Economics 250a

Lecture 1: A very quick overview of consumer choice.

1. Review of basic consumer theory
2. Functional form, aggregation, and separability
3. Discrete choice

Some recommended readings:

Angus Deaton and John Muellbauer, *Economics and Consumer Behavior*, Cambridge Press, 1980

Geoffrey Jehle and Philip Reny, *Advanced Microeconomic Theory* (2nd ed), Addision Wesley, 2001

John Chipman, "Aggregation and Estimation in the Theory of Demand" *History of Political Economy* 38 (annual supplement), pp. 106-125.

Kenneth Train. *Discrete Choice Methods with Simulation*, Cambridge Press 2003.

Kenneth A. Small and Harvey A. Rosen "Applied Welfare Economics with Discrete Choice Models." *Econometrica*, 49 (January 1981), pp. 105-130.

1. Review of basic consumer theory

a. Basic assumptions and the demand function.

We start with a choice set X (closed, bounded below); often (though not always) X=$\mathbb{R}^{\ell}_{+}$, the positive orthant. Preferences over alternatives in X are represented by $u : X \to \mathbb{R}$, twice continuously differentiable, strictly increasing, strictly quasi concave. (u is *s.q.c.* if $u(x^1) \geq u(x^2) \Rightarrow u(\alpha x^1 + (1-\alpha)x^2) > u(x^2)$ for $\alpha \in (0,1]$).

The demand function $x(p,I)$ maps from prices $p \in \mathbb{R}^{\ell}_{++}$ and income $I \in \mathbb{R}_+$ to a "preference maximal choice" $x \in X$:

$$x(p,I) = \arg\max_{x \in X} u(x) \quad s.t. \quad px \leq I.$$

Under the preceding assumptions $x(p,I)$ exists and is well-defined, and is continuous in (p,I) for all (p,I) such that the interior of the "budget set" $= \{x \in X : px \leq I\}$ is non-empty. $x(p,I)$ is also homogeneous of degree 0 (HD$_0$) in (p,I). An interior choice $x^0 = x(p^0, I^0)$ satisfies the first order conditions:

$$\begin{aligned} Du(x^0) &= \lambda^0 p^0 \\ p^0 x^0 &= I^0 \end{aligned}$$

for some number $\lambda^0 > 0$, and the second order conditions

$$t' D^2 u(x^0) t < 0 \text{ for all } t \neq 0 \text{ with } tp^0 = 0.$$

Assuming $u(x)$ is *s.q.c.*, the first order conditions are in fact both necessary and sufficient to characterize an interior "preference maximal" choice. (see Jehle and Reny, Theorem 1.4).

Using the implicit function theorem, the derivatives of $x(p,I)$ can be obtained by differentiating the $\ell+1$ first order conditions w.r.t. the $\ell+1$ endogenous variables (x,λ). We pursue this in subsection c.

b. Expenditure and indirect utility functions.

For $p \in \mathbb{R}^{\ell}_{++}$ and u in the range of $u(x)$, define

$$e(p,u) = \min_{x \in X} px \quad s.t. \quad u(x) \geq u,$$

and the associated cost-minimizing consumption choice (or "Hicksian demands"):

$$h(p,u) = \arg\min px \quad s.t. \quad u(x) \geq u.$$

The expenditure function $e(p,u)$ is a an example of a "support function" (see Mas Colell, Whinston and Green) and is therefore concave and HD_1 in prices. Under the previous assumptions on $u(x)$, $h(p,u)$ is a well defined function (i.e., the cost-minimizing choice exists and is unique) and using a basic fact about support functions, e is therefore differentiable in p with

$$D_p e(p,u) = h(p,u). \quad \text{(Sheppard's lemma)}.$$

Another very useful fact is that $h(p,u)$ is HD_0 in prices.

