

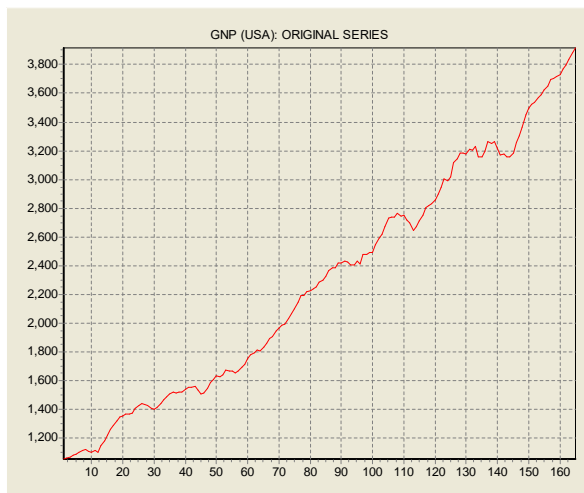
# 1 Introducción

## 1.1 Definición de series temporal y de proceso estocástico.

### SERIE TEMPORAL

- Punto de vista descriptivo: Es una sucesión ordenada en el tiempo de observaciones de una variable separadas entre si por el mismo lapso temporal.

PIB USA trimestral 1947-I a 1987.IV



Las series la podemos clasificar por tanto desde este punto de vista en cuanto a su periodicidad en anuales, trimestrales, mensuales, ...

- Punto de vista estocástico: Una serie temporal es la realización de un proceso estocástico.
- **Proceso estocástico:** Un proceso estocástico es un conjunto ordenado de variables aleatorias que depende de un indicador en el caso de las series temporales el indicador es el tiempo  $t$   $\{x_t\}$  ,  $\{x_{t_1}, x_{t_2}, x_{t_3}, \dots, x_{t_n}\}$  .

Entonces, una serie temporal es una realización de tamaño muestral 1 de cada una de las variables aleatorias que integran el proceso estocástico. Como sólo vamos a disponer de una única realización muestral por v.a., será necesario asumir que se cumplirán un conjunto de supuestos para

poder trabajar con las series temporales. Para poder extraer conclusiones a partir de la información de la series temporal (muestra) que sean válidas para el proceso estocástico (población). Estos supuestos o propiedades serán básicamente la estacionariedad y la ergodicidad.

## **1.2 Estacionariedad.**

Dado que un proceso estocástico es un conjunto de variables aleatorias y que las v.a. se caracterizan através de sus funciones de distribución y densidad y de sus momentos poblacionales (valor esperado, varianza, covarianzas, ...), caracterizemos los procesos estocásticos através de la Función de Distribución Conjunta del proceso estocástico y de los momentos muestrales:

$$\begin{aligned}
&F(x_{t_1}, x_{t_2}, x_{t_3}, \dots, x_{t_n}) \\
&E[x_{t_1}], E[x_{t_2}], \dots, E[x_{t_n}] \\
&VAR[x_{t_1}], VAR[x_{t_2}], \dots, VAR[x_{t_n}] \\
&COV[x_{t_1}x_{t_2}], COV[x_{t_1}x_{t_3}], \dots
\end{aligned}$$

Por ejemplo, supongamos que estamos analizando la serie temporal de PIB USA del gráfico anterior, como disponemos de 165 observaciones para caracterizar el proceso estocástico deberíamos sacar conclusiones sobre las v.a.'s integrantes de proceso estocástico a partir de realizaciones muestrales de tamaño 1. Es decir, por ejemplo a partir de una única observación obtener conclusiones sobre el valor esperado  $E[ ]$  y varianza  $VAR[ ]$  de una variable aleatoria.

- **Estacionariedad en sentido estricto:** Un proceso estocástico se dice que es estacionario en sentido estricto o fuerte si para cualquier subconjunto  $\{t_1, t_2, \dots, t_h\}$  se cumple:

$$F(x_{t_1}, x_{t_2}, \dots, x_{t_h}) = F(x_{t_1+k}, x_{t_2+k}, \dots, x_{t_h+k}).$$

Es decir, que la función de distribución conjunta es invariante ante desplazamiento en el tiempo. Por lo tanto, también serán invariante ante desplazamientos en el tiempo todos los momentos poblacionales dado que estos se definen a partir de la función de densidad. Esta propiedad es muy difícil de comprobar o casi imposible.

El valor esperado o media de un proceso  $x_t$  será:

$$\mu_t = E[x_t],$$

la varianza del proceso  $x_t$  como:

$$\sigma_t^2 = VAR[x_t] = E[(x_t - \mu_t)^2],$$

la covarianza entre  $x_{t_1}$  y  $x_{t_2}$  como:

$$\gamma(t_1, t_2) = COV[x_{t_1}x_{t_2}] = E[(x_{t_1} - \mu_{t_1})(x_{t_2} - \mu_{t_2})],$$

y finalmente la correlación entre  $x_{t_1}$  y  $x_{t_2}$  como:

$$\rho(t_1, t_2) = \frac{\gamma(t_1, t_2)}{\sqrt{\sigma_{t_1}^2} \sqrt{\sigma_{t_2}^2}}.$$

Para los procesos estacionarios en sentido estricto, dado que la distribución conjunta es invariante ante desplazamientos en el tiempo, o es la misma para todo  $t$ , el valor esperado  $E[x_t] = \mu_t = \mu$  será constante, siempre que  $E[|x_t|] < \infty$ . De igual manera si  $E[x_t^2] < \infty$  para procesos estacionarios en sentido estricto se cumplirá  $VAR[x_t] = \sigma_t^2 = \sigma^2$  para todo  $t$  y la varianza también será constante. Finalmente dado que para procesos estrictamente estacionarios se cumplirá  $F(x_{t_1}, x_{t_2}) = F(x_{t_1+k}, x_{t_2+k})$  para cualquier entero  $t_1$ ,  $t_2$  y  $k$ , también tendremos:

$$\begin{aligned}\gamma(t_1, t_2) &= \gamma(t_1 + k, t_2 + k) \\ \rho(t_1, t_2) &= \rho(t_1 + k, t_2 + k).\end{aligned}$$

Si usamos  $t_1 = t$  y  $t_2 = t - k$  tenemos:

$$\begin{aligned}
\gamma(t_1, t_2) &= \gamma(t, t - k) = \gamma(t - k, t) = \gamma(t, t + k) \\
&= \gamma(t + k, t) = \gamma_k \\
\rho(t_1, t_2) &= \rho(t, t - k) = \rho(t - k, t) = \rho(t, t + k) \\
&= \rho(t + k, t) = \rho_k.
\end{aligned}$$

Es decir, para procesos estrictamente estacionarios con los dos primeros momentos con respecto al origen ( $E[x_t]$  y  $E[x_t^2]$ ) la covarianza y correlación entre  $x_t$  y  $x_{t \pm k}$  sólo depende de la diferencia temporal  $k$  (lapso temporal que separa a  $x_t$  y  $x_{t \pm k}$ ).

En la práctica es muy difícil comprobar si un proceso es estacionario en sentido estricto o fuerte. En la práctica se trabaja con el concepto de estacionariedad en sentido débil. Un proceso se dice que es estacionario en sentido débil de orden  $n$  si todos sus momentos hasta el orden  $n$  son finitos e invariante con respecto al tiempo (es decir, que su valor no dependerá del periodo de tiempo en el se encuentren). Normalmente se trabaja con estacionariedad en sentido débil de segundo orden.

- **Estacionariedad en sentido débil (segundo orden)**: Se dice que un proceso estocástico es estacionario en sentido débil o de segundo orden cuando se cumple simultáneamente las tres siguientes condiciones:

$$\begin{aligned}
 E[x_t] &= E[x_{t-k}] = \mu \\
 VAR[x_t] &= VAR[x_{t-k}] = \sigma^2 \\
 COV[x_s, x_t] &= COV.[x_{s-k}, x_{t-k}] = \gamma_k \\
 \forall k &= \pm 1, \pm 2, ..
 \end{aligned}$$

La estacionariedad en sentido estricto implica estacionariedad en sentido débil, pero la existencia de estacionariedad en sentido débil sólo implica estacionariedad en sentido estricto cuando el proceso estocástico sigue una ley normal. En la práctica se trabaja con el concepto de estacionariedad en sentido débil de segundo orden que implica que el valor esperado y varianza son constantes y que la covarianza entre v.a.'s asociadas a distintas observaciones sólo depende del lapso temporal que las separa.



- **Ergodicidad:** se dice que un proceso es ergódico cuando se cumple:

$$\lim_{T \rightarrow \infty} \left( \frac{1}{T} \sum_{k=-(T-1)}^{T-1} \rho_k \right) = 0.$$

Donde  $T$  es el tamaño muestral. La ergodicidad permite limitar el grado de dependencia temporal y nos permite modelizar la dependencia temporal de un proceso estocástico utilizando un conjunto finito de parámetros. Para ilustrar esta condición es útil que necesitamos para que la media muestral calculada sobre la serie temporal sea un estimador consistente de la media poblacional o valor

esperado de un proceso estacionario:

$$\begin{aligned}\bar{x} &= \frac{1}{T} \sum_{t=1}^T x_t \\ E[\bar{x}] &= \frac{1}{T} E \left[ \sum_{t=1}^T x_t \right] = \frac{1}{T} E[x_1 + x_2 + \cdots + x_T] \\ &= \frac{T}{T} \mu = \mu \\ VAR[\bar{x}] &= \frac{1}{T^2} VAR \left[ \sum_{t=1}^T x_t \right] = \frac{1}{T^2} VAR[x_1 + \cdots + x_T] \\ &= \frac{1}{T^2} \sum_{t=1}^T \sum_{s=1}^T COV[x_t x_s] \\ &= \frac{\gamma_0}{T^2} \sum_{t=1}^T \sum_{s=1}^T \rho(t-s) \\ &= \frac{\gamma_0}{T^2} \sum_{k=-(T-1)}^{T-1} (T - |k|) \rho_k \\ &= \frac{\gamma_0}{T} \sum_{k=-(T-1)}^{T-1} \left( 1 - \frac{|k|}{T} \right) \rho_k \\ k &= (t - s) .\end{aligned}$$

Entonces se puede ver que para que  $\bar{x}$  sea un estimador consistente, es decir, que  $VAR[\bar{x}]$  tienda a zero a medida que el tamaño muestral tienda infinito ( $\lim_{T \rightarrow \infty} VAR[\bar{x}] = 0$ ) lo que necesitamos es que  $\left( \frac{1}{T} \sum_{k=-(T-1)}^{T-1} \rho_k \right)$  tienda a cero cuando se incrementa el tamaño muestral.

### 1.3 Funciones de autocovarianzas, autocorrelación simple y parcial.

A la covarianza entre dos variables aleatorias integrantes de un mismo proceso estocástico la llamaremos autovarianza  $\gamma_k$  y dado que asumiremos que los procesos son estacionarios,  $k$  se refiere al lapso temporal que separa a las dos observaciones:

$$\begin{aligned}
\gamma_k &= COV [x_t, x_{t-k}] \\
&= E [(x_t - E [x_t]) (x_{t-k} - E [x_{t-k}])] \\
&= E [(x_t - \mu) (x_{t-k} - \mu)]
\end{aligned}$$

donde :

$$E [x_t] = E [x_{t-k}] = \mu.$$

Para los procesos estacionarios se cumplen las siguientes propiedades:

$$\begin{aligned}
\gamma_0 &= VAR [x_t] = \sigma^2 \\
\gamma_k &= \gamma_{-k} \\
|\gamma_k| &\leq \gamma_0 \quad \forall k \geq 1.
\end{aligned}$$

Al conjunto de todas las autocovarianzas que se pueden calcular para una serie temporal o proceso estocástico se le llama **función de autocovarianzas**  $\gamma_0, \gamma_1, \gamma_2, \dots, \gamma_k, \gamma_{k+1}, \dots$ . Una propiedad importante de la función

de autocovarianzas es que es semidefinda positiva en el sentido de que:

$$\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j \gamma_{|t_i - t_j|} \geq 0 \quad (1)$$

para cualquier conjunto de puntos en el tiempo  $t_1, t_2, \dots, t_n$  y conjunto de números reales  $\alpha_1, \alpha_2, \dots, \alpha_n$ . Si definimos la variable aleatoria  $X = \sum_{i=1}^n \alpha_i x_{t_i}$  tenemos:

$$0 \leq VAR[X] = \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j \gamma_{|t_i - t_j|}.$$

A partir de la autocovarianza de orden  $k$  ( $\gamma_k$ ) se define la autocorrelación de orden  $k$  ( $\rho_k$ ) como:

$$\rho_k = \frac{COV[x_t, x_{t-k}]}{\sqrt{VAR[x_t]} \sqrt{VAR[x_{t-k}]}} = \frac{\gamma_k}{\gamma_0}. \quad (2)$$

Y es fácil ver que se cumple:

$$\begin{aligned}\rho_0 &= 1 \\ \rho_k &= \rho_{-k} \\ -1 &\leq \rho_k \leq 1\end{aligned}$$

