Demand Estimation with Finitely Many Consumers

Jonas Lieber
University of Chicago

Thomas Wiemann
University of Chicago

June 24, 2023

Introduction

Discrete choice demand model:

- ▷ Consumers choose a product if it maximizes their latent utility
- ▷ T1EV-assumption for latent utility shocks results in logit-model

Estimation w/ endogenous prices (Berry, 1994; Berry et al., 1995):

- ▷ Equate market shares & conditional choice probabilities (CCPs)
- ▷ 1-1 mapping between shares and latent demand shocks
- ▷ NFP estimator based on *nonlinear* demand inversion

But market shares are often estimated from consumer choices

- ▷ Issue 1: Inference may be invalid
- ▷ Issue 2: Demand estimators cannot always be computed

Zero-Market-Share Example: Dominick's Finer Foods Data

Table 1: Selected Product Categories

Category	Average Number of UPC's in a Store/Week Pair	Percent of Total Sale of Top 20% UPC's	Percent of Zero Sales
Beer	179	87.18%	50.45%
Cereals	212	72.08%	27.14%
Dish Detergent	115	69.04%	42.39%
Frozen Juices	94	75.16%	23.54%
Laundry Detergents	200	65.52%	50.46%
Paper Towels	56	83.56%	48.27%
Soft Drinks	537	91.21%	38.54%
Soaps	140	77.26%	44.39%
Toothbrushes	137	73.69%	58.63%
Bathroom Tissues	50	84.06%	28.14%

Notes. Excerpt of Table 2 of Gandhi et al. (2013).

Contribution to Literature

This paper:

- ▷ Views discrepancy between shares and CCPs as a finite-sample issue
- ▷ Uses concentration inequalities to bound deviations of shares/CCPs
- ▷ Develops feasible consistent estimator w/ estimated market shares
- ▷ Provides inference in settings w/ non-negligible sampling error

Literature:

1. Random coefficient logit model estimation w/ endogenous prices: Berry et al. (1995), Dubé et al. (2012), ...
2. Estimation and inference w/ estimated market shares: Berry et al. (2004), Gandhi et al. (2013), Freyberger (2015), ...
3. Zero-market-share problem: Quan and Williams (2018), Dubé et al. (2021), Hortaçsu et al. (2021), Gandhi et al. (2023)

Outline

1. Discrete choice model w/ endogenous prices
2. Issues from estimated market shares
3. Demand estimation w/ finitely many consumers
   - ▷ EZ-MPEC
   - ▷ Inference
4. Simulations (work in progress)

Outline

1. **Discrete choice model w/ endogenous prices**

2. Issues from estimated market shares

3. Demand estimation w/ finitely many consumers
   - ▷ EZ-MPEC
   - ▷ Inference

4. Simulations (work in progress)

Random coefficient logit model

Assumption 1

Consumers choose an alternative from $\mathcal{J} \equiv \{0, \ldots, J\}$ via

$$Y_i = \underset{j \in \mathcal{J}}{\arg\max}\ X_j^\top \beta_i + \xi_j + \varepsilon_{i,j}.$$

Latent utility shocks $\varepsilon_{i,j}$ are i.i.d. T1EV. Customer preference parameters β_i are i.i.d. multivariate normal with parameter $\theta \equiv (\mu, \Sigma) \in \Theta$.

Here:

- $X = (X_1^\top, \ldots, X_J^\top) \equiv$ product characteristics (e.g., prices)
- $\xi = (\xi_1, \ldots, \xi_J) \equiv$ latent demand shocks
- $\varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{iJ}) \equiv$ additively separable latent utility shocks

Endogeneity concern: $E[\xi_j | X_j] \neq 0$

Random coefficient logit model (Contd.)

Integrating over ε_i and β_i results in

$$\Pr(Y_i = j | X, \xi; \theta) = \int \frac{\exp\left(X_j \beta + \xi_j\right)}{1 + \sum_{k=1}^{J} \exp\left(X_k \beta + \xi_k\right)} dF(\beta; \theta), \quad \forall j \qquad (1)$$

Let $\pi(X, \xi; \theta)$ denote $J \times 1$ vector of CCPs

Berry (1994) proves demand inversion:

