Applied Panel Data Analysis – Lecture 3

Christopher F. Parmeter

AGRODEP
March 10-14th, 2015
Addis Ababa, Ethiopia

- Cover estimation of the unobserved effect model under the fixed effects framework
- Develop intuition for how within estimator works
- Discuss why strict exogeneity is needed for the covariates

- Our last lecture paid careful attention to the unobserved effects panel data model
- To begin our discussion we will focus on building an estimator for the fixed effects framework
- Recall that with the fixed effects framework we are assuming that $Cov(x_{it}, c_i) \neq 0$

- Once again, the unobserved effects model with individual, time constant heterogeneity is

$$y_{it} = x'_{it}\beta + c_i + \varepsilon_{it} \tag{1}$$

- We discussed an example in the previous lecture on removing c_i by differencing between two time periods
- A natural question is "When we have more than two time periods which period should we difference with?"
- We could difference with x_{i1} always or we could difference with x_{it-1} or x_{it-2} (though in this case you would unnecessarily lose observations)

- An alternative to deciding how to difference is to use time averaging
- Notice that $\bar{c}_i = T^{-1} \sum_{t=1}^{T} c_i = c_i$
- So if we time average each variable, for each individual, we could then difference and c_i would be eliminated
- We will use the notation $\bar{z}_{i\cdot} = T^{-1} \sum_{t=1}^{T} z_{it}$ where z could be either y, ε or one of our x variables

- Our time averaged version of the unobserved effects model becomes

$$\bar{y}_{i\cdot} = \bar{x}_{i\cdot}'\beta + c_i + \bar{\varepsilon}_{i\cdot} \tag{2}$$

- Subtracting (2) from (1) yields

$$\begin{aligned} y_{it} - \bar{y}_{i\cdot} &= (x_{it} - \bar{x}_{i\cdot})'\,\beta + c_i - c_i + \varepsilon_{it} - \bar{\varepsilon}_{i\cdot} \\ &= (x_{it} - \bar{x}_{i\cdot})'\,\beta + \varepsilon_{it} - \bar{\varepsilon}_{i\cdot} \end{aligned} \tag{3}$$

- Notice from (3) that c_i is no longer present
- This suggests if we run the regression of time demeaned y on time demeaned x we can recover an estimate of β

- Now this setup might appear cumbersome because we have to calculate $\bar{y}_{i\cdot}$ and $\bar{x}_{i\cdot}$ for each individual
- With a large micro panel this can quickly get expensive
- An alternative way to think about the unobserved effects model assuming nonzero correlation is that c_i represents a unique intercept for each individual
- Instead of a common intercept for all individuals, here each individual has a (potentially) different starting point

- A model where there are different intercepts should be familiar to you
- Consider a wage regression with a dummy variable for gender
- Consider production function estimation with a dummy variable for old and young farmers
- Consider a profit model with a dummy variable for public and private companies
- Notice that in each example the model contained a dummy variable that split the sample into two groups
- Here we have a dummy variable that splits our sample into N groups, one for each individual

- Lets collect this set of N dummy variables into a matrix called C
- What does C look like?
- C is an $NT \times T$ matrix that can be written as $I_N \otimes \imath_T$ where I_N is the $N \times N$ identify matrix and $\imath_T$ is a $T \times 1$ vector of all ones
- The symbol $\otimes$ stands for the Kronecker product

- The Kronecker product may be unfamiliar to you so lets have a brief review
- For A, an $m \times r$ matrix, and B a $p \times q$ matrix, $A \otimes B$ is a $mp \times rq$ matrix

$$A \otimes B = \begin{bmatrix} a_{11}B & \cdots & a_{1r}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \cdots & a_{mr}B \end{bmatrix} \tag{4}$$

- Notice that the Kronecker product does not require A and B to be conformable as would be necessary with standard matrix multiplication and addition

- Some basic properties of the Kronecker product that we will use are:
  - $(A \otimes B)' = A' \otimes B'$
  - $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$
  - $(A \otimes B)(D \otimes F) = AD \otimes BF$
- There are many other associative, bilinearity and commutative properties, but they will not appear in our discussion

- We will define several additional matrices that will help us to succinctly formulate the estimator for the unobserved effects model
- Let y denote the $NT \times 1$ vector of regressands, ε denote the $NT \times 1$ vector of error terms and X denote the $NT \times K$ vector of regressors
- Define $P = C(C'C)^{-1}C'$ and $Q = I_{NT} - P$
- P and Q will be important in our construction of an estimator for the unobserved effects model

