Panel Data

Christopher Taber

University of Wisconsin

September 8, 2018

Lets start by reviewing the asymptotic results for OLS when the data is iid

Without worrying about every detail we assume

$$Y_i = X_i'\beta + u_i$$

where $E(X_i u_i) = 0$ and we do not have perfect multicollinearity then

$$\begin{aligned}
\widehat{\beta} &= \left(\frac{1}{N}\sum_i X_i X_i'\right)\frac{1}{N}\sum_i X_i Y_i \\
&= \beta + \left(\frac{1}{N}\sum_i X_i X_i'\right)^{-1}\frac{1}{N}\sum_i X_i u_i
\end{aligned}$$

Consistency

$$\left(\frac{1}{N}\sum_i X_i X_i'\right)^{-1} \approx \left(E\left[X_i X_i'\right]\right)^{-1}$$

$$\frac{1}{N}\sum_i X_i u_i \approx 0$$

Thus

$$\widehat{\beta} \approx \beta$$

Asymptotic Variance

Multiply by $\sqrt{N}$ then

$$\sqrt{N}\left(\widehat{\beta}-\beta\right)=\left(\frac{1}{N}\sum_i X_i X_i'\right)^{-1}\left[\frac{1}{\sqrt{N}}\sum_i X_i u_i\right]$$

The CLT on term in brackets says

$$\frac{1}{\sqrt{N}}\sum_i X_i u_i \sim N\left(0, E\left[X_i X_i' u_i^2\right]\right)$$

so

$$\sqrt{N}\left(\widehat{\beta}-\beta\right) \sim N\left(0, \left(E\left[X_i X_i'\right]\right)^{-1} E\left[X_i X_i' u_i^2\right]\left(E\left[X_i X_i'\right]\right)^{-1}\right)$$

Approximation

To take the finite approximation of this we say

$$Var\left(\sqrt{N}\left(\widehat{\beta}-\beta\right)\right) \approx \left(E\left[X_i X_i'\right]\right)^{-1} E\left[X_i X_i' u_i^2\right]\left(E\left[X_i X_i'\right]\right)^{-1}$$

so

$$\begin{aligned} Var\left(\widehat{\beta}\right) &\approx \frac{\left(E\left[X_i X_i'\right]\right)^{-1} E\left[X_i X_i' u_i^2\right]\left(E\left[X_i X_i'\right]\right)^{-1}}{N} \\ &\approx \frac{\left(\frac{1}{N}\sum_i X_i X_i'\right)^{-1}\left[\frac{1}{N}\sum_i X_i X_i' \widehat{u}_i^2\right]\left(\frac{1}{N}\sum_i X_i X_i'\right)^{-1}}{N} \\ &\approx \left(\sum_i X_i X_i'\right)^{-1}\left[\sum_i X_i X_i' \widehat{u}_i^2\right]\left(\sum_i X_i X_i'\right)^{-1} \end{aligned}$$

Panel Data

Suppose now that we have data for N units (people, countries, firms, states,schools...)

However, the complication is that for each unit we have more than one observation

Assume we have T_i observations for that unit i (typically time periods, but could be members of a family, students in a classroom...)

We analyze the the model assuming that

- N is large, so consistency occurs as N grows
- T_i is small, so as N grows, T_i stays constant

Lets start with the regression model

$$Y_{it} = X_{it}'\beta + u_{it}$$

We still maintain Assumption 2 that

$$E(u_{it}X_{it}) = 0$$

We also continue to assume independence **across individuals** but not across time within an individual.

Consistency

What happens if we just run a regression with the data this way

We will still get consistency of the model since

$$
\begin{aligned}
\widehat{\beta} &= \left(\sum_{i=1}^{N} \sum_{t=1}^{T_i} X_{it} X_{it}' \right)^{-1} \left(\sum_{i=1}^{N} \sum_{t=1}^{T_i} X_{it} Y_{it} \right) \\
&= \left(\sum_{i=1}^{N} \sum_{t=1}^{T_i} X_{it} X_{it}' \right)^{-1} \left(\sum_{i=1}^{N} \sum_{t=1}^{T_i} X_{it} \left(X_{it}' \beta + u_{it} \right) \right) \\
&= \beta + \left(\frac{\sum_{i=1}^{N} \sum_{t=1}^{T_i} X_{it} X_{it}'}{\sum_{i=1}^{N} T_i} \right)^{-1} \left(\frac{\sum_{i=1}^{N} \sum_{t=1}^{T_i} X_{it} u_{it}}{\sum_{i=1}^{N} T_i} \right) \\
&\approx \beta
\end{aligned}
$$

However, we don't buy the assumption of no serial correlation so our asymptotic variance from before is not right

In particular we still believe that error terms are uncorrelated across individuals, but not within individuals

That is the assumption that

$$cov(u_{it}, u_{jt}) = 0$$

for $j \neq i$ seems fine as in the cross section

However the idea that

$$cov(u_{it}, u_{i\tau}) = 0$$

for $t \neq \tau$ seems crazy, so the assumptions of the classical linear regression model are not satisfied

How big a problem is this?

Let's think about this for a simple example

We want to estimate the sample mean which is analogous to estimating the intercept in a regression model ($X_{it} = 1$)

$$Y_{it} = \beta_0 + u_{it}$$

Now I want to put some structure on u_{it}.

