Department of Economics **University of Manitoba**

ECON 7010: Econometrics I
Sample Midterm

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

NOTE: The purpose of this sample midterm is to prepare you for the format and style of the midterm. This sample is longer (and probably more difficult) than the real one. Some of the topics may not be relevant to the real midterm.

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.)

Question 1:
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.

Question 2:
The "power" of any "consistent" statistical test:
(a) Is usually one minus the probability of a "Type II" error.
(b) Is equal to the significance level chosen for the test, when the null hypothesis is true.
(c) Increases (or at least does not decrease) as the sample size grows.
(d) All of the above.

Question 3:
When applying the Instrumental Variables estimator for the coefficient vector in a linear regression model:
(a) We require at least as many instruments as there are regressors.
(b) We require at least as many instruments as there are regressors, and the instruments should be non-random.
(c) We require at least as many instruments as there are regressors, and the instruments should be uncorrelated with the errors (at least asymptotically).
(d) We require the same number of instruments as there are regressors, and the instruments should be uncorrelated with the errors (at least asymptotically).

Question 4:
The coefficient of determination (R^2) for the usual least squares regression model:
(a) Always lies strictly between zero and unity in value.
(b) Must increase when we add one or more regressors to the model.
(c) Is an unbiased estimator of the squared population correlation coefficient between y and the single regressor, when $k = 1$.

(d) Cannot be less than the "adjusted" coefficient of determination (R^2).

Question 5:
The Gauss-Markov Theorem tells us that, under appropriate assumptions, the least squares estimator of β in the usual linear regression model:
(a) Is a linear estimator, and therefore is "best".
(b) Has the smallest bias among all possible linear estimators for this parameter vector.
(c) Is most efficient among all possible linear and unbiased estimators of this parameter.
(d) Is most efficient among all possible unbiased estimators that have a Normal sampling distribution.

PART B:
State whether each of the following is TRUE or FALSE, and BRIEFLY explain your answer. Each question is worth ***6 marks****. Of these,* ***5 marks are given for the explanation****.*

Question 6:

(a) I am testing the hypothesis that one of the coefficients in a linear regression model takes the value unity, against a 2-sided alternative hypothesis. I reject the null hypothesis at the 1% significance level. Without any more information, you know for sure that I will also reject the null hypothesis at the 5% and 10% significance levels.

(b) The Hausman and Wu tests are designed to test if the instruments that we plan to use meet the requirement of being uncorrelated with the errors (at least asymptotically). In each case, a rejection of the null hypothesis would lead us to use OLS instead of I.V. estimation.

(c) Suppose that all of the usual assumptions for our regression model hold, except that the errors are non-normal. Let b_i denote the usual OLS estimator of β_i, the *i*th element of β. Then $[b_1/b_2]$ is an unbiased and weakly consistent estimator of $[\beta_1/\beta_2]$.

(d) Suppose that the second regressor in a *k*-regressor multiple linear regression OLS regression model is a variable that takes the value "1" for the first and last observations, but is zero for all other observations. Then, $e_1 = e_n$, where the e_i*'s* are the OLS residuals.

(e) One connection between the variance of an estimator and the mean squared error of that estimator is that the mean squared error cannot be less than the variance.

PART C: *Answer all questions.*

Question 7:
Consider a simple regression model with a single regressor, fitted through the origin:

$$y_i = \beta x_i + \varepsilon_i \qquad ; \qquad i = 1,2,3, \ldots\ldots, n.$$

Consider the following estimator for β: $\hat{\beta} = (\bar{y}/\bar{x})$, where $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ and $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

(a) Is this estimator a 'linear estimator'?
(3 marks)
(b) Prove that this estimator is unbiased.
(3 marks)

(c) Derive the variance of $\hat{\beta}$.
(4 marks)
(d) Let $\hat{\varepsilon}$ be the residual vector associated with $\hat{\beta}$. Does $X'\hat{\varepsilon} = 0$, necessarily?
(4 marks)
(e) What is the value of $X'\hat{\varepsilon}$ if the regressor takes the value "one" for all observations?
(4 marks)
(f) What is the sum of the residuals associated with the estimator $\hat{\beta}$?
(4 marks)
(g) When we estimate the model in this way, does the fitted line pass through the sample mean of the data?
(3 marks)
Total: 25 marks

Question 8:
Consider the linear multiple regression model, $y = X\beta + \varepsilon$, where all of the usual assumptions are satisfied, ***except*** that the regressors are random and correlated with the errors (even asymptotically). That is, $plim\left(\frac{1}{n}X'\varepsilon\right) = \gamma \neq 0$, and γ is finite. We have available a set of k instrumental variables which form the columns of the (n x k) matrix Z. The X and Z matrices satisfy the following conditions:

(i) $plim\left(\frac{1}{n}X'X\right) = Q_{XX}$; positive-definite & finite.

