FINM3123 Introduction to Econometrics

Chapter 6 Exercises

1. Suppose we want to estimate the effects of alcohol consumption ($alcohol$) on college grade point average ($colGPA$). In addition to collecting information on grade point averages and alcohol usage, we also obtain attendance information (say, percentage of lectures attended, called $attend$). A standardized test score (say, SAT) and high school GPA ($hsGPA$) are also available.

   (a) Should we include $attend$ along with $alcohol$ as explanatory variables in a multiple regression model? (Think about how you would interpret $\beta_{alcohol}$.)

   (b) Should SAT and $hsGPA$ be included as explanatory variables? Explain.

2. Data were collected from a random sample of 220 home sales from a community in 2003. Let $Price$ denote the selling price (in \$1000), BDR denote the number of bedrooms, $Bath$ denote the number of bathrooms, $Hsize$ denote the size of the house (in square feet), $Lsize$ denote the lot size (in square feet), Age denote the age of the house (in years), and $Poor$ denote a binary variable that is equal to 1 if the condition of the house is reported as "poor." An estimated regression yields

$$\begin{aligned}\hat{Price} &= 119.2 + 0.485BDR + 23.4Bath + 0.156Hsize + 0.002Lsize \\ &\quad +0.090Age - 48.8Poor, \bar{R}^2 = 0.72,\end{aligned}$$

   (a) Suppose that a homeowner converts part of an existing family room in her house into a new bathroom. What is the expected increase in the value of the house?

   (b) Suppose that a homeowner adds a new bathroom to her house, which increases the size of the house by 100 square feet. What is the expected increase in the value of the house?

   (c) What is the loss in value if a homeowner lets his house run down so that its condition becomes "poor"?

   (d) Compute the R^2 for the regression.

3. The data set NBASAL.RData (or NBASAL.xls) contains salary information and career statistics for 269 players in the National Basketball Association (NBA).

   (a) Estimate a model relating points-per-game ($points$) to years in the league ($exper$), age, and years played in college ($coll$). Include a quadratic in $exper$; the other variables should appear in level form. Report the results in the usual way.

   (b) Holding college years and age fixed, at what value of experience does the next year of experience actually reduce point-per-game? Does this make sense?

   (c) Why do you think $coll$ has a negative and statistically significant coefficient? (*Hint:* NBA players can be drafted before finishing their college careers and even directly out of high school.)

(d) Add a quadratic in *age* to the equation. Is it needed? What does this appear to imply about the effects of age, once experience and education are controlled for?

(e) Now regress $\log(wage)$ on $points, exper, exper^2, age$, and $coll$. Report the results in the usual format.

(f) Test whether *age* and *coll* are jointly significant in the regression from part (e). What does this imply about whether age and education have separate effects on wage, once productivity and seniority are accounted for?

4. Use the subset of 401KSUBS.RData (or 401KSUBS.xls) with $fsize = 1$; this restricts the analysis to single-person households.

(a) What is the youngest age of people in this sample? How many people are at that age?

(b) In the model

$$nettfa = \beta_0 + \beta_1 inc + \beta_2 age + \beta_3 age^2 + u,$$

what is the literal interpretation of β_2? By itself, is it of much interest?

(c) Estimate the model from part (b) and report the results in standard form. Are you concerned that the coefficient on *age* is negative? Explain.

(d) Because the youngest people in the sample are 25, it makes sense to think that, for a given level of income, the lowest average amount of net total financial assets is at age 25. Recall that the partial effect of *age* on $nettfa$ is $\beta_2 + 2\beta_3 age$, so the partial effect at age 25 is $\beta_2 + 2\beta_3(25) = \beta_2 + 50\beta_3$; call this θ_2. Find $\hat{\theta}_2$ and obtain the two-sided $p-$value for testing $H_0 : \theta_2 = 0$. You should conclude that $\hat{\theta}_2$ is small and very statistically insignificant. [*Hint*: One way to do this is to estimate the model $nettfa = \alpha_0 + \beta_1 inc + \theta_2 age + \beta_3 (age - 25)^2 + u$, where the intercept, α_0, is different from β_0. There are other ways, too.]

(e) Because the evidence against $H_0 : \theta_2 = 0$ is very weak, set it to zero and estimate the model

$$nettfa = \alpha_0 + \beta_1 inc + \beta_3 (age - 25)^2 + u.$$

In terms of goodness-of-fit, does this model fit better than that in part (b)?

(f) For the estimated equation in part (e), set $inc = 30$ (roughly, the average value) and graph the relationship between $nettfa$ and *age*, but only for $age \geq 25$. Describe what you see.

(g) Check to see whether including a quadratic in *inc* is necessary.