Lecture 3
Unobserved choice sets

Rachel Griffith

CES Lectures, April 2017

1. The general problem of unobserved choice sets
2. Sovinsky Goeree (2008) "Limited information and advertising in the US personal computer industry" Econometrica
3. Gaynor, Propper and Seiler (2016) "Free to Choose? Reform, Choice, and Consideration Sets in the National Health Service" American Economic Review

Demand estimation relies of revealed preference

- Identification of demand parameters rests on revealed preference arguments:
  - someone chooses A over B
  - we infer that they prefer A to B
  - to do this we need to know that they **could have** chosen B if they instead preferred B to A
- cited on the Wikipedia page on revealed preference:
  - *If there exist only an apple and an orange, and an orange is picked, then one can definitely say that an orange is revealed preferred to an apple. In the real world, when it is observed that a consumer purchased an orange, it is impossible to say what good or set of goods or behavioral options were discarded in preference of purchasing an orange. In this sense, preference is not revealed at all in the sense of ordinal utility.*
    Koszegi and Rabin (2007, AER)

Rich theoretical literature suggesting choice sets are heterogenous and constrained

- Consumers might face constrained choices for many reasons, and these are likely to differ across consumers
  - limited information, search, time constraints, potentially combined with firm behaviour (e.g. advertising)
    - Eliaz and Spiegler (2011, REStud); De Los Santos et. al (2012, *AER*); Reutskaja, Camerer, and Rangel (2011, *AER*)
  - limited or rational (in)attention
    - Masatlioglu et al (2012, *AER*); Manzini and Mariotti (2014, *Econometrica*)
  - self control problems and commitment
    - Bernheim and Rangel (2009, QJE), DellaVigna (2009, JEL)
- Our ability to explore their empirical relevance is limited

Implications of heterogeneity in choice sets for demand estimation

- Heterogeneity in choice sets can be useful for identification
  - if variation is exogenous and we observe the variation in choice sets
  - or if we know and can model the process of choice set formation
- but more often it is problematic
  - when it is unobserved (by the econometrician) or when choice set formation is endogenous

Simple example to gain intuition for the problem

- a consumer can either buy ($Y_i = 1$) or not buy ($Y_i = 0$)
- payoff to buying is U_{i1}; payoff to not buying is U_{i0}
- some consumers have the option to buy ($CS_i^\star = \{0, 1\}$) and others don't ($CS_i^\star = \{0\}$)
- what we observe in the data

$$Y_i = \begin{cases} 1 & \text{if } U_{i1} \geq U_{i0}, \text{ and } CS_i^\star = \{0,1\} \\ 0 & \text{if } \begin{cases} U_{i1} < U_{i0} \text{ and } CS_i^\star = \{0,1\} \\ \text{or } CS_i^\star = \{0\} \end{cases} \end{cases}$$

Simple example to gain intuition for the problem

- In this example the probability that i buys is:

$$\Pr(Y_i = 1) = \underbrace{\Pr[U_{i1} > U_{i0} | CS_i^\star = \{0,1\}]}_{\text{probability that buy given have choice}} \overbrace{\Pr[CS_i^\star = \{0,1\}]}^{\text{probability that have choice}}$$

- revealed preference arguments allow us to infer preferences based on this part
- this part is a problem and causes bias in parameter estimates

Discrete choice demand model with heterogenous choice sets

- Market with $j = 1, \ldots, J$ products
- $i = 1, \ldots, I$ *consumers*
- $t = 1 \ldots, T$ *choice situations*
- The choice i makes in t: Y_{it}
  - the sequence of choices i makes, $Y_i = (Y_{i1}, \ldots, Y_{iT})$
- The true choice set of i in t: $CS^{\star}_{it}$
  - unobserved to the econometrician
  - the sequence of choice sets: $\mathcal{CS}^{\star}_i = (CS^{\star}_{i1}, \ldots, CS^{\star}_{iT})$

Discrete choice demand model with heterogenous choice sets

- We are interested in the consequences of mistakenly assuming that i chooses from some larger choice set that contains options that were not in fact in their choice set
  - denote this larger set S_{it}
  - and the sequence of these: $\mathcal{S}_i = (S_{i1}, \ldots, S_{iT})$
  - (for notational convenience assume S_t the same for all i)
  - so that the sequence of these is: $\mathcal{S} = (S_1, \ldots, S_T)$

