QUEEN'S UNIVERSITY AT KINGSTON
Department of Economics

ECONOMICS 351* - Section A

Introductory Econometrics

Fall Term 1998 **FINAL EXAMINATION** M.G. Abbott

DATE: **Friday December 18, 1998.**

TIME: **Three (3) hours (180 minutes); 2:00 p.m. - 5:00 p.m.**

INSTRUCTIONS: The examination is divided into two parts.

PART A contains three questions; students are required to answer **ANY TWO** of the three questions 1, 2, and 3 in Part A.

PART B contains two questions; students are required to answer **BOTH** of the two questions 4 and 5 in Part B.

Answer all questions in the exam booklets provided. Be sure your ***name*** and ***student number*** are printed clearly on the front of all exam booklets used.

Do not write answers to questions on the front page of the first exam booklet.

Please label clearly each of your answers in the exam booklets with the appropriate number and letter.

Please write legibly. GOOD LUCK! Happy Holidays.

MARKING: The marks for each question are indicated in parentheses immediately above each question. **Total marks** for the exam **equal 200**.

PART A: Questions 1, 2, and 3 (40 marks for each question) .... **80 marks**
Answer ***any two*** of Questions 1, 2, and 3.

PART B: Question 4 (50 marks) and Question 5 (70 marks) .... **120 marks**
Answer ***all parts of both*** Questions 4 and 5.

TOTAL MARKS .... **200 marks**

PART A (80 marks)

Instructions: Answer **ANY TWO (2)** of the three questions 1, 2, and 3 in this part. Total marks for each question equal 40; marks for each part are given in parentheses.

(40 marks)

1. Consider the simple (two-variable) linear regression model for which the population regression equation can be written in conventional notation as:

$$Y_i = \beta_1 + \beta_2 X_i + u_i \qquad \textbf{(1)}$$

where Y_i and X_i are observable variables, β_1 and β_2 and unknown (constant) regression coefficients, and u_i is an unobservable random error term. The Ordinary Least Squares (OLS) sample regression equation corresponding to regression equation (1) is

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i \qquad (i = 1, ..., N) \qquad \textbf{(2)}$$

where $\hat{\beta}_1$ is the OLS estimator of the intercept coefficient β_1, $\hat{\beta}_2$ is the OLS estimator of the slope coefficient β_2, $\hat{u}_i$ is the OLS residual for the i-th sample observation, and N is sample size (the number of observations in the sample).

(15 marks)

(a) Stating explicitly all required assumptions, derive the expression (or formula) for $\text{Var}(\hat{\beta}_2)$, the variance of the OLS slope coefficient estimator $\hat{\beta}_2$. How do you compute an unbiased estimator of $\text{Var}(\hat{\beta}_2)$?

(15 marks)

(b) Give a general definition of a t-statistic. Starting from this definition, derive the t-statistic for $\hat{\beta}_2$ in OLS sample regression equation (2). State the assumptions required for the derivation.

(10 marks)

(c) Explain how you would use the results for OLS sample regression (2) to compute a test of the null hypothesis H_0: $\beta_2 \leq 0$ against the alternative hypothesis H_1: $\beta_2 > 0$ at significance level α. Show how you calculate the required test statistic, and state its null distribution. Define the p-value for the calculated test statistic, and state how you would use the p-value approach to decide between rejection and non-rejection of the null hypothesis at the chosen significance level α.

(40 marks)

2. Consider the multiple linear regression model for which the population regression equation can be written in conventional notation as:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \qquad \textbf{(1)}$$

where Y_i, X_{2i}, and X_{3i} are observable variables; β_1, β_2, and β_3 are unknown (constant) regression coefficients; and u_i is an unobservable random error term. The Ordinary Least Squares (OLS) sample regression equation corresponding to regression equation (1) is

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{u}_i \qquad (i = 1, ..., N) \qquad \textbf{(2)}$$

where $\hat{\beta}_1$ is the OLS estimator of the intercept coefficient β_1, $\hat{\beta}_2$ is the OLS estimator of the slope coefficient β_2, $\hat{\beta}_3$ is the OLS estimator of the slope coefficient β_3, $\hat{u}_i$ is the OLS residual for the i-th sample observation, and N is sample size (the number of observations in the sample).

