

Estimating Nonseparable Selection Models: A Functional Contraction Approach*

Fan Wu Yi Xin

October 21, 2026

Abstract

We propose a novel method for estimating nonseparable selection models. We show that, for a given selection function and under suitable contraction conditions, 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 stochastic winning rules and incomplete bid data, and Roy-type 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, CUF, Peking University, Tsinghua University, UC Irvine, University of Michigan, USC, 2025 Asian Summer School in Econometrics and Statistics, Econometric Society World Congress 2025, DSE Conference 2025, and BSE Summer Forum 2026. 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 and under suitable contraction conditions, the potential outcome distributions are nonparametrically identified from the selected outcome distributions and can be recovered using a simple iterative algorithm. Consequently, the 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.

Formally, we consider a discrete choice problem in which each price-bearing 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. Our framework also permits 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 the price-bearing 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 under the contraction condition. Importantly, the contraction condition is automatically satisfied with at most two price-bearing alternatives, including a binary decision between one product and the outside option. 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 using the instrumental variable approach from the nonclassical measurement error literature (Hu, 2008; Hu and Schennach, 2008). 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.

Theorems 3 and 4 establish consistency of the proposed estimators and semiparametric asymptotic normality of the selection-parameter estimator. Corollary 2 gives the convergence rate of the recovered outcome CDFs and, under an additional first-step functional central limit theorem, their Gaussian process limit. 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 [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 (see, among many others, [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).

Our approach provides an alternative framework that differs from existing methods along several dimensions. 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. By contrast, many existing approaches that estimate conditional mean models restrict covariates to affecting only the location of the outcome distribution. Notable exceptions include Arellano and Bonhomme (2017), who develop a sample-selection correction for quantile regression by modeling the copula of the error terms in the outcome and selection equations, and Chernozhukov et al. (2025), who propose 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.

Third, our approach does not require the conventional exclusion restriction used in many sample selection models, namely a variable that shifts the choice probability without directly affecting outcomes. Such variables are central to identification in many two-step procedures and may not always be readily available in practice (see Vella (1998) and Honoré and Hu (2020) for further discussion). Furthermore, our method does not rely on identification-at-infinity arguments.¹

Instead, we assume that the potential outcomes are conditionally independent given the observed and unobserved characteristics, and our method relies on instruments satisfying the conditional independence requirements of the nonclassical measurement-error literature, as in Hu (2008) and Hu and Schennach (2008), to estimate the selected outcome distributions conditional on the unobservable in the first

¹Relatedly, D’Haultfoeulle and Maurel (2013) and D’Haultfoeulle 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. Lee and Lewbel (2013) study nonparametric identification in competing risks models without exclusion restrictions or identification-at-infinity (or near-zero) arguments. Honoré and Hu (2020) derive identified sets for sample selection models without exclusion restrictions.

step. These instruments serve a different role from the exclusion restrictions commonly used in sample selection models. Neither identifying strategy is uniformly weaker than the other; rather, they represent alternative sources of identifying variation that may be more or less plausible depending on the application.

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 the presence of price discrimination or personalized pricing.²

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 and competing risks models (see, e.g., [Athey and Haile, 2002](#); [Komarova, 2013](#); [Meilijson, 1981](#); [Guerre and Luo, 2019](#)).³ In these models, the selection rule is deterministic (e.g., 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](#)). Because selection is probabilistic rather than determined by an extremal outcome, classical order-statistic techniques are not applicable. Instead, our identification strategy relies on a fixed-point characterization of the selection process. 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 \(2007\)](#)).

²A 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.

³[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.

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

2 Model and Main Results

In this section, 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 includes unobserved components, although the presence of unobservables introduces additional challenges for estimation, which we address in Section 3. Because all results in this section are stated conditional on (x, x^*) , we omit these variables from the notation to simplify exposition.

Consider a discrete choice problem with complete choice set $\mathcal{J} = \{0, 1, \dots, J\}$, where 0 denotes the outside option and $1, \dots, J$ are the price-bearing alternatives. For $j = 1, \dots, J$, let

$$\mathcal{P}_j = [\underline{p}_j, \bar{p}_j], \quad \mathcal{P} = \prod_{j=1}^J \mathcal{P}_j,$$

where $\underline{p}_j < \bar{p}_j$ are fixed and finite. Let $G_j \in \Delta(\mathcal{P}_j)$ denote the offered price distribution of alternative j , where $\Delta(A)$ is the set of probability measures on A . Write $G = (G_1, \dots, G_J)$ for the collection of marginals and, when integrating over the price vector, identify it with the product measure $\bigotimes_{j=1}^J G_j$. Unless otherwise indicated, indices in products of price measures and in price-distribution operators run from 1 to J ; sums over the complete choice set also include 0.

We assume that the offered prices $p_j \sim G_j$ are independently distributed across alternatives, *conditional on x and x^** . This assumption allows prices across alternatives to be correlated conditional on observables. For example, if x^* captures a worker's unobserved productivity, wage offers from different firms may appear correlated when conditioning only on observable characteristics. However, the assumption imposes a more restrictive dependence structure: any residual correlation across prices must be mediated by x^* , rather than allowing prices to be arbitrarily correlated conditional on

observables. Nevertheless, x^* is not restricted to be one-dimensional and may instead represent a high-dimensional vector of latent characteristics. As the conditioning information becomes richer, more of the dependence across alternatives is captured by the observed and latent characteristics, leaving less residual dependence to be ruled out by the conditional independence assumption.

A *selection function* is $f = (f_0, f_1, \dots, f_J)$, where $f_j(\mathbf{p})$ is the probability of choosing $j \in \mathcal{J}$ given the vector of inside prices $\mathbf{p} = (p_1, \dots, p_J)$. For $j = 1, \dots, J$, assume that f_j is continuous and strictly positive on $\mathcal{P}$. The outside probability is

$$f_0(\mathbf{p}) = 1 - \sum_{j=1}^J f_j(\mathbf{p}) \geq 0, \quad \sum_{j=0}^J f_j(\mathbf{p}) = 1.$$

Thus an outside choice is allowed without introducing an additional price coordinate. The contraction and price-identification results below concern only $j = 1, \dots, J$ and require strict positivity only for these alternatives. The case without an outside option is obtained by setting $f_0 \equiv 0$ and omitting alternative 0 from the choice criterion.

The strict positivity assumption ensures that every price realization has a positive probability of generating an observed choice and therefore can, in principle, be observed in the data. To see why this is important, suppose instead that there exists an interval $A \subset [\underline{p}_j, \bar{p}_j]$ such that whenever $p_j \in A$, the probability of choosing alternative j is zero. Then prices in A are effectively censored from the data; that is, the data contain no information about the frequency with which such prices are offered. As a result, G_j cannot be identified on A without imposing additional restrictions.⁴

The selection function f is taken as a primitive of our framework. Nevertheless, it can be microfounded through a standard utility-maximization problem. Under particular assumptions on the distribution of utility shocks, familiar discrete choice models such as the multinomial logit and probit arise as special cases. We also want to emphasize that the selection function f can freely depend on observable and unobservable characteristics (x, x^*) . This flexibility allows the model to capture

⁴The strict positivity assumption 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. The strict positivity assumption excludes deterministic selection rules, such as standard auction, competing-risks, and pure Roy models, in which some alternatives are selected with probability zero for certain realizations of $\mathbf{p}$. Identification in these settings typically relies on arguments based on ordered statistics, which are well established in the literature. See the discussion in the handbook chapter by [Athey and Haile \(2007\)](#).

dependence between the potential outcome distribution and the selection mechanism. For example, unobserved productivity may simultaneously influence wage offers and labor-force participation decisions, thereby generating endogenous selection.

The *offered* price distribution (or *potential* outcome distribution) G_j is generally not directly observed in the data, yet it is often the primary object of interest. Instead, data typically reveal only the *selected* (or *accepted*) outcome distribution. For example, researchers may observe the distribution of transacted prices or accepted wage offers, but not the distribution of all offered prices or wage offers. Let $G_j^s \in \Delta([p_j, \bar{p}_j])$ represent the price distribution of alternative j conditional on j being selected. Similarly, $G^s = \prod_{j=1}^J G_j^s$ denotes the collection of selected outcome distributions across the price-bearing alternatives.

In the remainder of this section, we establish identification of the potential outcome distributions G from the selected outcome distributions G^s , holding the selection function f fixed. We then turn in Section 3 to the empirically relevant case in which f is specified up to a finite-dimensional parameter vector that is estimated from observed choice behavior.

2.1 A Fixed-Point Characterization of Potential Outcome Distributions

We begin with a density-based formulation that highlights the main ideas most transparently. Throughout Section 2.1, we assume that all relevant distributions admit densities with full support and use lower-case letters to denote densities corresponding to probability measures denoted by upper-case letters. In particular, let g_j and g_j^s denote the densities of the probability measures G_j and G_j^s , respectively. Similarly, let $g = \prod_{j=1}^J g_j$ and $g^s = \prod_{j=1}^J g_j^s$ denote the collections of potential and selected outcome densities across the price-bearing alternatives, respectively. A measure-theoretic formulation accommodating general probability measures, including discrete distributions, continuous distributions, and arbitrary mixtures thereof, is provided in Section 2.2.

The density of selected outcomes g_j^s can be derived from the density of potential outcomes g_j using Bayes' rule:

$$g_j^s(p_j) \propto \Pr_j(p_j; g) g_j(p_j), \text{ for } j = 1, \dots, J. \quad (1)$$

In Equation (1), $\Pr_j(p_j; g)$ represents the probability of selecting alternative j conditional on p_j and is given by

$$\Pr_j(p_j; g) = \int_{\mathbf{p}_{-j}} f_j(p_j, \mathbf{p}_{-j}) \prod_{k \neq j} g_k(p_k) d\mathbf{p}_{-j}, \quad (2)$$

where $\mathbf{p}_{-j} = (p_1, \dots, p_{j-1}, p_{j+1}, \dots, p_J)$ denotes the vector of prices excluding j 's price. Because we assume that prices are independently distributed across alternatives conditional on (x, x^*) , the joint distribution of $\mathbf{p}_{-j}$ factors into the product of the marginal densities, $\prod_{k \neq j} g_k(p_k)$.

Collecting Equations (1) and (2) across the price-bearing alternatives $j = 1, \dots, J$ defines a forward mapping from the potential outcome distributions to the selected outcome distributions, where the selection mechanism is governed by the selection function f through the conditional choice probabilities. Our objective, however, is the inverse: to recover the unobserved potential outcome distributions g from the observed selected outcome distributions g^s .

Whether such recovery is possible depends critically on the information contained in the observed data. In this regard, the conditional independence assumption plays a key role in identification. Because the data reveal only the price of the selected alternative, they contain no direct information about the dependence structure of prices across alternatives. As a result, any correlation among competing alternatives is generally unidentifiable.⁵ For example, if prices tend to move together across alternatives, this dependence would not be detectable from the observed data. The conditional independence assumption circumvents this problem by restricting the joint distribution of prices to be determined by the marginal distributions. As a result, for a given selection function, the system defined by Equations (1) and (2) contains as many unknown objects as identifying restrictions: both g and g^s consist of J densities. Even so, identification remains challenging because it requires recovering a collection of infinite-dimensional objects from a nonlinear system of equations.

To address this challenge, we now derive an inversion of the forward mapping defined by Equations (1) and (2). In particular, dividing both sides of Equation (1)

⁵See Heckman and Honore (1990) for an early discussion of this nonidentification result; see also the survey chapter by French and Taber (2011) for a related discussion.

by the selection probability $\Pr_j(p_j; g)$ yields

$$g_j(p_j) \propto g_j^s(p_j) / \Pr_j(p_j; g), \text{ for } j = 1, \dots, J. \quad (3)$$

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, $\Pr_j(p_j; g)$, 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.

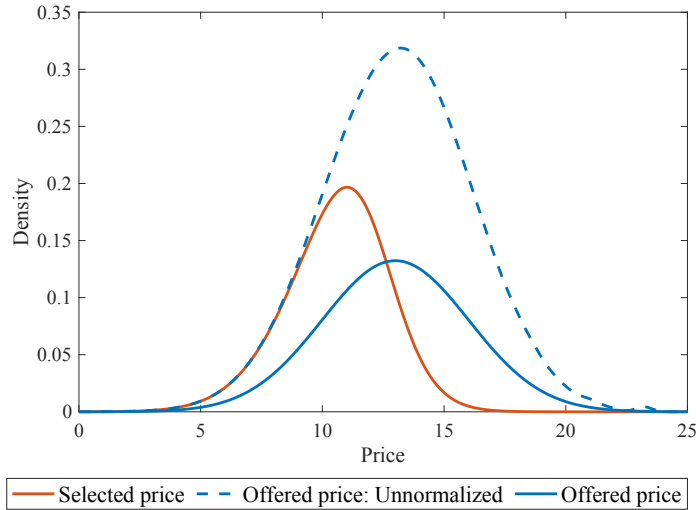


Figure 1: Densities of offered and selected prices. We draw offered prices from $\mathcal{N}(13, 9)$, and the probability that the agent given price p chooses this alternative is given by $\exp(10 - p) / (0.1 + \exp(10 - p))$.

Equation (3) plays a central role in our analysis. 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 (3) would be straightforward. 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 (3) 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 $T: \prod_j \Delta(\underline{p}_j, \bar{p}_j) \rightarrow \prod_j \Delta(\underline{p}_j, \bar{p}_j)$ as follows.⁶

$$(T\psi)_j(p_j) = \frac{g_j^s(p_j) / \Pr_j(p_j; \psi)}{\int_{\underline{p}_j}^{\bar{p}_j} g_j^s(p) / \Pr_j(p; \psi) dp}, \text{ for } j = 1, \dots, J. \quad (4)$$

where $\psi = (\psi_1, \psi_2, \dots, \psi_J) \in \prod_j \Delta(\underline{p}_j, \bar{p}_j)$. The denominator is the normalizing constant that ensures $(T\psi)_j$ integrates to one.

Importantly, if the conjecture ψ is correct, i.e., $\psi = g$, then the selection probability $\Pr_j(\cdot; \psi)$ is correct, 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 $\psi, \phi \in \prod_j \Delta(\underline{p}_j, \bar{p}_j)$,

$$D(T\psi, T\phi) \leq \rho D(\psi, \phi), \quad (5)$$

whenever $D(\psi, \phi) < \infty$, for an extended metric D . The fixed-point argument is applied on a complete finite-distance component; after one update every initial conjecture lies in the component determined by the selected distributions. We now introduce a suitable metric.

Hilbert's Projective Metric For two densities $\psi_j, \phi_j \in \Delta(\underline{p}_j, \bar{p}_j)$, we measure their distance using Hilbert's projective metric. If ψ_j and ϕ_j are strictly positive on the same support,

$$d_H(\psi_j, \phi_j) = \ln \left(\frac{\sup_{p \in [\underline{p}_j, \bar{p}_j]} \frac{\psi_j(p)}{\phi_j(p)}}{\inf_{p \in [\underline{p}_j, \bar{p}_j]} \frac{\psi_j(p)}{\phi_j(p)}} \right), \quad (6)$$

otherwise set $d_H(\psi_j, \phi_j) = +\infty$.