By far the most common model is the "Random Effects" Model

$$u_{it} = \theta_i + \varepsilon_{it}$$

where

- ε_{it} is i.i.d. across i and t
- θ_i is i.i.d. across i
- they are uncorrelated with each other

Note that θ_i has a nice interpretation in this model: it is the "permanent" component of the error term

It stays with an individual their whole life

ε_{it} is called the "transitory" component as it lasts just one period

Let

$$\begin{aligned}\sigma_\theta^2 &= var(\theta_i) \\ \sigma_\varepsilon^2 &= var(\varepsilon_{it})\end{aligned}$$

Notice that with these models:

- If $j \neq i$, then

$$\begin{aligned} cov(u_{it}, u_{j\tau}) &= cov(\theta_i + \varepsilon_{it}, \theta_j + \varepsilon_{j\tau}) \\ &= 0 \end{aligned}$$

  even if $\tau = t$

-

$$\begin{aligned} var(u_{it}) &= var\left(\theta_i + \varepsilon_{it}\right) \\ &= \sigma_\theta^2 + \sigma_\varepsilon^2 \end{aligned}$$

- for $t \neq \tau$,

$$\begin{aligned} cov\left(u_{it}, u_{i\tau}\right) &= cov(\theta_i + \varepsilon_{it}, \theta_i + \varepsilon_{i\tau}) \\ &= \sigma_\theta^2 \end{aligned}$$

Notice that one implication of this is that

$$cov\left(u_{it}, u_{it+1}\right) = cov\left(u_{it}, u_{it+10}\right)$$

The Var/Cov of the error terms is "block diagonal"

Consider the case in which $T_i = 2$ for all obs

$$\begin{bmatrix} \sigma_\theta^2+\sigma_\varepsilon^2 & \sigma_\theta^2 & 0 & 0 & \cdots & 0 & 0 \\ \sigma_\theta^2 & \sigma_\theta^2+\sigma_\varepsilon^2 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \sigma_\theta^2+\sigma_\varepsilon^2 & \sigma_\theta^2 & \cdots & 0 & 0 \\ 0 & 0 & \sigma_\theta^2 & \sigma_\theta^2+\sigma_\varepsilon^2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & \sigma_\theta^2+\sigma_\varepsilon^2 & \sigma_\theta^2 \\ 0 & 0 & 0 & 0 & \cdots & \sigma_\theta^2 & \sigma_\theta^2+\sigma_\varepsilon^2 \end{bmatrix}$$

Now to see why this is important I will do the following

- First calculate what you would get if you ignored the panel data aspect
- Show what you should get if you did it correctly
- Discuss methods for getting the right answer

Sticking with the $T_i = 2$ case, the estimator is:

$$\widehat{\beta}_0 = \frac{1}{2N} \sum_{i=1}^{N} \sum_{t=1}^{2} Y_{it}$$

To get standard errors we could use our estimate of the variance we derived before:

$$\begin{aligned}
\widehat{Var\left(\widehat{\beta}\right)} &= \left(\sum_{i=1}^{N}\sum_{t=1}^{2} X_{it}X_{it}'\right)^{-1}\left[\sum_{i=1}^{N}\sum_{t=1}^{2} X_{it}X_{it}'\widehat{u}_{it}^{2}\right]\left(\sum_{i=1}^{N}\sum_{t=1}^{2} X_{it}X_{it}'\right)^{-1} \\
&\approx \frac{1}{2N}\left[\sum_{i=1}^{N}\sum_{t=1}^{2}\left(\theta_i+\varepsilon_{it}\right)^2\right]\frac{1}{2N} \\
&\approx \frac{E\left(\theta_i+\varepsilon_{it}\right)^2}{2N} \\
&= \frac{\sigma_\theta^2+\sigma_\varepsilon^2}{2N}
\end{aligned}$$

It turn out this is not right

The actual variance of $\widehat{\beta}_0$ is

$$\begin{aligned}
Var\left(\widehat{\beta}_0\right) =& Var\left(\frac{1}{2N}\sum_{i=1}^{N}\sum_{t=1}^{2}Y_{it}\right)\\
=& Var\left(\frac{1}{2N}\sum_{i=1}^{N}\sum_{t=1}^{2}\left(\beta_0+\theta_i+\varepsilon_{it}\right)\right)\\
=& E\left(\left[\frac{1}{2N}\sum_{i=1}^{N}\sum_{t=1}^{2}\left(\theta_i+\varepsilon_{it}\right)\right]^2\right)\\
=& \frac{1}{4N^2}\sum_{i_1=1}^{N}\sum_{t_1=1}^{2}\sum_{i_2=1}^{N}\sum_{t_2=1}^{2}E\left[\left(\theta_{i_1}+\varepsilon_{i_1t_1}\right)\left(\theta_{i_2}+\varepsilon_{i_2t_2}\right)\right]\\
=& \frac{1}{4N^2}\sum_{i_1=1}^{N}\sum_{t_1=1}^{2}\sum_{t_2=1}^{2}E\left[\left(\theta_{i_1}+\varepsilon_{i_1t_1}\right)\left(\theta_{i_1}+\varepsilon_{i_1t_2}\right)\right]\\
&+\frac{1}{4N^2}\sum_{i_1=1}^{N}\sum_{t_1=1}^{2}\sum_{i_2\neq i_1}\sum_{t_2=1}^{2}E\left[\left(\theta_{i_1}+\varepsilon_{i_1t_1}\right)\left(\theta_{i_2}+\varepsilon_{i_2t_2}\right)\right]\\
=& \frac{1}{4N^2}\sum_{i_1=1}^{N}\left[\left(\sigma_\theta^2+\sigma_\varepsilon^2\right)+\left(\sigma_\theta^2\right)+\left(\sigma_\theta^2\right)+\left(\sigma_\theta^2+\sigma_\varepsilon^2\right)\right]\\
=& \frac{\sigma_\theta^2}{N}+\frac{\sigma_\varepsilon^2}{2N}
\end{aligned}$$

So we should get

$$\frac{\sigma_\theta^2}{N} + \frac{\sigma_\varepsilon^2}{2N}$$

But if we ignore the panel nature of the data we would get

$$\frac{\sigma_\theta^2}{2N} + \frac{\sigma_\varepsilon^2}{2N}$$

Thus what we had before was wrong-but in an intuitive way:

