

**Econometrics I**  
**Midterm Exam 2**  
**Fall 1996**  
**Total Points: 100**  
**Time: 1 hr. and 15 min.**

**Answer all questions. Note that each question has different weight. Good Luck!!**

1. State whether the following statements are true or false. Briefly justify your answer. (7 points each)

- (a) If heteroskedasticity is present then the conventional  $t$  and  $F$  tests are invalid.
- (b) When autocorrelation is present, the OLS estimators are biased and inconsistent.
- (c) If the regressors ( $X$ ) are random but independent of the error term ( $u$ ) then the OLS estimator of  $\beta$  and  $\sigma^2$  are unbiased.
- (d) In the presence of exact multicollinearity one can only estimate a linear combination of the  $\beta$ 's.

2. A researcher is investigating the determinants of earning and estimates the simple regression

$$Y = \beta_1 + \beta_2 X_2 + u \tag{1}$$

where  $Y$  is earning and  $X_2$  is years of schooling. The true relationship is

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u \tag{2}$$

where  $X_3$  is ability. It is believed that ability and schooling are positively correlated.

- (a) Let the OLS estimators of  $\beta_2$  from (1) and (2) be labeled as  $b$  and  $b^*$ . Find these estimators and show that  $b$  is biased. Calculate the bias of  $b$  explicitly. Can you say something about the nature of the bias? (15 points)
- (b) Is there any problem in testing hypotheses regarding  $\beta_2$  from the model in (1) knowing that it is misspecified? Explain. (8 points)

3. Consider the model

$$Y = X\beta + u$$

where  $E(u) = 0$  and  $E(uu') = \sigma^2\Omega$ . All other standard assumptions are satisfied by the above model. Assume that  $\Omega$  is known. (6 points each)

- (a) Derive the covariance matrix of the OLS and GLS estimators of  $\beta$ .
- (b) Derive the covariance matrix of the OLS residual  $e$ .
- (c) Derive the covariance matrix of the GLS residual  $e_G$ .
- (d) Derive the covariance matrix of  $e$  and  $e_G$ .
- (e) Can you think of a situation where  $\Omega$  can be estimated from the data? Explain.

4. Consider the following two variable regression model

$$Y_t = \beta X_t + u_t \tag{1}$$

$$u_t = \rho u_{t-1} + \epsilon_t \tag{2}$$

where  $|\rho| < 1$  and it is known.  $\epsilon_t$  is i.i.d. normal with zero mean and constant variance.

- (a) Derive the OLS estimator of  $\beta$  after the Cochran-Orcutt transformation. (9 points)
- (b) Derive the OLS estimator of  $\beta$  after the Prais-Winsten transformation. (10 points)

**Econometrics I**  
**Midterm Exam 2**  
**Fall 1998**  
**Total Points: 100**  
**Time: 1 hr. and 20 min.**

**Answer all parts. Note that each part has different weight. Good Luck!!**

1. Let the true regression model be

$$y = \beta_1 x_1 + \beta_2 x_2 + u \tag{1}$$

where  $y, x_1, x_2$  are measured in deviation forms. However, one mistakenly estimates the model

$$y = \beta_1 x_1 + v \tag{2}$$

- (a) Derive the expression for bias in the OLS estimator of  $\beta_1$  obtained from (2). When is it unbiased? **8 points.**
- (b) Interpret the bias term and its significance in hypothesis testing. **5 points.**

2. Consider the following simple regression model

$$Y_{ig} = \beta X_{ig} + \epsilon_{ig}, \tag{1}$$

where  $E(\epsilon_{ig}) = 0$ , and  $Var(\epsilon_{ig}) = \sigma^2$ . The subscript  $i$  refers to  $i$ th individual in the  $g$ th group. The data consists of 5 group means which are to be used to estimate

$$\bar{Y}_g = \beta \bar{X}_g + \bar{\epsilon}_g, \quad g = 1, \dots, 5. \tag{2}$$

