Lecture 22: Longer Panels

Pierre Biscaye

Fall 2022

Agenda

1. 2 period panel review
2. First differences estimation
3. Strict exogeneity
4. Fixed effects vs. first differences

2 period panel review

- Last time we focused on panel data with 2 periods, where the structural model looks something like

$$y_{it} = \beta_0 + \beta_1 x_{1it} + ... + \beta_k x_{kit} + \alpha_i + \delta_t + u_{it} \tag{1}$$

- The α_i term would be a source of OVB if we didn't have access to panel data, but we presented two approaches to deal with it.
- Unit *fixed effects* control directly for each cross-sectional unit. Implement by including α_i as a vector of dummy variables.
- *First differences* subtract values within units from adjacent (in time) observations, differencing away the α_i.

$$y_{it} = \beta_0 + \beta_1 x_{1it} + ... + \beta_k x_{kit} + \alpha_i + \delta_t + u_{it} \quad FE \tag{2}$$

$$\Delta y = \delta_0 + \beta_1 \Delta x_{1i} + ... + \beta_k \Delta x_{ki} + \Delta u_i \quad FD \tag{3}$$

- Both approaches control for time-invariant characteristics of the cross-sectional units, removing a potentially important source of OVB.

MLR implications of using panel data methods

- For MLR3, need all x variables to vary over time, as well as across units.
  - Otherwise, the x variable is colinear with the fixed effects/dropped by the first difference.
  - If x varies for only *some* units, then the effect of x on y is only identified for those units. Need to think critically about whether the population for which x varies is differs from the overall population.
- With FD, MLR4: $E[\Delta u_i|\Delta x_i] = 0$.
- With FE, MLR4: $E[u_{it}|x_{it}, \alpha_i, \delta_t] = 0$.
- Interpretation is we need there to be no omitted variables whose *changes* are correlated with *changes* in x and *changes* in y.
- If there is a variable in u that changes over time in a manner correlated with changes in some x, that bias will *not* be addressed with panel data methods.

What if we have more than $T > 2$ time periods?

$$
\begin{aligned}
y_{i1} &= \beta_0 + \beta_1 x_{1i1} + ... + \beta_k x_{ki1} + \alpha_i + u_{i1} \\
y_{i2} &= \beta_0 + \beta_1 x_{1i2} + ... + \beta_k x_{ki2} + \alpha_i + \delta_2 + u_{i2} \\
y_{i3} &= \beta_0 + \beta_1 x_{1i3} + ... + \beta_k x_{ki3} + \alpha_i + \delta_3 + u_{i3} \\
&... \\
y_{iT} &= \beta_0 + \beta_1 x_{1iT} + ... + \beta_k x_{kiT} + \alpha_i + \delta_T + u_{iT}
\end{aligned}
$$

We can still first difference to eliminate α_i

$$
\begin{aligned}
y_{i2} - y_{i1} = \delta_2 + \beta_1(x_{1i2} - x_{1i1}) + \dots \\
+\beta_k(x_{ki2} - x_{ki1}) + u_{i2} - u_{i1} \qquad (4) \\
y_{i3} - y_{i2} = \delta_3 - \delta_2 + \beta_1(x_{1i3} - x_{1i2}) + \dots \\
+\beta_k(x_{ki3} - x_{ki2}) + u_{i3} - u_{i2} \qquad (5)
\end{aligned}
$$

- The intercept varies in each equation/year, so when stacking these and writing it in first differences notation, will need to include year dummy variables.

$$\Delta y_{it} = \delta_0 + \delta_2 d_2 + \delta_3 d_3 + \dots \qquad (6)$$

$$+\beta_1 \Delta x_{1it} + \dots + \beta_k \Delta x_{kit} + u_{it} \qquad (7)$$

- To Jupyter!

