

IDENTIFICACIÓN DE MODELOS ARFIMA^a

IDENTIFICATION IN ARFIMA MODELS

ELKIN CASTAÑO^{b c}

Recibido 04-09-2015, aceptado 03-02-2016, versión final 29-02-2016.

Artículo Investigación

RESUMEN: Desde la introducción de los modelos fraccionalmente integrados ARFIMA para series de tiempo con memoria larga, ha surgido un gran interés en el estudio de sus propiedades y áreas de aplicación. En este modelo, el grado de la memoria está definido por el parámetro de diferenciación fraccional, el cual toma valores en un intervalo continuo de números reales. Para realizar la estimación de este parámetro y probar la existencia de memoria larga, se han propuesto distintos procedimientos en la literatura. Ahora bien, generalmente no basta con conocer si hay memoria larga en la serie de tiempo, sino que es necesario estimar adecuadamente el valor del parámetro de diferenciación fraccional, del cual depende la dinámica de largo plazo de la serie, y de la componente ARMA asociada al comportamiento de corto plazo. Esta estimación requiere de la especificación correcta del modelo ARFIMA. El objetivo de este artículo es el de implementar un proceso de identificación del modelo ARFIMA para series estacionarias a partir de un procedimiento paramétrico propuesto, y comparar su desempeño con métodos semiparamétricos propuestos en la literatura. Los resultados, obtenidos a través de un estudio de simulación, muestran que el procedimiento propuesto tiene, en general, un mejor desempeño.

PALABRAS CLAVE: *Integración fraccional, memoria larga, métodos paramétricos, métodos semiparamétricos, persistencia.*

ABSTRACT: Since the introduction of ARFIMA models for fractionally integrated time series with long memory, there has been great interest in the study of their properties and application areas. In this model, the degree of memory is defined by the fractional differencing parameter, which takes values in a continuous range of real numbers. In order to estimate this parameter and prove the existence of long memory, they have been proposed various methods in the literature. But usually it is not enough to know if there is long memory in time series, it is necessary to properly assess the value of the fractional differencing parameter, which depends on the long-term dynamics of the series, and the associated component ARMA short-term behavior. This estimate requires the correct specification of the ARFIMA model. The purpose of this paper is to implement a process of identification for the ARFIMA model based in a parametric procedure, and compare their performance with semi-parametric methods proposed in the literature. The results obtained through a simulation study show that the proposed method has generally improved performance.

^aCastaño, E. (2016). Identificación de modelos ARFIMA *Revista de la Facultad de Ciencias*, 5 (1), 12–37. DOI: <https://doi.org/10.15446/rev.fac.cienc.v5n1.52890>

^bEscuela de Estadística. Universidad Nacional de Colombia, Sede Medellín. elkincv@gmail.com

^cDepartamento de Economía. Universidad de Antioquia, Medellín.

KEYWORDS: Fractional integration, long memory, parametric method, semiparametric method, persistence.

1. INTRODUCCIÓN

La evidencia empírica sobre datos con memoria larga se remonta mucho tiempo atrás, siendo el trabajo de Hurst (1951) en el campo de la hidrología, el ejemplo más conocido. En los últimos años ha habido un gran interés en el estudio de las propiedades de las series de tiempo con memoria larga de la clase ARFIMA. Beran (1992) señala que se ha encontrado evidencia de memoria larga en series de tiempo de otras ciencias tales como Economía, Finanzas, Astronomía, Agricultura, Química, Meteorología, Medio Ambiente, Biología, Telecomunicaciones y Geología. Muchos de los desarrollos formales sobre estimación y contrastes de hipótesis en estos modelos son relativamente recientes.

Pérez & Ruiz (2002) señalan que desde un punto de vista empírico, la propiedad de memoria larga suele relacionarse con la persistencia que muestran las autocorrelaciones muestrales de algunas series de tiempo estacionarias, las cuales decrecen a un ritmo muy lento pero finalmente convergen hacia cero. Este comportamiento no es compatible ni con el de los modelos estacionarios autorregresivos y de medias móviles ARMA, en los cuales las autocorrelaciones decrecen exponencialmente, ni con el grado extremo de persistencia de los modelos integrados no estacionarios ARIMA.

Granger (1980) y Granger & Joyeux (1980) advierten que la práctica habitual de diferenciar una serie de tiempo aparentemente no estacionaria hasta obtener estacionariedad, puede tener consecuencias negativas en la correcta modelación de algunas series de tiempo. Muchas series aparentemente no estacionarias suelen diferenciarse para conseguir una serie estacionaria. Sin embargo, la serie diferenciada se convierte en una serie en la cual se eliminó la componente de bajas frecuencias, que es fundamental en las predicciones a largo plazo. Dichos autores señalan que para modelar este tipo de series, la diferenciación entera es “excesiva”(sobrediferenciación) pero no diferenciar tampoco es adecuado (subdiferenciación). Los siguientes gráficos corresponden a series simuladas generadas por procesos de memoria larga ARFIMA estacionarios y no estacionarios, ilustran la posibilidad de pensar que fueron generadas por series con raíces unitarias y proceder a estacionarizarlas usando diferenciación entera.

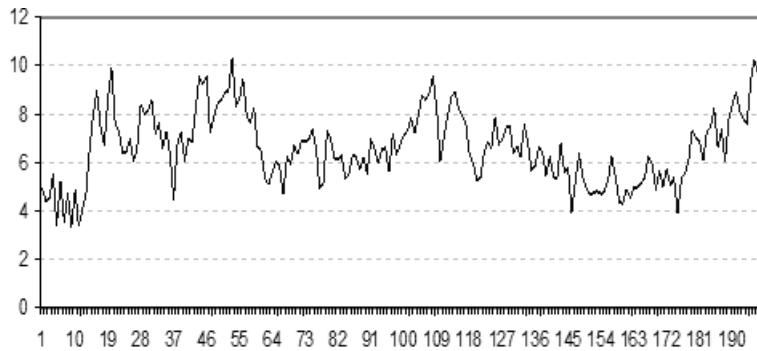


Figura 1: Simulación de un proceso ARFIMA estacionario

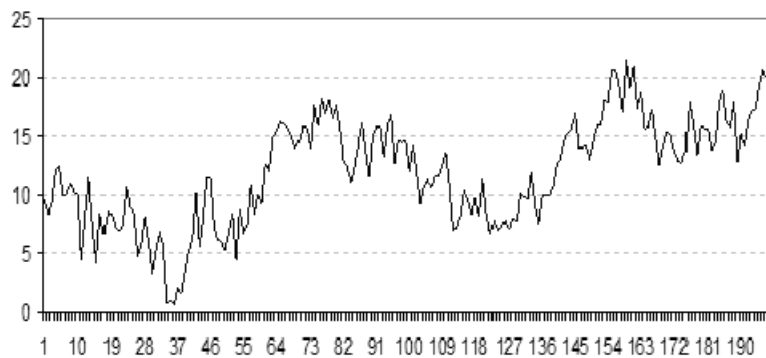


Figura 2: Simulación de un proceso ARFIMA no estacionario

Para tratar esta clase de series de tiempo, Granger (1980), Granger & Joyeux (1980) y Hosking (1981) proponen una clase de procesos intermedios en los que el orden de integración es fraccional. En estos modelos, la memoria de la serie es intermedia entre la memoria corta de los modelos ARMA, y la memoria persistente de los modelos ARIMA, lo cual permite que las innovaciones de dichas series tengan efectos transitorios que perduran durante mucho tiempo pero que finalmente terminan desapareciendo. Este comportamiento es diferente al encontrado en los modelos estacionarios ARMA, en los cuales el efecto de la innovación es transitoria y desaparece rápidamente en forma exponencial, y al encontrado en los modelos no estacionarios ARIMA en los que las innovaciones tienen efectos permanentes. Estos modelos son denominados procesos autorregresivos y de medias móviles fraccionalmente integrados, y son denotados como $ARFIMA(p, d, q)$, donde d es un número real. En este proceso el parámetro d describe las propiedades dinámicas en el largo plazo, mientras que la estructura de dependencia en el corto plazo es explicada por medio de los parámetros de la componente $ARMA(p, q)$.

