

**Econ 327 - Spring 2010 -- Introduction to Econometrics**  
**Dr. R. Pete Parcells – Assignment One**  
**(Due in class on Tuesday Feb 2, 2010)**  
**You may work in groups of no more than 2 persons**  
**Report (with all names) must be written and turned in during class**

---

**Statistical Review Assignment (Monte Carlo)**

Here is a preview of questions which need to be answered to complete this assignment. They are repeated at the end of this assignment.

Consider the following and write a brief report on your findings and bring to class.

1. How does the mean slope estimate compare with the true known value of the parameter?
  2. How, if at all, is the variance of these estimates of  $\beta$  related to the variance of the X values?
  3. Does the distribution of the values as illustrated in the histogram appear to be normal? How different do you think your answers would have been if you had created 1000, or 10000, rather than 100 samples of 25 observations?
  4. Explain what is meant by the Sampling Distribution of the Least Squares Slope estimate.
  5. Consult your (statistics) notes and textbooks on the theoretical properties of the sampling distribution of the slope coefficient, its mean, and variance. Consider questions of bias.
- 

Monte Carlo studies are an important part of modern econometrics. Researchers can investigate the properties (especially the small sample properties) of estimators and also test procedures where results cannot be derived theoretically.

Using a computer, a large number of artificial or simulated data sets can be created according to a known data generation process. Then an estimator or test procedure can be applied to the artificial data so the pattern of results obtained can be analyzed and compared with the (known) features that were designed into the data. In this way, investigators can get a measure of the extent of any inherent biases in the estimators or in the power of the test procedures under various conditions.

Students of economics and econometrics ought to be aware of the important contribution that such studies are having to the development of the subject. There is great educational value in students conducting their own simple Monte Carlo exercises. This point has been recognized by the authors of such recent texts as Kennedy (*A Guide to Econometrics*).

The purpose of this assignment is not only to teach students about the use of Monte Carlo studies (in a research context) as a method for discovering properties of estimators in situations where they cannot be derived analytically, but also to help students understand concepts (such as the

sampling distribution of a least squares estimator) which may be difficult to grasp when they have to rely solely on their imagination.

**MS Excel**© provides a very convenient environment to use for simple Monte Carlo experiments. It provides built-in functions that can be used to generate the random disturbances for the data generating process. It contains built-in least squares regression estimation procedures (or matrix tools if it is required to construct other estimators or test statistics). In addition, it can provide summary statistics and graphical displays to assess patterns in the results obtained.

It might be argued that other econometrics software such as Eviews might be a more natural tool for a researcher to use for Monte Carlo studies. However, it can make sense for a student to work with a tool with which (s)he is already familiar and which is simple to use - unless it is insufficient for the task at hand. Feel free to write a short program in Eviews to carry out this assignment if you wish to do so.

One of the main purposes of this exercise, is to help students understand the meaning of the sampling distribution of a least squares regression estimator, (such as  $\beta$ ) and the way in which the properties of the sampling distribution reflect the characteristics of the regression model itself. In addition, it is expected that students will learn something of the “flavor” of Monte Carlo studies in general.

There are four stages in a Monte Carlo study.

- 1) The first stage is to construct a model of the data generation process. Assume that it takes the form

$$Y_i = \alpha + \beta X_i + u_i \dots\dots \text{(equation 1)}$$

with  $\alpha=20$  and  $\beta=0.6$  and where  $u_i$  is  $N(0,1)$ .

A fixed set of 25 values for X are given as shown in Table 1. Any set of values could of course be used, but these values have a clearly recognizable mean and variance.

**Table 1**

i	1	2	3	4	5	6	7	8	9	10	11	12	13
X <sub>i</sub>	88	89	90	91	92	93	94	95	96	97	98	99	100
i	14	15	16	17	18	19	20	21	22	23	24	25	
X <sub>i</sub>	101	102	103	104	105	106	107	108	109	110	111	112	

- 2) The second stage is to create sets of data. Use a computer’s random number generator to generate 100 sets of 25 values of the random disturbance  $u$  by taking random drawings from the standard normal distribution (mean = 0, standard deviation = 1). In Excel, this

can be achieved by using the Random Number Generation from the Data, Data Analysis, Random Number Generation option.

Create 100 sets of 25 Y values to go with the 25 X values and the u values, based on equation 1. This process is relatively straightforward, if a little tedious. You need to enter the formula for the first observation of Y in each sample and then use the spreadsheet's copying feature to produce the other 24 data points for each set of Y values. Of course the X values are designed to remain fixed over the different samples.

- 3) The artificial data sets are now used to estimate the parameters of the model. Run 100 regressions, using each of the samples of 25 observations on X and Y, and extract the estimated slope coefficient (100 of them) and put them into a table.
- 4) Analyze the results. Examine the distribution of the set of coefficient estimates obtained, calculating the mean and variance. Produce a frequency table and histogram of the values.

**Consider the following and write a brief report on your findings and print out the necessary graphs and bring to class.**

- A) How does the mean slope estimate compare with the true know value of the coefficient parameter (0.6)?
- B) How, if at all, is the variance of these estimates of  $\beta$  related to the variance of the X values? (What is the formula for the Variance the  $\beta$ , and do the X values change?)
- C) Does the distribution of the values as illustrated in the histogram appear to be normal? How different do you think your answers would have been if you had created 1000, or 10000, rather than 100 samples of 25 observations?
- D) Explain what is meant by the Sampling Distribution of the Least Squares Slope estimate.
- E) Consult your (statistics) notes and textbooks on the theoretical properties of the sampling distribution of the slope coefficient, its mean, and variance. Consider questions of bias. Write up your understanding of estimated slope bias (small and large sample properties).
- F) What was the benefit of implementing this Monte Carlo study of coefficient estimation?