9. Panel data methods

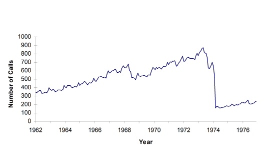
LPO.7870: Research Design and Data Analysis II

Sean P. Corcoran

Last time

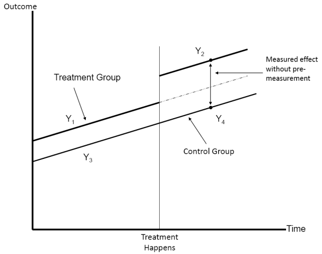
Time series and difference-in-differences research designs:

- Data structures: cross-section vs. time series vs. panel.
- Time series data can be used for *forecasting*.
- Sometimes time series data can be used to estimate the impact of a policy change or other event (*interrupted time series*).
  - Conditions under which an ITS can make strong claims to causal inference, and threats to internal validity.



Last time

Using difference-in-differences to estimate the impact of a policy change or other event:



Critical role of the *parallel trends assumption*.

Tonight

More adventures in panel data!

- A few techniques for working with panel data in Stata.
- The value of panel data for addressing (certain types of) omitted variables bias.
- Fixed effects regressions
- Fixed effects regression with time effects

Example publications: Jack et al. (2023) on COVID school closures and Hillman et al. (2015) on performance funding in higher education.

Tonight's sample datasets

We will refer to three datasets tonight (all on Github):

1. `Census_states_1970_2000.dta`: a few state-level Census variables, 1970-2000.
2. `fatality.dta`: state-by-year data on traffic fatalities, drinking and driving laws, etc. From S&W chapter 10.
3. `Texas_elementary_panel_2004_2007.dta`: school-level data from Texas.

You can follow along in the .do file for this lecture.

Working with panel data

Panel data

Panel or **longitudinal data** includes many cross-sectional units (i) observed at multiple points in time (t).

state_name	year	expp_co2	personal_i~c
Alaska	2005	12606.64	36911
Alaska	2006	13115.29	38951
Alaska	2007	15148.9	41316
Alabama	2005	8439.994	29681
Alabama	2006	8936.927	31208
Alabama	2007	9491.025	32528
Arkansas	2005	8821.27	27858
Arkansas	2006	8929.807	29385
Arkansas	2007	8977.985	31353
Arizona	2005	7156.699	32223
Arizona	2006	7785.889	34326
Arizona	2007	7995.629	35441
California	2005	9118.798	38731
California	2006	9526.36	41518
California	2007	10042.94	43211
Colorado	2005	8970.674	38795
Colorado	2006	8817.396	41181
Colorado	2007	9469.688	42724
Connecticut	2005	14359.31	48134

Panel data

Cross sectional units may be students, teachers, schools, hospitals, neighborhoods, counties, states, etc.

- A panel has N cross-sectional units and T time periods ($T \geq 2$)
- A **balanced panel** has exactly $N \times T$ observations (T time observations for all N units)
- An **unbalanced panel** has fewer than T observations for some units
- Examples: National Longitudinal Survey of Youth (NLSY), NELS-88, Beginning Teacher Longitudinal Survey (BTLS)

A *pooled cross-section* is like a panel but has different cross-sectional units in every time period (e.g., Current Population Survey, NAEP).

Panel data - long

Panel data in *long* format, N students in $T = 4$ years:

studentID	year	readscore	mathscore	incomecat	...
1	1999	75	82	3	
1	2000	78	84	4	
1	2001	80	90	4	
1	2002	78	91	3	
2	1999	91	92	2	
2	2000	94	92	2	
2	2001	80	85	2	
2	2002	87	83	2	
3	1999	62	50	5	
3	2000	70	47	5	
3	2001	75	55	4	
3	2002	73	60	5	

Panel data - wide

Panel data in *wide* format, N students in $T = 4$ years:

studentID	read99	math99	inc99	read00	math00	inc00	read01	...
1	75	82	3	78	84	4	80	
2	91	92	2	94	92	2	80	
3	62	50	5	70	47	5	75	
4	...	...	...					

Many longitudinal datasets in education research come in “wide” format, but regression analysis typically requires it be in “long format.” Stata has a `reshape` command that can help you move back and forth.

