JANUARY/FEBRUARY 2016 ECONOMETRICS ECS4863

# QUESTIONS 1 & 2 DETAILED ANSWERS

**DEFINING KEY CONCEPTS IN THE QUESTION PAPER**

***MODELLING LONG RUN RELATIONSHIPS***

*Stationarity and unit root testing*

- A stationary series can be defined as one with a constant mean, constant variance and constant autocovarience. The use of non-stationary data can lead to spurious regression – normally very high R- squared greater the Durbin Watson test.

*Testing for unit root*

- ADF
- Philips Perron

*Cointegration*

- Cointegration is an important tool for modeling the long-run relationships in time series data. Economic theory suggests that many times series variables move together in the long run or fluctuating around a long run equilibrium and any divergence between variables is a short run phenomenon. Cointegration occurs when two or more non- stationary time series:
    - Have a long run equilibrium
    - Move together such that their linear combination results in a stationary series
    - Share underlying stochastic trend

*Test for cointegration*

- Residual based test
- Johansen technique

*Error Correction Model*

- Cointegration implies that the time series will be connecting through an error correction model. The error correction model:
    - Reflects long run relationships of variables
    - Includes short run dynamic adjustment mechanism that describes how variables adjust when they are out of equilibrium

***LIMITED DEPENDENT VARIABLE MODELS***

*Linear probability models*

There are many situations in research where the dependent variable is qualitative. The qualitative information will coded as a dummy variable and the situation would be referred a limited dependent variable. In our case in question 2 the dependent variable is binary where 1 is accepted into an honours module and 0 is not accepted into an honours module. The use of OLS is not perfect when estimation models with a limited dependent variable.

*Logit and profit approaches*

These are approaches used to overcome the limitation of the LPM that it can produce probabilities that are negative and greater than 1. They do this by using a function that transform the regression model so that fitted values are bounded between 0 and 1 interval.

**1a) Use the ADF test to test all four variables for unit roots. Provide your answers in the table below (Hint: please remember to log variables before performing the tests):**

| Variable | Model | Lags | ADF test statistic | Prob | Interpretation |
|---|---|---|---|---|---|
| LNS | Trend and Intercept | AIC 1 | -2.080231 | 0.5418 | Non Stationary |
| | Intercept | AIC 1 | -2.080231 | 0.7841 | Non Stationary |
| | None | AIC 1 | -2.080231 | 0.9526 | Non Stationary |
| DLNS | Trend and Intercept | AIC 1 | -3.457736 | 0.0571 | Non Stationary |
| | Intercept | AIC 1 | -3.487169 | 0.0131 | Stationary |
| | None | AIC 1 | -3.487169 | 0.0023 | Stationary |
| LGDP | Trend and Intercept | AIC 1 | -0.025106 | 0.6983 | Non stationary |
| | Intercept | AIC 1 | -1.777274 | 0.9508 | Non stationary |
| | None | AIC 1 | 3.384360 | 0.9997 | Non stationary |
| DLGDP | Trend and Intercept | AIC 1 | -4.627472 | 0.00031 | Stationary |
| | Intercept | AIC 1 | -4.671391 | 0.0005 | Stationary |
| | None | AIC 1 | -4.671391 | 0.0049 | Stationary |
| LLC | Trend and Intercept | AIC 1 | -1.217220 | 0.8942 | Non Stationary |
| | Intercept | AIC 1 | -1.876704 | 0.3398 | Non Stationary |
| | None | AIC 1 | -1.876704 | 0.7859 | Non Stationary |
| DLLC | Trend and Intercept | AIC 1 | -4.817874 | 0.0018 | Stationary |
| | Intercept | AIC 1 | -4.396242 | 0.0011 | Stationary |
| | None | AIC 1 | -0.778159 | 0.3728 | Non Stationary |
| LLP | Trend and Intercept | AIC 1 | -2.665692 | 0.2554 | Non Stationary |
| | Intercept | AIC 1 | -1.391756 | 05767 | Non Stationary |
| | None | AIC 1 | -1.391756 | 0.8465 | Non Stationary |
| DLLP | Trend and Intercept | AIC 1 | -2.887812 | 0.1770 | Non Stationary |
| | Intercept | AIC 1 | -2.922536 | 0.0516 | Non Stationary |
| | None | AIC 1 | -2.869805 | 0.0052 | Stationary |