Discrete choice demand model with heterogenous choice sets

- Denote the preferences that govern choice by θ_i and γ_i
  - $\Pr[Y_i = j | \mathcal{CS}_i^\star = c, \theta_i]$
    - probability of making a sequence of choices, $Y_i = j$, given the consumer is matched to choice sequence, $\mathcal{CS}_i^\star$
    - revealed preference arguments allow us to make inference about consumer preferences based on this
  - $\Pr[\mathcal{CS}_i^\star = c | \gamma_i]$
    - the process matching consumers to their (unobserved) choice sequence
    - revealed preference arguments are not informative here
- Probability of making a sequence of choices, $Y_i = j$, is $\Pr[Y_i = j | \mathcal{CS}_i^\star = c, \theta_i] \Pr[\mathcal{CS}_i^\star = c | \gamma_i]$

Discrete choice demand with heterogenous choice sets

- Let indirect utility from choosing i in t:

$$U_{ijt} = V_{ijt}\left(X_{ijt}, \theta\right) + \varepsilon_{ijt}$$

- assume ε_{ijt} is distributed Type I Extreme Value *conditional* on the specific sequence of choice sets to which i is matched
- Thus for any choice set, c, that the consumer is matched to

$$\Pr\left[Y_i = j | \mathcal{CS}_i^\star = c, \theta\right] = \prod_{t=1}^{T} \frac{\exp\left(V_{ijt}\left(X_{ijt}, \theta\right)\right)}{\sum_{m \in CS_{it}^\star = c_t} \exp\left(V_{imt}\left(X_{ijt}, \theta\right)\right)}$$

Bias

- If econometrician **incorrectly** specifies choice as larger set $\mathcal{S} = s$:

$$\Pr\left[Y_i = j | \mathcal{S} = s, \theta\right] = \prod_{t=1}^{T} \frac{\exp\left(V_{ijt}\left(X_{ijt}, \theta\right)\right)}{\sum_{m \in S_t = s_t} \exp\left(V_{imt}\left(X_{ijt}, \theta\right)\right)}$$

- This leads to bias

$$\Pr\left[Y_i = j | \mathcal{S} = s, \theta\right] = \underbrace{\Pr\left[Y_i = j | \mathcal{CS}_i^\star = c, \theta\right]}_{\text{probability that buy given have choice}} \overbrace{\Pr\left[Y_i \in \mathcal{CS}_i^\star = c | \mathcal{S} = s, \theta\right]}^{\text{problem}}$$

$$= \prod_{t=1}^{T} \frac{\exp\left(V_{ijt}\left(X_{ijt}, \theta\right)\right)}{\sum_{m \in CS_{it}^\star = c} \exp\left(V_{imt}\left(X_{ijt}, \theta\right)\right)} \frac{\sum_{m \in CS_{it}^\star = c} \exp\left(V_{imt}\left(X_{ijt}, \theta\right)\right)}{\sum_{r \in S_t = s_t} \exp\left(V_{irt}\left(X_{ijt}, \theta\right)\right)}$$

Bias

- Estimation of θ based on "too large" choice set $\mathcal{S} = s$ will be biased if $\Pr[Y_i \in \mathcal{CS}_i^\star = c | \mathcal{S} = s, \theta]$ is important
  - this is the probability that, if i could actually choose from S, the choice they would make would be in their true choice set
  - if all the products that i likes are in their choice set then this will be close to 1 and there will be no bias
  - if when facing the bigger choice set $\mathcal{S} = s$, i would make a choice *not* in $\mathcal{CS}_i^\star = c$, then bias will be larger
  - if we (mistakenly) include alternatives in estimation that i really likes, but were not available, then we will get larger bias
- The size of bias is increasing in
  - the degree of differentiation
  - the share of individuals with constrained choice set
  - the extent of the constraint

Some approaches that **do not** correct the bias

- Add product (j) specific constants to the model
  - this doesn't control for the bias term, which is individual specific
- Add random coefficients to the model
  - relies on distribution of random coefficients being independent of the Xs
  - we can clearly see that this won't hold by looking at the bias term