The choice of this metric is deliberate and closely tailored to the structure of the operator T . In particular, the usefulness of Hilbert's projective metric in our setting comes from its invariance to positive rescaling. Since the operator T updates each density by reweighting g_j^s by the reciprocal of selection probability $\Pr_j(\cdot; \psi)$ and then renormalizing, both the common factor g_j^s and the normalizing constants cancel out

⁶Here, densities are identified with their induced probability measures when specifying the domain of T .

when computing the projective distance. Therefore, for any two conjectured densities ψ and ϕ ,

$$\begin{aligned}
d_H((T\psi)_j, (T\phi)_j) &= \ln \left(\frac{\sup_{p \in [\underline{p}_j, \bar{p}_j]} \frac{(T\psi)_j(p)}{(T\phi)_j(p)}}{\inf_{p \in [\underline{p}_j, \bar{p}_j]} \frac{(T\psi)_j(p)}{(T\phi)_j(p)}} \right) \\
&= \ln \left(\frac{\sup_{p \in [\underline{p}_j, \bar{p}_j]} \frac{g_j^s(p) / \Pr_j(p; \psi)}{\text{norm. const. for } \psi} \frac{\text{norm. const. for } \phi}{g_j^s(p) / \Pr_j(p; \phi)}}{\inf_{p \in [\underline{p}_j, \bar{p}_j]} \frac{g_j^s(p) / \Pr_j(p; \psi)}{\text{norm. const. for } \psi} \frac{\text{norm. const. for } \phi}{g_j^s(p) / \Pr_j(p; \phi)}} \right) \\
&= \ln \left(\frac{\sup_{p \in [\underline{p}_j, \bar{p}_j]} \frac{\Pr_j(p; \phi)}{\Pr_j(p; \psi)}}{\inf_{p \in [\underline{p}_j, \bar{p}_j]} \frac{\Pr_j(p; \phi)}{\Pr_j(p; \psi)}} \right) = d_H(\Pr_j(\cdot; \phi), \Pr_j(\cdot; \psi)).
\end{aligned}$$

This equation reveals an important isometry: measuring the distance between $(T\psi)_j$ and $(T\phi)_j$ is equivalent to measuring the distance between the corresponding selection probabilities under Hilbert's projective metric. Consequently, the contraction properties of the operator T can be characterized by studying how perturbations in the conjectured densities affect the implied choice probabilities.

2.2 General Probability Measure Formulation

The density formulation in Section 2.1 is useful for illustrating the key ideas but is not essential for the results. We now present the corresponding measure-theoretic formulation for arbitrary probability measures and adopt this framework for the rest of the paper.

As before, the selected outcome distribution G_j^s can be derived from the potential outcome distribution G_j via Bayes' rule. For any Borel set $A \subseteq [\underline{p}_j, \bar{p}_j]$, the counterpart of Equation (1) for general probability measures is:

$$G_j^s(A) = \frac{\int_A \Pr_j(p_j; G) dG_j(p_j)}{\int_{\underline{p}_j}^{\bar{p}_j} \Pr_j(p_j; G) dG_j(p_j)}, \text{ for } j = 1, \dots, J, \quad (7)$$

where

$$\Pr_j(p_j; G) = \int_{\mathbf{p}_{-j}} f_j(p_j, \mathbf{p}_{-j}) \prod_{k \neq j} dG_k(p_k). \quad (8)$$

Equation (7) defines a mapping $F: \prod_j \Delta([\underline{p}_j, \bar{p}_j]) \rightarrow \prod_j \Delta([\underline{p}_j, \bar{p}_j])$ that maps the potential outcome distribution G to the selected outcome distribution G^s .

The inversion argument developed in the density-based framework extends directly to the present setting. In particular, the counterpart of Equation (3) is

$$G_j(A) = \frac{\int_A dG_j^s(p_j) / \Pr_j(p_j; G)}{\int_{\underline{p}_j}^{\bar{p}_j} dG_j^s(p_j) / \Pr_j(p_j; G)}, \text{ for } j = 1, \dots, J. \quad (9)$$

Equation (9) defines an operator $T: \prod_j \Delta([\underline{p}_j, \bar{p}_j]) \rightarrow \prod_j \Delta([\underline{p}_j, \bar{p}_j])$ as follows.

$$(T\Psi)_j(A) = \frac{\int_A dG_j^s(p_j) / \Pr_j(p_j; \Psi)}{\int_{\underline{p}_j}^{\bar{p}_j} dG_j^s(p_j) / \Pr_j(p_j; \Psi)}, \text{ for } j = 1, \dots, J, \quad (10)$$

where $\Psi = (\Psi_1, \Psi_2, \dots, \Psi_J) \in \prod_j \Delta([\underline{p}_j, \bar{p}_j])$.

For two probability measures $\Psi_j, \Phi_j \in \Delta([\underline{p}_j, \bar{p}_j])$, let $d_H(\Psi_j, \Phi_j)$ denote their distance under the general form of Hilbert's projective metric; see Appendix A.1 for details. To account for the J price-bearing alternatives, we define a metric in the space $\prod_j \Delta([\underline{p}_j, \bar{p}_j])$ by taking the maximum distance among all alternatives:

$$D(\Psi, \Phi) = \max_{1 \leq j \leq J} d_H(\Psi_j, \Phi_j)$$

for any $\Psi, \Phi \in \prod_{j=1}^J \Delta(\mathcal{P}_j)$. Henceforth, the product of the J inside probability spaces is equipped with this extended metric D .

2.3 Functional Contraction

We now provide a primitive sufficient condition under which the Bayes-inversion operator T is a contraction. We first define the *relative selection probability variation*:

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

Intuitively, Δ_j measures how much the relative selection probability of alternative j between two own-price realizations, p_j and p'_j , can vary as competitors' prices change from $\mathbf{p}_{-j}$ to $\mathbf{p}'_{-j}$.⁷

⁷In Perron-Frobenius theory, Δ_j is also known as the projective diameter. Geometrically, Δ_j measures the departure of f_j from multiplicative separability between p_j and $\mathbf{p}_{-j}$. If $f_j(p_j, \mathbf{p}_{-j}) = a(p_j)b(\mathbf{p}_{-j})$, then $\Delta_j = 0$.

Let

$$\rho = (J - 1) \max_{1 \leq j \leq J} \tanh \left(\frac{\Delta_j}{4} \right),$$

where $\tanh(\cdot)$ denotes the hyperbolic tangent function. Since $\Delta_j \geq 0$ is finite under the maintained assumption that f_j is continuous and strictly positive on compact support, it follows that $\tanh(\Delta_j/4) \in [0, 1)$.

Theorem 1 (Contraction). *If $\rho < 1$, the operator T is a contraction with modulus at most ρ . In particular, the condition holds when $J \leq 2$, whether or not an outside option is present.*

Proof. See Appendix A.2. □

Theorem 1 establishes a key identification result for selection models. Whenever $\rho < 1$ (which is automatically satisfied when $J \leq 2$), the Bayes-inversion operator T admits a unique fixed point. Since the offered price distribution G is characterized as the fixed point of T , it follows that G is uniquely determined by G^s given the selection function f . Thus G is identified given 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.⁸ We formally state the identification results in the following corollary.

Corollary 1 (Identification). *Suppose $\rho < 1$. Then, for every selected price distribution G^s , there exists a unique offered price distribution G consistent with G^s given the selection function f . Moreover, for any initial distribution $\Psi \in \prod_j \Delta([p_j, \bar{p}_j])$, $G = \lim_{n \rightarrow \infty} T^n \Psi$. Hence, the offered price distribution G is nonparametrically identified from G^s given f .*

Corollary 1 provides a *constructive* identification result for the offered price distribution G . Take any initial distribution $\Psi \in \prod_j \Delta([p_j, \bar{p}_j])$, by Theorem 1,

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

⁸Theorem 1 applies more generally to any selection function f that is strictly positive and has finite projective diameter. Thus, the continuity and compact-support assumptions on the selection function f are not essential to this theorem.

where $D(T\Psi, G)$ is finite.⁹ This implies

$$\begin{aligned}\lim_{n \rightarrow \infty} D(T^n \Psi, G) &= 0, \\ \lim_{n \rightarrow \infty} T^n \Psi &= G.\end{aligned}$$

Thus, starting from any initial conjecture, repeated application of T converges geometrically to the potential outcome distributions associated with the selection function.

Interpreting the Contraction Condition. We now provide some economic intuition for the factors that determine the strength of the contraction in Theorem 1.

First, the factor $J - 1$ counts the other unknown price marginals entering the selection probability for a price-bearing alternative. Adding the outside option does not add a price marginal and therefore does not change this factor. With two price-bearing alternatives ($J = 2$), the contraction bound is automatically less than one, even if the outside option is also available. With one price-bearing alternative and the outside option ($J = 1$), there is no other price distribution to integrate out: T is independent of the conjecture and recovery takes a single Bayes update. This covers binary purchase and labor-force participation decisions.

Another key primitive underlying the contraction condition is the relative selection probability variation Δ_j . This quantity can be related to a familiar semi-elasticity measure. Specifically, if $\ln f_j$ is continuously differentiable in p_j , define

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

By the mean value theorem,

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

The quantity M_j measures the maximum variation in the own-price semi-elasticity of alternative j across different configurations of competitors' prices. It is small when the own-price responsiveness of alternative j remains stable as competitors' prices vary. This is likely to occur when alternative j is highly differentiated or only weakly

⁹Appendix A.1 shows that $D(T\Psi, G)$ is finite for any initial distribution Ψ under the maintained assumption that f_j is continuous and strictly positive on compact support.

substitutable with other alternatives, so that changes in competitors' prices have little effect on consumers' sensitivity to p_j . In the extreme case where the own-price semi-elasticity of alternative j is independent of competitors' prices, $M_j = 0$, which in turn implies $\Delta_j = 0$.

In addition to the strength of interaction effects, as captured by the selection function f_j , the relative selection probability variation also depends on the range over which prices vary. Holding the selection function fixed, Δ_j is weakly increasing as the price support expands. Consequently, even when competitors' prices affect the responsiveness of demand to product j 's own price, the resulting Δ_j may remain small if prices vary over a sufficiently narrow range. Conversely, the same degree of interaction can generate a larger Δ_j when prices span a wider interval.

Finally, note that the contraction condition in Theorem 1 is sufficient rather than necessary. The operator T may still be a contraction even when the upper bound exceeds one. Moreover, the contraction modulus established in the theorem is a global bound. Consequently, even when the operator is not globally contractive, it may still be locally contractive in a neighborhood of the fixed point.

This distinction is particularly important for empirical applications. While the global bound is driven by worst-case conjectures in Equation (5), estimation in practice is primarily concerned with regions near the economically relevant distributions implied by the observed data. Moreover, numerical inversion procedures based on fixed-point iterations are typically initialized from economically plausible distributions. As a result, local contraction may be sufficient to ensure stable recovery of the offered price distributions, even in settings where the global sufficient condition cannot be verified.

2.4 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 relative selection probability variation. 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 $j = 1, \dots, J$ and $p_j \in [\underline{p}_j, \bar{p}_j]$, $\frac{\partial \ln f_j(p_j, \mathbf{p}_{-j})}{\partial p_j}$ is weakly increasing in each p_k with $k \neq j$.

Under log supermodularity, the relative selection probability variation in Theorem 1 is attained at the boundary. The result is stated below. Let

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

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.3. □

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. 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.

2.5 Identification without Contraction

The preceding results establish identification by proving that the Bayes-inversion operator is a contraction. Identification itself, however, requires only that the forward mapping F be injective. A restriction on the structure of the selection function can deliver this uniqueness without imposing the contraction condition.

Proposition 1 (Identification under a common kernel). *Suppose the given selection function admits the representation*

$$f_j(\mathbf{p}) = a_j(p_j)K(\mathbf{p}), \quad j = 1, \dots, J, \quad (11)$$

where $a_j : \mathcal{P}_j \rightarrow (0, \infty)$ and $K : \mathcal{P} \rightarrow (0, \infty)$ are continuous. The forward mapping F defined by (7) is globally injective for every finite J . Thus the offered price distributions are nonparametrically identified from the selected price distributions G^s , given f . In particular, the result applies to the multinomial-logit rule

$$f_j(\mathbf{p}) = \frac{\exp(v_j(p_j))}{a_0 + \sum_{k=1}^J \exp(v_k(p_k))}, \quad j = 1, \dots, J, \quad (12)$$

where each $v_j : \mathcal{P}_j \rightarrow \mathbb{R}$ is continuous and $a_0 \geq 0$ is a given outside-option weight.¹⁰

Proof. See Appendix A.4. □

The common-kernel restriction concerns the relationship between selection probabilities across alternatives. In particular,

$$\frac{f_j(\mathbf{p})}{f_k(\mathbf{p})} = \frac{a_j(p_j)}{a_k(p_k)}, \quad j, k \in \{1, \dots, J\}.$$

It does not require f_j to be multiplicatively separable between p_j and $\mathbf{p}_{-j}$: the common function K may depend nonseparably on the entire price vector. Proposition 1 establishes identification under multinomial-logit selection rule for any finite number of alternatives. This implication concerns the displayed conditional logit probabilities. Further averaging over unobserved heterogeneity need not preserve the common-kernel representation; a mixed-logit specification must therefore be checked separately [under what conditions do a mixed logit model satisfy this?].

[Here we probably say a little bit more that Proposition 1 does not impose the assumption that $\rho < 1$, but impose a different type of assumption on f , which the most commonly used logit model can satisfy. For logit model, Proposition 1 tells us about the uniqueness, but does not guarantee contraction, Theorem 2 guarantees contraction under stronger assumption on ρ .] Proposition 1 separates identification from the convergence of a particular inversion algorithm. It establishes uniqueness of the offered distributions for model-generated selected distributions, but does not by itself establish convergence of the iterates $T^m \Psi$. The contraction conditions in Theorems 1 and 2 provide both uniqueness and geometric convergence for a different, more generally specified class of selection functions. The estimation and asymptotic results in Section 3 continue to use their stated contraction and regularity conditions.

¹⁰Taking

$$a_j(p_j) = \exp(v_j(p_j)), \quad K(\mathbf{p}) = \left(a_0 + \sum_{k=1}^J \exp(v_k(p_k)) \right)^{-1}$$

verifies (11).

3 Estimation

We now turn to estimation of the offered price distributions G and the selection function f . For each individual, the sampling frame records the choice $j_i \in \mathcal{J} = \{0, 1, \dots, J\}$ and observable characteristics $x_i \in X$; the selected price p_i is observed only when $j_i \geq 1$. For a generic observation ω , j denotes its realized choice; a dummy alternative index is written as k when needed to distinguish the two roles. Let x_{ij} be the observed characteristics of alternative j , and write $x_i = (x'_{i1}, \dots, x'_{iJ})'$. Let $x_i^* \in X^*$ be an unobservable characteristic that may affect both the selection decision and potential outcome distributions. Note that the unobserved characteristic need not be one-dimensional. More generally, x^* may represent a vector of latent characteristics. In this case, X^* denotes the corresponding product space of latent attributes.

Our theoretical results are developed under a highly flexible specification of the selection mechanism f . In principle, the framework could be combined with parametric, semiparametric, or nonparametric approaches to estimating f . In practice, however, the selection mechanism may depend on the entire vector of prices and product characteristics, as well as consumer heterogeneity, rendering fully nonparametric estimation challenging. We therefore adopt the standard differentiated-products demand framework as in [Berry \(1994\)](#) and [Berry et al. \(1995\)](#), which provides a tractable and widely used parameterization of the selection probabilities. This parametric specification allows us to propose a simple estimation procedure while substantially reducing the data requirements needed to recover f , a feature that is particularly valuable when observed choice variation is limited.

In particular, 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 $\varepsilon_i = (\varepsilon_{i0}, \dots, \varepsilon_{iJ})$ follows a known joint distribution, such as Type 1 extreme value. The outside utility is $u_{i0} = v_0(x_i, x_i^*, \varepsilon_{i0}; \theta)$, with no price argument. Each individual chooses the alternative in $\mathcal{J}$ 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:

$$u_{ij} = \gamma p_{ij} + x'_{ij}\beta + \xi_j + x_i^* \kappa_j + \varepsilon_{ij}, \quad j = 1, 2, \dots, J, \quad (13)$$

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 $\boldsymbol{\xi} = (\xi_1, \dots, \xi_J)$ and $\boldsymbol{\kappa} = (\kappa_1, \dots, \kappa_J)$.

3.1 Two-Step Estimation Strategy

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

Step 1: Estimating Selected Outcome Distribution We first estimate the selected outcome distributions G^s conditional on (x, x^*) . When all relevant covariates are observed, these distributions can be estimated directly, for example by empirical or conditional distribution estimators. When x^* is unobserved, additional information is required.

