Introduction to Econometrics 2: Recitation 6

Seung-hun Lee

Columbia University

February 26th, 2020

Random Effects Model

Setup

- Consider the following data generating process

$$y_{it} = x_{it}'\beta_1 + \alpha_i + u_{it}, \quad (i = 1, .., n \text{ and } \; t = 1, .., T)$$

  where we assume that $E(\alpha_i) = 0, E(\alpha_i u_{it}) = 0$, u_{it} is IID across i and t and independent of x_{it}.
- A key assumption that separates random effects from fixed effects is α_i and x_{it} is now **uncorrelated**.
- Assuming strict exogeneity, POLS becomes consistent
- This is because

$$\hat{\beta}_{POLS} - \beta_1 = \left(\sum_{i=1}^{n} \sum_{t=1}^{T} x_{it} x_{it}' \right)^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} x_{it} v_{it}$$

  where $v_{it} = \alpha_i + u_{it}$. The $\frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} x_{it} v_{it}$ converges in probability to 0

Random Effects Model

GLS estimation: Motivation

- However, this is not really the best we can do.
- The reason is that the composite error from different time periods are serially correlated. This can be seen by

$$
\begin{aligned}
cov(v_{it}, v_{is}) &= cov(\alpha_i + u_{it}, \alpha_i + u_{is}) \\
&= E[(\alpha_i + u_{it})(\alpha_i + u_{is})] - E[\alpha_i + u_{it}]E[\alpha_i + u_{is}] \\
&= E[\alpha_i^2 + \alpha_i u_{it} + \alpha_i u_{is} + u_{it} u_{is}] - (0+0)(0+0) \\
&= E[\alpha_i^2] + E[\alpha_i u_{it}] + E[\alpha_i u_{is}] + E[u_{it} u_{is}] \\
&= var(\alpha_i)
\end{aligned}
$$

- Therefore, an alternative method - GLS - would allow for the most efficient estimation

Random Effects Model

GLS estimation: Derivation

- The key to constructing a GLS is to find a variance-covariance matrix for $\mathbf{v}_i$ from

$$\mathbf{y}_i = \mathbf{X}_i \beta_1 + \mathbf{v}_i$$

where $\mathbf{y}_i = \begin{pmatrix} y_{i1} \\ ... \\ y_{iT} \end{pmatrix}$, $\mathbf{X}_i = \begin{pmatrix} x'_{i1} \\ ... \\ x'_{iT} \end{pmatrix}$, $\mathbf{v}_i = \begin{pmatrix} v_{i1} \\ ... \\ v_{iT} \end{pmatrix} = \begin{pmatrix} \alpha_i + u_{i1} \\ ... \\ \alpha_i + u_{iT} \end{pmatrix}$.

- Define 1_T as a T-dimensional column vector of 1's. Then,

$$\mathbf{v}_i = 1_T \alpha_i + \mathbf{u}_i$$

- Variance matrix is obtained by the $E[(X - E(X))(X - E(X))']$
- So in our case..(next slide)

Random Effects Model

GLS estimation: Derivation

- Once variance-covariance matrix V is obtained, we premultiply $V^{-1/2}$ to get

$$V^{-1/2}\mathbf{y}_i = V^{-1/2}\mathbf{X}_i\beta_1 + V^{-1/2}\mathbf{v}_i$$

- The resulting GLS estimator would be

$$\left(\sum_{i=1}^{n}(V^{-1/2}\mathbf{X}_i)'(V^{-1/2}\mathbf{X}_i)\right)^{-1}\sum_{i=1}^{n}(V^{-1/2}\mathbf{X}_i)'(V^{-1/2}\mathbf{y}_i) =$$

$$\left(\sum_{i=1}^{n}\mathbf{X}_i'V^{-1}\mathbf{X}_i\right)^{-1}\sum_{i=1}^{n}\mathbf{X}_i'V^{-1}\mathbf{y}_i$$

