FACULTY/COLLEGE	College of Business and Economics
SCHOOL	School of Economics
CAMPUS(ES)	APK
MODULE NAME	Econometrics 3A
MODULE CODE	ECM03A3
SEMESTER	First
ASSESSMENT OPPORTUNITY, MONTH AND YEAR	Special Summative Assessment Opportunity August 2022

ASSESSMENT DATE	August 2022	**SESSION**	
ASSESSOR(S)	Dr M Pretorius and Dr C Saba		
MODERATOR(S)	Ms B Thobejane (UJ) and Ms N Cattaneo (Rhodes University)		
DURATION	3 hours (180 min)	**TOTAL MARKS**	150

NUMBER OF PAGES OF QUESTION PAPER (Including cover page)	5

INSTRUCTIONS:

- This is a closed-book assessment.
- There are five questions (A to E). **Answer all questions.**
- Read the questions carefully and answer only what is required.
- Number your answers clearly and correctly as per the question paper.
- Write neatly and legibly on both sides of the paper in the answer book, starting on the first page.
- Calculators may be used.
- Please round your final answers off to 2 decimal places.

QUESTION A **[50 MARKS]**

1. Are the following statements true or false? If a statement is false, discuss why this is the case. (24)

a. Regression analysis is concerned with the study of the dependence of the dependent variable, on one or more explanatory variables, with a view to estimating the population mean in terms of the known values of the explanatory variables. True
b. In statistical relationships we essentially deal with stochastic variables. True
c. A statistical relationship in itself cannot logically imply causation. True
d. A population regression curve is the locus of the conditional means of the dependent variable for the fixed values of the explanatory variable/s. True
e. A linear population regression function is linear in the parameters. True
f. Given the assumptions of the classical linear regression model, the least-squares estimators, in the class of unbiased linear estimators, have minimum variance, they are therefore BLUE. True
g. The R^2 statistic measures the proportion of the total variation in Y explained by the regression model. True
h. The classical normal linear regression model differs from the classical linear regression model in that it specifically assumes that the disturbance term entering the regression model is normally distributed. True
i. The p value and the size of a test statistic mean the same thing. True
j. The t test of significance requires that the sampling distributions of the estimators follow the normal distribution. True
k. If you regress the actual Y on the estimated Y, the intercept and slope values will be 0 and 1, respectively. True
l. In the two-variable PRF, the slope coefficient β_2 is zero and the intercept β_1 is estimated by the sample mean $\hat{Y}$. True

2. Describe any five assumptions underlying the method of least squares. (10)
Any 5:

a. Linear regression model, or linear in the parameters.

b. Fixed X values or X values independent of the error term. Here, this means we require zero covariance between U_i and each X variables. cov (U_i, X_{2i}) = cov (U_i, X_{3i}) = 0

c. Zero mean value of disturbance U_i. E(U_i |X_{2i}, X_{3i})=0 for each i

d. Homoscedasticity or constant variance of U_i. var (U_i)= σ^2

e. No autocorrelation, or serial correlation, between the disturbances. cov (U_i, U_j) =0 i = j

f. The number of observations n must be greater than the number of parameters to be estimated, which is 3 in our current case.

g. There must be variation in the values of the X variables.

h. No exact collinearity between the X variables. No exact linear relationship between X_2 and X_3

I. There is no specification bias. The model is correctly specified.

3. Highlight six reasons why the variance of u_i may not be constant. (6)

a. Error learning models
b. Discretionary income
c. Improved data collection techniques
d. Presence of outliers
e. Specification error (omission of important variable)
f. Skewness in the distribution of one or more explanatory variables
g. Incorrect transformation and functional form

4. What are the five practical consequences of multicollinearity? (5)

1. Although BLUE, the OLS estimators have large variances and covariances, making precise estimation difficult. (1mark)
2. Because of consequence 1, the confidence intervals tend to be much wider, leading to the acceptance of the "zero null hypothesis" (i.e., the true population coefficient is zero) more readily. (1marks)
3. Also because of consequence 1, the t ratio of one or more coefficients tends to be statistically insignificant. (1marks)
4. Although the t ratio of one or more coefficients is statistically insignificant, R^2, the overall measure of goodness of fit, can be very high. (1marks)

5. The OLS estimators and their standard errors can be sensitive to small changes in the data. (1marks)

5. Name any five methods of detecting heteroscedasticity in a regression model. (5)

a. Graphical method
b. Park Test
c. Glejser Test
d. Spearman's rank correlation Test
e. Goldfeld-Quandt Test
f. Breusch-Pagan-Godfrey Test
g. White's general heteroscedasticity test

QUESTION B **[26 MARKS]**

Data has been provided for the real Gross Domestic Product (GDP) of South Africa as well as the population estimates of the country in the Excel worksheet "Question B". Make use of the data where necessary and answer the following questions:

1. Do you think there is a relationship between GDP and population? Discuss the economic relationship between these variables. (3)
   Yes, a positive relationship is expected – as the population of a country increases, it is expected that the GDP of a country will also increase as there are more people available to contribute to the economy. Mark were allocated if arguments made economic sense.
2. Estimate the correlation between GDP and population and interpret. Does the result correspond to your expectation in 1? (2)

GDP	1	0.96427
POPULATION	0.96427	1

Strong positive correlation.

