Empirical Exercise 6.1

Zheng Tian

[2016-05-04 Wed]

This file include answers and R codes for completing Empirical Exercise 6.1 in Introduction to Econometrics (3rd edition) by Stock and Watson.

1 Instructions

1.1 Read the data

As usual, we first read the data and load the `AER` package

```
library(AER)
library(foreign)
teachingdata <- read.dta("TeachingRatings.dta")
summary(teachingdata)
```

```
   minority           age            female         onecredit
Min.   :0.0000   Min.   :29.00   Min.   :0.0000   Min.   :0.00000
1st Qu.:0.0000   1st Qu.:42.00   1st Qu.:0.0000   1st Qu.:0.00000
Median :0.0000   Median :48.00   Median :0.0000   Median :0.00000
Mean   :0.1382   Mean   :48.37   Mean   :0.4212   Mean   :0.05832
3rd Qu.:0.0000   3rd Qu.:57.00   3rd Qu.:1.0000   3rd Qu.:0.00000
Max.   :1.0000   Max.   :73.00   Max.   :1.0000   Max.   :1.00000
    beauty          course_eval        intro          nnenglish
Min.   :-1.45049   Min.   :2.100   Min.   :0.0000   Min.   :0.00000
1st Qu.:-0.65627   1st Qu.:3.600   1st Qu.:0.0000   1st Qu.:0.00000
Median :-0.06801   Median :4.000   Median :0.0000   Median :0.00000
Mean   : 0.00000   Mean   :3.998   Mean   :0.3391   Mean   :0.06048
3rd Qu.: 0.54560   3rd Qu.:4.400   3rd Qu.:1.0000   3rd Qu.:0.00000
Max.   : 1.97002   Max.   :5.000   Max.   :1.0000   Max.   :1.00000
```

1.2 Run regressions

Run a simple regression of `Course_Eval` on `Beauty`

We first run a simple linear regression model of `Course_Eval` on `Beauty`.

```
mod1 <- lm(course_eval ~ beauty, data = teachingdata)
summary(mod1)
```

Run a multiple regression model

Then we run a multiple regression model.

```
mod2 <- lm(course_eval ~ beauty + intro + onecredit + female
           + minority + nnenglish, data = teachingdata)
summary(mod2)
```

Re-run the multiple regression model to test the FWL theorem

```
mod2.a <- lm(course_eval ~ intro + onecredit + female
             + minority + nnenglish, data = teachingdata)
mod2.b <- lm(beauty ~ intro + onecredit + female
             + minority + nnenglish, data = teachingdata)
mod2.c <- lm(resid(mod2.a) ~ resid(mod2.b) - 1)
summary(mod2.c)
```

1.3 Prediction

Prediction can be made with `predict()`. First, we need to define a `data.frame` object for Professor Smith. Then, use it in the function.

```
smith <- data.frame(minority = 1, female = 0,
                    beauty = mean(teachingdata$beauty),
                    nnenglish = 0, intro = 0, onecredit = 0)
smith.hat <- predict(mod2, smith)
```

2 Answers

2.1 Answers for a and b

The results of the simple linear regression model and the multiple regression model are summarized in Table 1.

The coefficient on `beauty` in the simple linear regression model is `0.133`, while that in the multiple regression model is `0.1656`. The difference between the two estimated coefficients are not very large, which may imply that the omitted variable bias in the simple regression model is not very serious.

2.2 Answer for c

The estimation following the three steps in the FWL theorem yields the estimate of the coefficient on `beauty` is `0.1656`, which is identical to that in the multiple regression model shown above.

Table 1: The OLS Estimation of the Simple and Multiple Regressions

	Dependent variable:	
	course-eval	
	(1)	(2)
beauty	0.1330***	0.1656***
	(0.0322)	(0.0307)
intro		0.0113
		(0.0545)
onecredit		0.6345***
		(0.1113)
female		−0.1735***
		(0.0493)
minority		−0.1666**
		(0.0763)
nnenglish		−0.2442**
		(0.1070)
Constant	3.9983***	4.0683***
	(0.0253)	(0.0375)
Observations	463	463
R^2	0.0357	0.1546
Adjusted R^2	0.0336	0.1435
Residual Std. Error	0.5455 (df = 461)	0.5135 (df = 456)
F Statistic	17.0847*** (df = 1; 461)	13.9036*** (df = 6; 456)
Note:		*p<0.1; **p<0.05; ***p<0.01

2.3 Answer for d

The predicted course evaluation for Professor Smith is `3.9017`, calculated as $4.0683 + 0.1656 \times 0 + 0.0113 \times 0 + 0.6345 \times 0 - 0.1735 \times 0 - 0.1666 \times 1 - 0.2442 \times 0 = 3.9017$.

3 Appendix: R codes

```
library(AER)
library(foreign)
teachingdata <- read.dta("TeachingRatings.dta")
summary(teachingdata)

# simple regression
mod1 <- lm(course_eval ~ beauty, data = teachingdata)
summary(mod1)

# multiple regression
mod2 <- lm(course_eval ~ beauty + intro + onecredit + female
           + minority + nnenglish, data = teachingdata)
summary(mod2)

# FWL regressions
mod2.a <- lm(course_eval ~ intro + onecredit + female
             + minority + nnenglish, data = teachingdata)
mod2.b <- lm(beauty ~ intro + onecredit + female
             + minority + nnenglish, data = teachingdata)
mod2.c <- lm(resid(mod2.a) ~ resid(mod2.b) - 1)
summary(mod2.c)

# prediction
smith <- data.frame(minority = 1, female = 0,
                    beauty = mean(teachingdata$beauty),
                    nnenglish = 0, intro = 0, onecredit = 0)
smith.hat <- predict(mod2, smith)

b1.mod1 <- coef(mod1)[2]
b1.mod2 <- coef(mod2)[2]
b1.mod2c <- coef(mod2.c)[1]

library(stargazer)
stargazer(mod1, mod2,
          title = "The OLS Estimation of the Simple and Multiple Regressions",
          label = "tab:results_ab",
          dep.var.labels = "course_eval",
          digits = 4, no.space = TRUE)
```