Lecture 3 February 2008

I. Household Decision Models and Children

One of the most important applications of household decision models is to the issue of children. Presumably, mothers and fathers make some joint decision about the allocation of family resources to their children. Whether they allocate "too little" to children is a question of much policy interest. We can illustrate some of the issues in a highly simplified model in which partners 1 and 2 have incomes y_1 and y_2, and individual consumption levels c_1 and c_2. Let k denote total spending on child-related expenses. In a unitary model we posit a family utility function $U(c_1, c_2, k)$, and a budget constraint $c_1 + c_2 + k - y_1 - y_2 = 0$. This approach implies optimal choices of the form $c_1 = c_1^*(y_1 + y_2)$, $c_2 = c_2^*(y_1 + y_2)$, $k = k^*(y_1 + y_2)$. Note that the choices depend only on pooled income $y_1 + y_2$, and not on "who earns what". This prediction has led to a variety of tests of the unitary model that involve testing whether y_1 and y_2 have the same impact on child spending. The classic study is

Duncan Thomas. "Intra Household Resource Allocation: An Inferential Approach". *JHR* 25 (1990).

More recent studies have looked at experimental manipulations of the income of one spouse – e.g. the Progressa experiment in Mexico, which sent income subsidies to the female heads of households in the treatment group.

What kind of model would NOT show pooling? Intuitively, it seems that a non-cooperative family in which the two spouses have different prefernces might exhibit such behavior. Suppose 1 and 2 have preference functions $U^1(c_1, k)$ and $U^2(c_2, k)$, and contribute amounts $k_1$and k_2 to child expenditures. In a non-cooperative model, each would choose c_i and k_i subject to the k-choice of the other:

$$\max_{c_i, k_i} U^i(c_i, k_i + k_{\sim i}) \quad s.t. \quad c_i + k_i = y_i.$$

(Such a model might apply to separated parents, for example). In this setup, child spending is a public good. Thus, we expect under-spending relative to a Pareto-efficient allocation. The f.o.c.'s for partner i can be re-written as

$$mrs_i = \frac{\partial U^i(c_i, k_i + k_{\sim i})/\partial k}{\partial U^i(c_i, k_i + k_{\sim i})/\partial c} = 1$$

where mrs_i is i's marginal rate of substitution (or "willingness to pay", the preferred term in Chicago) for the last unit of child spending. Since a dollar of spending on children costs $1, the Samuelson condition implies that at a Pareto efficient allocation

$$\frac{\partial U^1(c_1, k_1 + k_2)/\partial k}{\partial U^1(c_1, k_1 + k_2)/\partial c} + \frac{\partial U^2(c_2, k_1 + k_2)/\partial k}{\partial U^2(c_2, k_1 + k_2)/\partial c} = 1$$

Compared to the efficient solution, which has $mrs_1 + mrs_2 = 1$, the non-cooperative family has $mrs_1 + mrs_2 = 2$, a fairly large discprepancy!

Nevertheless, even a non-cooperative family might have spending outcomes that satisfy income-pooling. To see this, consider a simple Cobb-Douglas example:

$$U^i(c_i, k) = c_i^{\alpha_i} k^{1-\alpha_i}$$

Given incomes y_1 and y_2, the "reaction functions" of the two spouses are

$$\begin{aligned} k_1 &= (1-\alpha_1)y_1 - \alpha_1 k_2 \\ k_2 &= (1-\alpha_2)y_2 - \alpha_2 k_1 \end{aligned}$$

At the equilibrium:

$$k = k_1 + k_2 = \frac{(1-\alpha_1)(1-\alpha_2)}{1-\alpha_1\alpha_2}(y_1+y_2)$$

which satisfies income pooling, even when $\alpha_1 \neq \alpha_2$.

A third variety of family decision-making models is the so-called "efficient bargaining" models of Chiappori et al. Chiaporri and various co-authors have proposed that households maximize a weighted sum of the utilities of the two spouses, where the weight can depend on such things as relative income of the spouses, their relative attractiveness in the marriage market, etc.. Blundell, Chiappori and Meghir (*JPE* 2005) apply this framework to a case where children are a public good (similar to the one we are discussing, but with the added complication of endogenous labor supply). In this framework, a family solves:

$$\max_{c_1,c_2,k} \lambda U^1(c_1,k) + (1-\lambda)U^2(c_2,k) \;\; s.t. \;\; c_1+c_2+k = y_1+y_2,$$