Stata panel commands

Stata has many useful commands specifically designed for panel data. (They begin with `xt`). These require that you first tell Stata that the data are a panel using `xtset`:

- `xtset` *`panelvar timevar`*
- The *panelvar* must be numeric. If it is not, you can use encode: **`encode`** *`panelvar`*, `gen(`*`panelvar2`*`)`
- It is possible to tell Stata in the `xtset` options what units of time the data represent—e.g., years, quarters, minutes (useful for some purposes—I don't usually do this)
- `xtset` alone will report back the panel settings

Why use panel data?

Panel data offers researchers a number of advantages:

- Can help us answer questions not possible with a strictly cross-sectional or time-series design.
- Can generate *measures* not possible with cross-sectional or time series data (e.g., growth, work/teaching spells)
  - If 75% of women are working in year *t*, does this reflect 75% of women working at any given point, or 75% of women who work all the time?
- Allows us to address omitted variables bias that would exist in a cross-sectional research design. ← We will focus on this.

Example: vehicle fatalities

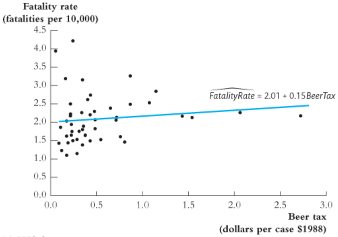
The `fatality.dta` dataset is a panel with state × year observations spanning 1982-1987. It can be used to estimate the effects of state policies related to drinking and driving on traffic fatalities (S&W ch. 10)

- Is this data in long or wide format? How do you know?
- How many cross-sectional units are there? How many time periods?
- Is the panel balanced or unbalanced?

The key outcome variable here is *mrall*. Multiply it by 10,000 to get the vehicle fatality rate per 10,000 persons. We will begin by looking at the relationship between the state tax on beer and fatality rates (*beertax*).

Example: vehicle fatalities

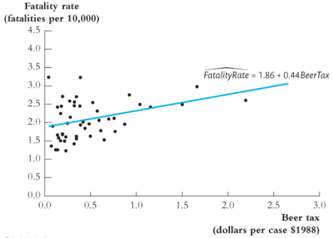
Cross-sectional analysis of fatality rate and beer tax in 1982:



(a) 1982 data

Example: vehicle fatalities

Cross-sectional analysis of fatality rate and beer tax in 1988:



(b) 1988 data

Does this association make sense to you? Why or why not?

Example: vehicle fatalities

The estimate of the slope β_1 is an unbiased estimate of the causal effect of the beer tax on fatalities **only if** there is no omitted variable bias ($E(u|X) = 0$).

What concerns do you have about omitted variables in this context?

Panel data with two time periods

Omitted variables bias revisited

Interpretation of regression coefficients as causal is often complicated by omitted variables bias. Example:

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

with $E(u_i|X_i) \neq 0$ because we believe there are omitted factors correlated with both X_i and Y_i. We can attempt to control for other factors in a multiple regression, but this may not be sufficient.

In practice, we are often concerned about omitted variables that we <u>cannot measure</u> or <u>cannot observe</u> ("unobservables").

Unobservables

Unobserved factors

Suppose there is an **unobserved** factor associated with each unit (C_i) that affects the outcome and is correlated with the explanatory variable of interest (X_i):

$$Y_i = \beta_0 + \beta_1 X_i + C_i + u_i$$

C_i could represent the effects of ability, health, motivation, intelligence, parental resources, managerial quality, organizational culture, state/local policies or regulations, etc. Sometimes called "unobserved heterogeneity."

Important: these are factors that do not change (or change slowly) over time.

Two period "before and after" model

Suppose we have two time periods (T=2) for each cross-sectional unit *i*, and assume the model above applies in both periods:

$$Y_{i2} = \beta_0 + \beta_1 X_{i2} + C_i + u_{i2}$$
$$Y_{i1} = \beta_0 + \beta_1 X_{i1} + C_i + u_{i1}$$

Now subtract period 1 from period 2 for the "first difference":

$$\Delta Y_i = \beta_1 \Delta X_i + \Delta u_i$$

The Δs are the before-to-after <u>change</u>.