Approaches to dealing with this

- Collect data that indicates which products are in each consumer's choice sets
- Chamberlain (1980) fixed effect logit
- Approaches deriving from Manski (1977)
  - Sovinsky Goeree (2008, Econometrica), van Nierop et al (2010, JMR), Draganska and Klapper (2011, JMR)
  - jointly model both choice set formation and the purchase decision given a choice set
  - use information on the matching process to integrate out unobserved choice set heterogeneity

Chamberlain (1980)'s Fixed Effect Logit

- It is convenient to rewrite the bias

$$\Pr\left[Y_i = j | \mathcal{S} = s, \theta\right] = \underbrace{\Pr\left[Y_i = j | \mathcal{CS}_i^\star = c, \theta\right]}_{\text{probability that buy given have choice}} \overbrace{\Pr\left[Y_i \in \mathcal{CS}_i^\star = c | \mathcal{S} = s, \theta\right]}^{\text{problem}}$$

$$= \prod_{t=1}^{T} \frac{\exp\left(V_{ijt}\left(X_{ijt}, \theta\right)\right)}{\sum_{m \in CS_{it}^\star = c} \exp\left(V_{imt}\left(X_{ijt}, \theta\right)\right)} \frac{\sum_{m \in CS_{it}^\star = c} \exp\left(V_{imt}\left(X_{ijt}, \theta\right)\right)}{\sum_{r \in S_t = s_t} \exp\left(V_{irt}\left(X_{ijt}, \theta\right)\right)}$$

- in this form

$$\Pr\left[Y_i = j | \mathcal{S} = s, \theta\right] == \prod_{t=1}^{T} \frac{\exp\left(V_{ijt}\left(X_{ijt}, \theta\right) - ln(\pi_i)\right)}{\sum_{m \in S_t^\star = s_t} \exp\left(V_{imt}\left(X_{ijt}, \theta\right)\right)}$$

- this is a fixed-effect (in a non-linear context)

Chamberlain (1980)'s Fixed Effect Logit

- Chamberlain's Fixed Effect Logit "differences" out the bias term
- Conditions on observed choice switches
  - consider a market with products $\{a, b, c, d, e, ...\}$
    - if we observe i choosing a in period 1 and b in period 2, so $Y_i = (a, b)$
    - we know a and b are in i's true unobserved choice set
    - i could have chosen either (a, b) or (b, a)
  - the Fixed Effect Logit estimates the probability:

$$\Pr[Y_i = (a, b) | \{(a, b), (b, a)\}, \theta']$$

  - individual-alternative specific effects difference out
  - **Relies on choice set stability**
  - **Identifies only coefficients on time varying Xs**

Chamberlain (1980)'s Fixed Effect Logit

- This enable us to point-identify preference parameters θ
  - but we cannot point-identify elasticities

$$\xi_{jj} = \beta_{\mathrm{p}} \mathrm{p}_j \left(1 - \frac{\exp(\boldsymbol{\beta X}_j)}{\sum_{l \in CS^\star} \exp(\boldsymbol{\beta X}_l)} \right) \qquad \xi_{jk} = -\beta_{\mathrm{p}} \mathrm{p}_k \left(\frac{\exp(\boldsymbol{\beta X}_k)}{\sum_{l \in CS^\star} \exp(\boldsymbol{\beta X}_l)} \right)$$

  - elasticities are functions of the full choice set, $CS_i^\star$, which are not observable without imposing more structure/further assumptions about the choice set formation process

Approaches deriving from Manski (1977)

- Jointly model both the choice set formation process and the purchase decision given a choice set
- Manski (1977): the *unconditional* probability of i selecting choice sequence j can be written as:

$$\Pr[Y_i = j|\theta, \gamma] = \sum_{c \in C_i^\star} \Pr[Y_i = j|\mathcal{CS}_i^\star = c, \theta] \Pr[\mathcal{CS}_i^\star = c|\gamma],$$

  where $C_i^\star$ is the collection of sets of possible choice sequences to which consumer type i can be matched
- By having information on the **matching process** between consumer types and choice sets, researchers can integrate out unobserved choice set heterogeneity in a matter analogous to that routinely done with unobserved preference heterogeneity