where λ represents the relative "bargaining weight" of spouse 1. Since the budget constraint only depends on $y_1 + y_2$ income sources don't matter, and optimal child expenditure looks like $k = k^*(y_1+y_2, \lambda)$. The "trick", however, is that λ can potentially depend on y_1 and y_2. BCM show that if $\lambda = \lambda(z)$, and c and k are "normal" goods for both spouses, then a shift in z that raises the relative bargaining power of spouse 1 will raise spending on children if 1's mrs function is more responsive to increases in private consumption than 2's (quite a mouthful). This can be shown by differentiating the first order conditions for the co-operative maximization problem (these are: the Samuelson condition; the condition $\lambda\partial U^1/\partial c = (1-\lambda)\partial U^2/\partial c$; and the budget constraint).

Since the theory is silent on the determinates of λ, this class of models gives researchers a lot of flexibility in explaining how various things affect family outcomes. Its a matter of taste whether this is a "good thing". Moreover, welfare analysis is hard, because factors that shift λ shift the preferences of decision-making units. If the social welfare function gives one set of spouses (e.g., mothers) a higher relative weight than the family bargain at some value of the $z\prime s$, than targeted transfers will be welfare-enhancing. In this case, a finding that income sources matter suggests that there may be a way to change the welfare of children, even though bargaining is "efficient".

II. Discrete Choice Demand Models

In many settings, agents choose among a discrete set of alternatives. In labor economics, classic examples are: education levels, occupations and location. In IO the classic applications are to durable goods (cars, appliances). Typically, in labor economics the analyst observes a set of individuals (i=1..N) their characteristics X_i and the choice $j(i)$ that each made from the set of alternatives {1,2,..J}. In IO it is more often the case that one observes the market share of choice j – that is, the fraction of all consumers (in a given market) who selected the jth choice. In some applications (such as the paper by Hastings, Kane, and Staiger) we observe the preference rankings that individuals apply to some subset of choices (e.g., they report their "top three" choices").

The basic idea in discrete choice models is that individual i assigns utility u_{ij} to choice j, and selects the choice with highest utility. The event that i chooses j is denoted by the indicator $d_{ij} = 1$ (where $\sum_{j=1}^{J} d_{ij} = 1$). A good starting point is

$$u_{ij} = X_i\beta_j + Z_j\gamma_i + \epsilon_{ij} = v_{ij} + \epsilon_{ij}$$

which allows individual characteristics to have choice specific effects, and choice characteristics to have individual-specific effects. The term ϵ_{ij} can be interpreted as as unobserved component of tastes that is known to the agent but unknown to the analyst. Treating $\{\epsilon_{ij}\}$ as randomly distributed across the population,

$$P(d_{ij} = 1|X_i, Z_j) = P(v_{ij} + \epsilon_{ij} > v_{ik} + \epsilon_{ik} \quad for \quad all \quad k \neq j).$$

Observations:

(1) only relative utilities matter. If we add θ to every value of v_{ij} choices are the same

(2) scale is arbitrary. If we rescale $v_{ij} \to \lambda v_{ij}$, $\epsilon_{ij} \to \lambda\epsilon_{ij}$, choices are the same

(3) u_{ij}represents the indirect utility assigned by the agent to choice j. In general, then, u_{ij}should depend on income and the price of choice j...

(4) a very standard assumption is that there is an underlying quasi-linear direct utility function of a numeraire good n and the choice characteristics:

$$U^i(n, d_{ij}) = \alpha n + \phi_i(Z_j) + \epsilon_{ij}$$

If the jth choice has price p_j and agent i has income y_i the indirect utility of choice j is

$$u_{ij} = \alpha(y_i - p_j) + \phi_i(Z_j) + \epsilon_{ij}$$

which (using observation (1) is equivalent to)

$$u_{ij} = -\alpha p_j + \phi_i(Z_j) + \epsilon_{ij}$$

Quasi-linearity is appropriate for choice over "small" things (like brand of cereal) but is hard to justify for larger purchases (like cars) and is really problematic for houses. Quasi-linearity is convenient for calculating "willingness to pay", however. For example, suppose we assume $\phi_i(Z_j) = Z_j\gamma$ (ignoring any heterogeneity in γ for now). Then the marginal willingness to pay for the kth characteristic in Z is γ_k/α. As you recall from consumer demand theory, when preferences are quasi-linear the demands for characteristics Z have no income effects. This makes welfare evaluation extremely simple.