- ▷ $\forall (X, \theta)$ and $s \in (0,1)^J$, there exists a unique solution to

$$\pi(X, \xi; \theta) - s = 0$$

Denoted: $\xi(s, X; \theta)$

BLP Estimator

Assumption 2

There exists a vector of instruments Z such that

$$E\left[Z^\top \xi\right] = 0.$$

When $\{(S_t, X_t, Z_t)\}_{t=1}^T \overset{iid}{\sim} (\pi(X, \xi; \theta), X, Z)$, this motivates

$$\hat{\theta}_T^{blp} \equiv \arg\min_{\theta \in \Theta} \left\| \sum_{t=1}^{T} Z_t^\top \xi(S_t, X_t; \theta) \right\|_2^2$$

Note: Typically weighted by $W_T^{1/2} = (\frac{1}{T} \sum_{t=1}^T Z_t Z_t^\top)^{-1/2}$

Estimated Market Shares

But: Population-level market shares S_t are seldom observed directly

Assumption 3

Observed market shares are sample averages of n consumer choices:

$$S_j^{(n)} \equiv \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_j(Y_i), \quad \forall j,$$

and $S^{(n)} = (S_1^{(n)}, \ldots, S_J^{(n)})$.

Observables we consider:

Assumption 4

The data is $(S_t^{(n)}, X_t, Z_t) \overset{iid}{\sim} (S^{(n)}, X, Z)$, $\forall t = 1, \ldots, T$.

Note: $E[S_t^{(n)}] = S_t$

Outline

1. Discrete choice model w/ endogenous prices
2. **Issues from estimated market shares**
3. Demand estimation w/ finitely many consumers
   - ▷ EZ-MPEC
   - ▷ Inference
4. Simulations (work in progress)

Issues from Estimated Market Shares

Using $S_t^{(n)}$ rather than S_t in estimation causes two key issues

Subtle issue: Incidental parameter problem

- ▷ Estimation error $S_t^{(n)} - S_t$ does not "cancel out"
- ▷ Freyberger (2015): $\sqrt{T}(\hat{\theta}_T^{blp} - \theta_0)$ with $S_t^{(n)}$ unbounded unless

$$\sqrt{T} = o(n)$$

- ▷ # consumers must grow sufficiently quickly relative to # markets

Salient issue: Zero-market-share problem

- ▷ Demand inversion only defined for $s \in (0,1)^J$ but $\operatorname{supp} S_t^{(n)} = [0,1]^J$
- ▷ $\hat{\theta}_T^{blp}$ *cannot* be computed when $\exists j, t : S_{j,t}^{(n)} = 0$

Illustration: Infeasible MPEC Estimation

MPEC estimator of Dubé et al. (2012) equivalent to $\hat{\theta}_T^{blp}$:

$$\begin{aligned} \min_{(\theta,\xi)} \quad & \left\| \sum_{t=1}^{T} Z_t^\top \xi_t \right\|_2^2 \\ \text{s.t.} \quad & S_{j,t} = \int \frac{\exp\left(X_{j,t}\beta + \xi_{j,t}\right)}{1 + \sum_{k=1}^{J} \exp\left(X_{k,t}\beta + \xi_{k,t}\right)} dF(\beta;\theta), \ \forall t, j \end{aligned} \tag{2}$$

Replacing $S_{j,t}$ with $S_{j,t}^{(n)}$ when $\exists t, j : S_{j,t}^{(n)} = 0$ implies no feasible solution

- ▷ "Zero-market-share" problem

In practice: Ad-hoc data manipulation to force-fit BLP estimator:

- ▷ Remove j, t with $S_{j,t}^{(n)} = 0$ or add $\varepsilon > 0$ to zero-valued $S_{j,t}^{(n)}$
- ▷ When $\sqrt{T} = o(n)$, asymp. inference with ad-hoc solutions is ok

Or: Deviate from BLP (e.g., Dubé et al., 2021; Gandhi et al., 2023)

Outline

1. Discrete choice model w/ endogenous prices
2. Issues from estimated market shares
3. **Demand estimation w/ finitely many consumers**
   - ▹ **EZ-MPEC**
   - ▹ Inference
4. Simulations (work in progress)

MPEC with **E**stimated/**Z**ero-Valued Shares (**EZ**-MPEC)

We propose $\hat{\theta}_T^{ezmpec}$ minimizer to:

$$\begin{aligned} \min_{(\theta,\xi)} \quad & \left\| \sum_{t=1}^{T} Z_t^\top \xi_t \cdot D_{t,n,T}^{(\alpha_{n,T})} \right\|_2^2 \\ \text{s.t.} \quad & S_{j,t}^{(n)} \in C_{n,T}^{j}(X_t, \xi_t; \theta, \alpha_{n,T}),\ \forall t, j. \end{aligned}$$