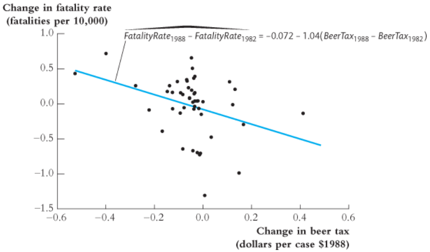
Two period "before and after" model

$$\Delta Y_i = \beta_1 \Delta X_i + \Delta u_i$$

- Because C_i does not change over time, it differences out.
- By focusing on *changes* in Y over time, we hold constant unobservables (C_i) that do not change over time.
- Notice the intercept β_0 also differences out. The presumption is that Y would not otherwise change if X does change. If we want to allow for this possibility, we can include the intercept.

Example: vehicle fatalities

Regress *change* in fatality rate (1982-1988) on *change* in beer tax:



Does this association make more sense?

Example: vehicle fatalities

A few notes on the above:

- The effect size implied here (-1.04) is quite meaningful.
- Assumption: there are unobserved factors (C_i) specific to each state that do not change over time. Because they do not change, they do not contribute to any change in traffic fatality rates over time.
- Might we also be concerned about omitted variables here? If so, what type?

The `fatality.dta` panel includes more years than just 1982 and 1988. It seems we might want to take advantage of them!

Fixed effects regression

Fixed effects regression

Let's imagine there is a slope associated with our unobserved factor C_i:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \beta_2 C_i + u_{it}$$

Because C_i does not change over time, there are effectively n intercepts, one for each cross sectional unit:

$$\alpha_i = \beta_0 + \beta_2 C_i$$

The **fixed effects regression** can be written as:

$$Y_{it} = \beta_1 X_{it} + \alpha_i + u_{it}$$

The intercepts **absorb** the influences of all omitted variables that vary across i but are constant over time (at least in your study).

Fixed effects regression

In the traffic fatalities example, the α_i represent state-specific intercepts. An alternative way of writing the fixed effects regression is:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \gamma_2 D2_i + \gamma_3 D3_i + ... + \gamma_n Dn_i + u_{it}$$

The $D2_i, ..., Dn_i$ are dummy variables indicating the different cross-sectional units (e.g., states). Need to (arbitrarily) exclude one dummy variable (e.g., $D1$) to avoid perfect multicollinearity.

Fixed effects regression

A few notes on the above:

- While every unit (e.g., state) has its own intercept, we are estimating a common slope β_1 for X.
- The separate intercepts can be thought of as the "effect" for that unit (e.g., the "effect" of being in Tennessee). This is why they're called fixed effects!
- As with any regression, you can control for other explanatory variables (X_2, X_3, etc.).

Estimating in Stata

We are estimating the intercept β_0, slope β_1, and $(N-1)$ intercepts, the "fixed effects." This can be done by manually including $N-1$ dummy variables in the regression. For example:

- reg *mrall beertax* i.*state*
- areg is equivalent but hides the $(N-1)$ intercepts (easier to read)
- areg *mrall beertax*, absorb(*state*)

Example: vehicle fatalities

Fixed effects regression for vehicle fatality rates:

```
. areg mrall beertax, absorb(state)

Linear regression, absorbing indicators         Number of obs     =        336
Absorbed variable: state                        No. of categories =         48
                                                F(   1,     287)  =      12.19
                                                Prob > F          =     0.0006
                                                R-squared         =     0.9050
                                                Adj R-squared     =     0.8891
                                                Root MSE          =     0.1899

------------------------------------------------------------------------------
       mrall |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     beertax |  -.6558736     .18785    -3.49   0.001    -1.025612   -.2861352
       _cons |   2.377075   .0969699    24.51   0.000     2.186212    2.567937
------------------------------------------------------------------------------
F test of absorbed indicators: F(47, 287) = 52.179            Prob > F = 0.000
```

More on fixed effects regression

A few more important points:

- Fixed effects regression mostly relies on the OLS assumptions that you saw earlier, including no omitted variables correlated with X and Y.

- In a panel context, this is expanded a bit: there can be no correlation between an omitted variable *in any time period* and X *in any time period* for each cross-sectional unit. This can be a concern if we think something unobservable happened in time $t-1$ that led to change in X at period t. Examples?

