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	Supplementary Summative Assessment Opportunity July 2022

ASSESSMENT DATE	July 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)	6

INSTRUCTIONS:

- This is a closed-book assessment.
- There are four questions (A to D). **Answer all questions.**
- Read the questions carefully and only answer 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. State whether the following statements are true or false. (20)
   a. A type II error occurs when rejecting a true hypothesis. False
   b. Since the correlation between two variables, Y and X, can range from -1 to +1, this also means that $cov(Y, X)$ also lies between these limits. False
   c. In a regression model that contains the intercept, the sum of the residuals is always zero. True
   d. Despite perfect multicollinearity, OLS estimators are BLUE. False
   e. You will not obtain a high R^2 value in a multiple regression if all the partial slope coefficients are individually statistically insignificant based on the usual t-test. False
   f. In the presence of heteroscedasticity, the usual OLS method always overestimates the standard errors of estimators. False
   g. The OLS residuals will show a distinct pattern if a regression model is misspecified. True
   h. Excluding an important variable/s from a regression model may give a significant Durbin-Watson d statistic. True
   i. The estimators are no longer BLUE if autocorrelation is present in a model. True
   j. If perfect collinearity exists in a model, estimating the regression coefficients is still possible. False

2. Explain what the difference is between an error term and a residual. (5)
   A regression model can never be a completely accurate description of reality. Therefore, there is bound to be some difference between the actual values of the regressand and its values estimated from the chosen model. This difference is simply the stochastic error term. The residual is the sample counterpart of the stochastic error term.
   The error is the deviation from a sample observation and the true regression line and the residual is the deviation from the estimated regression line.
3. Describe the nine assumptions of the Classical Linear Regression Model (CLRM). (9)

   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.

4. Highlight the six rule of thumb approaches when you have to find a remedy to the problem of multicollinearity. (6)
   a. A priori information
   b. Combining cross sectional and time series data
   c. Dropping a variable(s) and specification bias
   d. Transformation of variables
   e. Additional or new data
   f. Reducing collinearity in polynomial regression
5. Name six sources of autocorrelation. (6)
   a. Inertia/Sluggishness
   b. Omitted variables
   c. Incorrect functional form
   d. Cobweb phenomenon
   e. Use of lags
   f. Manipulation of data
   g. Data transformation
   h. Non stationarity
6. Explain four consequences of using OLS in the presence of autocorrelation. (4)
   a. Autocorrelation affects standard errors and variances.
   b. The residual variance is likely to underestimate the true variance base on point (a) above.
   c. Likely to overestimate R^2
   d. Usual t- and F-tests of significance are no longer valid and will give misleading conclusions about the statistical significance of the estimated coefficients.

QUESTION B **[50 MARKS]**

Consider the following estimation outputs and statistics for Model 1. Answer the questions that follow:

Model 1			
Dependent Variable: GDP			
Method: Least Squares			
Sample: 2001 2017			
Included observations: 17			
Variable	**Coefficient**	**Std. Error**	**Prob.**
C	-2011351	355922.1	0
POPULATION	92443.92	6940.324	0
R-squared	0.922045	F-statistic	177.4178
Adjusted R-squared	0.916848	Prob(F-statistic)	0

Correlation	POPULATION	GDP
POPULATION	1	0.960231555
GDP	0.960231555	1
	POPULATION	**GDP**
Mean	51.16201	2718266
Median	50.85038	2748008
Maximum	57.42396	3144539
Minimum	45.92083	2081837
Std. Dev.	3.632143	349675.6
Observations	17	17

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 and spend in the economy.
   Marks were allocated if arguments made economic sense.
2. Interpret the correlation coefficient of GDP and the population variable. Does this statistic correspond to what you expected in 1? (3)
   0.96 which indicates a very strong positive correlation between the variables. This corresponds to expectations in 1.
3. Interpret the slope coefficient of the estimated regression between GDP and population. Does this relationship correspond to what you expected in 1? (3)
   If population increases with one unit, GDP will increase with 92443.92 units, ceteris paribus. Once more this corresponds to the expectation in 1.
4. It is stated that "2010 = 100" for the GDP variable used in the analyses above – explain what this means? (3)
   It means that 2010 was the base year for the GDP variable. The GDP variable is thus stated in real terms – it has been adjusted for inflation.
5. Interpret two measures of central tendency for the Population variable. (4)
   The average population in the sample was 51.16201.
   The middle value when the data was ordered in ascending/descending values is 50.85038.
6. Interpret three measures of dispersion for the GDP variable. (6)
   The difference between the maximum and the minimum value of GDP is R1062702.
   The standard deviation (349675.6) shows that 68% of the values will lie within one standard deviation from the mean [2368590.4; 3067941.6].
   The variance is the standard deviation squared which is 122273025235.36.
7. Construct a 95% confidence interval for the Population coefficient of Model 1. Show all your steps. (4)

$$\hat{\beta}_2 \pm t_{\frac{\alpha}{2}}\, se(\hat{\beta}_2)$$

[Critical t-value: df = 17 – 2 = 15, make use of Excel: TINV(0.05,15)=2.13145]

$$92443.92 \pm (2.13145 * 6940.324) = 92443.92 \pm 14792.9535898$$

[77650.97; 107236.87]

