Applied Panel Data Analysis – Lecture 4

Christopher F. Parmeter

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

- Cover estimation of the unobserved effect model under the random effects framework
- Develop intuition for how the generalized least squares estimator works in this context
- Learn about the between estimator
- Discuss why strict exogeneity is needed for the covariates

- Our unobserved effects model is identical for the random effects framework for the unobserved effects model as with the fixed effects framework

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

- The key difference is that in the random effects framework, we assume not only that $E[\varepsilon_{it}|x_{is}, c_i] = 0$ for $s = 1, \ldots, T$, but $E[c_i|x_{is}] = E[c_i] = 0$ for $s = 1, \ldots, T$
- This last condition is the important distinction between the random and fixed effects framework, as we discussed in Lecture 2

- The assumption of no correlation between the observable covariates, x_{it} and the unobservable, individual specific heterogeneity, c_i means that we do not need to control for its presence when we estimate β in (1)
- However, by placing c_i in the error term we now have what is known as a composed error or a one-way error component
- Typically standard OLS estimation when there is a composed error will not produce an estimator with appealing statistical properties

- To see this more clearly rewrite the model in (1) as

$$y_{it} = x_{it}'\beta + u_{it} \tag{2}$$

  where $u_{it} = c_i + \varepsilon_{it}$
- Note that if $E[\varepsilon\varepsilon'|X] = \sigma^2 I_{NT}$, then $E[uu'|X] \neq \sigma^2 I_{NT}$
- This is because there is correlation between u_{it} and u_{is}, $s \neq t$ due to the presence of c_i
- In essence, we have introduced serial correlation amongst some errors when we migrate from the fixed effects framework to the random effects framework

- Given that our error no longer has constant variance and zero serial correlation, OLS estimation of (2) will not produce an efficient estimator
- However, we can derive the exact generalized least squares estimator (GLS) since we know the form of the variance-covariance structure
- Recall that when the variance-covariance structure of the error term from a model is Ω, the GLS estimator is $\hat{\beta}_{GLS} = \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}y$
- The OLS estimator for the random effects framework of the unobserved effects model is nothing more than GLS

- To construct the GLS estimator for the random effects framework we need to determine the structure of the variance-covariance matrix of u
- Given that our individual, unobserved heterogeneity is in the error term it will help to think of the c_is as random variables that come from some distribution
- We will assume that $c_i \sim IID(0, \sigma_c^2)$
- To help distinguish the variance parameter for ε we will also assume that $\varepsilon_{it} \sim IID(0, \sigma_\varepsilon^2)$

- To determine the structure of the variance-covariance of u we need to determine the expectation of four different terms
- Recall that the variance-covariance matrix is $E[uu'|X]$ and uu' contains elements of the form $u_{it}u_{js}$ for $s, t = 1, \ldots, T$ and $i, j = 1, \ldots, N$
- When $i = j$ and $s = t$ we have $E[u_{it}^2|X] = E[c_i^2] + 2E[c_i\varepsilon_{it}] + E[\varepsilon_{it}^2] = \sigma_c^2 + 0 + \sigma_\varepsilon^2$
- The middle term is zero since we assume that both c and ε are IID

- Now, when $i = j$ but $s \neq t$ then we have $E[u_{it}u_{is}|X] = E[c_i^2] + E[c_i\varepsilon_{it}] + E[c_i\varepsilon_{is}] + E[\varepsilon_{it}\varepsilon_{is}] = \sigma_c^2 + 0 + 0 + 0$
- The last three terms are zero since we assume that both c and ε are IID

- Lastly, when $i \neq j$ we have $E[u_{it}u_{js}|X] = E[c_i c_j] + E[c_i \varepsilon_{js}] + E[c_j \varepsilon_{it}] + E[\varepsilon_{it}\varepsilon_{js}] = 0 + 0 + 0 + 0$
- All of the terms are zero since we assume that both c and ε are IID

- Let the variance-covariance for u_i be defined as

$$\Omega_i = \begin{bmatrix} \sigma_c^2 + \sigma_\varepsilon^2 & \sigma_c^2 & \cdots & \sigma_c^2 \\ \sigma_c^2 & \sigma_c^2 + \sigma_\varepsilon^2 & \cdots & \sigma_c^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_c^2 & \sigma_c^2 & \cdots & \sigma_c^2 + \sigma_\varepsilon^2 \end{bmatrix} \tag{3}$$

