Economics 250c September 2008

Mixed logit models, continued

1. Prologue - ecological inference (borrowing from David Freedman)
2. Berry Levinsohn Pakes (BLP) - overview

1. Ecological Inference

The problem of ecological inference arises when we do not observe individual decision makers' choices, only the outcomes for a group. The classic application is to voting behavior. Suppose we are interested in estimating the relative preference of minority voters for Democrats. We observe the fraction of minorities m_i in each voting district (indexed by i), as well as the Democrat vote share in the district as a whole y_i. Let p_i and q_i denote the (unobserved) fractions of minorities and whites who vote for Democrats in district i, and let $\delta_i = p_i - q_i$. Write:

$$\begin{aligned} y_i &= m_i p_i + (1-m_i)q_i \qquad (1) \\ &= q_i + m_i\delta_i \\ &= q + m_i\delta + (q_i - q) + m_i(\delta_i - \delta) \end{aligned}$$

where p and q are the population means for p_i and q_i and $\delta = p - q$ is the mean difference in preferences. It is temping to define $\xi_i = (q_i - q) + m_i(\delta_i - \delta)$ and consider fitting the "model"

$$y_i = q + m_i\delta + \xi_i \ .$$

This is the so-called ecological regression, fit by relating district-wide vote shares to the local fraction of minorities. Notice that the "true model" has a random intercept (q_i) and a random slope (δ_i). Thus:

$$p\lim \widehat{\delta}_{ols} = \delta + cov(q_i - q, m_i)/var(m_i) + cov(m_i(\delta_i - \delta), m_i)/var(m_i)$$

Formally this looks just like the bias formula for the estimated return to schooling when the true earnings generating function has a person-specific intercept and slope (Card, 1999). In general OLS will be inconsistent unless there is no correlation between the preferences of whites and the minority share (the first bias term) and also no correlation between minorities' relative preference for Democrats ($\delta_i - \delta$) and the fraction of Democrats in the district. These are called the "constancy" assumptions in the early literature on ecological inference (Goodman, 1953).

In classical (1950-65) labor economics the problem of ecological inference arose in interpreting the correlation between average wages in a city (or industry) and the fraction of union members in the city (or industry), in the days before micro-level data with information on wages and union status was available. These regressions almost always showed a very large union wage "effect". H.G. Lewis dismissed these estimates, arguing that union power (the equivalent of δ) was stronger in areas with higher union densities, leading to an upward bias.

One of the earliest applications of "bounding" in social sciences was to the ecological inference problem. Observe that (1) implies

$$q_i = \frac{y_i - m_i p_i}{1 - m_i}.$$

Since $0 \leq p_i \leq 1$:

$$\frac{y_i}{1-m_i} \leq q_i \leq \frac{y_i - m_i}{1-m_i}$$

(Duncan and Davis, ASR, 1953). This is called the "method of bounds."

Most recently, King (1997) has proposed a solution to the ecological inference problem that looks a lot like a mixed logit model. Returning to (1), King assumes that the pair (p_i, q_i) are i.i.d. distributed across districts with some bivariate distribution function $F(p, q|\mu, \Sigma)$. (For example, King considers a bivariate truncated normal distribution). For district i the predicted vote share, conditional on m_i is

$$s_i^P(m_i) = \int\int [q + m_i(p - q)] f(p, q|\mu, \Sigma) dp dq \tag{2}$$

The integral can be evaluated by simulation, using the method discussed in Lecture 3. One could estimate the parameters (μ, Σ) by weighted least squares, using as a weight for the ith district the estimated sampling error of the observed vote share in that district. The key indentifying assumption in (2) is that (p, q) are distributed independently from m_i. In principle this can be generalized.

The empirical performance of King's method has been much debated. Freedman evaluated the method by using it to estimate the fractions of natives and immigrants with incomes over a certain bound, treating the observed units as the individually identified neighborhoods in the CPS. In this case, the "truth" is known, and King's method shows some bias. Freedman notes that in many applications, if you are interested in estimating the average difference between two groups whose shares vary across areas, you might do better to assume a "pure neighborhood model": $p_i = q_i = y_i$. In this case, if N_i is the total population of district i, we estimate the means p and q by:

$$\begin{aligned} p &= \sum_i \frac{m_i N_i}{M} y_i, \quad q = \sum_i \frac{(1 - m_i) N_i}{N - M} y_i \\ where \ \ N &= \sum_i N_i \ \text{ and } \ M = \sum_i m_i N_i. \end{aligned}$$

2. BLP's Method

We will summarize BLP's method, building on Nevo's "User's Guide". The set up is the same as at the end of Lecture 3, though with a slightly more complicated model of the individual heterogeneity. Specifically, we allow the income coefficient (the M.U. of income) to vary across individuals. We also allow for the presence of "demographic variables" – a vector D_{im} that shifts preferences, and for which we have information on the marginal distribution in each market. For concreteness, think of D_{im} as a dummy for minorities (or more generally as a set of dummies for a partition of the space of demographic variables), and assume that we know the mean minority share in market m. The utility that individual i in market $m = 1...M$ assigns to choice $j = 1...J$ is:

$$\begin{aligned} u_{imj} &= \alpha_{im}(y_{im} - p_{mj}) + X_j \beta_{im} + \xi_{mj} + \epsilon_{imj}, \\ with \ \ u_{im0} &= \epsilon_{im0} \ \text{ for the "no purchase" option, and} \\ \alpha_{im} &= \alpha + \pi_\alpha D_{im} + v_{1im}, \\ \beta_{im} &= \beta + \pi_\beta D_{im} + v_{2im}. \end{aligned}$$