A pesar de que por medio de estas dos componentes los procesos ARFIMA parecen representar adecuadamente series de tiempo en muchas áreas del conocimiento, su identificación en el trabajo aplicado puede resultar difícil debido a que ellas tienden a confundirse en dicho proceso. Ahora

bien, la selección errónea del modelo puede conducir a estimaciones sistemáticamente sesgadas en el parámetro de memoria larga, lo que podría conducir a decisiones equivocadas sobre la persistencia de la serie de tiempo. Algunos autores muestran que la escogencia errónea de la componente de corto plazo puede conducir a una estimación muy equivocada del parámetro d . Por otro lado, un sesgo en la estimación del parámetro d puede afectar también la selección de la componente de corto plazo del modelo ARFIMA.

Este artículo propone un procedimiento para la identificación correcta cuando el proceso generador de la serie de tiempo es estacionario de memoria larga. Dicho procedimiento parece comportarse mejor que algunos otros de los sugeridos en la literatura.

En este artículo, en la sección 2 se describe el modelo ARFIMA; en la sección 3 se presenta la metodología de identificación. La descripción del estudio de simulación y sus resultados se encuentran en la sección 4. En la sección 5 se presenta la aplicación del procedimiento a una serie simulada. Finalmente, en la sección 6 se presentan las conclusiones.

2. EL MODELO ARFIMA

Se dice que un proceso estocástico Z_t sigue un proceso ARFIMA(p, d, q) si es una solución a la ecuación:

$$\phi(B)(1-B)^d Z_t = \theta_0 + \theta(B)a_t, \quad t = 1, \dots, T, \quad (1)$$

donde $\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ y $\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$ son, respectivamente, los polinomios autorregresivo y de medias móviles de orden p y q de un proceso ARMA, cuyos ceros están fuera del círculo unidad y no tienen raíces comunes; d y θ_0 son números reales, d es llamado el parámetro de diferenciación fraccional, a_t son variables aleatorias no observables independientes e idénticamente distribuidas con media cero y varianza finita σ_a^2 , y

$$(1-B)^d = 1 - dB - (1/2!)d(1-d)B^2 - (1/3!)d(1-d)(2-d)B^3 \dots, \quad (2)$$

es el operador de diferencia fraccional, definido para $d > -1$.

Hosking (1981) mostró que si $d > -0.5$ y todas las raíces de $\theta(B) = 0$ caen fuera del círculo unidad, Z_t es un proceso invertible; si $d < 0.5$ y todas las raíces de $\phi(B) = 0$ caen fuera del círculo unidad, Z_t es un proceso estacionario. Por tanto, el proceso ARFIMA(p, d, q) es estacionario e invertible si todas las raíces de $\theta(B) = 0$ y de $\phi(B) = 0$ caen fuera del círculo unidad y $-0.5 < d < 0.5$.

Hosking (1981) prueba que cuando $0 < d < 0.5$ existe una constante positiva tal que:

$$\rho_k \approx Ck^{2d-1}, \quad (3)$$

esto implica que la función de autocorrelación decae hiperbólicamente hacia cero y no es absolutamente sumable, es decir, $\sum_{k=-\infty}^{\infty} |\rho_k|$ no converge. La expresión exacta de las autocorrelaciones se deriva en Hosking (1981).

Cuando $d < 0.5$, es decir, Z_t es un proceso estacionario, la función de densidad espectral del proceso ARFIMA es:

$$f_Z(\omega) = |1 - e^{-i\omega}|^{-2d} f_W(\omega) = \left[2 \operatorname{sen} \left(\frac{\omega}{2} \right) \right]^{-2d} f_W(\omega), \quad 0 \leq \omega \leq \pi, \quad (4)$$

donde $f_W(\omega) = \frac{\sigma_a^2}{2\pi} \left| \frac{\theta(e^{-i\omega})}{\phi(e^{-i\omega})} \right|^2$ es la densidad espectral del proceso ARMA(p, q), $W_t = (1 - B)^d Z_t$.

Para valores $0 < d < 0.5$, la densidad espectral es una función decreciente de ω no acotada en el origen, y está dominada por las frecuencias bajas. En este caso el proceso ARFIMA tiene memoria larga. Esto muestra la relación directa que hay entre la persistencia de las correlaciones rezagadas lejanas y la dinámica del espectro en las frecuencias bajas.

Si $d < 0$, la densidad espectral se anula en el origen y está dominada por las frecuencias altas. En este caso, se dice que el proceso ARFIMA presenta dependencia negativa o antipersistencia, y el proceso tiene memoria corta.

Si $d > 0.5$, el proceso es, en general, no estacionario. Sin embargo, si $0.5 < d < 1.0$ el proceso es no estacionario pero tiene reversión a la media, mientras que si $d \geq 1$ el proceso es no estacionario sin reversión a la media.

3. METODOLOGÍA PARA LA IDENTIFICACIÓN DEL MODELO ARFIMA(p, d, q)

El proceso ARFIMA(p, d, q) proporciona una gran flexibilidad en la modelación del comportamiento dinámico de una serie temporal. Permite describir simultáneamente las propiedades dinámicas en el largo plazo a través del parámetro d, y la correlación serial en el corto plazo a través de los parámetros de la componente ARMA del modelo.

Debido a que en un modelo ARFIMA la dependencia de largo plazo está concentrada en el parámetro de diferenciación fraccional, la estimación precisa del parámetro d es muy importante en el trabajo aplicado. Además, Boes *et al.* (1989) señalan que la elección de una especificación incorrecta para el modelo conduce a estimadores sesgados del parámetro d sobre todo en muestras pequeñas, lo que implicaría conclusiones erróneas sobre la persistencia de la serie.

Para la estimación del parámetro de diferenciación fraccional, Cheung (1990) muestra que varios métodos basados en máxima verosimilitud presentan un comportamiento superior a los métodos semiparamétricos como el sugerido por Geweke & Porter-Hudak (1983). Sin embargo, en contraste con los métodos semiparamétricos, para la aplicación en el mundo real de los procedimientos de máxima verosimilitud es necesario que el modelo sea especificado completamente. Schmidt & Tschernig (1993) señalan que “Es de esperarse que tanto la selección como la estimación esté plagada por los mismos problemas de muestras pequeñas. Aún así, este tema no ha recibido mucha atención (para una excepción, véase Hauser (1997)).”

En la literatura se sugieren algunos procedimientos para la identificación del modelo ARFIMA. A continuación se presentan algunos de los métodos propuestos.

3.1. Procedimiento basado en el empleo de criterios de información

Schmidt & Tschernig (1993) presentan un estudio Monte Carlo en el cual emplean los criterios de información de Akaike, el criterio de Schwarz, el criterio de información Bayesiano y el criterio de Hannan-Quinn para seleccionar el modelo correcto. En general, ellos concluyen que la correcta identificación del modelo por este método sólo se logra en una pequeña fracción de los experimentos. Esa fracción depende del conjunto de modelos alternativos considerados. También encuentran que la selección del verdadero modelo puede no asegurarse aún para muestras moderadamente grandes, aunque la probabilidad de éxito aumenta sustancialmente con el incremento de la longitud de las series analizadas. También concluyen que el criterio de Schwarz tiene el mejor comportamiento para identificar un proceso de ruido blanco fraccional (donde no existe componente ARMA).

Para modelos ARFIMA más generales, donde hay componente de corto plazo ARMA, sus resultados demuestran que el comportamiento de todos los criterios cae sustancialmente, y que el criterio de Schwarz pierde su liderazgo frente al AIC, mientras que el criterio de Hannan-Quinn presenta un comportamiento estable a través de todos los modelos considerados en los experimentos.

3.2. Procedimiento basado en la estimación inicial del parámetro de diferenciación fraccional

Algunos autores tales como Wei (2006), Reisen *et al.* (2001), Castaño *et al.* (2008), proponen un procedimiento de identificación de un modelo ARFIMA similar al proceso de identificación en un modelo ARIMA sugerido por Box & Jenkins (1976). Este procedimiento se basa en determinar primero a d , a continuación diferenciar la serie empleando este valor de d , y finalmente usar la serie diferenciada para seleccionar los órdenes adecuados p y q , para la componente de corto plazo. Específicamente, Reisen *et al.* (2001) sugieren un procedimiento iterativo cuyas etapas se describen a continuación, para una serie de tiempo X_t .

- 1) Estime d en el modelo ARFIMA(p, d, q). Denote el estimador por \hat{d} .