<i>Group</i>	1	2	3	4	5	<i>All</i>
$n_g$	10	20	15	30	25	100
$\bar{Y}_g$	2	4	3	1	5	--
$\bar{X}_g$	1	4	2	1	6	--

- (a) Derive algebraic expressions for the OLS and GLS estimators of  $\beta$ . Prove that both OLS and GLS estimators are unbiased. **10 points.**
- (b) Derive algebraic expressions for  $\text{Var}(b_{OLS})$  and  $\text{Var}(b_{GLS})$ . **10 points.**
- (c) Use the above data to compute the variances of OLS and GLS estimators of  $\beta$  assuming  $\sigma^2 = 1$ . Show that the GLS estimator is more efficient. **10 points.**

3. Consider the following regression model

$$Y_t = \beta_0 + \beta_1 X_t + u_t \quad (1)$$

$$u_t = \rho u_{t-1} + \epsilon_t, \quad (2)$$

where  $|\rho| < 1$ , and  $\epsilon_t$  is i.i.d.  $(0, \sigma_\epsilon^2)$ . The following are 4 observations on  $Y_t$ : 100, 210, 360, and 400. The corresponding  $X_t$  values are: 14, 20, 24, and 30.

- (a) Assume  $\rho = 0.5$ . Calculate OLS and GLS estimators of  $\beta_1$ . Do not drop the first observation in deriving the GLS estimator. **10 points.**
- (b) Find the variance of the GLS estimator of  $\beta_1$  assuming  $\rho = 0.5$ . **10 points.**
- (c) Now assume that  $\rho$  is not known. Test the hypothesis that  $\rho = 0$  using the DW test statistic. **7 points.**
- (d) Assuming  $\rho$  unknown, use the Breusch and Godfrey test to test the hypothesis that  $\rho = 0$ . **10 points.**

4. Consider the following regression model

$$Y_t = \beta X_t + u_t \quad (1)$$

where the error term  $u$  is heteroskedastic the form of which is unknown. You are given the following four values of  $X$ : 1, 2, -3, 0. The corresponding values of  $Y$  are: 4, 3, -6, -1.

- (a) Calculate  $b_{OLS}$  and the variance of  $b_{OLS}$ . **10 points.**
- (b) Calculate the heteroskedastic consistent estimator of  $V(b_{OLS})$ . **10 points.**

**Econometrics I**  
**Midterm Exam II**  
**Spring 1998**  
**Total Points: 100**  
**Time: 1 hr. 15 min.**

Answer all questions. Note that each question has different weight. Good Luck!!

1. Five sample observations on  $X$  and  $Y$  are

$$X = \begin{pmatrix} 1 & 4 \\ 1 & 1 \\ 1 & 5 \\ 1 & 8 \\ 1 & 2 \end{pmatrix}, \quad Y = \begin{pmatrix} 6 \\ 3 \\ 12 \\ 15 \\ 4 \end{pmatrix}.$$

- (i) Based on the above data estimate  $\alpha$  and  $\beta$  from the model  $Y_t = \alpha + \beta X_t + u_t$ ,  $t = 1, \dots, 5$ , using the OLS method. Find  $R^2$  of the regression. **10 + 5 points.**
- (ii) Assume that  $u$  is heteroskeastic, the form of which is not known. Find the heteroskedasticity corrected standard errors (White correction) of the OLS estimator of  $\alpha$  and  $\beta$ . **10 points.**
- (iii) Test the presence of heteroskedasticity using the Bruesch-Pagan test at the 5% level of significance. Assume that  $\sigma_t^2 = h(X_t, \gamma)$ , i.e., the heteroskedasticity is explained by the  $X$  variable. (Note: You don't need to know the form of  $h(X_t, \gamma)$ ). **10 points.**
- (iv) Now assume that  $u_t = \rho u_{t-1} + \epsilon_t$ ,  $|\rho| < 1$  and  $\epsilon \sim i.i.d.(0, \sigma_\epsilon^2)$ . The model you would like to estimate is  $Y_t = \alpha + \beta X_t + u_t$ .
  - (a) If  $\rho = .8$  how would you estimate  $\alpha$  and  $\beta$  *efficiently*? Calculate the estimated values of  $\alpha$  and  $\beta$ . **12 points.**
  - (b) If  $\rho$  is *not known* how would you estimate  $\alpha$  and  $\beta$  ? Calculate the estimated values of  $\alpha$  and  $\beta$  using a method of your choice. **10 points.**