- Note that we can provide a more intuitive characterization of P by using its formal definition

$$
\begin{aligned}
P =& C(C'C)^{-1}C' \\
=& (I_N \otimes \imath_T)((I_N \otimes \imath_T)' I_N \otimes \imath_T)^{-1}(I_N \otimes \imath_T)' \\
=& (I_N \otimes \imath_T)(I_N \otimes \imath_T' \imath_T)^{-1}(I_N \otimes \imath_T') \\
=& (I_N \otimes \imath_T)(I_N \otimes T)^{-1}(I_N \otimes \imath_T') \\
=& (I_N \otimes \imath_T)(I_N \otimes T^{-1})(I_N \otimes \imath_T') \\
=& I_N \otimes \imath_T \imath_T' T^{-1} \\
=& I_N \otimes T^{-1} J_T = I_N \otimes \bar{J}_T
\end{aligned} \tag{5}
$$

- Here J_T is a $T \times T$ matrix of all ones and $\bar{J}_T$ is a $T \times T$ matrix where each element is T^{-1}
- Notice that

$$Pz = \bar{z} = \left(\underbrace{z_{1\cdot}, \ldots, z_{1\cdot}}_{T \text{ times}}, \underbrace{z_{2\cdot} \ldots, z_{2\cdot}}_{T \text{ times}}, \ldots, \ldots, \underbrace{z_{N\cdot}, \ldots, z_{N\cdot}}_{T \text{ times}} \right)$$

- We also have $Qz = z - Pz$ will demean each variable
- It is useful to point out that $Q' = Q$, $P' = P$, $QQ = Q$ and $PP = P$
- In words, both Q and P are symmetric and idempotent
- Further $QP = 0$ and $Q + P = I_{NT}$

- Now, rewrite our unobserved effects model in (1) in matrix form as

$$y = X\beta + C\alpha + \varepsilon \tag{6}$$

- Premultiplication of (6) by Q (remember we want to time demean) yields

$$\begin{aligned} Qy =& QX\beta + QC\alpha + Q\varepsilon \\ =& QX\beta + Q\varepsilon \end{aligned} \tag{7}$$

- The presence of C has disappeared given that since Q demeans, and C is time constant, then demeaning a constant yields 0

- If we were to use the notation $\tilde{z} = Qz$, then (7) can be succinctly written as

$$\tilde{y} = \tilde{X}\beta + \tilde{\varepsilon} \tag{8}$$

- The elegance of (8) is that aside from the awkward $\tilde{\varepsilon}$ this looks like a standard OLS regression of $\tilde{y}$ on $\tilde{X}$
- If $E(\varepsilon\varepsilon'|X, C) = \sigma^2 I_{NT}$, then $E(\tilde{\varepsilon}\tilde{\varepsilon}'|X) = \sigma^2 Q \neq \sigma^2_{NT}$
- Thus, a generalized least squares estimator will be appropriate
- Remember that even though we are estimating β from (8) we must interpret β from (7)

- Our estimator for β that controls for the unobserved, fixed individual heterogeneity is

$$\tilde{\beta} = \left(\tilde{X}'\tilde{X}\right)^{-1} \tilde{X}'\tilde{y} \tag{9}$$

- Standard results also yield

$$Var(\tilde{\beta}) = \sigma^2 \left(\tilde{X}'\tilde{X}\right)^{-1} \tag{10}$$

- An estimator for σ^2 is

$$\widehat{\sigma}^2 = (N(T-1) - K)^{-1} \hat{\tilde{\varepsilon}}' \hat{\tilde{\varepsilon}} \tag{11}$$

where $\hat{\tilde{\varepsilon}} = \tilde{y} - \tilde{X}\tilde{\beta}$

- The unobservable individual effects can be recovered as

$$\tilde{c} = Py - PX\tilde{\beta} \tag{12}$$

- That is, $\tilde{c}_i = \bar{y}_{i\cdot} - x'_{i\cdot}\beta$

- The matrix of dummy variables C has prompted $\tilde{\beta}$ to be referred to as the least squares dummy variable estimator (LSDV)
- The more common parlance is to refer to $\tilde{\beta}$ as the within estimator or the fixed effects estimator

- What is the intuition for how the LSDV/within estimator works?
- β is identified off of variation across individual variation over time
- Recall the pooled OLS estimator identifies β off of variation in the covariates in general
- Consider the simple unobserved effects model with a single covariate

$$y_{it} = \beta x_{it} + c_i + \varepsilon_{it} \tag{13}$$

- The pooled OLS estimator (ignoring c_i) produces

$$\hat{\beta} = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} (y_{it} - \bar{y})(x_{it} - \bar{x})}{\sum_{i=1}^{N} \sum_{t=1}^{T} (x_{it} - \bar{x})^2}$$