We follow the proxy-variable approach of [Hu \(2008\)](#) and its continuous-variable counterpart in [Hu and Schennach \(2008\)](#). Let z_1 and z_2 be two instruments, and suppose that $\omega_i = (x_i, j_i, p_i, z_{1i}, z_{2i})$, $i = 1, \dots, n$, are i.i.d. When additional proxies are used, they are included in ω_i as well. All first-step influence representations refer to this complete observed record. Prices take values in the compact intervals $\mathcal{P}_j$ introduced in Section 2. We do not impose finite support on the covariates or instruments. Write H for a conditional probability measure and h for a conditional density, or a probability mass function for a discrete coordinate. Here p_i is recorded

as missing for $j_i = 0$ and is not used in any outside-choice contribution below. For inside choices $j = 1, \dots, J$, the proxy restriction is

$$\begin{aligned} & h_{p, z_1 | z_2, x, j}(p, z_1 \mid z_2, x, j) \\ &= \int_{X^*} h_{p | x^*, x, j}(p \mid x^*, x, j) h_{z_1 | x^*, x, j}(z_1 \mid x^*, x, j) dH_{x^* | z_2, x, j}(x^* \mid z_2, x, j). \end{aligned} \quad (14)$$

The integral becomes a sum when the latent variable is discrete. The same restriction can be stated directly in terms of conditional distributions, without requiring the existence of densities. It expresses the conditional independence of the price and the proxies given the latent type and (x, j) .¹¹ The factorization alone does not establish identification: the relevant rank or completeness, informativeness, and normalization conditions of the proxy model are also required. We refer to [Hu \(2008\)](#); [Hu and Schennach \(2008\)](#) for these conditions in their respective settings.

In practice, finding such instruments is often feasible. For example, in the insurance pricing setting studied by [Cosconati et al. \(2025\)](#), x^* represents a consumer's unobserved risk type, which affects the premiums offered by insurers. Realized claims can provide proxies for this type. In labor applications, measures of work performance or test scores may provide proxies for unobserved productivity.

Use a first-step estimator appropriate for the specified proxy model, such as a suitably regularized sieve likelihood estimator. Its selected price component yields conditional CDF estimates

$$\widehat{G}_j^s(t \mid x, x^*) = \widehat{H}_{p | x^*, x, j}([p_j, t] \mid x^*, x, j), \quad t \in \mathcal{P}_j.$$

The other input to the second step is the latent-type distribution conditional on (x, j) ,

¹¹An alternative specification uses three instruments (z_1, z_2, z_3) satisfying

$$\begin{aligned} & h_{z_3, z_1 | z_2, x, j, p}(z_3, z_1 \mid z_2, x, j, p) \\ &= \int_{X^*} h_{z_3 | x^*, x, j, p}(z_3 \mid x^*, x, j, p) h_{z_1 | x^*, x, j, p}(z_1 \mid x^*, x, j, p) dH_{x^* | z_2, x, j, p}(x^* \mid z_2, x, j, p). \end{aligned}$$

This specification permits the instruments to depend on price while maintaining the required conditional independence among the instruments. The appropriate proxy-model identification conditions must still be verified.

for every $j \in \mathcal{J}$, including $j = 0$.¹² Its first-step estimate is obtained by

$$\widehat{H}_{x^*|x,j}(A | x, j) = \int \widehat{H}_{x^*|z_2,x,j}(A | z_2, x, j) d\widehat{H}_{z_2|x,j}(z_2 | x, j).$$

For the remainder of the estimation analysis, let

$$h = (G^s, H_{x^*|x,j}), \quad \hat{h} = (\widehat{G}^s, \widehat{H}_{x^*|x,j})$$

denote these two derived nuisance objects, and let h_0 denote their true values. Other first-step estimates affect the second step through $\hat{h}$. Identification of the proxy model does not by itself establish the statistical properties of $\hat{h}$; the consistency, rate, and influence-function conditions needed below are stated explicitly in Section 3.2.

Step 2: Estimating Selection Function Parameters and Offered Price Distribution Given $\hat{h}$, we use a weighted log-choice criterion and define

$$\hat{\theta} \in \arg \max_{\theta \in \Theta} \hat{Q}_n(\theta), \quad (15)$$

where

$$\hat{Q}_n(\theta) = \frac{1}{n} \sum_{i=1}^n \int_{X^*} \log \pi_{j_i}(x_i, x^*; \theta, \widehat{G}^s) d\widehat{H}_{x^*|x,j}(x^* | x_i, j_i), \quad (16)$$

¹²When outside option is present and x^* is unobserved, the first step must also identify $H_{x^*|x,j}(\cdot | x, 0)$ from nonprice information on outside choosers. An outside share alone does not identify these latent-type weights. One sufficient route is an appropriately identified three-proxy model among outside choosers, with conditional independence given $(x, x^*, j = 0)$ and a latent-type normalization common to the inside-choice models. For example, replacing p in the three-measurement factorization by an additional nonprice proxy z_3 gives, among observations with $j = 0$,

$$\begin{aligned} & h_{z_3, z_1 | z_2, x, j}(z_3, z_1 | z_2, x, 0) \\ &= \int_{X^*} h_{z_3 | x^*, x, j}(z_3 | x^*, x, 0) h_{z_1 | x^*, x, j}(z_1 | x^*, x, 0) dH_{x^* | z_2, x, j}(x^* | z_2, x, 0). \end{aligned}$$

The corresponding rank or completeness and normalization conditions must be verified. No price for alternative 0 is used. The available covariate and proxy data must support the conditional weights used in the criterion, rather than merely an unconditional outside count. With no latent covariate, these additional weights are unnecessary.

and

$$\begin{aligned} \pi_j(x, x^*; \theta, G^s) \\ = \int_{\mathcal{P}} f_j(\mathbf{p}; x, x^*, \theta) d(F^{-1}(G^s(x, x^*); \theta, x, x^*))(\mathbf{p}), \quad j = 0, \dots, J. \end{aligned} \quad (17)$$

Integration with respect to F^{-1} in (17) means integration with respect to the product of its recovered marginal probability measures. We use π_j for this integrated choice probability, while $\Pr_j(p_j; G)$ continues to denote the choice probability conditional on the alternative's own price (with the conditioning characteristics and parameters suppressed). For discrete x^* , (16) reduces to the corresponding weighted sum. This is a two-step semiparametric M -estimator: the latent-type weights are estimated in the first step and held fixed when optimizing over θ .

Equation (17) is defined for every $j = 0, 1, \dots, J$, but its product integral contains only the J inside price marginals. In particular,

$$\pi_0(x, x^*; \theta, G^s) = 1 - \sum_{j=1}^J \pi_j(x, x^*; \theta, G^s).$$

The observations with $j_i = 0$ contribute their outside log choice probability to (16), using their own first-step weights $\hat{H}_{x^*|x,j}(\cdot | x_i, 0)$. No selected CDF for alternative 0 is required. This complete probability vector is used in the population identification and zero-mean score arguments below.

The inverse in (17) is computed by the contraction algorithm of Theorem 1. Write $\hat{T}_{\theta, \hat{h}}$ for the operator constructed with θ and $\widehat{G}^s$, and let $\hat{T}_{\theta, \hat{h}}^\infty \Psi$ denote its fixed point, obtained from any initial probability distribution Ψ .¹³ In the inner loop, the fixed point determines the choice probabilities in (17). The outer loop maximizes (16) over θ . After estimating θ , the offered price distributions are recovered by

$$\hat{G} = \hat{T}_{\hat{\theta}, \hat{h}}^\infty \Psi = F^{-1}(\widehat{G}^s; \hat{\theta}).$$

The dependence on (x, x^*) is suppressed in this last display.

¹³The finite-iteration approximation and the numerical accuracy needed for inference are discussed at the end of Section 3.2 and proved in Appendix A.7.

3.2 Consistency and Asymptotic Normality

We establish consistency and semiparametric asymptotic normality. The main technical issue is the dependence of the criterion on the functional inverse F^{-1} .

Proposition 2. *Suppose $\rho < 1$. For a fixed selection function, F is a homeomorphism in the extended Hilbert metric D . Moreover,*

$$\begin{aligned} D(F(G), F(G')) &\leq (1 + \rho)D(G, G'), \\ D(F^{-1}(G^s), F^{-1}(G^{st})) &\leq \frac{D(G^s, G^{st})}{1 - \rho}. \end{aligned}$$

The bounds are informative on the corresponding finite-distance components.

Proof. See Appendix A.5. □

These inequalities concern the metric D , not the uniform distance between CDFs. An empirical measure and a continuous population measure need not be mutually absolutely continuous, so their Hilbert distance can be infinite. For the statistical analysis we therefore use CDF distances, retaining Hilbert contraction to control the linearized fixed-point equation.

Probability measures are identified with their CDFs whenever a function norm is used. For a vector v of conditional CDFs or CDF perturbations, write

$$\|v\|_\infty = \sup_{(x, x^*) \in X \times X^*} \max_{1 \leq j \leq J} \sup_{p_j \in \mathcal{P}_j} |v_j(p_j \mid x, x^*)|. \quad (18)$$

When (x, x^*) is fixed and suppressed, the same notation denotes only the maximum over j and supremum over p_j . The functions are defined on the maintained conditioning domain $X \times X^*$; identification is interpreted under the actual joint distribution of (x, x^*) , which need not have full product support. For $h = (G^s, H)$, with $H = H_{x^*|x, j}$, define

$$\|h - h_0\|_{\mathcal{H}} = \|G^s - G_0^s\|_\infty + \sup_{x \in X, j \in \mathcal{J}} \|H(\cdot \mid x, j) - H_0(\cdot \mid x, j)\|_{\text{TV}}. \quad (19)$$

Here $\|\nu\|_{\text{TV}} = \sup_{|a| \leq 1} |\int a d\nu|$ for a finite signed measure ν .¹⁴ Total variation is imposed only on the latent-type mixing measure, including the weights conditional

¹⁴We work with fixed measurable versions of the conditional distributions on the maintained conditioning domain. Uniform consistency on this domain is a first-step requirement, not an implication

on $j = 0$, not on selected price CDFs. For a density estimator of the mixing measure, this is an L^1 condition. With no latent covariate, the mixing component is omitted.

Assumption 2 (Uniform regularity and first-step consistency).

- (i) $\Theta \subset \mathbb{R}^{d_\theta}$ is compact and contains θ_0 . The observations ω_i are i.i.d.
- (ii) For $j \in \mathcal{J}$, each $f_j(\mathbf{p}; x, x^*, \theta)$ is measurable in (x, x^*) and continuous in $(\mathbf{p}, \theta)$. On the fixed compact price rectangle $\mathcal{P}$ and an open neighborhood of Θ , the mixed derivatives

$$\partial_{\mathbf{p}}^\alpha \partial_\theta^\beta f_j(\mathbf{p}; x, x^*, \theta), \quad \alpha \in \{0, 1\}^J, \quad |\beta| \leq 2,$$

exist, are uniformly continuous in $(\mathbf{p}, \theta)$ uniformly over (x, x^*) , and are uniformly bounded over all indices and arguments. There exists $\varepsilon > 0$ such that

$$f_j(\mathbf{p}; x, x^*, \theta) \geq \varepsilon \quad \text{for all } j \in \mathcal{J}, (x, x^*) \in X \times X^*, \mathbf{p} \in \mathcal{P}, \theta \in \Theta.$$

- (iii) For some $\bar{\rho} \in (0, 1)$,

$$\sup_{(x, x^*) \in X \times X^*, \theta \in \Theta} \rho(x, x^*, \theta) \leq \bar{\rho} < 1,$$

where $\rho(x, x^*, \theta)$ is the bound in Theorem 1.

- (iv) The first-step estimates are probability measures on the maintained conditional supports and satisfy $\|\hat{h} - h_0\|_{\mathcal{H}} = o_p(1)$.

The price derivatives permit integration by parts in each coordinate and provide a sufficient condition for CDF-space differentiation without smoothness or parametric restrictions on offered price densities. The bounds must hold uniformly over (x, x^*) as well as prices and parameters.¹⁵

of almost-sure identification alone. When measurability issues arise, o_p , O_p , and convergence in probability are interpreted in outer probability; functional weak convergence is understood in the outer-expectation sense.

¹⁵The first-step consistency condition includes $H_{x^*|x,j}(\cdot | x, 0)$ when the outside option is available, and must be verified for the specified proxy model and estimator. If the outside option is absent ($f_0 \equiv 0$), omit 0 from the choice sums, the positivity condition, and the mixing-measure norm; the same proofs then apply to the reduced choice set. In contrast, when the outside option is present, its positive log-probability contribution is included throughout.

Write G_0 and G_0^s for the true collections of inside offered and selected distributions. Define

$$\ell(\omega; \theta, h) = \int_{X^*} \log \pi_j(x, x^*; \theta, G^s) dH(x^* | x, j). \quad (20)$$

Then $\hat{Q}_n(\theta) = \mathbb{P}_n \ell(\theta, \hat{h})$, where $\mathbb{P}_n a = n^{-1} \sum_{i=1}^n a(\omega_i)$. The population criterion is

$$\begin{aligned} Q_0(\theta) &= \mathbb{E}[\ell(\omega; \theta, h_0)] \\ &= \mathbb{E}_{x, x^*} \left[\sum_{k=0}^J \pi_k(x, x^*; \theta_0, G_0^s) \log \pi_k(x, x^*; \theta, G_0^s) \right]. \end{aligned} \quad (21)$$

The last equality follows by iterated expectations using the true latent-type distribution conditional on (x, j) , including $j = 0$.

Assumption 3 (Identification of the selection parameters). For $\theta \in \Theta$, if

$$\pi_j(x, x^*; \theta, G_0^s) = \pi_j(x, x^*; \theta_0, G_0^s) \quad \text{for all } j \in \mathcal{J}$$

for almost every (x, x^*) under its joint distribution, then $\theta = \theta_0$.

Contraction identifies the inside offered distributions conditional on any candidate selection parameter. Assumption 3 requires that the resulting complete choice-probability vector identify the remaining parameter θ . A difference at a single zero-probability conditioning value is not sufficient for identification through this criterion. This requirement is analogous to uniqueness conditions in structural demand estimation: in [Berry et al. \(1995\)](#), an inner fixed point recovers mean utilities, and the remaining parameters are identified by population moments. Here the inner fixed point recovers offered price distributions. Additional instruments, such as exogenous cost shifters, can provide further identifying moments in applications.

Lemma 2 in the appendix proves local Lipschitz continuity and differentiability of the inverse in the CDF norm. Combining its uniform bounds with [Newey and McFadden \(1994, Lemma 2.4 and Theorem 2.1\)](#) yields the following result.

Theorem 3 (Consistency). *Under Assumptions 2 and 3,*

$$\hat{\theta} \xrightarrow{p} \theta_0, \quad \|\hat{G} - G_0\|_\infty \xrightarrow{p} 0, \quad \hat{G} = \hat{T}_{\hat{\theta}, \hat{h}}^\infty \Psi.$$

Proof. See Appendix A.5. □

For asymptotic normality, define the second-step score by

$$\mathbf{g}(\omega; \theta, h) = \nabla_{\theta} \ell(\omega; \theta, h), \quad M_0 = \mathbb{E}[\nabla_{\theta} \mathbf{g}(\omega; \theta_0, h_0)]. \quad (22)$$

We use $D_h \mathbf{g}(\omega; \theta_0, h_0)$ for its Fréchet derivative in the nuisance norm (19). Specifically, this is a bounded linear map on nuisance perturbations such that, through the local extension used in the appendix,

$$\mathbf{g}(\omega; \theta_0, h_0 + v) - \mathbf{g}(\omega; \theta_0, h_0) - D_h \mathbf{g}(\omega; \theta_0, h_0)[v] = o(\|v\|_{\mathcal{H}}) \quad \text{as } \|v\|_{\mathcal{H}} \rightarrow 0.$$

Appendix A.6 establishes its existence and a uniform quadratic remainder. In expressions involving $\hat{h}$, the notation $\mathbb{E}_{\omega}$ means integration over a fresh observation with the true law while holding the realized first-step estimate fixed. It is not expectation over the training sample used to construct $\hat{h}$.

Assumption 4 (Semiparametric first-step conditions).

