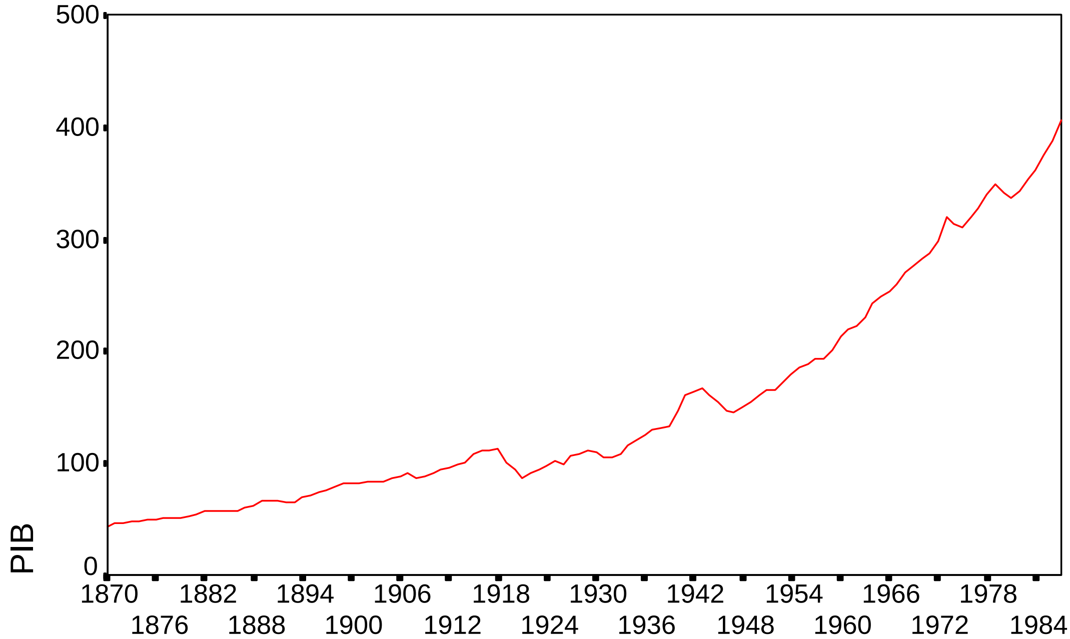
Capítulo 8. Introducción al análisis de series temporales.

8.1.- Caso 5: Análisis de algunas series temporales por la metodología Box-Jenkins.

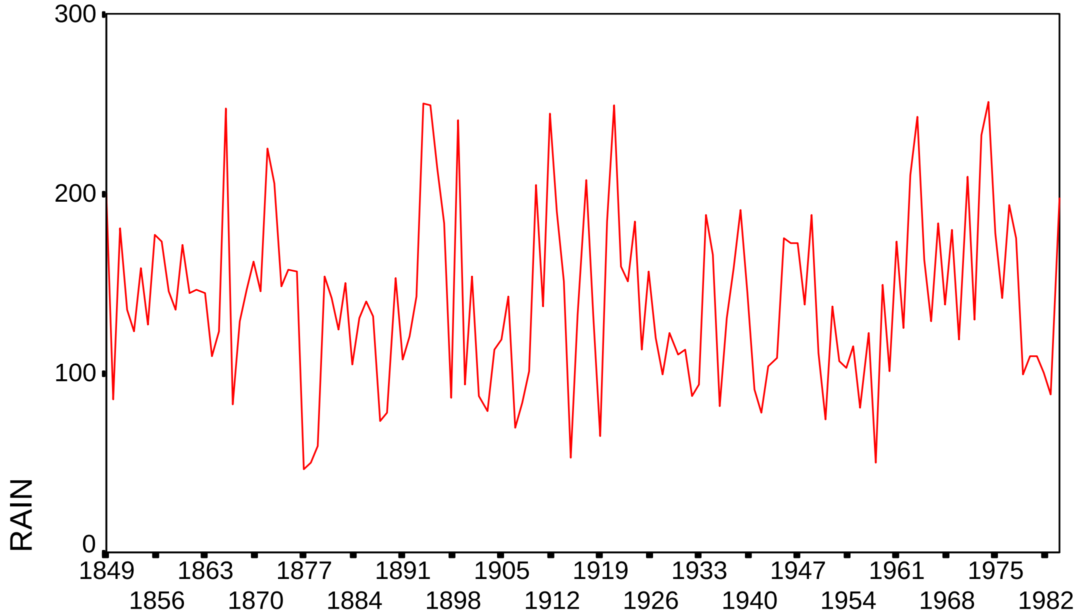
Una *serie temporal* puede definirse como una sucesión ordenada en el tiempo de valores de una variable. Aunque el tiempo es en realidad una variable continua, en la práctica, y al menos por lo que se refiere a este libro, utilizaremos las mediciones discretas correspondientes a periodos aproximadamente equidistantes en el tiempo.

Representaremos gráficamente la serie temporal del PIB de Gran Bretaña (un número índice con base 1913=100) en el periodo comprendido entre 1870 y 1987 (puede encontrarse en el fichero de SPSS, CASO5A.SAV):

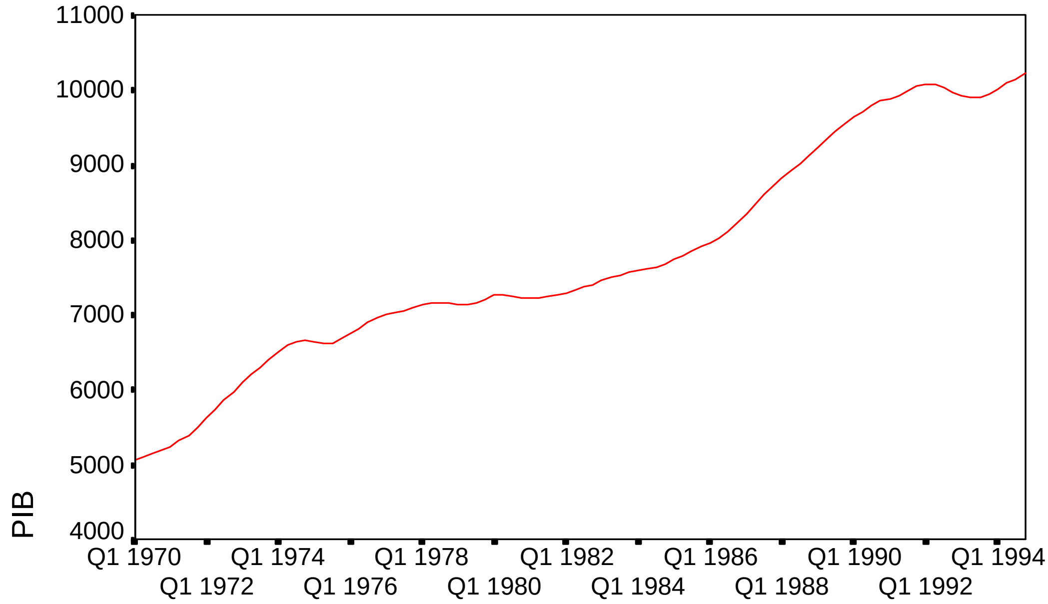


En este caso los valores del PIB se observaron anualmente.

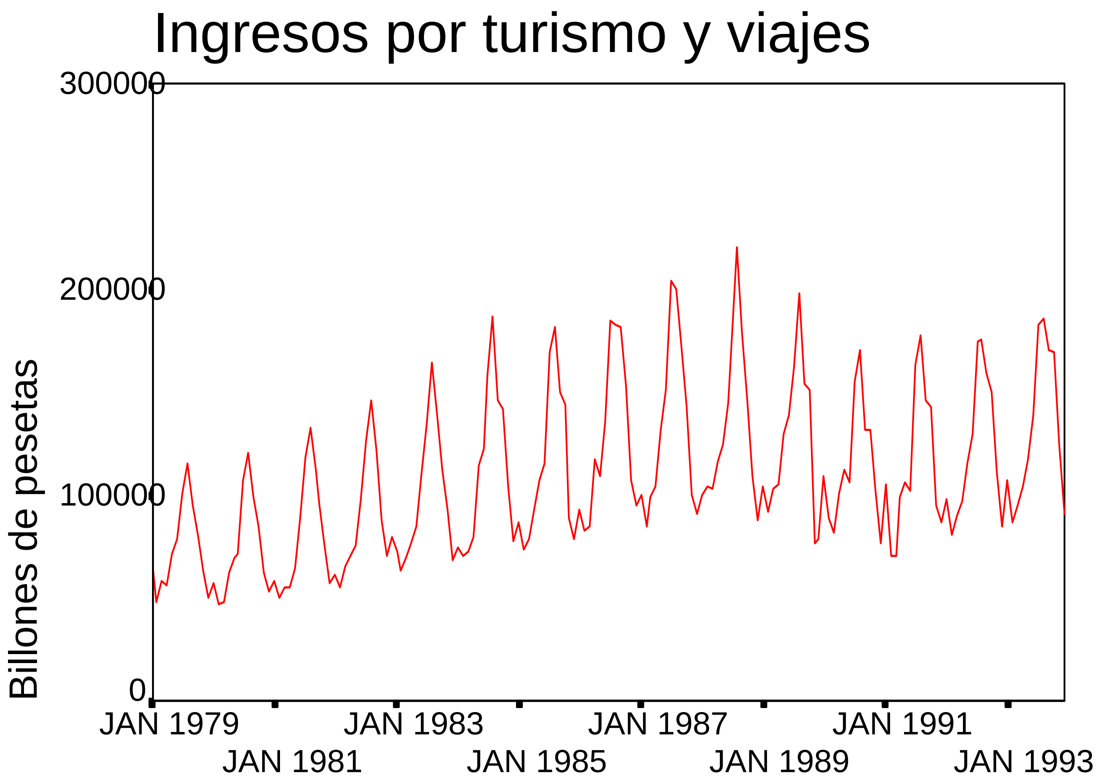
A continuación representamos la serie de datos anuales de la variable lluvia en Brasil (en centilitros) entre 1849 y 1984 (fichero CASO5B.SAV).

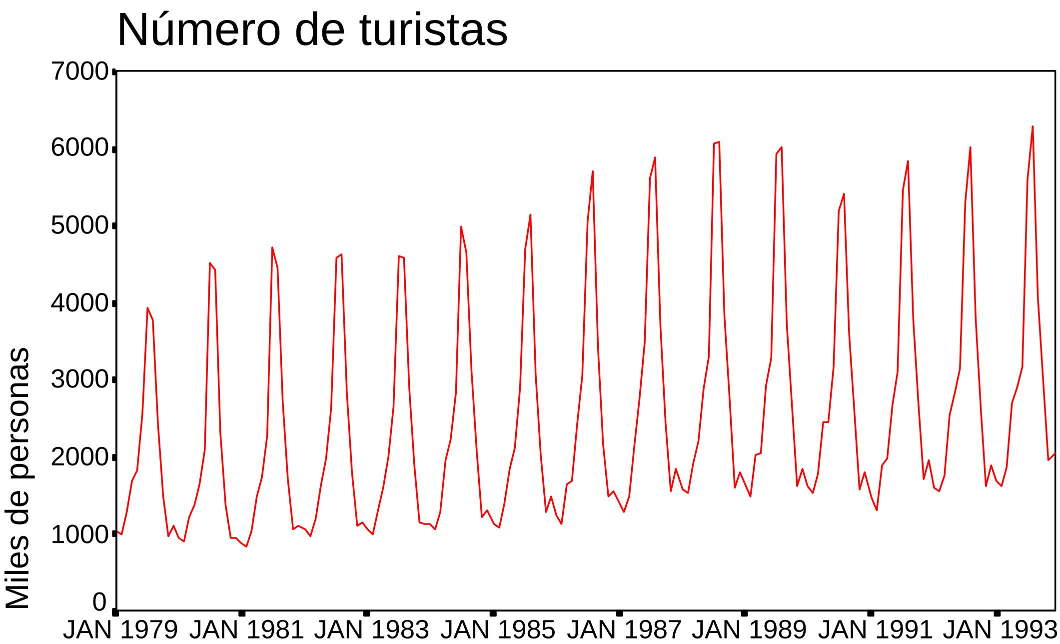


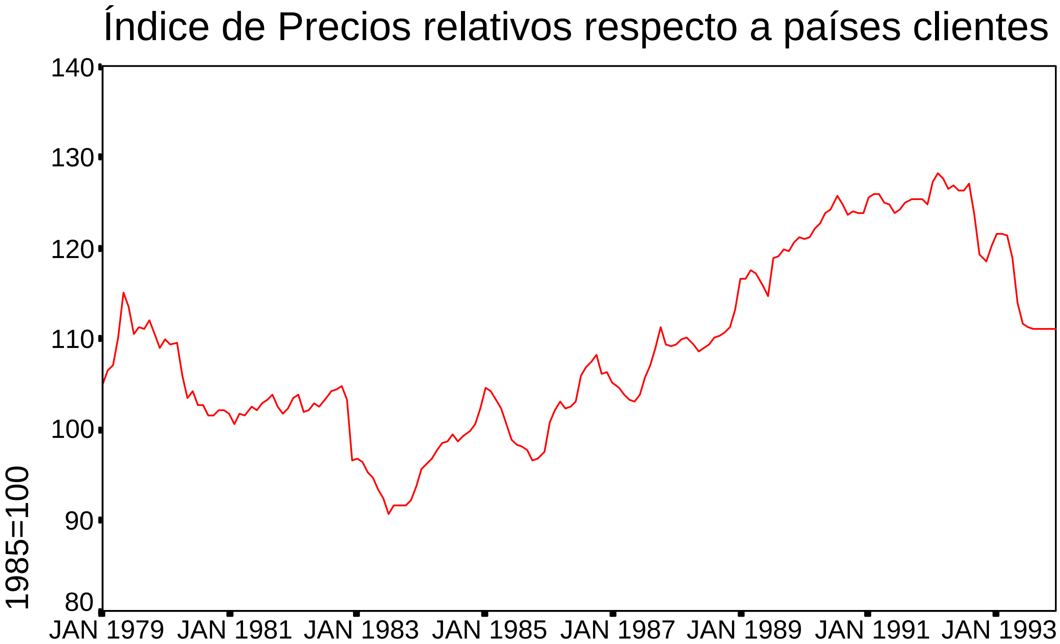
La variable PIB español (a precios de mercado y en pesetas de 1986) fue observada como una serie temporal trimestral (fichero CASO5C.SAV).

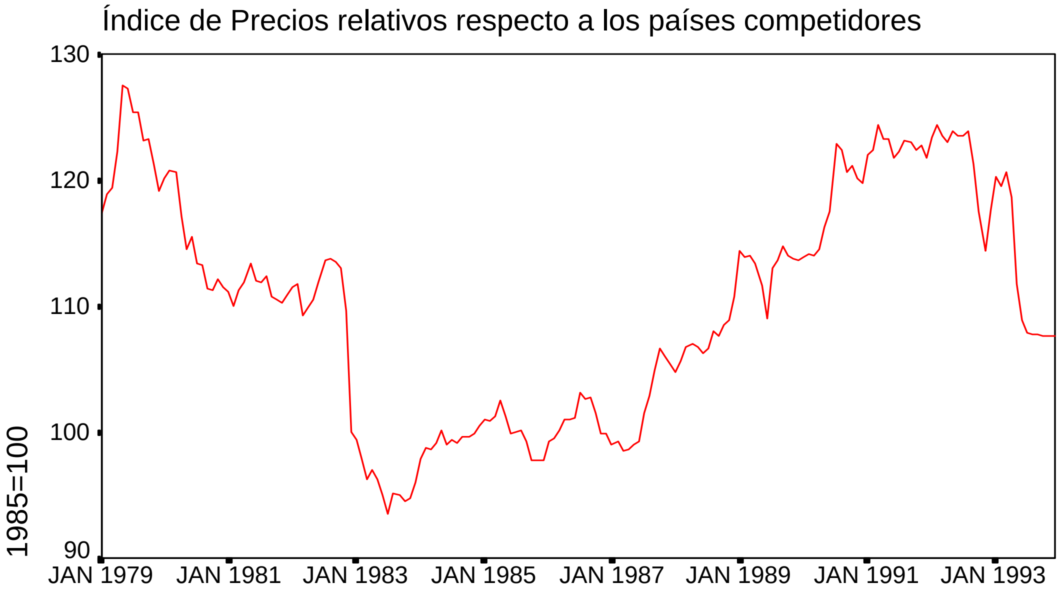


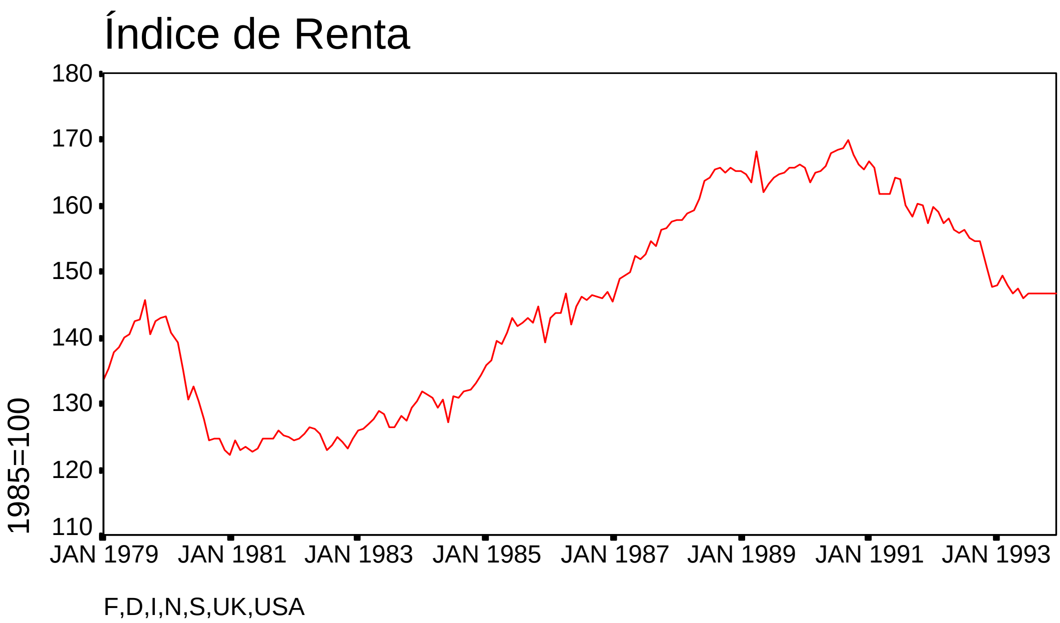
Finalmente, el fichero CASO5D.SAV contiene cinco series mensuales (Enero de 1979 – Diciembre de 1993): ingresos por turismo y viajes (TE) en billones de pesetas constantes de 1985; número de turistas (NT) en miles de personas; índice de precios relativos respecto a los países competidores (PRM) 1985=100; índice de precios relativos respecto a los países clientes (PRC) 1985=100; índice de renta de países clientes (INC) 1985=100. Los países considerados como competidores fueron Egipto, Francia, Grecia, Italia, Marruecos, Portugal, Túnez y Turquía, mientras que los países clientes fueron Alemania, Francia, Holanda, Italia, Portugal, Reino Unido, Suecia y USA. El índice de renta fue construido a partir de las series desestacionalizadas de los índices de producción industrial (IPI) de Alemania, Francia, Holanda, Italia, Reino Unido, Suecia y USA. Detalles adicionales pueden encontrarse en **González y Moral** (**1995**).




Número de turistas
Miles de personas
7000
6000
5000
4000
3000
2000
1000
0
JAN 1979 JAN 1981 JAN 1983 JAN 1985 JAN 1987 JAN 1989 JAN 1991 JAN 1993


Índice de Precios relativos respecto a países clientes
1985=100
140
130
120
110
100
90
80
JAN 1979 JAN 1981 JAN 1983 JAN 1985 JAN 1987 JAN 1989 JAN 1991 JAN 1993

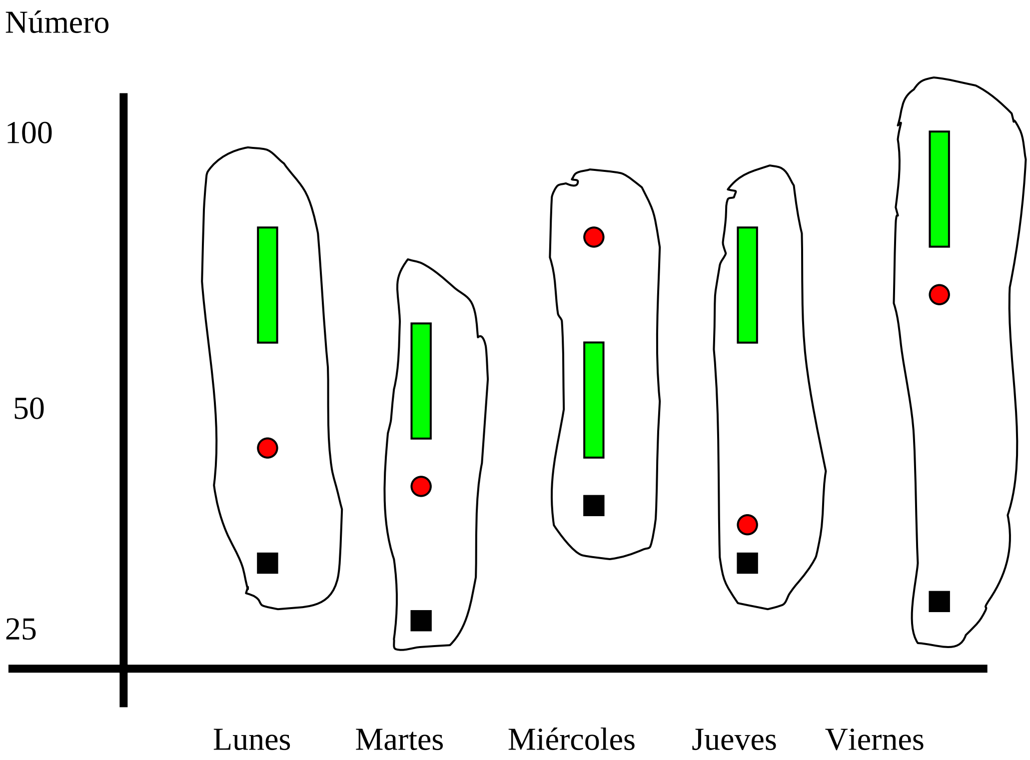




En este capítulo intentamos describir algunos de los métodos estadísticos utilizados para analizar una serie temporal sin tener en cuenta otras variables que puedan influir en la misma, métodos denominados *univariantes*. En particular nos interesa la *aproximación estocástica* (frente a la *determinista*) es decir aquélla que supone que la serie temporal tiene un carácter probabilístico, por cuanto ha sido generada por alguna variable aleatoria con una distribución de probabilidad determinada aunque habitualmente desconocida (el lector interesado en otros métodos y otras aproximaciones puede recurrir a **Otero** (**1993**) y **Murillo** (**1994**) entre otros).

