Universiti Tun Hussein Onn Malaysia

UNIVERSITI TUN HUSSEIN ONN MALAYSIA

FINAL EXAMINATION
SEMESTER II
SESSION 2023/2024

COURSE NAME	:	ECONOMETRICS
COURSE CODE	:	BWB 31603
PROGRAMME CODE	:	BWQ
EXAMINATION DATE	:	JULY 2024
DURATION	:	3 HOURS
INSTRUCTIONS	:	1. ANSWER ALL QUESTIONS 2. THIS FINAL EXAMINATION IS CONDUCTED VIA ☐ Open book ☒ Closed book 3. STUDENTS ARE **PROHIBITED** TO CONSULT THEIR OWN MATERIAL OR ANY EXTERNAL RESOURCES DURING THE EXAMINATION CONDUCTED VIA CLOSED BOOK

THIS QUESTION PAPER CONSISTS OF **TWELVE (12)** PAGES.

Part A Answer all the questions.

Q1. Which one of the following is **NOT** an example of functional form misspecification?

A. Using a linear specification when y scales as a function of the squares of x.

B. Using a linear specification when a double-logarithmic model would be more appropriate.

C. Modelling y as a function of x when it scales as a function of $1/x$.

D. Excluding a relevant variable from a linear regression model.

Q2. If a relevant variable is omitted from a regression equation, the consequences would be that:

i. The standard errors would be biased.

ii. If the excluded variable is uncorrelated with all the included variables, all the slope coefficients will be inconsistent.

iii. If the excluded variable is uncorrelated with all the included variables, all the intercept coefficients will be inconsistent.

iv. If the excluded variable is uncorrelated with all the included variables, all the slope and intercept coefficients will be consistent and unbiased but inefficient.

A. iii and iv

B. ii and iii

C. i, ii, and iii

D. i, ii, iii, and iv

Q3. Consider a simple regression model, $Y = \beta_0 + \beta_1 x + u_i$. To obtain consistent estimators of β_0 and β_1, when x and u are correlated, a new variable z is introduced into the model, which satisfies the following two conditions: $Cov(z, x) \neq 0$ and $Cov\,(z, u) = 0$. The variable z is called a (n) _____ variable.

A. dummy

B. instrumental

C. lagged dependent

D. random

Q4. Consider a simple regression model, $Y = \beta_0 + \beta_1 x + u$. Suppose z is an instrument for x. Which of the following conditions denotes instrument exogeneity?

A. $Cov(z,u) > 0$

B. $Cov(z,x) > 0$

C. $Cov(z,u) = 0$

D. $Cov(z,x) = 0$

Q5. Consider a simple regression model, $Y = \beta_0 + \beta_1 x + u$. Suppose z is an instrument for x. Which of the following conditions denotes instrument relevance?

A. $Cov(z,u) > 0$

B. $Cov(z,x) > 0$

C. $Cov(z,x) \neq 0$

D. $Cov(z,x) = 0$

Q6. Consider a simple regression model, $Y = \beta_0 + \beta_1 x + u$. Suppose z is an instrument for x. Which of the following statements is true?

A. The condition $Cov(z,u) = 0$ can be tested statistically.

B. The condition $Cov(z,x) \neq 0$ cannot be tested statistically.

C. The instrumental variables estimator is always biased if $Cov(x,u) \neq 0$.

D. The ordinary least squares estimator is unbiased if $Cov(x,u) \neq 0$.

Q7. Consider a simple regression model, $Y = \beta_0 + \beta_1 x + u$. The variable z is a poor instrument for x if _____.

A. there is a high correlation between z and x

B. there is a low correlation between z and x

C. there is a high correlation between z and u

D. there is a low correlation between z and u

Q8. Which of the following conditions requires the two-stage least squares estimation method?

A. There are perfect linear relationships among the instrumental variables.

B. There is a strong correlation between each instrumental variable and the error term.

C. The conditional variance of the error term depends on an exogenous explanatory variable.

D. The error term has zero mean.

Q9. Which of the following is **TRUE** of two-stage least squares estimators?

