Random Effects Model and Dynamic Panels

Matteo Paradisi
(EIEF and NBER)

Applied Micro - Lecture 4

Random Effects Models

- Less Problematic than FE: unobserved heterogeneity uncorrelated with observables
- Find consistent and efficient estimates
- Most efficient estimator: simple GLS on the model in levels
- We call it random-effects estimator

Random Effects Models

- Start with the usual model

$$\begin{aligned} y_{it} &= x_{it}\beta + \eta_i + \varepsilon_{it} \\ &= x_{it}\beta + u_{it} \end{aligned}$$

- where $u_{it} = \eta_i + \varepsilon_{it}$
- Assumption:

$$E\left(u_{it} | x_{it}\right) = 0$$

- Unoberved heterogeneity uncorrelated with observables
- Hence, omitting u_{it} does not impede identification with OLS!

Pooled OLS

- The OLS of the model is

$$\hat{\beta}_{POLS} = \left[\sum_{i=1}^{N} X_i' X_i\right]^{-1} \left[\sum_{i=1}^{N} X_i' y_i\right]$$

- This is called a pooled OLS (POLS)
- It is consistent
- This is the same model repeated at each time point
- Is POLS the best estimator for this model?
- No, there is serial correlation. Hence, not efficient!

Serial Correlation in the Model

- The term η_i appears in multiple periods
- Let's make the following assumptions

$$\text{Var}(\varepsilon_{it}) = \sigma_\varepsilon^2 \ \forall i, \forall t$$
$$\text{Cov}(\varepsilon_{it}, \varepsilon_{is}) = 0 \ \ \forall i, \forall t \neq s$$
$$\text{Var}(\eta_i) = \sigma_\eta^2 \ \forall i$$

- Notice that

$$\text{Var}(u_{it}) = \sigma_\eta^2 + \sigma_\varepsilon^2 \ \forall i, \forall t$$
$$\text{Cov}(u_{it}, u_{is}) = \sigma_\eta^2 \ \forall i, \forall t \neq s$$

Serial Correlation in the Model

- In a more compact way

$$\mathsf{Var}\left(\mathsf{u_i}\right) = \mathsf{E}\left(\mathsf{u_i u_i'}\right) = \mathsf{E}\left[\begin{pmatrix} \mathsf{u_{i1}} \\ \mathsf{u_{i2}} \\ \vdots \\ \mathsf{u_{iT}} \end{pmatrix} \begin{pmatrix} \mathsf{u_{i1}} & \mathsf{u_{i2}} & \dots & \mathsf{u_{iT}} \end{pmatrix}\right]$$

$$= \begin{bmatrix} \sigma_\eta^2 + \sigma_\varepsilon^2 & \dots & \sigma_\eta^2 \\ \vdots & \ddots & \vdots \\ \sigma_\eta^2 & \dots & \sigma_\eta^2 + \sigma_\varepsilon^2 \end{bmatrix} = \Omega$$

- We will use Ω in the GLS estimator

A GLS: The Random Effects Estimator

- GLS is the most efficient estimation method
- Let's apply it to the model

$$\hat{\beta}_{RE} = \left[\sum_{i=1}^{N} X_i' \Omega^{-1} X_i\right]^{-1} \left[\sum_{i=1}^{N} X_i' \Omega^{-1} y_i\right]$$

- We need to know Ω to implement the formula
- Ω is a function of σ_η^2 and σ_ε^2
- Hence, we need consistent estimates for these parameters

Balestra-Nerlove Estimator

- Since there are different ways to estimate σ_η^2 and σ_ε^2, there are multiple RE estimators
- One of the most popular is the Balestra-Nerlove
- Estimator for σ_ε^2

$$\sigma_\varepsilon^2 = \frac{1}{N(T-1)} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(y_{it} - x_{it}\hat{\beta}_{FE}\right)^2$$

Estimating σ_η^2

- One solution: estimate ηs and compute variance
- However, computationally intense
- AND when T small, ηs are imprecisely estimated (more on this later)

Estimating σ_η^2

- We need an alternative
- Notice the following

$$\begin{aligned} Var(\bar{u}_i) &= Var\left(\frac{1}{T}\sum_{t=1}^{T} u_{it}\right) = Var\left[\frac{1}{T}\sum_{t=1}^{T}(\eta_i + \varepsilon_{it})\right] \\ &= Var\left(\eta_i + \frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}\right) = \sigma_\eta^2 + \frac{\sigma_\varepsilon^2}{T} \end{aligned}$$

- Hence, to find σ_η^2, we need a consistent estimate of $Var(\bar{u}_i)$
- Once we have it, we invert the formula

$$\hat{\sigma}_\eta^2 = \widehat{Var(\bar{u}_i)} - \frac{\sigma_\varepsilon^2}{T}$$

