Week 14: FE Vs Lagged Dependent Variable

JJ Chen

April 17, 2015

Today's Plan

- Correlated Random Effects Panel Data Models
- FE and Linear Dynamic Model

Recap: RE Model

- We are interested in the individual effect model for Y_{it}

$$\mathrm{E}(\mathrm{Y}_{it}) = \mathrm{X}_{it}'\boldsymbol{\beta} + \alpha_i$$

or

$$\mathrm{Y}_{it} = \mathrm{X}_{it}'\boldsymbol{\beta} + \alpha_i + \varepsilon_{it}.$$

- In the RE model, α_i is assumed to be random, which implies that α_i is uncorrelated with X_{it}
- This assumption allows us to estimate the individual effect model using a GLS estimator because we know the correlation structure of the composite error term $\eta_{it} = \alpha_i + \varepsilon_{it}$
- Rewrite the model in Matrix form: $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\eta}$
  - $\hat{\beta}_{GLS} = (\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{y}$, where $\boldsymbol{\Omega} = \mathrm{V}(\boldsymbol{\eta} \mid \boldsymbol{X})$

Correlated RE Model: Mundlak Projection

- Correlated RE Model let α_i be correlated with X_{it} and still estimate the model through GLS
- The key step is the Mundlak projection:

$$\mathrm{E}(\alpha_i \mid \mathrm{X}_{it}) = \bar{\mathrm{X}}_i' \boldsymbol{\delta}$$

or

$$\alpha_i = \bar{\mathrm{X}}_i' \boldsymbol{\delta} + \upsilon_i$$

- We then back to the RE set up:

$$\begin{aligned}
\mathrm{Y}_{it} &= \mathrm{X}_{it}' \boldsymbol{\beta} + \alpha_i + \varepsilon_{it} \\
&= \mathrm{X}_{it}' \boldsymbol{\beta} + \bar{\mathrm{X}}_i' \boldsymbol{\delta} + \upsilon_i + \varepsilon_{it} \\
&= (\mathrm{X}_{it} - \bar{\mathrm{X}})' \boldsymbol{\beta} + \bar{\mathrm{X}}_i' (\boldsymbol{\beta} + \boldsymbol{\delta}) + \upsilon_i + \varepsilon_{it} \\
&= \mathrm{Z}_{it}' \boldsymbol{\gamma} + \upsilon_i + \varepsilon_{it}
\end{aligned}$$

Correlated RE Model: Chamberlain Project

- We can also use a more flexible specification for α_i:

$$\alpha_i = \sum_{t=1}^{T} \beta_t \mathrm{X}_{it} + \upsilon_i$$

Recap: FE Regression

- In reality, when we run a serious FE regression

$$\mathrm{Y}_{it} = \alpha + \rho \mathrm{D}_{it} + \mathrm{A}_i' \boldsymbol{\gamma} + \mathrm{I}_t' \boldsymbol{\delta} + \mathrm{X}_{it}' \boldsymbol{\beta} + e_{it},$$

we're trying to get a naive conditional mean comparisons:

$$\mathrm{E}(\mathrm{Y}_{it} \mid \mathrm{A}_i, \mathrm{I}_t, \mathrm{X}_{it}, \mathrm{D}_{it} = 1) - \mathrm{E}(\mathrm{Y}_{it} \mid \mathrm{A}_i, \mathrm{I}_t, \mathrm{X}_{it}, \mathrm{D}_{it} = 0),$$

and we really want to measure is a causal relationship between Y_{it} and D_{it},

$$\mathrm{E}(\mathrm{Y}_{1it} \mid \mathrm{A}_i, \mathrm{I}_t, \mathrm{X}_{it}, \mathrm{D}_{it} = 1) - \mathrm{E}(\mathrm{Y}_{0it} \mid \mathrm{A}_i, \mathrm{I}_t, \mathrm{X}_{it}, \mathrm{D}_{it} = 1).$$

Recap: FE Model Assumptions I

- Let's first review the ID assumptions for FE models
- CIA for Y_{0it}: $\mathrm{Y}_{0it} \perp\!\!\!\perp \mathrm{D}_{it} \mid \mathrm{A}_i, \mathrm{I}_t, \mathrm{X}_{it}$
  - Is this strict exogeneity?
- CIA implies Mean Independence Assumption for Y_{0it}:

$$\mathrm{E}(\mathrm{Y}_{0it} \mid \mathrm{A}_i, \mathrm{I}_t, \mathrm{X}_{it}, \mathrm{D}_{it}) = \mathrm{E}(\mathrm{Y}_{0it} \mid \mathrm{A}_i, \mathrm{I}_t, \mathrm{X}_{it}),$$

or to write it slightly differently,

$$\mathrm{E}(\mathrm{Y}_{0it} \mid \mathrm{A}_i, \mathrm{I}_t, \mathrm{X}_{it}, \mathrm{D}_{it} = 1) = \mathrm{E}(\mathrm{Y}_{0it} \mid \mathrm{A}_i, \mathrm{I}_t, \mathrm{X}_{it}, \mathrm{D}_{it} = 0).$$

- Time-varing treatment, such as union status, is as good as ramdomly assigned conditional on the covariates.