The "inverse" of the expenditure function (as a function of u, holding constant p) is the indirect utility function

$$v(p,I) = \max_{x \in X} u(x) \quad s.t. \quad px \leq I.$$

With $u(x)$ strictly increasing

$$\begin{aligned} v(p, e(p,u)) &= u \\ e(p, v(p,I)) &= I. \end{aligned}$$

The indirect utility function $v(p,I)$ is strictly quasi-convex in prices, and HD_0 in (p,I). From the envelope theorem, $DV_I(p,I) = \lambda(p,I)$ (the λ from the first order conditions). We also have "Roy's identity":

$$x(p,I) = \frac{-DV_p(p,I)}{DV_I(p,I)}.$$

Finally, a very useful relationship:

$$h(p,u) = x(p, e(p,u)).$$

This holds identically in p, and assuming $x(p,I)$ is differentiable at some (p^0, I^0), h is differentiable at (p^0, u^0) (where $u^0 = v(p^0, I^0)$) with

$$\begin{aligned} \frac{\partial h_i(p^0,u^0)}{\partial p_j} &= \frac{\partial x_i(p^0,I^0)}{\partial p_j} + \frac{\partial x_i(p^0,I^0)}{\partial I}\frac{\partial e(p^0,u^0)}{\partial p_j} \\ &= \frac{\partial x_i(p^0,I^0)}{\partial p_j} + \frac{\partial x_i(p^0,I^0)}{\partial I} x_j(p^0,I^0)'. \end{aligned}$$

Re-arranging this we have the "Slutsky decomposition":

$$\frac{\partial x_i(p^0,I^0)}{\partial p_j} = \frac{\partial h_i(p^0,u^0)}{\partial p_j} - \frac{\partial x_i(p^0,I^0)}{\partial I} x_j(p^0,I^0).$$

The change in the demand for good i with respect to a change in p_j consists of two terms: the "substitution effect" captured by the change in the Hicksian demand, and the "income effect" which is proportional to the amount of good j originally purchased. The matrix of terms

$$S_{ij} = \frac{\partial x_i(p^0, I^0)}{\partial p_j} + \frac{\partial x_i(p^0, I^0)}{\partial I} x_j(p^0, I^0)$$

is called the "Slutsky matrix" (and can be defined knowing only the Marshallian demand functions). The Slutsky decomposition shows that

$$S_{ij} = \frac{\partial h_i(p^0, u^0)}{\partial p_j} = \frac{\partial^2 e(p^0, u^0)}{\partial p_i \partial p_j}$$

has to be symmetric (from Young's theorem) and negative semi-definite (from concavity of e).

c. Using the expenditure function

The expenditure function has many uses in applied micro. One is in the analysis of changes in the cost of living. Consider a consumer who is observed in a base year facing prices p^0 and making consumption choices x^0. Define $u^0 = u(x^0)$. When prices change to p^1, the consumer would need minimum income $e(p^1, u^0)$ to achieve the same utility. So the rise in the "true cost of living index" is

$$COL(p^1; p^0) = \frac{e(p^1, u^0)}{e(p^0, u^0)}.$$

The standard Laspayre's index, on the other hand, is

$$L(p^1; p^0) = \frac{p^1 x^0}{p^0 x^0}.$$

Using a second-order expansion for e:

$$\begin{aligned} e(p^1, u^0) &\approx e(p^0, u^0) + (p^1 - p^0)' \frac{\partial e(p^0, u^0)}{\partial p} + \frac{1}{2}(p^1 - p^0)' \frac{\partial^2 e(p^0, u^0)}{\partial p \partial p'}(p^1 - p^0) \\ &= p^1 x^0 + \frac{1}{2}(p^1 - p^0)' S (p^1 - p^0) \end{aligned}$$

where we are using the facts that

$$\begin{aligned} e(p^0, u^0) &= p^0 x^0, \\ \frac{\partial e(p^0, u^0)}{\partial p} &= h(p^0, u^0) = x^0, \end{aligned}$$

and that the Slutsky matrix S is just the Hessian of the expenditure function. Combining terms and substituting we get:

$$COL(p^1; p^0) = L(p^1; p^0) + \frac{1}{2} \frac{\Delta p' S \Delta p}{p^0 x^0}$$

and since S is negative semi definite, $COL \leq L$, with a gap that depends on the magnitude of the Slutsky (substitution) terms. (With Leontief preferences $S = 0$ and the Laspeyre's index is exact). The gap is known as the "substitution bias" in the Laspeyre's index formula. We can also use this expression to show that when

$$\Delta p = \lambda p^0 \text{ for some positive scalar } \lambda$$

(i.e., all prices rise by the same proportion λ) then there is no bias. The reason is that Hicksian demands are homogeneous of degree 0. By Euler's theorem, then,

$$\frac{\partial h(p^0, u^0)}{\partial p'} p^0 = 0 \Rightarrow Sp^0 = 0.$$

(The Slutsky matrix always has maximal rank $n-1$, where n is the number of commodities. In particular, the original price vector p^0 is in the null space of S).

