

PRUEBAS DE BONDAD DE AJUSTE PARA EL MOVIMIENTO BROWNIANO

GOODNESS OF FIT TEST FOR BROWNIAN MOVEMENT

José A. Villaseñor-Alva y Elizabeth González-Estrada

Estadística. Campus Montecillo. Colegio de Postgraduados. 56230. Montecillo, Estado de México.
(jvillasr@colpos.mx)

RESUMEN

En el análisis de series de datos financieros que ocurren en el tiempo, es importante verificar la suposición de que los datos pueden modelarse por un proceso de movimiento Browniano con el propósito de calcular el precio de instrumentos financieros de riesgo, llamados derivados. Con base en las propiedades del proceso Browniano, se proponen tres pruebas de bondad de ajuste del tipo unión-intersección. Se realizó un estudio de simulación de Monte Carlo para comparar las potencias de las pruebas para algunas hipótesis alternativas dadas de procesos no Brownianos, y diferentes tamaños de series de datos. Los resultados obtenidos muestran que, en general, las tres pruebas propuestas preservan el tamaño fijado; la prueba basada en la combinación de rachas, Shapiro-Wilk y t resultó ser en general la más potente respecto a las alternativas estudiadas y para casi todos los tamaños de series de datos considerados. Se hicieron aplicaciones a conjuntos de datos financieros.

Palabras clave: Aleatoriedad, datos financieros, procesos Gaussianos, pruebas estadísticas.

INTRODUCCIÓN

Un modelo muy usual en la teoría de procesos estocásticos es el movimiento Browniano. Un proceso estocástico $\{X_t, t \geq 0\}$ es un movimiento Browniano si para $0 < s < t$ la variable aleatoria (va) $X_t - X_s$ tiene distribución $N(0, \sigma^2(t-s))$, y es independiente de X_u , $0 < u < s < t$.

En las aplicaciones de este modelo, por ejemplo en el área de finanzas, es de interés contar con metodología estadística que permita probar si una realización de un número finito de datos obtenidos en el tiempo soporta la hipótesis de movimiento Browniano.

En este trabajo se proponen tres pruebas de bondad de ajuste para probar la hipótesis de que una realización de tamaño n de un proceso estocástico $\{X_t, t \geq 0\}$ sigue un movimiento Browniano. Estas pruebas se comparan mediante sus funciones de potencia por medio de simulación de Monte Carlo bajo algunas hipótesis alternativas de interés y diferentes valores de n . Las pruebas se aplican a conjuntos de datos reales.

ABSTRACT

In time series analyses of financial data it is important to verify the assumption that the data can be modeled by a Brownian movement process to calculate the price of the risky financial instruments called derivatives. Based on the properties of the Brownian process, three union-intersection-type goodness-of-fit tests are proposed. A Monte-Carlo simulation study was conducted to compare the power of the tests for some given alternative hypotheses of non-Brownian processes and different sizes of data series. The findings show that, in general, the three proposed tests preserve the assigned size; the test based on a combination of runs, Shapiro-Wilk and t, showed to be, in general, the most powerful against the studied alternatives and for almost all the data series sizes considered. Applications were performed on financial data sets.

Key words: randomness, financial data, Gaussian processes, statistical tests.

INTRODUCTION

A very usual model in the theory of stochastic processes is Brownian movement. A stochastic process $\{X_t, t \geq 0\}$ is a Brownian movement if, for $0 < s < t$, the random variable (va) $X_t - X_s$ has distribution $N(0, \sigma^2(t-s))$, and is independent of X_u , $0 < u < s < t$.

In the applications of this model, for example in the area of finance, it is of interest to have a statistical methodology that would prove whether a realization of a finite number of data obtained over time supports the hypothesis of Brownian movement.

In this study three goodness-of-fit tests are proposed to test the hypothesis that an n -size realization of a stochastic process $\{X_t, t \geq 0\}$ follows a Brownian movement. These tests are compared by their power functions through Monte Carlo simulations under some alternative hypotheses of interest and different n values. These tests are applied to sets of real data.

When discrete realizations are considered, Brownian movement is also called random walk, with independent increments, identically distributed, normal, and with mean and constant variance.

En realizaciones discretas, el movimiento Browniano es llamado caminata aleatoria con incrementos independientes e idénticamente distribuidos, normales, con media y varianza constantes.

En series de tiempo existe interés en probar la hipótesis nula de que un proceso dado es una caminata aleatoria, contra la alternativa de que el proceso sigue un modelo ARMA. La hipótesis nula también es llamada de la raíz unitaria. En esta situación, si la hipótesis nula se rechaza, hay evidencia de que el proceso es de covarianza estacionaria.

Este problema ha sido estudiado por Hall (1994), Leyburne y McCabe (1994), Ng y Perron (1995), Lumsdaine y Papell (1997), Elliot *et al.* (1996), Elliott (1999), Elliot y Janssonb (2003).

Una de las pruebas que han recibido gran atención en Econometría es la de Dickey y Fuller (1981), la cual es una prueba de razón de verosimilitudes generalizada para probar la hipótesis de raíz unitaria contra la alternativa de que el modelo que siguen los datos es un proceso autorregresivo de orden p , bajo el supuesto de que el proceso de interés es un proceso Gaussiano. En esta prueba se tiene una hipótesis alternativa especificada, por lo cual no se trata de una prueba de bondad de ajuste.

MATERIALES Y MÉTODOS

Supóngase que la sucesión de $X_1, X_2, \dots$ sigue un movimiento Browniano. Entonces, de acuerdo con su definición, las variables aleatorias (va) $X_t - X_{t-1}$ y $X_s - X_{s-1}$, para toda $s \leq t-1$, son iid $N(0, \sigma^2)$. Es decir, si se define $e_t = X_t - X_{t-1}$, las v.a.'s X_t , $t=1, 2, \dots$, admiten la representación:

$$X_t = X_{t-1} - e_t \text{ iid } N(0, \sigma^2) \quad (1)$$

Por tanto, el modelo (1) para X_t es otra representación del movimiento Browniano, el cual también es conocido como caminata aleatoria con incrementos independientes e idénticamente distribuidos $N(0, \sigma^2)$.

Son dos las versiones más usadas de la caminata aleatoria (modelo 1) en las que no se requiere la condición de normalidad. Una es la caminata aleatoria donde las va $\{e_t\}$ son independientes, y otra, aún más general, cuando las $\{e_t\}$ no están correlacionadas.

Existen varias investigaciones en las que se estudia el problema de probar si una sucesión finita de variables aleatorias sigue una caminata aleatoria. Sin embargo, la mayor parte de ellas consideran caminatas aleatorias con incrementos independientes o no correlacionados, lo que no es equivalente a probar que se tiene un movimiento Browniano.

Un proceso estocástico $\{X_t, t \in T\}$ es una colección de variables aleatorias indexada por un parámetro de tiempo (t) que toma valores en un conjunto T . Si $T = \{t : -\infty < t < \infty\}$, se dice que $\{X_t, t \in T\}$ es un proceso estocástico con parámetro de tiempo continuo.

In time series, there is interest in proving the null hypothesis that the given process is a random walk, against the alternative that the process follows an ARMA model. The null hypothesis is also called the unit root. In this situation, if the null hypothesis is rejected, there is evidence that the process is one of stationary covariance.

This problem has been studied by Hall (1994), Leyburne and McCabe (1994), Ng and Perron (1995), Lumsdaine and Papell (1997), Elliot *et al.* (1996), Elliot (1999), Elliot and Janssonb (2003).

One of the tests that have received much attention in econometrics is that of Dicker and Fuller (1981), a generalized likelihood ratio test to prove the unit root hypothesis against the alternative that the model that the data follow is an autoregressive process of p order, under the assumption that the process of interest is Gaussian. In this test a specific alternative hypothesis exists, thus it is not a goodness-of-fit test.

MATERIALS AND METHODS

Assume that the succession of $X_1, X_2, \dots$ follows a Brownian movement. Then, according to its definition, the random variables (va) $X_t - X_{t-1}$ and $X_s - X_{s-1}$, for all $s \leq t-1$, are iid $N(0, \sigma^2)$; that is, if $e_t = X_t - X_{t-1}$, the va X_t , $t=1, 2, \dots$, admit the representation:

$$X_t = X_{t-1} - e_t \text{ iid } N(0, \sigma^2) \quad (1)$$

Therefore, model (1) for X_t is another representation of Brownian movement, which is also known as random walk with independent, identically distributed increments $N(0, \sigma^2)$.

Two are the most used versions of the random walk (model 1); with these the condition of normality is not required. One is the random walk in which the va $\{e_t\}$ are independent and the other, even more general, when the $\{e_t\}$ are not correlated.

There are several studies in which the problem of proving whether a finite succession of random variables follows a random walk. Most of them, however, consider random walks with independent or non-correlated increments; this is not equivalent to proving that we are dealing with a Brownian movement.

