STRATHMORE INSTITUTE OF MATHEMATICAL SCIENCES
BACHELOR OF BUSINESS SCIENCE: ACTUARIAL SCIENCE, FINANCIAL ECONOMICS AND FINANCIAL ENGINEERING

END OF SEMESTER EXAMINATION

BSE 2205: INTERMEDIATE ECONOMETRICS

Date: 6^{th} December, 2022 Time: 2.5 hours

Instructions

1. This examination consists of **Five** questions.
2. Answer **Question One**(Compulsory) and **any other two** questions.

Question 1

(a) Consider the model $Y = X\beta + U$ where $Y = \begin{bmatrix} y_1 & y_2 & y_3 \dots & y_n \end{bmatrix}'$,

$$X = \begin{bmatrix} 1 & x_{21} & x_{31} & \dots & x_{K1} \\ 1 & x_{22} & x_{32} & \dots & x_{K2} \\ 1 & x_{23} & x_{33} & \dots & x_{K3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{2n} & x_{3n} & \dots & x_{Kn} \end{bmatrix}, \beta = \begin{bmatrix} \beta_1 & \beta_2 & \beta_3 & \dots \beta_K \end{bmatrix} \text{ and } \begin{bmatrix} U_1 & U_2 & U_3 \dots & U_n \end{bmatrix}'$$

(i) If U is the residual, derive $\hat{\beta}_{OLS}$ using matrix algebra {3 marks}

(ii) Show that $\hat{\beta}_{OLS}$ is unbiased {3 marks}

(iii) Derive the expression for $var(\hat{\beta}_{OLS})$ {2 marks}

(iv) If $Y = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 \end{bmatrix}'$ and $x_2 = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \end{bmatrix}'$ find $\hat{\beta}_{OLS}$ using the expression derived in 1a(i) above {3 marks}

(v) If $\sigma^2 = \frac{6}{25}$ find $var(\hat{\beta}_{OLS})$ using the expression in 1a(ii) above {3 marks}

(vi) What t-statistic is associated with the slope and intercept parameters given the estimates in 1a(iv) and 1a(v) above? {3 marks}

(b) Consider the model $y_i = \beta_1 + \beta_2 x_{2i} + \epsilon_i$. Required:

(i) State the OLS assumption that would be violated if x_{2i} was endogenous {2 marks}

(ii) What would be the consequences of the endogeneity of x_{2i} ? {4 marks}

(iii) One of the methods of dealing with endogeneity requires that we understand how to derive $\hat{\beta}_{OLS}$ but with the generalized method of moments (GMM). Derive $\hat{\beta}_{GMM}$ for the equation in 1(b) above {5 marks}

(iv) Suggest the solution for endogeneity implied in 1b(iii) above {2 marks}

[30 marks]

Question 2

You are given the following data sampling process $y_i = \beta_1 + \beta_2 x_{2i} + \epsilon_i$ where:

$$\epsilon_i = \sqrt{x_{2i}} * u_i$$
$$u_i \overset{iid}{\sim} N(0,1)$$

and x_{2i} is a non-stochastic positive variable.

(a) Show that this model is heteroskedastic {2 marks}

(b) If the empirical information is $Y = \begin{bmatrix} 4 & 2 & 5 & 7 \end{bmatrix}'$ and $x_{2i} = \begin{bmatrix} 1 & 1 & 4 & 4 \end{bmatrix}'$. Estimate $\hat{\beta}_{OLS}$ {2 marks}

(c) What are the characteristics of $\hat{\beta}_{OLS}$? {2 marks}

(d) Discuss how you would transform the data so that you could remove the heteroskedasticity {2 marks}

(e) Now estimate the model with the empirical information given in section (b), but by GLS. {4 marks}

(f) Show that in this case $var(\hat{\beta}_{GLS}) = \begin{bmatrix} \frac{10}{9} & -\frac{4}{9} \\ -\frac{4}{9} & \frac{5}{18} \end{bmatrix}$ {2 marks}

(g) Supply the robust standard errors that can be used to test the significance of β_1 and β_2 in 2(c) above {2 marks}

(h) The figure below shows edited stata output detailing results for a heteroscedasticity test. Use it to test whether the reference estimation was heteroscedastic. {2 marks}

```
. hettest

Breusch–Pagan/Cook–Weisberg test for heteroskedasticity
Assumption: Normal error terms
Variable: Fitted values of price

H0: Constant variance

    chi2(1) = 105.75
Prob > chi2 = 0.0000
```

(i) Benji ran the following regression $\text{Price}_i = \beta_1 + \beta_2 \text{lotsize} + u_i$. If this regression was heteroscedastic and price and lot-size are positively related, sketch the distribution of price around the line of best fit. {2 marks}

[20 marks]

Question 3

Consider the following data sampling process

$$\begin{aligned} Y_t &= \beta x_t + \epsilon_t \text{ where} \\ \epsilon_t &= 0.6\epsilon_{t-1} + U_t \\ U_t &\overset{iid}{\sim} N(0,1) \end{aligned}$$