2. Consider the standard linear multiple regression model given by

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t, \quad t = 1, \dots, 10. \quad (1)$$

Suppose that in deviation form:

$$x'x = \begin{pmatrix} 22 & 15 \\ 15 & 11 \end{pmatrix}, \quad x'y = \begin{pmatrix} 35 \\ 26 \end{pmatrix}, \quad y'y = 71.$$

- (a) Find the OLS estimator of  $\beta_2$  and  $\beta_3$ . **10 points.**
- (b) Find  $R^2$ . **5 points.**
- (c) It is often stated that multicollinearity is a serious problem if  $R^2 < R_k^2$  where  $R^2$  is the goodness-of-fit measure of the regression equation in (1) and  $R_k^2$  is goodness-of-fit for the auxiliary regression  $X_k$  on an intercept and  $X_2, \dots, X_{k-1}$ . Do you find serious multicollinearity problem, using this rule? **6 points.**

3. Consider the following simple regression model

$$Y_{ig} = \beta X_{ig} + \epsilon_{ig}, \quad (1)$$

where  $E(\epsilon_{ig}) = 0$ , and  $Var(\epsilon_{ig}) = \sigma^2$ . The subscript  $i$  refers to  $i$ th individual in the  $g$ th group. The data consists of 5 group means which are to be used to estimate

$$\bar{Y}_g = \beta \bar{X}_g + \bar{\epsilon}_g, \quad g = 1, \dots, 5. \quad (2)$$

<i>Group</i>	1	2	3	4	5	<i>All</i>
---	---	---	---	---	---	---
$n_g$	10	20	15	30	25	100
$\bar{Y}_g$	2	4	3	1	5	---
$\bar{X}_g$	1	4	2	1	6	---

- (a) Estimate  $\beta$  using both OLS and GLS procedures. **12 points.**
- (b) Compute the variances of OLS and GLS estimators of  $\beta$  assuming  $\sigma^2 = 1$ . Show that the GLS estimator is more efficient. **10 points.**

**Econometrics I**  
**Midterm Exam II**  
**Spring 1999**  
**Total Points: 100**  
**Time: 1 hr. 15 min.**

**Answer all questions. Note that each question has different weight. Good Luck!!**

1. Suppose

$$S = \alpha_1 + \alpha_2 Ed + \alpha_3 IQ + \alpha_4 EX + \alpha_5 Sex + \alpha_6 DF + \alpha_7 DE + u, \quad (1)$$

where S is salary, Ed is years of education, IQ is IQ level, EX is years of experience, Sex is one for males and zero for females, DF is one for French-only speakers and zero otherwise, DE is one for English-only speakers and zero otherwise. Given a sample of N individuals who speak only French, only English or are bilingual:

- (a) Explain how you would test for discrimination against females (in the sense that *ceteris paribus* females earn less than males. **4 points**.
- (b) Explain how you would measure and test the payoff to someone of becoming bilingual given that his or her mother tongue is (i) French, (ii) English. **8 points**.
- (c) Explain how you would test the hypothesis that the two payoffs in part (b) above are equal. **4 points**.
- (d) Explain how you would test the hypothesis that a French-only male earn as much as an English-only female. **4 points**.
- (e) Explain how you would test if the influence of experience is greater for males than for females. **4 points**.

