

MODELOS UNIVARIANTES LINEALES NO ESTACIONARIOS

Tere García Muñoz (tgarciam@ugr.es)
Román Salmerón Gómez (romansg@ugr.es)
Alessio Gaggero (alessiogaggero@ugr.es)

Universidad de Granada
Departamento de Métodos Cuantitativos para la Economía y la Empresa

Tema 3– Econometría 3
Grado en Económicas

Una serie temporal es no estacionaria en media cuando tiene tendencia creciente o decreciente, o cambios de nivel. Es decir, una serie no estacionaria no se mueve de forma estable alrededor de una constante.

Ejemplo

Paseo aleatorio: $Y_t = Y_{t-1} + \varepsilon_t$, (AR(1) con $\delta = 0$ y $\phi = 1$)



Figura: Gráfico de un paseo aleatorio

Otra pista de que la serie no es estacionaria se puede encontrar en la FAC porque decrece de forma muy lenta.

Ejemplo

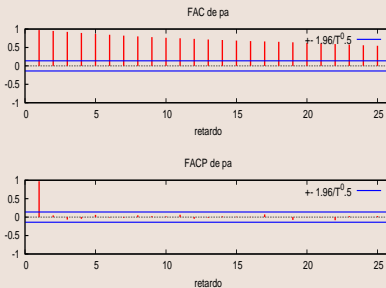


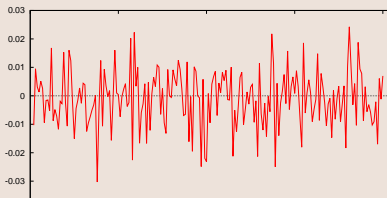
Figura: FAC y FACP de un paseo aleatorio

La tendencia puede eliminarse tomando diferencias regulares: $\nabla^d Y_t$
En la práctica, a lo sumo es necesario una ($d = 1$) o dos ($d = 2$) diferencias regulares para obtener un comportamiento estable.

$$\nabla^d Y_t \rightarrow ARMA(p, q) \Leftrightarrow Y_t \rightarrow ARIMA(p, d, q)$$

Ejemplo

Sobre el paseo aleatorio: $Y_t = Y_{t-1} + \varepsilon_t$. Dado que $\phi = 1$, se sabe que no es un proceso estacionario pero su primera diferencia: $\nabla Y_t = \varepsilon_t \Leftrightarrow Y_t - Y_{t-1} = \varepsilon_t$, es un ruido blanco y, por tanto, estacionario.



Ejemplo

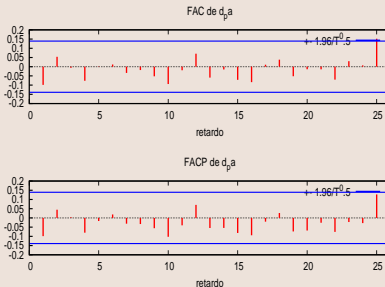


Figura: FAC y FACP de las primeras diferencias

Si una serie es no estacionaria en varianza, el procedimiento habitual para convertirla en estacionaria consiste en transformar la variable tomando logaritmos y pasar a analizar la variable transformada, $\ln Y_t$, en lugar de la original, Y_t .

Ejemplo

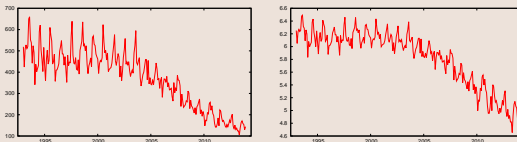


Figura: Número de fallecidos desde Enero de 1993 a Diciembre de 2013 y su logaritmo

Es importante destacar que esta transformación además de inducir estacionariedad en varianza también induce normalidad en los datos en algunos casos. Si es necesario inducir estacionariedad en media y varianza, primero se aplicarán

Estacionalidad

- Modelos ARIMA multiplicativos
- Estimación de los modelos ARIMA
- Diagnos del modelo
- Selección de modelos
- Predicción puntual en modelos ARIMA
- Actualización de predicciones en modelos ARIMA
- Predicción mediante intervalos de confianza en modelos ARIMA

Proceso estacional autorregresivo

Proceso estacional de media móvil

Proceso estacional autorregresivo y de media móvil

Cuando la serie temporal presenta una pauta de comportamiento que se repite con una periodicidad inferior a un año (meses, trimestres, cuatrimestres,...) se dice que la serie presenta **estacionalidad**.

A la periodicidad de los datos se le denota por s ($s = 12$ para meses, $s = 4$ para trimestres, $s = 3$ para cuatrimestres y $s = 2$ para semestres).

Durante la modelización de un proceso ARIMA se debe tener en cuenta el comportamiento estacional ya que si, por ejemplo, se trabaja con datos mensuales, las observaciones cada 12 meses estarán correladas. Por tanto, además de estudiar el comportamiento regular de la serie (relación entre observaciones sucesivas) hay que tener en cuenta la relación lineal existente entre observaciones del mismo mes en años consecutivos (comportamiento estacional).

La identificación y especificación del modelo para la estructura estacional se hace exactamente igual a la descrita hasta ahora, con la novedad de que en la FAC y FACP hay que prestar especial atención a los múltiplos de s : $s, 2s, 3s, 4s, \dots$

Sea el proceso estacional puro autorregresivo de orden 1 con periodicidad s , $AR(1)_s$:

$$Y_t = \Phi Y_{t-s} + \varepsilon_t \Leftrightarrow (1 - \Phi B^s) Y_t = \varepsilon_t,$$

donde ε_t es ruido blanco.