Multinomial Logit

The probability statement for the event $d_{ij} = 1$ involves a J-1 dimensional integral. For up to 3 choices, it is conventional to assume the ϵ_{ij}'s are normally distributed. Beyond that, the probability has to be evaluated by simulation methods. The usual approach for J>3 is multinomial logit (MNL). This is an extremely convenient form for a host of reasons that we will be exploring in the remainder of the lecture. A key lesson in structural microeconometrics is "know your logit".

A random variable ϵ with support on $(-\infty, +\infty)$ is distributed as EV-Type I if F(ϵ)=$e^{-e^{-\epsilon}}$. See Imbens and Wooldridge (Lecture 11) for a graph of the pdf of the EV-I vs. a standard normal. EV-I (a.k.a. Gumbell) has mode at 0, mean of τ =0.577 (Euler's constant) and variance of $\pi^2/6 \approx$1.65. In the 1970s McFadden showed that when the random components ϵ_{ij} of the indirect utilities associated with different choices are distributed as independent EV-I's,

$$P(d_{ij} = 1) = \frac{\exp(v_{ij})}{\sum_{k=1}^{J} \exp(v_{ik})}$$

In the case of only 2 choices, this boils down to a "logit". (A proof is presented in the Imbens-Wooldridge lecture). Consistent with observation (1) above, if we add a constant to each element of v_{ij} it cancels out of the numerator and denominator of the probability statement. This is an extremely convenient functional form!

A key feature of MNL is the so-called "IIA" (independence of irrelevant alternatives) property. If choices are generated by MNL

$$\frac{P(d_{i1} = 1)}{P(d_{i2} = 1)} = \frac{\exp(v_{i1})}{\exp(v_{i2})}$$

which says that the relative probability of choices 1 and 2 does not depend on the attributes of the other choices (they are "irrelevant"). This will not hold if a 3rd potential choice is available that is (say) very close to choice 2 and far away from choice 1. Then, when the 3rd is available demand for choice 2 will fall relative to 1, whereas when choice 3 is unavailable, people who would choice 2 or 3 all flock to 2. Some authors (e.g., Luce, 1959) have argued that if the consumer and choice characteristics are all fully specified then IIA "makes sense". See McFadden's Nobel Lecture (AER, 2001) for more on the history of IIA-related reasoning.

In some applications IIA is a critical plus! For example, suppose we want to forecast the demand for a product that does not exist, but whose characteristics are known. Suppose demand for products j=1...J are given by a MNL model with $u_{ij} = Z_j(\gamma_0 + \gamma_1 X_i) + \epsilon_{ij}$ (Here, we are allowing an interaction between consumer characteristics X_i and product characteristics Z_j – for example, number of seats in a car and number of kids in a family). In this case, if we can estimate the γ coefficients we can predict the demand for product J+1.

Another place where IIA really helps is in modeling choices when the choice set is very large (e.g., residential location). Suppose we observe individual i making choice j (e.g., they have chosen to live in Census tract j in a given metro area). Under IIA, we don't need to model all the choices that were potentially available: we can randomly select a subset of other choices (say, 3 alternatives), combine them with the one that was actually selected, and estimate the model as if each person had 4 choices and selected 1. This idea (introduced in a paper on residential choice by McFadden in 1978) is widely used in many applications. (The efficiency of this "conditional" likelihood is enhanced by including more alternatives in the choice set).

A third place where IIA helps is in interpreting preference rankings over varying choice sets (as in H-K-S). Suppose parent i is asked to rank 3 schools in order. IIA says that we can write the likelihood for the 3 choices as

$$P(1-2-3) = P(1st|3\ available)P(2nd|remaining\ 2).$$

Combining the previous two ideas, suppose we need to develop a likelihood for the top 3 stated choices over a very large choice set. Then we could augment each person's 3 choices with K others, randomly selected, and use a likelihood of the form:

$$P(1st|1st, 2nd, 3rd, K\ others) \times P(2nd|2nd, 3rd, K\ others) \times P(3rd|3rd, K\ others).$$

This would be a convenient way to estimate a model of school choice given an ordered list of colleges that each student applied to.

