Modelos tradicionales: conceptos iniciales

Tere García Muñoz (Grupos A y B, tgarciam@ugr.es)
Alessio Gaggero (Grupo C, alessiogaggero@ugr.es)

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

Tema 1 – Econometría 3
Grado en Económicas

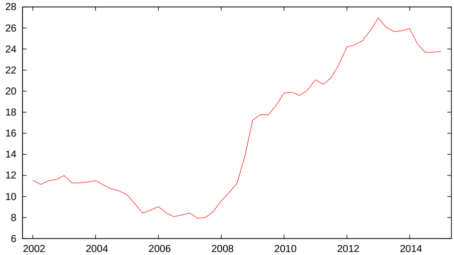
Índice

Introducción

Hasta ahora (en los cursos sobre Econometría previos) se ha tratado la estimación de modelos econométricos explicando el comportamiento de una variable endógena (dependiente o explicada) a partir de un conjunto de variables exógenas (independientes, explicativas o regresores).

En este tema se va a introducir una metodología (la de Box y Jenkins) que es habitualmente usada en el análisis de series económicas y en la que el comportamiento de una variable se explica usando sólo su propio pasado.

Tasa de paro en España desde el primer trimestre de 2002 al primer trimestre de 2015



Introducción

Objetivo

Introducir el análisis de series temporales en tiempo discreto (mediante modelos estocásticos lineales) para explicar la estructura y prever la evolución de una variable observada a lo largo del tiempo.

- Los datos son observados (tomados) en intervalos regulares de tiempo (tiempo discreto: minutos, horas, días, semanas, meses, trimestres, cuatrimestres, semestres, años, etc). Es decir, los datos están ordenados cronológicamente (no son datos de sección cruzada ni de panel).
- El enfoque se denomina univariante ya que se usa únicamente la información que aporta la propia serie. No se consideran variables exógenas.
- Esta metodología, introducida por Box y Jenkins en 1970 con la publicación del libro *"Time Series Analysis: Forecasting and Control"*, trata de predecir la evolución de la serie usando sólo su evolución pasada y es útil para la previsión a corto plazo.

Introducción

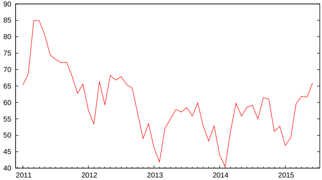
Este enfoque pone más atención a los datos que la Econometría *tradicional*, dejando que sean los datos *los que hablen*. Van a ser fundamentales herramientas no sólo estadísticas sino (como veremos) también gráficas.

Algunos ejemplos:

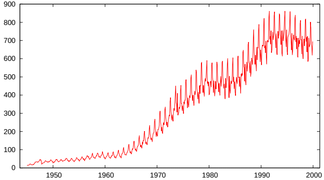
- Variables macroeconómicas como los índices de precios (IPC, vivienda, etc.), tasas de paro, afiliación a la Seguridad Social, PIB o índices bursátiles.
- Variables microeconómicas como las ventas de una empresa, gastos de una familia o precios de un producto.
- Variables ambientales como la velocidad del viento, grado de humedad, temperaturas o precipitaciones en un determinado lugar.
- Variables socio-demográficas: nacimientos, matrimonios, defunciones, separaciones, etc.
- Variables médicas como la presión arterial, tensión, peso o altura de personas.

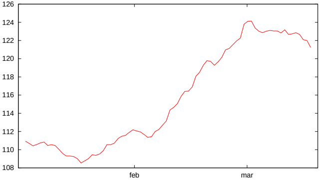
Introducción

Número de pasajeros mensuales aeropuerto de granada de enero 2011 a junio de 2015 y valor mensual del euribor de enero de 2005 a junio de 2015



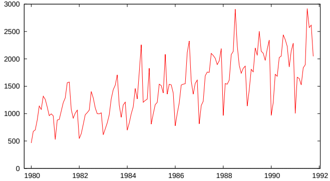
Consumo de gasolina mensual de enero de 1945 a diciembre de 1999 y datos diarios del índice Dow Jones de enero a marzo de 1950

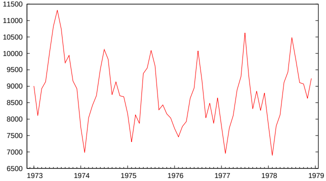




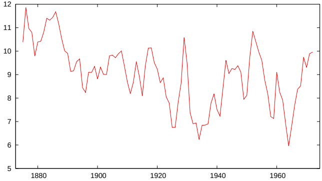
Introducción

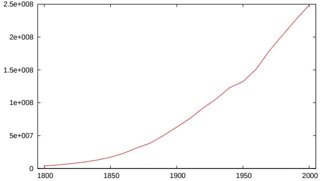
Ventas mensuales de vino tinto en Australia de enero de 1980 a octubre de 1991 y accidentes mortales mensuales de enero de 1973 a diciembre de 1978





Nivel anual del lago Huron de 1875 a 1972 y población en USA de 1800 a 2000 cada 10 años





Índice

Proceso estocástico

Definición

Se define un proceso estocástico como una sucesión de variables aleatorias $\{Y_t\}$, $t \in \mathbb{Z}$.

El subíndice t se interpreta normalmente como el periodo (instante de tiempo) al que corresponde la variable aleatoria Y_t. Generalmente se considera que $t = 1, \ldots, T$.

En las ciencias no experimentales (como la Economía) la definición de proceso estocástico es puramente conceptual: se considera que una serie temporal como una realización de tamaño uno de un proceso estocástico (una única fila de la tabla).

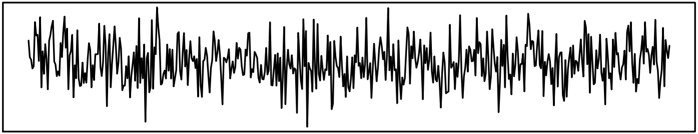
Y_1	Y_2	$\ldots$	Y_T
$Y_1(1)$	$Y_2(1)$	$\ldots$	$Y_T(1)$
$Y_1(2)$	$Y_2(2)$	$\ldots$	$Y_T(2)$
$Y_1(3)$	$Y_2(3)$	$\ldots$	$Y_T(3)$
$\vdots$	$\vdots$		$\vdots$
$Y_1(n)$	$Y_2(n)$	$\ldots$	$Y_T(n)$

Proceso estocástico

Ejemplo

Un **ruido blanco** $\{\varepsilon_t\}$ es un proceso estocastico que se define como una secuencia de variables aleatorias **incorreladas** que tienen **esperanza cero** y **varianza constante**.
Se denomina *blanco* por analogía a la luz blanca y significa que todas las posibles oscilaciones están presentes con la misma intensidad.
Sus oscilaciones son rápidas.

Ruido Blanco



Proceso estocástico

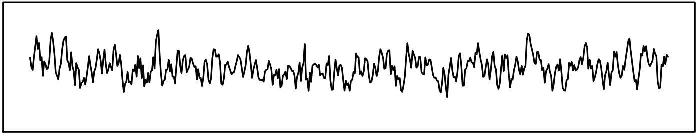
Ejemplo

A partir del ruido blanco anterior creamos un **proceso de medias móviles**:

$$v_t = \frac{1}{3}(\varepsilon_{t-1} + \varepsilon_t + \varepsilon_{t+1})$$

Es una versión suavizada del ruido blanco anterior.
Sus oscilaciones son más lentas.

Media Móvil



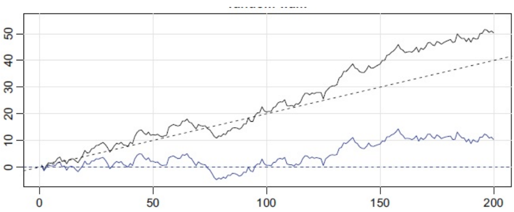
Proceso estocástico

Ejemplo

A partir del ruido blanco anterior creamos un proceso denominado **paseo aleatorio con constante**:

$$x_t = \delta + x_{t-1} + \varepsilon_t$$

Si $\delta = 0$ se le denomina paseo aleatorio. Esta denominación viene del hecho de que el valor en el instante t depende del valor anterior más un movimiento completamente aleatorio determinado por el ruido blanco.



Distribuciones marginales

Definición

Se llama **función de medias** de un proceso a una función del tiempo que proporciona las esperanzas de las distribuciones marginales para cada instante:

$$E[Y_t] = \mu_t, \quad t = 1, ..., T,$$

donde T es el número de observaciones disponibles. Si todas las variables tienen la misma media, entonces la función de medias es constante, las realizaciones de este proceso no mostrarán ninguna tendencia y se dice que es un proceso estable en media.

Función de autocovarianzas

Definición

Se define la función de varianzas de un proceso estocástico $\{Y_t\}$ a la que proporciona las varianzas en cada instante temporal:

$$Var(Y_t) = \sigma_t^2 = E[(Y_t - E(Y_t))^2].$$

Se dice que el proceso es estable en varianza si éstas son constantes en el tiempo.

Definición

Se define la función de autovarianzas de un proceso estocástico $\{Y_t\}$ a la que a las que describe las covarianzas entre dos variables del proceso en dos instantes cualesquiera:

$$Cov(Y_t, Y_{t+k}) = \gamma_{t,t+k} = E[(Y_t - E(Y_t))(Y_{t+k} - E(Y_{t+k}))].$$

Cuando $k = 0$, $\gamma_{t,t} = Var(Y_t)$.