Here:

- $C_{n,T} \equiv$ closed probabilistic bounds on $S_{j,t}^{(n)} - S_{j,t}$
- $\alpha_{n,T} \equiv$ hyperparameter s.t. $1 - \alpha_{n,T}$ is the uniform coverage rate
- $D_{t,n,T}^{(\alpha_{n,T})} \equiv$ indicators for whether $\{0, 1\}$ is in the confidence band

Proposition 1

Fix $\alpha \in (0,1)$. Then with Hoeffding's inequality, it holds that

$$Pr\left(\exists j,t:\ |S_{j,t}^{(n)} - \pi_j(X_t,\xi_t;\theta_0)| \geq \sqrt{\frac{\log\left(\frac{2J}{1-\sqrt[T]{1-\alpha}}\right)}{2n}}\right) \leq \alpha,$$

$\forall n \in \mathbb{N}_{++}$. With Binomial quantiles, it holds that

$$Pr\Bigg(\exists j,t:\ S_{j,t}^{(n)} \notin \Bigg[\frac{1}{n}F_{\text{Bin}}^{-1}\left(\frac{1-\sqrt[T]{1-\alpha}}{J},\pi_j(X_t,\xi_t;\theta_0),n\right), \frac{1}{n}F_{\text{Bin}}^{-1}\left(1-\frac{1-\sqrt[T]{1-\alpha}}{J},\pi_j(X_t,\xi_t;\theta_0),n\right)\Bigg]\Bigg) \leq \alpha,$$

$\forall n \in \mathbb{N}_{++}$, where F_{Bin}^{-1} denotes the quantile function of the Binomial distribution. For $J=1$, the second inequality holds with equality.

EZ-MPEC Estimation w/ Hoeffding's Bounds

When using Hoeffding's bounds, $\hat{\theta}_T^{ezmpec}$ is

$$\min_{(\theta,\xi)} \quad \left\| \sum_{t=1}^{T} Z_t^\top \xi_t \cdot \left(\mathbb{1}\left\{ S_{j,t}^{(n)} \in (\delta_{n,T}(\alpha_{n,T}),\, 1-\delta_{n,T}(\alpha_{n,T})) \right\} \right)_{j=1}^{J} \right\|_2^2$$

$$\text{s.t.} \quad S_{j,t}^{(n)} - \pi_j(X_t, \xi_t; \theta) \le \delta_{n,T}(\alpha_{n,T}),\ \forall j, t$$

$$\pi_j(X_t, \xi_t; \theta) - S_{j,t}^{(n)} \le \delta_{n,T}(\alpha_{n,T}),\ \forall j, t$$

where

$$\delta_{n,T}(\alpha) \equiv \sqrt{\frac{\log\left(\frac{2J}{1-\sqrt[T]{1-\alpha}}\right)}{2n}}$$

Basically MPEC but w/ inequality constraints

Consistency

In addition assume:

- $\exists \gamma \in (0,1)$ such that $P(\pi(X, \xi; \theta_0) \in [\gamma, 1-\gamma]^J) = 1$
- $\Theta, \operatorname{supp} X, \operatorname{supp} Z$ are compact
- full rank and boundedness of $\frac{1}{T} \sum_{t=1}^{T} Z_t^\top Z_t$
- θ_0 is identified from the moment condition $E[Z^\top \xi] = 0$ details

Theorem 1

If $\alpha_{n,T} \in (0,1) : \alpha_{n,T} = o(1)$ and $\log(T) = o(n)$, then as $n, T \to \infty$,

$$\sup_{\tilde{\theta} \in \Theta^*_{n,T}} \left\| \tilde{\theta} - \theta_0 \right\| \xrightarrow{p} 0,$$

where $\Theta^*_{n,T}$ denotes the $\arg\min$ of the EZ-MPEC estimator with hyperparameter $\alpha_{n,T}$, and θ_0 are the true demand parameters.

