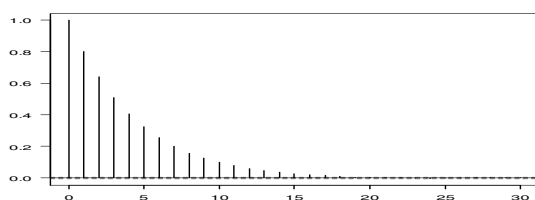


Series de Tiempo de Memoria Larga

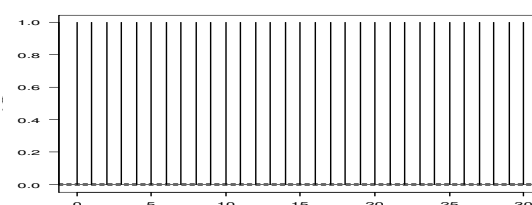
Introducción

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

ACF de un proceso ARMA



ACF de un proceso ARIMA



ACF de un proceso ARFIMA

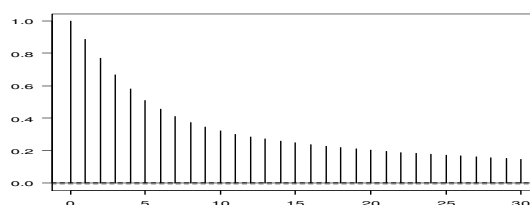


Figura 1: Comparación de las funciones de autocorrelación

- Formalmente, siguiendo la notación de Robinson (1994b) y Baillie (1996), se dice que una serie Z_t estacionaria con función de autocorrelación ρ_k

tiene memoria larga si las autocorrelaciones no son absolutamente sumables, es decir si $\sum_{k=0}^{\infty} |\rho_k| = \infty$

- Además de esta definición, existen otras posibles caracterizaciones de la propiedad de memoria larga.

- ✓ Por ejemplo, se dice que un proceso estacionario Z_t tiene memoria larga si los coeficientes de su correspondiente representación de

Wold, $Z_t = \sum_{j=0}^{\infty} \psi_j a_{t-j}$, siendo a_t ruido blanco, verifican la condición

$\sum_{j=0}^{\infty} \psi_j^2 < \infty$, pero no son absolutamente sumables, es decir

$$\sum_{j=0}^{\infty} |\psi_j| = \infty.$$

- Es importante señalar, Beran (1994), que estas definiciones son definiciones asintóticas: sólo nos informan del comportamiento de las correlaciones cuando $k \rightarrow \infty$.

- ✓ La propiedad de memoria larga sólo determina la tasa de convergencia hacia cero de las correlaciones, pero no la magnitud de cada una de ellas.

- ✓ Cada correlación individual puede ser arbitrariamente pequeña, pero el decrecimiento debe ser lento.

- La evidencia empírica sobre datos con memoria larga se remonta a mucho tiempo atrás; quizás el ejemplo más conocido sea el trabajo de Hurst (1951), en el campo de la hidrología. Sin embargo, muchos de los trabajos aplicados a otras ciencias y, sobre todo, muchos de los desarrollos formales sobre estimación y contrastes de hipótesis son relativamente recientes.

- ✓ El interés por los modelos con memoria larga para la modelación de series económicas surge a partir de los trabajos de Granger (1980) y Granger y Joyeux (1980).
- Granger (1980) y Granger y Joyeux (1980) advierten que la práctica habitual de diferenciar una serie hasta conseguir estacionariedad, puede tener consecuencias negativas en la correcta modelación de algunas series.
 - ✓ Muchas series económicas aparentemente no estacionarias suelen diferenciarse para conseguir varianza finita. Sin embargo, la serie diferenciada se convierte a menudo en una serie en la cual se ha eliminado de la serie original la componente de bajas frecuencias, que es muy importante en las predicciones a largo plazo.
 - ✓ Para modelar este tipo de series, la diferenciación es “excesiva” (sobrediferenciación) pero la no diferenciación tampoco es adecuada.
- Para cubrir este vacío entre los casos extremos de modelos con raíces unitarias, típicamente utilizados para series no estacionarias, y modelos estacionarios que imponen un decrecimiento exponencial de las autocorrelaciones, Granger (1980), Granger y Joyeux (1980) y Hosking (1981) proponen una clase de procesos intermedios en los que el **orden de integración es fraccionario**.
- Introducen los procesos autorregresivos y de medias móviles fraccionalmente integrados ARFIMA(p,d,q), donde d es un número real. Al permitir que el orden de integración d sea un número no entero positivo, estos modelos actúan como un “puente” entre los procesos con raíces unitarias ARIMA (d=1) y los procesos ARMA estacionarios que no requieren integración (d=0)

- Como se verá más adelante, los procesos ARFIMA producen memoria larga si $0 < d < 1/2$, en el sentido de que para ese rango de valores se verifica que las autocorrelaciones decrecen a un ritmo muy lento pero finalmente convergen hacia cero.
- Los modelos ARFIMA proporcionan una gran flexibilidad en la interpretación de la persistencia, entendida ésta como *el efecto que tiene un cambio unitario en el término de error del modelo (shock, innovación) sobre las predicciones del nivel futuro de la serie*.
 - ✓ Mientras en los procesos integrados ($d=1$), el efecto de un shock persiste indefinidamente, en un proceso fraccionalmente integrado con $0 < d < 1$, el efecto de un shock acaba desapareciendo y la serie revierte finalmente a su media, incluso en el intervalo $1/2 \leq d < 1$, donde el proceso no es estacionario.
 - ✓ En este sentido suele decirse que un ARFIMA no estacionario es “menos no estacionario” que los procesos con raíces unitarias.
- Entre la extensa literatura existente sobre procesos con memoria larga, tanto a nivel de resultados teóricos como de aplicaciones empíricas, se destacan las revisiones de Robinson (1994b) y Baillie (1996), Beran (1994) y Palma (2007). En ellos se revisan las propiedades más importantes de los procesos con memoria larga, en especial de los ARFIMA, y se discuten los métodos de estimación y contrastes de hipótesis que de ellos se derivan. Se incluyen asimismo referencias de los trabajos aplicados más relevantes, en particular aquellos relativos a la modelación de series macroeconómicas, como Diebold y Rudebush (1989) o Sowell (1992b), y aquellos aplicados a la modelación de series financieras, como Lo (1991), Cheung (1993a) o Baillie y Bollerslev (1994), entre otros.

Definición y Propiedades de los Modelos ARFIMA

- Un proceso Z_t es integrado de orden d , denotado por $Z_t \sim I(d)$, si:

$$(1 - B)^d Z_t = u_t \quad (1)$$

donde u_t es un proceso estacionario.

- ✓ Si $d=0$, Z_t es un proceso estacionario y decimos que es un proceso $I(0)$.
- ✓ Si $d=1$, Z_t tiene una raíz unitaria y decimos que es un proceso $I(1)$.
- ✓ Si d es otro número entero positivo, Z_t tiene d raíces unitarias y decimos que es un proceso $I(d)$.
- ✓ Finalmente, si d es un número real no entero, decimos que Z_t es un proceso fraccionalmente integrado. Si además el proceso u_t es un ARMA(p, q) de la forma $\phi(B)u_t = \theta(B)a_t$, con a_t ruido blanco, decimos que Z_t es un ARMA fraccionalmente integrado o ARFIMA(p, d, q).
- El más sencillo de los procesos fraccionalmente integrados es el ARFIMA(0, d , 0) o **ruido blanco fraccionalmente integrado**, en el cual el proceso u_t es ruido blanco ($u_t = a_t$). En este caso, la serie Z_t se modela según la ecuación:

$$(1 - B)^d Z_t = a_t$$

donde a_t es ruido blanco de media cero y varianza σ_a^2 , d es un número fraccionario y el polinomio $(1 - B)^d$ viene definido, para cualquier $d > -1$, por la serie binomial:

$$(1 - B)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-B)^k = 1 - dB - \frac{d(1-d)}{2!} B^2 - \frac{d(1-d)(2-d)}{3!} B^3 - \dots \quad (2)$$

- Para los valores del parámetro d en el intervalo $-1/2 < d < 1/2$ el proceso ARFIMA(0,d,0) es estacionario e invertible.
- ✓ Cuando $d > -1/2$, Hosking (1981) demuestra que Z_t es invertible, y tiene una representación $AR(\infty)$ de la forma:

$$(1-B)^d Z_t = \sum_{j=0}^{\infty} \pi_j Z_{t-j} = a_t \quad (3)$$

donde π_j son los coeficientes del polinomio $(1-B)^d$ en la expresión (2),

$$\pi_j = \frac{d(1-d)(2-d)\dots(j-1-d)}{j!} = \frac{\Gamma(j-d)}{\Gamma(-d)\Gamma(j+1)}$$

y donde $\Gamma(\cdot)$ denota la función gamma.

- ✓ Para los valores $d < 1/2$, Hosking (1981) demuestra que Z_t es estacionario y tiene una representación $MA(\infty)$ de la forma:

$$Z_t = (1-B)^{-d} a_t = \sum_{j=0}^{\infty} \psi_j a_{t-j} \quad (4)$$

$$\text{con } \psi_j = \frac{\Gamma(j+d)}{\Gamma(d)\Gamma(j+1)}$$

Utilizando la aproximación $\Gamma(a+x)/\Gamma(b+x) \approx x^{a-b}$, se comprueba que cuando $k \rightarrow \infty$, $\psi_k \approx \frac{k^{d-1}}{\Gamma(d)}$, y por tanto los coeficientes de la representación de Wold del proceso ARFIMA(0,d,0) convergen hiperbólicamente hacia cero, en contraposición con el decrecimiento exponencial típico de los procesos ARMA.

- La propiedad que más caracteriza a los procesos ARFIMA estacionarios ($d < 1/2$) es el comportamiento asintótico de su función de autocorrelación, o equivalentemente, el de su densidad espectral¹ en las bajas frecuencias. A continuación definiremos el concepto de densidad espectral.

Análisis de una serie de tiempo en el dominio de las frecuencias.

