Estimation of Pure Characteristics Demand Models with Pricing

Che-Lin Su

The University of Chicago
Booth School of Business

joint work with Jong-Shi Pang and Yu-Ching Lee

London Business School
January 19, 2015

Roadmap of the Talk

- Research Agenda
- Overview of Demand Estimation
  - Multinomial logic demand models by McFadden (1973, 1978)
  - Random-coefficients logic demand model by Berry, Levinsohn and Pakes (1995)
  - Pure characteristics demand model by Berry and Pakes (2007)
- MPCC Formulation for Estimating Pure Characteristics Models
  - Existence of a feasible solution
  - Existence of an optimal solution
- Profit Optimization Problem of a Firm
- Numerical Results

Research Agenda

- Methodology Development
  - Constrained Optimization Approaches to Estimation of Structural Models. Su and Judd (2012). **Econometrica**.
  - Improving the Numerical Performance of BLP Static and Dynamic Discrete Choice Random Coefficients Demand Estimation. Dubé, Fox and Su (2012). **Econometrica**.
  - Estimating Discrete-Choice Games of Incomplete Information: Simple Static Examples. Su (2014). **Quantitative Marketing and Economics**.
  - Estimating Dynamic Discrete-Choice Games of Incomplete Information. Egesdal, Lai and Su (2014). **Quantitative Economics**.
  - Estimation of Pure Characteristics Demand Models with Pricing. Pang, Su and Lee (2014). Under third round review at **Operations Research** (minor revision).

Research Agenda

- Empirical Applications
  - Structural Estimation of Callers' Delay Sensitivity in Call Centers. Aksin-Karaesmen, Ata, Emadi and Su (2013). **Management Science**.
  - Impact of Delay Announcements in Call Centers: An Empirical Approach. Aksin-Karaesmen, Ata, Emadi and Su (2014). Under revision for the third round review at **Operations Research**.
  - Real Options Models for Investment Decision on Small Hydropower Plants Using Project Level Data. Fleten, Paarsch and Su.
  - Empirical Counterpart of Impulse Control Models. Keppo and Su.

Demand Estimation

- An important research area in empirical industrial organization, quantitative marketing, and revenue management
- Useful for measuring marketing power of firms, consumer welfare, optimal pricing of firms, etc, in antitrust and merger analysis
- **Goal:** Infer consumers' preferences from observed data
  - Observed data: product characteristics (price, brand, size, quality, etc), individual consumer's purchase decisions or market shares of products
  - Consumers' preferences on observed product characteristics (price, brand, size, quality, etc)
- Model: need a model to map from product characteristics and consumers' preferences to consumers' purchase decisions (or market shares of products)
- Estimation: To achieve this goal, one needs to solve an optimization problem

Part I

Multi-nomial Logit Demand Models

Multi-nomial Logit Model – McFadden (1973, 1978)

- For simplicity, assume only one market
- Product: $j = 1, \dots, J$
  - $x_j \in R^K$, $p_j \in R$: observed product characteristics and price
  - no-purchase option: $j = 0$
- Consumer: $i = 1, \dots, I$
  - homogeneous consumer preferences over (x_j, p_j): $(\beta^x, \beta^p) \in R^{K+1}$
- Consumer i's utility from purchasing product j

$$\begin{aligned} u_{ij} &\triangleq x_j^T \beta^x - \beta^p p_j + \varepsilon_{ij} \\ u_{i0} &\triangleq \varepsilon_{i0} \end{aligned}$$

  - ε_{ij}: i.i.d error term with type-I extreme value distribution
  - Consumers are differentiated by the random shock $\varepsilon_{i,j}$

Multi-nomial Logit Model – McFadden (1973, 1978)