Función de autocorrelación

Definición

Se define la función de autocorrelación simple (FAC) de un proceso estocástico $\{Y_t\}$ como la función que describe las correlación entre dos variables del proceso en dos instantes cualesquiera:

$$\rho_{t,t+k} = \frac{Cov(Y_t, Y_{t+k})}{\sqrt{Var(Y_t)}\sqrt{Var(Y_{t+k})}} = \frac{\gamma_{t,t+k}}{\sqrt{\gamma_{t,t}\gamma_{t+k,t+k}}}.$$

Esta función propociona medidas adimensionales de la dependencia lineal entre variables.
Se mueve entre -1 y 1.

Función de autocorrelación parcial

Definición

Se define la función de autocorrelación parcial (*FACP*) de un proceso estocástico $\{Y_t\}$ como la correlación lineal existente entre dos variables del proceso en dos instantes cualesquiera pero eliminando el efecto que tienen los retardos intermedios sobre ellas:

$$Corr(Y_t, Y_{t+k}/Y_{t+1}, Y_{t+2}, \ldots, Y_{t+k-1}).$$

El primer valor de la *FACP* corresponde a la correlación entre Y_t e Y_{t+1} sin que haya que eliminar la influencia de ningún retardo intermedio (ya que no existen). Por tanto, para $k = 1$ coinciden los valores de la *FAC* y *FACP* de cualquier proceso estocástico.

Índice

Series Temporales

Una serie temporal es una muestra de un proceso estocástico, es una muestra de tamaño 1 de cada una de las T variables que conforman el proceso. Para que una series temporal quede caracterizada debemos conocer las características del proceso estocástico que la genera.

Caracterización de un proceso estocástico: a partir de las funciones de distribución o a partir de los momentos. Puesto que el primer procedimiento es complicado, se usa el segundo a pesar de que lleve a una caracterización más incompleta.

$$E[Y_1],\ E[Y_2],\ \ldots,\ E[Y_T], \begin{pmatrix} Var(Y_1) & Cov(Y_1,Y_2) & \ldots & Cov(Y_1,Y_T) \\ Cov(Y_2,Y_1) & Var(Y_2) & \ldots & Cov(Y_2,Y_T) \\ \vdots & \vdots & & \vdots \\ Cov(Y_T,Y_1) & Cov(Y_T,Y_2) & \ldots & Var(Y_T) \end{pmatrix}$$

¡¡En tal caso hay que estimar T medias distintas y $\frac{T(T+1)}{2}$ elementos distintos de la matriz de varianzas-covarianzas, disponiendo de T observaciones!!

Por tanto, para poder efectuar inferencia sobre un proceso estocástico deben imponerse una serie de restricciones o supuestos. Las habituales son considerar que la serie temporal es **estacionaria** y **ergódica**.

Objetivo

Especificar un modelo estadístico (una ecuación matemática) que explique sus movimientos a partir de la propia serie temporal retardada. El modelo que se obtenga debe cumplir tres requisitos:

- **Describir** los cambios de la serie con respecto al tiempo (tendencia, estacionalidad, variaciones cíclicas y variaciones irregulares).
- **Predecir** los valores futuros.
- **Contrastar** hipótesis sobre las características a la que se refieren los valores de la serie.

Etapas del análisis:

- **Identificación:** Deteminar la especificación del modelo.
- **Estimación:** Estimar los parámetros del modelo especificado en la fase anterior.
- **Validación/Diagnosis:** Analizar la validez del modelo estimado.
- **Predicción:** Realizar predicciones con el/los modelos elegidos y validados.

Etapas del análisis

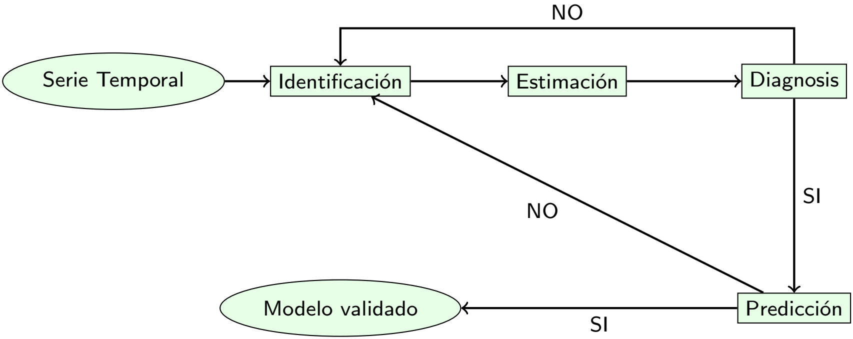


Figura: Fases en la modelización de una serie temporal

Operadores de retardo y de diferencias

Definición

El operador retardo, denotado por B, aplicado a una variable temporal la retarda en un periodo: $BY_t = Y_{t-1}$
$B^k Y_t = Y_{t-k}$ para $k \geq 1$
$Bc = c$, $c =$constante.

Definición

El operador diferencia regular, denotado por ∇, aplicado a una variable temporal la transforma de la siguiente manera: $\nabla Y_t = (1 - B)Y_t = Y_t - Y_{t-1}$
$\nabla^k Y_t = (1 - B)^k Y_t, k \geq 1$

Índice

Ruido Blanco

Definición

Un **ruido blanco** $\{\varepsilon_t\}$ es un proceso estocastico que se define como una secuencia de variables aleatorias **incorreladas** que tienen **esperanza cero** y **varianza constante**:

$$E[\varepsilon_t] = 0$$

$$Var[\varepsilon_t] = E[\varepsilon_t^2] = \sigma_\varepsilon^2$$

$$Cov[\varepsilon_t, \varepsilon_s] = E[\varepsilon_t \varepsilon_s] = 0$$

Principales propiedades:

- No hay correlación entre términos
- Valores pasados no ayudan a pronosticar valores futuros.

Se llamará ruido blanco gaussiano (o normal) si la distribución de $\{\varepsilon_t\}$ es normal.

Índice

Estacionariedad

La estacionariedad conlleva que la serie tenga un comportamiento estable a lo largo del tiempo. Para que esto sea así, una serie temporal debe cumplir las siguientes características:

- No tener tendencia: no presentar un crecimiento o decrecimiento sistemático ni cambios de nivel.
- Ser homocedástica: a lo largo del tiempo las oscilaciones deben tenr una amplitud más o menos constante.
- En caso de existir, deben detectarse comportamientos estacionales para su correcta modelización.
- La estructura de dependencia debe mantenerse constante (si una observación influye en la posterior, que esto ocurra siempre)
- Para explicar el comportamiento de la serie deben tener mayor influencia las observaciones más cercanas.

Estacionariedad

Definición

Un proceso estocástico $\{Y_t\}$ se dice estacionario en sentido estricto si la función de distribución del proceso, F, permanece invariante con respecto al tiempo. Es decir, si se verifica que: $F(Y_{t_1}, Y_{t_2}, \ldots, Y_{t_l}) = F(Y_{t_1+k}, Y_{t_2+k}, \ldots, Y_{t_l+k}), \forall k$.

El concepto de estacionariedad en sentido estricto implica el cumplimiento de condiciones díficil de manejar en la práctica, por tal motivo se suele considerar un concepto menos exigente:

Definición

Un proceso estocástico $\{Y_t\}$ se dice estacionario en sentido débil cuando todos sus momentos de primer y segundo orden son invariantes en el tiempo.
Es decir, $E[Y_t] = \mu$ y $Cov(Y_t, Y_{t+h}) = \gamma_h$, $\forall t, h$.

Bajo el supuesto de estacionariedad, el problema de conocer el proceso estocástico que genera a la serie temporal se reduce a conocer la media, la varianza y las autocovarianzas del proceso.

Función de autocovarianzas

La función de autocovarianzas de un proceso estocástico estacionario es una función del número de periodos de separación entre las variables k:

$$\gamma_{t,t+k} = \gamma_k, k = 0, \pm 1, \pm 2, ...$$

Sus propiedades son:

- $\gamma_0 = Var(Y_k) \geq 0$
- $\gamma_k = \gamma_{-k}$, para todo k entero (la función de autocovarianzas se calcula solo para los valores no negativos del retardo k)
- $|\gamma_k| \leq \gamma_0$, para todo k entero

Función de autocorrelación simple

La FAC de un proceso estacionario depende únicamente del desfase entre los dos instantes:

$$\rho_k = \frac{\gamma_k}{\sqrt{\gamma_0}\sqrt{\gamma_0}} = \frac{\gamma_k}{\gamma_0}$$

Se suele representar mediante un gráfico de barras denominado **correlograma**. Sus principales propiedades son:

- $\rho_0 = 1$
- $\rho_k = \rho_{-k}$ (la FAC se representa únicamente para valores positvos).
- $|\rho_k| \leq 1$
- $\rho_k \to 0$ cuando $k \to \infty$.

