QUEEN'S UNIVERSITY AT KINGSTON
Department of Economics

**ECONOMICS 351* - Section A**

**Introductory Econometrics**

Fall Term 1997 **FINAL EXAMINATION** M.G. Abbott

DATE: **Friday December 12, 1997.**

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:** | (2 of 3 questions; 40 marks for each question)<br>Questions 1, 2, and 3; answer **any two**. | .................. | **80 marks** |
| **PART B:** | (2 of 2 questions; 60 marks for each question)<br>Questions 4 and 5; answer all parts of **both**. | .................. | **120 marks** |
| **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, $\beta_1$ and $\beta_2$ 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 $\beta_1$, $\hat{\beta}_2$ is the OLS estimator of the slope coefficient $\beta_2$, $\hat{u}_i$ 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, prove that the OLS slope coefficient estimator $\hat{\beta}_2$ is an unbiased estimator of the slope coefficient $\beta_2$. Include in your answer a definition of unbiasedness.

**(15 marks)**

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

**(10 marks)**

**(c)** State the error normality assumption, and explain why it is important.

**(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, $\beta_1$, $\beta_2$, and $\beta_3$ 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 $\beta_1$, $\hat{\beta}_2$ is the OLS estimator of the slope coefficient $\beta_2$, $\hat{\beta}_3$ is the OLS estimator of the slope coefficient $\beta_3$, $\hat{u}_i$ is the OLS residual for the i-th sample observation, and N is sample size (the number of observations in the sample).

**(5 marks)**

**(a)** State the Ordinary Least Squares (OLS) estimation criterion.

**(15 marks)**

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

**(10 marks)**

**(c)** Derive the OLS decomposition equation for $TSS \equiv \sum_{i=1}^{N} y_i^2 = \sum_{i=1}^{N} (Y_i - \bar{Y})^2$, the total sum-of-squares of the observed $Y_i$ values around their sample mean $\bar{Y}$ in sample regression equation (2). State all the computational properties of the OLS sample regression equation (2) on which the OLS decomposition equation depends.

**(10 marks)**

**(d)** State the Gauss-Markov theorem, and explain what it means.

**(40 marks)**

**3.** You are given a random sample of observations on N individual persons, $N_f$ of whom are identified as females and $N_m$ of whom are identified as males. The sample data provide observations on the observable variables $Y_i$, $X_{2i}$, $X_{3i}$, and $F_i$, where $F_i$ is a female indicator (or dummy) variable defined as follows:

$$\begin{aligned} F_i &= 1 \quad \text{if person i is female} \\ &= 0 \quad \text{if person i is male.} \end{aligned}$$

The population regression equation for females is:

$$Y_i = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + u_{1i} \qquad u_{1i} \sim N\left(0, \sigma_1^2\right) \qquad \textbf{(1)}$$

where $\alpha_1$, $\alpha_2$, and $\alpha_3$ are the female regression coefficients and $u_{1i}$ is an unobservable random error term. The population regression equation for males is:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_{2i} \qquad u_{2i} \sim N\left(0, \sigma_2^2\right) \qquad \textbf{(2)}$$

where $\beta_1$, $\beta_2$, and $\beta_3$ are the male regression coefficients and $u_{2i}$ is an unobservable random error term.

**(20 marks)**

**(a)** Explain fully how you would use the female indicator variable $F_i$ to test the following hypothesis:

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

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

State the regression equation you use to perform the test and define its regression coefficients; explain how you calculate the required test statistic; and state the decision rule you use to decide between rejection and nonrejection of the null hypothesis.

**(20 marks)**

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

**(1)** $H_0$: $\alpha_j = \beta_j \quad \forall \ j = 2, 3$ against $H_1$: $\alpha_j \neq \beta_j \quad j = 2, 3$;

**(2)** $H_0$: $\alpha_2 = \beta_2$ against $H_1$: $\alpha_2 \neq \beta_2$.

For each hypothesis test, explain how you calculate the required test statistic and state the decision rule you use to decide between rejection and nonrejection of the null hypothesis.

## PART B (120 marks)

***Instructions:*** Answer all parts of **BOTH** questions 4 and 5 in this part. Total marks for each question equal 60; 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 how the total weekly sales revenues of a chain of fast-food hamburger restaurants ($TR_t$) are related to the price per unit of the chain's hamburgers in week t ($P_t$) and the chain's weekly advertising expenditures in week t ($A_t$). The sample data consist of 52 weekly observations on the variables $TR_t$, $P_t$, and $A_t$, which are defined as follows:

$TR_t \equiv$ total sales revenues in week t, measured in thousands of dollars per week;
$P_t \equiv$ the price of hamburgers in week t, measured in dollars per hamburger;
$A_t \equiv$ advertising expenditures in week t, measured in thousands of dollars per week.

Using the sample data described above, you begin the investigation by estimating the following regression equation:

$$TR_t = \beta_1 + \beta_2 P_t + \beta_3 A_t + u_{1t} \qquad \textbf{(1)}$$

where the $\beta_j$ $(j = 1, 2, 3)$ are regression coefficients and $u_{1t}$ is a random error term.