- Discrete choice: each consumer purchases one and only one product (including no-purchase option)
- Decision rule: a consumer purchases the product that gives her the highest utility

$$d_{ij} = 1 \quad \text{if } u_{ij} \geq u_{ij'}, \quad \forall j' \neq j;$$

otherwise, $d_{ij} = 0$

- Choice probability: Given (x, p) and $\Theta \triangleq (\beta^x; \beta^p)$,

$$\mathbf{Pr}(d_{ij} = 1 | \Theta) = \frac{\exp\left(x_j^T \beta^x - \beta^p p_j\right)}{1 + \sum_{j'=1}^{J} \exp\left(x_{j'}^T \beta^x - \beta^p p_{j'}\right)}$$

Multi-nomial Logit Model – McFadden (1973, 1978)

- Observed data: $Z = \{x_j, p_j, d_{ij}\}_{i,j}$
  - Observed product characteristics and price: x and p
  - Observed individual purchased decision: $\{d_{ij}\}_{i,j}$ with $d_{ij} \in \{0, 1\}$
- **Statistical goal**: estimate $\Theta \triangleq (\beta^x; \beta^p)$ from observed data Z via maximum likelihood estimation

$$\max_{\Theta} L(\Theta; Z) = \sum_{i=1}^{I} \sum_{j=1}^{J} d_{ij} * \log[\mathbf{Pr}(d_{ij} = 1 | \Theta)],$$

where $\mathbf{Pr}(d_{ij} = 1 | \Theta) = \dfrac{\exp\left(x_j^T \beta^x - \beta^p p_j\right)}{1 + \sum_{j'=1}^{J} \exp\left(x_{j'}^T \beta^x - \beta^p p_{j'}\right)}$

- Applications in OM: Hansen and Martin (1996), Talluri and van Ryzin (2004), Vulcano, van Ryzin and Ratliff (2012), Li, Rusmevichientong and Topaloglu (2014), Gallego and Topalopglu (2014)

Part II

Random-Coefficients Demand Estimation

Three Main Concerns with the MNL Models

- Homogeneous consumer preferences
  - In reality, consumers have heterogeneous preferences over product characteristics and price
- Need to observe each individual's purchase decision in the data
  - Often we observe only aggregate market share data
- Logit error term implies that each product has a positive market share

Berry, Levinsohn and Pakes (BLP, 1995) propose a model and a method that address the first two concerns

Random-Coefficients Logit Demand: BLP (1995)

- Berry, Levinsohn and Pakes (BLP, 1995) consists of an economic model and a GMM estimator
- Demand estimation with a large number of differentiated products
  - characteristics approach
  - applicable when only aggregate market share data available
  - flexible substitution patterns / price elasticities
  - control for price endogeneity
- Computational algorithm to construct moment conditions from a non-linear model
- Useful for measuring market power, welfare, optimal pricing, etc.
- Used extensively in empirical IO and marketing: Nevo (2001), Petrin (2002), Dubé (2003–2009), etc.

Random-Coefficients Logit Demand

- Utility of consumer i from purchasing product j

$$
\begin{aligned}
u_{ij} &= \beta_i^0 + x_j^T \beta_i^x - \beta_i^p p_j + \xi_j + \varepsilon_{ij} \\
u_{i0} &= 0 + \varepsilon_{i0}
\end{aligned}
$$

- product characteristics: $x_j \in R^K$, $p_j \in R$, $\xi_j \in R$
  - x_j, p_j observed; $cov(\xi_j, p_j) \neq 0$
  - ξ_j: not observed – not in data
- β_i: random coefficients/individual-specific taste to be estimated
  - Distribution: $\beta_i \sim F_\beta(\beta; \Theta)$
  - BLP's statistical goal: estimate Θ in parametric distribution
- error term ε_{ij}: Type I E.V. shock (i.e., Logit)
- Consumer i picks product j if $u_{ij} \geq u_{ij'}, \quad \forall j' \neq j$
- Observe market share S_j

Market Share Equations

- Predicted market shares

$$s_j(x, p, \xi, ; \Theta) = \int_{\left\{\beta_i, \varepsilon_j | u_{ij} \geq u_{ij'}, \forall j' \neq j\right\}} dF_\beta(\beta; \Theta) dF_\varepsilon(\varepsilon)$$