- If $T = 2$ the "before and after" (difference) regression and the fixed effects regression will give you the same answer.

One possible omitted variable in a panel context: time trends.

Regression with time fixed effects

Time fixed effects

Let's ignore C_i for a moment and suppose there is a common time factor S_t that does not vary across i:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \beta_2 S_t + u_{it}$$

Because S_t does not vary across the i, there are effectively T intercepts, one for each time period:

$$\alpha_t = \beta_0 + \beta_2 S_t$$

The **time fixed effects regression** can be written as:

$$Y_{it} = \beta_1 X_{it} + \lambda_t + u_{it}$$

The λs represent the "effect" of being in time t.

Combining unit and time fixed effects

We can estimate a regression with both unit and time fixed effects:

$$Y_{it} = \beta_1 X_{it} + \alpha_i + \lambda_t + u_{it}$$

This eliminates omitted variables bias from both unobserved variables that are constant over time (C_i) and from unobserved variables that change over time but are common to all units (S_t).

Example: vehicle fatalities

Fixed effects regression—with time effects—for vehicle fatality rates:

```
. areg mrall beertax i.year, absorb(state)

Linear regression, absorbing indicators         Number of obs     =        336
Absorbed variable: state                        No. of categories =         48
                                                F(   7,     281)  =       3.50
                                                Prob > F          =     0.0013
                                                R-squared         =     0.9089
                                                Adj R-squared     =     0.8914
                                                Root MSE          =     0.1879

------------------------------------------------------------------------------
       mrall |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     beertax |  -.6399799   .1973768    -3.24   0.001    -1.028505   -.2514551
             |
        year |
       1983  |  -.0799029   .0383537    -2.08   0.038       -.1554   -.0044058
       1984  |  -.0724206   .0383517    -1.89   0.060    -.1479136    .0030725
       1985  |  -.1239763   .0384418    -3.23   0.001    -.1996468   -.0483058
       1986  |  -.0378645   .0385879    -0.98   0.327    -.1138225    .0380936
       1987  |  -.0509021   .0389737    -1.31   0.193    -.1276196    .0258155
       1988  |  -.0518038   .0396235    -1.31   0.192    -.1298003    .0261927
             |
       _cons |    2.42847   .1081198    22.46   0.000     2.215643    2.641298
-------------+----------------------------------------------------------------
F test of absorbed indicators: F(47, 281) = 53.193            Prob > F = 0.000
```


Example: vehicle fatalities

Following Stock & Watson chapter 10, we can add additional explanatory variables to this regression. See the .do file for this lecture.

Example: class size and passing rates

An example using *Texas_elementary_panel_2004_2007.dta*. These data are school $\times$ year in Texas.

- *ca311tar* is the average passing rate—rename this *avgpassing*
- Use this command to calculate the average class size in a school: `egen avgclass=rowmean(cpctg01a-cpctgmea)`.
- Estimate an OLS regression of the average passing rate on average class size **for 2007 only** and interpret. Does the slope β_1 make sense to you?
- Now estimate a fixed effects regression with school fixed effects (use `areg`) and interpret.
- Finally, include time fixed effects (continue to use `areg`) and interpret.

Fixed effects: last words of advice

Fixed effects models rely on *changes over time within units*.

- Obviously, we need changes over time for this to work! Can't estimate fixed effects models for factors that don't change.
- External validity: since the model depends entirely on changes over time, are the changes (and units that experience changes) generalizable?
  - Example: school switchers (e.g., charter school effects). Results may only generalize to students who switch schools.
- Cross-period correlation: are there unobserved factors in one time period (e.g., $t-1$) that produce changes in X in another period (e.g., t)?
  - Example: was there a precipitating event that led a family to switch from a public to a charter school?

Publication examples

Jack et al., (2023)

Pandemic Schooling Mode and Student Test Scores: Evidence from US School Districts

- Data: district level test scores (passing rates) for 2016-2019 and 2021, plus district demographic data and other county-level controls.
- Instruction mode from the COVID-19 School Data Hub (public): schooling mode classified as in person, remote, or hybrid.

