IAC801. Homework 2, Due: May 15, 6pm

April 30, 2020

Qestion 1. Multiple regression model (10 points, each 1 point unless indicated otherwise)

This question is an exercise of interpreting multiple regression result.

(1) Provide population model for the results below.

Dependent Variable: log(*salary*)			
Independent Variables	**(1)**	**(2)**	**(3)**
log(*sales*)	.224 (.027)	.158 (.040)	.188 (.040)
log(*mktval*)	——	.112 (.050)	.100 (.049)
profmarg	——	−.0023 (.0022)	−.0022 (.0021)
ceoten	——	——	.0171 (.0055)
comten	——	——	−.0092 (.0033)
intercept	4.94 (0.20)	4.62 (0.25)	4.57 (0.25)
Observations	177	177	177
R-squared	.281	.304	.353

The variable mktval is market value of the firm, profmarg is profit as a percentage of sales, ceoten is years as CEO with the current company, and comten is total years with the company.

(2) Interpret on the effect of profmarg on CEO salary. Use both log approximation as well as precise formula to get the effect.

(3) Does market value have a significant effect? Explain.

(4) Interpret the coefficients on ceoten and comten. Are these explanatory variables statistically significant?

(5) What do you make of the fact that longer tenure with the company, holding the other factors fixed, is associated with a lower salary?

(6) Also graphically show the relationship between comten and log(salary).

(7) Compared the *R*-squared values between columns (2) and (3) in the above table.

(8) Provide 95% conf. interval of $\hat{\beta}_{ceoten}$.

(9) Explain omitted variable bias by comparing the results in columns (1) and (2).

(10) Suppose progmarg is not relevant in the explanation of outcome. What is the sign of corr(log(sales), log(market))? Justify your answer.

Qestion 2. Multiple regression with quadratic term (15 points, each 1 point unless indicated otherwise)

Use the data in HTV.RAW to answer this question.

(1) Estimate the regression model:

$$eudc = \beta_0 + \beta_1 motheduc + \beta_2 fatheduc + \beta_3 abil + \beta_4 abil^2 + u$$

by OLS and report the results in the usual form.

(2) Generate the result table as Question 1. In the firmst column estimate the simple regression model where abil is the only explanatory variable. In the second comlumn, additionally include mother and father education variables. In the third column, includ all explanatory variables.

(3) Test the null hypothesis that educ is linearly related to abil against the alternative that the relationship is quadratic.

(4) Using the results in column (3), test $H_0 : \beta_1 = \beta_2$ against a two-sided alternative. What is the p-value of the test?

(5) Add the two college tuition variables to the regression from the results in column (3) and determine whether they are jointly statistically significant.

(6) What are the determinants of variance of OLS estimator?

(7) What is the correlation between tuit17 and tuit18? Explain why using the average of the tuition over the two years might be preferred to adding each separately. What happens when you do use the average?

(8) Do the findings for the average tuition variable in part (iv) make sense when interpreted causally? What might be going on?

(9) What is the marginal effect of ability?

(10) Report the result with beta coefficient. Explain about beta coefficient. Interpret the results for the marginal effect of abil.

(11) How many different values are taken on by educ in the sample? Does educ have a continuous distribution?

(12) Plot a histogram of educ with a normal distribution overlay.

(13) Does the distribution of educ appear anything close to normal?

(14) What are the CLM assumptions(A1~A5 and one more assumption)?

(15) Which of the CLM assumptions seems clearly violated in the model?

Qestion 3. Log transformation and interaction terms (10 points, each 1 point unless indicated otherwise)

Consider the log(price) equation we used for multiple regression:

$$log(price) = \beta_0 + \beta_1 log(lotsize) + \beta_2 log(sqrft) + \beta_3 bdrms + \beta_4 bdrms * log(lotsize) + u \quad (1)$$

Use the housing price data in HPRICE1.RAW for this exercise

(1) Estimate the model using OLS regression.

(2) Report the results in the usual OLS format and generate the table as in Question 1. In column (1), use only bdrms as explanatory variable. In column (2), include all except interaction term as explanatory variables. In column (3), include all explanatory variables.