- With logit errors ε

$$s_j(x, p, \xi, ; \Theta) = \int_\beta \frac{\exp(\beta^0 + x_j^T \beta^x - \beta^p p_j + \xi_j)}{1 + \sum_{k=1}^J \exp(\beta^0 + x_k^T \beta^x - \beta^p p_k + \xi_k)} dF_\beta(\beta; \Theta)$$

- Approximate market shares by sample average approximation

$$\hat{s}_j(x, p, \xi, ; \Theta) = \frac{1}{ns} \sum_{r=1}^{ns} \frac{\exp(\beta^{0r} + x_j^T \beta^{xr} - \beta^{pr} p_j + \xi_j)}{1 + \sum_{k=1}^J \exp(\beta^{0r} + x_k^T \beta^{xr} - \beta^{pr} p_k + \xi_k)}$$

- Sample average approximation of the market share equations

$$\hat{s}_j(x, p, \xi, ; \Theta) = S_j, \; \forall j \in J$$

Random-Coefficients Logit Demand: GMM Estimator

- Assume $E\left[\xi_j z_j\right] = 0$ for some vector of instruments z_j
  - Empirical analog $g\left(\Theta\right) = \frac{1}{J}\sum_j \xi_j(\Theta)^T z_j$
- Data: $\{(x_j, p_j, S_j, z_j)_{j \in J}\}$
- Minimize GMM objective function

$$Q(\Theta) = g\left(\Theta\right)^T W g\left(\Theta\right)$$

- Cannot compute $\xi_j(\Theta)$ analytically
  - "Invert" ξ_t from system of predicted market shares numerically

$$\begin{aligned} S &= s\left(x, p, \xi; \Theta\right) \\ \Rightarrow \quad \xi\left(\Theta\right) &= s^{-1}\left(x, p, S; \Theta\right) \end{aligned}$$

- BLP show the inversion of share equations for $\xi(\Theta)$ is a contraction-mapping

BLP/NFXP Estimation Algorithm

- Similar to the Nested-Fixed-Point algorithm in Rust (1987)
- Outer loop: $\min_{\Theta} g(\Theta)^T W g(\Theta)$
  - Guess Θ parameters to compute $g(\Theta) = \frac{1}{J} \sum_{j=1}^{J} \xi_j(\Theta)^T z_j$
  - Stop when $\|\nabla_\Theta (g(\Theta)^T W g(\Theta))\| \leq \epsilon_{\text{out}}$

BLP/NFXP Estimation Algorithm

- Similar to the Nested-Fixed-Point algorithm in Rust (1987)
- Outer loop: $\min_{\Theta} \; g(\Theta)^T W g(\Theta)$
  - Guess Θ parameters to compute $g(\Theta) = \frac{1}{J} \sum_{j=1}^{J} \xi_j(\Theta)^T z_j$
  - Stop when $\|\nabla_\Theta (g(\Theta)^T W g(\Theta))\| \leq \epsilon_{\text{out}}$
- Inner loop: compute $\xi_t(\Theta)$ for a given Θ
  - Solve $s(x, p, \xi; \Theta) = S$ for ξ by contraction mapping:

$$\xi^{h+1} = \xi^h + \log\, S - \log\, s(x, p, \xi; \Theta)$$

  - Stop when $\|\xi^{h+1} - \xi^h\| \leq \epsilon_{\text{in}}$
  - Denote the approximated demand shock by $\xi(\Theta, \epsilon_{\text{in}})$
- Stopping rules: need to choose tolerance/stopping criterion for both inner loop (ϵ_{in}) and outer loop (ϵ_{out})

Constrained Optimization Applied to BLP Estimation

- Su and Judd (2012) present a constrained optimization formulation for general structural estimation models
- Following the same reformulation idea, DFS (2012) give a constrained optimization formulation for random-coefficients demand models:

$$\begin{array}{ll} \min\limits_{(\Theta,\xi)} & \xi^T Z W Z^T \xi \\ \text{subject to} & s(\xi, \Theta) = S \end{array}$$