- We are now in a position to derive the full variance-covariance structure
- For the random effects framework the variance-covariance structure of the unobserved effects model is

$$\Omega = \begin{bmatrix} \Omega_i & 0 & \cdots & 0 \\ 0 & \Omega_i & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Omega_i \end{bmatrix} \tag{4}$$

- Ω can be written succinctly as $\Omega = I_N \otimes \Omega_i$

- Note that Ω_i does not actually depend on i, this is purely for notational convenience
- Also, Ω_i is a $T \times T$ matrix that can be written as $\sigma_c^2 J_T + \sigma_\varepsilon^2 I_T$
- Using properties of the Kronecker product we have

$$\begin{aligned} \Omega = I_N \otimes \Omega_i = & I_N \otimes \left(\sigma_c^2 J_T + \sigma_\varepsilon^2 I_T\right) \\ = & \sigma_c^2 \left(I_N \otimes J_T\right) + \sigma_\varepsilon^2 \left(I_N \otimes I_T\right) \end{aligned} \quad (5)$$

- Currently an unfortunate consequence of representing the variance-covariance structure in full matrix form is that we need Ω^{-1}
- Ω is an $NT \times NT$ matrix, which for typical panels is large
- Obtaining the inverse of matrices beyond a 1000×1000 are difficult and time consuming for standard machines

- To invert Ω we use the trick of Wansbeek and Kapteyn (1982)
- Their proposal is to write Ω as

$$\begin{aligned}\Omega =&\sigma_c^2 \left(I_N \otimes J_T\right) + \sigma_\varepsilon^2 \left(I_N \otimes I_T\right) \\ =&\sigma_c^2 \left(I_N \otimes T\bar{J}_T\right) + \sigma_\varepsilon^2 \left(I_N \otimes \left(I_T + \bar{J}_T - \bar{J}_T\right)\right) \\ =&T\sigma_c^2 \left(I_N \otimes \bar{J}_T\right) + \sigma_\varepsilon^2 \left(I_N \otimes +\bar{J}_T\right) + \sigma_\varepsilon^2 \left(I_N \otimes \left(I_T - \bar{J}_T\right)\right) \\ =&\left(T\sigma_c^2 + \sigma_\varepsilon^2\right) \left(I_N \otimes \bar{J}_T\right) + \sigma_\varepsilon^2 \left(I_N \otimes E_T\right) \qquad (6)\end{aligned}$$

where $E_T = I_T - \bar{J}_T$

- The key here is to notice that $I_N \otimes \bar{J}_T = P$ and $I_N \otimes E_T = Q$, our symmetric and idempotent matrices that appeared when we derived the fixed effects estimator
- Let $\sigma_1^2 = T\sigma_c^2 + \sigma_\varepsilon^2$
- We now have the simple characterization

$$\Omega = \sigma_1^2 P + \sigma_\varepsilon^2 Q \tag{7}$$

- The form of Ω in (7) is known as the spectral decomposition representation
- The benefit of this decomposition is that we have

$$\Omega^{-1} = \frac{1}{\sigma_1^2} P + \frac{1}{\sigma_\varepsilon^2} Q \tag{8}$$

- We can see that this is the correct form for Ω^{-1} as

$$\begin{aligned}\Omega^{-1}\Omega &= \left(\frac{1}{\sigma_1^2}P + \frac{1}{\sigma_\varepsilon^2}Q\right)\left(\sigma_1^2 P + \sigma_\varepsilon^2 Q\right) \\ &= \frac{\sigma_1^2}{\sigma_1^2}PP + \frac{\sigma_\varepsilon^2}{\sigma_1^2}PQ + \frac{\sigma_c^2}{\sigma_\varepsilon^2}QP + \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2}QQ \\ &= P + 0 + 0 + Q = I\end{aligned}$$

- In fact, it holds more generally that $\Omega^r = \left(\sigma_1^2\right)^r P + \left(\sigma_\varepsilon^2\right)^r Q$

