ECON312 Problem Set 5

Futing Chen, Hongfan Chen, Will Parker

05/06/2020

Contents


```r
library(tidyverse)
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

	n	mean	sd	median	min	max
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.

```
# 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

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

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

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)
```

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

```r
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.

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

```r
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.

```r
cov_matrix <- vcov(linear_model)

beta <- linear_model$coefficients

R <- c(0,0,0, 0, 1, -1, 0)
```

```r
n_sample <- for_regression %>% nrow()

t(R)%*%beta
```

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

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

```r
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.