Al ser un proceso autorregresivo es invertible para cualquier valor de Φ y será estacionario si y solo si las s raíces del polinomio $(1 - \Phi B^s)$ son, en modulo, mayores que la unidad. Suponiendo que el proceso es estacionario, se obtiene:

Media: $\mu = E[Y_t] = 0$

Varianza:

$$\begin{aligned} \gamma_0 &= \text{Var}(Y_t) = E[Y_t^2] = \Phi^2 E[Y_{t-s}^2] + E[\varepsilon_t^2] + 2\Phi E[Y_{t-s}\varepsilon_t] \\ &= \Phi^2 \gamma_0 + \sigma_\varepsilon^2 \\ \gamma_0 &= \frac{\sigma_\varepsilon^2}{1 - \Phi^2}. \end{aligned}$$

Función de autocovarianzas:

$$\gamma_s = E[Y_t Y_{t-s}] = \Phi E[Y_{t-s} Y_{t-s}] + E[\varepsilon_t Y_{t-s}] = \Phi \gamma_0,$$

$$\gamma_{2s} = E[Y_t Y_{t-2s}] = \Phi E[Y_{t-s} Y_{t-2s}] + E[\varepsilon_t Y_{t-2s}] = \Phi \gamma_s = \Phi^2 \gamma_0,$$

$$\gamma_{js} = \Phi^j \gamma_0.$$

Dado que Y_t se relaciona solo con las variables retardadas Y_{t-s}, Y_{t-2s}, \dots para los retardos distintos de s o sus múltiplos se tiene que:

$$\begin{aligned} h \neq s \cdot j, \quad \gamma_h &= E[Y_t Y_{t-h}] = E[(\Phi Y_{t-s} + \varepsilon_t) Y_{t-h}] \\ &= \Phi E[Y_{t-s} Y_{t-h}] + E[\varepsilon_t Y_{t-h}] = 0. \end{aligned}$$

Función de autocorrelación simple:

Dado que $\rho_j = \frac{\gamma_j}{\gamma_0}$, la FAC, para $j = 1, 2, 3, \dots$, será:

$$\rho_0 = 1, \quad \rho_h = \Phi^j, \quad h = s \cdot j, \quad \rho_h = 0, \quad h \neq s \cdot j.$$

Función de autocorrelación parcial:

Atendiendo a la expresión del modelo $AR(1)_s$ parece claro que la única influencia directa sobre Y_t la tiene Y_{t-s} , luego el único valor no nulo de la FACP será para $j = s$:

$$\rho_s^p = \rho_s = \Phi, \quad \rho_j^p = 0, \quad \forall j \neq s.$$

Para el **proceso** $AR(P)_s$:

$$Y_t = \Phi_1 Y_{t-s} + \Phi_2 Y_{t-2s} + \dots + \Phi_P Y_{t-Ps} + \varepsilon_t,$$

la FAC decrece con valores no nulos en los retardos estacionales, mientras que la FACP será nula excepto para los valores ρ_h^p con $h = s, 2s, \dots, Ps$.

Sea el proceso estacional de medias moviles de orden 1 con periodicidad s , $MA(1)_s$:

$$Y_t = \varepsilon_t + \Theta\varepsilon_{t-s} \Leftrightarrow Y_t = (1 + \Theta B^s)\varepsilon_t,$$

donde ε_t es ruido blanco.

Como es un proceso de medias moviles es estacionario. El proceso será invertible si las s raíces del polinomio $(1 + \Theta B^s)$ son, en módulo, mayores que uno. Sus principales características son:

Media: $\mu = E[Y_t] = 0$

Varianza:

$$\gamma_0 = \text{Var}(Y_t) = E[Y_t^2] = E[\varepsilon_t^2] + \Theta^2 E[\varepsilon_{t-s}^2] + 2\Theta E[\varepsilon_t \varepsilon_{t-s}] = (1 + \Theta^2)\sigma_\varepsilon^2$$

Función de autocovarianzas:

$$\begin{aligned}
 h \neq s, \quad \gamma_h &= E[Y_t Y_{t-h}] = E[(\varepsilon_t + \Theta \varepsilon_{t-s})(\varepsilon_{t-h} + \Theta \varepsilon_{t-s-h})] \\
 &= E[\varepsilon_t \varepsilon_{t-h} + \Theta \varepsilon_t \varepsilon_{t-s-h} + \Theta \varepsilon_{t-s} \varepsilon_{t-h} + \Theta^2 \varepsilon_{t-s} \varepsilon_{t-s-h}] = 0, \\
 h = s, \quad \gamma_h &= E[Y_t Y_{t-h}] = E[(\varepsilon_t + \Theta \varepsilon_{t-s})(\varepsilon_{t-h} + \Theta \varepsilon_{t-s-h})] \\
 &= E[\varepsilon_t \varepsilon_{t-s} + \Theta \varepsilon_t \varepsilon_{t-2s} + \Theta \varepsilon_{t-s}^2 + \Theta^2 \varepsilon_{t-s} \varepsilon_{t-2s}] = \Theta \sigma_\varepsilon^2.
 \end{aligned}$$

Función de autocorrelación simple

$$\rho_0 = 1, \quad \rho_h = \frac{\Theta}{1 + \Theta^2}, \quad h = s, \quad \rho_h = 0, \quad h \neq s.$$