A. The two-stage least squares estimator is equal to the instrumental variable estimator if R^2 equals 1.

B. The two-stage least squares estimators are biased if the regression model exhibits multicollinearity.

C. The two-stage least squares estimators have lower variance than the ordinary least squares estimators.

D. The two-stage least squares estimators have large standard errors when R^2 lies close to zero.

Q10. Which of the following tests are used to compare different instrumental variable estimates of the same parameter?

A. Overidentifying restrictions.

B. Endogeneity.

C. Heteroskedasticity.

D. Serial correlation.

Q11. Which of the following is **TRUE**?

A. Functional form misspecification can occur if the level of a variable is used when the logarithm is more appropriate.

B. Functional form misspecification occurs only if a key variable is uncorrelated with the error term.

C. Functional form misspecification does not lead to bias in ordinary least squares estimators.

D. Functional form misspecification does not lead to inconsistency in the ordinary least square estimators.

Q12. Which of the following is **TRUE** about a regression model that suffers from functional form misspecification?

A. a key variable is binary.

B. the dependent variable is binary.

C. an interaction term is omitted.

D. the coefficient of a key variable is zero.

Q13. Which of the following statements is **TRUE** of simultaneity in econometrics?

A. strictly exogenous explanatory variables determine the dependent variable through a step-by-step process.

B. the error term correlates with the dependent and explanatory variables.

C. one or more of the explanatory variables is jointly determined with the dependent variable.

D. both serial correlation and heteroskedasticity are present in a hypothesised model.

Q14. Which of the following statements is **TRUE** about the characteristic of structural equations?

A. A structural equation should contain an equal number of dependent and independent variables.

B. A structural equation should contain an equal number of endogenous and exogenous variables.

C. A structural equation should have a behavioural, ceteris paribus interpretation.

D. A structural equation should not contain structural errors.

Q15. The following simultaneous equations describe the demand and supply for a particular good in a competitive market. $Q_i = \alpha_1 P_i + \beta_1 z_{i1} + u_{i1}$ and $Q_i = \alpha_2 P_i + \beta_2 z_{i2} + u_{i2}$. Which of the following are the endogenous variables in these equations?

A. P_i, z_{i1} and z_{i2}

B. P_i and Q_i

C. z_{i1} and z_{i2}

D. u_{i1} and u_{i2}

Q16. Which of the following statements **CORRECTLY** explains the endogenous variable in a simultaneous equation model?

A. The values of endogenous variables are determined outside the system.

B. There can be fewer equations in the system than endogenous variables.

C. Reduced form equations will not contain any endogenous variables on the right-hand side of the equation.

D. Reduced form equations will contain only endogenous variables on the right-hand side of the equation.

Q17. What would be the consequences for the estimators if the parameters were estimated separately for each model in a simultaneous equation system using OLS?

A. Biased but consistent.

B. Unbiased and consistent

C. Biased and inconsistent

D. It is impossible to apply OLS to equations that are part of simultaneous models.

Q18. Which of the following estimation techniques are available to estimate over-identified systems of simultaneous equations?

i. Ordinary least squares

ii. Indirect least squares

iii. Two-stage least squares

iv. Instrumental variables

A. iii only

B. iii and iv only

C. ii, iii, iv only

D. All of the above

Q19. If you model the income of individuals on a constant, a binary variable ("Male") takes on the value one for males and is zero otherwise. Another binary variable ("Female") takes on the value one for females and is zero otherwise. Because females typically earn less than males, you would expect that _____.

A. both coefficients are to have the same distance from the constant, one above and the other below.

B. none of the OLS estimators exists because there is perfect multicollinearity.

C. the coefficient for males to have a positive sign and for females, a negative sign.

D. this is to yield a difference in means statistic.

Q20. For a one-sided Durbin-Watson (DW) test with a null hypothesis of no positive serial correlation against an alternative hypothesis of positive serial correlation. Then _____.

A. the border value of ρ under the null is one.

B. the border value of ρ under the null is zero.

C. the border value of ρ under the null is -1.

D. None of the above is a true statement.

(20 marks)

Part B: Answer all the questions.

Q1 Explain how heteroskedasticity influences the OLS estimates of our model and what steps would be taken to mitigate this issue throughout the estimation procedure.

