A Computer Program to Run a Monte Carlo Experiment: A Dickey-Fuller Distribution

Un programa de computadora para ejecutar un experimento de Monte Carlo: una Distribución Dickey-Fuller

Carlos Guerrero de Lizardi*

> "But with this miraculous development of the ENIAC—along with the applications Stan must have been pondering—it occurred to him that statistical techniques should be resuscitated, and he discussed this idea with von Neumann. Thus was triggered the spark that led to the Monte Carlo method." Nicholas Metropolis (1987, p. 126).

> "The obvious implications of these results are that applied econometricians should not worry about spurious regressions only when dealing with I(1), unit root, processes. Thus, a strategy of first testing if a series contains a unit root before entering into a regression is not relevant". Clive W. J. Granger (2003, p. 560).

Resumen

Se presenta un programa para realizar un experimento de Monte Carlo. Como ejemplo se utiliza una distribución de Dickey-Fuller. Al evitar el uso de matrices el código propuesto es más fácil de ejecutar que el diseñado por, entre otros, Brooks (2002) o Fantazzini (2007). Se presentan algunas notas respecto a la técnica de Monte Carlo y sobre las pruebas de raíces unitarias. Al final se comparan los valores críticos obtenidos con los reportados por Brooks (2002), Charemza and Deadman (1992), Enders (2004), y Patterson (2000).

Palabras clave:

- Trasporte urbano
- Condiciones laborales
- Velocidad, Pasajeros
- Competencia, Accidentes

Abstract

We present a computer program to run a Monte Carlo experiment. We use as example a Dickey-Fuller distribution. Avoiding the use of matrices, the proposed program is easier to put into practice than the code designed by, among others, Brooks (2002) or Fantazzini (2007). Some remarks about the Monte Carlo method and unit root tests are included. At the end we compare our critical values with the ones in Brooks (2002), Charemza and Deadman (1992), Enders (2004), and Patterson (2000).

Keywords:

- Urban Transportation
- Working Conditions
- Speed, Passengers
- Competition, Accidents

JEL: C15, C22, C87

1. Unit roots always cause trouble

It came as a bit of shock when econometricians realized that the "*t*" and the Durbin-Watson statistics did not retain its traditional characteristics in the presence on nonstationary data, i.e. regressions involving unit root process may give non-sense results. Following Bierens (2003), it is correct to say that, if y_t and x_t are mutually independent unit root processes, i.e. y_t is independent of x_{t-j} for all t and j, then OLS regression of y_t on x_t for t=1,…,n, with or without an intercept, will yield a significant estimate of the slope parameter

* Facultad de Economía de la Universidad Nacional Autónoma de México, cguerrero@economia.unam.mx.

if n is large: the absolute value of the t-value of the slope converges in probability to ∞ if $n \to \infty$. We then might conclude that y_t depends on x_t, while in reality the $y_t s$ are independent of the $x_t s$. Phillips (1986) was able to show that, in such a case, *DW* statistic tends to zero. Hence, adding lagged dependent and independent variables would make the misspecified problem worse. By the way, the first simulation on the topic was by Granger and Newbold (1974). They generated two random walks, each one had only 50 terms and 100 repetitions were used! In this sense (Granger, 2003, p. 559), "it seems that spurious regression occurs at all sample sizes."

How does one test for non-stationarity? In first place a variable is said to be integrated of order d, written I(d), if it must be differenced d times to be made stationary. Thus a stationary variable is integrated of order zero, written I(0), a variable which must be differenced once to become stationary is said to be I(1) integrated of order one, and so on. Economic variables, which include financial ones, are seldom integrated of order greater than two.

Consider the simplest example of an I(1) variable, a random walk without drift. Let $y_t = y_{t-1} + e_t$, where e_t is a stationary error term, i.e. , e_t is I(0). Here y_t can be seen to be I(1) because $\Delta y_t = e_t$, which is I(0) Now let this relationship be expressed as $y_t = \rho * y_{t-1} + e_t$. If $|\rho|<1$, then y is I(0) i.e., stationary, but if $\rho=1$ then y_t is I(1), i.e., nonstationary. In this sense, typically formal tests of stationarity are test for $\rho=1$, and because of this are referred to as tests for a unit root. By the way, the case of $|\rho|>1$ is ruled out as being unreasonable because it would cause the series y_t to explode. In other words, for an I(2) process the remote past is more influential that the recent one, which makes little sense.

In terms of our economics common-sense, the differences between a stationary, or "short memory" variable, and an I(1) or "long memory" one, are clues:

1. A stationary time series has a mean and there is a tendency for the series to return to that mean, whereas an integrated one tends to wander "widely".
2. Stationary variables tend to be "erratic", whereas integrated variables tend to exhibit some sort of smooth behavior (because of its trend).
3. A stationary variable has a finite variance, shocks are transitory, and its autocorrelations ρk die out as k grows, whereas an integrated series has an infinite variance, i.e. it grows over time, shocks are permanent, and its autocorrelations tend to one (Patterson, 2000).