8.2.- Conceptos preliminares: procesos estocásticos y requisitos para la inferencia.

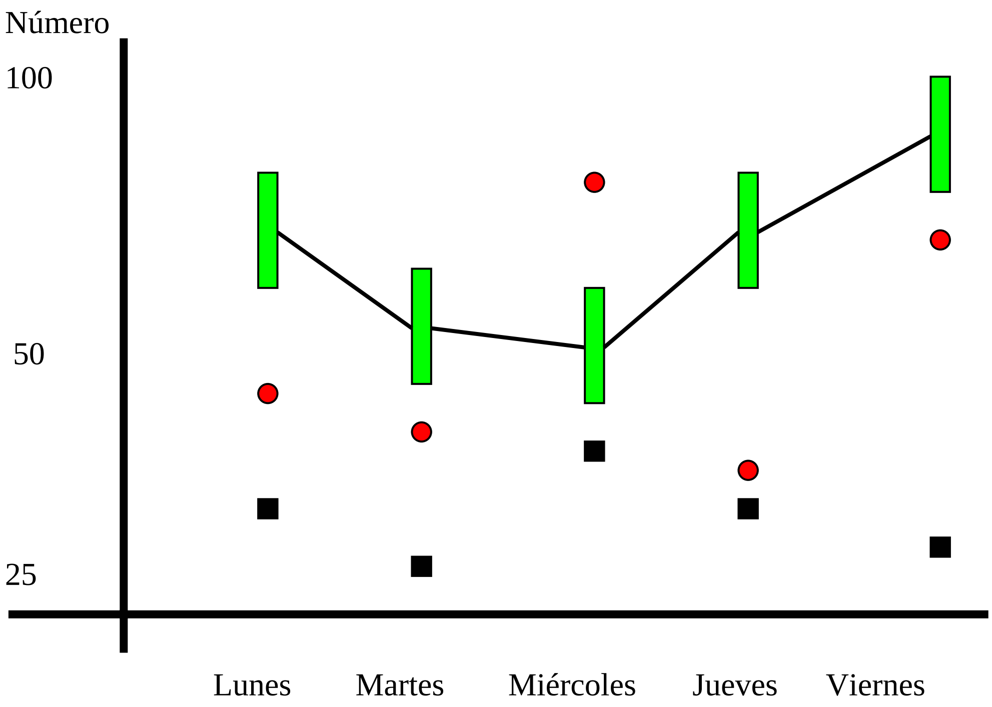
Bajo la aproximación estocástica una *serie temporal*, x_t, se define como una *realización muestral* de un *proceso estocástico*. Un proceso estocástico, $\{X_t\}$, t=1,2,...; es un conjunto de variables aleatorias, v.a., X_t, ordenadas según un parámetro temporal t. Consideremos por ejemplo la variable aleatoria “nacimientos en tres hospitales de Barcelona”:



Hospital del Mar, Hospital Clínico y Hospital de la Vall d’Hebró v.a.

Hemos representado el proceso estocástico {nacimientos en tres hospitales de Barcelona} para t=Lunes, Martes, Miércoles, Jueves, Viernes.

Definimos la serie temporal “nacimientos en el hospital de la Vall d’Hebró de Barcelona” como una realización muestral de este proceso estocástico, por ejemplo,



Hemos representado con una línea continua la serie temporal mencionada antes. Nótese que hemos supuesto que las tres series corresponden al mismo proceso estocástico. En el caso más habitual, sin embargo, sólo observaremos una única serie temporal para cada proceso estocástico.

El objetivo del análisis de series temporales es el de realizar inferencias sobre el proceso estocástico desconocido a partir de la serie temporal observada. El problema es que prácticamente todas las variables aleatorias, cuando se ordenan según un parámetro temporal, pueden ser consideradas un proceso estocástico y que, por otra parte, únicamente disponemos de una realización muestral del proceso (una única serie temporal). Por este motivo debemos imponer una serie de *condiciones* o requisitos que permitan realizar las inferencias de interés.

Linealidad

Entre todos los procesos estocásticos nos interesan aquéllos *lineales* o *gaussianos* (o *normales*). La distribución normal es la única distribución de probabilidad que permite la caracterización del proceso mediante el conocimiento de sólo dos de sus momentos, la media, μ_t=E(X_t) y la varianza, γ_{0t}=Var(X_t). Debido a esto únicamente nos preocuparemos de las inferencias sobre estos dos momentos.

Nótese que aunque lo hemos reducido bastante, el problema persiste, por cuanto el proceso podría tener infinitas medias y/o varianzas haciendo impracticable la inferencia.

Estacionariedad

Un proceso estocástico es *estacionario* en *sentido estricto*, o *fuertemente estacionario*, si las variables aleatorias que lo componen poseen la misma función de distribución de probabilidad, con independencia del momento del tiempo considerado y del número de variables aleatorias constitutivas del proceso. Se trata de una restricción demasiado fuerte por cuanto implica que las características del proceso estocástico no sufren alteración al considerar momentos del tiempo diferentes.

Por este motivo se define un proceso *estacionario en sentido amplio*, o *débilmente estacionario*, y a partir de ahora estacionario, como aquel con una media y una varianza constantes en el tiempo y con covarianzas ($\gamma_{t,t-k}$=Cov(X_t,X_{t-k})) que sólo dependen del lapso temporal considerado (k) pero no del tiempo (t).

$$\mu_t = \mu < \infty \qquad \gamma_{0t} = \gamma_0 < \infty \qquad \gamma_{t,t-k} = \gamma_k < \infty$$

En procesos estocásticos lineales la estacionariedad en sentido amplio implica la estacionariedad en sentido estricto.

La mayoría de los procesos no son estacionarios, pero, como veremos, se puede eliminar la tendencia y estabilizar la varianza para transformarlos en estacionarios.

Ergodicidad

El problema se ha reducido al de realizar inferencias únicamente sobre la media, la varianza y las covarianzas del proceso estocástico, parámetros invariantes en el tiempo, utilizando para ello la información suministrada por la serie temporal. Pero puesto que únicamente disponemos de una serie para cada proceso, deberíamos poder garantizar la idoneidad de los estimadores de los momentos; precisamente la *ergodicidad* lo permite. Según esta propiedad los momentos muestrales son estimadores *consistentes* de los momentos poblacionales.

Así la media del proceso puede ser estimada utilizando la media de la serie temporal, es decir:

$$\hat{\mu} = \frac{\sum_{t=1}^{T} x_t}{T} \qquad t = 1,2,\ldots,T$$

la varianza se puede estimar a través de la varianza de la serie:

$$\hat{\gamma}_0 = \frac{\sum_{t=1}^{T} \left(x_t - \hat{\mu}\right)^2}{T} \qquad t = 1,2,\ldots,T$$

y las covarianzas (denominadas *autocovarianzas*, por cuanto se refieren a covarianzas de la misma variable):

$$\hat{\gamma}_k = \frac{\sum_{t=1}^{T} \left(x_t - \hat{\mu}\right)\left(x_{t-k} - \hat{\mu}\right)}{T} \qquad t = 1,2,\ldots,T$$

Así, la autocovarianza de primer orden se puede expresar:

$$\hat{\gamma}_1 = \frac{\sum_{t=1}^{T} \left(x_t - \hat{\mu}\right)\left(x_{t-1} - \hat{\mu}\right)}{T} \qquad t = 1,2,\ldots,T$$

la de orden dos:

$$\hat{\gamma}_2 = \frac{\sum_{t=1}^{T}(x_t - \hat{\mu})(x_{t-2} - \hat{\mu})}{T} \qquad t = 1,2,...,T$$

y nótese que la de orden cero no es más que la varianza:

$$\hat{\gamma}_0 = \frac{\sum_{t=1}^{T}(x_t - \hat{\mu})(x_{t-0} - \hat{\mu})}{T} = \frac{\sum_{t=1}^{T}(x_t - \hat{\mu})^2}{T} \qquad t = 1,2,...,T$$

Definiremos, por último, los denominados *coeficientes de autocorrelación*:

$$\rho_k = \frac{Cov(X_t, X_{t-k})}{\sqrt{Var(X_t)}\sqrt{Var(X_{t-k})}} = \frac{\gamma_{t,t-k}}{\sqrt{\gamma_{0t}}\sqrt{\gamma_{0t-k}}}$$

Y cuando se trata de procesos estacionarios:

$$\rho_k = \frac{Cov(X_t, X_{t-k})}{\sqrt{Var(X_t)}\sqrt{Var(X_t)}} = \frac{\gamma_k}{\gamma_0}$$

que pueden ser estimados utilizando los coeficientes de autocorrelación muestral:

$$\hat{\rho}_k = \frac{\sum_{t=1}^{T}(x_t - \hat{\mu})(x_{t-k} - \hat{\mu})}{\sum_{t=1}^{T}(x_t - \hat{\mu})^2} \qquad t = 1,2,..,T \qquad -1 \leq \hat{\rho}_k \leq 1$$

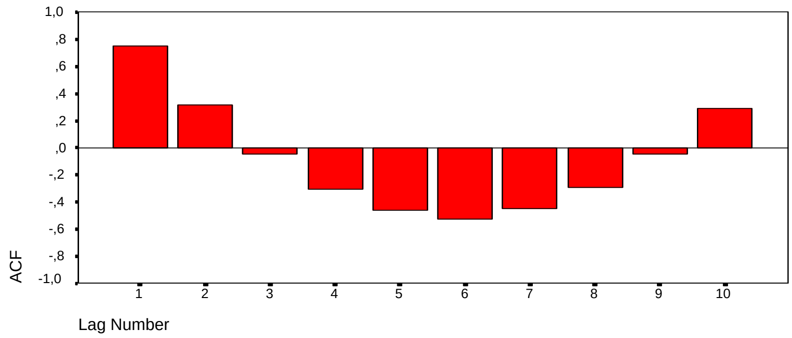
El coeficiente de autocorrelación de primer orden, por ejemplo, puede escribirse:

$$\hat{\rho}_1 = \frac{\sum_{t=1}^{T}(x_t - \hat{\mu})(x_{t-1} - \hat{\mu})}{\sum_{t=1}^{T}(x_t - \hat{\mu})^2} \qquad t = 1,2,..,T$$

y así para el resto, mientras que el de orden cero es siempre igual a 1, $\rho_0=\gamma_0/\gamma_0=1$.

La representación gráfica de los coeficientes de autocorrelación se denomina *Función de Autocorrelación Simple* (*FAS*), o *ACF* en sus siglas en inglés, y constituye un instrumento de análisis de series temporales de gran interés práctico. Por ejemplo,

k:	0	1	2	3	4	5	6	7	8	9	10
ρ_k	1,000	0,751	0,318	-0,046	-0,302	-0,457	-0,525	-0,449	-0,288	-0,044	0,294



8.3.- Procesos elementales: ruido blanco y camino aleatorio.

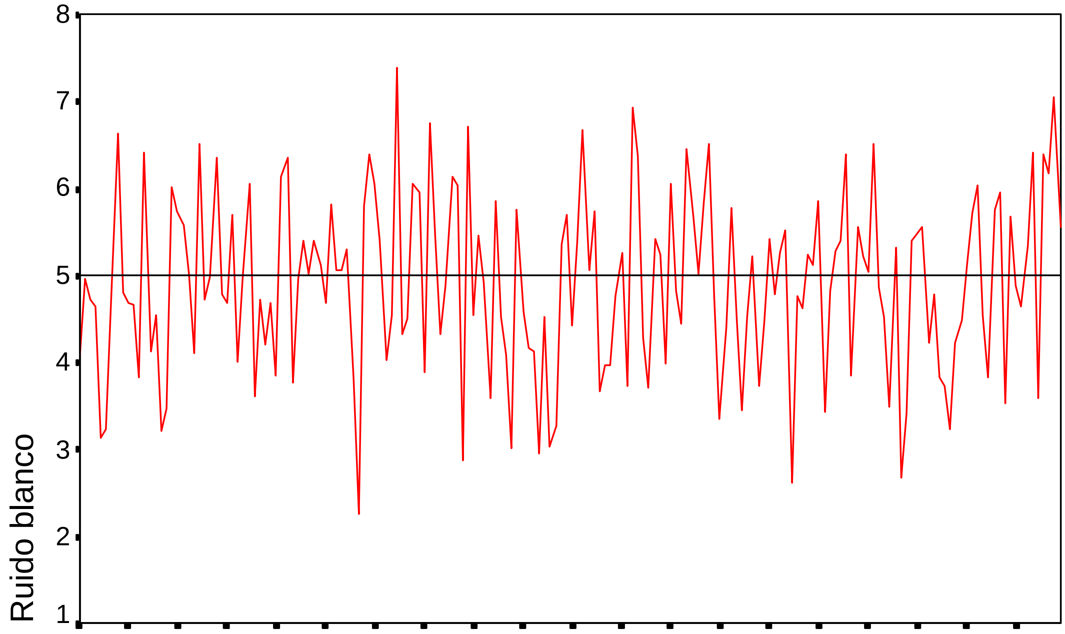
Proceso ruido blanco o puramente aleatorio

El proceso ruido blanco puede ser expresado como:

$$X_t = a + \varepsilon_t$$

siendo *a* una constante y ε_t una variable aleatoria idéntica e independientemente distribuida (es decir $Cov(\varepsilon_t,\varepsilon_{t-k}) = 0$ y $Cov(\varepsilon_{t-k},X_{t-k}) = 0$, $\forall k$) según una distribución normal con media nula, $\varepsilon_t \sim iid\ N(0,\sigma^2_\varepsilon)$, donde σ^2_ε es una constante.

La representación gráfica de un ruido blanco es la de una variable que oscila aleatoriamente en torno a su media *a* con varianza constante.



El término de perturbación aleatoria de un modelo de regresión lineal, por ejemplo, debería ser un ruido blanco si el modelo cumpliese todas las hipótesis básicas (véase Capítulo 2).

El proceso ruido blanco es *puramente* estacionario por cuanto sus momentos no varían con el tiempo. Así:

<u>Media</u>

$$\mu = E(X_t) = E(a + \varepsilon_t) = a + E(\varepsilon_t) = a \qquad \textit{constante}$$

<u>Varianza</u>

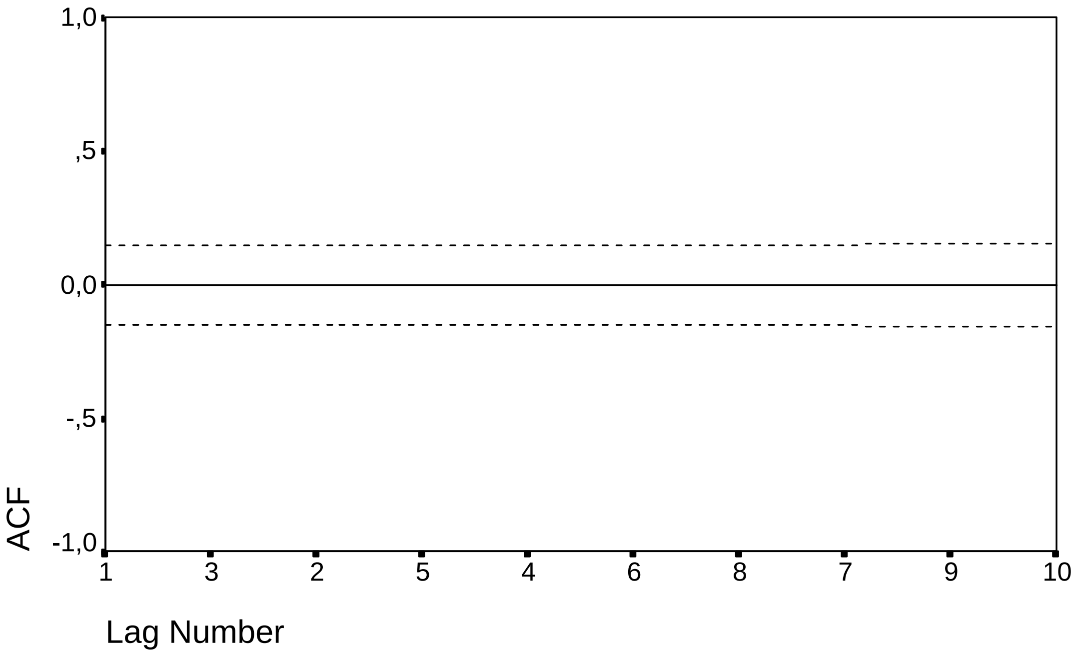
$$\gamma_0 = Var(X_t) = E(X_t - E(X_t))^2 = E(a + \varepsilon_t - a)^2 = E(\varepsilon_t)^2 = \sigma_\varepsilon^2 \qquad \textit{constante}$$

puesto que $\sigma^2_\varepsilon = \mathrm{Var}(\varepsilon_t) = \mathrm{E}(\varepsilon_t - \mathrm{E}(\varepsilon_t))^2 = \mathrm{E}(\varepsilon_t)^2 = \mathrm{E}(\varepsilon_t^2)$.

<u>Autocovarianzas</u>

$$\gamma_k = Cov(X_t, X_{t-k}) = E\left[(X_t - E(X_t))(X_{t-k} - E(X_{t-k}))\right] = E(\varepsilon_t \varepsilon_{t-k}) = 0 \qquad \forall k$$

Por lo que todos los coeficientes de autocorrelación son iguales a cero (excepto el de orden 0), $\rho_k=0$ $\forall k>0$. Por tanto, la función de autocorrelación simple, FAS, es *blanca*:



Hemos dibujado las bandas de confianza de los coeficientes de autocorrelación, las cuales comentaremos más adelante.

Proceso camino aleatorio

Con *deriva*

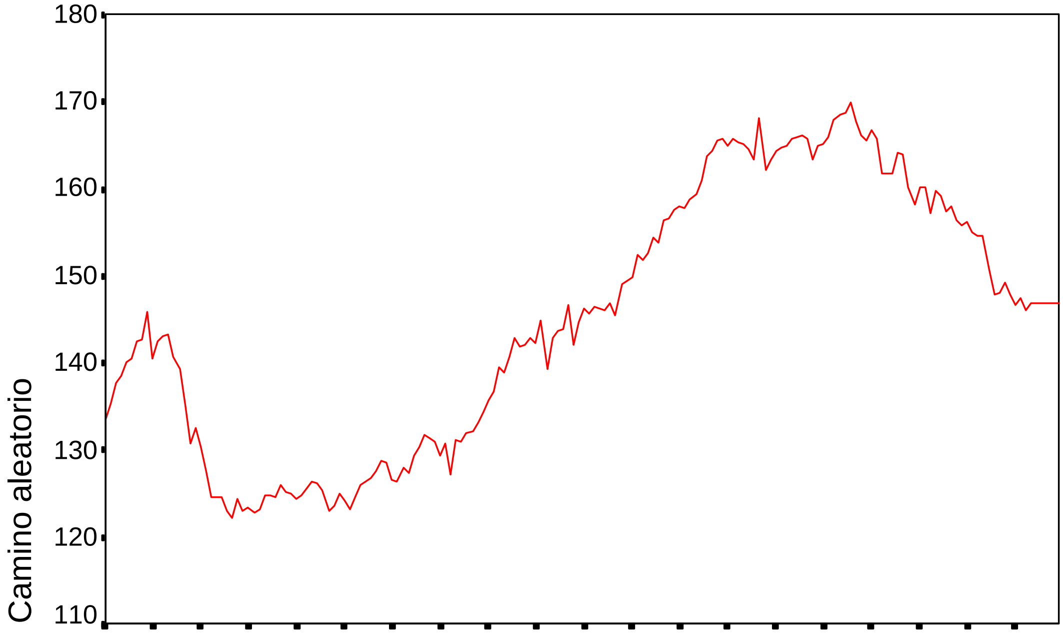
$$X_t = a + X_{t-1} + \varepsilon_t$$

Sin *deriva*

$$X_t = X_{t-1} + \varepsilon_t$$

siendo *a* una constante y ε_t~iid $N(0,\sigma^2_\varepsilon)$, donde σ^2_ε es una constante.

Se trata de un proceso (puramente) no estacionario, con una representación gráfica que presenta una media y/o varianza que varían en el tiempo. Por ejemplo:



Los momentos del proceso camino aleatorio no son constantes en el tiempo:

<u>Media</u>

$$\mu = E(X_t) = E(a + X_{t-1} + \varepsilon_t) = a + E(X_{t-1}) = a + E(a + X_{t-2} + \varepsilon_{t-1}) = 2a + E(X_{t-2}) = ... = Ta$$

la cual depende de T.