El análisis de una serie de tiempo presentado hasta ahora usa las funciones de autocorrelación y autocorrelación parcial para estudiar la evolución de una serie de tiempo a través de modelos paramétricos. Este análisis es llamado *análisis en el dominio del tiempo*. Una aproximación alternativa, la cual trata de describir la fluctuación de una serie de tiempo en términos de su comportamiento en ondas sinusoidales en las distintas frecuencias es llamado *análisis en el dominio de la frecuencia*.

Densidad espectral: Sea Z_t una proceso estacionario de valor real con función de autocovarianza absolutamente sumable, γ_k .

La transformada de Fourier de γ_k existe y está dada por

$$f(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_k e^{-i\omega k}$$

donde $-\pi \leq \omega \leq \pi$. $f(\omega)$ es llamada espectro o densidad espectral del proceso y ω es la frecuencia. El análisis en el dominio de la frecuencia se basa en $f(\omega)$. La función de autocovarianza puede ser recuperada por medio de la inversa de la transformada de Fourier

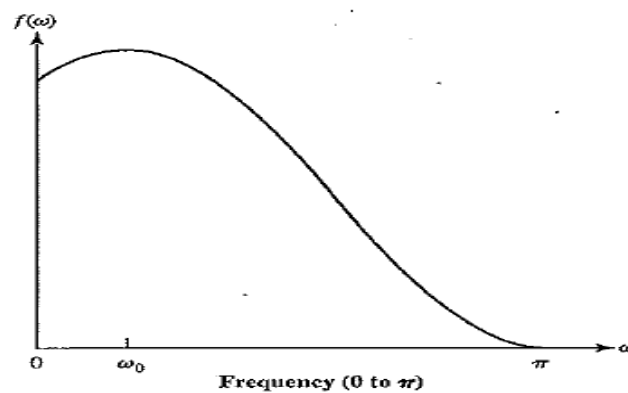
$$\gamma_k = \int_{-\pi}^{\pi} f(\omega) e^{i\omega k} d\omega.$$

¹ Ver, por ejemplo, Wei (2006).

Este resultado muestra que ambos análisis son equivalentes.

La función $f(\omega)$ tiene propiedades importantes.

- i) $f(\omega)$ es una función de valor real no negativa.
- ii) $f(\omega) = f(\omega + 2\pi)$. Por tanto $f(\omega)$ es periódica con período 2π .
- iii) $f(\omega) = f(-\omega)$, es decir $f(\omega)$ es una función par. Por tanto es simétrica y su gráfica solamente se presenta para $0 \leq \omega \leq \pi$. La siguiente gráfica presenta un ejemplo de una densidad espectral.

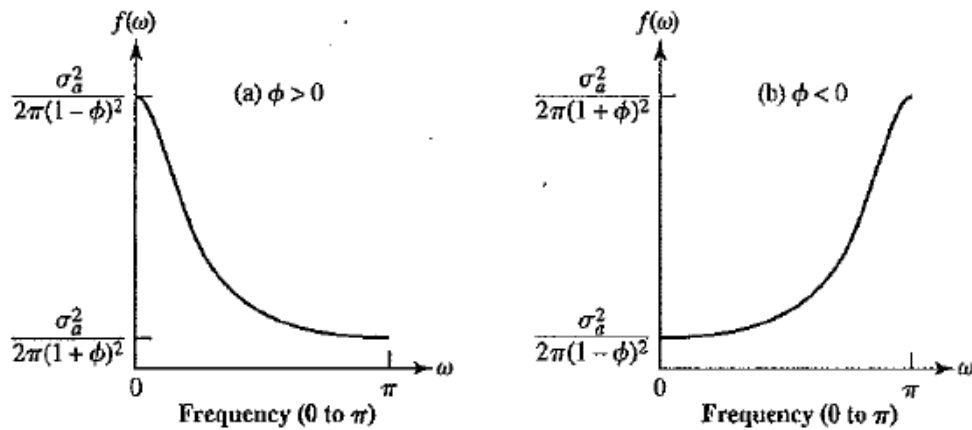


- iv) De la inversa de la transformada de Fourier, para $k=0$ se obtiene,

$$\text{Var}(Z_t) = \gamma_0 = \int_{-\pi}^{\pi} f(\omega) d\omega,$$

lo cual muestra que el espectro $f(\omega)$ puede ser interpretado como la descomposición de la varianza del proceso. El término $f(\omega)d\omega$ puede ser interpretado como la contribución a la varianza atribuible a la componente del proceso con frecuencias en el intervalo $(\omega, \omega + d\omega)$. Un pico en el espectro indica una importante contribución a la varianza de las componentes en las frecuencias en el intervalo correspondiente. La gráfica anterior muestra que las componentes cerca a la frecuencia ω_0 (bajas frecuencias) tiene la mayor importancia y las componentes de alta frecuencia cerca de π son de poca importancia. Esto significa que la serie tiene un largo período. Si el pico correspondiera a frecuencias

cerca a π , la serie tendría un corto período. Por ejemplo considere el proceso AR(1), $(1 - \phi B)Z_t = a_t$, Las siguientes gráficas muestran el comportamiento del espectro.



Si $\phi > 0$, entonces la serie está positivamente correlacionada, el espectro está dominado por componentes de bajas frecuencias, la serie tiene un período largo y la serie es suave. Si $\phi < 0$, entonces la serie está negativamente correlacionada, el espectro está dominado por componentes de altas frecuencias, la serie tiene un período corto y la serie es poco suave.

- ✓ Volviendo al proceso ARFIMA(0,d,0), según la ecuación (4) la serie Z_t puede interpretarse como el proceso resultante de un sistema lineal donde el proceso de entrada a_t , es un ruido blanco. Para este proceso de ruido blanco a_t , se puede mostrar que su función de densidad espectral es $f_a(\omega) = \frac{\sigma_a^2}{2\pi}$. Por tanto, la densidad espectral de Z_t puede obtenerse a partir de la de a_t de la siguiente forma (Ver Priestley, 1981 o Wei, 2006):

$$f(\omega) = f_a(\omega) |1 - e^{i\omega}|^{-2d} = \frac{\sigma_a^2}{2\pi} [2\sin(\omega/2)]^{-2d} \quad (5)$$

y cuando $\omega \rightarrow 0^+$ se tiene la siguiente aproximación:

$$f(\omega) = \frac{\sigma_a^2}{2\pi} \omega^{-2d}$$

De esta expresión se deduce que:

⇒ para valores $0 < d < 1/2$, la densidad espectral es una función decreciente de ω no acotada en el origen, y por tanto el proceso tiene memoria larga.

⇒ 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.

- ✓ Sobre el comportamiento de las autocorrelaciones, Hosking (1981) deriva las siguientes expresiones para la función de autocovarianza y la varianza del proceso ARFIMA(0,d,0), respectivamente:

$$\gamma_k = \sigma_a^2 \frac{\Gamma(1-2d)\Gamma(k+d)}{\Gamma(d)\Gamma(1-d)\Gamma(k+1-d)}, \text{ si } k \geq 1$$

$$\gamma_0 = \sigma_a^2 \frac{\Gamma(1-2d)}{\{\Gamma(1-d)\}^2}$$

por tanto, la función de autocorrelación es

$$\rho_k = \frac{\Gamma(1-d)\Gamma(k+d)}{\Gamma(d)\Gamma(k+1-d)}, \text{ si } k \geq 1$$

y aplicando nuevamente $\Gamma(a+x)/\Gamma(b+x) \approx x^{a-b}$, resulta que cuando $k \rightarrow \infty$,

$$\rho_k \approx \frac{\Gamma(1-d)}{\Gamma(d)} k^{2d-1}, \text{ si } k \geq 1$$

lo que indica que cuando $0 < d < 1/2$ la función de autocorrelación del proceso ARFIMA(0,d,0) decrece a un ritmo hiperbólico lento, a diferencia de los procesos ARMA, cuya función de autocorrelación decrece exponencialmente.

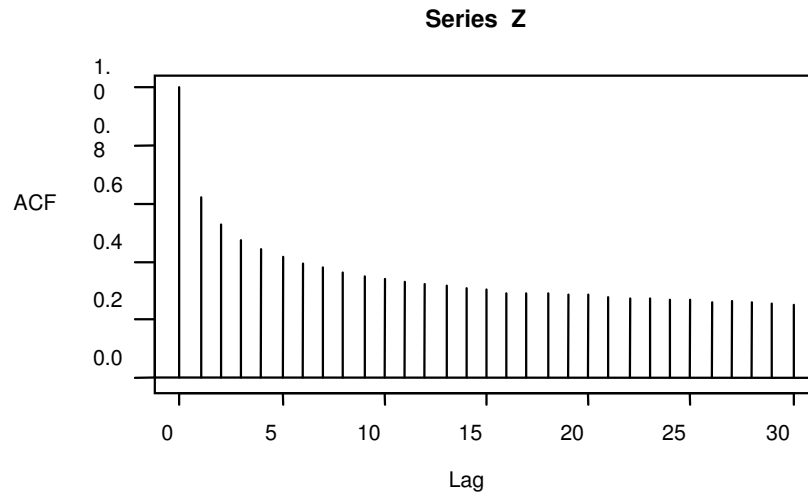
- Para visualizar esta dependencia a largo plazo típica de los procesos ARFIMA, el cuadro 1 compara las autocorrelaciones de un proceso ARFIMA(0,1/3,0) con las de un proceso AR(1) de la forma $Z_t = \phi Z_{t-1} + a_t$, con $\phi = 0.5$. Aunque en ambos casos la autocorrelación de orden uno es igual a 0.5, se aprecia claramente una gran diferencia en el ritmo al que las correlaciones convergen hacia cero: mientras que en el proceso AR(1) la autocorrelación de orden 25 es 2.98×10^{-8} , en el proceso fraccionalmente integrado es 0.173, y sólo se reduce a 0.109 en el retardo 100.