Corresponds to expectation in 1.

3. Calculate GDP per capita for 2006 quarter 1 (the GDP data is in terms of millions of rands). Interpret this number. (2)
   GDP / Population - The average GDP per head in South Africa during 2006 quarter is R5866-19.
4. What were South Africa's mean and median population statistics during the sample period? (2)

	POPULATION
Mean	30 918 535
Median	29 724 000

5. Give and interpret three measures of dispersion for the GDP variable in the sample period. (6)
   Standard deviation = 54177.58 – 68% of the data will lie within one standard deviation of the mean [111202.5; 54177.58]

   Variance – standard deviation squared = 2.94E+09
   Range = Max – Min = 285 431 – 66 045 = 219 386
6. Consider the following regression output and answer the questions that follow.

Dependent Variable: GDP
Method: Least Squares
Sample: 1960Q1 2006Q1
Included observations: 185

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-15942.46	3831.346		0.0000
POPULATION	0.005865	0.000119		0.0000
R-squared	0.929823			

a. Interpret the coefficient of the independent variable. (2)
   If population increases with one unit (person), GDP increases with 0.006 units (R5856).

b. Is the population coefficient statistically significant? Explain by making use of the t-statistic (you may make use of the rule of thumb). (4)
   - H_0: $\beta_2 = 0$
   - H_1: $\beta_2 \neq 0$
   - Reject if the test statistic (calculated/eviews) > 2 (big samples, df > 20 and level of significance = 0.05)
   - Test statistic (from Eviews) = 0.005865/0.000119 = 49.29
   - Therefore reject H_0 and conclude that population is statistically significant.

c. Why are you allowed to make use of the rule of thumb in the previous question? (2)
   Because the degrees of freedom for this model is greater than 20 (it is 185-2 = 183).

d. Is this a good model? Explain. (3)

Yes, this is a good model – the R^2 is 0.929823 which means that 92.98% of the variation in GDP is explained by population.

QUESTION C **[24 MARKS]**

Data is provided for retail trade (Retail) and the Producer Price Index (PPI) in the Excel worksheet "Question C". Answer the following questions (make use of the data where necessary):

Retail depends on PPI.

1. Specify the econometric model. (2)
   $Retail = \beta_1 + \beta_2 PPI + u_i$
2. What do you think will the relationship be between the two variables? Why? (3)
   We expect a positive relationship – increased producer inflation translates into increased prices. According to the law of supply as prices increase, quantity supplied will also increase. Greater supply will translate into increased retail trade.
3. Estimate the model and write down your regression. (2)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	11109.19	6167.556	1.801230	0.0794
PPI	183.9713	38.12732	4.825184	0.0000

Retail = 11109.19 + 183.97PPI

4. Interpret the slope coefficient. Does the model confirm your expectation in 2? (3)
   If PPI increases with one unit, Retail increases with 183.9713 units. The positive relationship expected in 2 is confirmed in the regression model.
5. Calculate the confidence interval for the slope coefficient. Make use of a 5% level of significance and interpret your result. Show all your calculations. (5)
   $\hat{\beta}_2 \pm t_{\frac{\alpha}{2}}\, se(\hat{\beta}_2)$
   [Critical t-value: df = 41 – 2 = 39, make use of Excel: TINV(0.05,39)=2.022691]
   $$183.9713 \pm (2.022691 * 38.12732) = 183.9713 \pm 77.11978$$
   $$[106.8515;\ 261.0911]$$
6. Is the slope coefficient statistically significant? Explain by referring to the t-statistic. (4)
   - H_0: $\beta_2 = 0$
   - H_1: $\beta_2 \neq 0$
   - Reject if the test statistic (calculated/eviews) > critical value (table/excel)
   - Test statistic (from Eviews) = 4.825184
- Critical value (from tables/Excel) = 2.022691 Therefore reject H0 and conclude that PPI is statistically significant zero.
7. Is this a good model? Use an appropriate statistic to prove your answer. (3)
   R2 = 0.373820, no, this is not a very good model. Only 37.38% of the variation in retail is explained by PPI.
8. Suggest two other variables that could be added as explanatory variables in this model. (2)

Marks were allocated if the variables suggested made economic sense.

QUESTION D **[18 MARKS]**

The aim of this question is to determine what factors play a role in determining the hourly wage of individuals in South Africa. The following equation is estimated based on the information of 63 individuals.