2. Data from two sub-periods with 50 observations each produced the following moment matrices (note that X contains a column of ones and another variable)

$$\text{Sample 1} \quad X'X = \begin{pmatrix} 50 & 300 \\ 300 & 2100 \end{pmatrix}, Y'X = (300 \quad 2000), Y'Y = 2100.$$

$$\text{Sample 2} \quad X'X = \begin{pmatrix} 50 & 300 \\ 300 & 2100 \end{pmatrix}, Y'X = (300 \quad 2200), Y'Y = 2500.$$

- (a) Use OLS to estimate the coefficients (intercept and slope) for both sub-periods. **12 points.**
- (b) Test the hypothesis that the coefficients (intercept and slope) in both sub-periods are the same ( $\beta^1 = \beta^2$ ), assuming that variances are the same for both sub-periods ( $\sigma_1^2 = \sigma_2^2$ ). **10 points.**
- (c) Test the hypothesis that the variances are the same for both sub-periods ( $\sigma_1^2 = \sigma_2^2$ ), assuming that the coefficients (intercept and slope) in both sub-periods are different. **10 points.**
- (d) Assume that the coefficients (intercept and slope) in both sub-periods are the same. Estimate  $\sigma_1^2$  and  $\sigma_2^2$  using the residuals from the restricted model in part (b). **10 points.**
- (e) Find the estimated GLS estimator of the intercept and slope coefficients, using the estimated values of  $\sigma_1^2$  and  $\sigma_2^2$  from part (d). **10 points.**



3. Consider the following simple regression model

$$Y_i = \beta_1 + \beta_2 X + u, \quad (1)$$

where  $u \sim i.i.d.(0, \sigma^2)$ .

- (a) Find the OLS estimator of  $\beta_2$  and prove that it is consistent, assuming  $plim(\frac{1}{n} \sum x^2) = \sigma_x^2 = \text{constant}$ . **No matrix formulation please. 8 points.**
- (b) Find the OLS estimator of  $\beta_1$  and prove that it is consistent, assuming  $plim(\frac{1}{n} \sum x) = \mu_x = \text{constant}$ . **No matrix formulation please. 8 points.**
- (c) If  $E(u_i^2) = \sigma_i^2$ , where  $\sigma_i^2, i = 1, \dots, n$  are unknown, how would you estimate  $\beta_1$  and  $\beta_2$ ? How would you test hypothesis on  $\beta_2$ ? Explain. **6 points.**

**Econometrics I**  
**Midterm Exam 2**  
**Fall 1999**  
**Total Points: 100**  
**Time: 1 hr. and 20 min.**

**Answer all parts. Note that each part has different weight. Good Luck!!**

1. Consider the following regression model with a scalar ( $X$ )

$$Y = \beta X + u, \quad u \sim (0, \sigma^2 X^2). \quad (1)$$

- (a) **(10 points)** Find  $b$  and  $b_G$ , the OLS and GLS estimators of  $\beta$ , and show that the GLS estimator is more efficient ( $V(b) > V(b_G)$ ).
- (b) **(8 points)** Consider another estimator of  $\beta$ , viz.,  $\tilde{b} = \bar{Y}/\bar{X}$ . Show that  $V(\tilde{b}) > V(b_G)$ .
- (c) **(12 points)** You are given the following 4 observations on  $Y = (1, 2, -3, 0)$  and  $X = (4, 3, -6, -1)$ . Find the OLS and GLS estimators of  $\beta$ . Compute the estimated variance of  $b_G$ ,  $\widehat{V}(b_G)$ .

**For the rest of the problems assume that the form of heteroskedasticity is NOT known.**

- (d) **(10 points)** You are given the following 4 observations on  $Y = (1, 2, -3, 0)$  and  $X = (4, 3, -6, -1)$ . Compute the heteroskedasticity-consistent estimator of  $V(b)$ .
- (e) **(15 points)** If the form of heteroskedasticity is given by  $h(X) = (\gamma_0 + \gamma_1 X)^2$ , how would you estimate  $\beta$  taking heteroskedasticity into account (assuming  $\gamma_0$  and  $\gamma_1$  unknown)? Use the data in part (d) and compute the GLS estimator of  $\beta$ .

