Unobserved Heterogeneity

Tyler Ransom

Duke University

June 21, 2012

Plan

Introduce and discuss what unobserved heterogeneity is and why it's important Go through different ways of correcting for unobs het

- Some will be familiar
- You will see all of these (and more!) in your 2nd year

Unobserved Heterogeneity

Researchers estimate models of the form $y_i = X_i\beta + \varepsilon_i$ where y_i is some outcome variable and ε_i is assumed to be uncorrelated with X_i.
In actuality, the X_i and ε_i are likely to be correlated

- This is usually because there are important variables explaining y that are not included in X (generally because of data limitations)

Unobserved Heterogeneity

- What we mean by "unobserved heterogeneity" is that there is something the researcher can't observe, which is important in answering the research question
  - The terms "unobserved heterogeneity" and "endogeneity" can be used interchangeably
  - "unobserved heterogeneity" is the preferred term in 2nd year modules
- Failing to correct for unobserved heterogeneity in some way will bias the estimates of the b (the parameters of interest)
- Correcting for unobserved heterogeneity is >50% of the work in estimation
- We care about correcting for it because at the end of the day we want to recover a causal relationship for what we observe. We can't do this without making such a correction
- There are many, many ways to correct for unobserved heterogeneity (we'll discuss just a few today)

Examples of Unobserved Heterogeneity

IO

- Demand estimation: products have unobservable attributes that are correlated to their demand (e.g. iPad is cooler than other tablets, even above and beyond any observable trait like screen resolution or app market size)

Education

- Teacher quality: Teachers have interactions with students that can't be quantified in any test score metric, but these interactions directly affect student outcomes (and are correlated with X)

Labor

- Workers have unobservable attributes (like innate ability) that are directly related to productivity, but which are also correlated with X (e.g. more ability $\Rightarrow$ more education)

Correcting for Unobserved Heterogeneity

Ways of correcting for unobserved heterogeneity

- Instrumental Variables
- Fixed/Random Effects
- Nonlinear Models
  - Nonparametric methods
  - Parametric methods

We'll go through each of these (along with their relative advantages/disadvantages) in more depth

Instrumental Variables

- $y_i = X_i\beta + \varepsilon_i$
- X_i correlated with ε_i
- Find some other set of variables Z_i which is correlated with X_i but not with ε_i
- Use GMM to estimate β
- Pros:
  - Simple to implement
  - Doesn't require panel data
  - Doesn't necessarily require a linear model (though computation intensifies if instruments are used in a nonlinear case)
- Cons:
  - Hard to find instruments that satisfy the required conditions
  - Overly simplistic in terms of describing behavior

How estimates can change with panel data

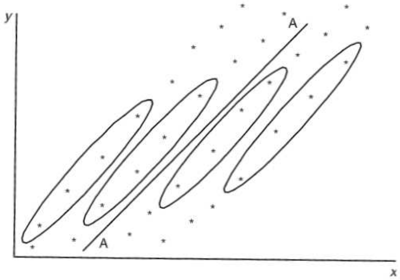


Figure 18.1 Panel data showing four observations on each of four individuals.

Source: Kennedy (2008)

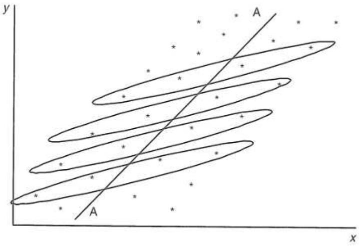


Figure 18.2 Panel data showing four observations on each of four individuals, with positive correlation between x and the intercept.