- If $\sigma_\theta^2 = 0$ then error term only ε_{it} which is i.i.d. and we would be fine
- If $\sigma_\varepsilon^2 = 0$ then $Y_{i1} = Y_{i2}$. We are acting as if we have 2N obervations, but we really only have N observations
- In general it will be somewhere in the middle, but we have understated the size of our standard errors

So what do we do about this? Generalizing to the regression case, if I were teaching this course 20 years ago I would have said:

1. Regress Y on X, that gives a consistent estimate of β
2. Calculate residuals from regression
3. Estimate σ_θ^2 and σ_ε^2 from residuals
4. Construct block diagonal variance/covariance matrix from these estimates
5. Run Feasible GLS

These days the solution is to “cluster” our standard errors instead

There is a very easy way to think about this to me

We have shown

$$\widehat{\beta}-\beta=\left(\sum_{i=1}^{N}\sum_{t=1}^{T_i}X_{it}X_{it}'\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T_i}X_{it}u_{it}\right)$$

Now define

$$\vartheta_i\equiv\sum_{t=1}^{T_i}X_{it}u_{it}$$

this is iid

Then we can write

$$\sqrt{N}\left(\widehat{\beta}-\beta\right)=\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T_i}X_{it}X_{it}'\right)^{-1}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\vartheta_i\right)$$

then using the Central Limit Theorem we know

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \vartheta_i \sim N\left(0, Var\left(\vartheta_i\right)\right)$$

Thus

$$\sqrt{N}\left(\widehat{\beta} - \beta\right) \sim N\left(0, \left(E\left[\sum_{t=1}^{T_i} X_{it} X_{it}\right]\right)^{-1} Var\left(\vartheta_i\right)\left(E\left[\sum_{t=1}^{T_i} X_{it} X_{it}\right]\right)^{-1}\right)$$

we approximate this as

$$
\begin{aligned}
Var\left(\widehat{\beta}\right) \approx & \frac{\left(E\left[\sum_{t=1}^{T_i} X_{it}X_{it}\right]\right)^{-1} E\left(\vartheta_i\vartheta_i'\right)\left(E\left[\sum_{t=1}^{T_i} X_{it}X_{it}\right]\right)^{-1}}{N} \\
\approx & \frac{\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T_i} X_{it}X_{it}'\right)^{-1}\left(\frac{1}{N}\sum_i\left[\sum_{t=1}^{T_i} X_{it}\widehat{u}_{it}\right]\left[\sum_{t=1}^{T_i} X_{it}'\widehat{u}_{it}\right]\right)\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T_i} X_{it}X_{it}'\right)^{-1}}{N} \\
\approx & \left(\sum_{i=1}^{N}\sum_{t=1}^{T_i} X_{it}X_{it}'\right)^{-1}\left(\sum_i\left[\sum_{t=1}^{T_i} X_{it}\widehat{u}_{it}\right]\left[\sum_{t=1}^{T_i} X_{it}'\widehat{u}_{it}\right]\right)\left(\sum_{i=1}^{N}\sum_{t=1}^{T_i} X_{it}X_{it}'\right)^{-1}
\end{aligned}
$$

This is a generalization of the heteroskedastic robust standard errors.

Rather than allowing $X_{it}u_{it}$ to have an arbitrary variance, now we are allowing $[X_{i1}u_{i1}, X_{i2}u_{i2}, ..., X_{iT_i}u_{iT_i}]$ to have an arbitrary variance/covariance matrix

We impose that the Var/Cov of the error terms is “block diagonal” but thats it

Consider the case in which $T_i = 2$ for all obs again

$$\begin{bmatrix} \sigma_{111} & \sigma_{112} & 0 & 0 & \cdots & 0 & 0 \\ \sigma_{112} & \sigma_{122} & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \sigma_{211} & \sigma_{212} & \cdots & 0 & 0 \\ 0 & 0 & \sigma_{212} & \sigma_{222} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & \sigma_{N11} & \sigma_{N12} \\ 0 & 0 & 0 & 0 & \cdots & \sigma_{N12} & \sigma_{N22} \end{bmatrix}$$

where

$$\sigma_{it\tau} = cov(u_{it}, u_{i\tau})$$

Lets verify that this gives us the right answer for the random effect case we discussed above with $T_i = 2$

Here we just have an intercept so $X_i = 1$.

So

$$\left(\sum_{i=1}^{N}\sum_{t=1}^{2} X_{it}X_{it}'\right)^{-1}\left[\sum_{i}\left[\sum_{t=1}^{2} X_{it}\widehat{u}_{it}\right]\left[\sum_{t=1}^{2} X_{it}'\widehat{u}_{it}\right]\right]\left(\sum_{i=1}^{N}\sum_{t=1}^{2} X_{it}X_{it}'\right)^{-1}$$

$$=(2N)^{-1}\left[\sum_{i}\left[u_{i1}+u_{i2}\right]\left[u_{i1}+u_{i2}\right]\right](2N)^{-1}$$

$$=\frac{1}{4N^2}\sum_{i}\left[\theta_i+\varepsilon_{i1}+\theta_i+\varepsilon_{i2}\right]\left[\theta_i+\varepsilon_{i1}+\theta_i+\varepsilon_{i2}\right]$$

$$\approx\frac{1}{4N}\left[4\sigma_\theta^2+2\sigma_\varepsilon^2\right]$$

$$=\frac{\sigma_\theta^2}{N}+\frac{\sigma_\varepsilon^2}{2N}$$

We can implement this using the cluster command in stata:

```
reg y x, cluster(i)
```

Fixed Effects

Why does anyone bother to collect panel data?

