Lecture 5: Advanced Panel Data

Chris Conlon

NYU Stern

March 8, 2019

Causal FE

Recall the FE Assumptions

$$y_{it} = x_{it}'\beta + \eta_i + \varepsilon_{it}$$

- η_i is a fixed effect.
- To estimate everything consistently, we need $E[\varepsilon_{it}|x_{it}, \eta_i] = 0$
- Mostly this is not true. Instead usually treat η_i as a control variable or nuisance parameter.
  - A nuisance parameter is one that we estimate but don't care about interpreting.
  - If we only care about β then η_i is a nuisance parameter.
- With a control or nuisance parameter we only require that $E[\varepsilon_{it}|\eta_i] = E[\varepsilon_{it}|x_{it}, \eta_i]$ conditional mean independence.

Recall the FE Assumptions

$$y_{it} = x_{it}'\beta + \eta_i + \varepsilon_{it}$$

- η_i is a fixed effect.
- To estimate everything consistently, we need $E[\varepsilon_{it}|x_{it}, \eta_i] = 0$
- Mostly this is not true. Instead usually treat η_i as a control variable or nuisance parameter.
  - A nuisance parameter is one that we estimate but don't care about interpreting.
  - If we only care about β then η_i is a nuisance parameter.
- With a control or nuisance parameter we only require that $E[\varepsilon_{it}|\eta_i] = E[\varepsilon_{it}|x_{it}, \eta_i]$ conditional mean independence.
- Once we condition on η_i it is as if ε_{it} and x_{it} are uncorrelated.

Causal FE

- We can get away with conditional mean independence if we don't care about η_i.
- But suppose that we care about $\widehat{\eta_i}$?
  - Teacher FE
  - Physician/Hospital FE
  - Location/County FE
  - Suppose we take someone from the 10*th* percentile and move them to the 90*th* percentile

Causal FE

- Now we have to really believe $E[\varepsilon_{it}|x_{it}, \eta_i] = 0$
- We should worry about the conventional omitted variable bias problem.
- Suppose there exists a variable w_{it} so that:

$$y_{it} = x_{it}'\beta + w_{it}'\gamma + \eta_i + \varepsilon_{it}$$

- Recall the conditions for OVB
  - w_{it} is correlated with x_{it}
  - w_{it} is a determinant of y_{it}
- New one: w_{it} is correlated with η_i
  - This is easy to satisfy!
  - w_{it} needs to be uncorrelated with anything about the individual i.

Example: Test Scores

- Students s, Teachers t
- Want to measure effect of Teachers on Test Scores

$$TestScore_{st} = \beta x_s + \gamma w_t + \eta_t + \varepsilon_{st}$$

- We observe some features of students but not all of them (parent's education, household income, language spoken at home).
- We also observe some school specific variables w_t but not all of them (district spending per pupil, % free lunch, etc.).
- But we don't observe other things (jackhammering outside the classroom, which students have disruptive home lives,etc.).
    - If the mean of those things varies across teachers $\rightarrow$ we are screwed!
    - Can't get an accurate estimate of η_i.

Example: Test Scores

We need a better design:

- We probably need random assignment of students to teachers.
- Ideally we would be able to control for student and school unobservables.
- Might want to see many students match with many teachers.

Dynamic Panel Data

Dynamic Panel

- Suppose that we also want to include a lagged $y_{i,t-1}$

$$y_{it} = \rho y_{i,t-1} + x_{it}'\beta + \eta_i + \varepsilon_{it}$$

- We can treat η_i as a random effect or a fixed effect.

Dynamic Panel: Nickell (1981) Bias

Consider the within transform

$$(y_{it} - \overline{y}_i) = \rho(y_{i,t-1} - \overline{y}_i) + (x_{it} - \overline{x}_i)'\beta + (\varepsilon_{it} - \overline{\varepsilon}_i)$$

- This eliminates the fixed effect.
- But $Cov(y_{i,t-1} - \overline{y}_i, \varepsilon_{it} - \overline{\varepsilon}_i) \neq 0$. Why?
  - Both contain past and future values
  - There is a direct relationship between y and ε
  - Bias does not disappear as $N \to \infty$ (it does as $T \to \infty$).
  - For small T, dynamic panel model is inconsistent.

Dynamic Panel: Bias Alternative

$$y_{it} = \rho y_{i,t-1} + x_{it}'\beta + \eta_i + \varepsilon_{it}$$

- We require the following assumption (strict exogeneity):

$$E\left(\varepsilon_{it} | x_{i1}, \ldots, x_{iT}, \eta_i\right) = 0, \quad t = 1, \ldots, T$$

- But what about y_{it-1}?
  - It is correlated with $\varepsilon_{i,t-1}$ and η_i (by construction).
  - With serial correlation it is correlated with ε_{it}
  - This is the usual endogeneity concern.

Dynamic Panel: Differenced Model (Anderson-Hsiao)

How do we deal with endogeneity? With instruments!

$$y_{it} = \rho y_{i,t-1} + x_{it}'\beta + \eta_i + \varepsilon_{it}$$

Consider the first differences (s is a dummy time index):

$$E\left[x_{is}\left(\Delta y_{it} - \rho\Delta y_{i(t-1)} - \Delta x_{it}'\beta\right)\right] = 0$$

Idea:

- Under strict exogeneity of x_{it} we can use both lags and leads as instruments for $y_{i,t-1}$
- Excluded Instruments $x_{i,s}$ do not have a direct effect on $\Delta y_{i,t-1}$.
- These moments work even in presence of serially correlated errors.