- Whereas the within estimator produces

$$\tilde{\beta} = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} (y_{it} - \bar{y}_{i\cdot})(x_{it} - \bar{x}_{i\cdot})}{\sum_{i=1}^{N} \sum_{t=1}^{T} (x_{it} - \bar{x}_{i\cdot})^2}$$

- It should be clear that the pooled OLS estimator is based on the overall variation in y and x
- The within estimator is based completely off of individual specific variation
- What does this mean?
- We need within individual variation of the covariates to produce an estimator that has meaning
- Sometimes you will hear/read that "identification is off of individual specific variation"

- For large micro panels you would always want to use the within estimator as opposed to directly estimating (6)
- The reason for this is the fact that instead of inverting a $K \times K$ matrix (which is easy) you would be inverting an $(N + K) \times (N + K)$ matrix (which is not easy when N is large)

- Note also that if you have any elements in X that are time constant, that Q eliminates them
- This is the same issue we discussed in lecture 2 when we introduced c_i
- We cannot separate c_i (which is time constant) from another, observable time constant variable
- Note that time constant here means a variable is constant across time for all individuals; it is fine to have a variable which varies for some individuals and is constant for others

- For the fixed effects estimator to be consistent we need strict exogeneity conditional on the unobserved effect, $E(\varepsilon_{it}|x_i s, c_i) = 0$, $s = 1, \ldots, T$
- Given the time averaging that we use to eliminate c_i relaxing this assumption would not ensure that x and ε are uncorrelated for a given individual, a necessary condition to have both an unbiased and consistent estimator
- We also want to think about T being fixed and $N \longrightarrow \infty$

- The estimator for the unobserved effects is inconsistent as $N \longrightarrow \infty$
- Why? As N grows larger, we have more unobserved effects to estimate
- This is known as the incidental parameters problem (see Lancaster, 2000)
- We need $T \longrightarrow \infty$ for the estimator of the unobserved effects to be consistent

- We have seen that one unfortunate consequence with the inclusion of time constant individual specific unobservable effects is that we cannot recover the impact of observable time constant individual specific effects
- However, we can discern if the impact of these observable time constant variables has changed over time
- We do this by interacting our time constant variables with time period dummy variables

- Let the matrix D_T denote $\imath_N \otimes I_{T-1}$, the matrix of time period effects (only for T_1); further partition our matrix of covariates so that the time constant variables are collected separately as Z, where Z is $N(T-1) \times K_z$
- Then, instead of including Z in our unobserved effects model, we include $D_T Z$
- Our new model is

$$y = X\beta + (D_T Z)\gamma + C\alpha + \varepsilon \tag{14}$$

- Given that the elements of $D_T Z$ vary over individual and time, when we premultiply by Q to remove C, Z will still appear in the model

- Now, we must be careful because γ_{jt} does not represent the effect of Z_{ij} on y, rather it is the effect of Z_{ij} on y in period t relative to the baseline period
- In our setup the baseline period is the last period, period T
- Thus, we can estimate how a particular time constant variable's impact is changing over time

- The variance-covariance of $\tilde{\beta}$ in (10) is incorrect if the assumption of a constant conditional variance
- This assumption could fail if we have time specific heteroskedasticity, if there is classic heteroskedasticity, or if there is serial correlation in the error terms (this is important with a moderately sized T)
- Can use the insights of White (1980) to construct a variance-covariance estimator for β that is robust to all forms of heteroskedasticity and auto-correlation

- Following Arellano (1987) use $\tilde{X}$ and $\hat{\tilde{\varepsilon}}$ in place of X and $\hat{\varepsilon}$ in the standard formula for constructing a robust variance-covariance estimator for the pooled estimator
- Recall the sandwich form of the variance-covariance estimator $(A^{-1}BA^{-1})$ from Lecture 1, which yields

$$Var(\tilde{\beta}) = (\tilde{X}'\tilde{X})^{-1}\left[\sum_{i=1}^{N}\tilde{X}_i'\hat{\tilde{\varepsilon}}_i\hat{\tilde{\varepsilon}}_i'\tilde{X}_i\right](\tilde{X}'\tilde{X})^{-1} \quad (15)$$

- Notice this allows for arbitrary heteroskedasticity across individuals and serial correlation amongst the errors terms within an individual; still restricts errors across individuals to be independent

- Developed the intuition for the within estimator of the unobserved effects panel data model when we assumed a fixed effects specification
- Need $N \longrightarrow \infty$ for consistency of the estimator
- Time demeaning eliminates the unobserved individual effect
- Incidental parameters problem requires $T \longrightarrow \infty$ for consistent estimation of these unobserved, individual specific effects