Al conjunto de todas las autocorrelaciones que pueden calcular se le llama **función de autocorrelación simple (FAS)**  $\rho_1, \rho_2, \dots, \rho_k, \rho_{k+1}, \dots$ . A partir de (1) también se puede establecer que:

$$\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j \rho_{|t_i - t_j|} \geq 0$$

También es de utilidad definir los coeficientes de autocorrelación parcial como:

$$\phi_{k,k} = \text{Corr} \left[ x_t, x_{t-k} | x_{t-1}, \dots, x_{t-(k-1)} \right] .$$

Es decir, estos coeficientes, a diferencia de los coeficientes de correlación simple  $\rho_k$  que miden la correlación entre

dos v.a.'s que distan entre si  $k$  períodos, miden las correlacion entre  $x_t$  y  $x_{t-k}$  pero descontando el efecto de  $x_{t-1}, \dots, x_{t-(k-1)}$ . Una forma sencilla de entender la idea es através de la siguiente regresión:

$$x_t = \phi_{k,1}x_{t-1} + \phi_{k,2}x_{t-2} + \dots + \phi_{k,k}x_{t-k} + u_t$$

El parametro que medirá el efecto que tiene una variación de  $x_{t-k}$  sobre  $x_t$ , descontando el efecto de las observaciones intermedias es  $\phi_{k,k}$ . Si multiplicamos la expression anterior por  $x_{t-j}$  y tomamos esperanzas (asumiendo que  $E[x_t] = 0$ ) es fácil ver que:

$$\begin{aligned} x_{t-j}x_t &= \phi_{k,1}x_{t-j}x_{t-1} + \dots + \phi_{k,k}x_{t-j}x_{t-k} \\ &\quad + x_{t-j}u_t \\ E[x_{t-j}x_t] &= \phi_{k,1}E[x_{t-j}x_{t-1}] + \dots + \phi_{k,k}E[x_{t-j}x_{t-k}] \\ &\quad + E[x_{t-j}u_t] \\ \gamma_j &= \phi_{k,1}\gamma_{j-1} + \phi_{k,2}\gamma_{j-2} + \dots + \phi_{k,k}\gamma_{k-j} \\ \rho_j &= \phi_{k,1}\rho_{j-1} + \phi_{k,2}\rho_{j-2} + \dots + \phi_{k,k}\rho_{k-j}, \end{aligned}$$

asumiendo que  $E[x_{t-k}u_t] = 0$ . Es fácil ver que existe una relación entre los coeficientes de autocorrelación simple  $\rho_j$  y parcial  $\phi_{j,j}$ . Dando valores a  $j$  obtenemos el sistema:

$$\begin{aligned}\rho_1 &= \phi_{k,1} + \phi_{k,2}\rho_1 + \dots + \phi_{k,k}\rho_{k-1} \\ \rho_2 &= \phi_{k,1}\rho_1 + \phi_{k,2} + \dots + \phi_{k,k}\rho_{k-2} \\ &\vdots \\ \rho_k &= \phi_{k,1}\rho_{k-1} + \phi_{k,2}\rho_{k-2} + \dots + \phi_{k,k}\end{aligned}$$

Que expresado en términos matriciales responde a:

$$\begin{bmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \\ \vdots \\ \rho_k \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{k-2} \\ \rho_2 & \rho_1 & 1 & \dots & \rho_{k-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \dots & 1 \end{bmatrix} \times \begin{bmatrix} \phi_{k,1} \\ \phi_{k,2} \\ \phi_{k,3} \\ \vdots \\ \phi_{k,k} \end{bmatrix}.$$



$$\begin{bmatrix} \phi_{k,1} \\ \phi_{k,2} \\ \phi_{k,3} \\ \vdots \\ \phi_{k,k} \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{k-2} \\ \rho_2 & \rho_1 & 1 & \dots & \rho_{k-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \dots & 1 \end{bmatrix}^{-1} \times \begin{bmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \\ \vdots \\ \rho_k \end{bmatrix}.$$

Aplicando la regla de Cramer:

$$\begin{aligned} \phi_{1,1} &= \rho_1 \\ \phi_{k,k} &= \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_1 \\ \rho_1 & 1 & \rho_1 & \dots & \rho_2 \\ \rho_2 & \rho_1 & 1 & \dots & \rho_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \dots & \rho_k \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{k-2} \\ \rho_2 & \rho_1 & 1 & \dots & \rho_{k-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \dots & 1 \end{vmatrix}}. \end{aligned} \quad (3)$$

Al conjunto de todos los coeficientes de autocorrelación parcial se le llama **función de autocorrelación parcial (FAP)**.

Tal y como iremos viendo las funciones autocorrelación simple y parcial se representan gráficamente, lo que recibe el nombre de correlograma. Nosotros identificaremos el tipo de proceso estocástico que sigue una serie en base a analizar el correlograma de dicha serie. Pero utilizaremos los correlogramas muestrales que los obtendremos teniendo en cuenta que estimaremos las autocovarianzas utilizando:

$$\begin{aligned}\hat{\gamma}_k &= \frac{1}{T-k} \sum (x_t - \bar{x}) (x_{t-k} - \bar{x}) \\ \hat{\gamma}_0 &= \frac{1}{T} \sum (x_t - \bar{x})^2 \\ \bar{x} &= \frac{1}{T} \sum x_t,\end{aligned}$$

y con posterioridad utilizando (2) y (3), pero sustituyendo  $\gamma_j$  por  $\hat{\gamma}_j$  y  $\rho_j$  por  $\hat{\rho}_j$ .

## 1.4 Procesos de Ruido Blanco y Paseo Aleatorio.

- Se dice que una serie sigue un proceso de ruido blanco cuando responde al siguiente esquema:

$$\begin{aligned}x_t &= \varepsilon_t \\ \varepsilon_t &\sim iid(0, \sigma_\varepsilon^2)\end{aligned}$$

donde  $iid(0, \sigma_\varepsilon^2)$  significa idénticamente e independientemente distribuidos con valor esperado 0 y varianza  $\sigma_\varepsilon^2$ . Es fácil ver que para un ruido blanco se cumple:

$$\begin{aligned}E[x_t] &= E[\varepsilon_t] = 0 \\ \gamma_0 &= VAR[x_t] = VAR[\varepsilon_t] = \sigma_\varepsilon^2 \\ \gamma_k &= COV[x_t x_{t-k}] = COV[\varepsilon_t \varepsilon_{t-k}] = 0 \\ \rho_k &= \frac{\gamma_k}{\gamma_0} = 0 \\ \forall k &\geq 1.\end{aligned}$$

Por lo tanto presenta un correlograma con todos los coeficientes iguales a cero. Este proceso es el proceso estacionario más sencillo. Y lo utilizaremos para definir el resto de procesos que analizaremos.

- Se dice que una serie sigue un proceso de paseo aleatorio cuando responde al siguiente esquema:

$$\begin{aligned}x_t &= x_{t-1} + \varepsilon_t \\ \varepsilon_t &\sim iid(0, \sigma_\varepsilon^2) .\end{aligned}\tag{4}$$

Tal y como veremos a continuación es el proceso no estacionario más sencillo que se puede definir. Si sustituimos recursivamente en (4) es sencillo ver que:

$$\begin{aligned}
x_t &= x_{t-1} + \varepsilon_t \\
&= (x_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\
&= (x_{t-3} + \varepsilon_{t-2}) + \varepsilon_{t-1} + \varepsilon_t \\
&\vdots \\
x_t &= x_0 + \sum_{i=1}^t \varepsilon_i
\end{aligned}$$

Donde  $x_0$  es una observación inicial y usualmente se asume que es una constante o igual a cero ( $x_0 = 0$ ). Para el valor esperado tenemos:

$$\begin{aligned}
E[x_t] &= E\left[x_0 + \sum_{i=1}^t \varepsilon_i\right] \\
&= x_0 + E[\varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_t] \\
&= x_0 + E[\varepsilon_1] + E[\varepsilon_2] + \cdots + E[\varepsilon_t] \\
&= x_0 + 0 + 0 + \cdots + 0 = x_0.
\end{aligned}$$

Y en cuanto a la varianza:

$$\begin{aligned}VAR[x_t] &= VAR\left[x_0 + \sum_{i=1}^t \varepsilon_i\right] = VAR\left[\sum_{i=1}^t \varepsilon_i\right] \\&= VAR[\varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_t] \\&= VAR[\varepsilon_1] + VAR[\varepsilon_2] + \cdots + VAR[\varepsilon_t] \\&= \sigma_\varepsilon^2 + \sigma_\varepsilon^2 + \cdots + \sigma_\varepsilon^2 = t\sigma_\varepsilon^2.\end{aligned}$$

Por lo tanto no es estacionario, dado que, la varianza depende del período temporal en el que nos encontramos.

## 2 Modelos ARIMA

Antes de ver los procesos definiremos el operador de retardos  $L$ :

$$\begin{aligned} Lx_t &= x_{t-1} \\ L^k x_t &= x_{t-k}, \end{aligned}$$

es decir, este operador retarda tantos períodos la serie a la que se aplica como exponente tenga.

### **Polinomio en el operador de retardos:**

$$\alpha_k(L) = (1 + a_1L + a_2L^2 + \cdots + a_pL^k)$$

Es algebraicamente similar a un polinomio  $(1 + a_1z + a_2z^2 + \cdots + a_pz^k)$  donde  $z$  es un escalar, la diferencia estriba en que el polinomio  $(1 + a_1z + a_2z^2 + \cdots + a_pz^k)$  está referido a un escalar (incognita) y el polinomio en el operador de retardos  $(1 + a_1L + a_2L^2 + \cdots + a_pL^k)$  está referido a un operador que retarda las series temporales.

$$\begin{aligned}\alpha_p(L) x_t &= \left(1 + a_1 L + a_2 L^2 + \cdots + a_p L^p\right) x_t \\ &= x_t + a_1 x_{t-1} + a_2 x_{t-2} + \cdots + a_p x_{t-p}\end{aligned}$$

La condición de estabilidad de polinomio en el operador de retardos consiste en las raíces de la ecuación (soluciones):

$$\left(1 + a_1 L + a_2 L^2 + \cdots + a_p L^p\right) = 0$$

caigan fuera del círculo unidad:

$$\left|L_j\right| > 1 \quad j = 1, \cdots, p$$

Es preciso tener presente que los valores de  $L$  que solucionan  $\left(1 + a_1 L + a_2 L^2 + \cdots + a_p L^p\right) = 0$   $L_j$  son los valores recíprocos o inversos de los valores  $z_j$  que solucionan:



$$\begin{aligned} \left( z^p + a_1 z^{p-1} + a_2 z^{p-2} + \cdots + a_p \right) &= 0 \\ z_j &= \frac{1}{L_j} \quad j = 1, \cdots, p \end{aligned}$$

Entonces la condición de estabilidad puede ser definida en términos de que los inversos de las raíces caigan en el círculo unidad:

$$|z_j| < 1 \quad j = 1, \cdots, p$$

El polinomio en el operador de retardos  $(1 + a_1 L + a_2 L^2 + \cdots + a_p L^p)$  se dice que es estable, si todas sus raíces caen fuera del círculo unidad  $(|L_j| > 1)$  o alternatively, si los inversos de las raíces caen dentro del círculo unidad  $(|z_j| < 1)$ .

