BIRKBECK
(University of London)

MSc EXAMINATION FOR INTERNAL STUDENTS

Department of Economics, Mathematics and Statistics

ECONOMETRICS/ ECONOMETRICS for PG Certificate/ FINANCIAL ECONOMETRICS 1

EMEC026S7/BUEM007H7/BUEM045H7

Friday, 11 January 2019, 2.00 pm - 4.10 pm (includes 10 minutes reading time).

The paper is divided into two sections. There are three questions in section A and three questions in section B.

Answer **ONE** question from SECTION A, and **ONE** question from SECTION B, and **ONE** further question from either section. All questions carry the same weight; the relative weight of sub-questions is indicated in square brackets.

Non-programmable calculators are allowed.

In some questions, you may require 5% critical values for the Chi-squared distribution, to carry out tests. These critical values, CV, for various degrees of freedom, DoF, are:

DoF	1	2	3	4	5	6	7	8	9
CV	3.84	5.99	7.81	9.49	11.07	12.59	14.07	15.51	16.92

Section A

1. Consider the model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \qquad \mathbf{u} \sim N(0, \sigma^2\mathbf{I}),$$

where $\mathbf{y}$ is a $T \times 1$ vector of observations on a dependent variable, $\mathbf{X}$ is a $T \times k$ full-rank matrix of observations on exogenous variables, $\boldsymbol{\beta}$ is a $k \times 1$ vector of unknown parameters, and $\mathbf{u}$ is a $T \times 1$ vector of unobserved normally distributed disturbances, with $E(\mathbf{u}) = 0$,and with $E(\mathbf{u}\ \mathbf{u}') = \sigma^2\mathbf{I}$.

(a) [25%] Derive the maximum likelihood estimators of $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma^2)$.

(b) [25%] Derive the variance-covariance matrix for $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma^2)$.

(c) [25%] Show that the maximum likelihood estimator of $\boldsymbol{\beta}$, $\widehat{\boldsymbol{\beta}}$, is unbiased and the maximum likelihood estimator of σ^2, $\widehat{\sigma}^2$ is biased. What is the bias of $\widehat{\sigma}^2$?

(d) [25%] What is the maximised value of the log-likelihood as a function of the maximum likelihood estimator of σ^2? How would you use it to test hypotheses about β?

2. Consider the structural form of a simultaneous equations model for two variables

$$\mathbf{B}\mathbf{y}_t = \boldsymbol{\gamma} + \boldsymbol{\Gamma}\mathbf{y}_{t-1} + \mathbf{u}_t;\ E(\mathbf{u}_t\mathbf{u}_t') = \boldsymbol{\Sigma}.$$

$$\begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{bmatrix} \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$$

where $E(u_{1t}) = E(u_{2t}) = 0$; $E(u_{1t}^2) = \sigma_{11}$; $E(u_{2t}^2) = \sigma_{22}$; $E(u_{1t}u_{2t}) = \sigma_{12}$; $E(u_{it}u_{j,t-h})\ 0$; $i, j = 1, 2,\ h \neq 0$.

(a) [25%] Call the reduced form

$$\mathbf{y}_t = \mathbf{a} + \mathbf{A}_1\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t;\ E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t') = \Omega.$$

(i) Express $\mathbf{a}, \mathbf{A}_1, \boldsymbol{\varepsilon}_t, \Omega.$ in terms of $\mathbf{B}, \boldsymbol{\gamma}, \boldsymbol{\Gamma}, \mathbf{u}_t, \boldsymbol{\Sigma}.$ (ii) How would you estimate the two equations of the reduced form?

(b) [25%] Consider the restrictions $\beta_{11} = 1$; $\beta_{22} = 1$; $\gamma_{12} = 0$. (i) Show that, by the order condition for identification, the first equation is identified and the second is not. (ii) Explain how you would estimate the first equation, setting out all the steps involved.

(c) [25%] Again assuming $\beta_{11} = 1$; $\beta_{22} = 1$; $\gamma_{12} = 0$, under what conditions would the rank condition for the identification of the first equation fail.

(d) [25%] Consider the alternative restrictions $\beta_{11} = 1$; $\beta_{22} = 1$;; $\beta_{12} = 0$; $\sigma_{12} = 0$. Discuss the identification of the two equations. How would you estimate any identified equations?

3. In the equations below ε_t is a white noise process, with $E(\varepsilon_t) = 0$; $E(\varepsilon_t^2) = \sigma^2$; $E(\varepsilon_t \varepsilon_{t-h}) = 0, h \neq 0$. The variance of y_t, $Var(y_t) = E(y_t - E(y_t))^2$ and the autocovariances of y_t are $Cov(y_t, y_{t-i}) = E(y_t - E(y_t))(y_{t-i} - E(y_{t-i}))$.