Wage = 31 + 1.42Education + 0.18Experience – 5.6Female – 0.04Female*Experience

Where
Wage – Earnings per hour in rand
Education – Number of years of formal schooling
Experience – Number of years of formal work experience
Female – Dummy = 1 if worker is female, Dummy = 0 if worker is male

1. Assume that all estimators are statistically significant. Explain how the explanatory variables have different impacts on the wages of males and females. (Hint: derive the different equations for male and female workers and formulate your answer accordingly). (12)
   WAGE = 31 + 1.42EDUCATION + 0.18EXPERIENCE – 5.60FEMALE – 0.04FEMALE*EXPERIENCE

   Male workers: WAGE = 31 + 1.42EDUCATION + 0.18EXPERIENCE
   Female workers: WAGE = (31 – 5.60) + 1.42EDUCATION + (0.18-0.04) EXPERIENCE

   The intercept of the Male equation is R5.60 more than for Female workers.
   Both males and females receive an additional R1.42 for every additional year of formal schooling
   Females receives R0.04 less than males for every additional year of work experience
2. Assume that the logarithmic format of the original relationship is as follows:

   Log (Wage) = 4.8 + 1.23 Log (Education) + 0.74 Log (Experience)

   Interpret the estimated coefficients of Log (Experience) and Log (Education). (6)
   For every 1% increase in years of formal work experience, the wage of the average worker increases with 0.74%, ceteris paribus.
   For every 1% increase in the years of education, the wage of the average worker increases by 1.23%, ceteris paribus.

QUESTION E **[32 MARKS]**

Make use of the Excel file "Question E" to answer the following questions. The data ranges from 1970 to 2016 and is supplied for South Africa.

N_t is the demand for labour in the agriculture sector RW_t is the real wage for the agriculture sector

$Output_t$ is the real output in the agriculture sector

1. Estimate the following model and write down your regression results: (3)
$$\log(N_t) = \alpha_1 + \alpha_2 \log(RW_t) + \alpha_3 \log(Output_t) + u_t$$
$$log(N_t) = 17.29 - 0.74 logRW_t + 0.04\, log(Output_t)$$
2. Interpret the partial slope coefficients of $\log(RW_t)$ and $\log(Output_t)$. (4)
A one percent increase in real wages will reduce the demand for labour in the agriculture sector by 0.74% holding real output constant.
A one percent increase in real output will increase the demand for labour in the agriculture sector by 0.04% holding real wages constant.
3. Is this a good model? Explain by making use of an appropriate statistic. Show all your steps. (3)
It is a good model, because 95% of variations in demand for labour are explained by real wages and real output jointly.
4. Are the coefficients of the explanatory variables statistically significant? Explain. (4)
The coefficient of real wages is significant (1) at the 1 per cent level of significance (using the p-value).
The coefficient of real output is insignificant (1) at the 10 per cent level of significance (using the p-value).
5. Test for the overall significance of this regression by making use of an appropriate statistic. Interpret your results. (3)
$H_0: \alpha_2 = \alpha_3 = 0$
Using the p-value of the F-statistic 406.05, we find that overall the model is statistically significant at 1 per cent level of significance.
6. Use the Jarque-Bera test to determine whether the residuals are normally distributed. Show all your steps. (4)
H_0: the residuals are normally distributed
Using the Jacque-Bera test result is 1.62 with a corresponding p-value of 0.44. Therefore fail to reject the null hypothesis and conclude that the residuals are normally distributed.
7. Is there a structural break in the regression in 1994? Motivate your answer by making use of the Chow test. Show all your steps. (4)

Chow Breakpoint Test: 1994
Null Hypothesis: No breaks at specified breakpoints
Varying regressors: All equation variables
Equation Sample: 1970 2016

F-statistic	39.16550	Prob. F(3,41)	0.0000
Log likelihood ratio	63.55154	Prob. Chi-Square(3)	0.0000
Wald Statistic	117.4965	Prob. Chi-Square(3)	0.0000

Based on the Chow test with F-stat of 39.17 and p-value of zero, we reject the null hypothesis and conclude that there was a structural break in 1994.

8. Test for the presence of heteroscedasticity by making use of the Breusch-Pagan-Godfrey test and the White test. Show all your steps. (7)

H_o: There is homoscedasticity

Breusch-Pagan-Godfrey test – The probability of the chi-square of Obs*R-squared is 0.95

White's test – F-statistic 0.67 and pvalue 0.98

Breusch-Pagan-Godfrey and White test confirm there is homoscedasticity in the regression residuals, thus we fail to reject the null hypothesis in both cases.

Heteroskedasticity Test: Breusch-Pagan-Godfrey
Null hypothesis: Homoskedasticity

F-statistic	0.051312	Prob. F(2,44)	0.9500
Obs*R-squared	0.109367	Prob. Chi-Square(2)	0.9468
Scaled explained SS	0.054685	Prob. Chi-Square(2)	0.9730

Heteroskedasticity Test: White
Null hypothesis: Homoskedasticity

F-statistic	0.119312	Prob. F(5,41)	0.9874
Obs*R-squared	0.674054	Prob. Chi-Square(5)	0.9844
Scaled explained SS	0.337035	Prob. Chi-Square(5)	0.9969

-----o0o-----