Finalmente es interesante tener en cuenta que un polinomio se puede descomponer de la siguiente manera:

$$\begin{aligned}
 (1 + a_1L + a_2L^2 + \cdots + a_pL^p) &= \prod_{j=1}^p (1 + z_jL) \\
 &= \prod_{j=1}^p \left(1 + \frac{1}{L_j}L\right).
 \end{aligned}$$

## 2.1 Procesos de medias móviles (MA) y procesos autoregresivos.

En el análisis de series temporales existen dos maneras muy útiles de expresar los procesos de series temporales. Uno es escribir un proceso  $x_t$  como combinación lineal de una secuencia de v.a. incorrelacionadas entre si (ruido blanco):

$$\begin{aligned}
 x_t &= \varepsilon_t + \psi_1\varepsilon_{t-1} + \psi_2\varepsilon_{t-2} + \cdots & (5) \\
 &= \sum_{j=0}^{\infty} \psi_j\varepsilon_{t-j} & \psi_0 = 1 \\
 \varepsilon_t &\sim iid(0, \sigma_\varepsilon^2)
 \end{aligned}$$

Utilizando el operador de retardos:

$$\begin{aligned}x_t &= \varepsilon_t + \psi_1 L \varepsilon_t + \psi_2 L^2 \varepsilon_t + \cdots \\&= \left(1 + \psi_1 L + \psi_2 L^2 + \cdots\right) \varepsilon_t \\&\quad \psi(L) \varepsilon_t.\end{aligned}$$

$$\psi(L) = 1 + \sum_{j=1}^{\infty} \psi_j L^j = \sum_{j=0}^{\infty} \psi_j L^j \quad \psi_0 = 1.$$

Podemos comprobar que:

$$\begin{aligned}E[x_t] &= E[\varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \cdots] = 0 \\VAR[x_t] &= VAR[\varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \cdots] \\&= \sigma_{\varepsilon}^2 \sum_{j=0}^{\infty} \psi_j^2.\end{aligned}$$

Tambien se puede ver que:

$$E[\varepsilon_t x_{t-j}] = \begin{cases} \sigma_{\varepsilon}^2 & \text{for } j = 0 \\ 0 & \text{for } j > 0. \end{cases}$$

Y en cuanto a la función de autocovarianzas y autocorrelación, se puede ver que:

$$\begin{aligned}
 \gamma_k &= E \left[ \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \sum_{l=0}^{\infty} \psi_l \varepsilon_{t-k-l} \right] \\
 &= E \left[ \sum_{j=0}^{\infty} \sum_{l=0}^{\infty} \psi_j \psi_l \varepsilon_{t-j} \varepsilon_{t-k-l} \right] \\
 &= \sigma_{\varepsilon}^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+k} \\
 \rho_k &= \frac{\sum_{i=0}^{\infty} \psi_i \psi_{i+k}}{\sum_{i=0}^{\infty} \psi_i^2}.
 \end{aligned} \tag{6}$$

Para que el proceso sea estacionario la varianza debe de ser constante y finita por lo tanto es necesario que  $\sigma_{\varepsilon}^2 \sum_{j=0}^{\infty} \psi_j^2 < \infty$ , por lo tanto bastará con que  $\sum_{j=0}^{\infty} \psi_j^2 < \infty$ . También se puede ver a través de la desigualdad de Cauchy–Schwarz que esta condición es la que necesitamos para que las funciones de autocovarianzas y autocorrelación también sean finitas, dado que:

$$|\gamma_k| = |COV[x_t x_{t-k}]| \leq (VAR[x_t] VAR[x_{t-k}])^{1/2} = \sigma_{\varepsilon}^2 \sum_{j=0}^{\infty} \psi_j^2$$

A la representación (5) de un proceso se le conoce con el nombre de representación Medía Móvil (MA), en concreto es una representación  $MA(\infty)$ . A cualquier proceso que admita una representación (5) se le conoce con el nombre de proceso lineal.

Dada la función de autocovarianzas (secuencia de autocovarianzas  $\gamma_k$  para  $k = 0, \pm 1, \pm 2, \dots$ ), la función generatriz de autocovarianzas se define como:

$$\gamma(L) = \sum_{k=-\infty}^{\infty} \gamma_k L^k$$

donde  $\gamma_0$  es la varianza del proceso y el coeficiente asociado a  $L^0$  y  $\gamma_k$  es la autocovarianza de orden  $k$  y el coeficiente asociado a  $L^k$  y  $L^{-k}$ . Utilizando (6) podemos

escribir:

$$\begin{aligned}
\gamma(L) &= \sum_{k=-\infty}^{\infty} \sigma_{\varepsilon}^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+k} L^k = \\
&= \sigma_{\varepsilon}^2 \sum_{k=-\infty}^{\infty} \sum_{i=0}^{\infty} \psi_i \psi_{i+k} L^k \\
&= \sigma_{\varepsilon}^2 \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \psi_i \psi_j L^{j-i} \quad j = i + k \\
&= \sigma_{\varepsilon}^2 \sum_{i=0}^{\infty} \psi_i L^i \sum_{j=0}^{\infty} \psi_j L^{-j} \\
&= \sigma_{\varepsilon}^2 \psi(L) \psi(L^{-1}) \\
&= \sigma_{\varepsilon}^2 \psi(L) \psi(F) \quad F = L^{-1}.
\end{aligned} \tag{7}$$

La función generatriz de autocovarianzas nos proporciona un método sencillo y muy conveniente de calcular la varianzas y autocovarianzas para los procesos lineales.

Otra forma muy útil de escribir un proceso estocástica, es mediante su representación autorregresiva (AR), en la que se expresa el proceso  $x_t$  como una combinación lineal

de sus valores pasados y de un proceso de ruido blanco:

$$\begin{aligned} (1 + \pi_1 L + \pi_2 L^2 + \dots) x_t &= \varepsilon_t \\ \pi(L) x_t &= \varepsilon_t \\ \pi(L) &= 1 + \sum_{j=1}^{\infty} \pi_j L^j. \end{aligned}$$

Box y Jenkins llamaron a los procesos que admitían una representación  $AR(\infty)$  y que cumplen  $1 + \sum_{j=1}^{\infty} |\pi_j| < \infty$  procesos invertibles. Para que un proceso lineal  $x_t = \psi(L) \varepsilon_t$  sea invertible, es decir pueda ser escrito como un  $AR(\infty)$  tal que  $1 + \sum_{j=1}^{\infty} |\pi_j| < \infty$  el polinomio  $\psi(L)$  tiene que ser estable, es decir, las raíces de  $\psi(L) = 0$  tienen que caer fuera del círculo unidad. De igual manera no todos los procesos invertibles son estacionarios para que un proceso invertible  $\pi(L) x_t = \varepsilon_t$  sea estacionario el polinomio  $\pi(L)$  tiene que ser estable. Lo habitual es trabajar con procesos AR y MA finitos, en esos casos los coeficientes se suelen denotar  $\pi_i = -\phi_i$  y  $\psi_i = -\theta_i$  dejando  $\pi_i$  y  $\psi_i$  para los procesos de orden infinito.

## 2.2 Procesos de medias móviles MA(q).

Una serie sigue un proceso de medias móviles de orden  $q$  si responde al siguiente esquema:

$$x_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}$$

$$x_t = \varepsilon_t - \theta_1 L \varepsilon_t - \cdots - \theta_q L^q \varepsilon_t$$

$$x_t = (1 - \theta_1 L - \cdots - \theta_q L^q) \varepsilon_t$$

$$x_t = \theta_q(L) \varepsilon_t$$

donde :

$$\theta_q(L) = (1 - \theta_1 L - \cdots - \theta_q L^q).$$

$$\varepsilon_t \sim iid(0, \sigma_\varepsilon^2).$$

Los procesos MA finitos son siempre estacionarios al ser combinaciones lineales de procesos ruido blanco y dado que se cumple  $\sum_{j=0}^{\infty} \theta_j^2 < \infty$ . Pero para este tipo de procesos hay que comprobar si son invertibles, es decir, se pueden expresar como un  $AR(\infty)$   $\pi(L)x_t = \varepsilon_t$  con



$1 + \sum_{j=1}^{\infty} |\pi_j| < \infty$ , es decir un proceso  $AR(\infty)$  convergente. Un proceso MA es invertible si el polinomio en el operador de retardos  $\theta_q(L)$  asociado al proceso es estable, es decir si al calcular las raíces del polinomio:

$$\theta_q(L) = (1 - \theta_1 L - \dots - \theta_q L^q) = 0 \quad (8)$$

caen fuera del círculo unidad (los valores de  $L$  que satisfacen (8) cumplen  $|L| > 1$ ). Los procesos MA no invertibles no admiten una representación Autorregresiva convergente tal y como veremos mas adelante.

### 2.2.1 MA(1).

Un proceso MA(1) responde al esquema:

$$x_t = \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

$$x_t = \varepsilon_t - \theta_1 L \varepsilon_t$$

$$x_t = (1 - \theta_1 L) \varepsilon_t$$

$$x_t = \theta_1(L) \varepsilon_t$$

donde :

$$\theta_1(L) = (1 - \theta_1 L).$$

Al calcular las raíces del polinomio  $(1 - \theta_1 L)$ , se ve que:

$$(1 - \theta_1 L) = 0$$

$$L = 1/\theta_1$$

$$|\theta_1| < 1 \Rightarrow |L| > 1.$$

Por tanto la condición de invertibilidad para un proceso MA(1) es  $|\theta_1| < 1$ .

En cuanto al valor esperado se puede ver que:

$$\begin{aligned}
 E[x_t] &= E[\varepsilon_t - \theta_1 \varepsilon_{t-1}] \\
 &= E[\varepsilon_t] - \theta_1 E[\varepsilon_{t-1}] \\
 &= 0 - \theta_1 \times 0 = 0
 \end{aligned}$$

La varianza:

$$\begin{aligned}
 VAR[x_t] &= VAR[\varepsilon_t - \theta_1 \varepsilon_{t-1}] \\
 &= VAR[\varepsilon_t] + \theta_1^2 VAR[\varepsilon_{t-1}] \\
 &\quad \sigma_\varepsilon^2 (1 + \theta_1^2).
 \end{aligned}$$

Las autocovarianzas:

$$\begin{aligned}
 \gamma_1 &= Cov[x_t, x_{t-1}] = E[x_t, x_{t-1}] \\
 &= E[(\varepsilon_t - \theta_1 \varepsilon_{t-1})(\varepsilon_{t-1} - \theta_1 \varepsilon_{t-2})] \\
 &= E[\varepsilon_t \varepsilon_{t-1} - \theta_1 \varepsilon_t \varepsilon_{t-2} - \theta_1 \varepsilon_{t-1}^2 + \theta_1^2 \varepsilon_{t-1} \varepsilon_{t-2}] \\
 &= -\sigma_\varepsilon^2 \theta_1.
 \end{aligned}$$

$$\begin{aligned}
\gamma_2 &= E [(\varepsilon_t - \theta_1 \varepsilon_{t-1}) (\varepsilon_{t-2} - \theta_1 \varepsilon_{t-3})] \\
&= E [\varepsilon_t \varepsilon_{t-2} - \theta_1 \varepsilon_t \varepsilon_{t-3} - \theta_1 \varepsilon_{t-1} \varepsilon_{t-2} + \theta_1^2 \varepsilon_{t-1} \varepsilon_{t-3}] \\
&= 0. \\
\gamma_k &= 0 \quad \forall k > 1.
\end{aligned}$$

A partir de lo anterior, la función de autocorrelación simple responde a:

$$\rho_k = \begin{cases} \frac{-\theta_1}{(1+\theta_1^2)} & \text{para } k = 1 \\ 0 & \text{para } k > 1 \end{cases}.$$

En cuanto a la función de autocorrelación parcial puede ser obtenida a partir de la función de autocorrelación simple y responde a:

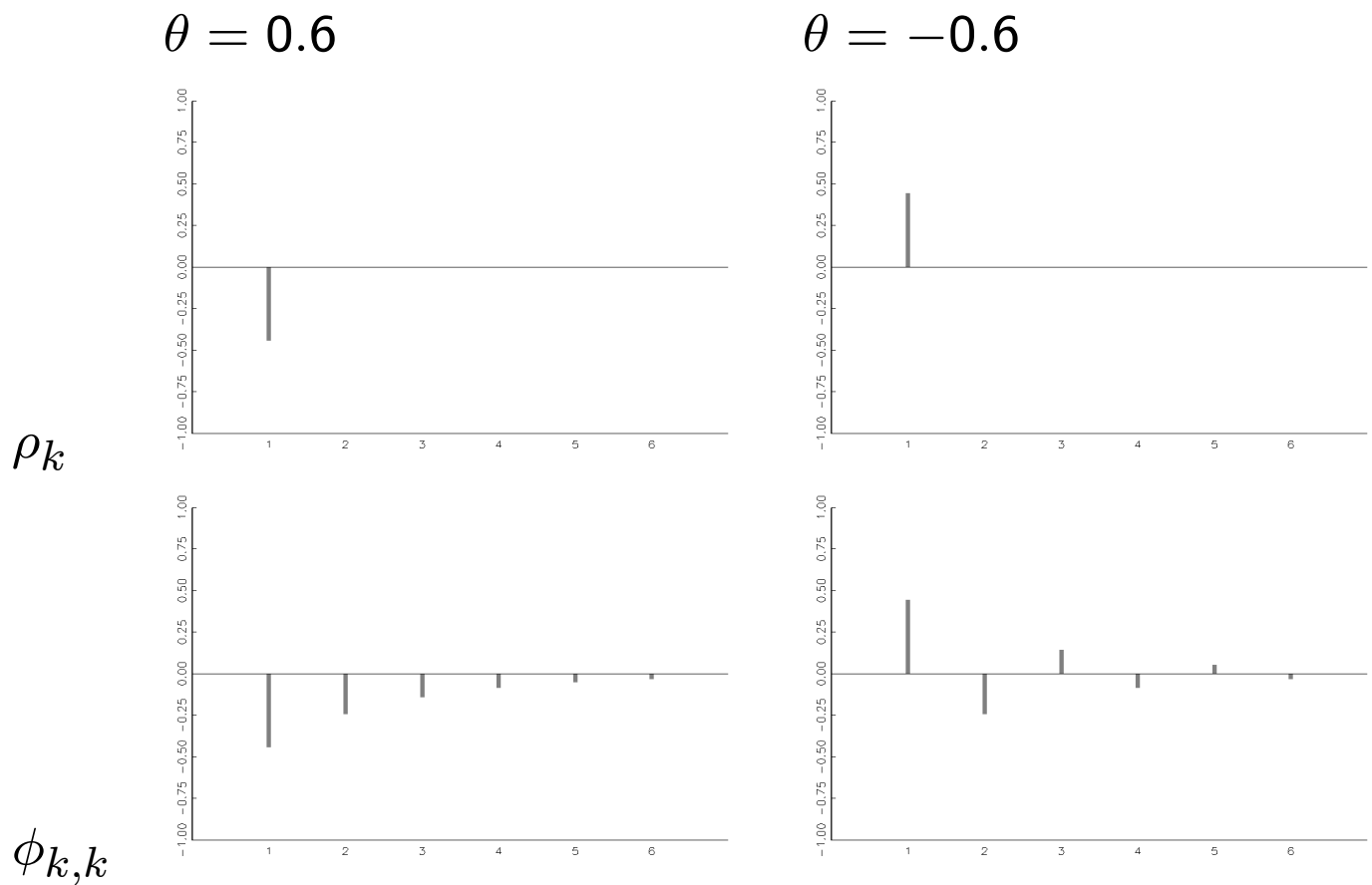
$$\phi_{kk} = \frac{-\theta_1^k (1 - \theta_1^2)}{(1 - \theta_1^{2(k+1)})}.$$