You are told that x is exogenous and are also given the following matrices:
$X'X = \begin{bmatrix} 20 & 10 \\ 10 & 10 \end{bmatrix}'$, $X'y = \begin{bmatrix} 86.6 \\ 68.4 \end{bmatrix}'$, $X'\Psi X = \begin{bmatrix} 72.5 & 36.25 \\ 36.25 & 32.55 \end{bmatrix}'$, $X'\Psi^{-1}X = \begin{bmatrix} 5.75 & 2.875 \\ 2.875 & 3.8125 \end{bmatrix}'$
and $X'\Psi^{-1}y = \begin{bmatrix} 25.475 \\ 25.29375 \end{bmatrix}'$ where $\sigma^2\Psi$ is E($\epsilon\epsilon'$)
Required:

(a) Assumme that $\epsilon_t \sim N(\mu_\epsilon, \sigma^2_\epsilon)$ for every t. Show that $\mu_\epsilon = 0$ and var$(\epsilon_t) = \frac{25}{16}$ {3 marks}

(b) What is the shape and dimension of Ψ? (You don't have to write it out in full) {4 marks}

(c) Estimate β_1 and β_2 using OLS {3 marks}

(d) Discuss the characteristics of $\hat{\beta}_{OLS}$ {5 marks}

(e) Estimate the true value of variance-covariance matrix of $\hat{\beta}_{OLS}$ {3 mark}

(f) Test the null hypothesis that $\beta_2 = 0$ using your OLS estimator of β_2 {2 marks}

[20 marks]

Question 4

You are given the following model $y_i = \beta_1 + \beta_2 x_{2i} + \epsilon_i$. The predictor x_{2i} is a categorical variable for marital status. A respondent is either married or not married.

(a) Given the nature of x_{2i} state the dummies that can be constructed from this categorical variable {1 mark}

(b) If $n = 5$ construct matrix X given the dummies in 4(a) above {2 marks}

(c) Using the matrix X, identify two problems that one would experience if the variables in X were to be used as independent variables in a regression {4 marks}

(d) State the three possible models that an investigator who is interested on the effect of marital status on wages would estimate to avoid the problems in 4(c) above {3 marks}

(e) For one of the models with a constant find the conditional expectation of wages$_i$|married$_i$ = 1 and wages$_i$|married$_i$ = 0 {2 marks}

(f) Use the findings 4(e) above to provide the theoretical interpretation of β_2 {1 mark}

(g) After running the three models in 4(d) you obtain the following results. Use the results to answer the following questions:

VARIABLES	Married	Not Married	Both Dummies
Married	1.166***		6.313***
	(0.112)		(0.0775)
Not Married		-1.166***	5.147***
		(0.112)	(0.0812)
Constant	5.147***	6.313***	
	(0.0812)	(0.0775)	
Observations	3,294	3,294	3,294
R-squared	0.032	0.032	0.764
Standard errors in parentheses			
*** p<0.01, ** p<0.05, * p<0.1			

(i) Interpret the parameters in all the three models {3 marks}

(ii) Does marriage boost earnings? {1 mark}

(h) The following results show the variance inflation factor for a given regression. What problem is this researcher experiencing? What solution would you advise this researcher to explore to deal with the problem? {3 marks}

```
. vif

Variable |       VIF       1/VIF
---------+----------------------
     lp1 |     47.35    0.021118
     lp2 |     45.21    0.022118
     lnm |     14.66    0.068208
---------+----------------------
Mean VIF |     35.74
```

[20 marks]

Question 5

Figure 1 shows the estimates for the regression model

$$\text{Home Ownership}_i = \beta_1 + \beta_2 \text{income}_i + \beta_3 \text{level of education}_i + \epsilon_i$$

Figure 1: Influence of income and level of education on home ownership

Model		**Probit**	**Logit**	**LPM**
Goodness of fit	Log likelihood	-20.532	-20.460	-
	LR test chi(2)	13.99***	14.13***	-
	F(2, 37)			8.62***
	Pseudo R square	0.254	0.257	
	R square	-	-	0.3179
Income	**Coefficient**	0.195**	0.328**	0.067**
	Z statistic	2.02	2.03	2.32
	P value	0.043	0.042	0.026
	Marginal effects	0.057**	0.055**	-
	Z statistic	2.39	2.54	-
	P value	0.017	0.011	-
Education	**Coefficient**	-0.023	-0.014	-0.012
	Z statistic	-0.07	-0.03	-0.11
	P value	0.946	0.980	0.913
	Marginal effects	-0.007	-0.002	-
	Z statistic	-0.07	-0.03	-
	P value	-0.946	0.980	-
Intercept	**Constant**	-2.557***	-4.407***	-0.373
	Z statistic	-3.20	-2.95	-1.60
	P value	0.001	0.003	0.118
Key	***significant at 1%			
	**significant At 5%			

(a) Write the complete expressions for the probability density function (p.d.f.) and cumulative distribution fucntion (c.d.f.) {3 marks}

(b) What is the probability that a family owns a house i.e. $P\{y_i = 1|x_i\}$ and the probability that a family does not own a house i.e. $P\{y_i = 0|x_i\}$ {3 marks}

(c) What are the shorting comings of the LPM model vis a vis the probit and logit model? {4 marks}

(d) Interpret the LPM coefficient of for income (income was measured in thousand Kenya shillings){3 marks}

(e) How does this coefficient compare with the marginal effects of the probit and logit model {3 marks}

(f) What is the effect of education on the probability of owning a home {2 marks}

(g) Should the variable education be dropped from the model {2 marks}

[20 marks]

Be More be Strathmore! ☺