Panel Data - Part I

Cross-Section: One observation on many individuals

Time-Series: Many observations on one individual.

Panel: Many observations on many individuals.

What's so special about panel data?

(1) can control for <u>fixed</u> observable and unobservable characteristics of an individual.

⇒ reduces (eliminates?) <u>OVB</u> but can exacerbate <u>attenuation bias</u> by reducing signal even though reduces noise (i.e. throw the baby (signal) out with the bath water (noise, OVB)).

(2) Difference-in-Difference: simple, easy, can do with panel data.

(3) Errors: not only are we worried about errors being correlated <u>across</u> individuals, we're now really worried about errors being correlated <u>within</u> an individual

over time (i.e., serially correlated).
$\Rightarrow$ CLUSTERING! (also "random effects," but clustering is better).

What does panel data look like?

i	t	y	$X1$	D
1	1991	10	5	0
1	1992	10	6	0
1	1993	20	6	1
1	1994	20	7	1
2	1991	10	7	0
2	1992	12	7	0
2	1993	11	7	0
2	1994	11	7	0

What is fixed effects? Add indiv. fixed-effects.

i	$\bar{t}$	$\bar{y}$	$\overline{X1}$	$\bar{D}$
1	1992.5	15	6	.5
2	1992.5	11	7	0

i	$t-\bar{t}$	$y-\bar{y}$	$X1-\bar{X}$	$D-\bar{D}$
1	-1.5	-5	-1	-.5
1	-.5	-5	0	-.5
1	.5	5	0	.5
1	1.5	5	1	.5
2	-1.5	-1	0	0
2	-.5	1	0	0
2	.5	0	0	0
2	1.5	0	0	0

Now let's also add in time fixed-effects,

$\overline{(t-\bar{t})}$	$\overline{(y-\bar{y})}$	$\overline{(X1-\overline{X1})}$	$\overline{(D-\bar{D})}$
-1.5	-3	-0.5	-.25
-.5	-2	0	-.25
.5	2.5	0	.25
1.5	2.5	0.5	.25

i	$(t-\bar{t})-\overline{(t-\bar{t})}$	$(y-\bar{y})-\overline{(y-\bar{y})}$	$(X1-\overline{X1})-\overline{(X1-\overline{X1})}$	$(D-\bar{D})-\overline{(D-\bar{D})}$
1	0	-2	-0.5	-.25
1	0	-3	0	-.25
1	0	2.5	0	.25
1	0	2.5	0.5	.25
2	0	2	0.5	.25
2	0	3	0	.25
2	0	-2.5	0	-.25
2	0	-2.5	-0.5	-.25

There is no longer a time effect. We ~~got~~ rid of it!
~ we "differenced" out any common effects due to time.

BIG ASSUMPTION: Both i's should be experiencing the same time effects over the years.

i.e. (1) We should all be experiencing the same wage growth over time. (Probably).

(2) We should all be experiencing the same absolute increase in wages (\$10,000) over time (no!).

Be careful about your use of fixed effects! What assumptions are you making?

$$y_{it} = \alpha + \beta x_{it} + \rho D_{it} + \underbrace{\eta_{it}}_{\alpha_i + \varepsilon_{it}}$$

Assumptions:

① $E(D_{it}\,\eta_{it}) \neq 0$ because $E(D_{it}\,a_i) \neq 0$

② $E(D_{it}\,\varepsilon_{it}) = 0$

So control for this "fixed-effect" and OVB disappears:

$$y_i = \alpha_i + \beta x_{it} + \rho D_{it} + \varepsilon_{it}$$

Aside: What is random effects?

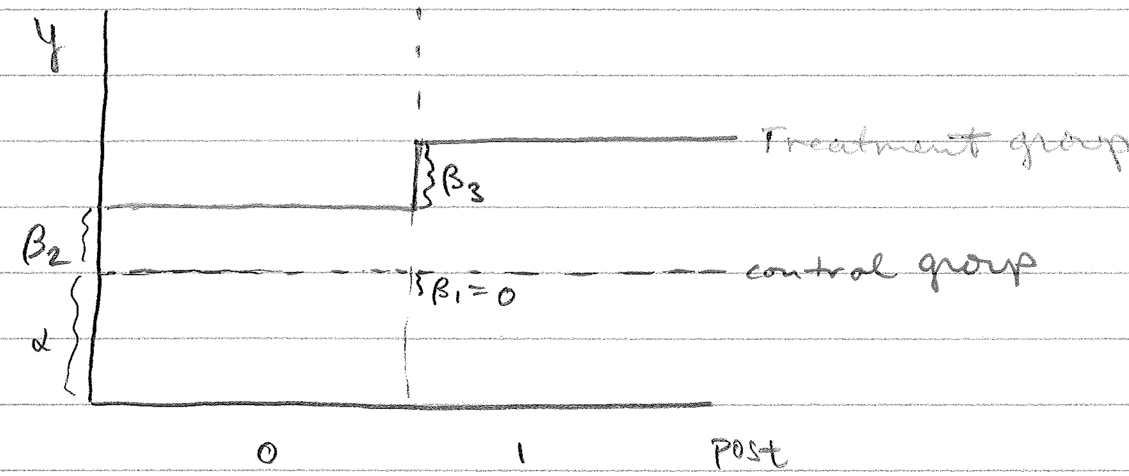
Assumptions:

① $E(D_{it}\,\eta_{it}) = 0$ because $E(D_{it}\,a_i) = 0$.

If a_i is "random", then $E(D_{it}\,a_i) = 0$. Why include random effects? Might allow for more efficient estimation, i.e., "better" standard errors, but it's best to just cluster standard errors.

---

(2) Difference-in-Difference:



$$y_{it} = \alpha + \beta_1 post_t + \beta_2 Treat_i + \underbrace{\beta_3}_{\text{effect of Treatment.}} Treat_i \times Post_t + \varepsilon_{it}$$

- Note that $\beta_2 Treat_i$ acts like a fixed-effect for the treatment group.

- Note that $\beta_1 Post_t$ acts like a time fixed-effect.