Función de autocorrelación parcial

La FACP en el retardo k mide la influencia neta y directa que tendría Y_{t-k} sobre Y_t, es decir, se prescinde de las influencias que sobre Y_t tienen $Y_{t-1}, Y_{t-2}, \ldots, Y_{t-k+1}$. Siguiendo esta definición, la forma de calcular el coeficiente de la FACP en el retardo k es obtener el coeficiente ϕ_{kk} de la siguiente regresión:

$$\tilde{Y}_t = \phi_{k1}\tilde{Y}_{t-1} + \phi_{k2}\tilde{Y}_{t-2} + \cdots + \phi_{kk}\tilde{Y}_{t-k} + v_t,$$

$\tilde{Y}_t = Y_t - E[Y_t]$
v_t es un proceso de media cero incorrelado con Y_{t-j} para todo j.
A partir de esta ecuación de regresión se puede calcular la expresión del coeficiente ϕ_{kk} de la FACP a partir de los coeficientes de la FAC.
Multiplicando a ambos lados de la regresión por $\tilde{Y}_{t-j}$ para $j = 1, ..., k$ se obtiene:

$$\tilde{Y}_t\tilde{Y}_{t-j} = \phi_{k1}\tilde{Y}_{t-1}\tilde{Y}_{t-j} + \phi_{k2}\tilde{Y}_{t-2}\tilde{Y}_{t-j} + \cdots + \phi_{kk}\tilde{Y}_{t-k}\tilde{Y}_{t-j} + v_t\tilde{Y}_{t-j}, \quad j =, 1, 2, \ldots, k,$$

Función de autocorrelación parcial

Tomando esperanzas:

$$\gamma_j = \phi_{k1}\gamma_{j-1} + \phi_{k2}\gamma_{j-2} + \cdots + \phi_{kk}\gamma_{j-k} + \underbrace{E[v_t \tilde{Y}_{t-j}]}_{0}, \quad j =, 1, 2, \ldots, k,$$

y dividiendo por γ_0:

$$\rho_j = \phi_{k1}\rho_{j-1} + \phi_{k2}\rho_{j-2} + \cdots + \phi_{kk}\rho_{j-k}, \quad j =, 1, 2, \ldots, k$$

Teniendo en cuenta que $\rho_0 = 1$ y que $\rho_k = \rho_{-k}$, estas k ecuaciones quedarían como:

$$\begin{aligned}
\rho_1 &= \phi_{k1} + \phi_{k2}\rho_1 + \cdots + \phi_{kk}\rho_{k-1} \\
\rho_2 &= \phi_{k1}\rho_1 + \phi_{k2} + \cdots + \phi_{kk}\rho_{k-2} \\
\vdots & \quad\ \vdots \qquad\quad \vdots \\
\rho_k &= \phi_{k1}\rho_{k-1} + \phi_{k2}\rho_{k-2} + \cdots + \phi_{kk}
\end{aligned}$$

Función de autocorrelación parcial

y utilizando la regla de Cramer se obtienen los valores de ϕ_{kk} para $k = 1, 2, \ldots$ Por ejemplo, denotando por $\rho_k^p (= \phi_{kk})$ a los coeficientes de la FACP, los tres primeros coeficientes se obtendrían de la siguiente manera:

Para $k = 1$: $\rho_1 = \phi_{11} \Rightarrow \rho_1^p = \phi_{11} = \rho_1$

Para $k = 2$:

$$\begin{array}{rcl} \rho_1 & = & \phi_{21} + \phi_{22}\rho_1 \\ \rho_2 & = & \phi_{21}\rho_1 + \phi_{22} \end{array} \Rightarrow \rho_2^p = \phi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}$$

Para $k = 3$:

$$\begin{array}{rcl} \rho_1 & = & \phi_{31} + \phi_{32}\rho_1 + \phi_{33}\rho_2 \\ \rho_2 & = & \phi_{31}\rho_1 + \phi_{32} + \phi_{33}\rho_1 \\ \rho_3 & = & \phi_{31}\rho_2 + \phi_{32}\rho_1 + \phi_{33} \end{array} \Rightarrow \rho_3^p = \phi_{33} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}}$$

Índice

Estimaciones muestrales

Proposición

Bajo estacionariedad en sentido debil, se pueden estimar los primeros momentos poblacionales del proceso mediante los correspondientes momentos muestrales de la serie temporal:

- Media muestral: $\hat{\mu}_Y = \bar{Y} = \frac{1}{T}\sum_{t=1}^{T} Y_t$.
- Varianza muestral: $\hat{\sigma}^2_Y = \hat{\gamma}_0 = \frac{1}{T}\sum_{t=1}^{T}(Y_t - \bar{Y})^2$.
- Autocovarianza muestral en el retardo k:

$$\hat{\gamma}_k = \frac{1}{T-k}\sum_{t=k+1}^{T}(Y_t - \bar{Y})(Y_{t-k} - \bar{Y}), \quad k =, 1, 2, \ldots$$

Estimaciones muestrales

Proposición (continuación

- Coeficiente de autocorrelación muestral en el retardo k:

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} = \frac{\frac{1}{T-k} \sum_{t=k+1}^{T} (Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\frac{1}{T} \sum_{t=1}^{T} (Y_t - \bar{Y})^2}, \quad k =, 1, 2, \ldots$$

Ahora bien, cuando se trabaja con estos estimadores es necesario garantizar que el proceso estocástico que genera la serie verifique unas condiciones tales que los estimadores (generados con la serie temporal) tengan asintóticamente (cuando el número de observaciones es muy alto) las mismas propiedades que los estimadores generados con distintas réplicas del proceso. A dichas condiciones se les denomina condiciones de ergodicidad.

Índice

Ergodicidad

Definición

Se dice que un proceso estacionario en sentido débil es **ergódico respecto de la media** si el estimador $\hat{\mu}_Y$ calculado con la serie temporal para estimar la media del proceso converge en media cuadrática al estimador de la media definido sobre una muestra de réplicas independientes del proceso.

Definición

Se dice que un proceso estacionario en sentido débil es **ergódico respecto de las autocovarianzas** si los estimadores $\hat{\gamma}_k$ calculados con la serie temporal para estimar las autocovarianzas del proceso convergen en media cuadrática a los estimadores de las autocovarianzas definidos sobre una muestra de réplicas independientes del proceso.

Ergodicidad

Si un proceso es estacionario y ergódico, los estimadores presentados anteriormente verifican:

- $\hat{\mu}_Y$ es un estimador consistente (en media cuadrática) de la media del proceso.
- $\hat{\gamma}_0$ es un estimador consistente (en media cuadrática) de la varianza del proceso.
- $\hat{\gamma}_k$ son estimadores consistentes (en media cuadrática) de las autocovarianzas del proceso.
- $\hat{\rho}_k$ son estimadores consistentes (en media cuadrática) de los coeficientes de autocorrelación del proceso.

Cuando aumenta el número de observaciones de una serie temporal surge otro problema: el número de parámetros desconocidos a estimar también aumenta. La ergodicidad conlleva que variables del proceso suficientemente separadas en el tiempo no estén correladas o, dicho de otra manera, la influencia de una variable aleatoria sobre otra muy alejada en el tiempo se puede considerar nula.

Modelos univariantes lineales estacionarios

Tere García Muñoz (Grupos A y B, tgarciam@ugr.es)
Alessio Gaggero (Grupo C, alessiogaggero@ugr.es)

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

Tema 2– Econometría 3
Grado en Económicas

AR(p)

En un proceso autorregresivo la variable en el periodo t es explicada por las observaciones de ella misma en los periodos anteriores más un término de perturbación. La expresión de un proceso autorregresivo de orden p ($AR(p)$) es:

$$Y_t = \delta + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \varepsilon_t$$

donde

- δ es una constante
- $\phi_1, \phi_2, ..., \phi_p$ son denominados parámetros autorregresivos
- ε_t es un ruido blanco

Utilizando el operador retardo B:

$$(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p) Y_t = \delta + \varepsilon_t$$

AR(1)

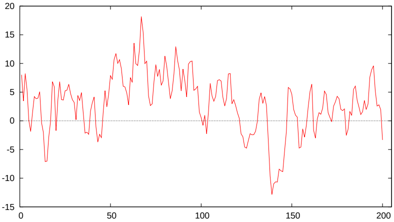
Definición

Un proceso estocástico $\{y_t\}$ sigue un modelo autorregresivo de orden 1, denotado por AR(1), si se puede escribir como

$$Y_t = \delta + \phi Y_{t-1} + \varepsilon_t \Leftrightarrow \phi_p(B) Y_t = \delta + \varepsilon_t$$

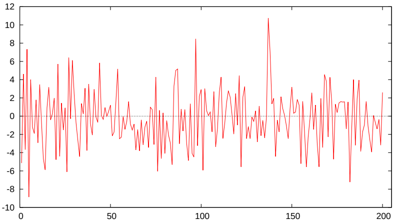
donde δ y ϕ son constantes y ε_t es un ruido blanco.

Representación de un proceso AR(1) con $\delta = 0$, $\phi = 0{,}8$ e $Y_0 = 5$



AR(1)

Representación de un proceso AR(1) con $\delta = 0$, $\phi = -0{,}3$ e $Y_0 = -2$



Proposición

- ε_t está incorrelado con los valores previos del proceso Y_t
- ε_t está correlado con los valores futuros del proceso Y_t

AR(1). Condición de Estacionariedad

Estacionariedad en Media
Si el proceso es estacionario en media se cumple que $E[Y_{t-i}] = \mu$, $\forall i \geq 0$. Entonces:

$$
\begin{aligned}
E[Y_t] &= E[\delta + \phi Y_{t-1} + \varepsilon_t] = \delta + E[\phi Y_{t-1}] + \underbrace{E[\varepsilon_t]}_{0} = \delta + \phi E[Y_{t-1}] \\
&\Rightarrow \mu - \phi\mu = \delta \Rightarrow (1-\phi)\mu = \delta \Rightarrow E[Y_t] = \mu = \frac{\delta}{1-\phi}.
\end{aligned}
$$

Esta expresión es finita si y solo si $\phi \neq 1$.

