BUSS975 Causal Inference in Financial Research

Panel Data

Professor Ji-Woong Chung
Korea University

Outline

Motivate how panel data is helpful

Fixed effects model

- Benefits [There are many]
- Costs [There are some...]

Random effects model

First differences

Lagged y models

Outline

Motivate how panel data is helpful

Fixed effects model

- Benefits [There are many]
- Costs [There are some...]

Random effects model

First differences

Lagged y models

Motivation (Part 1)

- Omitted variables pose a substantial hurdle in our ability to make causal inferences.
- What's worse ... many of these variables are inherently unobservable to researchers.

Motivation (Part 2)

Example: Firm-level estimation where leverage is debt/assets for firm i, operating in industry j in year t, and profit is net income/assets.

$$leverage_{i,j,t} = \beta_0 + \beta_1 profit_{i,j,t} + u_{i,j,t}$$

What might be some unobservable omitted variables in this estimation?

Motivation (Part 3)

- Possible unobserved variables include:
  - Managerial talent and/or risk aversion
  - Industry supply/demand shocks
  - Cost of capital
  - Investment opportunities
  - and so on...
- Easy to think of ways these might be affect leverage and be correlated with profits

Motivation (Part 4)

- Using observations from various geographical regions (e.g., state or country) opens even more possibilities.
- What unobserved variables might be related to a firm's location?
- Example: Differences in local economic environments, such as institutions, protection of property rights, financial development, investor sentiment, etc.

Motivation (Part 5)

- Sometimes, we can control for unobservable variables using proxy variables.
- But what assumption was required for a proxy variable to provide consistent estimates on the other parameters?
- Answer: the proxy variable must be a sufficiently good proxy such that the unobserved variable cannot be correlated with other explanatory variables after controlling for the proxy variable... This might be hard to find

Panel Data to the Rescue

- Thankfully, panel data can help us with a particular type of unobserved variable ...
- What type of unobserved variable does panel data help us with, and why?
    - Panel data can help with unobserved time-invariant variables.
    - Actually, it allows for controlling unobserved variables that do not vary within groups of observations.

Outline

Motivate how panel data is helpful

Fixed effects model

- Benefits [There are many]
- Costs [There are some...]

Random effects model

First differences

Lagged y models

Panel Data

- Panel data is defined when you have multiple observations per unit of observation i (e.g., observe each firm over multiple years).
- Examples:
  - 5,000 firms over a twenty-year period ($N = 5,000$, $T = 20$).
  - 1,000 CEOs over a ten-year period ($N = 1,000$, $T = 10$).
  - These are balanced panels.

Time-Invariant Unobserved Variable

- Consider the following model:

$$y_{i,t} = \alpha + \beta x_{i,t} + f_i + u_{i,t}$$

- f_i is the unobserved, time-invariant variable.
    - $E(u_{i,t}) = 0$
    - $corr(x_{i,t}, f_i) \neq 0$: If we don't control for f_i, we will have OVB.
    - $corr(f_i, u_{i,t}) = 0$
    - $corr(x_{i,t}, u_{i,s}) = 0$ for all s, t: This is stronger assumption than we usually make. It's called strict exogeneity.

If we ignore f_i, we get OVB

- If we estimate the model without controlling for f_i:

  $$y_{i,t} = \alpha + \beta x_{i,t} + v_{i,t}$$

  where $v_{i,t} = f_i + u_{i,t}$.
- x now correlated with the disturbance v through its correlation with the unobserved variable f_i, causing omitted variable bias.
- Easy to show

  $$\hat{\beta}_{OLS} = \beta + \frac{\sigma_{x,f}}{\sigma_x^2}\delta$$
- This is standard OVB: the coefficient from the regression of the omitted variable f_i on x, multiplied by the true coefficient on f_i.

Can solve this by transforming data (Part 1)

- First, notice that if you take the population mean of the dependent variable for each unit of observation i, you get:

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

where $\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{i,t}$, and similarly for $\bar{x}_i$ and $\bar{u}_i$.