(ii) $plim\left(\frac{1}{n}Z'Z\right) = Q_{ZZ}$; positive-definite & finite.

(iii) $plim\left(\frac{1}{n}Z'X\right) = Q_{ZX}$; positive-definite & finite.

(iv) $plim\left(\frac{1}{n}Z'\varepsilon\right) = 0$.

In this case the I.V. estimator for β is $b_{IV} = (Z'X)^{-1}Z'y$.

(a) Show that the I.V. residual vector is $e_{IV} = W\varepsilon$, where $W = [I - X(Z'X)^{-1}Z']$.
6 marks

(b) Show that the sum of the squares of these I.V. residuals can be written as:
$$\varepsilon'\varepsilon - \varepsilon'X(Z'X)^{-1}Z'\varepsilon - \varepsilon'Z(X'Z)^{-1}X'\varepsilon + \varepsilon'Z(X'Z)^{-1}X'X(Z'X)^{-1}Z'\varepsilon$$
6 marks

(c) Using assumptions (i) – (iv) and Khintchine's Theorem, prove that $s_{IV}^2 = (e_{IV}'e_{IV})/n$ is a weakly consistent estimator of σ^2 (the variance of the error term in the model).
[**Hint:** the proof follows the same lines as the proof of the weak consistency of s^2 for the OLS case.]
13 marks
Total: 25 marks

Question 9:

The following EViews results are for a model explaining the length of time between the filing of a patent claim for a new invention in the U.S.A., and the award of the patent right to the applicant. (Only successful applications are used.) The variables are: TIME = number of years from application to award of patent; CITES = number of citations to earlier work made in the application; CLAIMS = number of claims that are made regarding the originality of the new invention; APPYEAR = year in which the application was lodged (1963 to 1998); CAT1 to CAT5 = dummy variables that indicate the technological category that the invention falls into. (There are 6 such categories in all.)

Dependent Variable: LOG(TIME)
Method: Least Squares
Date: 10/22/07 Time: 16:22
Sample: 1 to 2,923,922 IF APPYEAR>=1963 AND CLAIMS>0
Included observations: 1,983,420

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	23.24623	0.093471	248.6990	0.0000
CITES	0.003776	4.03E-05	93.57141	0.0000
CLAIMS	0.002786	3.22E-05	86.44149	0.0000
APPYEAR	-0.011506	4.71E-05	-244.1827	0.0000
CAT1	0.056065	0.001001	56.02982	0.0000
CAT2	0.243224	0.001208	201.2811	0.0000
CAT3	0.179390		135.3954	0.0000
CAT4	0.078158	0.001051	74.37198	0.0000
CAT5	0.036124	0.000978	36.92930	0.0000

R-squared	0.052029	Mean dependent var	0.527984
Adjusted R-squared	0.052025	S.D. dependent var	0.464929
S.E. of regression		Akaike info criterion	1.252712
Sum squared resid	406426.7	Schwarz criterion	1.252769
Log likelihood	-1242318.	F-statistic	13607.29
Durbin-Watson stat	1.476812	Prob(F-statistic)	0.000000

(a) Do the signs of the estimated coefficients of CLAIMS and CITES seem sensible? What might explain why the coefficient on APPYEAR is negative?

(5 marks)

(b) Why are the R-squared and adjusted R-squared values almost identical? Interpret what the R-squared value tells us.

(4 marks)

(c) Calculate the values for the “S.E. (standard error) of regression” and the standard error for the estimated coefficient of the CAT3 dummy variables.

(4 marks)

(d) Construct a 99% confidence interval for the coefficient of CAT1, and carefully interpret its meaning.

(6 marks)

(e) Predict the length of time you would expect a (successful) patent application to take if it fell into technological category 1, made 12 claims and 8 citations, and the application was made in 1986.

(3 marks)

Total: 22 marks