Dynamic Panel Data

Motivation

- Consider a setting where

$$y_{it} = \rho y_{it-1} + \alpha_i + u_{it}$$

  where $t = 0, 1, ..., T$ and $i = 1, .., n$
- Assume that $E[\alpha_i] = 0, E[u_{it}] = 0, E[\alpha_i u_{it}] = 0 \ \forall i, t$ and $E[u_{it} u_{is}] = 0$ for $t \neq s$
- Also, the initial observation y_{i0} satisfies $E[y_{i0} u_{it}] = 0 \ (t \geq 1)$ and that $E|u_{it}|^{2+\delta} \leq M < \infty (\delta > 0)$.
- Also note that observations are independent across i.
- The lagged dependent variable enters as the regressor.
- All of the previous methods - random effects and fixed effects - are inconsistent.

Dynamic Panel Data

Inconsistency of POLS

- Let $v_{it} = \alpha_i + u_{it}$.
- The OLS estimates would be

$$\hat{\rho} - \rho = \left(\sum_{i=1}^{n} \sum_{t=0}^{T} y_{it-1}^2 \right)^{-1} \sum_{i=1}^{n} \sum_{t=0}^{T} y_{it-1}(\alpha_i + u_{it})$$

- Since $y_{it-1} = \rho y_{it-2} + \alpha_i + u_{it-1}$, the $\frac{1}{nT} \sum_{i=1}^{n} \sum_{t=0}^{T} y_{it-1}\alpha_i$ term does not converge in probability to 0.

Dynamic Panel Data

Inconsistency of FD

- Even if we get rid of α_i by using the first difference at $T = 2$, ρ estimates are still inconsistent.
- Start with

$$\Delta y_{i2} = \rho \Delta y_{i1} + \Delta u_{i2}$$

- We can now show that the regressor and the error term are correlated, since

$$\begin{aligned} cov(\Delta y_{i1}, \Delta u_{i2}) &= E[(y_{i1} - y_{i0})(u_{i2} - u_{i1})] \\ &= E[y_{i1} u_{i2}] - E[y_{i1} u_{i1}] - E[y_{i0} u_{i2}] + E[y_{i0} u_{i1}] \end{aligned}$$

- Because $E[y_{i1} u_{i1}]$ contains terms from the 1st period, this contains term that is nonzero.

Dynamic Panel Data

Inconsistency of WE

- Even for the within estimator, which is written as

$$y_{it} - \frac{1}{T}\sum_{i=1}^{T} y_{it} = \rho\left(y_{it-1} - \frac{1}{T}\sum_{i=1}^{T} y_{it-1}\right) + u_{it} - \frac{1}{T}\sum_{i=1}^{T} u_{it}$$

  The regressor contains $y_{i0}, .., y_{iT-1}$ and residuals contain $u_{i1}, .., u_{iT}$
- There are overlapping time periods, implying that the regressor becomes endogenous.
- Since LSDV is numerically identical to WE, this also is inconsistent

Dynamic Panel Data

Anderson-Hsiao Estimator

- First difference the original equation and obtain

$$\Delta y_{it} = \rho \Delta y_{it-1} + \Delta u_{it}$$

  where possible values of i remain the same but $t = 2, ..., T$
- The suggested IV is that Δy_{it-1} be instrumented with y_{it-2}
  - **Relevancy**: $\Delta y_{it-1} = y_{it-1} - y_{it-2}$ contains y_{it-2}
  - **Exogeneity**: Note that

$$\begin{aligned} cov(y_{it-2}, \Delta u_{it}) &= E[y_{it-2}, u_{it} - u_{it-1}] \\ &= E[y_{it-2} u_{it}] - E[y_{it-2} u_{it-1}] = 0 \end{aligned}$$

    Since we are assuming $E[y_{i0} u_{it}] = 0$ for $t \geq 1$, we can expand to

$$\begin{aligned} E[y_{i1} u_{it}] \ (t \geq 2) &= E[(\rho y_{i0} + \alpha_i + u_{i1}) u_{it}] \\ &= \rho E[y_{i0} u_{it}] + E[\alpha_i u_{it}] + E[u_{i1} u_{it}] = 0 \end{aligned}$$

    Therefore, we can generalize to $E[y_{is} u_{it}] = 0$ for $t \geq s + 1$.