- (i) $\theta_0 \in \text{int}(\Theta)$ and M_0 is nonsingular.
- (ii) $\|\hat{h} - h_0\|_{\mathcal{H}} = o_p(n^{-1/4})$.
- (iii) The linearized score satisfies
$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ D_h \mathbf{g}(\omega_i; \theta_0, h_0)[\hat{h} - h_0] - \mathbb{E}_{\omega} \left[D_h \mathbf{g}(\omega; \theta_0, h_0)[\hat{h} - h_0] \right] \right\} = o_p(1).$$
- (iv) There is a measurable function $\varphi(\omega) \in \mathbb{R}^{d_{\theta}}$ with $\mathbb{E}\varphi(\omega) = 0$ and $\mathbb{E}\|\varphi(\omega)\|^2 < \infty$ such that

$$\sqrt{n} \mathbb{E}_{\omega} [D_h \mathbf{g}(\omega; \theta_0, h_0)[\hat{h} - h_0]] = \frac{1}{\sqrt{n}} \sum_{i=1}^n \varphi(\omega_i) + o_p(1). \quad (23)$$

These conditions implement Newey and McFadden (1994, Theorems 8.1 and 8.2). Part (ii) controls the quadratic remainder, part (iii) is stochastic equicontinuity of the linearized score, and part (iv) describes the first-order effect of nuisance estimation. Part (ii) alone does not imply part (iv). The influence representation is a substantive condition on the first-step estimator and this particular score functional; it is not

implied by contraction or proxy identification. No $\sqrt{n}$ limit for the entire infinite-dimensional $\hat{h}$ is imposed.

Theorem 4 (Asymptotic Normality). *Suppose Assumptions 2, 3, and 4 hold. Then*

$$\begin{aligned}\sqrt{n}(\hat{\theta} - \theta_0) &= -M_0^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \{\mathbf{g}(\omega_i; \theta_0, h_0) + \varphi(\omega_i)\} + o_p(1) \\ &\xrightarrow{d} \mathcal{N}(0, V),\end{aligned}\tag{24}$$

where

$$\begin{aligned}V &= M_0^{-1} \Omega_0 (M_0^{-1})', \\ \Omega_0 &= \mathbb{E}[\{\mathbf{g}(\omega; \theta_0, h_0) + \varphi(\omega)\} \{\mathbf{g}(\omega; \theta_0, h_0) + \varphi(\omega)\}'].\end{aligned}\tag{25}$$

Proof. See Appendix A.6. □

The influence function includes estimation of both selected CDFs and latent-type weights, including outside-type weights and same-sample dependence. Appendix A.6 gives a sandwich variance estimator conditional on suitable estimates of this influence function.

The entire offered-price CDF requires a separate statement. Let $\mathcal{R}(G^s, \theta) = F^{-1}(G^s; \theta)$ act on conditional collections indexed by (x, x^*) . Write $D_{G^s} \mathcal{R}$ and $D_\theta \mathcal{R}$ for its bounded linear derivatives at (G_0^s, θ_0) .

Corollary 2 (Rates and functional limits for offered price distributions). *Under the conditions of Theorem 4:*

(i) *If $\|\widehat{G}^s - G_0^s\|_\infty = O_p(r_n)$ for $r_n \rightarrow 0$, then*

$$\|\hat{G} - G_0\|_\infty = O_p(r_n + n^{-1/2}).$$

(ii) *If, in addition, the joint functional central limit theorem*

$$(\sqrt{n}(\widehat{G}^s - G_0^s), \sqrt{n}(\hat{\theta} - \theta_0)) \rightsquigarrow (\mathbb{Z}_{G^s}, Z_\theta)\tag{26}$$

holds in the product of the CDF supremum-norm space and $\mathbb{R}^{d_\theta}$, with a tight, centered, jointly Gaussian limit, then

$$\sqrt{n}(\hat{G} - G_0) \rightsquigarrow D_{G^s} \mathcal{R}[\mathbb{Z}_{G^s}] + D_\theta \mathcal{R}[Z_\theta].\tag{27}$$

The distributional conclusions also hold at a fixed (x, x^*) when the corresponding first-step rate or joint functional limit is imposed at that conditioning value, while the conditions for $\hat{\theta}$ remain those of Theorem 4.

Proof. See Appendix A.6. □

Continuous price support therefore does not preclude a $\sqrt{n}$ Gaussian limit for an estimated CDF. Such a limit is inherited from a suitable first-step functional limit, not from Theorem 4 alone. Conditional distribution estimators or latent-variable first steps may instead converge at a slower rate.

Finite-iteration approximation. Let $\hat{Q}_{n,m}$ use $\hat{T}_{\theta, \hat{h}}^m \Psi$ in place of the exact inverse, with initialization independent of θ , and let $\hat{\theta}_m$ maximize this criterion. Appendix A.7 establishes

$$\sup_{\theta \in \Theta} |\hat{Q}_{n,m}(\theta) - \hat{Q}_n(\theta)| \leq C\bar{\rho}^{m-1},$$

and, on a neighborhood $\mathcal{N}$ of θ_0 ,

$$\sup_{\theta \in \mathcal{N}} \|\nabla_{\theta} \hat{Q}_{n,m}(\theta) - \nabla_{\theta} \hat{Q}_n(\theta)\| \leq Cm\bar{\rho}^{m-1}.$$

Thus $m(n) \rightarrow \infty$ suffices for consistency. The limits in Theorem 4 and Corollary 2(ii) are unchanged if

$$\sqrt{n} m(n) \bar{\rho}^{m(n)-1} \rightarrow 0. \tag{28}$$

A sufficient condition is

$$\liminf_{n \rightarrow \infty} \frac{m(n)}{\log n} > \frac{1}{2 \log(1/\bar{\rho})}.$$

Numerical optimization and any additional integration approximation must also satisfy the stated criterion or score accuracy requirements.

4 Monte Carlo Simulations

To examine how our estimators for θ and the offered price distribution G perform in finite samples, we conduct the original Monte Carlo experiment with two price-bearing alternatives ($J = 2$) and no outside option ($f_0 \equiv 0$). Alternative 0 and its log-probability contribution are omitted in this special case; the first-step conditions

are imposed only on choices 1 and 2. The contraction condition $\rho < 1$ is automatically satisfied. 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 $\kappa = 0.1$.¹⁶ Let $j_i \in \{1, 2\}$ 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 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: $\ln(p_{ij}) = \exp((\delta_{0j} + \delta_{1j}x_{i2})(\delta_{2j}x_i^* + \eta_{ij}))$, 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 $\delta_{02} = 0.1, \delta_{12} = 0.3, \delta_{22} = 0.1, \sigma_2 = 0.2$.

¹⁶Note 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 conventional excluded variable to exogenously shift the selection probability. 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. As shown, our estimator performs well in finite samples without an excluded variable in the selection model.

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.¹⁷

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 $z_i \sim \text{Poisson}(x_i^*)$ 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 a parametric relationship between the instrument and the unobservable, only one instrument is needed.

We assume that the econometrician observes $(j_i, x_{i1}, x_{i2}, p_i, z_i)$, where p_i denotes the price of the chosen alternative. Using these data, we apply the procedure described in Section 3.1 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.¹⁸ 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.¹⁹ For each design, we run 500 simulations with

¹⁷Although the offered price distributions have unbounded support, in the implementation we approximate the support by the realized price range. These numerical experiments illustrate finite-sample performance for unbounded DGPs; the fixed-compact-support asymptotic results do not, by themselves, justify this expanding-range approximation.

¹⁸We estimate the cumulative distribution function of prices at 300 grid points.

¹⁹Although [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.

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.

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

For the cumulative distribution functions of $\log(\text{price})$, 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, 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 $\log(\text{price})$ 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 $\log(\text{price})$ 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.

Our method requires the econometrician to correctly specify the functional form of 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 $\log(\text{price})$ 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

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 IBias² 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 IBias² 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.

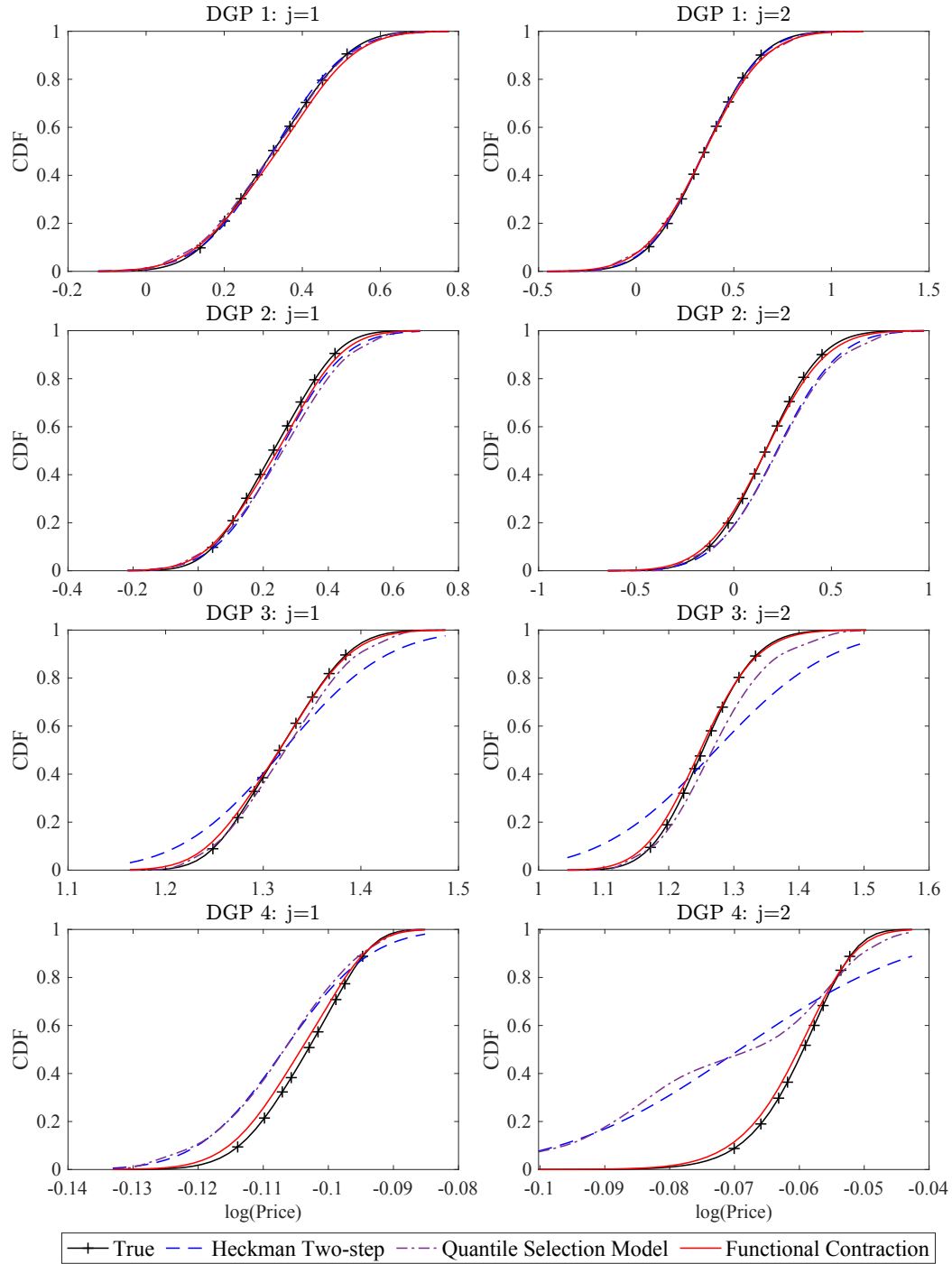


Figure 2: CDFs of $\log(\text{price})$ for alternatives 1 and 2, conditional on $x_{i2} = 0.25$. The black curve with “+” markers, the blue dashed curve, the purple dash-dotted curve, and the red solid curve correspond to the true CDF, the Heckman two-step estimate, the quantile selection estimate, and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

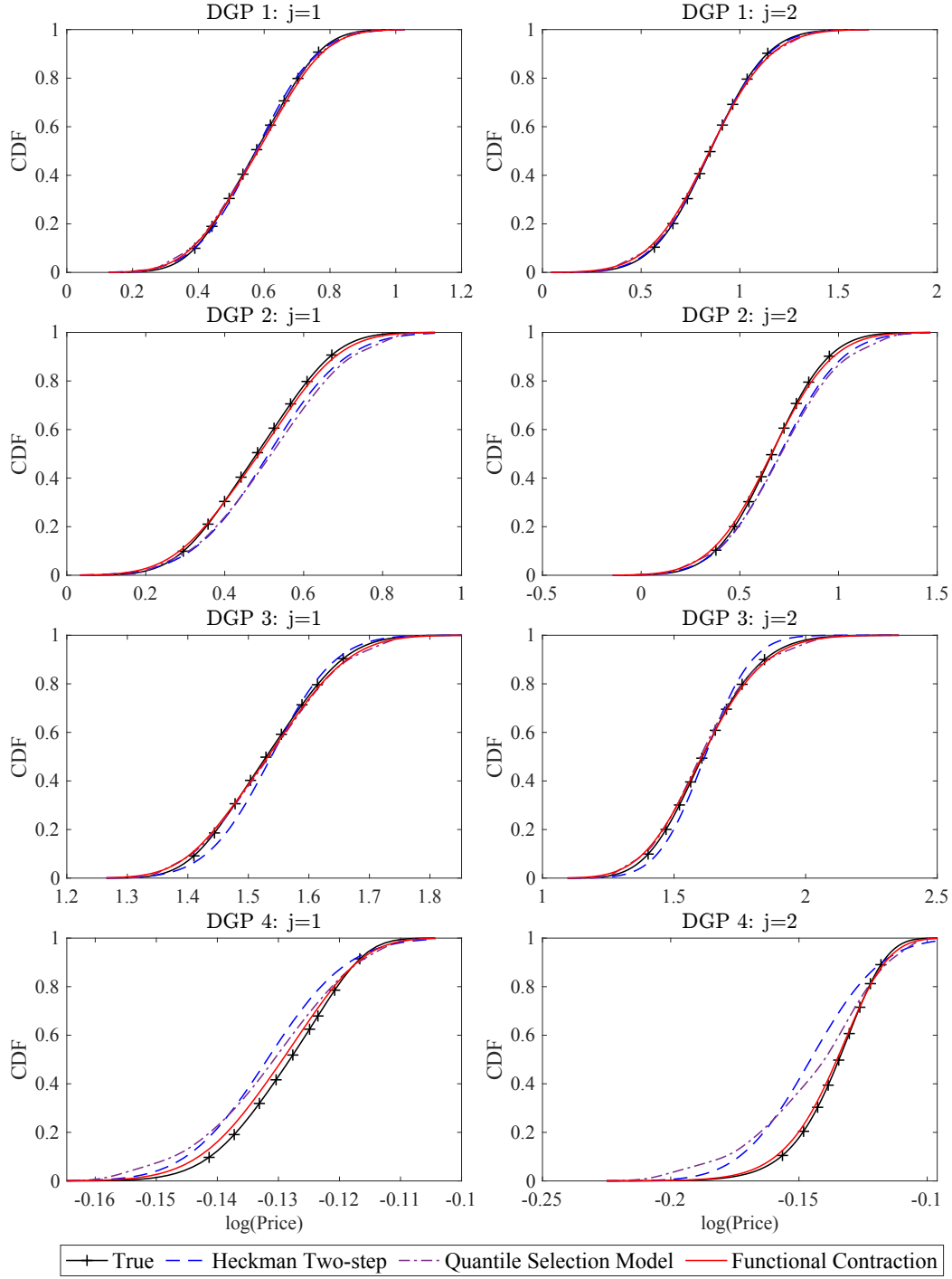


Figure 3: CDFs of $\log(\text{price})$ for alternatives 1 and 2, conditional on $x_{i2} = 0.75$. The black curve with “+” markers, the blue dashed curve, the purple dash–dotted curve, and the red solid curve correspond to the true CDF, the Heckman two-step estimate, the quantile selection estimate, and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

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 may be relatively insensitive to moderate misspecification of the selection function. This property could be useful in practice, particularly when limited prior information is available regarding the appropriate functional form.

Finally, we briefly discuss how the functional contraction performs computationally in practice. 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 attractive for applied work.