(15 marks)

(a) Derive the OLS normal equations for regression equation (1) from the OLS estimation criterion.

(15 marks)

(b) Explain how, using only the results from OLS estimation of regression equation (1), you would compute an F-test of the following hypothesis:

H_0: $\beta_2 = 0$ *and* $\beta_3 = 0$

H_1: $\beta_2 \neq 0$ *and/or* $\beta_3 \neq 0$.

Explain in words the meaning of the null hypothesis H_0 and the alternative hypothesis H_1. Give *two* alternative formulas for the required F-statistic, defining all terms they contain. What is the null distribution of the F-statistic? Define the p-value for the calculated F-statistic, and state how you would use the p-value approach to decide between rejection and non-rejection of the null hypothesis at some chosen significance level α.

(10 marks)

(c) Explain how you would change the specification of regression equation (1) to allow the marginal effect on Y_i of X_{2i} to depend on the value of X_{3i} and the marginal effect on Y_i of X_{3i} to depend on the value of X_{2i}. Write the re-specified equation you would estimate. Write the expression for the marginal effect on Y_i of X_{2i}, and the expression for the marginal effect on Y_i of X_{3i}, implied by your re-specified model. How would you test the original model given by regression equation (1) against your re-specified model?

(40 marks)

3. You are conducting an empirical investigation of the monthly wages of male and female employees of a large corporation. You are given for this project a random sample of observations on N employees of the corporation, N_f of whom are females and N_m of whom are males. The sample data provide observations on the observable variables Y_i, S_i, T_i, and M_i. The first three variables are defined as follows:

Y_i = the monthly wage rate of employee i, measured in dollars per month;

S_i = the number of years of formal schooling completed by employee i;

T_i = the number of years employee i has worked for the corporation.

The variable M_i is a male indicator (or dummy) variable defined as follows:

$$M_i = 1 \quad \text{if employee i is male}$$
$$= 0 \quad \text{if employee i is female.}$$

The wage regression equation for ***male*** **employees** is:

$$Y_i = \beta_1 + \beta_2 S_i + \beta_3 T_i + \beta_4 T_i^2 + u_{mi} \qquad u_{mi} \sim N(0, \sigma_m^2) \qquad \textbf{(1)}$$

where β_1, β_2, β_3 and β_4 are the male regression coefficients and u_{mi} is an unobservable random error term. The wage regression equation for ***female*** **employees** is:

$$Y_i = \alpha_1 + \alpha_2 S_i + \alpha_3 T_i + \alpha_4 T_i^2 + u_{fi} \qquad u_{fi} \sim N(0, \sigma_f^2) \qquad \textbf{(2)}$$

where α_1, α_2, α_3 and α_4 are the female regression coefficients and u_{fi} is an unobservable random error term.

(20 marks)

(a) Explain fully how you would use the male indicator variable M_i to test the following hypothesis:

H_0: $\alpha_j = \beta_j \quad \forall\ j = 1, 2, 3, 4$

H_1: $\alpha_j \neq \beta_j \quad j = 1, 2, 3, 4.$

Explain in words the meaning of the null hypothesis H_0 and the alternative hypothesis H_1. Write the regression equation you use to perform the test and interpret its regression coefficients. Explain how you calculate the required test statistic. State the decision rule you use to decide between rejection and non-rejection of the null hypothesis.

3. (continued)

(10 marks)

(b) Use your answer to part (a) of this question to explain how you would test the following hypothesis:

H_0: $\alpha_3 = \beta_3$ *and* $\alpha_4 = \beta_4$

H_1: $\alpha_3 \neq \beta_3$ *and/or* $\alpha_4 \neq \beta_4$.

