Recitation 7: Fixed effects and Dynamic panels

Seung-hun Lee

Columbia University
Introduction to Econometrics II Recitation

March 7th, 2022

Fixed effects regression

Fixed effects: Omitting c_i induces OVB

- The unobserved fixed effect c_i is correlated with x_{it}.

$$y_{it} = x_{it}'\beta + c_i + e_{it}$$

- POLS and RE no longer consistent: Relied on the fact that c_i is uncorrelated with x_{it}
- Therefore, it is essential that we minimize the role of c_i in the above equation in order to get the consistent estimation. $\rightarrow$ we do that by purging it using transformations
- There are three ways to approach fixed effects estimation
  - Within estimation (WE)
  - Least square dummy variables (LSDV)
  - First difference (FD)

WE weeds out c_i with demeaning!

- Write $\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}$ and similarly for other variables to get $\bar{y}_i = \bar{x}_i'\beta + c_i + \bar{e}_i$
- Even if y_i is average across time, we still get c_i itself
- Subtract the cross-sectional equation from the original data generating process to get

$$\ddot{y}_{it} = \ddot{x}_{it}'\beta + \ddot{e}_{it}, \quad (i = 1, .., n, \text{ and } t = 1, .., T)$$

  where $\ddot{y}_{it} = y_{it} - \bar{y}_i$
- The within estimator is obtained by taking an OLS to above equation

$$\hat{\beta}_{WE} = \left(\sum_{i=1}^{n}\sum_{t=1}^{T} \ddot{x}_{it}\ddot{x}_{it}'\right)^{-1} \sum_{i=1}^{n}\sum_{t=1}^{T} \ddot{x}_{it}\ddot{y}_{it} = \left(\sum_{i=1}^{n} \ddot{X}_i'\ddot{X}_i\right)^{-1} \sum_{i=1}^{n} \ddot{X}_i'\ddot{y}_i$$

WE leads to consistent estimates

FE assumptions

FE1 We assume strict exogeneity $E[e_{it}|X_i, c_i] = 0$

FE2 $\text{rank}\left(E[\ddot{X}_i'\ddot{X}_i]\right) = \text{rank}\left(\sum_{t=1}^{T} E[\ddot{x}_{it}'\ddot{x}_{it}]\right) = k$ (full column rank)

FE3 Conditionally spherical variance matrix: $E[e_i e_i'|X_i, c_i] = \sigma_e^2 I_T$

- We rewrite the WE estimator os

$$\hat{\beta}_{WE} - \beta = \left(\sum_{i=1}^{n}\sum_{t=1}^{T}\ddot{x}_{it}\ddot{x}_{it}'\right)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}\ddot{x}_{it}\ddot{e}_{it}$$

- $E[\ddot{x}_{it}\ddot{e}_{it}] = E[x_{it}e_{it}] - E[x_{it}\bar{e}_i] - E[\bar{x}_i e_{it}] + E[\bar{x}_i\bar{e}_i]$
- Since $\bar{x}_i, \bar{e}_i$ incorporates regressors and errors from all time periods, applying strict exogeneity (and strict exogeneity only) reduces the above equation to 0

Asymptotic distribution of WE

- Variance-covariance of $\ddot{e}_{it}$ is

$$
\begin{aligned}
E[\ddot{e}_{it}^2] &= E[(e_{it} - \bar{e}_i)^2] = E[e_{it}^2] - 2E[e_{it}\bar{e}_i] + E[\bar{e}_i^2] \\
&= \sigma_e^2 - \frac{2}{T}\sigma_e^2 + \frac{1}{T}\sigma_e^2 = \sigma_e^2\left(1 - \frac{1}{T}\right) \\
E[\ddot{e}_{it}\ddot{e}_{is}] &= E[(e_{it} - \bar{e}_i)(e_{is} - \bar{e}_i)] = E[e_{it}e_{is}] - E[e_{it}\bar{e}_i] - E[e_{is}\bar{e}_i] + E[\bar{e}_i^2] \\
&= 0 - \frac{1}{T}\sigma_e^2 - \frac{1}{T}\sigma_e^2 + \frac{1}{T}\sigma_e^2 = -\frac{1}{T}\sigma_e^2
\end{aligned}
$$

