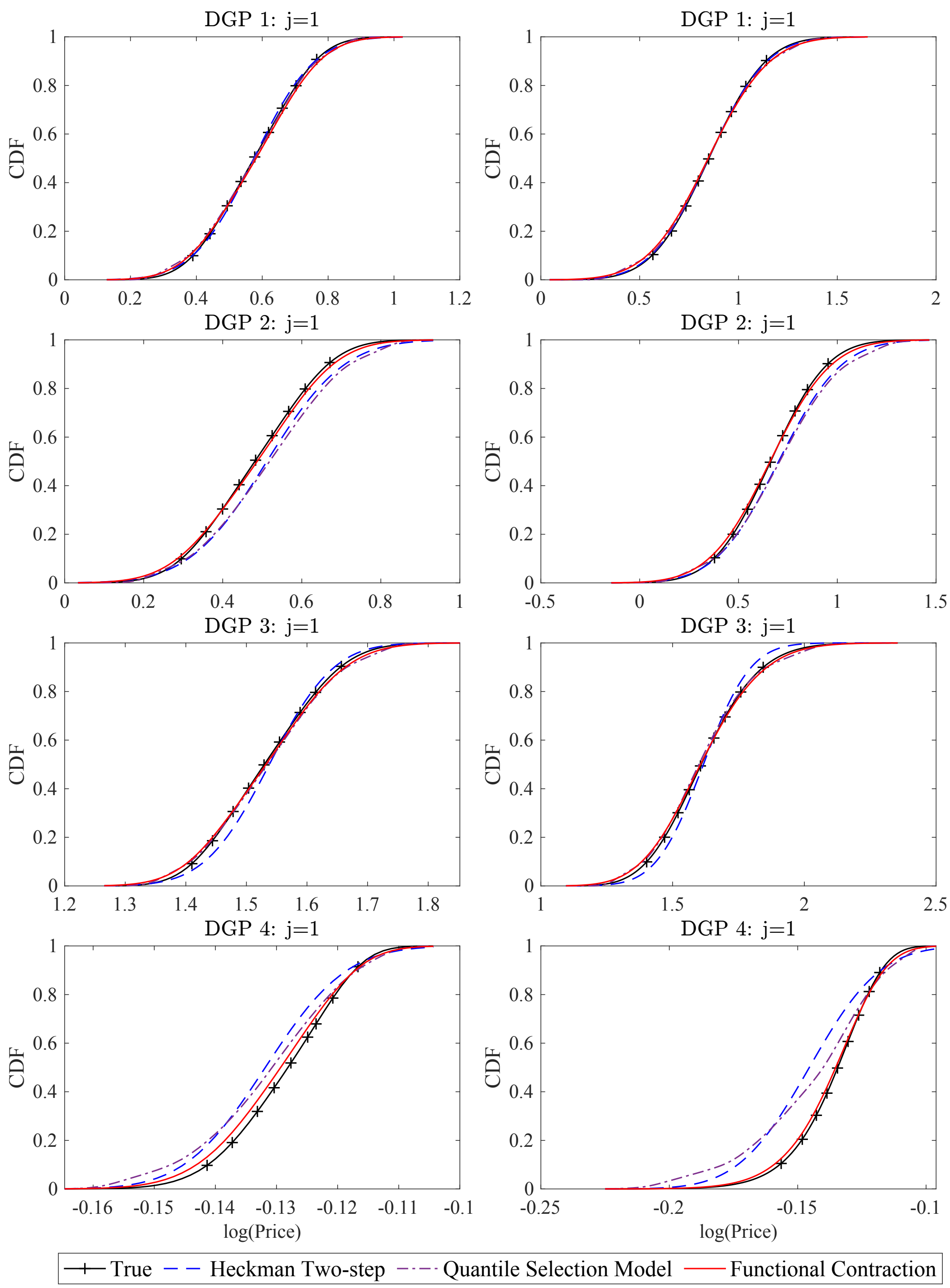

Figure 3: CDFs of $log(price)$ for alternatives 1 and 2, conditional on $x_{i2} = 0.75$. The black curve with "+" markers, the blue dashed curve, the purple dash–dotted curve, and the red solid curve correspond to the true CDF, the Heckman two-step estimate, the quantile selection estimate, and the functional contraction estimate, respectively, based on 500 Monte Carlo replications with a sample size of 2,000.

the selection function. To evaluate how the estimator performs under misspecification, we conduct a series of Monte Carlo simulations in which the econometrician assumes that ε follows a logistic distribution, while in truth it is generated from a normal distribution. In Tables 6–7 in Appendix B, we report the estimation results for the utility parameters and CDFs of $log(price)$ under this misspecification. For the utility parameters, we rescale the estimates by the scale parameter of the logit model to make them comparable to those in the original probit specification. After this adjustment, the biases are small. The estimator for the offered price distributions also performs well: the integrated squared biases and mean squared errors of the CDFs remain close to those in Tables 2 and 3. These results suggest that the estimator of the offered price distributions is robust to misspecification of the selection function, an appealing feature in practice, particularly when the econometrician has limited prior information about the correct functional form.

Finally, we briefly discuss how the functional contraction performs computationally in practice. We compute ρ^* for all four simulation designs and find that it is below 1 in every case. The average numbers of iterations needed to reach convergence (with a tolerance of 10^{-5}) are 3.8, 3.8, 3.4, and 2.1 for DGPs 1 through 4, respectively (averaged over 500 replications). These results indicate that the proposed estimator is computationally efficient, converges rapidly, and remains stable across a range of data generating processes, making it well suited for applied work.

6 Discussion of Empirical Applications

Our estimator introduced in Section 4 is broadly applicable to a wide range of empirical settings. It effectively addresses the challenge of selection bias that arises when only the outcomes of chosen alternatives are observed. The method has three features that are particularly important for empirical applications. First, it imposes no parametric or separability restrictions on the potential outcome distributions and allows them to vary flexibly across alternatives. Second, the framework accommodates unobservable characteristics in both the outcome distributions and the selection model, capturing selection on unobservables. Third, the selection function can incorporate alternative-specific unobserved heterogeneity and does not require an excluded variable, which is desirable in many empirical settings.

An important empirical application that illustrates these advantages is consumer demand estimation in markets where only transaction prices are observed. In classic differentiated product demand estimation pioneered by Berry (1994) and Berry et al. (1995), the price of a product is often assumed to be uniform across all consumers (e.g., the list price of a vehicle). But this assumption does not hold in contexts involving price discrimination or personalized pricing (D'Haultfœuille et al., 2019; Sagl, 2023; Buchholz et al., 2020; Dubé and Misra, 2023), discount negotiation (Goldberg, 1996; Allen et al., 2014), or risk-based pricing (Crawford et al., 2018; Cosconati et al., 2025). In these contexts, researchers can relatively easily gather data on the transaction prices consumers pay, but it is challenging to gain access to competing prices offered to consumers.

In a companion paper with coauthors (Cosconati et al., 2025), we apply our method to estimate demand and insurance companies' information technology in the auto insurance market, where only the transaction prices of selected insurance plans are observed. In this market, insurance companies employ risk-based pricing. For each consumer, an insurance company generates a noisy estimate of their risk type and prices accordingly. Our goal is to quantify the heterogeneity in insurers' information technology, as measured by the dispersion of their risk estimates. Since the shape of the offered price distribution reflects the distribution of risk estimates, allowing for flexible estimation of the offered price distribution is crucial.

In this application, we assume that the offered prices across different firms are independent conditional on observable characteristics and the consumer's true unobserved risk type. At the same time, the consumer's risk type may also influence their preferences over insurance products. For example, higher-risk consumers may prefer insurers with higher service quality. We therefore allow the true risk type to affect both the pricing distributions and the utility parameters. Our data include realized claim records for each consumer over multiple years, and we use these records as instruments for the latent risk type in the first-step estimation.

We nonparametrically estimate each insurance company's offered price distribution using our functional contraction approach. In Figure 4, we plot the CDFs of offered price for several firms based on estimates in Cosconati et al. (2025). The distributions differ substantially across firms, indicating significant heterogeneity in their pricing strategies. Building on this result, we estimate each firm's information precision parameter using supply-side model restrictions. These estimates provide important insights for analyzing competition under heterogeneous information structures in this market.

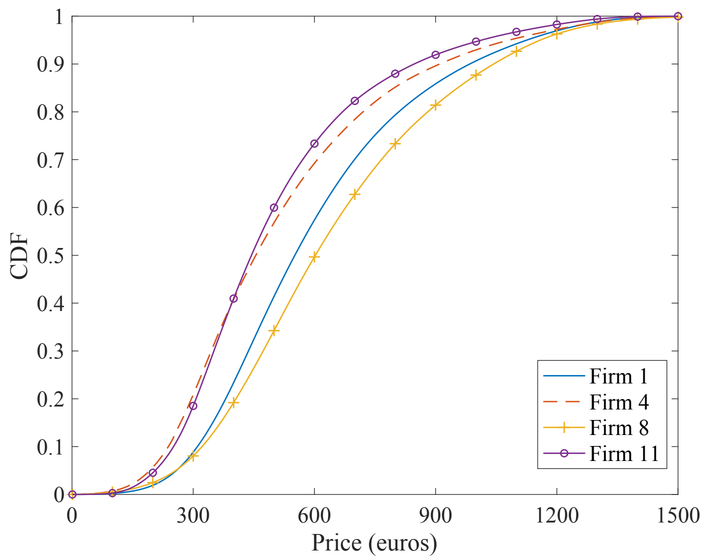

Figure 4: CDFs of the offered price distributions for firms 1, 4, 8, and 11 based on estimates from Cosconati et al. (2025). The CDFs are averaged across different characteristic groups.