Función de autocorrelación parcial: Teniendo en cuenta que el inverso del polinomio $(1 + \Theta B^s)$ es:

$$\frac{1}{1 + \Theta B^s} = 1 + \Theta B^s + \Theta^2 B^{2s} + \dots$$

el modelo $MA(1)_s$ puede expresarse como un $AR(\infty)_s$:

$$(1 + \Theta B^s + \Theta^2 B^{2s} + \dots) Y_{t-js} = \varepsilon_t \Leftrightarrow Y_t = -\Theta Y_{t-s} - \Theta^2 Y_{t-2s} - \Theta^3 Y_{t-3s} - \dots + \varepsilon_t,$$

y entonces la FACP será nula excepto para los valores

$$\rho_h^p = -\Theta^j, h = s \cdot j, j = 1, 2, 3, \dots$$

Para el **proceso** $MA(Q)_s$:

$$Y_t = \varepsilon_t + \Theta_1 \varepsilon_{t-s} + \Theta_2 \varepsilon_{t-2s} + \dots + \Theta_Q \varepsilon_{t-Qs},$$

la FAC será nula excepto en los retardos estacionales, mientras que la FACP decrece con valores no nulos en los retardos estacionales.

Estacionalidad

Modelos ARIMA multiplicativos
 Estimación de los modelos ARIMA
 Diagnóstico del modelo
 Selección de modelos
 Predicción puntual en modelos ARIMA
 Actualización de predicciones en modelos ARIMA
 Predicción mediante intervalos de confianza en modelos ARIMA

Proceso estacional autorregresivo
 Proceso estacional de media móvil
 Proceso estacional autorregresivo y de media móvil

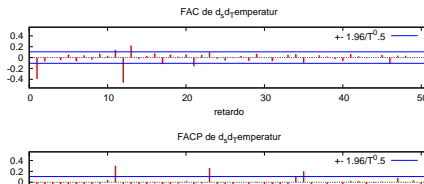
Sea el proceso estacional autorregresivo y de medias móviles de orden (1,1) y periodicidad s , $ARMA(1, 1)_s$:

$$Y_t = \Phi Y_{t-s} + \varepsilon_t + \Theta \varepsilon_{t-s} \Leftrightarrow (1 - \Phi B^s) Y_t = (1 + \Theta B^s) \varepsilon_t,$$

donde ε_t es ruido blanco.

La FAC y FACP son funciones infinitas que decrecen tomando valores no nulos en los retardos estacionales $s, 2s, 3s, \dots$, es decir, son análogas a las de los modelos $ARMA(p, q)$.

A la hora de identificar modelos, la identificación de la parte regular se hace mirando los primeros retardos de los correlogramas y la identificación de la parte estacional mirando los retardos estacionales ($s, 2s, \dots$).



Para eliminar la persistencia estacional fuerte se toman diferencias estacionales:
 $\nabla_s = 1 - B^s$ donde s es el periodo estacional.

La forma más eficiente de incluir la estacionalidad en los modelos de series temporales es de forma multiplicativa, así un proceso estocástico $\{Y_t\}$ sigue un modelo $ARIMA(p, d, q) \times (P, D, Q)_s$ si responde a la siguiente expresión:

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)(1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_P B^{Ps}) \nabla^d \nabla_s^D Y_t = \delta + (1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q)(1 + \Theta_1 B^s + \Theta_2 B^{2s} + \dots + \Theta_Q B^{Qs}) \varepsilon_t,$$

donde:

- $\nabla_s = (1 - B^s)$ es la diferencia estacional y $D = 0, 1$
- $\nabla = (1 - B)$ es la diferencia regular y $d = 0, 1, 2$

El orden en que se apliquen las diferencias regular y estacional es indiferente. Si además se consideran logaritmos para inducir estacionariedad en varianza, el modelo se expresaría de la siguiente manera:

$$\phi(B)\Phi(B^s)\nabla^d \nabla_s^D \ln Y_t = \delta + \theta(B)\Theta(B^s)\varepsilon_t.$$

Tras la identificación y especificación del modelo, el siguiente paso es la estimación de los parámetros del mismo. Con tal objetivo se pueden aplicar distintas técnicas tales como:

- Métodos de los momentos.
- Método de máxima verosimilitud: condicional y exacta (usada por la gran mayoría de paquetes informáticos).
- Estimación no lineal.
- Estimación por Mínimos Cuadrados Ordinarios (cuando el término de error no tenga autocorrelación: modelos AR).

- Los parámetros estimados han de ser estadísticamente distintos de cero.
- Los parámetros estimados han de verificar las condiciones de estacionariedad y/o invertibilidad, es decir, las raíces de los polinomios característicos han de ser, en módulo, mayores que la unidad.
- Sobreparametrización: en un modelo adecuado no deben existir factores en la parte AR ni en la parte MA susceptibles de cancelarse. Por ejemplo,

$$(1 - 1,2B + 0,35B^2)\nabla Y_t = (1 - 0,7B)\epsilon_t$$

$$(1 - 0,7B)(1 - 0,5B)\nabla Y_t = (1 - 0,7B)\epsilon_t.$$

- Si se sobrediferencia (se diferencia la serie más veces de las necesarias), también se sobreparametriza el modelo. Así, si la parte MA no es invertible podría indicar que se ha sobrediferenciado y que el modelo correcto es más sencillo.
- Principio de parsimonia: el modelo ha de tener el mínimo número de parámetros posible.

