Concordia University
Department of Economics

Econ 324 **Ivan Tchinkov**
Economic Data Analysis **Fall 2016**

Midterm Examination

FIRST NAME: __________________________ **LAST NAME:**__________________________

STUDENT NUMBER: __________________________

I. True/False/Uncertain - Briefly explain. No credit without an explanation (8 marks each).

1. One is more likely to reject the null in the presence of heteroskedasticity.

   True. Since heteroskedasticity would inflate the standard errors, but running OLS without correcting for it would result in standard errors that do not take this heteroskedasticity into consideration, i.e. are too small. Then the t-stats are too high and thus more likely to reject a null hypothesis.

2. The Stata code for implementing an ECM (3,2) model is: *reg l(0/2)d.y l(0/1)d.x time*

   False, it is missing the error correction term *l.resid.*

3. A random walk with drift has both a deterministic and a stochastic trend.

   True, see Lecture 6, 4.3.

4. With two I(1) variables that are not cointegrated, there is a spurious regression problem.

   True. Two I(1) variables are unit root, non-stationary ones. If no cointegration exists, then there is a spurious regression problem and one needs to work with their differences which would be stationary.

5. To find the long run multiplier with unit root variables, one needs to run an ADL(p,q) model in its rewritten version.

   False. This is how one finds the multiplier with stationary variables.

II. Problems - You have to show your work. No credit without an explanation (30 marks each).

1. The following investment demand equation for the US was estimated using annual data for 1964-1994 period:

$$\hat{I}_t = 32.952 + 0.164GDP_t - 5.638R_{t-1}$$

- I_t - total gross private domestic investment in year t.
- GDP_t - gross domestic product in year t.
- R_t - interest rate in year t.
- $N = 31$

- $\overline{R}^2 = .9$
- DW=0.976
- $d_L = 1.30$, $d_U = 1.57$ a 5% significance level.

(a) Test for first order serial correlation in the equation error terms at the 5% level. Why is this problem likely to occur in a model of this sort? (7 marks)

The model has an estimated Durbin-Watson d-statistic of 0.976. The upper and lower limits of the DW d-statistic for 31 observations and two independent variables at a 5% level of significance are 1.30 and 1.57 respectively. Therefore, we are faced with positive serial correlation. It is likely to occur since this is a time series regression.

(b) Depending on your answer in (a), specify if you think that there is an econometric problem present here and its consequences. (7 marks)

Serial correlation that is not the result of a missing important variable ("pure" serial correlation) leaves the coefficients unbiased (since there is no lagged dependent variable on the RHS of the equation), while it augments the variances. Therefore, the OLS estimates are not of a minimum-variance (inefficient). Furthermore, OLS estimators are not the most appropriate to use as they underestimate the variances, which results in obtaining unusually high t-ratios, F-ratio, and R squared.

(c) It is suggested that you re-estimate the equation as a first-difference model, which you do and get the following results:

$$\hat{I}_t^* = 35.872 + 0.168GDP_t^* - 10.897R_{t-1}^*$$

- $I_t^* = (I_t - \rho I_{t-1})$.
- $GDP_t^* = (GDP_t - \rho GDP_{t-1})$.
- $R_t^* = (R_t - \rho R_{t-1})$.
- $N = 30$
- $\overline{R}^2 = .833$
- $\hat{\rho} = 0.4073356$

What is this remedy called? Explain the steps involved to get the results presented here. (8 marks)

The remedy used here is called Generalized Least Squares. The serial correlation coefficient $\hat{\rho} = 0.4073356$ has been estimated using the Cochrane-Orcutt method using the residuals of the original equation shown at the beginning and running a regression of them as follows:

$$e_t = \rho e_{t-1} + v_t$$

(d) Comment on the results of (c) in comparison to the original results. Which equation do you prefer? (8 marks)

Comparing the original and the final equations, we observe the following:.