Transforming the Data (Part 2)

- Now, if we subtract the group mean $\bar{y}_i$ from $y_{i,t}$, we get:[1]

$$y_{i,t} - \bar{y}_i = \beta(x_{i,t} - \bar{x}_i) + (u_{i,t} - \bar{u}_i)$$

- The unobserved variable f_i is gone (as is the constant), because it is time-invariant.
- With strict exogeneity, it is easy to see that $(x_{i,t} - \bar{x}_i)$ is uncorrelated with the new disturbance $(u_{i,t} - \bar{u}_i)$, allowing us to estimate the model without omitted variable bias.
  - $corr(x_{i,t}, u_{i,s}) = 0$ for all s, t

[1] Note that the degrees of freedom will take account of the N means that were estimated, one for each individual. Thus, unlike pooled OLS where the number of degrees of freedom would be $(NT - k)$, the degrees of freedom for the FE estimator will be $(NT - k - N)$

Fixed Effects (or Within) Estimator

- OLS estimation of transformed model will yield a consistent estimate of β.
- This transformation is called the "within transformation" because it demeans all variables within their group.
  - In this case, the "group" was each cross-section of observations over time for each firm
  - This is also called the FE estimator

Unobserved Heterogeneity

- The unobserved variable f_i captures all unobserved variables that do not vary within the group.
- This is often called "unobserved heterogeneity."

Least Squares Dummy Variable (LSDV)

- Another way to perform FE estimation is by adding indicator (dummy) variables.

$$y_{i,t} = \alpha + \beta x_{i,t} + f_i + u_{i,t}$$

- Create a dummy variable for each group i, and add it to the regression.
- Now, to estimate this, we can just treat each f_i as a parameter to be estimated
- This is known as the least squares dummy variable (LSDV) model.
- We get consistent estimates and SE that are identical to what we'd get with within estimator

LSDV – Practical Advice

- Because the dummy variables will be collinear with the constant, one of them will be dropped in the estimation
  - Therefore, don't try to interpret the intercept; it is just the average y when all the x's are equal to zero for the group corresponding to the dropped dummy variable
  - In Stata, `xtreg, fe`: the reported intercept is just average of individual specific intercepts

LSDV versus FE (Part 1)

- Can show that LSDV and FE are identical, using partial regression results
  - Remember, to control for some variable z, we can regress y onto both x and z, or we can just partial z out from both y and x before regressing y on x (i.e., regress residuals from regression of y on z onto residual from regression of x on z)
  - The demeaned variables are the residuals from a regression of them onto the group dummies!

LSDV versus FE (Part 2)

- Reported R^2 will be larger with LSDV.
  - All the dummy variables will explain a lot of the variation in y, driving up RF^2
  - Within R^2 reported for FE estimator just reports what proportion of the within variation in y that is explained by the within variation in x
  - The within R^2 is usually of more interest to us

R-squared with FE – Practical Advice

- The within R^2 is usually of more interest, as it describes the explanatory power of the x's after partialling out the FE.
  - The get within R^2, use `xtreg, fe`
- Reporting overall adjusted-R2 is sometimes useful
  - To get the overall R^2, use the `areg` command in Stata.
  - The "overall R^2" reported by `xtreg` does not include variation explained by FE, but the R^2 reported by `areg` does

Outline

Motivate how panel data is helpful

Fixed effects model

- Benefits [There are many]
- Costs [There are some...]

Random effects model

First differences

Lagged y models

FE Estimator – Benefits (Part 1)

- There are many benefits of the FE estimator:
  - Allows for arbitrary correlation between each fixed effect, f_i, and each x within group i.
    - Very general and does not impose much structure on the underlying data.
  - Intuitive interpretation; the coefficient is identified using only changes within cross-sections.

FE Estimator – Benefits (Part 2)

- The FE estimator is very flexible and can control for many types of unobserved heterogeneities:
    - Add year fixed effects (FE) if worried about unobserved heterogeneity across time (e.g., macroeconomic shocks).
    - Add CEO FE if worried about unobserved heterogeneity across CEOs (e.g., talent, risk aversion).
    - Add industry-by-year FE if worried about unobserved heterogeneity across industries over time (e.g., investment opportunities, demand shocks).