Another important use of the expenditure function is to calculate equivalent and compensating variations associated with price changes, and to calculate the "excess burden" of a tax (see problem set #1).

d. Integrability

For an arbitrary "candidate" demand function $x(p, I)$ that has the "right dimensions," is homogeneous of degree 0, and is continuously differentiable we can define the matrix of substitution effects:

$$S(p, I) = D_p x(p, I) + D_I x(p, I) \cdot x(p, I)'$$

If $x(p, I)$ was generated by a utility maximizing consumer, $S(p, I)$ would have to be symmetric and n.s.d. The "integrability theorem" says that the converse if true: if you start with a candidate demand function $x(p, I)$ that has the right dimensions, is homogeneous of degree 0, and is continuously differentiable, and if the associated $S(p, I)$ matrix is everywhere symmetric negative semi definite you can "integrate" the demand function to find a utility function that generates $x(p, I)$. The basic idea is to find a solution to the set of differential equations

$$\frac{\partial e(p, u)}{\partial p_j} = x_j(p, e(p, u)) \equiv f_j(p).$$

A sufficient condition for a solution function $e(p, u)$ to exist for a system of p.d.e.'s like this is that

$$\frac{\partial f_j(p)}{\partial p_k} = \frac{\partial f_k(p)}{\partial p_j}$$

(this is known as Frobenius' theorem). Carrying out the differentiation you will see this requires Slutsky symmetry for the candidate demand functions. For the solution function $e(p, u)$ to be a valid expenditure function it has to be concave. This requires that the Slutsky matrix obtained from the candidate demands is negative semi definite.

e. Derivation of the Slutsky Decomposition from the First Order Conditions

Although it is a little "algebra-intensive" some additional insights can be gained by differentiating the first order conditions and looking at the implications for the derivatives of demand. We will assume for this section that $u(x)$ is "strictly concave" – in particular that $D^2 u(x)$ is negative definite. We'll come back to discuss the "cardinalization" of utility at the end of the section.

Start with the first order conditions

$$\begin{aligned} Du(x) &= \lambda p \\ px - I &= 0. \end{aligned}$$

Differentiate w.r.t (p, I) to get an $\ell+1$ by $\ell+1$ system of equations:

$$\begin{bmatrix} D^2u(x) & p \\ p' & 0 \end{bmatrix} \cdot \begin{bmatrix} \frac{\partial x}{\partial p'} & \frac{\partial x}{\partial I} \\ -\frac{\partial \lambda}{\partial p'} & -\frac{\partial \lambda}{\partial I} \end{bmatrix} = \begin{bmatrix} \lambda I_{\ell,\ell} & 0 \\ -x' & 1 \end{bmatrix}$$

where $I_{\ell,\ell}$ is the $\ell \times \ell$ identity matrix (apologies for the dual use of "I"). Thus

$$\begin{bmatrix} \frac{\partial x}{\partial p'} & \frac{\partial x}{\partial I} \\ -\frac{\partial \lambda}{\partial p'} & -\frac{\partial \lambda}{\partial I} \end{bmatrix} = \begin{bmatrix} D^2u(x) & p \\ p' & 0 \end{bmatrix}^{-1} \begin{bmatrix} \lambda I_{\ell,\ell} & 0 \\ -x' & 1 \end{bmatrix}.$$