- Bad news: Hessian of the Lagrangian is dense

Constrained Optimization Applied to BLP Estimation

- Introducing additional variable g and constraint $Z^T\xi = g$

$$\begin{array}{ll} \min\limits_{(\Theta,\xi,g,)} & g^T W g \\ \text{subject to} & s(\delta;\Theta) = S \\ & Z^T\xi = g \end{array}$$

- Advantages:
  - The Hessian of the objective function is now sparse

Monte Carlo in DFS (2012): Simulated Data Setup

- $$\begin{bmatrix} x_{1,j,t} \\ x_{2,j,t} \\ x_{3,j,t} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & -0.8 & 0.3 \\ -0.8 & 1 & 0.3 \\ 0.3 & 0.3 & 1 \end{bmatrix} \right)$$
- $\xi_{j,t} \sim N(0,1)$
- $p_{j,t} = |0.5 \cdot \xi_{j,t} + e_{j,t}| + 1.1 \cdot \left| \sum_{k=1}^{3} x_{k,j,t} \right|$
- $z_{j,t,d} \sim N\left(\frac{1}{4} p_{j,t}, 1\right)$, $D = 6$ instruments
- $F_\beta(\beta; \Theta)$: 5 independent normal distributions ($K = 3$ attributes, price and the intercept)
- $\beta_i = \left\{\beta_i^0, \beta_i^1, \beta_i^2, \beta_i^3, \beta_i^p\right\}$: $E[\beta_i] = \{0.1, 1.5, 1.5, 0.5, -3\}$ and $\mathrm{Var}[\beta_\mathrm{i}] = \{0.5, 0.5, 0.5, 0.5, 0.2\}$

Speeds, # Convergences and Finite-Sample Performance

$T = 50, J = 25, nn = 1000$, 20 replications, 5 starting points/replication

Intercept $E\left[\beta_i^0\right]$	Lipsch. Const	Alg.	CPU (min)	Elasticities			Out. Share
				Bias	RMSE	Value	
-2	0.891	NFP	21.7	-0.077	0.14	-10.4	0.91
		Cons. Opt.	18.3	-0.076	0.14		
-1	0.928	NFP	28.3	-0.078	0.15	-10.5	0.86
		Cons. Opt.	16.3	-0.077	0.15		
0	0.955	NFP	41.7	-0.079	0.16	-10.6	0.79
		Cons. Opt.	15.2	-0.079	0.16		
1	0.974	NFP	71.7	-0.083	0.16	-10.7	0.69
		Cons. Opt.	11.8	-0.083	0.17		
2	0.986	NFP	103.0	-0.085	0.17	-10.8	0.58
		Cons. Opt.	13.5	-0.085	0.17		
3	0.993	NFP	167.0	-0.088	0.17	-11.0	0.46
		Cons. Opt.	10.7	-0.088	0.17		
4	0.997	NFP	300.0	-0.091	0.16	-11.0	0.35
		Cons. Opt.	12.7	-0.090	0.16		

Lessons Learned

- For low Lipschitz constant, NFXP and Constrained Optimization about the same speed
- For high Lipschitz constant, NFXP becomes very slow
    - 1 hour per run for Intercept $= 4$
    - Reminder: you need to use 100 starting points or more if you want to find a good solution
- Cons. Opt. speed relatively invariant to Lipschitz constant
    - No contraction mapping in Cons. Opt.

of Function/Gradient/Hessian Evals and # Contraction Mapping Iterations

Intercept $E\left[\beta_i^0\right]$	Alg.	Func Eval	Grad/Hess Eval	Contraction Iter
-2	NFP	80	58	10,400
	Cons. Opt.	184	126	
-1	NFP	82	60	17,100
	Cons. Opt.	274	144	
0	NFP	77	56	29,200
	Cons. Opt.	195	113	
1	NFP	71	54	55,000
	Cons. Opt.	148	94	
2	NFP	68	50	84,000
	Cons. Opt.	188	107	
3	NFP	68	49	146,000
	Cons. Opt.	144	85	
4	NFP	81	50	262,000
	Cons. Opt.	158	100	