5 Discussion of Empirical Applications

The estimator introduced in Section 3 is broadly applicable to a wide range of empirical settings. It offers an alternative approach to addressing the challenge of selection bias that arises when only the outcomes of chosen alternatives are observed. The method has several features that can be useful in 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 to exogenously shift the choice probability. Such an excluded variable may not be readily available in certain empirical settings.

A natural application of the proposed method 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 [Cosconati et al. \(2025\)](#), we apply the 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.

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 this 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

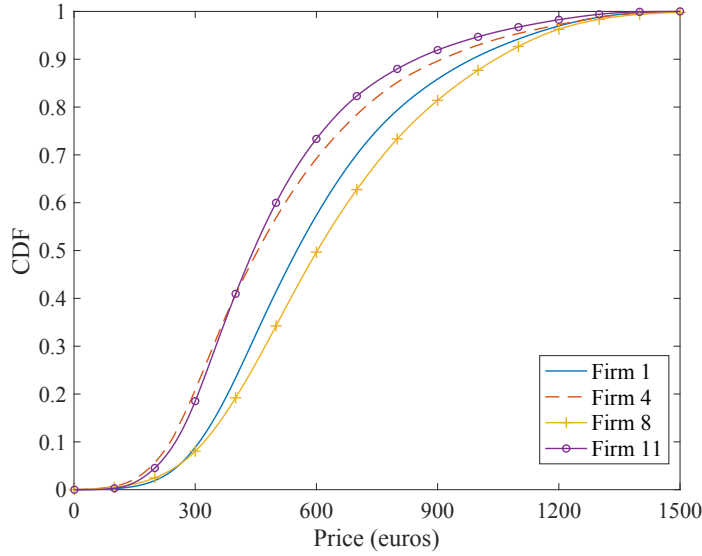


Figure 4: CDFs of the offered price distributions for firms 1, 4, 8, and 11 based on estimates from [Cosconati et al. \(2025\)](#). The CDFs are averaged across different characteristic groups.

winning bids and the winner’s identity are observed, even in the presence of bidder asymmetry.²⁰ Auctions in many settings have used the scoring rule that departs from the pure price-based criterion by accounting for quality differences,²¹ and our framework can flexibly accommodate these multi-attribute scoring mechanisms with both observable and unobservable components. A similar application arises in Roy-type models, where our method can recover the distribution of potential wages flexibly when only realized wages in the chosen sector are observed. Unlike identification strategies that rely on excluded variables in the selection equation, our approach instead requires proxy variables to recover the selected outcome distribution conditional on the latent characteristics. For example, in studies of occupational choice and wage inequality, cognitive and noncognitive abilities may be treated as latent characteristics that jointly affect occupational choices and wages, while standardized test scores or other skill assessments can be used as proxies for these latent characteristics.

²⁰Flexibly accommodating bidder asymmetries is a well-known challenge in auction models (see discussions in [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](#)).

²¹See, for example, [Asker and Cantillon \(2008\)](#); [Lewis and Bajari \(2011\)](#); [Nakabayashi \(2013\)](#); [Yoganarasimhan \(2016\)](#); [Takahashi \(2018\)](#); [Krasnokutskaya et al. \(2020\)](#); [Allen et al. \(2024\)](#).

6 Conclusion

We introduce a new method for estimating nonseparable selection models. We show that, for a given selection function and under suitable contraction conditions, the 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 establishing sufficient conditions under which the 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. Under suitable regularity conditions, we establish consistency of both components and semiparametric asymptotic normality of the selection-parameter estimator without imposing finite price support. The recovered outcome CDFs inherit the first-step convergence rate and admit Gaussian process limits when the corresponding first-step functional central limit theorem holds.

Our method has several features that are relevant for empirical applications. 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. Third, it does not require an excluded variable that shifts the selection probability without entering the outcome equation, nor does it rely on identification-at-infinity arguments. However, our approach does require a different class of instruments satisfying the conditional independence assumptions commonly imposed in the nonclassical measurement-error literature. Finally, the framework 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 [Cosconati et al. \(2025\)](#). The approach is straightforward to implement and computationally efficient, making it 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,

multi-attribute 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.
- 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’HAULTFÈUILLE, 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’HAULTFÈUILLE, X. AND A. MAUREL (2013): “Another look at the identification at infinity of sample selection models,” *Econometric Theory*, 29, 213–224.
- D’HAULTFÈUILLE, 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.
- FRENCH, E. AND C. TABER (2011): “Identification of models of the labor market,” in *Handbook of labor economics*, Elsevier, vol. 4, 537–617.
- 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.
- HECKMAN, J. J. AND B. E. HONORE (1990): “The empirical content of the Roy model,” *Econometrica: Journal of the Econometric Society*, 1121–1149.
- HENDRICKS, K. AND R. H. PORTER (1988): “An empirical study of an auction with asymmetric information,” *The American Economic Review*, 865–883.
- HONORÉ, B. E. AND L. HU (2020): “Selection without exclusion,” *Econometrica*, 88, 1007–1029.
- 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 non-classical 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.
- LEE, S. AND A. LEWBEL (2013): “Nonparametric identification of accelerated failure time competing risks models,” *Econometric Theory*, 29, 905–919.
- 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.
- NEWKEY, 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.
- NEWKEY, 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

All proofs are developed using the measure-theoretic formulation of the model presented in Section 2.2.

A.1 Hilbert's Projective Metric

For two probability measures $\Psi_j, \Phi_j \in \Delta([p_j, \bar{p}_j])$, we measure their distance using Hilbert's projective metric. If Ψ_j and Φ_j are mutually absolutely continuous,

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

where $\frac{d\Psi_j}{d\Phi_j}(p)$ denotes the Radon-Nikodym derivative. If Ψ_j and Φ_j are not mutually absolutely continuous, set $d_H(\Psi_j, \Phi_j) = +\infty$. When both Ψ_j and Φ_j have continuous densities, the Radon-Nikodym derivative simplifies to the ratio of densities as in Equation (6).

Recall that two probability measures Ψ_j and Φ_j are equivalent, denoted $\Psi_j \sim \Phi_j$, if they are mutually absolutely continuous. Since the selection function f_j is strictly positive, $\Pr_j(\cdot; G)$ is also strictly positive. It follows from Equation (7) that $G_j \sim G_j^s$. When $\Psi_j \sim \Phi_j$, the Radon-Nikodym derivative,

$$\frac{d\Psi_j}{d\Phi_j} : [p_j, \bar{p}_j] \rightarrow (0, \infty),$$

exists, as guaranteed by the Radon-Nikodym Theorem. Note that

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

Given our operator T in Equation (10), for all $\Psi_j, \Phi_j \in \Delta([p_j, \bar{p}_j])$,

$$(T\Psi)_j \sim G_j^s \sim (T\Phi)_j.$$

Moreover, since $f_j > 0$ is continuous with compact support, $\Pr_j$ is bounded away from 0. Thus, $d_H((T\Psi)_j, G_j)$, $d_H((T\Psi)_j, G_j^s)$ and $d_H((T\Psi)_j, (T\Phi)_j)$ are all finite.

A.2 Proof of Theorem 1

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

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

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).

The metric space $(\mathcal{X}(G_j), d_H)$ is complete. Consequently,

$$\mathcal{X}(G) = \prod_{j=1}^J \mathcal{X}(G_j)$$

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

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

where

$$\text{osc}_G(v) = \text{ess sup}_G v - \text{ess inf}_G v.$$

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,

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

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,

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

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

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

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

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

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

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

We first show that Φ is a probability measure. Since

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

for all sufficiently large n ,

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

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,

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

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

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,

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

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

Since the full space

$$\mathcal{X}(G) = \prod_{j=1}^J \mathcal{X}(G_j)$$

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

$$D(\Psi, \Phi) = \max_{1 \leq j \leq J} d_H(\Psi_j, \Phi_j).$$

□

Proof of Theorem 1. If $J = 1$, then $\text{Pr}_1(p_1; \Psi) = f_1(p_1)$ is independent of Ψ , so T is constant and the assertions follow immediately. In the rest of the proof take $J \geq 2$.

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

$$d(T\Psi)_j(p_j) = \frac{dG_j^s(p_j)/\text{Pr}_j(p_j; \Psi)}{\int dG_j^s(t)/\text{Pr}_j(t; \Psi)}.$$

Therefore,

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

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,

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

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

Write

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

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

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

to the cone of positive functions on $[\underline{p}_j, \bar{p}_j]$. Indeed, for any nonzero finite nonnegative measure Ψ_{-j} on $\prod_{k \neq j} [\underline{p}_k, \bar{p}_k]$, 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.²²

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

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

²²[Lemmens and Nussbaum \(2012\)](#) provide a detailed discussion on Perron-Frobenius theory and Birkhoff theorem.

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

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

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

$$d_H(\Pr_j(\cdot; \Psi), \Pr_j(\cdot; \Phi)) \leq \tanh\left(\frac{\Delta_j}{4}\right) d_H(\Psi_{-j}, \Phi_{-j}). \quad (\text{A.1})$$

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,

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

and similarly,

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

Hence,

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

This inequality is used only for pairs at finite input distance. No convention for $0 \cdot \infty$ is needed; arbitrary initial conjectures are handled by the first-update argument below.

Therefore, for each j ,

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

Taking the maximum over j yields

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

Thus, if

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

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

To obtain existence and uniqueness of the fixed point, fix the selected distribution G^s and consider the finite-distance component

$$\mathcal{X}(G^s) = \prod_{j=1}^J \left\{ \Psi_j \in \Delta([p_j, \bar{p}_j]) : d_H(\Psi_j, G_j^s) < \infty \right\}.$$

By Lemma 1, $\mathcal{X}(G^s)$ is complete under the metric D . Moreover, for any

$$\Psi \in \prod_{j=1}^J \Delta([p_j, \bar{p}_j]),$$

the update $T\Psi$ satisfies

$$(T\Psi)_j \sim G_j^s \quad \text{for all } 1 \leq j \leq J.$$

Since f_j is continuous and strictly positive on the compact support, the selection probability $\Pr_j(\cdot; \Psi)$ is bounded above and bounded away from zero. Hence

$$d_H((T\Psi)_j, G_j^s) < \infty \quad \text{for all } 1 \leq j \leq J,$$

and therefore

$$T\Psi \in \mathcal{X}(G^s).$$

Thus, T maps into $\mathcal{X}(G^s)$.

The contraction inequality established above therefore applies on the complete metric space $\mathcal{X}(G^s)$. By the Banach fixed-point theorem, T has a unique fixed point in $\mathcal{X}(G^s)$. Denote it by G . The Bayes identity implies $F(G) = G^s$, and any probability fixed point must lie in this same finite-distance component.

Finally, if the initial conjecture

$$\Psi \in \prod_{j=1}^J \Delta(\underline{p}_j, \bar{p}_j]$$

does not lie in $\mathcal{X}(G^s)$, then its first iterate $T\Psi$ does. Applying the contraction result from the second iterate onward yields

$$D(T^n\Psi, G) \leq \rho^{n-1} D(T\Psi, G) \rightarrow 0.$$

Hence, for any initial conjecture Ψ ,

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

□

A.3 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

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

Fix $j \in \{1, \dots, J\}$. For any $p_j, p'_j \in [\underline{p}_j, \bar{p}_j]$ and $\mathbf{p}_{-j}, \mathbf{p}'_{-j} \in \prod_{k \neq j} [\underline{p}_k, \bar{p}_k]$, consider

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

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

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

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

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

It follows that

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

The upper bound is attained by taking

$$p_j = \bar{p}_j, \quad p'_j = \underline{p}_j, \quad \mathbf{p}_{-j} = \bar{\mathbf{p}}_{-j}, \quad \mathbf{p}'_{-j} = \underline{\mathbf{p}}_{-j}.$$

Hence,

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

By Theorem 1,

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

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.4 Proof of Proposition 1

Proof. Fix the conditioning characteristics and the selection function. Suppose that two collections of offered distributions, $G = (G_1, \dots, G_J)$ and $G' = (G'_1, \dots, G'_J)$, satisfy

$$F(G) = F(G') = G^s.$$

We show that $G_j = G'_j$ for every j .

For any collection of offered distributions G , define

$$Z_G = \int_{\mathcal{P}} K(\mathbf{p}) \prod_{k=1}^J dG_k(p_k), \quad dQ^G(\mathbf{p}) = \frac{K(\mathbf{p})}{Z_G} \prod_{k=1}^J dG_k(p_k).$$

Continuity and strict positivity of K on the compact set $\mathcal{P}$ imply that Z_G is finite and strictly positive. Thus Q^G is a probability measure. Let Q_j^G denote its j th marginal. By construction,

$$dQ_j^G(p_j) = \frac{\int_{\prod_{k \neq j} \mathcal{P}_k} K(p_j, \mathbf{p}_{-j}) \prod_{k \neq j} dG_k(p_k)}{Z_G} dG_j(p_j).$$

Using (11) in the forward mapping (7) gives

$$dF_j(G)(p_j) = \frac{a_j(p_j) dQ_j^G(p_j)}{\int_{\mathcal{P}_j} a_j(t) dQ_j^G(t)}.$$

Consequently, the common selected distribution G_j^s determines the auxiliary marginal:

$$dQ_j^G(p_j) = dQ_j^{G'}(p_j) = \frac{a_j(p_j)^{-1} dG_j^s(p_j)}{\int_{\mathcal{P}_j} a_j(t)^{-1} dG_j^s(t)}, \quad j = 1, \dots, J. \quad (\text{A.2})$$

In particular, this recovery uses only the conditional selected distributions and the given selection function, not choice shares.

We next establish that equality of these marginals implies $Q^G = Q^{G'}$. Write

$$\underline{K} = \min_{\mathbf{p} \in \mathcal{P}} K(\mathbf{p}) > 0, \quad \overline{K} = \max_{\mathbf{p} \in \mathcal{P}} K(\mathbf{p}) < \infty.$$

The marginal formula above implies

$$\frac{K}{\overline{K}} \leq \frac{dQ_j^G}{dG_j} \leq \frac{\overline{K}}{\underline{K}} \quad G_j\text{-almost everywhere,}$$

and the same bounds hold with G replaced by G' . Since $Q_j^G = Q_j^{G'}$, the measures G_j and G'_j are mutually absolutely continuous. Moreover, their density ratio

$$r_j(p_j) := \frac{dG_j}{dG'_j}(p_j)$$

satisfies

$$\left(\frac{K}{\overline{K}}\right)^2 \leq r_j \leq \left(\frac{\overline{K}}{\underline{K}}\right)^2 \quad G'_j\text{-almost everywhere.}$$

Indeed, equality of the auxiliary marginals gives

$$r_j = \frac{dQ_j^{G'}/dG'_j}{dQ_j^G/dG_j} \quad G'_j\text{-almost everywhere.}$$

Thus $\log r_j \in L^\infty(G'_j)$.