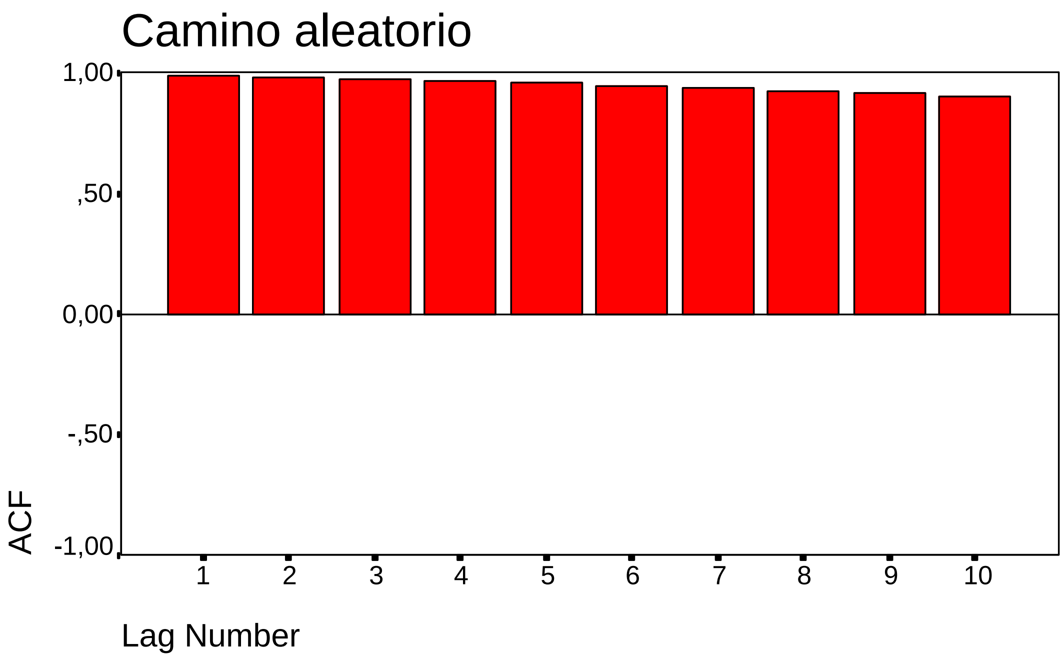
<u>Varianza</u>

Suponiendo que el proceso no tiene deriva para facilitar los cálculos:

$$\gamma_0 = Var(X_t) = E(X_t - E(X_t))^2 = E(X_{t-1} + \varepsilon_t - E(X_{t-1}))^2 = \gamma_{t-k} + k\sigma_\varepsilon{}^2$$

que también depende de T.

Las autocovarianzas también varían en el tiempo, implicando coeficientes de autocorrelación próximos a la unidad y con un decrecimiento muy lento conforme aumenta el lapso temporal, k.



En este caso se dice que la FAS está *muy cargada*, que es la situación típica de una serie no estacionaria en media.

8.4.- Procesos estocásticos estacionarios, lineales y ergódicos. Modelos lineales de series temporales.

Para analizar series temporales, realizaciones muestrales de procesos estacionarios, se utilizan tres tipos de *modelos lineales*: el autorregresivo (AR), el de medias móviles (MA) y el modelo mixto (ARMA).

Modelos Autorregresivos (AR)

AR(1)

El modelo autorregresivo de primer orden, AR(1) puede ser expresado como:

$$X_t = a + \phi X_{t-1} + \varepsilon_t$$

siendo *a* una constante, ϕ un parámetro desconocido, y ε_t~iid $N(0,\sigma^2_\varepsilon)$, donde σ^2_ε es una constante.

Una expresión alternativa se obtiene utilizando el denominado *operador de retardo* L (o B), mediante el cual: $LX_t=X_{t-1}$ ó $L^2X_t=X_{t-2}$ y, en general, $L^hX_t=X_{t-h}$.

$$X_t = a + \phi X_{t-1} + \varepsilon_t = a + \phi L X_t + \varepsilon_t$$

$$X_t - \phi L X = a + \varepsilon_t$$

$$(1-\phi L) X_t = a + \varepsilon_t$$

<u>Media</u>

$$\mu = E(X_t) = E(a + \phi X_{t-1} + \varepsilon_t) = a + \phi E(X_{t-1}) + E(\varepsilon_t) = a + \phi\mu$$

$$\mu = \frac{a}{1-\phi}$$

por cuanto se trata de un proceso estacionario, $E(X_t)=E(X_{t-k})$, $\forall k$.

Nótese que el parámetro phi debe ser forzosamente distinto que la unidad, $\phi\neq1$, por cuanto en caso contrario la media es infinita, $\mu=\infty$ (como en un camino aleatorio).

<u>Varianza</u>

Por simplicidad supondremos que el AR(1) tiene media nula a=0 y, por tanto, μ=0.

$$\gamma_0 = Var(X_t) = E(\phi X_{t-1} + \varepsilon_t)^2 = E(\phi^2 X_{t-1}^2 + \varepsilon_t^2 + 2\phi X_{t-1}\varepsilon_t) = \phi^2 E(X_{t-1}^2) + E(\varepsilon_t^2) + 2\phi E(X_{t-1}\varepsilon_t) = \phi^2\gamma_0 + \sigma^2_\varepsilon$$

$$\gamma_0 = \frac{\sigma^2_\varepsilon}{1-\phi^2}$$

Nótese que debe cumplirse que $|\phi|<1$ para que la media y la varianza sean constantes no infinitas y para que esta última no sea negativa. Esta condición se denomina de *estacionariedad*.

<u>Autocovarianzas</u>

$$\gamma_k = Cov(X_t, X_{t-k}) = E[(X_t - E(X_t))(X_{t-k} - E(X_{t-k}))] = E(X_t X_{t-k}) =$$
$$E[(\phi X_{t-1} + \varepsilon_t)X_{t-k}] = \phi E(X_{t-1}X_{t-k}) + E(\varepsilon_t X_{t-k}) = \phi\gamma_{k-1}$$

Así,

$$\gamma_1 = \phi\gamma_0$$
$$\gamma_2 = \phi\gamma_1 = \phi^2\gamma_0$$
$$\gamma_3 = \phi\gamma_2 = \phi^3\gamma_0$$
$$\ldots$$
$$\gamma_k = \phi^k\gamma_0 \qquad \forall k > 0$$

<u>Coeficientes de autocorrelación</u>

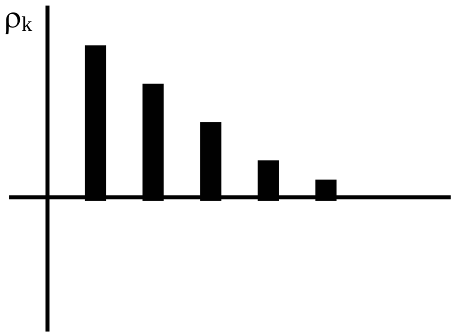
Por tanto,

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\phi^k\gamma_0}{\gamma_0} = \phi^k$$

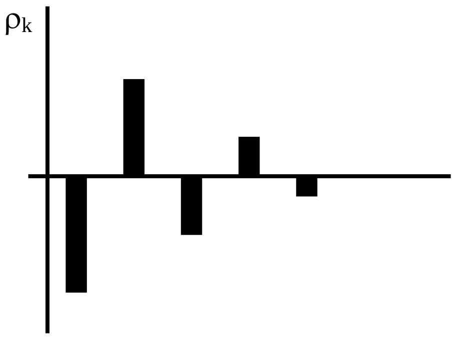
$$\rho_1 = \phi$$
$$\rho_2 = \phi^2$$
$$\ldots$$
$$\rho_k = \phi^k$$

La FAS de un AR(1) presenta un rápido decrecimiento:

a) Exponencial (ϕ>0)



b) Alternante (ϕ<0)



AR(2)

$$X_t = a + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t$$

siendo *a* una constante, ϕ_1 y ϕ_2 parámetros desconocidos, y ε_t~iid $N(0,\sigma^2_\varepsilon)$, donde σ^2_ε es una constante.

De forma alternativa:

$$(1-\phi_1 L - \phi_2 L^2)X_t = a + \varepsilon_t$$

Media

$$\mu = E(X_t) = E(a + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t) = a + \phi_1 E(X_{t-1}) + \phi_2 E(X_{t-2}) + E(\varepsilon_t) = a + \phi_1 \mu + \phi_2 \mu$$

$$\mu = \frac{a}{1 - \phi_1 - \phi_2}$$

En este caso se debe cumplir $\phi_1 + \phi_2 \neq 1$ para que $\mu \neq \infty$.

Varianza

Por simplicidad supondremos que el AR(2) tiene media nula a=0 y, por tanto, $\mu=0$.

$$\gamma_0 = Var(X_t) = E\left[(\phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t) X_t\right] = E(\phi_1 X_{t-1} X_t + \phi_2 X_{t-2} X_t + \varepsilon_t X_t) =$$
$$= \phi_1 E(X_{t-1} X_t) + \phi_2 E(X_{t-2} X_t) + E(\varepsilon_t X_t) = \phi_1 \gamma_1 + \phi_2 \gamma_2 + E(\varepsilon_t(\phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t)) = \phi_1 \gamma_1 + \phi_2 \gamma_2 + \sigma_\varepsilon^2$$

$$\gamma_0 = \phi_1 \gamma_1 + \phi_2 \gamma_2 + \sigma_\varepsilon^2$$

o lo que es lo mismo

$$\gamma_0 = \phi_1 \rho_1 \gamma_0 + \phi_2 \rho_2 \gamma_0 + \sigma_\varepsilon^2 \qquad \gamma_0(1 - \phi_1 \rho_1 - \phi_2 \rho_2) = \sigma_\varepsilon^2$$

$$\gamma_0 = \frac{\sigma_\varepsilon^2}{1 - \phi_1 \rho_1 - \phi_2 \rho_2}$$

Las condiciones de estacionariedad para un AR(2) son: a) $\phi_1+\phi_2<1$; b) $\phi_2-\phi_1<1$ y c) $|\phi_1|<1$ y $|\phi_2|<1$.

Autocovarianzas

$$\gamma_k = Cov(X_t, X_{t-k}) = E\left[(\phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t) X_{t-k}\right] =$$
$$= \phi_1 E(X_{t-1} X_{t-k}) + \phi_2 E(X_{t-2} X_{t-k}) + E(\varepsilon_t X_{t-k}) = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2}$$

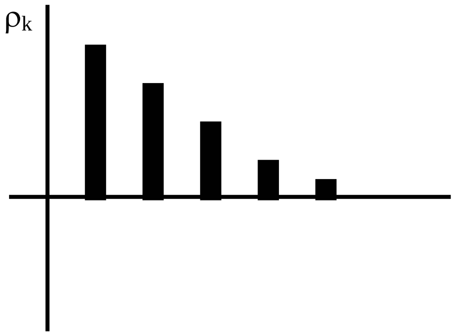
Coeficientes de autocorrelación

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2}}{\gamma_0} = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2}$$
$$\rho_1 = \phi_1 + \phi_2 \rho_1 \qquad (\rho_0 = 0 \;\; y \;\; \rho_{-1} = \rho_1)$$

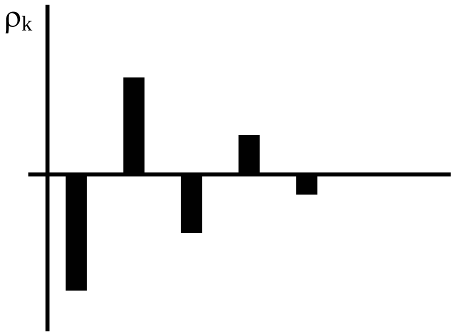
$$\rho_1 = \frac{\phi_1}{1 - \phi_2}$$
$$\rho_2 = \phi_1 \rho_1 + \phi_2 \rho_0 = \phi_1 \frac{\phi_1}{1-\phi_2} + \phi_2 = \frac{\phi_1^2}{1-\phi_2} + \phi_2$$
...
$$\rho_k \neq 0 \quad \forall k > 0$$

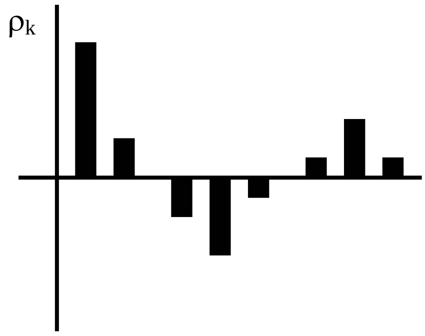
La FAS de un AR(2), como la del AR(1), también presenta un rápido decrecimiento:

a) Exponencial



b) Alternante





De hecho todos los AR presentan funciones de autocovarianza muy similares, lo que implica, como acabamos de ver, un decrecimiento en la FAS de todos ellos.

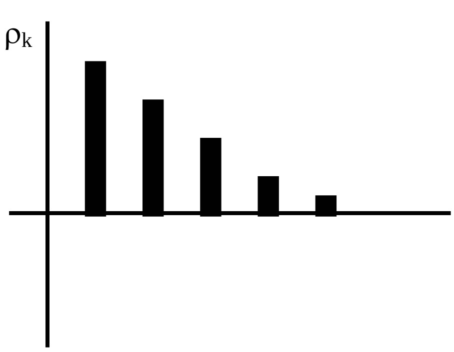
A fin de distinguir entre distintos modelos AR definimos la *función de autocorrelación parcial* (*FAP*), *PACF* en sus siglas en inglés. La FAP podría entenderse como la representación de los parámetros ϕ_{kk} en el modelo AR(k), los cuales se obtienen resolviendo recursivamente las ecuaciones implicadas por el modelo (véase **Otero** (**1993**) para más información).

Por ejemplo, en un AR(1) el coeficiente de correlación parcial, ϕ_{11}, se obtendría resolviendo la siguiente ecuación:

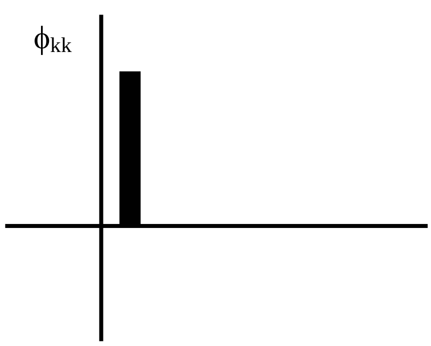
$$X_t=\phi_{11}\ X_{t-1}+\varepsilon_t$$

Por lo que la FAS y la FAP de un AR(1) presentan la siguiente forma:

a) Parámetro positivo, $\phi>0$

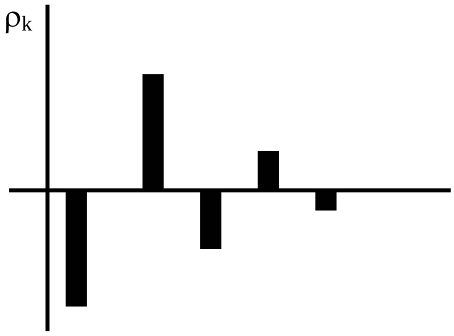


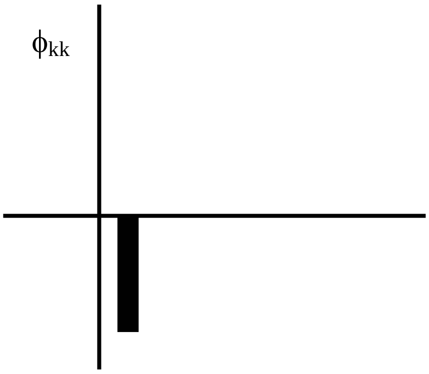
ACF



PACF

b) Parámetro negativo, ϕ<0





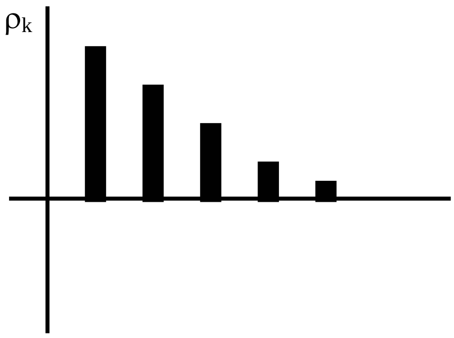
ACF PACF

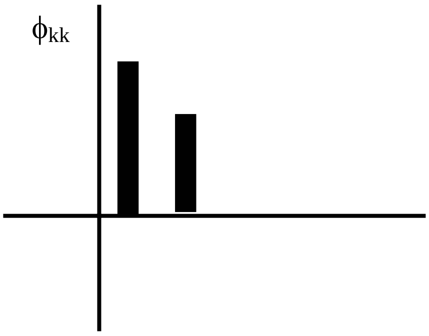
Los coeficientes de autocorrelación parcial de un AR(2), ϕ_{11} y ϕ_{22}, se obtienen resolviendo simultáneamente las siguientes ecuaciones:

$$X_t=\phi_{11}\, X_{t-1}+\varepsilon_t$$

$$X_t=\phi_{12}\, X_{t-1}+\phi_{22}\, X_{t-1}+\varepsilon_t$$

Presentando FAS y FAP como las que siguen (entre otras):

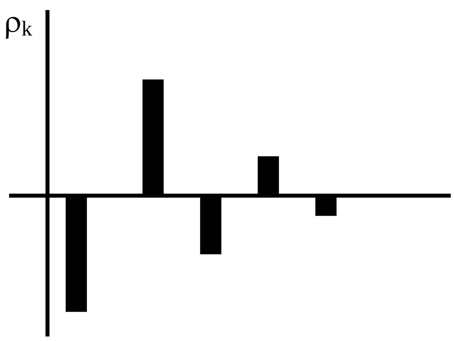


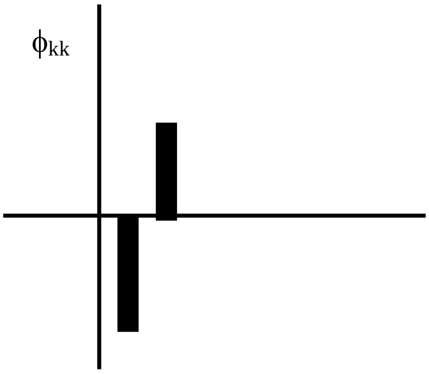


ACF PACF

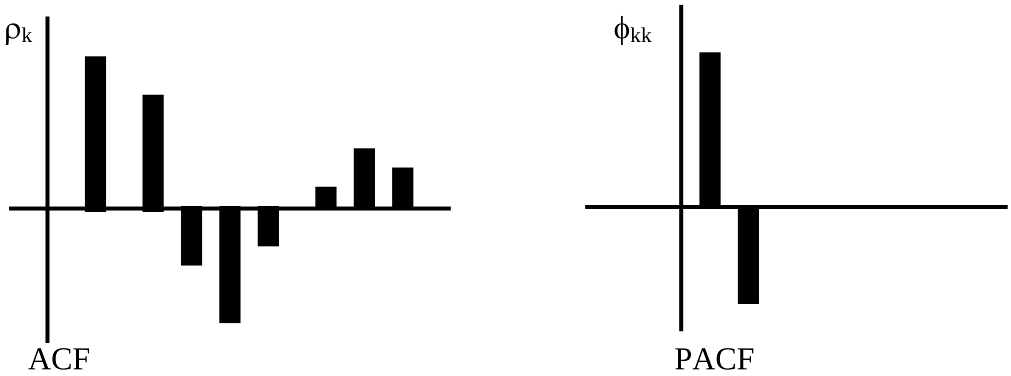
O bien,

ACF PACF





Y finalmente,



En general, un AR(p) puede expresarse:

$$X_t = a + \phi_1 X_{t-1} + \phi_2 X_{t-2} + ... + \phi_p X_{t-p} + \varepsilon_t$$

o de forma alternativa:

$$(1 - \phi_1 L - \phi_2 L^2 + ... + \phi_p L^p) X_t = a + \varepsilon_t$$

a es una constante; ϕ_1, ..., ϕ_p parámetros desconocidos; ε_t~iidN(0,σ^2_ε) con σ^2_ε constante.

Cualquier AR(p) tiene una media y una varianza constante y funciones de autocovarianza que implican una FAS decreciente, mientras que la FAP presenta p *palos*.

Modelos Medias Móviles (MA)

MA(1)

El modelo media móvil de primer orden, MA(1) puede ser expresado como sigue:

$$X_t = a + \varepsilon_t - \theta\varepsilon_{t-1}$$

o bien:

$$X_t = a + (1 - \theta L)\varepsilon_t$$

siendo *a* una constante,θ un parámetro deconocido, y ε_t~iid N(0,σ^2_ε) con σ^2_ε constante.

Media

$$\mu = E(X_t) = E(a + \varepsilon_t + \theta\varepsilon_{t-1}) = a$$

Varianza

Por simplicidad supondremos que el MA(1) tiene media nula a=0 y, por tanto, μ=0.

$$\gamma_0 = Var(X_t) = E(\varepsilon_t - \theta\varepsilon_{t-1})^2 = E(\varepsilon_t)^2 + \theta^2 E(\varepsilon_{t-1})^2 - 2\theta E(\varepsilon_t\varepsilon_{t-1}) = \sigma_\varepsilon^2(1+\theta^2)$$

Nótese que la varianza es siempre constante y positiva, por lo que el proceso MA(1) es siempre estacionario (de hecho todos los MA). Sin embargo debe cumplir que |θ|<1 para que sea *invertible* (como veremos más adelante).