Retardo (k)	$\rho(k)$ [ARFIMA(0,1/3,0)]	$\rho(k)$ [AR(1), con $\phi=0.5$]
1	0.500	0.500
2	0.400	0.250
3	0.350	0.125
4	0.318	0.063
5	0.295	0.031
10	0.235	0.001
25	0.173	2.98×10^{-8}
50	0.137	8.88×10^{-16}
100	0.109	7.89×10^{-31}

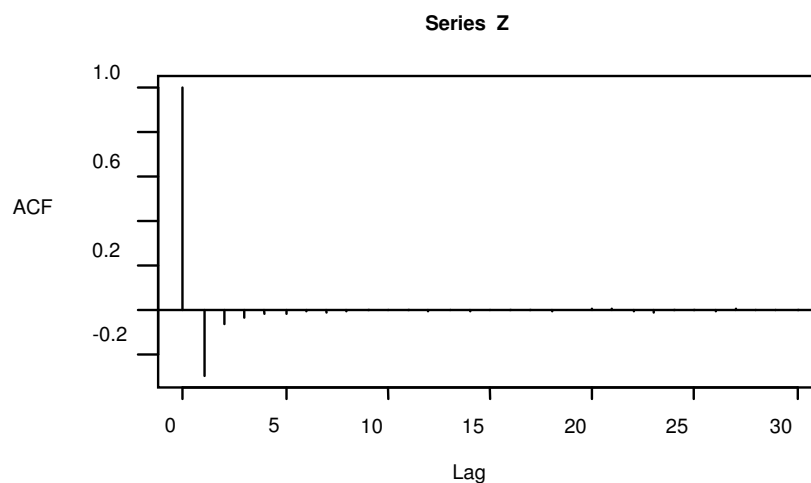
Resumiendo, las propiedades más importantes del proceso ARFIMA(0,d,0) son las siguientes.

- ✓ En primer lugar, si $0 < d < 1/2$, las autocorrelaciones son todas positivas y decaen lentamente, a un ritmo hiperbólico k^{2d-1} , y no son sumables. Estas propiedades son precisamente las que caracterizan a los modelos con memoria larga, y por tanto podemos concluir que en el intervalo $0 < d < 1/2$, el proceso ARFIMA(0,d,0) es un proceso estacionario con memoria larga. La

siguiente gráfica presenta la ACF de una ARFIMA(0,d,0) con $d=0.4$



- ✓ Por el contrario, cuando $-1/2 < d < 0$, Z_t es un proceso estacionario con memoria corta ya que las autocorrelaciones son todas negativas y absolutamente sumables. La siguiente gráfica presenta la ACF de una ARFIMA(0,d,0) con $d=-0.4$.



Aunque este rango de valores no tiene mucho interés en sí mismo, puede resultar muy útil en la modelación de las primeras diferencias de modelos ARFIMA(0,d,0) no estacionarios con orden

de integración $1/2 < d < 1$. Como veremos más adelante, los procesos fraccionalmente integrados con $1/2 < d < 1$ permiten modelar el comportamiento de series no estacionarias pero que finalmente revierten hacia la media, propiedad que no podrían modelar los procesos con raíces unitarias.

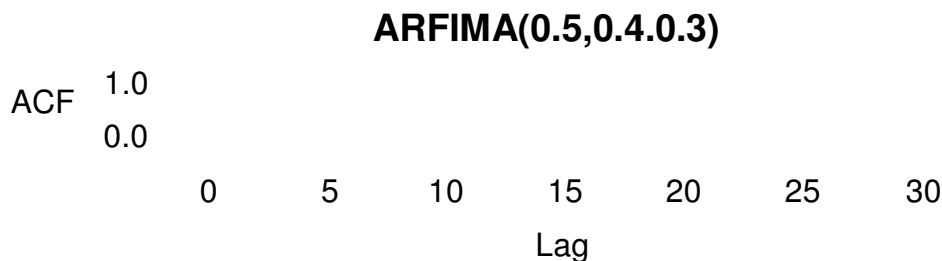
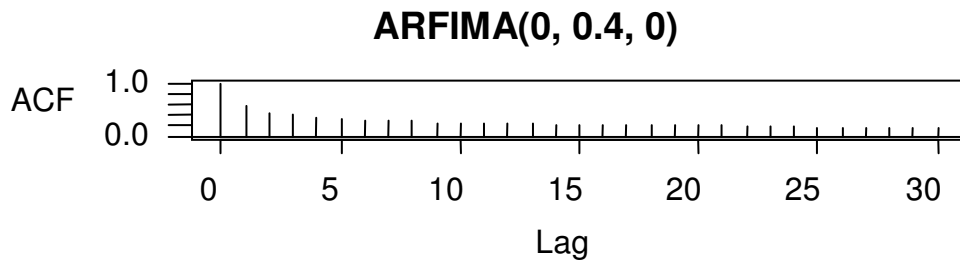
- El modelo ARFIMA(0,d,0) se puede generalizar de forma inmediata haciendo que en la ecuación $(1-B)^d Z_t = u_t$ el proceso u_t sea un proceso autorregresivo y de medias móviles ARMA(p,q) en lugar de un ruido blanco. En este caso diremos que Z_t es un proceso ARMA fraccionalmente integrado o ARFIMA(p,d,q). Este proceso admite una representación de la forma:

$$\phi(B)(1-B)^d Z_t = \theta(B)a_t \quad (6)$$

donde $\phi(B)$ y $\theta(B)$ son los polinomios de retardos de orden p y q, respectivamente, del proceso ARMA, cuyas raíces están fuera del círculo unidad, no hay raíces comunes y el parámetro d y la perturbación a_t son como en el caso anterior.

- 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 en el corto plazo, a través de los parámetros de la parte ARMA del modelo.
- El comportamiento a largo plazo del proceso ARFIMA(p,d,q) es muy similar al del ARFIMA(0,d,0), ya que para observaciones muy distantes, los efectos de los parámetros ARMA son prácticamente despreciables. En concreto, Hosking (1981) demuestra que si $d < 1/2$ y todas las raíces

de $\phi(B)=0$ están fuera del círculo unidad, Z_t es estacionario, y si $d>-1/2$ y todas las raíces de $\theta(B)=0$ están fuera del círculo unidad, Z_t es invertible.



- En el caso estacionario ($d < 1/2$), Hosking (1981) demuestra que si $d < 1/2$, existe una constante positiva C_ρ tal que $\rho_k \approx C_\rho k^{2d-1}$, y por tanto, si $0 < d < 1/2$, las autocorrelaciones decrecen hiperbólicamente y no son sumables. La expresión exacta de las autocorrelaciones que se deriva en Hosking (1981) es bastante complicada (Sowell, 1992a y Cheng, 1994, simplifican su cálculo).
- Una de las propiedades que distingue a los modelos ARFIMA de los modelos ARMA de memoria corta y de los modelos integrados ARIMA, es la forma en que el *output* responde a un cambio unitario en la perturbación del modelo:
 - ✓ En un modelo ARIMA, el efecto de un shock sobre el nivel futuro de la serie es persistente y no hay reversión a la media.

✓ En un modelo ARFIMA, el efecto acaba desapareciendo, aunque a un ritmo más lento que el exponencial de los procesos ARMA.

- De acuerdo con (3) cuando $d > -1/2$,

$$(1-B)^d = \sum_{j=0}^{\infty} \pi_j B^j \quad (7)$$

donde

$$\pi_j = \frac{\Gamma(j-d)}{\Gamma(-d)\Gamma(j+1)}$$

En la muestra la estructura infinita representada por (7) es aproximada reemplazando $(1-B)^d$ por

$$(1-B)_t^d = \sum_{j=0}^{t-1} \pi_j B^j$$

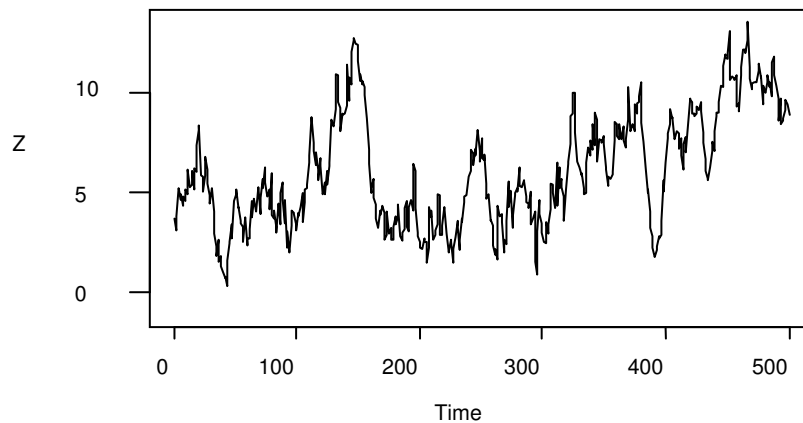
Es decir, los infinitos rezagos son truncados al inicio de la muestra disponible. Esto implica que la diferencia fraccional d , es calculada como,

$$(1-B)^d Z_t = \sum_{j=0}^{t-1} \pi_j B^j Z_t$$

- Simulación de modelos ARFIMA usando R.

Librería fracdiff: Simulación y gráfica de $n=500$ datos de un proceso ARFIMA(1,d,1) con $\phi_1=0.8$ $\theta_1=-0.3$, constante=2 y $d=0.25$, con errores $N(0,1)$.

```
z<-fracdiff.sim(500, ar=0.8, ma=-0.3, mu=2, d=.25)
plot(as.ts(z$series))
```



Identificación de modelos ARFIMA

- Es más complicada que la identificación en modelos ARIMA debido a la característica de decrecimiento hiperbólico de su ACF.
- Sin embargo, si existe un número suficientemente grande de observaciones y el proceso es invertible, un procedimiento que podría ser útil en la identificación del modelo consta de los siguientes pasos (Castaño, 2006) y Castaño, Gómez y Gallón(2008):

Paso1: Estime un modelo $ARFIMA(p^*, d, 0)$ donde el orden autorregresivo p^* es lo suficientemente alto para aproximar adecuadamente la componente $ARIMA(p, d)$ del proceso.

Paso 2: Sea \hat{d} la estimación del orden de diferenciación fraccional del modelo anterior. Obtenga $(1-B)^{\hat{d}} Z_t$, es decir la diferencia fraccional de orden \hat{d} de Z_t .

Paso 3: Use los procedimientos tradicionales de identificación de un modelo ARMA sobre la serie $(1-B)^{\hat{d}} Z_t$, tales como la ACF, PACF y EACF muestrales.

Paso 4. Una vez identificado el modelo ARMA(p,q) de $(1-B)^d Z_t$, estime el modelo completo ARFIMA(p,d,q).