Application to Market Shares

Consider a model of choice where consumer i in market m assigns indirect utility u_{imj} to choice j:

$$\begin{aligned} u_{imj} &= \alpha(y_{im} - p_{mj}) + X_j\beta + \xi_{mj} + \epsilon_{imj} \\ &= \alpha y_{im} + \delta_{mj} + \epsilon_{imj}, \quad \text{where } \delta_{mj} \equiv X_j\beta - \alpha p_{mj} + \xi_{mj}. \end{aligned}$$

Here, ξ_{jm} represents a shared error component that shifts the demand of all consumers in market m. Assume in addition that consumer i has the "outside option" of not buying any of the choices, in which case utility is $\alpha y_{im} + \epsilon_{im0}$ (i.e., $\delta_{m0} = 0$). Assuming that the ϵ's are all EV-I:

$$P(d_{imj} = 1) = \frac{\exp(\delta_{mj})}{\sum_{k=0}^{J}\exp(\delta_{mk})} = \frac{\exp(\delta_{mj})}{1 + \sum_{k=1}^{J}\exp(\delta_{mk})} \quad .$$

If we have data on the fractions of consumers who choose each option in market m, then these market shares S_{mj} are consistent estimates of the probabilities $P(d_{imj} = 1)$. Berry (1994) introduced the idea of using the market shares to infer the β's in the presence of endogenous price-setting. Before preceding, note that if

$$S_{mj} = \frac{\exp(X_j\beta - \alpha p_{mj} + \xi_{mk})}{1 + \sum_{k=1}^{J}\exp(X_k\beta - \alpha p_{mk} + \xi_{mk})}$$

then

$$\begin{aligned} \partial S_{mj}/\partial p_{mj} &= \frac{-\alpha\exp(\delta_{mj})}{1 + \sum_{k=1}^{J}\exp(\delta_{mk})} + \frac{\alpha\exp(\delta_{mj})^2}{\{1 + \sum_{k=1}^{J}\exp(\delta_{mk})\}^2} \\ &= -\alpha S_{mj}(1 - S_{mj}) \end{aligned}$$

and thus the own-elasticity of demand is

$$\partial S_{mj}/\partial p_{mj} \times (p_{mj}/S_{mj}) = -\alpha p_{mj}(1 - S_{mj}).$$

Likewise,

$$\partial S_{mj}/\partial p_{mk} = \alpha S_{mj}S_{mk}$$

and thus the cross-price elasticity of demand is

$$\partial S_{mj}/\partial p_{mk} \times (p_{mk}/S_{mj}) = \alpha p_{mk}S_{mk}.$$

Given market shares, all the own- and cross-price elasticities depend on only 1 parameter, α. Obviously, this is too restrictive for most applications.

Nevertheless, suppose we wanted to estimate α from market share data. A concern is that in markets where $\xi_{mj} > 0$, the producer of product j will set a higher price, leading to a simultaneity bias in direct estimation of the MNL. Berry noted that:

$$\begin{aligned}
\log S_{jm} &= \delta_{mj} - \log\{1 + \sum_{k=1}^{J} \exp(\delta_{mk})\} \\
\log S_{j0} &= -\log\{1 + \sum_{k=1}^{J} \exp(\delta_{mk})\} \\
&\Rightarrow \log(S_{jm}/S_{j0}) = \delta_{mj} = X_j\beta - \alpha p_{mj} + \xi_{mj}.
\end{aligned}$$

He proposed taking the estimates of S_{jm} and $S_{j0}$and forming an estimate $\widehat{\delta}_{mj}$, then estimating:

$$\widehat{\delta}_{mj} = X_j\beta - \alpha p_{mj} + \xi_{mj} + (\widehat{\delta}_{mj} - \delta_{mj})$$

by IV, using instruments for p_{mj} that are arguably orthogonal to ξ_{mj}. Instruments that have been proposed include:

(a) prices in other markets for the same product (Hausman, Nevo)

(b) characteristics of other products in the same market (Bresnahan, BLP).

We'll come back to these ideas next lecture when we discuss BLP, which extends the Berry idea to the "mixed logit" case. In labor economics, the analogue of "market shares" could be: (i) the fraction of a cohort (or the fraction in a given local labor market) who choose different levels of education; (ii) the fraction of people who choose different neighborhoods or schools...