Explain in words the meaning of the null hypothesis H_0. Write the *restricted* regression equation implied by the null hypothesis H_0. Explain how you calculate the required test statistic, and state the decision rule you use to decide between rejection and non-rejection of the null hypothesis.

(10 marks)

(c) Use your answer to part (a) of this question to explain how you would test the following hypothesis:

H_0: $\alpha_2 = \beta_2$

H_1: $\alpha_2 \neq \beta_2$.

Explain in words the meaning of the null hypothesis H_0. Write the *restricted* regression equation implied by the null hypothesis H_0. Explain how you calculate the required test statistic, and state the decision rule you use to decide between rejection and non-rejection of the null hypothesis.

PART B (120 marks)

Instructions: Answer all parts of **BOTH** questions 4 and 5 in this part. Question 4 is worth a total of 50 marks, and question 5 is worth a total of 70 marks. Marks for each part are given in parentheses. **Show explicitly all formulas and calculations**.

Both questions 4 and 5 in Part B pertain to an econometric investigation of the demand for bus travel and its determinants. The sample data consist of N = 40 observations on the following variables for a sample of 40 cities.

$BUS_i \equiv$ bus travel in city i, measured in thousands of passenger hours per year;
$INC_i \equiv$ average income per capita in city i, measured in dollars per person per year;
$POP_i \equiv$ population of city i, measured in thousands of persons;
$DEN_i \equiv$ population density in city i, measured in persons per square mile;
$FARE_i \equiv$ bus fare in city i, measured in dollars;
$GASP_i \equiv$ average gasoline price in city i, measured in dollars per gallon.

(50 marks)

4. Using the sample data described above, you begin the investigation by estimating the following LOG-LOG (double-log) regression equation:

$$\ln BUS_i = \beta_1 + \beta_2 \ln INC_i + \beta_3 \ln POP_i + \beta_4 \ln DEN_i + \beta_5 \ln FARE_i + \beta_6 \ln GASP_i + u_{1i} \quad \textbf{(1)}$$

where the $\beta_j\ (j = 1, 2, \ldots, 6)$ are regression coefficients, $\ln X_i$ denotes the natural logarithm of the variable X_i, and u_{1i} is a random error term. OLS estimation of regression equation (1) yields the following OLS coefficient estimates (with estimated standard errors in parentheses below the coefficient estimates) and summary statistics:

OLS Estimates of Equation (1)

	Constant (β_1)	$\ln INC_i$ (β_2)	$\ln POP_i$ (β_3)	$\ln DEN_i$ (β_4)	$\ln FARE_i$ (β_5)	$\ln GASP_i$ (β_6)
$\hat{\beta}_j$ $\left(\hat{se}(\hat{\beta}_j)\right)$	39.15 (9.332)	-4.850 (1.033)	0.8768 (0.1669)	1.074 (0.2348)	0.4282 (0.3887)	-1.767 (2.459)

Summary Statistics -- OLS Estimates of Equation (1):

$$RSS_{(1)} = \sum_{i=1}^{N} \hat{u}_{1i}^2 = 17.9744; \qquad TSS_{(1)} = \sum_{i=1}^{N} \left(\ln BUS_i - \overline{\ln BUS}\right)^2 = 52.2564; \qquad N = 40$$

where $RSS_{(1)}$ is the Residual Sum-of-Squares and $TSS_{(1)}$ is the Total Sum-of-Squares for OLS estimation of regression equation (1).

4. (continued)

(10 marks)

(a) Interpret the slope coefficients in regression equation (1); that is, explain what they mean. Illustrate your interpretation by explaining the numerical value of $\hat{\beta}_2$, the coefficient estimate for $\ln \text{INC}_i$ in regression equation (1).

(10 marks)

(b) Use the above results from OLS estimation of regression equation (1) to test the *individual* significance of the slope coefficient estimates for $\ln \text{INC}_i$ and $\ln \text{FARE}_i$. For each test, state the null and alternative hypotheses, show how you calculate the required test statistic, and state its null distribution. Which of these two slope coefficient estimates are individually significant at the 5 percent significance level? Which of these two slope coefficient estimates are individually significant at the 1 percent significance level?

