Problem Set 4 (Multivariate regression 2)

ECON 441, 2019 Fall

Please solve all problems below, and do not forget to submit the Stata outputs somehow (copy here, print it separately, or upload it as an attachment to the submission on Canvas). This is NOT a group exercise, so everyone needs to write up their own solutions.

1 Wooldridge, Chapter 4, Exercise 2 (20 pts)

Consider an equation to explain salaries of CEOs in terms of annual firm sales, return on equity (*roe*, in percentage points), and return on the firm's stock (*ros* in percentage points):

$$\log(salary) = \beta_0 + \beta_1 \log(sales) + \beta_2 roe + \beta_3 ros + U. \tag{1}$$

1. In terms of model parameters ($\sim$ the β-coefficients) state the null hypothesis that, after controlling for *sales* and *roe*, *ros* has no effect on CEO salary. State the alternative that better stock market performance *increases* (!) CEO salary.

2. Using CEOSAL1, the following equation was obtained by OLS:

$$\widehat{\log(salary)} = \underset{(0.31)}{4.32} + \underset{(0.033)}{0.310 \log(sales)} + \underset{(0.0045)}{0.0191 roe} + \underset{(0.00059)}{0.00022 ros}$$

   You can see the corresponding (heteroskedasticity-robust/White) standard errors in parentheses below the coefficients. In addition, you know from the output that $n = 210$ and $R^2 = 0.291$. By what percentage is *salary* predicted to increase if *ros* increases by 50 percentage points? (Answer with a whole interpretation sentence!) Does *ros* have a practically large effect on *salary*? (To phrase the question differently: Is the effect economically significant?)

3. Test the null hypothesis that *ros* has no effect on salary against the alternative that *ros* has a positive effect. Carry out the test on the 10% significance level. (Help to get the critical value: $\Phi^{-1}(0.9) = 1.282, \Phi^{-1}(0.8) = 0.842, \Phi^{-1}(0.95) = 1.645$.)

4. Would you include *ros* in a final model explaining CEO compensation in terms of firm performance? Explain.

2 Wooldridge, Chapter 4, Exercise 11 (10 pts)

SEE BOOK (pp 144-145). "Comment on the effect of..." means that you should look at the sign of the coefficient (if it is what you would expect from your economic intuition), but also look at economic and statistical significance.

The following table was created using the data in CEOSAL2, where standard errors are below the coefficients in parentheses. The independent variable is $\log(salary)$.

Independent variables	(1)	(2)	(3)
log($sales$)	0.224	0.158	0.188
	(0.027)	(0.040)	(0.040)
log($mktval$)	—	0.112	0.100
		(0.050)	(0.049)
$Profmarg$	—	-0.0023	-0.0022
		(0.0022)	(0.0021)
$Ceoten$	—	—	0.0171
			(0.0055)
$comten$	—	—	-0.0092
			(0.0033)
$intercept/constant$	4.94	4.62	4.57
	(0.20)	(0.25)	(0.25)
Observations	177	177	177
R^2	0.281	0.304	0.353

The variable $mktval$ is the 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.

1. Comment on the effect of $profmarg$ on CEO salary.

2. Does the market values has a significant effect (at 5%, 1% - please use the critical values above). Please do not forget to write down the null and alternative hypothesis. Explain your findings.

3. Interpret the coefficient on *ceoten* and *comten*. Are these explanatory variables statistically significant at 5%, 10%? Please do not forget to write down the null and alternative hypothesis.

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

3 Wooldridge, Chapter 4, Exercise 13 (10 pts)

SEE BOOK (p 144). Again, when he asks "What happens when ..." in part (ii), you need to comment on changes you observe in the sign, magnitude (economic significance) and statistical significance of the other coefficients in the regression.

If we do not specify the significance level of the test (the α), assume it is at the 5% level, also if not mentioned, the test is against the 2-sided alternative by default. We used the data to estimate model parameters that highlight the relationship between educational outcomes in 4th grade and socio-economic variables. We have 4 model specifications below. The LHS variable is the 4th grade score on the math test, $math4$.

