Estimating Nonseparable Selection Models: A Functional Contraction Approach

Fan Wu HSBC PKU

Yi Xin Caltech

December 10, 2025

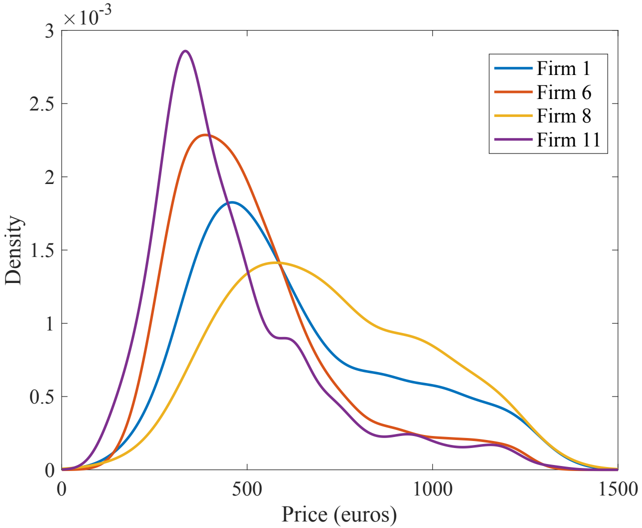
Motivation

- Sample selection issues: data is not representative due to a nonrandom selection process.
- Commonly observed in empirical settings:
  - **Consumer demand**: it is challenging to gain access to competing prices offered to consumers. (Goldberg, 1996; Cicala, 2015; Crawford et al., 2018; Allen et al., 2019; Salz, 2022; Sagl, 2023; Cosconati et al., 2024)
  - **Auctions**: bids are missing, either due to the auction's structure or incomplete data.
    - Allen et al. (2024): FDIC auctions for insolvent banks, only the winning and the second-highest bids are recorded.
    - Washington State DoT publishes only the three lowest bids.
  - **Roy models**: widely used in the literature to study decisions such as what occupation to pursue, whether to join a union.

Application in Insurance Demand

- An empirical problem we face when estimating demand for auto insurance! (REStud forthcoming)
- Insurers use risk-based pricing, leading to significant variation in prices across consumers.
- Insurers rely on heterogeneous pricing strategies (they collect different consumer information, use different algorithms).
- The shape of the offered price distribution has real economic meanings: it reflects the firm's pricing strategies, and risk-rating technology.
- Flexible estimation of the offered price distribution is essential for analyzing competition and supply-side heterogeneity.

Offered Price Distributions



This Paper

- We nonparametrically recover the outcome distributions before selection. The previous literature

This Paper

- We nonparametrically recover the outcome distributions before selection. The previous literature
  - Nonparametric: recovers only (difference between) the mean of outcome (Das, Newey, Vella 2003).

This Paper

- We nonparametrically recover the outcome distributions before selection. The previous literature
  - Nonparametric: recovers only (difference between) the mean of outcome (Das, Newey, Vella 2003).
  - Parametric: recovers the distribution (Arellano & Bonhomme 2017)

This Paper

- We nonparametrically recover the outcome distributions before selection. The previous literature
  - Nonparametric: recovers only (difference between) the mean of outcome (Das, Newey, Vella 2003).
  - Parametric: recovers the distribution (Arellano & Bonhomme 2017)
- We achieve this by constructing an operator whose fixed point is the potential outcome distributions, and showing that it is a functional contraction.

This Paper

- We nonparametrically recover the outcome distributions before selection. The previous literature
  - Nonparametric: recovers only (difference between) the mean of outcome (Das, Newey, Vella 2003).
  - Parametric: recovers the distribution (Arellano & Bonhomme 2017)
- We achieve this by constructing an operator whose fixed point is the potential outcome distributions, and showing that it is a functional contraction.
- We also propose an estimator for the parameters in the selection model. Establish consistency and asymptotic normality.

This Paper

- We nonparametrically recover the outcome distributions before selection. The previous literature
  - Nonparametric: recovers only (difference between) the mean of outcome (Das, Newey, Vella 2003).
  - Parametric: recovers the distribution (Arellano & Bonhomme 2017)
- We achieve this by constructing an operator whose fixed point is the potential outcome distributions, and showing that it is a functional contraction.
- We also propose an estimator for the parameters in the selection model. Establish consistency and asymptotic normality.
- We make use of all the information: the entire selected distribution, rather than first and second moments.

This Paper

- We nonparametrically recover the outcome distributions before selection. The previous literature
  - Nonparametric: recovers only (difference between) the mean of outcome (Das, Newey, Vella 2003).
  - Parametric: recovers the distribution (Arellano & Bonhomme 2017)
- We achieve this by constructing an operator whose fixed point is the potential outcome distributions, and showing that it is a functional contraction.
- We also propose an estimator for the parameters in the selection model. Establish consistency and asymptotic normality.
- We make use of all the information: the entire selected distribution, rather than first and second moments.
- No instruments needed to shift the choice probabilities.

A Discrete Choice Problem

- A finite set of alternatives $\mathcal{J} = \{1, \cdots, J\}$.
- Let $\pi_j^o \in \Delta([\underline{p}_j, \overline{p}_j])$ represent the offered price density associated with alternative j.
- All analyses are conditional on characteristics x, x^* which we omit.
- p_j are independent conditional on x, x^*.
  - In auto-insurance application, x^* also includes driver's true risk (estimable ex-post).
- The *offered* price distribution: $\pi^o = \prod_{j \in \mathcal{J}} \pi_j^o$

Selection Function

- A *selection function* $f = (f_1, f_2, \cdots, f_J)$, where f_j maps a price vector $\boldsymbol{p} = (p_1, \cdots, p_J)$ to a strictly positive probability of selecting j:

$$f_j \in \mathscr{C}^1 \colon \prod_j [\underline{p}_j, \overline{p}_j] \to (0, 1).$$

- $\sum_{j \in \mathcal{J}} f_j \leq 1$. The inequality allows for an outside option.