(10 marks)

(c) Use the results from OLS estimation of regression equation (1) to test the *joint* significance of the slope coefficient estimates in equation (1) at the 1 percent significance level (i.e., for significance level $\alpha = 0.01$). State the null and alternative hypotheses, show how you calculate the required test statistic, and state its null distribution. State the decision rule you use, and the inference you would draw from the test.

(10 marks)

(d) Use the results from OLS estimation of regression equation (1) to compute the two-sided 95 percent confidence interval for the slope coefficient β_3 of the regressor $\ln \text{POP}_i$. Explain how the confidence interval for β_3 is interpreted.

(10 marks)

(e) Use the results from OLS estimation of regression equation (1) to test the null hypothesis H_0: $\beta_4 \le 0$ against the alternative hypothesis H_1: $\beta_4 > 0$ at the 1 percent significance level (i.e., for significance level $\alpha = 0.01$). Show how you calculate the required test statistic, and state its null distribution. State the decision rule you use, and the inference you would draw from the test. What would you conclude from the results of the test?

(70 marks)

5. In light of your analysis of regression equation (1) in question 4, you decide to specify and estimate the regression equation

$$\ln BUS_i = \beta_1 + \beta_2 \ln INC_i + \beta_3 \ln POP_i + \beta_4 \ln DEN_i + u_{2i} \qquad \textbf{(2)}$$

where the $\beta_j\ (j = 1, 2, \ldots, 4)$ are regression coefficients and u_{2i} is a random error term.
OLS estimation of regression equation (2) on the sample of N = 40 observations yields the following OLS coefficient estimates (with estimated standard errors in parentheses below the coefficient estimates) and summary statistics:

OLS Estimates of Equation (2)

	Constant (β_1)	$\ln INC_i$ (β_2)	$\ln POP_i$ (β_3)	$\ln DEN_i$ (β_4)	$\ln FARE_i$ (β_5)	$\ln GASP_i$ (β_6)
$\hat{\beta}_j$ $\left(\hat{se}(\hat{\beta}_j)\right)$	39.29 (9.265)	-4.740 (1.018)	0.8555 (0.1627)	0.9607 (0.2030)	---	---

Summary Statistics -- OLS Estimates of Equation (2):

$$RSS_{(2)} = \sum_{i=1}^{N} \hat{u}_{2i}^2 = 18.7636\,; \qquad TSS_{(2)} = \sum_{i=1}^{N} \left(\ln BUS_i - \overline{\ln BUS}\right)^2 = 52.2564\,; \qquad N = 40$$

where $RSS_{(2)}$ is the Residual Sum-of-Squares and $TSS_{(2)}$ is the Total Sum-of-Squares for OLS estimation of regression equation (2).

The corresponding summary statistics from OLS estimation of regression equation (1) in question 4 are repeated here for ease of reference:

$$RSS_{(1)} = \sum_{i=1}^{N} \hat{u}_{1i}^2 = 17.9744\,; \qquad TSS_{(1)} = \sum_{i=1}^{N} \left(\ln BUS_i - \overline{\ln BUS}\right)^2 = 52.2564\,; \qquad N = 40$$

where $RSS_{(1)}$ is the Residual Sum-of-Squares and $TSS_{(1)}$ is the Total Sum-of-Squares for OLS estimation of regression equation (1) in question 4.

(10 marks)

(a) Use the above information to compute the value of R^2 obtained from OLS estimation of regression equation (1), and the value of R^2 obtained from OLS estimation of regression equation (2). Can these values of R^2 be used to determine which of the sample regression equations provides the better fit to the sample data? Explain why or why not.

5. (continued)

(10 marks)

(b) Identify an alternative goodness-of-fit measure to the R^2 that could be used to compare the goodness-of-fit of the sample regression equations corresponding to equations (1) and (2). Use the above information to compute the values of this alternative goodness-of-fit measure for the OLS estimates of regression equations (1) and (2). Which of the two OLS sample regression equations provides the better fit to the sample data?