FE Estimator – Tangent (Part 1)

- The FE estimator is very general and applies to any scenario where observations can be grouped together.
- Examples:
  - Firms can be grouped by industry.
  - CEO observations (which may span multiple firms) can be grouped by CEO-firm combinations.
- Grouping units i across time is the most common textbook example, but the concept can be generalized to other scenarios.

FE Estimator – Tangent (Part 2)

- Once you can construct groups, you can remove any unobserved group-level heterogeneity by adding group fixed effects (FE).
- Consistency requires there to be many groups[2]

[2] $\hat{\beta}_n^{FE} = \frac{\sum_i \sum_t \dot{x}_{i,t} \dot{y}_{i,t}}{\sum_i \sum_t \dot{x}_{i,t}^2} = \frac{\sum_i \sum_t \dot{x}_{i,t}(\beta \dot{x}_{i,t} + \dot{u}_{i,t})}{\sum_i \sum_t \dot{x}_{i,t}^2} = \beta + \frac{\sum_i \sum_t \dot{x}_{i,t} \dot{u}_{i,t}}{\sum_i \sum_t \dot{x}_{i,t}^2} \xrightarrow{p} \beta + E(xu) = \beta$ as $n \to \infty$ by the WLLN. $\dot{x}_{i,t} \equiv x_{i,t} - \bar{x}_i$.

Asymptotic Properties As $n \to \infty$ and T fixed (short panels), $\hat{\beta}_n^{FE}$ is asymptotically normal.

Write: $\sqrt{n}(\hat{\beta}_n^{FE} - \beta) = \left(\frac{1}{n}\sum_i \sum_t \dot{x}_{i,t}^2\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_i \sum_t \dot{x}_{i,t}\dot{u}_{i,t}\right)$

Two tricks:

1. $\sum_t \dot{x}_{i,t}\dot{u}_{i,t} = \sum_t \dot{x}_{i,t}u_{i,t} - \bar{u}_i \sum_t \dot{x}_{i,t} = \sum_t \dot{x}_{i,t}u_{i,t}$ $(\because \sum_t \dot{x}_{i,t} = 0)$
2. Let $\dot{x}_i \equiv (\dot{x}_{i,1}, \dot{x}_{i,2}, ..., \dot{x}_{i,T})'$. We can write $\dot{x}_i'\dot{x}_i = \sum_t \dot{x}_{i,t}^2$ and $\dot{x}_i' u_i = \sum_t \dot{x}_{i,t} u_i$

Then $\sqrt{n}(\hat{\beta}_n^{FE} - \beta) = \left(\frac{1}{n}\sum_i \dot{x}_i'\dot{x}_i\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_i \dot{x}_i' u_i\right)$.

cont'd By the WLLN, $\frac{1}{n}\sum_i \dot{x}_i'\dot{x}_i \xrightarrow{p} \Sigma_{\dot{x}} \equiv E(\dot{x}_i'\dot{x}_i) = \sum_t E(\dot{x}_{i,t}\dot{x}_{i,t}')$

By the CLT, $\frac{1}{\sqrt{n}}\sum_i \dot{x}_i' u_i = \sqrt{n}\left(\frac{1}{n}\sum_i \dot{x}_i' u_i\right) \xrightarrow{d} N(0,\Omega)$, where
$\Omega = Var(\dot{x}_i' u_i) = E[\dot{x}_i' u_i u_i' \dot{x}_i]$

By the CMT, $\sqrt{n}(\hat{\beta}_n^{FE} - \beta) \xrightarrow{d} N(0, \mathbb{V}^{FE})$, where $\mathbb{V}^{FE} = \Sigma_{\dot{x}}^{-1}\Omega\Sigma_{\dot{x}}^{-1}$

Note: The robust consistent estimator of the asymptotic variance is
$\hat{\mathbb{V}}^{FE} = \left(\frac{1}{n}\sum_i \dot{x}_i'\dot{x}_i\right)^{-1}\left(\frac{1}{n}\sum_i \dot{x}_i'\hat{u}_i\hat{u}_i'\dot{x}_i\right)\left(\frac{1}{n}\sum_i \dot{x}_i'\dot{x}_i\right)^{-1}$, where
$\hat{u}_i = \dot{y}_i - \dot{x}_i\hat{\beta}_n^{FE}$.
It can be shown that $\hat{\mathbb{V}}^{FE} \xrightarrow{p} \mathbb{V}^{FE}$ as $n \to \infty$.