Speed for Varying # of Markets, Products, Draws

T	J	nn	Lipsch. Const.	Alg	Runs	CPU (hrs)	Outside Share
100	25	1000	0.999	NFP Cons. Opt.	80% 100%	10.9 0.3	0.45
250	25	1000	0.997	NFP Cons. Opt.	100% 100%	22.3 1.2	0.71
500	25	1000	0.998	NFP Cons. Opt.	80% 100%	65.6 2.5	0.65
100	25	3000	0.999	NFP Cons. Opt.	80% 100%	42.3 1.0	0.46
250	25	3000	0.997	NFP Cons. Opt.	100% 100%	80.0 3.0	0.71
25	100	1000	0.993	NFP Cons. Opt.	100% 100%	5.7 0.5	0.28
25	250	1000	0.999	NFP Cons. Opt.	100% 100%	28.4 2.3	0.07

of Function/Gradient/Hessian Evals and # Contraction Mapping Iterations

T	J	nn	Alg	# Iter.	Func. Eval.	Grad Eval.	Contrac. Mapping
100	25	1000	NFP Cons. Opt.	68 84	130 98	69 85	372,278
250	25	1000	NFP Cons. Opt.	58 118	82 172	59 119	246,000
500	25	1000	NFP Cons. Opt.	52 123	99 195	53 124	280,980
100	25	3000	NFP Cons. Opt.	60 83	171 114	61 84	479,578
250	25	3000	NFP Cons. Opt.	55 102	68 135	56 103	204,000
25	100	1000	NFP Cons. Opt.	54 97	71 145	55 98	198,114
25	250	1000	NFP Cons. Opt.	60 75	126 103	61 76	359,741

Part III

Pure Characteristics Demand Models

Three Main Concerns with the MNL Models

- Homogeneous consumer preferences
  - In reality, consumers have heterogeneous preferences over product characteristics and price
- Need to observe each individual's purchase decision in the data
  - Often we observe only aggregate market share data
- Logit error term implies that each product has a positive market share

Berry, Levinsohn and Pakes (BLP, 1995) propose a model and a method that address the first two concerns

Berry and Pakes (2007) modify the BLP model to address the third concern

Random-Coefficients Logit Demand in BLP (1995)

- Utility of consumer i from purchasing product j

$$
\begin{aligned}
u_{ij} &= \beta_i^0 + x_j^T \beta_i^x - \beta_i^p p_j + \xi_j + \varepsilon_{ij} \\
u_{i0} &= 0 + \varepsilon_{i0}
\end{aligned}
$$

- product characteristics: $x_j \in R^K$, $p_j \in R$, $\xi_j \in R$
  - x_j, p_j observed; $cov(\xi_j, p_j) \neq 0$
  - ξ_j: not observed – not in data
- β_i: random coefficients/individual-specific taste to be estimated
  - Distribution: $\beta_i \sim F_\beta(\beta; \Theta)$
  - BLP's statistical goal: estimate Θ in parametric distribution
- error term ε_{ij}: Type I E.V. shock (i.e., Logit)
- Consumer i picks product j if $u_{ij} \geq u_{ij'}, \quad \forall j' \neq j$
- Observe market share S_j

Pure Characteristics Models in Berry and Pakes (2007)

- Berry and Pakes (2007) removes the logit error term ε_{ij} from the utility function
- Utility of consumer i from purchasing product j

$$
\begin{aligned}
u_{ij} &= \beta_i^0 + x_j^T \beta_i^x - \beta_i^p p_j + \xi_j \\
u_{i0} &= 0
\end{aligned}
$$

- product characteristics: $x_j \in R^K$, $p_j \in R$, $\xi_j \in R$
  - x_j, p_j observed; $cov(\xi_j, p_j) \neq 0$
  - ξ_j: not observed – not in data
- β_i: random coefficients/individual-specific taste to be estimated
  - Distribution: $\beta_i \sim F_\beta(\beta; \Theta)$
  - BLP's statistical goal: estimate Θ in parametric distribution
- Consumer i picks product j if $u_{ij} \geq u_{ij'}, \quad \forall j' \neq j$
- Observe market share S_j