Autocovarianzas

$$\gamma_1 = E(X_t X_{t-1}) = E\left[(\varepsilon_t - \theta\varepsilon_{t-1})(\varepsilon_{t-1} - \theta\varepsilon_{t-2})\right] = -\theta E(\varepsilon_{t-1})^2 = -\theta\sigma_\varepsilon^2$$

$$\gamma_2 = E(X_t X_{t-2}) = E\left[(\varepsilon_t - \theta\varepsilon_{t-1})(\varepsilon_{t-2} - \theta\varepsilon_{t-3})\right] = 0 \qquad \gamma_k = 0 \quad \forall k > 1$$

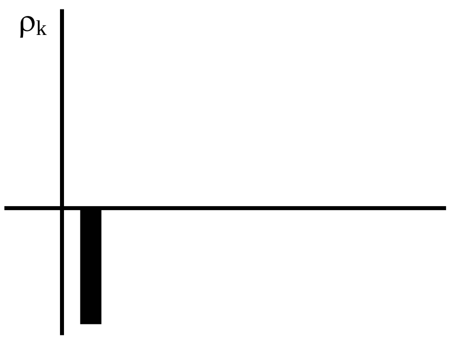
Coeficientes de autocorrelación

Por tanto,

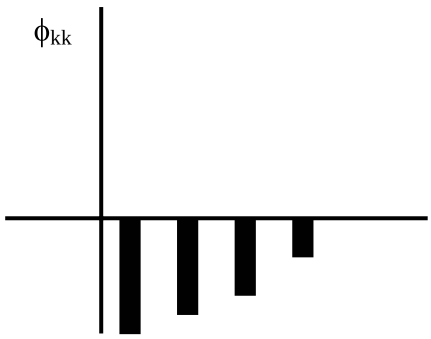
$$\rho_1 = \frac{\gamma_k}{\gamma_0} = \frac{-\theta\sigma_\varepsilon^2}{(1+\theta^2)\sigma_\varepsilon^2} = \frac{-\theta}{1+\theta^2}$$
$$\rho_k = 0 \quad \forall k > 1$$

La FAS de un MA(1) presenta un único palo y la FAP un rápido decrecimiento (al revés que un AR(1):

a) Parámetro positivo, θ>0

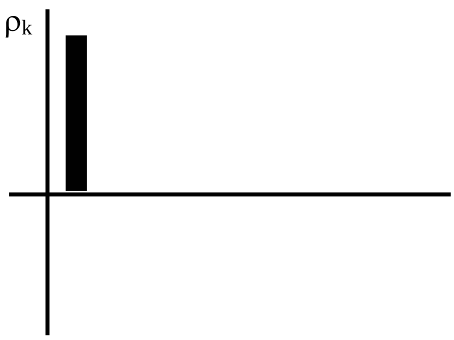


ACF

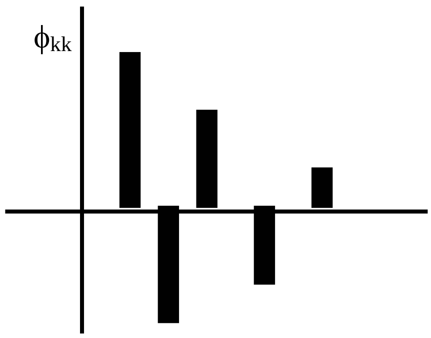


PACF

b) Parámetro negativo, θ<0



ACF



PACF

MA(2)

$$X_t = a + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2}$$

o bien:

$$X_t = a + (1 - \theta_1 L - \theta_2 L^2)\varepsilon_t$$

siendo *a* una constante, θ_1 y θ_2 parámetros desconocidos, y $\varepsilon_t \sim$iid $N(0,\sigma^2_\varepsilon)$, donde σ^2_ε es una constante.

<u>Media</u>

$$\mu = E(X_t) = a$$

<u>Varianza</u>

$$\gamma_0 = Var(X_t) = \sigma_\varepsilon^2 (1 + \theta_1^2 + \theta_2^2)$$

Las condiciones de invertibilidad en un MA(2) son: a) $\theta_1+\theta_2<1$; b) $\theta_2-\theta_1<1$ y c) $|\theta_q|<1$ y $|\theta_2|<1$.

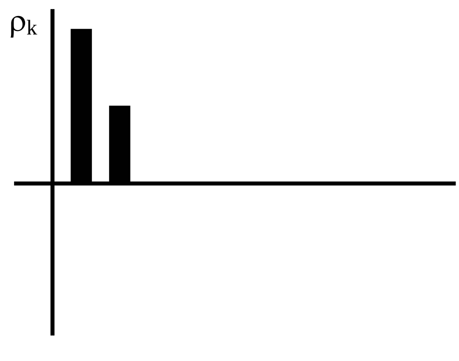
<u>Autocovarianzas</u>

$$\gamma_1 = \sigma_\varepsilon^2 (-\theta_1 + \theta_1\theta_2)$$
$$\gamma_2 = -\theta_2\sigma_\varepsilon^2$$
$$\gamma_k = 0 \quad \forall k > 2$$

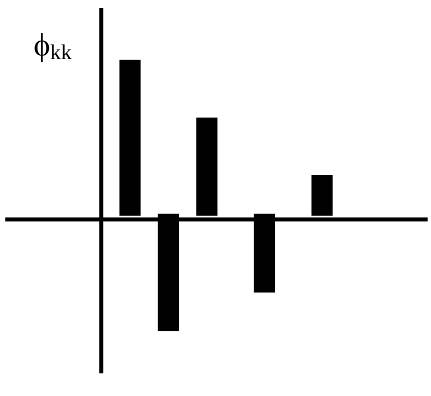
<u>Coeficientes de autocorrelación</u>

$$\rho_1 = \frac{-\theta_1 + \theta_1\theta_2}{1 + \theta_1^2 + \theta_2^2}$$
$$\rho_2 = \frac{-\theta_2}{1 + \theta_1^2 + \theta_2^2}$$
$$\rho_k = 0 \quad \forall k > 2$$

La FAP de un MA(2) (como la de todos los MA) es decreciente y la FAS presenta dos palos (al contrario que un AR(2)). Por ejemplo,



ACF



PACF

Finalmente, un MA(q)

$$X_t = a + \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - ... - \theta_q\varepsilon_{t-q}$$

o de forma alternativa:

$$X_t = a + (1 - \theta_1 L - \theta_2 L^2 - ... - \theta_q L^q)\varepsilon_t$$

siendo *a* una constante, θ_1, θ_2, ..., θ_q parámetros desconocidos, y ε_t~iid $N(0,\sigma^2_\varepsilon)$, donde σ^2_ε es una constante.

Un MA(q) tiene una media y una varianza constante y funciones de autocovarianza que, al contrario que un AR(p), implican una FAP decreciente, mientras que la FAS presenta q *palos*.

Hemos dejado al lector interesado la derivación de los momentos del MA(2), así como los del modelo mixto ARMA(p,q) que veremos a continuación, para lo cual puede proceder tal y como acabamos de ver.

Modelos mixtos ARMA(p,q)

Los modelos mixtos presentan bien una FAS y una FAP decrecientes (el caso más habitual) bien una FAS y una FAP con unos pocos palos.

En cuanto a su expresión analítica, mostramos algunos ejemplos:

ARMA(1,1)

$$X_t = a + \phi_1 X_{t-1} + \varepsilon_t - \theta_1\varepsilon_{t-1} \qquad (1 - \phi_1 L)X_t = (1 - \theta_1 L)\varepsilon_t$$

ARMA(2,1)

$$X_t = a + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t - \theta_1\varepsilon_{t-1} \qquad (1 - \phi_1 L - \phi_2 L^2)X_t = (1 - \theta_1 L)\varepsilon_t$$

ARMA(p,q)

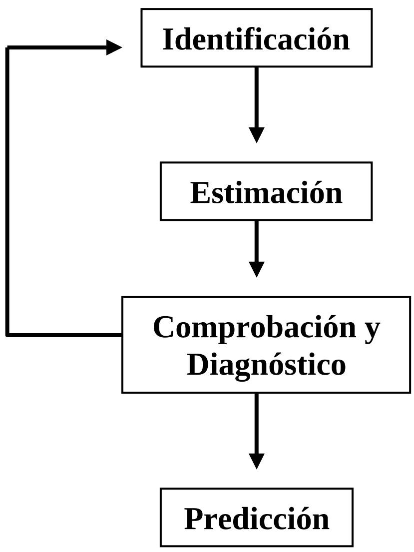
$$X_t = a + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q}$$

$$(1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p) X_t = a + (1 - \theta_1 L - \theta_2 L^2 - \ldots - \theta_q L^q) \varepsilon_t$$

$$\Phi(L) X_t = a + \Theta(L) \varepsilon_t$$

8.5.- Metodología Box-Jenkins.

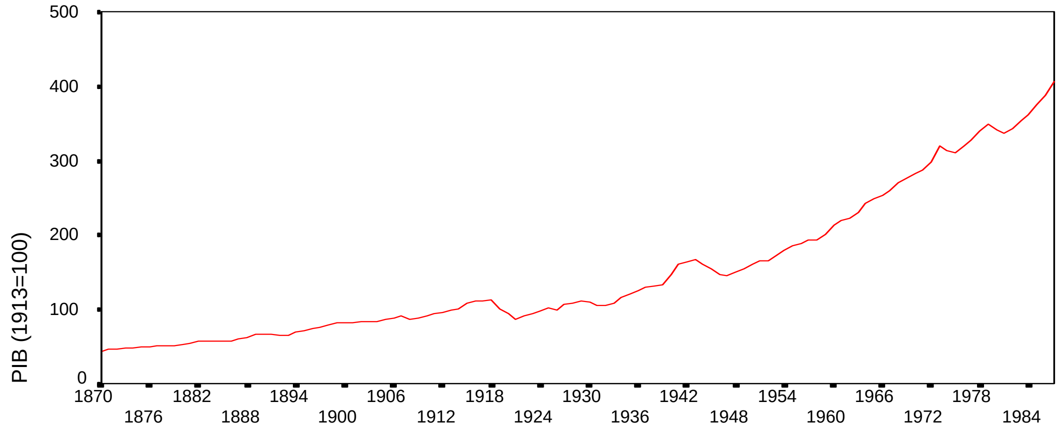
La utilización práctica de los modelos lineales de series temporales se desarrolló especialmente a partir de la propuesta metodológica de **Box y Jenkins** (**1977**). Estos autores propusieron un mecanismo circular de tratamiento de este tipo de modelos con objeto de conocer, de modo relativamente sencillo, cuál de entre los posibles elementos de aquella familia de modelos ejerce una mejor representación de la serie estudiada.



En la etapa de *identificación* se trata de especificar algún modelo lineal (AR, MA y/o mixto ARMA) para la serie temporal que aproxime el proceso estocástico que presumiblemente dio lugar a los datos disponibles.

Dentro de esta etapa, debemos primero proceder al análisis de la estacionariedad de la serie, es decir comprobar que ésta tenga una media y una varianza constantes en el tiempo, por cuanto los modelos lineales se definieron únicamente sobre procesos estocásticos estacionarios.

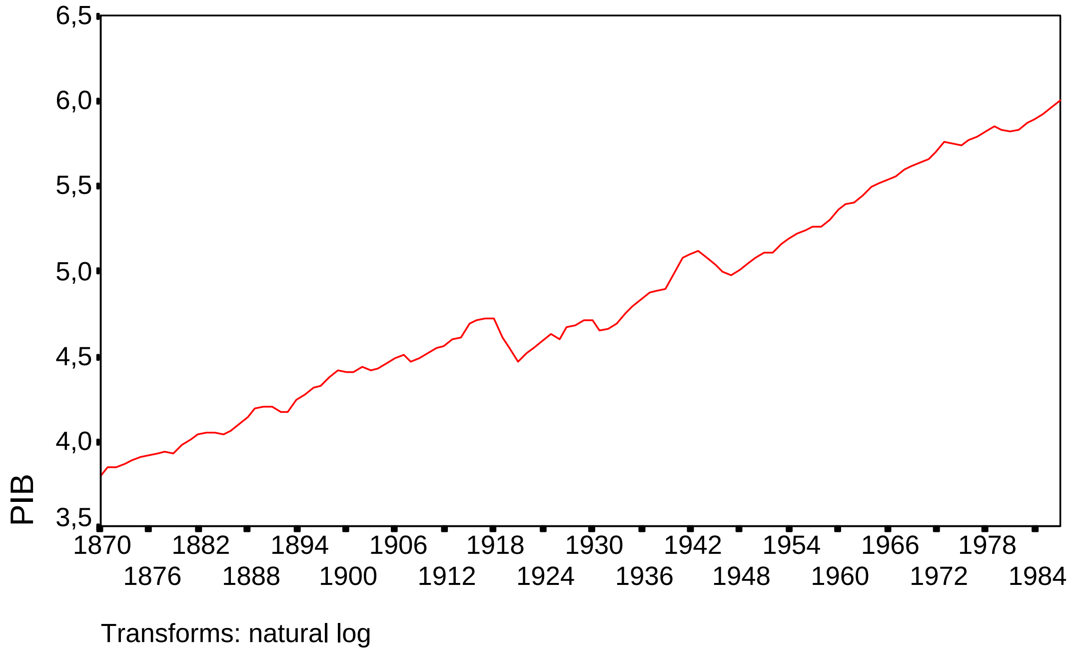
Aplicaremos la metodología Box-Jenkins a distintas series temporales. Sugerimos que el lector intente analizar el resto de series contenidas en los ficheros CASO5.SAV. Mostramos de nuevo la serie temporal del PIB de Gran Bretaña (puede encontrarse en el fichero CASO5A.SAV).



Como vemos esta serie no presenta una conducta estacionaria en media puesto que la serie crece. Nótese que gráficamente son difíciles de apreciar posibles fluctuaciones en la varianza, lo que evidenciaría no estacionariedad en la misma. El siguiente descriptivo muestra sin embargo que tanto la media como la varianza del PIB fueron crecientes en el tiempo.

```
Summaries of      PIB   By levels of      DECADAS

                                                  Mean     Std Dev    Variance
For Entire Population                         150,4153     96,8845   9386,5996

DECADAS              1870-1880                 49,0600      2,2936      5,2604
DECADAS              1881-1890                 58,7100      3,7943     14,3966
DECADAS              1891-1900                 71,9300      6,2115     38,5823
DECADAS              1901-1910                 85,6556      3,1405      9,8628
DECADAS              1911-1920                102,0100      8,9384     79,8943
DECADAS              1921-1930                 98,6000      6,8238     46,5644
DECADAS              1931-1940                117,1300     10,0615    101,2334
DECADAS              1941-1950                153,4100     10,4871    109,9788
DECADAS              1951-1960                176,6200     13,8615    192,1418
DECADAS              1961-1970                236,6600     22,4691    504,8582
DECADAS              1971-1980                307,7700     20,7272    429,6157
DECADAS              1981-1987                362,1222     23,9780    574,9444

Total Cases = 118
```

Cuando una serie muestra una conducta no estacionaria debe ser transformada adecuadamente. Si la varianza de la serie crece (como en nuestro caso) o decrece de forma más o menos continuada, ésta puede estabilizarse tomando logaritmos neperianos.

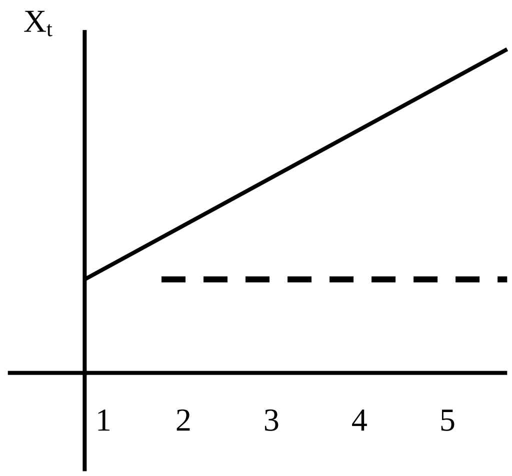


Se observará como la serie se ha hecho más estacionaria en varianza, reduciéndose algo la dispersión en torno a la tendencia creciente (sobre todo a partir de la Segunda Guerra Mundial).

En el caso de una media no estacionaria, la transformación es algo más complicada. Una serie que no sea estacionaria en media debe *diferenciarse*, es decir aplicar a la misma el operador diferencias, $\Delta=(1-L)$, así $\Delta X_t=(1-L)X_t= X_t - X_{t-1}$.

Mostraremos un sencillo ejemplo utilizando una serie no estacionaria en media:

t	$\mathbf{X_t}$	$\mathbf{X_{t-1}}$	$\mathbf{\Delta X_t = X_t - X_{t-1}}$
1	2	---	---
2	4	2	2
3	6	4	2
4	8	6	2
5	10	8	2



Al tomar diferencias hemos *eliminado* la tendencia de la serie, haciéndola estacionaria.

Teóricamente basta una diferencia cuando la serie presenta una tendencia lineal, como en nuestro sencillo ejemplo. Comportamientos no lineales requerirán diferencias de orden superior. Por ejemplo *dos* diferencias (ó diferencia de la diferencia, $\Delta(\Delta X_t)= \Delta^2 X_t$) es decir $\Delta^2 X_t=(1-L)^2 X_t$. Nótese que operando resulta una expresión muy complicada $\Delta^2 X_t=(1-L)^2 X_t = X_t+X_{t-2}-2X_{t-1}$.

En la práctica podemos utilizar el siguiente descriptivo:

```
Variable        Mean     Std Dev    Variance

Ln(PIB)           4,82         ,62         ,38
ΔLn(PIB)           ,02         ,03         ,00
Δ²Ln(PIB)          ,00         ,04         ,00
```

El orden de diferenciación adecuado sería el que implicase una varianza menor, una diferencia del logaritmo en nuestro caso. En la realidad, sin embargo, no siempre es así, por lo que este instrumento debe considerarse únicamente como auxiliar.

Proponemos, antes de intentar identificar ningún modelo lineal, escoger el par de *correlogramas*, FAS y FAP, más *limpio*, es decir el que presente una estructura menos cargada y, por tanto, más sencilla de cara a una posterior identificación del modelo.

Dibujamos en todos casos las *bandas de confianza*, dentro de las cuales los coeficientes de autocorrelación simple o parcial no son estadísticamente significativos al 95%. Dichas bandas se construyen utilizando las siguientes expresiones:

FAS

$$Bandas = \pm 1{,}96\sqrt{Var(\hat{\rho}_k)}$$

donde la varianza de los coeficientes de autocorrelación (aproximación de Barlett) es:

$$Var(\hat{\rho}_k) = \frac{1}{T}\left(1 + 2\sum_{i=1}^{K} \hat{\rho}_k{}^2\right)$$

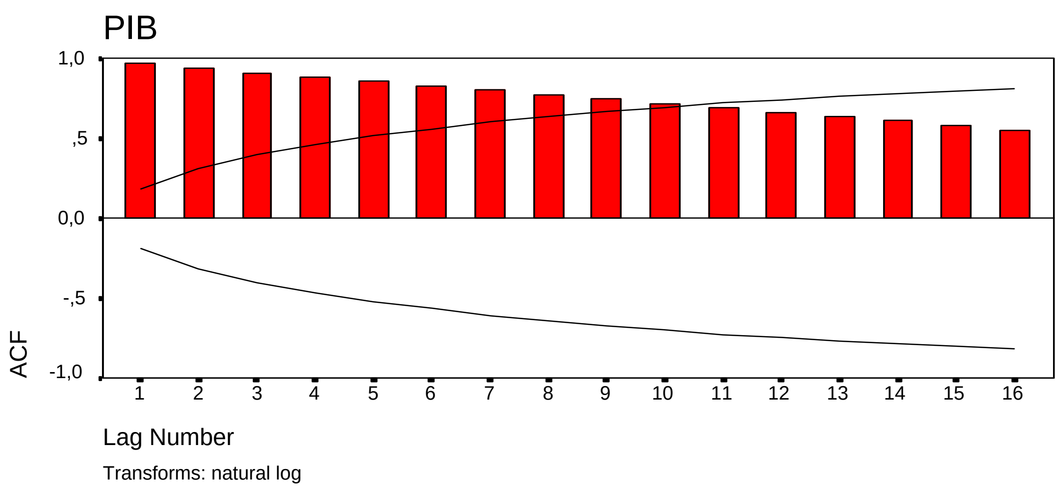
Por lo que las bandas presentan una forma de campana:

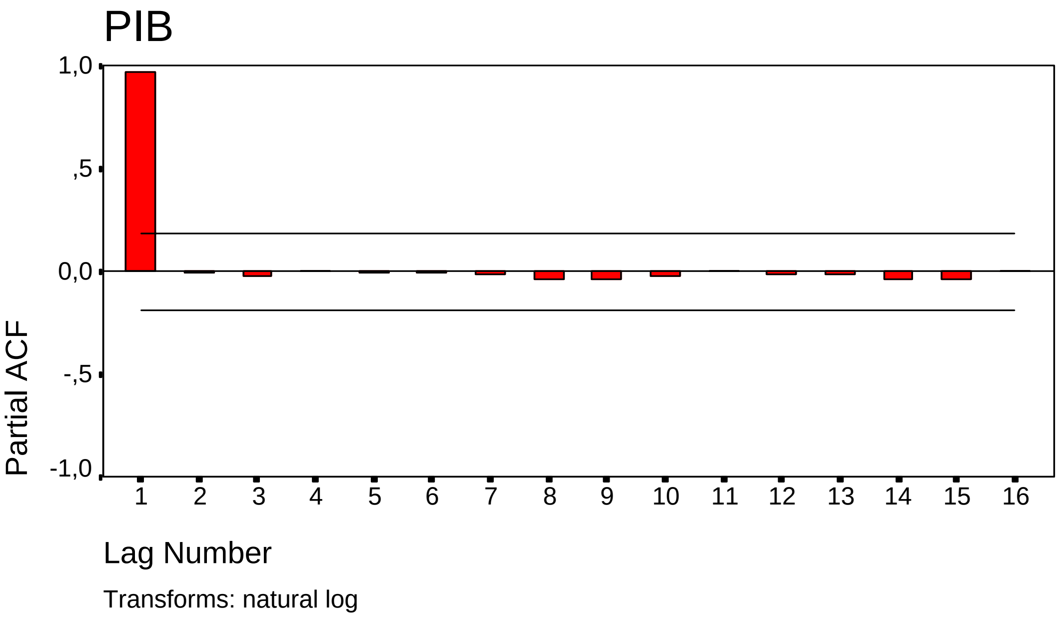
FAP

$$Bandas = \pm 1{,}96\frac{1}{\sqrt{T}}$$

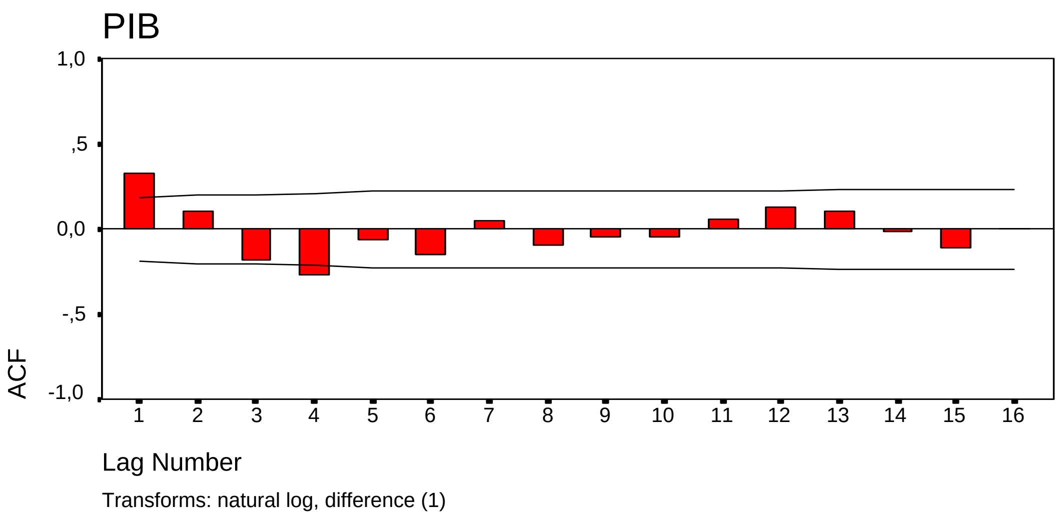
Las bandas de confianza de la FAP son líneas rectas.