Outline

Motivate how panel data is helpful

Fixed effects model

- Benefits [There are many]
- Costs [There are some...]

Random effects model

First differences

Lagged y models

FE Estimator – Costs

- Can't identify variables that don't vary within groups
- Subject to potentially large measurement error bias
- Can be hard to estimate in some cases
- Miscellaneous issues

FE Cost #1 – Can't estimate some var

- If no within-group variation in the independent variable, x, of interest, can't disentangle it from group FE
    - It is collinear with group FE; and will be dropped by computer or swept out in the within transformation

FE Cost #1 – Example

- Consider a CEO-level estimation where log(*totalpay*) includes year, CEO, and firm FE.

$$\begin{aligned}\ln(totalpay)_{ijt} &= \alpha + \beta_1 \ln(firmsize)_{ijt} + \beta_2 volatility_{ijt} \\ &\quad + \beta_3 female_i + \delta_t + f_i + \lambda_j + u_{ijt}\end{aligned}$$

- What coefficient can't be estimated?
  - Variables that do not vary within the group cannot be estimated; i.e., it is collinear with the CEO fixed effect (e.g., gender).

FE Cost #1 – Practical Advice

- Be careful of this!
  - Programs like `xreg` are good about dropping the female variable and not reporting an estimate ...
  - But, if you create dummy variables yourself and input them yourself, the estimation might drop one of them rather than the female indicator
  - I.e., you'll get an estimate for β_3, but it has no meaning! It's just a random intercept value that depends entirely on the random FE dropped by Stata

FE Cost #1 – Any Solution?

- Instrumental variables can provide a potential solution for variables that do not vary within groups.
- See Hausman and Taylor (Econometrica 1981)
- We will discuss this later.

FE Cost #2 – Measurement Error (Part 1)

- Measurement error of independent variable (and resulting biases) can be amplified
  - Think of there being two types of variation
  - Good (meaningful) variation
  - Noise variation because we don't perfectly measure the underlying variable of interest
- Adding FE can sweep out a lot of the good variation; fraction of remaining variation coming from noise goes up [What will this do?]

FE Cost #2 – Measurement Error (Part 2)

- Answer: Attenuation bias on mismeasured variable will go up!
  - Practical advice: Be careful in interpreting 'zero' coefficients on potentially mismeasured regressors; might just be attenuation bias!
  - And remember, sign of bias on other coefficients will be generally difficult to know

FE Cost #2 – Measurement Error (Part 3)

- Problem can also apply even when all variables are perfectly measured [How?]
    - Answer: Adding FE might throw out relevant variation; e.g., y in firm FE model might respond to sustained changes in x, rather than transitory changes[3]
    - With FE you'd only have the transitory variation leftover; might find x uncorrelated with y in FE estimation even though sustained changes in x is most important determinant of y

[3] McKinnish, T. 2008. Panel Data Models and Transitory Fluctuations in the Explanatory Variable In Modeling and Evaluating Treatment Effects in Econometrics

FE Cost #2 – Any solution?

- Admittedly, measurement error, in general, is difficult to address
- For examples on how to deal with measurement error, see following papers
  - Griliches and Hausman (JoE 1986)
  - Biorn (Econometric Reviews 2000)
  - Erickson and Whited (JPE 2000, RFS 2012)
  - Almeida, Campello, and Galvao (RFS 2010)

ME in Panel Data Remember, with ME in independent variable, we have

$$plim(\hat{\beta}) = \beta \frac{\text{var}(x^*)}{\text{var}(x^*) + \text{var}(e)}$$

In first-difference estimator:

$$plim(\hat{\beta}) = \beta \frac{\text{var}(\Delta x^*)}{\text{var}(\Delta x^*) + \text{var}(\Delta e)}$$

