Lecture 18

10-28-2021

Example 1: Eckhouse (2021)

"Metrics Management and Bureaucratic Accountability: Evidence from Policing"

Government depend on agents to enforce decisions. They may build in metrics to assess these. Eckhouse shows how metrics management encourages bureaucrats to prioritize work with rapid measurable results and deprioritize complex and uncertain projects

She focuses on police departments and the use of CompStat

Example 1: Eckhouse (2021)

Example 1: Eckhouse (2021)

TABLE 5 Effect of CompStat on Share of Rapes Declared Unfounded

	Share Unfounded					
	Rape		Auto Theft		Murder	
	(1)	(2)	(3)	(4)	(5)	(6)
CompStat	0.020**	0.016**	0.005	0.003	−0.006	−0.007
	(0.005)	(0.005)	(0.003)	(0.003)	(0.005)	(0.005)
Total population		−0.000**		−0.000*		−0.000
		(0.000)		(0.000)		(0.000)
Black population (%)		0.311†		0.201+		0.196
		(0.176)		(0.107)		(0.157)
White population (%)		−0.242		0.366**		0.246†
		(0.154)		(0.091)		(0.138)
Observations	900	900	900	900	900	900

Note: †p<0.1;*p<0.05; **p<0.01
All regressions include year and agency fixed effects.
8 cities report no data on unfoundedness, and are excluded from this analysis.

Two Way Fixed Effects

We have been discussing Fixed Effects this week.

Today we are going to move into an area that has had a flurry of research over the last three years: Two Way Fixed Effects

Much of this has been related to the Difference in Differences literature, which we will return to later in the semester

Revisiting the Interpretation of FE

We saw that we transformed the outcome variable and demeaned the explanatory variables.

The time fixed covariates drop out of the model.

Revisiting the Interpretation of FE

We can also think of a data subsetting approach to fixed effects.

Within the data for one case all variation must occur over time, so a regression coefficient within this subset must the average effect of a unit increase in x on y as each variable changes over time.

Two Way Fixed Effects

Under a TWFE approach we have a model regression with dummies for time and cross sectional units.

$$Y_{it} = \alpha_i + \alpha_t + \beta D_{it} + \epsilon_{it}$$

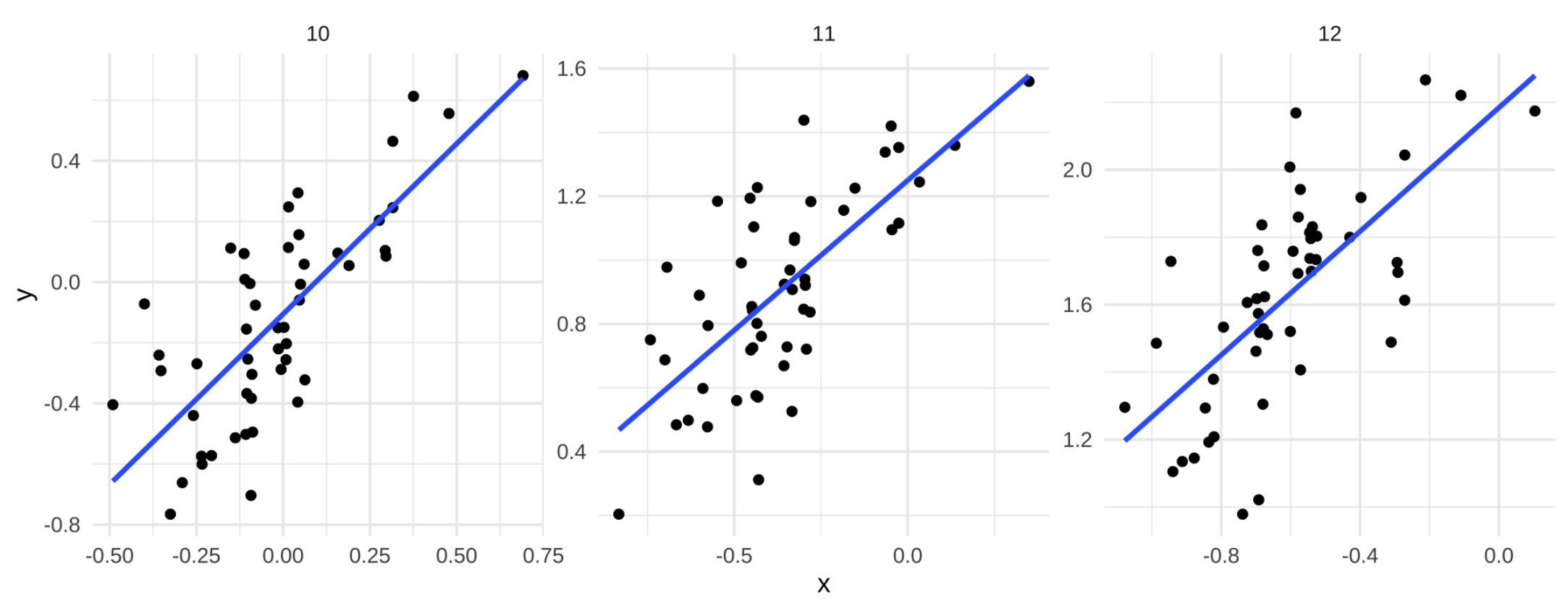
This is a weighted average like any regression, but a particular kind of weighted regression. It is a weighted average of all possible two-group/two period difference in difference estimators in the data.