Source: Kennedy (2008)

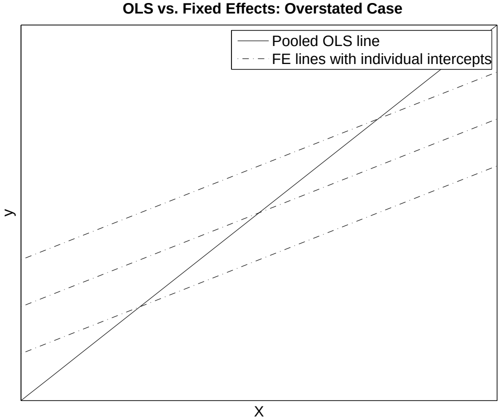
Fixed Effects, mathematically

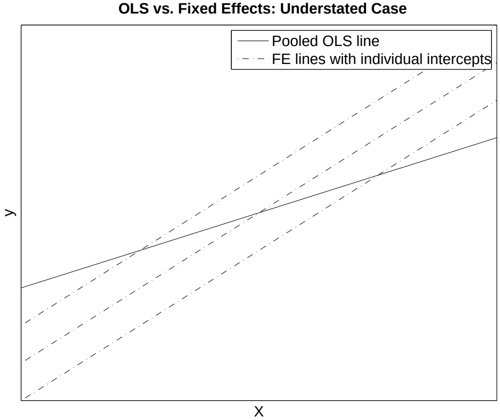
- $y_{it} = X_{it}\beta + \alpha_i + \varepsilon_{it}$
- Decompose the error term into a persistent component (α_i) and a transitory component (ε_{it}) – both mean-zero
- Assume ε_{it} is independent of X_{it}
- α_i can be correlated with X_{it}
- Use a difference within persons to recover beta:

$$\hat{\beta} = \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \ddot{X}'_{it} \ddot{X}_{it} \right)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \ddot{X}'_{it} \ddot{y}_{it}$$

  where $\ddot{X}_{it} = X_{it} - \overline{X}_i$ and $\ddot{y}_{it} = y_{it} - \overline{y}_i$
- The α_i disappears because it is time-invariant

Fixed Effects, graphically





Fixed Effects

- Can also express the fixed effects estimator in a matrix formula:

  $$\hat{\beta} = \left(X'QX\right)^{-1} X'Qy$$

  where
  $Q = I_N \otimes \left(I_T - \frac{1}{T} l_T l_T'\right)$
  $\otimes$ is the Kronecker Product
  l_T is a T-length vector of ones
  I_N is an $N \times N$ identity matrix
- Notes: This is much more difficult (but not impossible) to implement for panels where T varies across persons (i.e. unbalanced panels). It is also computationally intensive for large $N * T$, since Q is a very wide (block diagonal) matrix.

Standard Errors of Fixed Effects Estimates

The variance matrix of $\hat{\beta}$ for fixed effects is

$$\hat{\sigma}_{\varepsilon}^{2}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\ddot{X}_{it}'\ddot{X}_{it}\right)^{-1}$$

where

$$\hat{\sigma}_{\varepsilon}^{2}=\frac{\mathrm{SSR}}{N(T-1)-K}$$

and where SSR is derived from the sum of the squared residuals from pooled OLS:

$$\sum_{i}\sum_{t}\left(y_{it}-X_{it}\hat{\beta}_{POLS}\right)^{2}$$

Note: in matrix form, can estimate

$$\mathrm{SE}_{\hat{\beta}_{FE}}=\hat{\sigma}_{\varepsilon}^{2}\left(X'QX\right)^{-1}$$

Fixed Effects

Pros:

- Easy to compute
- Corrects for permanent unobserved heterogeneity in a straightforward fashion

Cons:

- Requires panel data
- β can only be attached to a time-varying X (all time-invariant X's get differenced out)
- Specification must be linear
- Must assume that heterogeneity is permanent

Random Effects

- $y_{it} = X_{it}\beta + \alpha_i + \varepsilon_{it}$
- Decompose the error term into a persistent component (α_i) and a transitory component (ε_{it}) – both mean-zero
- Assume ε_{it} and α_i are each independent of X_{it} and each other for all t
- Use variance of ε_{it} and α_i within persons to recover beta:

$$\hat{\beta} = \left(\sum_{i=1}^{N} X_i' \hat{\Omega}^{-1} X_i \right)^{-1} \sum_{i=1}^{N} X_i' \hat{\Omega}^{-1} y_i$$

where $\hat{\Omega}^{-1}$ is a matrix of $\hat{\sigma}_\alpha^2$ and $\hat{\sigma}_\varepsilon^2$ (variances of the error term components, see Wooldridge p. 259)