Los **residuos** resultantes deben comportarse como un proceso de **ruido blanco**.

- La serie de los residuos ha de ser estacionaria en media y en varianza. Así, por ejemplo, si los residuos tienen tendencia (creciente o decreciente) habría que diferenciar alguna vez más. Si no es estacionaria en varianza, habría que replantearse una transformación de Box-Cox:

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

- Las FAC y FACP de los residuos no deben presentar ninguna estructura, es decir, los coeficientes han de ser estadísticamente cero. Si a partir de estas funciones se observa que los residuos presentan estructura autorregresiva habría que reespecificar el modelo teniéndolo en cuenta, igual si presenta una estructura de medias móviles.

- Usar los estadísticos de Ljung-Box o Box-Pierce para analizar si los residuos del modelo son ruido blanco.

$$\left. \begin{array}{l} H_0 : \text{los primeros } m \text{ coeficientes de autocorrelación} \\ \text{son conjuntamente cero} \\ H_1 : \text{existe algún coeficiente de autocorrelación no nulo} \end{array} \right\}.$$

En el caso del contraste de Ljung-Box se rechaza la hipótesis nula si

$$Q_{exp} = T(T+2) \sum_{k=1}^m \frac{\hat{\rho}_k^2}{T-k} > \chi_{m-p-q}^2(\alpha)$$

En el caso del contraste de Box-Pierce se rechaza la hipótesis nula si:

$$Q_{exp} = T \sum_{k=1}^m \hat{\rho}_k^2 > \chi_m^2(\alpha).$$

Tras la diagnosis es posible que se consideren válidos más de un modelo para una misma serie temporal. Para ello se recurre a los criterios de selección de modelos: criterios de información de Akaike (AIC), el bayesiano de Schwarz (BIC) y el de Hannan-Quinn (HQC).

Teniendo en cuenta que la verosimilitud de un modelo viene dada por

$$\mathcal{L} = -\frac{T}{2} \cdot (1 + \ln(2 \cdot \pi) - \ln(T)) - \frac{T}{2} \cdot \ln(SCR)$$

donde T es el número de observaciones disponibles y SCR es la suma de cuadrados de los residuos, los criterios de información se definen como:

- Criterio de Akaike: $AIC = -2 \cdot \mathcal{L} + 2 \cdot p$.
- Criterio de Schwarz: $BIC = -2 \cdot \mathcal{L} + p \cdot \ln(T)$.
- Criterio de Hannan-Quinn: $HQC = -2 \cdot \mathcal{L} + 2 \cdot p \cdot \ln(\ln(T))$.

donde p es el número de parámetros que tiene el modelo y que actúa como un factor de penalización.

Utilizando estos criterios se **escogería aquel modelo con un menor valor de AIC, BIC o HQC.**

Y_1, \dots, Y_T y $\varepsilon_1, \dots, \varepsilon_T$ son conocidas.

$\hat{Y}_T(k)$ = predictor de Y_{T+k} suponiendo información hasta el periodo T

A k se le denomina **horizonte de predicción**.

El predictor que minimiza el error cuadrático medio de predicción $E_T[(Y_{T+k} - \hat{Y}_T(k))^2]$ es la esperanza de la variable Y_{T+k} condicionada a la información disponible:

$$\hat{Y}_T(k) = E[Y_{T+k}/Y_1, \dots, Y_T] \equiv E_T[Y_{T+k}].$$

Adviértase que:

- $E_T[Y_{T+s}] = Y_{T+s}$ para $s \leq 0$ ya que estos valores han sido observados.
- $E_T[Y_{T+s}] = \hat{Y}_T(s)$ para $s > 0$.
- $E_T[\varepsilon_{T+s}] = \varepsilon_{T+s}$ para $s \leq 0$ ya que estos valores son conocidos
- $E_T[\varepsilon_{T+s}] = 0$ para $s > 0$ por ser ε_t ruido blanco.

$$AR(1): Y_t = \delta + \phi Y_{t-1} + \varepsilon_t.$$

$$k = 1: \hat{Y}_T(1) = E_T[Y_{T+1}] = E_T[\delta + \phi Y_T + \varepsilon_{T+1}] = \delta + \phi Y_T.$$

$$\begin{aligned} k = 2: \hat{Y}_T(2) &= E_T[Y_{T+2}] = E_T[\delta + \phi Y_{T+1} + \varepsilon_{T+2}] = \delta + \phi E_T[Y_{T+1}] \\ &= \delta + \phi \hat{Y}_T(1) = \delta + \phi(\delta + \phi Y_T) = \delta(1 + \phi) + \phi^2 Y_T. \end{aligned}$$

$$\begin{aligned} k = 3: \hat{Y}_T(3) &= E_T[Y_{T+3}] = E_T[\delta + \phi Y_{T+2} + \varepsilon_{T+3}] = \delta + \phi E_T[Y_{T+2}] \\ &= \delta + \phi \hat{Y}_T(2) = \delta + \phi(\delta(1 + \phi) + \phi^2 Y_T) \\ &= \delta(1 + \phi + \phi^2) + \phi^3 Y_T \end{aligned}$$

$$k \geq 4: \hat{Y}_T(k) = \delta + \phi \hat{Y}_T(k-1) = \delta(1 + \phi + \phi^2 + \dots + \phi^{k-1}) + \phi^k Y_T.$$

Si el proceso es estacionario ($|\phi| < 1$):

