ECO 480 Econometrics 1

Take-Home Exam 2

Due Wednesday, 3/10/2021

**Statement of Academic Honesty*:*** *On the first page of your Word-file before you answer the assigned questions, you must type in the following passage, then put your name, sign, and date:*

**For this exam, I make the following truthful statements:**

**1. I have not received any non-instructor approved assistance, I have not given any non-instructor approved assistance to another student taking this exam, including discussing the exam with students in another section of the course.**

**2. I did not plagiarize someone else’s work and turn it in as my own.**

**3. I understand that acts of academic dishonesty may be penalized to the full extent allowed by the University at Buffalo Student Conduct Code, including receiving a failing grade for the course with a transcript notation and being expelled from the university. I recognize that I am responsible for understanding the provisions of the University at Buffalo Student Conduct Code as they relate to this academic exercise.**

**Name: \_\_Hao Wu\_\_\_\_ Sign: \_\_Hao Wu\_\_\_\_\_ Date: \_\_\_\_\_03/08/2021\_\_\_\_**

***Important Note: If you want to receive a grade for your take-home exam, you must have this statement. If you do not have this statement, I will not accept your work and your grade would be zero.***

Instruction: You have one week to complete this exam. This exam consists of data analysis. You may NOT discuss the exam questions with your classmates. No late work will be accepted and all the files must be submitted via UBLearns. Make sure you upload it to the correct submission slot because no credit will be given for incorrect submissions.

Important: It is extremely important to write a clean well-commented program for transparency and replication purposes in *all* empirical work. You should always be able to reproduce your result from raw data to support your claim.

There are 3 items to hand in: (1) Typed write-up (i.e., word-file) answering the assigned questions, reporting your results, and interpreting your findings; if the question asks for graphs or tables, these must be in the word-file in an organized manner with your interpretation, (2) do-file (i.e., program file), and (3) log-file (i.e., output file that shows the results). You MUST use Stata. For questions involving data analysis, you will NOT get any credit if you do not provide a program code and the output. You may not use Excel. Do not submit any undigested log-file that contains errors. Put all your answers in the word file and do NOT say “please see log-file (or do-file) for answers. You will not receive any credit for answers that is not stated in the word file.

1. (10 points) Consider the relationship between variables Q and R. Write down a complete econometric model where if R were to increase from 1 to 2, 2 to 3, or 3 to 4, Q would change by 𝛾, gamma (the same amount in all three cases). If R were to be equal to 0, then we would expect Q to be equal to 𝜃, theta. [Hint: Do not forget the subscripts.]

Solution:

We can briefly know as the R were to increase, Q would change by 𝛾, so that 𝛾 is the slope of our econometric model equation, and the intercept is 𝜃, because if R were to be equal to 0, Q to be equal to 𝜃*.* Q is the dependent variable; gamma is the value the variable R increases by r. The econometric model can be given as:

Q = 𝜃 + 𝛾R

)

1. (15 points) Create your own dataset that has X and Y values with 10 observations.

|  |  |
| --- | --- |
| X | Y |
| 5 | 4 |
| 2 | 1 |
| 7 | 5 |
| 9 | 9 |
| 27 | 18 |
| 10 | 10 |
| 34 | 24 |
| 17 | 13 |
| 20 | 17 |
| 30 | 20 |

1. (10) Manually calculate the OLS regression equation using the data you created. Show your work. [Hint: Just like what we did in class together, but you create your own dataset.]

We need calculate X2 and XY for each (X, Y) point then sum up.

∑X = 161; ∑Y = 121; ∑X2 = 3733; ∑XY = 2701

Slope:

= N ∑(XY) − ∑X ∑Y) ÷ N ∑ (X2) – (∑X)2

= (10×2701−161×121) ÷ (10 × 3733 – 25921)

= (2701 – 19481) ÷ (3733− 25921)

= -16780÷ -22188

= 0.7563

Intercept:

= (∑Y – Slope ∑X) ÷ N

= (121 – 0.7563×161) ÷10

= (121– 121.7643) ÷ 10

= 0.07643

So that the OLS regression equation given as:

Y = 0.07643 + 0.7563X

b. (5) Input your data into Stata and run the regression. Did you get the same regression equation? Why or why not?

|  |  |
| --- | --- |
|  | (1) |
|  | B |
| A | 0.660\*\*\* |
|  | (14.08) |
|  |  |
| \_cons | 1.475 |
|  | (1.63) |
| *N* | 10 |

*t* statistics in parentheses

\* *p* < 0.05, \*\* *p* < 0.01, \*\*\* *p* < 0.001

The equation get from Stata is given as:

Y = 1.475 + 0.66X

That is different with my regression equation.