From a practical point of view, our iterative procedure to numerically solve for the offered price distributions given demand parameters is easy to implement and performs well in practice. In our empirical application using data from 11 insurers, the iterative algorithm converges very quickly, typically requiring only 6–7 iterations.

The usefulness of our method is not limited to consumer demand. It can also be applied to auction models and Roy models, where similar selection issues arise. For example, in multi-attribute auctions, our approach can be used to nonparametrically recover the full bid distribution and the auctioneer's scoring weights when only the winning bids and the winner's identity are observed, even in the presence of bidder asymmetry.[18] Auctions in many settings have used the scoring rule that departs from the pure price-based criterion by accounting for quality differences,[19] and our framework can flexibly accommodate these multi-attribute scoring mechanisms with both observable and unobservable components. A similar application arises in Roy

[18] Flexibly accommodating bidder asymmetries is a well-known challenge in auction models (Athey and Haile, 2007). Bidder asymmetries may arise from factors such as distance to the contract location (Flambard and Perrigne, 2006), information advantages (Hendricks and Porter, 1988; De Silva et al., 2009), varying risk attitudes (Campo, 2012), or strategic sophistication (Hortaçsu et al., 2019).

[19] See, for example, Asker and Cantillon (2008); Lewis and Bajari (2011); Nakabayashi (2013); Yoganarasimhan (2016); Takahashi (2018); Krasnokutskaya et al. (2020); Allen et al. (2024).

models, where our method can recover the distribution of potential wages when only realized wages in the chosen sector are observed. Our framework enables researchers to recover these distributions flexibly and without relying on excluded variables in the selection equation, which are frequently difficult to justify in applied settings. This capability provides a valuable tool for studying key questions in labor economics, such as occupational choice and wage inequality.

7 Conclusion

We introduce a novel method for estimating nonseparable selection models. We show that for a given selection function, potential outcome distributions are nonparametrically identified from the distribution of selected outcomes and can be recovered using a simple iterative algorithm. We achieve this by constructing an operator whose fixed point is the potential outcome distributions and proving that this operator is a functional contraction. Building on this theoretical result, we propose a two-step estimation strategy for both the selection function and potential outcome distributions. The consistency and asymptotic normality of the proposed estimators are established.

Our method has several important features. First, we allow the outcome equation to be fully nonparametric and nonseparable in error terms, and we recover the entire distribution of potential outcomes rather than focusing on specific moments or quantiles. In essence, we correct for sample selection bias by examining how the bias is *systematically* generated by the selection model. Second, our approach allows for fully heterogeneous effects of covariates on outcomes, which is a crucial feature for empirical analysis, as discussed in Chernozhukov et al. (2025). Another key advantage of our approach is that it does not require an excluded variable that shifts the selection probability without entering the outcome equation, which are often challenging to find in empirical settings, or on identification-at-infinity arguments. Finally, our approach also accommodates asymmetry in outcome distributions across alternatives and flexibly incorporates unobserved alternative-specific heterogeneity in the selection model.

We find that the proposed estimation strategy performs well in both simulations and real-world data applications; see our demand estimation using insurance market data in the companion paper Cosconati et al. (2025). The approach is straightforward to implement and computationally efficient, making it highly appealing to empirical researchers. More broadly, the estimator can be applied in a wide range of settings in which only selected outcomes are observed, including consumer demand models with only transaction prices, auctions with incomplete bid data, and various selection models in labor economics. Our method is particularly valuable in applications where the entire distribution of outcomes is of interest.

References

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

Allen, J., R. Clark, B. Hickman, and E. Richert (2024): "Resolving failed banks: Uncertainty, multiple bidding and auction design," *Review of Economic Studies*, 91, 1201–1242.

Allen, J., R. Clark, and J.-F. Houde (2014): "Price dispersion in mortgage markets," *The Journal of Industrial Economics*, 62, 377–416.

——— (2019): "Search frictions and market power in negotiated-price markets," *Journal of Political Economy*, 127, 1550–1598.

Andrews, D. W. and M. M. Schafgans (1998): "Semiparametric estimation of the intercept of a sample selection model," *The Review of Economic Studies*, 65, 497–517.

Arellano, M. and S. Bonhomme (2017): "Quantile selection models with an application to understanding changes in wage inequality," *Econometrica*, 85, 1–28.

Asker, J. and E. Cantillon (2008): "Properties of scoring auctions," *The RAND Journal of Economics*, 39, 69–85.

Athey, S. and P. A. Haile (2002): "Identification of standard auction models," *Econometrica*, 70, 2107–2140.

——— (2007): "Nonparametric approaches to auctions," *Handbook of econometrics*, 6, 3847–3965.

Baricz, Á. (2008): "Mills' ratio: Monotonicity patterns and functional inequalities," *Journal of Mathematical Analysis and Applications*, 340, 1362–1370.

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

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

Buchholz, N., L. Doval, J. Kastl, F. Matějka, and T. Salz (2020): "The value of time: Evidence from auctioned cab rides," *CEPR Discussion Paper No. DP14666*.

Bushell, P. J. (1973): "Hilbert's metric and positive contraction mappings in a Banach space," *Archive for Rational Mechanics and Analysis*, 52, 330–338.

Campo, S. (2012): "Risk aversion and asymmetry in procurement auctions: Identification, estimation and application to construction procurements," *Journal of Econometrics*, 168, 96–107.

Chen, S. and S. Khan (2003): "Semiparametric estimation of a heteroskedastic sample selection model," *Econometric Theory*, 19, 1040–1064.

Chernozhukov, V., I. Fernández-Val, and S. Luo (2025): "Distribution regression with sample selection and UK wage decomposition," *Journal of Political Economy*, 133, 3952–3992.

Cicala, S. (2015): "When does regulation distort costs? Lessons from fuel procurement in us electricity generation," *American Economic Review*, 105, 411–444.

Cosconati, M., Y. Xin, F. Wu, and Y. Jin (2025): "Competing under information heterogeneity: Evidence from auto insurance," *The Review of Economic Studies*, forthcoming.

Crawford, G. S., N. Pavanini, and F. Schivardi (2018): "Asymmetric information and imperfect competition in lending markets," *American Economic Review*, 108, 1659–1701.

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

De Silva, D. G., G. Kosmopoulou, and C. Lamarche (2009): "The effect of information on the bidding and survival of entrants in procurement auctions," *Journal of Public Economics*, 93, 56–72.

Dubé, J.-P. and S. Misra (2023): "Personalized pricing and consumer welfare," *Journal of Political Economy*, 131, 131–189.

D'Haultfœuille, X., I. Durrmeyer, and P. Février (2019): "Automobile prices in market equilibrium with unobserved price discrimination," *The Review of Economic Studies*, 86, 1973–1998.

D'Haultfoeuille, X. and A. Maurel (2013): "Another look at the identification at infinity of sample selection models," *Econometric Theory*, 29, 213–224.

D'Haultfœuille, X., A. Maurel, and Y. Zhang (2018): "Extremal quantile regressions for selection models and the black–white wage gap," *Journal of Econometrics*, 203, 129–142.

Fernández-Val, I., A. van Vuuren, and F. Vella (2024): "Nonseparable sample selection models with censored selection rules," *Journal of Econometrics*, 240, 105088.

Flambard, V. and I. Perrigne (2006): "Asymmetry in procurement auctions: Evidence from snow removal contracts," *The Economic Journal*, 116, 1014–1036.

Goldberg, P. K. (1996): "Dealer price discrimination in new car purchases: Evidence from the consumer expenditure survey," *Journal of Political Economy*, 104, 622–654.

Gronau, R. (1974): "Wage comparisons–A selectivity bias," *Journal of Political Economy*, 82, 1119–1143.

Guerre, E. and Y. Luo (2019): "Nonparametric identification of first-price auction with unobserved competition: A density discontinuity framework," *arXiv preprint arXiv:1908.05476*.

Heckman, J. J. (1974): "Shadow prices, market wages, and labor supply," *Econometrica: journal of the econometric society*, 679–694.

——— (1976): "The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models," in *Annals of economic and social measurement, volume 5, number 4*, NBER, 475–492.

——— (1979): "Sample selection bias as a specification error," *Econometrica*, 47, 153–161.

Hendricks, K. and R. H. Porter (1988): "An empirical study of an auction with asymmetric information," *The American Economic Review*, 865–883.

Hortaçsu, A., F. Luco, S. L. Puller, and D. Zhu (2019): "Does strategic ability affect efficiency? Evidence from electricity markets," *American Economic Review*, 109, 4302–4342.

Hu, Y. (2008): "Identification and estimation of nonlinear models with misclassification error using instrumental variables: A general solution," *Journal of Econometrics*, 144, 27–61.

Hu, Y. and S. M. Schennach (2008): "Instrumental variable treatment of nonclassical measurement error models," *Econometrica*, 76, 195–216.

Komarova, T. (2013): "A new approach to identifying generalized competing risks models with application to second-price auctions," *Quantitative Economics*, 4, 269–328.

Krasnokutskaya, E., K. Song, and X. Tang (2020): "The role of quality in internet service markets," *Journal of Political Economy*, 128, 75–117.

Lee, L.-F. (1982): "Some approaches to the correction of selectivity bias," *The Review of Economic Studies*, 49, 355–372.

——— (1983): "Generalized econometric models with selectivity," *Econometrica: Journal of the Econometric Society*, 507–512.

Lemmens, B. and R. Nussbaum (2012): *Nonlinear Perron-Frobenius Theory*, vol. 189, Cambridge University Press.

Lewis, G. and P. Bajari (2011): "Procurement contracting with time incentives: Theory and evidence," *The Quarterly Journal of Economics*, 126, 1173–1211.

