
Concordia University
Department of Economics

Econ 324
Economic Data Analysis

Ivan Tchinkov
Fall 2014

Midterm Examination

FIRST NAME: _____ LAST NAME: _____

STUDENT NUMBER: _____

I. True/False/Uncertain - Briefly explain. No credit without an explanation (2 marks each).

1. The standard deviation, $\frac{\sigma}{\sqrt{n}}$, is an estimate of the square root of the variance of the sampling distribution of the sample mean.

False. The standard error, not the standard deviation, is an estimate of the square root of the variance of the sampling distribution of the sample mean. Also, the expression for the standard error is $\frac{s}{\sqrt{n}}$.

2. The EViews code for running an OLS regression of *salary* on *years_exp*, *points* and a left-handed dummy, *left_dummy* with our hockey data is: *ls salary c years_exp points @expand(handed)*.

False. With the *@expand* command one has to remove the constant, *c*, from the code, otherwise there is perfect multicollinearity and the regression cannot be run.

3. Type II error will never be committed if one always rejects the null hypothesis.

True. If one always rejects the null, then one will never fail to reject it, including never failing to reject a false null and hence will never commit a Type II error.

4. If a big percentage of a doctor's patients die, then she is a bad doctor.

False/Uncertain. Correlation does not necessarily imply causation. It might be that the sickest patients go to the best doctor and most of them end up dying not because of the doctor, but because of their condition.

5. The higher the correlation between 2 explanatory variables, the higher the standard errors of their estimated coefficients.

True. From the formula for standard errors with 2 regressors $s.e.(\hat{\beta}_1) = \sqrt{\frac{\sum e_i^2 / (n-3)}{\sum (X_i - \bar{X})^2 (1-r_{12}^2)}}$, as r_{12} gets bigger, so does the standard error.

6. The double-log functional form is used when one expects a non-constant elasticity of Y with respect to X .

False. The slope coefficient from a double-log functional form is β_i , i.e. it is a constant and it represents the elasticity of Y with respect to X , i.e. $\beta_i = \frac{\partial \ln Y_i}{\partial \ln X_i} = \frac{\partial Y_i}{\partial X} \frac{X_i}{Y_i} = \eta_{YX}$.

7. Excluding an irrelevant variable from a regression will cause omitted variable bias if it is correlated with the included ones.

False. Regardless of whether it is correlated or not, there is no omitted variable bias, since the variable is irrelevant, i.e. its slope coefficient is zero.

8. Generalized Least Squares (GLS) is a way to correct for heteroskedasticity.
False. It corrects serial correlation, not heteroskedasticity.
9. The linear probability model avoid the unboundedness problem of the slope coefficient not being between 0 and 1.
False. The unboundedness problem is that the predicted value of the dummy dependent variable cannot be assured to be between 0 and 1 in the linear probability model. To avoid this problem, we use Logit.
10. The White test for heteroskedasticity can be used when one knows the variable Z_i that is causing that problem.
False. The White test is a general test for heteroskedasticity, it can be used without knowing what is causing the problem.

II. Problems - You have to show your work. No credit without an explanation (15 marks each).

1. There is a growing interest among political economists in the factors, which influence civil strife (e.g. civil wars, guerrilla wars, terrorism, etc.). Some researchers feel that civil strife may be related to environmental factors (e.g. shortages of clean water and good farmland, deforestation, etc.) as well as political and economic ones. In order to investigate this hypothesis they have collected data from $N = 91$ countries on the following variables:

Y = number of lives lost due to civil strife in the last 10 years.

X_1 = average annual GDP growth over the last 10 years (measured as a percentage so that a value of, say, 2 means average annual GDP growth of 2%).

X_2 = an index of environmental scarcity. (It is a number between 0 and 100 and higher values of this variable indicate more environmental scarcity problems.)

D = a dummy variable which equals 1 if the country is a democracy, and equals 0 otherwise.

Results from a regression of Y on X_1 , X_2 and D are given below in OUTPUT 1.

