Empirical Exercise 7.1

Zheng Tian

[2016-05-31 Tue]

This file include answers and R codes for completing Empirical Exercise 7.1 in Introduction to Econometrics (3rd edition) by Stock and Watson.

1 Answers

The estimation results for two regression models are reported in Table 1.

Table 1: Regression Results

	Dependent variable:	
	ahe	
	(1)	(2)
age	0.605***	0.585***
	(0.040)	(0.037)
female		−3.664***
		(0.208)
bachelor		8.083***
		(0.213)
Constant	1.082	−0.636
	(1.167)	(1.083)
Observations	7,711	7,711
R^2	0.029	0.200
Adjusted R^2	0.029	0.199
Residual Std. Error	9.992	9.072
F Statistic	230.432***	641.492***
Note:	*p<0.1; **p<0.05; ***p<0.01	

a The estimated coefficient on *Age* is `0.605`, which is significant at the 1% level. The estimated intercept is `1.082`, which is insignificant at the 10% level.

b The estimated marginal effect of *Age* on *AHE* is `0.585` dollars per year, which is significant at the 1% level. The confidence interval is (`0.514`, `0.657`).

c The results are quite similar. Evidently the regression in (a) does not suffer from important omitted variable bias.

d Bob's predicted average hourly earnings is `14.58` dollars. Alexis's predicted average hourly earnings is `21.34` dollars.

e The regression in (b) fits the data much better. Gender and education are important predictors of earnings. The R^2 and $\bar{R}^2$ are similar because the sample size is very large relative to the number of regressors.

f Gender and education are important. The t-statistic for *Female* is `-17.643`, the t-statistic for *Bachelor* is `37.993`, and the F-statistic for both *Female* and *Bachelor* is `822.467`, which are all significant at the 1% level.

g The omitted variables must have non-zero coefficients and must correlated with the included regressor. From (f) Female and Bachelor have non-zero coefficients; yet there does not seem to be important omitted variable bias, suggesting that the correlation of Age and Female and Age and Bachelor is small.

2 R codes

```
# This file contains the answer for the empirical exercise 7.1 in Chapter 8
# Introduction to Econometrics by Stock and Watson
library(AER)
library(foreign)

cpsdata <- read.dta("./data/cps08.dta")
summary(cpsdata)

# a ---------------------------------------------------------------------
fm.a <- ahe ~ age
ols.a <- lm(fm.a, data = cpsdata)
b0.a <- coef(ols.a)[1]
b1.a <- coef(ols.a)[2]
summ.a <- summary(ols.a)
rsq.a <- summ.a$r.squared
adrsq.a <- summ.a$adj.r.squared
se.a <- sqrt(diag(vcovHC(ols.a, type = "HC1")))

# b ---------------------------------------------------------------------
fm.b <- ahe ~ age + female + bachelor
ols.b <- lm(fm.b, data = cpsdata)
b1.b <- coef(ols.b)[2]
ci.b <- confint(ols.b, "age")
summ.b <- summary(ols.b)
rsq.b <- summ.b$r.squared
adrsq.b <- summ.b$adj.r.squared

# with heteroskedasticity-robust se
se.b <- sqrt(diag(vcovHC(ols.b, type = "HC1")))
```

```
ci.b.rob <- matrix(
  c(b1.b - qnorm(0.975) * se.b[2], b1.b + qnorm(0.975) * se.b[2]),
  nrow = 1)
colnames(ci.b.rob) <- c("2.5%", "97.5%")
rownames(ci.b.rob) <- "age"

library(stargazer)
stargazer(ols.a, ols.b, se = list(se.a, se.b),
          title = "Regression Results", style = "aer",
          no.space = TRUE)


# d ------------------------------------------------------------------------
bob <- data.frame(age = 26, female = 0, bachelor = 0)
bob.hat <- predict(ols.b, bob)
alexis <- data.frame(age = 30, female = 1, bachelor = 1)
alexis.hat <- predict(ols.b, alexis)

# e ------------------------------------------------------------------------
cmp.rsq <- rbind(c(rsq.a, adrsq.a),
                 c(rsq.b, adrsq.b))
rownames(cmp.rsq) <- c("a", "b")
colnames(cmp.rsq) <- c("R sqd", "Adj. R sqd")

# f ------------------------------------------------------------------------
# test female
tstat.b <- coeftest(ols.b, vcovHC, df = Inf)
b.female <- tstat.b[3, 1]
t.female <- tstat.b[3, 3]
p.female <- tstat.b[3, 4]

b.bach <- tstat.b[4, 1]
t.bach <- tstat.b[4, 3]
p.bach <- tstat.b[4, 4]

fstat.b <- linearHypothesis(ols.b,
                            hypothesis.matrix = rbind(c(0, 0, 1, 0),
                                                      c(0, 0, 0, 1)),
                            rhs = c(0, 0), test = "F")
F <- fstat.b[2, "F"]
F.p <- fstat.b[2, 6]
```