Unobserved Heterogeneity, State Dependence, and Health Plan Choices

Ariel Pakes, Jack Porter, Mark Shepard, and Sophie Calder-Wang

Presented by Andrew Capron

December 2, 2024

THE UNIVERSITY *of* NORTH CAROLINA *at* CHAPEL HILL

A Classic Identification Problem

The purpose of this paper is to provide a **new method** to disentangle **unobserved heterogeneity** from **state dependence**

A Classic Identification Problem

The purpose of this paper is to provide a **new method** to disentangle **unobserved heterogeneity** from **state dependence**

At a high-level: using individual-by-product FEs and switching choices, the authors create bounds on the magnitude of switching costs in health insurance markets, finding much smaller costs than standard models

Importantly, method does not identify *mechanisms* behind state dependence, only its empirical significance

Primary Contribution

Two broad strands of literature have tackled this problem:

1. Fully parametric utility model with explicit switching cost using lagged state
   - Heavily relies on valid initial conditions (first-choice problem)
2. Nonparametric approach using individual-by-product FEs to difference out unobserved heterogeneity
   - Requires finding cases where all product characteristics (e.g., prices, etc.) are constant over time

New approach: partial identification in model with flexible FEs using **changes in product characteristics** to identify state dependence (will provide both nonparametric and parametric approaches)

The General Framework

Fairly standard DDC model with individual-by-product FEs:

$$U_{d,i,t} = \left(-p_{d,i,t} - \kappa_0 \cdot \mathbb{1}\{y_{i,t-1} \neq d\}\right) \beta_i + \lambda_{d,i} + \varepsilon_{d,i,t}$$

- β_i: heterogeneous price sensitivity,
- $\lambda_{d,i}$: individual product preferences,
- κ_0: price-equivalent cost of switching
  - The main parameter of empirical interest: only partially identified through various restrictions on $\varepsilon_{d,i,t}$

Method relies on **revealed preference** to construct moment inequalities that identify bounds on κ_0

The Main Idea...

The identification strategy can be seen using a "simple" example:

- Suppose the relative price for d falls between periods $t-2$ and $t-1$; in period t, the relative price goes back up
- For consumers in $t-2$ with $y \neq d$, if $\mathbb{P}(y_{t-1} = d | y_{t-2} \neq d) > \mathbb{P}(y_t = d | y_{t-1} = d)$, then inference on upper bound of switching cost can be made (i.e., κ_0 small)
- Same goes for lower bound: probability of staying with worse product is greater than switching to it when product improves
- Because unobserved preferences (and price sensitivity) remain constant over time (**key assumption!**), these FEs are differenced out
- Tightness of bounds depends on sensitivity of switching to changes in product characteristics

Revealed Preference Moment Inequalities

If an individual chooses d in t and c in s (where $s < t$ and $\mathcal{D}_t \cap \mathcal{D}_s$), then revealed preference implies the following:

$$U_{d,i,t} - U_{c,i,t} \geq 0 \text{ and } U_{c,i,s} - U_{d,i,s} \geq 0$$

$$\begin{aligned}
\implies & \left(-[p_{d,i,t} - p_{c,i,t}] - [\mathbb{1}\{y_{i,t-1} \neq d\} - \mathbb{1}\{y_{i,t-1} \neq c\}]\kappa\right)\beta_i + \\
& [\lambda_{d,i} - \lambda_{c,i}] + [\varepsilon_{d,i,t} - \varepsilon_{c,i,t}] \geq 0 \text{ and} \\
& \left(-[p_{c,i,s} - p_{d,i,s}] - [\mathbb{1}\{y_{i,s-1} \neq c\} - \mathbb{1}\{y_{i,s-1} \neq d\}]\kappa\right)\beta_i + \\
& [\lambda_{c,i} - \lambda_{d,i}] + [\varepsilon_{c,i,s} - \varepsilon_{d,i,s}] \geq 0 \\
\implies & \beta_i^{-1} \Delta\Delta\varepsilon_{i,t,s}^{d,c} \geq \Delta\Delta p_{i,t,s}^{d,c} + \kappa \cdot sw_{i,t,s}^{d,c}
\end{aligned}$$