1. (20 points) Suppose (𝑌𝑖 , 𝑋𝑖) satisfy the three least squares assumptions we covered in class. A random sample of size 𝑛 = 1000 is drawn and yields the following equation:

𝑌̂ = 8.0 + 3.2𝑋

(6.0) (1.5)

1. (5) Test 𝐻0: 𝛽1 = 0 vs. 𝐻1: 𝛽1 ≠ 0 at the 5% level.

If 𝛽1 = 0

t = 𝛽1hat – β1=0 /SE (β1hat) = 3.2/1.5= 2.133

P | t | < 0.05

We can reject *H*0 on 5% level.

X is statistically significant at the 5% level.

1. (5) Construct a 95% confidence interval for 𝛽1.

Confidence interval for 𝛽1 = 3.2 ± (1.96 \*1.5) = [0.26, 6.14]

1. (5) Suppose you learned that 𝑌𝑖 and 𝑋𝑖 were independent. Would you be surprised? Why or why not?

Yes, It make no sense. It means 𝛽1 = 0; however, p value is less than 0.05. I cannot believe I just choose 5% or less of the sample.

1. (10) Suppose 𝑌𝑖 and 𝑋𝑖 are independent and many samples of size 𝑛 = 1000 are drawn, regressions estimated, and (a) and (b) answered. In what fraction of the samples would 𝐻0 from (a) be rejected? In what fraction of samples would the value 𝛽1 = 0 be included in the confidence interval from (b)? Explain your answers.

5% of samples H0 would be rejected, normally 95% confident interval will included the value 𝛽1 = 0

1. (15 points) In the 1980s, Tennessee conducted an experiment in which kindergarten students were randomly assigned to “regular” and “small” classes and given standardized tests at the end of the year. Regular classes contained approximately 24 students, and small classes contained approximately 15 students. Suppose in the population, the standardized tests have a mean score of 925 points and a standard deviation of 75 points. Let Small Class denote a binary variable equal to 1 if the student is assigned to a small class and equal to 0 otherwise and regress Test Score on Small Class.
2. (5) I would argue that the least squares assumption #1 does not hold for this regression. Do you agree? Why or why not?

I do not agree, the least squares assumption #1 is hold for this regression. Students were randomly selected; there has not other factor will affect the independent variable.

1. (5) Do you think that the regression errors are plausibly homoscedastic? Explain.

I cannot predict the regression errors are plausibly homoscedastic or not.

1. c. (5) Suppose the regression errors were homoscedastic and you computed heteroskedasticity-robust standard error. How would this affect the validity of the confidence interval? Explain

confident interval = 𝛽1 ± Z\*Standard error; so that it will not affect the validity of the confidence interval.

1. (35 points) For this question, we will use the dataset **Growth** again. We used this dataset together in class, which contains data on average growth rates from 1960 through 1995 for 65 countries, along with variables that are potentially related to growth. This dataset is already posted in the data folder. A detailed description is given in Growth Description.pdf. As we did in class, we will examine the relationship between growth and trade.
2. (5) As we saw in class, the regression function that includes Malta is steeper than the regression function that excludes Malta. Do a brief research on Malta and explain why the Malta trade share is so large. Based on your research please state whether Malta should be included or excluded from the analysis and explain your reasoning.

Malta should be excluded from the analysis if we make the analysis included malta. The growth and tradeshare will have a significant statistic; however, if we exclude it, the conclusion is totally different. That’s mean Malta trade share is a huge outlier.

1. (5) Exclude the data for Malta and run a regression of *Growth* on *TradeShare.* Interpret the slope coefficient.

|  |  |
| --- | --- |
|  | (1) |
|  | growth |
| tradeshare | 1.681 |
|  | P(0.094) |
|  |  |
| \_cons | 0.957 |
|  | (0.104) |
| *N* | 64 |