- $\mu = \frac{\delta}{1-\phi}$
- $1 + \phi + \phi^2 + \dots + \phi^{k-1} = \frac{1-\phi^k}{1-\phi}$

Por tanto:

$$\hat{Y}_T(k) = (1 - \phi^k)\mu + \phi^k Y_T.$$

$$\hat{Y}_T(k) = (1 - \phi^k)\mu + \phi^k Y_T.$$

Si el proceso es estacionario:

- La predicción es una combinación lineal de la última observación, Y_T , y la media estacionaria del proceso.
- Según se avanza al futuro (mayor es k), la última observación recibe una ponderación más pequeña ($|\phi| < 1$), por lo que aumenta la incertidumbre con la que se predice.
- Para un horizonte de predicción grande (cuando k tiende a infinito), la predicción será la media del proceso.

$$AR(2): Y_t = \delta + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t.$$

$$\begin{aligned}k = 1: \hat{Y}_T(1) &= E_T[Y_{T+1}] = E_T[\delta + \phi_1 Y_T + \phi_2 Y_{T-1} + \varepsilon_{T+1}] \\ &= \delta + \phi_1 Y_T + \phi_2 Y_{T-1}.\end{aligned}$$

$$\begin{aligned}k = 2: \hat{Y}_T(2) &= E_T[Y_{T+2}] = E_T[\delta + \phi_1 Y_{T+1} + \phi_2 Y_T + \varepsilon_{T+2}] \\ &= \delta + \phi_1 E_T[Y_{T+1}] + \phi_2 Y_T = \delta + \phi_1 \hat{Y}_T(1) + \phi_2 Y_T \\ &= \delta + \phi_1(\delta + \phi_1 Y_T + \phi_2 Y_{T-1}) + \phi_2 Y_T \\ &= \delta(1 + \phi_1) + (\phi_1^2 + \phi_2) Y_T + \phi_1 \phi_2 Y_{T-1}.\end{aligned}$$

$$\begin{aligned}k = 3: \hat{Y}_T(3) &= E_T[Y_{T+3}] = E_T[\delta + \phi_1 Y_{T+2} + \phi_2 Y_{T+1} + \varepsilon_{T+3}] \\ &= \delta + \phi_1 E_T[Y_{T+2}] + \phi_2 E_T[Y_{T+1}] = \delta + \phi_1 \hat{Y}_T(2) + \phi_2 \hat{Y}_T(1) \\ &= \dots \\ &= \delta(1 + \phi_1 + \phi_2 + \phi_1^2) + \phi_1(\phi_1^2 + 2\phi_2) Y_T + (\phi_1^2 \phi_2 + \phi_2^2) Y_{T-1}.\end{aligned}$$

La predicción para cualquier horizonte (valor de k) depende de los dos últimos valores observados. En un $AR(p)$ dependerá de las últimas p observaciones.

$MA(1): Y_t = \delta + \varepsilon_t + \theta\varepsilon_{t-1}.$

$$k = 1 : \hat{Y}_T(1) = E_T[Y_{T+1}] = E_T[\delta + \varepsilon_{T+1} + \theta\varepsilon_T] = \delta + \theta\varepsilon_T.$$

$$k = 2 : \hat{Y}_T(2) = E_T[Y_{T+2}] = E_T[\delta + \varepsilon_{T+2} + \theta\varepsilon_{T+1}] = \delta.$$

$$k \geq 3 : \hat{Y}_T(k) = E_T[Y_{T+k}] = E_T[\delta + \varepsilon_{T+k} + \theta\varepsilon_{T+k-1}] = \delta.$$

Con un $MA(1)$ sólo se puede predecir con un valor distinto a la media de la serie en un período hacia delante.

$$MA(2): Y_t = \delta + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2}.$$

$$\begin{aligned}k = 1 : \hat{Y}_T(1) &= E_T[Y_{T+1}] = E_T[\delta + \varepsilon_{T+1} + \theta_1\varepsilon_T + \theta_2\varepsilon_{T-1}] \\ &= \delta + \theta_1\varepsilon_T + \theta_2\varepsilon_{T-1}.\end{aligned}$$

$$k = 2 : \hat{Y}_T(2) = E_T[Y_{T+2}] = E_T[\delta + \varepsilon_{T+2} + \theta_1\varepsilon_{T+1} + \theta_2\varepsilon_T] = \delta + \theta_2\varepsilon_T.$$

$$k = 3 : \hat{Y}_T(3) = E_T[Y_{T+3}] = E_T[\delta + \varepsilon_{T+3} + \theta_1\varepsilon_{T+2} + \theta_2\varepsilon_{T+1}] = \delta.$$

$$k \geq 3 : \hat{Y}_T(k) = E_T[Y_{T+k}] = E_T[\delta + \varepsilon_{T+k} + \theta_1\varepsilon_{T+k-1} + \theta_2\varepsilon_{T+k-2}] = \delta$$

Con un $MA(2)$ sólo se pueden predecir valores distintos a la media de la serie en uno o dos períodos hacia delante.

En un $MA(q)$ para cualquier horizonte superior a q , la predicción es la media del proceso.

