Fixed Effects and Twins Estimates of Returns to Schooling

Jörn-Steffen Pischke

LSE

October 19, 2018

Fixed effects assumptions

- Regression (think of $S_i \in \{0,1\}$):

$$E(Y_{0i}|A_i, X_i, S_i) = E(Y_{0i}|A_i, X_i)$$

Conditional independence $\Rightarrow$ need to observe A_i

- Now we have panel data:

$$E(Y_{0it}|A_i, X_{it}, t, S_{it}) = E(Y_{0it}|A_i, X_{it}, t),$$

Notice: time subscripts, A_i does not have time subscript. Moreover assume linearity:

$$E(Y_{0it}|A_i, X_{it}, t) = \alpha + \lambda_t + \gamma A_i + X_{it}\delta, \tag{1}$$

and an additive and constant effect for the return to schooling

$$E(Y_{1it}|A_i, X_{it}, t) = E(Y_{0it}|A_i, X_{it}, t) + \rho.$$

The fixed effects model

Combining the two counterfatctuals

$$E(Y_{it}|A_i, X_{it}, t, S_{it}) = \alpha + \lambda_t + \rho S_{it} + \gamma A_i + X_{it}\delta$$

or

$$Y_{it} = \alpha + \lambda_t + \rho S_{it} + \gamma A_i + X_{it}\delta + \varepsilon_{it}.$$

Write

$$\alpha_i \equiv \alpha + \gamma A_i$$

so that

$$Y_{it} = \alpha_i + \lambda_t + \rho S_{it} + X_{it}\delta + \varepsilon_{it}.$$

This is a *fixed-effects model.* Given panel data, i.e. repeated observations on units i, the causal effect of schooling on wages can be estimated by treating α_i, the fixed effect, as a parameter to be estimated.

You get what you pay for

What did we put in:

- Linearity: counterfactual outcomes without schooling are linear in A_i.
  - Linearity in the regression model: pure convenience
  - Linearity in the fixed effects model: crucial assumption
- No time variation of A_i.

What do we get out:

- We do not need to observe the confounder A_i.

Twins

Instead of thinking about panel data as individuals over time, think about individuals in families. I.e. let i denote family or twin pair, and t sibling. The key assumption we are making is

$$A_{it} = A_i$$

i.e. ability is the same for both twins in the pair i. And, importantly, S_{it} has to vary within twin pairs. We know from the data that it does, although not tremendously much.
We can difference within twin pairs to get

$$\Delta Y_{it} = \Delta\lambda_t + \rho\Delta S_{it} + \Delta X_{it}\delta + \Delta\varepsilon_{it}.$$

The Twinsburg Twins Festival

Returns to schooling among identical twins

	1994 Data		1998 Data	
	Pooled OLS	First diff	Pooled OLS	First diff
Regressor	(1)	(2)	(3)	(4)
Years of schooling	0.084	0.092	0.110	0.070
	(0.014)	(0.024)	(0.009)	(0.019)
Age	0.088		0.104	
	(0.019)		(0.110)	
$Age^2/100$	−0.087		−0.106	
	(0.023)		(0.013)	
Female	−0.204		−0.318	
	(0.063)		(0.040)	
White	−0.410		−0.100	
	(0.127)		(0.072)	
Obs	298	149	680	340

Measurement error in the first difference estimator

It is easiest to consider measurement error in the first difference estimator and we just deal with the simple bivariate model

$$\Delta Y_{it} = \beta \Delta X_{it} + \Delta e_{it}.$$

Consider classical measurement error m_{it}. Proceeding like in the cross-sectional case gives

$$\widehat{\beta}_{FD} = \beta \frac{Var\left(\Delta X_{it}\right)}{Var\left(\Delta X_{it} + \Delta m_{it}\right)}.$$

The variance of the first difference

So we need $Var\left(\Delta X_{it}\right)$ and $Var\left(\Delta m_{it}\right)$:

$$Var\left(\Delta X_{it}\right) = Var(X_{it}) + Var(X_{it-1}) - 2Cov\left(X_{it}, X_{it-1}\right)$$

Assuming $Var\left(X_{it}\right) = Var(X_{it-1})$ (known as covariance stationarity), we get

$$Var\left(\Delta X_{it}\right) = 2Var(X_{it}) - 2\rho_X Var(X_{it}) = (1-\rho_X)\, 2Var(X_{it})$$

where ρ_X is the first-order autocorrelation coefficient of X_{it}.