A stochastic process $\{X_t, t \in T\}$ is a collection of random variables indexed by a time (t) parameter that takes values from a set T . If $T = \{t : -\infty < t < \infty\}$, it is said that $\{X_t, t \in T\}$ is a stochastic process with a continuous time parameter.

For a set of time values $\{t_0, t_1, \dots, t_n\}$, the collection of va $X_{t_0}, X_{t_1}, \dots, X_{t_n}$ is called a sample trajectory of $n+1$ size of the stochastic process $\{X_t, t \in T\}$.

Below are presented the definitions of some stochastic processes closely related to Brownian movement.

It is said that the stochastic process $\{X_t, t \in T\}$ has independent increments if, for any times $t_0 < t_1 < \dots < t_m$, the m va $X_{t_1} - X_{t_0}, \dots, X_{t_m} - X_{t_{m-1}}$ are independent and that the process has stationary increments if the va $X_{t+h} - X_{s+h}$ and $X_t - X_s$ have the

Para un conjunto $\{t_0, t_1, \dots, t_n\}$ de valores en el tiempo, la colección de $X_{t_0}, X_{t_1}, \dots, X_{t_n}$ es llamada una trayectoria muestral de tamaño $n+1$ del proceso estocástico $\{X_t, t \in T\}$.

A continuación se presentan las definiciones de algunos procesos estocásticos estrechamente relacionados con el movimiento Browniano.

Se dice que el proceso estocástico $\{X_t, t \in T\}$ tiene incrementos independientes si para cualesquiera tiempos $t_0 < t_1 < \dots < t_m$ las m va $X_{t_1} - X_{t_0}, \dots, X_{t_m} - X_{t_{m-1}}$ son independientes y se dice que el proceso tiene incrementos estacionarios si las va $X_{t+h} - X_s$ y $X_t - X_s$ tienen la misma distribución, para toda $h > 0$ y $0 < s < t$. Es decir que $\{X_t, t \in T\}$ tiene incrementos independientes si los cambios en el valor del proceso sobre intervalos de tiempo que no se traslapan son independientes, y tiene incrementos estacionarios si la distribución del cambio en el valor del proceso entre cualesquiera dos puntos del tiempo depende sólo de la distancia entre ellos.

Una sucesión de variables aleatorias $\{X_t\}_{t=1}^n$, $n \geq 1$, es una martingala si y sólo si $E\{X_t\} < \infty$ y $E\{X_{t+1} | X_t, \dots, X_1\} = X_t$.

Se dice que un proceso estocástico $\{X_t, t \geq 0\}$ es Gaussiano si para cualquier n y cualquier subconjunto $\{t_1, t_2, \dots, t_n\}$ de valores positivos, las va $X_{t_1}, X_{t_2}, \dots, X_{t_n}$ tienen distribución conjunta normal.

Un proceso estocástico $\{X_t, t \geq 0\}$ es un movimiento Browniano si es tal que para $X_0 = 0$,

- $\{X_t, t \geq 0\}$ tiene incrementos independientes y estacionarios,
- $X_t - X_s$ tiene distribución $N(0, \sigma^2(t-s))$, $0 < s < t$.

El movimiento Browniano es un proceso estocástico con parámetro de tiempo continuo, $T = \{t : 0 \leq t < \infty\}$, y espacio de estados continuo, $S = \{s : -\infty \leq s < \infty\}$. Es posible verificar que un movimiento Browniano es un proceso Gaussiano y una martingala. Ross (1996) hace una revisión detallada de las propiedades del proceso Browniano.

La siguiente caracterización del movimiento Browniano es la base de la metodología que se propone en este trabajo.

Teorema: Sean $X_1, X_2, \dots, X_n$ observaciones del proceso estocástico $\{X_t, t \geq 0\}$ registradas en tiempos equidistantes. Las va $X_1, X_2, \dots, X_n$ siguen un movimiento Browniano si y sólo si las va $e_1 = X_1 - X_0, \dots, e_n = X_n - X_{n-1}$ son iid $N(0, \sigma^2)$.

Note que las va $e_1, e_2, \dots, e_n$ son una trayectoria muestral del proceso de primeras diferencias de $\{X_t, t \geq 0\}$, denotado como $\{e_t, t \geq 0\}$, donde $e_t = X_t - X_{t-1}$.

Por el teorema anterior, probar que las va $X_1, X_2, \dots, X_n$ siguen un movimiento Browniano es equivalente a probar que las va $e_1 = X_1 - X_0, \dots, e_n = X_n - X_{n-1}$ son iid $N(0, \sigma^2)$. Esto es, se propone probar por separado las hipótesis de aleatoriedad (iid), normalidad y media cero sobre las primeras diferencias de las va $X_1, X_2, \dots, X_n$. Para ello, se consideran la prueba de rachas y la de McLeod-Li para la hipótesis de aleatoriedad (iid), la de Royston (Shapiro-Wilk) y la de Kolmogorov-Smirnov para normalidad y la prueba de t para media cero. A continuación se presenta una breve descripción de estas pruebas.

same distribution for all $h > 0$ and $0 < s < t$. That is, $\{X_t, t \in T\}$ has independent increments if the changes in the value of the process over non-overlapping time intervals are independent, and it has stationary increments if the distribution of the change in the value of the process between any two points of time depends only on the distance between them.

A succession of random variables $\{X_t\}_{t=1}^n$, $n \geq 1$, is a martingale if, and only if, $E\{X_t\} < \infty$ and $E\{X_{t+1} | X_t, \dots, X_1\} = X_t$.

A stochastic process $\{X_t, t \geq 0\}$ is Gaussian if, for any n and any subset $\{t_1, t_2, \dots, t_n\}$ of positive values, the va $X_{t_1}, X_{t_2}, \dots, X_{t_n}$ have a normal joint distribution.

A stochastic process $\{X_t, t \geq 0\}$ is a Brownian movement if $X_0 = 0$.

- $\{X_t, t \geq 0\}$ has independent and stationary increments,
- $X_t - X_s$ has $N(0, \sigma^2(t-s))$, $0 < s < t$ distribution.

Brownian movement is a stochastic process with continuous time parameter, $T = \{t : 0 \leq t < \infty\}$, and continuous stage spaces, $S = \{s : -\infty \leq s < \infty\}$. It is possible to verify whether or not a Brownian movement is a Gaussian process and a martingale. Ross (1996) reviewed in detail the properties of the Brownian process.

The following characterization of Brownian movement is the basis of the methodology that is proposed in this paper.

Theorem: Let $X_1, X_2, \dots, X_n$ be observations of the stochastic process $\{X_t, t \geq 0\}$ recorded at equidistant times. The va $X_1, X_2, \dots, X_n$ follow a Brownian movement if and only if the va $e_1 = X_1 - X_0, \dots, e_n = X_n - X_{n-1}$ are iid $N(0, \sigma^2)$.

Note that the va $e_1, e_2, \dots, e_n$ are a sample trajectory of the process of the first differences of $\{X_t, t \geq 0\}$ denoted as $\{e_t, t \geq 0\}$, where $e_t = X_t - X_{t-1}$.

With the above theorem, proving that the va $X_1, X_2, \dots, X_n$ follow a Brownian movement is equivalent to proving that the va $e_1 = X_1 - X_0, \dots, e_n = X_n - X_{n-1}$ are iid $N(0, \sigma^2)$. That is, it is proposed that the randomness hypothesis (iid), normality and mean zero over the first differences of the va $X_1, X_2, \dots, X_n$ be proved separately. To this end, the runs and the McLeod-Li tests are considered for the randomness hypothesis (iid), the Royston (Shapiro-Wilk) and the Kolmogorov-Smirnov test for normality, and the t test for zero mean. Below, a brief description of these tests is presented.

Runs test

The runs test is the most powerful for testing the randomness hypothesis against any alternative that is a stationary Markov process (Lehmann, 1997, p. 177). Campbell *et al.* (1997) and Taylor (1986), among others, mention that the runs test is commonly used in financial econometrics to prove randomness.

Let $\{e_1, e_2, \dots, e_n\}$ be an n -size sample trajectory of the stochastic process $\{e_t, t \geq 0\}$. There is interest in proving the null hypothesis that the va $e_1, e_2, \dots, e_n$ are iid, against the alternative that they are not.

La prueba de rachas

La prueba de rachas es la más potente para probar la hipótesis de aleatoriedad contra cualquier alternativa que es un proceso de Markov estacionario (Lehmann, 1997, p. 177). Campbell, *et al.* (1997) y Taylor (1986), entre otros, mencionan que la prueba de rachas se usa comúnmente en la econometría financiera para probar aleatoriedad.