AR(1). Condición de Estacionariedad

Estacionariedad en Varianza
Si el proceso es estacionario en varianza se cumple que $Var[Y_{t-i}] = \gamma_0$, $\forall i \geq 0$. Entonces:

$$Var[Y_t] = Var[\delta + \phi Y_{t-1} + \varepsilon_t] = Var[\phi Y_{t-1}] + Var[\varepsilon_t] + 2 \cdot \underbrace{Cov(\phi Y_{t-1}, \varepsilon_t)}_{0}.$$

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

$$\gamma_0 > 0, \sigma_\varepsilon^2 > 0 \Leftrightarrow 1 - \phi^2 > 0 \Leftrightarrow |\phi| < 1$$

$$1 - \phi B = 0 \Leftrightarrow B = \frac{1}{\phi} \Leftrightarrow |B| > 1$$

AR(1). Condición de Estacionariedad

Función de autocovarianzas

Para poder operar con mayor facilidad se transformará el modelo como sigue:

$$E[Y_t] = \frac{\delta}{1-\phi} \Rightarrow \delta = E[Y_t](1-\phi).$$

Sustituyendo esta expresión en la ecuación:

$$\begin{aligned} Y_t &= \delta + \phi Y_{t-1} + \varepsilon_t = E[Y_t](1-\phi) + \phi Y_{t-1} + \varepsilon_t \\ &= E[Y_t] + \phi(Y_{t-1} - E[Y_t]) + \varepsilon_t, \end{aligned}$$

y entonces:

$$Y_t - E[Y_t] = \phi(Y_{t-1} - E[Y_t]) + \varepsilon_t. \qquad (1)$$

Si se denota $Y_t - E[Y_t]$ como $\tilde{Y}_t$:

$$\tilde{Y}_t = \phi \tilde{Y}_{t-1} + \varepsilon_t.$$

AR(1). Condición de Estacionariedad

Función de autocovarianzas

Proposición

Los momentos de primer y segundo orden de $\tilde{Y}_t$ verifican:

- $E[\tilde{Y}_t] = 0.$
- $Var[\tilde{Y}_t] = Var[Y_t] = \gamma_0.$
- $\widetilde{\gamma}_k = \gamma_k.$

Por tanto, los momentos de segundo orden de $\tilde{Y}_t$ e Y_t coinciden por lo que la obtención de la función de autocovarianzas de Y_t se puede realizar, de forma más sencilla, a partir de la función de autocovarianzas de $\tilde{Y}_t$.

AR(1). Condición de Estacionariedad

Función de autocovarianzas

$$
\begin{aligned}
\gamma_0 &= Var[Y_t] = Var[\tilde{Y}_t] = E[(\tilde{Y}_t - \underbrace{E[\tilde{Y}_t]}_{0})^2] = E[(\phi\tilde{Y}_{t-1} + \varepsilon_t)^2] \\
&= E[\phi^2\tilde{Y}_{t-1}^2 + 2\phi\tilde{Y}_{t-1}\varepsilon_t + \varepsilon_t^2] \\
&= \phi^2\underbrace{E[\tilde{Y}_{t-1}^2]}_{\gamma_0} + 2\phi\underbrace{E[\tilde{Y}_{t-1}\varepsilon_t]}_{0} + \underbrace{E[\varepsilon_t^2]}_{\sigma_\varepsilon^2} = \phi^2\gamma_0 + \sigma_\varepsilon^2 \\
&\Rightarrow \gamma_0 - \phi^2\gamma_0 = \sigma_\varepsilon^2 \Rightarrow \gamma_0 = \frac{\sigma_\varepsilon^2}{1-\phi^2}. \\
\gamma_1 &= Cov(Y_t, Y_{t-1}) = Cov(\tilde{Y}_t, \tilde{Y}_{t-1}) = E[\tilde{Y}_t\tilde{Y}_{t-1}] - \underbrace{E[\tilde{Y}_t]}_{0}\underbrace{E[\tilde{Y}_{t-1}]}_{0} \\
&= E[(\phi\tilde{Y}_{t-1} + \varepsilon_t)\tilde{Y}_{t-1}] = E[\phi\tilde{Y}_{t-1}^2 + \tilde{Y}_{t-1}\varepsilon_t] \\
&= \phi\underbrace{E[\tilde{Y}_{t-1}^2]}_{\gamma_0} + \underbrace{E[\tilde{Y}_{t-1}\varepsilon_t]}_{0} = \phi\gamma_0.
\end{aligned}
$$

AR(1). Condición de Estacionariedad

Función de autocovarianzas

$$\begin{aligned}
\gamma_2 &= Cov(Y_t, Y_{t-2}) = Cov(\tilde{Y}_t, \tilde{Y}_{t-2}) = E[\tilde{Y}_t \tilde{Y}_{t-2}] - \underbrace{E[\tilde{Y}_t]}_{0}\underbrace{E[\tilde{Y}_{t-2}]}_{0} \\
&= E[(\phi \tilde{Y}_{t-1} + \varepsilon_t)\tilde{Y}_{t-2}] = E[\phi \tilde{Y}_{t-1}\tilde{Y}_{t-2} + \tilde{Y}_{t-2}\varepsilon_t] \\
&= \phi \underbrace{E[\tilde{Y}_{t-1}\tilde{Y}_{t-2}]}_{\gamma_1} + \underbrace{E[\tilde{Y}_{t-2}\varepsilon_t]}_{0} = \phi \underbrace{\gamma_1}_{\phi\gamma_0} = \phi^2 \gamma_0.
\end{aligned}$$

$$\begin{aligned}
\gamma_k &= Cov(Y_t, Y_{t-k}) = Cov(\tilde{Y}_t, \tilde{Y}_{t-k}) = E[\tilde{Y}_t \tilde{Y}_{t-k}] - \underbrace{E[\tilde{Y}_t]}_{0}\underbrace{E[\tilde{Y}_{t-k}]}_{0} \\
&= E[(\phi \tilde{Y}_{t-1} + \varepsilon_t)\tilde{Y}_{t-k}] = E[\phi \tilde{Y}_{t-1}\tilde{Y}_{t-k} + \tilde{Y}_{t-k}\varepsilon_t] \\
&= \phi \underbrace{E[\tilde{Y}_{t-1}\tilde{Y}_{t-k}]}_{\gamma_{k-1}} + \underbrace{E[\tilde{Y}_{t-k}\varepsilon_t]}_{0} = \phi\gamma_{k-1} = \phi^k \gamma_0, \qquad \forall k \geq 1.
\end{aligned}$$

Bajo la condición de estacionariedad obtenida para la varianza ($|\phi| < 1$ ó $|B| > 1$), se tiene que la función de autocovarianzas no depende del tiempo y el proceso es estacionario.

Función de autocorrelación simple

A partir de la función de autocorrelación, γ_k, suponiendo estacionariedad se tiene la siguiente *fas*:

$$\rho_0 = 1, \quad \rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\phi^k \gamma_0}{\gamma_0} = \phi^k, k \geq 1.$$

Y_t está correlada con cualquier variable retardada Y_{t-k}. Por esta razón se dice que el proceso AR(1) tiene memoria infinita. Aunque si el proceso es estacionario, es decir si $|\phi| < 1$, dicha correlación es más débil a medida que aumenta el retardo.

Función de autocorrelación simple

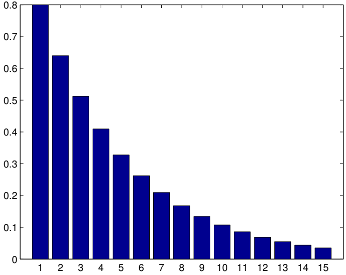


Figura: FAC para $\phi = 0{,}8$

Función de autocorrelación simple

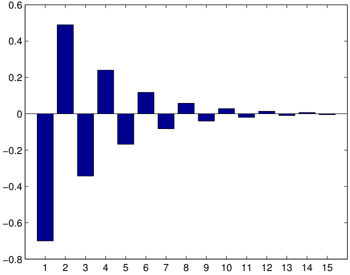


Figura: FAC para $\phi = -0{,}7$

Función de autocorrelación parcial

La FACP mide para cada retardo k el efecto directo de la variable retardada Y_{t-k} sobre Y_t con independencia de los efectos indirectos.

$$Y_t = \delta + \phi Y_{t-1} + \varepsilon_t = Y_t = \delta + \phi(\delta + \phi Y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t$$

Se observa que Y_{t-2} influye a Y_t utilizando como intermediaria a Y_{t-1}, es decir, si Y_{t-1} no existiera Y_{t-2} no podría influir en Y_t. Por tanto, no existe efecto directo de Y_{t-2} sobre Y_t y el coeficiente de autocorrelación de orden 2 sería igual a cero. Es fácil deducir que lo mismo ocurriría con los coeficientes de orden superior a 2.
La FACP para un proceso AR(1) es:

$$\rho_1^p = \rho_1 = \phi, \quad \rho_k^p = 0, \qquad \forall k \geq 2.$$

Principales características de un AR(1) estacionario

- El parametro autoregresivo debe verificar que $|\phi| < 1$
- La *FAC* decae hacia cero (en valor absoluto) de forma rápida.
- La *FACP* se anula para retardos iguales o superiores a 2.

Ejemplo (en el libro)

Dado el proceso $Y_t = 2{,}5 + 0{,}6 \cdot Y_{t-1} + \varepsilon_t$, donde ε_t es ruido blanco, calcular su media, varianza, covarianza, función de autocorrelación simple y los tres primeros coeficientes de la función de autocorrelación parcial.

AR(2)

Definición

Un proceso estocástico $\{Y_t\}$ sigue un modelo autorregresivo de orden 2, denotado por AR(2), si se puede escribir como

$$Y_t = \delta + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \epsilon_t \Leftrightarrow (1 - \phi_1 B - \phi_2 B^2) Y_t = \delta + \varepsilon_t$$

donde δ, ϕ_1 y ϕ_2 son constantes y ε_t es un ruido blanco.