where $sw_{i,t,s}^{d,c} = (\mathbb{1}\{y_{i,t-1} \neq d\} - \mathbb{1}\{y_{i,t-1} \neq c\}) - (\mathbb{1}\{y_{i,s-1} \neq d\} - \mathbb{1}\{y_{i,s-1} \neq c\})$

and $\Delta\Delta x_{i,t,s}^{d,c} = [(x_{d,i,t} - x_{c,i,t}) - (x_{d,i,s} - x_{c,i,s})]$ for any variable x

Important Assumption Across All Specifications

Conditional Independence Assumption:

$$\varepsilon_{i,t} | p_i, y_{i,t-1}, y_{i,t-2}, \ldots, y_{i,0}, \beta_i, \lambda_i \sim \varepsilon_{i,t} | \beta_i, \lambda_i$$

- All serial dependence in choices – not associated with covariates (prices) or individual fixed effects – is modeled through the state dependence parameters
- After accounting for state dependence (κ_0) and fixed effects (λ_i), any further patterns in choices are assumed to be random and uncorrelated error

Increasing Levels of Structure

Three assumptions on error term are considered:

- Nonparametric:
  1. Conditional median of $\beta_i^{-1}\Delta\Delta\varepsilon_{i,t,s}^{d,c} = 0$
  2. Stationarity assumption $\varepsilon_{i,t}|\lambda_i, \beta_i \sim \varepsilon_{i,1}|\lambda_i, \beta_i$
- Nonparametric approaches used to infer bounds on switching costs; use the bounds to explore degree of state dependence
- Parametric distributions:
  3. $\varepsilon_{i,t} \sim F(\cdot|\theta)$; unspecified distribution and logistic

Nonparametric I: Conditional Median Assumption

- Denoting $\Delta\Delta\varepsilon_{i,t}|p_i, y_{i,s-1} \sim \mathcal{F}^{d,c}_{t,s}(\cdot|p_i, y_{i,s-1})$, we have:

$$\mathbb{P}(y_{i,t} = d, y_{i,s} = c|p_i, y_{i,s-1}) \leq 1 - \mathcal{F}^{d,c}_{t,s}(\Delta\Delta p^{d,c}_{i,t,s} + \kappa_0 \cdot sw^{d,c}_{i,t,s})$$

- LHS estimate can be obtained directly from data
- RHS is a **theoretical bound** and depends on κ_0, but LHS and cond. median $[\Delta\Delta\varepsilon] = 0$ assumption can be used to infer bounds on κ_0
- Intuition (for upper bound, i.e., $sw > 0$):
  - If switching costs are high, then less likely to see switching and LHS probability should be small
  - Since RHS $\downarrow$ in κ_0, if switching is observed often and LHS $>$ RHS, then κ_0 is overestimated, providing an upper bound
  - Lower bound established in same way for $sw < 0$ (RHS $\uparrow$ in κ_0)

Nonparametric I: Bounds

- If LHS < 0.5, then $\Delta\Delta p_{i,t,s}^{d,c} + \kappa \cdot sw_{i,t,s}^{d,c} < 0 \implies 1 - \mathcal{F}_{t,s}^{d,c}(\cdot) > 0.5$ and inequality is satisfied $\forall\ \kappa$
- Thus the method provides no power to reject any values of κ
- If LHS > 0.5, then cond. median assumption implies that $\Delta\Delta p_{i,t,s}^{d,c} + \kappa \cdot sw_{i,t,s}^{d,c} < 0$ which produces the upper (lower) bound on κ_0 when $sw > 0$ (< 0)
- LHS always < 0.5 for authors' application, so this method is toothless here...