Meilijson, I. (1981): "Estimation of the lifetime distribution of the parts from the autopsy statistics of the machine," *Journal of Applied Probability*, 18, 829–838.

Nakabayashi, J. (2013): "Small business set-asides in procurement auctions: An empirical analysis," *Journal of Public Economics*, 100, 28–44.

Newey, W. K. (2007): "Nonparametric continuous/discrete choice models," *International Economic Review*, 48, 1429–1439.

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

Newey, W. K. and D. McFadden (1994): "Large sample estimation and hypothesis testing," *Handbook of Econometrics*, 4, 2111–2245.

Roy, A. D. (1951): "Some thoughts on the distribution of earnings," *Oxford Economic Papers*, 3, 135–146.

Sagl, S. (2023): "Dispersion, discrimination, and the price of your pickup," *Working paper.*

Takahashi, H. (2018): "Strategic design under uncertain evaluations: Structural analysis of design-build auctions," *The RAND Journal of Economics*, 49, 594–618.

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

Yoganarasimhan, H. (2016): "Estimation of beauty contest auctions," *Marketing Science*, 35, 27–54.

A Omitted Proofs

A.1 Proof of Theorem 1

Fix a probability measure $G_j \in \Delta([\underline{p}_j, \overline{p}_j])$ and define

$$\mathcal{X}(G_j) = \left\{ \Psi \in \Delta([\underline{p}_j, \overline{p}_j]) : d_H(\Psi, G_j) < \infty \right\}.$$

Equivalently, $\Psi \in \mathcal{X}(G_j)$ if and only if $\Psi \sim G_j$ and

$$\log \frac{d\Psi}{dG_j} \in L^\infty(G_j).$$

Lemma 1 (Completeness of the finite-distance component)**.** *Metric space $(\mathcal{X}_j(G_j), d_H)$ is complete. Consequently,*

$$\mathcal{X}(G) = \prod_{j \in \mathcal{J}} \mathcal{X}_j(G_j)$$

is complete under the max metric D.

Proof. Consider the one-dimensional component first. Let $(\Phi_n)_{n \geq 1}$ be a Cauchy sequence in $(\mathcal{X}(G_j), d_H)$. Then

$$\log \frac{d\Phi_n}{dG_j} \in L^\infty(G_j)$$

and

$$\int \exp\left(\log \frac{d\Phi_n}{dG_j}\right) dG_j = 1$$

for all n.

For any n, m, we have

$$d_H(\Phi_n, \Phi_m) = \mathrm{osc}_{G_j}\left(\log \frac{d\Phi_n}{dG_j} - \log \frac{d\Phi_m}{dG_j}\right),$$

where

$$\mathrm{osc}_G(v) = \operatorname*{ess\,sup}_G v - \operatorname*{ess\,inf}_G v.$$

Moreover,

$$\int \exp\left(\log \frac{d\Phi_n}{dG_j} - \log \frac{d\Phi_m}{dG_j}\right) d\Phi_m = \int \frac{d\Phi_n}{d\Phi_m}\, d\Phi_m = 1.$$

Hence,

$$\operatorname*{ess\,inf}_{G_j}\left(\log\frac{d\Phi_n}{dG_j}-\log\frac{d\Phi_m}{dG_j}\right)\le 0\le \operatorname*{ess\,sup}_{G_j}\left(\log\frac{d\Phi_n}{dG_j}-\log\frac{d\Phi_m}{dG_j}\right),$$

where the essential infimum and supremum can equivalently be taken with respect to G_j or Φ_m, since $\Phi_m \sim G_j$. Therefore,

$$\left\|\log\frac{d\Phi_n}{dG_j}-\log\frac{d\Phi_m}{dG_j}\right\|_\infty \le \operatorname{osc}_{G_j}\left(\log\frac{d\Phi_n}{dG_j}-\log\frac{d\Phi_m}{dG_j}\right)=d_H(\Phi_n,\Phi_m).$$

Since (Φ_n) is Cauchy under d_H, it follows that

$$\left(\log\frac{d\Phi_n}{dG_j}\right)_{n\ge 1}$$

is a Cauchy sequence in $L^\infty(G_j)$. Because $L^\infty(G_j)$ is a Banach space, there exists $\ell \in L^\infty(G_j)$ such that

$$\left\|\log\frac{d\Phi_n}{dG_j}-\ell\right\|_\infty \to 0.$$

Define a measure Φ on $[\underline{p}_j, \overline{p}_j]$ by

$$d\Phi = e^\ell dG_j.$$

We first show that Φ is a probability measure. Since

$$\left\|\log\frac{d\Phi_n}{dG_j}-\ell\right\|_\infty \to 0,$$

for all sufficiently large n,

$$\exp\left(\log\frac{d\Phi_n}{dG_j}\right)\le e^{\ell+1}.$$

The function $e^{\ell+1}$ is G_j-integrable because $\ell \in L^\infty(G_j)$ and G_j is a probability measure. Hence, by dominated convergence,

$$\int e^\ell dG_j = \lim_{n\to\infty}\int \exp\left(\log\frac{d\Phi_n}{dG_j}\right)dG_j = 1.$$

Thus $\Phi \in \Delta([\underline{p}_j, \overline{p}_j])$.

Since $\ell \in L^\infty(G_j)$, the density $e^\ell = d\Phi/dG_j$ is bounded above and bounded away from zero. Hence $\Phi \sim G_j$ and

$$d_H(\Phi, G_j) < \infty.$$

Therefore $\Phi \in \mathcal{X}(G_j)$.

Finally,

$$\begin{aligned} d_H(\Phi_n, \Phi) &= \operatorname{osc}_{G_j}\left(\log \frac{d\Phi_n}{dG_j} - \log \frac{d\Phi}{dG_j}\right) \\ &= \operatorname{osc}_{G_j}\left(\log \frac{d\Phi_n}{dG_j} - \ell\right) \\ &\leq 2 \left\| \log \frac{d\Phi_n}{dG_j} - \ell \right\|_\infty \to 0. \end{aligned}$$

Therefore $(\mathcal{X}(G_j), d_H)$ is complete.

Since the full space

$$\mathcal{X}(G) = \prod_{j \in \mathcal{J}} \mathcal{X}(G_j)$$

is a finite product of complete metric spaces, it is complete under the max metric

$$D(\Psi, \Phi) = \max_{j \in \mathcal{J}} d_H(\Psi_j, \Phi_j).$$

□

Proof of Theorem 1. Fix $j \in \mathcal{J}$. By the definition of T,

$$d(T\Psi)_j(p_j) = \frac{d\tilde{G}_j(p_j)/Pr_j(p_j;\Psi)}{\int d\tilde{G}_j(y)/Pr_j(y;\Psi)}.$$

Therefore,

$$\frac{d(T\Psi)_j}{d(T\Phi)_j}(p_j) = C_j(\Psi, \Phi)\frac{Pr_j(p_j;\Phi)}{Pr_j(p_j;\Psi)},$$

where $C_j(\Psi, \Phi) > 0$ is a normalizing constant that does not depend on p_j. Since Hilbert's projective metric is invariant to multiplication by positive constants, the normalizing constant cancels out. Hence,

$$d_H((T\Psi)_j, (T\Phi)_j) \leq \operatorname{osc}_{p_j} \log \frac{Pr_j(p_j;\Psi)}{Pr_j(p_j;\Phi)}.$$

where $\operatorname{osc} g = \sup g - \inf g$. The inequality becomes equality when $\tilde{G}_j$ has full support on $[\underline{p}_j, \bar{p}_j]$.

Write

$$\Psi_{-j} = \bigotimes_{k \neq j} \Psi_k, \qquad \Phi_{-j} = \bigotimes_{k \neq j} \Phi_k.$$

For fixed j, the map

$$\Psi_{-j} \mapsto Pr_j(\cdot; \Psi)$$

is the restriction to product probability measures of a positive linear operator from the cone of finite nonnegative measures on

$$\prod_{k \neq j} [\underline{p}_k, \overline{p}_k]$$

to the cone of positive functions on $[\underline{p}_j, \overline{p}_j]$. Indeed, for any finite positive measure Ψ_{-j} on $\prod_{k \neq j} [\underline{p}_k, \overline{p}_k]$, define

$$Pr_j(p_j; \Psi_{-j}) = \int f_j(p_j, p_{-j}) \, d\Psi_{-j}(p_{-j}).$$

This map is linear in Ψ_{-j} and positive because $f_j > 0$. To invoke Birkhoff contraction theorem, the first step is to bound the projective diameter of this positive linear operator. See Bushell (1973) for an excellent introduction on Hilbert's metric and Birkhoff theorem.[20]

Take any two finite positive measures Φ_{-j} and Ψ_{-j} on $\prod_{k \neq j} [\underline{p}_k, \bar{p}_k]$. For any p_j, p'_j,

$$\frac{Pr_j(p_j; \Phi_{-j}) Pr_j(p'_j; \Psi_{-j})}{Pr_j(p_j; \Psi_{-j}) Pr_j(p'_j; \Phi_{-j})} = \frac{\iint f_j(p_j, \boldsymbol{p}_{-j}) f_j(p'_j, \boldsymbol{p}'_{-j}) \, d\Phi_{-j}(\boldsymbol{p}_{-j}) d\Psi_{-j}(\boldsymbol{p}'_{-j})}{\iint f_j(p_j, \boldsymbol{p}'_{-j}) f_j(p'_j, \boldsymbol{p}_{-j}) \, d\Phi_{-j}(\boldsymbol{p}_{-j}) d\Psi_{-j}(\boldsymbol{p}'_{-j})}.$$

By the definition of Δ_j, for all $p_j, p'_j, \boldsymbol{p}_{-j}, \boldsymbol{p}'_{-j}$,

$$f_j(p_j, \boldsymbol{p}_{-j}) f_j(p'_j, \boldsymbol{p}'_{-j}) \leq \exp(\Delta_j) f_j(p_j, \boldsymbol{p}'_{-j}) f_j(p'_j, \boldsymbol{p}_{-j}).$$

[20] Lemmens and Nussbaum (2012) provides a detailed discussion on Perron-Frobenius theory and Birkhoff theorem.