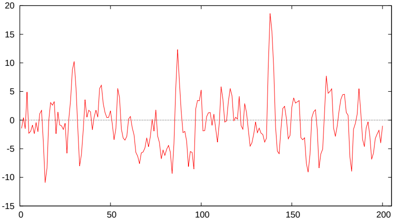


Figura: Proceso AR(2) con ($\phi_1 = 0{,}8$, $\phi_2 = -0{,}3$)

AR(2). Condiciones de Estacionariedad

Estacionariedad en Media
Si el proceso es estacionario en media se cumple que $E[Y_{t-i}] = \mu$, $\forall i \geq 0$. Entonces:

$$\begin{aligned} E[Y_t] &= E[\delta + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t] = \delta + E[\phi_1 Y_{t-1}] + E[\phi_2 Y_{t-2}] + \underbrace{E[\varepsilon_t]}_{0} \\ &= \delta + \phi_1 E[Y_{t-1}] + \phi_2 E[Y_{t-2}] \Rightarrow \mu - \phi_1 \mu - \phi_2 \mu = \delta \\ &\Rightarrow \mu = \frac{\delta}{1 - \phi_1 - \phi_2}. \end{aligned} \tag{2}$$

Para que la media sea finita es necesario imponer que $\phi_1 + \phi_2 \neq 1$.

AR(2). Condición de Estacionariedad

Estacionariedad en Varianza

La expresión de la varianza depende de las autocovarianzas de ordenes 1 y 2 (demostración en libro). Estos tres parámetros se obtienen resolviendo este sistema de ecuaciones (Ecuaciones de Yule-Walker):

$$\begin{cases} \gamma_0 = \phi_1\gamma_1 + \phi_2\gamma_2 + \sigma_\varepsilon^2 \\ \gamma_1 = \phi_1\gamma_0 + \phi_2\gamma_1 \\ \gamma_2 = \phi_1\gamma_1 + \phi_2\gamma_0 \end{cases}$$

Obteniéndose:

$$\gamma_0 = \tfrac{1-\phi_2}{(1+\phi_2)}\tfrac{\sigma_\varepsilon^2}{[(1-\phi_2)^2-\phi_1^2]}, \quad \gamma_1 = \tfrac{\phi_1}{1-\phi_2}\gamma_0, \quad \gamma_2 = (\tfrac{\phi_1^2}{1-\phi_2} + \phi_2)\gamma_0.$$

Para que la varianza sea positiva se debe cumplir que $|\phi_2| < 1$, $\phi_1 + \phi_2 < 1$ y $\phi_2 - \phi_1 < 1$.

AR(2). Condición de Estacionariedad

Si se analiza la estacionariedad utilizando el polinomio característico

$$1 - \phi_1 B - \phi_2 B^2$$

que se cumplan las condiciones de estacionariedad es equivalente a comprobar que **los módulos de las dos raíces del polinomio característico sean mayores que uno**.
Las soluciones (raíces) de la ecuación de segundo grado $1 - \phi_1 B - \phi_2 B^2 = 0$ son:

$$B_1 = \frac{\phi_1 - \sqrt{\phi_1^2 + 4\phi_2}}{-2\phi_2}, \quad B_2 = \frac{\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}}{-2\phi_2}.$$

Así pues, el proceso $AR(2)$ será estacionario si se cumplen las condiciones:

$$\left| \frac{\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}}{-2\phi_2} \right| > 1, \quad \left| \frac{\phi_1 - \sqrt{\phi_1^2 + 4\phi_2}}{-2\phi_2} \right| > 1.$$

AR(2)

Función de autocovarianzas

$$\gamma_0 = \frac{1-\phi_2}{(1+\phi_2)} \frac{\sigma_\varepsilon^2}{[(1-\phi_2)^2-\phi_1^2]}$$
$$\gamma_1 = \frac{\phi_1}{1-\phi_2}\gamma_0$$
$$\gamma_2 = \left(\frac{\phi_1^2}{1-\phi_2} + \phi_2\right)\gamma_0$$
$$\vdots$$
$$\gamma_k = \phi_1\gamma_{k-1} + \phi_2\gamma_{k-2}, \quad \forall k \geq 3$$

Función de autocorrelación

$$\rho_0 = \frac{\gamma_0}{\gamma_0} = 1$$
$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{\frac{\phi_1\gamma_0}{1-\phi_2}}{\gamma_0} = \frac{\phi_1}{1-\phi_2}$$
$$\rho_2 = \frac{\gamma_2}{\gamma_0} = \frac{\phi_1\gamma_1+\phi_2\gamma_0}{\gamma_0} = \frac{\phi_1^2}{1-\phi_2} + \phi_2$$
$$\vdots$$
$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2}$$

AR(2)

Función de autocorrelación parcial

La FACP estima el efecto directo que tiene Y_{t-k} sobre Y_t, excluyendo el efecto indirecto de los valores intermedios $Y_{t-k+1}, Y_{t-k+2}, \ldots, Y_{t-1}$.

$$Y_t = \delta + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t,$$

Se observa que Y_{t-1} e Y_{t-2} tienen efectos directos sobre Y_t, sin embargo Y_{t-3} y las variables con retardo más alto necesitan de Y_{t-1} e Y_{t-2} para influir sobre Y_t, es decir, no tienen efectos directos. Por tanto, para un AR(2) los coeficientes ρ_k^p para $k \geq 3$ serán iguales a cero.

Los dos primeros coeficientes parciales se pueden obtener a partir de las ecuaciones:

$$\rho_1^p = \rho_1 = \frac{\phi_1}{1-\phi_2}, \quad \rho_2^p = \frac{\rho_2 - \rho_1^2}{1-\rho_1^2}$$

Principales características de un AR(2) estacionario

- La FAC decae de forma rápida hacia cero.
- La forma de decrecimiento de la FAC depende de si las raices del polinomio característico son reales o complejas:
  - Si toman valores reales, la FAC decrecerá de forma similar a los procesos AR(1).
  - Si toman valores complejos, la FAC mostrará oscilaciones sinusoidales decrecientes.
- La FACP se anula para retardos superiores o iguales a 3.

Ejemplo (en el libro)

Dado el proceso

$$Y_t = 1 - 1{,}42 \cdot Y_{t-1} - 0{,}5 \cdot Y_{t-2} + \varepsilon_t$$

donde ε_t es ruido blanco con $\sigma_\varepsilon^2 = 0{,}5$, calcular su media, varianza, covarianza, función de autocorrelación simple y los dos primeros coeficientes de la autocorrelación parcial.

AR(p)

Se dice que un proceso estocástico $\{Y_t\}$ sigue un proceso autorregresivo de orden p, $AR(p)$, si se puede escribir como:

$$Y_t = \delta + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \varepsilon_t,$$

donde $\delta, \phi_1, \phi_2, \ldots, \phi_p$ son constantes y ε_t es ruido blanco con media cero y varianza σ_ε^2.

El proceso *AR(p)* puede escribirse utilizando el operador retardo B de la siguiente manera:

$$(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p) Y_t = \delta + \varepsilon_t,$$

donde $1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$ se denomina el polinomio característico.

Condiciones de estacionariedad

Un proceso $AR(p)$ es estacionario si el módulo de cada una de las p raíces del polinomio característico es mayor que uno.

Principales características de un AR(p) estacionario

- La FAC decae hacia cero de forma rápida.
- La FACP se anula para retardos superiores a p.

Ejemplo (en el libro)

Dado el proceso $Y_t = 2{,}5 - 1{,}8Y_{t-1} - 0{,}81Y_{t-2} + \varepsilon_t$ donde ε_t es ruido blanco. Se pide:

a) Comprobar si es estacionario.
b) Calcular la media del proceso.
c) Calcular la varianza del proceso sabiendo que $\sigma_\varepsilon^2 = 2$.
d) Calcular los coeficientes de retardos 1, 2 y 3 de la FAC.
e) Calcular la FACP.

MA(q)

Un proceso se denomina de medias móviles cuando explica el valor de una determinada variable en un periodo t en función de un término independiente y una sucesión de perturbaciones correspondiente a períodos pasados ponderados convenientemente.

La expresión genérica de un proceso de medias móviles *MA(q)* es la siguiente:

$$Y_t = \delta + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q},$$

donde

- δ es una constante
- $\theta_1, \theta_2, \ldots, \theta_q$ son denominados parámetros de media móvil
- ε_t es ruido blanco

Utilizando el operador retardo B:

$$Y_t = \delta + (1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q)\varepsilon_t,$$

MA(q)

Media

$$E[Y_t] = E[\delta] + \underbrace{E[\varepsilon_t]}_{0} + \theta_1 \underbrace{E[\varepsilon_{t-1}]}_{0} + \theta_2 \underbrace{E[\varepsilon_{t-2}]}_{0} + \cdots + \theta_q \underbrace{E[\varepsilon_{t-q}]}_{0} = \delta. \qquad (3)$$

Denotando por $\tilde{Y}_t = Y_t - E[Y_t] = Y_t - \delta$, el proceso $MA(q)$ se puede expresar de la siguiente manera:

$$\tilde{Y}_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q},$$

Condición de invertibilidad

Definicion

Un proceso Y_t es invertible si puede ser representado como un proceso autoregresivo de orden finito o un proceso autoregresivo de orden infinito pero convergente.

