Review Exercises

July 24, 2009, 10:00-12:00

Question 1 (Textbook 7.4)
An equation explaining chief executive officer salary is
\[
\log(salary) = 4.59 + .257 \log(sales) + .011 \text{ roe} + .158 \text{ finance} \\
\quad + .181 \text{ consprod} - .283 \text{ utility} \\
\quad (.30) \quad (.032) \quad (.004) \quad (.089) \\
\quad (.085) \quad (.099)
\]
\[n = 209, R^2 = .357.\]

The data used are in CEOSAL1.RAW, where finance, consprod, and utility are binary variables indicating the financial, consumer products, and utilities industries. The omitted industry is transportation.

(i) Compute the approximate percentage difference in estimated salary between the utility and transportation industries, holding sales and roe fixed. Is the difference statistically significant at the 1% level?

(ii) Use equation \(100 \cdot [\exp(\beta_1) - 1]\) to obtain the exact percentage difference in estimated salary between the utility and transportation industries and compare this with the answer obtained in part (i).

(iii) What is the approximate percentage difference in estimated salary between the consumer products and finance industries? Write an equation that would allow you to test whether the difference is statistically significant.

Remark: Generally, if \(\beta_j\) is the coefficient on a dummy variable, say \(X_j\), when \(\log(y)\) is the dependent variable, the exact percentage difference in the predicted \(y\) when \(X_j=1\) versus when \(X_j=0\) is \(100 \cdot [\exp(\beta_1) - 1]\).

Question 2 (Textbook 7.3)
Using the data in GPA2.RAW, the following equation was estimated:
\[
sat = 1028.10 + 19.30 hsize - 2.19 hsize^2 - 45.09 female \\
\quad (6.29) \quad (3.83) \quad (0.53) \quad (4.29) \\
- 169.81 black + 62.31 female-black \\
\quad (12.71) \quad (18.15)
\]
\[n = 4137, R^2 = .0858.\]
The variable sat is the combined SAT score, \( hsize \) is size of the student’s high school graduating class, in hundreds, female is a gender dummy variable, and black is a race dummy variable equal to one for blacks, and zero otherwise.

(i) Is there strong evidence that \( hsize^2 \) should be included in the model? From this equation, what is the optimal high school size?

(ii) Holding \( hsize \) fixed, what is the estimated difference in SAT score between nonblack females and nonblack males? How statistically significant is this estimated difference?

(iii) What is the estimated difference in SAT score between nonblack males and black males? Test the null hypothesis that there is no difference between their scores, against the alternative that there is a difference.

(iv) What is the estimated difference in SAT score between black females and nonblack females? What would you need to do to test whether the difference is statistically significant?

**Question 3 (Textbook 8.5)**

The variable \( smokes \) is a binary variable equal to one if a person smokes, and zero otherwise. Using the data in SMOKE.RAW, we estimate a linear probability model for \( smokes \):

\[
smokes = 0.656 - 0.069 \log(cigpric) + 0.012 \log(income) - 0.029 \text{educ} \\
\quad \quad (0.855) (0.204) \quad (0.026) \quad (0.006) \\
\quad \quad [0.856] [0.207] \quad [0.026] \quad [0.006] \\
\quad + 0.020 \text{age} - 0.00026 \text{age}^2 - 0.101 \text{restaurn} - 0.026 \text{white} \\
\quad \quad (0.006) \quad (0.00006) \quad (0.039) \quad (0.052) \\
\quad \quad [0.005] \quad [0.00006] \quad [0.038] \quad [0.050] \\
n = 807, R^2 = 0.062.
\]

The variable white equals one if the respondent is white, and zero otherwise; the other independent variables are defined as

- \( cigs = \text{number of cigarettes smoked per day} \)
- \( income = \text{annual income} \)
- \( cigpric = \text{the per pack price of cigarettes (in cents)} \)
- \( educ = \text{years of schooling} \)
- \( age = \text{measured in years} \)
- \( restaurn = \text{a binary indicator equal to unity if the person resides in a state with restaurant smoking restrictions} \)

Both the usual and heteroskedasticity-robust standard errors are reported.

- a. Are there any important differences between the two sets of standard errors?
- b. Holding other factors fixed, if education increases by four years, what happens to the estimated probability of smoking?
- c. At what point does another year of age reduce the probability of smoking?
- d. Interpret the coefficient on the binary variable \( restaurn \) (a dummy variable equal to one if the person lives in a state with restaurant smoking restrictions).
Person number 206 in the data set has the following characteristics: \( \text{cigpric} = 67.44 \), \( \text{income} = 6,500 \), \( \text{educ} = 16 \), \( \text{age} = 77 \), \( \text{restaurn} = 0 \), \( \text{white} = 0 \) and \( \text{smokes} = 0 \). Compute the predicted probability of smoking for this person and comment on the result.

**Question 4 (Textbook 11.6):**

Let \( \text{hy6}_t \) denote the three-month holding yield (in percent) from buying a six-month T-bill at time \((t-1)\) and selling it at time \( t \) (three months hence) as a three-month T-bill. Let \( \text{hy3}_{t-1} \) be the three-month holding yield from buying a three-month bill at time \((t-1)\). At time \((t-1)\), \( \text{hy3}_{t-1} \) is known, whereas \( \text{hy6}_t \) is unknown because \( p_3 \) (the price of three-month T-bills) is unknown at time \((t-1)\). The expectations hypothesis (EH) says that these two different three-month investments should be the same, on average. Mathematically, we can write this as a conditional expectation:

\[
\mathbb{E}(\text{hy6}_t | I_{t-1}) = \text{hy3}_{t-1}
\]

where \( I_{t-1} \) denotes all observable information up through time \( t-1 \). This suggests estimating the model

\[
\text{hy6}_t = \beta_0 + \beta_1 \text{hy3}_{t-1} + u_t
\]

and testing \( H_0: \beta_1 = 1 \)(we can also test \( H_0: \beta_0 = 0 \), but we often allow for a term premium for buying assets with different maturities, so that \( \beta_0 \neq 0 \))

(i) Estimating the previous equation by OLS using the data in INTQRT.RAW (spaced every three months) gives

\[
\text{hy6}_t = -0.058 + 1.104 \text{hy3}_{t-1}
\]

\[
\begin{align*}
(0.07) & \quad (0.039) \\
n=123, \; R^2 = .866
\end{align*}
\]

Do you reject \( H_0: \beta_1 = 1 \) against \( H_0: \beta_1 \neq 1 \) at the 1% significance level? Does the estimate seem practically different from one?

(ii) Another implication of the EH is that no other variables dated as \((t-1)\) or earlier should help explain \( \text{hy6}_t \), once \( \text{hy3}_{t-1} \) has been controlled for. Including one lag of the spread between six-month and three-month, T-bill rates gives

\[
\text{hy6}_t = -0.123 + 1.053 \text{hy3}_{t-1} + 0.48(r_6_{t-1} - r_3_{t-1})
\]

\[
\begin{align*}
(0.67) & \quad (0.039) & \quad (0.109)
\end{align*}
\]
n=123, $R^2=.885$

Now is the coefficient on $hy_{3,t-1}$ statistically different from one? Is the lagged spread term
significant? According to this equation, if, at time $(t-1)$, $r6$ is above $r3$, should you invest in
six-month or three-month, T-bills?

(iii) The sample correlation between $hy3$ and $hy_{r3,t-1}$ is .914. Why might this raise some
concerns with the previous analysis?

(iv) How would you test for seasonality in the equation estimated in part (ii)?