Sea $\{e_1, e_2, \dots, e_n\}$ una trayectoria muestral de tamaño n del proceso estocástico $\{e_t, t \geq 0\}$. Se tiene interés en probar la hipótesis nula de que las $e_1, e_2, \dots, e_n$ son iid, contra la alternativa de que no lo son. Para esto, una posibilidad es utilizar el número total de rachas en la secuencia $E_1, E_2, \dots, E_n$, donde $E_i = \begin{cases} 1, & \text{si } e_i < 0 \\ 0, & \text{si } e_i \geq 0 \end{cases}$, $i=1, 2, \dots, n$, y una racha se define como una sucesión de uno o más símbolos idénticos, los cuales son seguidos y precedidos por uno diferente.

Note que $E_1, E_2, \dots, E_n$ es una secuencia ordenada en el tiempo de elementos de dos tipos (0 o 1), y que si la secuencia es aleatoria se espera que no haya muchas ni pocas rachas. Sea R la va que denota el número total de rachas. Bajo la hipótesis nula, la distribución de probabilidad de R es:

$$f_R(r) = P(R=r \mid H_0 \text{ es cierta})$$

$$= \begin{cases} \frac{2 \binom{n_1-1}{r/2-1} \binom{n_2-1}{r/2-1}}{\binom{n_1+n_2}{n_1}} & \text{si } r \text{ es par} \\ \frac{\binom{n_1-1}{\frac{r-1}{2}} \binom{n_2-1}{\frac{r-3}{2}} + \binom{n_1-1}{\frac{r-3}{2}} \binom{n_2-1}{\frac{r-1}{2}}}{\binom{n_1+n_2}{n_1}} & \text{si } r \text{ es impar} \end{cases}$$

donde r es el número total de rachas en la secuencia y es igual a la suma del número de rachas del tipo uno (r_1) y el número de rachas del tipo dos (r_2), n_1 es el número de elementos del tipo uno, y n_2 es el número de elementos del segundo tipo.

Los puntos críticos de la distribución de R pueden calcularse mediante la fórmula anterior para $f_R(r)$. Tablas de la distribución de R bajo H_0 se encuentran en Conover (1999).

Si $n_1 > 20$ o $n_2 > 20$, los cuantiles de R se pueden aproximar por $R_\alpha = \frac{2n_1n_2}{n_1+n_2} + 1 + \Phi_\alpha \sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2(n_1+n_2-1)}}$, donde Φ_α es el cuantil α de la distribución normal estándar.

Para probar la hipótesis nula de aleatoriedad por medio de la prueba de rachas, cada una de las observaciones se clasifica como 0 o 1 dependiendo de si es mayor o menor que 0, se calcula el número total de rachas y si $R \leq R_{\alpha/2}$ o $R_{1-\alpha/2} < R$, donde $R_{1-\alpha/2}$ y $R_{\alpha/2}$ son los cuantiles $1-\alpha/2$ y $\alpha/2$ de la distribución de R bajo la hipótesis nula, entonces se rechaza H_0 . Esta prueba es de tamaño α .

To this end, one possibility is to use the total number of runs in the sequence $E_1, E_2, \dots, E_n$, where $E_i = \begin{cases} 1, & \text{si } e_i < 0 \\ 0, & \text{si } e_i \geq 0 \end{cases}$, $i=1, 2, \dots, n$, and a run is defined as a succession of one or more identical symbols, which are followed and preceded by one that is different.

Note that $E_1, E_2, \dots, E_n$ is a time order sequence of elements of two types (0 or 1), and that if the sequence is random, it is expected that there would not be many, nor few, runs. Let R be the va that denotes the total number of runs. Under the null hypothesis, the R probability distribution is:

$$f_R(r) = P(R=r \mid H_0 \text{ is true})$$

$$= \begin{cases} \frac{2 \binom{n_1-1}{r/2-1} \binom{n_2-1}{r/2-1}}{\binom{n_1+n_2}{n_1}} & \text{if } r \text{ is even} \\ \frac{\binom{n_1-1}{\frac{r-1}{2}} \binom{n_2-1}{\frac{r-3}{2}} + \binom{n_1-1}{\frac{r-3}{2}} \binom{n_2-1}{\frac{r-1}{2}}}{\binom{n_1+n_2}{n_1}} & \text{if } r \text{ is odd} \end{cases}$$

where r is the total number of runs in the sequence and is equal to the sum of the number of type-one runs (r_1) and the number of type-two runs (r_2), n_1 is the number of type-one elements, and n_2 is the number of type-two elements.

The critical points of the distribution of R can be calculated using the above formula for $f_R(r)$. R distribution tables under H_0 are found in Conover (1999).

If $n_1 > 20$ or $n_2 > 20$, the R quantiles can be approximated by $R_\alpha = \frac{2n_1n_2}{n_1+n_2} + 1 + \Phi_\alpha \sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2(n_1+n_2-1)}}$, where Φ_α is the α quantile of normal standard distribution.

To test the null hypothesis of randomness by means of the runs test, each of the observations is classified as 0 or 1 depending on whether it is greater or less than 0. The total number of runs is calculated, and if $R \leq R_{\alpha/2}$ or $R_{1-\alpha/2} < R$, where $R_{1-\alpha/2}$ y $R_{\alpha/2}$ are quantiles of $1-\alpha/2$ and $\alpha/2$ of the R distribution under the null hypothesis, then H_0 is rejected. This is an α -size test.

The McLeod-Li test

Brockwell and Davis (1996) and Gouriéroux (1997) suggest that the McLeod-Li statistic can be used to prove the null hypothesis H_0 ; the va $e_1, e_2, \dots, e_n$ are a random sample of the normal distribution.

Let $\{e_1, e_2, \dots, e_n\}$ be an n -size sample trajectory of the stochastic process $\{e_t, t \geq 0\}$. It is necessary to prove the null hypothesis of the va $e_1, e_2, \dots, e_n$ are iid $N(\mu, \sigma^2)$, against the alternative that they are not.

La prueba de McLeod-Li

Brockwell y Davis (1996) y Gouriéroux (1997) sugieren que la estadística de McLeod-Li puede usarse para probar la hipótesis nula H_0 : las $e_1, e_2, \dots, e_n$ son una muestra aleatoria de la distribución normal.

Sea $\{e_1, e_2, \dots, e_n\}$ una trayectoria muestral de tamaño n del proceso estocástico $\{e_t, t \geq 0\}$. Se quiere probar la hipótesis nula de que las $e_1, e_2, \dots, e_n$ son iid $N(\mu, \sigma^2)$, contra la alternativa de que no lo son.

$$\text{Sean} \quad \bar{e}^* = \frac{1}{n} \sum_{t=1}^n e_t^2 \quad \text{y} \quad \hat{\rho}(h) = \sum_{t=1}^{n-h} (e_t^2 - \bar{e}^*)$$

$$\left(e_{t+h}^2 - \bar{e}^* \right) / \sum_{t=1}^n (e_t^2 - \bar{e}^*)^2, \quad h=1, 2, \dots, m. \text{ La estadística de McLeod-}$$

Li se define como $ML = n(n+2) \sum_{h=1}^m \frac{\hat{\rho}^2(h)}{n-h}$; el número de retrasos m puede ser seleccionado convenientemente. Bajo H_0 y el supuesto de que el proceso $\{e_t, t \geq 0\}$ es de covarianza estacionaria, cuando $n \rightarrow \infty$, $ML \xrightarrow{d} X \sim \chi_m^2$. Así, se rechaza la hipótesis nula de que $e_1, e_2, \dots, e_n$ es una muestra aleatoria de la distribución normal, con un tamaño α si $ML > \chi_{m, 1-\alpha}^2$, donde $\chi_{m, 1-\alpha}^2$ es el cuantil $1-\alpha$ de la distribución χ_m^2 (Brockwell y Davis, 1996).

La prueba de Royston

Esta prueba (Royston, 1992) es una extensión de la de Shapiro-Wilk (1965) para normalidad con tamaños de muestra menores o iguales a 2000. Para muestras pequeñas la prueba de Shapiro-Wilk es de las más potentes (Seier, 2002).

Sea $\{e_1, e_2, \dots, e_n\}$ una trayectoria muestral de tamaño n del proceso estocástico $\{e_t, t \geq 0\}$, donde las $e_1, e_2, \dots, e_n$ son iid y sean $e_{(1)} \leq e_{(2)} \leq \dots \leq e_{(n)}$ las estadísticas de orden correspondientes.