A Simple Example

- Two alternatives $j \in \{1, 2\}$. An outcome equation:

$$p_j = x'\beta_j + \eta_j(x^*), \quad p_1 \perp p_2 | (x, x^*) \quad \text{(outcome conditionally independent)}.$$

- Utility for each alternative:

$$u_j = \gamma p_j + x^* \kappa_j + \varepsilon_j.$$

- The selection function:

$$f_1(p_1, p_2, x, x^*) = \int 1\{u_1 \geq u_2\} d\mu(\varepsilon_1, \varepsilon_2).$$

- Compared to the "reduced-form" selection model:

$$u_j = \gamma(\underbrace{x'\beta_j + \eta_j(x^*)}_{p_j}) + x^*\kappa_j + \varepsilon_j = x'\gamma\beta_j + (\underbrace{\gamma\eta_j(x^*) + x^*\kappa_j + \varepsilon_j}_{\text{composite error}}).$$

The Selected Price Distribution

- The probability of selecting j conditional on p_j

$$Pr_j(p_j;\pi^o) = \int_{\boldsymbol{p}_{-j}} f_j(p_j,\boldsymbol{p}_{-j}) \prod_{k\neq j} \pi_k^o(p_k)dp_k,$$

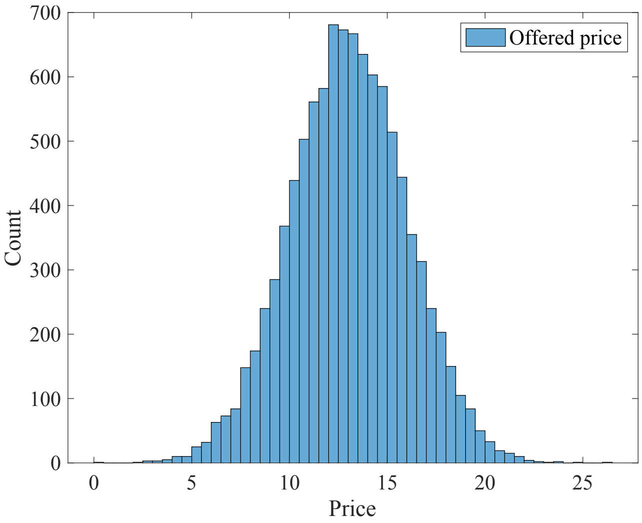
- Let $\pi_j^s \in \Delta([\underline{p}_j,\overline{p}_j])$ denote the j's selected price distribution.

$$\pi_j^s(p_j) \propto \pi_j^o(p_j) \times Pr_j(p_j;\pi^o).$$

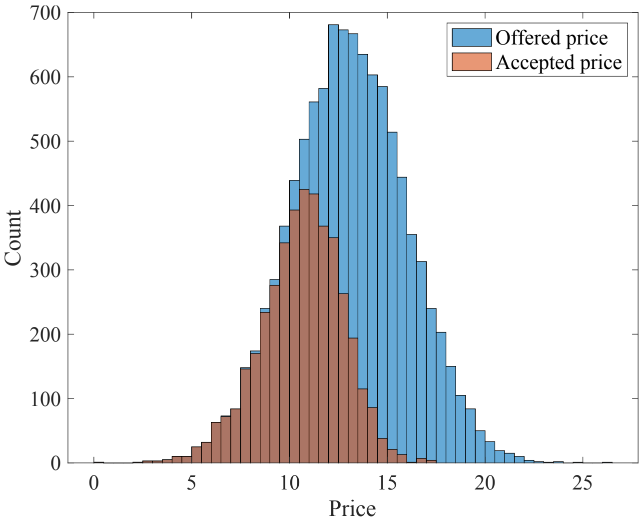
  $f_j > 0$, π_j^o and π_j^s share the same support $[\underline{p}_j,\overline{p}_j]$.

- We call $\pi^s = \prod_{j\in\mathcal{J}} \pi_j^s$ the *selected* price distribution.

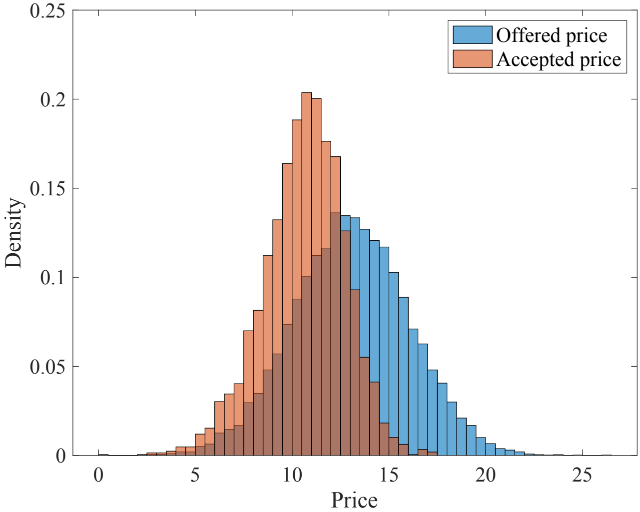
Graphical Illustration: $\pi_j^o(p_j)$



Graphical Illustration: $\pi_j^o(p_j) \times Pr_j(p_j; \pi^o)$



Graphical Illustration: $\frac{\pi_j^o(p_j)\times Pr_j(p_j;\pi^o)}{\int \pi_j^o(p)\times Pr_j(p;\pi^o)dp}$.