**(60 marks)**

**4.** OLS estimation of regression equation (1) yields the following OLS sample regression equation and summary statistics (with estimated standard errors in parentheses below the coefficient estimates):

$$\begin{aligned} TR_t &= \underset{(6.483)}{104.8} - \underset{(3.191)}{6.642}\, P_t + \underset{(0.1669)}{2.984}\, A_t + \hat{u}_{1t} \end{aligned} \qquad (1^*)$$

$$RSS_{(1)} = \sum_{t=1}^{N} \hat{u}_{1t}^2 = 1805.2; \quad TSS_{(1)} = \sum_{t=1}^{N} \left(TR_t - \overline{TR}\right)^2 = 13{,}581.4; \quad N = 52 \qquad (1')$$

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

**(10 marks)**

**(a)** Interpret the slope coefficient estimates in OLS sample regression equation (1*); that is, explain what they mean.

**4. (continued)**

**(10 marks)**

**(b)** Use the above results for OLS sample regression equation (1*) to test the individual significance of the slope coefficient estimates $\hat{\beta}_2$ and $\hat{\beta}_3$. For each test, state the null and alternative hypotheses, show how you calculate the required test statistic, and state its null distribution. Which slope coefficient estimates are individually significant at the 5 percent significance level? Which slope coefficient estimates are individually significant at the 1 percent significance level?

**(10 marks)**

**(c)** Use the results for OLS sample regression equation (1*) to compute the two-sided 95 percent confidence interval for the slope coefficient $\beta_2$ of the regressor $P_t$. Explain briefly how the confidence interval for $\beta_2$ is interpreted.

**(10 marks)**

**(d)** Use the results for OLS sample regression equation (1*) to test the ***joint*** significance of the slope coefficient estimates $\hat{\beta}_2$ and $\hat{\beta}_3$ 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)**

**(e)** Use the results for OLS sample regression equation (1*) to test the null hypothesis $H_0$: $\beta_3 \geq 2$ against the alternative hypothesis $H_1$: $\beta_3 < 2$ at the 5 percent significance level (i.e., for significance level $\alpha = 0.05$). 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. Briefly indicate how you interpret the results of the test.

**(10 marks)**

**(f)** A sales manager for the hamburger chain claims that the marginal effect on $TR_t$ of $P_t$ is twice the size and opposite in sign to the marginal effect on $TR_t$ of $A_t$. That is, the sales manager claims that $\beta_2 = -2\beta_3$. Use the results for OLS sample regression equation (1*) to test this hypothesis at the 5 percent significance level (i.e., for significance level $\alpha = 0.05$). In addition to the information given above for sample regression equation (1*), you are told that the estimated covariance of the coefficient estimates $\hat{\beta}_2$ and $\hat{\beta}_3$ is $\hat{\mathrm{Cov}}\left(\hat{\beta}_2, \hat{\beta}_3\right) = -0.05402$, and that the Residual Sum-of-Squares for the restricted OLS sample regression equation that imposes the restriction $\beta_2 = -2\beta_3$ is $RSS = 1806.8$. 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.

**(60 marks)**

**5.** You now decide to assess the specification of regression equation (1) above. You propose to do this by estimating the regression equation

$$TR_t = \beta_1 + \beta_2 P_t + \beta_3 A_t + \beta_4 P_t^2 + \beta_5 A_t^2 + \beta_6 P_t A_t + u_{2t} \tag{2}$$

where the $\beta_j \;\; (j = 1, \ldots, 6)$ are regression coefficients and $u_{2t}$ is a random error term. OLS estimation of regression equation (2) yields, among other things, the following results:

$$RSS_{(2)} = \sum_{t=1}^{N} \hat{u}_{2t}^2 = 1754.6; \quad TSS_{(2)} = \sum_{t=1}^{N} \left(TR_t - \overline{TR}\right)^2 = 13{,}581.4; \quad N = 52 \tag{2$'$}$$

where $RSS_{(2)}$ is the Residual Sum-of-Squares and $TSS_{(2)}$ is the Total Sum-of-Squares from OLS estimation of equation (2). The corresponding results from OLS estimation of regression equation (1) in question 4 are repeated here for ease of reference:

$$RSS_{(1)} = \sum_{t=1}^{N} \hat{u}_{1t}^2 = 1805.2; \quad TSS_{(1)} = \sum_{t=1}^{N} \left(TR_t - \overline{TR}\right)^2 = 13{,}581.4; \quad N = 52 \tag{1$'$}$$

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

**(10 marks)**

**(a)** Give expressions for the marginal effects on $TR_t$ of the regressors $P_t$ and $A_t$ in regression equation (2). How do the marginal effects on $TR_t$ of the regressors $P_t$ and $A_t$ in equation (2) differ from those in equation (1)?

**(10 marks)**

**(b)** 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.

**(10 marks)**

**(c)** Identify an alternative goodness-of-fit measure 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 regression equations (1) and (2). Which of the two OLS sample regression equations provides the better fit to the sample data?

**5. (continued)**

**(20 marks)**

**(d)** State the coefficient restrictions that regression equation (1) imposes on regression equation (2). 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)**

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

Given the results of the hypothesis test performed in part (d) of this question, which set of estimation results would you choose to report to the executives of the hamburger chain, those for equation (1) or those for equation (2)?

## 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:

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

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

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 $E\left(Y_0 \middle| X_0\right) = \beta_1 + \beta_2 X_0$,

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

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

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

## Selected Formulas (continued)

**For the Multiple (Three-Variable) Linear Regression Model**

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

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