Dynamic Panel Data

Anderson-Hsiao Estimator

- Define

$$\Delta\mathbf{y}_i = \rho\Delta\mathbf{y}_{i,-1} + \Delta\mathbf{u}_i$$

where $\Delta\mathbf{y}_i = \begin{pmatrix} \Delta y_{i2} \\ \dots \\ \Delta y_{iT} \end{pmatrix}$, $\Delta\mathbf{y}_{i,-1} = \begin{pmatrix} \Delta y_{i1} \\ \dots \\ \Delta y_{iT-1} \end{pmatrix}$ and $\Delta\mathbf{u}_i = \begin{pmatrix} \Delta u_{i2} \\ \dots \\ \Delta u_{iT} \end{pmatrix}$.

- The matrix of instruments Z_i would be

$$Z_i = \begin{pmatrix} y_{i0} & 0 & 0 & \dots \\ 0 & y_{i1} & 0 & \dots \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & y_{iT-2} \end{pmatrix} \in \mathbb{R}^{(T-1)\times(T-1)}$$

- So we solve the sample analogue of $E[Z_i'\Delta\mathbf{u}_i] = 0$ to obtain

$$\hat{\rho} = \left(\sum_{i=1}^{n} Z_i'\Delta\mathbf{y}_{i,-1}\right)^{-1} \sum_{i=1}^{n} Z_i'\Delta\mathbf{y}_i$$

Dynamic Panel Data

Arellano-Bond Estimator: No regressors

- Arellano and Bond (1991) suggests that to instrument for Δy_{it-1}, we use $y_{i0}, ..., y_{it-2}$ as instruments
  - **Relevancy:** It should be clear why y_{it-2} is relevant. As for others, since $y_{it-1} = \rho y_{it-2} + u_{it-1}$ and $y_{it-2} = \rho y_{it-3} + u_{it-2}$ We can write recursively that

$$y_{it-1} = \rho^2 y_{it-3} + \rho u_{it-2} + u_{it-1}$$

    ... and so on. Therefore, we can verify relevancy.
  - **Exogeneity:** Note that $cov(y_{is}, \Delta u_{it})$ for any $s < t$ is 0, as we have shown above. So exogeneity holds as well.

Dynamic Panel Data

Arellano-Bond Estimator: No regressors

- The generalized approach using matrix will be similar except for the instrument matrix Z_i.

$$Z_i = \begin{pmatrix} y_{i0} & 0 & 0 & \dots \\ 0 & (y_{i0}, y_{i1}) & 0 & \dots \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & (y_{i0}, \dots, y_{iT-2}) \end{pmatrix} \in \mathbb{R}^{(T-1)\times \frac{T(T-1)}{2}}$$

- This is an overidentified case. So we would need to use a GMM criterion with a weight matrix W_n. This would result in

$$\hat{\rho} = \arg\min_{\rho} \left\{ n \times \frac{1}{n} \sum_{i=1}^{n} (Z_i' \Delta \mathbf{y}_i - \rho Z_i' \Delta \mathbf{y}_{i,-1})' W_n \frac{1}{n} \sum_{i=1}^{n} (Z_i' \Delta \mathbf{y}_i - \rho Z_i' \Delta \mathbf{y}_{i,-1}) \right\}$$

$$\Longrightarrow \left[\left(\sum_{i=1}^{n} \Delta \mathbf{y}_{i,-1}' Z_i \right) W_n \left(\sum_{i=1}^{n} Z_i' \Delta \mathbf{y}_{i,-1} \right) \right]^{-1} \left(\sum_{i=1}^{n} \Delta \mathbf{y}_{i,-1}' Z_i \right) W_n \left(\sum_{i=1}^{n} Z_i' \Delta \mathbf{y}_i \right)$$

Weighting Matrix Selection

- Which W_n leads to the lowest variance possible?
- If $g(Z_i, \rho)$ is the moment condition, the following would qualify as the most efficient weighting matrix.