Outline

1. Discrete choice model w/ endogenous prices
2. Issues from estimated market shares
3. **Demand estimation w/ finitely many consumers**
   - ▷ EZ-MPEC
   - ▷ **Inference**
4. Simulations (work in progress)

Inference

Consider hypothesis tests of the form

$$H_0 : \theta_0 = \theta \qquad \text{versus} \qquad H_1 : \theta_0 \neq \theta$$

For this purpose, we propose the test statistic $\widehat{B}_T$ given by

$$\widehat{B}_T(\alpha) \equiv \min_{\{\tilde{\xi}_t\}_{t=1}^T} \quad \left\| \left(\sum_{t=1}^T \hat{\xi}_t^\top Z_t Z_t^\top \hat{\xi}_t \cdot D_{t,n,T}^{(\alpha)} \right)^{-\frac{1}{2}} \left(\sum_{t=1}^T Z_t^\top \tilde{\xi}_t \cdot D_{t,n,T}^{(\alpha)} \right) \right\|_2^2$$

$$\text{s.t.} \quad S_{j,t}^{(n)} \in C_{n,T}^j(X_t, \tilde{\xi}_t; \theta, \alpha), \ \forall t, j$$

Here, $(\hat{\xi}_t)_{t=1}^T$ are consistent first-step estimates

▷ E.g., from EZ-MPEC or the ad-hoc BLP estimator

Inference (Contd.)

Theorem 2

Let the assumptions of Theorem 1 hold. Then, under H_0, $\forall \tau, \alpha \in (0, \frac{1}{2})$,

$$\limsup_{n,T\to\infty} E\left[\mathbb{1}\{\widehat{B}_T(\alpha) > c_K^{1-\tau}\}\right] \leq \tau(1-\alpha) + \alpha,$$

where $c_K^{1-\tau}$ is the $1-\tau$ quantile of a χ^2 distribution with K d.o.f.

Two alternatives for test at significance level $\tilde{\tau}$:

- ▷ Take $\tau, \alpha : \tau(1-\alpha) + \alpha = \tilde{\tau}$
- ▷ Suppose $\alpha = \alpha_{n,T} = o(1)$ and take $\tau = \tilde{\tau}$

Importantly: Valid inference when $\log(T) = o(n)$

- ▷ Improvement over $\sqrt{T} = o(n)$ of $\hat{\theta}_T^{blp}$ with $S_t^{(n)}$ (Freyberger, 2015)
- ▷ Allows for non-negligible sampling error

Confidence intervals constructed via test inversion

Subvector Inference

Confidence interval for θ_0 requires grid-search

- ▷ Computationally demanding for many product characteristics

For a function $R : \Theta \to \mathbb{R}^d$ and vector $r \in \mathbb{R}^d$, consider

$$H_0 : R(\theta_0) = r \qquad \text{versus} \qquad H_1 : R^\top \theta_0 \neq r$$

$$\begin{aligned} \tilde{B}_T(\alpha) \equiv \min_{\tilde{\theta}, \{\tilde{\xi}_t\}_{t=1}^T} \quad & \left\| \left(\sum_{t=1}^T \hat{\xi}_t^\top Z_t Z_t^\top \hat{\xi}_t \cdot D_{t,n,T}^{(\alpha)} \right)^{-\frac{1}{2}} \left(\sum_{t=1}^T Z_t^\top \tilde{\xi}_t \cdot D_{t,n,T}^{(\alpha)} \right) \right\|_2^2 \\ \text{s.t.} \quad & S_{j,t}^{(n)} \in C_{n,T}^j(X_t, \tilde{\xi}_t; \tilde{\theta}, \alpha), \ \forall t, j \\ & R(\tilde{\theta}) = r \end{aligned}$$

Then as before: $\limsup_{n,T\to\infty} E\left[\mathbb{1}\{\tilde{B}_T(\alpha) > c_K^{1-\tau}\}\right] \leq \tau + \alpha$

Subvector inference: $H_0 : R^\top \theta_0 = r$

Outline

1. Discrete choice model w/ endogenous prices
2. Issues from estimated market shares
3. Demand estimation w/ finitely many consumers
   - ▷ EZ-MPEC
   - ▷ Inference
4. **Simulations** (work in progress)

Monte Carlo Simulations

Monte Carlo:

- $T = 50$, $J = 5$, n varying
- 3 product characteristics $X_{j,t}$ & prices $p_{j,t}$
- prices $p_{j,t}$ correlated w/ $\xi_{j,t}$
- excluded instruments $Z_{j,t}$ included w/ series expansion

Results based on 1,000 repetitions

Simulation w/o Random Coefficients

First: β_i fixed.