(4 marks)

Q2 Consider the earnings model:

$$Wage_i = \beta_0 + \beta_1 Exper_i + \beta_2 Edu_i + u_i$$

where $Wage$ is measured in dollars per hour, $Exper$ is work experience in years, and $Educ$ is the number of years of schooling. **Figure Q2.1** shows the OLS regression results for N=100 males each year.

```
  Source |       SS       df       MS              Number of obs =     100
---------+------------------------------           F(  2,    97) =   16.47
   Model |  2057.5037     2  1028.75185            Prob > F      =  0.0000
Residual |  6059.71269    97  62.4712648           R-squared     =  0.2535
---------+------------------------------           Adj R-squared =  0.2381
   Total |  8117.21639    99  81.9920847           Root MSE      =  7.9039

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        Educ |   1.435782   .321546     4.47   0.000     .7976026    2.073962
       Exper |    .328525   .0658247    4.99   0.000     .1978813    .4591687
       _cons |  -11.91922   4.750254   -2.51   0.014    -21.34716   -2.491275
------------------------------------------------------------------------------
```

Figure Q2.1 STATA results from OLS estimation of the earnings model

(a) Revise the earnings model to examine the hypothesis that male labour union members earn more on average than their non-union male counterparts. Explain how to conduct a test to assess this hypothesis.

(4 marks)

(b) Revise the earnings model to test the hypothesis that men in a labour union earn more for each additional year of experience than men who are not members of the labour union.

(4 marks)

Q3 Assume that there is an estimated salary function where the log of salary is regressed with a set of continuous explanatory variables (in levels) and two binary variables, one for gender and the other for marital status. One of the explanatory variables is working experience.

(a) Interpret the working experience coefficient.

(3 marks)

(b) Specify the binary variables and an equation, where the default is a single male, without allowing for interaction between marital status and gender. Indicate the coefficients that measure the effect of a single male, single female, married male, and married female.

(7 marks)

(c) Allow for an interaction between the binary variables of gender and marital status. State the model with the various effects based on these variables. Explain why this model is better than the model in **Q3(b)**.

(10 marks)

Q4 Consider the following two models.

$$Y_1 = \gamma_1 Y_2 + \beta_{11} X_1 + \beta_{21} X_2 + \beta_{31} X_3 + \epsilon_1 \qquad (1)$$

$$Y_2 = \gamma_2 Y_1 + \beta_{12} X_1 + \beta_{22} X_2 + \beta_{32} X_3 + \epsilon_2 \qquad (2)$$

Write the structural model in a matrix form.

(6 marks)

Q5 Consider the following data on the KLSE index as the dependent variable. The independent variables are exchange rate, interest rate, gold price, consumer price index, money supply M3, industrial production, and crude oil price.