Utilizando la función generatriz de autocovarianzas para un MA(1) tenemos:

$$\begin{aligned}\gamma(L) &= \sigma_{\varepsilon}^2 \theta_1(L) \theta_1(F) \\ &= \sigma_{\varepsilon}^2 (1 - \theta_1 L) (1 - \theta_1 F) \\ &= \sigma_{\varepsilon}^2 (1 + \theta_1^2) + (L + F) [-\theta_1] \sigma_{\varepsilon}^2.\end{aligned}$$

Entonces llegamos a:

$$\begin{aligned}\gamma_0 &= \sigma_{\varepsilon}^2 (1 + \theta_1^2) \\ \gamma_1 &= -\theta_1 \sigma_{\varepsilon}^2 \\ \gamma_j &= 0 \quad j > 1.\end{aligned}$$



Resumiendo un proceso MA(1) se caracteriza por tener el primer coeficiente de autocorrelación distinto de cero y los coeficientes de autocorrelación parcial decrecen exponencialmente en valor absoluto hacia cero.

Para finalizar con el proceso MA(1) se puede comprobar que un proceso MA(1) invertible existe otro no invertible que es observacionalmente equivalente en el sentido de

que tienen la misma FAS y la FAP:

$$\begin{aligned}x_t &= (1 - \theta_1 L) \varepsilon_t \\x_t &= \left(1 - \frac{1}{\theta_1} L\right) \varepsilon_t.\end{aligned}$$

por eso prestaremos más atención a la no invertibilidad de los procesos MA cuando las raíces del polinomio  $\theta(L) = 0$  sean iguales a la unidad dado que serán indicio de sobre diferenciación.

### 2.2.2 MA(2)

Un proceso MA(2) responde al esquema:

$$\begin{aligned}x_t &= \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} \\x_t &= \varepsilon_t - \theta_1 L \varepsilon_t - \theta_2 L^2 \varepsilon_t \\x_t &= (1 - \theta_1 L - \theta_2 L^2) \varepsilon_t \\x_t &= \theta_2(L) \varepsilon_t\end{aligned}$$

donde :

$$\theta_2(L) = (1 - \theta_1 L - \theta_2 L^2).$$

Al calcular las raíces del polinomio  $(1 - \theta_1 L - \theta_2 L^2) = 0$ , tendríamos dos soluciones:

$$L_1 = \frac{-\theta_1 + \sqrt{\theta_1^2 + 4\theta_2}}{2\theta_2}$$
$$L_2 = \frac{-\theta_1 - \sqrt{\theta_1^2 + 4\theta_2}}{2\theta_2}$$

y se puede deducir que las condiciones que han de cumplir los parámetros para que el polinomio sea estable simultaneamente son:

$$\begin{aligned}\theta_2 - \theta_1 &< 1 \\ \theta_2 + \theta_1 &< 1 \\ |\theta_2| &< 1.\end{aligned}$$

En cuanto al valor esperado se puede ver que:



$$\begin{aligned}
E[x_t] &= E[\varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2}] \\
&= E[\varepsilon_t] - \theta_1 E[\varepsilon_{t-1}] - \theta_2 E[\varepsilon_{t-2}] \\
&= 0.
\end{aligned}$$

La varianza:

$$\begin{aligned}
VAR[x_t] &= VAR[\varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2}] \\
&= VAR[\varepsilon_t] + \theta_1^2 VAR[\varepsilon_{t-1}] + \theta_2^2 VAR[\varepsilon_{t-2}] \\
&= \sigma_\varepsilon^2 (1 + \theta_1^2 + \theta_2^2).
\end{aligned}$$

Las autocovarianzas:

$$\begin{aligned}
\gamma_1 &= Cov[x_t, x_{t-1}] = E[x_t, x_{t-1}] \\
&= E[(\varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2}) \\
&\quad (\varepsilon_{t-1} - \theta_1 \varepsilon_{t-2} - \theta_2 \varepsilon_{t-3})] \\
&= -\sigma_\varepsilon^2 \theta_1 + \sigma_\varepsilon^2 \theta_1 \theta_2.
\end{aligned}$$

$$\begin{aligned}
\gamma_2 &= E [(\varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2}) \\
&\quad (\varepsilon_{t-2} - \theta_1 \varepsilon_{t-3} - \theta_2 \varepsilon_{t-4})] \\
&= -\sigma_\varepsilon^2 \theta_2. \\
\gamma_k &= 0 \quad \forall k > 2.
\end{aligned}$$

Entonces es fácil ver que para la función de autocorrelación simple:

$$\rho_k = \begin{cases} \frac{-\theta_1 + \theta_1 \theta_2}{(1 + \theta_1^2 + \theta_2^2)} \text{ para } k = 1 \\ \frac{-\theta_2}{(1 + \theta_1^2 + \theta_2^2)} \text{ para } k = 2 \\ = 0 \text{ para } k > 2 \end{cases} .$$

Utilizando la función generatriz de autocovarianzas de un MA(2) obtenemos:

$$\begin{aligned}
\gamma(L) &= \sigma_\varepsilon^2 \theta_2(L) \theta_2(F) \\
&= \sigma_\varepsilon^2 (1 - \theta_1 L - \theta_2 L^2) (1 - \theta_1 F - \theta_2 F^2) \\
&= \sigma_\varepsilon^2 (1 + \theta_1^2 + \theta_2^2) + (L + F) [-\theta_1] \sigma_\varepsilon^2.
\end{aligned}$$

Entonces llegamos a:

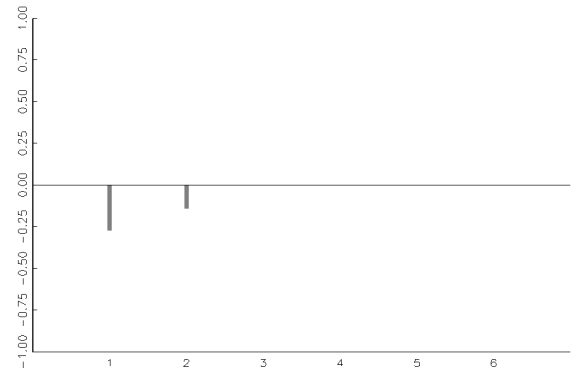
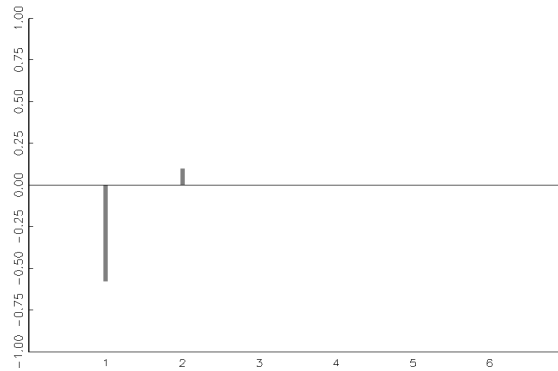
$$\begin{aligned}\gamma_0 &= \sigma_\varepsilon^2 (1 + \theta_1^2) \\ \gamma_1 &= (-\theta_1 + \theta_1\theta_2) \sigma_\varepsilon^2 \\ \gamma_2 &= -\theta_2\sigma_\varepsilon^2 \\ \gamma_j &= 0 \quad j > 2.\end{aligned}$$

Por tanto de forma equivalente al MA(1) un proceso MA(2) se caracteriza por tener los dos primeros coeficientes de la FAS distintos de cero, mientras que los coeficientes de la FAP irán disminuyendo exponencialmente en valor absoluto.

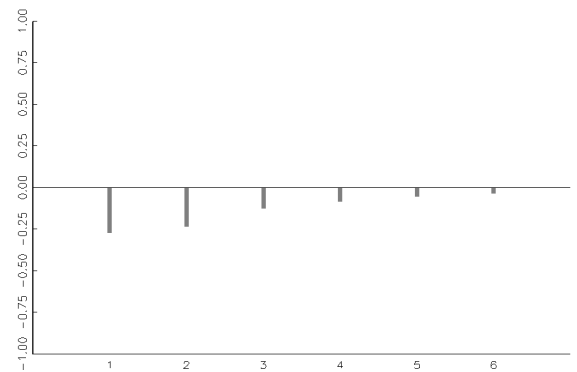
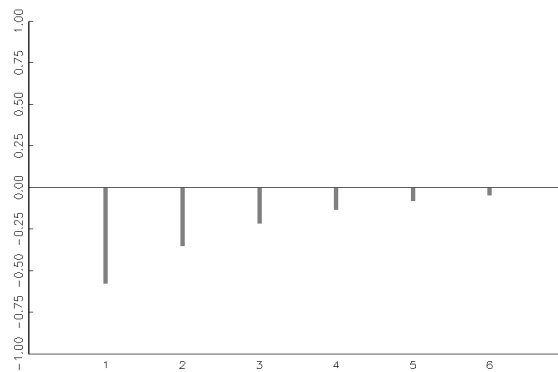
$$\theta_1 = 0.9 \quad \theta_2 = -0.18$$

$$\theta_1 = -1.3 \quad \theta_2 = -0.4$$

$\rho_k$



$\phi_{k,k}$



- Resumiendo los procesos MA(q) se caracterizan por tener una FAP con coeficientes que decrecen exponencialmente en valor absoluto hacia zero y por tener tantos coeficientes distintos de zero en la FAS como orden tenga el proceso. Los procesos más habituales son los de orden 1 y 2, y las condiciones de invertibilidad (estabilidad de los polinimios) son:

$$\text{MA}(1) \quad |\theta_1| < 1$$

$$\text{MA}(2) \quad \theta_2 - \theta_1 < 1; \quad \theta_2 + \theta_1 < 1; \quad |\theta_2| < 1$$

- Si un proceso  $\text{MA}(q)$  es invertible, admite una representación autorregresiva convergente, es decir, es posible expresarlo como un proceso autorregresivo infinito en el que los coeficientes iran decayendo. Para el caso particular de un  $\text{MA}(1)$  es posible escribir:

$$\begin{aligned}
 x_t &= (1 - \theta_1 L) \varepsilon_t \\
 \frac{x_t}{(1 - \theta_1 L)} &= \varepsilon_t \\
 \text{como} \quad &: \\
 \frac{1}{(1 - \theta_1 L)} &= 1 + \theta_1 L + \theta_1^2 L^2 + \dots \\
 &= 1 + \sum_{i=1}^{\infty} \theta_1^i L^i \\
 \text{entonces} \quad &: \\
 \left( 1 + \sum_{i=1}^{\infty} \theta_1^i L^i \right) x_t &= \varepsilon_t.
 \end{aligned}$$

Si se cumple  $|\theta_1| < 1$  se puede ver que llegara un momento en que los coeficientes del proceso  $AR(\infty)$  acabarán siendo prácticamente zero, dado que  $\left(1 + \sum_{i=1}^{\infty} |\theta_1^i|\right) < \infty$ . Es decir, tendrá una representación  $AR(\infty)$  convergente. En cualquier caso un proceso  $MA(q)$  siempre admite una representación  $AR(\infty)$ , pero si el proceso  $MA(q)$  es invetible la representación  $AR(\infty)$  será convergente.