Intuitivamente que un proceso sea invertible significa que puede explicarse en función de su propio pasado hasta un determinado retardo (AR finito) o que la influencia del pasado se diluye con el tiempo (AR infinito convergente).

MA(q)

De la expresión con el polinomio característico del proceso $MA(q)$:

$$\tilde{Y}_t = (1 + \theta_1 B + \cdots + \theta_q B^q)\varepsilon_t,$$

se obtiene:

$$(1 + \theta_1 B + \cdots + \theta_q B^q)^{-1}\tilde{Y}_t = \varepsilon_t.$$

Por las propiedades de los polinomios se sabe que el inverso de un polinomio finito es un polinomio infinito:

$$(1 + \theta_1 B + \cdots + \theta_q B^q)^{-1} = 1 + \varphi_1 B + \varphi_2 B^2 + \cdots ,$$

cuya expresión es convergente si y solo si las q **raíces del polinomio característico** $1 + \theta_1 B + \cdots + \theta_q B^q$ **son en módulo mayores que 1**, es decir, $|B_i| > 1$, $i = 1, ..., q$. A estas condiciones se les denomina **condiciones de invertibilidad** del proceso $MA(q)$.

MA(1)

Se dice que un proceso estocástico $\{Y_t\}$ sigue un proceso de medias móviles de orden 1, $MA(1)$, si se puede escribir como:

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

donde δ, θ son constantes y ε_t es ruido blanco con media cero y varianza σ_ε^2.

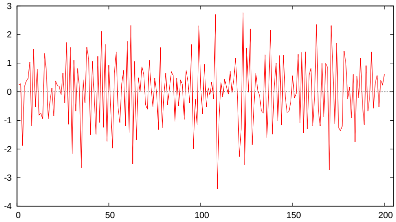


Figura: Proceso MA(1) con $\theta = -0{,}8$

MA(1). Estacionariedad

Como se ha visto, la media del proceso es δ y, por tanto, constante.
La varianza del proceso verifica:

$$\begin{aligned}\gamma_0 &= Var[Y_t] = Var[\delta + \varepsilon_t + \theta\varepsilon_{t-1}] = \underbrace{Var[\varepsilon_t]}_{\sigma_\varepsilon^2} + \theta^2 \underbrace{Var[\varepsilon_{t-1}]}_{\sigma_\varepsilon^2} \\ &\quad + 2\underbrace{Cov(\varepsilon_t, \theta\varepsilon_{t-1})}_{0} \\ \Rightarrow &\quad \gamma_0 = \sigma_\varepsilon^2 + \theta^2\sigma_\varepsilon^2 = (1+\theta^2)\sigma_\varepsilon^2.\end{aligned}$$

Como $\sigma_\varepsilon^2 > 0$ y $1 + \theta^2 > 0$, la varianza es positiva y constante.
Para confirmar la estacionariedad del proceso hace falta ver que la función de autocovarianzas no depende del tiempo.

MA(1). Estacionariedad

$$\begin{aligned}
\gamma_1 &= Cov(Y_t, Y_{t-1}) = Cov(\tilde{Y}_t, \tilde{Y}_{t-1}) = E[\tilde{Y}_t \tilde{Y}_{t-1}] - \underbrace{E[\tilde{Y}_t]}_{0}\underbrace{E[\tilde{Y}_{t-1}]}_{0} \\
&= E[\tilde{Y}_t \tilde{Y}_{t-1}] = E[(\varepsilon_t + \theta\varepsilon_{t-1})\tilde{Y}_{t-1}] = \underbrace{E[\varepsilon_t \tilde{Y}_{t-1}]}_{0} + \theta \underbrace{E[\varepsilon_{t-1}\tilde{Y}_{t-1}]}_{\sigma_\varepsilon^2} = \theta\sigma_\varepsilon^2.
\end{aligned}$$

$$\begin{aligned}
\gamma_2 &= Cov(Y_t, Y_{t-2}) = Cov(\tilde{Y}_t, \tilde{Y}_{t-2}) = E[\tilde{Y}_t \tilde{Y}_{t-2}] - \underbrace{E[\tilde{Y}_t]}_{0}\underbrace{E[\tilde{Y}_{t-2}]}_{0} \\
&= E[\tilde{Y}_t \tilde{Y}_{t-2}] = E[(\varepsilon_t + \theta\varepsilon_{t-1})\tilde{Y}_{t-2}] = \underbrace{E[\varepsilon_t \tilde{Y}_{t-2}]}_{0} + \theta \underbrace{E[\varepsilon_{t-1}\tilde{Y}_{t-2}]}_{0} = 0.
\end{aligned}$$

Razonando de igual manera se obtiene que todos los coeficientes con retardos superiores a dos son también nulos. Por tanto:

$$\gamma_0 = (1+\theta^2)\sigma_\varepsilon^2, \quad \gamma_1 = \theta\sigma_\varepsilon^2, \quad \gamma_k = 0, \ \forall \ k \geq 2.$$

No se necesita ninguna condición para obtener la estacionariedad.

MA(1)

Función de autocorrelación simple

$$\rho_0 = 1, \quad \rho_1 = \frac{\theta\sigma_\varepsilon^2}{(1+\theta^2)\sigma_\varepsilon^2} = \frac{\theta}{1+\theta^2}, \quad \rho_k = \frac{0}{\gamma_0} = 0, \ \forall \ k \geq 2$$

Función de autocorrelación parcial

Si el proceso es invertible, es decir, si la raíz de su polinomio característico $1+\theta B$ es mayor que 1 en valor absoluto, Y_t se puede expresar en función de su propio pasado.

Que el proceso $MA(1)$ se pueda expresar como un $AR(\infty)$ significa que existe un efecto directo entre la variable Y_t y todas las demás variables retardadas Y_{t-1}, Y_{t-2},... Por tanto, todos los coeficientes de la FACP son distintos de cero. Al ser el proceso invertible también se sabe que estos coeficientes decrecen a medida que aumenta el retardo, dicho con otras palabras, la influencia del pasado se debilita a medida que las variables se alejan en el tiempo.

Principales características de un MA(1) invertible

- Siempre es estacionario.
- La FAC sólo tiene un coeficiente no nulo en el retardo 1, siendo 0 en el resto.
- La FACP no se anula, pero tiene un comportamiento convergente a cero.

Ejemplo (en el libro)

Dado el proceso $Y_t = -1{,}5 + \varepsilon_t - 0{,}3 \cdot \varepsilon_{t-1}$, donde ε_t es ruido blanco con $\sigma_\varepsilon^2 = 0{,}2$, calcular su media, varianza, covarianza, función de autocorrelación simple y primeros tres coeficientes de la función de autocorrelación parcial.

MA(2)

Se dice que un proceso estocástico $\{Y_t\}$ sigue un proceso de medias móviles de orden 2, $MA(2)$, si se puede escribir como:

$$Y_t = \delta + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} \Leftrightarrow Y_t = \delta + (1 + \theta_1 B + \theta_2 B^2)\varepsilon_t$$

donde $\delta, \theta_1, \theta_2$ son constantes y ε_t es ruido blanco con media cero y varianza σ_ε^2.

Estacionariedad

La media de un proceso $MA(2)$ es igual a δ y, por tanto, constante.

$$\begin{aligned}\gamma_0 &= Var[Y_t] = Var[\delta + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2}] \\ &= \underbrace{Var[\varepsilon_t]}_{\sigma_\varepsilon^2} + \theta_1^2 \underbrace{Var[\varepsilon_{t-1}]}_{\sigma_\varepsilon^2} + \theta_2^2 \underbrace{Var[\varepsilon_{t-2}]}_{\sigma_\varepsilon^2} \\ &= \sigma_\varepsilon^2 + \theta_1^2\sigma_\varepsilon^2 + \theta_2^2\sigma_\varepsilon^2 = (1 + \theta_1^2 + \theta_2^2)\sigma_\varepsilon^2.\end{aligned}$$

Como $\sigma_\varepsilon^2 > 0$ y $1+\theta_1^2+\theta_2^2 > 0$, la expresión de la varianza es positiva y constante. Falta comprobar que la función de autocovarianzas no depende del tiempo, depende solo del retardo que separa a las variables.

MA(2)

La función de autocovarianzas de un proceso $MA(2)$ es:

$$\begin{aligned}
\gamma_0 &= (1+\theta_1^2+\theta_2^2)\sigma_\varepsilon^2 && (4)\\
\gamma_1 &= \theta_1(\theta_2+1)\sigma_\varepsilon^2 && (5)\\
\gamma_2 &= \theta_2\sigma_\varepsilon^2 && (6)\\
\gamma_k &= 0,\ \forall\ k \geq 3 && (7)
\end{aligned}$$

Como puede observarse, la función de autocovarianzas no depende del tiempo. Por tanto, no es necesario imponer ninguna condición para que el proceso $MA(2)$ sea estacionario.

Función de autocorrelación simple

$$\rho_0 = 1, \quad \rho_1 = \frac{\theta_1(\theta_2+1)\sigma_\varepsilon^2}{(1+\theta_1^2+\theta_2^2)\sigma_\varepsilon^2} = \frac{\theta_1(\theta_2+1)}{1+\theta_1^2+\theta_2^2},$$

$$\rho_2 = \frac{\theta_2\sigma_\varepsilon^2}{(1+\theta_1^2+\theta_2^2)\sigma_\varepsilon^2} = \frac{\theta_2}{1+\theta_1^2+\theta_2^2}, \quad \rho_k = \frac{0}{\gamma_0} = 0,\ \forall\ k \geq 3.$$

MA(2)

Función de autocorrelación parcial

Al igual que para el proceso $MA(1)$, si un proceso $MA(2)$ es invertible entonces se puede expresar en función de su propio pasado, por tanto, existe influencia directa de todas las variables retardadas Y_{t-1}, Y_{t-2},... sobre Y_t. Es decir, todos los coeficientes de la FACP serán distintos de cero.