They then constructed another variable, $Z = X_2 * D$, and ran a regression of Y on X_1 , X_2 , D and Z . Results from this regression are given below in OUTPUT 2.

Use these regression outputs to write a brief report on the effect of environmental, political and economic factors on civil strife. Your report should include:

- (a) A discussion of whether you would use OUTPUT 1 or OUTPUT 2 (or both) as the focus of a report. (2 marks)

Both regressions explain a high degree of variation in the dependent variable, and have an R^2 significantly different than zero.

However, output 2 seems better, because its R^2 adjusted is higher and it has all of the coefficient estimates significant at 5%. Output 1 has an insignificant coefficient on economic growth, which is contrary to economic theory and is a possible sign of the omitted variable bias.

Also, the coefficient on D is very high, again a possible sign for omitted variable bias.

Furthermore, economic intuition indicates that the effect of environmental factors on loss of life might differ for democracies and non-democracies.

All of this indicates a preference for output 2.

(b) A discussion of how coefficient estimates can be interpreted. (2 marks)

The coefficient estimates have the usual interpretation:

The intercept estimate suggests that there is higher loss of life due to civil strife in non-democracies, as compared to democracies by about 150 people for the past 10 years, for countries with comparable economic growth and environmental factors.

We also see that a one percent increase in the average economic growth has reduced the loss of life by about 108 people for the last 10 years.

Worse environmental factors for non-democracies increase loss of life, in particular a one-point increase in the index, increases loss of life by about 200 people for the past 10 years.

Finally, for democracies, there is virtually no effect of the environmental factors on loss of life due to civil strife.

(c) A discussion of the statistical significance of your results. (2 marks)

All coefficients from output 2 have the expected signs, and are statistically significant at 5 % and also at 1%, with the exception of the intercept dummy.

(d) A discussion of potential problems that the regression might face and how they can be fixed. (4 marks)

Since we have cross-sectional data, we would not expect to have any serial correlation, but to test for it we would have to compute the Durbin-Watson d-statistic, which is not provided and if a problem is found ($d \approx 0$), then we can use either Generalized Least Squares (GLS) with the Cochrane-Orcutt method or Newey-West standard errors.

To test for heteroskedasticity we would use the general White test and if a problem is found we would use either White heteroskedasticity-consistent standard errors, Weighted Least Squares (WLS) or a suitable redefinition of the variables.

To test for multicollinearity we would compute correlation matrices and examine whether any of them is bigger than .8 or Variance Inflation Factors (VIF) and see if any is bigger than 5. If so, we have a multicollinearity problem and we would need to get more data or drop of the correlated variables with a risk of omitted variable bias.

In addition, the equation might suffer from misspecification, to check we can use the RESET test and/or from omitted variable bias, in which case we have to think what other variables should theoretically be in the model. Also, we might have the incorrect functional form, so we need to think whether an alternative form makes more sense.

(e) A discussion of what light the results shed on the question of interest (i.e. the effect of environmental, political and economic factors on civil strife). (3 marks)

This suggests that increasing economic growth reduces the loss of life due to civil strife, thus economic factors are important, not only statistically, but also economically.

Environmental factors are also important in influencing the loss of life due to civil strife, but more so in non-democracies than in democracies, where the economic effect is virtually zero.

(f) Anything else that you would like to discuss, related to the problem.(2 marks)

This is an example only. Give points if people said something that makes sense here.

It might have been interesting to have an interaction term between economic growth and the dummy for democracy, to see if growth has a differential effect on democracies versus

non-democracies. Economic theory suggests that this should be the case, although maybe less so as compared to the interaction environmental factors term.