So far it just seems like a complication

Intuitively it seems like there may be some advantage

There is

Write the model as

$$Y_{it} = X_{it}'\beta + \theta_i + \varepsilon_{it}$$

In the random effects model we thought of θ_i as part of the error term so we assumed

$$E\left(\theta_i X_{it}\right) = 0$$

It turns out we don't need to assume this

We do need to assume that ε_{it} is uncorrelated with X_{it}-actually a bit stronger: assume that the vector of ε_{it} uncorrelated with the whole vector of X_{it} for each i (technically this is more than we need)

For a generic variable Z_{it} define

$$\bar{Z}_i \equiv \frac{1}{T_i} \sum_{t=1}^{T_i} Z_{it}$$

then notice that

$$\bar{Y}_i = \bar{X}_i' \beta + \theta_i + \bar{\varepsilon}_i$$

So

$$(Y_{it} - \bar{Y}_i) = (X_{it} - \bar{X}_i)' \beta + (\varepsilon_{it} - \bar{\varepsilon}_i)$$

We can get a consistent estimate of β by regressing $(Y_{it} - \bar{Y}_i)$ on $(X_{it} - \bar{X}_i)$.

The key thing is we didn't need to assume anything about the relationship between θ_i and X_i

This is numerically equivalent to putting a bunch of individual fixed effects into the model and then running the regressions

That is, let D_{it} be a $N \times 1$ vector of dummy variables so that for the j^{th}element:

$$D_{it}^{(j)} = \begin{cases} 1 & i = j \\ 0 & \text{otherwise} \end{cases}$$

and write the regression model as

$$Y_{it} = X_{it}\widehat{\beta} + D_{it}'\widehat{\delta} + \widehat{u}_{it}$$

You will get **exactly** the same $\widehat{\beta}$ running this regression as regressing $(Y_{it} - \bar{Y}_i)$ on $(X_{it} - \bar{X}_i)$

Model vs. Estimator

For me it is very important to distinguish the econometric model or data generating process from the method we use to estimate these models.

- The model is
$$Y_{it} = X_{it}\beta + \theta_i + u_{it}$$
- We can get consistent estimates of β by regressing Y_{it} on X_{it} and individual dummy variables

This is conceptually different than writing the model as

$$Y_{it} = X_{it}\beta + D_{it}'\theta + u_{it}$$

Technically they are the same thing but:

- The equation is strange because notationally the true data generating process for Y_{it} depends upon the sample
- More conceptually the model and the way we estimate them are separate issues-this mixes the two together

First Differencing

The other standard way of dealing with fixed effects is to "first difference" the data so we can write

$$Y_{it} - Y_{it-1} = \left(X_{it} - X_{it-1}\right)' \beta + \varepsilon_{it} - \varepsilon_{it-1}$$

Note that with only 2 periods this is equivalent to the standard fixed effect because

$$\begin{aligned} Y_{i2} - \bar{Y}_i &= Y_{i2} - \frac{Y_{i1} + Y_{i2}}{2} \\ &= \frac{Y_{i2} - Y_{i1}}{2} \end{aligned}$$

However, this is not the same as the regular fixed effect estimator when you have more than two periods

To see how they differ lets think about a simple "treatment effect" model with only the regressor T_{it}.

Assume that we have T periods for everyone, and that also for everyone

$$T_{it} = \begin{cases} 0 & t \leq \tau \\ 1 & t > \tau \end{cases}$$

Think of this as a new national program that begins at period $\tau + 1$

The standard fixed effect estimator is

$$\widehat{\alpha}_{FE} = \frac{scov((T_{it} - \bar{T}_i), (Y_{it} - \bar{Y}_i)}{svar(T_{it} - \bar{T}_i)}$$
$$= \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} (T_{it} - \bar{T}_i)(Y_{it} - \bar{Y}_i)}{\left(\sum_{i=1}^{N} \sum_{t=1}^{T} (T_{it} - \bar{T}_i)^2\right)}$$

Let

$$\bar{Y}_A = \frac{1}{N(T-\tau)} \sum_{i=1}^{N} \sum_{t=\tau+1}^{T} Y_{it}$$
$$\bar{Y}_B = \frac{1}{N\tau} \sum_{i=1}^{N} \sum_{t=1}^{\tau} Y_{it}$$

The numerator is

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\left(T_{it}-\frac{T-\tau}{T}\right)\left(Y_{it}-\bar{Y}_i\right)$$
$$=\sum_{i=1}^{N}\left[\sum_{t=1}^{\tau}\left(T_{it}-\frac{T-\tau}{T}\right)Y_{it}+\sum_{t=\tau+1}^{T}\left(T_{it}-\frac{T-\tau}{T}\right)Y_{it}\right]$$
$$=-\tau\left(\frac{T-\tau}{T}\right)N\bar{Y}_B+\left(T-\tau\right)\frac{\tau}{T}N\bar{Y}_A$$
$$=\tau\left(\frac{T-\tau}{T}\right)N\left[\bar{Y}_A-\bar{Y}_B\right]$$

The denominator is

$$
\begin{aligned}
&\sum_{i=1}^{N}\sum_{t=1}^{T}\left(T_{it}-\frac{T-\tau}{T}\right)^2 \\
&\quad=\sum_{i=1}^{N}\left[\sum_{t=1}^{\tau}\left(-\frac{T-\tau}{T}\right)^2+\sum_{t=\tau+1}^{T}\left(1-\frac{T-\tau}{T}\right)^2\right] \\
&\quad=N\left[\tau\frac{T-\tau}{T}\frac{T-\tau}{T}+(T-\tau)\frac{\tau}{T}\frac{\tau}{T}\right] \\
&\quad=N\left[\frac{\tau T^2-2\tau^2 T+\tau^3}{T^2}+\frac{T\tau^2-\tau^3}{T^2}\right] \\
&\quad=N\left[\frac{\tau T^2-\tau^2 T}{T^2}\right] \\
&\quad=N\tau\left[\frac{T-\tau}{T}\right]
\end{aligned}
$$