Observaciones:

1. Más adelante veremos algunos métodos no paramétricos y semiparamétricos que permiten la estimación de d sin tener que especificar el modelo de corto plazo ARMA del modelo ARFIMA. En este caso el proceso de identificación se realiza como en el procedimiento anterior, excepto que no se realiza el paso 1.

2. Algunos autores como Schmidt y Tscherning (1994) proponen la identificación usando criterios de información. Encuentran que el criterio de Schwarz es el que mejor se comporta.

Ejemplo: El siguiente programa de R, contiene los pasos de la identificación de un modelo ARFIMA(1, 0.25, 0) simulado con n=600 observaciones.

```
=====
library(fracdiff)

# Simulación del modelo ARFIMA
z <- fracdiff.sim(600, ar=0.7, d=.25)
par(mfrow = c(2, 1))
plot(as.ts(z$series))
acf(z$series, lag.max=50)

# Verificación teórica del procedimiento
df_z <- diffseries(z$series, 0.25)
par(mfrow = c(3, 1))
plot(df_z, type="l")
acf(df_z, ci=0)
pacf(df_z, ci=0)

# Procedimiento de Identificación

# paso 1
modelo1 <- fracdiff(z$series, nar = 8, nma = 0, dtol = NULL, drange = c(0, 0.5), M = 100)
summary(modelo1)
d_est=coef(modelo1)[1]
resid=residuals(modelo1)

# paso 2: del paso 1 se obtuvo un d estimado almacenado en d_est
df_z <- diffseries(z$series, d_est)

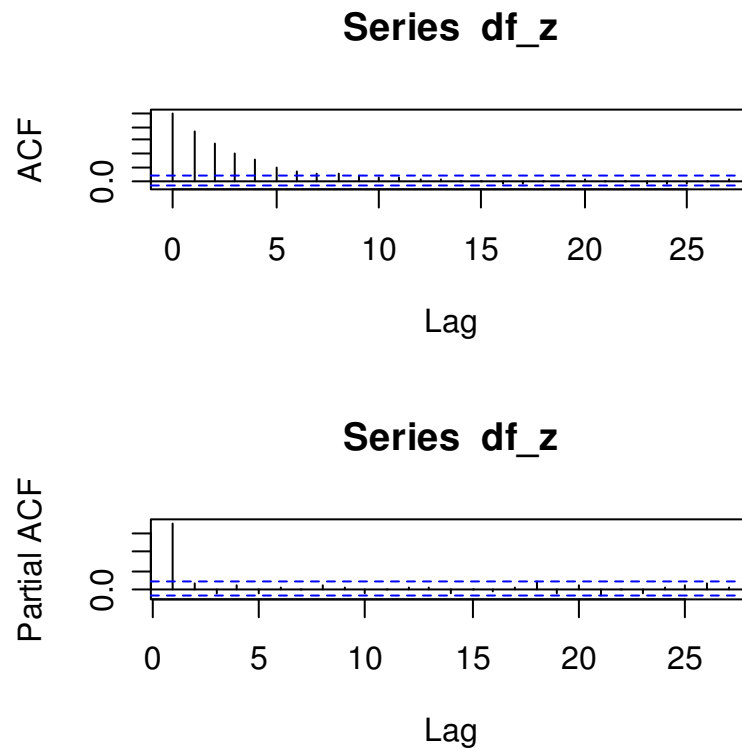
# paso 3
par(mfrow = c(2, 1))
acf(df_z, ci=0)
pacf(df_z, ci=0)
```

```
# los resultados son consistentes con un AR(1).

# paso 4 # Estimación del modelo identificado como ARFIMA(1,d,0)
modelo2 <- fracdiff(z$series, nar = 1, nma = 0, dtol = NULL, drange = c(0, 0.5), M = 100)
summary(modelo2)
=====
```

RESULTADOS:

1. Verificación teórica del procedimiento: Una vez diferenciada la serie por el verdadero orden $d=0.25$, se obtienen las siguientes ACF y PACF muestrales de $(1-B)^d Z_t$.



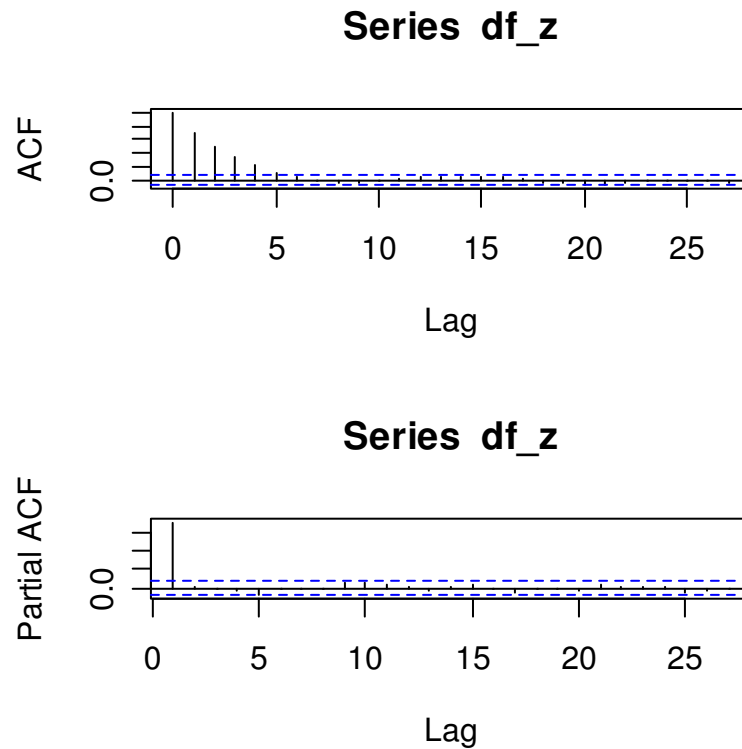
Los resultados indican que el verdadero modelo ARMA para $(1-B)^d Z_t$ corresponde a un AR(1). Por tanto el procedimiento propuesto funciona teóricamente.

2. Procedimiento de indentificación.

Paso 2: Del modelo preliminar estimado ARFIMA(8,d,0) se obtiene que

$\hat{d}=0.22$. A continuación obtenemos $(1-B)^{0.22}Z_t$.

Paso 3: identificamos el modelo ARMA contenido en $(1-B)^{0.22}Z_t$, por ejemplo, usando la AFC y la PACF muestrales. Se obtiene,



De los correlogramas anteriores se identifica un modelo AR(1) para $(1-B)^{0.22}Z_t$. Por tanto el modelo identificado es un ARFIMA(1,d,0).

Paso 4: estimación del modelo completo, ARFIMA(1,d,0).

```
$log.likelihood
[1] -842.2873
```

```
$d
[1] 0.2441921
```

```
$ar
[1] 0.7700853
```

```
$ma
[1] 0
```

```
$covariance.dpq
              d              ar1
d      6.063582e-05 -4.412309e-05
ar1 -4.412309e-05  7.121332e-04
```

```
$stderror.dpq
[1] 0.00778690 0.02668582
```

```
$correlation.dpq
              [,1]      [,2]
[1,] 1.0000000 -0.2123346
[2,] -0.2123346 1.0000000
```

Estimación de los Modelos ARFIMA

- Los métodos de estimación de modelos con memoria larga propuestos hasta el momento en la literatura pueden clasificarse en dos grandes grupos:
 - ✓ Métodos paramétricos, en los que se conoce de forma explícita toda la estructura de correlación, tanto en el corto como en el largo plazo, o se conoce la densidad espectral en todas las frecuencias.
 - ✓ Métodos semiparamétricos, en los que no se especifica un modelo completo de la estructura de autocovarianzas (o del espectro) si no sólo su comportamiento en los retardos alejados (o en las bajas frecuencias).
- Otros autores distinguen entre:
 - ✓ Métodos de estimación en dos etapas, en los que se estima en primer lugar el parámetro d y en una segunda etapa los parámetros ARMA del modelo, y

- ✓ Métodos de estimación en una etapa, que estiman conjuntamente, por máxima verosimilitud, todos los parámetros del modelo.
- El enfoque paramétrico tiene la ventaja de que permite caracterizar todas las autocorrelaciones y no sólo su decrecimiento hiperbólico. Además, los métodos propuestos se basan en maximizar la función de verosimilitud Gaussiana, y los estimadores máximo verosímiles (MV) resultantes tienen buenas propiedades: son consistentes y asintóticamente normales, y bajo hipótesis de normalidad, asintóticamente eficientes.
- Sin embargo, como señala Robinson (1995b), si el modelo paramétrico que se supone para $r(k)$ o $f(\omega)$ está mal especificado, los estimadores serán inconsistentes, y los métodos semiparamétricos pueden hacer un “mejor papel”, ya que son más robustos a una mala especificación del modelo. Además, los estimadores semiparamétricos también van a tener buenas propiedades asintóticas, aunque su tasa de convergencia será inferior a la \sqrt{T} típica de los estimadores paramétricos Gaussianos.
- El cuadro 2 resume los principales métodos de estimación que revisaremos en esta sección.
- Además de estos métodos existen otros que podríamos decir son minoritarios en aplicaciones empíricas.
- ✓ Li y McLeod (1986), por ejemplo, analizan las propiedades de los estimadores máximo verosímiles del modelo aproximado que se obtendría truncando el polinomio $(1-B)^d$ en la correspondiente representación $AR(\infty)$ del proceso ARFIMA dada en (3).

- ✓ Chung y Baillie (1993) analizan las propiedades en muestras finitas de los estimadores que resultan de minimizar una suma de cuadrados condicional en el dominio del tiempo.

Una breve descripción de ambos métodos puede encontrarse en Beran (1994) y Baillie (1996), respectivamente.