Let $D^2u(x) = U$ (assumed to be invertible) Then the partitioned inverse formula can be used to show

$$\begin{bmatrix} D^2u(x) & p \\ p' & 0 \end{bmatrix}^{-1} = (p'U^{-1}p)^{-1} \begin{bmatrix} (p'U^{-1}p)U^{-1} - U^{-1}pp'U^{-1} & U^{-1}p \\ p'U^{-1} & -1 \end{bmatrix}$$

With this in hand, it follows that

$$\begin{aligned} \frac{\partial \lambda}{\partial I} &= (p'U^{-1}p)^{-1} < 0, \\ \frac{\partial x}{\partial I} &= \frac{\partial \lambda}{\partial I} U^{-1}p, \\ \frac{\partial x}{\partial p'} &= \lambda U^{-1} - \lambda(\frac{\partial \lambda}{\partial I})^{-1}(\frac{\partial x}{\partial I})(\frac{\partial x}{\partial I})' - (\frac{\partial x}{\partial I})x'. \end{aligned}$$

From the Slutsky equation we know

$$\frac{\partial x}{\partial p'} = S(p, I) - (\frac{\partial x}{\partial I})x'.$$

Thus we have shown that

$$S(p, I) = \lambda U^{-1} - \lambda(\frac{\partial \lambda}{\partial I})^{-1}(\frac{\partial x}{\partial I})(\frac{\partial x}{\partial I})'$$

the matrix of substitution terms consists of two parts:λU^{-1} which directly reflects the second cross-partials of $u(x)$, and a second term that is related to the income effect. (Inspection of this equation shows that S has to be symmetric: its more work to show it is n.s.d., but it can be done).

Look back at the first order conditions and suppose we define $x^F(p, \lambda)$ as the choice of x that satisfies the f.o.c. as we vary p, holding constant λ, i.e., define $x^F(p, \lambda)$ implicitly by:

$$Du(x^F(p, \lambda)) = \lambda p.$$

These are known as the "Frisch" (or λ-constant) demands, and play an important role in the modeling of intertemporal choice. Differentiating w.r.t. prices:

$$D^2u() \cdot \frac{\partial x^F}{\partial p'} = \lambda I_{\ell,\ell} \Rightarrow \frac{\partial x^F}{\partial p'} = \lambda U^{-1}.$$

Thus, we can decompose

$$\frac{\partial h}{\partial p'} = S(p, I) = \frac{\partial x^F}{\partial p'} - \lambda(\frac{\partial \lambda}{\partial I})^{-1}(\frac{\partial x}{\partial I})(\frac{\partial x}{\partial I})'.$$

Focusing on the (i,j) element, we have:

$$\begin{aligned} \frac{p_j}{x_i}\frac{\partial h_i}{\partial p_j} &= \frac{p_j}{x_i}\frac{\partial x_i^F}{\partial p_j} - [\frac{I}{\lambda}(\frac{\partial \lambda}{\partial I})]^{-1}(\frac{I}{x_i}\frac{\partial x_i}{\partial I})(\frac{I}{x_j}\frac{\partial x_j}{\partial I})\frac{p_j x_j}{I} \\ \sigma_{ij} &= f_{ij} - \omega^{-1}\epsilon_i\epsilon_j w_j, \end{aligned}$$

where σ_{ij} is the compensated elasticity, f_{ij} is the Frisch elasticity, ϵ_j is the income elasticity of demand for good j, w_j is the share of the budget spent on good j, and

$$\omega = \frac{I}{\lambda}(\frac{\partial \lambda}{\partial I})$$

is the "income flexibility" of the marginal utility of income. We will show later on that in intertemporal choice problems with additive separability between periods, ω is related to the so-called "intertemporal substitution elasticity", a concept that is central to modern macro.

When there is no income effect in the demand for a good, the derivatives of the ordinary (or "Marshallian) demand function, the Hicksian demand function, and the Frisch demand function are all equal. More generally the three are different. For example, suppose $u(x) = \sum_{i=1}^{\ell} v_i(x_i)$, "additively separable" preferences. These are widely used in the study of intertemporal consumption. With additive separability D^2u is diagonal, and all the off-diagonal terms in the matrix of Frish derivatives $(= \lambda U^{-1})$ are zero. In this case, for $i \neq j$, we have

$$\sigma_{ij} = -\omega^{-1}\epsilon_i\epsilon_j w_j.$$

All the "cross substitution" effects in the compensated demands arise though re-adjustments of λ. This is why additive separability is so convenient – it means that all the impact of prices of other goods (or other periods, if each good is consumption in a different period) is channelled through λ.