Key Challenge

- In many empirical settings, we only observe the selected price distribution π^s.
- How to recover the offered price distribution π^o from the selected π^s?
- Both π^o and π^s are collections of J probability density functions.
- Given a selection function, the cardinality of unknowns and constraints are exactly the same.
- The key challenge is solving for a collection of infinite-dimensional objects entangled in a nonlinear system.

Roadmap

1. Model
2. Main Theoretical Results
3. Estimation
4. Monte Carlo Simulations

How to Recover Offered Price Distribution

- The probability of selecting j conditional on p_j,

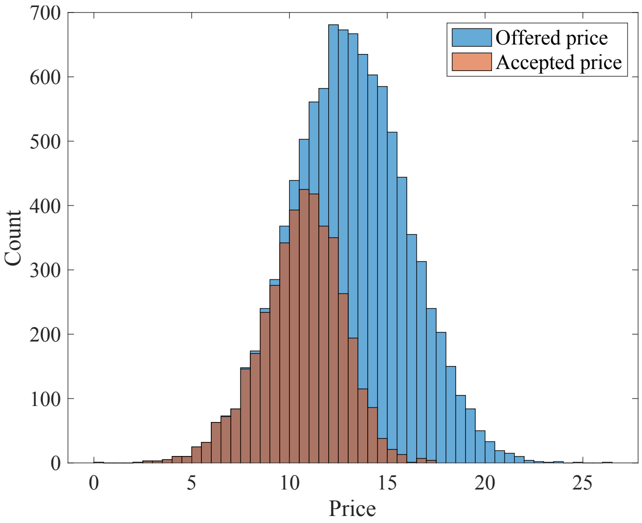
$$Pr_j(p_j;\pi^o) = \int_{\boldsymbol{p}_{-j}} f_j(p_j,\boldsymbol{p}_{-j}) \prod_{k\neq j} \pi_k^o(p_k) dp_k,$$

- We observe j's selected price distribution $\quad \pi_j^s(p_j) \propto \pi_j^o(p_j) \times Pr_j(p_j;\pi^o)$.
- Given $Pr_j(p_j;\pi^o)$, $\quad \pi_j^o(p_j) \propto \pi_j^s(p_j)/Pr_j(p_j;\pi^o)$.
- Plug in a conjecture $\psi = (\psi_1,\psi_2,\cdots,\psi_J) \in \prod_j \Delta([\underline{p}_j,\overline{p}_j])$ to obtain $Pr_j(p_j;\psi)$.
- Given $Pr_j(p_j;\psi)$, a new conjecture about j's offered price distribution $\propto \pi_j^s(p_j)/Pr_j(p_j;\psi)$.
- Define an operator $T\colon \prod_j \Delta([\underline{p}_j,\overline{p}_j]) \to \prod_j \Delta([\underline{p}_j,\overline{p}_j])$.

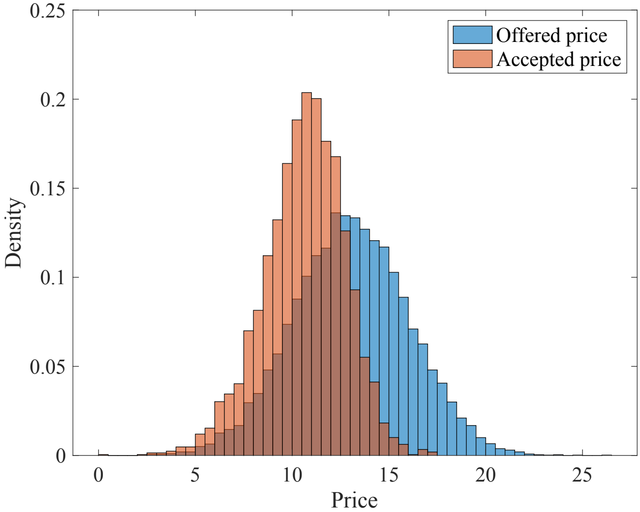
$$(T\psi)_j(p_j) \propto \pi_j^s(p_j)/Pr_j(p_j;\psi),$$

- If the conjecture is correct $\psi = \pi^o$, $T\pi^o = \pi^o$, a fixed point.

Graphical Illustration: $\pi_j^s(p_j)/Pr_j(p_j;\psi)$



Graphical Illustration: $\frac{\pi_j^s(p_j)/Pr_j(p_j;\psi)}{\int \pi_j^s(p)/Pr_j(p;\psi)dp}$



Functional Contraction

- Let

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

where M_j is *maximum semi-elasticity difference*:

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

Theorem 1

If $\rho < 1$, the operator T is a contraction with modulus less than ρ.