8. Is the Population variable statistically significant in Model 1? Explain by making use of the confidence interval that you constructed in 7. Show all your steps. (3)
   The confidence interval in 7 does not include the value zero and therefore, the population variable is statistically significant.
9. Are you allowed to make use of the "2-t" rule of thumb to determine individual statistical significance for the Population variable in Model 1? Explain. (2)
   No, we are not – there are not enough observations included in the regression. The degrees of freedom are too low.

Consider the following alternative specifications of the relationship between GDP and population. Answer the questions that follow:

Model 2			
Dependent Variable: GDP			
Variable	**Coefficient**	**Std. Error**	**Prob.**
POPULATION	53316.15	816.7534	0

Model 3			
Dependent Variable: Log (GDP)			
Variable	**Coefficient**	**Std. Error**	**Prob.**
C	7.680129	0.579813	0
Log (POPULATION)	1.812302	0.147414	0

Model 4			
Dependent Variable: Standardized GDP			
Variable	**Coefficient**	**Std. Error**	**Prob.**
Standardized POPULATION	53316.15	816.7534	0

Model 5			
Dependent Variable: GDP			
Variable	**Coefficient**	**Std. Error**	**Prob.**
C	7559685	289499	0
1/POPULATION	-2.47E+08	14707602	0

10. What type of model is model 2 known as? (1)
    Regression through the origin.

11. Interpret the slope coefficient in model 2. (2)

If population increases with one unit, GDP will increase with 53316.15 units

12. Considering the estimation results for model 1, do you think the specification of model 2 is justified? Explain. (3)

No it is not. The constant was statistically significant in Model 1 and therefore should not be left out.

13. What type of model is model 3 known as? (1)

Double log / Log-log / Log-linear model.

14. Interpret the slope coefficient in model 3. (2)

If population increases with one percent, GDP will increase with 1.81%.

15. What type of model is model 4 known as? (1)

Model with standardised variables

16. Interpret the slope coefficient in model 4. (2)

If population increases with one standard deviation unit, GDP increases with 53316.15 units

17. What type of model is model 5 known as? (1)

Reciprocal model

18. Interpret the slope coefficient in model 5. (2)

If population increases indefinitely, GDP will tend toward the constant of 7559685.

19. Considering the results in models 1 to 5, what is the relationship between GDP and Population? Which model do you consider to be the best one? Provide reasons for your answer. (4)

The relationship is positive.

3 Marks were allocated for sound motivations on which model students found best.

QUESTION C **[25 MARKS]**

The following regression identifies the factors influencing the salary of the average South African citizen. Answer the questions that follow.

Dependent Variable: Wage

Included Observations: 130

Variable	Coefficient	Std. Error	t-statistic
AGE	1785.945	56.90067	31.38707
AGE^2	-15.35752	0.676026	-22.71734
SCHL	4529.506	66.14925	68.47404
WKHP	1191.953	14.43777	82.55794
SCHL*DUMIMM	-1292.283	105.1906	-12.28515
DUMIMM	25764.13	1714.267	15.02924
DUMMALE	12927.43	6155.919	2.100000
DUMMAR	8358.173	287.9326	29.02823
DUMSELF	8880.847	1215.960	7.303567

C	-141649.2	1508.374	-93.90854
R-squared	0.258455	Mean dependent var	44223.74
Adjusted R-squared	0.258396	S.D. dependent var	55314.51
F-statistic	4.346437	Durbin-Watson stat	1.966950

Where:
WAGE: Weekly salary earned by an individual in Rands(R)
AGE: Age of individual in years
SCHL: Number of years of formal education
WKHP: Hours worked per week
DUMIMM: Dummy = 1 if person is an immigrant (not born in SA) and = 0 otherwise
DUMMALE: Dummy = 1 if person is male and = 0 otherwise
DUMSELF: Dummy = 1 if person works for him/herself and = 0 otherwise

1. Interpret the estimated coefficients of all the dummy variables. (9)

DUMIMM (25 764.13): Immigrants (not born in the SA) receive R25 764.13 more per week than the salary earned by workers born in SA– ceteris paribus
DUMMALE (12 927.43): Male workers receive R12 927.43 more in terms of weekly salary than female workers– ceteris paribus.
DUMSELF (8 880.85): People working for him/herself earn R8 880.85 more per week than people working for a boss and earning a salary–ceteris paribus.