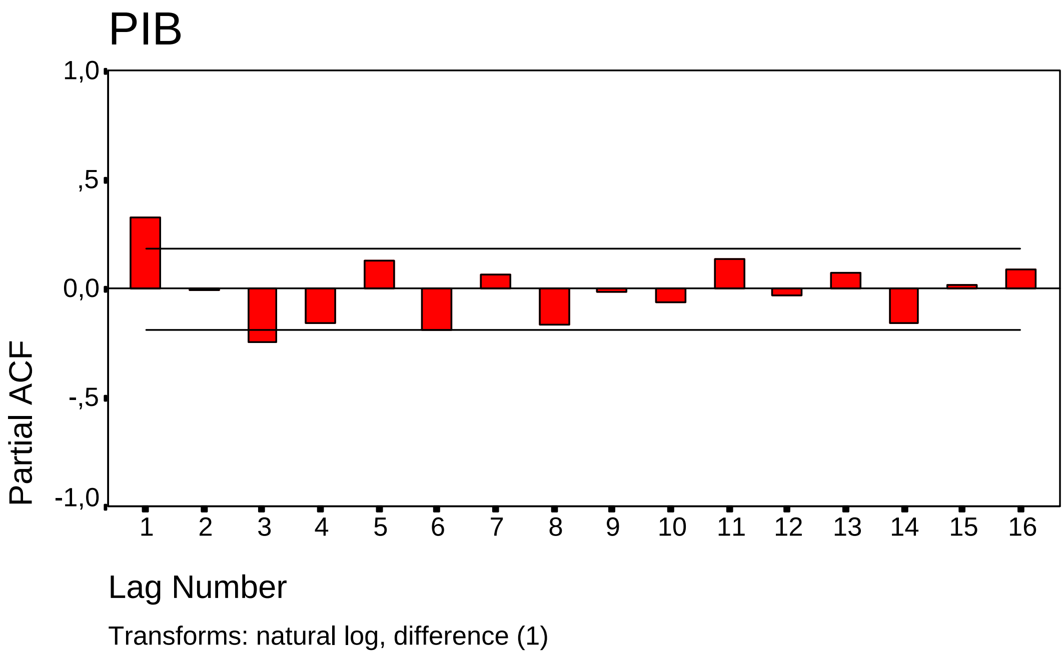
La FAS del logaritmo de la serie sin diferenciar está muy *cargada*, como la de un camino aleatorio..



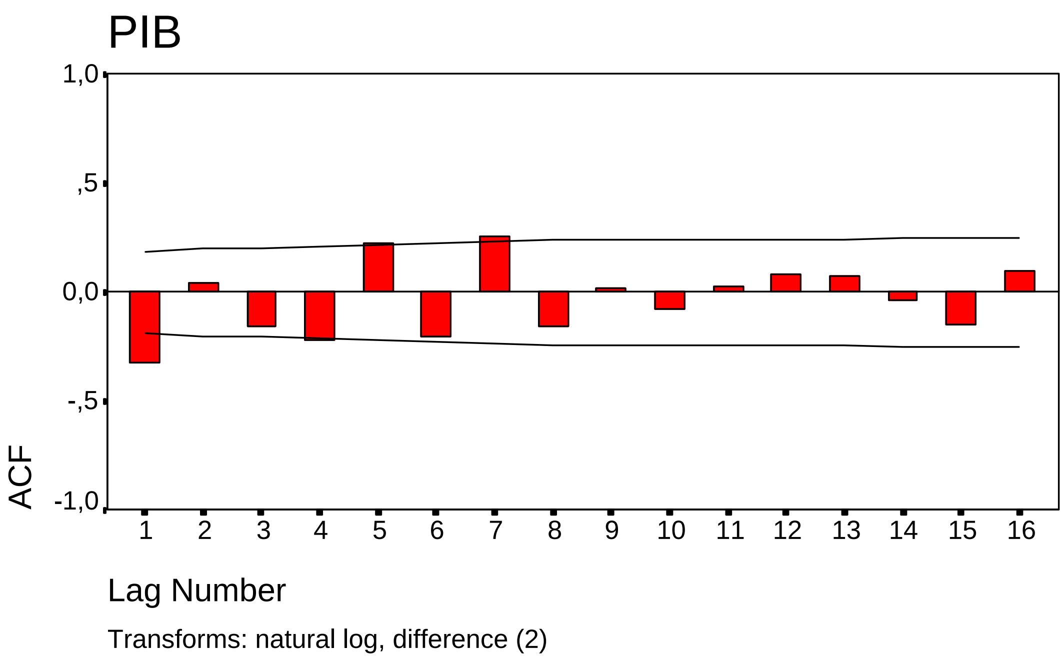


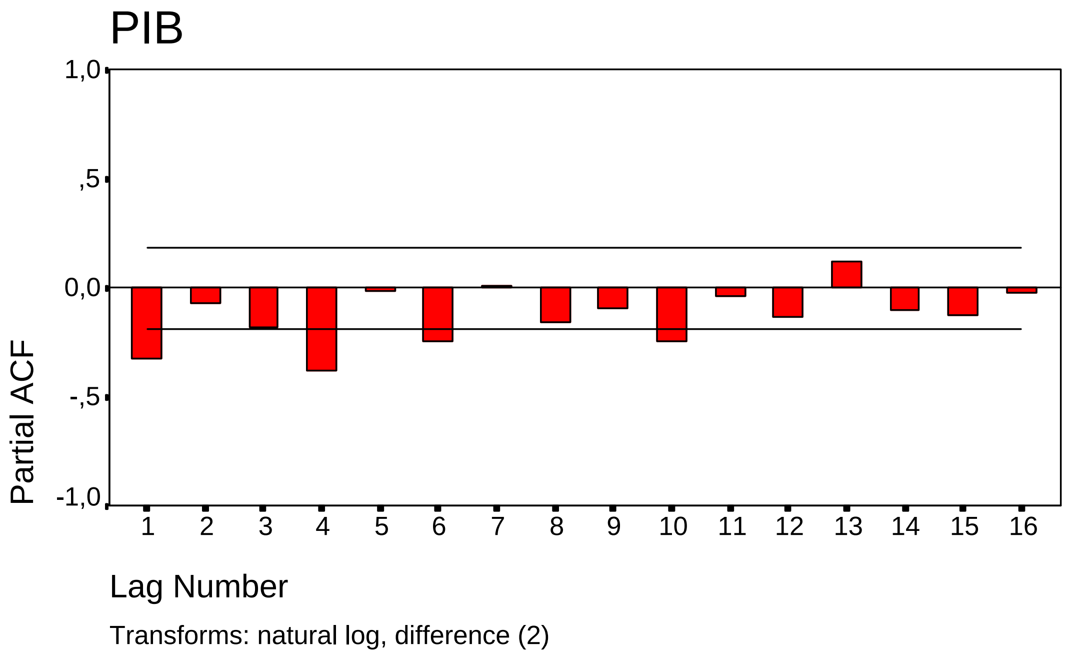
Resumiendo, los correlogramas del logaritmo de la serie sin diferenciar son los típicos de una serie no estacionaria en media, presentando una estructura muy cargada en la FAS y un palo muy próximo a la unidad en la FAP.





Los correlogramas de la serie diferenciada parecen ser bastante *limpios*.





Los correlogramas correspondientes a dos diferencias, particularmente la FAP, son menos limpios que los de una diferencia. Haciendo uso del *principio de parsimonia*, dominante en la metodología Box-Jenkins, preferimos los correlogramas correspondientes a una diferencia de la serie en logaritmos. Nótese que en este caso el orden de diferenciación más adecuado parece coincidir tanto en los descriptivos (varianza menor) como en los correlogramas (los más limpios).

Por el momento, no haremos caso de los retardos 3 (FAP) ni 4 (FAS). Ni en la FAS ni en la FAP se observa una estructura decreciente. Recordar que si la FAS mostrase una estructura decreciente se trataría de un AR de orden igual al número de coeficientes significativos en la FAP. En el caso de MA sería justo a la inversa. No nos encontramos en ninguna de las dos situaciones por lo que sugerimos, como primera aproximación, un modelo mixto, ARIMA(1,1,1). La letra I (correspondiente a la *d* en ARIMA(p,d,q)), indica que la serie ha debido ser diferenciada, una diferencia en nuestro caso (d=1).

Una vez hemos efectuado una apuesta tentativa por uno, o varios modelos iniciales, la siguiente etapa de la metodología Box-Jenkins es la de *estimación* de los parámetros del modelo.

En nuestro caso el problema que se plantea es el de estimar los parámetros ϕ (correspondiente al AR(1)) y θ (correspondiente al MA(1)) a partir del conjunto de 117 observaciones (118-1, observación que se pierde tras diferenciar). Si se admite que las ε_t se distribuyen normal e independientemente con media cero y varianza constante (como vimos) se puede obtener la función de verosimilitud condicional asociada a los parámetros ϕ,θ, y σ_ε^2. El logaritmo de esta función es en nuestro caso igual a:

$$L\left(\phi,\theta,\sigma_\varepsilon{}^2\right) = -T\ln\sigma_\varepsilon{}^2 - \frac{\sum_{t=1}^{T-d}\varepsilon_t 2}{2\sigma_\varepsilon{}^2} = -T\ln\sigma_\varepsilon{}^2 - \frac{\sum_{t=1}^{T-d}\left(\Delta PIB_t - \phi\Delta PIB_{t-1} + \theta\varepsilon_{t-1}\right)^2}{2\sigma_\varepsilon{}^2}$$

Obsérvese que existe un componente media móvil, componente que es inobservable, por lo que para maximizar esta expresión (o alternativamente minimizar el segundo término de la parte derecha de la igualdad) deberíamos realizar sustituciones sucesivas. Alternativamente podemos recurrir a métodos de estimación no lineales (véanse **Box y Jenkins** (**1977**) para profundizar en los problemas de estimación en modelos de series temporales y **Greene** (**1993**) para una introducción a métodos de estimación no lineales).

Los resultados de la estimación se muestran a continuación:

```
FINAL PARAMETERS:

Number of residuals  117
Standard error       ,03029419
Log likelihood       244,55253
AIC                  -483,10506
SBC                  -474,81854

            Analysis of Variance:

               DF  Adj. Sum of Squares    Residual Variance

Residuals     114           ,10472709           ,00091774
```

```
          Variables in the Model:

                    B          SEB      T-RATIO   APPROX. PROB.

AR1          ,32718395   ,26677943    1,2264212       ,22256717
MA1         -,00627004   ,28232042    -,0222089       ,98232025
CONSTANT     ,01915796   ,00417123    4,5928821       ,00001143


Correlation Matrix:

                AR1         MA1

AR1      1,0000000    ,9431552
MA1       ,9431552   1,0000000
```

La penúltima etapa de la metodología Box-Jenkins es la de *comprobación y diagnóstico* del modelo. Siguiendo un esquema de trabajo similar al que se efectúa en la inferencia en un modelo de regresión lineal, los modelos de series temporales se someten también a pruebas de validación de su capacidad de ajuste y predicción. El modelo quedará *validado*, o "mejor adaptado" a los datos, cuando pase todas y cada una de las siguientes pruebas.

Significación estadística de los parámetros

Cada uno de los parámetros autorregresivo o media móvil incluidos en el modelo debe ser estadísticamente significativo.

Como vemos, en nuestro caso ninguno de los dos parámetros es estadísticamente significativo, por lo que no podemos aceptar el modelo. Si nos fijamos más atentamente, sin embargo, uno de los parámetros parece redundante, puesto que la correlación entre los mismos es muy alta (,9431). Tras observar de nuevo la significación de los parámetros y puesto que el parámetro MA tiene una t de Student menor que la unidad proponemos un ARIMA(1,1,0).

```
FINAL PARAMETERS:

Number of residuals  117
Standard error       ,0301623
Log likelihood       244,56237
AIC                  -485,12474
SBC                  -479,60039

            Analysis of Variance:

               DF  Adj. Sum of Squares    Residual Variance

Residuals     115           ,10472813           ,00090976

           Variables in the Model:

                    B         SEB     T-RATIO   APPROX. PROB.

AR1         ,33307769   ,08809434   3,7809205       ,00024970
CONSTANT    ,01915985   ,00416342   4,6019486       ,00001094
```

El parámetro AR es estadísticamente significativo, por lo que el modelo pasa la primera prueba. La constante es un parámetro secundario en un modelo ARIMA, no siendo necesario que sea significativa (aunque lo es en este caso).

Cumplimiento de las condiciones de estacionariedad e invertibilidad

Los parámetros AR y MA del modelo no deben ser próximos a la unidad (o la suma de los mismos dentro de un AR o un MA en modelos de ordenes superior a uno), en caso contrario evidenciaría una sub o una sobre-diferenciación.

Si el parámetro AR fuese próximo a la unidad ($\phi \approx 1$) el modelo estaría sub-diferenciado,

$$(1-\phi L)X_t = a + \varepsilon_t \qquad (1-L)X_t = \Delta X_t = a + \varepsilon_t$$

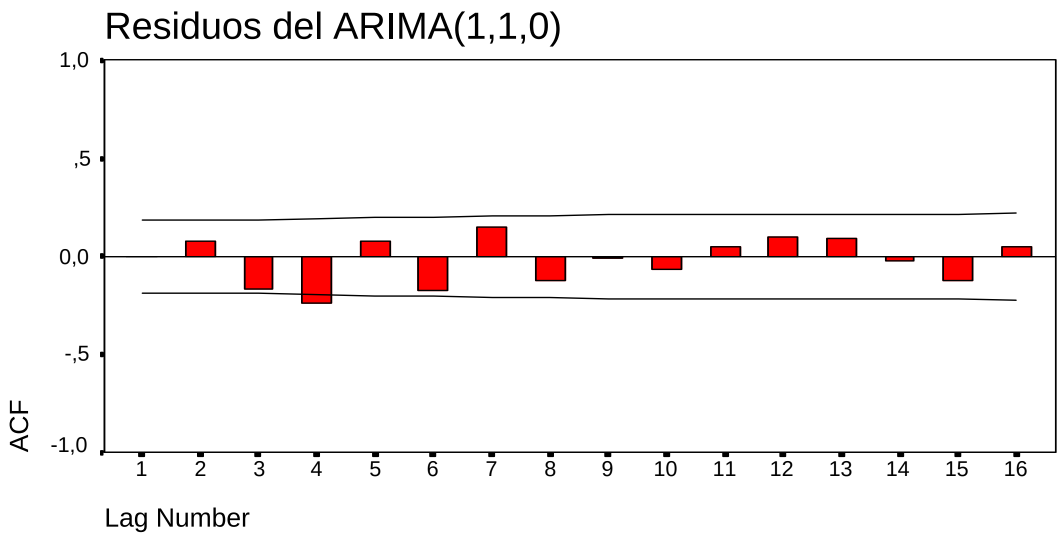
El modelo estaría sobre-diferenciado si fuese el MA el próximo a la unidad ($\theta \approx 1$) :

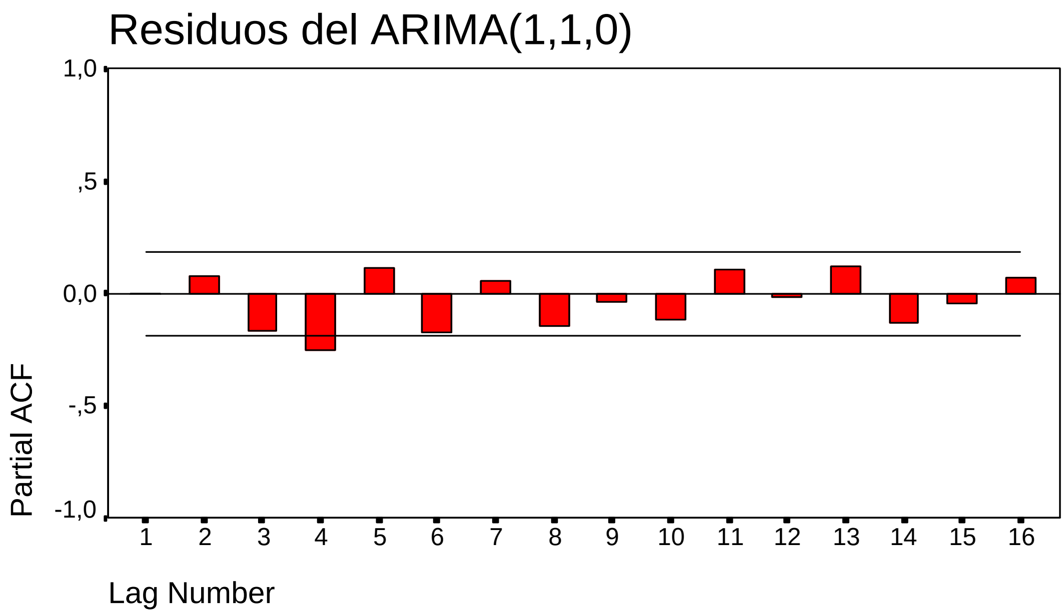
$$\Delta X_t = a + (1-\theta)\varepsilon_t \qquad (1-L)X_t = a + (1-L)\varepsilon_t \qquad \frac{1-L}{1-L} X_t = X_t = a + \varepsilon_t$$

En nuestro caso el parámetro AR está muy alejado de la unidad, por lo que el modelo pasa la segunda prueba.

Residuos del modelo: ruido blanco

Las funciones de correlación simple y parcial de los residuos del modelo no deben presentar valores significativos.





Excepto el coeficiente de autocorrelación de orden 4 (tanto en la FAS como en la FAP) los residuos del modelo se corresponden con un ruido blanco.

En la práctica suele ocurrir que algún coeficiente de autocorrelación en los correlogramas de los residuos sea significativo. Si tal coeficiente no se corresponde con retardos importantes (en series sin estacionalidad los tres primeros) el modelo puede considerarse ‘bueno’. Tales coeficientes, sin embargo, pueden ser controlados a fin de mejorar el ajuste, introduciéndolos como medias móviles en el modelo ARIMA. Por ejemplo,

ARIMA(1,1,0) original

$$(1-\phi L)\Delta Ln(PIB)_t = a + \varepsilon_t$$

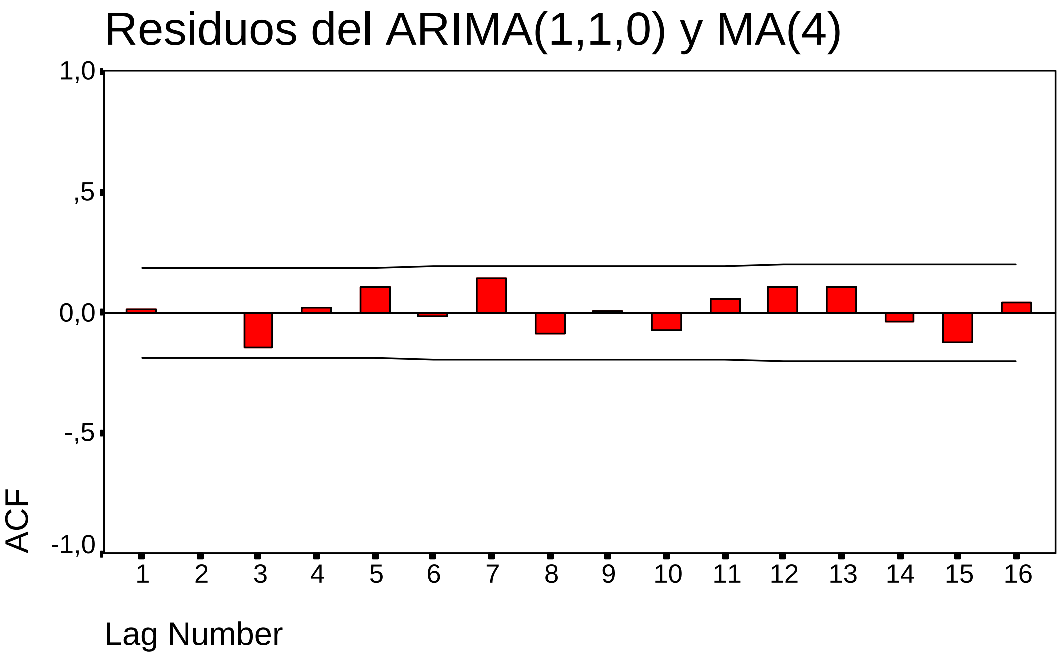
ARIMA(1,1,0) controlando el retardo 4

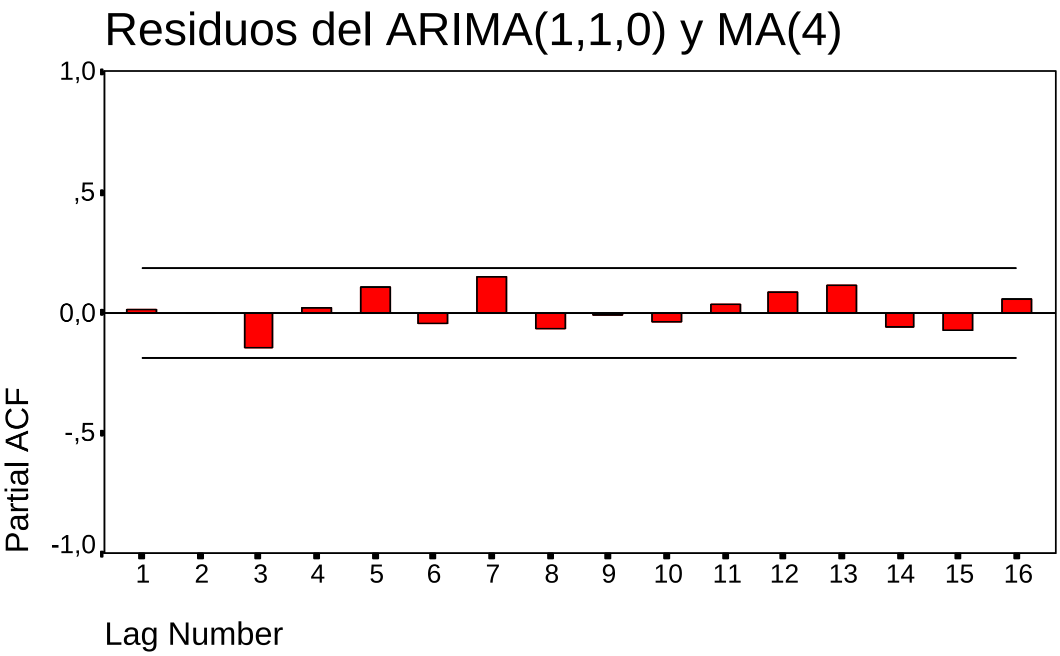
$$(1-\phi L)\Delta Ln(PIB)_t = a + \left(1-\theta_4 L^4\right)\varepsilon_t$$

En nuestro caso:

```
FINAL PARAMETERS:

Number of residuals  117
Standard error       ,0290504
Log likelihood       249,26935
AIC                  -492,53871
SBC                  -484,25219

            Analysis of Variance:

               DF  Adj. Sum of Squares    Residual Variance

Residuals     114           ,09662179           ,00084393

           Variables in the Model:

                    B         SEB     T-RATIO   APPROX. PROB.

AR1          ,32080363   ,08907044   3,6016844       ,00046990
MA4          ,30872867   ,09038730   3,4156200       ,00088266
CONSTANT     ,01892819   ,00276469   6,8464138       ,00000000
```

El modelo pasa las dos primeras pruebas (parámetros significativos y no próximos a la unidad) y, como vemos, también la tercera (residuos ruido blanco).