The Between-Group Model

- $\bar{u}_i$ is the time average of the error term of the model in levels
- can also be interpreted as residual of the between-group model
- model with one observation per individual, equal to time average of all observations

$$\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it} = \left(\frac{1}{T}\sum_{t=1}^{T} x_{it}\right)\beta + \eta_i + \left(\frac{1}{T}\sum_{t=1}^{T} \varepsilon_{it}\right)$$
$$= \bar{x}_i\beta + \bar{u}_i$$

The Between-Group Model

$$\bar{y}_i = \bar{x}_i \beta + \bar{u}_i$$

- The OLS estimator of this model is

$$\hat{\beta}_{BG} = \left[\sum_{i=1}^{N} \bar{x}_i' \bar{x}_i \right]^{-1} \left[\sum_{i=1}^{N} \bar{x}_i' \bar{y}_i \right]$$

- This estimator only exploits variation across individuals
- Does not exploit variation within individual over time

Estimating σ_η^2

- Using $\hat{\beta}_{BG}$, we can estimate $\bar{u}_i$ using

$$\hat{\bar{u}}_i = \bar{y}_i - \bar{x}_i \hat{\beta}_{BG}$$

- And get a consistent estimate of σ_η^2

$$\hat{\sigma}_\eta^2 = \left(\frac{1}{N} \sum_{i=1}^{N} \hat{\bar{u}}_i^2 \right) - \frac{\hat{\sigma}_\varepsilon^2}{T}$$
$$= \left[\frac{1}{N} \sum_{i=1}^{N} \left(\bar{y}_i - \bar{x}_i \hat{\beta}_{BG} \right)^2 \right] - \frac{\hat{\sigma}_\varepsilon^2}{T}$$

- Using $\hat{\sigma}_\eta^2$ and $\hat{\sigma}_\varepsilon^2$, we can estimate Ω and implement the GLS

Random vs Fixed-Effects: the Hausman Test

- Under the assumption H_0 that $E\left(u_{it}|x_{it}\right) = 0$:
  - FE is consistent, but not efficient
  - RE is both consistent and efficient
- Under assumption H_1 : $E\left(\varepsilon_{it}|x_{it}, \eta_i\right) = 0, \; E\left(u_{it}|x_{it}\right) \neq 0$:
  - FE is consistent
  - RE is not cosistent

Random vs Fixed-Effects: the Hausman Test

- Under H_0:

$$\begin{aligned} \hat{\beta}_{\text{FE}} &\to \beta \\ \hat{\beta}_{\text{RE}} &\to \beta \\ \hat{\beta}_{\text{RE}} - \hat{\beta}_{\text{FE}} &\to 0 \end{aligned}$$

- Under H_1:

$$\begin{aligned} \hat{\beta}_{\text{FE}} &\to \beta \\ \hat{\beta}_{\text{RE}} &\to \beta + \text{bias} \\ \hat{\beta}_{\text{RE}} - \hat{\beta}_{\text{FE}} &\to \text{bias} \end{aligned}$$

Random vs Fixed-Effects: the Hausman Test

- We can test the difference between the two estimators to investigate the presence of bias
- If there is no bias, then use RE since it is efficient
- Essentially, we test $H_0 : \text{Cov}(x_{it}, \eta_i) = 0$
- The statistics for the test is a quadratic form of the difference between the two estimators:

$$H = \left(\hat{\beta}_{RE} - \hat{\beta}_{FE}\right)' \left[\text{Var}\left(\hat{\beta}_{RE} - \hat{\beta}_{FE}\right)\right]^{-1} \left(\hat{\beta}_{RE} - \hat{\beta}_{FE}\right)$$

Random vs Fixed-Effects: the Hausman Test

$$H = \left(\hat{\beta}_{RE} - \hat{\beta}_{FE}\right)' \left[\text{Var}\left(\hat{\beta}_{RE} - \hat{\beta}_{FE}\right)\right]^{-1} \left(\hat{\beta}_{RE} - \hat{\beta}_{FE}\right) \sim^{a} \chi^2_K$$

- Since the estimators are asymptotically normally distributed, the H is asymptotically distributed as a χ^2
- Moreover, one can show that

$$\text{Var}\left(\hat{\beta}_{RE} - \hat{\beta}_{FE}\right) = \text{Var}\left(\hat{\beta}_{RE}\right) - \text{Var}\left(\hat{\beta}_{FE}\right)$$

Lagged Dependent Variable Models

Lagged Dependent Variable

- Think about the following finding
  - those who experienced recent drops in income apply more to training
  - known as Ahenfelter's Dip
- How do we model this situation?
- To control for pre-program dip (potential confounder), we must include the lagged dependent variable
- The reason is that y_{t-1} varies over time and not captured by FE

Easy Solution with Strong Assumption

- Assume exogeneity given lagged outcome, but no FE

$$E\left(u_{it} \middle| x_{it}, y_{t-1}\right) = 0$$

- The model would be

$$y_{it} = \alpha + x_{it}\beta + \delta y_{it-1} + u_{it}$$

- Under the exogeneity assumption above, we can identify this
- Two alternatives:
  - OLS: consistent, but not efficient
  - RE: consistent and efficient

More General Case: FE

- Assume exogeneity given lagged outcome AND FE

$$E\left(u_{it} | x_{it}, y_{t-1}, \eta_i\right) = 0$$

- The model would be

$$y_{it} = \alpha + x_{it}\beta + \delta y_{it-1} + \eta_i + \varepsilon_{it}$$

- This complicates the consistent estimation
- Let's see why!

