Economics 250c
Fall 2008, Lecture 6
This lecture will discuss two topics:
1. The mapping between choice probabilities and conditional valuations
2. Introduction to dynamic discrete choice problems

1. Choice probabilities and conditional valuations

a. Prologue
Consider the familiar two-sector choice model. Individual i can choose a job in one of two sectors, $j = 1, 2$, with

$$u_j = X_j\beta_j + \epsilon_j = v_j + \epsilon_j$$

A canonical example (from the ice age of labor economics) would be choosing a union or non-union job. In that case, the so-called "conditional valuation" v_j could represent the expected wage in sector j. The probability that i chooses sector 1 (denoted by $d_1 = 1$) is:

$$\begin{aligned} P(d_1 &= 1|v_1, v_2) = P(v_1 + \epsilon_1 > v_2 + \epsilon_2) \\ &= P(\epsilon_1 - \epsilon_2 > v_2 - v_1) \\ &= P(\xi > v_2 - v_1), \quad \text{where } \xi \equiv \epsilon_1 - \epsilon_2. \end{aligned}$$

In the standard bivarate-normal case: $(\epsilon_1, \epsilon_2)' \sim N(0, \Sigma)$, the difference ξ is also normally distributed with mean 0 and variance σ_ξ^2. Thus

$$p_1 = P(\xi > v_2 - v_1) = \Phi\left(\frac{v_1 - v_2}{\sigma_\xi}\right).$$

Moreover, ϵ_1 and ξ are jointly normally distributed, so

$$\begin{aligned} E[\epsilon_1|d_1 &= 1, v_1, v_2] = r_{\epsilon_1,\xi} \bullet E[\xi|\xi > v_2 - v_1] \quad (r_{\epsilon_1,\xi} \equiv cov[\epsilon_1, \xi]/var[\xi]) \\ &= r_{\epsilon_1,\xi} \bullet \sigma_\xi \bullet E[z|z > \frac{v_2 - v_1}{\sigma_\xi}] \qquad \text{(for z} \sim N(0,1)) \\ &= \rho_{\epsilon_1,\xi}\sigma_{\epsilon_1}\frac{\phi\left(\frac{v_2-v_1}{\sigma_\xi}\right)}{1-\Phi\left(\frac{v_2-v_1}{\sigma_\xi}\right)} = \rho_{\epsilon_1,\xi}\sigma_{\epsilon_1}\frac{\phi\left(\frac{v_1-v_2}{\sigma_\xi}\right)}{\Phi\left(\frac{v_1-v_2}{\sigma_\xi}\right)} = \rho_{\epsilon_1,\xi}\sigma_{\epsilon_1}\frac{\phi(\Phi^{-1}(p_1))}{p_1}, \end{aligned}$$

using the result that for a standard normal variate, E(z|z>a)=$\phi(a)/[1 - \Phi(a)]$. This says that in the standard bivariate normal selection model, we can write

$$E[\epsilon_1|d_1 = 1, v_1, v_2] = E[\epsilon_1|d_1 = 1, p_1]$$

In other words, p_1 incorporates all the relevant information about v_1, v_2 that is needed to evaluate the selectivity bias in the stochastic component of the payoff to choice j when choice j is taken

b. More general models

In fact, for the standard random-utility setup with any distribution for the ϵ_j's, there is a mapping between the v_j's (or, more precisely, the differences $v_1 - v_j$, $v_2 - v_j$,...$v_J - v_j$, for an arbitrary choice of the base j) and the choice probabilities. This result was noted by Hotz and

Miller (ReStud, 1993), and forms the basis for their "CCP" (conditional choice probability) approach to estimating dynamic choice models.