## 1 b) Test for cointegration between variables:

Date: 01/11/21 Time: 07:58
Sample (adjusted): 1973 2014
Included observations: 42 after adjustments
Trend assumption: Linear deterministic trend
Series: LNS LGDP LLC
Lags interval (in first differences): 1 to 2

Unrestricted Cointegration Rank Test (Trace)

| Hypothesized No. of CE(s) | Eigenvalue | Trace Statistic | 0.05 Critical Value | Prob.** |
|---|---|---|---|---|
| None | 0.302209 | 24.18886 | 29.79707 | 0.1926 |
| At most 1 | 0.147577 | 9.075745 | 15.49471 | 0.3584 |
| At most 2 | 0.054854 | 2.369482 | 3.841466 | 0.1237 |

Trace test indicates no cointegration at the 0.05 level
* denotes rejection of the hypothesis at the 0.05 level
**MacKinnon-Haug-Michelis (1999) p-values

Unrestricted Cointegration Rank Test (Maximum Eigenvalue)

| Hypothesized No. of CE(s) | Eigenvalue | Max-Eigen Statistic | 0.05 Critical Value | Prob.** |
|---|---|---|---|---|
| None | 0.302209 | 15.11311 | 21.13162 | 0.2811 |
| At most 1 | 0.147577 | 6.706263 | 14.26460 | 0.5244 |
| At most 2 | 0.054854 | 2.369482 | 3.841466 | 0.1237 |

Max-eigenvalue test indicates no cointegration at the 0.05 level
* denotes rejection of the hypothesis at the 0.05 level
**MacKinnon-Haug-Michelis (1999) p-values

Unrestricted Cointegrating Coefficients (normalized by b'*S11*b=I):

**(i) Estimate the following long-run cointegration equation and use your results to complete the table. (Remember to include an intercept term.**

LNS = f (LGDP, LLC)

Dependent Variable: LNS
Method: Least Squares
Date: 01/10/21 Time: 23:11
Sample: 1970 2014
Included observations: 45

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| LGDP | 1.298990 | 0.163023 | 7.968145 | 0.0000 |
| LLC | -0.129727 | 0.034002 | -3.815242 | 0.0004 |
| C | -14.06660 | 2.252832 | -6.243961 | 0.0000 |

| | | | |
|---|---|---|---|
| R-squared | 0.881356 | Mean dependent var | 4.241641 |
| Adjusted R-squared | 0.875706 | S.D. dependent var | 0.233948 |
| S.E. of regression | 0.082479 | Akaike info criterion | -2.088202 |
| Sum squared resid | 0.285718 | Schwarz criterion | -1.967757 |
| Log likelihood | 49.98453 | Hannan-Quinn criter. | -2.043301 |
| F-statistic | 156.0003 | Durbin-Watson stat | 0.284335 |
| Prob(F-statistic) | 0.000000 | | |

**ii) Interpretation of coefficients**

There is a positive relationship between GDP and demand for skilled labour, meaning that as the country expand economically the demand for skilled labor increases. A percentage change in GDP will result into a 1.298% increase in demand for skilled labor all things being equal.