Para probar la hipótesis nula H_0 : las $e_1, e_2, \dots, e_n$ tienen distribución $N(\mu, \sigma^2)$, contra la alternativa de que no la tienen Royston (1992) propuso la estadística: $z = \frac{w-\mu}{\sigma}$,

$$\text{donde} \quad w = \ln(1-W); \quad W = \left(\sum_{i=1}^n \tilde{a}_i e_{(i)} \right)^2 / \sum_{i=1}^n (e_{(i)} - \bar{e})^2;$$

$$\tilde{a}_i = \frac{\tilde{m}_i}{\sqrt{\phi}}, \quad i=1, 2, \dots, n-2 \quad \text{y} \quad n > 5; \quad \tilde{m}_i = \Phi^{-1} \left(\frac{i-3/8}{n+1/4} \right), \quad i=1, \dots, n;$$

$$\phi = \frac{\tilde{m} \tilde{m} - 2\tilde{m}_n - 2\tilde{m}_{n-1}}{1 - 2\tilde{a}_n - 2\tilde{a}_{n-1}}, \quad \tilde{a}_n = c_n + 0.221157x - 0.147981x^2 -$$

$$2.07119x^3 + 4.43685x^4 - 2.706056x^5;$$

$$\tilde{a}_{n-1} = c_{n-1} + 0.042981x - 0.293762x^2 - 1.752461x^3 + 5.682633x^4 - 3.582663x^5; \quad \tilde{c} = \tilde{m} / (\tilde{m} \tilde{m})^{1/2}$$

$$\mu = -1.5861 - 0.31082y - 0.083751y^2 + 0.0038915y^3,$$

$$\sigma = \exp\{-0.4803 - 0.082676y + 0.0030302y^2\} \quad \text{y} \quad y = \ln(n).$$

Usando z , se rechaza H_0 con un tamaño α si $\Phi(z) \geq (1-\alpha)$, donde $\Phi(x)$ es la función de distribución normal estándar.

$$\text{Let} \quad \bar{e}^* = \frac{1}{n} \sum_{t=1}^n e_t^2 \quad \text{y} \quad \hat{\rho}(h) = \sum_{t=1}^{n-h} (e_t^2 - \bar{e}^*)$$

$$\left(e_{t+h}^2 - \bar{e}^* \right) / \sum_{t=1}^n (e_t^2 - \bar{e}^*)^2, \quad h=1, 2, \dots, m. \text{ The McLeod-Li statistic is}$$

defined as $ML = n(n+2) \sum_{h=1}^m \frac{\hat{\rho}^2(h)}{n-h}$; the number of delays m can be

selected conveniently. Under H_0 and the assumption that the $\{e_t, t \geq 0\}$

process is stationary covariance, when $n \rightarrow \infty$, $ML \xrightarrow{d} X \sim \chi_m^2$. Thus, the null hypothesis that $e_1, e_2, \dots, e_n$ is a random sample of the normal distribution, with size α , is rejected if $ML > \chi_{m, 1-\alpha}^2$, where $\chi_{m, 1-\alpha}^2$ is the quantile $1-\alpha$ of the distribution χ_m^2 (Brockwell and Davis, 1996).

The Royston test

This test (Royston, 1992) is an extension of the Shapiro-Wilk (1965) test for normality with sample sizes less or equal to 2000. For small samples, the Shapiro-Wilk is one of the most powerful (Seier, 2002).

Let $\{e_1, e_2, \dots, e_n\}$ be a sample trajectory of n size of the stochastic process $\{e_t, t \geq 0\}$, where the $e_1, e_2, \dots, e_n$ are iid, and let $e_{(1)} \leq e_{(2)} \leq \dots \leq e_{(n)}$ be the statistics of corresponding order.

To prove the null hypothesis H_0 ; the $e_1, e_2, \dots, e_n$ have a $N(\mu, \sigma^2)$ distribution, against the alternative that it does not. Royston (1992) proposed the statistic $z = \frac{w-\mu}{\sigma}$, where

$$w = \ln(1-W); \quad W = \left(\sum_{i=1}^n \tilde{a}_i e_{(i)} \right)^2 / \sum_{i=1}^n (e_{(i)} - \bar{e})^2; \quad \tilde{a}_i = \frac{\tilde{m}_i}{\sqrt{\phi}},$$

$$i=1, 2, \dots, n-2 \quad \text{and} \quad n > 5; \quad \tilde{m}_i = \Phi^{-1} \left(\frac{i-3/8}{n+1/4} \right), \quad i=1, \dots, n;$$

$$\phi = \frac{\tilde{m} \tilde{m} - 2\tilde{m}_n - 2\tilde{m}_{n-1}}{1 - 2\tilde{a}_n - 2\tilde{a}_{n-1}}, \quad \tilde{a}_n = c_n + 0.221157x - 0.147981x^2 -$$

$$2.07119x^3 + 4.43685x^4 - 2.706056x^5;$$

$$\tilde{a}_{n-1} = c_{n-1} + 0.042981x - 0.293762x^2 - 1.752461x^3 + 5.682633x^4 - 3.582663x^5; \quad \tilde{c} = \tilde{m} / (\tilde{m} \tilde{m})^{1/2}$$

$$\mu = -1.5861 - 0.31082y - 0.083751y^2 + 0.0038915y^3,$$

$$\sigma = \exp\{-0.4803 - 0.082676y + 0.0030302y^2\} \quad \text{y} \quad y = \ln(n).$$

Using z , H_0 with size α is rejected if $\Phi(z) \geq (1-\alpha)$, where $\Phi(x)$ is the normal standard distribution function.

The Kolmogrov-Smirnov test

For large random sample sizes, the Kolmogorov-Smirnov test, even though it is not one of the most powerful (Seier, 2002), is one of the most used to test normality since it is found programmed in almost any known statistical data analysis software.

Let $\{e_1, e_2, \dots, e_n\}$ be a n -size sample trajectory of the stochastic process $\{e_t, t \geq 0\}$. When $e_1, e_2, \dots, e_n$ is a random sample of distribution $F(\cdot)$, it is desirable to prove the null hypothesis of normality.

La prueba de Kolmogorov-Smirnov

Para tamaños grandes de la muestra aleatoria la prueba de Kolmogorov-Smirnov, a pesar de que no es de las más potentes (Seier, 2002), es de las más usadas para probar normalidad, ya que se encuentra programada en casi cualquier paquete computacional conocido de análisis estadístico de datos.

Sea $\{e_1, e_2, \dots, e_n\}$ una trayectoria muestral de tamaño n del proceso estocástico $\{e_t, t \geq 0\}$. Se quiere probar la hipótesis nula de normalidad cuando $e_1, e_2, \dots, e_n$ es una muestra aleatoria de una función de distribución $F(\cdot)$.

Defínase $D_n = \sup_{-\infty < x < \infty} |F_n(x) - F(x)|$, donde $F_n(x)$ es la función de distribución empírica. Si H_0 es verdadera, $\sqrt{n}D_n$ se distribuye aproximadamente como $H(x) = \left[1 - 2 \sum_{j=1}^n (-1)^{j-1} \exp\{-2j^2 x^2\}\right] I_{(0,\infty)}(x)$, donde $H(x)$ no depende de $F(\cdot)$. Cuando H_0 es falsa, D_n será grande. Por tanto, se rechaza H_0 si $\sqrt{n}D_n > k_{1-\alpha}$, donde $k_{1-\alpha}$ es tal que $1 - H(k_{1-\alpha}) = \alpha$ y así $P(\sqrt{n}D_n > k_{1-\alpha} | H_0) \approx \alpha$.

La distribución exacta de D_n ha sido tabulada para varios valores de n . Los paquetes estadísticos S-Plus y SPSS, entre otros, permiten realizar la prueba de Kolmogorov-Smirnov.

Existen otras pruebas para normalidad y correlación serial, entre las cuales se encuentran las de Bera y Jarque (1982) y Jarque y Bera (1980, 1987). Para una revisión sistemática de éstas y otras pruebas, véase Baltagi (2001), Seier (2002), Thode (2002) y Nguyen y Dinh (2003).

Para probar la hipótesis nula de que una muestra aleatoria de una distribución normal tiene media cero se usa la prueba de t (Casella y Berger, 1990).

Metodología propuesta

Para probar la hipótesis nula H_0 de movimiento Browniano se proponen pruebas del tipo unión-intersección (Casella y Berger, 1990 p. 356). En esta situación, H_0 es la intersección de las tres hipótesis siguientes: H_{01} : las $e_1, e_2, \dots, e_n$ son iid, H_{02} : las $e_1, e_2, \dots, e_n$ son normales y H_{03} : las $e_1, e_2, \dots, e_n$ tienen media cero. Por tanto, se rechaza la hipótesis de movimiento Browniano si se rechaza al menos una de las tres hipótesis H_{0j} , $j=1,2,3$. Para esto se consideran los siguientes tres procedimientos de prueba.

Procedimiento	Pruebas que se combinan
P_1	McLeod-Li (ML) y la prueba de t
P_2	Rachas (R), Royston (W') y la prueba de t
P_3	Rachas (R), Kolmogorov-Smirnov (D) y la prueba de t

Por la desigualdad de Bonferroni (Casella y Berger, 1990, p. 523), cada procedimiento de prueba P_i es de tamaño menor o igual que α cuando la suma de los tamaños de las pruebas que integran al procedimiento sea menor o igual que α .