Integrating both sides with respect to $d\Phi_{-j}(\boldsymbol{p}_{-j})d\Psi_{-j}(\boldsymbol{p}'_{-j})$, we obtain

$$\frac{Pr_j(p_j;\Phi_{-j})Pr_j(p'_j;\Psi_{-j})}{Pr_j(p_j;\Psi_{-j})Pr_j(p'_j;\Phi_{-j})} \leq \exp(\Delta_j).$$

Taking logs and then the supremum over p_j, p'_j gives

$$d_H(Pr_j(\cdot;\Phi_{-j}), Pr_j(\cdot;\Psi_{-j})) \leq \Delta_j.$$

Thus, the projective diameter of the positive linear operator $\Psi_{-j} \mapsto Pr_j(\cdot;\Psi)$ is at most Δ_j.

By the Birkhoff–Hopf contraction theorem for positive linear operators under Hilbert's projective metric,

$$d_H(Pr_j(\cdot;\Psi), Pr_j(\cdot;\Phi)) \leq \tanh\left(\frac{\Delta_j}{4}\right) d_H(\Psi_{-j}, \Phi_{-j}). \tag{15}$$

Combining this with the previous bound on $d((T\Psi)_j, (T\Phi)_j)$, we have

$$d_H((T\Psi)_j, (T\Phi)_j) \leq \tanh\left(\frac{\Delta_j}{4}\right) d_H(\Psi_{-j}, \Phi_{-j}).$$

It remains to bound the distance between the product measures. If $D(\Psi, \Phi) < \infty$, then $\Psi_k \sim \Phi_k$ for every k, and

$$\frac{d\Psi_{-j}}{d\Phi_{-j}}(p_{-j}) = \prod_{k\neq j} \frac{d\Psi_k}{d\Phi_k}(p_k).$$

Therefore,

$$\log \operatorname{ess\,sup} \frac{d\Psi_{-j}}{d\Phi_{-j}} \leq \sum_{k\neq j} \log \operatorname{ess\,sup} \frac{d\Psi_k}{d\Phi_k},$$

and similarly,

$$\log \operatorname{ess\,sup} \frac{d\Phi_{-j}}{d\Psi_{-j}} \leq \sum_{k\neq j} \log \operatorname{ess\,sup} \frac{d\Phi_k}{d\Psi_k}.$$

Hence,

$$d_H(\Psi_{-j}, \Phi_{-j}) \leq \sum_{k\neq j} d_H(\Psi_k, \Phi_k) \leq (J-1)D(\Psi, \Phi).$$

If $D(\Psi, \Phi) = \infty$, the desired inequality is trivial under the extended metric convention.

Therefore, for each j,

$$d_H((T\Psi)_j,(T\Phi)_j) \le (J-1)\tanh\left(\frac{\Delta_j}{4}\right) D(\Psi,\Phi).$$

Taking the maximum over j yields

$$D(T\Psi,T\Phi) \le (J-1)\max_{j\in\mathcal{J}}\tanh\left(\frac{\Delta_j}{4}\right) D(\Psi,\Phi).$$

Thus, if

$$\rho_B = (J-1)\max_{j\in\mathcal{J}}\tanh\left(\frac{\Delta_j}{4}\right) < 1,$$

then T is a contraction under D with modulus at most ρ_B. □

Proof of Corollary 2. By Theorem 1, it suffices to show that

$$(J-1)\max_{j\in\mathcal{J}}\tanh\left(\frac{\Delta_j}{4}\right) \le \frac{J-1}{4}\max_{j\in\mathcal{J}}(\overline{p}_j-\underline{p}_j)M_j.$$

We first bound Δ_j. Fix $j\in\mathcal{J}$. For any $p_j,p'_j\in[\underline{p}_j,\overline{p}_j]$ and any $\boldsymbol{p}_{-j},\boldsymbol{p}'_{-j}\in\prod_{k\neq j}[\underline{p}_k,\overline{p}_k]$, consider

$$\ln\frac{f_j(p_j,\boldsymbol{p}_{-j})f_j(p'_j,\boldsymbol{p}'_{-j})}{f_j(p'_j,\boldsymbol{p}_{-j})f_j(p_j,\boldsymbol{p}'_{-j})}.$$

Since interchanging p_j and p'_j changes the sign of this expression, we may assume without loss of generality that $p_j\ge p'_j$. By the fundamental theorem of calculus,

$$\begin{aligned}&\ln\frac{f_j(p_j,\boldsymbol{p}_{-j})f_j(p'_j,\boldsymbol{p}'_{-j})}{f_j(p'_j,\boldsymbol{p}_{-j})f_j(p_j,\boldsymbol{p}'_{-j})}\\&=\int_{p'_j}^{p_j}\left[\frac{\partial\ln f_j(s,\boldsymbol{p}_{-j})}{\partial s}-\frac{\partial\ln f_j(s,\boldsymbol{p}'_{-j})}{\partial s}\right]ds.\end{aligned}$$

Therefore,

$$\begin{aligned}\left|\ln\frac{f_j(p_j,\boldsymbol{p}_{-j})f_j(p'_j,\boldsymbol{p}'_{-j})}{f_j(p'_j,\boldsymbol{p}_{-j})f_j(p_j,\boldsymbol{p}'_{-j})}\right| &\le \int_{p'_j}^{p_j}\left|\frac{\partial\ln f_j(s,\boldsymbol{p}_{-j})}{\partial s}-\frac{\partial\ln f_j(s,\boldsymbol{p}'_{-j})}{\partial s}\right|ds\\&\le (p_j-p'_j)M_j\\&\le(\overline{p}_j-\underline{p}_j)M_j.\end{aligned}$$

Taking the supremum over $p_j, p'_j, \boldsymbol{p}_{-j}, \boldsymbol{p}'_{-j}$ gives

$$\Delta_j \leq (\overline{p}_j - \underline{p}_j) M_j.$$

Since $\tanh(x/4) \leq x/4$ for all $x \geq 0$, we have

$$\tanh\left(\frac{\Delta_j}{4}\right) \leq \frac{\Delta_j}{4} \leq \frac{(\overline{p}_j - \underline{p}_j) M_j}{4}.$$

Hence,

$$(J-1) \max_{j \in \mathcal{J}} \tanh\left(\frac{\Delta_j}{4}\right) \leq \frac{J-1}{4} \max_{j \in \mathcal{J}} (\overline{p}_j - \underline{p}_j) M_j.$$

Thus, if

$$\frac{J-1}{4} \max_{j \in \mathcal{J}} (\overline{p}_j - \underline{p}_j) M_j < 1,$$

then the contraction condition in Theorem 1 holds. Therefore, T is a contraction. □

A.2 Proof of Theorem 2

Proof of Theorem 2. By Theorem 1, it suffices to show that, under Assumption 1, the projective diameter for each j satisfies

$$\Delta_j = \ln f_j(\overline{\boldsymbol{p}}) - \ln f_j(\underline{p}_j, \overline{\boldsymbol{p}}_{-j}) - \ln f_j(\overline{p}_j, \underline{\boldsymbol{p}}_{-j}) + \ln f_j(\underline{\boldsymbol{p}}).$$

Fix $j \in \mathcal{J}$. For any $p_j, p'_j \in [\underline{p}_j, \overline{p}_j]$ and $\boldsymbol{p}_{-j}, \boldsymbol{p}'_{-j} \in \prod_{k \neq j} [\underline{p}_k, \overline{p}_k]$, consider

$$\ln \frac{f_j(p_j, \boldsymbol{p}_{-j}) f_j(p'_j, \boldsymbol{p}'_{-j})}{f_j(p'_j, \boldsymbol{p}_{-j}) f_j(p_j, \boldsymbol{p}'_{-j})}.$$

Since interchanging p_j and p'_j changes the sign of this expression, we may assume without loss of generality that $p_j \geq p'_j$. Then

$$\begin{aligned} &\ln \frac{f_j(p_j, \boldsymbol{p}_{-j}) f_j(p'_j, \boldsymbol{p}'_{-j})}{f_j(p'_j, \boldsymbol{p}_{-j}) f_j(p_j, \boldsymbol{p}'_{-j})} \\ &= \int_{p'_j}^{p_j} \left[\frac{\partial \ln f_j(s, \boldsymbol{p}_{-j})}{\partial s} - \frac{\partial \ln f_j(s, \boldsymbol{p}'_{-j})}{\partial s} \right] ds. \end{aligned}$$

By Assumption 1, $\partial \ln f_j(s, \boldsymbol{p}_{-j}) / \partial s$ is weakly increasing in each component of $\boldsymbol{p}_{-j}$.