- So $\ddot{e}_{it}$ is conditionally homoskedastic and have negative serial correlation (minor problem due to nature of demeaning, Wooldridge 2010)
- The asymptotic distribution of the WE estimator is

$$\sqrt{n}(\hat{\beta}_{WE} - \beta) \sim N(0, E[\ddot{X}_i'\ddot{X}_i]^{-1}E[\ddot{X}_i'e_ie_i'\ddot{X}_i]E[\ddot{X}_i'\ddot{X}_i]^{-1})$$

Asymptotic variance estimate of WE

- If we impose **FE3**. Then the asymptotic variance is $\sigma_e^2 E[\ddot{X}_i' \ddot{X}_i]^{-1}$.
- The estimator of the asymptotic variance would be $\widehat{\sigma}_e^2 \left(n^{-1} \sum_{i=1}^{n} \ddot{X}_i' \ddot{X}_i\right)^{-1}$.
- To obtain $\widehat{\sigma}_e^2$, we start from our previous finding that $E[\ddot{e}_{it}^2] = \frac{T-1}{T}\sigma_e^2$. This implies that

$$\frac{1}{n(T-1)} \sum_{i=1}^{n} \sum_{t=1}^{T} E[\ddot{e}_{it}^2] = \sigma_e^2$$

- Then, we apply the (small-sample) correction by subtracting for k regressors. Thus, the estimate of σ_e^2 is

$$\widehat{\sigma}_e^2 = \frac{1}{n(T-1)-k} \sum_{i=1}^{n} \sum_{t=1}^{T} \widehat{\ddot{e}}_{it}$$

where $\widehat{\ddot{e}}_{it}$ is obtained from the OLS residual of the demeaned data generating process.

Matrix notation for WE

- Stack up the each observation: $y_i = X_i\beta + 1_T c_i + e_i$
- Define $Q_T \equiv I_T - 1_T(1_T'1_T)^{-1}1_T'$, where $(1_T'1_T)^{-1} = T^{-1}$ and $1_T1_T'$ is the T-dimensional square matrix of 1's as elements. (Symmetric and idempotent!)
- Now, premultiply Q_T to the indivudually-stacked data generating process to get $Q_T y_i = Q_T X_i\beta + Q_T 1_T c_i + Q_T e_i$
- A key feature is that $Q_T 1_T = 0$ and $Q_T y_i = \ddot{y}_i$. The latter is because

$$Q_T y_i = y_i - \frac{1}{T}\begin{pmatrix} 1 & 1 & \dots & 1 \\ \dots & \dots & \dots & \dots \\ 1 & 1 & \dots & 1 \end{pmatrix} \underbrace{\begin{pmatrix} y_{i1} \\ y_{i2} \\ \dots. \\ y_{iT} \end{pmatrix}}_{y_i} = y_i - \frac{1}{T}\begin{pmatrix} \sum_t y_{it} \\ \dots \\ \sum_t y_{it} \end{pmatrix} = \begin{pmatrix} y_{i1} - \frac{1}{T}\sum_t y_{it} \\ \dots \\ y_{iT} - \frac{1}{T}\sum_t y_{it} \end{pmatrix} = \ddot{y}_i$$

- Thus, $\hat{\beta}_{WE} = \left(\sum_{i=1}^n X_i' Q_T X_i\right)^{-1} \sum_{i=1}^n X_i' Q_T y_i$

LSDV: Panel is OLS + many dummy variables

- Consider c_i as a parameter to be estimated for each i (each i has distinct intercept)
- $Dk_i = 1$ if $i = k$th individual (0 otherwise) $\rightarrow$ Total of $N-1$ of such dummy variables

$$y_{it} = x_{it}'\beta + D1_i c_1 + \ldots + D(n-1)_i c_{n-1} + u_{it}$$

- Or stack observations for all population: Get y, X, e
  - Let c_i vary across i using Kroenecker product
    $$(I_n \otimes 1_T)c = \begin{pmatrix} 1_T & \ldots & 0 \\ \ldots & \ldots & \ldots \\ 0 & \ldots & 1_T \end{pmatrix} \begin{pmatrix} c_1 \\ \ldots \\ c_n \end{pmatrix} = (I_n \otimes 1_T)c \in \mathbb{R}^{nT\times n} \times \mathbb{R}^{n\times 1} = \mathbb{R}^{nT\times 1}$$
  - Combine what we know to get

$$y = X\beta + \underbrace{(I_n \otimes 1_T)}_{=D}c + e$$

We then use the Frisch-Waugh-Lovell theorem to get $\hat{\beta}_{LSDV} = (X'M_DX)^{-1}(X'M_Dy)$