- 2) Obtenga la diferencia fraccional $\widehat{U}_t = (1 - B)^{\widehat{d}} X_t$.
- 3) Usando el procedimiento de identificación de Box-Jenkins (o usando criterios de información) identifique p y q y estime los parámetros en el proceso ARMA(p, q), $\phi(B)\widehat{U}_t = \theta(B)a_t$.
- 4) Calcule la serie filtrada $\widehat{Y}_t = \frac{\widehat{\phi}(B)}{\widehat{\theta}(B)} X_t$.
- 5) Estime d en el modelo ARFIMA(0, d, 0), $(1 - B)^{\widehat{d}} \widehat{Y}_t = a_t$. El valor de \widehat{d} obtenido en esta etapa es el nuevo estimador de d.
- 6) Repita los pasos 2) a 5) hasta que los parámetros d, ϕ s y θ s converjan.

Los autores sugieren que los estimadores obtenidos en la primera iteración (pasos 1), 2) y 3)) son suficientes en la práctica.

Basados en una serie de tiempo de longitud n, a continuación se presentan dos estimadores semiparamétricos tradicionales para el parámetro de diferenciación fraccional y el método paramétrico propuesto.

a) El estimador GPH

Para procesos gaussianos estacionarios, la densidad espectral del proceso ARFIMA(p, d, q) está dada en la ecuación 3. Tomando logaritmo a ambos lados de la ecuación (3), se obtiene:

$$\ln f_Z(\omega) = \ln f_W(0) + d \ln |1 - e^{-i\omega}|^{-2} + \ln \left(\frac{f_W(\omega)}{f_W(0)} \right) \quad (5)$$

Reemplazando ω por las frecuencias de Fourier $\omega_j = 2\pi j/n, j = 1, 2, \dots, [n/2]$ y agregando el periodograma de $Z_t, I_Z(\omega_j)$, a ambos lados, se obtiene:

$$\ln I_Z(\omega_j) = \ln f_W(0) + d \ln |1 - e^{-i\omega}|^{-2} + \ln \left(\frac{f_W(\omega_j)}{f_W(0)} \right) + \ln \left(\frac{I_Z(\omega_j)}{f_Z(\omega_j)} \right), \quad (6)$$

donde

$$I_Z(\omega_j) = \frac{1}{2\pi} \left(\widehat{\gamma}(0) + 2 \sum_{l=1}^{n-1} \widehat{\gamma}(l) \cos(l\omega_j) \right) \quad (7)$$

siendo $\widehat{\gamma}(k) = \frac{1}{n} \sum_{t=1}^{n-k} (Z_t - \bar{Z})(Z_{t+k} - \bar{Z})$ para $k = 0, 1, 2, \dots, n-1$ la autocovarianza muestral del proceso Z_t . Para ω_j cercanas a cero, es decir, para $j = 1, 2, \dots, m \ll (n/2)$ y tal que $m/n \rightarrow 0$ cuando $n \rightarrow \infty$, se tiene que $\ln \left(\frac{f_W(\omega_j)}{f_W(0)} \right) \approx 0$ y la ecuación anterior se puede escribir como:

$$\Upsilon_j = c + dX_j + e_j, \quad (8)$$

donde $\Upsilon_j = \ln I_Z(\omega_j)$, $c = \ln f_W(0)$, $X_j = \ln|1 - e^{-i\omega}|^{-2} = \ln\left[\frac{1}{4[\text{sen}(\omega_j/2)]^2}\right]$ y la sucesión $e_j = \ln\left(\frac{I_Z(\omega_j)}{f_Z(\omega_j)}\right)$, son variables aleatorias i.i.d.

Geweke & Porter-Hudak (1983) sugieren obtener el estimador de d , usando OLS sobre la ecuación (4). Ellos mostraron que

$$\widehat{d} \xrightarrow{d} N\left(d, \frac{\pi^2}{6 \sum_{j=1}^m (X_j - \bar{X})^2}\right) \quad (9)$$

Con frecuencia se toma $m = n^\alpha$, $0 < \alpha < 1$.

b) El estimador SPR

Brockwell & Davis (2006) muestran que el periodograma no es un estimador consistente de la función de densidad espectral. Reisen (1994), propuso usar un estimador consistente, el cual es una versión suavizada del periodograma denominado el estimador SPR.

El estimador SPR se obtiene reemplazando el periodograma por el periodograma suavizado dado por

$$I_S(\omega) = \frac{1}{2\pi} \sum_{j=-v}^v \kappa\left(\frac{j}{v}\right) \widehat{\gamma}(j) \cos(j\omega) \quad (10)$$

donde $\kappa(\cdot)$ es la ventana de Parzen dada por

$$\kappa(u) = \begin{cases} 1 - 6u^2 + 6|u|^3, & \text{si } |u| \leq \frac{1}{2} \\ 2(1 - |u|)^3, & \text{si } \frac{1}{2} < |u| \leq \frac{1}{2} \\ 0, & \text{en otro caso} \end{cases} \quad (11)$$

El estimador SPR se obtiene aplicando OLS al modelo de regresión dado en la ecuación (4), usando el periodograma suavizado $I_S(\omega_j)$ en vez del periodograma $I_Z(\omega_j)$. El punto de truncamiento $v = n^\beta$, con $0 < \beta < 1$. Algunos autores, entre ellos Agiakloglou *et al.* (1993), señalan la pérdida de eficiencia del estimador GPH en muestras finitas. Señalan que cuando existe una componente AR(1) o MA(1) con parámetro cerca a la unidad, el estimador tiene un sesgo enorme y es muy ineficiente. Robinson (1995b) señala que el supuesto de normalidad del proceso es muy restrictivo. Para el caso del estimador SPR, Reisen *et al.* (2001) reportan pérdida de eficiencia para el caso donde existen componentes de corto plazo AR(1) o MA(1).

c) La metodología propuesta

Para series estacionarias, Castaño *et al.* (2008) proponen una prueba para investigar la existencia de memoria larga y obtener un estimador inicial para el parámetro d , basados en una aproximación autorregresiva finita de la componente de corto plazo de un modelo ARFIMA(p , d , q) estacionario e invertible. Especificando el modelo (1) alternativamente como:

$$(1 - B)^d \pi(B) Z_t = a_t \quad (12)$$

donde $\pi(B) = \theta_q^{-1}(B) \phi_p(B) = 1 - \pi_1 B - \pi_2 B^2 - \dots$, es la componente dual autorregresiva del modelo de corto plazo ARMA(p , q) del modelo ARFIMA(p , d , q), los autores proponen realizar la prueba de memoria corta ($d = 0$) contra la alternativa de memoria larga ($d > 0$) aproximando el polinomio infinito $\pi(B)$ por medio de un polinomio autorregresivo finito $\pi^*(B)$ donde $\pi^*(B) = 1 - \pi_1^* B - \pi_2^* B^2 - \dots - \pi_{p^*}^* B^{p^*}$ para un orden adecuado de p^* . La prueba se lleva a cabo realizando estimación de máxima verosimilitud en el modelo aproximado ARFIMA(p^* , d , 0)

$$(1 - B)^d \pi^*(B) Z_t = a_t \quad (13)$$

Basados en esta aproximación, el estadístico para probar la hipótesis nula de memoria corta $H_0 : d = 0$, contra la alternativa de memoria larga, $H_1 : d > 0$, está dado por:

$$t_d = \frac{\widehat{d}}{se(\widehat{d})} \xrightarrow{d} N(0, 1), \quad (14)$$

donde \widehat{d} es el estimador de máxima verosimilitud del parámetro d , y $se(\widehat{d})$ es su error estándar, obtenidos del modelo dado en (6).

Castaño *et al.* (2008) mostraron que usando una aproximación autorregresiva dada por el entero más próximo a $p^* = n^{1/3}$ (Véase Said y Dickey(1984), para una aproximación autorregresiva en modelos ARIMA), la prueba mantiene en general un tamaño promedio adecuado y una potencia mayor que las pruebas anteriormente mencionadas. Adicionalmente, Castaño *et al.* (2010), muestran que los resultados obtenidos por Castaño *et al.* (2008) pueden mejorarse en potencia y tamaño empleando una aproximación autorregresiva de $p^* = \lceil n^{1/4} \rceil$.

