ECON312 Problem Set 5

Futing Chen, Hongfan Chen, Will Parker

05/07/2020

Contents


```
library(dplyr)
library(knitr)
library(haven)
```

1. Describe the data

```r
data <- read_dta("PS5.dta")
psych::describe(data) %>%
  select(n, mean, sd, median, min, max) %>%
  kable(digits = 2)
```

	n	mean	sd	median	min	max
sheet	816	245.72	148.08	236.50	1.0	522.00
post	816	0.50	0.50	0.50	0.0	1.00
chain	816	2.11	1.11	2.00	1.0	4.00
state	816	0.81	0.39	1.00	0.0	1.00
empft	798	8.26	8.31	6.00	0.0	60.00
hrsopen	805	14.47	2.77	15.00	7.0	24.00
nregs	788	3.61	1.25	3.00	2.0	8.00
minwage	816	4.49	0.49	4.25	3.8	5.05
temp	816	0.50	0.50	1.00	0.0	1.00
d1	816	0.42	0.49	0.00	0.0	1.00
d2	816	0.19	0.40	0.00	0.0	1.00
d3	816	0.24	0.43	0.00	0.0	1.00
d4	816	0.14	0.35	0.00	0.0	1.00

2. Estimate the following regression on the sample of fast food restaurants in Feb-Mar 1992:

$$empft_{ikt} = \alpha + \gamma minwage_{kt} + \beta_1 nregs_{ikt} + \beta_2 hrsopen_{ikt} + \sum_{j=2}^{4} \eta_j d_j + \epsilon_{ikt}$$

i denotes restaurant, k denotes state, and $t = 0$ if the observation is from Feb-Mar and $t = 1$ if the observation is from Nov-Dec.

```r
# drop missing observations
for_regression <- data %>%
  select(empft, minwage,nregs, hrsopen, d2, d3, d4) %>%
  na.omit()
regress_formula <- formula(empft ~ minwage + nregs + hrsopen + d2 + d3 + d4)
linear_model <- lm(regress_formula, data = for_regression)
stargazer::stargazer(linear_model, header = FALSE)
```

3. Interpret the coefficient γ and calculate a 90% confidence interval.

$\gamma = 0.23$ is the average change in number of full time employees for a $1 increase in the minimum wage, adjusted for the other variables in the regression. The 90% CI is

```r
confint.lm(linear_model, level = 0.9)[2,]
```

```
##         5 %       95 %
## -0.7167418  1.1722597
```

Table 2:

	Dependent variable:
	empft
minwage	0.228
	(0.574)
nregs	0.404
	(0.311)
hrsopen	1.251***
	(0.160)
d2	1.174
	(1.136)
d3	−1.827**
	(0.847)
d4	4.174***
	(1.102)
Constant	−12.656***
	(3.699)
Observations	775
R^2	0.126
Adjusted R^2	0.119
Residual Std. Error	7.790 (df = 768)
F Statistic	18.494*** (df = 6; 768)
Note:	*p<0.1; **p<0.05; ***p<0.01

4.Use the Sum of squares table from the regression output to calculate the R^2 and the standard error of the regression (Root MSE).

```
SSE <- sum(linear_model$residuals^2)
SST <- var(for_regression$empft)*(for_regression %>% nrow() -1)
SSR <- SST - SSE
R2 <- SSR/SST
root_MSE <- sqrt(SSE/(for_regression %>% nrow()))
```

The $R^2 = 0.13$ and the $MSE = 7.8$

5. Give an economic interpretation of the coefficients η_2, η_3, η_4. What might explain the relatively large coefficient on -d4-?

These are fixed effects for each restaurant, specifically the average difference in number of employees relative to burger king. d-1 The large coefficient η_4 means that Wendy's employed 4.17 more people on average than Burger King, adjusted for the other variables in the regression. One possible explanation is that Wendy's has more customers and hence hires more employers.

6. Test $H0 : \eta_2 = \eta_3 = 0$

```
car::linearHypothesis(linear_model, c("d2 = 0", "d3 = 0"))
```

```
## Linear hypothesis test
##
## Hypothesis:
## d2 = 0
## d3 = 0
##
## Model 1: restricted model
## Model 2: empft ~ minwage + nregs + hrsopen + d2 + d3 + d4
##
##   Res.Df   RSS Df Sum of Sq      F  Pr(>F)
## 1    770 47093
## 2    768 46603  2    489.92 4.0369 0.01803 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We can reject the hypothesis $\eta_2 = \eta_3 = 0$ at a significane level of $\alpha = 0.5$

7. Test the hypothesis $H_0 : \eta_2 = \eta_3$ using the estimated covariance matrix of the coefficients. Verify your answer by running the test in Stata using and/or by performing an F-test. We now want to control for potential selection issues by using the panel structure of our data.

```
cov_matrix <- vcov(linear_model)
beta <- linear_model$coefficients
R <- c(0,0,0, 0, 1, -1, 0)
n_sample <- for_regression %>% nrow()
t(R)%*%beta
```

```
##          [,1]
## [1,] 3.001721
```

```
T_n <- n_sample*(R%*%beta)%*%solve(t(R)%*%cov_matrix%*%R)%*%t(R%*%beta)
```

```
car::linearHypothesis(linear_model, c("d2 = d3"))
```

```
## Linear hypothesis test
##
## Hypothesis:
## d2 - d3 = 0
##
## Model 1: restricted model
## Model 2: empft ~ minwage + nregs + hrsopen + d2 + d3 + d4
##
##   Res.Df   RSS Df Sum of Sq      F  Pr(>F)
## 1    769 47001
## 2    768 46603  1    398.45 6.5663 0.01058 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

8. Explain why the previous estimate of λ is likely to suffer from omitted variable bias.

λ suffer from omitted variable bias if some of the unobservables that affect the number of employees are correlated with minimum wage. The omitted variable bias is likely because one can imagine many such unobservables such as unmeasured conditions of state economies or demographics. For example, some states have higher per capita income or the age structure is younger. These characteristics are potentially positively correlated with $minwage$ and $empft$; then γ is upwardly biased.