Nickell Bias

- Write the model in first difference

$$\Delta y_{it} = \Delta x_{it}\beta + \delta \Delta y_{it-1} + \Delta \varepsilon_{it}$$

- Since $\Delta \varepsilon_{it} = \varepsilon_{it} - \varepsilon_{it-1}$ and $\Delta y_{it} = y_{it} - y_{it-1}$, there is correlation!
- This bias was first noted by Nickell (1981)
- Let's investigate the model with and without FE and lagged y

Simplified Model to Provide Intuition

- Let's take a very simple model:
  - 2 periods t and $t-1$
  - only one explanatory variable (treatment D_{it})
  - $D_{it-1} = 0$ for all i
- The model is

$$y_{it} = \beta D_{it} + \eta_i + \varepsilon_{it}$$

- Assume that ε_{it} is serially uncorrelated and uncorrelated with η_i and D_{it}
- Also (since $D_{it-1} = 0$):

$$y_{it-1} = \eta_i + \varepsilon_{it-1}$$

- η_i and ε_{it-1} are uncorrelated

Model with Lagged Variable Only

- Suppose we mistakenly estimate

$$y_{it} = \beta D_{it} + \delta y_{it-1} + \tilde{\varepsilon}_{it}$$

- The OLS delivers

$$\frac{Cov\left(y_{it}, \tilde{D}_{it}\right)}{Var\left(\tilde{D}_{it}\right)}$$

- where $\tilde{D}_{it} = D_{it} - \gamma y_{it-1}$ is the residual of regression of D_{it} on y_{it-1}
- Substitute η_i, you get

$$y_{it} = y_{it-1} + \beta D_{it} + \varepsilon_{it} - \varepsilon_{it-1}$$

Model with Lagged Variable Only

$$y_{it} = y_{it-1} + \beta D_{it} + \varepsilon_{it} - \varepsilon_{it-1}$$

- Hence, the estimator delivers

$$\frac{Cov\left(y_{it}, \tilde{D}_{it}\right)}{Var\left(\tilde{D}_{it}\right)} = \beta - \frac{Cov\left(\varepsilon_{it-1}, \tilde{D}_{it}\right)}{Var\left(\tilde{D}_{it}\right)}$$
$$= \beta - \frac{Cov\left(\varepsilon_{it-1}, D_{it} - \gamma y_{it-1}\right)}{Var\left(\tilde{D}_{it}\right)}$$
$$= \beta + \frac{\gamma\sigma_{\varepsilon}^{2}}{Var\left(\tilde{D}_{it}\right)}$$

- If trainees have lower y_{it-1}, $\gamma < 0$ and downward bias

Model with FE Only

- Suppose instead that the true model is

$$y_{it} = \alpha + \beta D_{it} + \theta y_{it-1} + \varepsilon_{it}$$

- with ε serially uncorrelated
- Suppose we ignore y_{it-1} and estimate first difference

$$\frac{Cov\left(y_{it} - y_{it-1}, D_{it} - D_{it-1}\right)}{Var\left(D_{it} - D_{it-1}\right)} = \frac{Cov\left(y_{it} - y_{it-1}, D_{it}\right)}{Var\left(D_{it}\right)}$$

- The model in first differences is

$$y_{it} - y_{it-1} = \alpha + \left(\theta - 1\right) y_{it-1} + \beta D_{it} + \varepsilon_{it}$$

Model with FE Only

$$y_{it} - y_{it-1} = \alpha + (\theta - 1)\, y_{it-1} + \beta D_{it} + \varepsilon_{it}$$

- Our estimate delivers

$$\frac{\text{Cov}\left(y_{it} - y_{it-1}, D_{it}\right)}{\text{Var}\left(D_{it}\right)} = \beta + (\theta - 1)\, \frac{\text{Cov}\left(y_{it-1}, D_{it}\right)}{\text{Var}\left(D_{it}\right)}$$

- Normally, $\theta < 1$ if the process is stationary
- Hence, if trainees have low y_{it-1}, the estimates are too big!

How Do We Solve the Problem?

- One solution is to use instruments for lagged variable
  - e.g. use y_{it-2} as an instrument for y_{it-1}
- Arellano Bond generalize the approach using all lags in GMM
- However, there is a problem with past lags:
  - we must require that y_{it-2} is uncorrelated with $\Delta\varepsilon_{it}$
  - this is unlikely since earnings are highly correlated over time
  - hence, $\Delta\varepsilon_{it}$ is serially correlated and cannot find a good instrument
- Recommendation: check alternative specifications for robustness