University of Victoria **Department of Economics**

ECON 546: Themes in Econometrics

Lab. Exercise 1

(13 January, 2010)

The purpose of this lab. class is to provide an introduction to the use of Monte Carlo simulation to evaluate some of the sampling properties of an estimator. What we will do is mimic the sampling distribution of the estimator, and then look at certain features of this distribution to get a measure of the corresponding estimator properties such as bias, variance, etc.

Suppose that we have a linear multiple regression model, satisfying all of the usual assumptions:

$$y = X\beta + \varepsilon \quad ; \qquad \varepsilon \sim N[0, \sigma^2 I_n]$$

where the regressors are non-random.

Recall that: (i) The OLS coefficient estimator, *b*, is unbiased.

(ii) $V(b) = \sigma^2 (X'X)^{-1}$.

Suppose that we did not know these results. We could learn something about these properties of *b* by conducting the following Monte Carlo experiment:

(1) Assign values to β and σ, and to the elements of the *X* matrix.
(2) Generate a sample of *n* random values from a $N[0, \sigma^2]$ distribution and assign these to the elements of ε.
(3) Generate an *n*-element vector, *y*, using *X*, β, and ε.
(4) Regress *y* on *X* by OLS, and store the value of *b*.
(5) Repeat steps (2) to (4) many, many times.
(6) Take all of the values of *b* and analyze the distribution of these values. This is an empirical approximation to the sampling distribution of the OLS estimator, **for this *X* matrix, and this choice of true parameter values.**

For example, the sample average of the many estimates of the second element of *b* will approximate the true expected value of this estimator, and should therefore be close to the value that we assigned to the second element of β, as this is an unbiased estimator. More generally, any discrepancy between the sample mean and the true parameter would represent any bias in the estimator, *etc.* Note that in a more general problem, our results may depend on the values assigned to the data and to the parameters. Then, we would have to experiment with different choices of these values. If the results do depend on these choices, then this poses a practical problem in using the Monte Carlo results, as we don't know the parameter values in practice. As we'll see later on, this is precisely where a different type of simulation experiment – a Bootstrap analysis – can help us.

Now let's use the EViews workfile, **S:\Social Sciences\Economics\ECON546\Lab1.wf1**, and the program file, **S:\Social Sciences\Economics\ECON546\Lab1.prg** to see how this all works. The program file looks like this to begin with:

```
' MONTE CARLO EXPERIMENT TO VERIFY CERTAIN PROPERTIES OF OLS
rndseed 123456
!n=20
smpl 1 !n
scalar beta1=10
scalar beta2=1
series x2=x^2
scalar sumx=@sum(x)
scalar sumx2=@xum(x2)
next
!nrep=10
vector(!nrep) b2
vector(!nrep) ratio
for !i=1 to !nrep
' NOTE THAT SIGMA = 1
series eps=@rnorm
series y=beta1+beta2*x+eps
ls y c x
b2(!i)=c(2)
ratio(!I)=c(1)/c(2)
next
smpl 1 !nrep
scalar truevarb2=!n/(!n*sumx2-sumx^2)
scalar true_std_dev_b2=@sqrt(truevarb2)
mtos(b2,b2s)
mtos(ratio,ratios)
for !k=1 to 3
series r!k=ratios^!k
scalar m!k=(1/!nrep)*@sum(r!k)
next
'scalar skew=
```

(a) Identify steps (1) – (4) of the above discussion in the code.

(b) Where do the repetitions noted in step (5) begin and end in the code?

(c) How many repetitions are being performed?

(d) Now, open the EViews workfile, then open the program file, and "Run" the latter. Note that various values that are calculated by the program are now stored in the workfile. Notice that the "echoing" of the OLS output to the monitor is both annoying and time-consuming – this can be avoided by changing the command **ls y c x** in the program code to **equation eq1.ls y c x**. Make this change and re-run the program.