, where $\text{var}(\Delta x^*) = \text{var}(x_t^*) - 2\text{cov}(x_t^*, x_{t-1}^*) + \text{var}(x_{t-1}^*)$. If x_t is stationary, $\text{var}(\Delta x^*) = 2\sigma_x^2 - 2\text{cov}(x_t^*, x_{t-1}^*) = 2\sigma_x^2(1-\rho)$. Define r to be the autocorrelation coefficient in u_t so we can write

$$plim(\hat{\beta}) = \beta \frac{2\sigma_x^2(1-\rho)}{2\sigma_x^2(1-\rho) + 2\sigma_u^2(1-r)} = \beta \frac{1}{1 + \frac{\sigma_u^2(1-r)}{\sigma_x^2(1-\rho)}}$$

When $r = \rho = 0$, traditional attenuation bias. When $r = 0$ only (ME is serially uncorrelated, but the signal is correlated), worrisome. When $r = 1$ (i.e., ME is fixed), FE eliminates ME.

FE Cost #3 – Computation issues (Part 1)

- Estimating a model with multiple types of FE can be computationally difficult
  - When more than one type of FE, you cannot remove both using within-transformation
  - Generally, you can only sweep one away with within-transformation; other FE dealt with by adding dummy variable to model
  - E.g., firm and year fixed effects [See next slide]

FE Cost #3 – Computation issues (Part 2)

- Consider below model:

$$y_{it} = \alpha + \beta x_{it} + \delta_t + f_i + u_{it}$$

  - To estimate this in Stata, we'd use a command something like the following ...

```
xtset firm
xi:  xtreg y x i.year, fe
```

FE Cost #3 – Computation issues (Part 3)

- Dummies not swept away in within-transformation are estimated
    - With year FE, this isn't problem because there aren't that many years of data
    - If had to estimate 1,000s of firm FE, however, it might be a problem
    - In fact, this is why we sweep away the firm FE rather than the year FE; there are more firms!

FE Cost #3 – Example

- But computational issues is becoming increasingly more problematic
  - E.g., if you try adding both firm and industry×year FE, you'll have a problem

FE Cost #3 – Any Solution?

- Yes, there are some potential solutions
  - We will come back to this in "Common Limitations and Errors" lecture.[4]

[4] Gormley and Matsa (2014) discusses some of these solutions in Section 4

FE – Some Remaining Issues

- Two more issues worth noting about FE
  - Predicted values of unobserved FE
  - Non-linear estimations with FE and the incidental parameter problem

Predicted values of FE (Part 1)

- Sometimes, predicted value of unobserved FE is of interest
- Can get predicted value using

$$\hat{f}_i = \bar{y}_i - \hat{\beta}\bar{x}_i, \quad \forall i$$

  - E.g., Bertrand and Schoar (QJE 2003) did this to back out CEO fixed effects
  - They show that the CEO FE are jointly statistically significant from zero, suggesting CEOs have 'styles' that affect their firms

Predicted values of FE (Part 2)

- But be careful with using these predicted values of the FE (using LSDV)[5]
- They are unbiased, but inconsistent
  - As sample size increases (and we get more groups), we have more parameters to estimate... never get the necessary asymptotics
  - We call this the **Incidental Parameters Problem**
  - Particularly problematic in non-linear models. In linear models, we can transform the data (within- or first difference) to estimate parameters of interest. We cannot in non-linear models.
  - Also, in most non-linear models, their inconsistency "contaminates" the estimation of the parameter of interest, which becomes inconsistent as well.[6]

[5] To be clear, $\hat{\beta}$ is consistent when T is fixed and $N \rightarrow \infty$. But $\hat{f}_i$'s are not.

[6] The fixed effects Poisson model doesn't suffer from the incidental parameter.

Predicted values of FE (Part 3)

- Moreover, doing an F-test to show they are statistically different from zero is only valid under rather strong assumptions
  - Need to assume errors, u, are distributed normally, homoskedastic, and serially uncorrelated
  - Fee, Hadlock, and Pierce (2011): Bertrand and Schoar (2003) manager style still appears when CEO-firm matches are scambled.
  - See Wooldridge (2010, Section 10.5.3) and for more details

Nonlinear Models with FE (Part 1)

- Because we don't get consistent estimates of the FE, we can't estimate nonlinear panel data models with FE
  - In practice, nonlinear models (e.g., Logit, Tobit, Probit) should not be estimated with many fixed effects.
  - They only give consistent estimates under rather strong and unrealistic assumptions