   (a) [25%] For the AR1 process

   $$y_t = \alpha + \rho y_{t-1} + \varepsilon_t,$$

   what are $E(y_t)$, $Var(y_t)$, $Cov(y_t, y_{t-1})$, and $Cov(y_t, y_{t-2})$?

   (b) [25%] For the MA1 process

   $$y_t = \alpha + \varepsilon_t + \mu \varepsilon_{t-1},$$

   what are $E(y_t)$, $Var(y_t)$, $Cov(y_t, y_{t-1})$, and $Cov(y_t, y_{t-2})$?

   (c) [25%] For the random walk with drift

   $$y_t = \alpha + y_{t-1} + \varepsilon_t,$$

   show that y_t can be written as the sum of a deterministic trend, a stochastic trend and an initial condition, y_0.

   (d) [25%] For the process

   $$\Delta y_t = \alpha + \rho \Delta y_{t-1} + u_t$$

   where u_t may be serially correlated, how would you implement an augmented Dickey-Fuller test to determine whether y_t was $I(1)$ or $I(2)$.

Section B

1. A VAR(1) was estimated on US annual data, 1873-2016 for short interest rates, R_t, long interest rates, RL_t, and inflation, INF_t, in % at annual rates. Trace tests for r, the number of cointegrating vectors, and 5% critical values are:

H_0	H_1	$trace$	CV
$r=0$	$r \geq 0$	113.1	29.8
$r=1$	$r \geq 2$	34.0	14.3
$r=2$	$r=3$	5.1	3.8

Assuming $r=2$, the estimated cointegrating relations (standard errors) are

$$\begin{aligned} Z_{1t} &= R_t \underset{(0.017)}{-1.33} INF_t \\ Z_{2t} &= RL_t \underset{(0.015)}{-1.31} INF_t \end{aligned}.$$

The VECM, suppressing constants is

$$\begin{aligned} \Delta R_t &= \underset{(0.09)}{-0.27} Z_{1,t-1} \underset{(0.09)}{+0.27} Z_{2,t-1} \\ \Delta RL_t &= \underset{(0.04)}{-0.10} Z_{1,t-1} \underset{(0.04)}{+0.11} Z_{2,t-1} \\ \Delta INF_t &= \underset{(0.30)}{-0.78} Z_{1,t-1} \underset{(0.09)}{+1.31} Z_{2,t-1} \end{aligned}$$

(a) [20%] How many cointegrating vectors do the tests indicate? How do you interpret the result?

(b) [20%] Write the estimated VECM as $\Delta \mathbf{y}_t = \boldsymbol{\alpha}\boldsymbol{\beta}\mathbf{y}_{t-1} + \mathbf{u}_t$ where $\boldsymbol{\alpha}$ is a 3×2 matrix and $\boldsymbol{\beta}$ a 2×3 matrix

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_{11} & \beta_{12} & \beta_{13} \\ \beta_{21} & \beta_{22} & \beta_{23} \end{bmatrix}.$$

How many restrictions are required to identify β? What are the restrictions imposed on the estimated cointegrating vectors above?

(c) [20%] In the VECM $\alpha_{11} \approx -\alpha_{11}$ and $\alpha_{21} \approx -\alpha_{21}$ while $\beta_{13} \approx \beta_{23}$. How would you interpret this pattern?

(d) [20%] Test the individual hypotheses $\beta_{13} = -1; \beta_{23} = -1$. A likelihood ratio test of the joint hypothesis gave a test statistic of 0.64. What is the distribution of this statistic? Is the joint hypothesis rejected?

(e) [20%] ADF tests with an intercept and lag order selected by the AIC gave test statistics of -1.92 for R_t, -1.55 for RL_t and -3.60 for INF_t. The critical value is -2.88. What order of integration do the tests suggest for each variable? How might you change the identifying assumptions in the light of this result and your answer to part (c).

2. US data for the change in the logarithm of the S&P500 stock market index, DLSP=LSP-LSP(-1), 1872-2015 were used to estimate ARIMA(1,1,1) and ARIMA(0,1,0) for LSP. The results are shown below.