We can still first difference to eliminate α_i

$$
\begin{aligned}
y_{i2} - y_{i1} = \delta_2 + \beta_1(x_{1i2} - x_{1i1}) + ... \\
+\beta_k(x_{ki2} - x_{ki1}) + u_{i2} - u_{i1} \qquad (4) \\
y_{i3} - y_{i2} = \delta_3 - \delta_2 + \beta_1(x_{1i3} - x_{1i2}) + ... \\
+\beta_k(x_{ki3} - x_{ki2}) + u_{i3} - u_{i2} \qquad (5)
\end{aligned}
$$

- The intercept varies in each equation/year, so when stacking these and writing it in first differences notation, will need to include year dummy variables.

$$\Delta y_{it} = \delta_0 + \delta_2 d_2 + \delta_3 d_3 + ... \qquad (6)$$

$$+\beta_1 \Delta x_{1it} + ... + \beta_k \Delta x_{kit} + u_{it} \qquad (7)$$

- To Jupyter!
- Challenge: u_{i2} appears on both 4 and 5.
- What does this imply for MLR4?

New MLR4: *Strict Exogeneity*

$$\text{Strict Exogeneity} : cov(x_{jit}, u_{is}) = 0 \, \text{for all} \, t, s, \text{and} j$$

- In other words, the unobserved error term in every period s is uncorrelated with all k of your x variables in every period t.
- So, your error today must be uncorrelated with your x yesterday, today, and tomorrow.
  - Can see why this is needed from the example above, where u_{i2} appears in equations with $x_{1i1}, x_{1i2}, x_{1i3}$ because of the differencing.
- This MLR4 is the same for FD and FE with more than 2 time periods.

What violates strict exogeneity?

- We've talked through issues where changes in x's relate to changes in u's.
- Those are still a problem that can violate MLR4, but this is a bit different.
- We also need to worry about x variables that relate to u's in different time period.
- One example: lagged dependent variables $y_{i,t-1}$.

Lagged Dependent Variables (LDVs)

- Panel data creates a lot of possibilities in terms of control variables.
- Lagged dependent variables $y_{i,t-1}$ can be attractive.
  - Idea: if we control for last year's y, then we hold constant whatever influenced our y last year.
  - If omitted variables are similar from year to year, this could take care of a lot of OVB.
- For example, suppose that we wanted to test whether the number of police officers influences crime.
- We suppose that there are important unobservables associated with policing and crime.
- One thing that might seem attractive is to hold last year's crime rate constant. The remaining variation in current policing is anything not explained by last year's crime rate.
- This can be great... but not in combination with FD or FE.

Lagged Dependent Variables with First Differences: algebraically

$$y_{it} = \gamma y_{it-1} + \beta_0 + \beta_1 x_{it} + \alpha_i + \delta_t + u_{it} \tag{8}$$

$$\Delta y_{it} = \gamma \Delta y_{it-1} + \beta_1 \Delta x_{it} + \Delta\delta_t + \Delta u_{it} \tag{9}$$

$$\Delta y_{it} = \gamma(y_{it-1} - y_{it-2}) + \beta_1 \Delta x_{it} + \Delta\delta_t + u_{it} - u_{it-1} \tag{10}$$

$$y_{it-1} = \beta_0 + \gamma y_{it-2} + \beta_1 x_{it-1} + \delta_{t-1} + u_{it-1} \tag{11}$$

$$\begin{aligned} \Delta y_{it} = \gamma((\gamma - 1) y_{it-2} + \beta_1 \Delta x_{it-1} + \delta_{t-1} + u_{it-1}) \\ + \beta_1 \Delta x_{it} + \Delta\delta_t + u_{it} - u_{it-1} \end{aligned} \tag{12}$$

Lagged Dependent Variables and strict exogeneity

- The idea behind LDVs is basically the same as the idea behind first differences: you want to hold constant something about the current level of the dependent variable.
- There is a concern then about doing both at the same time.
- The algebraic example above shows that the LDV term in a FD specification will be correlated with the error term.
- Violations of strict exogeneity often stem from the mistake of trying to control for current unobservables twice.
- While LDVs are useful in some settings, do not want to include them in FD or FE estimation.
  - One potentially useful setting: effect of education on wages.