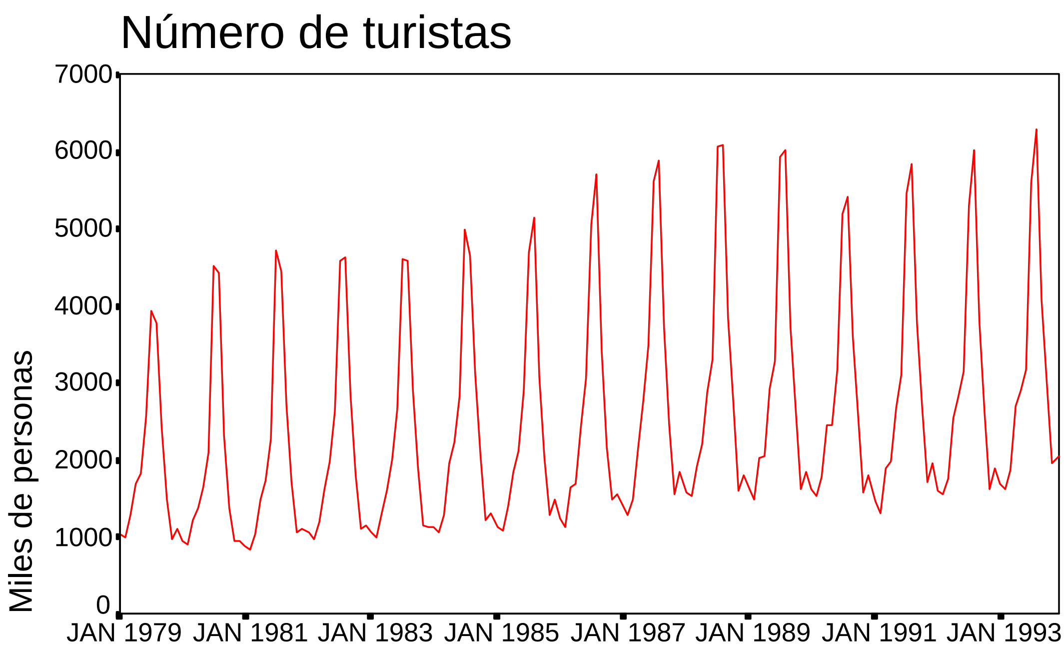

Residuos del ARIMA(1,1,0) y MA(4)
ACF
1,0
,5
0,0
-,5
-1,0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
Lag Number


Residuos del ARIMA(1,1,0) y MA(4)
Partial ACF
1,0
,5
0,0
-,5
-1,0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
Lag Number

Nótese que el error estándar del modelo final (ARIMA(1,1,0) con MA(4)) es menor que el del modelo original (ARIMA(1,1,0)), 0,02905 vs 0,03016. Como vimos, si la variable dependiente está en logaritmos el error estándar puede interpretarse como el error porcentual promedio cometido en el ajuste, un 2,91% en nuestro caso. Además, tanto el AIC (criterio de información de Akaike) como el SBC (criterio de información Bayesiano de Schwartz) son menores en el modelo final, lo que sugiere un mejor ajuste en este caso. El modelo final puede ser escrito como sigue:

$$(1-0{,}3208L)\Delta Ln(PIB)_t = 0{,}0189 + (1-0{,}3087L^4)\varepsilon_t$$

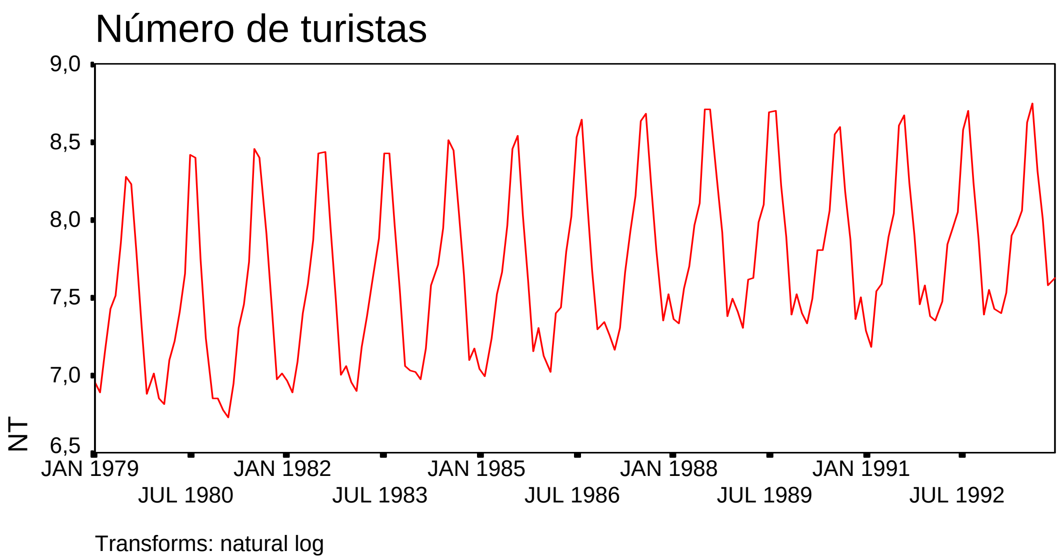
Analicemos ahora la serie del número de turistas, NT (puede encontrarse en el fichero CASO5D.SAV).



Hasta ahora hemos tratado con los modelos de series temporales en los que se han buscado pautas de comportamiento a través de las correlaciones simples y parciales como mecanismos para señalar la existencia de dependencias temporales *regulares*. Cuando las observaciones que dan lugar a la serie temporal son de frecuencia inferior al año, como en el caso de la serie mensual NT, suelen aparecer también relaciones entre los valores que se corresponden con periodos análogos (de mes de Agosto a mes de Agosto de cada año en una serie mensual, de viernes a viernes en otra semanal, etc). La necesidad de representar estas correlaciones estacionales da pie a los denominados modelos estacionales multiplicativos, o SARIMA.

En nuestro caso la serie NT presenta un claro comportamiento estacional que se repite periódicamente aproximadamente cada año, coincidiendo lógicamente con el periodo de mayor entrada de turistas (verano). El *periodo estacional* de la serie (denominado *s*) es igual a 12 en este caso (número de observaciones entre los picos estacionales). En general las series mensuales presentan estacionalidad de periodo 12, las series trimestrales de periodo 4, las series semestrales de periodo 2, las semanales de periodo 52, etc.

Parece claro que la serie no es estacionaria ni en media ni en varianza (véase también el descriptivo). Aplicando logaritmos se reduce bastante la dispersión.

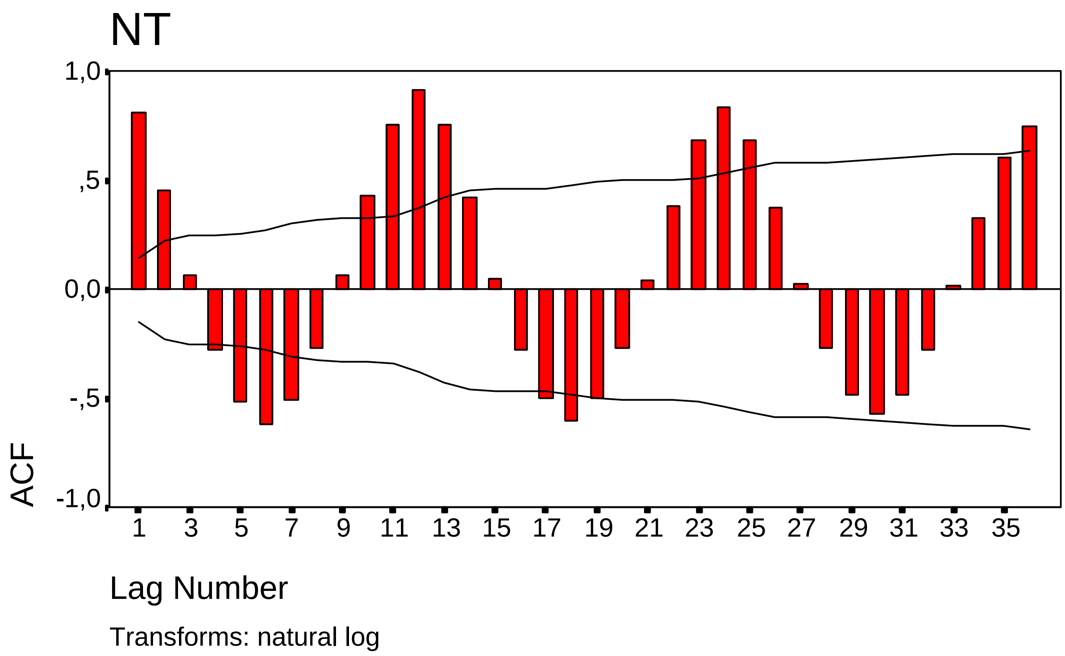


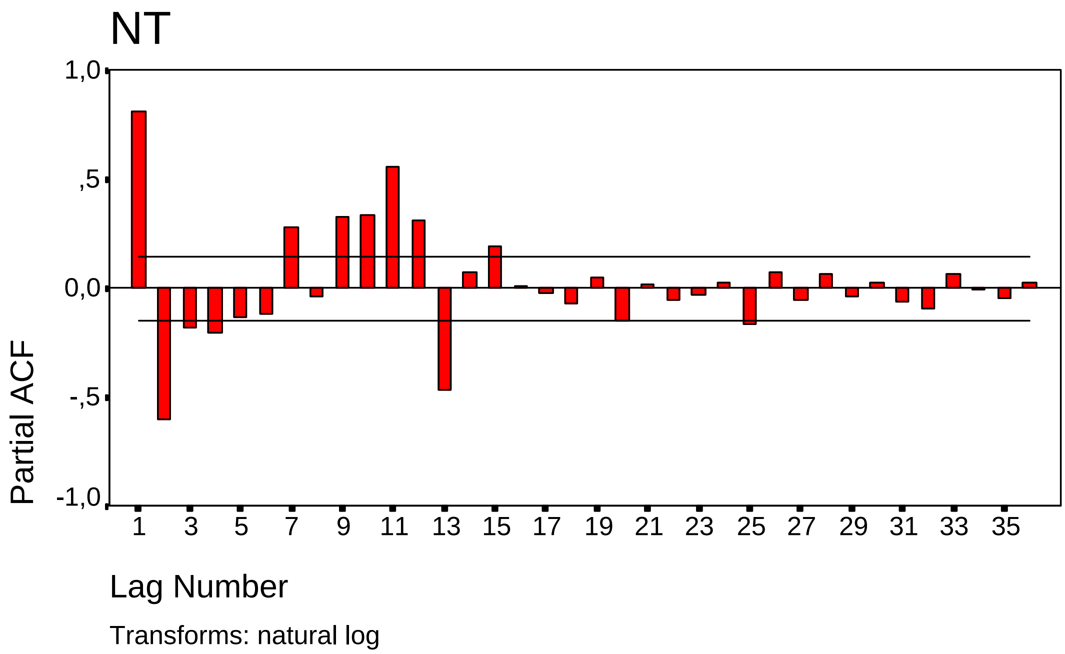
	Media	Varianza
TRIENIOS		
1979-1982		
NT	1944.64	1424362.1
1983-1985		
NT	2206.11	1639576.0
1986-1988		
NT	2570.66	2122963.0
1989-1991		
NT	2892.70	2207913.7
1992-1994		
NT	2932.70	2135644.3

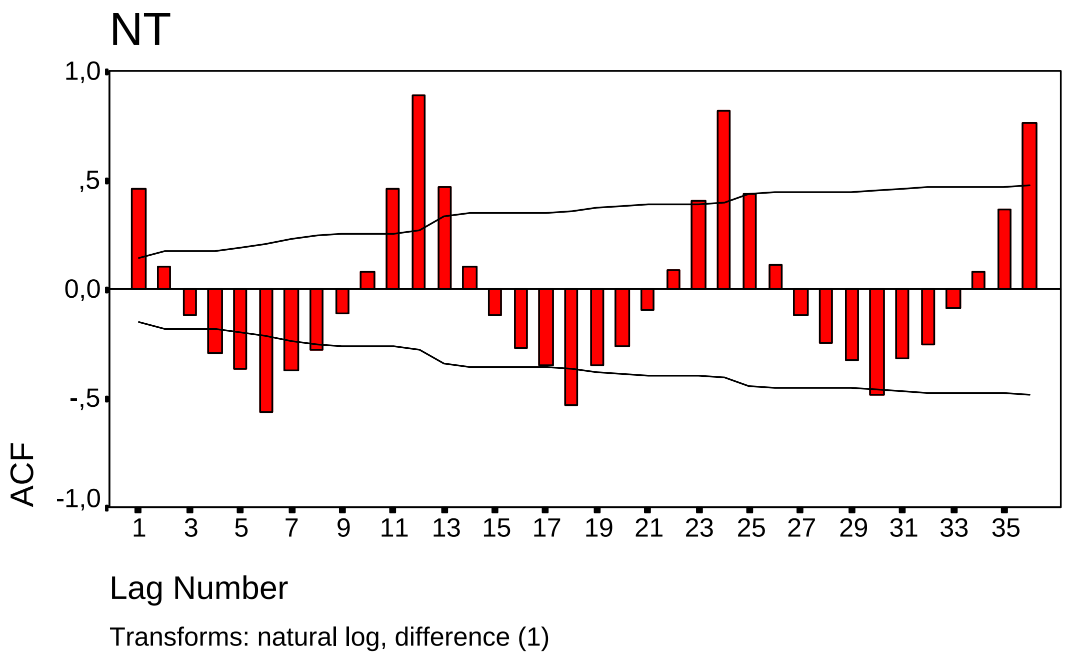
El problema que se presenta cuando analizamos series temporales con un posible componente estacional es que éste sea tan fuerte que implique una variación de la media (global) en el tiempo. En este caso se requerirán, D *diferencias estacionales*. Una diferencia estacional, por ejemplo (D=1), se obtiene aplicando a la serie el operador diferencias estacionales $\Delta_s X_t==(1-L^s)X_t$, así $\Delta_{12}X_t==(1-L^{12})X_t= X_t-X_{t-12}$ (nótese que no es lo mismo $\Delta_{12}X_t==(1-L^{12})X_t$ que $\Delta^{12}X_t==(1-L)^{12}X_t$).

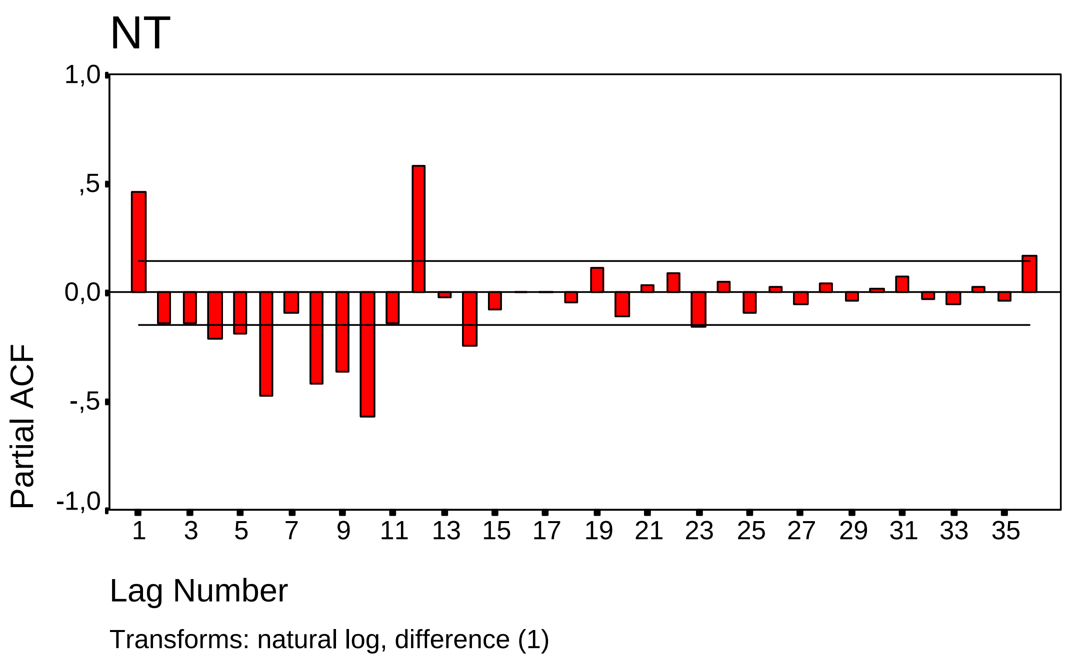
Variable	Mean	Std Dev	Variance
Ln(NT)	7,68	,53	,28
ΔLn(NT)	,00	,32	,10
Δ_sLn(NT)	,03	,09	,01
$\Delta\Delta_s$Ln(NT)	,00	,11	,01

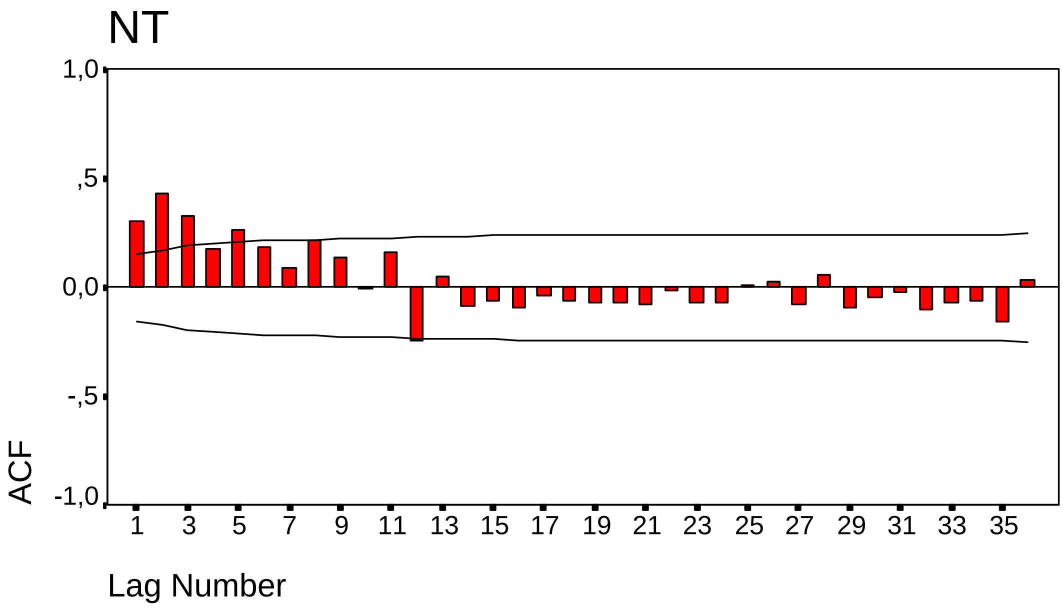
Este instrumento auxiliar nos sugiere que el mejor orden de diferenciación es el de una diferencia estacional de la serie en logaritmos.