2. The table below has the estimated coefficients and associated statistics for the demand for cigarettes in Turkey. The data are annual for the years 1960-1988 (29 observations). The definitions of the variables are as follows:

- $LQ$  = Logarithm of cigarette consumption per adult  
(dependent variable)
- $LY$  = Logarithm of per-capita real GNP in 1968 prices  
(in Turkish liras)
- $LP$  = Logarithm of real price of cigarettes in Turkish liras per kg.
- $D82$  = 1 for 1982 onward, 0 prior to that
- $D86$  = 1 for 1986 onward, 0 prior to that
- $LYD1$  =  $LY$  multiplied by  $D82$
- $LYD2$  =  $LY$  multiplied by  $D86$
- $LPD1$  =  $LP$  multiplied by  $D82$
- $LPD2$  =  $LP$  multiplied by  $D86$

	Variable	Model A	Model B	Model C	Model D
$\hat{\beta}_1$	constant	-4.997	-4.800	-4.186	-4.997
$\hat{\beta}_2$	D82	23.364	-0.108	-0.103	21.793
$\hat{\beta}_3$	D86	-36.259	-0.406	-0.103	-28.291
$\hat{\beta}_4$	LY	0.732	0.705	0.621	0.732
$\hat{\beta}_5$	LYD1	-2.798			-2.602
$\hat{\beta}_6$	LYD2	4.251			3.298
$\hat{\beta}_7$	LP	-0.371	-0.337	-0.201	-0.371
$\hat{\beta}_8$	LPD1	0.405	0.061		0.288
$\hat{\beta}_9$	LPD2	-0.236	0.288		
$\overline{R^2}$		0.921	0.859	0.852	0.921
$RSS$		0.018645	0.036444	0.04195	0.019427

(i) **(10 points)** Write down the estimated relationship for 1960-81, 1982-85, 1986-88, and calculate the income and price elasticities for each period.

**Note: Income elasticity**  $= \frac{\partial LQ}{\partial LY}$  **and price elasticity**  $= \frac{\partial LQ}{\partial LP}$ .

(ii) **(20 points)** Assume that Model A is the most general model. What sort of hypotheses can you test in Models B, C, and D? Consider all meaningful combinations of models. Explain these hypotheses in terms of income and price elasticities, and perform the tests. State any assumption you need to make.

(iii) **(15 points)** Now assume that the random term in the regression is  $N(0, \sigma^2)$ . Test the hypothesis that income and price elasticities are the same for all periods (i.e.,  $\beta_5 = \beta_6 = \beta_8 = \beta_9 = 0$ ) using the LR, W, and LM tests.

**Econometrics I**  
**Midterm Exam 2**  
**Spring 2000**  
**Total Points: 100**  
**Time: 1 hr. and 20 min.**

**Answer all parts. Note that each part has different weight. Good Luck!!**

1. Two research workers, working independently, considered the same regression model

$$Y = \alpha + \beta X + u, \quad (1)$$

which satisfied all classical assumptions. Some of the results for their samples (which were independently drawn) were:

Sample I:

$$X'X = \begin{pmatrix} 20 & 100 \\ 100 & 600 \end{pmatrix}, \sum Y = 500, b_I = 2.$$

Sample II:

$$X'X = \begin{pmatrix} 20 & 200 \\ 200 & 2400 \end{pmatrix}, \sum Y = 700, b_{II} = 2.5,$$

where  $b_I$  and  $b_{II}$  are the OLS estimator of  $\beta$  in Sample I and II, respectively.

When the researchers found out about each other's work, they decided to pool the data and present one pooled estimate of  $\beta$ .

- (a) **(10 points)** Find the pooled estimate of  $\beta$  (call it  $b_p$ ).
- (b) **(7 points)** Calculate the variance of  $b_p$  **assuming**  $\sigma^2$  – **the variance of  $u$  known.**
- (c) **(7 points)** Researcher I suggested the use of  $\tilde{b} = \frac{1}{2}(b_I + b_{II})$ . Calculate the variance of  $\tilde{b}$ . Assume that  $\sigma^2$  is known.
- (d) **(2 points)** Is  $b_p$  more efficient than  $\tilde{b}$  (i.e.,  $Var(b_p) < Var(\tilde{b})$ )?