Therefore, for every s,

$$\left|\frac{\partial \ln f_j(s,\boldsymbol{p}_{-j})}{\partial s} - \frac{\partial \ln f_j(s,\boldsymbol{p}'_{-j})}{\partial s}\right| \leq \frac{\partial \ln f_j(s,\overline{\boldsymbol{p}}_{-j})}{\partial s} - \frac{\partial \ln f_j(s,\underline{\boldsymbol{p}}_{-j})}{\partial s}.$$

It follows that

$$\begin{aligned}
&\left|\ln \frac{f_j(p_j,\boldsymbol{p}_{-j})f_j(p'_j,\boldsymbol{p}'_{-j})}{f_j(p'_j,\boldsymbol{p}_{-j})f_j(p_j,\boldsymbol{p}'_{-j})}\right| \\
&\leq \int_{\underline{p}_j}^{\overline{p}_j} \left[\frac{\partial \ln f_j(s,\overline{\boldsymbol{p}}_{-j})}{\partial s} - \frac{\partial \ln f_j(s,\underline{\boldsymbol{p}}_{-j})}{\partial s}\right] ds \\
&= \ln f_j(\overline{\boldsymbol{p}}) - \ln f_j(\underline{p}_j,\overline{\boldsymbol{p}}_{-j}) - \ln f_j(\overline{p}_j,\underline{\boldsymbol{p}}_{-j}) + \ln f_j(\underline{\boldsymbol{p}}).
\end{aligned}$$

The upper bound is attained by taking

$$p_j = \overline{p}_j, \qquad p'_j = \underline{p}_j, \qquad \boldsymbol{p}_{-j} = \overline{\boldsymbol{p}}_{-j}, \qquad \boldsymbol{p}'_{-j} = \underline{\boldsymbol{p}}_{-j}.$$

Hence,

$$\Delta_j = \ln f_j(\overline{\boldsymbol{p}}) - \ln f_j(\underline{p}_j,\overline{\boldsymbol{p}}_{-j}) - \ln f_j(\overline{p}_j,\underline{\boldsymbol{p}}_{-j}) + \ln f_j(\underline{\boldsymbol{p}}).$$

By Theorem 1,

$$D(T\Psi,T\Phi) \leq (J-1)\max_{j\in\mathcal{J}} \tanh\left(\frac{\Delta_j}{4}\right) D(\Psi,\Phi).$$

Substituting the expression for Δ_j derived above gives

$$D(T\Psi,T\Phi) \leq \rho^* D(\Psi,\Phi).$$

Therefore, if $\rho^* < 1$, the operator T is a contraction with modulus at most ρ^*.

□

A.3 Proof of Theorem 3

Proof of Proposition 1. Suppose $\rho < 1$. By Theorem 1, the operator T is a contraction. This implies that F is surjective, since for any $\tilde{G}$, we can take a $\Psi \in \prod_j \Delta([\underline{p}_j,\overline{p}_j])$,

$$F(\lim_{n\to\infty} T^n\Psi) = \tilde{G}.$$

Moreover, F is injective. Towards a contradiction, suppose F maps both $G_1 \neq G_2 \in \prod_j \Delta([\underline{p}_j, \overline{p}_j])$ to the same $\tilde{G}$. Then both G_1 and G_2 are fixed points for operator T, contradicting contraction.

The mapping F is continuous by Equation (1) and (2). Take two offered distributions G and G'. By Equation (2) and the definition of our metric,

$$\begin{aligned}
d(F(G)_j, F(G')_j) &= \ln \operatorname*{ess\,sup}_{p\in[\underline{p}_j,\overline{p}_j]} \left(\frac{dG_j}{dG'_j}(p)\frac{Pr_j(p;G)}{Pr_j(p;G')}\right) + \ln \operatorname*{ess\,sup}_{p\in[\underline{p}_j,\overline{p}_j]} \left(\frac{dG'_j}{dG_j}(p)\frac{Pr_j(p;G')}{Pr_j(p;G)}\right) \\
&\leq \ln \operatorname*{ess\,sup}_{p\in[\underline{p}_j,\overline{p}_j]} \frac{dG_j}{dG'_j}(p) + \ln \operatorname*{ess\,sup}_{p\in[\underline{p}_j,\overline{p}_j]} \frac{dG'_j}{dG_j}(p) \\
&+ \ln \sup_{p\in[\underline{p}_j,\overline{p}_j]} \left(\frac{Pr_j(p;G)}{Pr_j(p;G')}\right) + \ln \sup_{p\in[\underline{p}_j,\overline{p}_j]} \left(\frac{Pr_j(p;G')}{Pr_j(p;G)}\right) \\
&\leq D(G,G') + \rho D(G,G')
\end{aligned}$$

where the last inequality is by Equation (15). Consequently,

$$D(F(G), F(G')) \leq (1+\rho) D(G,G')$$

F is Lipschitz continuous with Lipschitz constant $1+\rho$.

Next, we show F^{-1} is Lipschitz continuous. Take two selected distributions $\tilde{G} \neq \tilde{G}' \in \prod_j \Delta([\underline{p}_j, \overline{p}_j])$ where $\tilde{G} = F(G)$. Let $T_{\tilde{G}}$ and $T_{\tilde{G}'}$ denote the corresponding operator T. Here we express dependence on the selected distribution. Note that

$$D(\tilde{G}, \tilde{G}') = D(T_{\tilde{G}}G, T_{\tilde{G}'}G) = D(G, T_{\tilde{G}'}G)$$

where the first equality is by the definition of the operator T and the metric D, while the second equality is by G being a fixed point of $T_{\tilde{G}}$. Observe that

$$D(T^k_{\tilde{G}'}G, T^{k+1}_{\tilde{G}'}G) \leq \rho^k D(G, T_{\tilde{G}'}G) = \rho^k D(\tilde{G}, \tilde{G}')$$

$$
\begin{aligned}
D(F^{-1}(\tilde{G}), F^{-1}(\tilde{G}')) = D(G, F^{-1}(\tilde{G}')) =& D(G, T^{\infty}_{\tilde{G}'}G) \\
\leq& \sum_{k=0}^{\infty} D(T^{k}_{\tilde{G}'}G, T^{k+1}_{\tilde{G}'}G) \\
\leq& \sum_{k=0}^{\infty} \rho^k D(\tilde{G}, \tilde{G}') \\
=& \frac{1}{1-\rho} D(\tilde{G}, \tilde{G}')
\end{aligned}
$$

where the first inequality is by triangular inequality. This proves that F^{-1} is Lipschitz continuous with Lipschitz constant $\frac{1}{1-\rho}$. □

We next prove the consistency result (Theorem 3). For proofs below, we shall suppress the dependence on variable x and x^*. The proof requires a combination of Lemma 2-3 below.

For the next lemma, we view $F^{-1}(\theta; \tilde{G})$ as a function of θ parametrized by $\tilde{G}$.

Lemma 2. *The function $F^{-1}(\theta; \tilde{G})$ is equicontinuous in θ, i.e., for all $\theta \in \Theta$, $\epsilon > 0$, there exists a $\delta > 0$ such that for all $|\theta' - \theta| < \delta$, $\tilde{G} \in \prod_j \Delta([\underline{p}_j, \overline{p}_j])$,*

$$
D(F^{-1}(\theta; \tilde{G}), F^{-1}(\theta'; \tilde{G})) \leq \epsilon.
$$

Proof of Lemma 2. Since the function f is continuous on a compact set $\prod_j [\underline{p}_j, \overline{p}_j] \times \Theta$ and $f_j > 0$, there exists $\underline{f} > 0$ such that for all $j \in \mathcal{J}$, $\theta \in \Theta$, $\boldsymbol{p} \in \prod_j [\underline{p}_j, \overline{p}_j]$,

$$
\underline{f} < f_j(\boldsymbol{p}; \theta).
$$

Consequently, for all $j \in \mathcal{J}$, $\theta \in \Theta$, $p_j \in [\underline{p}_j, \overline{p}_j]$, $G \in \prod_j \Delta([\underline{p}_j, \overline{p}_j])$,

$$
\underline{f} < Pr_j(p_j; G, \theta). \tag{16}
$$

Moreover, since the function f is continuous on a compact set $\prod_j [\underline{p}_j, \overline{p}_j] \times \Theta$, f is uniformly continuous. Thus, for any $\epsilon' > 0$, there exists a $\delta' > 0$ such that for all $j \in \mathcal{J}$, $\boldsymbol{p} \in \prod_j [\underline{p}_j, \overline{p}_j]$, $\theta, \theta' \in \Theta$ with $|\theta - \theta'| < \delta'$,

$$
|f_j(\boldsymbol{p}, \theta) - f_j(\boldsymbol{p}, \theta')| < \epsilon'.
$$

Therefore, for all $j \in \mathcal{J}$, $p_j \in [\underline{p}_j, \overline{p}_j]$, $G \in \prod_j \Delta([\underline{p}_j, \overline{p}_j])$, $\theta, \theta' \in \Theta$ with $|\theta - \theta'| < \delta'$,

$$\begin{aligned} &|Pr_j(p_j; G, \theta) - Pr_j(p_j; G, \theta')| \\ =&\left| \int_{\boldsymbol{p}_{-j}} [f_j(p_j, \boldsymbol{p}_{-j}; \theta) - f_j(p_j, \boldsymbol{p}_{-j}; \theta')] \prod_{k, k \neq j} dG_k(p_k) \right| < \epsilon'. \end{aligned} \tag{17}$$

Take an arbitrary $\tilde{G} \in \prod_j \Delta([\underline{p}_j, \overline{p}_j])$. Let $G_\theta = F^{-1}(\theta; \tilde{G})$. Let T_θ and $T_{\theta'}$ be the operator T associated with selected distribution $\tilde{G}$, when the parameter is θ and θ', respectively: for any $\Psi \in \prod_j \Delta([\underline{p}_j, \overline{p}_j])$,

$$(T_\theta \Psi)_j(p_j) = \frac{\int_{\underline{p}_j}^{p_j} d\tilde{G}_j(p) / Pr_j(p; \Psi, \theta)}{\int_{\underline{p}_j}^{\overline{p}_j} d\tilde{G}_j(p) / Pr_j(p; \Psi, \theta)}.$$