4. EXPERIMENTO MONTECARLO

El procedimiento para la identificación del modelo ARFIMA se basa en: i) simular la serie ARFIMA, ii) usar los tres procedimientos descritos anteriormente para estimar el parámetro d , iii) aplicar luego la diferenciación fraccional a la serie usando cada estimador, y iv) usar un procedimiento automático para la identificación del modelo ARMA que queda en la serie diferenciada. El procedimiento empleado en esta etapa es la función *auto.arima* del paquete *forecast* del programa computacional *R*. En este procedimiento, la estimación de los parámetros del modelo ARMA(p ,

q) seleccionado se realiza usando máxima verosimilitud condicional y luego máxima verosimilitud exacta. Las estimaciones obtenidas el método condicional son empleadas como valores iniciales en el método de estimación de máxima verosimilitud exacta, para obtener los estimadores definitivos de los parámetros. Para realizar las simulaciones de las series ARFIMA se utilizaron los paquetes *fracdiff* y *rugarch* de *R*. Se simularon los siguientes procesos:

Modelo 1. ARFIMA(0, d , 0) o ruido blanco fraccional, donde $d = 0.1, 0.25, 0.4, 0.45$.

Modelo 2. ARFIMA(1, d , 0), donde $\phi = 0.7, -0.7, d = 0.1, 0.25, 0.4, 0.45$.

Modelo 3. ARFIMA(0, d , 1), donde $\theta = 0.7, -0.7, d = 0.1, 0.25, 0.4, 0.45$.

Modelo 4. ARFIMA(1, d , 1), donde $\phi = 0.7, -0.7, \theta = 0.3, -0.3, d = 0.1, 0.25, 0.4, 0.45$.

En todos los casos $a_t \sim N(0, 1)$. Se emplearon realizaciones de tamaño $n=500$ y 1000 . Para cada caso el número de simulaciones fue de 1000 .

Las siguientes tablas presentan el comportamiento de los procedimientos GPH, SPR y el procedimiento propuesto (denominado Propuesta) para las 1000 simulaciones realizadas. Los resultados presentan el valor promedio del parámetro estimado (\bar{d}), y el número de éxitos en la identificación del verdadero modelo (éxitos), para cada procedimiento. La última columna muestra el promedio general de éxitos para cada tamaño muestral y el promedio del error cuadrático medio en raíz cuadrada, *recm*.

Modelo 1. ARFIMA(0, d, 0) o ruido blanco fraccional.

Tabla 1: Resultados para el modelo ARFIMA(0,d,0)

n = 500									
d = 0.10				d = 0.25					
Estimador	\bar{d}	recm	éxitos	Estimador	\bar{d}	recm	éxitos		
GPH	0.098	0.057	0.837	GPH	0.255	0.057	0.867		
SPR	0.061	0.135	0.466	SPR	0.217	0.135	0.508		
Propuesta	0.093	0.081	0.679	Propuesta	0.231	0.109	0.606		
d = 0.40				d = 0.45				Resultados globales	
Estimador	\bar{d}	recm	éxitos	Estimador	\bar{d}	recm	éxitos	éxitos	recm
GPH	0.405	0.059	0.785	GPH	0.455	0.063	0.618	0.777	0.059
SPR	0.371	0.141	0.447	SPR	0.423	0.150	0.327	0.437	0.140
Propuesta	0.378	0.122	0.608	Propuesta	0.409	0.140	0.655	0.637	0.113

n = 1000									
d = 0.10				d = 0.25					
Estimador	\bar{d}	recm	éxitos	Estimador	\bar{d}	recm	éxitos		
GPH	0.099	0.043	0.844	GPH	0.251	0.044	0.862		
SPR	0.071	0.109	0.485	SPR	0.224	0.109	0.461		
Propuesta	0.093	0.067	0.646	Propuesta	0.239	0.081	0.618		
d = 0.40				d = 0.45				Resultados globales	
Estimador	\bar{d}	recm	éxitos	Estimador	\bar{d}	recm	éxitos	prom éxitos	recm
GPH	0.403	0.045	0.802	GPH	0.457	0.045	0.685	0.798	0.044
SPR	0.382	0.115	0.461	SPR	0.432	0.117	0.350	0.439	0.113
Propuesta	0.394	0.093	0.596	Propuesta	0.428	0.112	0.648	0.627	0.088

Observando la última columna de los resultados globales por tamaño muestral, el GPH es el mejor procedimiento en términos generales, es decir, del número de éxitos en la identificación y del recm. El método propuesto le sigue, y el estimador SPR tiene el menor número de éxitos en la identificación, así como el recm mayor. Observando el promedio de las estimaciones del parámetro d (columna \bar{d}) el estimador SPR presenta la mayor subestimación. Cuando aumenta el tamaño muestral, los éxitos en la identificación aumentan ligeramente en el GPH y se mantienen aproximadamente iguales en el SPR y en el método propuesto. En cuanto a \bar{d} , el GPH presenta convergencia hacia los verdaderos valores. El SPR mejora su valor, pero conserva un sesgo negativo importante. El método propuesto, también se aproxima al verdadero valor, aunque en algunos casos evidencia un sesgo negativo, aunque menor que en el caso de SPR.

Tabla 2: Resultados para el modelo ARFIMA(1,d,0), $\phi=0.7$

n = 500									
d = 0.10				d = 0.25					
Estimador	\bar{d}	recm	éxitos	Estimador	\bar{d}	recm	éxitos		
GPH	0.577	0.481	0.057	GPH	0.727	0.481	0.000		
SPR	0.140	0.137	0.752	SPR	0.295	0.141	0.802		
Propuesta	0.091	0.098	0.828	Propuesta	0.210	0.152	0.711		
d = 0.40				d = 0.45				Resultados globales	
Estimador	\bar{d}	recm	éxitos	Estimador	\bar{d}	recm	éxitos	éxitos	recm
GPH	0.879	0.483	0.000	GPH	0.925	0.479	0.000	0.014	0.481
SPR	0.435	0.144	0.552	SPR	0.496	0.150	0.357	0.616	0.143
Propuesta	0.314	0.202	0.627	Propuesta	0.324	0.236	0.557	0.681	0.172

n = 1000									
d = 0.10				d = 0.25					
Estimador	\bar{d}	recm	éxitos	Estimador	\bar{d}	recm	éxitos		
GPH	0.528	0.430	0.178	GPH	0.298	0.143	0.816		
SPR	0.107	0.111	0.749	SPR	0.263	0.110	0.867		
Propuesta	0.091	0.083	0.865	Propuesta	0.213	0.125	0.745		
d = 0.40				d = 0.45				Resultados globales	
Estimador	\bar{d}	recm	éxitos	Estimador	\bar{d}	recm	éxitos	prom éxitos	recm
GPH	0.448	0.149	0.517	GPH	0.495	0.149	0.361	0.468	0.218
SPR	0.412	0.115	0.658	SPR	0.463	0.114	0.443	0.679	0.113
Propuesta	0.344	0.161	0.688	Propuesta	0.357	0.200	0.615	0.728	0.143

De acuerdo a los resultados globales por tamaño muestral, el estimador GPH presenta el menor número de éxitos en la identificación (solamente el 1.4 %, en promedio) y el mayor recm. El estimador SPR presenta el menor recm, mientras que el método propuesto presenta la mejor tasa de identificación. En cuanto a \bar{d} , para $n=500$, el GPH presenta un sesgo positivo importante, que crece con el tamaño de d , y en todos los casos $\bar{d} > 0.5$. El estimador SPR, para $n = 500$, presenta sesgo positivo, aunque de menor tamaño que en el caso del GPH. El método propuesto presenta un sesgo negativo. Cuando aumenta el tamaño muestral, los tres métodos disminuyen el recm, siendo el GPH el que obtiene la mayor ganancia; el número de éxitos en la identificación, aumenta significativamente en el GPH, pero sigue siendo inferior al de los otros métodos. En el SPR y el método propuesto, el sesgo disminuye y el número de éxitos aumenta, manteniéndose mejor el método propuesto. En cuanto a \bar{d} , los tres métodos presentan convergencia hacia los verdaderos valores.

