ECON 3150/4150: Repeated sampling and Monte Carlo simulation

Ragnar Nymoen

6 February 2013

1 Repeated sampling

Section 2.4.3 of the textbook by HGL is called *Repeated sampling.* The point is that by drawing repeated samples of size 40 from a large population of households, they repeat the food expenditure regression several times (they do 10 replications).The average of the 10 OLS estimates of the slope parameter is 9.7.The point is that the average 9.7 is an estimate of the *expectation of the OLS estimator* of the slope coefficient.

2 Monte Carlo simulation and analysis

Repeated sampling is costly in practice, but we can do something similar at a very low cost: We can use the computer to generate a number of data sets (call it M), each of sample length n, and then estimate the regression model on all M data sets and average all of the M estimates of the slope parameter β_1.

The result of the experiment is that we have M OLS estimates $\hat{\beta}_{1j}$ ($j = 1, 2, \ldots, M$). Since the data sets are independent (the computer takes care of that), the average

$$\tilde{\beta}_1 = \frac{\sum_{j=1}^{M} \hat{\beta}_{1j}}{M}.$$

will come close to the true $E(\hat{\beta}_1)$ when M is large. Note that we can make a reference to the *Law of large numbers* (Lecture 6) here: Because the experiment we perform on the computer satisfies the assumptions of the *Law of large numbers* we can use that

$$plim\ \tilde{\beta}_1 = E(\hat{\beta}_1),\ M \longrightarrow \infty$$

This means that by repeated sampling and estimation, we obtain a sequence of estimates $\tilde{\beta}_1$ that has the true expectation of the OLS estimator $\hat{\beta}_1$ as its probability limit. This is the main principle of a Monte Carlo analysis.

3 "Checking" theoretical results by Monte Carlo analysis

HGL mentions Monte Carlo experiment on page 68, as a way of "checking" and getting intuituion about the theoretical statistical results. HGL has also a very instructive appendix 2G about Monte Carlo simulations applied to the simple regression model (the case with fixed/non-random X, our *RM1*).

As an example, the theory of the regression model states that *if* the classical assumptions of the regression model hold, the OLS estimator $\hat{\beta}_1$ is *unbiased*:

$$E(\hat{\beta}_1 - \beta_1) = 0$$

We can use Monte Carlo analysis to check this theoretical result. To do that, we instruct the computer to generate data in accordance with the regression model with classical assumptions.

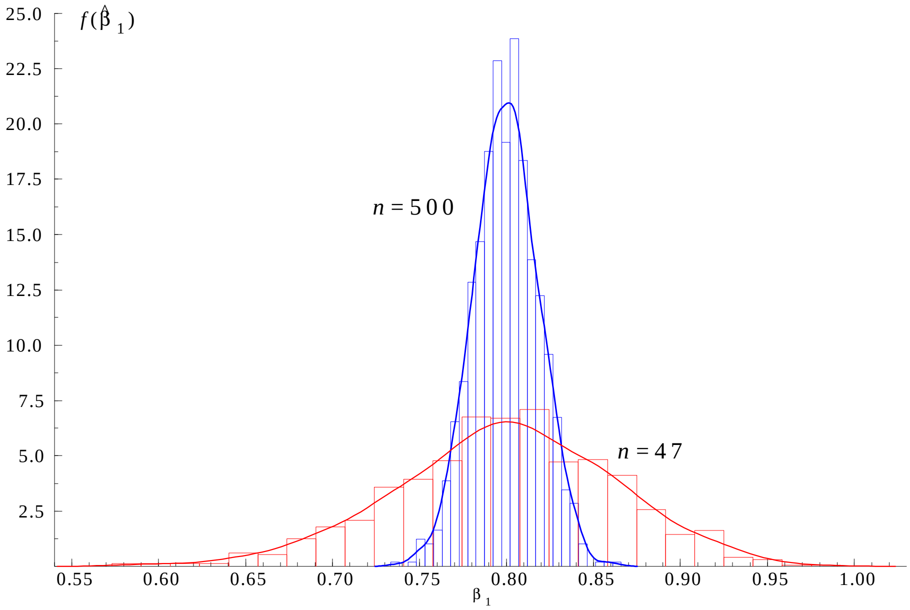
Since the our regression model then is the data generating process ("we have the true model"), we expect to find that $\overset{\sim}{\beta}_1 \approx \beta_1$ for a reasonably large M, (e.g., 20) and that the approximation only gets better if M becomes really large.

In so called *recursive* Monte Carlo analysis we keep the number of replications M fixed and calculate the whole sequence of $\tilde{\beta}_{1(n)}$ estimates (from small n to large n (e.g., $n = 100$), to check that the bias $\tilde{\beta}_{1(n)} - \beta_1$ disappears as the sample size grows larger. A plot of the bias $\tilde{\beta}_{1(n)} - \beta_1$ as a function of n will illustrate the consistency property:

$$plim\ \tilde{\beta}_{1(n)} = \beta_1\ n\ \longrightarrow \infty$$

of the OLS estimator under the classical assumptions of the regression model.

When the number of replications is large, it is also instructive to plot the distributions for different sample sizes:



Monte Carlo analysis: Two distributions (for $n = 47$and $n = 500$) of the OLS estimator $\hat{\hat{\beta}}_1$of the slope coefficient in the case where the classical assumptions of the regression model hold, and $X$is non-random. The true value of $\beta_1$is 0.80.

4 Monte Carlo analysis in the course

There will be a little bit of Monte Carlo analysis for you to do in the exercises to seminar 2 and 4.

In the lectures there will occasionally be plots like the one above (or similar for recursive Monte Carlo which entails plotting distributions for or *all* n, not only $n = 47$). The point will always be to illustrate analytical results (In this case that would be unbiasedness, variance of the estimator declining with n, and suggesting the asymptotic property of consistency).