The authors are interested in the effects of schooling mode (X) on passing rates (Y). What would be problematic about estimating a cross-sectional regression with these data for 2021? What might the advantages of a panel be?

Jack et al., (2023)

Table 2—Pairwise Correlations between In-Person Learning on District Demographic and Pandemic Variables

	Correlation (no fixed effects)		Correlation (state fixed effects)		Correlation (commute zone fixed effects)	
Previous pass rate	0.440	(0.066)	0.611	(0.062)	0.598	(0.053)
Share Black	−0.465	(0.039)	−0.752	(0.043)	−0.757	(0.041)
Share Hispanic	−0.442	(0.067)	−0.328	(0.063)	−0.296	(0.061)
Share FRPL	−0.160	(0.048)	−0.255	(0.048)	−0.365	(0.046)
Share ELL	−1.290	(0.121)	−0.879	(0.104)	−0.764	(0.099)
Avg. case rate	0.803	(0.199)	0.367	(0.107)	0.115	(0.051)
Repub. vote share	0.010	(0.000)	0.010	(0.000)	0.009	(0.001)

Notes: This table shows the pairwise correlations of the share of days in person during the 2020–2021 school year with district demographic and pandemic characteristics. We present the correlations of the sample overall, without fixed effects included ("no fixed effects"), with state-year fixed effects ("state fixed effects"), and with commuting zone fixed effects ("commute zone fixed effects"). The share in person measures the share of time during the 2020–2021 school year that the district offered full-time in-person instruction (rather than hybrid or virtual instruction). "Previous pass rate" represents the average pass rate on state

Jack et al., (2023)

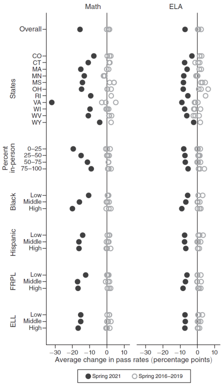


Figure 1. Average Change in Pass Rates on State Standardized Assessments in Spring 2021

Jack et al., (2023)

They use a fixed effects regression:

$$pass_{ict} = \alpha + \beta_1(\%inperson_{it}) + \beta_2(\%hybrid_{it}) + \gamma_{ct} + \delta_t + \nu_i + \Pi X_{ict} + u_{ict}$$

- i is district, t is year, and c is county
- $\%inperson$ and $\%hybrid$ are the percentage of time the district spend in person and in hybrid modes
- δ_t are the time fixed effects
- ν_i are the district fixed effects
- The γ_{ct} are effects for county specific trends

Jack et al., (2023)

Table 3—Schooling Mode and Changes in Pass Rates

	Math			ELA		
	(1)	(2)	(3)	(4)	(5)	(6)
	Pass rate	Pass rate	Pass rate	Pass rate	Pass rate	Pass rate
Panel A. Main specifications						
% in-person	0.140	0.134	0.128	0.0813	0.0828	0.0872
	(0.0137)	(0.0147)	(0.0156)	(0.0102)	(0.0105)	(0.0105)
% hybrid	0.0776	0.0722	0.0743	0.0608	0.0537	0.0637
	(0.0143)	(0.0148)	(0.0161)	(0.0116)	(0.00949)	(0.00994)
Observations	11,041	11,041	11,041	11,064	11,064	11,064
Commute zone × year	No	Yes	No	No	Yes	No
County × year	No	No	Yes	No	No	Yes
Panel B. Demographic interactions						
% in-person × 2021	0.0960	0.156	0.0872	0.0686	0.0784	0.0729
	(0.0174)	(0.0196)	(0.0388)	(0.0138)	(0.0123)	(0.0276)
% hybrid × 2021	0.0379	0.0907	0.0280	0.0381	0.0409	0.0360
	(0.0169)	(0.0205)	(0.0388)	(0.0129)	(0.0123)	(0.0279)
% Black × % in-person × 2021	0.0943			0.0193		
	(0.0398)			(0.0240)		
% Black × % hybrid × 2021	0.0855			0.0508		
	(0.0472)			(0.0279)		
% Hispanic × % in-person × 2021		−0.135			−0.0178	
		(0.0680)			(0.0482)	
% Hispanic × % hybrid × 2021		−0.0664			0.0247	
		(0.0734)			(0.0421)	
% FRPL × % in-person × 2021			0.0810			0.000259
			(0.0582)			(0.0371)
% FRPL × % hybrid × 2021			0.0689			0.0153
			(0.0605)			(0.0380)
Observations	11,041	11,041	9,620	11,064	11,064	9,643
Commute zone × year	Yes	Yes	Yes	Yes	Yes	Yes

