Sample Exam for EC 306

BRIEF SKETCHES OF ANSWERS GIVEN IN BOLD

This exam contains four questions. Please do all questions.

1. *Empirical Practice (20%)*

In a study of costs in the banking industry, data has been collected for 85 banks in America. Some banks specialize in lending money to households, other banks specialize in lending money to businesses. Similar issues hold with respect to their depositors. I am interested in investigating whether these bank characteristics affect their costs. Accordingly, I define my dependent and explanatory variables as follows:

Y = total costs per employee (in thousands of dollars per year).
X_1 = proportion of total loans which go to businesses (measured as a percentage so that a value of, say, 20 means 20% of loans are made to businesses).
X_2 = proportion of total deposits which come from households (measured as a percentage so that a value of, say, 20 means 20% of deposits come from households).
D = a dummy variable which equals 1 if the bank is a big bank (has more than 100 employees), = 0 otherwise.
I also constructed another variable, $Z = X_2 \times D$.

a) I ran a regression of Y on X_1, X_2, D and Z. Results from this regression are given below in the following fitted regression line:

Y =	960	$-109\times X_1$	$+ 120\times X_2$	$- 1.49\times D$	$-23\times Z$
	(6×10^{-7})	(0.008)	(0.04)	(0.03)	(0.002)

where the numbers in parentheses are P-values for testing the hypothesis that the coefficient equals zero.

i) How would you interpret (in words) the estimated coefficients in this model? What is the OLS estimate of the marginal effect of X_2 on Y?

The interpretation of any coefficient is: "if the explanatory variable changes by one unit, then the dependent variable tends to change by [insert coefficient here] units, holding other explanatory variables constant"

(getting units right and ceteris paribus idea is important for first class grade)

For dummy variable (D) this can be refined in the usual way: "individuals with D=1 have a regression with intercept 1.49 less than individuals with D=0" or in the lectures I used expected value operator to write out expected Y for D=1 and 0.

The presence of the interaction variable means the marginal effect of X2 on Y (ceteris paribus) is different for companies with D=0 versus D=1. To be precise, this marginal effect is 97 for big banks and 120 for small banks.

ii) Which of the statements you have just made are statistically significant at the 5% level? Which are significant at the 1% level?

All explanatory variables are significant at the 5% level, only intercept X1 and Z are significant at 1% level.

b) I then did a White test (using X_2 as the independent variable to explain the heteroskedasticity) and found a test statistic value of 5.02 (with a p-value of .025). I re-ran the previous regression using a heteroskedasticity consistent estimator (HCE) and obtained:

Y =	960	$-109\times X_1$	$+ 120\times X_2$	$- 10{,}449\times D$	$-203\times Z$
	(4×10^{-5})	(0.023)	(0.067)	(0.04)	(0.005)

where the numbers in parentheses are P-values for testing the hypothesis that the coefficient equals zero.

i) When presenting final results in a project, would you use my OLS results of part a) or my HCE results of part b)? Why?

Since the p-value for the heteroskedasticity test is less than .05, we accept the hypothesis that heteroskedasticity is present. Therefore, the variance of the OLS estimator in part a) was incorrect (and, although estimates were unbiased, p-values were wrong). So HCE (which uses the correct formula for the variance) will be better and should be used.

2. *Econometric Theory: Derivations and Proofs (30%)*

a) Consider the simple regression model with a single explanatory variable under the assumptions (for i=1,..,N):

$$y_i = \beta X_i + \varepsilon_i$$
$$E(\varepsilon_i) = 0$$
$$\text{var}(\varepsilon_i) = \sigma^2 \omega_i^2$$

and the errors are uncorrelated with one another. X_i is not a random variable.

The OLS estimator for β is given by:

$$\hat{\beta} = \frac{\sum_{i=1}^{N} X_i y_i}{\sum_{i=1}^{N} X_i^2}$$

i) Calculate the expected value of the OLS estimator. Is this estimator unbiased?

THIS IS STANDARD DERIVATION DONE IN LECTURES.
WRITE ESTIMATOR AS

$$\hat{\beta} = \beta + \frac{\sum_{i=1}^{N} X_i \varepsilon_i}{\sum_{i=1}^{N} X_i^2}$$

TAKE EXPECTED VALUE OF BOTH SIDES AND NOTE THAT, SINCE X IS NOT RANDOM, THE EXPECTED VALUE OF THE LAST TERM IS ZERO. HENCE OLS IS UNBIASED (EVEN UNDER HETEROSKEDASTICITY).

ii) Calculate the variance of this estimator. What does the Gauss-Markov theorem tell us about the size of this variance relative to the size of the variance of the generalised least squares (GLS) estimator?

TAKE VARIANCE OPERATOR OF BOTH SIDES OF EQUATION ABOVE, YOU GET

$$Var\hat{\beta} = \frac{\sum_{i=1}^{N} X^2{}_i \sigma_i^2}{\left(\sum_{i=1}^{N} X_i^2\right)^2}$$

GAUSS MARKOV THEOREM TELLS US (UNDER THE ASSUMPTIONS OF THIS QUESTION) THE GLS ESTIMATOR HAS THE SMALLEST VARIANCE OF ALL LINEAR UNBIASED ESTIMATORS. SINCE OLS IS LINEAR AND UNBIASED, IT MUST HAVE A LARGER VARIANCE THAN GLS.

b) Using the setup, assumptions and definitions given for part a) except that now there is no heteroskedasticity so that var(ε_i)=σ^2. Suppose an estimator for β is given by:

$$\tilde{\beta} = \frac{\sum_{i=1}^{N} X_i^2 y_i}{\sum_{i=1}^{N} X_i^3}$$

i) Calculate the expected value of this estimator. Is it an unbiased estimator of β?