## 2.3 Procesos Autorregresivos $AR(p)$

Una serie sigue un proceso autorregresivo de orden  $p$  si responde al siguiente esquema:

$$\begin{aligned}
x_t &= \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + \varepsilon_t \\
x_t - \phi_1 x_{t-1} - \cdots - \phi_p x_{t-p} &= \varepsilon_t \\
x_t - \phi_1 L x_t - \cdots - \phi_p L^p x_t &= \varepsilon_t \\
(1 - \phi_1 L - \cdots - \phi_p L^p) x_t &= \varepsilon_t \\
\phi_p(L) x_t &= \varepsilon_t \\
\text{donde } &: \\
\phi_p(L) &= (1 - \phi_1 L - \cdots - \phi_p L^p) \\
\varepsilon_t &\sim iid(0, \sigma_\varepsilon^2).
\end{aligned}$$

Para que un proceso AR sea estacionario el polinomio en el operador de retardos  $\phi_p(L)$  asociado al proceso ha de ser estable, es decir, al calcular las raíces del polinomio:

$$\phi_p(L) = (1 - \phi_1 L - \cdots - \phi_p L^p) = 0 \quad (9)$$

estas han de caer fuera del círculo unidad (los valores de  $L$  que satisfacen (9) cumplen  $|L| > 1$ ). Los procesos AR no estacionarios no pueden ser expresados como procesos MA( $\infty$ ) convergentes.

### 2.3.1 AR(1).

Un proceso AR(1) responde al esquema:

$$\begin{aligned}x_t &= \phi_1 x_{t-1} + \varepsilon_t \\x_t - \phi_1 x_{t-1} &= \varepsilon_t \\x_t - \phi_1 L x_t &= \varepsilon_t \\(1 - \phi_1 L) x_t &= \varepsilon_t \\\phi_1(L) x_t &= \varepsilon_t \\\text{donde } &: \\\phi_1(L) &= (1 - \phi_1 L).\end{aligned}$$

Al calcular las raíces del polinomio  $(1 - \phi_1 L)$ , se ve que:

$$\begin{aligned}(1 - \phi_1 L) &= 0 \\L &= 1/\phi_1 \\|\phi_1| < 1 &\Rightarrow |L| > 1.\end{aligned}$$



Por tanto la condición de estacionariedad para un proceso AR(1) es  $|\phi_1| < 1$ . Es fácil ver que un proceso AR(1) se puede reescribir como:

$$\begin{aligned}
 (1 - \phi_1 L) x_t &= \varepsilon_t \\
 x_t &= \frac{\varepsilon_t}{(1 - \phi_1 L)} \\
 \text{como} &: \\
 \frac{1}{(1 - \phi_1 L)} &= 1 + \phi_1 L + \phi_1^2 L^2 + \dots \\
 &= 1 + \sum_{i=1}^{\infty} \phi_1^i L^i \\
 \text{Entonces} &: \\
 x_t &= \varepsilon_t + \sum_{i=1}^{\infty} \phi_1^i \varepsilon_{t-i} \\
 &= \sum_{i=0}^{\infty} \phi_1^i \varepsilon_{t-i}.
 \end{aligned}$$

Por lo que obtendremos un proceso MA( $\infty$ ), que será convergente si el proceso AR(1) es estacionario  $|\phi_1| < 1$ ,

ya que se garantizara que los coeficientes del  $MA(\infty)$  vayan decreciendo en valor absoluto.

En cuanto al valor esperado:

$$\begin{aligned} E[x_t] &= E \left[ \sum_{i=0}^{\infty} \phi_1^i \varepsilon_{t-i} \right] \\ &= E \left[ \varepsilon_t + \phi_1 \varepsilon_{t-1} + \phi_1^2 \varepsilon_{t-2} + \cdots \right] \\ &= E[\varepsilon_t] + \phi_1 E[\varepsilon_{t-1}] + \phi_1^2 E[\varepsilon_{t-2}] + \cdots \\ &= 0. \end{aligned}$$

Para la varianza del proceso es fácil ver:

$$\begin{aligned}
VAR[x_t] &= VAR\left[\sum_{i=0}^{\infty} \phi_1^i \varepsilon_{t-i}\right] = E\left[\left(\sum_{i=0}^{\infty} \phi_1^i \varepsilon_{t-i}\right)^2\right] \\
&= E\left[\left(\varepsilon_t + \phi_1 \varepsilon_{t-1} + \phi_1^2 \varepsilon_{t-2} + \dots\right)^2\right] \\
&= E\left[\varepsilon_t^2\right] + \phi_1^2 E\left[\varepsilon_{t-1}^2\right] + \phi_1^4 E\left[\varepsilon_{t-2}^2\right] + \dots \\
&= \sigma_\varepsilon^2 + \phi_1^2 \sigma_\varepsilon^2 + \phi_1^4 \sigma_\varepsilon^2 + \dots \\
&= \sigma_\varepsilon^2 (1 + \phi_1^2 + \phi_1^4 + \dots) \\
&= \frac{\sigma_\varepsilon^2}{(1 - \phi_1^2)}.
\end{aligned}$$

Si el proceso no es estacionario  $|\phi_1| < 1$  la varianza no está bien definida. En cuanto a las autocovarianzas se puede ver que:

$$\begin{aligned}
x_t &= \phi_1 x_{t-1} + \varepsilon_t \\
x_t x_{t-k} &= \phi_1 x_{t-1} x_{t-k} + \varepsilon_t x_{t-k} \\
cov [x_t x_{t-k}] &= \phi_1 cov [x_{t-1} x_{t-k}] + cov [\varepsilon_t x_{t-k}] \\
\text{como} &: \\
cov [\varepsilon_t x_{t-k}] &= E [\varepsilon_t x_{t-k}] \\
&= E \left[ \varepsilon_t \left( \varepsilon_{t-k} + \phi_1 \varepsilon_{t-k-1} + \phi_1^2 \varepsilon_{t-k-2} + \cdots \right) \right] \\
&= 0 \\
\text{se obtiene} &: \\
\gamma_k &= \phi_1 \gamma_{k-1}.
\end{aligned}$$

Por lo tanto, para los coeficientes de autocorrelación simple tendremos:

$$\begin{aligned}
\frac{\gamma_k}{\gamma_0} &= \phi_1 \frac{\gamma_{k-1}}{\gamma_0} \\
\rho_k &= \phi_1 \rho_{k-1}.
\end{aligned}$$

Resulta interesante ver que:

$$\begin{aligned}
\gamma_1 &= \phi_1 \gamma_0 \\
\gamma_2 &= \phi_1 \gamma_1 = \phi_1^2 \gamma_0 \\
\gamma_3 &= \phi_1 \gamma_3 = \phi_1^3 \gamma_0 \\
&\vdots \\
\gamma_k &= \phi_1^k \gamma_0.
\end{aligned}$$

Entonces:

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\phi_1^k \gamma_0}{\gamma_0} = \phi_1^k.$$

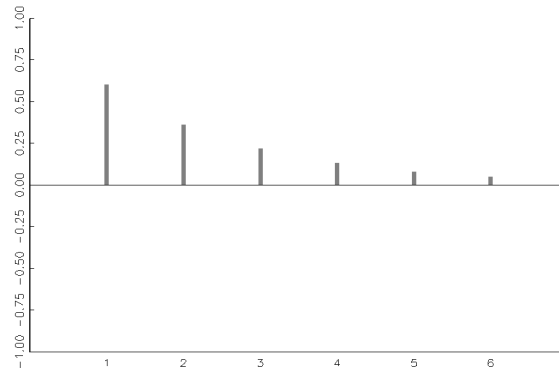
Finalmente utilizando la relación entre los coeficientes de la FAS y la FAP es fácil ver que:

$$\phi_{1,1} = \rho_1 = \phi_1$$

$$\begin{aligned}
\phi_{2,2} &= \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} \\
&= \frac{\begin{vmatrix} 1 & \phi_1 \\ \phi_1 & \phi_1^2 \end{vmatrix}}{\begin{vmatrix} 1 & \phi_1 \\ \phi_1 & 1 \end{vmatrix}} \\
&= \frac{\phi_1^2 - \phi_1^2}{1 - \phi_1^2} = 0 \\
\phi_{k,k} &= 0 \quad \forall k > 1.
\end{aligned}$$

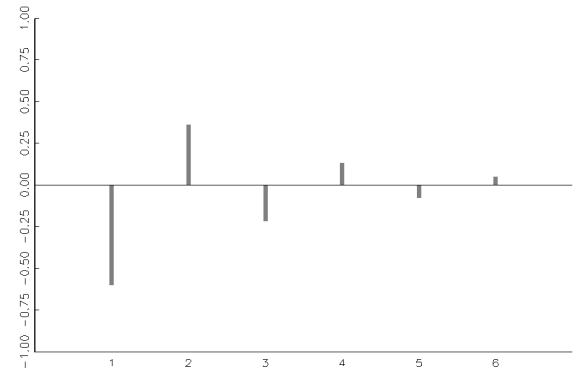
Por lo tanto un proceso AR(1) se caracterizará por tener una FAS con coeficientes que decaen exponencialmente en valor absoluto hacia cero y por tener el primer coeficiente de la FAP distinto de cero.

$$\phi_1 = 0.6$$

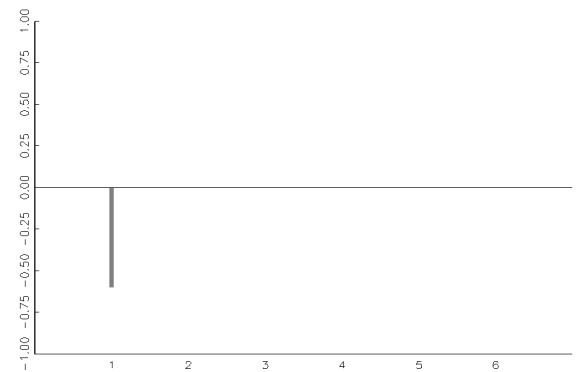
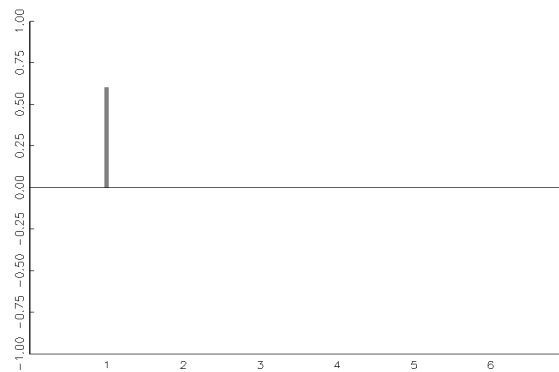


$\rho_k$

$$\phi_1 = -0.6$$



$\phi_{k,k}$



## 2.3.2 AR(2)

Un proceso AR(2) responde al esquema:

$$\begin{aligned}
x_t &= \phi_1 x_{t-1} + \phi_2 x_{t-2} + \varepsilon_t \\
(1 - \phi_1 L - \phi_2 L^2) x_t &= \varepsilon_t \\
\phi_2(L) x_t &= \varepsilon_t \\
\text{donde} \quad &: \\
\phi_2(L) &= (1 - \phi_1 L - \phi_2 L^2).
\end{aligned}$$

Al calcular las raíces del polinomio  $(1 - \phi_1 L - \phi_2 L^2) = 0$ , tendríamos dos soluciones y se puede deducir que las condiciones que han de cumplir los parámetros para que el polinomio sea estable simultaneamente son:

$$\begin{aligned}
\phi_2 - \phi_1 &< 1 \\
\phi_2 + \phi_1 &< 1 \\
|\phi_2| &< 1.
\end{aligned}$$

Finalmente se puede comprobar que los resultados obtenidos en el proceso AR(1) en términos de autocovarianzas y de



coeficientes de autocorrelación simple son generalizables al caso de un AR(p):

$$\begin{aligned}
 x_t &= \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + \varepsilon_t \\
 x_t x_{t-k} &= \phi_1 x_{t-1} x_{t-k} + \cdots + \phi_p x_{t-p} x_{t-k} + \varepsilon_t x_{t-k} \\
 cov [x_t x_{t-k}] &= \phi_1 cov [x_{t-1} x_{t-k}] + \cdots + \phi_p E [x_{t-p} x_{t-k}] \\
 &\quad + cov [\varepsilon_t x_{t-k}] \\
 \gamma_k &= \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \cdots + \phi_p \gamma_{k-p} \\
 \rho_k &= \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \cdots + \phi_p \rho_{k-p}.
 \end{aligned}$$

- Resumiendo los procesos AR(p) se caracterizan por tener una FAS con coeficientes que decrecen exponencialmente en valor absoluto hacia zero y por tener tantos coeficientes distintos de zero en la FAP como orden tenga el proceso. Los procesos más habituales son los de orden 1 y 2, y las condiciones de estacionariedad (estabilidad de los polinimios) son:

$$\text{AR}(1) \quad |\phi_1| < 1$$

$$\text{AR}(2) \quad \phi_2 - \phi_1 < 1; \quad \phi_2 + \phi_1 < 1; \quad |\phi_2| < 1.$$

## 2.4 Procesos ARMA.

Una serie sigue un proceso ARMA(p,q) cuando responde al siguiente esquema:

$$\left(1 - \phi_1 L - \dots - \phi_p L^p\right) x_t = \left(1 - \theta_1 L - \dots - \theta_q L^q\right) \varepsilon_t$$

$$\phi_p(L) x_t = \theta_q(L) \varepsilon_t$$

donde :

$$\phi_p(L) = \left(1 - \phi_1 L - \dots - \phi_p L^p\right)$$

$$\theta_q(L) = \left(1 - \theta_1 L - \dots - \theta_q L^q\right)$$

Es decir, presenta un parte autorregresiva de orden p y una de medias móviles de orden q. Para este tipo de procesos es preciso comprobar si la parte AR(p) es estacionaria y si la parte MA(q) es invertible. Los procesos ARMA se pueden expresar como procesos MA( $\infty$ ):

$$x_t = \frac{\theta_q(L)}{\phi_p(L)} \varepsilon_t = \psi_\infty(L) \varepsilon_t$$

donde :

$$\psi_\infty(L) = (1 + \psi_1 L + \psi_2 L^2 + \dots).$$

Si la parte AR(p) es estacionaria, entonces el proceso ARMA tendrá una representación MA( $\infty$ ) convergente. De forma equivalente los procesos ARMA se pueden expresar como procesos AR( $\infty$ ):

$$\begin{aligned} \frac{\phi_p(L)}{\theta_q(L)} x_t &= \varepsilon_t \\ \pi_\infty(L) x_t &= \varepsilon_t \end{aligned}$$

donde :

$$\begin{aligned} \pi_\infty(L) &= (1 + \pi_1 L + \pi_2 L^2 + \dots) \\ \pi_\infty(L) &= \psi_\infty(L)^{-1}. \end{aligned}$$

Si la parte MA(q) es invertible el proceso ARMA tendrá una representación AR( $\infty$ ) convergente.

