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

MARCH 2025 SPECIAL EXAMINATION

BSE 2205: INTERMEDIATE ECONOMETRICS

DATE: 2nd April 2025 Time: 2 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 {2 marks}

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

(iv) If $Y = \begin{bmatrix} 1 & 2 & 0 & 3 & 2 \end{bmatrix}'$ and $x_2 = \begin{bmatrix} 1 & 2 & 2 & 1 & 2 \end{bmatrix}'$ find $\hat{\beta}_{OLS}$ using the expression derived in 1a(i) above {3 marks}

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

(vi) What t-statistic is associated with the slope and intercept parameters given the estimates in 1a(iv) and 1a(v) above? {2 marks}

(b) Using a sample of 545 full-time workers, a researcher is interested in the question as to whether women are systematically underpaid compared with men. To answer this question the researcher estimates the model $\text{Wages}_i = \beta_1 + \beta_2 \text{Male}_i + \epsilon_i$ where the wages are measured in Kenya shillings per hour and Male_i is a dummy that sets to one for males and zero for females. Required:

(i) The conditional expectation of Wages_i if $\text{Male}_i = 1$ and $\epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$ {1 mark}

(ii) The conditional expectation of Wages_i if $\text{Male}_i = 0$ and $\epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$ {1 mark}

(iii) What is the interpretation of β_2 in this model given your findings in 1 b(i) and 1b(ii) {2 marks}

(iv) The researcher estimates the model and obtains the results in Figure 1:

Figure 1: Hourly wages explained from gender: OLS results

Variable	Estimate	Standard error	*t*-ratio
constant	5.09	0.58	8.78
male	0.82	0.15	5.47

$N = 545 \quad s = 2.17 \quad R^2 = 0.26$

use the findings in 1b(ii) and 1b(iii) above to interpret the intercept and the slope parameter {3 marks}

(v) Some gender activist complain about the researchers ommission of the females in the model $\text{Wages}_i = \beta_1 + \beta_2 \text{Male}_i + \epsilon_i$. To calm the activists the researcher constructs a female dummy and attempts to include it in the model as follows:

$$\text{Wages}_i = \beta_1 + \beta_2 \text{Male}_i + \beta_3 \text{Female}_i + \epsilon_i \tag{1}$$

Required:

(I) Specify the female dummy constructed by the researcher {2 marks}

(II) What estimation issue(s) is this researcher likely to experience in estimating equation 1 {2 marks}

(III) What are the causes of the issue(s) identified in 1b(v)(II) above {2 marks}

(IV) What three ways could the researcher use to deal with the issue(s) in 1b(v)(II) given the causes in 1b(v)(III) {3 marks}

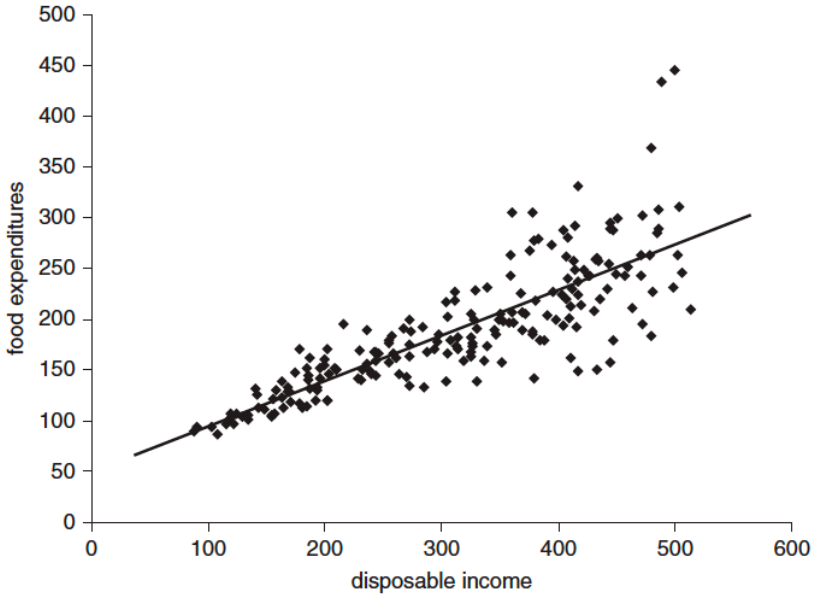
[30 marks]

