Problem Set 3

Program Evaluation and Causal Inference

Christian Bircher   Fenqi Guo   Mingrui Zhang   Wenjie Tu   Zunhan Zhang

Spring Semester 2021

Names are listed in alphabetical order.

Analysis of Experimental Data: Project STAR

```r
# load relevant libraries
library(haven) # read dta file
library(plm) # fixed effects regression
library(stargazer) # print summary statistics
library(car) # table output
library(lmtest) # coeftest
library(sandwich) # robust standard errors

# read data
d.star <- read_dta('STAR.dta')
```

1. Summary statistics

1(a)

```r
df <- table(d.star$girl)
rownames(df) <- c('Boys', 'Girls')
knitr::kable(df)
```

Var1	Freq
Boys	2954
Girls	2795

There are more boys.

1(b)

```r
# average test score
ts.ave <- mean(d.star$tscorek)

# median test score
ts.med <- median(d.star$tscorek)

# average test score in treatment group
ts.ave.treat <- mean(d.star$tscorek[d.star$sck==1])

# average test score in control group
ts.ave.control <- mean(d.star$tscorek[d.star$sck==0])

cat(sprintf('The average test score is %.2f
            \nThe median test score is %.2f
            \nThe average test score in treatment is %.2f
            \nThe average test score in control is %.2f',
            ts.ave, ts.med, ts.ave.treat, ts.ave.control))
```

The average test score is 461.20

The median test score is 457.50

The average test score in treatment is 466.03

The average test score in control is 459.11

1(c)

```r
# find the maximum value for totexpk_m
# convert experience in months into experience in years
exper.years <- max(d.star$totexpk_m)/12

cat(sprintf('The most experienced teacher has %.0f years of experience',
            exper.years))
```

The most experienced teacher has 27 years of experience

2. OLS Regression

2(a)

```r
# OLS
## model.ols and model.pool are exactly the same
model.ols <- lm(tscorek ~ sck, data=d.star)
model.pool <- plm(tscorek ~ sck, data=d.star, index='schidkn',
                  effect='individual', model='pooling')
```

```
stargazer(model.ols, model.pool, omit='Constant',
          keep.stat=c('n','ser'), header=F)
```

Table 2:

	Dependent variable:	
	tscorek	
	OLS	panel linear
	(1)	(2)
sck	6.913***	6.913***
	(1.058)	(1.058)
Observations	5,749	5,749
Residual Std. Error	36.802 (df = 5747)	
Note:	*p<0.1; **p<0.05; ***p<0.01	

Holding other variables constant, a kid in a small class would achieve 6.9 on average higher on the SAT test than a kid in a regular class. This effect is large in economic terms since the *p-value* for the treatment coefficient is smaller than 0.01. In other words, the treatment parameter is statistically different from zero and we can further conclude that such effect is also significant in an economic sense.

2(b)

```
# fixed effects
model.fe <- plm(tscorek ~ sck, data=d.star, index='schidkn',
                effect='individual', model='within')

stargazer(model.ols, model.fe, keep.stat=c('n', 'ser'),
          header=F, omit='Constant')
```

Table 3:

	Dependent variable:	
	tscorek	
	OLS	panel linear
	(1)	(2)
sck	6.913***	7.563***
	(1.058)	(0.949)
Observations	5,749	5,749
Residual Std. Error	36.802 (df = 5747)	
Note:	*p<0.1; **p<0.05; ***p<0.01	

```
# F test for individual effects
pFtest(model.pool, model.fe)
```

```
F test for individual effects
```

data: tscorek ~ sck F = 16.628, df1 = -78, df2 = 5747, p-value = NA alternative hypothesis: significant effects

I prefer the fixed-effect estimate for the class size effect. If we include school fixed effects, we can control for the omitted variables across different schools. Schools in large cities may have more advantaged students and teachers while schools in rural areas may have more disadvantaged students and teachers so different schools may have different effects on test scores. As the *F-test* shows, the "individual effects" (school effects) are significant and thus cannot be ignored.