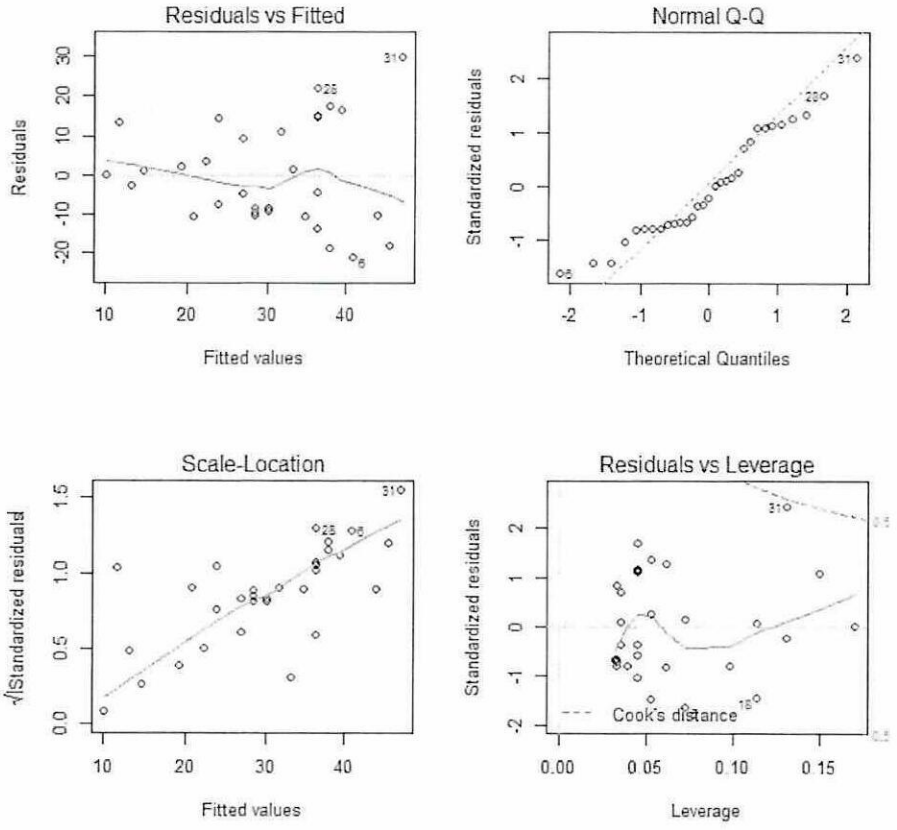

Figure Q5.1 Residual

(a) Based on **Figure Q5.1**, identify whether there is any evidence of heteroscedasticity and justify the answer.

(4 marks)

(b) Based on **Figure Q5.2,** write the null and hypothesis to conduct a White Test and conclude whether there is heteroscedasticity.

```
> skedastic::white_lm(model)
# A tibble: 1 x 5
  statistic  p.value parameter method       alternative
      <dbl>    <dbl>     <dbl> <chr>        <chr>
1      17.1 0.000197         2 White's Test greater
```

Figure Q5.2 White Test Output

(6 marks)

Q6 In basketball analytics, teams often analyse players' shooting percentages to evaluate their offensive efficiency and contributions to winning. Factors like field goal percentage and three-point shooting accuracy in the NBA significantly determine a team's success. All 30 NBA teams for the 2022 season. Shooting effectiveness is typically gauged by metrics such as "Effective Field Goal Percentage" (eFG%) and "Three-Point Percentage" (3P%). **Table Q6.1** shows the summary of the Distribution of Winning Percentage for the NBA in 2022

Table Q6.1

	Average	Standard Deviation	Percentile						
			10%	25%	40%	50%	60%	75%	90%
eFG	4.71	0.53	3.84	4.35	4.72	4.78	4.91	5.06	5.25
3P%	0.78	0.03	0.72	0.75	0.77	0.78	0.79	0.80	0.82
Winning %	0.50	0.08	0.40	0.43	0.46	0.48	0.49	0.59	0.60

The regression output is

$$\widehat{Winpct} = -0.19 - 0.099 \times eFG + 1.490 \times 3P\%$$

(0.08) (0.008) (0.126)

$$R^2 = 0.92, SER = 0.02$$

(a) Interpret the regression and whether the results are statistically significant.

(5 marks)

(b) There are two leagues in the NBA, League A (LA) and League B (LB). One significant difference is that the point guard in LA does not have to rebound. Instead, there is a "designated rebounder" in the line-up. You are concerned that, as a result, there is a different effect of playmaking and rebounding in the LA from the LB. To test this hypothesis, you allow the LA regression to have a different intercept and slopes from the LB regression. Therefore, create a binary variable for the League A (DLA) and estimate the following specification:

$$\widehat{Winpct} = -0.29 + 0.10 \times DLA - 0.100 \times eFG + 0.008 \times (DLA \times eFG) +$$

$$(0.12) \quad (0.24) \qquad (0.008) \qquad (0.018)$$

$$1.622 \times 3P\% - 0.187 \times (DLA \times 3P\%)$$

$$(0.163) \qquad (0.160)$$

$$\boldsymbol{R^2 = 0.92, SER = 0.02}$$

Write the regression model for the winning percentage in the LA and LB. Since all slopes and the intercept are allowed to vary between the two leagues, explain what the results would imply if all coefficients involving DLA were statistically significant.

(7 marks)

- END OF QUESTIONS -