Cuadro 2. Métodos de estimación para modelos con memoria larga

Métodos Paramétricos	<ul style="list-style-type: none"> • Máxima verosimilitud exacta en el dominio temporal • Aproximación de Whittle a la máxima verosimilitud en el dominio de las frecuencias
Métodos semiparamétricos	<ul style="list-style-type: none"> • Regresión mínimo-cuadrática en el dominio de las frecuencias para el logaritmo del periodograma • Estimador gaussiano semiparamétrico en el dominio de las frecuencias

- Los métodos de estimación paramétricos que revisaremos a continuación plantean la estimación conjunta de los parámetros del modelo maximizando la función de verosimilitud Gaussiana o alguna aproximación suya.
- ✓ Yajima (1985) obtuvo los primeros resultados para el más sencillo de los modelos fraccionalmente integrados, el ARFIMA(0,d,0) con perturbación a_t Gaussiana. Este autor demostró que el estimador del vector de parámetros $\theta = (\sigma_a^2, d)$ que se obtiene maximizando la función de verosimilitud exacta,

$$L(\theta; Z) = -\frac{T}{2} \ln(2\pi) - \frac{1}{2} \ln |\Sigma(\theta)| - \frac{1}{2} Z' \Sigma(\theta)^{-1} Z$$

donde T es el tamaño muestral, Z es el vector de observaciones y $\Sigma(\theta)$ es la matriz de varianzas covarianzas de Z , es consistente y asintóticamente normal.

- ✓ Dahlhaus (1989) extendió estos resultados a una clase más general de modelos Gaussianos estacionarios, en la que se incluye:

⇒ el modelo ARFIMA(p,d,q) con media desconocida μ y densidad espectral $f(\omega) = f(\omega; \theta)$.

⇒ Bajo ciertas condiciones, Dahlhaus (1989) demostró que si $0 < d < 1/2$ (proceso estacionario), el estimador máximo verosímil de θ , $\hat{\theta}_T$, es consistente y asintóticamente normal, y además es asintóticamente eficiente, en el sentido de que $\sqrt{T}(\hat{\theta}_T - \theta) \xrightarrow{d} N(0, W^{-1}(\theta))$, donde \xrightarrow{d} denota convergencia en distribución y $W(\theta)$ es el límite de la matriz de información de Fisher.

⇒ Además, estos resultados siguen siendo válidos si en la función de verosimilitud, $L(\theta, \mu; Z)$, se sustituye μ por un estimador consistente, por ejemplo la media muestral, y se maximiza con respecto a θ la función de verosimilitud resultante.

- ✓ Sowell (1992a) proporciona un algoritmo para calcular la matriz de varianzas covarianzas $\Sigma(\theta)$ y obtener posteriormente la función de verosimilitud $L(\theta, \mu; Z)$.

Aunque el método de estimación MV exacto es muy atractivo por las buenas propiedades de los estimadores que proporciona, el método es muy costoso computacionalmente.

- ⇒ Para maximizar la función de verosimilitud hay que emplear métodos numéricos que requieren calcular, en cada iteración, la inversa de la matriz de varianzas y covarianzas. Esto consume mucho tiempo, especialmente si la serie es larga o hay muchos parámetros.
- ⇒ Además, si d está próximo a $1/2$, la evaluación de la inversa de $\Sigma(\theta)$ puede ser muy inestable, ya que las covarianzas apenas cambian y la matriz llega a ser casi singular.

Para evitar estos problemas se han propuesto métodos alternativos que consisten en maximizar una aproximación a la función de verosimilitud Gaussiana $L(\theta, \mu; Z)$, antes definida.

- ✓ Fox y Taqqu (1986) consideran los estimadores MV aproximados de Whittle (1962), que se obtienen minimizando la función:

$$L_W(\theta; Z) = \frac{1}{4\pi} \int_{-\infty}^{\infty} \left[\ln f(\omega; \theta) + \frac{I_T(\omega)}{f(\omega; \theta)} \right] d\omega$$

donde el subíndice W hace referencia a Whittle, $f(\omega; \theta)$ es la función de densidad espectral de proceso ARFIMA(p,d,q) estacionario dada por

$$f(\omega; \theta) = \frac{\sigma_a^2}{2\pi} \left| \frac{\ddot{\theta}(e^{i\omega})}{\phi(e^{i\omega})} \right|^2 [2\sin(\omega/2)]^{-2d}$$

donde $\ddot{\theta}(e^{i\omega})$ y $\phi(e^{i\omega})$ son polinomios de orden q y p en potencias de $e^{i\omega}$ con los coeficientes de la componentes de corto plazo del

ARFIMA, e $I_T(\omega)$ es el periodograma² de z_1, z_2, \dots, z_T , definido como

$$I_T(\omega) = \frac{1}{2\pi T} \left| \sum_{t=1}^T (z_t - \bar{z}) e^{it\omega} \right|^2$$

En el caso de que Z_t sea un proceso estacionario gaussiano cuya densidad espectral admita una parametrización de la forma $f(\omega; \theta) = \theta_1 g(\omega; \theta^*)$, con $\theta = (\theta_1, \theta^*)$ y $\int_{-\pi}^{\pi} \ln g(\omega; \theta^*) d\omega = 0$, Fox y Taqu (1986) prueban la consistencia de los estimadores MV de Whittle de θ_1 y θ^* , y demuestran además que el estimador de θ^* , $\hat{\theta}_w^*$, es asintóticamente normal, pero no obtienen la distribución asintótica del estimador de θ_1 .

- ✓ Dahlhaus (1989) extiende este resultado y demuestra que para un proceso estacionario gaussiano con densidad espectral arbitraria $f(\omega; \theta)$, el estimador de Whittle del vector paramétrico completo, $\hat{\theta}_w$, es consistente y asintóticamente normal, y establece además condiciones para que dicho estimador sea asintóticamente eficiente. El proceso ARFIMA(p,d,q) gaussiano verifica las condiciones anteriores y por tanto el estimador MV de Whittle de d es consistente, asintóticamente normal y asintóticamente eficiente.
- ✓ Para una parametrización de la densidad espectral similar a la de Fox y Taqu(1986), Giritatis y Surgalis (1990) derivan un teorema central del límite para $\hat{\theta}_w^*$ en un caso más general en el que Z_t es un proceso lineal con memoria larga, no necesariamente gaussiano.

² Ver, por ejemplo, Wei (2006).

- Los estimadores aproximados de Whittle son atractivos porque reemplazan la evaluación de la inversa de la matriz de varianzas y covarianzas por transformadas de Fourier del vector de observaciones, y los estimadores tienen, bajo ciertas condiciones, la misma distribución asintótica que los estimadores MV exactos. Pero el método tampoco es conveniente desde un punto de vista computacional, ya que la integral en $L_w(\theta; Z)$ debe evaluarse por métodos numéricos y esto debe repetirse en cada iteración. La solución adoptada generalmente en estos casos es sustituir esta integral por una suma finita sobre las frecuencias de Fourier, de la forma:

$$\tilde{L}_w(\theta; Z) = \frac{1}{2T} \sum_{k=1}^{T-1} \left[\ln f(\omega_k; \theta) + \frac{I_T(\omega_k)}{f(\omega_k; \theta)} \right]$$

donde ω_k es la k-ésima frecuencia de Fourier, $\omega_k = \frac{2\pi k}{T}$, y donde T es

el número de observaciones. Minimizando $\tilde{L}_w(\theta; Z)$ respecto a θ se obtiene una versión discreta del estimador MV de Whittle en el dominio de las frecuencias, que denotaremos $\tilde{\theta}_w$. Desde un punto de vista computacional, $\tilde{\theta}_w$ es el estimador más conveniente, ya que los cálculos se pueden agilizar utilizando la transformada rápida de Fourier (ver Bloomfield, 1976) y la simetría de la función de densidad espectral y del periodograma. Además, según Dahlhaus (1989), este estimador tiene las mismas propiedades asintóticas que su análogo anterior $\hat{\theta}_w$, y por tanto es, bajo ciertas condiciones, consistente, asintóticamente normal y asintóticamente eficiente.

- El método que acabamos de describir se simplifica considerablemente si la densidad espectral admite una parametrización del tipo $f(\omega; \theta) = \theta_1 g(\omega; \theta^*)$, con $\theta = (\theta_1, \theta^*)$, en cuyo caso el estimador de θ^* se obtiene minimizando

$$Q(\theta^*; Z) = \sum_{k=1}^{T-1} \frac{I_T(\omega_k)}{g(\omega_k; \theta^*)}$$

y el estimador de θ_1 se calcula como $\hat{\theta}_1 = \frac{1}{T} \sum_{k=1}^{T-1} \frac{I_T(\omega_k)}{f(\omega_k; \hat{\theta}^*)}$.

- Según la ecuación (5), la densidad espectral del ARFIMA(0,d,0) es

$$f(\omega) = \frac{\sigma_a^2}{2\pi} [2\text{sen}(\omega/2)]^{-2d}$$

y por tanto su densidad espectral admite una parametrización de la forma $f(\omega; \theta) = \theta_1 g(\omega; d)$, con $\theta = (\theta_1, d)$ con $\theta_1 = \frac{\sigma_a^2}{2\pi}$, y por tanto el estimador de la varianza de la perturbación de este modelo será $\sigma_a^2 = 2\pi\hat{\theta}_1$. Cheung y Diebold (1994) comparan las propiedades en muestras finitas de la versión discreta del estimador aproximado de Whittle, $\tilde{\theta}_w$, y del estimador máximo verosímil exacto en el dominio temporal, $\hat{\theta}_T$, para procesos ARFIMA(0,d,0), y observan que si la media es conocida, $\hat{\theta}_T$ es por supuesto más eficiente, pero cuando la media debe estimarse, el estimador $\tilde{\theta}_w$ es preferible, especialmente en muestras grandes donde calcular el estimador MV exacto resulta impracticable. Sin embargo, Smith, Sowell y Zin (1997) proponen una corrección por el sesgo del estimador $\hat{\theta}_T$ que parece tener buenas propiedades en muestras finitas frente al estimador $\tilde{\theta}_w$.

- Los estimadores paramétricos Gaussianos que acabamos de revisar tienen propiedades muy deseables (consistencia, distribución asintótica normal y eficiencia), pero, como señala Robinson (1995b), en caso de que el modelo paramétrico establecido para $f(\omega)$ esté mal

especificado, los estimadores serán sesgados e inconsistentes, y de ahí el interés por estimadores semiparamétricos.