El procedimiento propuesto (P_i) para probar H_0 : las $e_1, e_2, \dots, e_n$ siguen un movimiento Browniano es el siguiente:

Define $D_n = \sup_{-\infty < x < \infty} |F_n(x) - F(x)|$, where $F_n(x)$ is the function of empirical distribution. If H_0 is true, $\sqrt{n}D_n$ is distributed

approximately as $H(x) = \left[1 - 2 \sum_{j=1}^n (-1)^{j-1} \exp\{-2j^2 x^2\}\right] I_{(0,\infty)}(x)$,

where $H(x)$ does not depend on $F(\cdot)$. When H_0 is false, D_n will be

large. Therefore, H_0 is rejected if $\sqrt{n}D_n > k_{1-\alpha}$, where $k_{1-\alpha}$ is such that $1 - H(k_{1-\alpha}) = \alpha$, and thus, $P(\sqrt{n}D_n > k_{1-\alpha} | H_0) \approx \alpha$.

The exact distribution of D_n has been tabulated for several values of n . The software S-Plus and SPSS, among others, include the Kolmogorov-Smirnov test.

There are other tests for normality and serial correlation, among which are found those of Bera and Jarque (1982) and Jarque and Bera (1980, 1987). For a systematic review of these and other tests, see Baltagi (2001), Seier (2002), Thode (2002) and Nguyen and Dinh (2003).

Proposed methodology

To test the Brownian movement null hypothesis H_0 , union-intersection type tests (Casella and Berger, 1990, p. 356) are proposed. In this situation, H_0 is the intersection of the following three hypotheses: H_{01} : the $e_1, e_2, \dots, e_n$ are iid, H_{02} : the $e_1, e_2, \dots, e_n$ are normal and H_{03} : the $e_1, e_2, \dots, e_n$ have mean zero. Therefore, the Brownian movement hypothesis is rejected unless one of the three hypotheses H_{0j} , $j=1,2,3$. For this, the following three test procedures are considered:

Procedure	Tests that combine
P_1	McLeod-Li (ML) and t test
P_2	Runs (R) Royston (W') and t test
P_3	Runs (R), Kolmogorov-Smirnov (D) and the t test

By the Bonferroni inequality (Casella and Berger, 1990, p. 523), each procedure of test P_i is smaller than or equal to α when the sum of the test sizes that integrate the procedure is smaller than or equal to α .

The proposed procedure (P_i) to prove H_0 : that the $X_1, X_2, \dots, X_n$ follow a Brownian movement is the following:

- Calculate $e_1 = X_1 - X_0, e_2 = X_2 - X_1, \dots, e_n = X_n - X_{n-1}$.
- Based on $e_1, e_2, \dots, e_n$, calculate the test statistics that constitute P_i .
- For a given size α , reject the null hypothesis if at least one of the test statistics that integrate P_i rejects the corresponding hypothesis.

It is convenient to note that because P_i is a union-intersection-type test, it is not required that a particular order be followed when conducting the tests of hypothesis H_{0j} , $j=1,2,3$. For example, when performing the test H_{01} with the runs test, the null hypothesis H_0 includes H_{02} and H_{03} , so that the conditions required for said test are satisfied.

- i. Calcular $e_1 = X_1 - X_0, e_2 = X_2 - X_1, \dots, e_n = X_n - X_{n-1}$.
- ii. Con base en $e_1, e_2, \dots, e_n$, calcular las estadísticas de prueba que componen a P_i .
- iii. Para un tamaño α dado, rechazar la hipótesis nula si al menos una de las estadísticas de prueba que integran P_i rechaza la hipótesis correspondiente.

Es conveniente hacer notar que por ser P_i una prueba de tipo unión-intersección, no se requiere seguir algún orden en particular al realizar las pruebas de las hipótesis H_{0j} , $j=1,2,3$. Por ejemplo, al realizar la prueba de H_{01} mediante la prueba de rachas, la hipótesis nula H_0 incluye H_{02} y H_{03} , por lo que las condiciones requeridas para dicha prueba son satisfechas.

RESULTADOS Y DISCUSIÓN

En esta sección se discute un estudio de simulación de Monte Carlo para el tamaño de los procedimientos de prueba propuestos para probar la hipótesis de movimiento Browniano, así como una comparación de la potencia de los mismos para algunas alternativas.

Con el propósito de investigar si el procedimiento de prueba P_i conserva el tamaño previsto, se utilizó el siguiente algoritmo:

1. Se simula una trayectoria muestral de tamaño n , $X_1, X_2, \dots, X_n$, de un movimiento Browniano, para una n seleccionada.
2. Se calculan los incrementos $e_1 = X_1 - X_0, e_2 = X_2 - X_1, \dots, e_n = X_n - X_{n-1}$.
3. Con base en $e_1, e_2, \dots, e_n$, se calculan las estadísticas de prueba que componen al procedimiento P_i .
4. Se rechaza la hipótesis nula con el procedimiento P_i si al menos una de las estadísticas de prueba que lo componen rechaza la hipótesis correspondiente.
5. Se asocia el número 1 al procedimiento P_i si éste rechaza la hipótesis nula o el número 0 si no la rechaza.
6. Los pasos anteriores se repiten J veces.
7. Para el procedimiento P_i , se divide por J la suma de los números asociados del paso 5.

La cantidad que resulta en el paso 7 es la probabilidad estimada de cometer el error tipo I cuando se usa el procedimiento P_i para probar la hipótesis de movimiento Browniano.

Para estudiar el comportamiento del procedimiento P_i respecto a su tamaño, se diseñó el siguiente experimento de simulación. Se consideraron diferentes tamaños de trayectorias muestrales, los tamaños pequeños, cuando $n=20, 30, 50$, y los grandes, cuando $n=100, 200$ y 300 . El número de repeticiones utilizado fue $J=2000$, y

RESULTS AND DISCUSSION

In this section a Monte Carlo simulation study is discussed for the size of the test procedures proposed to test the Brownian movement hypothesis, as well as a comparison of the power of the same for some alternatives.

With the purpose of investigating whether the P_i test procedure sustains the predicted size, the following algorithm was used:

1. An n -size, $X_1, X_2, \dots, X_n$, sample trajectory of a Brownian movement is simulated for a selected n .
2. The increments $e_1 = X_1 - X_0, e_2 = X_2 - X_1, \dots, e_n = X_n - X_{n-1}$ are calculated.
3. Based on $e_1, e_2, \dots, e_n$, the test statistics that compose the P_i procedure are calculated.
4. The null hypothesis is rejected with the P_i procedure if at least one of the component test statistics rejects the corresponding hypothesis.
5. The number 1 is associated to the P_i procedure if this rejects the null hypothesis, or the number 0 if it does not reject it.
6. The above steps are repeated J times.
7. For the P_i procedure, the sum of the associated numbers of step 5 is divided by J .

The amount that results in step 7 is the estimated probability of committing a type I error when the P_i procedure is used to test the Brownian movement hypothesis.

To study the behavior of the P_i procedure in terms of its size, the following simulation experiment was designed. Different sizes of sample trajectories are considered: small when $n=20, 30$ and 50 , and large when $n=100, 200$ and 300 . The number of replications used was $J=2000$, and the test size was $\alpha=0.1$. The sizes with which the tests composing each of the procedures were conducted are shown in Table 1.

To perform the Mcleod-Li test, $m=10$ was considered in the case of small sample trajectories and $m=20$ in the case of large sizes because in these cases the test preserves its size.

To perform the Kolmogorov-Smirnov test, the values for the parameters of normal distribution were $\mu=0$ and σ^2 , equal to the estimation obtained by the estimator of maximum likelihood $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n e_i^2$, because we are dealing with a union-intersection-type test.

It is valid to use the t test to prove H_{03} since the null hypothesis of Brownian movement includes the hypotheses of independence and normality of the various $e_1, e_2, \dots, e_n$.

el tamaño de la prueba es $\alpha=0.1$. Los tamaños con que se efectuaron las pruebas que componen cada uno de los procedimientos se presentan en el Cuadro 1.

Para realizar la prueba de McLeod-Li se consideró $m=10$, en el caso de tamaños de trayectorias muestrales pequeños y $m=20$, en el caso de tamaños grandes, debido a que en esos casos la prueba conserva su tamaño.

Para la prueba de Kolmogorov-Smirnov los valores para los parámetros de la distribución normal fueron $\mu=0$ y σ^2 , igual a la estimación obtenida por el estimador de máxima verosimilitud $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n e_i^2$, debido a que se trata de una prueba de tipo unión-intersección.