Market Share Equations in BP Model

- Predicted market shares

$$s_j(x,p,\xi,;\Theta) = \int_{\left\{\beta_i | u_{ij} \geq u_{ij'}, \forall j' \neq j\right\}} dF_\beta(\beta;\Theta)$$
$$= \mathbb{E}\left[\mathbf{1}\left\{u_{ij} \geq u_{ij'}, \forall j' \neq j\right\}\right]$$

- Approximate market shares by sample average approximation

$$\hat{s}_j(x,p,\xi,;\Theta)$$
$$= \frac{1}{ns}\sum_{r=1}^{ns}\mathbf{1}\left\{x_j^T\beta_i^x - \beta_i^p p_j + \xi_j \geq \max_{\ell=1,\ldots,J}\{x_\ell^T\beta_i^x - \beta_i^p p_\ell + \xi_\ell, 0\}\right\},$$

- Sample average approximation of the market share equations

$$\hat{s}_j(x,p,\xi,;\Theta) = S_j,\ \forall j \in J$$

GMM Estimator

- Assume $E\left[\xi_j z_j\right] = 0$ for some vector of instruments z_j
  - Empirical analog $g\left(\Theta\right) = \frac{1}{J}\sum_j \xi_j(\Theta)^T z_j$
- Data: $\{(x_j, p_j, S_j, z_j)_{j \in J}\}$
- Minimize GMM objective function

$$Q(\Theta) = g\left(\Theta\right)^T W g\left(\Theta\right)$$

- Cannot compute $\xi_j(\Theta)$ analytically
  - "Invert" ξ from system of predicted market shares numerically

$$\begin{array}{rcl} S & = & s\left(x, p, \xi; \Theta\right) \\ \Rightarrow \quad \xi\left(\Theta\right) & = & s^{-1}\left(x, p, S; \Theta\right) \end{array}$$

- BP (2007) show the inversion of share equations for $\xi(\Theta)$ exits under some conditions but the inversion is NOT a contraction-mapping

BP/NFXP Estimation Algorithm

- Outer loop: $\min_{\Theta} g(\Theta)^T W g(\Theta)$
  - Guess Θ parameters to compute $g(\Theta) = \frac{1}{J} \sum_{j=1}^{J} \xi_j(\Theta)^T z_j$
  - Stop when $\|\nabla_{\Theta}(g(\Theta)^T W g(\Theta))\| \leq \epsilon_{\text{out}}$

BP/NFXP Estimation Algorithm

- Outer loop: $\min_{\Theta} g(\Theta)^T W g(\Theta)$
  - Guess Θ parameters to compute $g(\Theta) = \frac{1}{J}\sum_{j=1}^{J} \xi_j(\Theta)^T z_j$
  - Stop when $\|\nabla_\Theta (g(\Theta)^T W g(\Theta))\| \leq \epsilon_{\text{out}}$
- Inner loop: compute $\xi(\Theta)$ for a given Θ
  - Solve $s(x, p, \xi; \Theta) = S$ for ξ
  - But HOW?
  - BP (2007) suggest combining contraction mapping iterations, homotopy method and Newton's method

Part IV

MPEC for Estimating Pure Characteristics Demand Models

Discontinuity in Market Share Equations and GMM Function

- The main difficulty in solving the estimation problem of the pure characteristics model formulated in Berry and Pakes (2007) is due to the use of an indicator function in the market share equations
  - Share equation in the population

$$s_j(x, p, \xi, ; \Theta) = \mathbb{E}\left[\mathbf{1}\left\{u_{ij} \geq u_{ij'}, \forall j' \neq j\right\}\right]$$

  - Share equation with sample average approximation

$$\hat{s}_j(x, p, \xi, ; \Theta) = \frac{1}{ns} \sum_{r=1}^{ns} \mathbf{1}\left\{u_{ij} \geq u_{ij'}, \forall j' \neq j\right\}$$