Modelo 2.2. ARFIMA(1,d,0), $\phi = -0.7$ Tabla 3: Resultados para el modelo ARFIMA(1,d,0), $\phi = -0.7$

n = 500									
d = 0.10				d = 0.25					
Estimador	\bar{d}	recm	éxitos	Estimador	\bar{d}	recm	éxitos		
GPH	-0.010	0.125	0.334	GPH	0.142	0.124	0.519		
SPR	0.058	0.137	0.521	SPR	0.205	0.142	0.559		
Propuesta	0.092	0.075	0.793	Propuesta	0.225	0.099	0.700		
d = 0.40				d = 0.45				Resultados globales	
Estimador	\bar{d}	recm	éxitos	Estimador	\bar{d}	recm	éxitos	prom éxitos	recm
GPH	0.296	0.119	0.539	GPH	0.345	0.121	0.525	0.479	0.122
SPR	0.366	0.142	0.522	SPR	0.414	0.144	0.434	0.509	0.141
Propuesta	0.376	0.117	0.693	Propuesta	0.398	0.152	0.743	0.732	0.111

n = 1000									
d = 0.10				d = 0.25					
Estimador	\bar{d}	recm	éxitos	Estimador	\bar{d}	recm	éxitos		
GPH	0.022	0.090	0.478	GPH	0.174	0.088	0.566		
SPR	0.075	0.109	0.524	SPR	0.219	0.119	0.539		
Propuesta	0.092	0.066	0.736	Propuesta	0.234	0.081	0.723		
d = 0.40				d = 0.45				Resultados globales	
Estimador	\bar{d}	recm	éxitos	Estimador	\bar{d}	recm	éxitos	prom éxitos	recm
GPH	0.324	0.088	0.527	GPH	0.376	0.086	0.529	0.525	0.088
SPR	0.875	0.120	0.490	SPR	0.427	0.117	0.415	0.492	0.116
Propuesta	0.891	0.087	0.699	Propuesta	0.416	0.126	0.729	0.722	0.090

Observando los resultados globales por tamaño muestral, el método propuesto presenta el mayor número de éxitos en la identificación y el menor recm. El SPR, tiene el mayor recm de los tres métodos. Cuando aumenta el tamaño muestral, los tres métodos disminuyen el recm; el número de éxitos en la identificación, aumenta en el GPH, mientras que para el SPR y el método propuesto permanecen aproximadamente iguales. En cuanto a \bar{d} , el GPH presenta un sesgo negativo importante, seguido por el SPR. Para los tres métodos, \bar{d} tiende al verdadero valor.

Modelo 3.1. ARFIMA(0,d,1), $\theta = 0.7$

Tabla 4: Resultados para el modelo ARFIMA(0,d,1), $\theta = 0.7$

n = 500									
d = 0.10				d = 0.25					
Estimador	\bar{d}	recm	éxitos	Estimador	\bar{d}	recm	éxitos		
GPH	0.025	0.121	0.542	GPH	0.359	0.124	0.500		
SPR	0.062	0.136	0.456	SPR	0.209	0.146	0.474		
Propuesta	0.159	0.119	0.559	Propuesta	0.298	0.120	0.518		
d = 0.40				d = 0.45				Resultados globales	
Estimador	\bar{d}	recm	éxitos	Estimador	\bar{d}	recm	éxitos	prom éxitos	recm
GPH	0.513	0.128	0.249	GPH	0.567	0.133	0.089	0.345	0.127
SPR	0.362	0.140	0.409	SPR	0.418	0.142	0.253	0.398	0.141
Propuesta	0.414	0.107	0.527	Propuesta	0.413	0.144	0.512	0.529	0.122

n = 1000									
d = 0.10				d = 0.25					
Estimador	\bar{d}	recm	éxitos	Estimador	\bar{d}	recm	éxitos		
GPH	0.180	0.090	0.550	GPH	0.330	0.091	0.497		
SPR	0.075	0.111	0.466	SPR	0.221	0.114	0.466		
Propuesta	0.074	0.067	0.616	Propuesta	0.211	0.094	0.532		
d = 0.40				d = 0.45				Resultados globales	
Estimador	\bar{d}	recm	éxitos	Estimador	\bar{d}	recm	éxitos	prom éxitos	recm
GPH	0.482	0.093	0.364	GPH	0.532	0.093	0.144	0.389	0.092
SPR	0.378	0.113	0.371	SPR	0.434	0.113	0.292	0.399	0.112
Propuesta	0.360	0.106	0.504	Propuesta	0.386	0.146	0.469	0.530	0.103

Los resultados globales por tamaño muestral muestran que el método propuesto presenta el mayor número de éxitos en la identificación, y obtiene un recm aproximadamente igual al del GPH. Cuando aumenta el tamaño muestral, los tres métodos disminuyen el recm; el número de éxitos en la identificación, aumenta en el GPH, mientras que para el SPR y el método propuesto permanecen aproximadamente iguales. Para los tres métodos, \bar{d} tiende al verdadero valor. En cuanto a \bar{d} , el GPH presenta un sesgo positivo importante. Para n=500 y d= 0.40 y 0.45, se obtiene $\bar{d} > 0.5$.

Modelo 3.2. ARFIMA(0,d,1), $\theta = -0.7$ Tabla 5: Resultados para el modelo ARFIMA(0,d,1), $\theta = -0.7$

n = 500									
d = 0.10				d = 0.25					
Estimador	\bar{d}	recm	éxitos	Estimador	\bar{d}	recm	éxitos		
GPH	-0.377	0.480	0.000	GPH	-0.229	0.483	0.000		
SPR	-0.016	0.175	0.426	SPR	0.131	0.181	0.789		
Propuesta	0.007	0.095	0.929	Propuesta	0.102	0.162	0.928		
d = 0.40				d = 0.45				Resultados globales	
Estimador	\bar{d}	recm	éxitos	Estimador	\bar{d}	recm	éxitos	prom éxitos	recm
GPH	-0.070	0.473	0.019	GPH	-0.024	0.478	0.040	0.015	0.479
SPR	0.298	0.170	0.870	SPR	0.339	0.178	0.821	0.727	0.176
Propuesta	0.274	0.149	0.937	Propuesta	0.340	0.144	0.933	0.932	0.187

n = 1000									
d = 0.10				d = 0.25					
Estimador	\bar{d}	recm	éxitos	Estimador	\bar{d}	recm	éxitos		
GPH	-0.328	0.430	0.000	GPH	-0.178	0.430	0.000		
SPR	0.034	0.125	0.598	SPR	0.184	0.124	0.904		
Propuesta	0.013	0.090	0.938	Propuesta	0.138	0.124	0.931		
d = 0.40				d = 0.45				Resultados globales	
Estimador	\bar{d}	recm	éxitos	Estimador	\bar{d}	recm	éxitos	prom éxitos	recm
GPH	0.367		0.798	GPH	0.025	0.428	0.050	0.212	0.357
SPR	0.334		0.908	SPR	0.391	0.131	0.786	0.799	0.127
Propuesta	0.269		0.950	Propuesta	0.371	0.109	0.930	0.937	0.116

Los resultados globales por tamaño muestral muestran que el método propuesto presenta el mayor número de éxitos en la identificación, y obtiene el menor recm. El GPH es el que presenta el menor número de éxitos en la identificación (para $n=500$, la tasa de éxitos es sólo de 1.5%), y el mayor recm. Para el GPH, cuando $n=500$, \bar{d} presenta sesgo importante y toma valores negativos; para $n=1000$ y valores bajos de d , también $\bar{d} < 0$. En el caso del SPR, para $n=500$ y $d=0.10$, $\bar{d} < 0$. Cuando aumenta el tamaño muestral, los tres métodos disminuyen el recm y aumentan el número de éxitos en la identificación.

Tabla 6: Resultados para el modelo ARFIMA(0,d,1), $\theta = -0.7$

n = 500									
d = 0.10				d = 0.25					
Estimador	\bar{d}	recm	éxitos	Estimador	\bar{d}	recm	éxitos		
GPH	0.651	0.554	0.001	GPH	0.793	0.547	0.000		
SPR	0.142	0.142	0.584	SPR	0.280	0.144	0.663		
Propuesta	0.094	0.106	0.721	Propuesta	0.191	0.158	0.670		
d = 0.40				d = 0.45				Resultados globales	
Estimador	\bar{d}	recm	éxitos	Estimador	\bar{d}	recm	éxitos	prom éxitos	recm
GPH	0.950	0.553	0.000	GPH	0.998	0.551	0.000	0.000	0.551
SPR	0.450	0.151	0.464	SPR	0.493	0.145	0.311	0.506	0.146
Propuesta	0.280	0.220	0.523	Propuesta	0.273	0.272	0.408	0.581	0.189

n = 1000									
d = 0.10				d = 0.25					
Estimador	\bar{d}	recm	éxitos	Estimador	\bar{d}	recm	éxitos		
GPH	0.581	0.483	0.016	GPH	0.733	0.485	0.000		
SPR	0.109	0.104	0.759	SPR	0.265	0.113	0.876		
Propuesta	0.084	0.082	0.837	Propuesta	0.204	0.133	0.779		
d = 0.40				d = 0.45				Resultados globales	
Estimador	\bar{d}	recm	éxitos	Estimador	\bar{d}	recm	éxitos	prom éxitos	recm
GPH	0.884	0.486	0.000	GPH	0.495	0.484	0.361	0.094	0.400
SPR	0.414	0.113	0.660	SPR	0.463	0.114	0.443	0.685	0.110
Propuesta	0.302	0.193	0.637	Propuesta	0.357	0.241	0.615	0.717	0.151

Los resultados globales por tamaño muestral muestran que el método propuesto presenta el mayor número de éxitos en la identificación. Su recm es el segundo en tamaño después del de SPR. El GPH tiene muy pobre desempeño, pues para el caso de $n=500$, no consigue identificar el modelo, y en el caso de $n=1000$, sólo lo logra el 9.4% de las veces. Además su recm es el mayor en todos los casos. Para el GPH, \bar{d} presenta sesgo positivo muy importante, presentado un rango de estimación fuera del intervalo de estacionaridad (excepto para el caso de $n=1000$, donde queda en el borde de la no estacionaridad). Cuando aumenta el tamaño muestral, los tres métodos disminuyen el recm y aumentan el número de éxitos en la identificación.