Assume we have J choices, with $u_j = v_j + \epsilon_j$, with v_j a set of functions whose form is known (up to a vector of unknown parameters), and $(\epsilon_1, \epsilon_2, ...\epsilon_J) \sim F(\epsilon_1, \epsilon_2, ...\epsilon_J)$. Choice 1 is selected when $v_1 + \epsilon_1 > v_k + \epsilon_k$, or $\epsilon_k < v_1 - v_k + \epsilon_1$ (for all $k = 2, ..J$), which has probability

$$\begin{aligned} p_1 &= \int_{-\infty}^{\infty} \int_{-\infty}^{v_1 - v_2 + \epsilon_1} ... \int_{-\infty}^{v_1 - v_J + \epsilon_1} f(\epsilon_1, \epsilon_2, ...\epsilon_J) d\epsilon_2 ... d\epsilon_J \; d\epsilon_1, \\ &= \phi_1(v_1 - v_2, v_1 - v_3, ...v_1 - v_J;\; F). \end{aligned}$$

Similarly for choices $2, 3, ...J$, we can write

$$p_j = \phi_j(v_j - v_1, v_j - v_3, ...v_j - v_J;\; F).$$

(From now on I will drop the dependence on F but that is implicit, and quite important, since the choice of F dictates the functional form of the $\phi_j' s$. Note that the functions ϕ_j have the property that

$$\phi_j(r_1, r_2, ...r_J) = \phi_j(r_1 - \Delta, r_2 - \Delta, ...r_J - \Delta) \quad \text{for any } \Delta.$$

They also sum to 1. Now consider the system of J-1 equations:

$$\begin{aligned} p_2 &= \phi_2(0, v_1 - v_2, ...v_1 - v_J) \\ p_3 &= \phi_3(0, v_1 - v_2, ...v_1 - v_J) \\ &... \\ p_J &= \phi_J(0, v_1 - v_2, ...v_1 - v_J). \end{aligned}$$

Hotz and Miller apply the inverse function theorem to this system and obtain J-1 solution functions

$$v_1 - v_k = \psi_{1k}(p_2, ...p_J).$$

Once you have the J-1 solution functions for any base choice (e.g., the first), you can easily translate to another (e.g., the second) by subtracting the appropriate row from all the others. E.g.:

$$v_2 - v_k = (v_1 - v_k) - (v_1 - v_2) = \psi_{1k}(p_2, ...p_J) - \psi_{12}(p_2, ...p_J).$$

Keeping in mind there are only J-1 underlying functions, we can write

$$v_j - v_k = \psi_{jk}(p), \quad \text{where } p = (p_1, ...p_J).$$

This shows that in general the choice probabilities can be mapped into the differences in the conditional valuations, relative to an arbitrary base.

Now lets consider the "selectivity bias" expressions:

$$\begin{aligned} E[\epsilon_1 | d_1 = 1, v_1..v_J] &= \frac{\int_{-\infty}^{\infty} \int_{-\infty}^{v_1 - v_2 + \epsilon_1} ... \int_{-\infty}^{v_1 - v_J + \epsilon_1} \epsilon_1 f(\epsilon_1, \epsilon_2, ...\epsilon_J) d\epsilon_2 ... d\epsilon_J d\epsilon_1}{P(d_1 = 1 | v_1..v_J)} \\ &= \frac{\int_{-\infty}^{\infty} \int_{-\infty}^{\psi_{12}(p) + \epsilon_1} ... \int_{-\infty}^{\psi_{1J}(p) + \epsilon_1} \epsilon_1 f(\epsilon_1, \epsilon_2, ...\epsilon_J) d\epsilon_2 ... d\epsilon_J d\epsilon_1}{p_1} \\ &= w_j(p). \end{aligned}$$

Note that in case there are only 2 choices, this says that regardless of the distribution of (ϵ_1, ϵ_2), one can write

$$E[\epsilon_1|d_1 = 1, v_1, v_2] = w_1(p_1).$$

This forms the basis for "semi-parametric" approaches to estimating the conditional valuation function in a selected sample. If we observe a noisy version of the payoff for choice 1 among those who choose 1, say $y = u_1 + \varsigma$ where ς is an independent measurement error, and we assume $v_1 = f(X, \beta)$ then we know

$$E[y|X, d_1 = 1, p_1] = f(X, \beta) + w_1(p_1)$$

One can approximate $w_1(p_1)$ by some flexible functional form, or one can find a way to "match" observations with nearly the same values of p_1. Obviously, there has to be variation in p_1 for observations with the same value of X.