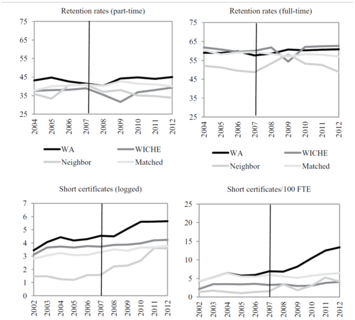
Notes: This table shows the relationship between district in-person share, hybrid share, and pass rates in math and ELA on state standardized assessments for students in grades 3–8. Virtual share is the reference group. In panel A, we present our results for state-year fixed effects in columns 1 and 4 for math and ELA, respectively, for commuting zone–year fixed effects in columns 2 and 5, and for county-year fixed effects in columns 3 and 6. All regressions are weighted by district enrollment and include district fixed effects, year fixed effects, state-year fixed effects, demographic controls (race/ethnicity shares, share of students eligible for FRPL, and share of ELLs), county-level

Hillman et al., (2015)

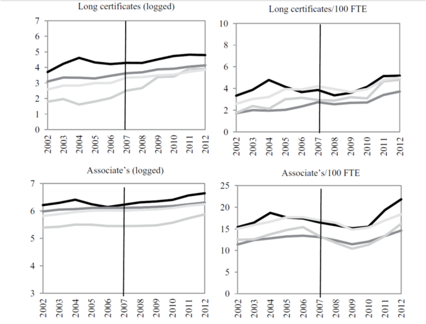
Evaluating the Impact of “New” Performance Funding in Higher Education

- Evaluates the Student Achievement Initiative (SAI) in Washington: 2007 statewide performance accountability system to improve retention and degree attainment in community colleges. Institutions were given financial support for degrees earned.
- Uses IPEDS 2002-2012 to estimate effects on community college retention, certificates/degrees earned
- Multiple comparison groups: (1) CCs in western states, (2) those in border states, (3) those in non-performance funding states.
- Institution fixed effects models

Hillman et al., (2015)



Hillman et al., (2015)



Hillman et al., (2015)

TABLE 3
Regression Estimates for SAI's Impact on Part-Time and Full-Time Retention Rates

	Part-time			Full-time		
	WICHE	Neighbors	Matched	WICHE	Neighbors	Matched
a. Main model						
Treat × Post	−0.49 (1.93)	−4.07 (2.43)	−2.22 (1.44)	−0.73 (1.65)	−1.66 (2.37)	−0.78 (1.16)
b. Year interactions						
Treat × Post × 2007	−2.78*** (0.82)	−5.87* (2.52)	−3.61*** (1.05)	−1.08 (0.86)	0.07 (0.94)	−2.15** (0.69)
Treat × Post × 2008	−1.18 (1.47)	−8.68** (3.08)	−2.85 (2.08)	−2.98** (1.03)	−8.35*** (1.85)	−2.27* (1.06)
Treat × Post × 2009	6.62*** (1.55)	−4.60 (3.04)	−2.07 (1.90)	6.80*** (1.26)	−8.85*** (2.07)	1.42 (1.24)
Treat × Post × 2010	1.06 (1.61)	−1.48 (2.99)	−0.71 (1.85)	−2.12 (1.25)	−1.99 (1.72)	0.39 (1.35)
Treat × Post × 2011	−0.48 (1.59)	−1.68 (3.32)	−0.96 (1.97)	−1.89 (1.24)	0.85 (2.17)	0.83 (1.52)
Treat × Post × 2012	−1.01 (1.56)	0.36 (3.45)	1.41 (2.10)	−1.96 (1.21)	4.65** (1.79)	1.76 (1.58)
Observations	1,854	441	855	1,854	441	855
Institutions	206	49	95	206	49	95
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Institution fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Main model, R^2	.47	.35	.47	.61	.50	.50
Year interactions model, R^2	.48	.36	.47	.62	.52	.50