- By the Banach fixed point theorem, any selected distribution π^s corresponds to a unique offered price distribution π^o.
- Take any $\psi \in \prod_j \Delta([\underline{p}_j, \overline{p}_j])$,

$$\lim_{n \to \infty} T^n \psi = \pi^o.$$

Small ρ for Logit

- An important and widely-used selection function is the logit model:

$$f_j(p_1, \cdots, p_J) = \frac{\exp(-\gamma p_j + \xi_j)}{\exp(-\gamma p_j + \xi_j) + \sum_{k \neq j} \exp(-\gamma p_k + \xi_k)}$$

- For a large J, when changing p_j, the main variation comes from the numerator, since the denominator is relatively large.
- Plug in the log derivative:

$$\frac{\partial \ln f_j}{\partial p_j} \approx \frac{\partial \ln \exp(-\gamma p_j + \xi_j)}{\partial p_j} = -\gamma$$

$$M_j \approx 0$$

- In our auto-insurance application, price range 1000 euros, $J = 11$, the iteration converges around 5 loops.

Functional Contraction

- Thus far, we have not imposed any structure on the selection function.
- Take the supreme over $\boldsymbol{p}_{-j}, \boldsymbol{p}'_{-j}$ to compute M_j

Assumption 1 (Log Supermodularity)

For all $j \in \mathcal{J}$ and $p_j \in [\underline{p}_j, \overline{p}_j]$, $\frac{\partial \ln f_j(p_j, \boldsymbol{p}_{-j})}{\partial p_j}$ is weakly increasing in each p_k with $k \neq j$.

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

Theorem 2

Suppose that Assumption 1 holds. If $\rho^* < 1$, the operator T is a contraction with modulus less than ρ^*.

Special Cases

- The logit model satisfies Assumption 1 and Theorem 2 applies.

- Theorem 2 also applies to the probit model with two alternatives:

$$f_j(p_j, p_k) = Pr(e_j - e_k > \gamma p_j - \gamma p_k), \quad (e_j, e_k) \sim N(0, \Sigma).$$

  but not for more alternatives.

Roadmap

1. Model
2. Main Theoretical Results
3. Estimation
4. Monte Carlo Simulations

Data and Primitives

- For each individual i, we observe:
  - choice
  - selected price
  - characteristics: x_{ij}, $x_i = (x'_{i1}, \cdots, x'_{iJ})' \in X$.
- **Primitives of interest**:
  - the offered price distribution π^o.
  - the selection function f.

Estimation with a Known Selection Function

- Suppose f is known:
  - **Step 1**: Obtain a nonparametric estimate of the selected price distribution, $\hat{\pi}^s(x)$.
  - **Step 2**: Plug in $\hat{\pi}^s$ into the operator T, initiate the iteration with $\psi \in \prod_j \Delta([\underline{p}_j, \overline{p}_j])$, keep iterating until convergence.
- We obtain an estimator for the offered price distribution π^o, denoted by $T^m\psi$.
- As the sample size goes large, $\hat{\pi}^s \xrightarrow{p} \pi^s$, does the recovered offered price distribution converges to the true offered price distribution?

Sampling Error

F denotes the mapping from π^o to π^s, $F(\pi^o) = \pi^s$

Proposition 1

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

- The inverse F^{-1} is well-defined and $\pi^o = F^{-1}(\pi^s)$.
- Since F^{-1} is continuous, we have

$$F^{-1}(\hat{\pi}^s) \xrightarrow{p} F^{-1}(\pi^s) \quad \text{as} \quad \hat{\pi}^s \xrightarrow{p} \pi^s.$$

- F^{-1} is Lipschitz continuous, $F^{-1}(\hat{\pi}^s)$ converges to π^o at the same rate as $\hat{\pi}^s$ converges to π^s.

Estimation with an Unknown Selection Function

- The selection function f is parameterized by θ, which comes from a multinomial choice model with an indirect utility given by

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

  The observed choice y_{ij} is defined as:

  $$y_{ij} = 1\{u_{ij} > u_{ik}, \forall k \neq j\}.$$

- For example, a widely used linear specification:

  $$u_{ij} = -\gamma p_{ij} + x_{ij}'\beta + \xi_j + \varepsilon_{ij},$$

  where $\theta = (\gamma, \beta, \boldsymbol{\xi})$, the vector of unobserved error terms $\varepsilon_i = (\varepsilon_{i1}, \cdots, \varepsilon_{iJ})$ is jointly distributed according to a known distribution.

Estimation Steps

- **Step 1**: Obtain a nonparametric estimate of the selected price distribution, $\hat{\pi}^s(x)$.

Estimation Steps

- **Step 1**: Obtain a nonparametric estimate of the selected price distribution, $\hat{\pi^s}(x)$.
- **Step 2**: For any θ, plugging $\hat{\pi^s}$ into the iterative procedure to obtain an estimate of the offered price distribution, $T^m_{\hat{\pi^s}(x),\theta}\psi$.

Estimation Steps

- **Step 1**: Obtain a nonparametric estimate of the selected price distribution, $\hat{\pi}^s(x)$.
- **Step 2**: For any θ, plugging $\hat{\pi}^s$ into the iterative procedure to obtain an estimate of the offered price distribution, $T^m_{\hat{\pi}^s(x),\theta}\psi$.
- **Step 3**: Construct the choice probability given θ and the offered price distribution:

$$Prob_j(x;\theta,\hat{\pi}^s,m) = \int_{\boldsymbol{p}} f_j(\boldsymbol{p};x,\theta)\, d\big(T^m_{\hat{\pi}^s(x),\theta}\psi\big)(\boldsymbol{p}).$$