La FACP tiene infinitos valores distintos de cero que decrecen exponencialmente a partir del primer valor hacia cero.

Principales características de un MA(2) invertible

- Siempre es estacionario.
- La FAC tiene sus dos primeros coeficientes distintos de cero.
- La FACP no se anula, pero tiene un comportamiento convergente a cero.

Ejemplo (en el libro)

Dado el proceso $Y_t = \varepsilon_t - 0{,}8 \cdot \varepsilon_{t-1} + 0{,}25 \cdot \varepsilon_{t-2}$, donde ε_t es ruido blanco con $\sigma_\varepsilon^2 = 0{,}4$, calcular su media, varianza, covarianza, función de autocorrelación simple y primeros tres coeficientes de la función de autocorrelación parcial.

Principales características de un MA(q) invertible

- Siempre es estacionario.
- La FAC tiene los q primeros coeficientes distintos de cero.
- La FACP no se anula, pero tiene un comportamiento convergente a cero.

Ejemplo (en el libro)

Considere el modelo $Y_t = 5 + \varepsilon_t - 0{,}8\varepsilon_{t-1} + 0{,}25\varepsilon_{t-2}$ donde ε_t es ruido blanco. Sabiendo que $\sigma_\varepsilon^2 = 2$ se pide:

1. Comprobar las condiciones de estacionariedad e invertibilidad.
2. Calcular la media y varianza del proceso.
3. Obtener la FAC.
4. Calcular los dos primeros términos de la FACP.
5. Hallar la representación $AR(\infty)$ del proceso.

ARMA(p,q)

Se dice que un proceso $\{Y_t\}$ admite una representación autorregresiva y de medias móviles de orden p, q respectivamente, si se puede modelizar a través de la ecuación:

$$Y_t = \delta + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}.$$

Este modelo se puede escribir en términos del operador de retardos:

$$\begin{aligned} Y_t - \phi_1 Y_{t-1} - \phi_2 Y_{t-2} - \cdots - \phi_p Y_{t-p} &= \delta + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} \\ (1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p) Y_t &= \delta + (1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q) \varepsilon_t \end{aligned}$$

ARMA(p,q)

Teorema

Un proceso autorregresivo de medias móviles finito *ARMA(p,q)* es estacionario sí y solo sí el módulo de cada una de las raíces del polinomio autorregresivo es superior a uno.

Teorema

Un proceso autorregresivo y de medias móviles finito *ARMA(p,q)* es invertible sí y solo sí el módulo de cada una de las raíces del polinomio de medias móviles es superior a uno.

El modelo *ARMA(p,q)*:

- Tiene media y varianza constantes.
- Los primeros q coeficientes de la FAC están dados por la parte *MA*. A partir del retardo q se producirá un decrecimiento que vendrá dado por la estructura *AR*.
- Los primeros p coeficientes de la FACP están dados por la parte *AR*. A partir del retardo p se producirá un decrecimiento que vendrá dado por la

ARMA(p,q)

Ejemplo (en el libro)

Considere el modelo $Y_t = 0{,}4Y_{t-1} + \varepsilon_t + 0{,}8\varepsilon_{t-1}$ donde ε_t es ruido blanco. Sabiendo que $\sigma_\varepsilon^2 = 1$ se pide:

1. Comprobar las condiciones de estacionariedad e invertibilidad.
2. Calcular la media y varianza del proceso.
3. Obtener los coeficientes de orden 1 y 2 de la FAC.
4. Calcular los coeficientes de orden 1 y 2 de la FACP.
5. Hallar la representación $AR(\infty)$ del proceso.

Ejemplo (en el libro)

Dado el proceso $(1 + 0{,}2 \cdot B) \cdot Y_t = (1 - 0{,}7 \cdot B) \cdot \varepsilon_t$, donde ε_t es ruido blanco con $\sigma_\varepsilon^2 = 0{,}1$, se pide contestar de forma razonada a las siguientes cuestiones:

a) ¿Es el proceso estacionario e invertible?

b) Calcular la función de autocorrelación simple para los retardos 1 y 2.

Teorema de Wold

Teorema de Wold

Cualquier proceso estacionario Y_t puede representarse como la suma de dos procesos mutuamente incorrelados:

$$Y_t = D_t + X_t,$$

donde D_t es un proceso determinista y X_t es un proceso $MA(\infty)$.

Sin tener en cuenta la parte determinista, el Teorema de Wold establece que un proceso estacionario puede expresarse como un proceso $MA(\infty)$:

$$Y_t = \varepsilon_t + \varphi_1 \varepsilon_{t-1} + \varphi_2 \varepsilon_{t-2} + \cdots = (1 + \varphi_1 B + \varphi_2 B^2 + \cdots)\varepsilon_t$$

En la práctica no es útil utilizar una expresión con infinitos términos pero un polinomio infinito se puede representar como el cociente de dos polinomios finitos:

$$Y_t = (1 + \varphi_1 B + \varphi_2 B^2 + \cdots)\varepsilon_t = \frac{\theta(B)}{\phi(B)}\varepsilon_t \Leftrightarrow \phi(B) Y_t = \theta(B)\varepsilon_t.$$

El Teorema de Wold permite que cualquier proceso estacionario sea descrito como un proceso $ARMA(p, q)$ o como sus casos particulares $AR(p)$ y $MA(q)$.

Identificación

Ya se conoce el tipo de modelos que pueden describir a los procesos estacionarios: **autorregresivos, de medias móviles y autoregresivos de medias móviles**. Ahora se trata de elegir qué tipo de modelo se ajustaría mejor a la serie temporal bajo estudio. Para ello se comparan las características muestrales de la serie con las características teóricas de los modelos. Concretamente se hacen comparaciones sobre las **funciones de autocorrelación simple y parcial**.

Estadísticos muestrales

	FAC	FACP
$AR(p)$	Decrecimiento exponencial y/o sinusoidal amortiguado	Nula para $k > p$
$MA(q)$	Nula para $k > q$	Decrecimiento exponencial y/o sinusoidal amortiguado
$ARMA(p, q)$	Decrecimiento exponencial y/o sinusoidal amortiguado	Decrecimiento exponencial y/o sinusoidal amortiguado

Cuadro: Comportamiento de la FAC y FACP para los procesos AR, MA y ARMA

Identificación

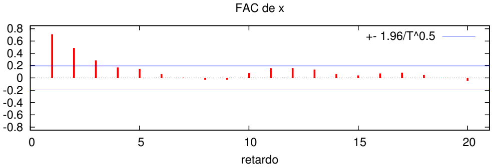
- Las primeras apuestas deberían ser modelos puros *AR* o *MA* y despues estudiar los modelos mixtos *ARMA*
- Si la primera decisión es sobre elegir un modelo *AR* o *MA*, se debe decidir qué función de autocorrelación muestral, la simple o la parcial, decrece:
  - Si se decide que la que decrece es la FAC muestral se apostaría por un modelo *AR* y el orden del modelo se decidiría viendo el número de coeficientes no nulos significativos en los primeros retardos de la FACP muestral
  - Si se decide que la función de autocorrelación muestral que decrece es la parcial, se apostaría por un modelo *MA* y su orden se decidiría viendo el número de primeros coeficientes significativos de la FAC muestral

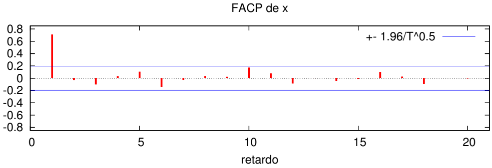
Identificación

Inferencia

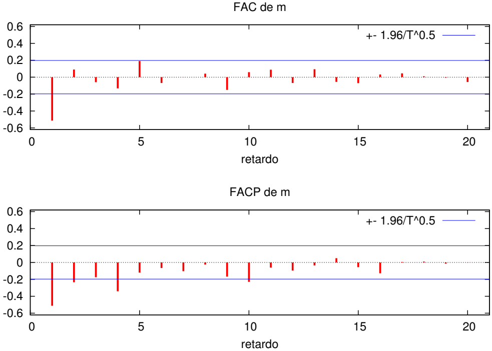
- Bajo la hipótesis nula de que $\rho_k^p = 0$: $\hat{\rho}_k^p \to N(0, \frac{1}{T})$
  Se rechazará la hipótesis nula a un nivel de significación del 5 % **si el estimador $\hat{\rho}_k^p$ está fuera del intervalo** $\pm\frac{2}{\sqrt{T}}$.
  Si se rechaza esta hipótesis nula para los p primeros coeficientes, se apostaría por describir la serie temporal bajo estudio con un proceso $AR(p)$.
- Bajo la hipótesis nula de que $\rho_k = 0$: $\hat{\rho}_k \to N(0, \frac{1}{T})$
  Se rechazará la hipótesis nula a un nivel de significación del 5 % **si el estimador $\hat{\rho}_k$ está fuera del intervalo** $\pm\frac{2}{\sqrt{T}}$.
  Si esto ocurre para los q primeros coeficientes, la serie bajo estudio se describiría mediante un proceso $MA(q)$.