First-differenced models: Difference with lags to get rid of c_i

- Assume $T \geq 2$. We work with

$$\Delta y_{it} = \Delta x_{it}'\beta + \Delta u_{it} \;\; (i = 1, ..., n \text{ and } t = 2, .., T)$$

  where $\Delta y_{it} = y_{i,t} - y_{i,t-1}$
- By taking an OLS, we can obtain

$$\hat{\beta}_{FD} = \left(\sum_{i=1}^{n} \sum_{t=2}^{T} \Delta x_{it} \Delta x_{it}' \right)^{-1} \sum_{i=1}^{n} \sum_{t=2}^{T} \Delta x_{it} \Delta y_{it}$$

FD is also consistent

FD assumptions

FD1 We assume strict exogeneity $E[e_{it}|X_i, c_i] = 0$

FD2 $\text{rank}\left(E[\Delta X_i' \Delta X_i]\right) = \text{rank}\left(\sum_{t=2}^{T} E[\Delta x_{it} \Delta x_{it}']\right) = k$ (full column rank)

FD3 Conditionally spherical variance matrix: $E[\Delta e_i \Delta e_i' | X_i, c_i] = \sigma^2_{\Delta e} I_{T-1}$

- Write

$$\hat{\beta}_{FD} - \beta = \left(\sum_{i=1}^{n} \sum_{t=2}^{T} \Delta x_{it} \Delta x_{it}' \right)^{-1} \sum_{i=1}^{n} \sum_{t=2}^{T} \Delta x_{it} \Delta e_{it}$$

- $\hat{\beta}_{FD}$ is consistent since $E[\Delta x_{it} \Delta e_{it}] = 0$

$$\begin{aligned} E[\Delta x_{it} \Delta e_{it}] &= E[(x_{it} - x_{i,t-1})(e_{it} - e_{i,t-1})] \\ &= E[x_{it} e_{it}] - E[x_{it} e_{i,t-1}] - E[x_{i,t-1} e_{it}] + E[x_{i,t-1} e_{i,t-1}] \\ &= 0 - 0 - 0 + 0 = 0 \end{aligned}$$

Difference between FD and WE?

- T observations for WE vs $T - 1$ for FD
- Structure of error terms: If e_{it} is free from serial correlation (or e_t is an IID), then taking a FD would introduce serial correlation.

$$\begin{aligned} cov(\Delta e_{it}, \Delta e_{i,t-1}) &= E[e_{it}e_{it-1}] - E[e_{it}e_{it-2}] - E[e_{it-1}e_{it-1}] + E[e_{it-1}e_{it-2}] \\ &= 0 - 0 - var(e_{it-1}) + 0 \neq 0 \end{aligned}$$

- If e_t is autocorrelated to begin with (e.g. random walk),

$$e_{it} = e_{it-1} + \eta_{it} \;\; (E[\eta_{it}] = 0, E[\eta_{it}\eta_{is}] = 0(s \neq t), var(e_{it}) = \sigma^2)$$

Thus, better eliminate this autocorrelation using FD

So the point of learning all these are...?

- We can show that for $T = 2$ FD and WE are identical
- LSDV and WE is numberically identical
  - You need to use properties of Kroenecker products
  - With that, you can transform $\hat{\beta}_{LSDV} = (X'M_D X)^{-1}(X'M_D y)$ to $\hat{\beta}_{WE} = \left(\sum_{i=1}^{n} X_i' Q_T X_i\right)^{-1} \sum_{i=1}^{n} X_i' Q_T y_i$
- Proving equivalence of FD and WE when $T = 2$ is not as hard
- The latter is time-consuming - take a look at my notes

Hausman - Taylor allows estimating impact of time-invariant variables

- For all FE estimates, we cannot put time-invariant variables as controls
  - Erased due to the transformation process (WE, FD)
  - Absorbed by a separate variable for unobserved individual effects (LSDV)
- Hausman and Taylor (1981) propose an estimation approach based on method of moments that can identify the effects of the time-invariant variables
- The data generating process as $y_{it} = z_i'\gamma + x_{it}'\beta + c_i + e_{it}$
- Assume $E[z_i e_{it}] = 0$, $E[x_{it} e_i] = 0$, $E[z_i c_i] = 0$, and $E[x_{it} c_i] \neq 0$
- Then, we have two moment conditions that we can work with

$$E[\ddot{x}_{it} e_{it}] = E[\ddot{x}_{it}(y_{it} - x_{it}'\beta - z_i'\gamma)] = 0$$
$$E[z_i e_{it}] = E[z_i(y_{it} - x_{it}'\beta - z_i'\gamma)] = 0$$

Obtaining γ estimates!