$$W_n = E[g(Z_i, \rho)g(Z_i, \rho)']^{-1}$$

- In our context, $g(Z_i, \rho)$ would be equivalent to $Z_i' \Delta u_i$
- So we need to find sample analogue of

$$E[Z_i' \Delta \mathbf{u}_i \Delta \mathbf{u}_i' Z_i] \implies \frac{1}{n} \sum_{i=1}^{n} Z_i \Delta \mathbf{u}_i \Delta \mathbf{u}_i' Z_i$$

Dynamic Panel Data

Weighting Matrix Selection: Homoskedastic errors

- $E[u_{it}^2] = \sigma_u^2$, so we write

$$E[\Delta \mathbf{u}_i \Delta \mathbf{u}_i] = \begin{bmatrix} 2\sigma_u^2 & -\sigma_u^2 & ... & 0 \\ -\sigma_u^2 & 2\sigma_u^2 & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & 2\sigma_u^2 \end{bmatrix}$$

- Define matrix $H \in \mathbb{R}^{(T-1)\times(T-1)}$ to have 2's in the diagonal elements, -1's in the immediate off-diagonals, and 0 everywhere else.
- This implies that the weighting matrix that we are looking for is

$$E[Z_i' \Delta \mathbf{u}_i \Delta \mathbf{u}_i' Z_i]^{-1} = \left(\frac{1}{n} \sum_{i=1}^{n} Z_i H Z_i \right)^{-1}$$

where σ_u^2 is taken out since scaling W_n by a scalar does not affect the value of the estimator.

Weighting Matrix Selection: Heteroskedastic errors

- We take an approach similar to the two-step GMM estimator
- The optimal weighting matrix in this case would be

$$W_n = \left(\frac{1}{n} \sum_{i=1}^{n} Z_i \Delta \tilde{\mathbf{u}}_i \Delta \tilde{\mathbf{u}}_i' Z_i \right)^{-1}$$

- $\Delta \tilde{\mathbf{u}}_i = \Delta \mathbf{y}_i - \tilde{\rho} \Delta \mathbf{y}_{i,-1}$, a residual from the preliminary estimator $\tilde{\rho}$
- The preliminary estimator could be either from $W_n = I_{T(T-1)/2}$ or from the one-step GMM estimator that we derived earlier.

Dynamic Panel Data

Overidentification test

- Because we are using more moment conditions that the number of endogenous variables, this is when we could test for an overidentification restriction.
- Suppose that W_n is the efficient weighting matrix. Then, similar to the GMM overidentification test, we are testing

$$H_0 : E[g(Z_i, \rho)] = 0, \quad H_1 : E[g(Z_i, \rho)] \neq 0$$

- We can construct the following test statistic

$$J = n\bar{g}_n(\hat{\rho})' W_n \bar{g}_n(\hat{\rho})$$

- Under H_0, J has a limiting distribution $\chi^2_{\left(\frac{T(T-1)}{2}-1\right)}$. We lose one degree of freedom since we have used one estimator for ρ.

Dynamic Panel Data

Arellano-Bond Estimator: With Regressors

- Now we generalize further by including regressors $x_{it} \in \mathbb{R}^k$.
- We can write the data generating process as

$$y_{it} = \rho y_{it-1} + \beta' x_{it} + \alpha_i + u_{it}$$

- There are two types of assumption we can put on the regressors

Predetermined vs. Strict Exogeneity

- We say that the regressors are **predetermined** if

$$E[u_{it}|x_{i0}, ..., x_{it}] = 0 \text{ for each } t = 0, ..., T$$

- We say that the regressors are **strictly exogenous** if

$$E[u_{it}|x_{i0}, ..., x_{iT}] = 0 \text{ for all } t = 0, ..., T$$

Dynamic Panel Data

Arellano-Bond Estimator: With Regressors

- We allow α_i to be correlated with x_{it} and y_{it-1}
- Take the first difference

$$\Delta y_{it} = \rho \Delta y_{it-1} + \beta' \Delta x_{it} + \Delta u_{it}$$

However, the equation is endogenous because of the two reasons

- Δy_{it-1} is endogenous as for the similar reasons as before
- Δx_{it} is endogenous under predetermined case since

$$\begin{aligned} E[\Delta x_{it} \Delta u_{it}] &= E[(x_{it} - x_{it-1})(u_{it} - u_{it-1})] \\ &= E[x_{it} u_{it}] - E[x_{it} u_{it-1}] - E[x_{it-1} u_{it}] + E[x_{it-1} u_{it-1}] \end{aligned}$$

While other terms are 0 by predetermined assumption, this assumption is silent about $E[x_{it} u_{it-1}]$. This is where the nonzero correlation is from.

