Queen Mary
University of London

May Examination Period 2022-23

ECN225 Econometrics 2 Duration: 2 hours

YOU ARE NOT PERMITTED TO READ THE CONTENTS OF THIS QUESTION PAPER UNTIL INSTRUCTED TO DO SO BY AN INVIGILATOR

Answer ALL questions. The questions carry equal marks. JUSTIFY ALL YOUR ANSWERS. Cross out any answers that you do not wish to be marked. There are tables of critical values from the standard normal, $F_{m,\infty}$, ADF, χ^2_m, and QLR distribution provided on pages 7-8.

Only nonprogrammable calculators are permitted in this examination. Please state on your answer book the name and type of machine used.

Complete all rough workings in the answer book and cross through any work that is not to be assessed.

Possession of unauthorised material at any time when under examination conditions is an assessment offence and can lead to expulsion from QMUL. Check now to ensure you do not have any notes, mobile phones, smartwatches or unauthorised electronic devices on your person. If you do, raise your hand and give them to an invigilator immediately.

It is also an offence to have any writing of any kind on your person, including on your body. If you are found to have hidden unauthorised material elsewhere, including toilets and cloakrooms it will be treated as being found in your possession. Unauthorised material found on your mobile phone or other electronic device will be considered the same as being in possession of paper notes. A mobile phone that causes a disruption in the exam is also an assessment offence.

EXAM PAPERS MUST NOT BE REMOVED FROM THE EXAM ROOM

Examiner: Dr. Jinu Lee

Question 1 (25 marks)

(a) One way to specify a nonlinear regression function is to use the natural logarithms of the dependent variable Y and/or independent variable X. Provide examples of applications in economics in which it would make sense to take the logarithm of Y, the logarithm of X, or the logarithm of both Y and X. **(3 marks)**

(b) For data on test scores and income, the following four results were obtained:

$$\widehat{TestScore} = \underset{(1.87)}{625.4} + \underset{(0.11)}{1.87} Income, \qquad \bar{R}^2 = 0.508$$

$$\widehat{TestScore} = \underset{(3.8)}{557.8} + \underset{(1.40)}{36.42} \ln(Income), \qquad \bar{R}^2 = 0.561$$

$$\ln(\widehat{TestScore}) = \underset{(0.003)}{6.439} + \underset{(0.00018)}{0.00284} Income, \qquad \bar{R}^2 = 0.497$$

$$\ln(\widehat{TestScore}) = \underset{(0.006)}{6.336} + \underset{(0.0021)}{0.055} \ln(Income), \qquad \bar{R}^2 = 0.557$$

where income is measured in thousands of pounds (£) and test scores are measured in points.
Can we use the $\bar{R}^2$ to compare the fit of the four models above? Which model should be used to analyse the effect of income on test scores, and why? **(10 marks)**

(c) Test for and interpret the estimated coefficient of the regressor for the model that you chose in part (b). **(4 marks)**

(d) By the model that you chose in part (b), further, discuss the claim that an income change has a larger effect on test scores in lower incomes than higher incomes. **(4 marks)**

(e) Suppose you would consider including higher powers of the existing regressor for the model that you chose in part (b) to improve the model. Briefly explain how you would verify the improvement over the original model. **(4 marks)**

Question 2 (25 marks)

(a) Define and interpret the j-th autocovariance and autocorrelation coefficients. **(4 marks)**

(b) Suppose that Y_t follows the stationary AR(1) model

$$Y_t = 1 + 0.3Y_{t-1} + u_t,$$

where errors u_t are independent and identically distributed with $E(u_t) = 0$ and $Var(u_t) = 4$. Compute the first two autocovariances of Y_t. **(8 marks)**

(c) A researcher needs to decide on how many lags to include in the AR(p) model. She estimates AR(p) models for $p = 1, \cdots, 6$ over the same period. She obtains the following results:

p	1	2	3	4	5	6
$\ln\left(\frac{SSR(p)}{T}\right)$	1.007	0.865	0.838	0.837	0.836	0.836
$(p+1)\frac{\ln(T)}{T}$	0.060	0.090	0.120	0.150	0.180	0.209

where $SSR(p)$ is the sum of squared residuals of the estimated AR(p) and T is the sample size. Using the Bayes information criterion (BIC), how does she estimate the number of lags? **(5 marks)**

(d) Suggest two other approaches for the choice of lag length? State their advantages and disadvantages with respect to the BIC criterion. **(4 marks)**