Breaking Out of IIA - Mixed Logit

There are two basic extensions of MNL that have been proposed: nested logit and other "generalized extreme value" (GEV) distributions for the ϵ's; and "mixed logit". We defer the former for two lectures and focus on the latter. The idea of mixed logit is that the population consists of a variety of consumers, each of whom have their own parameters. Thus, a given consumer (or class of consumers) follow the MNL model, but in the population as a whole we see a mixture of various types. Consider an extension of the market model just presented:

$$\begin{aligned}
u_{imj} &= \alpha(y_{im} - p_{mj}) + X_j(\beta + \nu_{im}) + \xi_{mj} + \epsilon_{imj} \\
&= \alpha y_{im} + \delta_{mj} + X_j\nu_{im} + \epsilon_{imj}
\end{aligned}$$

where ν_{im}represents a random vector summarizing the deviation of consumer i's preferences from those of the population as a whole. If $F(\nu_{im})$ represents the d.f. for ν_{im}, then

$$\begin{aligned}
S_{mj} &= \int P(d_{imj} = 1|\nu_{im})dF(\nu_{im}) \\
&= \int P_{imj}dF(\nu_{im})
\end{aligned}$$

where

$$P_{imj} = \frac{\exp(X_j(\beta + \nu_{im}) - \alpha p_{mj} + \xi_{mj})}{1 + \sum_{k=1}^{J} \exp(X_k(\beta + \nu_{im}) - \alpha p_{mj} + \xi_{mj})}.$$

Following the earlier development:

$$\partial S_{mj}/\partial p_{mj} = \int -\alpha P_{imj}(1 - P_{imj})dF(\nu_{im})$$

and

$$\partial S_{mj}/\partial p_{mk} = \int \alpha P_{imj} P_{imk} dF(\nu_{im})$$

Consider the own-price derivative. Notice that if some consumers always choose a particular product, then they will have $P_{imj}(1 - P_{imj})$ close to 0 for all j. Other consumers, for a certain range of prices, may be have a probability of purchasing choice j that is close to 1/2, maximizing $P_{imj}(1-P_{imj})$. The integral formula frees up the connection between the marginal response of market share to price and the absolute size of market share that arises in the MNL with no heterogeneity. A similar argument applies to the cross-price derivative.

Estimation of MNL

With micro data, the likelihood that consumer i in market m is observed making choice j is

$$\int \{ \frac{\exp(\delta_{mj} + X_j \nu_{im} + \epsilon_{imj})}{1 + \sum_{k=1}^{J} \exp(\delta_{mk} + X_k \nu_{im} + \epsilon_{imk})} \} dF(\nu_{im})$$

Assuming a particular distribution for ν_{im} across the population this can be estimated by simulation methods, yielding estimates of the δ's and the parameters of $F(\nu_{im})$. A second stage estimation of the type suggested by Berry could then be conducted.

Note that mixed logit can be used in any general choice setting where the analyst is trying to break out of a strict logit form. To aid in discussion of some of the general issues that arise, lets simplify the indirect utility model to the following form:

$$u_{ij} = X_{ij}\beta_i + \epsilon_{ij}$$

The probability that consumer i makes choice j is

$$P(d_{ij} = 1|X_{ij}) = \int \frac{\exp(X_{ij}\beta_i)}{\sum \exp(X_{ik}\beta_i)} dF(\beta_i).$$

Typically, the variation in β is restricted to a few key coefficients: for example, the coefficient on the price of alternative j, and some other crucial characteristic. Thus, $\beta = (\beta_1, \beta_2)$ where only $\beta_1$is random. A common mixing distribution is the normal or log-normal. The integral is evaluated as a sum across r=1...R "replications".

$$P(d_{ij} = 1|X_{ij}) \approx \sum_{r=1}^{R} \frac{\exp(X_{ij}\beta_r)}{\sum \exp(X_{ik}\beta_r)},$$

where in the rth replication $\beta_1 = \beta_{1r}$. Consider the case where β_1 is assumed to be normally distributed. Recall that to generate a random variate β_1 from a N(μ ,Σ) distribution, one draws a set (z_1,...z_K) of i.i.d. N(0,1) variates (where K is the number of elements of β_1), and transforms $\beta_{1r} = \mu + T(\Sigma)$z, where $T(\Sigma)$ is the Cholesky decomposition of Σ: $TT' = \Sigma$. In the estimation, sets of R z's are drawn once for each observation i, and held constant, as choices are made for the underlying parameters (μ, Σ, β_2). The reason the "replicants" are held constant is to ensure that if you evaluate the likelihood at a choice for (μ, Σ, β_2) you will get exactly the same value every time (that would not be true if one "resampled" the replicants each iteration). Moreover, if you evaluate at (μ, Σ, β_2).close to some initial values, the likelihood will not change "much".

Chiou and Walker (2007) have pointed out that when MNL is estimated by simulation, it can happen that the estimates converge even when the model is not identified. Behavior of standard monte-carlo simulation estimators is not entirely well understood!