Modelo 4.2. ARFIMA(1,d,1), $\phi = -0.7$, $\theta = -0.3$

Tabla 7: Resultados para el modelo ARFIMA(1,d,1), $\phi = -0.7$, $\theta = -0.3$

n = 500									
d = 0.10				d = 0.25					
Estimador	\bar{d}	recm	éxitos	Estimador	\bar{d}	recm	éxitos		
GPH	-0.143	0.250	0.000	GPH	0.007	0.250	0.069		
SPR	0.054	0.139	0.520	SPR	0.204	0.139	0.532		
Propuesta	0.083	0.067	0.680	Propuesta	0.230	0.084	0.682		
d = 0.40				d = 0.45				Resultados globales	
Estimador	\bar{d}	recm	éxitos	Estimador	\bar{d}	recm	éxitos	prom éxitos	recm
GPH	0.158	0.249	0.137	GPH	0.210	0.247	0.145	0.088	0.249
SPR	0.352	0.144	0.430	SPR	0.402	0.146	0.312	0.449	0.142
Propuesta	0.385	0.091	0.704	Propuesta	0.417	0.122	0.714	0.695	0.091

n = 1000									
d = 0.10				d = 0.25					
Estimador	\bar{d}	recm	éxitos	Estimador	\bar{d}	recm	éxitos		
GPH	-0.091	0.196	0.006	GPH	0.062	0.193	0.106		
SPR	0.065	0.114	0.656	SPR	0.220	0.119	0.737		
Propuesta	0.089	0.058	0.918	Propuesta	0.239	0.069	0.870		
d = 0.40				d = 0.45				Resultados globales	
Estimador	\bar{d}	recm	éxitos	Estimador	\bar{d}	recm	éxitos	prom éxitos	recm
GPH	0.215	0.189	0.104	GPH	0.264	0.188	0.098	0.079	0.193
SPR	0.369	0.119	0.626	SPR	0.423	0.116	0.511	0.633	0.117
Propuesta	0.396	0.072	0.899	Propuesta	0.425	0.107	0.856	0.886	0.078

Los resultados globales por tamaño muestral muestran que el método propuesto presenta el mayor número de éxitos en la identificación y su recm es el menor. El GPH tiene muy pobre desempeño, pues para el caso de $n=500$, la tasa de identificación correcta es sólo de 8.8%, mientras que cuando $n=1000$ es de 7.9%. Además posee el recm mayor. El GPH, también \bar{d} presenta sesgo negativo muy importante en general; observe que en el caso donde $d=0.1$, $\bar{d} < 0$. Cuando aumenta el tamaño muestral, los tres métodos disminuyen el recm y aumentan el número de éxitos en la identificación, excepto el GPH.

Resultados agregados

La Tabla 8 presenta los resultados agregados en términos de la proporción promedia de éxitos, para cada procedimiento para cada valor del parámetro d y $n=500, 1000$.

Tabla 8: Resultados agregados

n = 500			
d = 0.1		d = 0.25	
Estimador	Prom éxitos	Estimado	Prom éxitos
GPH	0.197	GPH	0.217
SPR	0.461	SPR	0.368
PROP	0.680	PROP	0.638
d = 0.40		d = 0.45	
Estimador	Prom éxitos	Estimado	Prom éxitos
GPH	0.194	GPH	0.162
SPR	0.507	SPR	0.404
PROP	0.617	PROP	0.606
n = 1000			
d = 0.1		d = 0.25	
Estimador	Prom éxitos	Estimado	Prom éxitos
GPH	0.230	GPH	0.316
SPR	0.537	SPR	0.639
PROP	0.722	PROP	0.681
d = 0.40		d = 0.45	
Estimador	Prom éxitos	Estimado	Prom éxitos
GPH	0.434	GPH	0.253
SPR	0.565	SPR	0.447
PROP	0.658	PROP	0.644

Los resultados muestran que la propuesta presenta, en promedio, mayor éxito en la identificación del verdadero modelo. Este número aumenta con el incremento de n . Los siguientes gráficos presentan la comparación de los resultados obtenidos para los procedimientos estudiados.

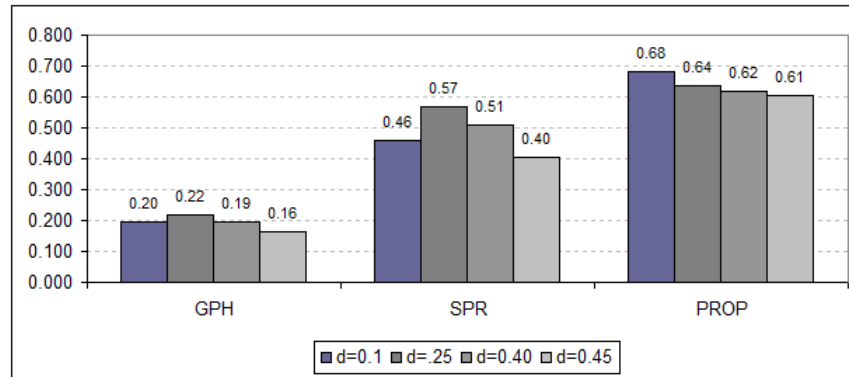


Figura 3: Porcentaje de éxitos para cada método para los distintos valores de d , $n=500$.

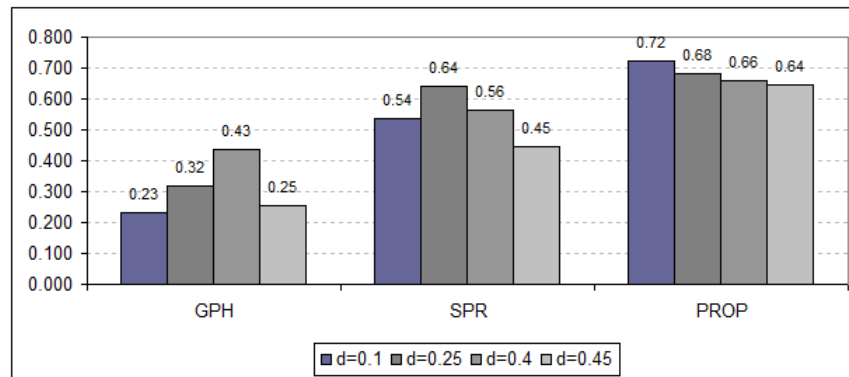


Figura 4: Porcentaje de éxitos para cada método para los distintos valores de d , $n=1000$.

Finalmente, promediando el número de éxitos y el $recm$ para las simulaciones de todos los modelos anteriores, se obtienen los siguientes resultados:

Tabla 9: Resultados totales

Estimador	Promedio de éxitos total	Promedio de $recm$ total
GPH	0.306	0.247
SPR	0.555	0.131
Propuesta	0.710	0.122

Se observa que en general, el método propuesto arroja la tasa más alta de identificación correcta y el menor $recm$. El GPH es el que tiene el desempeño más pobre.

5. APLICACIÓN DEL PROCEDIMIENTO A UNA SERIE SIMULADA

Para ilustrar el procedimiento de identificación propuesto, se simuló una realización de longitud $n=500$ de un proceso ARFIMA(1, d, 1) estacionario, con parámetros $d=0.4$, $\phi_1 = 0.7$, $\theta_1 = 0.3$ y a_t generada por un ruido blanco $N(0, 1)$. El siguiente gráfico presenta la realización obtenida, que puede ser confundida con la de un proceso no estacionario.

