ARE213 **Econometrics**

Spring 2006 UC Berkeley Department of Agricultural and Resource Economics

Panel Data II: Fixed Effects

In this lecture we consider the same setup, with a linear model:

$$Y_{it} = X_{it}'\beta + c_i + \varepsilon_{it},$$

with c_i an unobserved individual-specific, time-invariant component. However, compared to the random effects discussion we relax the assumption that c_i is independent of the observed covariates X_{it}. We continue to maintain the exogeneity assumption on the residuals:

$$\mathbb{E}[\varepsilon_{it}|X_{i1},\ldots,X_{iT},c_i] = 0,$$

and in fact for inference we make the stronger assumption

$$\mathbb{E}[\varepsilon_i\varepsilon_i'|c_i,X_i] = \sigma^2 \cdot I_T.$$

We consider a couple of estimators. The first is based on simply adding a N-dimensional vector of time-invariant covariates Z, with its jth element for unit i in period equal to $Z_{it,j}$:

$$Z_{it,j} = 1\{i=j\}.$$

Then if we define c to be the vector with typical element c_i, we can write

$$Y_{it} = X_{it}'\beta + Z_i'c + \varepsilon_{it}.$$

The first estimator is just the least squares estimator for this regression function:

$$\min_{c,\beta} \sum_{i,t} (Y_{it} - X_{it}'\beta - Z_i'c)^2.$$

The estimators for both β and c are unbiased. However, the estimators for c are not consistent. As we get more and more observations, we do not get more information about c_i. To see this, consider the special case without X_{it}. In that case $\hat{c}_i = \sum_t Y_{it}/T = \bar{Y}_i$, which has variance σ^2/T which does not go to zero as N goes to infinity. We call the estimator for β the dummy variable estimator, $\hat{\beta}_{dv}$. A nice feature is that we can also use its associated standard error. Here it is important that we estimate the error variance σ^2 as

$$\frac{1}{NT - K - N} \sum_{i,t} (Y_{it} - X'_{it}\beta - Z'_i c)^2,$$

and not just divide by N. The number of parameters, K (for β) and N (for c), increases with the sample size so the degrees of freedom correction is important. Implementing this estimator could be difficult if N is very large. The least squares estimator requires inverting a $(N + K) \times (N + K)$ dimensional matrix.

The other estimators employ various techniques to remove c_i from the regression function. For the second estimator define the unit averages

$$\bar{Y}_i = \frac{1}{T}\sum_{t=1}^{T} Y_{it}, \qquad \text{and } \bar{X}_i = \frac{1}{T}\sum_{t=1}^{T} X_{it},$$

as well as the deviations from the mean:

$$\ddot{Y}_{it} = Y_{it} - \bar{Y}_i, \qquad \text{and } \ddot{X}_{it} = X_{it} - \bar{X}_i.$$

Note that in general we can write the $T \times K$ matrix of deviations from the means, $\ddot{X}_i$, with tth row equal to $\ddot{X}'_{it}$, as

$$\ddot{X}_i = AX_i,$$

where

$$A = I_T - \iota_T \iota_T / T,$$

so that A is idempotent:

$$AA = (I_T - \iota_T \iota_T/T)(I_T - \iota_T \iota_T/T) = I_T - \iota_T \iota_T/T - \iota_T \iota_T/T + \iota_T \iota_T' \iota_T \iota_T'/T^2$$

$$= I_T - \iota_T \iota_T/T = A.$$

Then the "between" estimator $\hat{\beta}_b$ is based on the regression

$$\ddot{Y}_{it} = \ddot{X}_{it}'\beta + \ddot{\varepsilon}_{it}.$$

Because

$$\mathbb{E}[\ddot{\varepsilon}_{it}|X_i] = 0,$$

this estimator is again unbiased. In fact, some careful linear algebra shows that it is identical to the dummy variable estimator $\hat{\beta}_d = \hat{\beta}_{dv}$. In this case, however, we cannot use the ols variance to get the right standard errors. The residual from this regression is $\ddot{\varepsilon}_{it}$. It has the following properties:

$$\mathbb{E}[\ddot{\varepsilon}_{it}^2] = \mathbb{E}[(\varepsilon_{it} - \bar{\varepsilon}_i)^2] = \mathbb{E}[\varepsilon_{it}^2 - 2\varepsilon_{it}\bar{\varepsilon}_i + \bar{\varepsilon}_i^2]$$

$$= \sigma^2 - 2\sigma^2/T + \sigma_2/T = \sigma^2(1 - 1/T),$$

and

$$\mathbb{E}[\ddot{\varepsilon}_{it} \cdot \ddot{\varepsilon}_{is}] = \mathbb{E}[\varepsilon_{it} \cdot \varepsilon_{is} - \varepsilon_{it} \cdot \bar{\varepsilon}_i - \varepsilon_{is} \cdot \bar{\varepsilon}_i + \bar{\varepsilon}_i^2]$$

$$= 0 - \sigma^2/T - \sigma^2/^T + \sigma^2/T = -\sigma^2/T.$$

To get the correct variance, consider the difference between the estimator and β:

$$\hat{\beta}_b - \beta = \left(\frac{1}{N}\sum_{i=1}\ddot{X}_i'\ddot{X}_i\right)^{-1}\frac{1}{N}\sum_i \ddot{X}_i'\ddot{\varepsilon}_i.$$