The panel attenuation factor

Hence

$$\frac{Var\left(\Delta X_{it}\right)}{Var\left(\Delta X_{it}+\Delta m_{it}\right)} = \frac{\left(1-\rho_X\right)2Var(X_{it})}{\left(1-\rho_X\right)2Var(X_{it})+\left(1-\rho_m\right)2Var(m_{it})}$$

$$= \frac{Var(X_{it})}{Var(X_{it})+Var(m_{it})\frac{(1-\rho_m)}{(1-\rho_X)}}$$

$$= \frac{\lambda}{\lambda+(1-\lambda)\frac{(1-\rho_m)}{(1-\rho_X)}}$$

This new attenuation factor will be smaller than λ whenever $\rho_m < \rho_X$.

When is measurement error worse for the first difference estimator?

- This attenuation factor for the first difference estimator is smaller than λ whenever $\rho_m < \rho_x$.
- ρ_m: serial correlation in the measurement error. Probably low, much mismeasurement may be period by period noise.
- ρ_X: serial correlation in the signal. Probably high, many economic variables are highly persistent.

$\Rightarrow$ it is often reasonable to assume that attenuation bias from measurement error will be a worse problem when controlling for fixed effects (this happens both in the FD and deviations from means estimator).

A numerical example

Suppose

- $\lambda = 0.9$
- $\rho_m = 0.3$
- $\rho_X = 0.9$

$$\frac{Var(\Delta X_{it})}{Var(\Delta X_{it}) + Var(\Delta m_{it})} = \frac{\lambda}{\lambda + (1-\lambda)\frac{(1-\rho_m)}{(1-\rho_X)}}$$

$$= \frac{0.9}{0.9 + 0.1 \cdot \frac{0.7}{0.1}} = 0.563$$

Measurement error in the twins data

$$S_{it} = S_{it}^* + m_{it}$$

Ashenfelter and Krueger/Rouse also collected reports of schooling from siblings. So we have for sibling 1:

$$\begin{aligned} S_{i1}^1 &= S_{i1}^* + m_{i1}^1 \\ S_{i1}^2 &= S_{i1}^* + m_{i1}^2 \end{aligned}$$

where the superscript denotes which sibling reports schooling.
Then

$$Cov\left(S_{i1}^1, S_{i1}^2\right) = Var\left(S_{i1}^*\right) + Cov\left(m_{i1}^1, m_{i1}^2\right).$$

Assuming $Cov\left(m_{i1}^1, m_{i1}^2\right) = 0$, the correlation coefficient of the two observed schooling levels is equal to the attenuation factor for the levels equation.

Measurement error in the twins data

In the 1994 data

$$\rho\left(S_{i1}^{1}, S_{i1}^{2}\right) \simeq 0.9.$$

The same exercise can be done for the twin differenced data.

$$\rho\left(\Delta S_{i}^{own}, \Delta S_{i}^{sib}\right) \simeq 0.57$$

The attenuation factor within twin pair is substantially lower than the attenuation factor in the raw data.

Under the assumtion $Cov\left(m_{i1}^{1}, m_{i1}^{2}\right) = 0$, we can use IV with sibling schooling to get estimates free of bias from measurement error

Returns to schooling among identical twins

OLS and IV estimates using sibling reports as instrument

	1994 Data		1998 Data	
	Pooled	First diff	Pooled	First diff
Method	(1)	(2)	(3)	(4)
OLS	0.084	0.092	0.110	0.070
	(0.014)	(0.024)	(0.009)	(0.019)
IV	0.116	0.167	0.116	0.088
	(0.030)	(0.043)	(0.010)	(0.025)
Obs.	298	149	680	340

When are twins estimates less biased than OLS estimates?

