ECON3133 Introductory Econometrics

Chapter 3 Exercises

Solutions

1.

From equation (3.22) we have

$$\tilde{\beta}_1 = \frac{\sum_{i=1}^{n} \hat{r}_{i1} y_i}{\sum_{i=1}^{n} \hat{r}_{i1}^2},$$

where the $\hat{r}_{i1}$ are defined in the problem. As usual, we must plug in the true model for y_i:

$$\tilde{\beta}_1 = \frac{\sum_{i=1}^{n} \hat{r}_{i1}(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + u_i}{\sum_{i=1}^{n} \hat{r}_{i1}^2}.$$

The numerator of this expression simplifies because $\sum_{i=1}^{n} \hat{r}_{i1} = 0$, $\sum_{i=1}^{n} \hat{r}_{i1} x_{i2} = 0$, and $\sum_{i=1}^{n} \hat{r}_{i1} x_{i1} = \sum_{i=1}^{n} \hat{r}_{i1}^2$. These all follow from the fact that the $\hat{r}_{i1}$ are the residuals from the regression of x_{i1} on x_{i2}: the $\hat{r}_{i1}$ have zero sample average and are uncorrelated in sample with x_{i2}. So the numerator of $\tilde{\beta}_1$ can be expressed as

$$\beta_1 \sum_{i=1}^{n} \hat{r}_{i1}^2 + \beta_3 \sum_{i=1}^{n} \hat{r}_{i1} x_{i3} + \sum_{i=1}^{n} \hat{r}_{i1} u_i .$$

Putting these back over the denominator gives

$$\tilde{\beta}_1 = \beta_1 + \beta_3 \frac{\sum_{i=1}^{n} \hat{r}_{i1} x_{i3}}{\sum_{i=1}^{n} \hat{r}_{i1}^2} + \frac{\sum_{i=1}^{n} \hat{r}_{i1} u_i}{\sum_{i=1}^{n} \hat{r}_{i1}^2} .$$

Conditional on all sample values on x_1, x_2, and x_3, only the last term is random due to its dependence on u_i. But $\mathrm{E}(u_i) = 0$, and so

$$\mathrm{E}(\tilde{\beta}_1) = \beta_1 + \beta_3 \frac{\sum_{i=1}^{n} \hat{r}_{i1} x_{i3}}{\sum_{i=1}^{n} \hat{r}_{i1}^2} ,$$

which is what we wanted to show. Notice that the term multiplying β_3 is the regression coefficient from the simple regression of x_{i3} on $\hat{r}_{i1}$.

2. (i) The constant elasticity equation is

$$\widehat{log(salary)} = 4.62 + .162 \;\; log(sales) + .107 \;\; log(mktval)$$
$$n = 177, \;\; R^2 = .299.$$

(ii) We cannot include profits in logarithmic form because profits are negative for nine of the companies in the sample. When we add it in levels form we get

$$\widehat{log(salary)} = 4.69 + .161 \;\; log(sales) + .098 \;\; log(mktval) + .000036 profits$$
$$n = 177, R^2 = .299.$$

The coefficient on *profits* is very small. Here, *profits* are measured in millions, so if profits increase by \$1 billion, which means $\Delta profits = 1{,}000$ – a huge change – predicted salary increases by about only 3.6%. However, remember that we are holding sales and market value fixed.

Together, these variables (and we could drop *profits* without losing anything) explain almost 30% of the sample variation in log(*salary*). This is certainly not "most" of the variation.

(iii) Adding *ceoten* to the equation gives

$$\widehat{log(salary)} = 4.56 + .162\ \ log(sales) + .102\ \ log(mktval) + .000029 profits + .012 ceoten$$

$n = 177,\ \ R^2 = .318.$

This means that one more year as *CEO* increases predicted salary by about 1.2%.

(iv) The sample correlation between log(*mktval*) and *profits* is about .78, which is fairly high. As we know, this causes no bias in the OLS estimators, although it can cause their variances to be large. Given the fairly substantial correlation between market value and firm profits, it is not too surprising that the latter adds nothing to explaining CEO salaries. Also, *profits* is a short term measure of how the firm is doing while *mktval* is based on past, current, and expected future profitability.

3.

(i) The average of *prpblck* is .113 with standard deviation .182; the average of *income* is 47,053.78 with standard deviation 13,179.29. It is evident that *prpblck* is a proportion and that *income* is measured in dollars.

(ii) The results from the OLS regression are

$$\widehat{psoda} = .956 + .115 prpblck + .0000016 income$$

$$n = 401, \quad R^2 = .064.$$

If, say, *prpblck* increases by .10 (ten percentage points), the price of soda is estimated to increase by .0115 dollars, or about 1.2 cents. While this does not seem large, there are communities with no black population and others that are almost all black, in which case the difference in *psoda* is estimated to be almost 11.5 cents.

(iii) The simple regression estimate on *prpblck* is .065, so the simple regression estimate is actually lower. This is because *prpblck* and *income* are negatively correlated (-.43) and *income* has a positive coefficient in the multiple regression.

(iv) To get a constant elasticity, income should be in logarithmic form. I estimate the constant elasticity model:

$$\log\widehat{(psoda)} = -.794 + .122 prpblck + .077\log(income)$$

$$n = 401, \quad R^2 = .068.$$

If *prpblck* increases by .20, log(*psoda*) is estimated to increase by .20(.122) = .0244, or about 2.44 percent.

(v) $\hat{\beta}_{prpblck}$ falls to about .073 when *prppov* is added to the regression.

(vi) The correlation is about $-.84$, which makes sense because poverty rates are determined by income (but not directly in terms of median income).

(vii) There is no argument that they are highly correlated, but we are using them simply as controls to determine if the is price discrimination against blacks. In order to isolate the pure discrimination effect, we need to control for as many measures of income as we can; including both variables makes sense.