Taking differences with $T > 2$ panel data

- There's nothing unique about the *first* difference.
- *Any* difference would eliminate α_i.
- We could subtract off 2 periods ago,3 periods ago, ...
- One attractive idea: difference off the mean.

$$y_{it} = \beta_0 + \beta_1 x_{1it} + ... + \beta_k x_{kit} + \alpha_i + \delta_t + u_{it} \tag{13}$$

$$\bar{y_i} = \beta_0 + \beta_1 \bar{x_{1i}} + ... + \beta_k \bar{x_{ki}} + \bar{\delta} + \alpha_i \tag{14}$$

$$y_{it} - \bar{y_i} = \beta_1 (x_{1it} - \bar{x_{1i}}) + ... + \beta_k (x_{kit} - \bar{x_{ki}}) + \delta_t - \bar{\delta} + u_{it} \tag{15}$$

- This is called the *within* transformation or *fixed effects*.

Fixed effects

- Why does differencing the mean equate to include unit fixed effects?

$$y_{it} - \bar{y_i} = \beta_1(x_{1it} - \bar{x_{1i}}) + ... + \beta_k(x_{kit} - \bar{x_{ki}}) + \delta_t - \bar{\delta} + u_{it} \quad (16)$$

$$y_{it} = \beta_0 + \beta_1 x_{1it} + ... + \beta_k x_{kit} - (\beta_1 \bar{x}_{1i} + ... + \beta_k \bar{x}_{ki} + \bar{\delta}) + \delta_t + u_{it} \quad (17)$$

$$y_{it} = \beta_0 + \beta_1 x_{1it} + ... + \beta_k x_{kit} + \sum_i \alpha_i c_i + \delta_t + u_{it} \quad (18)$$

- These are equivalent: the means of all variables within units are fixed within those units. Therefore, including unit fixed effects dummies controls for all those means.
- When $T = 2$, fixed effects and first differences are identical. We saw this last lecture.
  - Think what it means to subtract the mean when $T = 2$.
- When $T > 2$, they are a little different.
- To Jupyter!

Fixed effects vs. first differences

- Both eliminate anything about a cross-sectional unit that doesn't change.
- Both require the same assumption of strict exogeneity:. $cov(x_{jit}, u_{is}) = 0$
- Our big concerns will remain that trends in x might be related to trends in u
- When we do fixed effects, we can recover α_i estimates.
  - This is a categorical variable: interpretations?
- Which is the better estimator?
  - Both are unbiased if MLR4 holds.
  - Which has lower variance? Depends on u.

u_{it} in panel data

- We've set aside an issue: is u_{it} distributed independently over time?

$$\Delta y_{it} = \delta_t + \beta_1 \Delta x_{1it} + ... + \beta_k \Delta x_{kit} + \Delta u_{it} \tag{19}$$

$$\Delta y_{i2} = \delta_2 + \beta_1 \Delta x_{1i2} + ... + \beta_k \Delta x_{ki2} + u_{i2} - u_{i1} \tag{20}$$

$$\Delta y_{i3} = \delta_3 + \beta_1 \Delta x_{1i3} + ... + \beta_k \Delta x_{ki3} + u_{i3} - u_{i2} \tag{21}$$

$$cov(u_{i3} - u_{i2}, u_{i2} - u_{i1}) \neq 0 \tag{22}$$

- Even if u_{it} are uncorrelated over time (which may be unlikely), the residuals in first difference estimation are correlated over time.

Our panel errors *will* be correlated over time

- This is referred to as *serial* correlation.
- What does this mean?
- We no longer have a random sample, even if cross-sectional i's are randomly sampled, because the observations over time are not independent.
- $\implies$ we have fewer *actual* observations than we appear to, in terms of the amount of independent information contributes to the estimation.
    - This matters for inference, since calculation of SEs depends on number of observations.
- We can correct for this, but intuitively when observations are correlated there is less we can learn from them.