Nonlinear Models with FE (Part 2)

- E.g., Probit with FE requires...
  - Unobserved f_i is to be distributed normally
  - f_i and $x_{i,t}$ to be independent
- And Logit with FE requires ...
  - No serial correlation of y after conditioning on the observable x and unobserved f
- For more details, see...
  - Wooldridge (2010), Sections 13.9.1, 15.8.2-3
  - Greene (2004) – uses simulation to show how bad

Outline

Motivate how panel data is helpful

Fixed effects model

- Benefits [There are many]
- Costs [There are some...]

Random effects model

First differences

Lagged y models

Random Effects (RE) Model (Part 1)

- Very similar model as FE

$$y_{it} = \alpha + \beta x_{it} + f_i + u_{it}$$

- But one big difference
  - The RE model assumes that unobserved heterogeneity f_i and the observed x's are uncorrelated.
  - What does this imply about consistency of OLS?
  - Is this a realistic assumption?

Random Effects (RE) Model (Part 2)

- Answer $\#1$ – That assumption means that OLS would give you consistent estimate of β!
- Then why bother?
  - Answer... potential efficiency gain relative to FE
  - FE is no longer most efficient estimator. If our assumption is correct, we can get more efficient estimate via generalized least squares [Note: can't just do OLS; it will be consistent as well but SE will be wrong since they ignore serial correlation]

Random Effects (RE) Model (Part 3)

- Answer #2 – The assumption that f and x are uncorrelated is likely unrealistic
  - The violation of this assumption is whole motivation behind why we do FE estimation!
  - Recall that correlation between unobserved variables, like managerial talent, demand shocks, etc., and x will cause omitted variable bias

Random Effects (RE) Model (Part 4)

- In practice, RE model is not very useful
- Angrist-Pischke (page 223) write,
  - Relative to fixed effects estimation, random effects requires stronger assumptions to hold
  - Even if right, asymptotic efficiency gain likely modest
  - And finite sample properties can be worse

Outline

Motivate how panel data is helpful

Fixed effects model

- Benefits [There are many]
- Costs [There are some...]

Random effects model

First differences

Lagged y models

First Differencing (FD) (Part 1)

- First differencing is another way to remove unobserved heterogeneities.
    - Rather than subtracting off the group mean of the variable from each variable, you instead subtract the lagged observation

First Differencing (FD) (Part 2)

- Notice that,

$$y_{i,t} = \alpha + \beta x_{i,t} + f_i + u_{i,t}$$
$$y_{i,t-1} = \alpha + \beta x_{i,t-1} + f_i + u_{i,t-1}$$

- From this, we can see that[7]

$$y_{i,t} - y_{i,t-1} = \beta(x_{i,t} - x_{i,t-1}) + (u_{i,t} - u_{i,t-1})$$

  - When will OLS estimate of this provide a consistent estimate of β?
  - Answer: With same strict exogeneity assumption of FE (i.e., $x_{i,t}$ and $u_{i,s}$ are uncorrelated for all t and s)

[7] Note: we'll lose on observation per cross-section because there won't be a lag

First Differences (without time)

- First differences can also be done even when observations within groups aren't ordered by time
  - Just order the data within groups in whatever way you want, and take 'differences'
  - Works, but admittedly, not usually done

FD vs FE (Part 1)

- Wooldridge (2010, 10-7)
- When just two observations per group (i.e. $T = 2$), they are identical to each other
- In other cases ($T > 2$), the two do not yield the same results. But both are consistent; difference is generally about efficiency
  - FE is more efficient if disturbances, $u_{i,t}$, are serially uncorrelated[8]
  - FD is more efficient if disturbances, $u_{i,t}$, are serially correlated[9]

[8] Intuition: taking first differences introduces correlation in $\Delta u_{i,t}$ as $E(\Delta u_{i,t} \Delta u_{i,t-1}) = E(u_{i,t} u_{i,t-1} - u_{i,t-1} u_{i,t-1} - u_{i,t} u_{i,t-2} - u_{i,t-1} u_{i,t-2}) = -Var(u_{i,t-1})$