Nonparametric II: Conditional Stationarity Assumption

- Stationarity assumption $\varepsilon_{i,t}|\lambda_i, \beta_i \sim \varepsilon_{i,1}|\lambda_i, \beta_i$
  - Common dynamic panel assumption
  - Error may differ between individuals and correlate across choices within-individual, but stationary over time
  - Essentially, just ensures persistence in choices attributable to κ_0 and λ_i
- Structural component of utility defined as:

  $SU_{d,i,t}(y_{i,t-1}, p_i; \lambda_i, \kappa) = U_{d,i,t} - \varepsilon_{d,i,t}$
- Then, $d_1(\cdot)$ is the choice whose SU improves most between s and t:

$$d_1(y_{i,t-1}, y_{i,s-1}, p_{i,t}, p_{i,s}; \kappa) = \max_{d \in \mathcal{D}} SU_{d,i,t}(\cdot) - SU_{d,i,s}(\cdot)$$

- For given κ, ordering of $d_1(\cdot), \ldots, d_D(\cdot)$ depends only on price differences and switching costs

Nonparametric II: Main Idea

- For $t > s$, if

$$\min_{d \in D_0} SU_{d,i,t}(\cdot|\lambda_i, \kappa_0) - SU_{d,i,s}(\cdot|\lambda_i, \kappa_0) \geq$$
$$\max_{c \notin D_0} SU_{c,i,t}(\cdot|\lambda_i, \kappa_0) - SU_{c,i,s}(\cdot|\lambda_i, \kappa_0)$$

then

$$\mathbb{P}(y_{i,t} \in D_0 | p_i, y_{i,t-1}, \lambda_i) \geq \mathbb{P}(y_{i,s} \in D_0 | p_i, y_{i,s-1}, \lambda_i)$$

- If the SU_i $\forall$ $d \in D_0$ improves by as much as all $c \notin D_0$ between s and t, then i will be more likely to choose any $d \in D_0$ at t than s regardless of λ_i

Nonparametric II: Main Idea

- For $t > s$, if

$$\min_{d \in D_0} SU_{d,i,t}(\cdot|\lambda_i, \kappa_0) - SU_{d,i,s}(\cdot|\lambda_i, \kappa_0) \geq$$
$$\max_{c \notin D_0} SU_{c,i,t}(\cdot|\lambda_i, \kappa_0) - SU_{c,i,s}(\cdot|\lambda_i, \kappa_0)$$

  then

$$\mathbb{P}(y_{i,t} \in D_0 | p_i, y_{i,t-1}, \lambda_i) \geq \mathbb{P}(y_{i,s} \in D_0 | p_i, y_{i,s-1}, \lambda_i)$$

- If the $SU_i \; \forall \; d \in D_0$ improves by as much as all $c \notin D_0$ between s and t, then i will be more likely to choose any $d \in D_0$ at t than s regardless of λ_i
- Why does this matter? Directly relates ΔSU to switching behavior
- Would provide test for values of κ, but for conditioning on λ_i

Nonparametric II: Main Idea (Continued)

- To overcome state dependence issue, some math is done (see paper) to arrive at:

$$\mathbb{P}(y_{i,t} = d_1^0 | p_i, y_{i,t-1} = d_1^0, y_{i,t-2}) \geq \mathbb{P}(y_{i,t-1} = d_1^0 | p_i, y_{i,t-2})$$

  where d_1^0 is $d_1 | \kappa = \kappa_0$

- **Main idea:** in the presence of switching costs, more likely to stick with same choice in t than to have made choice in $t-1$ to begin with
- Bounds can be formed: if κ_0 is too small, then inequality may be violated; too big and it becomes slack
- d_1^0 is the choice where utility changes are most favorable between s and t, so bounds are tightest when applied to d_1^0