It follows that Q^G and $Q^{G'}$ are also mutually absolutely continuous, and cancellation of the common kernel gives

$$\log \frac{dQ^G}{dQ^{G'}}(\mathbf{p}) = \log \frac{Z_{G'}}{Z_G} + \sum_{j=1}^J \log r_j(p_j).$$

All logarithms in this identity are bounded, so the following integrals are finite. Since both joint measures have total mass one and their corresponding marginals agree,

$$\begin{aligned} & \int_{\mathcal{P}} \log \frac{dQ^G}{dQ^{G'}}(\mathbf{p}) [dQ^G(\mathbf{p}) - dQ^{G'}(\mathbf{p})] \\ &= \sum_{j=1}^J \int_{\mathcal{P}_j} \log r_j(p_j) [dQ_j^G(p_j) - dQ_j^{G'}(p_j)] \\ &= 0. \end{aligned}$$

The integral on the left is the symmetrized relative entropy. Equivalently, it is

$$\int_{\mathcal{P}} \left(\frac{dQ^G}{dQ^{G'}} - 1 \right) \log \left(\frac{dQ^G}{dQ^{G'}} \right) dQ^{G'}.$$

For every $z > 0$, $(z - 1) \log z \geq 0$, with equality only at $z = 1$. Therefore,

$$\frac{dQ^G}{dQ^{G'}} = 1 \quad Q^{G'}\text{-almost everywhere,} \quad \text{and hence} \quad Q^G = Q^{G'}.$$

Finally, the offered product measure is uniquely recovered by removing the strictly positive common weight:

$$\prod_{j=1}^J dG_j(p_j) = \frac{K(\mathbf{p})^{-1} dQ^G(\mathbf{p})}{\int_{\mathcal{P}} K(\mathbf{q})^{-1} dQ^G(\mathbf{q})}.$$

The same formula applies to G' . Since $Q^G = Q^{G'}$, we obtain

$$\bigotimes_{j=1}^J G_j = \bigotimes_{j=1}^J G'_j.$$

Taking marginals yields $G_j = G'_j$ for every j , proving that F is injective. $\square$

A.5 Proof of Theorem 3

Proof of Proposition 2. By Theorem 1, for every inside selected collection G^s , the operator T_{G^s} has a unique fixed point. The Bayes identity (7) implies that this is the unique G with $F(G) = G^s$; hence F is bijective. For $D(G, G') < \infty$, reweighting and normalization give

$$\begin{aligned} d_H(F_j(G), F_j(G')) &\leq d_H(G_j, G'_j) + \text{osc}_{p_j} \log \frac{\Pr_j(p_j; G)}{\Pr_j(p_j; G')} \\ &\leq (1 + \rho) D(G, G'), \quad j = 1, \dots, J, \end{aligned}$$

using (A.1) and the product-measure bound. Taking the maximum over the inside indices proves the forward bound.

For $D(G^s, G^{s'}) < \infty$, let $G = F^{-1}(G^s)$ and $G' = F^{-1}(G^{s'})$. Strict positivity on compact support gives $D(G, G^s) < \infty$ and $D(G', G^{s'}) < \infty$, and thus $D(G, G') < \infty$.

Reweightings both selected measures by the same positive function preserves their projective distance, so

$$D(T_{G^s}G, T_{G^{s'}}G) = D(G^s, G^{s'}).$$

The two fixed-point equations and contraction imply

$$\begin{aligned} D(G, G') &\leq D(T_{G^s}G, T_{G^{s'}}G) + D(T_{G^{s'}}G, T_{G^{s'}}G') \\ &\leq D(G^s, G^{s'}) + \rho D(G, G'). \end{aligned}$$

Rearranging gives the inverse bound. The finite-distance components are preserved by F and its inverse, and these inequalities give continuity on each component and hence in the extended-metric topology. The outside option adds no coordinate to these spaces. $\square$

For the CDF-space argument, fix (x, x^*) and suppress it in expressions local to this conditioning pair. Let

$$\mathbb{D}_j = \{v_j : \mathcal{P}_j \rightarrow \mathbb{R} : v_j \text{ is càdlàg and } v_j(\bar{p}_j) = 0\}, \quad \mathbb{D} = \prod_{j=1}^J \mathbb{D}_j,$$

with $\|v\|_\infty = \max_{1 \leq j \leq J} \sup_{t \in \mathcal{P}_j} |v_j(t)|$. This is a Banach space: uniform limits preserve the càdlàg property and the endpoint constraint. Each perturbation is extended by zero to the left of $\underline{p}_j$, so $v_j(\underline{p}_j-) = 0$. This convention is automatic for differences of CDFs supported on $\mathcal{P}_j$ and does not require $v_j(\underline{p}_j) = 0$; different masses at the lower endpoint are allowed. For $q \in C^1(\mathcal{P}_j)$, define

$$\int_{[\underline{p}_j, t]} q(p) dv_j(p) := q(t)v_j(t) - \int_{\underline{p}_j}^t v_j(p)q'(p) dp. \quad (\text{A.3})$$

For bounded-variation perturbations this is the usual Stieltjes integral, including the atom at $\underline{p}_j$. For other elements of $\mathbb{D}_j$, the right-hand side defines its continuous extension. In particular,

$$\sup_{t \in \mathcal{P}_j} \left| \int_{[\underline{p}_j, t]} q dv_j \right| \leq (\|q\|_\infty + (\bar{p}_j - \underline{p}_j)\|q'\|_\infty) \|v_j\|_\infty. \quad (\text{A.4})$$

This makes the integral a bounded linear map in the CDF perturbation. All subse-

quent integrals against such perturbations use (A.3). Repetition in different coordinates defines the corresponding continuous multilinear product integrals.

Lemma 2 (CDF stability and differentiation of the inverse). *Under Assumption 2(ii)–(iii),*

$$\mathcal{R}(G^s, \theta; x, x^*) := F^{-1}(G^s; \theta, x, x^*)$$

is uniformly continuous on the domain

$$\left(\prod_{j=1}^J \Delta(\mathcal{P}_j) \right) \times \Theta,$$

using the CDF supremum norm and Euclidean parameter norm, uniformly over (x, x^*) . Around every probability input it admits a local extension to the affine CDF space and a parameter neighborhood.²³ The first and second Fréchet derivatives of this extension are continuous and bounded. Mixed derivatives of total order at most three, with at most two parameter derivatives, also exist, are continuous, and are locally bounded. The neighborhood sizes and bounds are uniform over (x, x^*) , $\theta \in \Theta$, and probability inputs. In particular, for sufficiently close inputs,

$$\begin{aligned} \|\mathcal{R}(G^{s'}, \theta'; x, x^*) - \mathcal{R}(G^s, \theta; x, x^*)\|_\infty \\ \leq C(\|G^{s'} - G^s\|_\infty + \|\theta' - \theta\|). \end{aligned} \tag{A.5}$$

For a CDF direction η and a parameter direction a ,

$$D\mathcal{R}[\eta, a] = (I - \mathcal{A})^{-1}\{\mathcal{B}\eta + D_\theta T[a]\}, \quad \mathcal{A} = D_\Psi T, \quad \mathcal{B} = D_{G^s} T, \tag{A.6}$$

where derivatives of T are evaluated at $\Psi = \mathcal{R}(G^s, \theta; x, x^*)$ and (G^s, θ) . The same statements apply to measurable collections indexed by (x, x^*) with norm (18).

²³For $j = 1, \dots, J$, the affine space is

$$\mathbb{A}_j = \{\Phi_j : \mathcal{P}_j \rightarrow \mathbb{R} : \Phi_j \text{ is càdlàg, } \Phi_j(\bar{p}_j) = 1\}, \quad \mathbb{A} = \prod_{j=1}^J \mathbb{A}_j,$$

with $\Phi_j(\underline{p}_j -) = 0$ by left extension. For any probability CDF G_j° , $\mathbb{A}_j = G_j^\circ + \mathbb{D}_j$. Elements need not be nondecreasing, lie in $[0, 1]$, or have bounded variation; “total mass one” refers to their endpoint normalization. A local extension of $\mathcal{R}$ is a map from an open neighborhood in $\mathbb{A} \times \mathbb{R}^{d_\theta}$ into $\mathbb{A}$ that agrees with the original inverse at nearby probability inputs. In the proof, T is first extended in both its distribution arguments, to a neighborhood in $\mathbb{A} \times \mathbb{A} \times \mathbb{R}^{d_\theta}$.

Proof. All constants and neighborhood sizes below are uniform over the conditioning pairs and probability inputs. We suppress (x, x^*) unless a supremum over these arguments is being displayed.

Step 1: Smooth extension of the Bayes operator. The original operator takes two probability-CDF collections and a parameter as inputs and returns a probability-CDF collection. We extend it locally to $\mathbb{A} \times \mathbb{A} \times \mathbb{R}^{d_\theta}$, with $\mathbb{A}$ defined in the lemma's footnote. For $\Psi = \Psi^\circ + v$, $v \in \mathbb{D}$, the product integral

$$\Pr_j(p_j; \Psi, \theta) = \int f_j(\mathbf{p}; \theta) \prod_{k \neq j} d\Psi_k(p_k), \quad j = 1, \dots, J,$$

is a polynomial of degree at most $J - 1$ in v . By repeated integration by parts, the derivatives in Assumption 2(ii) make it a continuous polynomial map into $C^1(\mathcal{P}_j)$. Each integrated coordinate uses at most one price derivative. Differentiating the resulting function in p_j uses one additional derivative in the own-price coordinate. The same argument applies after up to two parameter derivatives. On a uniformly bounded neighborhood of probability inputs all these multilinear maps have uniform bounds.

At probability inputs, $\varepsilon \leq \Pr_j \leq 1$, and

$$\int_{\mathcal{P}_j} \frac{dG_j^s(p)}{\Pr_j(p; \Psi, \theta)} \geq 1.$$

The preceding continuity estimates and (A.4) therefore provide a common small neighborhood on which both these positive quantities stay bounded away from zero. For an affine input Φ_j , integration means the extension of (A.3) to $\Phi_j = G_j^\circ + v_j$; it agrees with the usual integral when Φ_j is a CDF. Thus the same formula

$$T_j(\Psi; G^s, \theta)(t) = \frac{\int_{[\underline{p}_j, t]} \Pr_j(p; \Psi, \theta)^{-1} dG_j^s(p)}{\int_{\mathcal{P}_j} \Pr_j(p; \Psi, \theta)^{-1} dG_j^s(p)}$$

defines a local extension, still denoted T , into $\mathbb{A}$. It need not return a probability CDF at nonprobability inputs, but $T_j(\bar{p}_j) = 1$ still holds. Reciprocation is smooth on functions bounded away from zero in the Banach algebra $C^1(\mathcal{P}_j)$, and weighted integration is continuous bilinear in its weight and CDF argument. Consequently,

T is smooth in its distribution arguments and twice continuously differentiable in θ , with the uniform mixed derivative bounds stated in the lemma. This supplies an open domain for differentiation without assuming that the probability-CDF set has nonempty interior.

Step 2: Continuity of the probability-valued inverse. Fix an initial probability collection Ψ° independent of the inputs. Set $G^{(m)} = T^m(\Psi^\circ; G^s, \theta)$ and $G = \mathcal{R}(G^s, \theta)$. The first update and the fixed point are obtained by reweighting the same selected distribution by weights in $[1, 1/\varepsilon]$. Their projective distance is therefore uniformly bounded, and contraction gives, for $m \geq 1$,

$$D(G^{(m)}, G) \leq C\bar{\rho}^{m-1}.$$

For probability measures at finite projective distance d , normalization implies $e^{-d} \leq dU/dV \leq e^d$ and hence $\sup_t |U(t) - V(t)| \leq e^d - 1$. It follows that

$$\|G^{(m)} - G\|_\infty \leq C\bar{\rho}^{m-1}, \tag{A.7}$$

uniformly over the selected probability inputs, parameters, and conditioning pairs.

The common local neighborhoods and first-derivative bounds from Step 1 give uniform continuity of each finite iterate in (G^s, θ) . More explicitly, for each fixed m there are $C_m, \delta_m > 0$, independent of the base input, such that two inputs at distance $d < \delta_m$ have m th iterates at CDF distance at most $C_m d$. Applying (A.7) at both inputs gives

$$\|\mathcal{R}(G^{s'}, \theta') - \mathcal{R}(G^s, \theta)\|_\infty \leq 2C\bar{\rho}^{m-1} + C_m d, \quad d = \|G^{s'} - G^s\|_\infty + \|\theta' - \theta\|.$$

Given any output tolerance, first choose a finite m to make the first term small and then choose d to make the second small. This proves uniform continuity on $(\prod_j \Delta(\mathcal{P}_j)) \times \Theta$. No uniform bound on C_m as $m \rightarrow \infty$ is required. This step will identify the subsequent local implicit solution with the actual probability-valued inverse.

Step 3: First derivatives at a fixed point. Evaluate T at $G = \mathcal{R}(G^s, \theta)$. The Bayes identity is

$$dG_j^s(p) = \frac{\Pr_j(p; G, \theta)}{\pi_j(\theta, G^s)} dG_j(p), \quad j = 1, \dots, J,$$

where only the fixed conditioning arguments are suppressed. For a CDF direction $v \in \mathbb{D}$,

$$D_G \Pr_j(p_j; G, \theta)[v] = \sum_{k \neq j} \int f_j(\mathbf{p}; \theta) dv_k(p_k) \prod_{\ell \neq j, k} dG_\ell(p_\ell). \quad (\text{A.8})$$

Differentiating the numerator and denominator of the update gives

$$\begin{aligned} (\mathcal{A}v)_j(t) = & - \int_{[p_j, t]} \frac{D_G \Pr_j(p; G, \theta)[v]}{\Pr_j(p; G, \theta)} dG_j(p) \\ & + G_j(t) \int_{\mathcal{P}_j} \frac{D_G \Pr_j(p; G, \theta)[v]}{\Pr_j(p; G, \theta)} dG_j(p), \end{aligned} \quad (\text{A.9})$$

and

$$\begin{aligned} (\mathcal{B}\eta)_j(t) = \pi_j(\theta, G^s) \Big\{ & \int_{[p_j, t]} \frac{d\eta_j(p)}{\Pr_j(p; G, \theta)} \\ & - G_j(t) \int_{\mathcal{P}_j} \frac{d\eta_j(p)}{\Pr_j(p; G, \theta)} \Big\}. \end{aligned} \quad (\text{A.10})$$

For a parameter direction $a \in \mathbb{R}^{d_\theta}$,

$$D_\theta \Pr_j(p_j; G, \theta)[a] = \int \partial_\theta f_j(\mathbf{p}; \theta) a \prod_{k \neq j} dG_k(p_k),$$

where the distribution is held fixed in this partial derivative. For $(D_\theta T[a])_j$, use (A.9) and replace the distribution derivative $D_G \Pr_j[v]$ by the parameter derivative $D_\theta \Pr_j[a]$. All these maps are bounded linear operators by (A.4) and uniform positivity.

Step 4: A bounded inverse on the CDF space. We do not assert $\|\mathcal{A}\|_{\infty \rightarrow \infty} < 1$. Instead, consider the relative-density perturbation space

$$\mathbb{E}_G = \prod_{j=1}^J \{b_j \in L^\infty(G_j) : \int b_j dG_j = 0\}, \quad \|b\|_{\text{osc}} = \max_{1 \leq j \leq J} \text{osc}_{G_j}(b_j).$$

On this mean-zero space, $\|b\|_{L^\infty} \leq \|b\|_{\text{osc}} \leq 2\|b\|_{L^\infty}$, so the oscillation norm is com-

plete. The symbols b_j here denote density directions, not price endpoints. Define

$$(J_G b)_j(t) = \int_{[\underline{p}_j, t]} b_j(p) dG_j(p)$$

and

$$\begin{aligned} (\mathcal{L}_G v)_j(p) &= -\frac{D_G \Pr_j(p; G, \theta)[v]}{\Pr_j(p; G, \theta)} \\ &\quad + \int_{\mathcal{P}_j} \frac{D_G \Pr_j(r; G, \theta)[v]}{\Pr_j(r; G, \theta)} dG_j(r). \end{aligned}$$

Then $\mathcal{A} = J_G \mathcal{L}_G$, $\mathcal{L}_G : \mathbb{D} \rightarrow \mathbb{E}_G$, and

$$\|J_G b\|_\infty \leq \|b\|_{\text{osc}}, \quad \|\mathcal{L}_G v\|_{\text{osc}} \leq C \|v\|_\infty. \quad (\text{A.11})$$

For fixed $b \in \mathbb{E}_G$, take the probability path

$$dG_{j,\epsilon} = (1 + \epsilon b_j) dG_j$$

for sufficiently small $|\epsilon|$. Boundedness and the mean-zero constraint ensure positivity and total mass one. The selected collection G^s , the parameter θ , and (x, x^*) remain fixed throughout; in particular the same operator T is used and $TG = G$. In CDF coordinates,

$$G_{j,\epsilon}(t) - G_j(t) = \epsilon \int_{[\underline{p}_j, t]} b_j(p) dG_j(p) = \epsilon (J_G b)_j(t).$$

