Solutions to Tutorial Questions - Week 4

Statistics and Econometrics

Question 1

Consider a model where the return to education depends upon the amount of work experience (and vice versa):

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 educ \cdot exper + u$$

1. What is the return to another year of education?
2. State the null hypothesis that the return to education does not depend on the level of $exper$. What do you think is the appropriate alternative?
3. Use the data in wage2.RData to test the null hypothesis in part 2 against your stated alternative.
4. Predict the expected wage for an average person with $educ = 12$ and $exper = 10$.

Solutions

1. Holding $exper$ (and the elements in u) fixed, we have

   $$\Delta \log(wage) = \beta_1 \Delta educ + \beta_3 (\Delta educ) exper = (\beta_1 + \beta_3 exper) \Delta educ,$$

   or $\Delta \log(wage)/\Delta educ = \beta_1 + \beta_3 exper$, which is the approximate percentage change in $wage$ given one more year of education.

2. $H_0 : \beta_3 = 0$. If we think that education and experience interact positively – so that people with more experience are more productive when given another year of education – then $\beta_3 > 0$ is the appropriate alternative.

3.

```r
load("wage2.RData")
log.wage.model <- lm(log(wage) ~ educ + exper + educ:exper, data)
```

The estimated equation is

$$\widehat{\log(wage)} = \underset{(.24)}{5.95} + \underset{(.017)}{.044} educ - \underset{(.020)}{.021} exper + \underset{(.0015)}{.0032} educ \cdot exper,$$

$n = 935, R^2 = .135, \bar{R}^2 = .132$. The t statistic on the interaction term is about 2.095, which gives a p-value below .02 against $H_1 : \beta_3 > 0$. Therefore, we reject $H_0 : \beta_3 = 0$ against $H_1 : \beta_3 > 0$ at the 2% level.

4.

```r
newdata <- data.frame(educ = 12, exper = 10)
predicted.logwage <- predict(log.wage.model, newdata, interval = "none")
predicted.wage <- mean(exp(log.wage.model$residuals)) * exp(predicted.logwage)
predicted.wage
```

```
##        1
## 829.9023
```

The expected wage for an average person with $educ = 12$ and $exper = 10$ is predicted to be around 830.

Question 2

Use the data from jtrain.RData for this exercise.

1. Consider the simple regression model

$$\log(scrap) = \beta_0 + \beta_1 grant + u,$$

where *scrap* is the firm scrap rate (percentage of failed assemblies or material that cannot be repaired or restored, and is therefore condemned and discarded), and *grant* is a dummy variable indicating whether a firm received a job training grant. Can you think of some reaons why the unobserved factors in u might be correlated with *grant*?
2. Estimate the simple regression model using the data for 1988. (You should have 54 observations) Does receiving a job training grant significantly lower a firm's scrap rate?
3. Now, add as an explanatory variable $\log(scrap_{87})$. How does this change the estimated effect of *grant*? Interpret the coefficient on *grant*. Is it statistically significant at the 5% level against the one-sided alternative $H_1 : \beta_{grant} < 0$?
4. Test the null hypothesis that the parameter on $\log(scrap_{87})$ is one against the two-sided alternative. Report the p-value for the test.
5. Repeats parts 3 and 4, using heteroskedasticity-robust standard errors, and briefly discuss any notable differences.

Solutions

1. If the grants were awarded to firms based on firm or worker characteristics, grant could easily be correlated with such factors that affect productivity. In the simple regression model, these are contained in u.

2.

```
load("jtrain.RData")
data.88 <- data %>% filter(year == 1988) %>% select(lscrap, grant, lscrap_1) %>% na.omit
lscrap.m1 <- lm(lscrap ~ grant, data.88)
```

The simple regression estimates using the 1988 data are

$$\widehat{\log(scrap)} = \underset{(.241)}{.409} + \underset{(.406)}{.057}\, grant,$$

where $n = 54$, and $R^2 = .0004$. The coefficient on *grant* is actually positive, but not statistically different from zero.

3.

```
lscrap.m2 <- lm(lscrap ~ grant + lscrap_1, data.88)
-qt(0.95, 51)
```

```
## [1] -1.675285
```

When we add $\log(scrap_{87})$ to the equation, we obtain

$$\widehat{\log(scrap_{88})} = \underset{(.089)}{.021} - \underset{(.147)}{.254}\, grant_{88} + \underset{(.044)}{.831} \log(scrap_{87}),$$

where $n = 54$, $R^2 = .873$. The t statistic for $H_0 : \beta_{grant} = 0$ is $-.254/.147 \approx -1.73$. The 5% critical value for 51 df is around -1.68. Because $t = -1.73 < -1.68$, we reject H_0 in favor of $H_1 : \beta_{grant} < 0$ at the 5% level.

4.

```r
2 * pt(-3.84, 51)
```

```
## [1] 0.0003413515
```

The t statistic is $(.831 - 1)/.044 \approx -3.84$ (p-value ≈ 0.00034), which is a strong rejection of H_0.

5.

```r
cov <- vcovHC(lscrap.m2, type = "HC1")
robust.se <- sqrt(diag(cov))
stargazer(lscrap.m2, lscrap.m2, se = list(NULL, robust.se), column.labels = c("default", "robust"),
          header = FALSE, type = 'latex', title = "Question 2.5")
```

Table 1: Question 2.5

	Dependent variable:	
	lscrap	
	default	robust
	(1)	(2)
grant	-0.254^{*}	-0.254^{*}
	(0.147)	(0.146)
lscrap_1	0.831^{***}	0.831^{***}
	(0.044)	(0.074)
Constant	0.021	0.021
	(0.089)	(0.100)
Observations	54	54
R^2	0.873	0.873
Adjusted R^2	0.868	0.868
Residual Std. Error (df = 51)	0.513	0.513
F Statistic (df = 2; 51)	174.941^{***}	174.941^{***}
Note:	*p<0.1; **p<0.05; ***p<0.01	

```r
2 * pt(-2.28, 51)
```

[1] 0.026821

With the heteroskedasticity-robust standard error, the t statistic for $grant_{88}$ is $-.254/.146 \approx -1.74$, so the coefficient is slightly more significantly less than zero when we use the heteroskedasticity-robust standard error. The t statistic for $H_0 : \beta_{\log(scrap_{87})} = 1$ is $(.831 - 1)/.074 \approx -2.28$, which is notably smaller than before. We can reject null at 5% significance level, but not at 1% significance level.