- $\hat{\theta}_T^{blp}$ simplifies to TSLS w/ outcome $\log(S_{j,t})/log(S_{0,t})$

TSLS w/ $S_t^{(n)}$ is computed w/o observation for which $S_{j,t}^{(n)} = 0$

Table 2: MAE for DGP w/o Random Coefficients

n	TSLS w/ S_t (infeasible) (1)	TSLS w/ $S_t^{(n)}$ (feasible) (2)	EZ-MPEC (feasible) (3)	Share of $S_t^{(n)} = 0$ (4)
1000	0.088	0.260	0.185	0.095
2000	0.091	0.191	0.166	0.068
3000	0.084	0.167	0.154	0.054
4000	0.088	0.173	0.150	0.046
5000	0.092	0.150	0.158	0.040

Notes. Results based on 1,000 Monte Carlo simulations. Throughout, $T = 50$ and $J = 5$.

Simulation w/ Random Coefficients

Second: β_i is multivariate normal

MPEC w/ $S_t^{(n)}$ is computed w/o observation for which $S_{j,t}^{(n)} = 0$

Table 3: MAE for DGP w/ Random Coefficients

n	MPEC w/ S_t (infeasible) (1)	MPEC w/ $S_t^{(n)}$ (feasible) (2)	EZ-MPEC (feasible) (3)	Share of $S_t^{(n)} = 0$ (4)
1000	0.148	0.263	0.250	0.061
2000	0.151	0.222	0.205	0.040
3000	0.142	0.213	0.189	0.031
4000	0.146	0.199	0.178	0.026
5000	0.134	0.199	0.171	0.023

Notes. Results based on 1,000 Monte Carlo simulations. Throughout, $T = 50$ and $J = 5$.

Conclusion

This paper:

- ▷ Develops feasible consistent estimator for BLP model
- ▷ Provides valid inference in settings w/ non-negligble sampling error

Most relevant in settings with:

- ▷ Relatively few consumers per market
- ▷ Zero-valued market shares

Work in progress:

- ▷ More extensive simulation study
- ▷ Empirical application

References I

Berry, S., Levinsohn, J., and Pakes, A. (1995). Automobile Prices in Market Equilibrium. *Econometrica: Journal of the Econometric Society*, pages 841–890.

Berry, S., Linton, O. B., and Pakes, A. (2004). Limit theorems for estimating the parameters of differentiated product demand systems. *The Review of Economic Studies*, 71(3):613–654.

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

Dubé, J.-P., Fox, J. T., and Su, C.-L. (2012). Improving the numerical performance of static and dynamic aggregate discrete choice random coefficients demand estimation. *Econometrica*, 80(5):2231–2267.

Dubé, J.-P., Hortaçsu, A., and Joo, J. (2021). Random-coefficients logit demand estimation with zero-valued market shares. *Marketing Science*.

Freyberger, J. (2015). Asymptotic theory for differentiated products demand models with many markets. *Journal of Econometrics*, 185(1):162–181.

References II

Gandhi, A., Lu, Z., and Shi, X. (2013). Estimating Demand for Differentiated Products with Error in Market Shares.

Gandhi, A., Lu, Z., and Shi, X. (2023). Estimating demand for differentiated products with zeroes in market share data. *Quantitative Economics*, 14(2):381–418.

Hortaçsu, A., Natan, O. R., Parsley, H., Schwieg, T., and Williams, K. R. (2021). Incorporating search and sales information in demand estimation. Technical report, National Bureau of Economic Research.

Quan, T. W. and Williams, K. R. (2018). Product variety, across-market demand heterogeneity, and the value of online retail. *The RAND Journal of Economics*, 49(4):877–913.

Let G_T denote the (infeasible) objective function of $\hat{\theta}_T^{blp}$:

$$G_T(\theta) = \left(\frac{1}{T} \sum_{t=1}^{T} Z_t^\top \xi(S_t, X_t, \theta) \right)^\top W_T \left(\frac{1}{T} \sum_{t=1}^{T} Z_t^\top \xi(S_t, X_t, \theta) \right)$$

Assumption 5

$\forall \delta > 0$, $\exists M(\delta) > 0$, such that

$$\lim_{T \to \infty} \Pr \left(\inf_{\theta \in \Theta : \|\theta - \theta_0\| > \delta} \| G_T(\theta) - G_T(\theta_0) \| \geq M(\delta) \right) = 1.$$