*p*-values in parentheses

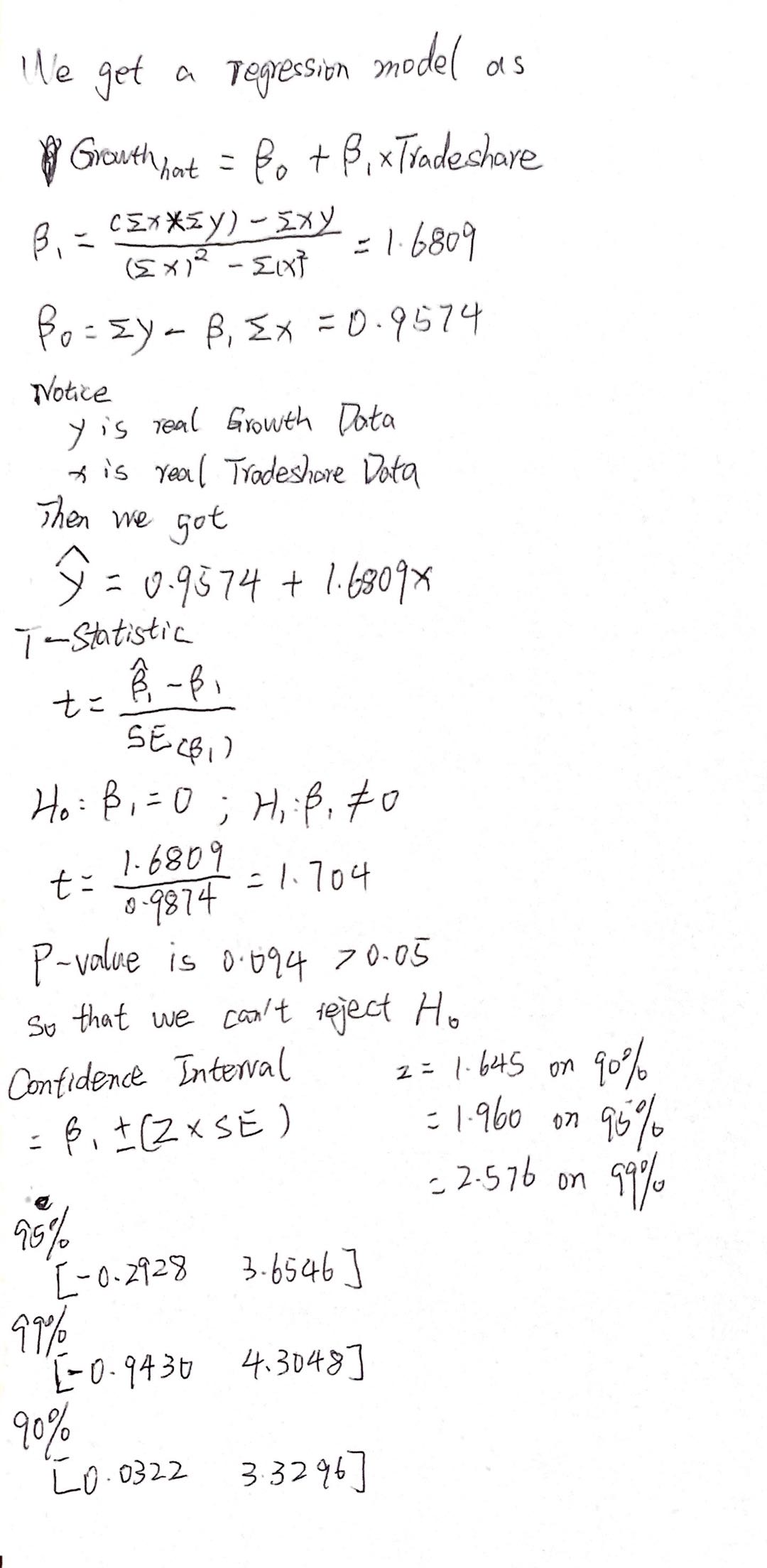
\* *p* < 0.05, \*\* *p* < 0.01, \*\*\* *p* < 0.001

The intercept is 0.957, and the slope coefficient is 1.681. Our confidence Interval is 95%.

1. (5) Is the estimate regression slope statistically significant at the 1%, 5%, or 10% level?

Estimate regression slope statistically significant at the 10% level.

1. (5) Manually calculate the t-statistic. Show your work.
2. (5) Manually construct a 99%, 95%, and 90% confidence interval and interpret these intervals. Show your work.



1. (10) Instead of two-sided alternative hypothesis, suppose that you change your alternative hypothesis 𝐻𝑎: 𝛽1>0. How would your t-statistics, p-value, and conclusion change?

T-statistic is very small, and the P values will become approach to 1. Conclusion doesn’t change because null hypothesis change.

6. [**Extra Credit for ECO 480, Required for ECO 580**] (35 points) Simulation is a powerful methodology for investigating the properties of econometric estimators and tests. The power of the method derives from being able to define and control the statistical environment in which the investigator specifies the data-generating process (DGP) and generate data used to investigate the properties. We are going to use this simulation method to examine the OLS estimator properties we learned in class. [Hint: Understand the Stata code I posted for this question.]