Fixed Effects has the same serial correlation issue

$$y_{it} - \bar{y}_i = \delta_t - \bar{\delta} + \beta_1(x_{1It} - \bar{x_{1i}}) + ... + \beta_k(x_{kit} - \bar{x_{ki}}) + u_{it} - \bar{u_i} \quad (23)$$

$$y_{it} - \frac{1}{T}\sum_{t'} y_{it'} = \delta_t - \frac{1}{T}\sum_{t'} \delta'_t + ... + u_{it} - \frac{1}{T}\sum_{t'} u_{it'} \quad (24)$$

$$cov(u_{i3} - \frac{1}{T}\sum_{t'} u_{it'}, u_{i2} - \frac{1}{T}\sum_{t'} u_{it'}) \neq 0 \quad (25)$$

- It is not possible to say which correlation is bigger, or which variance is smaller between FD and FE.
- In practice, it depends on how u_{it} is correlated over time
  - If u_{it} is similar to u_{it-1} ($u_{it} = u_{it-1} + \epsilon_{it}$), then first differences may be lower variance.
  - If u_{it} are close to random ($u_{it} = \epsilon_{it}$) then fixed effects is lower variance.

Errors in panel data

- Correlation between errors (serial correlation) means that we actually have fewer independent observations than it looks like we have.
- There is solution that is easy to implement in R: *clustered errors*.
- Clustering specifies a level of cluster and calculates standard errors only assuming different clusters are independent observations.
  - For example, in the county-year crime data, clustering at the county level will treat all observations for each county as a *single* independent observation for the purpose of SE calculation.
- This will tend to inflate your SEs (the count of observations is in the denominator): results are less precise.
  - Implications for inference.

Clustered errors in panel data

- Clustering specifies a level of cluster and calculates standard errors only assuming different clusters are independent observations.
- Usually with clustered errors, adding more time periods will reduce standard errors by less than adding more cross-sectional units.
  - Makes sense if all observations within an cluster are treated as one independent observation. What you need to do is increase the number of clusters.
- It is pretty much always a good idea to cluster errors at the level of the cross-sectional unit.
  - In some cases we might cluster at a higher level (e.g., state): if we think there is also a lot of correlation across units within that higher level of aggregation.
- To Jupyter!

Balanced and unbalanced panels

- In a balanced panel, each cross-sectional unit is observed in each time period.
- In our crime example, we had a balanced panel: we saw each county in each year.
- What if we have some missing data?

First differences struggle with unbalanced panels

Consider this case:

$$\begin{bmatrix} y_{i1} & x_{1i1} & x_{2i1} & x_{3i1} \\ y_{i2} & . & x_{2i2} & x_{3i2} \\ y_{i3} & x_{1i3} & x_{2i3} & x_{3i3} \end{bmatrix}$$

We cannot estimate any first difference regression that includes x_1 because it is missing in period two.

Fixed Effects work better

- Lose observations only if they are missing data for an included variable.
  - If have a unit with valid data in only one time period, it will be dropped entirely since it's no longer a "panel" observation.
- Should we worry about the unbalanced panel?
- Maybe...
- Why is the observation missing?
- If missing due to changes in x variables of interest, could be a concern
-

Example: impact of tariff on businesses

- Suppose we have data on firms and want to evaluate the effect of a steep tariff.
- What if the tariff drives firm 2 out of business?
  - Refer to units dropping out of a panel over time as *attrition*.

$$\begin{bmatrix} \textit{profits} & \textit{tariff} \\ y_{11} & 0 \\ y_{12} & 1 \\ y_{21} & 0 \\ . & 1 \\ y_{31} & 0 \\ y_{32} & 0 \end{bmatrix}$$

- Firm two gets dropped from the sample - MLR2 no longer holds.
- Don't observe what would be an extreme response to the tariff.

Fixed effects can help

- Fixed Effects lose fewer observations than first differences.
  - And hold constant anything about a cross-sectional unit that doesn't change over time - including, maybe, its likelihood of attrition.
- Our coefficients will be biased if attrition is related to changes in our x variables.
- We can often check this by evaluating attrition as an outcome and testing if any characteristics are correlated with it.