So the fixed effects estimator is just

$$\bar{Y}_A - \bar{Y}_B$$

Next consider the first differences estimator

$$\begin{aligned}
&\frac{\sum_{i=1}^{N} \sum_{t=1}^{T} \left(T_{it} - T_{it-1}\right)\left(Y_{it} - Y_{it-1}\right)}{\sum_{i=1}^{N} \sum_{t=2}^{T} \left(T_{it} - T_{it-1}\right)^2} \\
=&\frac{\sum_{i=1}^{N} \left(Y_{i\tau} - Y_{i\tau-1}\right)}{N} \\
=&\bar{Y}_{\tau} - \bar{Y}_{\tau-1}
\end{aligned}$$

Notice that you throw out all the data except right before and after the policy change.

Fixed Effects Versus Regression

Is fixed effects obviously better than regression (i.e. regressing Y_{it} on X_{it} like we talked about before)?

It kind of seems so, in regression we essentially need to assume that θ_i is uncorrelated with X_{it} but for fixed we don't, so that sounds obviously better

However its not quite that simple

Suppose that X_{it} does not vary across time at all: Gender, education, firm sector of economy, village location, etc

Then $(X_{it} - \bar{X}_i) = 0$ and we can't identify that coefficient at all

More generally we can divide data into within variance and between variance:

$$
\begin{aligned}
ss(X_{it}) &= \sum_i \sum_t \left(X_{it} - \overline{X}\right)^2 \\
&= \sum_i \sum_t \left(X_{it} - \overline{X}_i + \overline{X}_i - \overline{X}\right)^2 \\
&= \sum_i \sum_t \left(X_{it} - \overline{X}_i\right)^2 + \sum_i \sum_t \left(\overline{X}_i - \overline{X}\right)^2 \\
&\quad + \sum_i \sum_t 2\left(X_{it} - \overline{X}_i\right)\left(\overline{X}_i - \overline{X}\right)
\end{aligned}
$$

but

$$\sum_i \sum_t 2\left(X_{it}-\overline{X}_i\right)\left(\overline{X}_i-\overline{X}\right) = \sum_i 2\left(\overline{X}_i-\overline{X}\right)\sum_t \left(X_{it}-\overline{X}_i\right)$$

$$= \sum_i 2\left(\overline{X}_i-\overline{X}\right)\left[\sum_t X_{it} - T_i \frac{\sum_t X_{it}}{T_i}\right]$$

$$=0$$

Thus we can write the sample variance as

$$\sum_i \sum_t \left(X_{it}-\overline{X}_i\right)^2 + \sum_i T_i\left(\overline{X}_i-\overline{X}\right)^2$$

The first part is called “within variance” and the second is called “between variance”

The fixed effects estimator is sometimes called the within estimator because it only uses the within variance of X_i for estimation

Throwing out the between variation in X_i could be bad for two different reasons

1. It is inefficient. This is particulary apparent in the case in which X_{it} does not change in which case you can't even estimate it, but if the within variance is very small the standard errors will be very large
2. It might actually be better variation.
   That is

$$\begin{aligned}
\beta_{OLS} &= \beta + \frac{cov(X_{it}, \theta_i + \varepsilon_{it})}{var(X_{it})} \\
\beta_{FE} &= \beta + \frac{cov(X_{it} - \overline{X}_i, \varepsilon_{it} - \overline{\varepsilon}_i)}{var(X_{it} - \overline{X}_i)}
\end{aligned}$$

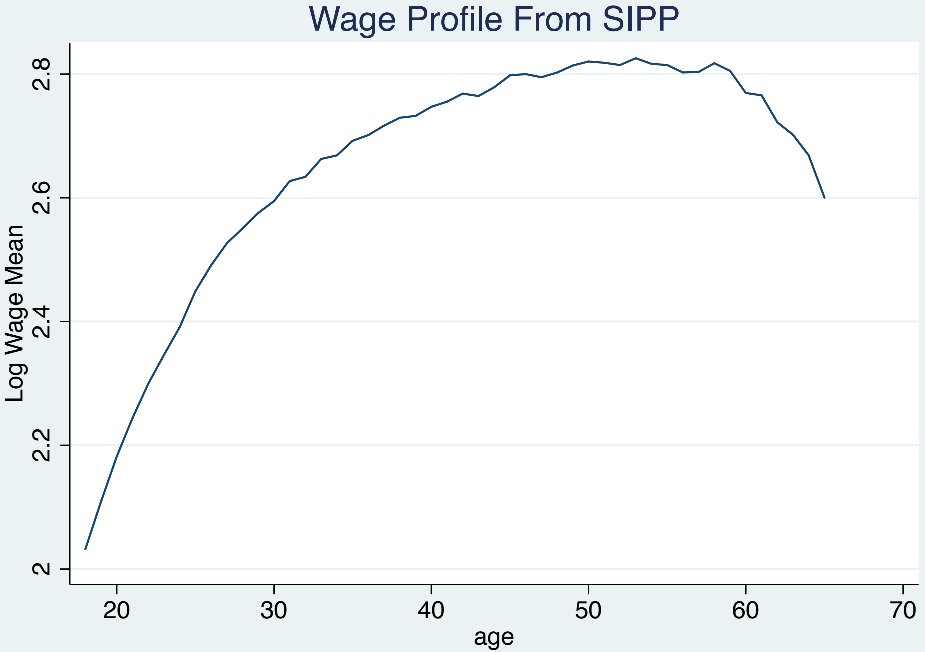
   We can't say for sure in which case the bias is worse

Lifecycle Wage Profile

Here is an example where fixed effects make sense.