- We are now in position to construct the GLS estimator for the random effects framework of the unobserved effects panel data model
- Our GLS estimator is

$$\hat{\beta}_{GLS} = \left(X' \left(\frac{1}{\sigma_1^2} P + \frac{1}{\sigma_\varepsilon^2} Q \right) X \right)^{-1} X' \left(\frac{1}{\sigma_1^2} P + \frac{1}{\sigma_\varepsilon^2} Q \right) y \quad (9)$$

- As it stands this estimator does not look intuitive; however, with some further algebraic manipulations we can recast this estimator in a similar fashion as the within estimator

- Note that $\Omega^{-1} = \Omega^{-1/2}\Omega^{-1/2}$
- Further, $\Omega^{-1/2} = \frac{1}{\sigma_1}P + \frac{1}{\sigma_\varepsilon}Q$
- We have

$$\begin{aligned}\hat{\beta}_{GLS} &= \left(X'\Omega^{-1/2}\Omega^{-1/2}X\right)^{-1} X'\Omega^{-1/2}\Omega^{-1/2}y \\ &= \left(X'\sigma_\varepsilon\Omega^{-1/2}\sigma_\varepsilon\Omega^{-1/2}X\right)^{-1} X'\sigma_\varepsilon\Omega^{-1/2}\sigma_\varepsilon\Omega^{-1/2}y \\ &= \left(\check{X}'\check{X}\right)^{-1} \check{X}'\check{y} \qquad (10)\end{aligned}$$

where $\check{z} = \sigma_\varepsilon\Omega^{-1/2}z$

- Lets think about what an element of $\check{z}$ looks like
- First $\sigma_\varepsilon \Omega^{-1/2} = Q + \frac{\sigma_\varepsilon}{\sigma_1} P$
- A row of this matrix has elements $1 - (1/T) + (\sigma_\varepsilon / T\sigma_1)$ for a given individual and 0s everywhere else
- Thus, we see that a typical element of $\check{z}_{it} = z_{it} - \bar{z}_{i.} + (\sigma_\varepsilon / T\sigma_1)\, \bar{z}_{i.}$
- Condensing on notation we have that $\check{z}_{it} = z_{it} - \theta \bar{z}_{i.}$ where $\theta = 1 - (\sigma_\varepsilon / \sigma_1)$

- This almost looks like the within estimator for the fixed effects framework
- Recall from lecture 3 that a typical transformed element there had $\tilde{z}_{it} = z_{it} - \bar{z}_i$.
- Here the difference is the presence of $(\sigma_\varepsilon/\sigma_1)$
- When this component is 0 we have that the estimators for the random effects and fixed effects frameworks are the same
- When is $\sigma_\varepsilon/\sigma_1 = 0$?
- We would need the variation in c to be orders of magnitude larger than the variation in the idiosyncratic shocks
- When is $\sigma_\varepsilon/\sigma_1 = 1$?
- We would have no variation in c, i.e. individual heterogeneity is identical, so we just have an intercept

- A different decomposition of the random effects estimator is both intuitive and will be useful for later discussions (such as when we discuss the Hausman test in Lecture 5)
- The Between estimator is rarely used in practice, but appears in many algebraic derivations and is useful for helping to gain perspective
- The Between estimator is the OLS estimator of the transformed unobserved effects model

$$Py = PX\beta + Pu \tag{11}$$

- The estimator, denoted $\hat{\beta}_{Between}$, is

$$\hat{\beta}_{Between} = \left(X'PX\right)^{-1} X'Py \tag{12}$$

- Maddala (1971) uses the between estimator to construct a useful decomposition for $\hat{\beta}_{GLS}$

- Consider the following system of $2NT$ observations

$$\begin{pmatrix} Qy \\ Py \end{pmatrix} = \begin{pmatrix} QX \\ PX \end{pmatrix} \beta + \begin{pmatrix} Qu \\ Pu \end{pmatrix} \tag{13}$$

- Maddala (1971) shows that GLS estimation of this system produces exactly the random effects estimator $\hat{\beta}_{GLS}$
- What is interesting about this formulation is that we can decompose the GLS estimator into 'within' and 'between' components