Standard Errors

3(a)

I do not think this is correct. Although treatment is randomly assigned, there are still some omitted variables (individual heterogeneity) even within the same school. Homoskedasticity assumption, or even mean-zero error assumption may not hold. If $Var(\epsilon|\boldsymbol{X}) \neq \sigma^2 \boldsymbol{I_n}$ means that the general *t test* and *F test* would become ineffective. $E(\epsilon|\boldsymbol{X}) \neq 0$ means that the OLS estimator would not be the best linear unbiased estimator anymore.

There are two solutions to such issue:

- Using the generalized least squares method (GLS)

- Using the ordinary least squares but with robust standard errors

Here in our case, we would prefer to use robust standard errors in OLS so that the treatment estimator would still be unbiased and the statistical tests would still be accurate with robust standard errors.

3(b)

```
# robust standard errors
cluster.vcov <- vcovHC(model.fe, type='HC3', cluster='group')
cluster.se <- sqrt(diag(cluster.vcov))

stargazer(model.fe, model.fe, se=list(NULL, cluster.se),
          column.labels=c('Homoskedasticity', 'Cluster'),
          header=F, keep.stat='n', font.size='small')
```

```
# print summary using standard errors
coeftest(model.fe)
```

```
t test of coefficients:

    Estimate Std. Error t value  Pr(>|t|)
sck  7.56275    0.94912  7.9682 1.932e-15 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Table 4:

	Dependent variable:	
	tscorek	
	Homoskedasticity	Cluster
	(1)	(2)
sck	7.563***	7.563***
	(0.949)	(1.798)
Observations	5,749	5,749
Note:	*p<0.1; **p<0.05; ***p<0.01	

```
# print summary using robust standard errors
coeftest(model.fe, vcov=vcovHC(model.fe,
                               method='arellano',
                               type='HC3',
                               cluster='group'))
```

```
t test of coefficients:

    Estimate Std. Error t value  Pr(>|t|)
sck   7.5628     1.7979  4.2065 2.634e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

As we can see, the estimate for *sck* in OLS with robust standard errors is almost the same as in OLS while the *t-value*s are different. Although we cannot reject $H_0 : \beta_{sck} = 0$ as strongly as in OLS case, we can still conclude that the estimate for the treatment effect is statistically different from zero even we take heterskedasticity into account. Children within the same school may have some heterogeneous variables such as innate ability. We want to test if the treatment effect is reliable but OLS estimation without robust standard errors would render the statistical tests ineffective. Therefore, it is important to use robust standard errors in our case.

Testing whether randomization worked

4(a)

```
model1 <- lm(sck ~ girl, data=d.star)
model2 <- lm(sck ~ freelunk, data=d.star)
model3 <- lm(sck ~ totexpk_m, data=d.star)

stargazer(model1, model2, model3, omit='Constant',
          keep.stat='n', font.size='small', header=F)
```

As we can see, the coefficients on *girl* and on *feelunk* are not statistically different from zero while the coefficient on $totexpk_m$ is statistically significant. We can draw a preliminary conclusion that $totextpk_m$ should be included to our model but some further steps should be taken to see if that is the case.

Table 5:

	(1)	(2)	(3)
	Dependent variable:		
	sck		
girl	−0.0004		
	(0.012)		
freelunk		−0.012	
		(0.012)	
totexpk_m			−0.0002***
			(0.0001)
Observations	5,749	5,749	5,749
Note:			*p<0.1; **p<0.05; ***p<0.01

4(b)

```
model4 <- lm(sck ~ girl + freelunk + totexpk_m, data=d.star)

stargazer(model4, keep.stat=c('n', 'f'), header=F,
          omit='Constant', font.size='small')
```

Table 6:

	Dependent variable:
	sck
girl	0.001
	(0.012)
freelunk	−0.014
	(0.012)
totexpk_m	−0.0002***
	(0.0001)
Observations	5,749
F Statistic	2.932** (df = 3; 5745)
Note:	*p<0.1; **p<0.05; ***p<0.01

F test

$$H_0 : \beta_{girl} = \beta_{freelunk} = \beta_{totexpk\_m} = 0$$

$$H_1 : \beta_{girl} \neq 0 \text{ or } \beta_{freelunk} \neq 0 \text{ or } \beta_{totexpk\_m} \neq 0$$

F-statistic is larger than 1 (with two stars) so we can reject the null hypothesis that there is no relationship between response variable and predictor variables ($H_0 : \beta_{girl} = \beta_{freelunk} = \beta_{totexpk_m} = 0$). In other words, any of *girls*, *freelunk*, or $totexpk_m$ can "explain" treatment status.