Question 2

Figure 2 shows the scatter and line of best fit for the regression model

$$\text{Food expenditure}_i = \beta_1 + \beta_2 \text{Disposable income}_i + \epsilon_i$$

Figure 2: Engel Curve



(a) What does the distribution of food expenditures around the line of best fit suggest about the behaviour of the error term and the estimated model {2 marks}

(b) What are the consequences of the behaviour of the error term stated in 2(a) above? {3 marks}

(c) Should one be worried about estimates of β_1 and β_2? {2 marks}

(d) To simulate the solution to the problem identified in 2(a) above, Myra estimates the model $Y = \beta_1 + \beta_2 x_{2i} + \epsilon_i$ where $\epsilon_i = x_{2i} * U_i$ and $U_i \overset{iid}{\sim} N(0,1)$. Show that Myra's model indeed simulates the problem identified in 2(a) above? {3 marks}

(e) Suggest a transformation to Myra that would solve the problem with her model {2 marks}

(f) If Myra's empirical information is $Y = \begin{bmatrix} 4 & 3 & -5 & -6 \end{bmatrix}'$ and $x_{2i} = \begin{bmatrix} 2 & 1 & -1 & -2 \end{bmatrix}'$. Estimate $\hat{\beta}_{OLS}$ {2 marks}

(g) Now estimate the model with the empirical information given in section (f), but by GLS. {2 marks}

(h) Show that in this case $var(\hat{\beta}_{GLS}) = \begin{bmatrix} \frac{2}{5} & 0 \\ 0 & \frac{1}{4} \end{bmatrix}$ {2 marks}

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

[20 marks]

Question 3

Consider the following data sampling process

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

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

(b) Show that this model has autocorrelation. What does the error covariance matrix look like, given that $n = 5$? {4 marks}

(c) Discuss how you would transform the data to remove the autocorrelation {5 marks}

(d) You are given the following empirical information: $Y = \begin{bmatrix} 3 & 4 & 5 & -1 & 4 \end{bmatrix}'$ and $x = \begin{bmatrix} 5 & 5 & 10 & 0 & 10 \end{bmatrix}'$. Estimate $\hat{\beta}_{GLS}$ {5 marks}

(e) What is the use of the estimates 3(d) above? {1 mark}

(f) State the formal test for autocorrelation together with its null and alternative hypothesis {2 marks}

[20 marks]

Question 4

You are given the following model $y_i = \beta_1 + \beta_2 x_{2i} + \epsilon_i$
You suspect, however, that $cov(x, \epsilon) \neq 0$. You think that Z is a valid instrument. You have the following empirical information at your disposal: $Y = \begin{bmatrix} 1 & 4 & 3 & 6 & 0 & 2 \end{bmatrix}'$ and $x_{2i} = \begin{bmatrix} 2 & 5 & 3 & 6 & 2 & 1 \end{bmatrix}'$ and $Y = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 & 0 \end{bmatrix}'$.

(a) Considering that $y_i = \beta_1 + \beta_2 x_{2i} + \epsilon_i$, derive an expression for $\hat{\beta}_{GMM}$ while clearly stating and simplfying each orthogonality condition {5 marks}

(b) Considering the empirical information provided and the fact that $\sigma_i^2 = 1 + 3Z_i$. Estimate the model $y_i = \beta_1 + \beta_2 x_{2i} + \epsilon_i$ using an appropriate GLS estimator assuming that $cov(x, \epsilon) = 0$ {5 marks}

(c) Estimate the model by IV using Z as an instrument for x. {5 marks}

(d) What would be the optimal GMM estimator for the IV model, given that $\sigma_i^2 = 1 + 3Z_i$? Explain {5 marks}

[20 marks]

Question 5

Figure 3 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 3: 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! ☺