DESPITE THE DIFFERENT SETUP, THE PROOF IS ACTUALLY QUITE SIMILAR TO THAT TO PART A, i). I WILL LEAVE YOU TO WORK IT OUT (IT IS AN UNBIASED ESTIMATOR)

ii) Is this estimator more efficient that the OLS estimator?

NO. THE GAUSS MARKOV THEOREM TELLS US THIS (SEE ANSWER TO PART A,ii))

3. *Understanding Econometric Theory (25%)*

a) Define the term "multicollinearity" and explain its importance for empirical practice..

TEXTBOOK MATERIAL. WRITE A (VERY BRIEF) ESSAY SUMMARIZING THIS MATERIAL.

b) Define the term "instrumental variable". Explain the importance of this concept for regression analysis.

TEXTBOOK MATERIAL. WRITE A (VERY BRIEF) ESSAY SUMMARIZING THIS MATERIAL.

4. *Time Series Econometrics (25%)*

I have collected data on two time series variables, X_t and Y_t and run various regressions using this data. Excel outputs for these regressions are below and labelled as "OUTPUT 1", "OUTPUT 2", "OUTPUT 3" and "OUTPUT 4". To be specific:

- OUTPUT 1 contains results from a regression of ΔY on one lag of Y. That is,

$$\Delta Y_t = \alpha + \beta \times Y_{t-1} + e_t .$$

- OUTPUT 2 contains results from a regression of ΔX on one lag of X.
- OUTPUT 3 contains results from the simple regression of Y on X.
- OUTPUT 4 takes the residuals, e, from the regression of Y on X (i.e. the one in OUTPUT 3) and regresses Δe on one lag of e.

i) Define and describe the Dickey-Fuller test. Can this test be done using any of the OUPTPUTS above? If yes, what does the Dickey-Fuller test tell you about the properties of Y and Y? You may assume that the 5% critical value for the Dickey-Fuller test is -2.89.

THE DICKEY FULLER TEST IS DESCRIBED ON PAGES 277-288 OF THE TEXTBOOK. OUTPUTS 1 AND 2 DO CONTAIN RELEVANT REGRESSIONS. SINCE THE T-STATS ARE SMALL (SMALLER THAN THE DICKEY FULLER CRITICAL VALUE MENTIONED ON PAGE 280) IN BOTH CASES WE CAN CONCLUDE THAT UNIT ROOTS ARE PRESENT IN BOTH X AND Y.

ii) Define and describe the Engle-Granger test for cointegration. Does cointegration seem to be present in this data set? You may assume that the 5% critical value for the Engle-Granger test is -3.39.

COINTEGRATION TESTING IS DICUSSED BEGINNING ON PAGE 312 OF THE TEXTBOOK. OUTPUT 4 CAN BE USED TO DO THE ENGLE GRANGER TEST IS. COMPARING -11.7749 TO THE ENGLE-GRANGER CRITICAL VALUE OF -3.33 WE CAN REJECT THE HYPOTHESIS THAT THE ERRORS HAVE A UNIT ROOT. THUS COINTEGRATION IS PRESENT

iii) Can you obtain an estimate of the long run multiplier from any of these OUTPUTS? If yes, what is the estimate of the long multiplier?

SINCE X AND Y ARE COINTEGRATED, OUTPUT 3 CAN BE USED TO GIVE US A MULTIPLIER OF 1.93891. NOTE, HOWEVER, THAT IF X AND Y WERE NOT COINTEGRATED, THEN OUTPUT 3 WOULD HAVE BEEN A SPURIOUS REGRESSION AND WE WOULD NOT HAVE BEEN ABLE TO USE IT TO CALCULATE THE MULTIPLIER.

OUTPUT 1

Regression Statistics	
Multiple R	0.100336
R Square	0.010067
Adjusted R Square	0.003957
Standard Error	0.149963
Observations	164

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.03705	0.03705	1.647497	0.201133
Residual	162	3.643202	0.022489		
Total	163	3.680253			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	0.209069	0.148738	1.405625	0.16175	-0.08465	0.502784	-0.08465	0.502784
Y-lagged	-0.01519	0.011833	-1.28355	0.201133	-0.03856	0.008179	-0.03856	0.008179

OUTPUT 2

Regression Statistics	
Multiple R	0.071587
R Square	0.005125
Adjusted R Square	-0.00102
Standard Error	0.010183
Observations	164

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	8.65E-05	8.65E-05	0.834485	0.362336
Residual	162	0.016798	0.000104		
Total	163	0.016885			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	0.011895	0.002202	5.400638	2.33E-07	0.007545	0.016244	0.007545	0.016244
X-lagged	-0.00148	0.00162	-0.9135	0.362336	-0.00468	0.001719	-0.00468	0.001719

OUTPUT 3

Regression Statistics	
Multiple R	0.993897
R Square	0.987831
Adjusted R Square	0.987755
Standard Error	0.109426
Observations	164

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	157.4576	157.4576	13149.98	5.1E-157
Residual	162	1.939785	0.011974		
Total	163	159.3974			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	9.99487	0.023857	418.9481	8.8E-248	9.947759	10.04198	9.947759	10.04198
X	1.93891	0.017434	114.6734	5.1E-157	1.964787	2.033641	1.964787	2.033641

OUTPUT 4

Regression Statistics	
Multiple R	0.680224
R Square	0.462705
Adjusted R Square	0.459368
Standard Error	0.10957
Observations	163

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	1.664557	1.664557	138.6493	1.75E-23
Residual	161	1.932888	0.012006		
Total	162	3.597445			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-0.00013	0.008583	-0.01468	0.988308	-0.01708	0.016824	-0.01708	0.016824
Resid(-1)	-0.9397	0.079805	-11.7749	1.75E-23	-1.0973	-0.7821	-1.0973	-0.7821