Sovinsky Goeree (2008)

- Considers the US personal computer market
- Consumers have limited consumer information, advertising influences the set of products from which consumers choose to purchase
- Specifies a model of the form:

$$\Pr[Y_{it} = j_t|\theta,\gamma] = \underbrace{\sum_{c\in C^j} \frac{\exp\left(V_{ij_t t}(X_{ijt},\theta)\right)}{\sum_{r\in c}\exp\left(V_{irt}(X_{ijt},\theta)\right)}}_{\text{probability that buy given have choice}} \overbrace{\left[\prod_{l\in c}\phi_{ilt}(\gamma)\prod_{k\notin c}(1-\phi_{ikt}(\gamma))\right]}^{\text{probability that have choice}}$$

where C_t^j is the collection of *all* the choice sets that include product j

Sovinsky Goeree (2008)

- The consideration of each product is assumed to be *independent* from the consideration of the other products

$$\Pr_{it}\left[CS_{it}^{\star} = c|\gamma\right] = \prod_{l \in c} \phi_{ilt}(\gamma) \prod_{k \notin c} (1 - \phi_{ikt}(\gamma))$$

- the number of PCs is large (over 2,000), so the non-parametric estimation of all the ϕ's is not feasible; Sovinksy Goeree's assumes that

$$\phi_{ilt}(\gamma) = \frac{\exp(W_{ilt}(\gamma))}{1 + \exp(W_{ilt}(\gamma))}$$

- where W is advertising

TABLE VI

ESTIMATED PERCENTAGE MARKUPS UNDER LIMITED AND FULL INFORMATION[a]

	Median Percentage Markup		
	Under Limited Information	Under Full Information	Change in Markups
Total industry	15%	5%	67%
Apple		2.5%	84%
iMac	22.1%	3.1%	
Power Mac	13.7%	2.0%	
PowerBook*	10.0%	1.6%	
Compaq		7.0%	69%
Armada 7xxx*	41.4%	3.5%	
Presario 2xxx	18.1%	2.6%	
Presario 1xxx*	15.2%	2.0%	
ProLinea	23.3%	7.0%	
Dell		1.8%	82%
Latitude XPI*	7.0%	1.4%	
Dimension	15.5%	2.4%	
Inspiron	9.4%	1.6%	
Gateway		1.7%	86%
Gateway Desk Series	12.8%	1.9%	
Gateway Portable Series	8.1%	1.5%	
HP		4.5%	72%
OmniBook 4xxx*	8.3%	5.7%	
Pavilion 6xxx	22.7%	3.1%	
Vectra 5xx	15.8%	6.8%	
IBM		2.0%	88%
Aptiva	16.0%	2.3%	
Thinkpad 7xxx*	7.4%	1.6%	

Issues

- potential for misspecification of choice set formation process
- even when correctly specified it is likely to suffer from a curse of dimensionality because the number of elements in $C_i^\star$ grows exponentially in the number of products J sold in the market
- if c is not in the support of the choice set distribution to which individual i can be matched to in period t there will still be some bias
- computing expectations over the universal set would not be a problem if we could estimate a truly flexible specification for ϕ that was able to accomodate $\Pr_{it}\left[CS_{it}^\star = c|\gamma\right] = 0$ whenever necessary
- the problem arises because we are not usually able to estimate a truly flexible model for ϕ, and we need to make additional assumptions

Gaynor, Propper and Seiler

- Expanding choice in healthcare is a popular reform
  - supporters believe it will increase competition and raise quality
  - opponents argue that patients either can't make choices or will make poor choices, so competition will lead to wasteful use of resources, focus on the wrong attributes of care or harm the poor and less educated
- Model demand for coronary artery bypass graft before and after a reform that gave patients choice
  - pre-reform they had constrained choice (their physician decided)
  - post-reform they had increased choice (constrains and incentives for physicians relaxed)