- IV: Valid IV for this procedure is $\ddot{x}_{it}$ and z_i. We can obtain β and γ estimates by a 2SLS procedure with $\ddot{x}_{it}$ and z_i as the set of IVs.
- Method of moments: β estimate is obtained with any FE methods
  - From the moment condition and including $\hat{\beta}$, we get

$$z_i'(y_{it} - x_{it}'\hat{\beta} - z_i\gamma) = z_i'(\bar{y}_i - \bar{x}_i'\hat{\beta} - z_i\gamma) = 0$$

  - $\bar{e}_i$ can be ruled out since $E[z_i e_{it}] = 0$
  - Combining the information we have, we use method of moments to get

$$\frac{1}{n}\sum_{i=1}^{n} z_i'(\bar{y}_i - \bar{x}_i'\hat{\beta} - z_i\gamma) = 0 \iff \frac{1}{n}\sum_{i=1}^{n} z_i(\bar{y}_i - \bar{x}_i'\hat{\beta}) = \frac{1}{n}\sum_{i=1}^{n} z_i z_i'\gamma$$

  - So $\hat{\gamma} = \left(\sum_{i=1}^{n} z_i z_i'\right)^{-1}\left(\sum_{i=1}^{n} z_i(\bar{y}_i - \bar{x}_i\hat{\beta})\right)$.
  - In practice, this is obtained by regressing $\hat{c}_i$ on z_i and thus $\hat{\gamma} = (Z'Z)^{-1}Z'\hat{c}$

FE vs RE? Use Hausman principles!

- FE: always consistent, but inefficient if $E[X_i' c_i] = 0$
- RE: consistent and efficient if $E[X_i' c_i] = 0$, but otherwise inconsistent.
- We create this test statistic for the null hypothesis of $H_0 : E[X_i' c_i] = 0$

$$H \equiv (\hat{\beta}_{FE} - \hat{\beta}_{RE})'[\widehat{V}_{\beta_{FE} - \beta_{RE}}]^{-1}(\hat{\beta}_{FE} - \hat{\beta}_{RE}) \sim \chi^2_{\text{dim}\,(X)}$$

  where we can write $\widehat{V}_{\beta_{FE} - \beta_{RE}} = \widehat{V}_{\beta_{FE}} - \widehat{V}_{\beta_{RE}}$
- If the null is not rejected, then using RE is acceptable. Otherwise, RE is inconsistent and FE should be preferred.

Generalizing the FE structure

- We can also include unobserved time effects δ_t: $y_{it} = x_{it}'\beta + c_i + \delta_t + e_{it}$
- The fixed effects estimator should get rid of both c_i and δ_t using a two-step demeaning
- Define

$$\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it},\ \bar{y}_t = \frac{1}{n}\sum_{i=1}^{n} y_{it},\ \bar{y} = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} y_{it}$$

- We take $\tilde{y}_{it} = y_{it} - \bar{y}_i - \bar{y}_t + \bar{y}$. That is because

$$\begin{aligned}\tilde{y}_{it} &= (\alpha_i + \gamma_t + x_{it}\beta + e_{it}) - (\alpha_i + \bar{\gamma} + \bar{x}_i\beta + \bar{e}_i) - (\bar{\alpha} + \gamma_t + \bar{x}_t\beta + \bar{e}_t) + (\bar{\alpha} + \bar{\gamma} + \bar{\bar{x}}\beta + \bar{\bar{e}}) \\ &= (x_{it} - \bar{x}_i - \bar{x}_t + \bar{\bar{x}})\beta + (e_{it} - \bar{e}_i - \bar{e}_t + \bar{\bar{e}}) \\ &= \tilde{x}_{it}\beta + \tilde{e}_{it}\end{aligned}$$

- POLS on the above equation would lead to consistent estimates of β.

Interactive FE: Most saturated set of FE possible

- v_{it} can be written as $v_{it} = \lambda_i f_t + e_t$, where f_t is a vector of factors, and λ_i is a vector of factor loadings
- TWFE is a special case: $\lambda_i = \begin{pmatrix} 1 \\ c_i \end{pmatrix}$ and $f_t = \begin{pmatrix} \delta_t \\ 1 \end{pmatrix}$
- We can model for unobservable individual effects that may vary over time by putting a fixed effect for each individual-time level at the cost of increasing computational burden.