Hillman et al., (2015)

TABLE 4
Regression Estimates for SAI's Impact on Short Certificates Awarded

	Short-term certificates (logged)			Short-term certificates (per 100 FTE)		
	WICHE	Neighbors	Matched	WICHE	Neighbors	Matched
a. Main model						
Treat × Post	0.33 (0.17)	−0.29 (0.27)	0.12 (0.17)	2.40** (0.91)	0.17 (0.84)	1.52* (0.65)
b. Year interactions						
Treat × Post × 2007	0.28 (0.18)	0.15 (0.27)	0.10 (0.13)	0.86* (0.36)	−0.26 (0.65)	0.56 (0.33)
Treat × Post × 2008	−0.03 (0.22)	−0.55 (0.33)	−0.43** (0.15)	0.72 (0.47)	−1.80 (1.16)	0.39 (0.42)
Treat × Post × 2009	0.27 (0.24)	−0.31 (0.33)	0.08 (0.16)	1.58** (0.53)	−0.94 (1.09)	0.87* (0.42)
Treat × Post × 2010	0.48* (0.23)	−0.31 (0.30)	0.26 (0.16)	3.12*** (0.51)	−0.03 (0.82)	1.73*** (0.43)
Treat × Post × 2011	0.45* (0.23)	−0.80* (0.31)	0.36* (0.16)	5.06*** (0.49)	1.13 (0.85)	3.58*** (0.45)
Treat × Post × 2012	0.47* (0.23)	−0.91** (0.29)	0.28 (0.17)	5.82*** (0.48)	2.42** (0.93)	4.15*** (0.46)
Observations	2,266	539	1,045	2,266	539	1,045
Institutions	206	49	95	206	49	95
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Institution fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Main model, R^2	.80	.76	.84	.71	.66	.84
Year interactions model, R^2	.80	.77	.84	.72	.69	.85

Hillman et al., (2015)

TABLE 5
Regression Estimates for SAI's Impact on Long Certificates Awarded

	Certificates (logged)			Certificates (per 100 FTE)		
	WICHE	Neighbors	Matched	WICHE	Neighbors	Matched
a. Main model						
Treat × Post	−0.41** (0.15)	−0.79** (0.27)	−0.60*** (0.17)	−1.14** (0.37)	−1.06 (0.55)	−0.96 (0.41)
b. Year interactions						
Treat × Post × 2007	−0.27 (0.16)	−0.41 (0.30)	−0.51** (0.19)	−0.75 (0.40)	−0.19 (0.63)	−0.69 (0.49)
Treat × Post × 2008	−0.47* (0.20)	−0.99* (0.39)	−0.76** (0.24)	−1.44** (0.48)	−1.71 (0.90)	−1.44 (0.59)
Treat × Post × 2009	−0.67** (0.22)	−1.29*** (0.36)	−0.74** (0.26)	−2.01*** (0.53)	−2.32** (0.86)	−1.33* (0.63)
Treat × Post × 2010	−0.53* (0.22)	−0.95** (0.35)	−0.55* (0.25)	−1.58** (0.51)	−1.20 (0.74)	−1.02 (0.62)
Treat × Post × 2011	−0.45* (0.23)	−0.93** (0.35)	−0.55* (0.25)	−1.08* (0.51)	−1.38 (0.72)	−0.94 (0.60)
Treat × Post × 2012	−0.52* (0.23)	−1.02** (0.35)	−0.74** (0.25)	−1.28* (0.52)	−1.34* (0.68)	−1.18* (0.59)
Observations	2,266	539	1,045	2,266	539	1,045
Institutions	206	49	95	206	49	95
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Institution fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Main model, R^2	.67	.63	.76	.65	.79	.78
Year interactions model, R^2	.67	.64	.75	.65	.79	.78

Note. Panel-corrected standard errors in parentheses. SAI = Student Achievement Initiative; FTE = full-time equivalent; WICHE = Western Interstate Commission for Higher Education.
$^*p < .05$. $^{**}p < .01$. $^{***}p < .001$.

Next time

Final topics!

- Panel data wrap-up
- Introduction to instrumental variables