[9] e.g., $u_{i,t} = u_{i,t-1} + \epsilon_{i,t}$

FD vs FE (Part 2)

- FE is more sensitive to nonnormality, heteroskedasticity, and serial correlation in errors.
- Both FD and FE are sensitive to classical measurement error in explanatory variables.
- If strict exogeneity is violated (i.e., $x_{i,t}$ is correlated with $u_{i,s}$ for $s \neq t$), FE might be better
    - As long as we believe $x_{i,t}$ and $u_{i,t}$ are uncorrelated, the FE's inconsistency shrinks to 0 at rate $1/T$; but FD gets no better with larger T (T is the $\#$ of observations per group)

FD vs FE (Part 3)

- Bottom line: not a bad idea to try both ...
  - If different, you should try to understand why
  - With an omitted variable or measurement error, you'll get diff. answers with FD and FE
  - In fact, Griliches and Hausman (1986) shows that because measurement error causes predictably different biases in FD and FE, you can (under certain circumstances) use the biased estimates to back out the true parameter.

Outline

Motivate how panel data is helpful

Fixed effects model

- Benefits [There are many]
- Costs [There are some...]

Random effects model

First differences

Lagged y models

Lagged Dependent Variables with FE

- Models with both lagged dependent variables and FE cannot easily be estimated using OLS or FE.

$$y_{i,t} = \alpha + \rho y_{i,t-1} + \beta x_{i,t} + f_i + u_{i,t}, \quad |\rho| < 1$$

- Same as before, but now true model contains lagged y as independent variable
  - Can't estimate with OLS even if x & f are uncorrelated
  - Can't estimate with FE

Lagged y & FE - Problem with OLS

- To see the problem with OLS, suppose you estimate the following:

$$y_{i,t} = \alpha + \rho y_{i,t-1} + \beta x_{i,t} + \underbrace{\nu_{i,t}}_{f_i + u_{i,t}}$$

- But $y_{i,t-1} = \alpha + \rho y_{i,t-2} + \beta x_{i,t-1} + f_i + u_{i,t-1}$
- Thus $y_{i,t-1}$ and $\nu_{i,t}$ are correlated because they both contain f_i.
- I.e., you get omitted variable bias

Lagged y & FE - Problem with FE

- Will skip the math, but it is always biased
  - Basic idea is that if you do a within transformation, the lagged mean of y, which will be on RHS of the model now, will always be negatively correlated with demeaned error, u
  - Note #1 – This is true even if there was no unobserved heterogeneity, f; FE with lagged values is always bad idea
  - Note #2: Same problem applies to FD
  - Problem, however, goes away as T goes to infinity

Then how do we estimate this? IV?

- Basically, you're going to need instrument; we will come back to this later...

Lagged y & FE - Bracketing

- Suppose you don't know which is correct
  - Lagged model
  - Or FE model
- Can show that estimate of $\beta > 0$ will
  - Be too high if lagged model is correct, but you incorrectly use FE model
  - Be too low if FE model is correct, but you incorrectly used lagged model

Lagged y & FE - Bracketing (Contd.)

- Use this to ‘bracket’ where true β is...
  - But sometimes, you won’t observe bracketing
  - Likely means your model is incorrect in other ways, or there is some severe finite sample bias

Summary (Part 1)

- Panel data allows us to control for certain types of unobserved variables.
  - FE estimator can control for these potential unobserved variables in very flexible way
  - Greatly reduces the scope for potential omitted variable biases we need to worry about
  - Random effects model is useless in most empirical corporate finance settings

Summary (Part 2)

- FE estimator, however, has weaknesses
  - Can't estimate variables that don't vary within groups [or at least, not without an instrument]
  - Could amplify any measurement error
  - For this reason, be cautious interpreting zero or small coefficients on possibly mismeasured variables
  - Can't be used in models with lagged values of the dependent variable [or at least, not without an IV]

Summary (Part 3)

- FE are generally not a good idea when estimating nonlinear models [e.g., Probit, Tobit, Logit]; estimates are inconsistent
- First differences can also remove unobserved heterogeneity
  - Largely just differs from FE in terms of relative efficiency; which depends on error structure