Dynamic Panel Data

Arellano-Bond Estimator: Predetermined Regressors

- Rewrite the differenced equation and stack in the vector form for each individual by writing

$$\Delta \mathbf{y}_i = \Delta \mathbf{w}_i \delta + \Delta \mathbf{u}_i$$

$$\Delta \mathbf{w}_i = [\Delta \mathbf{y}_{i,-1},\ \Delta \mathbf{x}_i] \in \mathbb{R}^{(T-1)\times(k+1)},\ \Delta \mathbf{x}_i = \begin{pmatrix} x_{i2} \\ \dots \\ x_{iT} \end{pmatrix},\ \delta = \begin{pmatrix} \rho \\ \beta' \end{pmatrix}$$

- I define

$$\mathbf{y}_{it} = \begin{pmatrix} y_{i0} \\ \dots \\ y_{it} \end{pmatrix}, \mathbf{x}_{it} = \begin{pmatrix} x'_{i0} \\ \dots \\ x'_{it} \end{pmatrix}$$

- The possible instruments for the endogenous variable Δx_{it} is any regressor up to period $t-1$, or $\mathbf{x}'_{it-1}$.
  - Exogeneity is satisfied since the error term is $u_{it} - u_{it-1}$, regressors $x_{i1}, \dots, x_{it-1}$ is not correlated to this by predeterminedness assumption
  - $\mathbf{x}'_{it-1}$ includes x_{it-1}, the IV's are relevant as well

Arellano-Bond Estimator: Predetermined Regressors

- Thus, we can write

$$Z_i = \begin{pmatrix} y_{i0}, \mathbf{x}'_{i1} & 0 & 0 & \dots & 0 \\ 0 & \mathbf{y}'_{i1}, \mathbf{x}'_{i2} & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \mathbf{y}'_{iT-2}, \mathbf{x}'_{iT-1} \end{pmatrix} \in \mathbb{R}^{(T-1)\times(T^2-1)}$$

- Then we take a GMM approach by using the moment condition $E[Z_i'\Delta\mathbf{u}_i] = 0$, which results in the following estimator

$$\hat{\delta} = \left[\left(\sum_{i=1}^{n} \Delta\mathbf{w}_i' Z_i \right) W_n \left(\sum_{i=1}^{n} Z_i' \Delta\mathbf{w}_i \right) \right]^{-1} \left(\sum_{i=1}^{n} \Delta\mathbf{w}_i' Z_i \right) W_n \left(\sum_{i=1}^{n} Z_i' \Delta\mathbf{y}_i \right)$$

Arellano-Bond Estimator: Strictly Exogenous Regressors

- Now the entire elements in $\mathbf{x}'_{iT}$ is a valid IV
  - Strict exogeneity prevents any correlation between error terms of all time periods
  - There are elements in $\mathbf{x}'_{iT}$ that are included in Δy_{it} for each t
- write the Z_i matrix as

$$Z_i = \begin{pmatrix} y_{i0}, \mathbf{x}'_{iT} & 0 & 0 & ... & 0 \\ 0 & \mathbf{y}'_{i1}, \mathbf{x}'_{iT} & 0 & ... & 0 \\ ... & ... & ... & ... & ... \\ 0 & 0 & 0 & ... & \mathbf{y}'_{iT-2}, \mathbf{x}'_{iT} \end{pmatrix} \in \mathbb{R}^{(T-1)\times\left(\frac{T(T-1)}{2}+(T+1)(T-1)\right)}$$

- We still work with the same moment condition $E[Z_i'\Delta\mathbf{u}_i] = 0$.
- The resulting estimator for δ shares the same expression as in the predetermined case, but with different matrix for Z_i.