- **The coefficients have the proper signs in both equations, however, they seem more realistic in the final equation.**
- **The coefficient of GDP has hardly changed, while the coefficient of Interest Rate did change significantly.**
- **The standard errors in the final equation should be larger than in the original equation, given that OLS tends to underestimate the variances.**
- **In the final equation we are losing one degree of freedom due to the model of first differences used here.**

- **The final equation should be preferred as it has serial correlation removed.**

2. Economists are interested in the relationship between the prices of organic (Y_t) and non-organic (X_t) oranges. They have collected data on X_t and Y_t and run various regressions using this data. Outputs for these regressions are below and labeled as OUTPUT 1, OUTPUT 2, OUTPUT 3 and OUTPUT 4. To be specific:

   - OUTPUT 1 contains results from a regression ofΔY on one lag of Y. That is, $\Delta Y_t = \alpha + \beta Y_{t-1} + \epsilon_t$.

   - 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.

   (a) Define and describe the Dickey-Fuller test. Can this test be done using any of the OUTPUTS below? If yes, what does the Dickey-Fuller test tell you about the properties of Y and X? (6 marks)

   The Dickey-Fuller test is used to test for unit root. It is based on testing the AR(p) model's slope coefficients. Using this test we can test for unit roots both the X and Y series, assuming that it is correctly specified in terms of lag length and time trend. Using OUTPUT 1, we test for beta equals zero. The t-statistic on the lagged output coefficient estimates is –1.28. This number is not more negative than the threshold value used in the rule of thumb for the Dickey-Fuller test – 2.89. Thus, we fail to reject the null of unit root, thus we have unit root in the Y variable.

   Testing for unit root the X variable involves looking at OUTPUT 2. The t-statistics for the coefficient estimate on X lagged one period is -.91, which again is not more negative than -2.89, thus we fail to reject the unit root null hypothesis and conclude that both X and Y series have unit root (again assuming the correct specification in terms of lag length and no time trend).

   (b) Would you expect the prices of organic and non-organic oranges to be cointegrated? Why or why not? (6 marks)

   Yes, since those are substitutes and the forces of supply and demand would hold them together, i.e. if organic oranges price gets too high, consumers switch to regular oranges thus bidding up their price.

   (c) OUTPUT 3 contains results from the regression of Y on X. Are these results reliable or spurious? (6 marks)

   We have already seen that both X and Y have unit roots. So, unless they are cointegrated, then we will run in the spurious regression problem, i.e. we will have a statistical significant coefficient estimate for X, and also significant and high R^2, when in fact the true coefficient on beta could be zero and zero R squared. Thus, we have the spurious regression problem UNLESS X and Y are cointegrated.

   We can use OUTPUT 4 to test for cointegration and we conclude that X and Y are cointegrated, since the t-statistic is –11.78, and is more negative than –2.89, which means that we can reject the unit root null and conclude that there is no unit root in the residuals, so that X and Y are cointegrated (assuming the correct specification in terms of lag length and no time trend). Thus OUTPUT 3 does not give spurious regression results.

   (d) Can you obtain an estimate of the long run multiplier from any of these OUTPUTS? If yes, what is the estimate of the long run multiplier? (6 marks)

   Normally, we cannot estimate the long run multiplier from the given information. But since X and Y are cointegrated, from OUTPUT 3 the estimate of the long run

multiplier is just the coefficient estimate beta, thus the estimate of the long run multiplier is 1.93891.

(e) Write down an error correction model and discuss the interpretation of the coefficients. Discuss whether the data described or calculated above (Y,X,ΔY,ΔX and e) could be used to estimate an error correction model. (6 marks)

The error correction model (ECM) is written as usual (with only one lag for each X and Y)

$$\Delta Y_t = \phi + \lambda e_{t-1} + \omega_0 \Delta X_t + \epsilon_t$$

where $e_{t-1} = Y_{t-1} - a - bX_{t-1}$.

The coefficient interpretations are exactly as in the text, i.e. the coefficient estimate of delta X gives the immediate change in Y for a given change in X.

The more interesting interpretation is on the coefficient estimate in front of lagged errors. It says that Y depends on the equilibrium error last period. Thus, suppose that delta X is zero, and that Y was above equilibrium last period, i.e. lagged error was positive. The lambda coefficient is negative (for stability reasons), thus if Y is above equilibrium last period, then equilibrium error last period is positive and lambda is negative, which means that Y will fall this period towards the equilibrium value. The lambda estimate is thus giving us the short run adjustment to equilibrium.

You can use the data discussed and calculated above to estimate the ECM.

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