Chapter 13

Simulation Methods

Simulation Methods in Econometrics and Finance

- **1. The Monte Carlo Method**

This technique is often used in econometrics when the properties of a particular estimation method are not known.

Examples from econometrics include:

1. Quantifying the simultaneous equations bias induced by treating an endogenous variable as exogenous.
2. Determining the appropriate critical values for a Dickey-Fuller test.

Simulations are also often extremely useful tools in finance, in situations such as:

1. The pricing of exotic options, where an analytical pricing formula is unavailable.
2. Determining the effect on financial markets of substantial changes in the macroeconomic environment.
3. "Stress-testing" risk management models to determine whether they generate capital requirements sufficient to cover losses in all situations.

Conducting Simulations Experiments

- In all of the above examples, the basic way that such a study would be conducted (with additional steps and modifications where necessary) is as follows.
  1. **Generate the data according to the desired data generating process (DGP), with the errors being drawn from some given distribution.**
  2. **Do the regression and calculate the test statistic.**
  3. **Save the test statistic or whatever parameter is of interest.**
  4. **Go back to stage 1 and repeat *N* times.**

- *N* is the number of replications, and should be as large as is feasible.
- The central idea behind Monte Carlo is that of **random sampling** from a given distribution.
- Therefore, if the number of replications is set too small, the results will be sensitive to “odd” combinations of random number of draws.

Random Number Generation

- We generate **numbers that are a continuous uniform(0,1)** according to the following recursion:

  $$y_{i+1} = (\boldsymbol{a}\, y_i + \boldsymbol{c}) \text{ modulo } \boldsymbol{m}$$

  then

  $$R_{i+1} = y_{i+1} / m \text{ for } i = 0, 1, \ldots, T$$

  for T random draws, where $\boldsymbol{y_0}$ is the seed (the initial value of y), $\boldsymbol{a}$ is a multiplier and $\boldsymbol{c}$ is an increment.
- An example of a **discrete uniform number generator** would be a die or a roulette wheel.
- Computers generate continuous uniform random number draws. These are pseudo-random numbers.
- The U (0,1) draws can be transformed into other standard distributions (e.g normal, t, etc).

Variance Reduction Techniques

- The sampling variation in a Monte Carlo study is measured by the standard error estimate, denoted S_x:
  $$S_x = \sqrt{\frac{\mathrm{var}(x)}{N}}$$
  where var(x) is the variance of the estimates of the quantity of interest over the N replications
- In order to achieve acceptable accuracy, the number of replications may have to be set at an infeasibly high level.
- An alternative way to reduce **Monte Carlo sampling error** is to use a **variance reduction technique**
- Two of the intuitively simplest and most widely used methods are **antithetic variates** and **control variates**.

Antithetic Variates

- One reason that a lot of replications are typically required of a Monte Carlo study is that it may take many, repeated sets of sampling before the entire probability space is adequately covered.
- What is really required is for successive replications to cover different parts of the probability space.
- The antithetic variate technique involves taking the complement of a set of random numbers and running a parallel simulation on those.
- **For each replication conducted using random draws u_t, an additional replication with errors given by - u_t is also used.**
- Suppose that the average value of the parameter of interest across 2 sets of Monte Carlo replications is given by

$$\bar{x} = (x_1 + x_2)/2$$

where x_1 and x_2 are the average parameter values for replications set 1 and 2 respectively.

Antithetic Variates (cont'd)

- The variance of will be given by:

$$\operatorname{var}(\bar{x}) = \frac{1}{4}\left(\operatorname{var}(x_1) + \operatorname{var}(x_2) + 2\operatorname{cov}(x_1, x_2)\right).$$

- If no antithetic variates are used:

$$\operatorname{var}(\bar{x}) = \frac{1}{4}\left(\operatorname{var}(x_1) + \operatorname{var}(x_2)\right).$$

- However, the use of antithetic variates would lead the covariance to be negative, and therefore the Monte Carlo sampling error to be reduced.

Control Variates

- The application of control variates involves employing a variable similar to that used in the simulation, but whose properties are known prior to the simulation.

- Denote the variable whose properties are known by y, and that whose properties are under simulation by x.

- The simulation is conducted on x and also on y, with the same sets of random number draws being employed in both cases.

Control Variates (cont'd)

- Denoting the simulation estimates of *x* and *y* using hats, a new estimate of *x* can be derived from:

$$x^* = y + (\hat{x} - \hat{y})$$

- Again, it can be shown that the Monte Carlo sampling error of this quantity, *x**, will typically be lower than that of *x*^.

- In fact, it can be shown that control variates will reduce the Monte Carlo variance if

$$Corr(\hat{x}, \hat{y}) > \frac{1}{2}\sqrt{\frac{Var(\hat{y})}{Var(\hat{x})}}$$

- For example, if the topic of interest is the pricing of an Asian option, the control variate approach would be given by

$$P_A^* = (\ \hat{P}_A - \hat{P}_{BS}\ ) + P_{BS}$$

An Example of the use of Monte Carlo Simulation in Econometrics: Deriving a Set of Critical Values for a Dickey-Fuller Test

- Recall, that the equation for a Dickey Fuller test applied to some series y_t is the regression

$$y_t = \phi y_{t-1} + u_t$$

- The test is one of H_0: $\phi = 1$ against H_1: $\phi \leq 1$. The relevant test statistic is given by

$$\tau = \frac{\hat{\phi} - 1}{SE(\hat{\phi})}$$

- Under the null hypothesis of a unit root, the test statistic does not follow a standard distribution, and therefore a simulation would be required to obtain the relevant critical values.