4(c)

```
model5 <- lm(tscorek ~ sck + girl + freelunk + totexpk_m, data=d.star)

stargazer(model.ols, model5, column.labels=c('2(a)', '4(c)'),
          header=F, keep.stat=c('n', 'f'),
          font.size='small', omit='Constant')
```

Table 7:

	Dependent variable:	
	tscorek	
	2(a)	4(c)
	(1)	(2)
sck	6.913***	6.912***
	(1.058)	(1.009)
girl		6.902***
		(0.927)
freelunk		−19.549***
		(0.929)
totexpk_m		0.049***
		(0.007)
Observations	5,749	5,749
F Statistic	42.714*** (df = 1; 5747)	156.008*** (df = 4; 5744)
Note:		*p<0.1; **p<0.05; ***p<0.01

The coefficients on all predictor variables are statistically different from zero. As we can see from column(1) to column(2), the estimate on *sck* decreases slightly and so does the standard error. In 2(a), we have omitted variable bias in the model and we should expect it to be positive.

$$2\text{(a):} \quad Y = \alpha_0 + \alpha_1 D + \eta$$
$$4\text{(c):} \quad Y = \beta_0 + \beta_1 D + \beta_2 W_1 + \beta_3 W_2 + \beta_4 W_3 + \epsilon$$

$$\mathbb{E}(\hat{\alpha}_1) = \beta_1 + \underbrace{\beta_2 \frac{Cov(D, W_1)}{Var(W_1)} + \beta_3 \frac{Cov(D, W_2)}{Var(W_1)} + \beta_4 \frac{Cov(D, W_3)}{Var(W_3)}}_{\text{omitted variable bias}}$$

- The estimate on *totextpk_m* is statistically different from zero and there is a weak positive relationship between the test score and the experience of a teacher. It also makes sense in causality. A child is expected to achieve a higher score as a teacher gets more experience. However, this causal effect is not so obvious.

- The estimate on *freelunk* is also statistically different from zero and there is a strong negative relationship between the test score and *freelunk*. Kids from poor households will get lower scores on the test since they do not have too much financial support from their families, which more or less may have some negative impacts on their academic performance. The causal effect makes sense statistically and intuitively.

- The estimate on *girl* is statistically different from zero and there is a positive relationship between the test score and the gender. A kid is more likely to achieve a higher score if this kid is a girl. In general, boys are more likely to be distracted from studying than girls so girls are expected to outperform boys on SAT test.

Heterogeneity in Treatment Effects

5(a)

```
# generate interaction terms
d.star$sck_girl <- d.star$sck*d.star$girl
d.star$sck_freelunk <- d.star$sck*d.star$freelunk
d.star$sck_totexpk_m <- d.star$sck*d.star$totexpk_m
```

```
# fixed effects with sck_girl interaction
model6 <- plm(tscorek ~ sck + girl + freelunk + totexpk_m +
                sck_girl + sck_freelunk + sck_totexpk_m, data=d.star,
            index='schidkn', effect='individual', model='within')

stargazer(model6, keep.stat='n', omit='Constant',
          font.size='small', header=F)
```

Table 8:

	Dependent variable:
	tscorek
sck	12.543***
	(2.293)
girl	6.915***
	(1.001)
freelunk	−19.034***
	(1.147)
totexpk_m	0.039***
	(0.008)
sck_girl	−3.069*
	(1.823)
sck_freelunk	1.076
	(1.844)
sck_totexpk_m	−0.037**
	(0.015)
Observations	5,749
Note:	*p<0.1; **p<0.05; ***p<0.01

Interactions

$$Y = \beta_0 + \beta_1 D + \beta_2 W_1 + \beta_3 W_2 + \beta_4 W_3 + \beta_5 D \cdot W_1 + \beta_6 D \cdot W_2 + \beta_7 D \cdot W_3 + \epsilon$$

- β_1 is the treatment effect on a boy without a free lunch. Holding the month of experience fixed, a boy without a free lunch in small class will achieve 12.543 higher on SAT test than he would have achieved in regular class.