2. Let the true model be

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u \quad (2)$$

which satisfies all the assumptions of the classical linear regression model.

- (a.1) (**8 points**) Show that  $\tilde{b}_2 = b_2 + b_3 b_{32}$  where  $\tilde{b}_2$  is the OLS estimate of the coefficient of  $X_2$  in the regression  $Y$  on  $X_2$ , and  $b_2, b_3$  are the OLS estimators of  $\beta_2, \beta_3$  from (2),  $b_{32}(= \sum x_2 x_3 / \sum x_3^2)$  is the OLS estimate of the coefficient of  $X_2$  in the regression  $X_3$  on  $X_2$ . (**Hint:** Start from the second normal equation, and then express it in deviation form using the first normal equation).
- (a.2) (**3 points**) The above result relates simple and multiple regression coefficients. Can you generalize this result to case of  $K$  regressors when one  $X$  variable (for example  $X_K$ ) is omitted from the estimated model?
- (b) (**5 points**) If  $X_3$  is omitted from the estimated model, show that  $E(\tilde{b}_2) = \beta_2 + \beta_3 b_{32}$  where  $\tilde{b}_2$  is the coefficient of  $X_2$  in the simple regression  $Y$  on  $X_2$ , and  $b_{32} = \sum x_2 x_3 / \sum x_3^2$  is the simple regression coefficient of  $X_3$  on  $X_2$ . (**Hint:** You can use the result from part (a) to minimize time).

3. Suppose you have data on wages, education, experience, and gender. You also have information on marijuana usage. The data on marijuana use was a response from the question: **On how many separate occasions last month did you smoke marijuana?**

- (a) (**4 points**) Write an equation that would allow you to estimate the effects of marijuana use on wages, while controlling for other factors. You should be able to make statements such as, *Smoking marijuana five more times per month is estimated to change wage by  $x\%$ .*
- (b) (**6 points**) Write a model that would allow you to test whether drug usage has different effects on wages for male and female. How would you test that there are no differences in the effects of drug usage for men and women? Explain.
- (c) (**8 points**) Suppose that you think it is better to measure marijuana usage by putting people in one of four categories: nonuser, light user (1 to 5 times per month), moderate user (6 to 10 times a month), and heavy user (more than 10 times per month). Now write a model that allows you to estimate the effects of marijuana usage on wage.
- (d) (**8 points**) Using the model in part (c), explain in detail how to test the null hypothesis that marijuana usage has no specific effect on wage.

(e) (**6 points**) Using the model in part (b), explain how to model and test the hypothesis that the effect of drug use on wage is *structurally different* for men and women. Be specific.

4. In the regression  $Y = \beta X + u$ , the regressor  $X$  is a scalar, and  $u \sim i.i.d(0, \sigma^2)$ . Consider the following three estimators of  $\beta$ :

$$\begin{aligned}\hat{\beta}_1 &= \frac{\bar{Y}}{\bar{X}} \\ \hat{\beta}_2 &= \frac{\sum XY}{\sum X^2} \\ \hat{\beta}_3 &= \frac{\sum(X - \bar{X})(Y - \bar{Y})}{\sum(X - \bar{X})^2}.\end{aligned}\tag{3}$$

(Assume that  $plim(\bar{X}) = \mu$ , and  $\sum X'u/\sqrt{n}$  converges in distribution to  $N(0, \sigma^2 q)$ , where  $q = plim(\sum X^2/n)$ . Note that if  $X = 1$ , then  $q = 1$ ).

- (a) (**18 points**) Show that all the above estimators are consistent. (**Hint:** use convergence in quadratic mean).
- (b) (**8 points**) Derive the limiting distributions of  $\sqrt{n}(\hat{\beta}_1 - \beta)$  and  $\sqrt{n}(\hat{\beta}_2 - \beta)$ .

