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 γ_t depends on x_t, while in reality the $\gamma_t s$ 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 γ_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 γ is I(0) i.e., stationary, but if $\rho=1$ then γ_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 γ_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).

2. Monte Carlo simulations

A Monte Carlo experiment attempts to replicate an actual data-generating process (DGP). The process is repeated numerous times so that the distribution of the desired parameters and sample statistics can be tabulated. Its reliability is warranted by the Law of Large Numbers: as the sample size grows sufficiently large, the sample statistic converges to the true one. Thus, the sample statistic is an unbiased estimate of the population one. The beauty of the simulation is that attributes of the constructed series are known to the researcher. It is well-known that a limitation of a Monte Carlo experiment is that the results are specific to the assumptions used to generate the simulated data. For example, if you modify the sample size, include or delete an additional parameter, or use an alternative initial condition, a new simulation needs to be performed.[1]

3. An Eviews program to run a Monte Carlo experiment: a Dickey-Fuller distribution

In order to generate a Dickey-Fuller distribution using a Monte Carlo approach, it is necessary to follow four steps:

1. Generate a sequence of (seudo) random numbers e_t based on a standard normal distribution.[2]
2. Generate the sequence $y_t = \rho * y_{t-1} + e_t$ (eq. 1), where $\rho = 1$. With the intention of minimize the influence of y_o, its value is fixed to zero and T = 500.
3. Estimate the model $\Delta y_t = \gamma * y_{t-1} + e_t$ (eq. 2), where $\gamma = \rho - 1$. Following Dickey and Fuller (1979, 1981), the "*t*"-statistic will be recorded as . Obtain its distribution is our goal. According to Patterson (2000, p. 228), even though $\gamma = 0$ by construction, "the presence of e_t a random disturbance term, will prevent us from reaching this conclusion with certainty from a particular dataset", that is, there will be a distribution of the with nonzero values occurring; if the estimation method is unbiased then this should be picked up if *T* is large enough "by an average of over the *T* samples equal to the value in the DGP". About eq. 2 Charemza and Deadman (1997, p. 99) remind us that, because it is a regression of an I(0) variable on a I(1)

1 "The term is reported to have originated with Metropolis and Ulam (1949). If it had been coined a little later, it might have been called the 'Las Vegas method' instead of the 'Monte Carlo method'" (Davidson and MacKinnon, 1993, p. 732). According to Dufour and Khalaf (2003), early uses in econometrics of Monte Carlo test techniques were Dwass (1957), Barnard (1963) and Birnbaum (1974).

2 Its mechanics is fully explained in Davidson and MacKinnon (1993).

variable, "not surprisingly, in such a case the *t*-ratio does not have a limiting normal distribution."

4. Repeat steps 1 to 3. By the way, Dickey and Fuller (1979, 1981) obtained 100 values for e_t, set $\gamma = 1$, $y_0 = 0$ and calculated, accordingly, 100 values for y_t.

The Eviews program to run the experiment is the following:

```
'Create a workfile undated, range 1 to 500.
!reps = 50000
for !i=1 to !reps
genr perturbacion{!i}=@nrnd
smpl 1 1
genr y{!i}=0
smpl 2 500
genr y{!i}=y{!i}(-1)+perturbacion{!i}
smpl 1 500
matrix(!reps,2) resultados
equation eq{!i}.ls D(y{!i})=c(1)*y{!i}(-1)
resultados(!i,1)=eq{!i}.@coefs(1)
resultados(!i,2)=eq{!i}.@tstats(1)
d perturbacion{!i}
d y{!i}
d eq{!i}
NEXT
'Export "resultados" to Excel.
'Create a workfile undated, range 1 to 50,000.
'Copy and paste from Excel to the workfile.
```