Es válido emplear la prueba de t para probar H_{03} debido a que la hipótesis nula de movimiento Browniano incluye las hipótesis de independencia y normalidad de las v.a's $e_1, e_2, \dots, e_n$.

Para la prueba de normalidad de Royston es necesario suponer que las v.a $e_1, e_2, \dots, e_n$ son iid con media $\mu=0$.

En la Figura 1 se presentan los resultados del experimento de simulación, los que muestran que para tamaños pequeños los tres procedimientos se comportan satisfactoriamente. Para tamaños grandes el procedimiento P_3 es el más consistente, aunque los otros dos no exceden por más del 10% el tamaño de prueba fijado.

Para investigar y comparar las potencias de los tres procedimientos, se diseñó el siguiente experimento. Se simularon trayectorias muestrales de tamaños 20, 30, 50, 100, 200 y 300 de algunos procesos alternativos no Brownianos. La potencia se estimó usando el siguiente algoritmo:

1. Se simula una trayectoria muestral de tamaño n , $X_1, X_2, \dots, X_n$, de uno de los procesos alternativos.
2. Se siguen los pasos 2 a 7 del algoritmo usado para la estimación del tamaño del procedimiento de prueba P_i .

La cantidad resultante en el paso 7 es la probabilidad estimada con que cada procedimiento rechaza la hipótesis nula dada la alternativa.

Se consideraron diferentes esquemas de simulación en los que se toman en cuenta hipótesis alternativas que se alejan de la hipótesis nula en diferentes direcciones

Cuadro 1. Tamaños con que se efectuaron las pruebas que componen cada uno de los procedimientos.
Table 1. Sizes with which tests constituting each procedure were conducted.

Procedimiento	ML	R	W'	K-S	T	α
P_1	0.075	-	-	-	0.025	0.1
P_2	-	0.05	0.025	-	0.025	0.1
P_3	-	0.05	-	0.025	0.025	0.1

For the Royston normality test, it is necessary to consider the assumptions that the v.a $e_1, e_2, \dots, e_n$ are iid with mean $\mu=0$.

Figure 1 shows the results of the simulation experiment. It can be seen that for small sizes of the realization of the process, the three procedures behave satisfactorily. For large sizes of the realization of the procedure, P_3 is the most consistent, although the other two procedures do not exceed the assigned test size by more than 10%.

To study and compare the powers of the three test procedures proposed, the following experiment was designed. Sample trajectories of sizes 20, 30, 50, 100, 200, and 300 of some alternative non-Brownian processes were simulated. The power was estimated using the following algorithm:

1. An n -size, $X_1, X_2, \dots, X_n$, sample trajectory of one of the alternative processes is simulated.
2. Steps 2 to 7 of the algorithm used for the estimation of the size of the procedure of test P_i are followed.

The quantity resulting in step 7 is the estimated probability with which each procedure rejects the null hypothesis, given the alternative.

Different simulation schemes were considered taking into account alternative hypotheses that move away from the null hypothesis in different directions, and for small ($n=20, 30, 50$) and large ($n=100, 200, 300$) sample trajectory sizes. Some of the alternative processes considered are:

1. The AR(1) process: $X_t = \phi X_{t-1} + Z_t$, Z_t iid $N(0,4)$, when $\phi=0.5$. Note that Brownian movement exists when $\phi=0.1$. In this case, the powers for large and small sizes are to be compared.

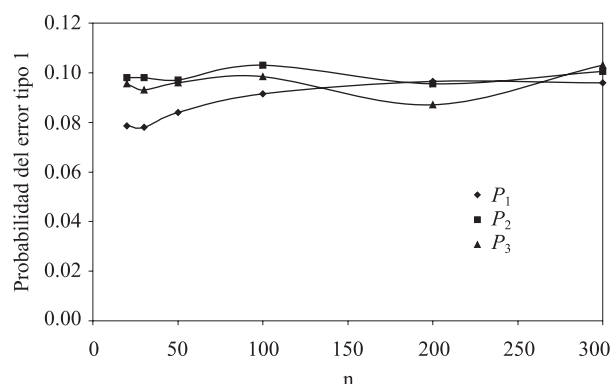


Figura 1. Probabilidades estimadas de cometer el error tipo I usando P_1 , P_2 y P_3 , con $\alpha=0.1$.

Figure 1. estimated probabilities of committing type 1 error using P_1 , P_2 and P_3 with $\alpha=0.1$.

y para tamaños de trayectorias muestrales pequeños ($n=20, 30, 50$) y grandes ($n=100, 200, 300$). Algunos de los procesos alternativos considerados son:

1. El proceso AR(1): $X_t = \phi X_{t-1} + Z_t$, Z_t iid $N(0,4)$, cuando $\phi=0.5$. Note que se tiene un movimiento Browniano cuando $\phi=1$. En este caso se pretende comparar las potencias para tamaños de la realización grandes y pequeños.
2. Los procesos autorregresivos de orden 1, AR(1): $X_t = \phi X_{t-1} + Z_t$, Z_t iid $N(0,4)$, con $\phi=0.1, 0.25, 0.5, 0.75, 0.9$ y $n=300$. En este caso se pretende analizar las potencias conforme el parámetro ϕ se aleja de 1.
3. Para investigar la sensibilidad de los procedimientos de prueba cuando los incrementos del proceso no son idénticamente distribuidos se consideró el siguiente proceso: $X_t = X_{t-1} + Z_t$, Z_t independientes con distribución $N(0, \sigma^2 t^2)$ cuando $\alpha=0.1$.
4. Para estudiar la sensibilidad de las pruebas respecto a procesos que se alejan de la normalidad se consideró el proceso $X_t = X_{t-1} + Z_t$, Z_t iid t-Student con 4 grados de libertad.
5. Para investigar la sensibilidad respecto a la no centralidad alrededor de cero de los incrementos, se consideró el modelo $X_t = X_{t-1} + Z_t$, Z_t iid $N(0.5, 1)$.

En las Figuras 2 a 6 se presentan los resultados obtenidos cuando se consideró cada hipótesis alternativa de 1 a 5.

Los resultados de la Figura 2 muestran que las potencias estimadas de los procedimientos P_2 y P_3 son prácticamente iguales, y muy superiores a la del procedimiento P_1 . Nótese que las potencias de P_2 y P_3 aumentaron a medida que aumentaron los tamaños de la realización.

Las funciones de potencias estimadas respecto al parámetro ϕ de la alternativa 1 de los procedimientos P_2 y P_3 son prácticamente iguales, y que éstas aumentan a medida que el valor de ϕ se aleja de la hipótesis nula $\phi=1$ para hipótesis alternativas $\phi<1$. El procedimiento P_1 presenta baja potencia respecto a P_2 y P_3 aún cuando ϕ toma valores alejados de 1 (Figura 3).

Los resultados de la Figura 4 muestran que para tamaños pequeños ($n=20, 30, 50$) la potencia del procedimiento de prueba P_2 es uniformemente mayor que la potencia de P_1 y P_3 , mientras que para tamaños grandes ($n=100, 200, 300$), la potencia de P_1 es uniformemente mayor que la potencia de P_2 y P_3 . Esto es, ninguno de los procedimientos de prueba es uniformemente más potente respecto a n . En este caso se observó que

2. The autoregressive processes of first order, AR(1): $X_t = \phi X_{t-1} + Z_t$, Z_t iid $N(0,4)$, with $\phi=0.1, 0.25, 0.5, 0.75, 0.9$ and $n=300$. In this case the powers are to be analyzed as the ϕ parameter moves away from 1.
3. To study the sensitivity of the test procedures when the increments of the process are not identically distributed, the following process was considered: independent $X_t = X_{t-1} + Z_t$, Z_t with distribution $N(0, \sigma^2 t^2)$ when $\alpha=0.1$.
4. To study the sensitivity of the tests with respect to processes that move away from normality, the process $X_t = X_{t-1} + Z_t$, Z_t iid t-Student with 4 degrees of freedom.
5. To study the sensitivity with respect to non-centrality around zero of the increments, the model $X_t = X_{t-1} + Z_t$, Z_t iid $N(0.5, 1)$ was considered.

Figures 2 to 6 present the results obtained when each alternative hypothesis, from 1 to 5, were considered.

The results of Figure 2 show that the estimated powers of the P_2 and P_3 procedures are practically equal and far superior to that of procedure P_1 . Note that the powers of P_2 and P_3 increase in the measure that sizes of the realization increase.

The estimated power functions with respect to the ϕ parameter of alternative 1 of procedures P_2 and P_3 are practically equal and that these increase in the measure that the value of ϕ moves away from the null hypothesis $\phi=1$, for alternative hypotheses $\phi<1$. The procedure P_1 has low power compared with P_2 and P_3 , even when ϕ takes values distant from 1 (Figure 3).