Simulating the Problem

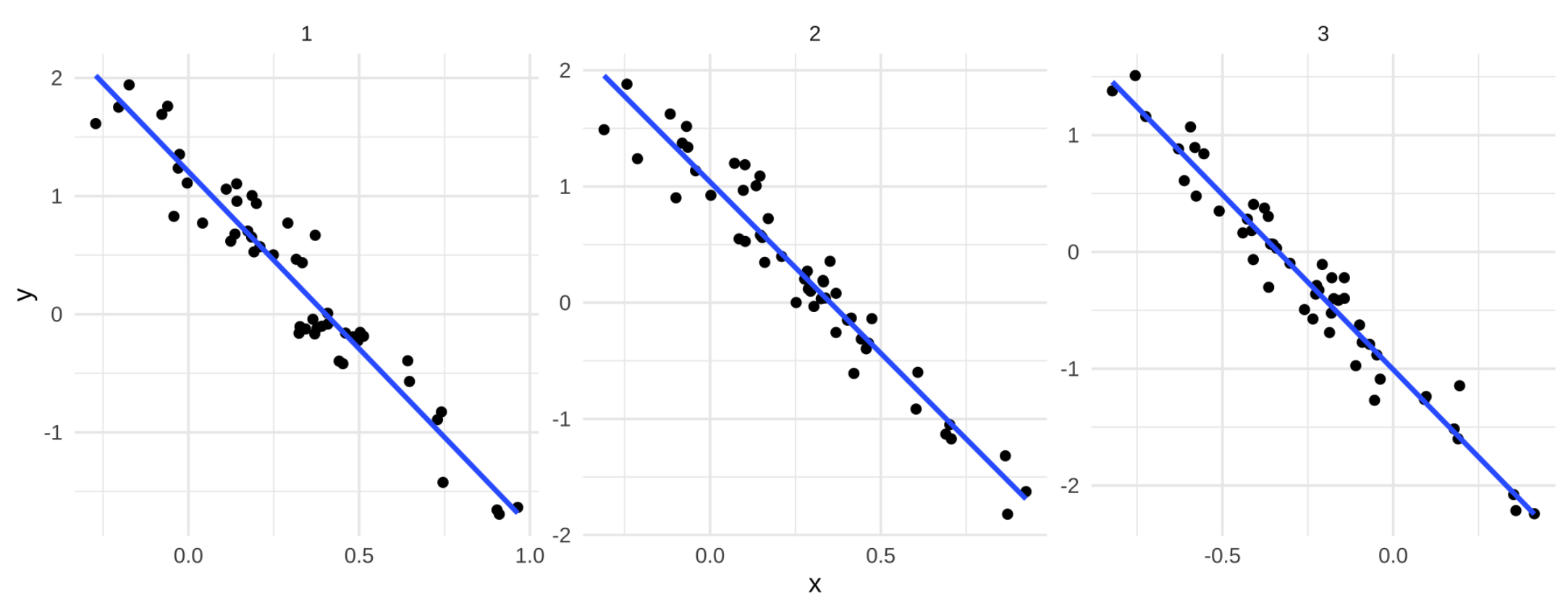
Consider the panel data set that we generate this way:

```
set.seed(42)
gen_data <- tw_data(N=50,
                    T=50,
                    case.eff.mean = 1,
                    cross.eff.mean = -3,
                    cross.eff.sd = 0,
                    case.eff.sd = 0,
                    noise.sd = .25)
```

Visualizing Relationship between CS and WC



Visualizing Relationship between CS and WC



TWFE

```
##          term  estimate  std.error statistic       p.value  conf.low conf.high
## 1 (Intercept) 1.3648606 0.03662145  37.26943 3.210375e-241 1.2930483  1.436673
## 2           x 0.9787602 0.02174064  45.01984 4.940656e-323 0.9361282  1.021392
##     df outcome
## 1 2449       y
## 2 2449       y
```

TWFE

```
##          term  estimate  std.error  statistic       p.value  conf.low conf.high
## 1 (Intercept)  1.205346 0.03543913   34.01172 5.279515e-208  1.135852  1.274840
## 2           x -2.993821 0.01715058 -174.56090  0.000000e+00 -3.027452 -2.960189
##     df outcome
## 1 2449       y
## 2 2449       y
```

TWFE

What happens when we put them together?

```
##          term  estimate  std.error statistic       p.value  conf.low conf.high
## 1 (Intercept)  1.208534 0.05070428  23.83494 6.981612e-113  1.109105  1.307962
## 2           x -3.024623 0.04936737 -61.26766  0.000000e+00 -3.121430 -2.927816
##     df outcome
## 1 2401       y
## 2 2401       y

##             term estimate std.error statistic p.value conf.low conf.high df
## 1 factor(case)18       NA        NA        NA      NA       NA        NA NA
##   outcome
## 1       y
```

R provides us the error message "1 coefficient not defined because the design matrix is rank deficient"

What's wrong with the Regression

When panel data has a single effect for cross-section and for within-case variation the TWFE coefficient is statistically undefined.

Software solves this problem by randomly dropping a variable. The coefficient that determines the effect is now relative to whatever fixed effect we happen to drop.

The solution to this is to generate our panel data differently to have a different effect of x on y for each time point and case.

How Concerned should we be?

Well it depends.

The β is equivalent to

$$\sum_i Y_i \left(\frac{\tilde{D}_{it}}{\sum_i \tilde{D}_{it}^2} \right)$$

(Jakiela 2021)

Here $\tilde{D}_{it}$ is the residual from the regression of the treatment indicator on the group and time fixed effects. In a balanced panel:

$$\tilde{D}_{it} = D_{it} - \bar{D}_t - \bar{D}_i + \bar{D}_a$$

How Concerned should we be?

In situations where treatment effects are heterogeneous, TWFE does not necessarily yield an unbiased estimate of any weighted ATE when treated units receive negative weights (Goodman-Bacon)

So what units get negative weights? The observations with below mean treatment intensity.