Estimation Steps

- **Step 1**: Obtain a nonparametric estimate of the selected price distribution, $\hat{\pi}^s(x)$.
- **Step 2**: For any θ, plugging $\hat{\pi}^s$ into the iterative procedure to obtain an estimate of the offered price distribution, $T^m_{\hat{\pi}^s(x),\theta}\psi$.
- **Step 3**: Construct the choice probability given θ and the offered price distribution:

$$Prob_j(x;\theta,\hat{\pi}^s,m) = \int_{\boldsymbol{p}} f_j(\boldsymbol{p};x,\theta)d\left(T^m_{\hat{\pi}^s(x),\theta}\psi\right)(\boldsymbol{p}).$$

- **Step 4**: Estimating θ by matching the choice probabilities with the market share:

$$\hat{\theta} = \arg\max_{\theta} \hat{Q}_n(\theta) = \arg\max_{\theta} \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{J} y_{ij} \ln Prob_j(x_i;\theta,\hat{\pi}^s,m(n)).$$

Estimation Steps

- **Step 1**: Obtain a nonparametric estimate of the selected price distribution, $\hat{\pi^s}(x)$.
- **Step 2**: For any θ, plugging $\hat{\pi^s}$ into the iterative procedure to obtain an estimate of the offered price distribution, $T^m_{\hat{\pi^s}(x),\theta}\psi$.
- **Step 3**: Construct the choice probability given θ and the offered price distribution:

$$Prob_j(x;\theta,\hat{\pi^s},m) = \int_{\boldsymbol{p}} f_j(\boldsymbol{p};x,\theta)d\left(T^m_{\hat{\pi^s}(x),\theta}\psi\right)(\boldsymbol{p}).$$

- **Step 4**: Estimating θ by matching the choice probabilities with the market share:

$$\hat{\theta} = \arg\max_{\theta} \hat{Q}_n(\theta) = \arg\max_{\theta} \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{J} y_{ij} \ln Prob_j(x_i;\theta,\hat{\pi^s},m(n)).$$

- **Step 5**: Once $\hat{\theta}$ is obtained, a plug-in estimator for π^o is given by $T^{m(n)}_{\hat{\pi^s},\hat{\theta}}\psi$.

Consistency

Assumption 2

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

Assumption 3 (Identification)

For the true parameter θ_0 *and offered distribution* π^o*, there does no exist* $\theta' \in \Theta$*,* $\theta' \neq \theta_0$*, offered price distributions* $\pi^{o\prime}$ *which generates all the same observable outcomes.*

Theorem 3 (Consistency)

Under Assumptions 2 and 3, $\hat{\theta} \xrightarrow{p} \theta_0$*,* $T^{m(n)}_{\hat{\pi}^s, \hat{\theta}} \psi \xrightarrow{p} \pi^o$*.*

Asymptotic Normality

Assumption 4

(i) supp(π^s) is finite. (ii) θ_0 is in the interior of Θ. (iii) f is twice continuously differentiable in θ. (iv) $\mathbb{E}\nabla_\theta \mathfrak{g}^*(z;\theta_0,\pi^s)$ is nonsingular.

Theorem 4 (Asymptotic Normality)

Suppose that Assumption 2, 3, and 4 hold. Then $\hat{\theta}$ is asymptotically normal and $\sqrt{n}(\hat{\theta}-\theta_0) \xrightarrow{d} \mathcal{N}(0,V)$. $T^{m(n)}_{\hat{\pi}^s,\hat{\theta}}\psi$ converges to π^o in probability at rate $\sqrt{n}$.

Roadmap

1. Model
2. Main Theoretical Results
3. Estimation
4. Monte Carlo Simulations

Setup

- We conduct a Monte Carlo simulation experiment with $J = 2$.
- Utility specification:

$$
\begin{aligned}
u_{i1} &= -\gamma \log(p_{i1}) + \xi_1 + \beta x_{i1} + \varepsilon_i, \quad \varepsilon_i \sim N(0,1) \\
u_{i2} &= -\gamma \log(p_{i2}) + \xi_2,
\end{aligned}
$$

  with ξ_1 normalized to 0.
- x_{i1} is an instrument that exogenously shifts the choice probability, which we do not need.

DGPs for Prices

- **DGP 1**: $\log(p_{ij}) = \delta_{0j} + \delta_j x_{i2} + \eta_{ij}$, Normal errors.
- **DGP 2**: $\log(p_{ij}) = \delta_{0j} + \delta_j x_{i2} + \eta_{ij}$, EV errors.
- **DGP 3**: $\log(p_{ij}) = (\delta_{0j} + \delta_j x_{i2})(1 + \eta_{ij})$, Normal errors with heteroskedasticity.
- **DGP 4**: $\log(p_{ij}) = \exp\left((\delta_{0j} + \delta_j x_{i2})(1 + \eta_{ij})\right)$, nonseparable errors.
- **DGP 5**: $\log(p_{ij}) = (\delta_{0j} + \delta_j x_{i2})(1 + \eta_{ij})^{-1}$, nonseparable errors.