(3) Interpret β_0, β_1, β_2 and β_3.

(4) Obtain the marginal effect of increasing the number of bdrms.

(5) To interpret β_1 as 1% increase in lotsize will increase price by β_1% for bdrms=3, transform variable in equation (1). Report the results in the usual OLS format

(6) Find the predicted value of log(price), when lotsize=20,000, sqrft=2,500, and bdrms=4.

(7) Using the follwing methods, find the predicted value of price at the same values of the explanatory variables.

$$\hat{y} = \hat{\alpha}_0 exp(\hat{logy}) \text{ and } \hat{\alpha}_0 = \frac{\Sigma_{i=1}^{n} exp(\hat{u}_i)}{n}$$

(8) For explaining variation in price, estimate the following model using OLS regression.

$$price = \beta_0 + \beta_1 lotsize + \beta_2 sqrft + \beta_3 bdrms + \beta_4 bdrms * lotsize + u \tag{2}$$

(9) Decide whether you prefer the model from equation (1) or the model of equation (2). Justify your answer.

(10) Obtain predicted price, when we plug in lotsize=10,000, sqrft=2,300, and bdrms=4; round this price to the nearest dollar.

Qestion 4. Effect of Attendance on Final Exam Performance (10 points)

A model to explain the standardized outcome on a final exam (stndfnl) in terms of percentage of classes attended, prior college grade point average, and ACT score is

$$\begin{aligned} stndfnl = \beta_0 + \beta_1 atndrte + \beta_2 priGPA + \beta_3 ACT + \beta_4 priGPA^2 \\ + \beta_5 ACT^2 + \beta_6 atndrte * priGPA + u \end{aligned} \tag{3}$$

(1) Estimate the model using OLS regression. Make sure using the standardized exam score for the reasons discussed in the lecture in April 29th: it is easier to interpret a student's performance relative to the rest of the class.

(2) The idea is that class attendance might have a different effect for students who have performed differently in the past. Which term is included to refelct this heterogeneity?

(3) Provide $\frac{\Delta stndfnl}{\Delta atndrte}$. Interpret this object.

(4) What is the effect of class attendance on final exam for students who are 1 SD below the mean.

(5) Provide $\frac{\Delta stndfnl}{\Delta priGPA}$. What is the partial effect when priGPA=2.59 and atndrte=82. Interpret your estimate.

(6) Estimate the model using OLS regression.

$$\begin{aligned} stndfnl = \beta_0 + \beta_1 atndrte + \beta_2 priGPA + \beta_3 ACT + \beta_4 (priGPA - 2.59)^2 \\ + \beta_5 ACT^2 + \beta_6 (atndrte - 82) * priGPA + u \end{aligned} \tag{4}$$

(7) Let $\theta_2 = \beta_2 + 2\beta_4(2.59) + \beta_6(82)$. Use this to obtain the standard error of $\hat{\theta}_2$from question (5).

(8) Suppose that, in place of priGPA(atndrte-82), you put (priGPA -2.59)*(atndrte-82). Now how do you interpret the coefficients on atndrte and priGPA?

$$\begin{aligned} stndfnl = \beta_0 + \beta_1 atndrte + \beta_2 priGPA + \beta_3 ACT + \beta_4 (priGPA - 2.59)^2 \\ + \beta_5 ACT^2 + \beta_6 (atndrte - 82) * (priGPA - 2.59) + u \end{aligned} \tag{5}$$

(9) Estimate in Stata to interpret all coefficient as beta coefficient.

(10) Consider the following regression model. Which one do you prefer between (5) and (6)? Justify your answer.

$$\begin{aligned} log(stndfnl) = \beta_0 + \beta_1 atndrte + \beta_2 priGPA + \beta_3 ACT + \beta_4 (priGPA - 2.59)^2 \\ + \beta_5 ACT^2 + \beta_6 (atndrte - 82) * (priGPA - 2.59) + u \end{aligned} \tag{6}$$