Gaynor, Propper and Seiler

- They do not know precisely the nature of the constraint pre-reform
  - they model demand post reform and recover the contraint by comparing the predicted choices using the model applied in the pre-reform period, and attribute the differences in predicted and actual choices to the constraint
- They find that introducing choice led to patients requiring heart bypass surgery became more responsive to the quality of care; this gave hospitals a greater incentive to improve quality (and they provide some reduced form evidence on this)

Hospital Choice

- Patient's utility

$$U_{ij} = \beta_{wi} W_{jt} + \beta_{zi} Z_{jt} + f(D_{ij}) + \xi_j + \epsilon_{ij}$$

i: patient, j: hospital
W: wait time, Z: quality, D: distance

- Physician utility

$$V_{ij} = g(D_{ij}) + \zeta_j + v_{ij}$$

  - physician utility determined by contracts, financial incentives and other unobservable factors captured by hospital fixed effects, distance related variables and an indicator for whether the hospital was located in the same administrative region (PCT) as the referring physician
  - the reform relaxed these incentives

Hospital Choice

- Physician offers hospital k to patient if

$$V_{ik} \geq max_{j \in J}(V_{ij}) - \lambda_i$$

J: all hospitals, $\lambda_i \geq 0$

  - All hospitals within a distance of λ_i in utility space are included in the choice set (offered to the patient)
    - λ_i captures extent to which physician cares about patient utility
    - $\lambda_i = 0 \implies$ patients preferences have no influence on the choice
    - $\lambda_i > 0 \implies$ physician might include multiple hospitals in the choice set, how many depends on the value of λ_i
    - GPS estimate λ_i, and allow it to vary across patient characteristics

Identification

- How do they separately identify patient and physican preferences
  - they identify them separately because they observe a change in the process by which choice sets are formed due to the reform
  - patients preferences are identified from the post-reform period, where choice is freed up
  - physician's preferences are identified from the pre-reform period
    - if patient choice pre-reform was unconstrained (and if patient preferences are stable over time) then the post-reform estimates should predict hospital choice before the reform (if patients had choice)
    - if instead post-reform preferences do not predict referral patterns in the pre-reform time period, it has to be the case that the way in which referrals were made changed over time

TABLE 5—STRUCTURAL PARAMETER ESTIMATES

	Coefficient	Standard error
Patient preferences		
Distance	−6.983	0.211
Closest hospital dummy	1.341	0.052
Mortality rate	−7.883	2.229
Mortality rate × high severity	−5.419	2.467
Mortality rate × high income	3.832	2.320
Waiting times	−1.528	1.887
Waiting times × high severity	−1.584	1.140
Waiting times × high income	6.262	1.196
Physician preferences		
Distance	−4.985	0.207
Closest hospital dummy	1.734	0.110
Within-pct dummy	1.309	0.308
Choice constraint parameters		
Constant	0.000	0.119
High severity	1.011	0.178
High income	0.000	0.113

Welfare

- Difficult to measure welfare effects for a number of reasons
  - want to find the impact of removing restriction; find surplus in constrained world (just physician's utility function) and in unconstrained world (physician's and patients)
  - problem is they don't observe choice set in constrained world, and need that to measure surplus; they simulate choice sets
  - there is no price mechanism so surplus expressed in utils; can't translate to monetary terms; instead translate to distance (find equivalent to 15km reduction in travel costs)

Comparison of Gaynor Propper Seiler and Goeree

- Goeree (2008) models the probability of inclusion into the choice set; implies that the inclusion probabilities are independent separately for each product
  - in GPS a very attractive hospital can lead to a smaller choice sets by pushing other hospitals out of the choice set
- Goeree (2008) allows positive probability that no product is in the choice set
  - in GPS this is not allowed, the physician has to offer the patient at least one hospital
- Why don't GPS include expected patient utility directly in the physician's utility function?
  - this is defined over all possible choice set permutations, e.g. if 29 possible options is $(2^{29} - 1)$=536,870,911

Summary

- Unobserved and endogenous choice sets are a potentially big problem
- We have some econometric tools to deal with them, but they are limited in application
- Exciting on-going research and an area with opportunities for new research
  - with Greg Crawford and Alessandro Iaria we are developing more general methods based on extending the "differencing" idea of Chamberlain's fixed effect model