The results in Figure 4 show that for small sizes ($n=20, 30, 50$), the power of the P_2 test is uniformly greater than the power of P_1 and P_3 , while for large sizes ($n=100, 200, 300$), the power of P_1 is uniformly greater than the power of P_2 and P_3 . That is, none of the test procedures is uniformly more powerful with respect to n . In this case, it was observed that the power of the three procedures increased in the measure that n increased.

P_3 is power uniformly greater than that of P_1 and, in turn, the power of P_2 is uniformly greater than that of P_3 , so that P_2 is the best in this case. Here, only P_2 and P_3 increased their power when n increased (Figure 5).

The estimated powers of the three test procedures are practically equal for all the considered n ; this is because in all of the procedures the t test is of size 0.025.

The procedure for test P_1 has low power when the alternatives have dependent increments, homogeneous variances and non-normal distributions (Figures 2, 3 and 5). Because this procedure is based on the McLeod-Li test to prove the randomness hypothesis and normality

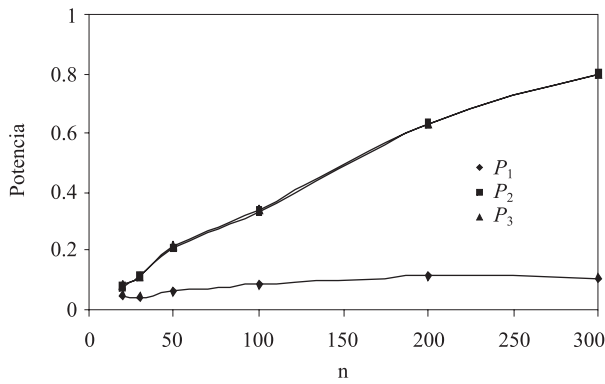


Figura 2. Potencia estimada de P_1 , P_2 y P_3 contra la hipótesis alternativa 1.
Figure 2. Estimated power of P_1 , P_2 and P_3 against alternative hypothesis 1.

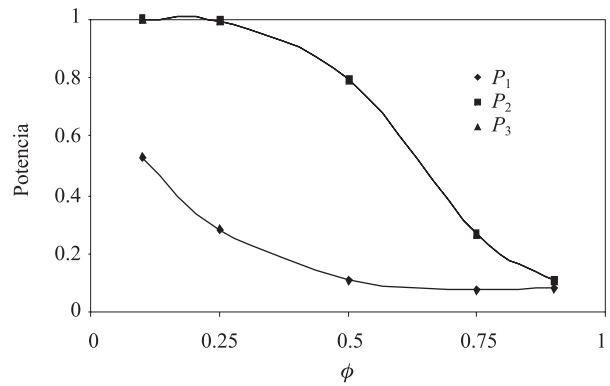


Figura 3. Potencia estimada de P_1 , P_2 y P_3 contra la hipótesis alternativa 2.
Figure 3. Estimated power of P_1 , P_2 and P_3 against alternative hypothesis 2.

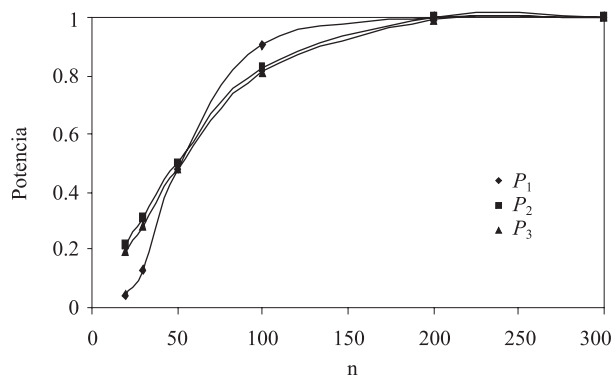


Figura 4. Potencia estimada de P_1 , P_2 y P_3 contra la hipótesis alternativa 3.
Figure 4. Estimated power of P_1 , P_2 and P_3 against alternative hypothesis 3.

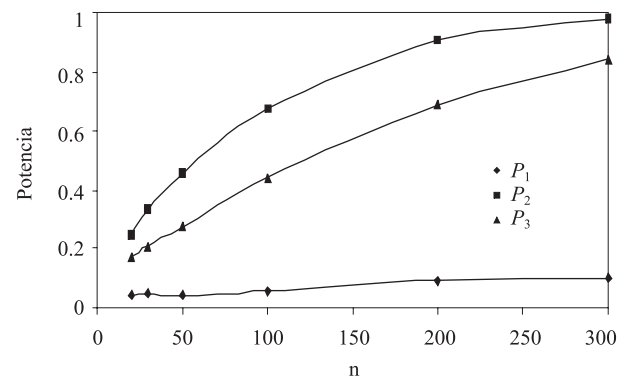


Figura 5. Potencia estimada de P_1 , P_2 y P_3 contra la hipótesis alternativa 4.
Figure 5. Estimated power of P_1 , P_2 and P_3 against alternative hypothesis 4.

las potencias de los tres procedimientos aumentaron a medida que aumentó n .

La potencia de P_3 es uniformemente mayor que la de P_1 , y a su vez la potencia de P_2 es uniformemente mayor que la de P_3 , por lo que P_2 es la mejor en este caso. Aquí, solamente P_2 y P_3 aumentaron su potencia cuando aumentó n (Figura 5).

La potencias estimadas de los tres procedimientos de prueba son prácticamente iguales para todos los n considerados, lo anterior porque en en todos los procedimientos la prueba de t es de tamaño 0.025.

El procedimiento de prueba P_1 tiene baja potencia cuando las alternativas son con incrementos dependientes, varianzas homogéneas y distribuciones no normales (Figuras 2, 3 y 5). Debido a que este procedimiento está basado en la prueba de McLeod-Li para probar la hipótesis de aleatoriedad y normalidad de las primeras diferencias, los resultados sugieren que esta prueba no es eficiente en general para probar que un conjunto de datos es una muestra aleatoria normal, lo que difiere de Brockwell y Davis(1996) y Gouriéroux (1997).

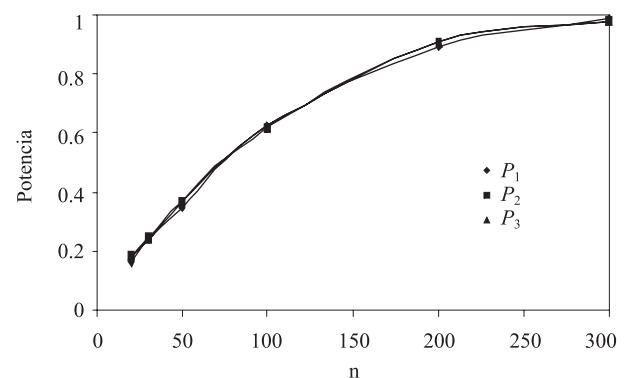


Figura 6. Potencia estimada de P_1 , P_2 y P_3 contra la hipótesis alternativa 5.
Figure 6. Estimated power of P_1 , P_2 and P_3 against alternative hypothesis 5.

of the first differences, the results suggest that this test is not, in general, efficient to prove that a set of data is a normal random sample, differing from Brockwell and Davis (1996) and Gouriéroux (1997).

Aplicaciones

La metodología propuesta se aplicó a los siguientes conjuntos de datos para probar la hipótesis de movimiento Browniano con base en una trayectoria muestral de tamaño n :

1. Precios de petróleo crudo (serie corta) durante el periodo 16/7/1996 a 21/4/1997 en los Estados Unidos ($n=275$).
2. Logaritmos de los precios de petróleo crudo ($n=275$).
3. Logaritmos de los precios de petróleo crudo (serie larga) durante el periodo 3/1/1995 a 19/11/1997 en los Estados Unidos ($n=752$).
4. Precios de gas en Estados Unidos del 1/1/1997 al 8/12/1997 ($n=244$).
5. Logaritmos de los precios de gas ($n=244$).
6. Precios de cierre diarios de las acciones de IBM durante el periodo 17/05/1961 a 2/12/1962 ($n=369$).
7. El logaritmo del IPC-cierre de la BMV del 2/1/2001 al 18/12/2002 ($n=490$).

Para mostrar las aplicaciones de la metodología propuesta se consideraron series cortas de datos en las cuales se observa cierta estacionaridad de los incrementos, lo cual se refleja en el hecho de que la hipótesis de movimiento Browniano no es rechazada (ver Cuadro 2). Mientras que en las series largas generalmente la estacionaridad llega a romperse cuando se presenta algún evento económico extraordinario.

En el Cuadro 2 se presentan los resultados obtenidos al utilizar los procedimientos P_1 , P_2 y P_3 para probar la hipótesis de que se tiene un movimiento Browniano con base en cada uno de los conjuntos de datos considerados. El número 1 indica que se rechaza la hipótesis nula y el número 0 que no se rechaza.

