FINM3123 Introduction to Econometrics

Chapter 4 Exercises

Solutions

1.

(i) H_0: $\beta_3 = 0$. H_1: $\beta_3 \neq 0$.

(ii) Other things equal, a larger population increases the demand for rental housing, which should increase rents. The demand for overall housing is higher when average income is higher, pushing up the cost of housing, including rental rates.

(iii) The coefficient on log(*pop*) is an elasticity. A correct statement is that "a 10% increase in population increases *rent* by .066(10) = .66%."

(iv) With $df = 64 - 4 = 60$, the 1% critical value for a two-tailed test is 2.660. The t statistic is about 3.29, which is well above the critical value. So β_3 is statistically different from zero at the 1% level.

2.

(i) We need to compute the F statistic for the overall significance of the regression with n = 142 and $k = 4$: $F = [.0395/(1 - .0395)](137/4) \approx 1.41$. The 5% critical value with 4 numerator *df* and using 120 for the numerator *df*, is 2.45, which is well above the value of F. Therefore, we fail to reject H_0: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ at the 10% level. No explanatory variable is individually significant at the 5% level. The largest absolute t statistic is on *dkr*, $t_{dkr} \approx 1.60$, which is not significant at the 5% level against a two-sided alternative.

(ii) The F statistic (with the same *df*) is now $[.0330/(1 - .0330)](137/4) \approx 1.17$, which is even lower than in part (i). None of the t statistics is significant at a reasonable level.

(iii) We probably should not use the logs, as the logarithm is not defined for firms that have zero for *dkr* or *eps*. Therefore, we would lose some firms in the regression.

(iv) It seems very weak. There are no significant t statistics at the 5% level (against a two-sided alternative), and the F statistics are insignificant in both cases. Plus, less than 4% of the variation in *return* is explained by the independent variables.

3.

(i) There are 2,017 single people in the sample of 9,275.

(ii) The estimated equation is

$$\widehat{nettfa} = \underset{(4.08)}{-43.04} + \underset{(.060)}{.799}\, inc + \underset{(.092)}{.843}\, age$$

$n = 2{,}017, \quad R^2 = .119.$

The coefficient on *inc* indicates that one more dollar in income (holding *age* fixed) is reflected in about 80 more cents in predicted *nettfa*; no surprise there. The coefficient on *age* means that, holding income fixed, if a person gets another year older, his/her *nettfa* is predicted to increase by about \$843. (Remember, *nettfa* is in thousands of dollars.) Again, this is not surprising.

(iii) The intercept is not very interesting as it gives the predicted *nettfa* for *inc* = 0 and *age* = 0. Clearly, there is no one with even close to these values in the relevant population.

(iv) The t statistic is $(.843 - 1)/.092 \approx -1.71$. Against the one-sided alternative H_1: $\beta_2 < 1$, the p-value is about .044. Therefore, we can reject H_0: $\beta_2 = 1$ at the 5% significance level (against the one-sided alternative).

(v) The slope coefficient on *inc* in the simple regression is about .821, which is not very different from the .799 obtained in part (ii). As it turns out, the correlation between *inc* and *age* in the sample of single people is only about .039, which helps explain why the simple and multiple regression estimates are not very different; refer back to page 84 of the text.

4.

(i) The results from the OLS regression, with standard errors in parentheses, are

$$\log(\widehat{psoda}) = -1.46 + .073\, prpblck + .137 \log(income) + .380\, prppov$$

$$\begin{array}{cccc} (0.29) & (.031) & (.027) & (.133) \end{array}$$

$n = 401, \quad R^2 = .087$

The p-value for testing H_0: $\beta_1 = 0$ against the two-sided alternative is about .018, so that we reject H_0 at the 5% level but not at the 1% level.

(ii) The correlation is about $-.84$, indicating a strong degree of multicollinearity. Yet each coefficient is very statistically significant: the t statistic for $\hat{\beta}_{\log(income)}$ is about 5.1 and that for $\hat{\beta}_{prppov}$ is about 2.86 (two-sided p-value = .004).

(iii) The OLS regression results when log($hseval$) is added are

$$\widehat{\log(psoda)} = \underset{(.29)}{-.84} + \underset{(.029)}{.098}\, prpblck - \underset{(.038)}{.053} \log(income)$$

$$+ \underset{(.134)}{.052}\, prppov + \underset{(.018)}{.121} \log(hseval)$$

$n = 401, \quad R^2 = .184$

The coefficient on log($hseval$) is an elasticity: a one percent increase in housing value, holding the other variables fixed, increases the predicted price by about .12 percent. The two-sided p-value is zero to three decimal places.

(iv) Adding log($hseval$) makes log($income$) and $prppov$ individually insignificant (at even the 15% significance level against a two-sided alternative for log($income$), and $prppov$ is does not have a t statistic even close to one in absolute value). Nevertheless, they are jointly significant at the 5% level because the outcome of the $F_{2,396}$ statistic is about 3.52 with p-value = .030. All of the control variables – log($income$), $prppov$, and log($hseval$) – are highly correlated, so it is not surprising that some are individually insignificant.

(v) Because the regression in (iii) contains the most controls, log($hseval$) is individually significant, and log($income$) and $prppov$ are jointly significant, (iii) seems the most reliable. It holds fixed three measure of income and affluence. Therefore, a reasonable estimate is that if the proportion of blacks increases by .10, *psoda* is estimated to increase by 1%, other factors held fixed.