Nonparametric II: Broader Idea

- The authors further extend this Lemma to consider transitions over *multiple periods* (e.g., $s < t-1$) and sequences of *choice sets*
- For $s = t-1$ and $D_0 = D_1$ (satisfying similar min/max SU condition),

$$\mathbb{P}(y_{i,t} \in D_0 | p_i, y_{i,t-1} \in D_0, y_{i,t-2}) \geq \mathbb{P}(y_{i,t-1} \in D_0 | p_i, y_{i,t-2})$$

- For $s < t-1$ and any D_0, D_1 (satisfying SU condition),

$$\mathbb{P}(y_{i,t} \in D_0 | p_i, y_{i,t-1} \in D_1, y_{i,s} \in D_0, y_{i,s-1}) \geq \mathbb{P}(y_{i,t-1} \in D_1, y_{i,s} \in D_0 | p_i, y_{i,s-1})$$

- "Optimal" (most improved) choice set in s and t (D_0) compared with set in $t-1$ (D_1)
- If switching costs are significant, more likely to stay even when alternative options improve
- **Importantly,** LHS $>$ RHS $\implies$ small switching costs; vice versa for LHS $<$ RHS
  - Once again allows us to establish bounds on κ_0

Empirical Application: Massachusetts Health Insurance

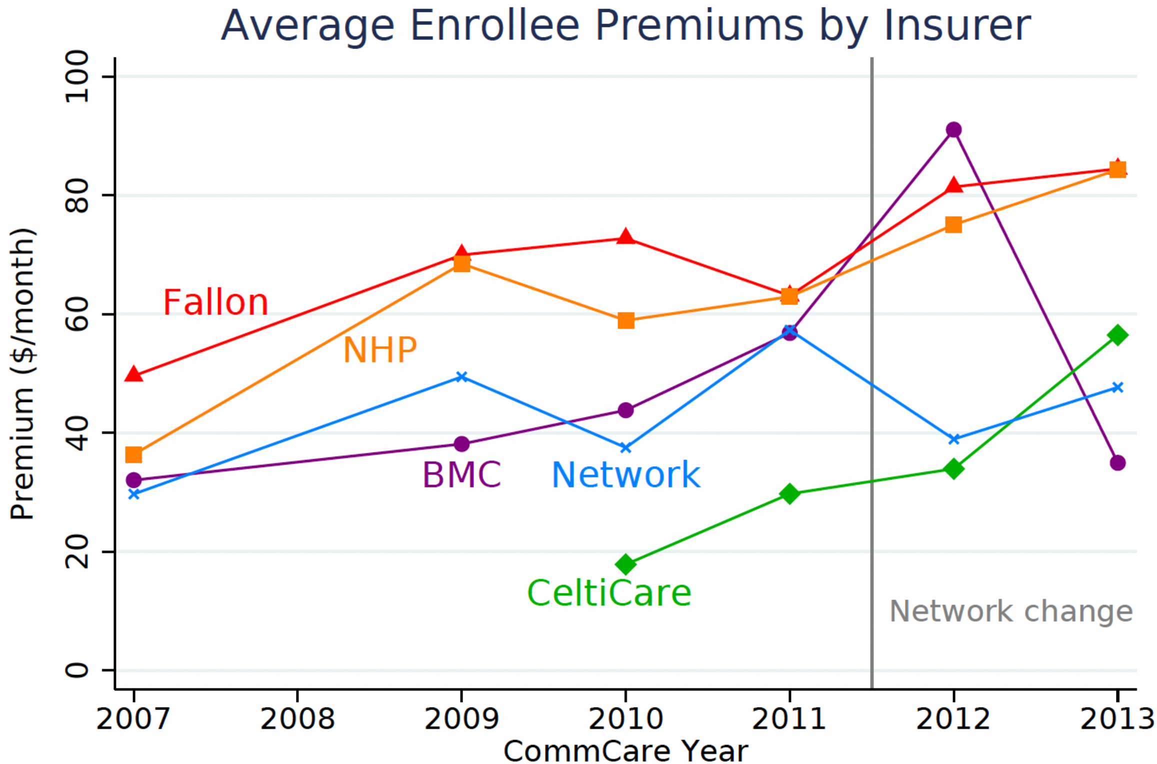
- "CommCare" program from 2009-2013
- Five insurers participate in exchange over sample period, with each only providing one plan by rule
  - Importantly, insurers enter and exit across panel, changing choice sets
  - Network treated as two plans (2009-2011, 2012-on) due to dropping Partners; Celticare entry in 2010; Fallon partial exit in 2010
- Subsidizing individuals earning less than 300% of Federal Poverty Level (no family coverage)
- Income groups are used to set subsidies are the *only source* of variation in the structural utility component over time:
  - 0-100% of poverty, 101-150%, 151-200%, 201-250%, and 251-300%
  - Prices can vary by region in 2009-10, but not thereafter