(e) Suppose you include four lagged values of an additional predictor, and the Granger-causality statistic on their estimated coefficients is 2.35. Briefly explain the Granger-causality test and interpret the result. **(4 marks)**

Question 3 (25 marks)

(a) Let X_t and Y_t be time series with a unit root, i.e. series that have a stochastic trend. Is it a good idea to run the OLS regression of Y_t on X_t? Why or why not? **(5 marks)**

(b) Suppose a researcher has a sample of inflation data up to the last quarter of the year 2020. In order to determine whether the inflation time series (Inf) contains a unit root, the following regression has been estimated,

$$\widehat{\Delta Inf}_t = \underset{(0.20)}{0.40} - \underset{(0.040)}{0.102} Inf_{t-1} - \underset{(0.077)}{0.22} \Delta Inf_{t-1} - \underset{(0.07)}{0.24} \Delta Inf_{t-2} + \underset{(0.07)}{0.18} \Delta Inf_{t-3},$$

where the number of lags was selected using the BIC criterion. Does the series (Inf) have a unit root? Explain and justify your answer. **(10 marks)**

(c) Suppose the researcher in part (b) wishes to forecast inflation in the first quarter of the year 2021 and considers the following regression:

$$\Delta Inf_t = \beta_0 + \beta_1 \Delta Inf_{t-1} + \beta_2 \Delta Inf_{t-2} + \beta_3 \Delta Inf_{t-3} + \beta_4 \Delta Inf_{t-4} + u_t.$$

After estimating the regression with the data and plugging in the last four available values of ΔInf_t, she obtains the forecast for the change in inflation, $\widehat{\Delta Inf}_{2021:I|2020:IV} = 0.4\%$. Since the last available value of inflation is $Inf_{2020:IV} = 3.5\%$, her final forecast is

$$\widehat{Inf}_{2021:I|2020:IV} = Inf_{2020:IV} + \widehat{\Delta Inf}_{2021:I|2020:IV} = 3.9\%.$$

Why does the researcher proceed in this way? Would it be possible to forecast the value of $Inf_{2021:I}$ directly from an AR(p) model for Inf_t? Explain your answer. **(5 marks)**

(d) Let u_t be independent and identically distributed random variables with $E(u_t) = 0$ and $Var(u_t) = \sigma^2$. Decide whether the following time series model is stationary or nonstationary:

$$Y_t = Y_{t-1} + u_t.$$

Explain and justify your answer. **(5 marks)**

Question 4 (25 marks)

A wage equation is analysed in a balanced panel of 595 observations on heads of households. The sample data are drawn from years 1976–1982 of a survey of income dynamic. The equation to be estimated is

$$\begin{aligned}\ln(Wage_{it}) &= \beta_1 + \beta_2 Exp_{it} + \beta_3 Exp_{it}^2 + \beta_4 Wks_{it} + \beta_5 Occ_{it} \\ &\quad + \beta_6 Ind_{it} + \beta_7 South_{it} + \beta_8 SMSA_{it} + \beta_9 MS_{it} \\ &\quad + \beta_{10} Union_{it} + \beta_{11} Fem_{it} + \beta_{12} Ed_{it} + \beta_{13} Blk_{it} + u_{it},\end{aligned}$$

where the variables are as follows:

Exp ...years of work experience,
Wks ...weeks worked,
Occ ...1 if blue-collar occupation, 0 if not,
Ind ...1 if the individual works in a manufacturing industry, 0 if not,
$South$ ...1 if the individual resides in the south, 0 if not,
$SMSA$ ...1 if the individual resides in a Standard Metropolitan Statistical Area (SMSA), 0 if not,
MS ...1 if the individual is married, 0 if not,
$Union$ ...1 if the individual wage is set by a union contract, 0 if not,
Fem ...1 if the individual is female, 0 if not,
Ed ...years of education,
Blk ...1 if the individual is black, 0 if not.

Note that variables Fem, Ed and Blk are time invariant (i.e. constant) in the sample. The average number of years of experience in the sample is 19.