**Econometrics I**  
**Midterm Exam II**  
**Fall 2000**  
**Total Points: 100**  
**Time: 1 hr. 15 min.**

**Answer all questions. Note that each question has different weight. Good Luck!!**

1. Suppose

$$S = \alpha_1 + \alpha_2 Ed + \alpha_3 IQ + \alpha_4 EX + \alpha_5 D_M + \alpha_6 D_F + \alpha_7 D_E + \alpha_8 D_M * D_F + \alpha_9 D_M * D_E + u, \quad (1)$$

where  $S$  is salary,  $Ed$  is years of education,  $IQ$  is IQ level,  $EX$  is years of experience,  $D_M$  is one for males and zero for females,  $D_F$  is one for French-only speakers and zero otherwise,  $D_E$  is one for English-only speakers and zero otherwise. Given a sample of  $N$  individuals who speak only French, only English or are bilingual:

- (a) Explain how you would measure and test the payoff to a male person becoming bilingual given that his mother tongue is (i) French, (ii) English. **5 points.**
- (b) Explain how you would measure and test the payoff to a female person becoming bilingual given that her mother tongue is (i) French, (ii) English. **5 points.**
- (c) Explain how you would test the hypothesis that the payoff to a male person becoming bilingual and a female becoming bilingual will be the same. **5 points.**
- (d) Explain how you would interpret the coefficients  $\alpha_8$  and  $\alpha_9$ . **5 points.**
- (e) Explain how you would test the hypothesis that a French-only male earn as much as an English-only female. **5 points.**
- (f) Based on this model, can you test whether the influence of experience is greater for males than for females? If not, how would you extend the model? Explain. **5 points.**



2. Data from two sub-periods with 5 observations each produced the following moment matrices (note that  $X$  contains a column of ones and another variable)

$$\text{Sample 1} \quad X'X = \begin{pmatrix} 5 & 30 \\ 30 & 210 \end{pmatrix}, Y'X = (30 \quad 200), Y'Y = 210.$$

$$\text{Sample 2} \quad X'X = \begin{pmatrix} 5 & 30 \\ 30 & 210 \end{pmatrix}, Y'X = (30 \quad 220), Y'Y = 250.$$

- (a) Use OLS to estimate the coefficients (intercept and slope) for both sub-periods. **12 points.**
- (b) Test the hypothesis that the coefficients (intercept and slope) in both sub-periods are the same ( $\beta^1 = \beta^2$ ), assuming that variances are the same for both sub-periods ( $\sigma_1^2 = \sigma_2^2$ ). **16 points.**
- (c) Use the LR test to test the hypothesis that the variances are the same for both sub-periods ( $\sigma_1^2 = \sigma_2^2$ ). Assume that the coefficients (intercept and slope) in both sub-periods are different. **6 points.**
- (d) Assume that the coefficients (intercept and slope) in both sub-periods are the same. Estimate  $\sigma_1^2$  and  $\sigma_2^2$  using the residuals from the restricted model in part (b). Use them to find the estimated GLS estimator of the intercept and slope coefficients. **8 points.**

3. Four sample observations on  $X$  and  $Y$  are

$$X = \begin{pmatrix} 1 & 4 \\ 1 & 1 \\ 1 & 5 \\ 1 & 2 \end{pmatrix}, Y = \begin{pmatrix} 6 \\ 3 \\ 12 \\ 4 \end{pmatrix}.$$

- (a) Based on the above data estimate  $\beta_1$  and  $\beta_2$  from the model  $Y_t = \beta_1 + \beta_2 X_t + u_t, t = 1, \dots, 4$ , using OLS technique. Find the standard errors of  $b_1$  and  $b_2$ . Assume that  $u_t \sim (0, \sigma^2)$  and  $\sigma^2$  is unknown. **5 points.**
- (b) Assume that  $\text{var}(u_t) = \sigma_t^2 = \sigma^2 X_t^2$ . Find the feasible GLS estimator of  $\beta_1$  and  $\beta_2$  and their standard errors. **12 points.**
- (c) Do you find your GLS estimators more efficient? Explain. **3 points.**
- (d) Find the heteroscedasticity-corrected standard errors of the OLS estimators of  $\beta_1$  and  $\beta_2$  assuming that the form of heteroscedasticity is *unknown*. **8 points.**