There is a negative relationship between labor costs and demand for skilled labour, meaning that as labor costs increases demand for skilled labor decreases. A percentage change in labor costs will result in 0.1297% decrease in demand for skilled labor.

iii) Yes the coefficients correspond to priori expectations because theoretically GDP is positively related to demand for skilled labor. Also costs are negatively related to demand for skilled labor both in theory and practice

### iv) Generate residual series

Null Hypothesis: RESID02 has a unit root
Exogenous: Constant
Lag Length: 1 (Automatic - based on SIC, maxlag=9)

| | | t-Statistic | Prob.* |
|---|---|---|---|
| Augmented Dickey-Fuller test statistic | | -2.776454 | 0.0701 |
| Test critical values: | 1% level | -3.592462 | |
| | 5% level | -2.931404 | |
| | 10% level | -2.603944 | |

*MacKinnon (1996) one-sided p-values.

At 1% and 5% levels of significant the residual series has a unit root meaning non stationary, however stationary at 10% level of significant the series is stationary.

v) Since residual series are not stationary at 5% level of significant we can conclude that there is no cointegration between variables. The results are in line with the cointegration test above.

### C) Build an Error Correction Model (ECM) for the demand for skilled labor

Dependent Variable: D(LNS)
Method: Least Squares
Date: 01/10/21 Time: 23:17
Sample (adjusted): 1971 2014
Included observations: 44 after adjustments

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| D(LGDP) | 0.892320 | 0.053556 | 16.66151 | 0.0000 |
| D(LLP) | -0.970114 | 0.032150 | -30.17447 | 0.0000 |
| RESID02(-1) | -0.036667 | 0.015163 | -2.418204 | 0.0202 |
| C | 0.002268 | 0.001746 | 1.298846 | 0.2014 |

| | | | |
|---|---|---|---|
| R-squared | 0.971835 | Mean dependent var | 0.019602 |
| Adjusted R-squared | 0.969722 | S.D. dependent var | 0.043222 |
| S.E. of regression | 0.007521 | Akaike info criterion | -6.855758 |
| Sum squared resid | 0.002263 | Schwarz criterion | -6.693559 |
| Log likelihood | 154.8267 | Hannan-Quinn criter. | -6.795607 |
| F-statistic | 460.0599 | Durbin-Watson stat | 2.891087 |
| Prob(F-statistic) | 0.000000 | | |

ii) The error correction term should be negative, significant and less than unit. In our case the error correction term is negative, less than 1 and statistically significant and 5% level of significance.

### d) Perform diagnostic checks on the ECM

| Test | Null Hypothesis | Test statistic | P-value | Conclusion |
|---|---|---|---|---|
| Jarque-Bera | $H_o$: Normally distributed residuals | JB = 27.50960 | 0.0000001 | We reject Ho. Residuals are not normally distributed |
| Ljung –Box Q | $H_o$: No serial correlation | LBQ(6) = 11.954 | 0.063 | We failed to reject Ho at 5% level of significant. Residuals are not serially correlated |
| Breusch- Godfrey LM TEST | $H_o$: No serial correlation | $nR^2(2)$ = 15.44363 | 0.0004 | We reject Ho. Residuals are serially correlated |
| ARCH-LM | $H_o$: No hetroscedasticity | $nR^2(2)$ =9.327868 | 0.0094 | We reject Ho. There is presence of Hetroscedasticity |
| White | $H_o$: No hetroscedasticity | $nR^2(no\ CT)$ = 9.327868 | 0.0000 | We reject Ho. There is presence of Hetroscedasticity |
| Ramsey RESET | $H_o$: No misspecification | LR (2) = 2.861188 | 0.2392 | We failed to reject Ho. No misspecification |



### ii) Given your conclusions on the diagnostic check of the ECM, do you think that this is an acceptable model

The model is not acceptable because it has violated some of the OLS assumptions. The model suffers from the problem of serial autocorrelation therefore estimates won't be BLUE (Best Linear Unbiased Estimators), and they won't be reliable enough. The model suffers from the problem of Hetroscedasticity. If errors are heteroscedastic it will be difficult to trust the standard errors of the OLS estimates. Hence, the confidence intervals will be either too narrow or too wide. This impact will forecasting and variance decomposition.