Discontinuity in Market Share Equations and GMM Function

- The main difficulty in solving the estimation problem of the pure characteristics model formulated in Berry and Pakes (2007) is due to the use of an indicator function in the market share equations
  - Share equation in the population

$$s_j(x, p, \xi, ; \Theta) = \mathbb{E}\left[\mathbf{1}\left\{u_{ij} \geq u_{ij'}, \forall j' \neq j\right\}\right]$$

  - Share equation with sample average approximation

$$\hat{s}_j(x, p, \xi, ; \Theta) = \frac{1}{ns} \sum_{r=1}^{ns} \mathbf{1}\left\{u_{ij} \geq u_{ij'}, \forall j' \neq j\right\}$$

- How do we address this discontinuity issue caused by the use of an indicator function?

Discontinuity in Market Share Equations and GMM Function

- The main difficulty in solving the estimation problem of the pure characteristics model formulated in Berry and Pakes (2007) is due to the use of an indicator function in the market share equations
  - Share equation in the population

$$s_j(x, p, \xi, ; \Theta) = \mathbb{E}\left[\mathbf{1}\left\{u_{ij} \geq u_{ij'}, \forall j' \neq j\right\}\right]$$

  - Share equation with sample average approximation

$$\hat{s}_j(x, p, \xi, ; \Theta) = \frac{1}{ns} \sum_{r=1}^{ns} \mathbf{1}\left\{u_{ij} \geq u_{ij'}, \forall j' \neq j\right\}$$

- How do we address this discontinuity issue caused by the use of an indicator function?
  - Do NOT use an indicator function to characterize a consumer's purchase decision

Consumer's Decision as an LP

- Let π_{ij} be the choice probability that that consumer i purchases product j, for $j = 0, \ldots, J$.
- Consumer i purchases the product (including the outside option) that gives her the highest utility
- Consumer i's decision problem can be formulated as a linear program

$$\begin{aligned} \underset{\{\pi_{ij}\}_{j=0,1,\ldots J}}{\text{maximize}} \quad & \sum_{j=1}^{J} \pi_{ij}(x_j^T \beta_i^x - \beta_i^p p_j + \xi_j) \\ \text{subject to} \quad & \sum_{j=0}^{J} \pi_{ij} = 1 \quad (\text{dual var: } \gamma_i) \\ & \pi_{ij} \geq 0, \quad j = 0, \ldots, J \end{aligned}$$

Smooth Reformulation of Approximated Share Equations

- Sample average approximated market share equations in the pure characteristics model are characterized by a linear complementarity problem (LCP)

$$
\begin{aligned}
\forall j : \quad & \frac{1}{N}\sum_{i=1}^{N} \pi_{ij} = S_j, \\
\forall\, i, j : \quad & 0 \le \pi_{ij} \perp \gamma_i - (x_j^T \beta_i^x - \beta_i^p p_j + \xi_j) \ge 0, \\
\forall\, i : \quad & 0 \le \pi_{i0} \perp \gamma_i \ge 0, \\
\forall\, i : \quad & \sum_{j=0}^{J} \pi_{ij} = 1
\end{aligned}
$$

$$(x \perp y \Longleftrightarrow x^T y = 0)$$

MPCC Formulation for Estimating Pure Characteristics Models

- Embedding the LCP constraints for market share equations in the estimation problem, we obtain an MPEC formulation

$$\begin{array}{rl} \underset{\{\Theta,\xi,\pi,\gamma\}}{\text{minimize}} & \xi' Z W Z' \xi \\ \text{subject to} & \dfrac{1}{N}\displaystyle\sum_{i=1}^{N} \pi_{ij} = S_j, \quad \forall j, \\ \forall\, i,j: & 0 \le \pi_{ij} \perp \gamma_i - (x_j^T \beta_i^x - \beta_i^p p_j + \xi_j) \ge 0, \\ \forall\, i: & 0 \le \pi_{i0} \perp \gamma_i \ge 0, \\ \forall\, i: & \displaystyle\sum_{j=0}^{J} \pi_{ij} = 1, \\ \forall i: & \beta_i \sim dF_\beta(\beta; \Theta) \end{array} \tag{1}$$

- **Proposition.** The MPCC problem (1) has a feasible solution.

