Department of Economics **University of Manitoba**

ECON 7010: Econometrics I
Midterm 2, Nov. 10, 2015

Instructor: **Ryan Godwin**
Instructions: **Answer ALL QUESTIONS, and put all answers in the booklet provided**
Time Allowed: **75 minutes (Total marks = 80 – *i.e., ~ one mark per minute*)**
Number of Pages: **3**

PART A: Short answer – **choose 4 out of 5**. [10 marks each]

1.) Explain the differences between the Wald and F-test, and when it is appropriate to use each.

2.) Using an initial value of $\theta_0 = 2$, calculate the first few iterations of the Newton algorithm to find the value of θ that minimizes the function:

$$f(\theta) = \theta^3 - 3\theta.$$

3.) Briefly describe how the Non-linear Least Squares (NLLS) estimator is derived, and how NLLS estimates are typically obtained.

4.) Explain how to compare the asymptotic efficiency of two mean-square consistent estimators, even when the asymptotic distribution degenerates to a "spike".

5.) Prove that: (i) the OLS estimator is inconsistent; but (ii) the IV estimator is consistent, when the error term and a regressor are *not* independent. State any assumptions you use.

PART B: Longer Answer – **choose 2 out of 3**. [20 marks each]

6) Suppose that we wish to estimate the standard multiple regression model, with fixed regressors:

$$y = X\beta + \varepsilon \quad ; \quad \varepsilon \sim N(0, \sigma^2 I)$$

subject to a set of "J" exact linear restrictions $R\beta = q$; where rank(R) = J. The Restricted Least Squares (RLS) estimator of β is:

$$b_* = b - (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(Rb - q),$$

Where b is the OLS estimator of β.

(a) If e^* is the RLS residual vector, and "e" is the OLS residual vector, prove that:

$$e_*'e_* = e'e + (Rb - q)'[R(X'X)^{-1}R']^{-1}(Rb - q).$$

(b) Explain why $e_*'e_*$ cannot be less than $e'e$.

(c) Under what condition would $e_*'e_* = e'e$?

(d) Explain how you could actually apply the RLS estimator, using just OLS, in the case where the model is: $y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i} + \varepsilon_i$, and the restrictions that we want to impose are: $\beta_1 + \beta_2 = 1$, and $\beta_3 = \beta_4$.

7) Suppose that we want to estimate a linear regression model:

$$y_i = \beta_1 + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + \varepsilon_i \quad ; \quad i = 1,2, \ldots, n$$

This model satisfies *all* of the usual assumptions, however, we are not provided with individual data for the n values of each of the variables. Instead, the data are the average group values among m groups. (This is sometimes called "clustering".) There are n_j observations in the j^{th} group; where $j = 1,2, \ldots, m$; **and where $\boldsymbol{n_j = j}$.** The data that are available are:

$$\bar{y}_j = \frac{1}{n_j}\sum_{j=1}^{n_j} y_j \quad ; \quad \bar{x}_{2j} = \frac{1}{n_j}\sum_{j=1}^{n_j} x_{2j} \quad ; \quad \ldots \quad ; \quad \bar{x}_{kj} = \frac{1}{n_j}\sum_{j=1}^{n_j} x_{kj} \quad ; \quad \text{and } n_i \quad ; \quad i = 1,2, \ldots, m.$$

This means that the model we have to estimate is actually:

$$\bar{y}_j = \beta_1 + \beta_2 \bar{x}_{2j} + \cdots + \beta_k \bar{x}_{kj} + \bar{\varepsilon}_j$$

(a) Derive the variance of $\bar{\varepsilon}_j$.

(b) Use your answer in part (a) to construct the GLS estimator of β.

(c) What would be the consequences of ignoring the information on the group sizes, n_j?

(d) Now, suppose that you do not know the group sizes, n_j. Explain some ways that you could deal with the problem of heteroscedasticity.

8) Consider the following model:

$$income_i = \beta_1 + \beta_2 education_i + \beta_3 age_i + \varepsilon_i$$

The data consist of 1000 observations total, 500 of which are for males and 500 for females. The variables are: *income* – yearly income; *education* – years of education; and *age*. In addition, there is a dummy variable, *male*, which equals "1" if the individual is male, and "0" otherwise.

a) Suppose that you divide the sample into two (by gender) and estimate the above model by OLS twice separately. The sum-of-squared-residuals for the "males regression" is $e_M' e_M = 0.06$, and for females is $e_F' e_F = 0.05$. Then, you estimate the "pooled model" (all 1000 observations), and get $e_*' e_* = 0.15$. Calculate the F statistic for determining if there are differences in the effects on income between males and females.

b) Describe an alternative way in which you could enact the test in part (a).

c) Suppose that the variable *education* is correlated with the error term. Explain why this is a problem.

d) Suppose that the variable *education* is correlated with the error term, but that you have an instrument: *distance from college*. Explain how you could use this variable in two-stage-least-squares (2SLS) in order to fix the problem.