(a) Regression (1) in Table 1 estimates the effect of additional education by regressing $\ln(Wage)$ on Exp, Exp^2, Wks, Ed and all the dummy variables listed above. Does the estimated regression suggest that additional years of schooling increases the wage? **(4 marks)**

(b) The researcher wishes to include fixed effects. To do so, she first drops regressors Fem, Ed and Blk. Why do these regressors have to be excluded when fixed effects are included? **(4 marks)**

(c) Do the results change when individual fixed effects are included? If the results do change, provide an explanation for why the results changed. **(4 marks)**

(d) In regression (3) in Table 1, appropriate time dummy variables are included to represent time fixed effects. Do the results further change compared to regression (2) in Table 1? Explain your answer. **(4 marks)**

(e) Which regression specification is most preferred and why, (1), (2), or (3)? Explain your answer. **(4 marks)**

(f) Using the results in regression (3) in Table 1, discuss the effect of increasing years of experience on the wage. **(5 marks)**

Continues on next page...

Table 1: Results for estimating regression models for fatalities (standard errors of the estimated coefficients in parentheses).

Regressor	(1)	(2)	(3)
Exp	0.0401 (0.0042)	0.113 (0.004)	0.112 (0.00197)
Exp^2	-0.000673 (0.000093)	-0.000418 (0.000060)	-0.000398 (0.0000595)
Wks	0.00422 (0.00152)	0.000836 (0.000685)	0.000652 (0.000684)
Occ	-0.140 (0.027)	-0.0215 (0.0147)	-0.0192 (0.0151)
Ind	0.0468 (0.0230)	0.0192 (0.0140)	0.0206 (0.0140)
$South$	-0.0557 (0.0253)	-0.0019 (0.0413)	0.00314 (0.0415)
$SMSA$	0.152 (0.024)	-0.0425 (0.0196)	-0.0423 (0.0189)
MS	0.0484 (0.0384)	-0.0297 (0.0150)	-0.0284 (0.0146)
$Union$	0.0926 (0.0243)	0.0328 (0.0198)	0.0299 (0.0195)
Ed	0.0567 (0.0052)	–	–
Fem	-0.368 (0.049)	–	–
Blk	-0.167 (0.045)	–	–
Individual Effects	No	Yes	Yes
Time Effects	No	No	Yes
$\bar{R}^2$	0.429	0.891	0.892

Table 2: Tests for joint significance of effects (null hypothesis: coefficients on effects are jointly zero).

Regression	Effects	F statistic	p value
(2)	Fixed effects	38.2	< 0.0001
(3)	Fixed effects	36.2	< 0.0001
(3)	Year effects	48.2	< 0.0001

End of Paper - An Appendix of 2 pages follows

Large Sample Critical Values for the t-statistic from the Standard Normal Distribution			
	Significance level		
	10%	5%	1%
2-Sided Test ($\neq$) Reject if $\lvert t \rvert$ is greater than	1.64	1.96	2.58
1-Sided Test (¿) Reject if t is greater than	1.28	1.64	2.33
1-Sided Test (¡) Reject if t is less than	-1.28	-1.64	-2.33

Critical Values for the $F_{m,\infty}$ Distribution			
	Significance level		
Degrees of freedom	10%	5%	1%
1	2.71	3.84	6.63
2	2.30	3.00	4.61
3	2.08	2.60	3.78
4	1.94	2.37	3.32
5	1.85	2.21	3.02
6	1.77	2.10	2.80
7	1.72	2.01	2.64
8	1.67	1.94	2.51
9	1.63	1.88	2.41
10	1.60	1.83	2.32

Large Sample Critical Values of the Augmented Dickey-Fuller Statistic			
	Significance level		
Deterministic regressors	10%	5%	1%
Intercept only	-2.57	-2.86	-3.43
Intercept and time trend	-3.12	-3.41	-3.96

Critical Values for the χ^2 Distribution			
	Significance level		
Degrees of freedom	10%	5%	1%
1	2.71	3.84	6.63
2	4.61	5.99	9.21
3	6.25	7.81	11.34
4	7.78	9.49	13.28
5	9.24	11.07	15.09
6	10.64	12.59	16.81
7	12.02	14.07	18.48
8	13.36	15.51	20.09
9	14.68	16.92	21.67
10	15.99	18.31	23.21

Critical Values of the QLR Statistic with 15% Trimming			
Number of Restrictions (q)	10%	5%	1%
1	7.12	8.68	12.16
2	5.00	5.86	7.78
3	4.09	4.71	6.02
4	3.59	4.09	5.12
5	3.02	3.66	4.53
6	2.84	3.37	4.12
7	2.69	3.15	3.82
8	2.58	2.98	3.57
9	2.48	2.84	3.38
10	2.40	2.71	3.23

End of Appendix