Cell and Moment Inequality Formation

- Important to focus on switching costs due to price changes, not in choice environment (drop these individuals)
- "Cells" are individuals with same $y_{i,s-1}$ and observed characteristics
  - Cartesian product of: couple of years; region; plan options; income group (level at which λ_i is defined)
- All cells with more than 20 members are used in estimation
  - Inequalities generated from by cells are summed across regions and plan availability to form "groups" used in estimation
  - Grouping into cells reduces variance in unobserved heterogeneity; focus on prices and state dependence
- Since κ_0 directly impacts SU, the number of moment inequalities depends on the value of κ being tested

Table 1: Summary Statistics for the Nonparametric Estimator

(s, t)	Number of Members	Number of Groups	Number of Members Above Cutoff	Number of Groups Above Cutoff	Minimum Number of Moments	Maximum Number of Moments
(2009, 2010)	19,550	96	17,349	66	1,494	3,671
(2010, 2011)	13,989	96	13,181	76	3,181	8,748
(2012, 2013)	47,266	120	44,438	100	4,251	11,225
Total	**80,805**	**312**	**74,968**	**242**	**8,926**	**23,644**



Average premiums for above-poverty enrollees only. Below poverty people pay $0.

- Starting in 2012, enrollees below federal poverty line ($\sim$ 50% of customers) can only purchase from two lowest priced plans

Table 2: Statistics on Enrollment and Switching for 2011 Enrollees over 2011-2013

	All 2011 Enrollees	**By 2011 Plan** BMC	All Other Plans
Number of Enrollees			
Total Enrollees in 2011	111,226	36,235	74,991
Leave Market before 2013	76,007	24,812	51,195
Stay in Market 2011-13	35,219	11,423	23,796
Switching Rates (among stayers in market)			
Switch Plans from 2011-2012	14.5%	19.0%	12.3%
Switch in 2012, Switch Back in 2013	2.5%	6.8%	0.5%
Switch in 2012, Do Not Switch Back 2013	11.9%	12.1%	11.8%

- BMC's sharp price decrease in 2013 led to 36% (6.8%/19%) of leavers to switch back (far more than rivals)