### 2.4.1 Cálculo del ratio entre polinomios

supongamos el caso siguiente:

$$\frac{(1 - \theta_1 L - \theta_2 L^2)}{(1 - \phi_1 L)} = \psi_\infty(L)$$

$$(1 - \theta_1 L - \theta_2 L^2) = (1 + \psi_1 L + \psi_2 L^2 + \psi_3 L^3 + \dots) \times (1 - \phi_1 L)$$

$$\begin{aligned} (1 - \theta_1 L - \theta_2 L^2) &= \\ &= (1 + [\psi_1 - \phi_1] L + [\psi_2 - \phi_1 \psi_1] L^2 + \\ &\quad [\psi_3 - \phi_1 \psi_2] L^3 + \dots) . \end{aligned}$$

Como tenemos una identidad ambos polinimios deben tener los mismos coeficientes:

$$\begin{aligned} -\theta_1 &= \psi_1 - \phi_1 \\ -\theta_2 &= \psi_2 - \phi_1 \psi_1 \\ 0 &= \psi_3 - \phi_1 \psi_2 \\ 0 &= \psi_4 - \phi_1 \psi_3 \\ &\vdots \end{aligned}$$

y por lo tanto tenemos:

$$\begin{aligned}\psi_1 &= -\theta_1 + \phi_1 \\ \psi_2 &= -\theta_2 + \phi_1\psi_1 \\ \psi_3 &= \phi_1\psi_2 \\ \psi_4 &= \phi_1\psi_3 \\ &\vdots\end{aligned}$$

## 2.4.2 Proceso ARMA(1,1)

Un proceso ARMA(1,1) responde a:

$$\begin{aligned}(1 - \phi_1 L) x_t &= (1 - \theta_1 L) \varepsilon_t \\ x_t &= \phi_1 x_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1}, \quad (10)\end{aligned}$$

que podemos escribir como un  $MA(\infty)$ :

$$x_t = \frac{(1 - \theta_1 L)}{(1 - \phi_1 L)} \varepsilon_t = \psi_\infty(L) \varepsilon_t$$

cuyos coeficientes se pueden obtener a partir de:

$$\begin{aligned}(1 - \theta_1 L) &= \psi_\infty(L) (1 - \phi_1 L) \\(1 - \theta_1 L) &= \left(1 + [\psi_1 - \phi_1] L + [\psi_2 - \phi_1 \psi_1] L^2 + \right. \\&\quad \left. [\psi_3 - \phi_1 \psi_2] L^3 + \dots \right)\end{aligned}$$

$$\begin{aligned}-\theta_1 &= \psi_1 - \phi_1 \\0 &= \psi_2 - \phi_1 \psi_1 \\0 &= \psi_3 - \phi_1 \psi_2 \\&\vdots\end{aligned}$$

$$\begin{aligned}\psi_1 &= -\theta_1 + \phi_1 \\\psi_2 &= \phi_1 (-\theta_1 + \phi_1) \\\psi_3 &= \phi_1^2 (-\theta_1 + \phi_1) \\&\vdots \\\psi_j &= \phi_1^j (-\theta_1 + \phi_1).\end{aligned}$$

Como se puede ver para que los coeficientes  $\psi_j$  acaben siendo cada vez más pequeños a media que el retardo

crece  $j$  lo importante es que  $|\phi_1| < 1$ . De forma equivalente la representación  $AR(\infty)$ : cuyos coeficientes se pueden obtener a partir de:

$$\pi_{\infty}(L) x_t = \frac{(1 - \phi_1 L)}{(1 - \theta_1 L)} x_t = \varepsilon_t$$

$$\begin{aligned} \pi_{\infty}(L) (1 - \theta_1 L) &= (1 - \phi_1 L) \\ (1 + [\pi_1 - \theta_1] L + [\pi_2 - \theta_1 \pi_1] L^2 + \dots) &= (1 - \phi_1 L) \end{aligned}$$

$$\begin{aligned} [\pi_1 - \theta_1] &= -\phi_1 \\ [\pi_2 - \theta_1 \pi_1] &= 0 \\ [\pi_3 - \theta_1 \pi_2] &= 0 \\ &\vdots \end{aligned}$$

$$\begin{aligned}
\pi_1 &= -\phi_1 + \theta_1 \\
\pi_2 &= \theta_1 (-\phi_1 + \theta_1) \\
\pi_3 &= \theta_1^2 (-\phi_1 + \theta_1) \\
&\vdots \\
\pi_j &= \theta_1^j (-\phi_1 + \theta_1).
\end{aligned}$$

Como se puede ver para que los coeficientes  $\pi_j$  acaben siendo cada vez más pequeños a media que el retardo crece  $j$  lo importante es que  $|\theta_1| < 1$

En cuanto a la FAS de un ARMA(1,1) se puede ver que partiendo de (10) multiplicando por  $x_{t-k}$  y tomando esperanzas:

$$\begin{aligned}
E[x_{t-k}x_t] &= \phi_1 E[x_{t-k}x_{t-1}] + E[x_{t-k}\varepsilon_t] - \theta_1 E[x_{t-k}\varepsilon_{t-1}] \\
\gamma_k &= \phi_1 \gamma_{k-1} + E[x_{t-k}\varepsilon_t] - \theta_1 E[x_{t-k}\varepsilon_{t-1}]. \quad (
\end{aligned}$$



Para  $k = 0$  tenemos:  $x_t = \phi_1 x_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1}$

$$\begin{aligned}\gamma_0 &= \phi_1 \gamma_1 + E[x_t \varepsilon_t] - \theta_1 E[x_t \varepsilon_{t-1}] \\ E[x_t \varepsilon_t] &= \sigma_\varepsilon^2 \\ E[x_t \varepsilon_{t-1}] &= \phi_1 E[x_{t-1} \varepsilon_{t-1}] + E[\varepsilon_t \varepsilon_{t-1}] - \theta_1 E[\varepsilon_{t-1}^2] \\ &= \phi_1 \sigma_\varepsilon^2 - \theta_1 \sigma_\varepsilon^2 \\ \gamma_0 &= \phi_1 \gamma_1 + \sigma_\varepsilon^2 - \theta_1 (\phi_1 - \theta_1) \sigma_\varepsilon^2.\end{aligned}\tag{12}$$

Para  $k = 1$ :

$$\begin{aligned}\gamma_1 &= \phi_1 \gamma_0 + E[x_{t-1} \varepsilon_t] - \theta_1 E[x_{t-1} \varepsilon_{t-1}] \\ E[x_{t-1} \varepsilon_t] &= 0 \\ E[x_{t-1} \varepsilon_{t-1}] &= \sigma_\varepsilon^2 \\ \gamma_1 &= \phi_1 \gamma_0 - \theta_1 \sigma_\varepsilon^2.\end{aligned}\tag{13}$$

Sustituyendo (13) en (12) tenemos:

$$\begin{aligned}\gamma_0 &= \phi_1^2 \gamma_0 - \phi_1 \theta_1 \sigma_\varepsilon^2 + \sigma_\varepsilon^2 - \theta_1 (\phi_1 - \theta_1) \sigma_\varepsilon^2 \\ \gamma_0 &= \frac{(1 - \theta_1^2 - 2\phi_1 \theta_1)}{(1 - \phi_1^2)} \sigma_\varepsilon^2.\end{aligned}$$

Luego:

$$\begin{aligned}\gamma_1 &= \phi_1 \frac{(1 - \theta_1^2 - 2\phi_1\theta_1)}{(1 - \phi_1^2)} \sigma_\varepsilon^2 - \theta_1 \sigma_\varepsilon^2. \\ &= \frac{(\phi_1 - \theta_1)(1 - \theta_1\phi_1)}{(1 - \phi_1^2)} \sigma_\varepsilon^2.\end{aligned}$$

A partir de (11) se puede ver que para  $k \geq 2$  tenemos:

$$\gamma_k = \phi_1 \gamma_{k-1} \quad k \geq 2.$$

Por lo tanto:

$$\rho_k = \begin{cases} \frac{(\phi_1 - \theta_1)(1 - \theta_1\phi_1)}{(1 - \theta_1^2 - 2\phi_1\theta_1)} & k = 1 \\ \phi_1 \rho_{k-1} & k > 1. \end{cases}$$

Por lo tanto tiene una FAS como la de un AR(1) cuando el efecto del MA(1) desaparece en el coeficiente de orden uno. Una situación equivalente se da en la FAP en la que se puede observar un decaimiento exponencial cuya forma precisa depende de los valores de los coeficientes  $\phi_1$  y  $\theta_1$ .

Los procesos ARMA, presentan un FAS como la de su parte AR y una FAP como su parte MA. Por lo tanto se

caracterizan por tener FAS y FAP que decrecen exponencialmente en valor absoluto hacia cero. Y no es posible determinar el orden. En cualquier caso el proceso más usual es el ARMA(1,1). Con posterioridad veremos el criterio de parquedad o parsiminio.

El siguiente cuadro es útil de cara a identificar los procesos que sigue una serie:

		FAS	
		Decrece	$1^o \text{ o } 2^o \neq 0$
FAP	Decrece	ARMA(1,1)	MA(1) o MA(2)
	$1^o \text{ o } 2^o \neq 0$	AR(1) o AR(2)	Imposible

## 2.5 Procesos ARIMA(p,d,q)

En primer lugar presentaremos el operador diferencias que se define como:

$$\begin{aligned}\Delta &= (1 - L) \\ \Delta^d &= (1 - L)^d \\ \Delta x_t &= (1 - L) x_t = x_t - x_{t-1}.\end{aligned}$$

Se dice que una serie sigue un proceso ARIMA(p,d,q) cuando despues de aplicarle el operador diferencias  $d$  veces sigue un proceso ARMA(p,q) estacionario e invertible:

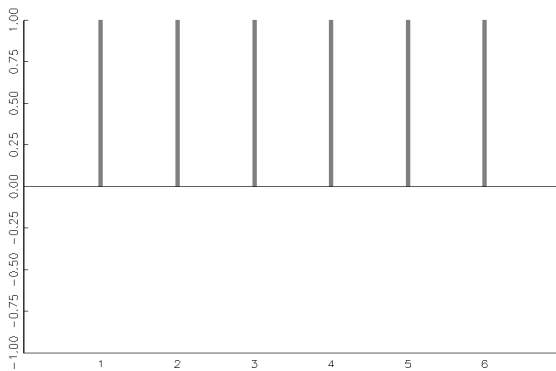
$$\begin{aligned}\phi_p(L) \Delta^d x_t &= \theta_q(L) \varepsilon_t \\ \Delta^d x_t &= z_t \\ \phi_p(L) z_t &= \theta_q(L) \varepsilon_t.\end{aligned}$$

Es decir, un proceso ARIMA(p,d,q) es un proceso no estacionario que necesita que se le aplique el operador diferencias  $d$  veces para que pueda trasformarse en un proceso estacionario. El proceso no estacionario más sencillo que conocemos es el paseo aleatorio (ARIMA(0,1,0)):

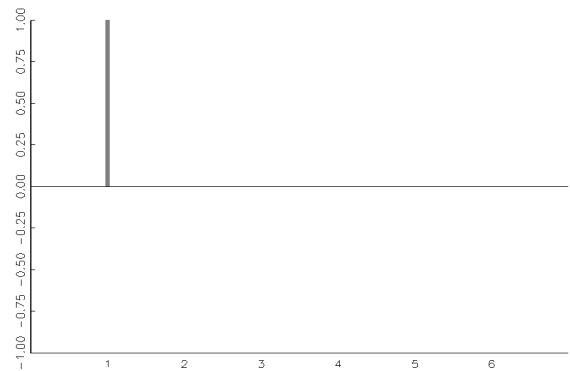
$$\begin{aligned}\Delta x_t &= \varepsilon_t \\ (1 - L) x_t &= \varepsilon_t \\ x_t - x_{t-1} &= \varepsilon_t \\ x_t &= x_{t-1} + \varepsilon_t\end{aligned}$$

La forma de determinar si una serie necesita ser diferenciada para convertirse en estacionaria, es que presente un correlograma como el de un proceso AR(1) con  $\phi = 1$ , es decir, si  $\rho_1$  y  $\phi_{11}$  están muy próximos a la unidad y los coeficientes de la FAS no decaen rápidamente estaremos ante un indicio de no estacionariedad.