We want to estimate the age profile-that is how wages vary across ages for high school men.

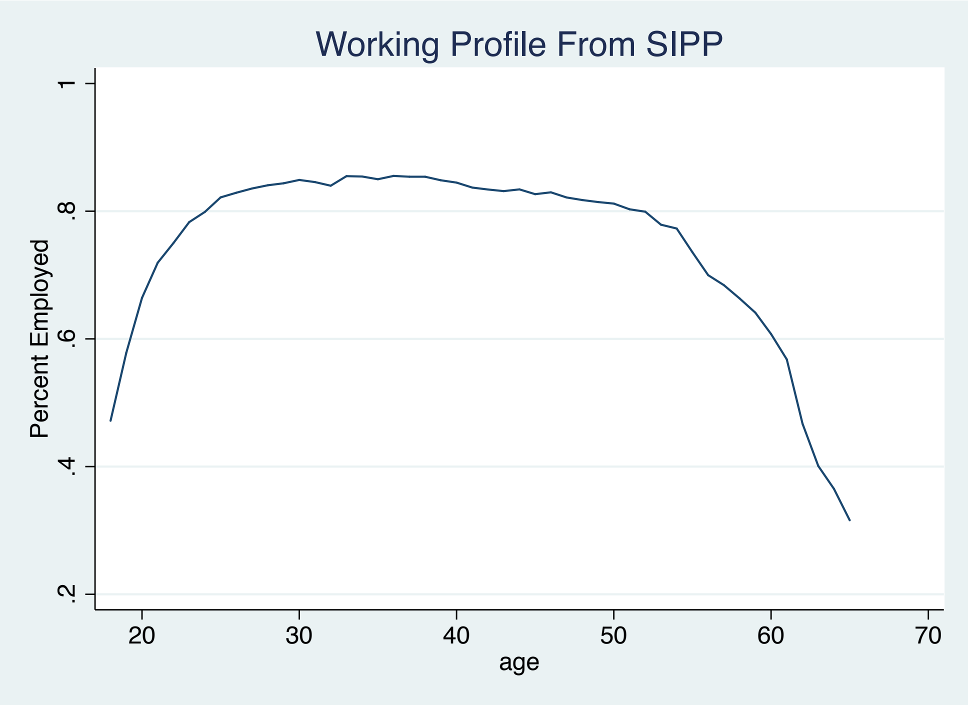
We just run a regression of log wages on age dummies using a short panel and here is what we get

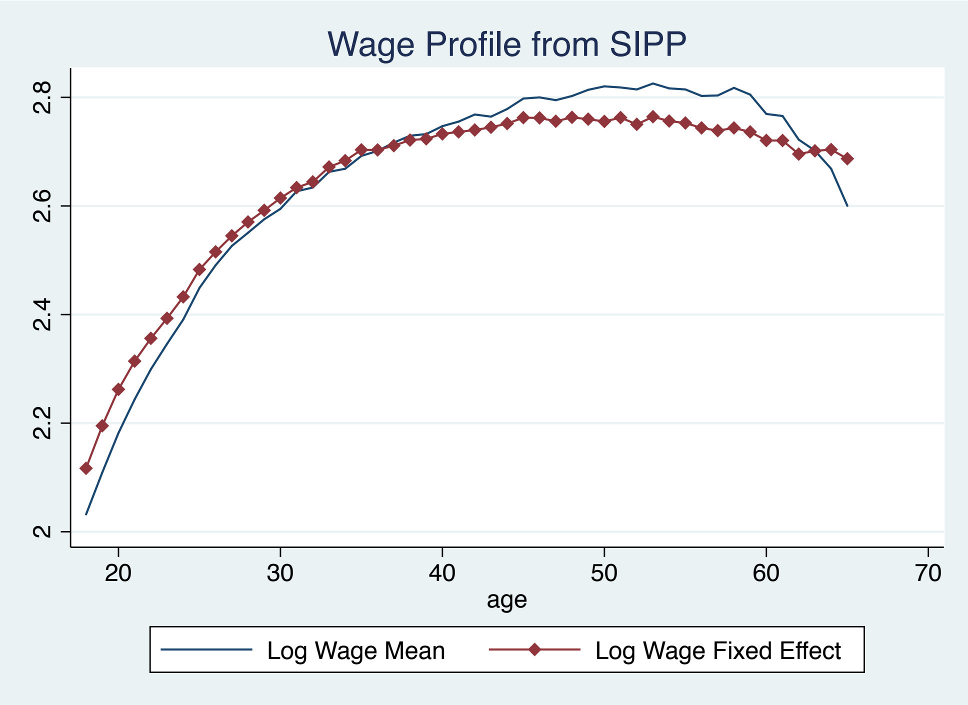

Wage Profile From SIPP
Log Wage Mean
2
2.2
2.4
2.6
2.8
20
30
40
50
60
70
age

Problems

- Cohort effects
- Selection effects

Fixed effects can help with both of these problems


Working Profile From SIPP
Percent Employed
.2
.4
.6
.8
1
20
30
40
50
60
70
age



Wage Profile from SIPP
2.8
2.6
2.4
2.2
2
20
30
40
50
60
70
age
Log Wage Mean
Log Wage Fixed Effect

Birth Weight

Almond, Chay, and Lee (QJE, 2005)

Their goal is to understand the effects of birth weight on health

$$h_{ij} = \alpha + \beta bw_{ij} + X_i'\gamma + a_i + \varepsilon_{ij}$$

where

- h_{ij} is health of newborn j of mother i
- bw_{ij} is birth weight
- a_i is mother specific effect

If we just run a regression of h_{ij} on bw_{ij} we get

$$\widehat{\beta}_{OLS} \approx \beta + \frac{cov\left(bw_{ij}, X_i'\gamma\right)}{var\left(bw_{ij}\right)} + \frac{cov\left(bw_{ij}, a_i\right)}{var\left(bw_{ij}\right)}$$

Solution: use twins

- Not always the same because often one has better access to nutrition
- They tend to be smaller to begin with so incidence of low birth weight is higher

Estimate model as

$$\Delta h_{ij} = \beta \Delta bw_{ij} + \Delta \varepsilon_{ij}$$

so assumption that $cov\left(\Delta bw_{ij}, \Delta \varepsilon_{ij}\right) = 0$ seems quite plausible.