NT
1,0
,5
0,0
-,5
-1,0
ACF
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35
Lag Number
Transforms: natural log

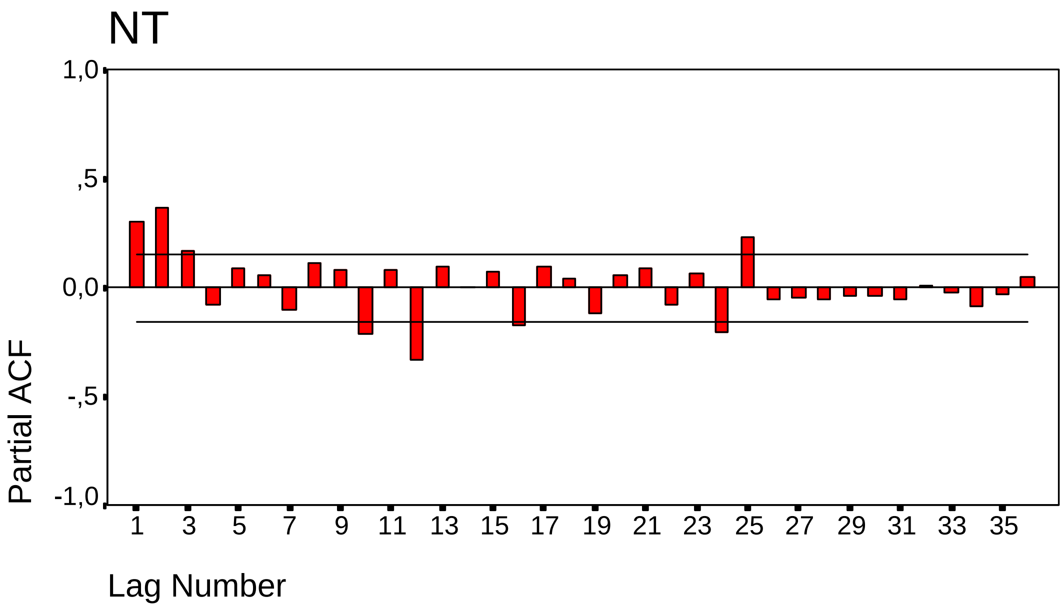

NT
1,0
,5
0,0
-,5
-1,0
Partial ACF
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35
Lag Number
Transforms: natural log


NT
1,0
,5
0,0
-,5
-1,0
ACF
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35
Lag Number
Transforms: natural log, difference (1)

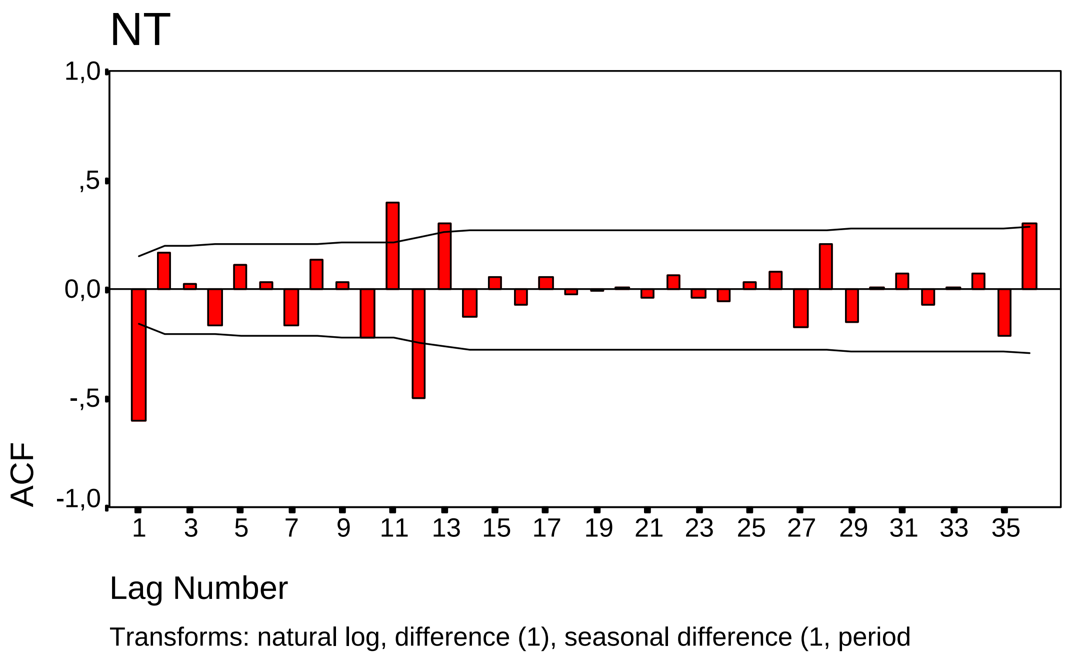

NT
1,0
,5
0,0
-,5
-1,0
Partial ACF
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35
Lag Number
Transforms: natural log, difference (1)

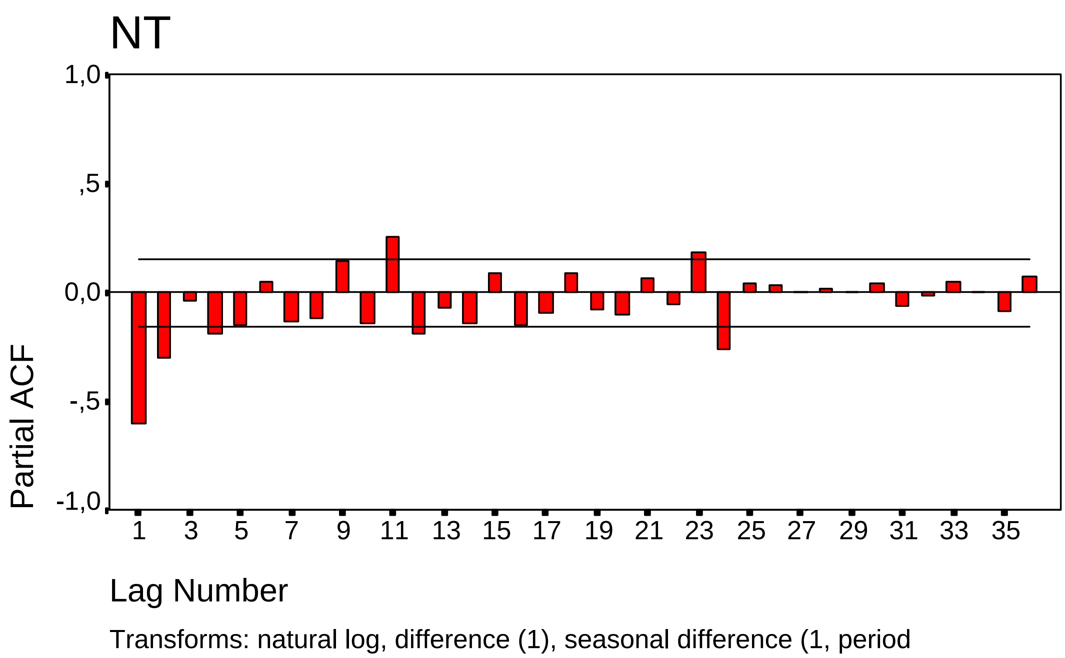


Transforms: natural log, seasonal difference (1, period 12)



Transforms: natural log, seasonal difference (1, period 12)





Nótese, sin embargo, que el par de correlogramas más limpios son estos dos últimos, los correspondientes a una diferencia regular y a una estacional de la serie en logaritmos. El modelo puede escribirse ARIMA(p,1,q)(P,1,Q)$_{12}$, correspondiendo el último término a la parte estacional de periodicidad 12.

Una vez determinado el mejor orden de diferenciación se debe proceder a identificar algún modelo lineal para la serie. En el caso de modelos estacionales se procede como si se dispusiese de dos series en una, la parte *regular* (correspondiendo los tres primeros retardos) y la parte *estacional* (correspondiéndole los retardos estacionales, en este caso el 12, el 24 y el 36; en series trimestrales el 4, 8, 12, 16, 18, etc).

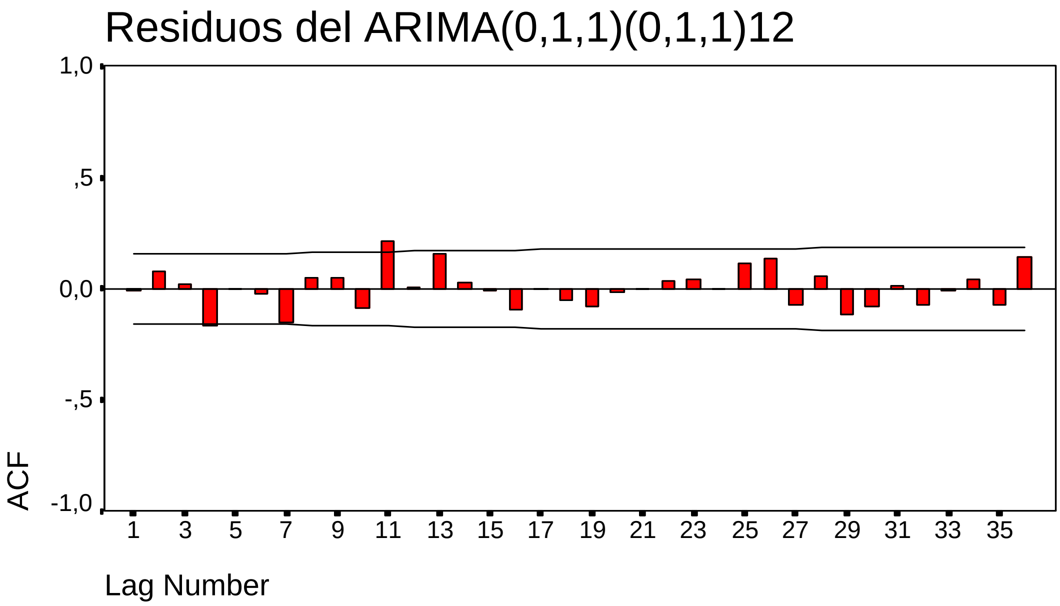
Volviendo a los dos últimos correlogramas y por lo que se refiere a la parte regular, vemos un único palo significativo en la FAS y una estructura decreciente en la FAP, por lo que proponemos un MA(1) para la misma, $ARIMA(0,1,1)(P,1,Q)_{12}$. Por lo que se refiere a la parte estacional (retardos 12, 24 y 36) vemos la misma estructura, un único palo significativo en la FAS (el 12) y una estructura decreciente en la FAP (12 y 24). Proponemos por tanto $ARIMA(0,1,1)(0,1,1)_{12}$ (este modelo se conoce como modelo *airline*).

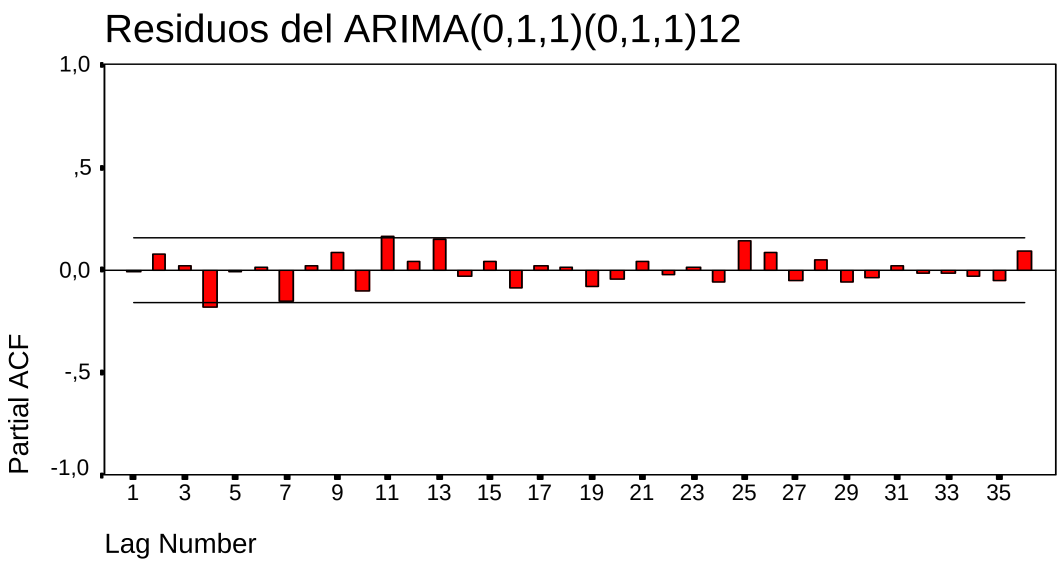
```
FINAL PARAMETERS:

Number of residuals  167
Standard error       ,06930079
Log likelihood       207,30757
AIC                  -408,61513
SBC                  -399,26115

            Analysis of Variance:

               DF  Adj. Sum of Squares    Residual Variance

Residuals     164            ,81646721           ,00480260

           Variables in the Model:

                     B          SEB     T-RATIO   APPROX. PROB.

MA1         ,67265575   ,05791637   11,614260       ,00000000
SMA1        ,60167289   ,07335746    8,201932       ,00000000
CONSTANT    ,00049733   ,00080278     ,619508       ,53644180
```

El modelo pasa las dos primeras pruebas, los parámetros son significativos (excepto la constante, pero es un parámetro secundario) y no son próximos a la unidad. Si se dudase de la proximidad a la unidad de alguno de ellos podría realizarse el siguiente contraste (por ejemplo con el MA(1) regular):

H_0: $\theta=1$ $|(0{,}6765-1)/0{,}0579|=5{,}652>1{,}96$ (por lo que se rechaza la hipótesis nula)

Los correlogramas de los residuos, con la salvedad de los coeficientes 4 y 11 (correspondientes a retardos no importantes, los tres primeros y los estacionales en series de periodicidad inferior al año) son aproximadamente los de un ruido blanco:





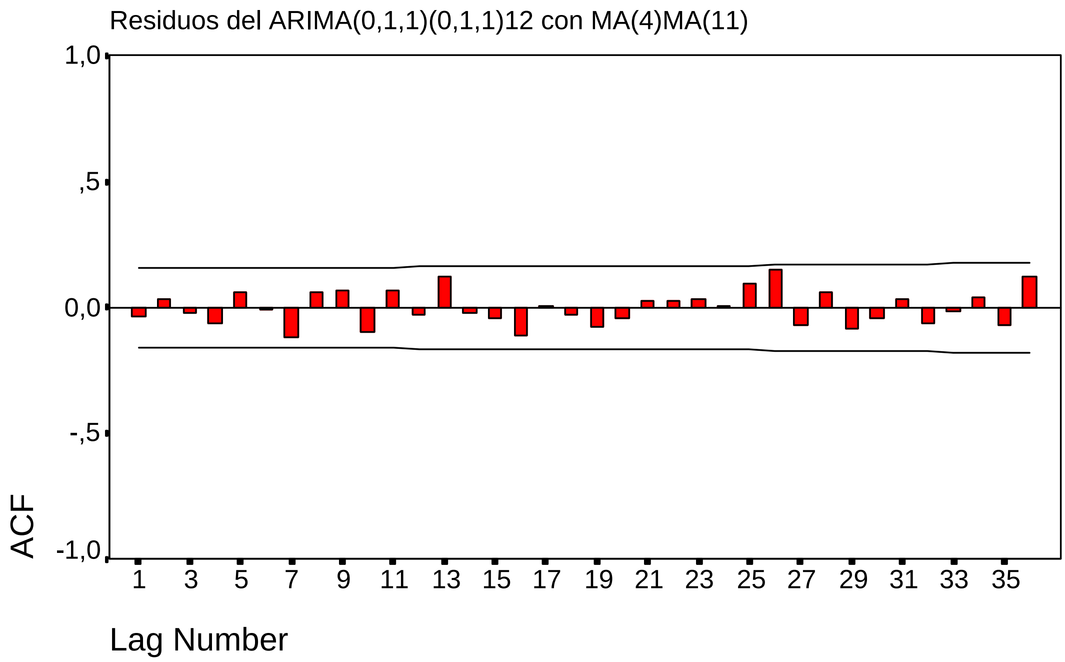
El modelo puede escribirse:

$$\Delta\Delta_{12}Ln(NT)_t=(1-0{,}6826L)(1-0{,}6017L^{12})\varepsilon_t$$

Observar en especial la parte estacional, un media móvil $(1-0{,}6017L^{12})$.

Como hicimos antes, intentaremos controlar las autocorrelaciones residuales con el objetivo de mejorar el ajuste, aún más si cabe:

```
FINAL PARAMETERS:

Number of residuals  167
Standard error       ,06755769
Log likelihood       211,7941
AIC                  -413,58819
SBC                  -397,99822

            Analysis of Variance:

               DF   Adj. Sum of Squares    Residual Variance

Residuals     162            ,77379059            ,00456404

           Variables in the Model:

                      B          SEB       T-RATIO   APPROX. PROB.

MA1          ,60693027   ,06191403    9,8027914       ,00000000
MA4          ,10923939   ,06107533    1,7886011       ,07554796
MA11        -,14871101   ,06361135   -2,3378061       ,02061976
SMA1         ,69457227   ,07483176    9,2817844       ,00000000
CONSTANT     ,00050327   ,00082975     ,6065269       ,54501390
```

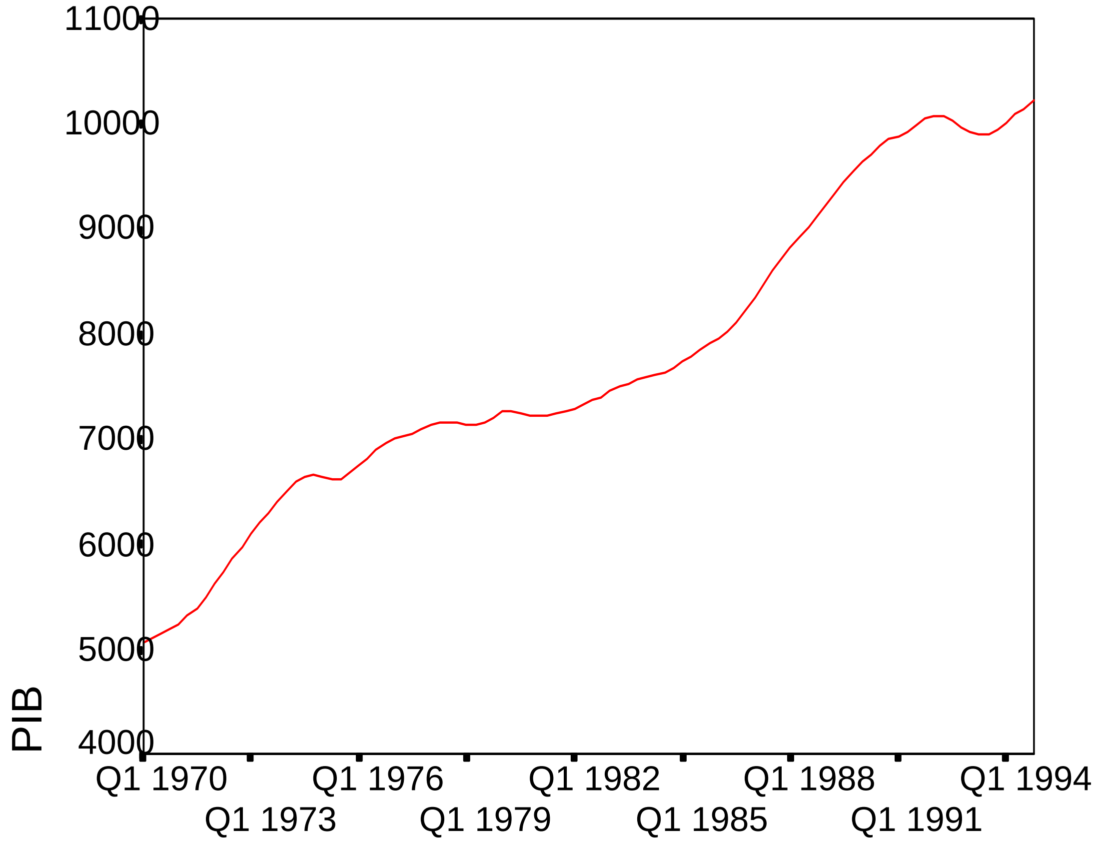


Como vemos el modelo pasa todas las pruebas, por lo que lo escribiremos:

$$\Delta\Delta_{12}Ln(NT)_t=(1-0,6069L)(1-0,1092L^4)(1+0,1487L^{11})(1-0,6946L^{12})\varepsilon_t$$

El error promedio de ajuste en este caso fue del 6,76%, frente a un 6,93% en el modelo previo, por lo que la bondad de ajuste apenas se ha modificado en este caso.

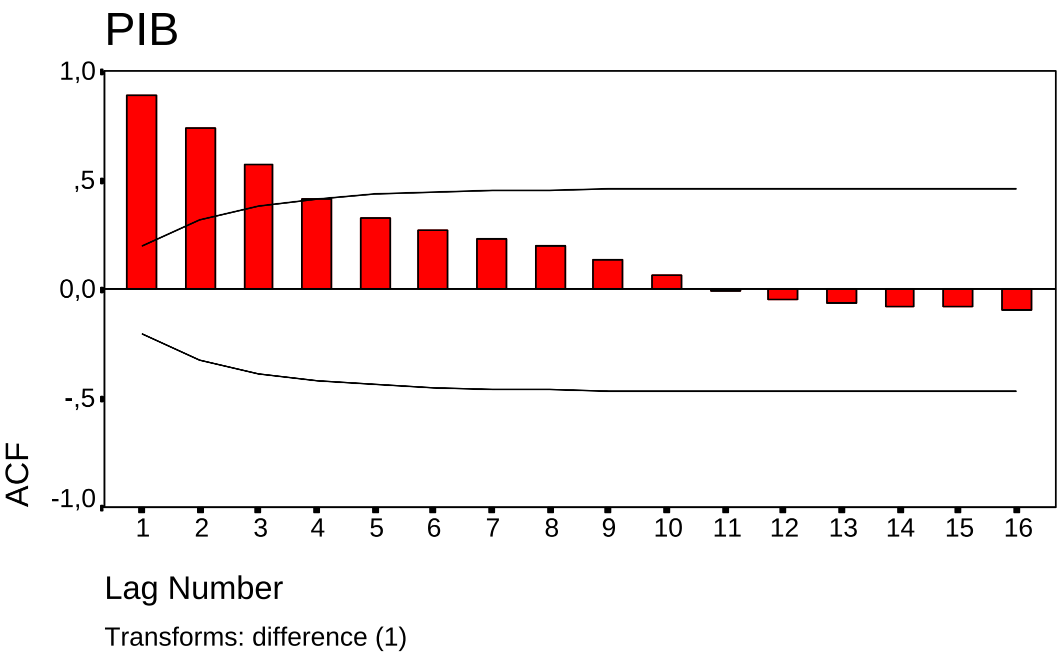
La serie trimestral del PIB español (CASO5C.SAV) presenta una tendencia creciente pero las fluctuaciones de la varianza no siguen una conducta sistemática, creciendo o decreciendo de forma no continua (véase el descriptivo), por lo que optamos por no tomar logaritmos neperianos.

