ECO383
Economics of Education
*In-Class Test – Sample Questions**

October 2011

Please fill in your full name and student number in the spaces below.

NAME:

STUDENT NUMBER:

You have 1 hour to complete this **closed-book** test. When the instructor asks everyone to stop, you must stop writing. This should be entirely your own work – **no conferring please**.

The test is worth X points. Please attempt all the questions, and be sure to read each question carefully. The number of points each question is worth is indicated in brackets.

You should answer the test directly on this booklet, in the spaces provided. Please write legibly, and answer in proper sentences, where sentences are required. If you need additional space, indicate clearly which question you are answering, and write on the reverse side of the page. [Approximate guidelines as to the length of the perfect answer are given in square brackets, where necessary.]

*Instructor: Robert McMillan, email: `mcmillan@chass.utoronto.ca`

Please write clearly in the spaces below (or overleaf if necessary) and answer in proper sentences (rather than in some shorthand code that only you may understand).

Question 1: Summations (worth 5 points)

Let the sample mean of a variable X be defined by

$$\bar{X} \equiv \frac{1}{N} \sum_{i=1}^{N} X_i,$$

given a sample of N realizations of the variable.

a) Can you re-write the following:

$$\sum_{i=1}^{N} \alpha X_i,$$

where α is a fixed parameter, as some function of the sample mean of X? Please set out your reasoning. (2 points)

b) Can you re-write the following:

$$\sum_{i=1}^{N} \alpha (X_i + \gamma Y_i),$$

where α and γ are fixed parameters, as some function of the sample means of X and of Y? Please set out your reasoning carefully. (3 points)

Question 2: 'Noise' (worth 14 points)

Suppose you have test score data on high school students over time. Let student performance on a test be given by the following very simple (purely random) model:

$$S_i = \alpha u_i, \tag{1}$$

where S_i is a student i's performance, u_i is a random error term that captures factors assumed to influence test scores in an unsystematic way, and α is a positive constant.

Suppose that the random error term follows a discrete distribution, taking on three possible values (irrespective of the student), given by

$u_i = 1$ with probability $= 0.25$; $u_i = 2$ with probability=0.5; $u_i = 3$ with probability=0.25.

a) What is the mean (or 'expected value') of the random error term u_i, given the above discrete distribution? Please show your calculations. [Hint: start by writing down the expression for the expected value of a discrete random variable.] (2 points)

b) What is the mean of student i's score? Again, please show your calculations. (3 points)

c) Write down an expression for the variance of student i's score. [Hint: again, start from the relevant formula, this time for the *variance* of a discrete random variable.] (3 points)

d) Given your answer to part c), what would happen to the variance as α became more positive? (2 points)

e) If the score were instead given by

$$S_i = 10 + \alpha u_i, \tag{2}$$

what would the expected value of the score be? Please explain, with reference to the relevant formula. [Hint: what is the expected value of a constant?] (4 points)

Question 3: the Mechanics of Curve Fitting (worth 16 points)

Suppose we are interested (as many researchers have been) in assessing whether schools that have smaller classes produce better academic outcomes.

To that end, you are given a sample of N students drawn from a variety of schools. For each student i, you observe the student's class size (C_i) and a test score (T_i). Imagine you plotted the sample data points on a graph, with class size, C, on the horizontal axis and test scores, T, on the vertical axis.

Suppose the underlying (or 'true') model is given by

$$T_i = \alpha + \beta C_i + u_i, \tag{3}$$

where u_i is a random error with mean zero, and α and β are unknown parameters to be estimated.

Suppose that researchers pick the slope parameters to minimize the least squares objective, given by

$$S = \sum_{i=1}^{N} \Big(e_i(a, b) \Big)^2,$$

where e_i is the residual associated with observation i, given the line parameterized by (a, b), and there are N observations in total.

a) What is the least squares formula for the intercept, a^{ols}? (2 points)

b) What is the least squares formula for the slope coefficient, b^{ols}? Write down the relevant expression accurately. (3 points)

c) **Derive** the least squares intercept *and* slope estimators using calculus, showing the relevant steps. (11 points)