c. Logit-based applications

As shown in Arcidiacono and Miller (2007), the form of the $\psi_{jk}(p)$ and $w_j(p)$ functions can be simplified a lot if F has a MNL, nested logit or GEV form. They consider a nested logit with J choices in R nests, (where the rth nest has K_r choices):

$$F(\epsilon_{11}, ...\epsilon_{1K_1},\ \ \epsilon_{21}, ...\epsilon_{2K_2},\ \epsilon_{R1}\ ...\epsilon_{RK_R}) = \exp[-H(e^{-\epsilon_{11}}, ..e^{-\epsilon_{RK_R}})],$$

$$H(y_{11}, .., y_{RK_R}) = \sum_{r=1}^{R}\left[\sum_{k=1}^{K_r} y_{rk}^{\delta_r}\right]^{1/\delta_r}.$$

(Note they parameterize the "CES-like" part with $\delta_r = 1/\lambda_r$ relative to our earlier presentation). For this model they show that

$$E[\epsilon_{sj}|d_{sj=1}] = \gamma - \frac{1}{\delta_s}\log p_{sj} - (1 - \frac{1}{\delta_s})\log p_s + \log\left(\sum_{r=1}^{R} p_r^{1-1/\delta_r}\left[\sum_{k=1}^{K_r} p_{rk}^{\delta_s/\delta_r}\right]^{1/\delta_s}\right),$$

where $\gamma = 0.577$ is Euler's constant, p_{sj} is the probability of choice j in nest s, and p_s is the overall probability of any choice in nest s. For the "easy" case where $\delta_s = \delta$ for all s, the sum inside the log( ) for the last term equals 1, and the expression simplifies to

$$E[\epsilon_{sj}|d_{sj=1}] = \gamma - \frac{1}{\delta}\log p_{sj} - (1 - \frac{1}{\delta})\log p_s \quad ,$$

which expresses the selection bias in terms of the overall probability of a choice in nest s and the specific probability of choice j in nest s. Finally, if $\delta = 1$ we get the simple MNL, and

$$E[\epsilon_{sj}|d_{sj=1}] = \gamma - \log p_{sj}$$

These are remarkably simple formulas that could be useful in forming "first pass" selection corrections in settings with multiple choices. For a very different derivation of a selection correction for inter-state migration that looks a lot like the simple nested logit correction, see G. Dahl, Econometrica, Nov. 2002. A question for further thinking: would it be possible to derive a correction for a mixed logit choice model?

2. Introduction to Dynamic Discrete Choice

a. Prologue

Consider an agent who faces a discrete choice problem, with the payoff to choice j:

$$u_j = v_j + \epsilon_j$$

where the ϵ_j are random variables, unknown at the present time to the agent. (This is different from the way we have been thinking about the ϵ's up to now). Suppose the agent can make a choice of j once the ϵ's are realized. In this case, her expected utility is:

$$E[\max_j(u_1, ...u_J)],$$

a construct which is abbreviated as "Emax" in the literature. Emax is closely related to the concept of option value. In particular, suppose the agent had to choose before she could see the ϵ's. Then she would select j to

$$\max_j(E(u_1), E(u_2)...E(u_J))$$

a construct which we could call maxE. The option value of being able to select j after see the ϵ's is:

$$E[\max(u_1, ...u_J)] - \max(E(u_1), E(u_2)...E(u_J)) \geq 0.$$

The key idea in dynamic discrete choice problems with uncertainty is that an agent has to plan ahead, knowing that when the next period comes around she will have additional information and will be able to make an Emax decision.