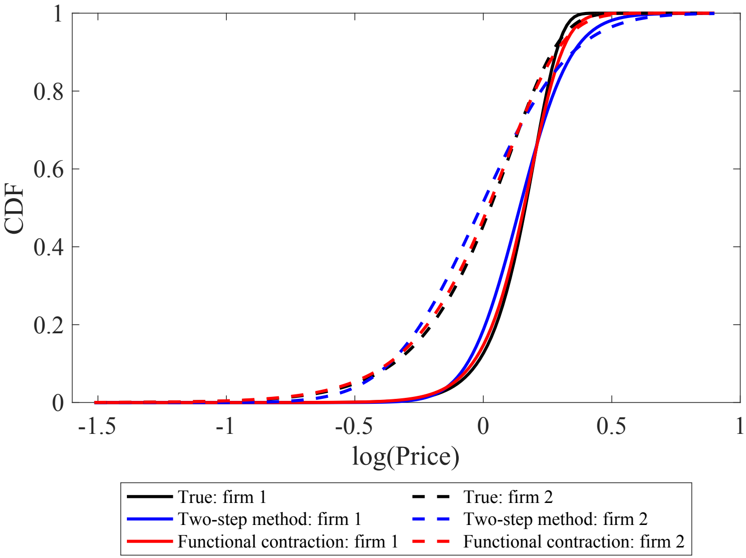
More Details

- Econometricians observe the choice, the price of the chosen alternative, and (x_{i1}, x_{i2}).
- We estimate $\theta = (\gamma, \xi_2, \beta)$, and nonparametric distributions of prices.
- 500 simulations, $N = 1000$, and 5000.
- We also implement the classic two-step method, assuming DGP 1 is the true DGP.

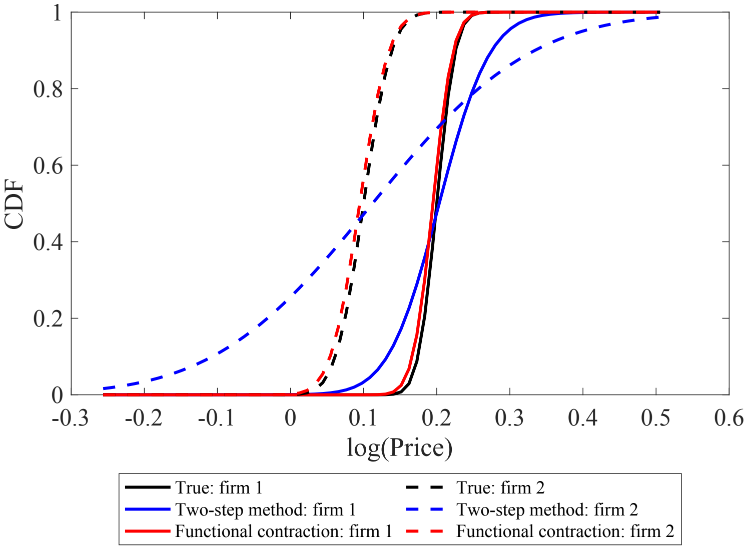
Results

	Functional Contraction			Two-Step Method		
	Bias	Std. Dev.	RMSE	Bias	Std. Dev.	RMSE
DGP 1						
γ	-0.0075	0.1958	0.1957	0.0027	0.2126	0.2124
ξ_2	0.0021	0.0721	0.0721	0.0003	0.0980	0.0979
β	-0.0010	0.0906	0.0905	0.0016	0.0929	0.0929
DGP 2						
γ	-0.0087	0.1990	0.1990	0.0196	0.2451	0.2457
ξ_2	0.0021	0.0728	0.0728	0.0183	0.1203	0.1215
β	-0.0049	0.0945	0.0946	0.0036	0.0960	0.0960
DGP 3						
γ	-0.0254	0.1603	0.1621	0.1704	0.2398	0.2940
ξ_2	-0.0006	0.0702	0.0701	0.0097	0.0860	0.0864
β	-0.0045	0.0930	0.0930	-0.0023	0.0947	0.0946
DGP 4						
γ	-0.0131	0.3485	0.3484	0.0368	0.3826	0.3840
ξ_2	-0.0016	0.0677	0.0676	-0.0044	0.0731	0.0731
β	-0.0045	0.0933	0.0933	-0.0040	0.0941	0.0941
DGP 5						
γ	0.0551	0.9650	0.9656	0.1873	0.7830	0.8044
ξ_2	-0.0023	0.0671	0.0671	0.0047	0.0675	0.0676
β	-0.0050	0.0886	0.0886	-0.0050	0.0885	0.0886