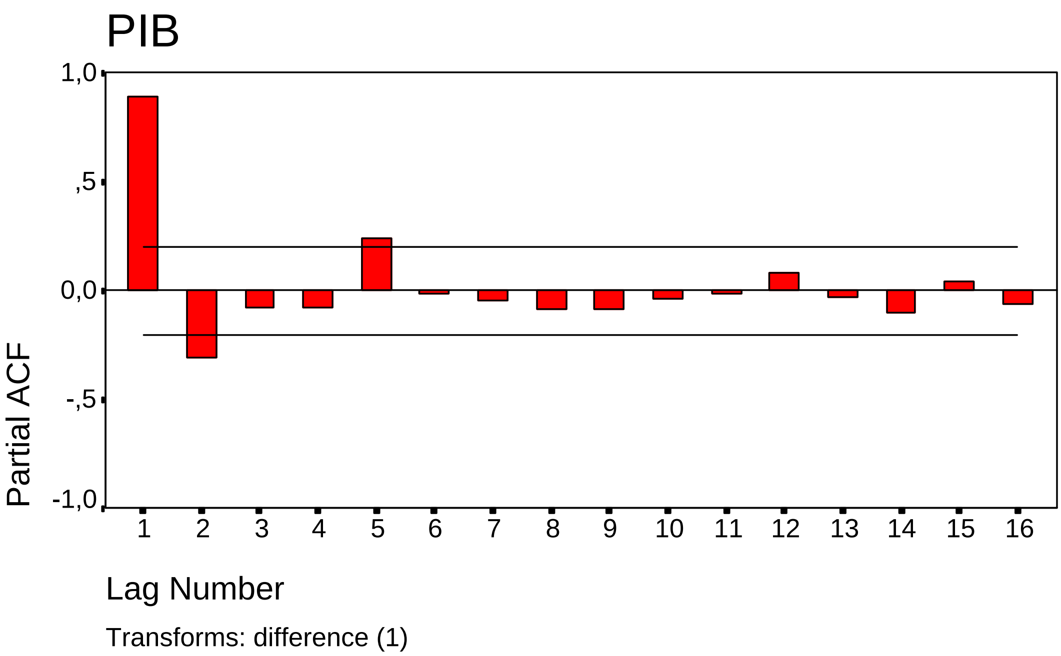


	Media	Varianza
TRIENIO		
1,00		
PIB	5432,72	96108,63
2,00		
PIB	6502,04	39370,84
3,00		
PIB	7026,87	19209,89
4,00		
PIB	7210,21	2997,51
5,00		
PIB	7300,73	4734,16
6,00		
PIB	7660,73	20475,76
7,00		
PIB	8531,79	161136,48
8,00		
PIB	9710,20	69046,78
9,00		
PIB	10033,86	9853,52

El instrumento auxiliar sugiere tomar únicamente una diferencia regular:

Variable	Mean	Std Dev	Variance
PIB	7748,66	1470,58	2162617,32
ΔPIB	52,21	44,47	1977,72
Δ_4PIB	208,33	165,68	27449,67
$\Delta\Delta_4$PIB	1,03	48,70	2371,94



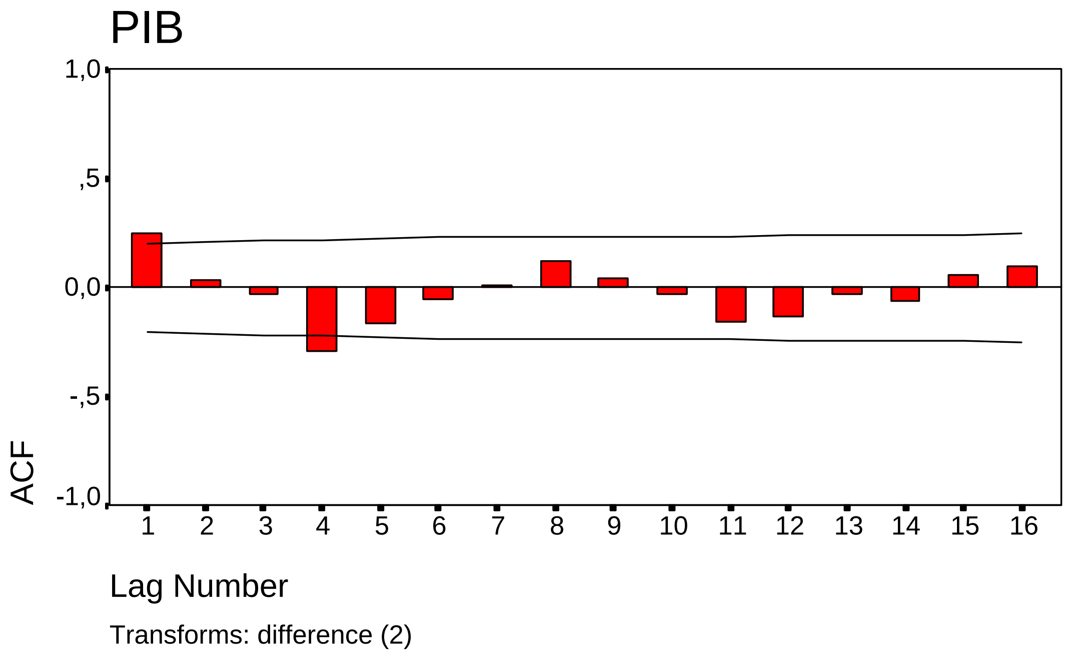


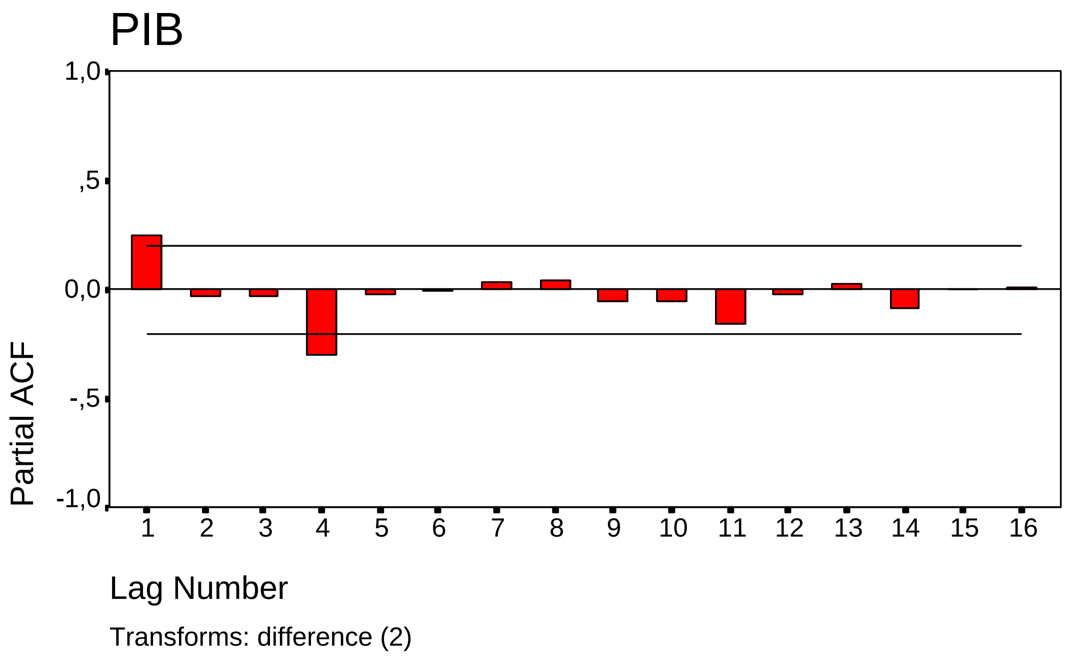
Si no hacemos caso del retardo 5 en la FAP (retardo no importante, por cuanto no está entre los tres primeros ni entre los estacionales, 4, 8, 12 y 16), los correlogramas podrían corresponder a los de un AR(2) únicamente en la parte regular (estructura decreciente en la FAS y dos palos significativos en la FAP).

ARIMA(2,1,0)(0,0,0)$_4$

```
FINAL PARAMETERS:

Number of residuals   99
Standard error        18,891642
Log likelihood        -430,79041
AIC                   867,58081
SBC                   875,36617

            Analysis of Variance:

              DF   Adj. Sum of Squares    Residual Variance

Residuals     96             34891,579            356,89415

           Variables in the Model:

                       B         SEB      T-RATIO   APPROX. PROB.

AR1             1,171826     ,096242    12,175807       ,00000000
AR2             -,311974     ,096248    -3,241360       ,00163589
CONSTANT       52,767721   13,042469     4,045838       ,00010567
```

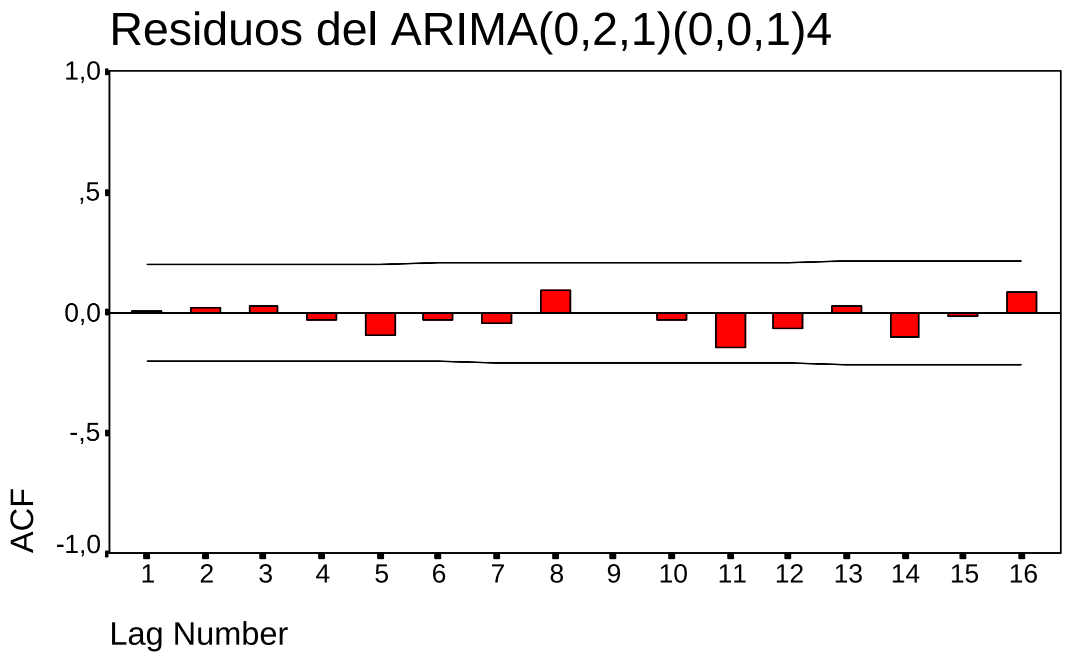
El modelo pasa la primera prueba (parámetros significativos) pero no la segunda, la suma de los parámetros AR es muy próxima a la unidad, sugiriendo una sub-diferenciación.





Tanto la parte regular como la estacional presentan una estructura muy similar, un único palo significativo en la FAS y en la FAP. Siguiendo el principio de parsimonia sugerimos un media móvil de orden uno para ambos componentes, ARIMA(0,2,1)(0,0,1)$_4$.

```
FINAL PARAMETERS:

Number of residuals   98
Standard error        19,169272
Log likelihood        -427,13703
AIC                   860,27405
SBC                   868,02895

            Analysis of Variance:

              DF   Adj. Sum of Squares    Residual Variance

Residuals     95             35024,038            367,46098

           Variables in the Model:

                          B          SEB       T-RATIO   APPROX. PROB.

MA1         -,23392028    ,1002599    -2,3331401        ,02175045
SMA1         ,25328113    ,1033348     2,4510737        ,01607152
CONSTANT     ,17239555   1,8058094      ,0954672        ,92414510
```



Como vemos el modelo pasa todas las pruebas, formulándose finalmente como:

$$\Delta^2 PIB_t = (1+0,2339L)(1-0,2533L^4)\varepsilon_t$$

8.6.- Predicción.

Cuando el modelo propuesto y estimado supera satisfactoriamente la etapa de *validación* entramos en la cuarta y última etapa, que no es otra que la de *predicción* de los valores futuros de la series.

La predicción se efectúa mediante un estadístico denominado *predictor óptimo* que, como vimos, minimiza el error cuadrático de predicción. La expresión formal de este predictor para una variable generada por un proceso ARIMA es la siguiente:

$$X_{T+h}^{*} = \phi_1 X_{T+h-1}^{*} + \phi_2 X_{T+h-2}^{*} + \ldots + \phi_p X_{T+h-p}^{*} + \varepsilon_{T+h}^{*} + \theta_1 \varepsilon_{T+h-1}^{*} + \theta_2 \varepsilon_{T+h-2}^{*} + \ldots + \theta_q \varepsilon_{T+h-q}^{*}$$

en donde X_{T+h} representa el valor predicho para X que se calcula con la información disponible hasta el periodo T y para h periodos en adelante, mientras que X_t^* y ε_t^* indican los valores de las variables X y ε que debemos incluir en la fórmula de predicción.

La construcción del predictor se efectúa según el esquema siguiente:

$$X_t^* = \begin{cases} X_t & para \quad t = 1,2,...,T \\ X_t^* & para \quad t = T+1, T+2,... \end{cases} \qquad \varepsilon_t^* = \begin{cases} e_t & para \quad t = 1,2,...,T \\ 0 & para \quad t = T+1, T+2,... \end{cases}$$

Cuando t se refiere a periodos de tiempo dentro de la muestra X^*, se sustituye por el valor observado correspondiente y ε^* por el residuo, que actúa como estimación del término de error desconocido. Por otra parte, cuando t se refiera a un periodo de tiempo fuera de la muestra, X^* se sustituye por la predicción obtenida justo anteriormente y ε^* por cero.

La predicción por punto se acompaña de intervalos de confianza:

$$\text{Predicción por intervalo} \qquad X_{T+h}^* \pm 1{,}96\sqrt{Var\left(X_{T+h}^*\right)}$$

siendo la varianza de la predicción:

$$Var\left(X_{T+h}^*\right) = \sigma_\varepsilon^2\left(1 + \Psi_1^2 + \Psi_2^2 + ... + \Psi_{h-1}^2\right)$$

en donde los coeficientes Ψ se obtienen a partir de la representación en forma de MA(∞) de cualquier modelo ARIMA.

Veamos por ejemplo la predicción a partir de un ARIMA(1,0,0).

$$X_t = \phi X_{t-1} + \varepsilon_t \qquad |\phi| < 1$$

El cual puede ser expresado como un MA(∞):

$$(1-\phi L)X_t = \varepsilon_t \qquad X_t = \frac{1}{(1-\phi L)}\varepsilon_t = \left(1+\phi L+\phi^2 L^2 + \ldots\right)\varepsilon_t$$

Puesto que la suma de una progresión geométrica con razón menor que la unidad, $|\phi|<1$, es decir $(1+\phi L+\phi^2L^2+...)$, no es más que $1/(1-\phi L)$. Nótese del mismo modo que cualquier MA *invertible* puede ser expresado como un AR(∞) (por lo que justificamos las condiciones de invertibilidad).

Volviendo a nuestro ejemplo de predicción: $\Psi_1=\phi, \Psi_2=\phi^2, ...$

Horizonte	Predictor puntual	Varianza de predicción
h=1	$X_{T+1}^*=\phi X_T$	σ_ε^2
h=2	$X_{T+2}^*=\phi X_{T+1}^*=\phi^2 X_T$	$\sigma_\varepsilon^2(1+\phi^2)$
h=3	$X_{T+3}^*=\phi X_{T+2}^*=\phi^3 X_T$	$\sigma_\varepsilon^2(1+\phi^2+\phi^4)$
...	...	...
h=∞	$X_{T+\infty}^*=\phi^\infty X_T=0$	$\sigma_\varepsilon^2/(1-\phi)$

Que reproduce, en la predicción para un horizonte infinito, la media y la varianza de la serie.

En cuanto a la predicción a partir de un ARIMA(0,0,1).

$$X_t = a + \varepsilon_t - \theta\varepsilon_{t-1} \qquad |\theta| < 1$$

$\Psi_1=\theta$ y $\Psi_k=0 \ \forall k>1$

Horizonte	Predictor puntual	Varianza de predicción
h=1	X_{T+1}*=a-θe_T	σ_ε^2
h=2	X_{T+2}*=a	$\sigma_\varepsilon^2(1+\theta^2)$
h=3	X_{T+3}*=a	$\sigma_\varepsilon^2(1+\theta^2)$
...	...	...
h=∞	$X_{T+\infty}$*=a	$\sigma_\varepsilon^2(1+\theta^2)$

Los procesos MA se dice que *no tienen memoria*, puesto que rápidamente (dependiendo del orden) convergen hacia la media y la varianza del proceso.

Por lo que respecta a un ARIMA(1,1,0) con un MA(4), el modelo final para la serie temporal del PIB inglés, por ejemplo:

$$(1-0{,}3208L)\Delta Ln(PIB)_t = 0{,}0189 + (1-0{,}3087L^4)\varepsilon_t$$

$$\Delta Ln(PIB)_t = 0{,}0189 + 0{,}3208\Delta Ln(PIB)_{t-1} + \varepsilon_t - 0{,}3087\varepsilon_{t-4}$$

$$\Delta Ln(PIB)_t = 0{,}0189+(1+0{,}3208L+0{,}3208^2L^2+...)(1-0{,}3087L^4)\varepsilon_t$$

$\Psi_1=0{,}3208$, $\Psi_2=0{,}3208^2=0{,}1029$, $\Psi_3=0{,}3208^3=0{,}0330$, $\Psi_4=0{,}3208^4-0{,}3087= -0{,}2981$

Horizonte	Predictor puntual	Varianza
h=1	$\Delta Ln(PIB)_{T+1}$*=0,0189+0,3208$\Delta Ln(PIB)_T$-0,3087e_{T-3}	σ_ε^2
h=2	$\Delta Ln(PIB)_{T+2}$*=0,0189+0,3208$\Delta Ln(PIB)_{T+1}$*-0,3087e_{T-2}	$\sigma_\varepsilon^2(1+0{,}3208^2)$

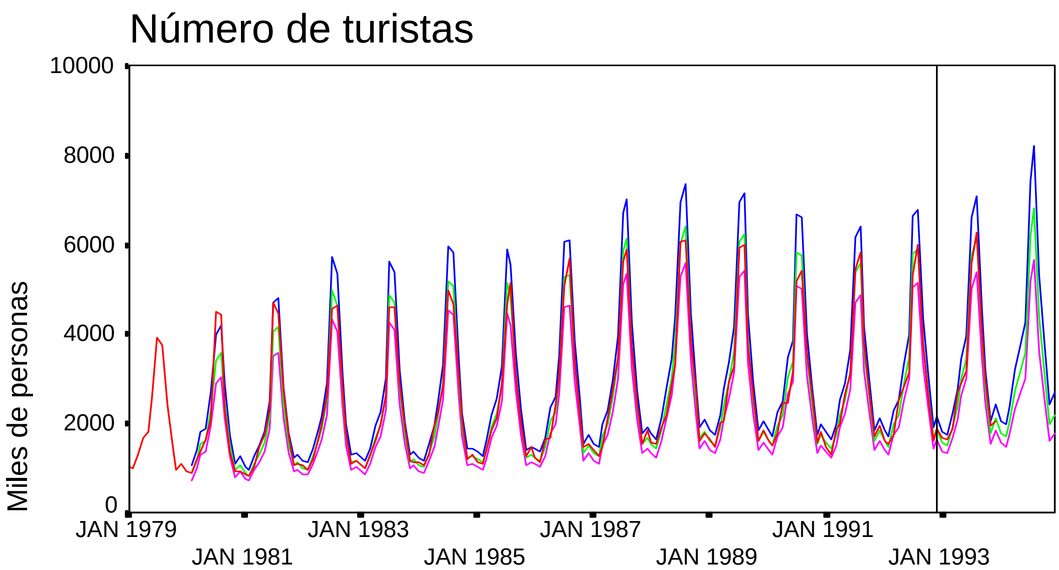
Es importante recordar que, en este caso, el valor de la predicción puntual del PIB no es igual al anti-logaritmo de la predicción, sino que se calcula como sigue:

$$PIB_{T+1}^* = \exp(Ln(PIB)_{T+1}^* + 0{,}5Var(Ln(PIB)_{T+1}^*))$$

Acabaremos el capítulo intentando realizar predicciones para el número de turistas extranjeros en España (en miles de personas) para 1993 y 1994, para lo cual y tras validar el modelo:

$$\Delta\Delta_{12}Ln(NT)_t=(1-0,6069L)(1-0,1092L^4)(1+0,1487L^{11})(1-0,6946L^{12})\varepsilon_t$$

calcularemos las predicciones puntuales y por intervalo utilizando las facilidades proporcionadas por el programa SPSS:



La capacidad predictiva puede valorarse utilizando los siguientes estadísticos:

```
Error Variable                      ERROR_Predicción
 Observed Variable                   NT
 N of Cases          Use             167
 Deg Freedom         Use             164
 Mean Error          Use             -2,2694
 Mean Abs Error      Use            141,9929
 Mean Pct Error      Use              -,0176
 Mean Abs Pct Err    Use               5,4027
 SSE (SCE)           Use          7509187,54
 MSE (ECM)           Use          45787,7289
 RMS (√ECM)          Use            213,9807
 Durbin-Watson       Use               1,6634
```

Comparar el error porcentual absoluto medio de predicción, MAPE, (5,4027%) con el error porcentual cometido en el ajuste (6,7557%), lo que evidencia una muy buena capacidad predictiva.