2. Calculate the turning point of this relationship and interpret your results. (9)

Wage = -141649.2 + 1785.945 AGE – 15.358 AGE^2 + 4529.51 SCHL .....
δWAGE/ δAGE = 1785.95 – 30.72 AGE
for turning point set equal to 0
AGE=1785.95/30.72
therefore AGE = 58.14 at turning point
So WAGE increases as AGE increases until individual reaches 58.13 years and then the wage starts to decline.

3. Use your results in 2. and calculate the increase in the salary of the average 60 year old individual and interpret your result. (7)

δWAGE/ δAGE(60) = 1785.95 – 30.72 (60)
= -57.25
At age 60 an individual will receive $57.25 less than at 59 – ceteris paribus. This then confirms the calculated turning point of around 58 years.

QUESTION D **[25 MARKS]**

Make use of the Excel file “Question D” to answer the following questions. The data ranges from 1950 to 2019.

q_x is quantity of exports for South Africa (dependent variable)
pl_x is price of exports for South Africa
xr is the exchange rate for South Africa
dum is a dummy variable (dum = 0 before the 2008 financial crisis and dum = 1 after the 2008 financial crisis

1. Estimate a regression with the quantity of exports as the dependent variable, price of exports, the exchange rate, a differential intercept dummy (dum), a multiplicative dummy (dum interacted with the exchange rate) and another multiplicative dummy (dum interacted with the price of exports). Specify your regression first before reporting the estimated regression equation. (3)

$Q\_X = b_1 + b_2*PL\_X + b_3*XR + b_4\ dum + b_5\ dum*XR + b_6\ dum*PL\_X + U_1$

$Q\_X = 245252.0 + 510212.7*PL\_X + 71318.54*XR - 181703.7*DUM - 41960.35*DUM*XR + 909254.6*DUM*PL\_X$

2. Give the interpretation of the differential intercept dummy (dum) coefficient. (2)

During the period after the 2008 financial crisis, the quantity of exports increased (alternatively, the intercept decreased by 181,703.7units) holding other variables constant.

3. Give the interpretation of the differential slope dummy (dum*xr) coefficient. (2)

During the period after the 2008 financial crisis, the differential slope with exchange rate decreased by 41,960.35 units, holding other variables constant.

4. Give the interpretation of the differential slope dummy (dum*pl_x) coefficient. (2)

During the period after the 2008 financial crisis, the differential slope with price of exports increased by 909,254.6 units, holding other variables constant.

5. Estimate a regression with the quantity of exports as the dependent variable, price of exports, the exchange rate, a differential intercept dummy (dum) and a multiplicative dummy (dum interacted with the exchange rate only). Write down the two regression equations – the one before the 2008 financial crisis and the one after the 2008 financial crisis. (4)

Before the crisis:

Q_X = 514850.5*PL_X + 71046.59*XR + 244597.6

After the crisis:

Q_X = 514850.52*PL_X + 27422,83*XR + 678406,91

6. Use auxiliary regressions to detect multicollinearity between the exchange rate, the price of exports and dum. Give your auxiliary regressions and report the adjusted R^2 statistic of the auxiliary regressions and conclude. (6)

Auxiliary regressions

XR = 9.32*PL_X + 4.82*DUM - 0.23 Adjusted R^2= 0.77

PL_X= 0.05*XR - 0.01*DUM + 0.17 Adjusted R^2= 0.64

DUM = -0.03*PL_X + 0.07XR - 0.10 Adjusted R^2= 0.60

Conclusion: Since the adjusted R^2 value of the main regression of 0.93 is much higher than the adjusted R^2 of the auxiliary regressions, we conclude that there is no multicollinearity among the explanatory variables.

7. Use variance inflation factors to detect multicollinearity among the exchange rate, the price of exports and dum. Report your regression table and conclude. (6)

Variance Inflation Factors
Sample: 1950 2019
Included observations: 70

Variable	Coefficient Variance	Uncentered VIF	Centered VIF

DUM	2.73E+09	3.086454	2.557348
PL_X	8.47E+09	9.553452	2.883565
XR	42184509	8.548873	4.503046
C	5.65E+08	3.727520	NA

Conclusion: Since the VIFs of the variables are lower than 10, we do not suspect multicollinearity problem.

-----o0o-----