Griliches (1979) and Bound and Solon (1999)

Think of our returns to schooling regression as

$$S_{it} = \alpha + \rho S_{it} + \gamma A_{it} + e_{it}$$

where now A_{it} can vary by sibling. What conditions do we need on A_{it} for within twin pair estimates to be less biased than OLS?

Consider the following error components model:

$$\begin{aligned} S_{it} &= A_{it} + H_{it} \quad Cov(A_{it}, H_{it}) = 0 \\ A_{it} &= A_i + \mu_{it} \quad Cov(A_i, \mu_{it}) = 0 \\ H_{it} &= H_i + \eta_{it} \quad Cov(H_i, \eta_{it}) = 0 \end{aligned}$$

What is A_{it} and H_{it}?

- A_{it} is a component of schooling correlated with earnings, i.e. a control variable in a regression like ability.
- H_{it} is a component of schooling uncorrelated with earnings, i.e. a potential instrument for schooling like the cost of schooling.

Within and OLS estimates in the error components model

$$\rho_{OLS} = \rho + \gamma \frac{Cov(A_{it}, S_{it})}{Var(S_{it})} = \rho + \gamma \frac{\sigma_A^2 + \sigma_\mu^2}{\sigma_A^2 + \sigma_\mu^2 + \sigma_H^2 + \sigma_\eta^2}$$

$$\rho_W = \rho + \gamma \frac{Cov(\Delta A_{it}, \Delta S_{it})}{Var(\Delta S_{it})} = \rho + \gamma \frac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma_\eta^2}$$

Define

$$\omega = \frac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma_\eta^2}, \psi = \frac{\sigma_A^2}{\sigma_A^2 + \sigma_H^2}, \theta = \frac{\sigma_\mu^2 + \sigma_\eta^2}{\sigma_A^2 + \sigma_\mu^2 + \sigma_H^2 + \sigma_\eta^2}.$$

When are twins estimates less biased than OLS estimates?

Then we have

$$\begin{aligned} \rho_{OLS} &= \rho + \gamma \left[\theta \omega + (1 - \theta) \psi \right] \\ \rho_W &= \rho + \gamma \omega \end{aligned}$$

Since all coeffficents are positive, within twin pair estimates are less biased if

$$\begin{aligned} \theta \omega + (1 - \theta) \psi &> \omega \\ \psi &> \omega. \end{aligned}$$

What does this mean?

- ψ the relative variance of "ability" in the variance of the common family component of schooling
- ω the relative variance of "ability" in the variance of the idiosyncratic components of schooling

The within estimates are less biased if the relative variance of "ability" is bigger in the family component than in the idiosyncratic component.
The standard assumption leading to the within twin estimates is

$$\sigma_\mu^2 = 0, \sigma_\eta^2 > 0$$

but why would this necessarily be the case?

Sandewall, Cesarini, and Johannesson (2014)

They have data from Sweden on

- 890 male twin pairs, or 1780 men
- Two measures of schooling: self-reported and from administrative records
- An ability measure from a test given at the time of military conscription

Within twin pair estimates of returns to schooling and ability

Sandewall, Cesarini, and Johannesson (2014)

Dep.Var.	Schooling		Earnings		
Method	FE	FE	FE-IV	FE	FE-IV
	(1)	(2)	(3)	(4)	(5)
Schooling	—	0.024	0.034	0.021	0.029
		(0.008)	(0.013)	(0.008)	(0.013)
Ability	0.517	—	—	0.078	0.074
	(0.135)			(0.026)	(0.026)
Obs.	1780	1780	1780	1780	1780