Dependent Variable: DLSP
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 11/07/18 Time: 15:43
Sample: 1873 2015
Included observations: 143
Convergence achieved after 13 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.063533	0.027546	2.306425	0.0226
DLSP(-1)	-0.497774	0.270885	-1.837582	0.0683
MA(1)	0.675063	0.235344	2.868415	0.0048
SIGMASQ	0.029778	0.003037	9.804551	0.0000

R-squared	0.039636	Mean dependent var	0.042195
Adjusted R-squared	0.018909	S.D. dependent var	0.176707
S.E. of regression	0.175029	Akaike info criterion	-0.615909
Sum squared resid	4.258262	Schwarz criterion	-0.533032
Log likelihood	48.03747	Hannan-Quinn criter.	-0.582231
F-statistic	1.912271	Durbin-Watson stat	2.118860
Prob(F-statistic)	0.130406		
Inverted MA Roots	-.68		

Dependent Variable: DLSP
Method: Least Squares
Date: 11/07/18 Time: 15:52
Sample (adjusted): 1872 2015
Included observations: 144 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.042529	0.014678	2.897529	0.0044

R-squared	0.000000	Mean dependent var	0.042529
Adjusted R-squared	0.000000	S.D. dependent var	0.176134
S.E. of regression	0.176134	Akaike info criterion	-0.628223
Sum squared resid	4.436316	Schwarz criterion	-0.607599
Log likelihood	46.23206	Hannan-Quinn criter.	-0.619843
Durbin-Watson stat	1.851124		

Questions are on the next page.

(a) [20%] Use a Likelihood Ratio test at the 5% level to test between the two models. Comment on the view that because the AR and MA terms are significant the ARIMA(1,1,1) should be preferred.

(b) [20%] Explain the difference between a joint test that the two parameters of the ARIMA(1,1,1) model are zero and two individual tests. Explain why the joint and individual tests might conflict in this case.

(c) [20%] Explain how the Akaike information criterion, AIC, and the Schwarz Bayesian information criterion, BIC, are calculated.What is the difference between them? Which model would each of them choose?

(d) [20%] The values for LSP were 2014: 7.508; 2015: 7.615; 2016: 7.559. The 2015 residual was 0.03278. Use the two models to forecast LSP for 2016 and compare their errors.

(e) [20%] Briefly explain the non-linear estimation procedure used to obtain the maximum likelihood estimates of the ARIMA model.

3. US data for the logarithm of dividends, d_t, and the logarithm of earnings, e_t for US firms in the S&P500 were used to estimate ARDL and ECM models over the period 1872-2014. The models are

$$\begin{aligned} ARDL &: d_t = \alpha_0 + \alpha_1 d_{t-1} + \beta_0 e_t + \beta_1 e_{t-1} + u_t, \\ ECM &: \Delta d_t = a_0 + a_1 d_{t-1} + b_0 \Delta e_t + b_1 e_{t-1} + u_t, \end{aligned}$$

the estimated coefficients (standard errors) are

$$\begin{array}{lccccccccccl} d_t = & -0.13 & +0.65 & d_{t-1} & +0.20 & e_t & +0.12 & e_{t-1} & +\hat{u}_t & & R^2 = 0.997, \\ & (0.02) & (0.03) & & (0.03) & & (0.03) & & & & SER = 0.086 \\ \Delta d_t = & -0.13 & -0.35 & d_{t-1} & +0.20 & \Delta e_t & +0.32 & e_{t-1} & +\hat{u}_t & & R^2 = 0.526, \\ & (0.02) & (0.03) & & (0.03) & & (0.03) & & & & SER = 0.086 \end{array}$$

(a) [20%] Explain the relationship between the ARDL and the ECM parameter estimates. It is argued that the ARDL is a better model because the R^2 is higher. Do you agree with this.

(b) [20%] Calculate the long-run elasticity of earnings to dividends in each case.

(c) [20%] Wald tests on the ECM for the hypotheses (i) $a_1 + b_1 = 0$ and (ii) $a_1/b_1 + 1 =$ give chi squared test statistics of 27.1 and 48.6 respectively. Interpret the hypotheses. What are the 5% critical values? Use the critical values at the beginning of the exam.. Are the hypotheses rejected? What does this example reveal about Wald tests?

(d) [20%] Diagnostic tests on the ECM and their p values were: heteroskedasticity (regressing the squared residuals on the regressors) $p = 0.0035$; normality (skewness-kurtosis): $p = 0.0000$; up to second order serial correlation: $p = 0.0672$. In which cases do the tests indicate failure of the assumptions at the 5% level.? What are the consequences of the failures?

(e) [20%] Explain the null hypothesis of a RESET test and how it is calculated. The RESET p values for the ARDL are $p = 0.64$ and for the ECM $p = 0.99$. Comment on the results.