Minimal Example: Anderson-Hsiao

Imagine we have only $T = 3$ periods:

$$y_3 - y_2 = \alpha\left(y_2 - y_1\right) + \beta_0\left(x_3 - x_2\right) + \beta_1\left(x_2 - x_1\right) + \left(\varepsilon_3 - \varepsilon_2\right)$$

- $E\left(x_{is}\Delta\varepsilon_{i3}\right) = 0$ has three instruments (x_{i1}, x_{i2}, x_{i3}).
- The model is just identified with 3 parameters $(\alpha, \beta_0, \beta_1)$.
- The challenge with this approach is often that it suffers from weak instruments.

Study annual cigarette consumption with state-level data:

$$c_{it} = \theta c_{i,t-1} + \beta\theta c_{i,t-1} + \gamma p_{it} + \eta_i + \delta_t + v_{it}$$

A model of (forward looking) rational addiction:

- c_{it} = Annual per capita cigarette consumption in packs by state.
- p_{it} = Average cigarette price per pack.
- θ = Measure of the extent of addiction (for $\theta > 0$).
- β = Discount factor.
- Derived from forward looking model of habit formation FOC's.

Becker, Grossman, Murphy (1994)

$$c_{it} = \theta c_{i,t-1} + \beta\theta c_{i,t-1} + \gamma p_{it} + \eta_i + \delta_t + v_{it}$$

- Marginal utility of wealth can show up in γ or η_i.
- The errors v_{it} are unobserved life-cycle utility shifters, can be autocorrelated.
- Absent addiction $\theta = 0$ and serial correlation in prices, we would expect to find dependence over time in c_{it}.
- Conditional on $c_{i,t}|(c_{i,t-1}, c_{i,t+1})$ does not depend on $p_{i,t+1}$ or $p_{i,t+1}$.

Becker, Grossman, Murphy (1994)

$$c_{it} = \theta c_{i,t-1} + \beta\theta c_{i,t-1} + \gamma p_{it} + \eta_i + \delta_t + v_{it}$$

- Identify (θ, β, γ) from the assumption that prices are strictly exogenous
- Use lagged and future $p_{i,t+s}$ and $p_{i,t-s}$ as IV.
- Use the change in cigarette taxes.
- Consumers need to fully anticipate future price changes for this to work.

Becker, Grossman, Murphy (1994):Table

Becker, Grossman, Murphy (1994):Table

Dynamic Panel: Arellano Bond

The main idea is that the instruments come from within the model!

$$y_{it} = \rho y_{i,t-1} + x_{it}'\beta + \eta_i + \varepsilon_{it}$$

Consider the first differences (s is a dummy time index):

$$E\left[x_{is}\left(\Delta y_{it} - \rho \Delta y_{i(t-1)} - \Delta x_{it}'\beta\right)\right] = 0$$

Idea:

- Now relax strict exogeneity.
- Can still use x_{is} as contemporaneous exogenous instrument.
- What is an excluded instrument for $\Delta y_{i,t-1}$?
  - Needs to be relevant
  - Still needs to be exogenous: not a direct determinant

$$E\left[x_{is}\left(\Delta y_{it} - \rho \Delta y_{i(t-1)} - \Delta x_{it}' \beta\right)\right] = 0$$

Idea: Use higher lags of y_{it}:

- $[t = 2]$ or $[t = 1]$: no instruments,
- $[t = 3]$: valid instrument for $\Delta y_{i2} = (y_{i2} - y_{i1})$ is y_{i1}.
- $[t = 4]$: valid instruments for $\Delta y_{i3} = (y_{i3} - y_{i2})$ is (y_{i2}, y_{i1})
- $[t = 5]$: valid instruments for $\Delta y_{i5} = (y_{i5} - y_{i4})$ is $(y_{i1}, \ldots, y_{i4})$.
- $[t = T]$: valid instruments for $\Delta y_{iT} = (y_{iT} - y_{i,T-1})$ is $(y_{i1}, \ldots, y_{i,T-1})$.

Thus there are $T/(T-1)/2$ instruments

Dynamic Panel: Arellano Bond

$$E\left[\mathbf{y}_{is}\left(\Delta y_{it} - \rho \Delta y_{i(t-1)} - \Delta x_{it}'\beta\right)\right] = 0$$
$$E\left[\Delta x_{it}\left(\Delta y_{it} - \rho \Delta y_{i(t-1)} - \Delta x_{it}'\beta\right)\right] = 0$$

- $\mathbf{y}_{is} = [y_{i1}, \ldots, y_{i,t-2}]$ for $t > 2$.
- Levels instrument Differences
- Thus there are $T/(T-1)/2$ instruments
- We can estimate with linear IV GMM: `pgmm` or `dynpanel`.
- The common complain is that instruments are still weak.

More Moments: Blundell and Bond

$$E\left[\mathbf{y}_{is}\left(\Delta y_{it} - \rho\Delta y_{i(t-1)} - \Delta x_{it}'\beta\right)\right] = 0$$
$$E\left[\Delta y_{i,t-1}\left(y_{it} - \rho y_{i(t-1)} - x_{it}'\beta\eta_i\right)\right] = 0$$
$$E\left[\Delta x_{it}\left(\Delta y_{it} - \rho\Delta y_{i(t-1)} - \Delta x_{it}'\beta\right)\right] = 0$$

- Differences also instrument Levels!
- Important when $\rho \to 1$ or when σ_u/σ_ϵ becomes large.
- These can also pin down y_{i0}, etc.
- This is known as GMM-SYS.