Sea un proceso $ARMA(1, 1)$:

$$Y_t = \delta + \phi Y_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}.$$

La expresión del predictor para los distintos horizontes es:

$$k = 1: \hat{Y}_T(1) = E_T[Y_{T+1}] = E_T[\delta + \phi Y_T + \varepsilon_{T+1} + \theta \varepsilon_T] = \delta + \phi Y_T + \theta \varepsilon_T.$$

$$\begin{aligned}
 k = 2: \hat{Y}_T(2) &= E_T[Y_{T+2}] = E_T[\delta + \phi Y_{T+1} + \varepsilon_{T+2} + \theta \varepsilon_{T+1}] \\
 &= \delta + \phi E_T[Y_{T+1}] = \delta + \phi \hat{Y}_T(1) = \delta(1 + \phi) + \phi^2 Y_T + \phi \theta \varepsilon_T.
 \end{aligned}$$

$$\begin{aligned}
 k = 3: \hat{Y}_T(3) &= E_T[Y_{T+3}] = E_T[\delta + \phi Y_{T+2} + \varepsilon_{T+3} + \theta \varepsilon_{T+2}] \\
 &= \delta + \phi E_T[Y_{T+2}] = \delta + \phi \hat{Y}_T(2) \\
 &= \delta(1 + \phi + \phi^2) + \phi^3 Y_T + \phi^2 \theta \varepsilon_T.
 \end{aligned}$$

$$\begin{aligned}
 k \geq 3: \hat{Y}_T(k) &= E_T[Y_{T+k}] = \delta(1 + \phi + \phi^2 + \dots + \phi^{k-1}) + \phi^k Y_T - \phi^{k-1} \theta \varepsilon_T \\
 &= (1 - \phi^k) \mu + \phi^k Y_T + \phi^{k-1} \theta \varepsilon_T.
 \end{aligned}$$

Si el proceso es estacionario ($|\phi| < 1$), la predicción cuando el horizonte de predicción es muy alto (k tiende a infinito) es igual a la media del proceso.

$$ARIMA(p, d, q) : \phi_p(B)\nabla^d Y_t = \theta_q(B)\varepsilon_t.$$

$$MA(\infty) : Y_t = (\phi_p(B)\nabla^d)^{-1}\theta_q(B)\varepsilon_t.$$

Representación $MA(\infty)$ del proceso Y_t es: $Y_t = \delta + \varepsilon_t + \sum_{s=1}^{+\infty} \psi_s \varepsilon_{t-s}$

Suponiendo que se observa Y_{T+1} , el predictor a horizonte k teniendo información hasta el instante $T + 1$ es:

$$\hat{Y}_{T+1}(k) = \hat{Y}_T(k + 1) + \psi_k (Y_{T+1} - \hat{Y}_T(1)).$$

Para estudiar la calidad de la predicción realizada se analiza el **error de predicción** (diferencia entre el valor de la variable y la predicción realizada). Así, el error de predicción para el periodo $T + k$ sería:

$$e_T(k) = Y_{T+k} - \hat{Y}_T(k).$$

Además, su varianza ofrece una medida del error cometido. Más concretamente, su raíz cuadrada es usada para construir intervalos de confianza para las predicciones realizadas anteriormente mediante la expresión:

$$\hat{Y}_T(k) \pm 1,96\sqrt{\text{Var}(e_T(k))},$$

donde se ha supuesto que el ruido blanco es gaussiano y que el intervalo se ha construido con un 95 % de confianza.

Como se verá a continuación, las amplitudes del intervalo de confianza para las sucesivas predicciones de un modelo AR(p) crecen con el aumento de k , mientras que para un modelo MA(q) permanecen constantes a partir de $k = q + 1$.

$$AR(1) : Y_t = \delta + \phi Y_{t-1} + \varepsilon_t.$$

Error de predicción:

$$k = 1 : e_T(1) = Y_{T+1} - \hat{Y}_T(1) = (\delta + \phi Y_T + \varepsilon_{T+1}) - (\delta + \phi Y_T) = \varepsilon_{T+1}.$$

$$\begin{aligned} k = 2 : e_T(2) &= Y_{T+2} - \hat{Y}_T(2) \\ &= (\delta + \phi(\delta + \phi Y_T + \varepsilon_{T+1}) + \varepsilon_{T+2}) - (\delta(1 + \phi) + \phi^2 Y_T) \\ &= \varepsilon_{T+2} + \phi \varepsilon_{T+1}. \end{aligned}$$

$$\begin{aligned} k > 2 : e_T(k) &= Y_{T+k} - \hat{Y}_T(k) \\ &= \varepsilon_{T+k} + \phi \varepsilon_{T+k-1} + \phi^2 \varepsilon_{T+k-2} + \dots + \phi^{k-1} \varepsilon_{T+1}. \end{aligned}$$

Varianza del error de predicción:

$$k = 1 : \text{Var}(e_T(1)) = \sigma_\varepsilon^2.$$

$$\begin{aligned} k = 2 : \text{Var}(e_T(2)) &= E[e_T(2)^2] = E[\varepsilon_{T+2}^2 + \phi^2 \varepsilon_{T+1}^2 + 2\phi \varepsilon_{T+2} \varepsilon_{T+1}] \\ &= (1 + \phi^2) \sigma_\varepsilon^2. \end{aligned}$$

$$k > 2 : \text{Var}(e_T(k)) = (1 + \phi^2 + \phi^4 + \dots + \phi^{2(k-1)}) \sigma_\varepsilon^2 = \frac{1 - \phi^{2k}}{1 - \phi^2} \sigma_\varepsilon^2.$$

Si el proceso es estacionario, cuando k tiende a infinito la varianza del error de predicción tiende a $\frac{\sigma_\varepsilon^2}{1 - \phi^2}$ que es la varianza de un proceso $AR(1)$.