**e) Regardless of the results you obtained in question 1(d), suppose you still decide to create a model statement in EViews to combine your long run and ECM.**

The purpose of this step is to re-write the equation back to its levels and simulate it dynamically. The outcome is to create a new modelled variable of the dependent variable.

*Step 1*

**Long run**

$y_t = \quad y^\wedge{}_t \; + u^\wedge{}_t$ – The estimated cointegrating equation

We rewrite the equation so the residuals is on the left

$u_t = \quad y_t \; + y^\wedge{}_t$

*Step 2*

**ECM**

We specify the ECM model so that the differenced dependent variable $DLNS_t$ is the dependent variable in the equation. The purpose is to rewrite the equation back to levels.

By differencing we mean

$d(y_{t)} = \quad y_t - y_{t-1}$

$d(LNS_{t)} = \quad LNS_t - LNS_{t-1}$

Re- write the ECM so the LNS becomes a new dependent variable and the rest of the equation stays the same. We simply add $LNS_{t-1}$ at the end

*Step 3*

Get rid of the logs

NS = exp (logNS)

**i) Provide the missing values/variables in the model statement (please write your answer next to the correct option in the space provided below the statement):**

*LNS = 1.2989899682*LGDP - 0.129726760054*LLC - 14.0665958153 – Long run model*

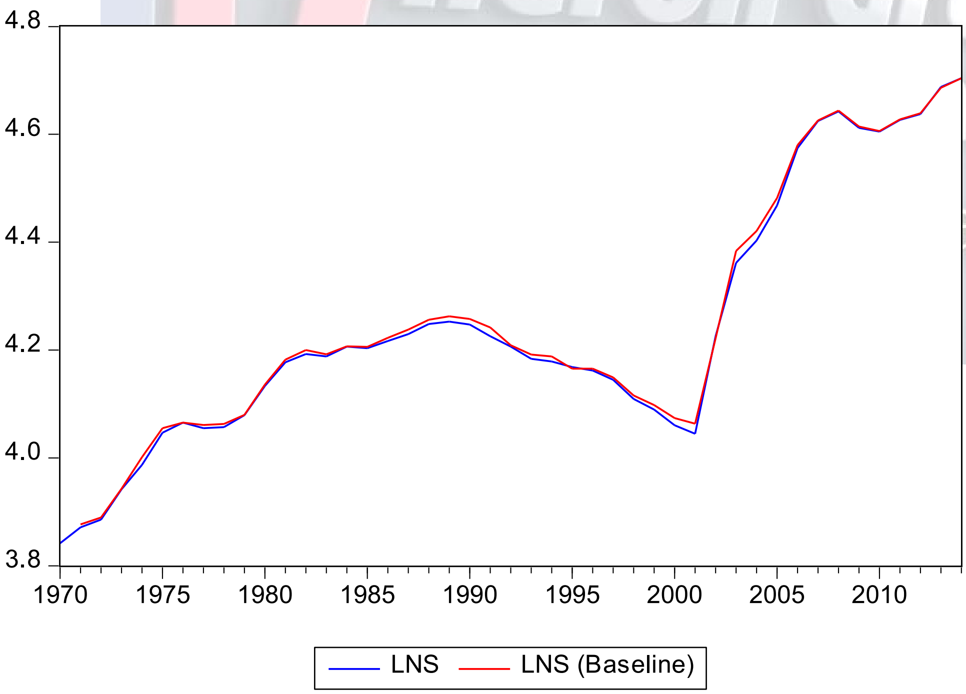
*D (LNS) = 0.892320312733*D (LGDP) - 0.970114223165*D (LLP) - 0.0366672785059*RESID02 (-1) + 0.00226826179642 – ECM Model*

*RELNS = LNS - 1.2989899682*LGDP - 0.129726760054*LLC - 14.0665958153*

*D (LNS) = 0.892320312733*D (LGDP) - 0.970114223165*D (LLP) - 0.0366672785059*RESID02 (-1) + 0.00226826179642 + LNS (-)*

NS = EXP(LNS)

**i) Graph the actual and estimated values for the demand for labour. Comment on the fit you observe. (Hint: copy/paste your graph of LNS and LNS (Baseline).)**




The fitted model tracks movements in the actual data. This is a very good fit of the observed variables. The model can be used for forecasting and scenario analysis.