TABLE I

Components of Variance for Birth Weight and Outcomes among Twins

Dependent variable	Mean squared error in OLS regressions				Ratio
	(1)	(2)	(3)	(4)	(4)/(3)
1989–1991 U. S. twins					
Birth weight	44.434	21.307	19.080	7.535	0.40
Mortality (1-year)	0.0356	0.0287	0.0219	0.0149	0.68
Mortality (1-day)	0.0183	0.0152	0.0102	0.0046	0.45
Mortality (28-day)	0.0283	0.0224	0.0158	0.0090	0.57
5-min. APGAR	1.9254	1.4078	1.1744	0.6510	0.55
Ventilator $\geq$30 min.	0.0370	0.0348	0.0338	0.0102	0.30
1995–2000 NY-NJ twins					
Hospital costs	14.410	—	—	2.958	—
Controls for					
Gestation length (linear)	No	Yes	—	—	
Gestation length dummies	No	No	Yes	—	
Mother fixed effects	No	No	No	Yes	

TABLE III

POOLED OLS AND TWINS FIXED EFFECTS ESTIMATES OF THE EFFECT OF BIRTH WEIGHT

Birth weight coefficient	Including congenital anomalies		Excluding congenital anomalies	
	Pooled OLS	Fixed effects	Pooled OLS	Fixed effects
Hospital costs	−29.95	−4.93	—	—
(in 2000 dollars)	(0.84)	(0.44)	—	—
	[−0.506]	[−0.083]	—	—
Adj. R^2	0.256	0.796	—	—
Sample size	44,410	44,410	—	—
Mortality, 1-year	−0.1168	−0.0222	−0.1069	−0.0082
(per 1000 births)	(0.0016)	(0.0016)	(0.0017)	(0.0012)
	[−0.412]	[−0.078]	[−0.377]	[−0.029]
Adj. R^2	0.169	0.585	0.164	0.629
Sample size	189,036	189,036	183,727	183,727
Mortality, 1-day	−0.0739	−0.0071	−0.0675	−0.0003
(per 1000 births)	(0.0015)	(0.0010)	(0.0015)	(0.0006)
	[−0.357]	[−0.034]	[−0.326]	[−0.001]
Adj. R^2	0.132	0.752	0.127	0.809
Sample size	189,036	189,036	183,727	183,727
Mortality, neonatal	−0.105	−0.0154	−0.0962	−0.0041
(per 1000 births)	(0.0016)	(0.0013)	(0.0016)	(0.0008)
	[−0.415]	[−0.061]	[−0.38]	[−0.016]
Adj. R^2	0.173	0.683	0.169	0.745
Sample size	189,036	189,036	183,727	183,727
5-min. APGAR score	0.1053	0.0117	0.1009	0.0069
(0–10 scale,	(0.0011)	(0.0012)	(0.0011)	(0.0011)
divided by 100)	[0.506]	[0.056]	[0.485]	[0.033]
Adj. R^2	0.255	0.663	0.248	0.673
Sample size	159,070	159,070	154,449	154,449
Ventilator incidence	−0.0837	−0.0039	−0.081	−0.002
(per 1000 births)	(0.0015)	(0.0017)	(0.0015)	(0.0016)
	[−0.228]	[−0.011]	[−0.221]	[−0.005]
Adj. R^2	0.052	0.706	0.05	0.716
Sample size	189,036	189,036	183,727	183,727
Ventilator ≥30 min.	−0.0724	0.0006	−0.0701	0.0016
(per 1000 births)	(0.0013)	(0.0013)	(0.0014)	(0.0012)
	[−0.252]	[0.002]	[−0.244]	[0.006]
Adj. R^2	0.063	0.724	0.062	0.739
Sample size	189,036	189,036	183,727	183,727

To look at things beyond linear case they divide the data into $K = 200$ different bins with

$$D_{ij}^{k} = 1\left(a_k \leq bw_{ij} < a_{k+1}\right)$$

Then use the fixed effect regression model

$$h_{ij} = \alpha + \sum_{k=1}^{K} \beta_k D_{ij}^{k} + X_i' \gamma + a_i + \varepsilon_{ij}$$

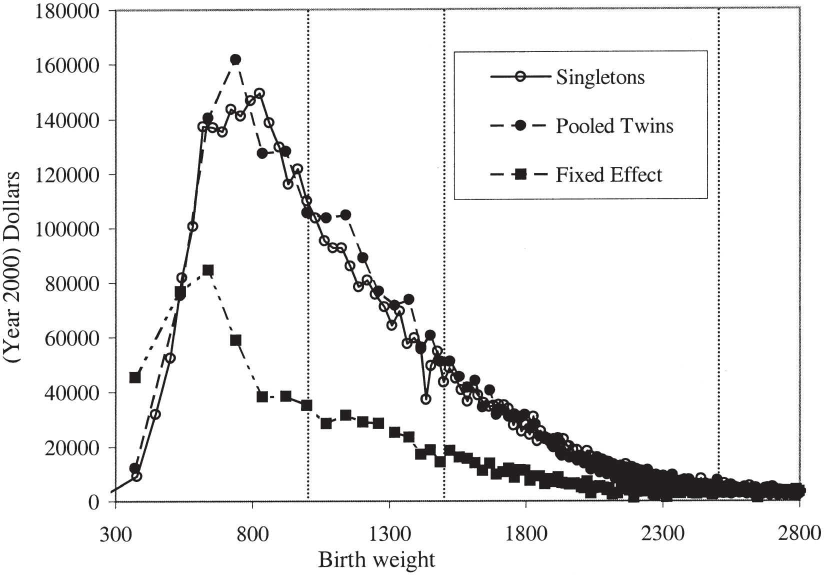


FIGURE Ia
Hospital Costs and Birth Weight
Note: 1995–2000 NY/NJ Hospital Discharge Microdata.