- To start note that the variance-covariance matrix of $\begin{pmatrix} Qu \\ Pu \end{pmatrix}$ is

$$\Sigma = \begin{bmatrix} \sigma_\varepsilon^2 Q & 0 \\ 0 & \sigma_1^2 P \end{bmatrix}$$

with inverse

$$\Sigma^{-1} = \begin{bmatrix} \frac{1}{\sigma_\varepsilon^2} Q & 0 \\ 0 & \frac{1}{\sigma_1^2} P \end{bmatrix}$$

- GLS estimation using this inverse matrix produces

$$\begin{aligned}\hat{\beta}_{GLS} &= \left(\frac{1}{\sigma_\varepsilon^2}X'QX + \frac{1}{\sigma_1^2}X'PX\right)^{-1}\left(\frac{1}{\sigma_\varepsilon^2}X'Qy + \frac{1}{\sigma_1^2}X'Py\right) \\ &= \left(X'QX + \frac{\sigma_\varepsilon^2}{\sigma_1^2}X'PX\right)^{-1}\left(X'Qy + \frac{\sigma_\varepsilon^2}{\sigma_1^2}X'Py\right) \\ &= \left(X'QX + \phi^2 X'PX\right)^{-1}\left(X'Qy + \phi^2 X'Py\right) \qquad (14)\end{aligned}$$

- This derivation can be further decomposed
- Let $W = \left(X'QX + \phi^2 X'PX\right)$
- We have

$$\begin{aligned}\hat{\beta}_{GLS} =& W^{-1}\Big(X'QX\left(X'QX\right)^{-1}X'Qy \\ & \qquad +\phi^2 X'PX\left(\phi^2 X'PX\right)^{-1}\phi^2 X'Py\Big) \\ =& W^{-1}\left(X'QX\tilde{\beta} + \phi^2 X'PX\hat{\beta}_{Between}\right) \\ =& W_1\tilde{\beta} + (I - W_1)\hat{\beta}_{Between} \qquad (15)\end{aligned}$$

- Now, an interesting question is what does the Between estimator capture/measure?
- Notice that the Between regression is

$$\bar{y}_{i\cdot} = \alpha + \bar{X}'_{i\cdot}\beta + \bar{u}_{i\cdot}$$

- Thus, β is identified off of time variation in the mean of each variable
- No person specific variation is used to estimate β
- Many consider this a serious limitation and it is partly the reason why the Between estimator is not used in practice

- As it stands the GLS estimator for the random effects framework is infeasible since σ_1 and σ_ε are unknown
- We can construct estimators of the unknown variance terms to produce a feasible GLS estimator
- How do we estimate σ_1 and σ_ε?

- To start, note that $Pu \sim D(0, \sigma_1^2 P)$ and $Qu \sim D(0, \sigma_\varepsilon^2 Q)$
- This suggests the estimators

$$\widehat{\sigma}_1^2 = \frac{u'Pu}{tr(P)} \tag{16}$$

and

$$\widehat{\sigma}_\varepsilon^2 = \frac{u'Qu}{tr(Q)} \tag{17}$$

- This follows directly from a similar setup in the cross-sectional case

- Note that $tr(A \otimes B) = tr(A)tr(B)$ so $tr(Q) = tr(I_N)tr(E_T) = N \cdot (T-1)$ and $tr(P) = tr(I_N)tr(\bar{J}_T) = N \cdot 1 = N$
- We have the solutions

$$\widehat{\sigma}_1^2 = \frac{T \sum_{i=1}^{N} \bar{u}_{i.}^2}{N} \quad (18)$$

and

$$\widehat{\sigma}_\varepsilon^2 = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} (u_{it} - \bar{u}_{i.})^2}{N(T-1)} \quad (19)$$

- Of course the estimators in (18) and (19) are still not functional because they rely on u, which is unobserved
- There have been several proposals for replacing u with an estimator
- The main papers in this area are Wallace and Hussain (1969), Amemiya (1971), Nerlove (1971) and Swamy and Aurora (1972)

- Wallace and Hussain (1969) proposed replacing u with the residuals obtained from OLS estimation of the unobserved effects panel data model
- Under the random effects framework the OLS estimator of β is a consistent estimator so the residuals are reasonable estimates for the unknown u
- The Wallace and Hussain (1969) procedure is
  - Step 1: Estimate the unobserved effects model using pooled OLS, obtain residuals
  - Step 2: Use residuals in place of u in (18) and (19)
  - Step 3: Use estimates of σ_1^2 and σ_ε^2 to construct Ω
  - Step 4: Obtain the GLS estimator