Recap: FE Model Assumption II

- To esimate a FE model, another key assumption is the linear additive structure for the potential outcome Y_{0it}:

$$\text{E}(\text{Y}_{0it} \mid \text{A}_i, \text{I}_t, \text{X}_{it}) = \alpha + \text{A}_i'\gamma + \text{I}_t'\boldsymbol{\delta} + \text{X}_{it}'\boldsymbol{\beta},$$

or to write it slightly differently

$$\begin{aligned} \text{E}(\text{Y}_{0it} \mid i, t, \text{X}_{it}) &= \alpha_i + \lambda_t + \text{X}_{it}'\boldsymbol{\beta} \\ \Longrightarrow \text{Y}_{0it} &= \text{X}_{it}'\boldsymbol{\beta} + \alpha_i + \lambda_t + [\text{Y}_{0it} - \text{E}(\text{Y}_{0it})] \\ \Longrightarrow \text{Y}_{0it} &= \text{X}_{it}'\boldsymbol{\beta} + \alpha_i + \lambda_t + \varepsilon_{it} \end{aligned}$$

- Why is the linear structure important for estimation?

Recap: FE Model Assumption III

- The last assumption is the additive homogenenous treatment effect:

$$\mathrm{E}(\mathrm{Y}_{1it} \mid i, t, \mathrm{X}_{it}, \mathrm{D}_{it} = 1) = \mathrm{E}(\mathrm{Y}_{0it} \mid i, t, \mathrm{X}_{it}, \mathrm{D}_{it} = 0) + \rho$$

or

$$\rho = \mathrm{E}(\mathrm{Y}_{1it} \mid i, t, \mathrm{X}_{it}, \mathrm{D}_{it} = 1) - \mathrm{E}(\mathrm{Y}_{0it} \mid i, t, \mathrm{X}_{it}, \mathrm{D}_{it} = 0)$$

Recap: FE Models

- The three assumptions implies the CEF of interest:

$$\mathrm{E}(\mathrm{Y}_{it} \mid i, t, \mathrm{X}_{it}, \mathrm{D}_{it}) = \alpha_i + \lambda_t + \mathrm{X}_{it}'\boldsymbol{\beta} + \rho \mathrm{D}_{it}.$$

- By Regression Justification III, we can run

$$\mathrm{Y}_{it} = \alpha_i + \lambda_t + \mathrm{X}_{it}'\boldsymbol{\beta} + \rho \mathrm{D}_{it} + e_{it},$$

or the equivalent LSDV regression

$$\mathrm{Y}_{it} = \alpha + \rho \mathrm{D}_{it} + \mathrm{A}_i'\boldsymbol{\gamma} + \mathrm{I}_t'\boldsymbol{\delta} + \mathrm{X}_{it}'\boldsymbol{\beta} + e_{it}.$$

Motivation for Lagged Models

- The Key ID assumption is that

$$\mathrm{Y}_{0it} \perp\!\!\!\perp \mathrm{D}_{it} \mid \mathrm{A}_i, \mathrm{I}_t, \mathrm{X}_{it}$$

  - which is equivalent to say that A_i, I_t, and X_{it} are important omitted variables if we fail to control for them in a regression
  - we can also connect to the parallel trend assumption, why?
- The Ashenfelter's dip we discuss in class suggests that the above CIA might not be plausible. Why?
- Other similar examples: remedial education program, school turnaround grants, educational tracking

ID Assumption for Lagged Models

- If we worry about sorting based on previous outcomes, which is common, we might want to consider another ID assumption:

$$\mathrm{Y}_{0it} \perp\!\!\!\perp \mathrm{D}_{it} \,|\mathrm{Y}_{i,t-h}, \mathrm{I}_t, \mathrm{X}_{it}.$$

  - In this case, we think the important omitted variable is pre-treatment trends
- The above ID assumption implies regression

$$\mathrm{Y}_{it} = \alpha + \theta \mathrm{Y}_{i,t-h} + \lambda_t + \rho \mathrm{D}_{it} + \mathrm{X}'_{it}\boldsymbol{\beta} + e_{it}.$$

- But what if you are not sure about whether there is ability bias or sorting based on previous outcomes?

Solution 1: Assume Both (Dynamic Panel Data Methods)

- If we assume both, then we are think the following ID assumptions:

$$\mathrm{Y}_{0it} \perp\!\!\!\perp \mathrm{D}_{it} \,|\mathrm{Y}_{i,t-h}, \mathrm{A}_i \mathrm{I}_t, \mathrm{X}_{it},$$

which implies a fixed effect dynamic regression

$$\mathrm{Y}_{it} = \theta \mathrm{Y}_{i,t-h} + \alpha_i + \lambda_t + \rho \mathrm{D}_{it} + \mathrm{X}'_{it}\boldsymbol{\beta} + e_{it}.$$
$$\implies \Delta \mathrm{Y}_{it} = \theta \Delta \mathrm{Y}_{i,t-h} + \Delta\lambda_t + \rho \Delta \mathrm{D}_{it} + \Delta \mathrm{X}'_{it}\boldsymbol{\beta} + \Delta e_{it}.$$

- But we can't estimate ρ consistently because of the Nickell bias
  - $\mathrm{Cov}(\Delta \mathrm{Y}_{i,t-h}, \Delta e_{it}) \neq 0$, HW6 from last semester

Solution 1: Assume Both (Dynamic Panel Data Methods)

- To estimate the difference model consistently in large sample, Anderson and Hsiao suggest using lagged outcomes as IVs
  - Stata command: `xtivreg, fd`
- Arellano and Bond later suggest using GMM so that one can exploit all information in the IVs and get a more effecicient esitimates
  - Stata command: `ssc install xtabond2`
- In class we know that the ER assumption for the IVs might not hold (serial correlation for the error process), and also weak IV problem
  - We will learn it again in Prof. Casey's class

Solution 2: Assume Either One

- MHE also suggests estimating both the FE model and lagged model and see if the result is sensitive
- FE and lagged model have a bracketing property
  - If FE is the true model, lagged regression estimate gives a lower bound
  - If lagged model is the true model, FE estimate gives a upper bound