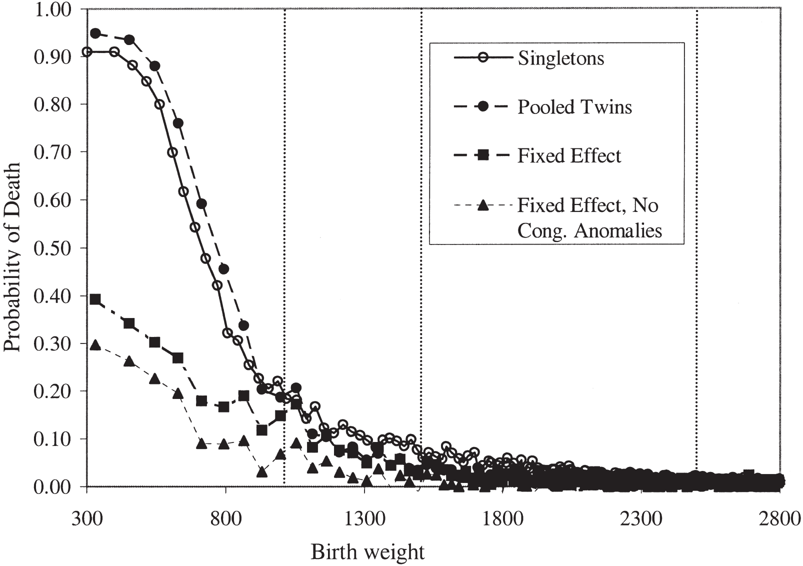


FIGURE Ib

Infant Mortality (1-year) and Birth Weight

Note: Linked Birth-Death certificate data, 1989.

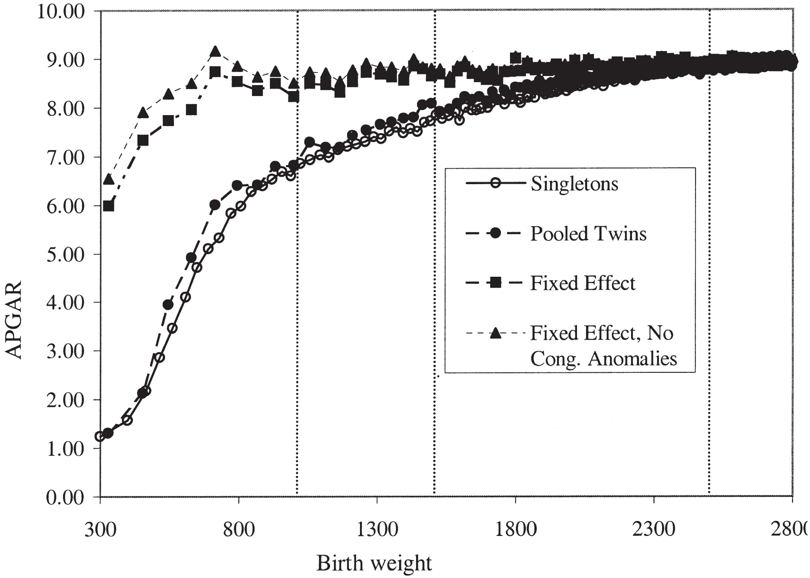


FIGURE IIa
Five-minute APGAR Score and Birth Weight
Note: Linked Birth-Death certificate data, 1989.

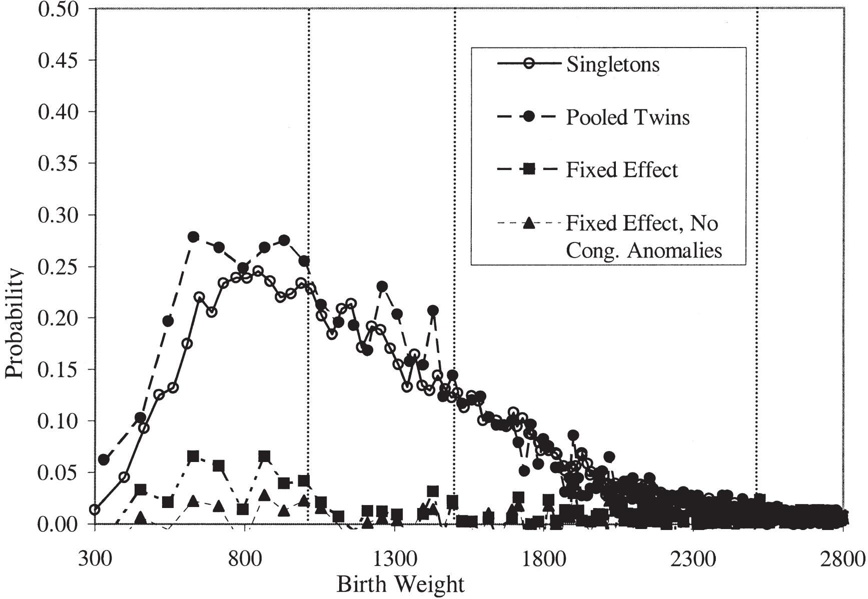


FIGURE IIb
Assisted Ventilation (30 minutes or more) and Birth Weight
Note: Linked Birth-Death certificate data, 1989.