- Una primera alternativa a los métodos de estimación paramétricos es el método propuesto por Geweke y Porter-Hudak (1983), GPH de aquí en adelante, basado en un modelo de regresión para el logaritmo del periodograma, o la versión mejorada del mismo, propuesta por Robinson (1995a), que proporciona un estimador más eficiente y establece con rigor la consistencia y su distribución asintótica. El estimador GPH explota la expresión tan sencilla que tiene la densidad espectral de los procesos fraccionalmente integrados en las bajas frecuencias. Según la expresión de la densidad espectral de los modelos ARFIMA(p,d,q) vista anteriormente, la densidad espectral del proceso Z_t en un modelo ARFIMA(p,d,q) puede escribirse como:

$$f(\omega) = f_u(\omega)[4\text{sen}^2(\omega/2)]^{-d}$$

donde $f_u(\omega)$ es la densidad espectral del proceso $u_t \sim \text{ARMA}(p,q)$,

$$f_u(\omega) = \frac{\sigma_a^2}{2\pi} \left| \frac{\ddot{\theta}(e^{i\omega})}{\phi(e^{i\omega})} \right|^2$$

Evaluando la expresión para $f(\omega)$ en las frecuencias de Fourier y tomando logaritmos, se obtiene una expresión aproximada de la forma:

$$\ln[I_T(\omega_j)] \cong \beta_0 - d\{\ln[4\text{sen}^2(\omega_j/2)]\} + \varepsilon_j$$

donde $j=1,2, \dots, m$ y $\omega_j = \frac{2\pi j}{T}$ es la j -ésima frecuencia de Fourier,

$I_T(\omega_j)$ es la j -ésima ordenada del periodograma, β_0 es una constante y

$m=g(T)$ es un entero menor que T y tendiendo hacia infinito más despacio que T .

La expresión puede interpretarse como un modelo de regresión en el dominio espectral, lo que sugiere estimar d por mínimos cuadrados ordinarios (MCO). GPH (1983) argumenta que las perturbaciones ε_j son asintóticamente independientes e igualmente distribuidas con media cero y varianza $\frac{\pi^2}{6}$.

- Las pruebas rigurosas de las propiedades del estimador GPH no se establecen sino hasta el trabajo de Robinson (1995a), que demuestra la consistencia y la normalidad asintótica para todo el rango de valores $-1/2 < d < 1/2$, y el mismo autor propone una versión modificada y más eficiente del mismo. El estimador propuesto por Robinson (1995a) es mucho más general que GPH en varios sentidos:
- ✓ En primer lugar, sólo requiere que la densidad espectral en las bajas frecuencias verifique la condición:

$$f(\omega) \sim C_f |\omega|^{-2d}, \text{ cuando } \omega \rightarrow 0^+$$

sin necesidad de especificar la forma concreta de $f(\omega)$. En este caso el modelo de regresión que se estima es:

$$\ln[I_T(\omega_j)] \cong \ln(C_f) - 2d(\ln|\omega_j|) + \ln[I_T(\omega_j)/f(\omega_j)]$$

con $j=1,2,\dots,m$ y donde ω_j , $I_T(\omega_j)$ y m son como en el caso anterior. Bajo hipótesis de normalidad para Z_t , Robinson (1995a) demuestra que el estimador que se obtiene estimando por mínimos

cuadrados ordinarios la ecuación anterior, utilizando sólo las m primeras frecuencias en torno al origen, es consistente y asintóticamente normal, aunque con una tasa de convergencia \sqrt{m} inferior a la \sqrt{T} típica de los estimadores paramétricos Gaussianos.

- ✓ En segundo lugar, todos los resultados de Robinson (1995a) se extienden para modelos multivariados, y su trabajo se convierte así en el primero en presentar resultados asintóticos para series multivariadas con memoria larga. Este método de estimación es muy sencillo de aplicar (cualquier computador hace una estimación MCO) y requiere pocos supuestos sobre el modelo. De hecho, el estimador GPH ha sido durante años el más utilizado en aplicaciones empíricas, al menos como punto de comparación para analizar las propiedades de otros estimadores. Sin embargo la pérdida de eficiencia puede llegar a ser extrema en muestras pequeñas.
- Agiakloglou, Newbold y Wohar (1993) analizan las propiedades de este estimador en muestras finitas y observan que cuando u_t es un AR(1) o un MA(1) con parámetro cercano a la unidad, **el estimador tiene un sesgo enorme y es muy ineficiente**. Otro problema adicional de este método es la elección del número m de frecuencias que se utilizan, y según Robinson (1995b), el mayor inconveniente es el carácter restrictivo que supone la hipótesis de normalidad sobre la serie Z_t . En dos trabajos recientes, Velasco (1998a) y Velasco (1998b) extiende los resultados de Robinson (1995a) a modelos no necesariamente Gaussianos y no estacionarios ($d \geq 1/2$), respectivamente.
- Como alternativa al estimador MCO en el dominio espectral, Robinson (1995b) propone un estimador gaussiano semiparamétrico para series cuyo espectro verifique la condición anterior de que

$$f(\omega) \sim C_f |\omega|^{-2d}, \text{ cuando } \omega \rightarrow 0^+$$

El estimador se obtiene maximizando una versión de la aproximación discreta de Whittle a la función de verosimilitud Gaussiana vista antes y dada por

$$\tilde{L}_w(\theta; Z) = \frac{1}{2T} \sum_{k=1}^{T-1} \left[\ln f(\omega_k; \theta) + \frac{I_T(\omega_k)}{f(\omega_k; \theta)} \right]$$

donde la suma sólo se realiza en un entorno de la frecuencia cero. Este estimador domina a los anteriores en varios aspectos: es más eficiente, no exige normalidad sobre Z_t y evita el recorte de las componentes en las bajas frecuencias así como la elección subjetiva del número m de ordenadas a promediar. Bajo ciertas condiciones, Robinson (1995b) prueba que en el rango $-1/2 < d < 1/2$, el estimador es consistente y asintóticamente normal, aunque nuevamente la tasa de convergencia es inferior a \sqrt{T} . Recientemente Velasco(1999) extiende estos resultados al caso no estacionario ($d \geq 1/2$).

Contrates de Hipótesis

- Existen diversos métodos para contrastar la existencia de memoria larga en una serie. El contraste propuesto por Geweke y Porter-Hudak (1983) se deriva de la estimación mínimo cuadrática del modelo de regresión en el dominio espectral definido en la ecuación

$$\ln[I_T(\lambda_j)] \cong \beta_0 - d \{ \ln[4 \text{sen}^2(\lambda_j/2)] \} + \varepsilon_j, \text{ con } j=1,2,\dots,m$$

La pendiente de este modelo es precisamente $-d$, y por tanto, para contrastar la hipótesis $H_0: d=0$ (memoria corta) frente a la alternativa

$H_1: d \neq 0$, se propone realizar el clásico contraste de significancia de la t de Student sobre el parámetro d .

- Aunque este contraste es muy sencillo de aplicar, Agiakloglou, Newbold y Wohar (1993) prueban que el estimador de d es muy sesgado cuando bajo H_0 existen componentes autorregresivas o de medias móviles con valores grandes de los parámetros, y por tanto, en estos casos, el contraste rechazará la hipótesis nula, aún siendo cierta, con mucha frecuencia. Esta misma conclusión se deriva de las simulaciones de Cheung (1993b), en las que se observa un sesgo del contraste hacia el no rechazo hacia alternativas del tipo $d > 0$ cuando el proceso generador de los datos es un $AR(1)$ con un valor grande del parámetro autorregresivo.
- Haubrich y Lo (1989) y Lo (1991) proponen un estadístico de contraste basado en una modificación del estadístico R/S (rango reescalado) introducido por Hurst (1951). El estadístico R/S es el rango de las sumas parciales de las desviaciones de una serie con respecto a su media normalizado por su desviación típica muestral. La modificación propuesta por Lo (1991) consiste en utilizar como factor de normalización otro estimador de la varianza que no sea la varianza muestral. En concreto, el estadístico de contraste que se propone es:

$$Q_T(q) = \frac{1}{S_T(q)} \left[\max_{1 \leq k \leq T} \sum_{i=1}^k (z_i - \bar{z}) - \min_{1 \leq k \leq T} \sum_{i=1}^k (z_i - \bar{z}) \right]$$

donde \bar{z} , es la media muestral de las T observaciones, y el factor de normalización, $S_T(q)$, es la raíz cuadrada de un estimador consistente de la varianza definido como:

$$S_T^2(q) = r_0 + 2 \sum_{j=1}^q w_j(q) r_j \text{ con } w_j(q) = 1 - \frac{j}{q+1} \text{ si } q < T$$

donde r_0 y r_j representan la varianza y la autocovarianza muestral de orden j de z_t , respectivamente,

$$r_0 = \frac{1}{T} \sum_{t=1}^T (z_t - \bar{z})^2 \quad y \quad r_j = \frac{1}{T} \sum_{t=1}^{T-j} (z_t - \bar{z})(z_{t+j} - \bar{z})$$

El estadístico $Q_T(q)$ se denomina “rango reescalado modificado”, y para $q=0$ se reduce al estadístico R/S de Hurst. Valores extremos de $Q_T(q)$ se consideran indicativos de la posible existencia de memoria larga en la serie.

Lo (1991) proporciona los valores críticos de su distribución para los niveles de significación habituales. Además presenta un estudio de Monte Carlo sobre el tamaño y la potencia del contraste en muestras finitas, del cual se deduce que el contraste tiende a ser conservador, al menos con los dos modelos de memoria corta analizados (IID y AR(1)), y tiene una potencia razonable, que aumenta con T y disminuye con q , frente a alternativas del tipo ARFIMA(0,d,0). Frente a esta clase de modelos, el contraste parece detectar mejor valores negativos de d que alternativas del tipo $d>0$. Las simulaciones de Cheung (1993b) sugieren que, al igual que GPH, este contraste es muy sensible a la existencia de correlación bajo la hipótesis nula: por ejemplo, un MA(1) con valores grandes del parámetro sesga el contraste hacia alternativas $d<0$.