Since $\log(1 + \epsilon b_j) = \epsilon b_j + O(\epsilon^2)$ in $L^\infty(G_j)$ for this fixed direction,

$$D(G_\epsilon, G) = |\epsilon| \|b\|_{\text{osc}} + o(|\epsilon|).$$

Expanding the product of perturbed marginal measures gives

$$\begin{aligned} \Pr_j(p_j; G_\epsilon, \theta) &= \int f_j(\mathbf{p}; \theta) \prod_{k \neq j} (1 + \epsilon b_k(p_k)) \prod_{k \neq j} dG_k(p_k) \\ &= \Pr_j(p_j; G, \theta) + \epsilon D_G \Pr_j(p_j; G, \theta) [J_G b] + O(\epsilon^2), \end{aligned}$$

uniformly over p_j and j for fixed b . Indeed, $d(J_G b)_k = b_k dG_k$, and all higher product

terms have at least two factors of ϵ . Uniform positivity therefore implies

$$\frac{\Pr_j(p; G, \theta)}{\Pr_j(p; G_\epsilon, \theta)} = 1 - \epsilon \frac{D_G \Pr_j(p; G, \theta)[J_G b]}{\Pr_j(p; G, \theta)} + O(\epsilon^2).$$

Substitution of the fixed-point Bayes identity into the update gives the exact relative-density formula

$$\frac{d(TG_\epsilon)_j}{dG_j}(p) = \frac{\Pr_j(p; G, \theta) / \Pr_j(p; G_\epsilon, \theta)}{\int_{\mathcal{P}_j} \frac{\Pr_j(r; G, \theta)}{\Pr_j(r; G_\epsilon, \theta)} dG_j(r)}.$$

The normalizing denominator equals

$$1 - \epsilon \int_{\mathcal{P}_j} \frac{D_G \Pr_j(r; G, \theta)[J_G b]}{\Pr_j(r; G, \theta)} dG_j(r) + O(\epsilon^2).$$

Expanding the ratio consequently yields

$$\begin{aligned} \frac{d(TG_\epsilon)_j}{dG_j}(p) &= 1 + \epsilon \left\{ - \frac{D_G \Pr_j(p; G, \theta)[J_G b]}{\Pr_j(p; G, \theta)} \right. \\ &\quad \left. + \int_{\mathcal{P}_j} \frac{D_G \Pr_j(r; G, \theta)[J_G b]}{\Pr_j(r; G, \theta)} dG_j(r) \right\} + O(\epsilon^2) \\ &= 1 + \epsilon(\mathcal{L}_G J_G b)_j(p) + O(\epsilon^2). \end{aligned}$$

Thus the output relative-density direction is $\mathcal{L}_G J_G b$, and the centering term makes it mean zero under G_j . More precisely,

$$\max_j \left\| \frac{d(TG_\epsilon)_j}{dG_j} - 1 - \epsilon(\mathcal{L}_G J_G b)_j \right\|_{L^\infty(G_j)} = o(|\epsilon|).$$

These are Radon–Nikodym densities relative to G_j , not Lebesgue densities. Taking logarithms preserves this first-order expansion. For bounded q, r ,

$$|\text{osc}_{G_j}(q + r) - \text{osc}_{G_j}(q)| \leq 2\|r\|_{L^\infty(G_j)}, \quad \text{osc}_{G_j}(\epsilon q) = |\epsilon| \text{osc}_{G_j}(q).$$

Hence

$$\begin{aligned} D(TG_\epsilon, G) &= \max_j \text{osc}_{G_j} \left(\log \frac{d(TG_\epsilon)_j}{dG_j} \right) \\ &= |\epsilon| \|\mathcal{L}_G J_G b\|_{\text{osc}} + o(|\epsilon|). \end{aligned}$$

Applying $D(TG_\epsilon, TG) \leq \bar{\rho}D(G_\epsilon, G)$ and $TG = G$, dividing by $|\epsilon|$, and letting $\epsilon \rightarrow 0$ gives

$$\|\mathcal{L}_G J_G b\|_{\text{osc}} \leq \bar{\rho} \|b\|_{\text{osc}}. \quad (\text{A.12})$$

This argument uses a first-order expansion of the distance; it does not differentiate an essential supremum or require differentiability of the metric itself.

For any $v \in \mathbb{D}$ and $m \geq 1$,

$$\mathcal{A}^m v = J_G (\mathcal{L}_G J_G)^{m-1} \mathcal{L}_G v.$$

Combining (A.11) and (A.12) gives

$$\|\mathcal{A}^m v\|_\infty \leq C \bar{\rho}^{m-1} \|v\|_\infty. \quad (\text{A.13})$$

The Neumann series therefore converges in operator norm, with

$$(I - \mathcal{A})^{-1} = \sum_{m=0}^{\infty} \mathcal{A}^m, \quad \|(I - \mathcal{A})^{-1}\| \leq 1 + \frac{C}{1 - \bar{\rho}}. \quad (\text{A.14})$$

Indeed multiplication of each finite partial sum by $I - \mathcal{A}$ gives $I - \mathcal{A}^{m+1}$, and the last term tends to zero in operator norm. This proves both left and right invertibility, not merely uniqueness of a formal derivative.

Step 5: Implicit differentiation and uniform bounds. Apply the Banach-space implicit function theorem to the extension of

$$\Psi - T(\Psi; G^s, \theta) = 0.$$

After translation by a fixed probability triple, this is a map from an open subset of $\mathbb{D} \times \mathbb{D} \times \mathbb{R}^{d_\theta}$ into $\mathbb{D}$. Step 1 gives differentiability on this open set, and Step 4 gives the bounded inverse of its derivative in Ψ . Step 2 identifies the local implicit solution with the original inverse at nearby probability inputs. Differentiation proves (A.6).

The inverse bound and local second-derivative bounds are uniform. Consequently, a common small ball can be chosen on which the linearization-based implicit-function construction is contractive. The input tolerance can also be chosen uniformly, since Step 2 places nearby probability solutions inside this ball. This proves (A.5) with a common constant.

For input directions (η_1, a_1) and (η_2, a_2) , twice differentiating the fixed-point identity yields

$$\begin{aligned} & (I - \mathcal{A})D^2\mathcal{R}[(\eta_1, a_1), (\eta_2, a_2)] \\ &= D^2T[(D\mathcal{R}[\eta_1, a_1], \eta_1, a_1), (D\mathcal{R}[\eta_2, a_2], \eta_2, a_2)]. \end{aligned}$$

The arguments on the right correspond, in order, to (Ψ, G^s, θ) . Repeating the differentiation for pure CDF directions and for mixed directions with at most two parameter derivatives proves the claimed third-order mixed bounds. Each new derivative is obtained by applying $(I - \mathcal{A})^{-1}$ to bounded derivatives of T and already bounded lower-order derivatives of $\mathcal{R}$. No third derivative solely in θ is required. The bounded second derivative gives a uniformly quadratic Taylor remainder. Pointwise constructions and their limits preserve measurability in (x, x^*) , and the common bounds yield the conditional-collection statements under (18). $\square$

Lemma 3. *Under Assumption 2, $\sup_{\theta \in \Theta} |\hat{Q}_n(\theta) - Q_0(\theta)| = o_p(1)$, and Q_0 is continuous.*

Proof. For every choice $j = 0, \dots, J$, uniform positivity gives

$$\varepsilon \leq \pi_j(x, x^*; \theta, G^s) \leq 1.$$

The product integral has bounded CDF derivatives, so Lemma 2 and the chain rule give

$$\begin{aligned} & \sup_{x, x^*, \theta, 0 \leq j \leq J} |\pi_j(x, x^*; \theta, G^s) - \pi_j(x, x^*; \theta, G_0^s)| \\ & \leq C \|G^s - G_0^s\|_\infty \end{aligned}$$

for sufficiently close CDF inputs. Splitting the difference in (20) into the change in log probability and the change in the mixing measure gives

$$\sup_{\omega, \theta} |\ell(\omega; \theta, h) - \ell(\omega; \theta, h_0)| \leq C \|h - h_0\|_{\mathcal{H}}. \quad (\text{A.15})$$

In particular,

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

The first term is $o_p(1)$ by first-step consistency. For fixed h_0 , the criterion is continuous in θ and bounded in absolute value by $|\log \varepsilon|$. Compactness and i.i.d. sampling allow application of [Newey and McFadden \(1994, Lemma 2.4\)](#); this controls the second term and proves continuity of Q_0 . $\square$

Proof of Theorem 3. For every conditioning pair, the vectors $(\pi_j(x, x^*; \theta, G_0^s))_{j=0}^J$ are positive probability vectors, including the outside choice. Thus

$$\begin{aligned} Q_0(\theta_0) - Q_0(\theta) &= \mathbb{E}_{x, x^*} \left[\sum_{j=0}^J \pi_j(x, x^*; \theta_0, G_0^s) \log \frac{\pi_j(x, x^*; \theta_0, G_0^s)}{\pi_j(x, x^*; \theta, G_0^s)} \right] \geq 0. \end{aligned}$$

The integrand is a Kullback–Leibler divergence. Equality requires agreement of the complete probability vectors almost surely under the joint distribution of (x, x^*) , and hence $\theta = \theta_0$ by Assumption 3. The outside term is essential when outside choices occur; it cannot be discarded while retaining the same unnormalized inside log probabilities.

Lemma 3, compactness, and [Newey and McFadden \(1994, Theorem 2.1\)](#) imply consistency of the exact maximizer. For an approximate maximizer, the population criterion shortfall is at most twice the uniform criterion approximation error plus the sample optimization gap, and is therefore $o_p(1)$. Continuity and uniqueness on compact Θ then give the same parameter consistency. On an event with probability tending to one, the CDF and parameter inputs lie in the common neighborhood of Lemma 2, so

$$\|\hat{G} - G_0\|_{\infty} \leq C\{\|\widehat{G}^s - G_0^s\|_{\infty} + \|\hat{\theta} - \theta_0\|\} = o_p(1).$$

This uses CDF distance, not the potentially infinite Hilbert distance between estimated and population distributions. $\square$

A.6 Proof of Theorem 4

We first make the nuisance derivative in (22) explicit. Fix (x, x^*) temporarily and suppress it where this causes no ambiguity. For a selected-CDF direction η , put $v = D_{G^s} \mathcal{R}[\eta] = (I - \mathcal{A})^{-1} \mathcal{B} \eta$. Differentiating the product integral in (17) gives, for every choice $j = 0, 1, \dots, J$,

$$D_{G^s} \pi_j[\eta] = \sum_{k=1}^J \int f_j(\mathbf{p}; x, x^*, \theta) dv_k(p_k) \prod_{\ell \neq k} dG_\ell(p_\ell). \quad (\text{A.16})$$

Here $G = \mathcal{R}(G^s, \theta; x, x^*)$, and the derivatives include the effect of changing the recovered inside distributions. Similarly, for a parameter direction a , put $v^a = D_\theta \mathcal{R}[a] = (I - \mathcal{A})^{-1} D_\theta T[a]$. Then

$$\begin{aligned} D_\theta \pi_j[a] &= \int \partial_\theta f_j(\mathbf{p}; x, x^*, \theta) a \prod_{k=1}^J dG_k(p_k) \\ &\quad + \sum_{k=1}^J \int f_j(\mathbf{p}; x, x^*, \theta) dv_k^a(p_k) \prod_{\ell \neq k} dG_\ell(p_\ell). \end{aligned} \quad (\text{A.17})$$

The sum over k always contains exactly J price marginals, including when $j = 0$. There is no price direction for the outside option.

For a nuisance direction (η, ν) , where $\nu(\cdot \mid x, j)$ is a signed mixing-measure perturbation of mass zero, the derivative at the truth is

$$\begin{aligned} D_h \mathbf{g}(\omega; \theta_0, h_0)[\eta, \nu] &= \int_{X^*} D_{G^s} \nabla_\theta \log \pi_j(x, x^*; \theta, G^s) [\eta(x, x^*)] \big|_{\theta=\theta_0, G^s=G_0^s} dH_0(x^* \mid x, j) \\ &\quad + \int_{X^*} \nabla_\theta \log \pi_j(x, x^*; \theta, G_0^s) \big|_{\theta=\theta_0} d\nu(x^* \mid x, j). \end{aligned} \quad (\text{A.18})$$

For fixed arguments, the derivative in the first integrand can be computed as

$$D_{G^s} \nabla_\theta \log \pi_j[\eta] = \nabla_\theta \left(\frac{D_{G^s} \pi_j[\eta]}{\pi_j} \right).$$

These formulas use (A.6), (A.16), and the mixed derivatives in Lemma 2, not a closed-form inverse.

We next prove the two bounds needed for the asymptotic expansion. The criterion is the individual second-step objective $\ell(\omega; \theta, h)$ in (20), with $h = (G^s, H)$, and its score is

$$\mathfrak{g}(\omega; \theta, h) = \int_{X^*} \nabla_{\theta} \log \pi_j(x, x^*; \theta, G^s) dH(x^* | x, j).$$

Choose a sufficiently small convex neighborhood $\mathcal{N}$ of θ_0 with closure contained in $\text{int}(\Theta)$, and a sufficiently small convex neighborhood $\mathcal{U}$ of h_0 in the extended nuisance space. The local extensions and all derivative bounds are uniform on these neighborhoods. For the mixing measures in a neighborhood of radius δ ,

$$\sup_{x, j} \|H(\cdot | x, j)\|_{\text{TV}} \leq 1 + \delta.$$

Bounds for the integrand derivatives. The map

$$(G, \theta) \mapsto \int_{\mathcal{P}} f_j(\mathbf{p}; x, x^*, \theta) \prod_{k=1}^J dG_k(p_k)$$

has bounded mixed derivatives of the required orders by the price smoothness assumption and the extended Stieltjes integrals. Composing with $G = \mathcal{R}(G^s, \theta; x, x^*)$ gives bounded mixed derivatives of π_j of total order at most three, with at most two parameter derivatives. At probability inputs, $\pi_j \geq \varepsilon$ for every $j \in \mathcal{J}$, including 0. Its extension is at least $\varepsilon/2$ after shrinking $\mathcal{U}$ if necessary.

In the next displays suppress the common arguments $(x, x^*; \theta, G^s)$, without introducing another notation for the choice probability. The chain rule gives

$$\begin{aligned} D_{G^s} \nabla_{\theta} \log \pi_j[\eta] &= \frac{\nabla_{\theta} D_{G^s} \pi_j[\eta]}{\pi_j} \\ &\quad - \frac{\nabla_{\theta} \pi_j}{\pi_j^2} D_{G^s} \pi_j[\eta], \end{aligned}$$

$$\begin{aligned} D_{G^s}^2 \nabla_{\theta} \log \pi_j[\eta_1, \eta_2] \\ = \nabla_{\theta} \left\{ \frac{D_{G^s}^2 \pi_j[\eta_1, \eta_2]}{\pi_j} - \frac{D_{G^s} \pi_j[\eta_1] D_{G^s} \pi_j[\eta_2]}{\pi_j^2} \right\}, \end{aligned}$$

and

$$D_{G^s} \nabla_{\theta}^2 \log \pi_j[\eta] = \nabla_{\theta}^2 \left\{ \frac{D_{G^s} \pi_j[\eta]}{\pi_j} \right\}.$$

The highest derivatives required are $\nabla_\theta D_{G^s}^2 \pi_j$ and $\nabla_\theta^2 D_{G^s} \pi_j$. All other terms involve lower-order derivatives and inverse powers of π_j . Uniform positivity and the derivative bounds consequently imply

$$\begin{aligned}\|\nabla_\theta \log \pi_j\| + \|\nabla_\theta^2 \log \pi_j\| &\leq C, \\ \|D_{G^s} \nabla_\theta \log \pi_j[\eta]\| &\leq C\|\eta\|_\infty, \\ \|D_{G^s}^2 \nabla_\theta \log \pi_j[\eta_1, \eta_2]\| &\leq C\|\eta_1\|_\infty \|\eta_2\|_\infty, \\ \|D_{G^s} \nabla_\theta^2 \log \pi_j[\eta]\| &\leq C\|\eta\|_\infty.\end{aligned}$$

The bounds are uniform over the conditioning pairs, choice indices, and indicated neighborhoods. A direction in an integrand is evaluated at (x, x^*) ; its local norm is bounded by the conditional-collection norm (18).