Mathematical Programs with Complementarity Constraints (MPCC)

$$\begin{aligned} \underset{\{z\}}{\text{minimize}} \quad & f(z) \\ \text{subject to} \quad & g(z) = 0 \\ & 0 \le z \perp h(z) \ge 0. \end{aligned}$$

- MPCCs are a class of challenging optimization problems that have been actively studied in the last 20 years
- Luo, Pang and Ralph (1996), Hu and DeMiguel (2009)

Firm's Price Optimization Problem under Pure Characteristics Models

- There are F firms, each producing a subset of J products
  - $\mathcal{J}_f$: set of products produced by firm f
  - $\mathcal{J}_f \cap \mathcal{J}_g = \emptyset$, for $f \neq g$, and $\cup_f \mathcal{J}_f = \{1, 2, \ldots, J\}$.
- Notations:
  - $\mathcal{M}$: the population in the market
  - mc_j: the marginal cost of producing product j
  - $p = (p_j)_j$, $p_f = (p_j)_{j \in \mathcal{J}_f}$ and $p_{-f} = p \setminus p_f = (p_j)_{j \notin \mathcal{J}}$
- Firm f's profit maximization problem

$$\max_{p_f} \Pi_f(p_f; p_{-f}) = \sum_{j \in \mathcal{F}_f} (p_j - mc_j) \mathcal{M} s_j(x, p, \xi; \Theta)$$

MPCC Formulation for Firm f's Pricing Problem

- Taken $(p_j)_{j \notin \mathcal{F}_f}$ as fixed and using the smooth reformulation of the share equation, we formulate firm f's pricing decision problem as

$$
\begin{array}{cl}
\underset{\{(p_j)_{j \in \mathcal{J}_f}, \pi^f, \gamma^f\}}{\text{maximize}} & \displaystyle\sum_{j \in \mathcal{J}_f} (p_j - mc_j) \mathcal{M} \left(\frac{1}{N} \sum_{i=1}^{N} \pi^f_{ij} \right) \\
\text{subject to} & \\
\forall\, i, \ell = 1, \ldots, J: & 0 \le \pi^f_{i\ell} \perp \gamma^f_i - (x_\ell^T \beta_i - \alpha_i p_\ell + \xi_\ell) \ge 0, \\
\forall\, i: & 0 \le \pi^f_{i0} \perp \gamma^f_i \ge 0, \\
\forall\, i: & \displaystyle\sum_{\ell=0}^{J} \pi^f_{i\ell} = 1.
\end{array}
$$

- This formulation allows us to properly study the firm's price optimization problem

Numerical Experimentation on Synthetic Data

- Use SNOPT and KNITRO to solve the MPCC estimation problem
- The KNITRO solution times were very fast on the problems solved
- In several runs, the sum of squared deviations (SSD) is added to the objective function:

$$\|\Theta - \Theta^*\|^2$$

  with the goal of recovering the incumbent parameters
- With the additional SSD, the solver recovered the incumbent values of the parameters
- The largest problem solved has 100 consumers, 10 products in each markets, 10 markets, and 5 firms competing in these markets. The solver successfully solved the QPCC with 121,100 linear complementarity constraints and 123,320 equalities after a little less than 2 hours

Further Research Questions

- The MPCC formulation for both the estimation problem and the pricing optimization problem is a sample average approximation of the original stochastic equilibrium problem
- Does a solution of the sample average approximation problem converge to a solution of the original stochastic equilibrium problem?
- This question is more difficult to answer because there might be multiple solutions to the equilibrium constraints in MPCC
    - Estimation: Chen, Sun, and Wets (2014)
    - Pricing: Chen, Su and Sun (2015)

Open Questions

- Computational methods for solving the GMM estimation problem formulated as a large-scale MPEC
  - 10000+ variables & (LCP) constraints
- Incorporating supply-side (Bertrand pricing problem) in the demand estimation?
  - constrained optimization with an EPEC as constraints
- Existence of an pricing equilibrium in the pricing model?