The critical values depend on the specification of the null and alternative hypotheses. The H_o: $\gamma = 0$ implying $y_t = \rho * y_{t-1} + e_t$ with $\rho = 1$, that is, $y_t \sim I(1)$. The alternative "should be chosen to maximize the power of the test in the likely direction of departure from the null. A two-sided alternative $\gamma \neq 0$, comprising $\gamma > 0$ and $\gamma < 0$ is *not* chosen in general because $\gamma > 0$ corresponds to $\rho > 1$ and in that case the process generating y_t is not stable; instead the one-side alternative H_a: $\gamma < 0$, that is $\rho < 1$, is chosen because departures from the null are expected to be in this direction corresponding to an I(0) process. Thus the critical values are *negative*, with sample values more negative than

the critical values leading to rejection of the null hypothesis in the direction of the one-side alternative" (Patterson 2000, p. 228). The empirical distribution of our Dickey-Fuller statistic and its descriptive statistics are shown in the following figures:

Figure 1
Histogram and estimated densities of the Dickey-Fuller statistic

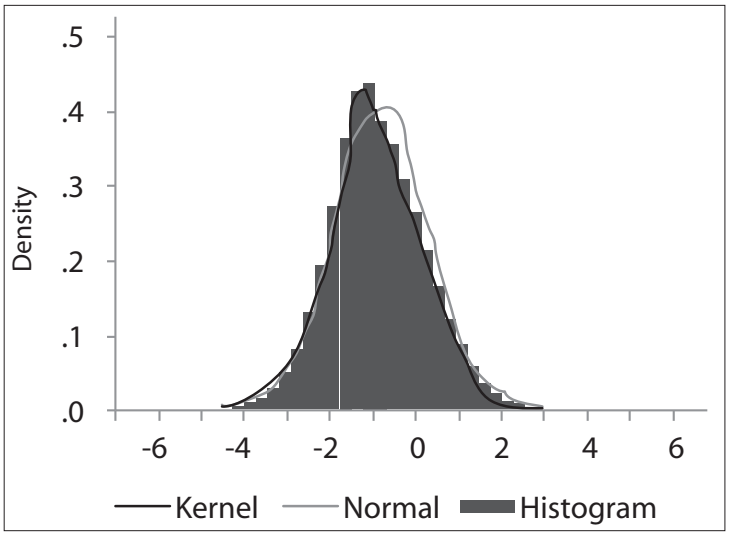
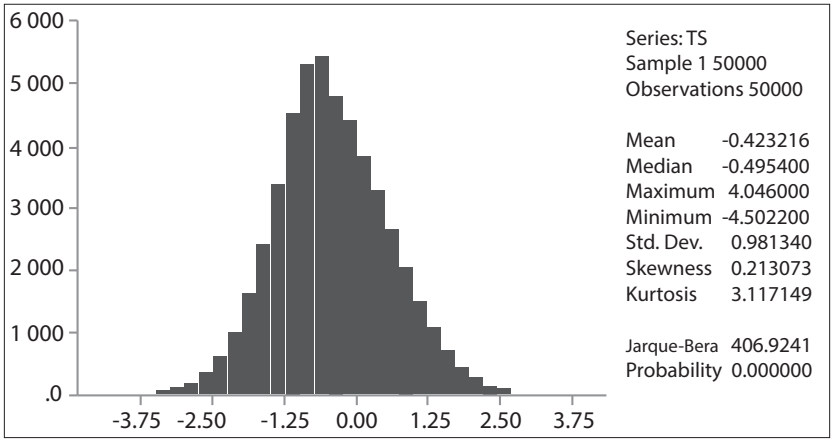


Figure 2
Descriptive statistics



It is clear that the simulated distribution is not like that of the *t*-distribution, which is symmetric and centered at zero. In Excel we sort $\hat{\tau}$ from the highest values to the lowest ones. The value of -1.9359 is the average between the 2500th and 2501st lowest values in the 50,000 replications, and may be regarded as the critical value at the level of significance of 5%.

As a final point, in the following table we compare our results with those of Brooks (2002), Charemza y Deadman (1992), Enders (2004), and Patterson (2000).

Table 1. Critical values, Dickey-Fuller distribution

	Sample size (T)	Replications	
Brooks (2002)	1000	50,000	-1.950
Charemza y Deadman (1992)	50	50,000	-1.949
Enders (2004)	100	10,000	-2.890
Patterson (2000)	500	25,000	-1.943

4. Final comments

It is an undeniable true the 3th law in econometrics proposed by Phillips (2003, p. 8), which I borrowed it as the title of the first section: "units roots always cause trouble". At the moment, you can find not only a good number of papers that propose unit root tests, but also on testing strategies (Perron 1988, Dolado, Jenkinson, and Sosvilla-Rivero 1990, Holden and Perman 1994, Enders 1995, Ayat and Burridge 2000, and Elder and Kennedy 2001). Following Bierens (2003), I recommend the two most frequently applied types of unit root tests, the Augmented Dickey Fuller and the Phillips-Perron tests, and the strategy proposed by Dolado et al. (1990).