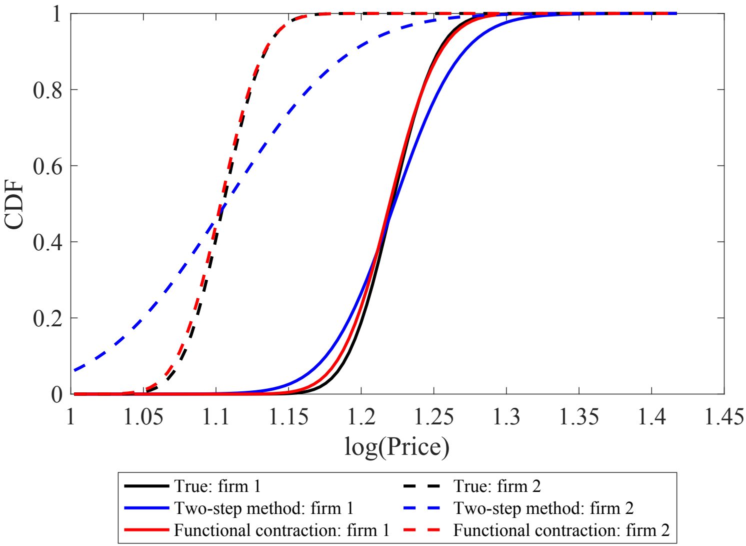
CDFs of log(price): DGP 2



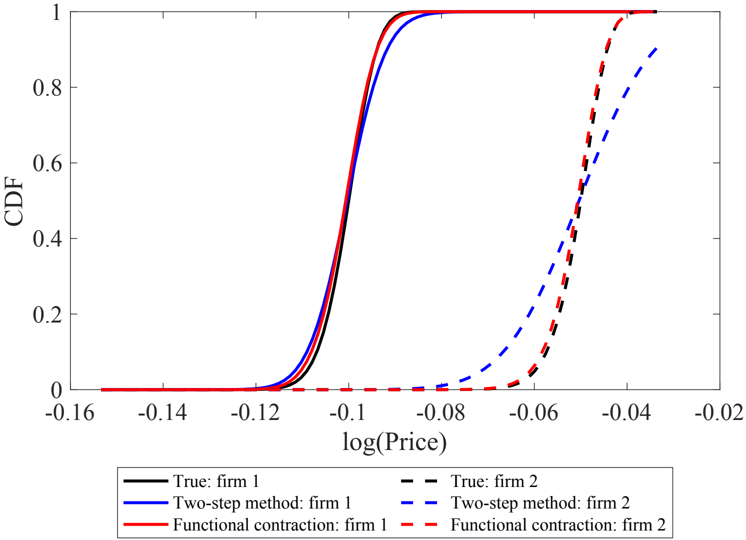
CDFs of log(price): DGP 3



CDFs of log(price): DGP 4



CDFs of log(price): DGP 5



Main Takeaways

- The biases of our estimator are small, and the standard deviation decreases as the sample size increases in all simulation designs.
- Our method allows for nonparametric estimation of the offered price distributions, so it outperforms the standard method in DGPs 2–5.
- For the two-step method, misspecifying the pricing equation also creates a severe bias in estimating the parameters in the selection function (DGPs 3–5).
- Our method does not require an excluded variable in the selection equation (we tried another simulation excluding x_{i1}). Results
- Even if the selection function is misspecified, our method performs well in estimating the price distribution. Results

Conclusions

- We propose a novel method for estimating selection models.
- Our method makes use of all available information on observed outcomes:
  - fully nonparametric estimation of potential outcome distribution, nonseparable in errors.
  - flexibly specified parametric selection rule, typically used in demand estimation, auctions, and labor studies.
  - no need of excluded variables to exogeneously shift the choice probabilities.
- Monte Carlo simulations show that our method performs well in finite samples. The iterative procedure converges very fast, appealing for empirical applications.

Constructing the Metric

- Recall that two probability measures $S, Q \in \Delta(Y)$ are equivalent, denoted $S \sim Q$, if they are absolutely continuous with respect to each other.
- When $S \sim Q$, the Radon-Nikodym derivative, $\frac{dS}{dQ} \colon Y \to (0, \infty)$, exists, as guaranteed by the Radon-Nikodym Theorem.

$$S = Q \quad \Leftrightarrow \quad \frac{dS}{dQ}(y) = 1 \qquad Q\text{-a.e.}$$

- If both S and Q have continuous densities, the Radon-Nikodym derivative is simply the ratio between densities, i.e., $\frac{dS}{dQ}(y) = \frac{s(y)}{q(y)}$.
- In space $\Delta(Y)$, we define a metric $d \colon \Delta(Y) \times \Delta(Y) \to [0, +\infty]$.

$$d(S, Q) = \begin{cases} \ln \operatorname{ess\,sup}_{y \in Y} \frac{dS}{dQ}(y) + \ln \operatorname{ess\,sup}_{y \in Y} \frac{dQ}{dS}(y), & \text{if } S \sim Q, \\ +\infty, & \text{otherwise} \end{cases}$$

Back

Removing the Excluded Variable

	$N = 1000$			$N = 5000$		
	Bias	Std. Dev.	RMSE	Bias	Std. Dev.	RMSE
			DGP 1			
γ	-0.0011	0.2082	0.2080	0.0018	0.0866	0.0865
ξ_2	0.0074	0.0570	0.0574	0.0000	0.0254	0.0253
			DGP 2			
γ	-0.0018	0.2066	0.2064	0.0024	0.1000	0.0999
ξ_2	0.0033	0.0535	0.0535	0.0028	0.0264	0.0265
			DGP 3			
γ	-0.0163	0.1581	0.1587	-0.0043	0.0728	0.0729
ξ_2	0.0061	0.0542	0.0544	0.0007	0.0238	0.0238
			DGP 4			
γ	0.0019	0.3660	0.3656	0.0059	0.1563	0.1563
ξ_2	0.0050	0.0498	0.0500	-0.0005	0.0225	0.0225
			DGP 5			
γ	0.0000	1.0797	1.0786	-0.0146	0.4409	0.4407
ξ_2	0.0014	0.0531	0.0530	-0.0007	0.0233	0.0233

