Department of Economics **University of Manitoba**

ECON 7010: Econometrics I
Midterm, Oct. 13, 2016

Instructor: **Ryan Godwin**
Instructions: **Answer ALL QUESTIONS, and put all answers in the booklet provided**
Time Allowed: **75 minutes (Total marks = 60)**
Number of Pages: **3**

PART A:

Select the most appropriate answer in each case. Each question is worth ***3 marks****. (No explanation is needed to obtain full marks, but it will be taken into account if given.)*

1.) The formula for the OLS estimator: $\boldsymbol{b} = (X'X)^{-1}X'\boldsymbol{y}$, is derived by:

a) Ensuring that $\boldsymbol{b}$ is a linear, unbiased and efficient estimator for $\boldsymbol{\beta}$.
b) Minimizing the sum of squared residuals.
c) Minimizing bias and variance.
d) Ensuring that $\boldsymbol{b}$ has the best fit (which also maximizes R^2).

2.) The Least Squares principle for estimating a regression model, $y = X\beta + \varepsilon$, where *all* of the usual assumptions are satisfied:

a) Produces unbiased and efficient estimators of both β and σ.
b) Involves minimizing the sum of the squares of elements of ε.
c) Produces an estimator for β that has a Normal sampling distribution, centered at β itself.
d) Produces an equal number of positive and negative residuals if the sample size is even.

3.) A p-value is:

a) The probability of calculating a test statistic more extreme than the one just calculated.
b) The maximum and minimum values for the test statistic, that won't be rejected in a hypothesis test.
c) The maximum and minimum values for the null hypothesis, that won't be rejected in a hypothesis test.
d) Equal to the probability of a type I error.

4.) The correct interpretation of a 95% confidence interval constructed around a regression coefficient is:

a) There's a 0.95 probability that the true value of the regression coefficient lies in the interval.
b) The interval includes the true value of the regression coefficient 95% of the time.
c) 95% of such intervals will contain the true value of the regression coefficient.
d) None of the above.

5.) When we prove that the OLS estimator is unbiased:

a) Assumption A2: full rank, is not needed.
b) Assumption A3: errors have zero mean, is not needed.
c) Assumption A4: homoskedasticity and non-autocorrelation, is not needed.
d) none of the above.

PART B: *Answer all questions.*

6.) Show that the OLS residuals sum to zero if the model includes an intercept.

[10 marks]

7.) Derive the variance-covariance matrix for the OLS estimator, **b**. State any assumptions that you use.

[10 marks]

8.) Consider the population model:

$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i,$

where the usual assumptions A1 to A6 hold. Consider the following estimator for β_1:

$$\tilde{\beta}_1 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - \bar{y}}{x_i - \bar{x}}\right)$$

a) Show that this estimator is unbiased.

b) Is this estimator linear?

c) What can say about the variance of this estimator as compared to the OLS estimator?

[15 marks]

9.) A common strategy for handling the case in which some data is missing for an observation is to add a "dummy" variable to the model that takes the value 1 for the observations with missing data and 0 for all other observations. Show that, in terms of the computation of the OLS estimator **b**, this "strategy" is equivalent to discarding the observations with missing data.

Hint 1: Without loss of generality, assume that only one observation has missing data, and that it is the first observation. Hence, the new variable added to the *X* matrix would look like:

$$\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Hint 2: Use the results for partitioned and partial regression, i.e., the *M* matrix.

[10 marks]

END.