- Robinson (1991) desarrolla un estadístico basado en el principio de los multiplicadores de Lagrange (LM) para contrastar la hipótesis de ruido blanco frente a una familia muy general de alternativas, que incluye a la clase de los procesos ARFIMA(0,d,0). En este caso, el estadístico admite una expresión muy sencilla de la forma

$$\Lambda_1 = \lambda_1^2, \text{ donde } \lambda_1 = \sqrt{\frac{6T}{\pi^2}} \sum_{j=1}^{T-1} \frac{r_j}{j},$$

donde, siendo r_j la autocorrelación muestral de orden j , $r_j = c_j/c_0$. Bajo la hipótesis de ruido blanco ($H_0: d=0$), λ_1 tiene una distribución asintótica $N(0,1)$, y por tanto, Λ_1 se distribuye asintóticamente como una χ^2 con un grado de libertad. Estos estadísticos permiten contrastar alternativas del tipo $d>0$, $d<0$ ó $d \neq 0$. Así, el contraste de un lado frente a la alternativa de memoria larga ($H_1: d>0$), consiste en rechazar H_0 si λ_1 es significativamente positivo respecto al correspondiente valor crítico de una $N(0,1)$, pero si lo que se quiere contrastar es la sobrediferenciación ($H_1: d<0$), lo apropiado es un contraste de un lado sobre la cola izquierda de la distribución $N(0,1)$. Finalmente, un contraste frente a la alternativa $H_1: d \neq 0$ rechazará la hipótesis de ruido blanco si Λ_1 toma valores significativamente grandes respecto a la distribución χ^2 con un grado de libertad.

- El contraste LM de Robinson (1991) presenta ciertas ventajas sobre los anteriores, ya que evita tener que estimar el parámetro d , y los estadísticos de contraste λ_1 y Λ_1 tienen una distribución asintótica conocida y sencilla. Además los estudios de Monte Carlo de Cheung (1993b) prueban que, en muestras finitas, este contraste detecta muy bien alternativas fraccionales, aunque es muy sensible a la presencia de correlación serial en los datos: un parámetro $AR(1)$ o $MA(1)$ en el proceso generador de los datos hace que el contraste rechace H_0 con muchísima más frecuencia que el nivel de significación nominal. Robinson (1994b) sugiere una forma de robustecer el estadístico Λ_1 contra la presencia de autocorrelación débil en la hipótesis nula, y Agiakloglou y Newbold (1994) presentan una generalización del mismo para contrastar $H_0: ARMA(p,q)$ frente a la alternativa $H_1: ARFIMA(p,d,q)$.

Para terminar mencionaremos brevemente algunos contrastes de raíces unitarias que pueden enmarcarse en el ámbito de los procesos fraccionalmente integrados, en cuyo caso la hipótesis de raíz unitaria se convierte en $H_0:d=1$.

- Sowell (1990) deriva la distribución asintótica del estadístico del contraste de Dickey-Fuller bajo el supuesto de que el proceso generador de los datos sea un proceso fraccionalmente integrado $I(d)$ no estacionario con $1/2 < d < 3/2$, y demuestra que el estadístico sólo tiene una distribución límite bien definida cuando $d=1$ (raíz unitaria). Diebold y Rudebush (1991) analizan las propiedades de este contraste en muestras finitas y comprueban que tiene muy poca potencia frente a alternativas fraccionales, y en consecuencia será muy difícil que nos permita discriminar entre un proceso integrado $I(1)$ y un proceso $I(d)$ con $0 < d < 1$. Dolado y Marmol (1997) extienden estos resultados al contraste secuencial de Dickey-Pantula, permitiendo así el análisis de procesos integrados con más de una raíz unitaria.
- Lee y Schmidt (1996) consideran el contraste de raíces unitarias KPSS de Kwiatkowski et al. (1992) como un contraste de memoria corta y prueban que es consistente frente a procesos con memoria larga estacionarios $I(d)$, con $-1/2 < d < 1/2$ y $d \neq 0$. Lee y Amsler (1997) extienden estos resultados a procesos fraccionalmente integrados no estacionarios ($1/2 < d < 1$) y encuentran que el contraste KPSS no puede distinguir de forma consistente entre estos procesos y un proceso con raíz unitaria.

Marmol (1998) revisa los resultados anteriores y prueba que los contrastes de un lado de Dickey-Fuller ($H_0:d=1$) y KPSS ($H_0:d=0$) son consistentes frente a alternativas fraccionales, para todo $d < 1$ y todo $d > 0$, respectivamente, pero la utilización combinada de ambos contrastes sólo con lleva a la decisión correcta, para todo $d > -1/2$, si éstos se realizan como contrastes de dos lados.

- Robinson (1994a) propone un contraste de los multiplicadores de Lagrange para la hipótesis de raíz unitaria, y otras formas de no estacionariedad, frente a alternativas muy generales que incluyen a los procesos fraccionalmente integrados. Este contraste, que extiende el contraste LM de ruido blanco frente a alternativas fraccionales de Robinson (1991), tiene propiedades estándar de eficiencia y tiene una distribución asintótica χ^2 bajo la hipótesis nula. Además, en muestras finitas, compara favorablemente con otros contrastes de raíces unitarias, y permite contrastar no sólo la hipótesis $d=1$ (raíz unitaria) sino otros valores del parámetro d . Robinson y Gil-Alana (1997) recogen una aplicación de este contraste a diversas series macroeconómicas.
- Recientemente, Lobato y Robinson (1998) han propuesto un contraste no paramétrico para la hipótesis $I(0)$ frente a alternativas de integración fraccionaria, aplicando el principio de los multiplicadores de Lagrange a la función objetivo del método de estimación semiparamétrico de Robinson (1995b). El contraste parece tener buenas propiedades de eficiencia frente a alternativas fraccionarias, comparado con otros contrastes semiparamétricos. Además el estadístico de contraste se computa fácilmente y, bajo ciertas condiciones, tiene una distribución asintótica χ^2 .

Pruebas de hipótesis y estimación en R.

Librería **fArma**: Contiene un grupo de funciones para investigar el comportamiento de memoria larga de un proceso de series de tiempo univariante. Entre ellas se encuentran funciones para:

- ✓ Simular un proceso ARFIMA(p,d,q): `farimaSim`
- ✓ Obtener las verdaderas autocorrelaciones: `farimaTrueacf`
- ✓ Calcular un estimador del efecto Hurst= $d+1/2$: `rsFit`, `perFit`, `whittleFit`

Librería **fracdiff**: Contiene funciones que permiten:

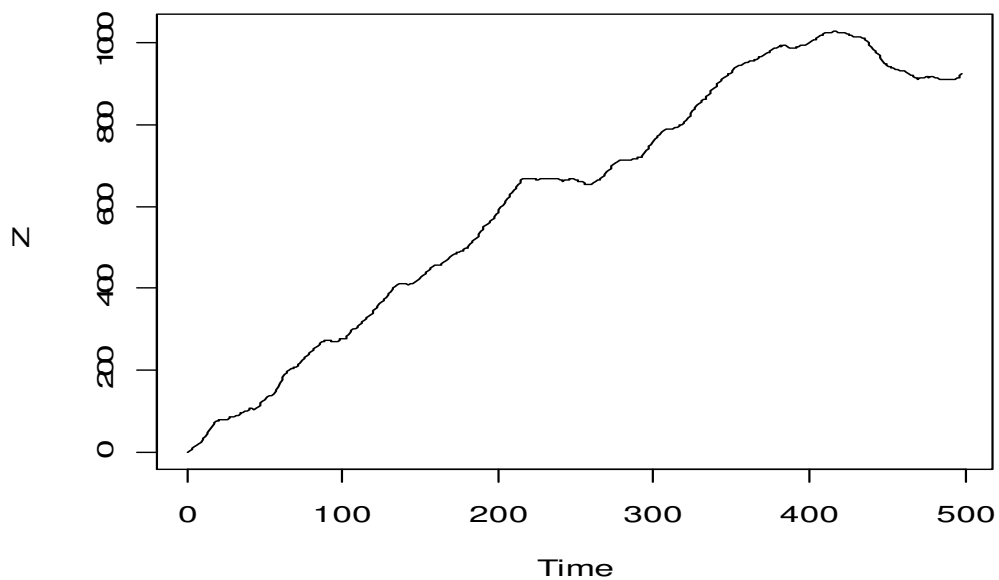
- ✓ La simulación de un ARFIMA(p,d,q): fracdiff.sim
- ✓ La estimación máximo verosímil de un ARFIMA(p,d,q): fracdiff
- ✓ La diferenciación fraccionaria de una serie: diffseries
- ✓ La estimación del parámetro fraccional (o memoria) d
en un ARFIMA(p,d,q) por el método GPH: fdGPH

Otros paquetes para la estimación de modelos ARFIMA:

Arfima package para Ox: Es un paquete para estimar, pronosticar y simular modelos ARFIMA y permite incluir regresores. Autores: JURGEN A. DOORNIK AND MARIUS OOMS.

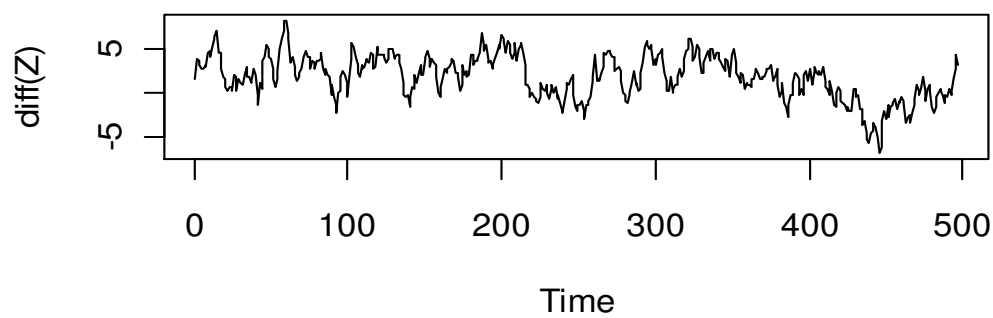
TSM (Time Series Modelling): Entre otras potencialidades estima, pronostica y simula modelos ARFIMA con o sin regresores. Corre bajo Ox. Autor: James Davidson.

Ejemplo: Identificación y estimación de un modelo ARFIMA no estacionario
Considere serie de tiempo presentada en la siguiente gráfica cuyos datos se encuentran en el archivo **arfm_noestac_norev.txt**.