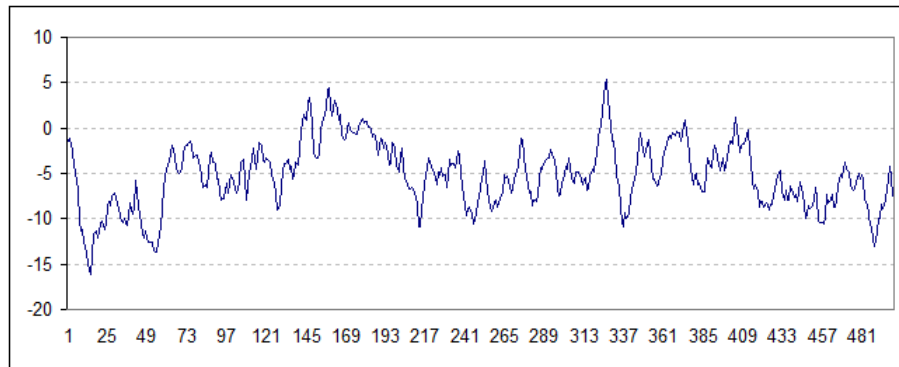


Figura 5: Realización de un proceso ARFIMA(1,d,1)

En la figura 6 se observa la función de autocorrelación muestral de la serie, la cual muestra un decrecimiento lento, típico de un proceso de memoria larga. Éste puede ser confundido con el comportamiento de la función de autocorrelación muestral de una realización de un proceso ARIMA.

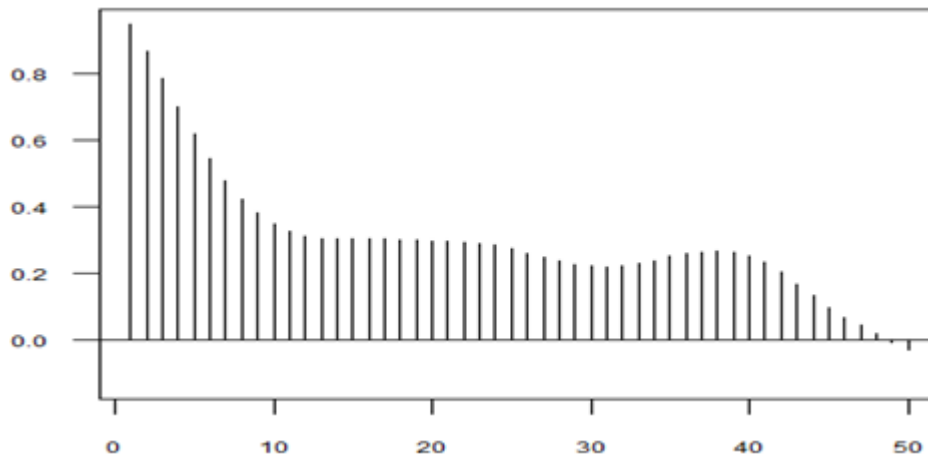


Figura 6: Autocorrelación muestras de la serie.

A continuación se presentan las etapas del procedimiento de identificación.

- 1) Estime el modelo aproximado ARFIMA(p^* , $d, 0$) donde p^* es la parte entera de $n^{1/4}$, es decir $p^*=4$. En este caso se usa la distribución normal en el proceso a_t . La tabla 10 presenta los resultados de la estimación. La estimación del parámetro d señala que la serie es estacionaria.

Tabla 10: Resultados de la estimación del modelo preliminar

Parámetro	Estimación	Error Est.	Valor t	Pr($> t $)
ar1	0.911104	0.108966	8.3614	0.000000
ar2	-0.258650	0.072060	-3.5893	0.000332
ar3	0.156400	0.061771	2.5319	0.011344
ar4	-0.058083	0.044934	-1.2926	0.196141
d	0.471814	0.098813	4.7698	0.000002
sigma	1.028464	0.032523	31.6228	0.000000

Algunos diagnósticos sobre el modelo ajustado.

- a) Prueba de adecuación del modelo Box-Ljung test:

$$Q(35) = 16.1807, \text{ valor-p} = 0.9938246$$

$$Q(25) = 12.1493, \text{ valor-p} = 0.9681026$$

$$Q(15) = 5.9231, \text{ valor-p} = 0.9489066$$

$$Q(10) = 4.3036, \text{ valor-p} = 0.8287455$$

$$Q(5) = 0.2646, \text{ valor-p} = 0.9665380$$

Los resultados muestran que el modelo aproxima bien la estructura de dependencia del corto plazo.

- b) Análisis de normalidad: A continuación se verifica el supuesto de normalidad usando el gráfico quantil-quantil y las pruebas de Shapiro-Wilk y Jarque-Bera.

Shapiro-Wilk Normality test

$$W = 0.9955, \text{ valor-p} = 0.155$$

Jarque-Bera Normality Test

$$X\text{-squared: } 4.3533, \text{ valor-p asintótico: } 0.1134$$

Los resultados no rechazan el supuesto de normalidad.

- 2) Diferenciación fraccional usando el estimador preliminar de d .

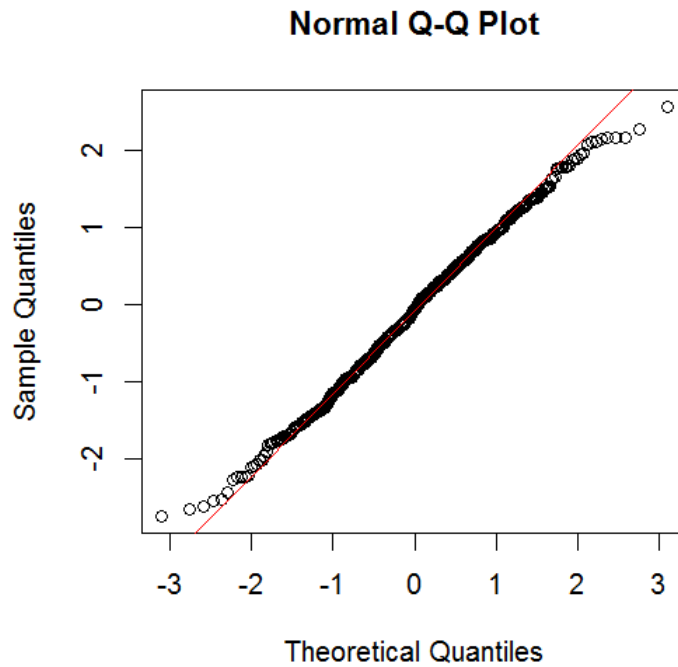


Figura 7: Gráfico cuantil-cuantil.

Se obtiene la serie diferenciada fraccionalmente, $Z_fracdif_t = (1 - B)^{\hat{d}}Z_t$ donde $\hat{d} = 0.471314$. La figura 8 presenta la nueva serie.

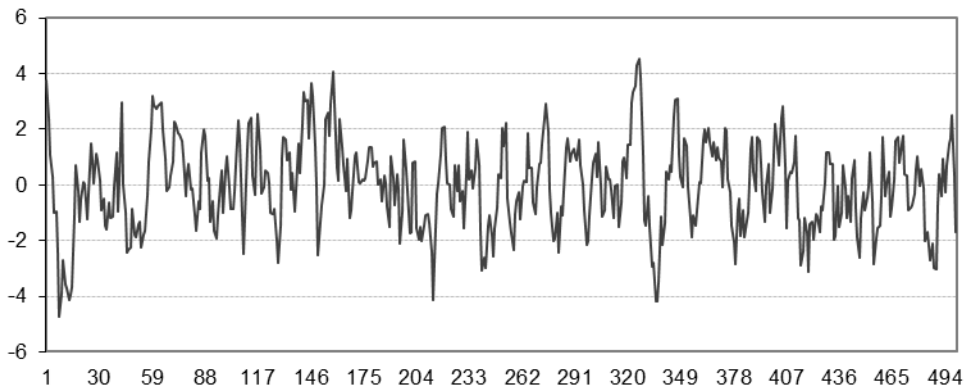


Figura 8: Serie diferenciada fraccionalmente.

3) Identificación de la componente ARMA de corto plazo.

Se emplearon distintas técnicas para identificar la componente ARMA.