By the definition of metric D,

$$D(T_\theta G_\theta, T_{\theta'} G_\theta) \leq \max_j \left[\sup_p \ln \frac{Pr_j(p; G_\theta, \theta)}{Pr_j(p; G_\theta, \theta')} + \sup_p \ln \frac{Pr_j(p; G_\theta, \theta')}{Pr_j(p; G_\theta, \theta)} \right].$$

By Equation (16) and (17), for all $\tilde{G} \in \prod_j \Delta([\underline{p}_j, \overline{p}_j])$, $\theta, \theta' \in \Theta$ with $|\theta - \theta'| < \delta'$,

$$D(T_\theta G_\theta, T_{\theta'} G_\theta) \leq 2 \ln \frac{\underline{f} + \epsilon'}{\underline{f}},$$

$$\begin{aligned} D(F^{-1}(\theta; \tilde{G}), F^{-1}(\theta'; \tilde{G})) =& D(G_\theta, T_{\theta'}^\infty G_\theta) \\ \leq& \sum_{k=0}^\infty D(T_{\theta'}^k G_\theta, T_{\theta'}^{k+1} G_\theta) \\ \leq& \sum_{k=0}^\infty \bar{\rho}^k D(G_\theta, T_{\theta'} G_\theta) \\ =& \frac{1}{1 - \bar{\rho}} D(T_\theta G_\theta, T_{\theta'} G_\theta) \\ \leq& \frac{2}{1 - \bar{\rho}} \ln \frac{\underline{f} + \epsilon'}{\underline{f}}. \end{aligned}$$

Finally, for any $\epsilon > 0$, let ϵ' be such that $\frac{2}{1-\bar{\rho}} \ln \frac{\underline{f}+\epsilon'}{\underline{f}} = \epsilon$. The δ' corresponding to this ϵ' is the desired δ in the statement of the Lemma.

□

Lemma 3. *$\hat{Q}_n(\theta)$ converges uniformly in probability to $Q_0(\theta)$.*

Proof of Lemma 3. By standard consistency argument of MLE (Theorem 2.5 in (Newey and McFadden, 1994)) and identification result (Theorem 1) in Hu (2008), $\hat{h} \xrightarrow{p} h_0$.

Let $\omega_i = (x_i, y_i, p_i, z_{1i}, z_{2i})$. For a generic first-step object

$$h = \left(h_{p|x^*,x,y}, h_{x^*|x,y}\right),$$

define the individual criterion

$$\ell(\omega, \theta, h) = \sum_{x^* \in X^*} h_{x^*|x,y}(x^* \mid x, y) \log \mathrm{Prob}_y \left(x, x^*; \theta, h_{p|x^*,x,y}\right).$$

Then the sample criterion can be written as

$$\hat{Q}_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \ell(\omega_i, \theta, \hat{h}),$$

and the population criterion is

$$Q_0(\theta) = \mathbb{E}\left[\ell(\omega, \theta, h_0)\right],$$

where the expectation is taken with respect to the true distribution of a generic observation ω.

For any $\theta \in \Theta$, we decompose

$$\begin{aligned}
\hat{Q}_n(\theta) - Q_0(\theta) &= \frac{1}{n} \sum_{i=1}^{n} \ell(\omega_i, \theta, \hat{h}) - \mathbb{E}\left[\ell(\omega, \theta, h_0)\right] \\
&= \frac{1}{n} \sum_{i=1}^{n} \left\{\ell(\omega_i, \theta, \hat{h}) - \ell(\omega_i, \theta, h_0)\right\} \\
&\quad + \left[\frac{1}{n} \sum_{i=1}^{n} \ell(\omega_i, \theta, h_0) - \mathbb{E}\left[\ell(\omega, \theta, h_0)\right]\right].
\end{aligned}$$

Therefore,

$$\sup_{\theta\in\Theta}\left|\hat{Q}_n(\theta)-Q_0(\theta)\right| \le \sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{i=1}^{n}\left\{\ell(\omega_i,\theta,\hat{h})-\ell(\omega_i,\theta,h_0)\right\}\right| + \sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{i=1}^{n}\ell(\omega_i,\theta,h_0)-\mathbb{E}\left[\ell(\omega,\theta,h_0)\right]\right|.$$

We first show that the first term is $o_p(1)$. By Assumption 2, $f_j(p;x,x^*,\theta)$ is continuous in (p,θ) on a compact set and is strictly positive. Hence, uniformly over j, x, x^*, p, and $\theta\in\Theta$, there exists a constant $\underline{f}>0$ such that

$$f_j(p;x,x^*,\theta)\ge \underline{f}.$$

$$\text{Prob}_j\left(x,x^*;\theta,h_{p|x^*,x,y}\right)\ge \underline{f}.$$

Thus the logarithm in $\ell(\omega,\theta,h)$ is uniformly bounded.

Note that h enters ℓ in two ways. First it enters linearly through the mixing weight. Second it enters through $Prob$. Then, by the logarithm in $\ell(\omega,\theta,h)$ being uniformly bounded, the first part influence of h on ℓ is Lipschitz. By the logarithm in $\ell(\omega,\theta,h)$ is uniformly bounded and map F^{-1} being Lipschitz continuous, the second part influence of h on ℓ is Lipschitz. Thus, $\ell(\omega,\theta,h)$ is Lipschitz continuous in h, uniformly over $\theta\in\Theta$ and over ω in the support of the data. Since $\|\hat{h}-h_0\|=o_p(1)$, we have

$$\sup_{\theta\in\Theta}\sup_{\omega}\left|\ell(\omega,\theta,\hat{h})-\ell(\omega,\theta,h_0)\right|=o_p(1).$$

Consequently,

$$\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{i=1}^{n}\left\{\ell(\omega_i,\theta,\hat{h})-\ell(\omega_i,\theta,h_0)\right\}\right| \le \sup_{\theta\in\Theta}\sup_{\omega}\left|\ell(\omega,\theta,\hat{h})-\ell(\omega,\theta,h_0)\right| = o_p(1).$$

It remains to control the second term. For fixed h_0, consider the class of functions

$$\mathcal{L}=\left\{\ell(\omega,\theta,h_0):\theta\in\Theta\right\}.$$

By Lemma 2 and the continuity of f_j, the function $\ell(\omega,\theta,h_0)$ is continuous in θ. Since

Θ is compact and the log choice probability is uniformly bounded, $\mathcal{L}$ is a bounded and continuous class indexed by a compact parameter space. Because $\{\omega_i\}_{i=1}^n$ are i.i.d., a standard uniform law of large numbers implies

$$\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{i=1}^{n}\ell(\omega_i,\theta,h_0)-\mathbb{E}\left[\ell(\omega,\theta,h_0)\right]\right|\xrightarrow{p}0.$$

For example, this follows from Lemma 2.4 of Newey and McFadden (1994).

Combining the two bounds yields

$$\sup_{\theta\in\Theta}\left|\hat{Q}_n(\theta)-Q_0(\theta)\right|\xrightarrow{p}0.$$

□

Proof of Theorem 3. We are ready to apply Theorem 2.1 in Newey and McFadden (1994). (1). By the identification assumption 3, $Q_0(\theta)$ is uniquely maximized at θ_0. (2). Θ is compact. (3). As $Prob_j(\theta;\tilde{G})$ is also bounded below by $\underline{f}>0$ and continuous in θ by Lemma 2, $Q_0(\theta)$ is continuous. (4). $\hat{Q}_n(\theta)$ converges uniformly in probability to $Q_0(\theta)$, by Lemma 3. Thus, $\hat{\theta}$ is consistent.

To see $\hat{T}^\infty\Psi\xrightarrow{p}G$, note that $\hat{T}^\infty\Psi=F^{-1}(\hat{h}_{p|y},\hat{\theta})$,

$$D(\hat{T}^\infty\Psi,G)\le D(F^{-1}(\hat{h}_{p|y},\hat{\theta}),F^{-1}(\hat{h}_{p|y},\theta_0))+D(F^{-1}(\hat{h}_{p|y},\theta_0),G).$$

The first term

$$D(F^{-1}(\hat{h}_{p|y},\hat{\theta}),F^{-1}(\hat{h}_{p|y},\theta_0))\xrightarrow{p}0,\quad\text{as}\quad\hat{\theta}\xrightarrow{p}\theta_0$$

since F^{-1} is continuous in θ by Lemma 2. The second term

$$D(F^{-1}(\hat{h}_{p|y},\theta_0),G)\xrightarrow{p}0,\quad\text{as}\quad\hat{h}_{p|y}\xrightarrow{p}\tilde{G}$$

since F is a homeomorphism by Proposition 1.

□

A.4 Proof of Theorem 4

Proof of Theorem 4. Our GMM estimator is

$$\frac{1}{n}\sum_{i=1}^{n}\tilde{\mathfrak{g}}(\omega_i,\hat{\theta},\hat{h})=0.$$

For this GMM estimator, we can directly invoke Theorem 6.1 in Newey and McFadden (1994). Note that our $\mathfrak{g}$ is their g and our h is their γ in Newey and McFadden (1994).

By the proof of Theorem 3, $\hat{\theta}\xrightarrow{p}\theta_0$. By standard argument of MLE and identification result in Hu (2008), $\hat{h}\xrightarrow{p}h_0$. By Assumption 4, (θ_0,h_0) is in the interior of $\Theta\times H$. Next, we verify that $\tilde{\mathfrak{g}}(\omega,\theta,h)$ is continuously differentiable in a neighborhood $\mathscr{N}$ of (θ_0,h_0).