Suppose we are interested in the effect of education on salary as expressed in the following model:

𝑆𝑎𝑙𝑎𝑟𝑦𝑖= 𝛽0+ 𝛽1𝐸𝑑𝑢𝑐𝑎𝑡𝑖𝑜𝑛𝑖+ 𝑢𝑖

For this problem, we are going to assume that the true model is

𝑆𝑎𝑙𝑎𝑟𝑦𝑖=12000+1000 𝐸𝑑𝑢𝑐𝑎𝑡𝑖𝑜𝑛𝑖+𝑢𝑖

The model indicates that the salary for each person is $12,000 plus $1,000 times the number of years of education plus the error term for the individual. Our goal is to explore how much our estimate of 𝛽𝐸𝑑𝑢𝑐𝑎𝑡𝑖𝑜𝑛̂ varies.

I posted a code that will simulate a data set with 100 observations. Values of education for each observation are between 0 and 16 years. The error term will be a normally distributed error term with a standard deviation of 10,000. [Hint: Understand the OLS properties.]

1. a. (5) Explain why the means of the estimated coefficients across the multiple simulations are. what they are.
2. We randomly choose sample, the standard deviation is unchanged; thus the means of the estimated coefficients across the multiple simulations.
3. b. (5) What are the minimum and maximum values of the estimated coefficients on education? Explain whether these values are inconsistent with our statement that OLS estimates are unbiased.

Variable | Obs Mean Std. Dev. Min Max

-------------+---------------------------------------------------------

\_b\_Ed | 50 1003.985 211.1072 457.5943 1544.242

\_b\_cons | 50 12152.51 2031.451 7491.454 16520.59

These values are consistent with our statement that OLS estimates are unbiased. We randomly choose 50 sample.

c. (5) Rerun the simulation with a larger sample size in each simulation. Specifically, set the sample size to 1,000 in each simulation. Compare the mean, minimum, and maximum of the estimated coefficients on education to the original results above. Briefly explain.

Variable | Obs Mean Std. Dev. Min Max

-------------+---------------------------------------------------------

\_b\_Ed | 50 1009.481 68.2596 860.0313 1200.722

\_b\_cons | 50 11894.36 648.8948 10573.36 13389.94

We got more sample means standard deviation will be less and outlier will less.

1. d. (5) Rerun the simulation with a smaller sample size in each simulation. Specifically, set the sample size to 20 in each simulation. Compare the mean, minimum, and maximum of the estimated coefficients on education to the original results above. Briefly explain.

Variable | Obs Mean Std. Dev. Min Max

-------------+---------------------------------------------------------

\_b\_Ed | 50 1140.196 465.6789 116.0028 2062.499

\_b\_cons | 50 10749.31 4329.906 1695.739 18838.25

Less sample size will produce large outlier, interval is huge.

1. e. (5) Reset the sample size to 100 for each simulation, and rerun the simulation with a smaller standard deviation (equal to 500) for each simulation. Compare the mean, minimum, and maximum of the estimated coefficients on education to the original results above. Briefly explain.

Variable | Obs Mean Std. Dev. Min Max

-------------+---------------------------------------------------------

\_b\_Ed | 50 1000.94 8.603732 983.2396 1022.538

\_b\_cons | 50 11989.66 90.11584 11822.84 12164.55

Standard deviation become less mean the interval of number which we choose is smaller.

1. f. (5) Keeping the sample size at 100 for each simulation, rerun the simulation with a larger standard deviation for each simulation. Specifically, set the standard deviation to 50,000 for each simulation. Compare the mean, minimum, and maximum of the estimated coefficients on education to the original results above. Briefly explain.

Variable | Obs Mean Std. Dev. Min Max

-------------+---------------------------------------------------------

\_b\_Ed | 50 1009.834 1092.942 -1907.754 4110.02

\_b\_cons | 50 11990.23 10197.74 -14523.34 38639.11

Absolutely, the interval of number which we choose is larger, because larger standard deviation for simulation.

1. g. (5) Revert to original model (sample size at 100 and standard deviation at 10,000). Now run 500 simulations. Summarize the distribution of the 𝛽𝐸𝑑𝑢𝑐𝑎𝑡𝑖𝑜𝑛̂ estimates as you’ve done so far, but now also plot the distribution of these coefficients using code provided. Describe the density plot in your own words.

Variable | Obs Mean Std. Dev. Min Max

-------------+---------------------------------------------------------

\_b\_Ed | 500 1006.871 216.8483 244.767 1717.585

b\_cons | 500 11920.24 2050.226 5476.07 18438.87