There is a "problem" with this decomposition: it is not invariant to transformations of $u(x)$ (whereas the Slutsky decomposition *is* invariant). In some contexts this is ok: for example, in the study of intertemporal consumption under uncertainty, one cannot renormalize the intertemporal utility function without losing intertemporal separability (and changing the degree of risk aversion). We will discuss in lecture 3 the relation between risk aversion and the elasticities of the Frisch demands.

2. Functional form, Aggregation, and Separability

a. The classic aggregation results

A longstanding question in demand theory is how to relate the demand system created by a collection of consumers to the underlying demands of the individual consumers. In particular, prior to the availability of micro data, economists asked whether there are assumptions on preferences such that aggregrate demand is generated by a "representative consumer" with "rationalizable" preferences. A nice reference is the 2006 paper by Chipman, which is not too technical but gives a good flavor of the results.

Preferences $(\succeq)$ are homothetic if $x^1 \succeq x^2 \Leftrightarrow \beta x^1 \succeq \beta x^2$ for any $\beta > 0$. With homothetic preferences there is only 1 indifference curve: any indifference curve is a "radial blow-up" of any other. It is intuitively obvious but surprisingly hard to prove that the demand system can be written as $x(p,I) = I \cdot h(p,1) = I \cdot \beta(p)$ iff preferences are homothetic. With homothetic preferences all income elasticities are equal to 1 - a restriction that appears to be false for many goods. A classic result is that with identical homothetic preferences, aggregate demand is "as if" there were a single consumer with the same preferences and the total income of all consumers. (The proof is easy and left as an exercise). A more subtle result is that if different consumers have different homothetic preferences, and each consumer has a fixed share of total aggregate income (as prices and total income are varied) then aggregate demand is "as if" there were a single consumer with some homothetic preference ordering (see Chipman).

Preferences are "quasi-homothetic" if they give rise to a demand function of the form:

$$x(p, I) = \alpha(p) + I \cdot \beta(p).$$

The classic example of quasi-homothetic preferences is "Stone-Geary" preferences (please excuse the re-use of α, β):

$$u(x_1, x_2, ...x_\ell) = (x_1 - \alpha_1)^{\beta_1}(x_2 - \alpha_2)^{\beta_2}...(x_\ell - \alpha_\ell)^{\beta_\ell} \text{ with } \sum_j \beta_j = 1.$$

These are "translated" Cobb Douglas preferences in which the consumer has "minimum needs" $\alpha_1, \alpha_2, ...\alpha_\ell$. The demand function for good i can be written as

$$x_i(p, I) = \alpha_i + \frac{\beta_i}{p_i}(I - p_1\alpha_1 - ..p_\ell\alpha_\ell)$$

while optimal expenditures on good i can be written as

$$e_i(p, I) \equiv p_i x_i(p, I) = p_i\alpha_i + \beta_i(I - p_1\alpha_1 - ..p_\ell\alpha_\ell)$$

which is linear in prices and incomes. (Hence, Stone Geary demands are also called the "linear expenditure system", LES).

If consumer k has quasi-homothetic preferences with demand function

$$x^k(p, I^k) = \alpha^k(p) + I^k \cdot \beta(p),$$

i.e., a person-specific $\alpha(p)$ function, but indentical $\beta(p)$ functions, then aggregate demand is

$$\begin{aligned} x(p, I) &= \sum_k x^k(p, I^k) = \alpha(p) + I \cdot \beta(p), \quad \text{where} \\ \alpha(p) &= \sum_k \alpha^k(p), \text{ and } \ I = \sum_k I^k \end{aligned}$$