Indeed Monte Carlo simulations "have revolutionized the way we approach statistical analysis" (Dufour and Khalaf, 2003, p. 494). We hope that the proposed program, easy to run, and the relevance of the used example, the Dickey-Fuller distribution, both serve as an introduction to the quoted literature.

References

Ayat, L., and P. Burridge (2000), "Unit root tests in the presence of uncertainty about the non-stochastic trend", *Journal of Econometrics*, 95:1, pp.71-96.

Barnard, G. A. (1963), "Comment on: The spectral analysis of point processes", by M. S. Bartlett. *Journal of the Royal Statistical Society*, Series B 25, 294.

Bierens, H. J. (2003). "Unit roots", *A companion to Theoretical Econometrics*, edited by Badi H. Baltagi, Blackwell Publishing.

Birnbaum, Z. W. (1974), "Computers and unconventional test-statistics", *Reliability and Biometry*, F. Proschan, and R. J. Serfling (eds.), SIAM.

Brooks, C. (2002). *Introductory Econometrics for Finance*, Cambridge University Press.

Charemza, W. W. and D. F. Deadman (1999), *New Directions in Econometric Practice*, Edward Elgar.

Davidson, R. and J. G. MacKinnon (1993), *Estimation and Inference in Econometrics*, Oxford University Press.

Dickey D. and W. A. Fuller (1981), "Likelihood ratio statistics for autoregressive time series with a unit root", *Econometrica* 49, July, pp. 1057-72.

Dickey, D. and W. A. Fuller (1979), "Distribution of the estimates for autoregressive time series with a unit root", *Journal of the American Statistical Association* 72, June, pp. 427-31.

Dolado, J., T. Jenkinson, and S. Sosvilla-Rivero (1990), "Cointegration and unit roots", *Journal of Economic Surveys*, 4: 3, pp. 249-73.

Dufour, J-M. and L. Khalaf (2003), "Monte Carlo test methods in econometrics", *A companion to Theoretical Econometrics*, edited by Badi H. Baltagi, Blackwell Publishing.

Dwass, M. (1957). "Modified randomization tests for nonparametric hypotheses", *Annals of Mathematical Statistics*, 28, pp. 181-7.

Elder, J. and P. E. Kennedy, (2001), "Testing for unit roots: what should students be taught?", *The Journal of Economic Education*, 32:2, pp. 137-46.

Enders, W. (2004), *Applied Econometric Time Series*, John Wiley&Sons.

Fantazzini, D. (2007), *Econometrics I:* OLS, University of Pavia, http://economia.unipv.it/pagp/pagine_personali/dean/slides%20E1_1_final.pdf.

Granger, C. W. J. (2003). "Spurious regressions in econometrics", *A Companion to Theoretical Econometrics*, edited by Badi H. Baltagi, Blackwell Publishing.

Granger, C. W. J., and P. Newbold (1974). "Spurious regression in econometrics", *Journal of Econometrics*, 2, pp. 111-20.

Holden, D., and R. Perman (1994), "Unit roots and cointegration for the economist", *Cointegration for the Applied Economist*, edited by B. B. Rao, pp. 47-112. St. Martin's Press.

Kennedy, P. (2003), *A Guide to Econometrics*, fifth edition, The MIT Press.

Metropolis, N. (1987), *Los Alamos Science*, Special Issue: Stanislaw Ulam 1909-1984, pp. 125-30.

Patterson, K. (2000), *An Introduction to Applied Econometrics: a Time Series Approach*, Palgrave.

Perron, P. (1988), "Trends and random walks in macroeconomic time series", *Journal of Economic Dynamics and Control*, 12:12, pp. 297-332.

Phillips, P. C. B. (2003), "Law and limits of econometrics", *Cowles Foundation Discussion Paper*, núm. 1397.