Independent variables	(1)	(2)	(3)	(4)
pctsgle	-0.833 (0.071)	-0.275 (0.0.117)	-0.274 (0.161)	-0.259 (0.117)
$free$	—	-0.402 (0.070)	-0.422 (0.071)	-0.420 (0.70)
lmedinc	—	—	-0.752 (5.358)	—
lexpp	—	—	9.01 (4.04)	8.80 (3.76)
intercept/constant	96.77 (1.60)	93.00 (1.63)	24.49 (59.24)	17.52 (32.25)
Observations R^2	299 0.380	299 0.459	299 0.472	299 0.472

1. Interpret the coefficient on the variable *prctsgle* in the first equation. Comment on what happens when $free$ is added as an explanatory variable.

2. Does expenditure per pupil, entered in a logarithmic form, have a statistically significant effect on performance? Interpret the coefficient estimate in the 3rd model. Is it economically significant ($\sim$ meaningful)?

3. If you had to choose among the four equations as your best estimate of the effect of *pctsgle* and obtain a 95% confidence interval of $\beta_{pctsgle}$, which one would you choose? Write down the 95% confidence interval. Is the coefficient estimate significantly larger than zero at 5%? (Write down the test steps.)

4 Wooldridge, Chapter 4, Exercise C9 (30 pts)

DO NOT FORGET TO USE WHITE STANDARD ERRORS FOR THIS EXERCISE! Use the DISCRIM data set to answer this question. These are ZIP-code level data on prices for various items at fast-food restaurants, along with characteristics of the zip code population, in New Jersey and Pennsylvania. The idea is to see whether fast food restaurants charge higher prices in areas with higher concentration of African-American inhabitants.

1. Estimate the following simple linear 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 $\hat{\beta}_1$ statistically significant at 5% against a 2-sided alternative? (Include a very brief description how you decided.) What about the 1% level?

2. What is the correlation between log($income$) and $prppov$? (Intuition fine, if do not want to use the Stata code. Give sign and strength.) Is each variable statistically significant in any case? Report the two-sided p-values? (=the p-values)

3. To the regression in part 1., add the variable log($hseval$). (Do not forget to check what this variable means.) Interpret this coefficient and report the 2-sided p-value for $H_0 : \beta_{\log(hseval)} = 0$.

4. In the regression part in 3., what happens to the individual statistical significance of log($income$) and $prppov$? Are these variables jointly significant? (Compute a p-value - meaning let Stata compute a p-value by using the right command.) What do you make of your answers?

5. 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? (Do not overthink this, just give a statistic that lets you decide. Also try to use the results above.)

5 Wooldridge, Chapter 4, C11 (30 pts)

DO NOT FORGET TO USE WHITE STANDARD ERRORS FOR THIS EXERCISE! Use data in HTV to answer this question. The data set includes information on wages, education, parents' education and several other variables for 1,230 working men in 1991.

1. Estimate the regression model

$$educ = \beta_0 + \beta_1 motheduc + \beta_2 fatheduc + \beta_3 abil + \beta_4 abil^2 + U$$

by OLS and report the results in the usual form. Test the null hypothesis that *educ* is linearly related to *abil* against the alternative that the relationship is quadratic. (Write up the null, choose the test, tell how you decided and why very briefly and what the decision means in terms of the semi-English question in the previous sentence.)

2. Using the equation in part (i), test $H_0 : \beta_1 = \beta_2$ against the two-sided alternative. What is the p-value of the test?

3. Add the two college tuition variables to the regression from part 1., and determine if they are jointly statistically significant.

4. What is the correlation between $tuit17$ and $tuit18$? (Again, sign + strength is enough from your intuition, but feel free to check the correlation in Stata!) Explain why using the average of the tuition over the two years might be preferred to adding them separately. What happens when you do use the average?

5. Do the findings for the average tuition variable in part 4., make sense when interpreted causally? What might be going on?