$$AR(2) : Y_t = \delta + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t.$$

Error de predicción:

$$\begin{aligned} k = 1 : e_T(1) &= Y_{T+1} - \hat{Y}_T(1) \\ &= (\delta + \phi_1 Y_T + \phi_2 Y_{T-1} + \varepsilon_{T+1}) - (\delta + \phi_1 Y_T + \phi_2 Y_{T-1}) = \varepsilon_{T+1}. \end{aligned}$$

$$\begin{aligned} k = 2 : e_T(2) &= Y_{T+2} - \hat{Y}_T(2) \\ &= (\delta + \phi_1(\delta + \phi_1 Y_T + \phi_2 Y_{T-1} + \varepsilon_{T+1}) + \phi_2 Y_T + \varepsilon_{T+2}) - \\ &\quad (\delta(1 + \phi_1) + (\phi_1^2 + \phi_2) Y_T + \phi_1 \phi_2 Y_{T-1}) \\ &= \varepsilon_{T+2} + \phi_1 \varepsilon_{T+1}. \end{aligned}$$

$$k = 3 : e_T(3) = Y_{T+3} - \hat{Y}_T(3) = \dots = \varepsilon_{T+3} + \phi_1 \varepsilon_{T+2} + (\phi_1^2 + \phi_2) \varepsilon_{T+1}.$$

Varianza del error de predicción:

$$k = 1 : \text{Var}(e_T(1)) = \sigma_\varepsilon^2.$$

$$k = 2 : \text{Var}(e_T(2)) = (1 + \phi_1^2) \sigma_\varepsilon^2.$$

$$\begin{aligned} k = 3 : \text{Var}(e_T(3)) &= E[e_T(3)^2] = E[\varepsilon_{T+3}^2] + \phi_1^2 E[\varepsilon_{T+2}^2] + (\phi_1^2 + \phi_2)^2 E[\varepsilon_{T+1}^2] \\ &= (1 + \phi_1^2 + (\phi_1^2 + \phi_2)^2) \sigma_\varepsilon^2. \end{aligned}$$

$$MA(1) : Y_t = \delta + \varepsilon_t + \theta\varepsilon_{t-1}$$

Error de predicción:

$$k = 1 : e_T(1) = Y_{T+1} - \hat{Y}_T(1) = (\delta + \varepsilon_{T+1} + \theta\varepsilon_T) - (\delta + \theta\varepsilon_T) = \varepsilon_{T+1}.$$

$$k = 2 : e_T(2) = Y_{T+2} - \hat{Y}_T(2) = (\delta + \varepsilon_{T+2} + \theta\varepsilon_{T+1}) - \delta = \varepsilon_{T+2} + \theta\varepsilon_{T+1}.$$

$$\begin{aligned} k > 2 : e_T(k) &= Y_{T+k} - \hat{Y}_T(k) = (\delta + \varepsilon_{T+k} + \theta\varepsilon_{T+k-1}) - \delta \\ &= \varepsilon_{T+k} + \theta\varepsilon_{T+k-1}. \end{aligned}$$

Varianza del error de predicción:

$$k = 1 : \text{Var}(e_T(1)) = \sigma_\varepsilon^2.$$

$$\begin{aligned} k = 2 : \text{Var}(e_T(2)) &= E[e_T(2)^2] = E[\varepsilon_{T+2}^2 + \theta^2\varepsilon_{T+1}^2 + 2\theta\varepsilon_{T+1}\varepsilon_{T+2}] \\ &= (1 + \theta^2)\sigma_\varepsilon^2. \end{aligned}$$

$$\begin{aligned} k > 2 : \text{Var}(e_T(k)) &= E[e_T(k)^2] = E[\varepsilon_{T+k}^2 + \theta^2\varepsilon_{T+k-1}^2 + 2\theta\varepsilon_{T+k-1}\varepsilon_{T+k}] \\ &= (1 + \theta^2)\sigma_\varepsilon^2. \end{aligned}$$

La varianza del error de predicción para $k \geq 2$ es la varianza del proceso $MA(1)$, es decir, alcanza su cota un periodo después de su orden.

$$MA(2) : Y_t = \delta + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}$$

Error de predicción:

$$\begin{aligned} k = 1 : e_T(1) &= Y_{T+1} - \hat{Y}_T(1) \\ &= (\delta + \varepsilon_{T+1} + \theta_1 \varepsilon_T + \theta_2 \varepsilon_{T-1}) - (\delta + \theta_1 \varepsilon_T + \theta_2 \varepsilon_{T-2}) = \varepsilon_{T+1}. \end{aligned}$$

$$\begin{aligned} k = 2 : e_T(2) &= Y_{T+2} - \hat{Y}_T(2) \\ &= (\delta + \varepsilon_{T+2} + \theta_1 \varepsilon_{T+1} + \theta_2 \varepsilon_T) - (\delta + \theta_2 \varepsilon_T) = \varepsilon_{T+2} + \theta_1 \varepsilon_{T+1}. \end{aligned}$$

$$\begin{aligned} k > 2 : e_T(k) &= Y_{T+k} - \hat{Y}_T(k) = (\delta + \varepsilon_{T+k} + \theta_1 \varepsilon_{T+k-1} + \theta_2 \varepsilon_{T+k-2}) - \delta \\ &= \varepsilon_{T+k} + \theta_1 \varepsilon_{T+k-1} + \theta_2 \varepsilon_{T+k-2}. \end{aligned}$$