An Example of the use of Monte Carlo Simulation in Econometrics: Deriving a Set of Critical Values for a Dickey-Fuller Test (cont'd)

The simulation would be conducted in the following steps:

- Construct the data generating process under the null hypothesis – that is, obtain a series for y that follows a unit root process. This would be done by
  - Draw a series of length T, the required number of observations, from a normal distribution. This will be the error series, so that $u_t \sim N(0,1)$.
  - Assume a first value for y, i.e. a value for y at time $t = 0$.
  - Construct the series for y recursively, starting with y_1, y_2, and so on:
    - » $y_1 = y_0 + u_1$
    - » $y_2 = y_1 + u_2$
    - » $y_3 = y_2 + u_3$
    - » …
    - » $y_T = y_{T-1} + u_T$

An Example of the use of Monte Carlo Simulation in Econometrics: Deriving a Set of Critical Values for a Dickey-Fuller Test (cont'd)

– Calculate the test statistic, τ.
– Repeat steps 1 and 2 N times to obtain N replications of the experiment. A distribution of values for τ will be obtained across the replications.
– Order the set of N values of τ from the lowest to the highest. The relevant 5% critical value will be the 5th percentile of this distribution.

An Example of how to simulate the Price of a Financial Option

- Although the example presented here is for a simple plain **vanilla call** option, the same basic approach would be used to value an **exotic option**. The steps are:
  - Specify a data generating process for the underlying asset. A random walk with drift model is usually assumed. Specify also the assumed size of the drift component and the assumed size of the volatility parameter. Specify also a strike price K, and a time to maturity, T.
  - Draw a series of length T, the required number of observations for the life of the option, from a normal distribution. This will be the error series, so that $u_t \sim N(0,1)$.
  - Form a series of observations of length T on the underlying asset.

An Example of how to simulate the Price of a Financial Option (cont'd)

  - Observe the price of the underlying asset at maturity observation T. For a call option, if the value of the underlying asset on maturity date, $P_T < K$, the option expires worthless for this replication. If the value of the underlying asset on maturity date, $P_T > K$, the option expires in the money, and has value on that date equal to $P_T - K$, which should be discounted back to the present day using the risk-free rate.
  - **Repeat steps 1 to 4 *N* times and take the average value of the option over the *N* replications. This average will be the price of the option.**
- The data for the underlying can be generated according to a more empirically realistic process.

2. Bootstrapping

- In finance, the bootstrap is often used instead of a pure simulation.
- This is mainly because financial asset returns do not follow the standard statistical distributions that are used in simulations.
- Bootstrapping provides an alternative where by definition, the properties of the artificially constructed series will be similar to those of the actual series.
- Bootstrapping is used to obtain a description of the properties of empirical estimators by using the sample data points themselves.
- **Bootstrapping is similar to pure simulation but the former involves sampling from real data rather than creating new data.**
- Suppose we have a sample of data, $\boldsymbol{y} = y_1, y_2, ..., y_T$ and we want to estimate some parameter θ.
- We can get an approximation to the statistical properties of $\hat{\theta}_T$ by studying a sample of bootstrap estimators.
- We do this by taking N samples of size T with replacement from $\boldsymbol{y}$ and re-calculating $\hat{\theta}$ with each new sample.
- We then get a series of $\hat{\theta}$ estimates.

Bootstrapping (cont'd)

- Advantage of bootstrapping: it allows the researcher to make inferences without making strong distributional assumptions.

- Instead of imposing a shape on the sampling distribution of the $\hat{\theta}$'s, bootstrapping involves empirically estimating the sampling distribution by looking at the variation of the statistic within sample.

- We draw a set of new samples with replacement from the sample and calculate the test statistic of interest from each of these.

- Call the test statistics calculated from the new samples $\hat{\theta}*$.

- The samples are likely to be quite different from each other.
- We thus get a distribution of $\hat{\theta}*$'s.

An Example of the Use of Bootstrapping in a Regression Context

- Consider a standard regression model, $y = \beta X + u$.
- The regression model can be bootstrapped in two ways.

1. Resample the Data

- Take the data, and sample the entire rows corresponding to observation i together.

  The steps would then be

  1. Generate a sample of size T from the original data by sampling with replacement from the whole rows taken together.
  2. Calculate $\hat{\beta}^*$, the coefficient matrix for this bootstrap sample.
  3. Go back to stage 1 and generate another sample of size T. Repeat this a total of N times.

- Problem with this approach is that we are sampling from the regressors.

An Example of the Use of Bootstrapping in a Regression Context (cont'd)

- The only random influence in the regression is the errors, u, so why not just bootstrap from those?

2. Resampling from the Residuals

The steps are

1. Estimate the model on the actual data, obtain the fitted values $\hat{y}$, and calculate the residuals, $\hat{u}$.
2. Take a sample of size T with replacement from these residuals (and call these $\hat{u}^*$), and generate a bootstrapped dependent variable:
$$y^* = \hat{y} + \hat{u}^*$$
3. Then regress the original X data on this new dependent variable to get a bootstrapped coefficient vector, $\hat{\beta}^*$.
4. Go back to stage 2, and repeat a total of N times.

Situations where the Bootstrap will be Ineffective

- If there are **extreme outliers** in the data, the conclusions of the bootstrap may be affected.

- Use of the bootstrap implicitly assumes that the **data from which the sampling is done are independent**. This would not hold, for example, if the data were auto-correlated.

Problems with Simulation (including bootstrapping)

1. It might be **computationally expensive**
2. The results might **not be precise**
3. The results are often **hard to replicate**
4. Simulation results are **experiment-specific**

- To conclude, simulation is an extremely useful tool that can be applied to an enormous variety of problems. The technique has grown in popularity over the past decade, and continues to do so. However, like all tools, it is dangerous in the wrong hands.