Back

Removing the Excluded Variable

	$N = 1000$		$N = 5000$	
	IBias^2	IMSE	IBias^2	IMSE
DGP 1				
$F_1(\cdot \mid x_{i2} = 0)$	0.0002	0.0018	0.0002	0.0004
$F_2(\cdot \mid x_{i2} = 0)$	0.0001	0.0003	0.0000	0.0001
$F_1(\cdot \mid x_{i2} = 1)$	0.0003	0.0019	0.0002	0.0005
$F_2(\cdot \mid x_{i2} = 1)$	0.0001	0.0007	0.0001	0.0002
DGP 2				
$F_1(\cdot \mid x_{i2} = 0)$	0.0004	0.0017	0.0003	0.0006
$F_2(\cdot \mid x_{i2} = 0)$	0.0002	0.0004	0.0001	0.0001
$F_1(\cdot \mid x_{i2} = 1)$	0.0005	0.0018	0.0003	0.0006
$F_2(\cdot \mid x_{i2} = 1)$	0.0002	0.0007	0.0001	0.0002
DGP 3				
$F_1(\cdot \mid x_{i2} = 0)$	0.0058	0.0073	0.0061	0.0064
$F_2(\cdot \mid x_{i2} = 0)$	0.0029	0.0031	0.0028	0.0028
$F_1(\cdot \mid x_{i2} = 1)$	0.0006	0.0021	0.0005	0.0008
$F_2(\cdot \mid x_{i2} = 1)$	0.0001	0.0007	0.0000	0.0002
DGP 4				
$F_1(\cdot \mid x_{i2} = 0)$	0.0006	0.0021	0.0006	0.0008
$F_2(\cdot \mid x_{i2} = 0)$	0.0008	0.0010	0.0007	0.0007
$F_1(\cdot \mid x_{i2} = 1)$	0.0004	0.0024	0.0003	0.0007
$F_2(\cdot \mid x_{i2} = 1)$	0.0001	0.0006	0.0000	0.0002
DGP 5				
$F_1(\cdot \mid x_{i2} = 0)$	0.0014	0.0025	0.0013	0.0016
$F_2(\cdot \mid x_{i2} = 0)$	0.0013	0.0015	0.0014	0.0014
$F_1(\cdot \mid x_{i2} = 1)$	0.0007	0.0033	0.0006	0.0012
$F_2(\cdot \mid x_{i2} = 1)$	0.0002	0.0006	0.0001	0.0002

Misspecifying the Selection Function

	$N = 1000$			$N = 5000$		
	Bias	Std. Dev.	RMSE	Bias	Std. Dev.	RMSE
			DGP 1			
γ	-0.0793	0.1826	0.1989	-0.0743	0.0806	0.1096
ξ_2	-0.0754	0.0714	0.1038	-0.0752	0.0342	0.0826
β	-0.0309	0.0856	0.0909	-0.0306	0.0392	0.0497
			DGP 2			
γ	-0.0748	0.1864	0.2007	-0.0667	0.0806	0.1045
ξ_2	-0.0714	0.0727	0.1018	-0.0742	0.0340	0.0816
β	-0.0315	0.0902	0.0954	-0.0282	0.0407	0.0495
			DGP 3			
γ	-0.1051	0.1475	0.1810	-0.0940	0.0650	0.1142
ξ_2	-0.0789	0.0698	0.1053	-0.0768	0.0317	0.0830
β	-0.0323	0.0887	0.0943	-0.0293	0.0398	0.0494
			DGP 4			
γ	-0.0746	0.3249	0.3330	-0.0584	0.1442	0.1554
ξ_2	-0.0776	0.0679	0.1030	-0.0755	0.0307	0.0814
β	-0.0263	0.0900	0.0937	-0.0222	0.0392	0.0450
			DGP 5			
γ	0.0029	0.9169	0.9160	-0.0606	0.4177	0.4217
ξ_2	-0.0827	0.0667	0.1063	-0.0801	0.0302	0.0856
β	-0.0315	0.0847	0.0902	-0.0266	0.0381	0.0465

Back

Misspecifying the Selection Function

	$N = 1000$		$N = 5000$	
	IBias^2	IMSE	IBias^2	IMSE
	DGP 1			
$F_1(\cdot\|x_{i2} = 0)$	0.0004	0.0029	0.0002	0.0007
$F_2(\cdot\|x_{i2} = 0)$	0.0001	0.0006	0.0001	0.0002
$F_1(\cdot\|x_{i2} = 1)$	0.0004	0.0032	0.0002	0.0009
$F_2(\cdot\|x_{i2} = 1)$	0.0002	0.0013	0.0001	0.0003
	DGP 2			
$F_1(\cdot\|x_{i2} = 0)$	0.0006	0.0033	0.0005	0.0010
$F_2(\cdot\|x_{i2} = 0)$	0.0002	0.0006	0.0001	0.0002
$F_1(\cdot\|x_{i2} = 1)$	0.0007	0.0037	0.0004	0.0010
$F_2(\cdot\|x_{i2} = 1)$	0.0003	0.0015	0.0001	0.0004
	DGP 3			
$F_1(\cdot\|x_{i2} = 0)$	0.0062	0.0087	0.0061	0.0066
$F_2(\cdot\|x_{i2} = 0)$	0.0028	0.0032	0.0028	0.0029
$F_1(\cdot\|x_{i2} = 1)$	0.0007	0.0033	0.0005	0.0011
$F_2(\cdot\|x_{i2} = 1)$	0.0002	0.0013	0.0001	0.0003
	DGP 4			
$F_1(\cdot\|x_{i2} = 0)$	0.0008	0.0034	0.0006	0.0012
$F_2(\cdot\|x_{i2} = 0)$	0.0008	0.0012	0.0007	0.0008
$F_1(\cdot\|x_{i2} = 1)$	0.0006	0.0046	0.0003	0.0012
$F_2(\cdot\|x_{i2} = 1)$	0.0002	0.0011	0.0001	0.0003
	DGP 5			
$F_1(\cdot\|x_{i2} = 0)$	0.0014	0.0034	0.0014	0.0019
$F_2(\cdot\|x_{i2} = 0)$	0.0014	0.0018	0.0014	0.0015
$F_1(\cdot\|x_{i2} = 1)$	0.0008	0.0058	0.0007	0.0018
$F_2(\cdot\|x_{i2} = 1)$	0.0002	0.0011	0.0001	0.0003