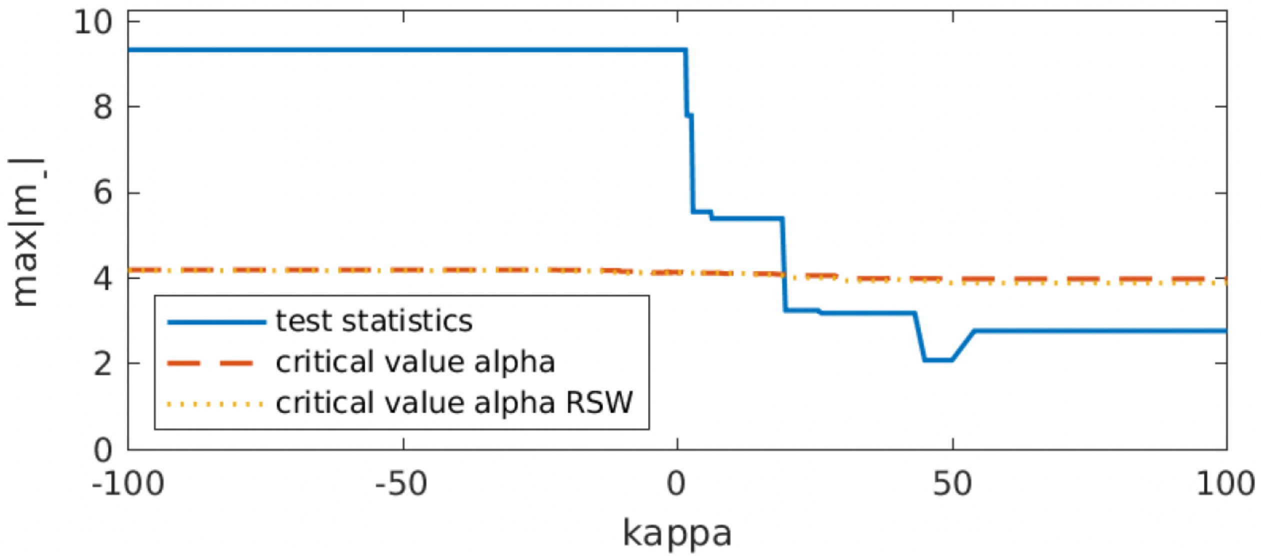
Simple Example of Nonparametric Approach With BMC

Back to nonparametric II and choice sets:

- D_0: BMC raises prices in 2012 but drops in 2013
- D_1: CeltiCare lowest price in 2013 but relatively higher in 2013
- Observe individuals switch to CeltiCare in 2012 but then back to BMC in 2013
- LHS probability is large ($t \to D_0$), meaning easy movement between plan sets
- In this case, LHS $>$ RHS $\implies \kappa_0$ small

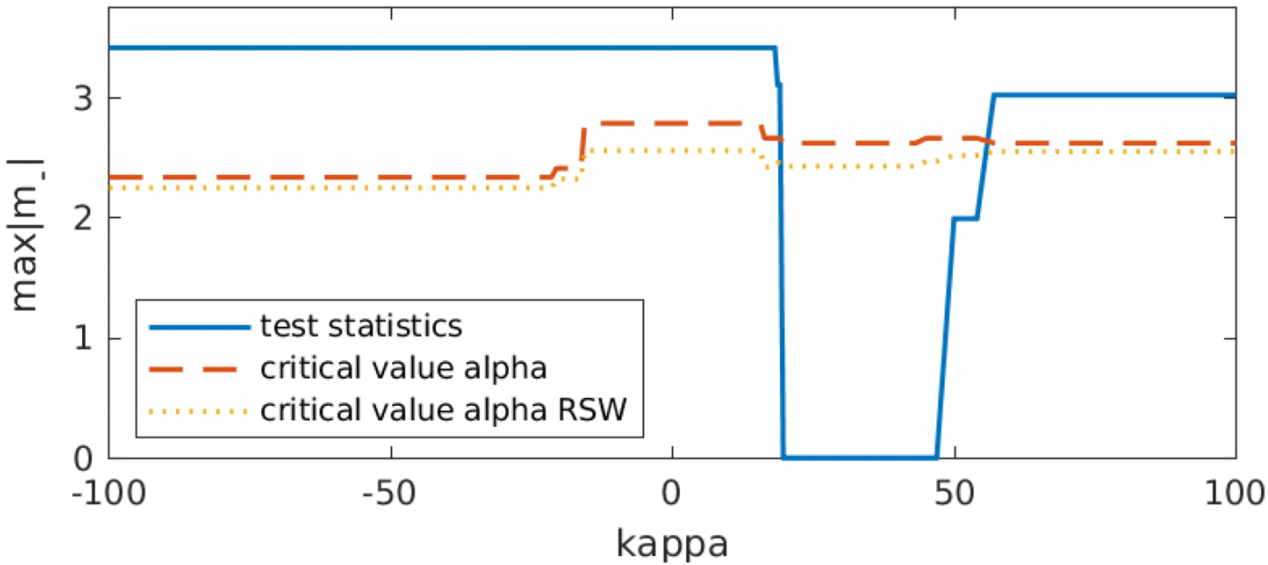
Nonparametric Estimates Using Table 1 Sample