**QUESTION 2**

**a) Consider the following statement "(the) use of coefficients of determination as a summary statistic should be avoided in models with qualitative dependent variables" (Aldrich and Nelson, Qualitative Response Model, *Journal of Economic Literature*, 1981, vol. 19:331-354).**

*i) List and briefly explain two possible problems when using OLS to estimate linear probability models (LPM's)*

The use of ordinary least squares (OLS) technique to estimate a model with a dummy dependent variable is known as creating a linear probability model (LPM), is not perfect because of the following reasons:

**Non normality of the error term**- error terms must be normally distributed for one to perform hypothesis tests after model estimations. The error term of an LPM follows a binomial distribution, it implies that the traditional t –tests and F-test overall significance are not valid.

**OLS model suffers the problem of heteroskedasticity** because the error term in an LPM is heteroskedastic and the variance of the error term is not constant as it depends on the value of the independent variables. *Note: no relationship between error term and explanatory variables*

*ii) What other measures of goodness of fit, apart from $R^2$, are available in binary regress and models? List any two and briefly explain how they work.*

The conventionally computed R2 is of limited value in the dichotomous response models

- McFadden R^2 – A rule of thumb – the McFadden's pseudo R2 ranging from 0.2 to 0.4 indicates very good model fit.
- Pearson's Chi-square test - Pearson's classical Chi-square test and Deviance test are well known, and they work very well when the covariates are categorical

*iii) Interpret your estimated coefficients. (Hint: remember to do the relevant adjustments to the coefficients.)*

Dependent Variable: HON
Method: ML - Binary Logit (Newton-Raphson / Marquardt steps)
Date: 01/11/21 Time: 11:51
Sample: 1 200
Included observations: 200
Convergence achieved after 5 iterations
Coefficient covariance computed using observed Hessian

| Variable | Coefficient | Std. Error | z-Statistic | Prob. |
|---|---|---|---|---|
| C | -10.83161 | 1.610748 | -6.724581 | 0.0000 |
| MATH | 0.118779 | 0.029822 | 3.982864 | 0.0001 |
| READ | 0.056716 | 0.025394 | 2.233432 | 0.0255 |

| | | | |
|---|---|---|---|
| McFadden R-squared | 0.273021 | Mean dependent var | 0.245000 |
| S.D. dependent var | 0.431166 | S.E. of regression | 0.364669 |
| Akaike info criterion | 0.839528 | Sum squared resid | 26.19779 |
| Schwarz criterion | 0.889003 | Log likelihood | -80.95280 |
| Hannan-Quinn criter. | 0.859550 | Deviance | 161.9056 |
| Restr. deviance | 222.7100 | Restr. log likelihood | -111.3550 |
| LR statistic | 60.80445 | Avg. log likelihood | -0.404764 |
| Prob(LR statistic) | 0.000000 | | |

| | | | |
|---|---|---|---|
| Obs with Dep=0 | 151 | Total obs | 200 |
| Obs with Dep=1 | 49 | | |

*To adjust the coefficients, we exponent the coefficient and subtract 1*

If the math mark obtained increase by 1% it will increase the probability of being accepted into an honors module by 13%

If the reading mark obtained increase by 1% it will in the probability of being accepted into an honors module by 5.84%

***iv) Which of the coefficients are statistically significant? Explain.***

Both coefficients are statistically significant because their p-value is less than 5%

***v) What is the effect of a percentage point increase in the mathematics mark obtained on the odds of being accepted into an honours class?***

An increase in 1 unit on Math score increases the odds of getting admission by 13%