9. Assume that $\epsilon_{ikt} = \mu_k + \zeta_t + u_{ikt}$ and that $E[u_{ikt}|X_{ikt}] = 0$ (where X_{ikt} is the vector of RHS-variables in (2) except -minwage-). Explain how you can then use the increase in the minimum wage in New Jersey and a difference-in-differences (DD) model to identify the effect of the minimum wage on employment. Give an example where the necessary assumption(s) are violated.

Since $\epsilon_{ikt} = \mu_k + \zeta_t + u_{ikt}$, we know

$$Y = \alpha + \gamma minwage_{kt} + X_{ikt}^{'}\beta + \mu_k + \zeta_t + u_{ikt}$$

Then $E[u_{ikt}|X_{ikt}] = 0$ implies

$$E[Y|post = 1, state = 1, X_{ikt}] - E[Y|post = 0, state = 1, X_{ikt}] = \gamma\Delta minwage + (\zeta_1 - \zeta_0)$$
$$E[Y|post = 1, state = 0, X_{ikt}] - E[Y|post = 0, state = 0, X_{ikt}] = (\zeta_1 - \zeta_0)$$

Thus,

$$\{E[Y|post = 1, state = 1, X_{ikt}] - E[Y|post = 0, state = 1, X_{ikt}]\}$$
$$-\{E[Y|post = 1, state = 0, X_{ikt}] - E[Y|post = 0, state = 0, X_{ikt}]\} = \gamma\Delta minwage$$

Hence, we can use the following DD model:

$$empft_{ikt} = \pi_0 + \delta_{kt} state_k * post_t + \pi_1 state_k + \pi_2 post_t + X_{ikt}^{'}\beta + u_{ikt}$$

and the DD estimator δ_{kt} will identify the effect of the minimum wage on employment.

The necessary assumption is common trend in the absence of intervation:

$$E[Y_1^0 - Y_0^0|D_1 = 1] = E[Y_1^0 - Y_0^0|D_1 = 0]$$

This would be violated if, for example, the employment of one state (say, New Jersey) benefited more from progressive federal policies than Pennsylvania.

10. Generate a table of means, a table of standard errors and a table of frequencies for -empft- in each state and each time period (post = 1 and post = 0).

```
with(data, tapply(empft, list(state=state,post=post), mean,na.rm=T) )
```

```
##      post
## state         0        1
##     0 10.311688 7.651316
##     1  7.732308 8.446875
```

```
with(data, tapply(empft, list(state=state,post=post), sd,na.rm=T) )
```

```
##      post
## state         0        1
##     0 10.805103 8.514309
##     1  7.974731 7.857189
```

```r
filter(data, !is.na(empft)) %>%
  with(., tapply(empft, list(state=state,post=post), length) )
```

```
##      post
## state   0   1
##     0  77  76
##     1 325 320
```

11. Using these statistics, calculate a DD estimate of the impact of the minimum wage law on employment.

```r
(8.446875-7.732308)-(7.651316-10.311688)
```

```
## [1] 3.374939
```

12. Specify and estimate the corresponding regression.

The regression corresponding to (11) is

$$empft_{ikt} = \pi_0 + \delta state_k * post_t + \pi_1 state_k + \pi_2 post_t + v_{ikt}$$

```r
DID1<-lm(empft~state*post,data=data); summary(DID1)
```

```
##
## Call:
## lm(formula = empft ~ state * post, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -10.312  -6.413  -2.447   3.553  52.268
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  10.3117     0.9451  10.911   <2e-16 ***
## state        -2.5794     1.0511  -2.454   0.0143 *
## post         -2.6604     1.3410  -1.984   0.0476 *
## state:post    3.3749     1.4915   2.263   0.0239 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.293 on 794 degrees of freedom
##   (18 observations deleted due to missingness)
## Multiple R-squared:  0.008244,   Adjusted R-squared:  0.004497
## F-statistic:   2.2 on 3 and 794 DF,  p-value: 0.08666
```

We can also add covariates and estimate

```r
DID2<-lm(empft~state*post+nregs+hrsopen+d2+d3+d4,data=data); summary(DID2)
```

```
##
## Call:
```

```
## lm(formula = empft ~ state * post + nregs + hrsopen + d2 + d3 +
##     d4, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -18.442  -5.130  -1.422   3.519  40.545
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -9.9380     2.7793  -3.576 0.000371 ***
## state        -2.3412     0.9898  -2.365 0.018258 *
## post         -2.8888     1.2650  -2.284 0.022668 *
## nregs         0.4474     0.3106   1.440 0.150216
## hrsopen       1.2514     0.1593   7.854 1.36e-14 ***
## d2            1.1988     1.1319   1.059 0.289909
## d3           -1.8780     0.8448  -2.223 0.026499 *
## d4            4.1681     1.0986   3.794 0.000160 ***
## state:post    3.6657     1.4096   2.601 0.009486 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.763 on 766 degrees of freedom
##   (41 observations deleted due to missingness)
## Multiple R-squared:  0.1344, Adjusted R-squared:  0.1253
## F-statistic: 14.87 on 8 and 766 DF,  p-value: < 2.2e-16
```