Thus aggregate demand is quasi-homothetic. Gorman (1961) showed that quasi-homothetic demands of the form

$$x(p, I) = \alpha(p) + I \cdot \beta(p).$$

are generated by a consumer with expenditure function

$$\begin{aligned} e(p, u) &= a(p) + u \cdot b(p), \text{ where } a(p) \text{ and } b(p) \text{ are HD}_1 \text{ in prices and} \\ \beta_j(p) &= \frac{1}{b(p)} \frac{\partial b(p)}{\partial p_j} \quad \text{and} \\ \alpha_j(p) &= \frac{\partial a(p)}{\partial p_j} - \beta_j(p) a(p) \end{aligned}$$

The proof follows by applying Sheppard's lemma to the conjectured expenditure function, and using the expressions for $a_j(p)$ and $b_j(p)$. (See Deaton and Muellbauer exercise 6.3 for hints on the converse, which is also true). Note that an expenditure function of the form

$$e(p,u) = a(p) + u \cdot b(p)$$

(known as "Gorman polar form") implies an indirect utility function of the form:

$$v(p,I) = (I - a(p))/b(p).$$

In the Gorman polar form the function $a(p)$ is a "subsistence spending amount", and $b(p)$ is a price index that deflates income over and above subsistence. Note that if $\beta(p)$ (from the demand function) is restricted to be the same across consumers then $b(p)$ (from the expenditure function) must the the same across consumers

For the Stone Geary case described above, the expenditure function is $\sum_j p_j \alpha_j + u \prod_j p_j^{\beta_j}$ which is in the Gorman polar form class. If everyone has the same $\beta's$ but person-specific $\alpha's$ aggregate demand will come from the Stone Geary class with the aggregate α simply summing the individual $\alpha's$.

Homothetic and quasi-homothetic preferences (with the same $\beta(p)$ function in the demand equation) are effectively the only ones that aggregate.

b. More on Functional Form

In most modern microeconometric work, the researcher chooses a functional form for $u(x)$ that incorporates observed and unobserved heterogeneity in a convenient way, and then develops an estimator for unknown parameters of $u(x)$. For example, one could choose Stone Geary preferences, and allow α to depend on observed characteristics (e.g., family size). It is not entirely clear how to account for unobserved sources of heterogeneity. In the pre-1980's literature it was common to "tack on" error components without really specifying how or where the errors came from. In the more recent literature a substantial premium is placed on developing a model that provides a complete description of the data generating process (dgp), by positing an explicit distribution for the unobserved components..

To illustrate, suppose a sample of data on individuals in different markets is available. In a given market, everyone pays the same prices (p), and consumer k has incomeI^k and observed expenditure amounts $(e_1^k, e_2^k, ...e_\ell^k)$ on products $1, 2, ...\ell$. The LES model asserts that the amount spent on good i by consumer k is

$$e_i^k(p,I) = p_i \alpha_i^k + \beta_i(I - p_1\alpha_1^k - ..p_\ell \alpha_\ell^k)$$

We might want to assume that the $\beta's$ are constant across the population and allow the $\alpha's$ to vary. So for example we could assume $\alpha_i^k = X^k\theta_i + \xi_i^k$, where $(\xi_1^k, \xi_2^k, ...\xi_\ell^k)$ represents a random vector that is known by consumer k but unobserved by the data analyst (and assumed to have some distribution across the population) A problem for the LES is that the cost function is only concave for people who have enough income to "cover" their minimum spending threshold $\sum_j p_j \alpha_j$: if the $\xi^{k\prime}s$ have unlimited support concavity will be violated (with high probability). Another problem is that all of the $\alpha's$ enter into each expenditure equation, leading to a rather messy model in which the stochastic term in expenditures is a price-weighted average of the $\xi^{k\prime}s$.

