FINM3123 Introduction to Econometrics

Chapter 4 Exercises

1. Are rent rates influenced by the student population in a college town? Let *rent* be the average monthly rent paid on rental units in a college town in the United States. Let *pop* denote the total city population, *avginc* the average city income, and *pctstu* the student population as a percentage of the total population. One model to test for a relationship is

$$\log(rent) = \beta_0 + \beta_1 \log(pop) + \beta_2 \log(avginc) + \beta_3 \, pctstu + u,$$

   i) State the null hypothesis that size of the student body relative to the population has no ceteris paribus effect on monthly rents. State the alternative that there is an effect.
   ii) What signs do you expect for β_1 and β_2?
   iii) The equation estimated using 1990 data for 64 college towns is

$$\widehat{\log(rent)} = \underset{(.844)}{.043} + \underset{(.039)}{.066} \log(pop) + \underset{(.081)}{.507} \log(avginc) + \underset{(.0017)}{.0056} \, pctstu$$

$$n = 64, R^2 = .458.$$

   What is wrong with the statement: "A 10% increase in population is associated with about a 6.6% increase in rent"?
   iv) Test the hypothesis stated in part (i) at the 1% level.

2. Regression analysis can be used to test whether the market efficiently uses information in valuing stocks. For concreteness, let *return* be the total return from holding a firm's stock over the four-year period from the end of 1990 to the end of 1994. The *efficient markets hypothesis* says that these returns should not be systematically related to information known in 1990. If firm characteristics known at the beginning of the period help to predict stock returns, then we could use this information in choosing stocks.

   For 1990, let *dkr* be a firm's debt to capital ratio, let *eps* denote the earnings per share, let *netinc* denote net income, and let *salary* denote total compensation for the CEO.
   i) The following estimated equation was obtained:

$$\widehat{return} = -14.37 + .321\, dkr + .043\, eps - .0051\, netinc + .0035\, salary$$
$$\quad (6.89) \quad (.201) \qquad (.078) \quad (.0047) \qquad (.0022)$$
$$n = 142,\ R^2 = .0395.$$

Test whether the explanatory variables are jointly significant at the 5% level. Is any explanatory variable individually significant?

ii) Now, reestimate the model using the log form for *netinc* and *salary*:

$$\widehat{return} = -36.30 + .327\, dkr + .069\, eps - 4.74 \log(netinc) + 7.24 \log(salary)$$
$$(39.37) \quad (.203) \qquad (.080) \quad (3.39) \qquad (6.31)$$
$$n = 142,\ R^2 = .0330.$$

Do any of your conclusions from part (i) change?

iii) In this sample, some firm have zero debt and others have negative earnings. Should we try to use log(*dkr*) or log(*eps*) in the model to see if these improve the fit? Explain.

iv) Overall, is the evidence for predictability of stock returns strong or weak?

3. The data set 401KSUBS.RData (or 401KSUBS.xls) contains information on net financial wealth (*nettfa*), age of the survey respondent (*age*), annual family income (*inc*), family size (*fsize*), and participation in certain pension plans for people in the United States. The wealth and income variables are both recorded in thousands of dollars. For this question, use only the data for single-person households (so *fsize* = 1).

i) How many single-person households are there in the data set?

ii) Use OLS to estimate the model

$$nettfa = \beta_0 + \beta_1 inc + \beta_2 age + u,$$

and report the results using the usual format. Be sure to use only the single-person households in the sample. Interpret the slope coefficients. Are there any surprises in the slope estimates?

iii) Does the intercept from the regression in part (ii) have an interesting meaning? Explain.

iv) Find the *p*-value for the test $H_0{:}\beta_2 = 1$ against $H_1{:}\beta_2 < 1$. Do you reject H_0 at the 1% significance level?

v) If you do a simple regression of *nettfa* on *inc*, is the estimated coefficient on *inc* much different from the estimate in part (ii)? Why or why not?

4. Use the data in DISCRIM.RData (or DISCRIM.xls) to answer this question.

   i) Use OLS to estimate the model

   $$\log(psoda) = \beta_0 + \beta_1\, prpblck + \beta_2 \log(income) + \beta_3\, prppov + u,$$

   and report the results in the usual form. Is $\widehat{\beta}_1$ statistically different from zero at the 5% level against a two-sided alternative? What about at the 1% level?

   ii) What is the correlation between log(*income*) and *prppov*? Is each variable statistically significant in any case? Report the two-sided *p*-values.

   iii) To the regression in part (i), add the variable log(*hseval*). Interpret its coefficient and report the two-sided *p*-value for H_0: $\beta_{\log(hseval)} = 0$.

   iv) In the regression in part (iii), what happens to the individual statistical significance of log(*income*) and *prppov*? Are these variables jointly significant? (Compute a *p*-value.) What do you make of your answers?

   v) Given the results of the previous regressions, which one would you report as most reliable in determining whether the racial makeup of a zip code influences local fast-food prices?