Como se observa en el cuadro anterior, tanto para los precios de petróleo crudo (serie corta) como para sus logaritmos (serie corta), ninguno de los 3 procedimientos de prueba rechaza la hipótesis de movimiento Browniano con $\alpha=0.1$.

Applications

The proposed methodology was applied to the following sets of data to obtain the Brownian movement hypothesis, based on a sample trajectory of n size:

1. Crude oil prices (short series) during the period 16/7/1996 to 21/4/1997 in the United States ($n=275$).
2. Logarithms of crude oil prices ($n=275$).
3. Logarithms of crude oil prices (long series) during the period 3/1/1995 to 19/11/1997 in the United States ($n=752$).
4. Gas prices in the United States from 1/1/1997 to 8/12/1997 ($n=244$).
5. Logarithms of gas prices ($n=244$).
6. Daily closing prices of IBM stock during the period 17/05/1961 to 2/12/1962 ($n=369$).
7. Logarithm of IPC-closing of BMW from 2/1/2001 to 18/12/2002 ($n=490$).

To demonstrate the applications of the proposed methodology, the six short series of data considered were observed with certain stationarity in the increments, which is reflected in the fact that the Brownian movement hypothesis is not rejected (see Table 2), while in the long series, generally stationarity breaks when some extraordinary economic event occurs.

Table 2 presents the results obtained when using procedures P_1 , P_2 and P_3 to prove the Brownian movement hypothesis based on each of the data sets considered. The number 1 indicates that the null hypothesis is rejected and the number 0, that it is not rejected.

As can be observed in the previous table, for both crude oil prices (short series) and their logarithms (short series), none of the three test procedures rejects the Brownian movement hypothesis with $\alpha=0.1$.

For gas prices, their logarithms and IBM stock prices, the Brownian movement hypothesis is rejected with $\alpha=0.1$. For the IPC data, procedure P_1 does not reject the Brownian movement hypothesis; procedures P_2 and P_3 reject H_0 .

Cuadro 2. Resultados ($p \leq 0.1$).
Table 2. Results ($p \leq 0.1$).

	Conjunto de datos	P_1	P_2	P_3
1	Precio de Petróleo (serie corta)	0	0	0
2	Logaritmo de precio de petróleo (serie corta)	0	0	0
3	Logaritmo de precio de petróleo (serie larga)	1	1	1
4	Precio de Gas	1	1	1
5	Logaritmo de precio de gas	1	1	1
6	Precios de IBM	1	1	1
7	IPC	0	1	1

Para los precios de gas, sus logaritmos y los precios de IBM, se rechaza la hipótesis de movimiento Browniano con $\alpha=0.1$. Para los datos del IPC, el procedimiento P_1 no rechaza la hipótesis de movimiento Browniano. Los procedimientos P_2 y P_3 rechazan H_0 .

Para los logaritmos de los precios de petróleo crudo (serie larga) las tres pruebas rechazan la hipótesis nula, lo cual es congruente con los resultados reportados por Ross (1999).

En su procedimiento de prueba, Ross (1999) clasifica los logaritmos de los precios en 4 categorías, y usando una tabla de contingencia 4×4 para probar la hipótesis de independencia, encuentra que la estadística de prueba Ji-cuadrada toma el valor 23.477, por lo que se rechaza dicha hipótesis ($p=0.00526$). Además, por medio de un análisis de varianza, también rechaza la hipótesis de que las primeras diferencias son idénticamente distribuidas ($p \leq 0.001$).

CONCLUSIONES

En general, el procedimiento P_1 conserva el tamaño de la prueba; sin embargo, presenta muy baja potencia en general, de donde se concluye que la prueba de McLeod-Li no es eficiente para probar que un conjunto de datos es una muestra aleatoria de la distribución normal.

El procedimiento P_2 , en general conserva el tamaño de la prueba, y presenta mayor potencia ante la mayoría de las alternativas consideradas, tanto para tamaños de trayectorias muestrales pequeños como grandes. Además, la potencia de P_2 aumentó conforme se aumentó n . Se recomienda usar el procedimiento P_2 , basado en las pruebas de Rachas, Shapiro-Wilk y t .

AGRADECIMIENTO

Los autores extienden su más cumplido agradecimiento a los editores asignados a este trabajo por sus atinados comentarios y sugerencias, que ayudaron a mejorar sustancialmente la presentación de la versión original de este artículo.

LITERATURA CITADA

- Baltagi, B. H. 2001. *A Companion to Theoretical Econometrics*. Blackwell Publisher, Oxford, UK. 709 p.
- Bera, A. K., and C. M. Jarque. 1982. Model specification tests: a simultaneous approach. *Journal of Econometrics* 20: 59-82.
- Brockwell, P. J., and R. A. Davis. 1996. *Introduction to Time Series Analysis and Forecasting*. Springer-Verlag, New York. 420 p.
- Campbell, J. Y., Lo, A. W., and A. C. Mackinlay. 1997. *The Econometrics of Financial Markets*. Princeton University Press. USA.
- Casella, G., and R. L. Berger. 1990. *Statistical Inference*. Brooks/Cole, USA.
- Conover, W. J. 1999. *Practical Nonparametric Statistics*. 3rd ed. J. Wiley and Sons Inc, New York. 584 p.

For the logarithms of crude oil prices (long series) the three tests reject the null hypothesis; this is congruent with the results reported by Ross (1999).

In his test procedure, Ross (1999) classifies the logarithms of prices in four categories and, using a 4×4 contingency table to prove the independence hypothesis, he finds that the test statistic χ^2 takes the value 23.477, and so this hypothesis is rejected ($p=0.00526$). Also, by means of an analysis of variance, the hypothesis that the first differences are identically distributed ($p \leq 0.001$) is also rejected.

CONCLUSIONS

In general, the P_1 procedure preserves test size; however, it is in general of very low power. Thus, it is concluded that the McLeod-Li test is not efficient to prove that a set of data is a random sample of the normal distribution.

The P_2 procedure, in general, preserves the test size and has greater power in most of the alternatives considered, for both small and large sample trajectories. Also, the power of P_2 increases as n increases. It is recommended that the P_2 procedure be used, based on the runs, Shapiro-Wilk and t tests.

—End of the English version—



- Dickey, D. A., and W. A. Fuller. 1981. Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica* 49: 1057-1072.
- Elliot, G. 1999. Efficient tests for a unit root when the initial observation is drawn from its unconditional distribution. *Int. Econ. Rev.* 40 (3): 767-783.
- Elliot, G., and M. Janssonb. 2003. Testing for unit roots with stationary covariates. *J. Econometrics*. 115: 75-89.
- Elliot, G., T. Rothenberg, and J. Stock. 1996. Efficient tests for an autoregressive unit root. *Econometrica*: 64 (4): 813-836.
- Gouriéroux, Ch. 1997. *ARCH Models and Financial Applications*. Springer-Verlag, New York, Inc.
- Hall, A. 1994. Testing for a unit root in time series with pretest data-based model selection. *J. Bus. and Econ. Stat.* 12: 461-470.
- Jarque, C. M., and A. K. Bera. 1980. Efficient test for normality, heteroscedasticity and serial independence of regression residuals. *Economics Letters* 11: 123-127.
- Jarque, C. M., and A. K. Bera. 1987. A test for normality of observations and regression residuals. *International Statistical Review* 55: 163-172.
- Lehmann, E. L. 1997. *Testing Statistical Hypotheses*. 2nd Ed. Springer-Verlag, New York. 600 p.
- Leyburne, S. J., and B. P. M. McCabe. 1994. A consistent test for a unit root. *J. Bus. and Econ. Stat.* 12(2): 157-166.
- Lumsdaine, R. L., and D. H. Papell. 1997. Multiple trend breaks and the unit-root hypothesis. *Rev. Econ. & Stat.* 79(2): 212-218.
- Ng, S., and P. Perron. 1995. Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag. *JASA*. 90: 268-281.

- Nguyen, T. T., and K. T. Dinh. 2003. Characterizations of normal distributions and EDF goodness-of-fit tests. *Metrika* 58: 149-157.
- Ross, S. M. 1996. *Stochastic Processes*. 2nd Ed. John Wiley & Sons, Inc. 510 p.
- Ross, S. M. 1999. *An Introduction to Mathematical Finance: Options and Other Topics*. Cambridge University Press, USA. 184 p.
- Royston, P. 1992. Approximating the Shapiro-Wilk W-test for non-normality. *Statistics and Computing* 2: 117-119.
- Seier, E. 2002. Comparison of tests for univariate normality. *InterStat*: 1.
- Shapiro, S. S., and M. B. Wilk. 1965. An analysis of variance test for normality. *Biometrika* 52: 591-611.
- Taylor, S. 1986. *Modelling Financial Time Series*. John Wiley and Sons, Chichester. 268 p.
- Thode Jr., H. C. 2002. *Testing for Normality*. Marcel Dekker, New York. 368 p.