Rather than model expenditures as a linear function of prices and incomes, in many cases it is more convenient to model expenditure shares ($w_i = p_i x_i / I$) as functions of the logs of prices and incomes. This is the approach taken by Deaton and Muellbauer in their famous "almost-ideal demand system" (known at least originally as the AIDS system). D+M begin by focusing on a class of expenditure functions that satisfy

$$\log e(p, u) = a(p) + ub(p).$$

An initial observation is that this choice gives rise to expenditure share equations of the form

$$w_i = A_i(p) + B_i(p) \log I.$$

(this can be seen by applying Sheppard's lemma and a little algebra). Running expenditure shares on the log of total income (or total expenditure) has a long history in applied demand analysis. D+M propose to use

$$\begin{aligned} a(p) &= \alpha_0 + \sum_j \alpha_j \log p_j + \frac{1}{2} \sum_j \sum_k \gamma^*_{jk} \log p_j \log p_k \\ b(p) &= \beta_0 \prod\nolimits_k p_k^{\beta_k}, \quad \text{with} \\ \sum_j \alpha_j &= 1 \quad \text{and} \quad \sum_j \gamma^*_{jk} = \sum_k \gamma^*_{jk} = \sum_j \beta_j = 0 \end{aligned}$$

Substituting the expressions for $a(p)$ and $b(p)$ leads to budget share equations of the form

$$w_i = \alpha_i + \sum_j \gamma_{ij} \log p_j + \beta_i \log(\frac{I}{P})$$

where P is a price index:

$$\log P = \alpha_0 + \sum_j \alpha_j \log p_j + \frac{1}{2} \sum_j \sum_k \gamma_{jk} \log p_j \log p_k$$

and $\gamma_{ij} = (\gamma^*_{ij} + \gamma^*_{ji})/2 = \gamma_{ji}$.

For many applications D-M note that the price index will be relatively well approximated by a simple index of the form

$$\log P = \sum_j w_j \log p_j$$

With this approximation, estimation is trivial. A question for further thought: how would you incorporate observed and unobserved heterogeneity into the AIDS system?

3. 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 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 $\sum_{j=1}^{J} d_{ij} = 1$).

The modern literature on discrete choice arises from seminal work by Dan McFadden in the early 1960s. The idea in this literature is that we (as econometricians) observe some characteristics of people (like their income and family size) and some characteristics of the choices (like the price of a given choice) but we admit from the beginning that there is some component of preferences that is unobserved by us, which we will treat as random. Thus, discrete choice models are often called "random utility" models. Utility is not random for the agents themselves (that is a "post-modern" idea) but is modeled as random.

A good starting point is a model where an agent has to choose 1 of a set of alternatives. Write the maximized utility that agent i will receive if she selects choice j (and optimally chooses all the other stuff she consumes, conditional on that choice) as:

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

This allows individual characteristics X_i to have choice specific effects (the $\beta_j's$), and the characteristics of the choices (Z_j) to be valued by different agents differently (the $\gamma_i's$).The term ϵ_{ij} is the unobserved component of tastes that is known to the agent but unknown to the analyst. Treating $\{\epsilon_{ij}\}$ as randomly distributed across the population, we get a *probability statement* for the event that i chooses j:

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

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 j^{th} 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. When preferences are quasi-linear the demands for characteristics Z have no income effects. This makes welfare evaluation extremely simple.

Small and Rosen (1981) consider a slightly more general version of this model, where agents can choose variable quantities of each "distinct good" (agents have to select one of the $j's$, but the amount they buy of that choice is potentially variable). They show that within this class of models, one can still define an expenditure function, that the derivative of the expenditure function w.r.t. the price of the *jth* choice is the compensated demand for that choice (which can be 0), and one can also define the equivalent and compensating variations associated with price changes.

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 (NBER "What's New In Econometrics", 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. 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.

Market Shares

The MNL model aggregates in a very convenient way. 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:

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

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 $P(d_{imj} = 1)$. Thus

$$\begin{aligned} \log S_{mj} &= \delta_{mj} - \log\{1 + \sum_{k=1}^{J} \exp(\delta_{mk})\} + sampling\ error \\ \log S_{m0} &= -\log\{1 + \sum_{k=1}^{J} \exp(\delta_{mk})\} + sampling\ error \\ &\Rightarrow \log(S_{mj}/S_{m0}) = \delta_{mj} = X_j\beta - \alpha p_{mj} + \xi_{mj}. \end{aligned}$$

So, the simple choice implies that the log of the market share of choice j, relative to the share in market m who choose the outside option, is linear in characteristics, price, and an error term that reflects the market-specific preferences of consumers for choice j. In applications, the share of the outside good in market m is sometimes hard to estimate.