Department of Economics **University of Manitoba**

ECON 7010: Econometrics I
FINAL EXAM, Dec. 19th, 2014

Instructor: **Ryan Godwin**
Instructions: **Put all answers in the booklet provided.**
Time Allowed: **3 hours.**
Number of Pages: **4**

There are a total of 100 marks.

PART A: *Short answer.* ***Answer 8 out of 10 questions.*** *Each question is worth 5 marks.*

1.) Describe how to implement White's test for heteroskedasticity. Is this test constructive? Why or why not?

2.) Explain what a "spurious regression" is.

3.) Suppose that all of the usual OLS assumptions are satisfied, except that the error term follows an AR(1) process. Explain how we might implement FGLS in this case.

4.) Describe how heteroskedasticity can affect OLS estimation, and the possible remedies to the problems.

5.) Show that an AR(1) process can be written as an MA(∞) process.

6.) Suppose that the R^2 from the full regression model is 0.5. You conduct an F-test, and decide to remove two variables from the model. You initially had 8 variables, and a sample size of 108. After removing the two variables, the R^2 is now 0.4. What was the value of the F-statistic?

7.) Suppose that we have a linear regression model, with *k* regressors:

$$y = X\beta + \varepsilon,$$

and we are concerned with *J* linear restrictions on β, of the form $R\beta = q$. Let *b* be the OLS estimator of β, and let *b** be the corresponding Restricted Least Squares (RLS) estimator. (*See your formulae sheet for the formula for the RLS estimator*.) Under what conditions is *b** an unbiased estimator of β?

8.) Carefully explain how to interpret a *p*-value, using an example if desired.

9.) Using suitable diagrams, describe how the Newton-Raphson algorithm works, and some of the problems that may arise in its application.

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

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

Choose 8 questions from above. For PART A, I will only mark the first 8 questions in your exam booklet.

PART B. ANSWER ALL QUESTIONS.

11.) Suppose that we want to estimate the following model by Instrumental Variables (IV) estimation, using a matrix of instruments, Z:

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where we are using the same number of instruments as we have regressors.

(a) List two conditions that we require the matrix of instruments to satisfy.

3 marks

(b) Assuming that these conditions are satisfied, prove that the IV estimator of the coefficient vector is (weakly) consistent.

6 marks

(c) Now suppose that in fact the true data-generating process is

$$\boldsymbol{y} = X\boldsymbol{\beta} + W\boldsymbol{\gamma} + \boldsymbol{\varepsilon}$$

where W is a matrix of observations for additional random regressors, such that

$$plim\left(\frac{1}{n}Z'W\right) \neq 0.$$

Prove that the IV estimator is now inconsistent.

6 marks

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

$$y_j = \beta_1 + \beta_2 x_{2j} + \cdots + \beta_k x_{kj} + \varepsilon_j \quad ; \quad j = 1,2,\ldots,n \tag{1}$$

This model satisfies *all* of the usual assumptions. The only problem is that we are not provided with individual data for the *n* values of each of the variables. Instead, the data have been gathered by conducting a survey across *m* groups of people, and then recording the group average values. (This is sometimes called "clustering".) There are different numbers of people (n_i) in each group, and we have this information as well. So, the data that are available are:

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

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

$$\bar{y}_i = \beta_1 + \beta_2 \bar{x}_{2i} + \cdots + \beta_k \bar{x}_{ki} + \bar{\varepsilon}_i \quad ; \quad i = 1,2,\ldots,m \tag{2}$$

(a) Show that the variance of $\bar{\varepsilon}_i$ in equation (2) is (σ^2/n_i), where σ^2 is the variance of each ε_j in equation (1). What are the other properties of the error term in equation (2)?

6 marks

(b) Expalin how you would estimate equation (2) by GLS. (What is the Ω matrix? What is the *P* matrix?) Why would it be preferable to use the GLS estimator, rather than applying OLS to equation (2)?

9 marks

13.) The exponential distribution may be used to describe the time between events in a Poisson process. A random variable, y_i, which follows an exponential distribution has probability density function (p.d.f.):

$$p(y_i|\lambda) = \lambda e^{-y_i\lambda}$$

and has mean:

$$E(y_i) = {}^{1}\!/_{\lambda}$$

a) Write down the log-likelihood function for this problem.

5 marks

b) Solve for the maximum likelihood estimator of λ. Check to ensure that your solution maximizes (and not minimizes) the likelihood function.

6 marks

c) Explain the intuition behind the likelihood ratio test.

4 marks

14.) Assume that *all* of the usual assumptions hold.

a) Prove that, in general, the OLS estimator is biased when a variable is excluded from the model. Under what special circumstances is OLS unbiased when a variable is excluded?

8 marks

b) Prove that if an irrelevant regressor is included in the model, the OLS estimator is unbiased.

7 marks

END