- a) Las funciones de autocorrelación (ACF) muestral y autocorrelación parcial (PACF) muestral de $Z_{fracdif}$, se presentan en la figura 9.

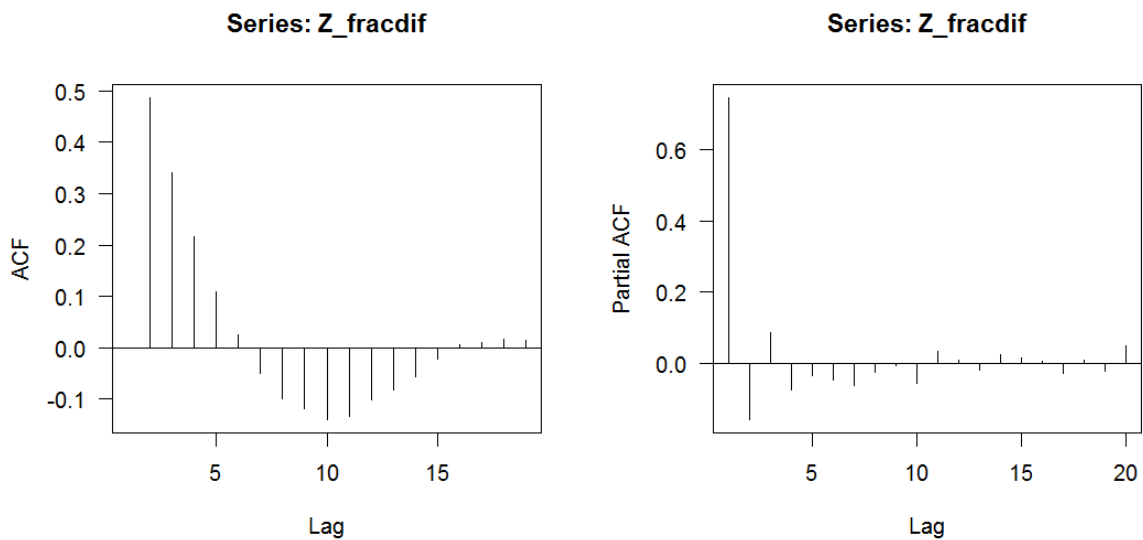


Figura 9: ACF y PACF muestrales de $Z_{fracdif}$

Tanto la ACF muestral como la PACF muestral muestran una tendencia a decrecer rápidamente, indicando la posible presencia de un proceso mixto, es decir con componente autorregresiva y de medias móviles. Sin embargo, este método no da una idea clara de los órdenes de dichas componentes.

- b) La función de autocorrelación extendida muestral, ESACF, proporciona la siguiente tabla.

El vértice de triángulo de ceros en negrilla (con menos distorsiones que otros posibles triángulos con vértices cero) señala la posibilidad de un $ARMA(1,1)$.

Tabla 11: Autocorrelación extendida muestral, EACF

AR/MA	0	1	2	3	4	5	6	7	8	9	10
0	x	x	x	x	x	o	o	x	x	x	x
1	x	o	o	x	x	o	o	o	o	o	o
2	x	o	o	o	o	o	o	o	o	o	o
3	x	x	x	o	o	o	o	o	o	o	o
4	x	o	x	o	o	o	o	o	o	o	o
5	x	x	x	o	x	o	o	o	o	o	o
6	x	x	x	o	x	o	o	o	o	o	o
7	x	x	x	o	x	o	o	o	o	o	o

c) Para buscar más evidencia sobre el ARMA(1,1), se empleó el criterio de información de Schwarz (SIC). La búsqueda proporciona la siguiente tabla (recortada a los primeros 10 modelos con menor SIC).

Tabla 12: Criterio de selección de modelos, SIC

Mod	p	q	SIC
9	1	1	2.937185
15	2	0	2.944254
16	2	1	2.946832
10	1	2	2.946977
22	3	0	2.949754
29	4	0	2.956050
11	1	3	2.958795
23	3	1	2.959102
17	2	2	2.959178
30	4	1	2.959623
18	2	3	2.959968
8	1	0	2.961265

El modelo con menor SIC (señalado en negrilla) es el ARMA(1,1). El empleo del procedimiento automático *auto.arima* del paquete *forecast* de *R*, también condujo al mismo modelo.

Con base en los resultados anteriores, el procedimiento de identificación conduce al modelo ARFIMA(1, d, 1), el cual fue el proceso que generó la realización analizada.

6. CONCLUSIONES

De los resultados obtenidos en el estudio de simulación, la metodología propuesta parece proporcionar mejoras importantes en la identificación de modelos ARFIMA estacionarios.

- I) A nivel agregado, la proporción de éxitos obtenidos en la identificación del modelo correcto usando el procedimiento propuesto es mayor que en los otros casos, mientras que su recm es menor.
- II) El estimador GPH presenta el mejor comportamiento en el caso de un proceso de ruido blanco fraccional. En los otros casos tiene el desempeño mas pobre.
- III) Todas las metodologías aumentan su capacidad para detectar el modelo correcto a medida que n crece.
- IV) La propuesta no es uniformemente mejor en todos los casos individuales.

La aplicación del procedimiento propuesto en la identificación del modelo de una serie que fue simulada, muestra su buen comportamiento.

Referencias

- Agiakloglou, C., Newbold, P. & Wohar, M. (1993). Bias in an estimator of the fractional difference parameter. *Journal of Time Series Analysis*, 14, 235–46.
- Beran, J. (1992). Statistical Methods for Data with Long-Range Dependence. *Statistical Science*, 7, 404–416.
- Boes, D.C., Davis, R.A. & Gupta, S.N. (1989). Parameter estimation in low order fractionally differenced ARMA processes. *Stochastic Hydrology and Hydraulics*, 3, 97–110.
- Box, G. E. P. & Jenkins, G. M. (1976). Time Series Analysis, Forecasting and Control. Holden-Day.
- Brockwell, P. J. & Davis, R. A. (2006). Time Series: Theory and Methods. Springer-Verlag.
- Castaño, E.; Gómez, K. & Gallón, S. (2008). Una nueva prueba para el parámetro de diferenciación fraccional. *Revista Colombiana de Estadística*, 31, 67–84.
- Castaño, E.; Gómez, K. & Gallón, S. (2010). Sesgos en estimación, tamaño y potencia de una prueba sobre el parámetro de memoria larga en modelos ARFIMA. *Lecturas de Economía*, 73, 131–146.

- Cheung, Y. (1990). Long memory in foreign exchange rates and Sampling Properties in Some Statistical Procedures of Long Memory Series. *PhD dissertation, University of Pennsylvania*.
- Geweke, J. & Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis*, 4(4), 221–238.
- Granger, C.W.J. (1980). Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics*, 14, 227–238.
- Granger, C.W.J. & R. Joyeux. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis*, 1, 15–29.
- Hauser, M. (1997). Semiparametric and Nonparametric Testing for Long Memory: A Monte Carlo Study. *Empirical Economics*, 22, 247–271.
- Hosking, J. R. M. (1981). Fractional differencing. *Biometrika*, 68(1), 165–176.
- Hurst, H.E. (1951). Methods for storage capacity of reservoirs. *Trans. Am. Soc. Civil Engineers*, 116, 770–799.
- Pérez, A. & Ruiz, E. (2002). Modelos de memoria larga para series económicas y financieras. *Investigaciones Económicas*, XXVI(3), 395–445.
- Reisen, Abraham & Lopes. (2001). Estimation of parameters in ARFIMA processes: a simulation study. *Communications in Statistics-Simulation and Computation*, 30(4), 787–803.
- Reisen, V. A. (1994). Estimation of the fractional difference parameter in the ARIMA(p, d, q) model using the smoothed periodogram. *Journal of Time Series Analysis*, 15(3), 335–350.
- Robinson, P.M. (1995b). Gaussian semiparametric estimation of long-range dependence. *Annals of Statistics*, 23, 1630–1661.
- Said, S.E. & Dickey D.A. (1984). Testing unit roots in autorregresive-Moving average model of unknown order. *Biometrika*, 71, 599–608.
- Schmidt, C. M. & Tschernig, R. (1993). Identification of fractional ARIMA models in the presence of long memory. *Artículo presentado en FAC Workshop on Economic Time Series Analysis and System Identification*.
- Wei, W.W.S. (2006). *Time Series Analysis, Univariate and Multivariate Methods*. Pearson Addison-Wesley.