- Amemiya (1971) shows that the Wallace and Hussain (1969) approach suffers some theoretical drawbacks
- Amemiya (1971) proposed replacing u with the residuals obtained from within estimation of the unobserved effects panel data model
- Under the random effects framework the within estimator of β is a consistent estimator so the residuals are reasonable estimates for the unknown u
- The Amemiya (1971) procedure is
  - Step 1: Estimate the unobserved effects model using the within estimator, obtain residuals
  - Step 2: Use residuals in place of u in (18) and (19)
  - Step 3: Use estimates of σ_1^2 and σ_ε^2 to construct Ω
  - Step 4: Obtain the GLS estimator

- Nerlove (1971) constructs σ_1^2 by using an estimator of σ_c^2
- Nerlove (1971) proposes estimating σ_c^2 using the estimated fixed effects from within estimation of the unobserved effects panel data model and estimating σ_ε^2 from the residuals sum or squares obtained by within estimation of the unobserved effects panel data model
- The Nerlove (1971) procedure is
  - Step 1: Estimate the unobserved effects model using the within estimator, obtain residuals and estimated fixed effects
  - Step 2: Construct $\hat{\sigma}_c^2 = \sum_{i=1}^{N} \left(\hat{c}_i - \bar{\hat{c}}\right)^2 / (N-1)$
  - Step 3: Construct $\hat{\sigma}_\varepsilon^2 = \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{\varepsilon}_{it}^2 / NT$
  - Step 4a: Use estimates of σ_c^2 and σ_ε^2 to construct $\hat{\sigma}_1^2$
  - Step4b: Use estimates of σ_1^2 and σ_ε^2 to construct Ω
  - Step 5: Obtain the GLS estimator

- Swamy and Arora (1972) proposed estimating σ_ε^2 and σ_1^2 using two different estimators
- The Swamy and Arora (1972) approach does not replace the errors in (18) and (19) but constructs entirely different estimators altogether
- They suggest using the residual variance estimator from the within model to estimate σ_ε^2 and the residual variance estimator from between model to estimate σ_1^2
- The Swamy and Arora (1972) procedure is
  - Step 1: Construct $\hat{\sigma}_\varepsilon^2$ from the residuals from within estimation of the unobserved effects model
  - Step 2: Construct $\hat{\sigma}_1^2$ from the residuals from between estimation of the unobserved effect model
  - Step 3: Use these estimates of σ_1^2 and σ_ε^2 to construct Ω
  - Step 4: Obtain the GLS estimator

- In practice there is no clear approach that one should favor
- One concern is what to do when one obtains a negative estimate of σ_c^2
- While an estimate of σ_1^2 is needed for GLS estimation, interest hinges on σ_c^2
- If $\hat{\sigma}_c^2 < 0$ this implies that $\hat{\sigma}_1^2 < \hat{\sigma}_\varepsilon^2$ which does not make sense
- Only Nerlove's (1971) approach guarantees a nonnegative estimate of σ_c^2

- An existing solution is to replace a negative estimate of $\hat{\sigma}_c^2$ with 0
- A simulation study by Maddala and Mount (1973) found that negative estimates of σ_c^2 occurred infrequently (in their simulated data) and was most prevalent when σ_c^2 was small
- It appears that this issue is not a serious problem; if you encounter it in applied work you can use an alternative approach to estimate the error component variances or simply replace $\hat{\sigma}_1^2$ with $\hat{\sigma}_\varepsilon^2$
- Further simulation studies by Baltagi (1981) find that there is little difference in the finite sample properties of the GLS estimator for β across the different approaches to estimating the unknown error variances

- Regardless of which estimation approach you use, make sure that you know which one is the default in your statistical software
- For example, in `R`, the `plm` command uses as a default the Swamy and Arora (1972) approach when the random effects estimator is chosen
- At a minimum you need to know which approach is used when using canned statistical software

- Discuss estimation of the unobserved effects model under the random effects framework
- Described the unique variance-covariance structure of the errors term in this model
- Proposed a GLS estimator that exploits this variance-covariance structure
- Learned several approaches to estimate the unknown parameters in the variance-covariance structure