Collecting terms, we have

$$
\begin{aligned}
u_{imj} &= \alpha_{im} y_{im} + \delta_{mj} + \mu_{imj} + \epsilon_{imj}, \quad \text{j=1...J} \\
&= \epsilon_{im0}, \qquad \text{j=0} \\
\delta_{mj} &= X_j \beta - \alpha p_{mj} + \xi_{mj} \\
\mu_{imj} &= -\pi_\alpha D_{im} p_{mj} - p_{mj} v_{1im} + X_j \pi_\beta D_{im} + X_j v_{2im}
\end{aligned}
$$

In this specification the random taste components (v_{1im}, v_{2im}) have a distribution with mean 0 and variance-covariance Σ. (Nevo writes the model in terms of a vector v_{im} that is assumed to have mean 0 and variance I). Assume the d.f. for (v_{im}, D_{im}) in market m is $F_v(v_{im})F_D(D_{im}|m)$. For example, F_v could be a multi-variate normal, and $F_D(D_{im}|m)$ could be a Bernoulli with mean D_m that is assumed to be known. A key assumption here (as in most applications of King's model) is that there is no "sorting" of people with unusually strong or weak preferences for different characteristics into different markets. In labor/public finance applications this could be problematic.

With this notation, the expected market share of choice j in market m is

$$
p_{mj} = \int\int \frac{\exp(\delta_{mj} + \mu_{imj})}{\sum_k \exp(\delta_{mk} + \mu_{ikj})} dF_v(v_{im}) dF(D_{im}|m).
$$

Note that in the case where D_{im} is a dummy for minority status, integration over the distribution of D_{im} amounts to taking a weighted sum for the cases with $D_{im} = 0$ and $D_{im} = 1$, with weights of $(1 - D_m)$ and D_m respectively.

Divide the parameters of the model into 2 groups: $\theta_1 = (\alpha, \beta)$ which only enter the "market level" component δ_{mj}, and $\theta_2 = (\pi_\alpha, \pi_\beta, cov(v_{im}))$ which enter the individual-specific term μ_{imj} and the distribution $F_v(v_{im})$. We can rewrite the previous equation as

$$
p_{mj} = p_{mj}(\delta_{m1}, \delta_{m2}, ...\delta_{mJ}, \theta_2) = \int\int \frac{\exp(\delta_{mj} + \mu_{imj}(\theta_2))}{\sum_k \exp(\delta_{mk} + \mu_{ikj}(\theta_2))} dF_v(v_{im}, \theta_2) dF(D_{im}|m).
$$

Notice that there are MJ observed market shares S_{mj} and MJ values for δ_{mj}. The system is invertible and, for any given choice of the individual-level parameters θ_2 there is an implied set of δ_{mj}'s that solve $p_{mj} = S_{mj}$, the actual market shares of choice j in market m. In the BLP-related literature, this is called the "inversion step". When there are no individual-specific components the "inversion step" is $\delta_{mj} = \log(S_{mj}/S_{m0})$ as noted in Lecture 3. BLP show that there is a relatively efficient iterative procedure that inverts the observed market shares into the δ_{mj}, for a given θ_2.

Assume that there are instruments Z_{mj} available for the price of choice j in market m. If we had the "right" value of θ_2, we would get the "right" values for the δ_{mj}'s, and then we could proceed as in Berry (1994) to estimate

$$
\delta_{mj} = X_j \beta - \alpha p_{mj} + \xi_{mj}
$$

by IV, using the Z_{mj} as instruments. Since we don't know the "right" value for θ_2, BLP's idea is to search over θ_2, and for each choice get the minimized value of the "GMM" objective function" associated with the IV "regression":

$$
\min_{\theta_1} (\widehat{\delta}_{mj} - X_j \beta + \alpha p_{mj}) Z \Omega Z' (\widehat{\delta}_{mj} - X_j \beta + \alpha p_{mj})
$$

where Z is the vector of instruments and Ω is a weighting matrix. Note that for $\Omega = (Z'Z)^{-1}$ this is the 2sls minimand from an IV regression of $\widehat{\delta}_{mj}$ on X_j and p_{mj} using the instrument set Z.

As we will discuss in the next lecture, if micro data are available on the choices actually made by individuals in market m, the whole thing is a lot easier: in that case, one estimates a mixed logit model with $choice \times market$ effects δ_{mj}, and in the second stage fits the relation of these to the X's and the p's. With only market share data, however, the division of the observed variation in market shares across markets into the parts due to micro level variation (in v_{im} and D_{im}) and to "market level" variation (in p_{mj}) is a lot more difficult (and tenuous).