Figure 2: Nonparametric Estimates of Switching Costs (κ_0)



- Blue line crosses at $19.6, providing a lower bound
- Never crosses back at above critical values, providing no upper bound (due to number of inequalities limiting power)

Estimates Using "Endogenously Selected" Inequalities



- When inequalities are selected around sharp price increases (i.e., BMC in $t = 2013$), bounds become [\$19.6, \$57]

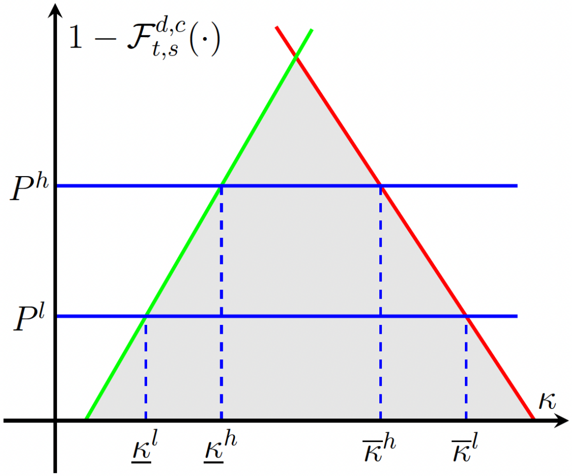
Next, A Comparison To The Parametric Approach

Rewrite utility as follows:

$$U_{d,i,t} = (-p_{d,i,t}\gamma_0 - \mathbb{1}\{y_{i,t-1} \neq d\}\delta_0)\beta(x_i) + \lambda_{d,i} + \varepsilon_{d,i,t}$$

- Assumption: $\beta_i = \beta(x_i)$, where x_i are time-invariant
- Still focused on $\kappa_0 = \delta_0/\gamma_0$
- Apply the same revealed preference logic from *Nonparametric I*:

Figure 4: Identified set for κ_0



Parametric Statistics When Assuming Gumbel Errors

Table 3: Summary Statistics for the Parametric Estimator

(s, t)	Number of Members	Number of Groups	Number of Members Above Cutoff	Number of Groups Above Cutoff	Number of Moments
(2009, 2010)	59,322	100	32,738	69	248
(2009, 2011)	39,955	100	24,740	78	296
(2010, 2011)	59,629	100	34,300	83	217
(2012, 2013)	69,441	125	43,138	99	522
Total	**228,347**	**425**	**134,916**	**329**	**1,283**

- Logit errors once again guarantee meaningful upper and lower bounds on κ_0 due to formation of traditional choice probabilities describing ratio of odds (d at t and c at $t-1$ versus c at t and d at $t-1$)