Nuisance derivatives of the score. Take $v_r = (\eta_r, \nu_r)$, $r = 1, 2$. Differentiating the integral and its mixing measure gives

$$\begin{aligned}D_h^2 \mathfrak{g}(\omega; \theta, h)[v_1, v_2] &= \int_{X^*} D_{G^s}^2 \nabla_\theta \log \pi_j[\eta_1, \eta_2] dH \\ &\quad + \int_{X^*} D_{G^s} \nabla_\theta \log \pi_j[\eta_1] d\nu_2 \\ &\quad + \int_{X^*} D_{G^s} \nabla_\theta \log \pi_j[\eta_2] d\nu_1,\end{aligned}$$

where the integrands use $(x, x^*; \theta, G^s)$ and the measures are conditional on (x, j) . There is no term with two derivatives of the mixing measure, because ℓ is linear in H . The two cross terms need not vanish. Similarly, for $v = (\eta, \nu)$,

$$\begin{aligned}D_h \nabla_\theta \mathfrak{g}(\omega; \theta, h)[v] &= \int_{X^*} D_{G^s} \nabla_\theta^2 \log \pi_j[\eta] dH \\ &\quad + \int_{X^*} \nabla_\theta^2 \log \pi_j d\nu.\end{aligned}$$

For a bounded vector- or matrix-valued integrand a ,

$$\left\| \int a(x^*) d\nu(x^*) \right\| \leq \sup_{x^*} \|a(x^*)\| \|\nu\|_{\text{TV}}.$$

Combining this with the preceding derivative bounds gives

$$\begin{aligned}
& \|D_h^2 \mathbf{g}(\omega; \theta, h)[v_1, v_2]\| \\
& \leq C \left\{ \|\eta_1\|_\infty \|\eta_2\|_\infty + \|\eta_1\|_\infty \sup_{x,k} \|\nu_2(\cdot \mid x, k)\|_{\text{TV}} \right. \\
& \quad \left. + \|\eta_2\|_\infty \sup_{x,k} \|\nu_1(\cdot \mid x, k)\|_{\text{TV}} \right\} \\
& \leq C \|v_1\|_{\mathcal{H}} \|v_2\|_{\mathcal{H}},
\end{aligned}$$

and

$$\|D_h \nabla_\theta \mathbf{g}(\omega; \theta, h)[v]\| \leq C \|v\|_{\mathcal{H}}.$$

These constants are uniform over ω , $\theta \in \mathcal{N}$, and $h \in \mathcal{U}$.

Taylor expansion along a nuisance path. Let $\Delta h = h - h_0$ and $h_t = h_0 + t\Delta h$, $0 \leq t \leq 1$, with the segment contained in $\mathcal{U}$. Taylor's formula with integral remainder gives

$$\begin{aligned}
& \mathbf{g}(\omega; \theta_0, h) - \mathbf{g}(\omega; \theta_0, h_0) - D_h \mathbf{g}(\omega; \theta_0, h_0)[\Delta h] \\
& = \int_0^1 (1-t) D_h^2 \mathbf{g}(\omega; \theta_0, h_t)[\Delta h, \Delta h] dt.
\end{aligned}$$

The uniform bilinear bound implies

$$\begin{aligned}
& \sup_{\omega} \|\mathbf{g}(\omega; \theta_0, h) - \mathbf{g}(\omega; \theta_0, h_0) - D_h \mathbf{g}(\omega; \theta_0, h_0)[h - h_0]\| \\
& \leq C \|h - h_0\|_{\mathcal{H}}^2.
\end{aligned} \tag{A.19}$$

The factor $\int_0^1 (1-t) dt = 1/2$ is absorbed into C . For each fixed $\theta \in \mathcal{N}$, the fundamental theorem of calculus similarly gives

$$\begin{aligned}
& \nabla_\theta \mathbf{g}(\omega; \theta, h) - \nabla_\theta \mathbf{g}(\omega; \theta, h_0) \\
& = \int_0^1 D_h \nabla_\theta \mathbf{g}(\omega; \theta, h_t)[\Delta h] dt.
\end{aligned}$$

Thus

$$\sup_{\omega, \theta \in \mathcal{N}} \|\nabla_\theta \mathbf{g}(\omega; \theta, h) - \nabla_\theta \mathbf{g}(\omega; \theta, h_0)\| \leq C \|h - h_0\|_{\mathcal{H}}. \tag{A.20}$$

Since $\mathbf{g} = \nabla_\theta \ell$, the two derivative orders are

$$D_h^2 \mathbf{g} = D_h^2 \nabla_\theta \ell, \quad D_h \nabla_\theta \mathbf{g} = D_h \nabla_\theta^2 \ell.$$

The first bound uses two nuisance derivatives and one parameter derivative of the individual criterion; the second uses one nuisance derivative and two parameter derivatives. The same calculations bound the first two parameter derivatives of ℓ and give their continuity in θ ; no pure third parameter derivative is needed.

Proof of Theorem 4. Theorem 3 gives consistency. By iterated expectations and the complete-choice identity,

$$\begin{aligned} & \mathbb{E}_\omega \mathbf{g}(\omega; \theta_0, h_0) \\ &= \mathbb{E}_{x, x^*} \left[\sum_{k=0}^J \pi_k(x, x^*; \theta_0, G_0^s) \nabla_\theta \log \pi_k(x, x^*; \theta, G_0^s) \Big|_{\theta=\theta_0} \right] \\ &= \mathbb{E}_{x, x^*} \left[\nabla_\theta \sum_{k=0}^J \pi_k(x, x^*; \theta, G_0^s) \Big|_{\theta=\theta_0} \right] = 0. \end{aligned}$$

The outside contribution is included, whereas all price integrations still involve only inside alternatives. Uniform derivative bounds justify interchanging differentiation and expectation and imply $\mathbb{E} \|\mathbf{g}(\omega; \theta_0, h_0)\|^2 < \infty$.

By (A.19) and Assumption 4(ii),

$$\begin{aligned} \sqrt{n} \mathbb{P}_n \mathbf{g}(\theta_0, \hat{h}) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{g}(\omega_i; \theta_0, h_0) \\ &\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^n D_h \mathbf{g}(\omega_i; \theta_0, h_0) [\hat{h} - h_0] + o_p(1). \end{aligned} \tag{A.21}$$

The remainder has norm at most $C\sqrt{n} \|\hat{h} - h_0\|_{\mathcal{H}}^2 = o_p(1)$. Assumption 4(iii) replaces the second sum by

$$\sqrt{n} \mathbb{E}_\omega [D_h \mathbf{g}(\omega; \theta_0, h_0) [\hat{h} - h_0]] + o_p(1),$$

where only the fresh observation is integrated out. Part (iv) then gives

$$\sqrt{n} \mathbb{P}_n \mathbf{g}(\theta_0, \hat{h}) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \{\mathbf{g}(\omega_i; \theta_0, h_0) + \varphi(\omega_i)\} + o_p(1). \quad (\text{A.22})$$

This is the score decomposition in [Newey and McFadden \(1994, Theorem 8.1\)](#).

The Jacobian bound [\(A.20\)](#) permits replacing $\hat{h}$ by h_0 in the sample Jacobian uniformly on a small compact convex neighborhood $\mathcal{N}_0 \subset \mathcal{N}$. At h_0 the Jacobian is continuous in θ and uniformly bounded. By [Newey and McFadden \(1994, Lemma 2.4\)](#),

$$\sup_{\theta \in \mathcal{N}_0} \left\| \mathbb{P}_n \nabla_{\theta} \mathbf{g}(\theta, \hat{h}) - \mathbb{E}_{\omega} \nabla_{\theta} \mathbf{g}(\omega; \theta, h_0) \right\| = o_p(1).$$

With probability tending to one, an exact maximizer is interior. Use the integral mean-value expansion

$$\mathbb{P}_n \mathbf{g}(\hat{\theta}, \hat{h}) = \mathbb{P}_n \mathbf{g}(\theta_0, \hat{h}) + M_n(\hat{\theta} - \theta_0),$$

where

$$M_n = \int_0^1 \mathbb{P}_n \nabla_{\theta} \mathbf{g}(\theta_0 + t(\hat{\theta} - \theta_0), \hat{h}) dt \xrightarrow{p} M_0.$$

The left-hand side is zero for an exact maximizer, and may be $o_p(n^{-1/2})$ for the stated numerical estimator. Since M_0 is nonsingular, solving this equation and using [\(A.22\)](#) gives [\(24\)](#). Its summands are i.i.d., mean zero, and square integrable. The multivariate central limit theorem yields [\(25\)](#), as in [Newey and McFadden \(1994, Theorem 8.2\)](#). $\square$

Variance estimation. Suppose the influence estimates satisfy

$$\frac{1}{n} \sum_{i=1}^n \|\hat{\varphi}_i - \varphi(\omega_i)\|^2 = o_p(1).$$

Let $\hat{M} = \mathbb{P}_n \nabla_{\theta} \mathbf{g}(\hat{\theta}, \hat{h})$ and define

$$\begin{aligned} \hat{\Omega} &= \frac{1}{n} \sum_{i=1}^n \{\mathbf{g}(\omega_i; \hat{\theta}, \hat{h}) + \hat{\varphi}_i\} \{\mathbf{g}(\omega_i; \hat{\theta}, \hat{h}) + \hat{\varphi}_i\}', \\ \hat{V} &= \hat{M}^{-1} \hat{\Omega} (\hat{M}^{-1})'. \end{aligned}$$

The Jacobian argument gives $\hat{M} \xrightarrow{p} M_0$. Consistency and the uniform score derivative bounds imply that the sample mean squared difference between the estimated summands and $\mathbf{g}(\omega_i; \theta_0, h_0) + \varphi(\omega_i)$ is $o_p(1)$. Cauchy–Schwarz and the law of large numbers give $\hat{\Omega} \xrightarrow{p} \Omega_0$ and $\hat{V} \xrightarrow{p} V$. Centering the estimated summands by their sample mean yields the same limit. Constructing $\hat{\varphi}_i$ remains a first-step-specific problem and is not supplied by contraction alone.

Proof of Corollary 2. The bounded second derivatives in Lemma 2 give, with derivatives evaluated at (G_0^s, θ_0) ,

$$\begin{aligned}\hat{G} - G_0 &= D_{G^s} \mathcal{R}[\widehat{G^s} - G_0^s] + D_\theta \mathcal{R}[\hat{\theta} - \theta_0] + R_n, \\ \|R_n\|_\infty &\leq C(\|\widehat{G^s} - G_0^s\|_\infty + \|\hat{\theta} - \theta_0\|)^2\end{aligned}\tag{A.23}$$

on an event with probability tending to one. At each conditioning pair,

$$D_{G^s} \mathcal{R} = (I - \mathcal{A})^{-1} \mathcal{B}, \quad D_\theta \mathcal{R} = (I - \mathcal{A})^{-1} D_\theta T.$$

Part (i) follows from the local CDF Lipschitz bound and Theorem 4. Under (26), both input errors are $O_p(n^{-1/2})$, so $\sqrt{n}\|R_n\|_\infty = o_p(1)$. The continuous mapping theorem for the bounded linear derivative gives (27). At a fixed (x, x^*) , the same proof uses the first-step CDF rate or joint functional convergence at that pair, along with the already established parameter result. $\square$

A.7 Finite-Iteration Approximation

Fix the selected probability measures, (x, x^*) , and an initial probability collection Ψ independent of θ ; suppress the fixed conditioning arguments below. Put $G_\theta^{(m)} = T_{\theta, G^s}^m \Psi$ and $G_\theta = \mathcal{R}(G^s, \theta)$. The uniform CDF bound (A.7), bounded price derivatives of f_j , and positivity give an error at most $C\bar{\rho}^{m-1}$ in each log choice probability, for $j = 0, \dots, J$. For $j = 0$ the same argument integrates f_0 over the inside prices. Integrating against the estimated mixing probability measures and averaging over observations proves the criterion bound.

For the score bound, use logarithmic coordinates relative to G^s after the first

update. Each coordinate is a class $[w_j] \in L^\infty(G_j^s)/\mathbb{R}$, with norm $\text{osc}_{G_j^s} w_j$, representing

$$dG_j(w_j) = \frac{e^{w_j} dG_j^s}{\int e^{w_j} dG_j^s}.$$

The product of these quotient spaces is Banach under the maximum oscillation norm. The update is

$$\mathcal{S}_\theta(w)_j = [-\log \text{Pr}_j(\cdot; G(w), \theta)], \quad j = 1, \dots, J.$$

It is well defined independently of the additive constants in the representatives. The logarithmic chart is an isometry for the projective distance, and Theorem 1 implies $\|D_w \mathcal{S}_\theta\| \leq \bar{\rho}$. The first update has representative $w_{\theta,j}^{(1)} = -\log \text{Pr}_j(\cdot; \Psi, \theta)$. Every iterate and the fixed point w_θ has a representative between 0 and $\log(1/\varepsilon)$.

The parameter and second mixed derivatives of $\mathcal{S}_\theta$ are uniformly bounded. One may choose the centered representatives $\int w_j dG_j^s = 0$ to perform this calculation. Derivatives of normalized exponential weights are finite products of bounded centered directions, and the selection-probability denominators are at least ε . The bounds are independent of atoms and support gaps in G_j^s . Contraction gives

$$\|w_\theta^{(m)} - w_\theta\| \leq C\bar{\rho}^{m-1},$$

and differentiation of the fixed-point equation gives $\|\partial_\theta w_\theta\| \leq C/(1 - \bar{\rho})$. Define only the derivative error

$$d_m = \|\partial_\theta w_\theta^{(m)} - \partial_\theta w_\theta\|.$$

Differentiating the iteration and subtracting the fixed-point derivative identity yields

$$d_{m+1} \leq \bar{\rho}d_m + C\|w_\theta^{(m)} - w_\theta\| \leq \bar{\rho}d_m + C\bar{\rho}^{m-1}.$$

Since d_1 is uniformly bounded and $\bar{\rho} \in (0, 1)$,

$$d_m \leq Cm\bar{\rho}^{m-1},$$

after enlarging a constant depending on $\bar{\rho}$.

For each $j = 0, \dots, J$, the function

$$(w, \theta) \mapsto \log \int_{\mathcal{P}} f_j(\mathbf{p}; \theta) \prod_{k=1}^J dG_k(w_k)(p_k)$$

has uniformly bounded first and second derivatives on the relevant set. The parameter derivatives along an iterate and the fixed point therefore differ by at most $C\{\|w_\theta^{(m)} - w_\theta\| + d_m\}$. Integrating against the estimated latent-type weights, including those for outside choices, proves

$$\sup_{\theta \in \mathcal{N}} \|\nabla_\theta \hat{Q}_{n,m}(\theta) - \nabla_\theta \hat{Q}_n(\theta)\| \leq Cm\bar{\rho}^{m-1}.$$

If $m(n) \rightarrow \infty$, the approximate criterion converges uniformly to Q_0 , so its global or $o_p(1)$ -approximate maximizer is consistent. For asymptotic equivalence, use the score bound and the nonsingular limiting Jacobian from Theorem 4. An integral mean-value expansion of the exact score between the consistent exact and approximate maximizers gives

$$\|\hat{\theta}_{m(n)} - \hat{\theta}\| = O_p(m(n)\bar{\rho}^{m(n)-1}).$$

The expansion is performed inside a common convex neighborhood of θ_0 , where the averaged exact Jacobian converges to M_0 . For numerical optimizers with $o_p(n^{-1/2})$ score residuals, add an $o_p(n^{-1/2})$ term. Uniform CDF stability and the iteration bound also give

$$\|\hat{T}_{\hat{\theta}_{m(n)}, \hat{h}}^{m(n)} \Psi - \hat{T}_{\hat{\theta}, \hat{h}}^\infty \Psi\|_\infty = O_p(m(n)\bar{\rho}^{m(n)-1}),$$

with the same possible numerical residual. Condition (28) makes both differences negligible at the $\sqrt{n}$ scale. The strict lower bound on $m(n)/\log n$ implies this condition: choose a smaller constant still exceeding $1/(2\log(1/\bar{\rho}))$, and use eventual monotonicity of $m\bar{\rho}^{m-1}$ to bound it at that constant times $\log n$.

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 $\log(\text{price})$: 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 IBias² 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 $\log(\text{price})$: 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 $\text{Logistic}(0, 1)$. The IBias² 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.