$\rho_k$



$\phi_{kk}$

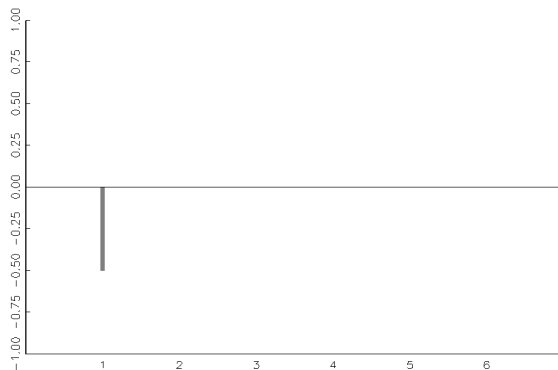


Existe el peligro de sobre-diferenciar, es decir, aplicar el operador  $\Delta$  cuando no es necesario, en este caso es fácil ver que aparecerá un proceso MA(1) no invertible, tal y como se muestra al aplicar  $\Delta$  a un proceso ruido blanco:

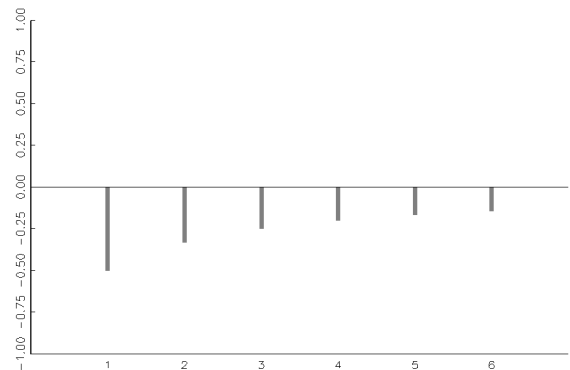
$$\begin{aligned}x_t &= \varepsilon_t \\(1 - L)x_t &= (1 - L)\varepsilon_t \\ \theta_1 &= 1.\end{aligned}$$

Esta situación, es claramente identificable, dado que la serie sobre-diferenciada presentará un esquema demasiado claro de proceso MA(1) donde  $\rho_1 = \phi_{11} = -\theta_1 / (1 + \theta_1^2) = -0.5$ :

$\rho_k$



$\phi_{kk}$



Una vez que se ha determinado el orden de diferenciación, la variable transformada  $\Delta^d x_t = z_t$  seguirá un procesos ARMA(p,q) estacionario e invertible, por lo tanto podrá ser expresado como un AR( $\infty$ ) o un MA( $\infty$ ) convergente:

$$z_t = \frac{\theta_q(L)}{\phi_p(L)} \varepsilon_t = \psi_\infty(L) \varepsilon_t$$

$$\frac{\phi_p(L)}{\theta_q(L)} z_t = \varepsilon_t \Rightarrow \pi_\infty(L) z_t = \varepsilon_t.$$

Los procesos ARIMA(p,d,q) engloban a todos los anteriores, a modo de ejemplo:

$$\begin{aligned} ARIMA(p, 0, 0) &\Rightarrow AR(p) \\ ARIMA(0, 0, q) &\Rightarrow MA(q) \\ ARIMA(p, 0, d) &\Rightarrow ARMA(p, d) \\ ARIMA(0, d, 0) &\Rightarrow I(d). \end{aligned}$$

- Finalmente si un proceso ARIMA tiene parte AR tendremos que comprobar es si estacionaria y si tiene parte MA tendremos que comprobar si es invertible.

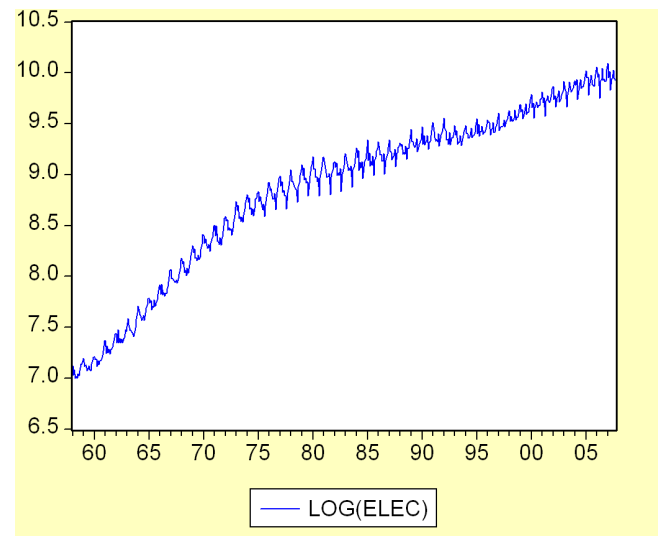
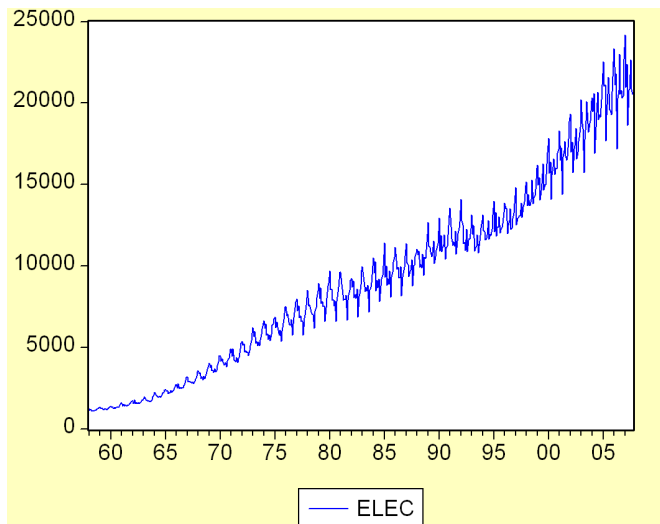
## 2.6 Transformación de Box-Cox

Lo más habitual es trabajar con la serie transformada en logaritmos neperianos, dado que, es un cambio de escala que nos permite homogeneizar la volatilidad que pueda presentar la serie. Esta transformación se corresponde como  $\lambda = 0$ , no transformar la serie se conoce como transformación  $\lambda = 1$ . La transformación de Box-Cox consiste en:

$$x_t^\lambda = \begin{cases} \lambda \neq 0 & \frac{x_t^\lambda - 1}{\lambda} \\ \lambda = 0 & \ln(x_t) \end{cases}$$



En la práctica se trabaja con  $\lambda = 1$ , no transformación y con  $\lambda = 0 \ln(x_t)$ .



### 3 Metodología Box-Jenkins

Box y Jenkins reconocen en su libro las siguientes etapas a la hora de aplicar los modelos ARIMA a las series temporales:

- *Identificación:* Consiste en determinar los ordenes  $p, d, q, P, D$  y  $Q$  y  $\lambda$  del modelo  $ARIMA(p, d, q) \times SARIMA(P, D, Q)$ .
  - Determinamos  $\lambda$  en base al gráfico de la serie.
  - Con el gráfico, la FAS y la FAP muestrales determinar los ordenes de diferenciación  $d$  y  $D$ .
  - Con la FAS y la FAP muestrales determinar  $p, q, P$  y  $Q$ .
  - comprobar que los procesos AR son estacionarios y los MA invertibles. En caso contrario replantear los ordenes de diferenciación.
- *Estimación:* Una vez determinado el modelo  $ARIMA(p, d, q) \times SARIMA(P, D, Q)$ . estimarlo.
  - comprobar que los procesos AR son estacionarios y los MA invertibles. En caso contrario replantear los ordenes de diferenciación.

- comprobar si los estimadores asociados a los parámetros son significativos.
- *Chequeo*: comprobar que los residuos del modelo estimado se comportan como ruido blanco.
  - Inspección de los correlogramas y utilización de los contrastes  $Q$  de Box-Pierce y Ljung-Box.
  - Selección entre modelos alternativos mediante criterios de información AIC y SBC.
- *Predicción*

### 3.1 Identificación

- Tanto a la hora de identificar, como de estimar un Modelo ARIMA, se utilizará el criterio de parsimonia o parquedad, es decir, será preferible el modelo con el menor número de parametros.

Brockwell y Davis (2002) muestran para el caso de un proceso  $MA(q)$  que la distribución asintótica de de los coeficiente de autocorrelación simple para  $k > q$  responde a:

$$\hat{\rho}_k \overset{A}{\sim} N \left( 0, \frac{1}{T} \left( 1 + 2 \sum_{j=1}^q \rho_j^2 \right) \right) \quad k > q.$$

Basandóse en lo anterior algunos programas (STATA, SCA, ... ) usan como intervalo de confianza para la FAS y la FAP:

$$FAS \quad \pm 1.96 \sqrt{\frac{1}{T} \left( 1 + 2 \sum_{j=1}^{k-1} \hat{\rho}_j^2 \right)}$$

$$FAP \quad \pm 1.96 \sqrt{\frac{1}{T}}.$$

Mientras que otros programas (EViews, GRETL, ...) usan como intervalos de confianza:

$$FAS \quad \pm 1.96\sqrt{\frac{1}{T}}$$

$$FAP \quad \pm 1.96\sqrt{\frac{1}{T}}.$$

Los valores fuera de las bandas de confianza pueden ser considerados como estadísticamente diferentes de cero.

## 3.2 Estimación

Por lo que respecta a la estimación de procesos ARMA, se puede ver que dado que estamos ante una serie temporal la función de densidad conjunta de las  $T$  observaciones del proceso  $Y_T = (y_T, y_{T-1}, \dots, y_1)$  puede ser expresada de la siguiente manera:

$$f(Y_T) = f(y_T|y_{T-1}, \dots, y_1) f(y_{T-1}|y_{T-2}, \dots, y_1) \dots \\ \times f(y_3|y_2, y_1) f(y_2|y_1) f(y_1)$$

donde  $f(y_2|y_1)$  es la función de densidad de  $y_2$  condicionada a que previamente hemos tenido  $y_1$ . Si asumimos normalidad tendríamos:

$$f(y_t|y_{t-1}, \dots, y_1) = \left(2\pi\sigma^2 v_{t|t-1}\right)^{-1/2} \\ \times \exp\left(-\frac{1}{2} \frac{(y_t - y_{t|t-1})^2}{\sigma^2 v_{t|t-1}}\right)$$

$$E[y_t|y_{t-1}, \dots, y_1] = y_{t|t-1}$$

$$VAR[y_t|y_{t-1}, \dots, y_1] = \sigma^2 v_{t|t-1}$$

y podremos escribir para la función de densidad conjunta:

$$f(Y_T) = \prod_{t=1}^T \sigma^{-1} v_{t|t-1}^{-1/2} (2\pi)^{-1/2} \exp\left(-\frac{1}{2} \sum_{t=1}^T \frac{(y_t - y_{t|t-1})^2}{\sigma^2 v_{t|t-1}}\right)$$

Y para la función soporte (el logaritmo neperiano de la función de Verosimilitud) tendremos:

$$\ln L = -\frac{T}{2} \ln(\sigma^2) - \frac{1}{2} \sum_{t=1}^T \ln(v_{t|t-1}) - \frac{1}{2\sigma^2} \sum_{t=1}^T \frac{e_{t|t-1}^2}{v_{t|t-1}} \quad (14)$$

donde  $e_{t|t-1} = y_t - y_{t|t-1}$  es el error de predicción un período hacia adelante. Tanto las varianzas condicionales  $v_{t|t-1}$  como los errores de predicción un período hacia adelante  $e_{t|t-1}$  dependen de los parámetros del proceso. La maximización de la función soporte se realiza mediante algoritmo de optimización no lineal y también se emplea el filtro de Kalman dado que permite evaluar de forma rápida la función de verosimilitud de cualquier proceso ARMA calculando los errores de predicción a un periodo  $e_{t|t-1}$  y las varianzas condicionales. Los anteriores procedimientos para poder ser utilizados necesitan valores iniciales de los parámetros de interés del proceso ARMA(p,q). Un método sencillo para obtener valores iniciales es el **algoritmo de Hannan y Rissanen**. Este

algoritmo parte de la idea de que un proceso ARMA se puede expresar como un proceso  $AR(\infty)$ :

$$\begin{aligned}\phi_p(L) y_t &= \theta_q(L) \varepsilon_t & \frac{\phi_p(L)}{\theta_q(L)} y_t &= \varepsilon_t \\ \pi_\infty(L) y_t &= \varepsilon_t & \left(1 + \pi_1 L + \pi_2 L^2 + \pi_3 L^3 + \dots\right) y_t &= \varepsilon_t\end{aligned}$$

Si el proceso es invertible admitirá una representación  $AR(\infty)$  convergente. Es decir, llegará un momento en que los coeficientes  $\pi_j$  serán muy pequeños. Entonces se trunca el polinomio  $\pi_\infty(L)$  a un orden finito  $r$  tal que  $r = \max\{\ln(T-d)/2, 2 \max\{p, q\}\}$  y se estima el siguiente modelo por MCO:

$$\left(1 + \pi_1 L + \pi_2 L^2 + \dots + \pi_r L^r\right) y_t = \varepsilon_t, \quad (15)$$

a partir de los estimadores MCO  $\hat{\pi}_j$  podemos obtener una estimación de los los residuos  $e_t$  que son unos estimadores consistentes de  $\varepsilon_t$  :



$$\left(1 + \hat{\pi}_1 L + \hat{\pi}_2 L^2 + \dots + \hat{\pi}_r L^r\right) y_t = e_t. \quad (16)$$

Con lo que es posible estimar la siguiente regresión por MCO y obtener estimadores de los parámetros de interés

$$y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q} + u_t. \quad (17)$$

De esta forma se puede obtener una primera estimación (o valores iniciales) de los coeficientes de los polinomios  $\phi_p(L)$  y  $\theta_q(L)$ . Este procedimiento se puede repetir de forma iterativa hasta que se alcance un criterio de convergencia.