Varianza del error de predicción:

$$k = 1 : \text{Var}(e_T(1)) = \sigma_\varepsilon^2.$$

$$\begin{aligned} k = 2 : \text{Var}(e_T(2)) &= E[e_T(2)^2] = E[\varepsilon_{T+2}^2 + \theta_1^2 \varepsilon_{T+1}^2 + 2\theta_1 \varepsilon_{T+1} \varepsilon_{T+2}] \\ &= (1 + \theta_1^2) \sigma_\varepsilon^2. \end{aligned}$$

$$\begin{aligned} k > 2 : \text{Var}(e_T(k)) &= E[e_T(k)^2] = E[\varepsilon_{T+k}^2] + \theta_1^2 E[\varepsilon_{T+k-1}^2] + \theta_2^2 E[\varepsilon_{T+k-2}^2] \\ &= (1 + \theta_1^2 + \theta_2^2) \sigma_\varepsilon^2. \end{aligned}$$

$$ARMA(1, 1) : Y_t = \delta + \phi Y_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}$$

Error de predicción:

$$\begin{aligned} k = 1 : e_T(1) &= Y_{T+1} - \hat{Y}_T(1) \\ &= (\delta + \phi Y_T + \varepsilon_{T+1} + \theta \varepsilon_T) - (\delta + \phi Y_T + \theta \varepsilon_T) = \varepsilon_{T+1}. \end{aligned}$$

$$\begin{aligned} k = 2 : e_T(2) &= Y_{T+2} - \hat{Y}_T(2) \\ &= (\delta + \phi Y_{T+1} + \varepsilon_{T+2} + \theta \varepsilon_{T+1}) - (\delta(1 + \phi) + \phi^2 Y_T + \phi \theta \varepsilon_T) \\ &= \dots = \varepsilon_{T+2} - (\phi + \theta) \varepsilon_{T+1}. \end{aligned}$$

$$\begin{aligned} k = 3 : e_T(3) &= Y_{T+3} - \hat{Y}_T(3) \\ &= (\delta + \phi Y_{T+2} + \varepsilon_{T+3} + \theta \varepsilon_{T+2}) - (\delta(1 + \phi + \phi^2) + \phi^3 Y_T + \phi^2 \theta \varepsilon_T) \\ &= \dots = \varepsilon_{T+3} + (\phi + \theta) \varepsilon_{T+2} + \phi(\phi + \theta) \varepsilon_{T+1}. \end{aligned}$$

Varianza del error de predicción:

$$k = 1 : \text{Var}(e_T(1)) = \sigma_\varepsilon^2.$$

$$\begin{aligned} k = 2 : \text{Var}(e_T(2)) &= E[\varepsilon_{T+2}^2 + (\phi - \theta)^2 \varepsilon_{T+1}^2 - 2(\phi + \theta)\varepsilon_{T+1}\varepsilon_{T+2}] \\ &= (1 + (\phi + \theta)^2)\sigma_\varepsilon^2. \end{aligned}$$

$$\begin{aligned} k = 3 : \text{Var}(e_T(3)) &= E[\varepsilon_{T+3}^2] + (\phi - \theta)^2 E[\varepsilon_{T+2}^2] + \phi^2(\phi + \theta)^2 E[\varepsilon_{T+1}^2] \\ &= (1 + (\phi + \theta)^2(1 + \phi^2))\sigma_\varepsilon^2. \end{aligned}$$

Calculo de la varianza del error de predicción en cualquier ARIMA

$$\text{MA}(+\infty): Y_t = \delta + \varepsilon_t + \sum_{s=1}^{+\infty} \psi_s \varepsilon_{t-s}$$

$$\text{Error de predicción: } e_T(k) = Y_{T+k} - \hat{Y}_T(k) = \varepsilon_{T+k} + \sum_{s=1}^{k-1} \psi_s \varepsilon_{T+k-s}$$

$$\text{Varianza error de predicción: } \text{Var}(e_T(k)) = (1 + \psi_1^2 + \dots + \psi_{k-1}^2) \sigma_\varepsilon^2$$

Intervalo de confianza al 95 % para la predicción en un proceso ARIMA estacionario:

$$\hat{Y}_T(k) \pm 1,96 \sqrt{\text{Var}(e_T(k))}$$

donde $\text{Var}(e_T(k)) = (1 + \psi_1^2 + \dots + \psi_{k-1}^2) \sigma_\varepsilon^2$ siendo $\psi_1, \dots, \psi_{k-1}$ los primeros $k - 1$ coeficientes de la representación $\text{MA}(+\infty)$ de dicho proceso.

Capacidad predictiva del modelo

Se pueden comparar las predicciones realizadas con el verdadero valor de la observación y obtener así una medida de la capacidad predictiva del modelo. Considerando que de las T observaciones disponibles se usan las m primeras para estimar el modelo, algunos de los estadísticos básicos que pueden usarse con tal objetivo son:

Error Absoluto Medio (EAM):
$$EAM = \frac{1}{T - m} \cdot \sum_{t=T-m+1}^T |Y_t - \hat{Y}_t|.$$

Error Cuadrático Medio (ECM):
$$ECM = \frac{1}{T - m} \cdot \sum_{t=T-m+1}^T (Y_t - \hat{Y}_t)^2.$$

Valores del EAM o ECM próximos a cero son indicativos de una buena capacidad predictiva.