Parametric Results Without Fixed Effects

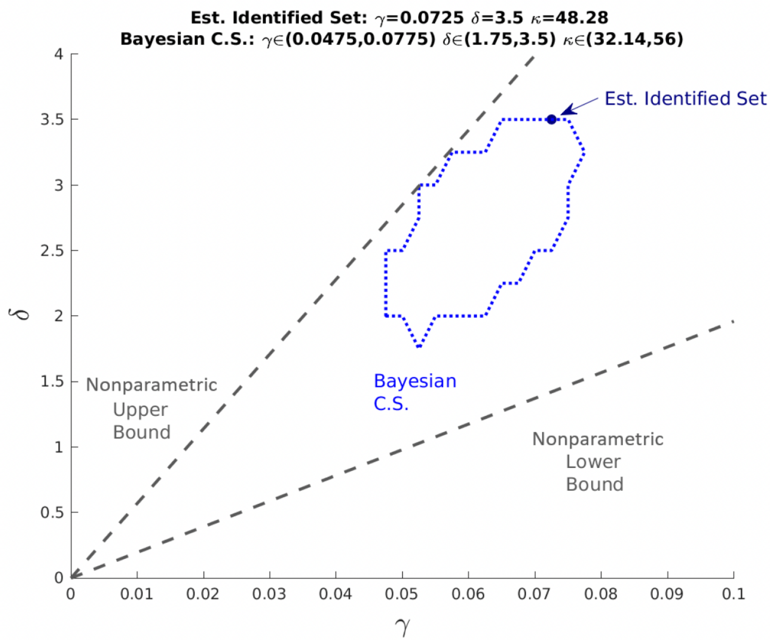
Table 4: Multinomial Logit Estimation

	Simple (1)	Plan Dummies (2)	Detailed Plan Dum. (3)	Detailed Plan Dum. + Network (4)	Plan Dum. + Random Effects (5)	Include New Enr (6)	New Enr + Random Effects (7)
Normalize $\epsilon_{i,d,t}$ to EV1							
Switching Cost (δ)	-4.086	-4.196	-4.156	-4.120	-4.480	-3.974	-4.478
	(0.006)	(0.007)	(0.007)	(0.007)	(0.014)	(0.006)	(0.010)
Price Coefficient (β)	-0.036	-0.049	-0.051	-0.053	-0.051	-0.036	-0.041
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Hospital Network Utility	—	—	—	0.137	—	—	—
				(0.007)			
Prev. Used Hospitals Covered	—	—	—	0.804	—	—	—
				(0.019)			
Prev. Used x Partners Hosp.	—	—	—	0.974	—	—	—
				(0.026)			
Normalize β to 1							
Switching Cost ($\kappa = \delta/\beta$)	114.03	86.35	80.76	78.13	87.88	110.79	110.17
	(0.55)	(0.34)	(0.32)	(0.30)	(0.35)	(0.43)	(0.44)
Plan Dummies	—	Yes	Yes	Yes	Yes	Yes	Yes
Plan x (Area, Age-Sex, Illness)	—	—	Yes	Yes	—	—	—
Plan Random Effects	—	—	—	—	Yes	—	Yes
N Parameters	2	7	249	252	11	7	12
N Individuals x Years	2,623,699	2,623,699	2,623,699	2,623,699	2,623,699	3,832,629	3,832,629

- Restrict attention to "Switching Costs" row (typo: β should be γ)

Parametric Results Allowing For Fixed Effects

Figure 5: Identified set and Bayesian confidence set for γ_0 and δ_0.



- Using same algorithm as nonparametric II, but now with Bayesian approach to generate bounds on (δ_0, γ_0), model yields bounds of [$32,$56] (almost identical to nonparametric u.b. of $57)

Simple Counterfactual on BMC Prices

Table 5: Counterfactual Comparisons

Specification	2011 market shares	2012 status-quo	2012 counterfactual	2012 *% diff*	2013 status-quo	2013 counterfactual	2013 *% diff*
Market shares without imposing κ							
Plan FE	0.357	0.289	0.321	*11.0*	0.266	0.304	*14.2*
Plan $\times$ Region FE	0.357	0.289	0.320	*10.6*	0.266	0.298	*12.2*
Plan FE + RE	0.357	0.282	0.324	*15.1*	0.289	0.311	*7.6*
Market shares imposing κ							
Plan FE	0.357	0.186	0.306	*64.1*	0.399	0.326	*-18.3*
Plan $\times$ Region FE	0.357	0.205	0.313	*52.8*	0.381	0.318	*-16.6*
Plan FE + RE	0.357	0.183	0.305	*66.8*	0.410	0.329	*-19.7*
Premium	58.4	91.1	62.9		41.5	65.3	

- Using the average (actual) prices across each income group for CFs, model predicts **large fluctuations** in market shares when switching costs are imposed