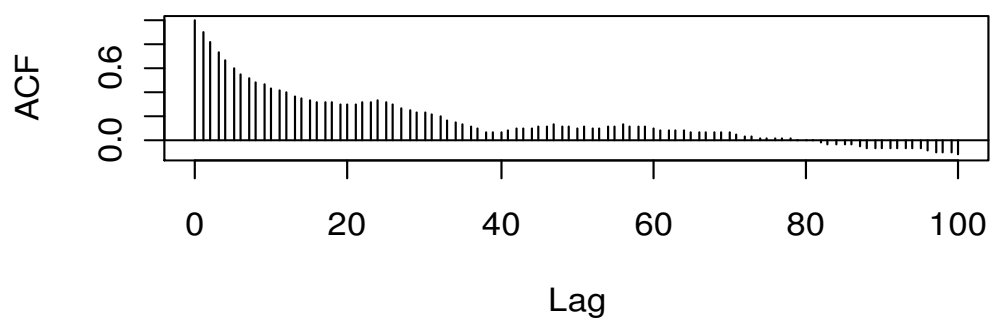


Proceso de identificación del modelo que genera los datos.

1. Diferencie una vez la serie Z_t . Es decir, obtenga la serie $(1-B) Z_t$. A continuación se presenta la serie diferenciada. Obtenga su ACF muestral.



Series $\text{diff}(Z)$



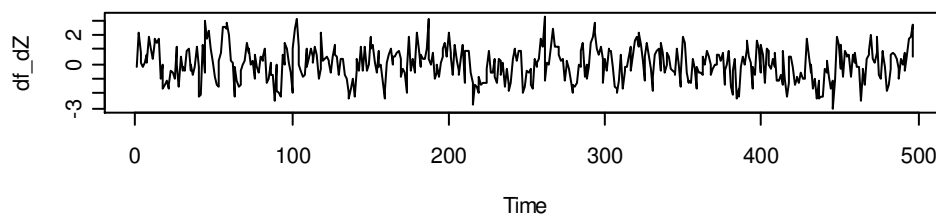
La serie diferenciada parece tener memoria larga.

2. Para la serie diferenciada estime el parámetro de diferenciación fraccional d usando, por ejemplo, el estimador GPH. Se obtiene

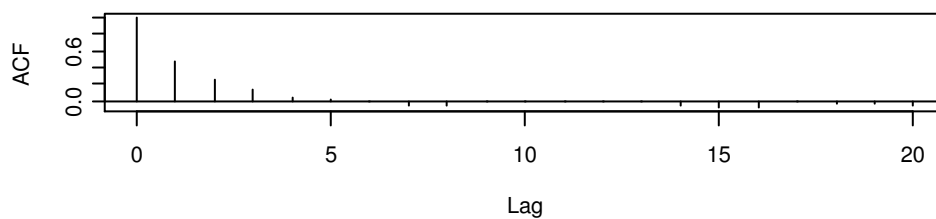
$\hat{d}_{GPH} = 0.52$. En la prueba de $H_0: d=0$ contra $H_1: d>0$ el estadístico de la prueba es $t = \hat{d}_{GPH} / se(\hat{d}_{GPH}) = 3.055$. para $\alpha = 0.05$ el valor crítico de la normal es -1.644854 . Por tanto se rechaza H_0 . Parece haber memoria larga.

3. Para la serie diferenciada $(1-B)Z_t$ diferénciela fraccionalmente usando

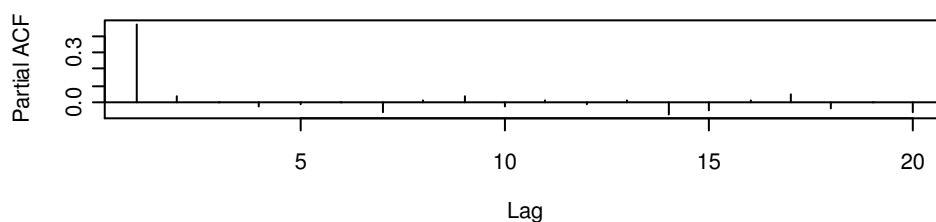
$\hat{d}_{GPH} = 0.52$. Obtenga su ACF y PACF muestrales.



Series df_dZ



Series df_dZ



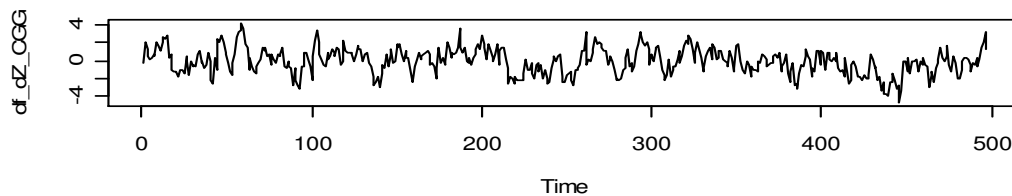
La componente de memoria corta de $(1-B)Z_t$ parece ser un $AR(1)$.

Entonces el modelo identificado para la serie es

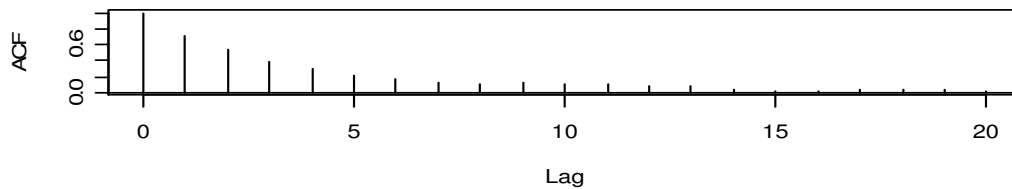
$$(1 - \phi B)(1 - B)^{d^*} Z_t = a_t, \text{ con } d^* > 1$$

Por tanto la serie parece provenir de un modelo ARFIMA no estacionario y sin reversión a la media.

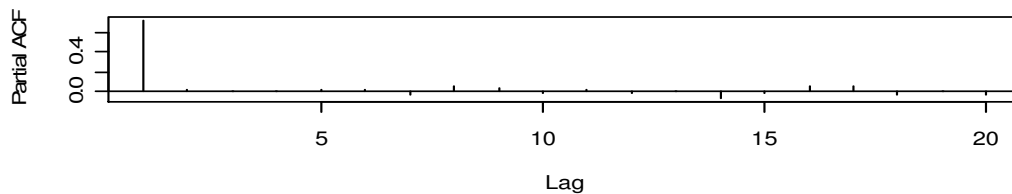
Otra forma de tratar la identificación es usar la estimación de d por medio del modelo aproximado ARFIMA($p^*, d, 0$) de la propuesta de Castaño, Gómez y Gallón. Usando un orden AR(5) para aproximar la componente de corto plazo se obtiene que el estimador de d es $\hat{d}_{CGG} = 0.27$ y que es estadísticamente significativo. Observe que en este caso el estimador satisface la condición de estacionaridad del proceso $(1-B)Z_t$. La gráfica y las ACF y PACF muestrales de la serie $(1-B)Z_t$ fraccionalmente diferenciada se encuentran a continuación.



Series df_dZ_CGG



Series df_dZ_CGG



Este procedimiento evidencia más claramente la existencia de una componente de corto plazo AR(1).

Script de R

```
library(fracdiff)
# lectura de datos
(serie=read.table("E:/UNal/Series de tiempo II/arfima_noestac_norev_n.txt",
header=TRUE))
attach(serie)

# gráfica de la serie y ACF
par(mfrow=c(2,1))
plot.ts(Z)
acf(Z, lag.max=100, ci=0)

# identificación A.
# Cuando la serie es claramente no estacionaria
# 1. Diferencie una vez la serie, i.e. use el operador (1-B)
dZ=diff(Z)
par(mfrow=c(2,1))
plot.ts(diff(Z))
acf(diff(Z), lag.max=100, ci=0)

# 2. obtenga una estimación de d usando GPH y pruebe la hipótesis
# de memoria larga
(d_GPH=fdGPH(dZ))
(t_GPH=d_GPH$d/d_GPH$sd.as)
(perc_normal_05=qnorm(0.05))
Conclusión: se rechaza Ho, i.e. hay memoria larga
# 3. Diferencie fraccionalmente la serie grafique la serie diferenciada
# obtenga la ACF y PACF
df_dZ=diffseries(dZ,d_GPH$d)
par(mfrow=c(3,1))
plot.ts(df_dZ)
acf(df_dZ, lag.max=20, ci=0)
pacf(df_dZ, lag.max=20, ci=0)
# Conclusión: El modelo parece ser ARFIMA(1,d,0) con d>1, es decir no
# estacionario y sin reversión a la media.

# identificación B.

# Tiene las mismas etapas que la identificación A, pero en lugar de usar el
# estimador GPH se estima un modelo ARFIMA(p*,d,0) para obtener un
# estimador preliminar de d (propuesta de Castaño, Gómez y Gallón).

mod_arfi=fracdiff(dZ, nar=5)
summary(mod_arfi) # Se observa que el estimador de d es positivo y
# estadísticamente significativo. Además es menor que 0.5
```

```
(d_CGG=coef(mod_arfi)[1])

df_dZ_CGG=diffseries(dZ,d_CGG)
par(mfrow=c(3,1))
plot.ts(df_dZ_CGG)
acf(df_dZ_CGG, lag.max=20, ci=0)
pacf(df_dZ_CGG, lag.max=20, ci=0)
# Conclusión: Con más evidencia se ve que el modelo parece ser ARFIMA(1,d,0)
# con d>1, es decir no estacionario y sin reversión a la media.
```

Lecturas:

Castaño, E. Gómez, K. y Gallón S. (2008), Una nueva prueba par el parámetro de diferenciación fraccional. Revista Colombiana de Estadística, Vol 31, No. 1, pp 67-84.

Beran, J. (1994), Statistitcs for long memory process. Chapman & Hall/CRC: New York.

Boutahar, M., Marimoutou, V. and Nourira, L. (2005) Estimation methods of the long memory parameter : Monte Carlo analysis and application. Documento de trabajo, GREQAM, Universités d'Aix-Marseille II et III.

Chan, N.H, and Palma, W. Estimation of Long-memory Time Series: A Survey of Likelihood-based Methods.

Schmidt, C.M. and Tscherning, R. (1994). The Identification of Fractional ARIMA Models.