Una vez se ha obtenido la estimación de los parámetros de interés, y después de haber comprobado las condiciones de estacionariedad e invertibilidad, es interesante realizar los contrastes de significación individual de los parámetros que basándonos en las propiedades de los estimadores MV de los parámetros de interés se puede realizar utilizando:

$$\begin{aligned}
\hat{B} &= \left\{ \hat{\phi}_1 \cdots \hat{\phi}_p, \hat{\theta}_1 \cdots \hat{\theta}_q, \hat{\Phi}_1 \cdots \hat{\Phi}_P, \hat{\Theta}_1 \cdots \hat{\Theta}_Q \right\}' \\
&= \left\{ \hat{\beta}_1 \cdots \hat{\beta}_j \cdots \hat{\beta}_k \right\} \\
\hat{B} &\sim N \left( B, VAR \left( \hat{B} \right) \right) \\
&\quad \begin{cases} H_0 : \beta_j = 0 \\ H_1 : \beta_j \neq 0 \end{cases} \\
t_{\beta_j} &= \frac{\hat{\beta}_j}{\hat{\sigma}_{\hat{\beta}_j}} \sim t_{T-k}.
\end{aligned}$$

El procedimiento de Newton-Raphson de optimización numérica se basa en utilizar recursivamente la siguiente expresión:

$$\hat{B}^{j+1} = \hat{B}^j - \left[ \mathbf{H}^j \right]^{-1} \mathbf{g}^j.$$

Se parte de un vector inicial de estimadores  $\hat{B}^0$  y se va iterando la anterior expresión hasta que la diferencia entre dos estimaciones consecutivas  $(\hat{B}^{j+1} - \hat{B}^j)$  no excede un umbral de tolerancia prefijado. Mientras que  $\mathbf{g}^j$  es el vector de primeras derivas de la cantidad que se desea maximizar o minimizar con respecto a los parámetros

de interés y  $\mathbf{H}^j$  es la matrix de segundas derivadas (ambas evaluadas en los parámetros de interés). Por ejemplo en el caso de un proceso MA(2) maximizar (14) es equivalente a minimizar:

$$\sum \varepsilon_t^2 = \sum (x_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2})^2,$$

por lo que:

$$\begin{aligned} \frac{\partial \sum \varepsilon_t^2}{\partial \theta_1} &= 2 \sum (x_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}) \varepsilon_{t-1} \\ \frac{\partial \sum \varepsilon_t^2}{\partial \theta_2} &= 2 \sum (x_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}) \varepsilon_{t-2} \\ \frac{\partial^2 \sum \varepsilon_t^2}{\partial^2 \theta_1} &= 2 \sum \varepsilon_{t-1}^2 \\ \frac{\partial^2 \sum \varepsilon_t^2}{\partial^2 \theta_2} &= 2 \sum \varepsilon_{t-2}^2 \\ \frac{\partial^2 \sum \varepsilon_t^2}{\partial \theta_1 \partial \theta_2} &= 2 \sum \varepsilon_{t-1} \varepsilon_{t-2}. \end{aligned}$$

En cualquier caso para poder implementarlo se necesita una estimación de  $\varepsilon_t$  que se puede obtener a partir de la representación  $\pi(L)x_t = \varepsilon_t$  en la que los valores de los coeficientes del polinomio  $\pi(L)$  los obtenemos a partir

de los valores de  $\theta_1$  y  $\theta_2$  estimados en la iteración anterior.

### 3.3 Chequeo

Además de examinar los correlogramas para comprobar si nos hemos dejado algo sin especificar en el modelo, es importante concentrarse en los primeros coeficientes de la FAS y la FAP. Existe un estadístico que permite contrastar la siguiente hipótesis nula:

$$\begin{aligned} H_0 &: \rho_1 = \rho_2 = \cdots = \rho_M = 0 \\ H_1 &: \text{No } H_0. \end{aligned}$$

Es decir bajo  $H_0$  los  $M$  primeros coeficientes de autocorrelación simple  $\rho_j$  son iguales a cero, por lo tanto

responden a un esquema de ruido blanco. El estadístico de prueba fue propuesto inicialmente por Box y Pierce siendo igual a:

$$Q_{BP} = T \sum_{j=1}^M \hat{\rho}_j^2 \sim \chi_{M-k}^2.$$

Este contraste está sesgado hacia el no rechazo de la  $H_0$  para muestras pequeñas. Por lo que **Ljung y Box** propusieron la alternativa:

$$Q_{LB} = T(T+2) \sum_{j=1}^M (T-j)^{-1} \hat{\rho}_j^2 \sim \chi_{M-k}^2.$$

La regla de decisión consiste en rechazar  $H_0$  siempre que el estadístico de prueba sea mayor o igual que el valor de tablas  $\chi_{M-k,1-\alpha}^2$ , para un nivel de significación  $\alpha$ .

Finalmente para elegir entre modelos alternativos la práctica habitual es utilizar los criterios de información de AIC y SC:

$$\begin{aligned}AIC &= -2\frac{\ln L}{T} + 2\frac{k}{T} = \ln(\hat{\sigma}^2) + 2\frac{k}{T} \\SC &= -2\frac{\ln L}{T} + 2\frac{k \ln(T)}{T} = \ln(\hat{\sigma}^2) + 2\frac{k \ln(T)}{T}.\end{aligned}$$

Es preferible el modelo con el AIC o el SC más bajo (teniendo en cuenta le signo).

### **3.4 Predicción puntual y por intervalo.**

Centremos en un modelo ARIMA:

$$\begin{aligned}
\phi_p(L) \Delta^d x_t &= \theta_q(L) \varepsilon_t \\
\varphi(L) x_t &= \theta_q(L) \varepsilon_t \\
\varphi(L) &= (1 - \varphi_1 L - \dots - \varphi_{p+d} L^{p+d}) \\
\theta_q(L) &= (1 - \theta_1 L - \dots - \theta_q L^q)
\end{aligned}$$

Por lo tanto podemos escribir:

$$\begin{aligned}
x_t &= \varphi_1 x_{t-1} + \dots + \varphi_{p+d} x_{t-(p+d)} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \quad (18) \\
\psi(L) &= \varphi(L)^{-1} \theta_q(L) \\
x_t &= \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \dots
\end{aligned}$$

La predicción puntual con información hasta el período  $T$ ,  $h$  períodos hacia adelante, se denota  $\hat{x}_T(h)$ , el criterio utilizado para obtener un predictor óptimo es minimizar:

$$E \left[ (\hat{x}_T(h) - x_{T+h})^2 | I_T \right]$$

$$I_T = (x_T, x_{T-1}, \dots, x_1).$$

Se puede demostrar que reponde a:

$$\hat{x}_T(h) = E[x_{T+h} | I_T]$$

Entonces para obtener  $\hat{x}_T(h)$  tomamos como punto de partida (18) y asumiendo que:

- los parámetros son conocidos, a efectos prácticos los sustituimos por sus estimadores,
- que conocemos  $\varepsilon_t$  para el periodo muestral, y que para el período extramuestral se comporta como un ruido blanco:



$$E \left[ \varepsilon_{T+j} | I_T \right] = \begin{cases} \varepsilon_{T+j} = x_{T+j} - \hat{x}_{T+j-1}(1) & j \leq 0 \\ 0 & j > 0 \end{cases}$$

Entonces tendremos:

$$\begin{aligned} \hat{x}_T(h) &= \varphi_1 \hat{x}_T(h-1) + \cdots + \varphi_{h-1} \hat{x}_T(1) + \\ &\quad \varphi_h x_T + \cdots + \varphi_{p+d} x_{T+h-p-d} + \\ &\quad \theta_h \varepsilon_T - \theta_{h+1} \varepsilon_{T-1} - \cdots - \theta_q \varepsilon_{T+h-q} \\ &= \varphi_1 \hat{x}_T(h-1) + \cdots + \varphi_{h-1} \hat{x}_T(1) + \\ &\quad \varphi_h x_T + \cdots + \varphi_{p+d} x_{T+h-p-d} + \\ &\quad \theta_h (x_T - \hat{x}_{T-1}(1)) - \theta_{h+1} (x_{T-1} - \hat{x}_{T-2}(1)) \\ &\quad - \cdots - \theta_q (x_{T+h-q} - \hat{x}_{T+h-q-1}(1)). \end{aligned}$$

En la siguiente tabla se recoge como se calcularían las predicciones puntuales a 1 y 2 períodos para cuatro casos sencillos:

	$\hat{x}_T (1)$
MA(1)	$-\theta_1 (x_T - \hat{x}_{T-1} (1))$
AR(1)	$\phi_1 x_T$
IMA(1,1)	$x_T - \theta_1 (x_T - \hat{x}_{T-1} (1))$
ARI(1,1)	$x_T + \phi_1 x_T - \phi_1 x_{T-1}$

	$\hat{x}_T (2)$
MA(1)	0
AR(1)	$\phi_1 \hat{x}_T (1)$
IMA(1,1)	$\hat{x}_T (1)$
ARI(1,1)	$\hat{x}_T (1) + \phi_1 \hat{x}_T (1) - \phi_1 x_T$

Finalmente en cuanto a la predicción por intervalo, partimos del error de predicción que definimos como:

$$e_T (h) = x_{T+h} - \hat{x}_T (h)$$

donde :

$$x_{T+h} = \varepsilon_{T+h} + \psi_1 \varepsilon_{T+h-1} + \cdots \\ + \psi_{h-1} \varepsilon_{T+1} + \psi_h \varepsilon_T + \cdots$$

$$\hat{x}_T (h) = \psi_h \varepsilon_T + \psi_{h+1} \varepsilon_{T-1} + \psi_{h+2} \varepsilon_{T-2} + \cdots$$

entonces :

$$e_T (h) = \varepsilon_{T+h} + \psi_1 \varepsilon_{T+h-1} + \cdots + \psi_{h-1} \varepsilon_{T+1}$$

Bajo el supuesto de normalidad de las innovaciones  $\varepsilon_t$ , es posible llegar a que:

$$\begin{aligned}\varepsilon_{T+h} &\sim N(0, \sigma_\varepsilon^2) \\ e_T(h) &\sim N\left(B, \sigma_\varepsilon^2 \left[1 + \sum_{i=1}^{h-1} \psi_i^2\right]\right).\end{aligned}$$

Entonces sólo queda tipificar y tener en cuenta que  $e_T(h) = x_{T+h} - \hat{x}_T(h)$  para obtener un intervalo de confianza en torno a  $x_{T+h}$ .

$$\frac{x_{T+h} - \hat{x}_T(h)}{sd(\varepsilon_{T+h})} \sim N(0, 1)$$

$$\begin{aligned}&P\left(\hat{x}_T(h) - z_{1-\frac{\alpha}{2}}sd(\varepsilon_{T+h}) \leq x_{T+h}\right. \\&\quad \left.\leq \hat{x}_T(h) + z_{1-\frac{\alpha}{2}}sd(\varepsilon_{T+h})\right) \\&= 1 - \alpha\end{aligned}$$

donde:  $sd(\varepsilon_{T+h}) = \sqrt{\frac{2}{\varepsilon} \left[ 1 + \sum_{i=1}^{h-1} \psi_i^2 \right]}$

Es interesante ver como el intervalo se va ampliando a medida que se incrementa el valor de  $h$ .