First, note that $\sum_i \ddot{X}_i'\ddot{\varepsilon}_i = \sum_i X_i'A'A\varepsilon_i = \sum_i X_i'A'\varepsilon_i \sum_i \ddot{X}_i\varepsilon_i$. Thus the variance of the between estimator is

$$V(\hat{\beta}_b) = \left(\frac{1}{N}\sum_{i=1} \ddot{X}_i'\ddot{X}_i\right)^{-1} \left(\frac{1}{N^2}\sum_{i=1} \ddot{X}_i'\mathbb{E}[\varepsilon_i\varepsilon_i']\ddot{X}_i\right) \left(\frac{1}{N}\sum_{i=1} \ddot{X}_i'\ddot{X}_i\right)^{-1}$$

$$= \sigma^2 \cdot \left(\sum_{i=1} \ddot{X}_i'\ddot{X}_i\right)^{-1}.$$

In order to get the correct variance we just need to estimate σ^2. Let us look at the residuals from the between regression. The expectation of their square is equal to $\sigma^2(1-1/T)$. Hence we can just average their squares, and then multiply by $T/(T-1)$ to get a consistent estimator.

Another estimator estimator is based on differencing. Define

$$\Delta Y_{it} = Y_{it} - Y_{it-1},\ \Delta X_{it} = X_{it} - X_{it-1}\ \ \text{and} Delta\varepsilon_{it} = \varepsilon_{it} - \varepsilon_{it-1}.$$

Then we can write

$$\Delta Y_{it} = \Delta' X_{it} + \Delta\varepsilon_{it},$$

for $t = 2, \ldots, T$. OLS for this regression is consistent. If the original errors ε_{it} are uncorrelated we now get a more complicated covariance matrix with the first-differenced errors correlated.

Fixed effect methods do not easily extend to nonlinear models. There are some exceptions where it is still possible to remove the individual component without distributional or independence assumptions. A famous example is Chamberlain's fixed effect logit model. Consider the case with two periods. Conditional on the individual effect we have

$$\Pr(Y_{it} = 1|X_{it}, c_i) = \frac{\exp(\beta' X_{it} + c_i)}{1 + \exp(\beta' X_{it} + c_i)}.$$

Chamberlain suggests looking at units with $Y_{i1} \neq Y_{i2}$. Then

$$\Pr(Y_{i1} = 1 | X_{i1}, X_{i2}, Y_{i1} \neq Y_{i2}) = \mathbb{E}\left[\Pr(Y_{i1} = 1 | X_{i1}, X_{i2}, Y_{i1} \neq Y_{i2}, c_i) \right| X_{i1}, X_{i2}, Y_{i1} \neq Y_{i2}]$$

$$= \mathbb{E}\left[\left. \frac{\Pr(Y_{i1} = 1, Y_{i2} = 0 | X_{i1}, X_{i2}, c_i)}{\Pr(Y_{i1} = 1, Y_{i2} = 0 | X_{i1}, X_{i2}, c_i) + \Pr(Y_{i1} = 0, Y_{i2} = 1 | X_{i1}, X_{i2}, c_i)} \right| X_{i1}, X_{i2} \right]$$

$$= \frac{\frac{\exp(\beta' X_{i1} + c_i)}{1 + \exp(\beta' X_{i1} + c_i)} \cdot \frac{1}{1 + \exp(\beta' X_{i2} + c_i)}}{\frac{\exp(\beta' X_{i1} + c_i)}{1 + \exp(\beta' X_{i1} + c_i)} \cdot \frac{1}{1 + \exp(\beta' X_{i2} + c_i)} + \frac{1}{1 + \exp(\beta' X_{i1} + c_i)} \cdot \frac{\exp(\beta' X_{i2} + c_i)}{1 + \exp(\beta' X_{i2} + c_i)}}$$

$$= \frac{\exp(\beta' X_{i1} + c_i)}{\exp(\beta' X_{i1} + c_i) + \exp(\beta' X_{i2} + c_i)}$$

$$= \frac{\exp(\beta'(X_{i1} - X_{i2}))}{1 + \exp(\beta'(X_{i1} - X_{i2}))},$$

which means we can just do a standard logistic regression for the observations with $Y_{i1} \neq Y_{i2}$ using the first difference of the regressors. This does not work for the probit model, demonstrating that this is a very special case.