For the case where $\epsilon_j \sim EV1$, we can use the expression for $E[\epsilon_j|d_{j=1}]$ presented above to derive a simple expression for Emax. In particular

$$\begin{aligned}
E[\max(u_1, ...u_J)] &= \Sigma_j \; p_j(v_j + E[\epsilon_j|d_{j=1}]) \\
&= \Sigma_j \; p_j(v_j + \gamma - \log p_j) \\
&= \gamma + \Sigma_j \; p_j(v_j - \log\left(\frac{\exp v_j}{\Sigma_k \exp v_k}\right)) \\
&= \gamma + \Sigma_j \; p_j \log(\Sigma_k \exp v_k) \\
&= \gamma + \log(\Sigma_k \exp v_k).
\end{aligned}$$

b. A basic example

Let's consider a T-period problem where there is state variable $X_t \in \{X_1, ...X_N\}$, a choice vector in each period $d_t = (d_{1t}, d_{2t}, ...d_{Jt})'$, (where $d_{jt} = 1$ means choice j was selected in period t), a "flow payoff" if choice j is made in period t:

$$v_{jt}(X_t) + \epsilon_{jt}$$

and a transition equation relating the state and choice in period t to the state (or the p.d.f. over possible states) in period $t + 1$:

$$P(X_{t+1}|X_t, d_t).$$

In some simple examples, such as Ebenstein's fertility model that we'll consider next lecture, the evolution of states is non-stochastic. In that example, the state is represented by the number and gender of children, and the choices in each stage are whether to conceive, whether to administer an ultrasound (if pregnant), and whether to abort the fetus (if the ultrasound reveals a girl). It is assumed that $P(X_{t+1}|X_t, d_t)$ is known.

In the last period (T), the state is X_T, and the agent has to solve

$$\max_j v_{jT}(X_T) + \epsilon_{jT} \quad .$$

Looking forward from period $T-1$, the expected utility associated with a particular value for X_T is

$$E \max[v_{jT}(X_T) + \epsilon_{jT}] = \gamma + \log \Sigma_j \exp(v_{jT}(X_T)).$$

In period $T-1$ the agent has to solve

$$\max_j v_{jT-1}(X_{T-1}) + \epsilon_{jT-1} + \beta \sum_{n=1}^{N} [\log \Sigma_j \exp(v_{jT}(X_n))] \, P(X_n|X_{T-1}, d_{jT-1} = 1).$$

(Note that we can drop the constant γ). Now pull together the non-random parts by defining

$$\psi_{jT-1}(X_{T-1}) = v_{jT-1}(X_{T-1}) + \beta \sum_{n=1}^{N} [\log \Sigma_j \exp(v_{jT}(X_n))] \, P(X_n|X_{T-1}, d_{jT-1} = 1)$$

and write the $T-1$ problem as

$$\max_j \psi_{jT-1}(X_{T-1}) + \epsilon_{jT-1}$$

Using the Emax formula, the expected utility from T-1 forward is (ignoring the constant):

$$\log \Sigma_j \exp(\psi_{jT-1}(X_{T-1})).$$

So, in period $T-2$ the agent has to solve

$$\max_j v_{jT-2}(X_{T-2}) + \epsilon_{jT-2} + \beta \sum_{n=1}^{N} \left[\log \Sigma_j \exp(\psi_{jT-1}(X_n))\right] P(X_n|X_{T-2}, d_{jT-2} = 1).$$

Again, collecting the non-random parts:

$$\psi_{jT-2}(X_{T-2}) = v_{jT-2}(X_{T-1}) + \beta \sum_{n=1}^{N} \left[\log \Sigma_j \exp(\psi_{jT-1}(X_n))\right] P(X_n|X_{T-2}, d_{jT-2} = 1)$$

the problem at T-2 can be written as

$$\max_j \psi_{jT-2}(X_{T-2}) + \epsilon_{jT-2}$$

Preceding backward in the same manner it is possible to define the objective function at t=1.