First, we verify that $\mathfrak{g}(\omega,\theta,h)$ is continuously differentiable in θ. It suffices to show that $Prob(\theta,h_{p|y})$ is twice continuously differentiable in θ. As f is twice continuously differentiable in θ, we only need to show that $F^{-1}(\tilde{G},\theta)$ is twice continuously differentiable in θ. By Equation (1), (2), $F(G,\theta)$ is infinitely continuously differentiable in G. By f being twice continuously differentiable in θ, $F(G,\theta)$ is twice continuously differentiable in θ. Moreover, matrix $\nabla_G F(G,\theta)$ is non-singular by Assumption 4. Thus, by the implicit function theorem,

$$\nabla_\theta F^{-1}(\tilde{G},\theta)=-\big[\nabla_G F(G,\theta)\big]^{-1}\nabla_\theta F(G,\theta)$$

and F^{-1} is twice continuously differentiable in θ.

Next, we verify that $\mathfrak{g}(\omega,\theta,h)$ is continuously differentiable in h. It suffices to show that $Prob(\theta,h_{p|y})$ is continuously differentiable in $h_{p|y}$, which follows if $F^{-1}(\tilde{G},\theta)$ is continuously differentiable in $\tilde{G}$. Since $F(G,\theta)$ is continuously differentiable in G, and $\nabla_G F(G,\theta)$ is nonsingular by Assumption 4, the implicit function theorem implies that $F^{-1}(\tilde{G},\theta)$ is continuously differentiable in $\tilde{G}$. Moreover, for $G=F^{-1}(\tilde{G},\theta)$,

$$\nabla_{\tilde{G}}F^{-1}(\tilde{G},\theta)=[\nabla_G F(G,\theta)]^{-1}.$$

Additionally, $\mathfrak{m}$ is infinitely continuously differentiable in all parameters θ,h. Consequently, we have shown that $\tilde{\mathfrak{g}}(\omega,\theta,h)$ is continuously differentiable in θ,h.

In addition,

$$\mathbb{E}[\tilde{\mathfrak{g}}(\omega,\theta_0,h_0)]=0$$

by the first-order condition of MLE and Q_0. Since for each observation ω, the value of Equation (11) is strictly positive, $||\mathfrak{m}(\omega, h_0)||$ is finite. Since $f_j \geq \underline{f} > 0$ is bounded from 0 and $Prob(\theta, h_{p|y})$ is continuously differentiable in θ, $||\mathfrak{g}(\omega, \theta_0, h_0)||$ is finite for each ω. Since there are only finite possible realizations of ω,

$$\mathbb{E}[||\tilde{\mathfrak{g}}(\omega, \theta_0, h_0)||^2] < \infty$$

By $\tilde{\mathfrak{g}}(\omega, \theta, h)$ being continuously differentiable in (θ, h) and a finite number of possible values of ω,

$$\mathbb{E}[\sup_{(\theta,h)\in\mathscr{N}} ||\nabla_{\theta,h}\tilde{\mathfrak{g}}(\omega, \theta, h)||] < \infty.$$

The last condition we need is that $\mathbb{E}\nabla_{\theta,h}\tilde{\mathfrak{g}}(\omega; \theta_0, h_0)$ is nonsingular, which is in Assumption 4.

We can write down the variance matrix V by Theorem 6.1 in Newey and McFadden (1994).

$$V = \big(\mathbb{E}\nabla_\theta \mathfrak{g}(\omega, \theta_0, h_0)\big)^{-1} \times \mathbb{E}(\mathscr{A}(\omega)\mathscr{A}(\omega)') \times \left(\big(\mathbb{E}\nabla_\theta \mathfrak{g}(\omega, \theta_0, h_0)\big)^{-1}\right)'$$

where

$$\mathscr{A}(\omega) = \mathfrak{g}(\omega, \theta_0, h_0) - \mathbb{E}[\nabla_h \mathfrak{g}(\omega, \theta_0, h_0)]\bigg(\mathbb{E}[\nabla_h \mathfrak{m}(\omega, h_0)]\bigg)^{-1} m(\omega, h_0).$$

To see the convergence rate of $\hat{T}^{\infty}_{\hat{\theta},\hat{h}}\Psi$, note that

$$D(\hat{T}^{\infty}\Psi, G) \leq D(F^{-1}(\hat{h}_{p|y}, \hat{\theta}), F^{-1}(\hat{h}_{p|y}, \theta_0)) + D(F^{-1}(\hat{h}_{p|y}, \theta_0), G).$$

By the proof above, F^{-1} is continuously differentiable in θ. Moreover, as Θ is compact, $F^{-1}(\tilde{G}, \theta)$ is Lipschitz continuous in θ. As $\hat{\theta} \xrightarrow{p} \theta_0$ at rate $\sqrt{n}$, the first term

$$D(F^{-1}(\hat{h}_{p|y}, \hat{\theta}), F^{-1}(\hat{h}_{p|y}, \theta_0))$$

is $O_p(n^{-1/2})$. By Theorem 6.1 of Newey and McFadden (1994), $\hat{h}_{p|y} \xrightarrow{p} \tilde{G}$ at rate $\sqrt{n}$. By Proposition 1, $F^{-1}(\tilde{G}, \theta)$ is Lipschitz continuous in $\tilde{G}$. Thus, the second term is $O_p(n^{-1/2})$.

At last, we show $\sqrt{n}$-asymptotic normality of $\hat{T}^{\infty}_{\hat{\theta},\hat{h}}\Psi = F^{-1}(\hat{h}_{p|y}, \hat{\theta})$. We follow the influence function approach in Newey and McFadden (1994). (See their Section 6.1, especially Equation 6.6, for a detailed treatment.) We introduce several useful definitions.

$$\mathfrak{M} = \mathbb{E}[\nabla_h \mathfrak{m}(\omega, h_0)],$$

$$\mathfrak{G}_h = \mathbb{E}[\nabla_h \mathfrak{g}(\omega; \theta_0, h_0)],$$

$$\mathfrak{G}_\theta = \mathbb{E}[\nabla_\theta \mathfrak{g}(\omega; \theta_0, h_0)].$$

Under the finite-support assumption, after parameterizing all probability mass functions by free coordinates, $(\tilde{G}, \theta) \mapsto F^{-1}(\tilde{G}, \theta)$ is a finite-dimensional continuously differentiable map in a neighborhood of $(\tilde{G}, \theta_0)$. Moreover,

$$\hat{h}_{p|y} - \tilde{G} = O_p(n^{-1/2}), \qquad \hat{\theta} - \theta_0 = O_p(n^{-1/2}).$$

Hence, by the delta method, the error term is

$$\begin{aligned}
&\sqrt{n}[F^{-1}(\hat{h}_{p|y}, \hat{\theta}) - F^{-1}(\tilde{G}, \theta_0)] \\
=&\sqrt{n}[\nabla_{\tilde{G}} F^{-1}(\tilde{G}, \theta_0)(\hat{h}_{p|y} - \tilde{G}) + \nabla_\theta F^{-1}(\tilde{G}, \theta_0)(\hat{\theta} - \theta_0)] + o_p(1) \\
=&\nabla_{\tilde{G}} F^{-1}(\tilde{G}, \theta_0)\sqrt{n}(\hat{h}_{p|y} - \tilde{G}) + \nabla_\theta F^{-1}(\tilde{G}, \theta_0)\sqrt{n}(\hat{\theta} - \theta_0) + o_p(1) \\
=&\nabla_{\tilde{G}} F^{-1}(\tilde{G}, \theta_0)\left[\sum_{i=1}^{n} \frac{-\mathfrak{M}^{-1}\mathfrak{m}(\omega_i, h_0)}{\sqrt{n}}\bigg|_{h_{p|y}} + o_p(1)\right] \\
&+ \nabla_\theta F^{-1}(\tilde{G}, \theta_0)\left[-\mathfrak{G}_\theta^{-1}\sum_{i=1}^{n} \frac{\mathfrak{g}(\omega_i; \theta_0, h_0) - \mathfrak{G}_h \mathfrak{M}^{-1}\mathfrak{m}(\omega_i, h_0)}{\sqrt{n}} + o_p(1)\right] + o_p(1) \\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Bigg\{-\nabla_{\tilde{G}} F^{-1}(\tilde{G}, \theta_0)(\mathfrak{M}^{-1})\bigg|_{h_{p|y}} \mathfrak{m}(\omega_i, h_0) \\
&-\nabla_\theta F^{-1}(\tilde{G}, \theta_0)\mathfrak{G}_\theta^{-1}[\mathfrak{g}(\omega_i; \theta_0, h_0) - \mathfrak{G}_h \mathfrak{M}^{-1}\mathfrak{m}(\omega_i, h_0)]\Bigg\} + o_p(1)
\end{aligned}$$

where $\Big|_{h_{p|y}}$ means we take the first $|h_{p|y}|$ rows (the two objects on which we apply $\Big|_{h_{p|y}}$ has $|h|$ rows where $|h|$ denotes the number of entries of vector h). The first equality is by F^{-1} being continuously differentiable in both $\tilde{G}$ and θ, $\sqrt{n}$-consistency of $\hat{h}$ and $\hat{\theta}$; the third equality is by plugging in the influence function of $\hat{h}$ and $\hat{\theta}$, given by Equation 6.6 of Newey and McFadden (1994). The conclusion follows by applying the central limit theorem to the term in the last equality.