(e) Now increase the number of repetitions to 1,000, and re-run the program. Interpret the results relating to the series B2S and TRUE_STD_DEV_B2.

(f) Increase the number of repetitions to 5,000, and re-run the program. Interpret the results.

(g) Reduce the sample size to 10, and increase it to 200. You will see that this OLS estimator is mean-square consistent.

(h) Now focus on the series called "RATIOS". It contains data for the sampling distribution of (b_1 / b_2), where b_i is the OLS estimator of β_i. Keep the sample size at 20, but try different numbers of repetitions of the experiment, up to 20,000.

(i) What is the true value of (β_1 / β_2)? Does the mean of the sampling distribution of (b_1 / b_2) "settle down" to this value as the number of repetitions is increased? Why not? Does the median of this sampling distribution "settle down" to this value? Can you see why the ratio (b_1 / b_2) might be called a "median-unbiased) estimator of (β_1 / β_2)?

(j) Discuss how the sampling distribution of (b_1 / b_2) departs from normality.

(k) Can you verify the value of the skewness associated with this distribution?

Population Moments

r^{th} Central Moment = $E[X^r] = \mu_r'$

Mean = $E[X] = \mu_1'$

r^{th} Moment about Mean = $E[(X-\mu_1')^r] = \mu_r$

$$\begin{aligned}\text{var.} = \mu_2 &= E[(X-\mu_1')^2] = E[X^2 + (\mu_1')^2 - 2X\mu_1'] \\ &= \mu_2' - (\mu_1')^2\end{aligned}$$

$$\begin{aligned}Skew = [\mu_3/(\mu_2)^{3/2}] &= E(X-\mu_1')^3/[E(X-\mu_1')^2]^{3/2} \\ &= [\mu_3' - 3\mu_2'\mu_1' + 2(\mu_1')^3]/[\mu_2' - (\mu_1')^2]^{3/2}\end{aligned}$$

$$\begin{aligned}Kurtosis = [\mu_4/(\mu_2)^2] &= E(X-\mu_1')^4/[E(X-\mu_1')^2]^2 \\ &= [\mu_4' - 4\mu_3'\mu_1' + 6\mu_2'(\mu_1')^2 - 3(\mu_1')^4]/[\mu_2' - (\mu_1')^2]^2\end{aligned}$$

Note that both Skew and Kurtosis are *unit-less*.

For any Normal distribution, *Skew* = 0 and *Kurtosis* = 3.
For the Jarque-Bera test, H_0: *Skew* = 0 and *Kurtosis* = 3 and H_A: Not H_0. There are 2 restrictions under H_0, which is why the JB test statistics is asymptotically $\chi^2_{(2)}$ if H_0 is true.

Sample Moments

r^{th} Sample Central Moment = $\frac{1}{n}\sum_{i=1}^{n} x_i^r = m_r'$

Sample Mean = $\frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x} = m_1'$

r^{th} Sample Moment about Mean = $\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^r = m_r$

$$\begin{aligned}Sample\ \text{var.} = m_2 &= \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2 = \frac{1}{n}\sum_{i=1}^{n}x_i^2 - (\frac{1}{n}\sum_{i=1}^{n}x_i)^2 \\ &= m_2' - (m_1')^2\end{aligned}$$

$$\begin{aligned}Sample\ Skew = [m_3/(m_2)^{3/2}] &= \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^3/[\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2]^{3/2} \\ &= [m_3' - 3m_2'm_1' + 2(m_1')^3]/[m_2' - (m_1')^2]^{3/2}\end{aligned}$$

$$\begin{aligned}Sample\ Kurtosis = [m_4/(m_2)^2] &= \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^4/[\sum_{i=1}^{n}(x_i-\bar{x})^2]^2 \\ &= [m_4' - 4m_3'm_1' + 6m_2'(m_1')^2 - 3(m_1')^4]/[m_2' - (m_1')^2]^2\end{aligned}$$