This study resource was
shared via CourseHero.com

OUTPUT 1

<i>Regression Statistics</i>						
Multiple R	0.812332					
R Square	0.73235					
Adjusted R Square	0.726569					
Standard Error	2721.902					
Observations	91					
ANOVA						
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>	
Regression	3	3.2E+09	1.07E+09	143.9793	1.25E-33	
Residual	87	6.45E+08	7408751			
Total	90	3.84E+09				
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
Intercept	5168.421	876.7405	5.895041	6.96E-08	3425.803	6911.039
X1	-161.399	287.6673	-0.56106	0.576198	-733.169	410.3714
X2	120.1529	10.64446	11.28783	1E-18	98.99584	141.3099
D	-10449.5	593.5834	-17.6042	2.04E-30	-11629.4	-9269.74

OUTPUT 2

<i>Regression Statistics</i>						
Multiple R	0.929763					
R Square	0.889527					
Adjusted R Square	0.879505					
Standard Error	145.4312					
Observations	91					
<i>ANOVA</i>						
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>	
Regression	4	3.84E+09	9.61E+08	45423.43	4.6E-142	
Residual	86	1818919	21150.22			
Total	90	3.84E+09				
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
Intercept	960.0885	52.69869	18.21845	2.88E-31	855.3271	1064.85
X1	-108.82	15.37302	-7.07864	3.71E-10	-139.381	-78.2596
X2	201.5658	0.735909	273.9004	2.9E-128	200.1029	203.0288
D	-147.564	67.06875	-2.2002	0.030474	-280.893	-14.2363
Z	-201.327	1.154891	-174.326	2E-111	-203.623	-199.031

2. Suppose that a fad for oats (resulting from the announcement of the health benefits of oat bran) has made you toying with the idea of becoming a broker in the oat market. Before spending your money, you decide to build a simple model of supply and demand for oats:

$$Q_{Dt} = \beta_0 + \beta_1 P_t + \beta_2 YD_t + \epsilon_{Dt}$$

$$Q_{St} = \alpha_0 + \alpha_1 P_t + \alpha_2 W_t + \epsilon_{St}$$

$$Q_{Dt} = Q_{St}$$

where

Q_{Dt} = the quantity of oats demanded in time period t.

Q_{St} = the quantity of oats supplied in time period t.

P_t = the price of oats in time period t.

W_t = the average oat farmer wages in time period t.

YD_t = disposable income in time period t.

- (a) There is no left-hand-side variable that appears on the right-hand-side of either equation. Does this mean that OLS will have no simultaneity bias? Why or why not? (3 marks)

OLS estimation will still encounter simultaneity bias because price and quantity are simultaneously determined. Not all endogenous variables will appear on the left-hand side of a structural equation.

- (b) Are your equations structural or reduced-form? (2 marks)

Those are structural equations, since they display the underlying theory of supply and demand, i.e. demand is affected by disposable income and supply by the farmers' wages as a cost of production. Both depend on price as well.

- (c) How many and what are the endogenous and exogenous/predetermined variables? (2 marks)

Endogenous: Q_{Dt} , Q_{St} and P_t - 3 endogenous variables.

Exogenous/predetermined: YD_t and W_t - 2 exogenous/predetermined variables.

- (d) Are the demand and supply equations over-, under- or just identified? (2 marks)

Applying the order condition for identification we get:

Demand equation: exactly identified - 2 predetermined variables (YD_t and W_t) in the system and 2 slope coefficients to estimate (β_1 and β_2).

Supply equation: exactly identified - 2 predetermined variables (YD_t and W_t) in the system and 2 slope coefficients to estimate (α_1 and α_2).

- (e) You expect that when P_t goes up, Q_{Dt} will fall. Does this mean that any simultaneity bias in the demand equation will be negative, instead of the positive bias we typically have with OLS estimation of simultaneous systems? Explain your answer. (3 marks)

The direction of the bias depends on the correlation between the error term and the right-hand-side endogenous variable. If the correlation between the error term and price is positive, as it most likely is, then the simultaneity bias will also be positive.

- (f) Carefully outline how you will apply 2SLS to this system. How many equations (including reduced-form) would you have to estimate and which variables will be in which equation. (3 marks)

Three equations will be run.

Stage one: $P = f(YD, W)$ (reduced-form).

Stage two: $Q_D = f(\hat{P}, YD)$ and $Q_S = f(\hat{P}, W)$.