□

B Additional Tables

Table 4: Simulation Results for Utility Parameters: Removing the Excluded Variable

	DGP 1					
	$N = 2000$			$N = 5000$		
	Bias	Std. Dev.	RMSE	Bias	Std. Dev.	RMSE
γ	-0.0658	0.1743	0.1861	-0.0358	0.1159	0.1212
κ	-0.0223	0.0527	0.0572	-0.0081	0.0351	0.0360
ξ_2	-0.0162	0.0458	0.0485	-0.0090	0.0309	0.0322
	DGP 2					
	$N = 2000$			$N = 5000$		
	Bias	Std. Dev.	RMSE	Bias	Std. Dev.	RMSE
γ	-0.0708	0.1636	0.1781	-0.0452	0.1091	0.1180
κ	-0.0192	0.0546	0.0579	-0.0070	0.0342	0.0349
ξ_2	-0.0130	0.0381	0.0402	-0.0083	0.0252	0.0265
	DGP 3					
	$N = 2000$			$N = 5000$		
	Bias	Std. Dev.	RMSE	Bias	Std. Dev.	RMSE
γ	-0.0623	0.2467	0.2542	-0.0262	0.1634	0.1653
κ	-0.0151	0.0516	0.0537	-0.0060	0.0345	0.0350
ξ_2	-0.0050	0.0313	0.0317	-0.0023	0.0201	0.0202
	DGP 4					
	$N = 2000$			$N = 5000$		
	Bias	Std. Dev.	RMSE	Bias	Std. Dev.	RMSE
γ	0.0453	0.8115	0.8119	0.0062	0.5276	0.5271
κ	-0.0138	0.0501	0.0519	-0.0068	0.0323	0.0330
ξ_2	-0.0001	0.0308	0.0308	-0.0003	0.0189	0.0189

Note: In these specifications, we remove the excluded variable x_{i1} from the selection function, so the parameter β in u_{i1} is not estimated.

Table 5: Simulation Results for CDF of $log(price)$: Removing the Excluded Variable

	DGP 1			
	$j=1$		$j=2$	
	IBias^2	IMSE	IBias^2	IMSE
$x_{i2}=0$	0.0004	0.0019	0.0001	0.0006
$x_{i2}=0.25$	0.0002	0.0016	0.0001	0.0007
$x_{i2}=0.5$	0.0001	0.0014	0.0001	0.0008
$x_{i2}=0.75$	0.0001	0.0013	0.0002	0.0009
$x_{i2}=1$	0.0001	0.0011	0.0001	0.0009
	DGP 2			
	$j=1$		$j=2$	
	IBias^2	IMSE	IBias^2	IMSE
$x_{i2}=0$	0.0003	0.0019	0.0001	0.0006
$x_{i2}=0.25$	0.0003	0.0018	0.0001	0.0007
$x_{i2}=0.5$	0.0002	0.0016	0.0001	0.0007
$x_{i2}=0.75$	0.0001	0.0015	0.0002	0.0008
$x_{i2}=1$	0.0001	0.0011	0.0001	0.0008
	DGP 3			
	$j=1$		$j=2$	
	IBias^2	IMSE	IBias^2	IMSE
$x_{i2}=0$	0.0007	0.0023	0.0029	0.0033
$x_{i2}=0.25$	0.0002	0.0015	0.0005	0.0010
$x_{i2}=0.5$	0.0001	0.0015	0.0002	0.0007
$x_{i2}=0.75$	0.0001	0.0015	0.0002	0.0008
$x_{i2}=1$	0.0001	0.0014	0.0001	0.0009
	DGP 4			
	$j=1$		$j=2$	
	IBias^2	IMSE	IBias^2	IMSE
$x_{i2}=0$	0.0014	0.0025	0.0011	0.0016
$x_{i2}=0.25$	0.0013	0.0023	0.0009	0.0014
$x_{i2}=0.5$	0.0012	0.0022	0.0005	0.0010
$x_{i2}=0.75$	0.0011	0.0023	0.0002	0.0008
$x_{i2}=1$	0.0005	0.0019	0.0001	0.0006

Note: In these specifications, we remove the excluded variable from the selection function. The IBias^2 of a function h is calculated as follows. Let $\hat{h}_r$ be the estimate of h from the r-th simulated dataset, and $\bar{h}(p) = \frac{1}{R}\sum_{r=1}^{R}\hat{h}_r(p)$ be the point-wise average over R simulations. The integrated squared bias is calculated by numerically integrating the point-wise squared bias $(\bar{h}(p) - h(p))^2$ over the distribution of p. The integrated MSE is computed in a similar way. The values reported in each row correspond to the price distributions conditional on a given value of x_{i2}. The results shown in this table are based on a 500 Monte Carlo replications with a sample size of 2,000. Corresponding results for a sample size of 5,000 are available upon request.

Table 6: Simulation Results for Utility Parameters: Misspecifying the Selection Function

	DGP 1					
	$N = 2000$			$N = 5000$		
	Bias	Std. Dev.	RMSE	Bias	Std. Dev.	RMSE
γ	-0.0851	0.1829	0.2016	-0.0540	0.1123	0.1245
β	0.0009	0.0634	0.0633	0.0052	0.0378	0.0381
κ	-0.0210	0.0499	0.0541	-0.0080	0.0356	0.0365
ξ_2	-0.0193	0.0596	0.0626	-0.0083	0.0364	0.0373
	DGP 2					
	$N = 2000$			$N = 5000$		
	Bias	Std. Dev.	RMSE	Bias	Std. Dev.	RMSE
γ	-0.0834	0.1739	0.1927	-0.0540	0.1103	0.1228
β	0.0028	0.0642	0.0642	0.0062	0.0387	0.0391
κ	-0.0167	0.0497	0.0523	-0.0054	0.0351	0.0355
ξ_2	-0.0116	0.0532	0.0544	-0.0034	0.0333	0.0334
	DGP 3					
	$N = 2000$			$N = 5000$		
	Bias	Std. Dev.	RMSE	Bias	Std. Dev.	RMSE
γ	-0.0970	0.2606	0.2779	-0.0569	0.1558	0.1657
β	0.0053	0.0652	0.0653	0.0079	0.0374	0.0382
κ	-0.0179	0.0493	0.0524	-0.0061	0.0336	0.0341
ξ_2	-0.0013	0.0474	0.0473	0.0030	0.0288	0.0289
	DGP 4					
	$N = 2000$			$N = 5000$		
	Bias	Std. Dev.	RMSE	Bias	Std. Dev.	RMSE
γ	0.1083	0.8175	0.8239	0.0813	0.4889	0.4951
β	0.0053	0.0624	0.0626	0.0086	0.0367	0.0376
κ	-0.0183	0.0480	0.0514	-0.0091	0.0340	0.0351
ξ_2	0.0047	0.0466	0.0468	0.0065	0.0289	0.0296

Note: In these specifications, we misspecify the selection model in estimation, assuming that the error term ε_i is drawn from $Logistic(0, 1)$. For the utility parameters, we rescale the estimates by the scale parameter of the logit model to make them comparable to those in the original probit specification.

Table 7: Simulation Results for CDF of $log(price)$: Misspecifying the Selection Function

	DGP 1			
	$j = 1$		$j = 2$	
	IBias^2	IMSE	IBias^2	IMSE
$x_{i2} = 0$	0.0004	0.0016	0.0002	0.0009
$x_{i2} = 0.25$	0.0004	0.0015	0.0002	0.0009
$x_{i2} = 0.5$	0.0002	0.0012	0.0002	0.0010
$x_{i2} = 0.75$	0.0002	0.0010	0.0002	0.0011
$x_{i2} = 1$	0.0001	0.0010	0.0002	0.0012
	DGP 2			
	$j = 1$		$j = 2$	
	IBias^2	IMSE	IBias^2	IMSE
$x_{i2} = 0$	0.0005	0.0016	0.0002	0.0008
$x_{i2} = 0.25$	0.0005	0.0017	0.0002	0.0008
$x_{i2} = 0.5$	0.0003	0.0013	0.0002	0.0009
$x_{i2} = 0.75$	0.0002	0.0011	0.0002	0.0010
$x_{i2} = 1$	0.0001	0.0010	0.0003	0.0011
	DGP 3			
	$j = 1$		$j = 2$	
	IBias^2	IMSE	IBias^2	IMSE
$x_{i2} = 0$	0.0007	0.0019	0.0031	0.0036
$x_{i2} = 0.25$	0.0002	0.0013	0.0005	0.0011
$x_{i2} = 0.5$	0.0002	0.0013	0.0003	0.0010
$x_{i2} = 0.75$	0.0002	0.0012	0.0002	0.0009
$x_{i2} = 1$	0.0002	0.0013	0.0002	0.0011
	DGP 4			
	$j = 1$		$j = 2$	
	IBias^2	IMSE	IBias^2	IMSE
$x_{i2} = 0$	0.0014	0.0023	0.0011	0.0016
$x_{i2} = 0.25$	0.0014	0.0024	0.0009	0.0015
$x_{i2} = 0.5$	0.0011	0.0021	0.0005	0.0011
$x_{i2} = 0.75$	0.0012	0.0021	0.0003	0.0009
$x_{i2} = 1$	0.0005	0.0018	0.0002	0.0007

Note: In these specifications, we misspecify the selection model, assuming that the error term ε_i is drawn from $Logistic(0, 1)$. The IBias^2 of a function h is calculated as follows. Let $\hat{h}_r$ be the estimate of h from the r-th simulated dataset, and $\bar{h}(p) = \frac{1}{R}\sum_{r=1}^{R} \hat{h}_r(p)$ be the point-wise average over R simulations. The integrated squared bias is calculated by numerically integrating the point-wise squared bias $(\bar{h}(p) - h(p))^2$ over the distribution of p. The integrated MSE is computed in a similar way. The values reported in each row correspond to the price distributions conditional on a given value of x_{i2}. The results shown in this table are based on a 500 Monte Carlo replications with a sample size of 2,000. Corresponding results for a sample size of 5,000 are available upon request.