- $(\beta_1 + \beta_5)$ is the treatment effect on a girl without a free lunch. Holding the month of experience fixed, a girl without a free lunch in small class will achieve 9.474 (12.543-3.069) higher on SAT test than she would have achieved in regular class.

- $(\beta_1 + \beta_6)$ is the treatment effect on a boy with a free lunch. Holding the month of experience fixed, a boy with a free lunch in small class will achieve 13.619 (12.543+1.076) higher on SAT test than he would have achieved in regular class.

- $(\beta_1 + \beta_7)$ is the treatment effect on a boy without a free lunch. For each additional month of experience that a teacher has, a boy without a free lunch in small class will achieve 12.506 (12.543-0.037) higher on SAT test than he would have achieved in regular class.

Treatment effect is the largest in the group where boys get a free lunch (poor household).

F-test for the joint hypotheses that the treatment effect is the same for everyone:

$$H_0 : \beta_5 = 0 \text{ and } \beta_6 = 0 \text{ and } \beta_7 = 0$$
$$H_1 : \beta_5 \neq 0 \text{ or } \beta_6 \neq 0 \text{ or } \beta_7 = 0$$

```
# heteroskedasticity-robust F-test
linearHypothesis(model6, c('sck_girl=0',
                           'sck_freelunk=0',
                           'sck_totexpk_m=0'),
                 test='F', white.adjusted='hc3')
```

```
Linear hypothesis test

Hypothesis:
sck_girl = 0
sck_freelunk = 0
sck_totexpk_m = 0

Model 1: restricted model
Model 2: tscorek ~ sck + girl + freelunk + totexpk_m + sck_girl + sck_freelunk +
    sck_totexpk_m

  Res.Df Df      F  Pr(>F)
1   5666
2   5663  3 3.3398 0.01847 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The p-value for *F-statistic* is smaller than 5% significance level so we can reject the null hypothesis. The treatment effect is not the same for everyone.

5(b)

$$ATE = \beta_1\overline{D} + \beta_5\overline{DW_1} + \beta_6\overline{DW_2} + \beta_7\overline{DW_3}$$

```r
# compute ATE by averaging the individual treatment effect
beta1 <- model6$coefficients['sck']
beta5 <- model6$coefficients['sck_girl']
beta6 <- model6$coefficients['sck_freelunk']
beta7 <- model6$coefficients['sck_totexpk_m']
te <- beta1 + beta5*d.star$girl +
      beta6*d.star$freelunk +
      beta7*d.star$totexpk_m

# print summary statistics for ATE
summary(te)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 -2.378   5.282   7.714   7.485   9.671  13.618
```

```r
stargazer(model.fe, model5, header=F, keep='sck',
          keep.stat='n', font.size='small',
          column.labels=c('3(b)', '4(c)'))
```

Table 9:

	Dependent variable:	
	tscorek	
	panel linear	*OLS*
	3(b)	4(c)
	(1)	(2)
sck	7.563***	6.912***
	(0.949)	(1.009)
Observations	5,749	5,749
Note:	*p<0.1; **p<0.05; ***p<0.01	

```r
cat(sprintf('The estimate for ATE in 3(b) is %.2f
            \nThe estimate for ATE in 4(c) is %.2f
            \nThe estimate for ATE in 5(b) is %.2f',
            model.fe$coefficients['sck'],
            model5$coefficients['sck'],
            summary(te)['Mean']))
```

The estimate for ATE in 3(b) is 7.56

The estimate for ATE in 4(c) is 6.91

The estimate for ATE in 5(b) is 7.48