Random Effects $\hat{\Omega}$ Matrix

The matrix $\hat{\Omega}$ looks like

$$\hat{\Omega}_{T\times T} = \begin{bmatrix} \hat{\sigma}_\alpha^2 + \hat{\sigma}_\varepsilon^2 & \hat{\sigma}_\alpha^2 & \cdots & \hat{\sigma}_\alpha^2 \\ \hat{\sigma}_\alpha^2 & \hat{\sigma}_\alpha^2 + \hat{\sigma}_\varepsilon^2 & & \vdots \\ \vdots & & \ddots & \hat{\sigma}_\alpha^2 \\ \hat{\sigma}_\alpha^2 & \cdots & & \hat{\sigma}_\alpha^2 + \hat{\sigma}_\varepsilon^2 \end{bmatrix}$$

where

$$\hat{\sigma}_\alpha^2 = \frac{1}{NT(T-1)/2 - K} \sum_{i=1}^{N} \sum_{t=1}^{T-1} \sum_{s=t+1}^{T} \hat{\hat{u}}_{it} \hat{\hat{u}}_{is}$$

($\hat{\hat{u}}_{it}$ are pooled OLS residuals), and

$$\hat{\sigma}_\varepsilon^2 = \left(\frac{1}{NT-K} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{\hat{u}}_{it}^2 \right) - \hat{\sigma}_\alpha^2$$

Standard Errors of Random Effects Estimates

The (robust) variance matrix of $\hat{\beta}_{RE}$ (as suggested by Wooldridge (pp. 160, 262) is

$$\left(\sum_{i=1}^{N} X_i'\hat{\Omega}^{-1}X_i\right)^{-1}\left(\sum_{i=1}^{N} X_i'\hat{\Omega}^{-1}\hat{u}_i\hat{u}_i'\hat{\Omega}^{-1}X_i\right)\left(\sum_{i=1}^{N} X_i'\hat{\Omega}^{-1}X_i\right)^{-1}$$

where where

$$\hat{u}_i = y_i - X_i\hat{\beta}_{RE}$$

and where all vectors are of length T_i (i.e. number of panels within the individual)

Random Effects

Pros:

- Easy to compute
- No differencing, so β can be attached to time-invariant X's

Cons:

- Requires panel data
- Requires both error terms to be uncorrelated with X and each other – strong assumption
- Specification must be linear
- Must assume that heterogeneity is permanent

Fixed Effects vs. Random Effects

- Random effects is more efficient, but fixed effects is more robust
- i.e. if α_i and ε_{it} are actually independent, RE will give smaller standard errors
- However, if this is not true, RE is inconsistent so FE should be used
- Researcher can use a Hausman test to see if RE should be used
- In practice, most papers use FE because it's easier to understand, requires fewer assumptions, and is more robust to misspecification

Unobserved Heterogeneity in Nonlinear Models

- So far, we've only looked at unobserved heterogeneity in linear models
- Once the model becomes nonlinear (e.g. simple probit), it becomes increasingly difficult to correct for unobserved heterogeneity because differencing won't work
- Can't easily do fixed effects in nonlinear models (no differencing, and dummy for each person $\Rightarrow$ incidental parameters problem)
- Random effects is easier to compute, but requires additional assumptions and is inconsistent if the effect is actually correlated with X

Unobserved Heterogeneity in MLE Models

- Most MLE models assume a random effects specification for the unobserved heterogeneity
- Assume a distribution of the RE (e.g. normal, discrete-non-parametric)
- Integrate out the unobserved effect from the likelihood while recovering parameters that characterize the assumed distribution
- Some fancier models can estimate heterogeneity that is both permanent and transitory (something FE/RE can't do)
  - This requires more elaborate assumptions
- You'll learn more about the finer points of this in your 2nd year modules

Unobserved Heterogeneity in Perspective

- Estimation results that have corrected for unobserved heterogeneity are more important than any other estimation results
- You will not be able to publish anything without saying *something* about unobserved heterogeneity
- At the end of the day, it doesn't matter too much *how* you make the correction, just *that* you make the correction