Dynamic panel models

DPD: None of the static panel methods lead to consistent estimates!

- For POLS, where $v_{it} = c_i + e_{it}$, The OLS estimates would be

$$\hat{\rho} - \rho = \left(\frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} y_{it-1}^2 \right)^{-1} \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} y_{it-1}(c_i + e_{it})$$

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

- For FD,

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

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

  Because of $E[y_{i1} e_{i1}]$ Δy_{i1} is an endogenous regressor.

DPD: None of the static panel methods lead to consistent estimates!

- Even for the within estimator, which is written as

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

  The regressor contains $y_{i0}, .., y_{iT-1}$ and residuals contain $e_{i1}, .., e_{iT}$. There are overlapping time periods, implying that the regressor becomes endogenous.

We also need different exogeneity assumption

DPD assumptions

DP1 Sequential exogeneity: $E[e_{it}|w_{it}, ..., w_{i1}, c_i] = 0$ for each t

- Or $E[w_{is}e_{it}] = 0$ for $s \in \{1, .., t\},\ t \in \{1, .., T\}$

DP2 Dynamic completeness: $E[y_{it}|x_t, y_{it-1}, x_{it-1}, y_{it-2}, ..., c_i] = E[y_{it}|x_t, y_{it-1}, c_i]$. This implies that x_t, y_{it-1} are all the lags needed and no information is lost by not including further lags. This implies that $E[e_{it}|x_t, y_{it-1}, x_{it-1}, y_{it-2}, ..., c_i] = 0$, or no residual correlation

- In our context, fine to use the two interchangeably

Sequential exogeneity allows feedback effects!

- It is difficult to take a stance on strict exogeneity, as the inclusion of the lagged variable introduces feedbacks in which x_{it} can be affected by past values of y_{it}, say y_{it-1}.
- In such case, the past shock e_{it-1} can affect values of x_{it}.
- Thus a flexible exogeneity assumption that takes into account these feedback effects are needed.
- The sequential exogeneity assumption implies the following
  - For $s \leq t$, $E[w'_{is}e_{it}] = 0$
  - For $s < t$, $E[e_{is}e_{it}] = 0$ for all possible values of t

Solution is to use instrument from within the model

- Internal instrument approach: Let $y_t = \alpha y_{t-1} + e_t, \;\; e_t \sim MA(1) = u_t + \theta u_{t-1}, \; u_t \sim WN$
- Here,

$$
\begin{aligned}
E[y_{t-1} e_t] &= E[(\alpha y_{t-2} + e_{t-1}) e_t] \\
&= E[e_{t-1} e_t] \\
&= E[(u_{t-1} + \theta u_{t-2})(u_t + \theta u_{t-1})] = \theta var(u_{t-1})
\end{aligned}
$$

- So y_{t-2} to instrument y_{t-1}?

$$
\begin{aligned}
E[y_{t-2} e_t] &= E[(\alpha y_{t-3} + e_{t-2}) e_t] \\
&= E[e_{t-2} e_t] \\
&= E[(u_{t-2} + \theta u_{t-3})(u_t + \theta u_{t-1})] = 0
\end{aligned}
$$

Just-identified solution: Anderson-Hsiao estimator

- First difference the DGP and obtain

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

  where possible values of i remain the same but $t = 2, ..., T$
- Δy_{it-1} can be instrumented with y_{it-2}
  - **Relevancy**: $\Delta y_{it-1} = y_{it-1} - y_{it-2}$.
  - **Exogeneity**: $cov(y_{it-2}, \Delta e_{it}) = E[y_{it-2}, e_{it} - e_{it-1}] = E[y_{it-2}e_{it}] - E[y_{it-2}e_{it-1}] = 0$
- The number of moment conditions equal the number of endogenous variables $\rightarrow$ use MM approach (exact equation in the notes)

Overidentified solution: Arellano-Bond estimator

- In a same first-difference equation, instrument for Δy_{it-1}, with $y_{i0}, \ldots, y_{it-2}$
  - **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

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

    ... and so on. Therefore, we can verify relevancy.
  - **Exogeneity:** Note that $cov(y_{is}, \Delta e_{it})$ for any $s < t - 1$ is 0, as we have shown above. So exogeneity holds as well.
- This is an overidentified case $\rightarrow$ GMM approach (refer to the notes for the estimators)