(10 marks)

(c) State the coefficient restrictions that regression equation (2) imposes on regression equation (1). Use the results given above from OLS estimation of regression equation (2), together with the relevant results from OLS estimation of equation (1), to perform a test of these coefficient restrictions at the 5 percent significance level (i.e., for significance level $\alpha = 0.05$). State the null and alternative hypotheses, show how you calculate the required test statistic, and state its null distribution. State the decision rule you use, and the inference you would draw from the test.

(10 marks)

(d) Compare the statistical properties of the OLS coefficient estimates of regression equation (1) and regression equation (2) under each of the following two assumptions:

(1) that the coefficient restrictions equation (2) imposes on equation (1) are ***true***;
(2) that the coefficient restrictions equation (2) imposes on equation (1) are ***false***.

Given the results of the hypothesis test performed in part (c) of this question, which set of coefficient estimates would you choose, those for equation (1) or those for equation (2)?

(10 marks)

(e) Use the results from OLS estimation of regression equation (2) to test the proposition that the coefficient β_4 of $\ln DEN_i$ is equal to 1. Perform this test at the 5 percent significance level (i.e., for significance level $\alpha = 0.05$). State the null and alternative hypotheses, and explain in words what they mean. Show how you calculate the required test statistic, and state its null distribution. State the decision rule you use, and the inference you would draw from the test.

(10 marks)

(f) Use the results from OLS estimation of regression equation (2) to test the proposition that the coefficient β_2 of $\ln INC_i$ is *less than* -2. Perform this test at the 5 percent significance level (i.e., for significance level $\alpha = 0.05$). State the null and alternative hypotheses, and explain in words what they mean. Show how you calculate the required test statistic, and state its null distribution. State the decision rule you use, and the inference you would draw from the test.

5. (continued)

(10 marks)

(g) Use the results from OLS estimation of regression equation (2) to test the proposition that the coefficient β_3 of $\ln POP_i$ *equals* the coefficient β_4 of $\ln DEN_i$. Perform this test at the 5 percent significance level (i.e., for significance level $\alpha = 0.05$). In addition to the information given above for the OLS estimates of regression equation (2), you are told that the estimated covariance of the coefficient estimates $\hat{\beta}_3$ and $\hat{\beta}_4$ is $\hat{Cov}(\hat{\beta}_3, \hat{\beta}_4) = -0.006523$. State the null and alternative hypotheses, and explain in words what they mean. Show how you calculate the required test statistic, and state its null distribution. State the decision rule you use, and the inference you would draw from the test.

Selected Formulas

For the Simple (Two-Variable) Linear Regression Model

$$Y_i = \beta_1 + \beta_2 X_i + u_i \qquad (i = 1, \ldots, N)$$

- Deviations from sample means are defined as:

$$y_i \equiv Y_i - \overline{Y}; \qquad x_i \equiv X_i - \overline{X};$$

where

$\overline{Y} = \Sigma_i Y_i / N = \dfrac{\Sigma_i Y_i}{N}$ is the sample mean of the Y_i values;

$\overline{X} = \Sigma_i X_i / N = \dfrac{\Sigma_i X_i}{N}$ is the sample mean of the X_i values.

- Formulas for the variance of the OLS intercept coefficient estimator $\hat{\beta}_1$ and the covariance of the OLS coefficient estimators $\hat{\beta}_1$ and $\hat{\beta}_2$ in the two-variable linear regression model:

$$\text{Var}(\hat{\beta}_1) = \frac{\sigma^2 \Sigma_i X_i^2}{N\,\Sigma_i (X_i - \overline{X})^2} = \frac{\sigma^2 \Sigma_i X_i^2}{N\,\Sigma_i x_i^2};$$

$$\text{Cov}(\hat{\beta}_1, \hat{\beta}_2) = -\overline{X}\left(\frac{\sigma^2}{\Sigma_i x_i^2}\right).$$

- Formulas for the variance of the conditional predictor $\hat{Y}_0 = \hat{\beta}_1 + \hat{\beta}_2 X_0$:

  - When $\hat{Y}_0$ is used as a mean predictor of $E(Y_0 \mid X_0) = \beta_1 + \beta_2 X_0$,

$$\text{Var}(\hat{Y}_0^m) = \sigma^2 \left[\frac{1}{N} + \frac{(X_0 - \overline{X})^2}{\Sigma_i x_i^2}\right].$$

  - When $\hat{Y}_0$ is used as an individual predictor of $Y_0 \mid X_0 = \beta_1 + \beta_2 X_0 + u_0$,

$$\text{Var}(\hat{Y}_0) = \sigma^2 + \sigma^2 \left[\frac{1}{N} + \frac{(X_0 - \overline{X})^2}{\Sigma_i x_i^2}\right].$$

Selected Formulas (continued)

<u>For the Multiple (Three-Variable) Linear Regression Model</u>

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \qquad (i = 1, \ldots, N)$$

- Deviations from sample means are defined as:

$$y_i \equiv Y_i - \overline{Y}; \qquad x_{2i} \equiv X_{2i} - \overline{X}_2; \qquad x_{3i} \equiv X_{3i} - \overline{X}_3;$$

where

$\overline{Y} = \Sigma_i Y_i / N = \dfrac{\Sigma_i Y_i}{N}$ is the sample mean of the Y_i values;

$\overline{X}_{2i} = \Sigma_i X_{2i} / N = \dfrac{\Sigma_i X_{2i}}{N}$ is the sample mean of the X_{2i} values;

$\overline{X}_{3i} = \Sigma_i X_{3i} / N = \dfrac{\Sigma_i X_{3i}}{N}$ is the sample mean of the X_{3i} values.

- The OLS slope coefficient estimators $\hat{\beta}_2$ and $\hat{\beta}_3$ in deviation-from-means form are:

$$\hat{\beta}_2 = \frac{\left(\Sigma_i x_{3i}^2\right)\left(\Sigma_i x_{2i} y_i\right) - \left(\Sigma_i x_{2i} x_{3i}\right)\left(\Sigma_i x_{3i} y_i\right)}{\left(\Sigma_i x_{2i}^2\right)\left(\Sigma_i x_{3i}^2\right) - \left(\Sigma_i x_{2i} x_{3i}\right)^2};$$

$$\hat{\beta}_3 = \frac{\left(\Sigma_i x_{2i}^2\right)\left(\Sigma_i x_{3i} y_i\right) - \left(\Sigma_i x_{2i} x_{3i}\right)\left(\Sigma_i x_{2i} y_i\right)}{\left(\Sigma_i x_{2i}^2\right)\left(\Sigma_i x_{3i}^2\right) - \left(\Sigma_i x_{2i} x_{3i}\right)^2}.$$

- Formulas for the variances and covariances of the slope coefficient estimators $\hat{\beta}_2$ and $\hat{\beta}_3$:

$$\mathrm{Var}\left(\hat{\beta}_2\right) = \frac{\sigma^2 \Sigma_i x_{3i}^2}{\left(\Sigma_i x_{2i}^2\right)\left(\Sigma_i x_{3i}^2\right) - \left(\Sigma_i x_{2i} x_{3i}\right)^2};$$

$$\mathrm{Var}\left(\hat{\beta}_3\right) = \frac{\sigma^2 \Sigma_i x_{2i}^2}{\left(\Sigma_i x_{2i}^2\right)\left(\Sigma_i x_{3i}^2\right) - \left(\Sigma_i x_{2i} x_{3i}\right)^2};$$

$$\mathrm{Cov}\left(\hat{\beta}_2, \hat{\beta}_3\right) = \frac{\sigma^2 \Sigma_i x_{2i} x_{3i}}{\left(\Sigma_i x_{2i}^2\right)\left(\Sigma_i x_{3i}^2\right) - \left(\Sigma_i x_{2i} x_{3i}\right)^2}.$$

Percentage Points of the t-Distribution

Selected Upper Percentage Points of the F-Distribution