Identificación





Identificación



Modelos univariantes lineales no estacionarios

Tere García Muñoz (Grupos A y B, tgarciam@ugr.es)
Alessio Gaggero (Grupo C, 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

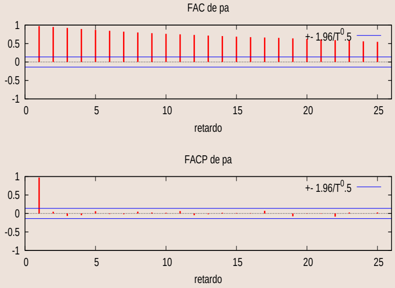


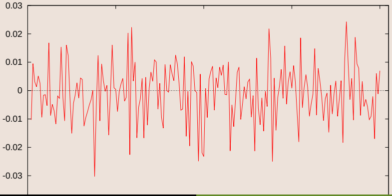
Figura: FAC y FACP de un paseo aleatorio

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.

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

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

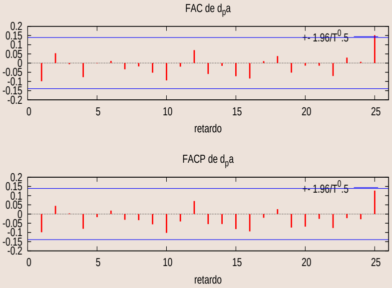


Figura: FAC y FACP de las primeras diferencias

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

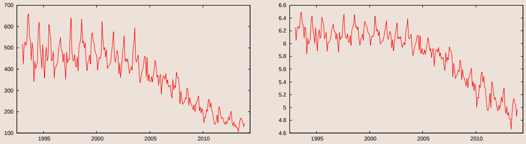


Figura: Número de fallecidos desde Enero de 1993 a Diciembre de 2013 y su logaritmo

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

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: $s, 2s, 3s, 4s, \ldots$.

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:

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

Función de autocovarianzas:

$$
\begin{aligned}
\gamma_s &= E[Y_t Y_{t-s}] = \Phi E[Y_{t-s} Y_{t-s}] + E[\varepsilon_t Y_{t-s}] = \Phi \gamma_0, \\
\gamma_{2s} &= E[Y_t Y_{t-2s}] = \Phi E[Y_{t-s} Y_{t-2s}] + E[\varepsilon_t Y_{t-2s}] = \Phi \gamma_s = \Phi^2 \gamma_0, \\
\gamma_{js} &= \Phi^j \gamma_0.
\end{aligned}
$$

Dado que Y_t se relaciona solo con las variables retardadas Y_{t-s}, Y_{t-2s},... 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 $j = 1, 2, 3, \ldots$, será:

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

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 $j = s$:

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

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

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

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 $h = s, 2s, \ldots, Ps$.

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:

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

Función de autocovarianzas:

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

Función de autocorrelación simple

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

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

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

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

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

y entonces la FACP será nula excepto para los valores

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

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

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

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

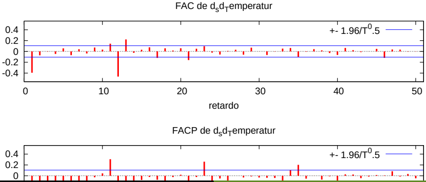
Sea el proceso estacional autorregresivo y de medias moviles de orden (1,1) y periodicidad s, $ARMA(1,1)_s$:

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

donde ε_t es ruido blanco.

La FAC y FACP son funciones infinitas que decrecen tomando valores no nulos en los retardos estacionales $s, 2s, 3s, \cdots$, 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, $2s$,...).



Para eliminar la persistencia estacional fuerte se tomas 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:

$$(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p)(1 - \Phi_1 B^s - \Phi_2 B^{2s} - \cdots - \Phi_P B^{Ps})\nabla^d \nabla_s^D Y_t =$$

$$\delta + (1 + \theta_1 B + \theta_2 B + \cdots + \theta_q B^q)(1 + \Theta_1 B^s + \Theta_2 B^{2s} + \cdots + \Theta_Q B^{Qs})\varepsilon_t,$$

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 subceptibles de cancelarse. Por ejemplo,

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

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

- 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:

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

- 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

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

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

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

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 verosimilud de un modelo viene dada por

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

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

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

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

Y_1,...,Y_T y ε_1,...,ε_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:

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

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.

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

$$
\begin{aligned}
k = 1 : \hat{Y}_T(1) &= E_T[Y_{T+1}] = E_T[\delta + \phi Y_T + \varepsilon_{T+1}] = \delta + \phi Y_T. \\
k = 2 : \hat{Y}_T(2) &= E_T[Y_{T+2}] = E_T[\delta + \phi Y_{T+1} + \varepsilon_{T+2}] = \delta + \phi E_T[Y_{T+1}] \\
&= \delta + \phi \hat{Y}_T(1) = \delta + \phi(\delta + \phi Y_T) = \delta(1+\phi) + \phi^2 Y_T. \\
k = 3 : \hat{Y}_T(3) &= E_T[Y_{T+3}] = E_T[\delta + \phi Y_{T+2} + \varepsilon_{T+3}] = \delta + \phi E_T[Y_{T+2}] \\
&= \delta + \phi \hat{Y}_T(2) = \delta + \phi(\delta(1+\phi) + \phi^2 Y_T) \\
&= \delta(1 + \phi + \phi^2) + \phi^3 Y_T \\
k \geq 4 : \hat{Y}_T(k) &= \delta + \phi \hat{Y}_T(k-1) = \delta(1 + \phi + \phi^2 + \cdots + \phi^{k-1}) + \phi^k Y_T.
\end{aligned}
$$

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

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

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.

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

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

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.

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

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

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

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

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

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:

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

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:

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

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:

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

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:

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

Varianza del error de predicción:

$$\begin{aligned}
k = 1: Var(e_T(1)) &= \sigma_\varepsilon^2. \\
k = 2: Var(e_T(2)) &= E[e_T(2)^2] = E[\varepsilon_{T+2}^2 + \phi^2\varepsilon_{T+1}^2 + 2\phi\varepsilon_{T+2}\varepsilon_{T+1}] \\
&= (1+\phi^2)\sigma_\varepsilon^2. \\
k > 2: Var(e_T(k)) &= (1+\phi^2+\phi^4+\cdots+\phi^{2(k-1)})\sigma_\varepsilon^2 = \frac{1-\phi^{2k}}{1-\phi^2}\sigma_\varepsilon^2.
\end{aligned}$$

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:

$$\begin{aligned}
k = 1 : e_T(1) &= Y_{T+1} - \hat{Y}_T(1) \\
&= (\delta + \phi_1 Y_T + \phi_2 Y_{T-1} + \varepsilon_{T+1}) - (\delta + \phi_1 Y_T + \phi_2 Y_{T-1}) = \varepsilon_{T+1}. \\
k = 2 : e_T(2) &= Y_{T+2} - \hat{Y}_T(2) \\
&= (\delta + \phi_1(\delta + \phi_1 Y_T + \phi_2 Y_{T-1} + \varepsilon_{T+1}) + \phi_2 Y_T + \varepsilon_{T+2}) - \\
&\quad (\delta(1+\phi_1) + (\phi_1^2 + \phi_2) Y_T + \phi_1\phi_2 Y_{T-1}) \\
&= \varepsilon_{T+2} + \phi_1 \varepsilon_{T+1}. \\
k = 3 : e_T(3) &= Y_{T+3} - \hat{Y}_T(3) = \cdots = \varepsilon_{T+3} + \phi_1 \varepsilon_{T+2} + (\phi_1^2 + \phi_2)\epsilon_{T+1}.
\end{aligned}$$

Varianza del error de predicción:

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

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

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

Varianza del error de predicción:

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

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:

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

Varianza del error de predicción:

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

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

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

Varianza del error de predicción:

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

Calculo de la varianza del error de predicción en cualquier ARIMA

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

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

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

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

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

donde $Var\left(e_T(k)\right) = \left(1 + \psi_1^2 + \cdots + \psi_{k-1}^2\right)\sigma_\varepsilon^2$ siendo $\psi_1, \ldots, \psi_{k-1}$ los primeros $k-1$ coeficientes de la representación 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): $EAM = \dfrac{1}{T-m} \cdot \displaystyle\sum_{t=T-m+1}^{T} |Y_t - \hat{Y}_t|.$

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

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

Modelos de función de transferencia: modelos ARMAX

Tere García Muñoz (Grupos A y B, tgarciam@ugr.es)
Alessio Gaggero (Grupo C, alessiogaggero@ugr.es)

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

Tema 4– Econometría 3
Grado en Económicas

MODELOS DE FUNCIÓN DE TRANSFERENCIA O MODELOS ARMAX:

- Generalización del análisis unidimensional de series temporales: incorporar al modelo otras variables exógenas como explicativas.
- Puesto que estas variables pueden aparecer retardadas se introducen **relaciones de carácter dinámico**.
- Y, recibe el nombre de **output**, y las explicativas, X, de **input**.
- Algunos ejemplos de interés en Economía: variables impulso y escalón.

Consideremos X_t, Y_t dos series debidamente transformadas para ser **estacionarias** (en el caso de que no lo fueran).
El modelo ARMAX o modelo de función de transferencia se puede expresar como:

$$\begin{aligned} Y_t &= \left(\nu_0 + \nu_1 B + \nu_2 B^2 + \nu_3 B^3 + \cdots\right) X_t + \eta_t \\ &= \nu_0 X_t + \nu_1 X_{t-1} + \nu_2 X_{t-2} + \nu_3 X_{t-3} + \cdots + \eta_t. \end{aligned}$$

- η_t representa al ruido, se supone que es estacionario e independiente del *input* X_t.
- ν_j se denominan **coeficientes de respuesta al impulso**: describen el peso que tienen los valores pasados de X_t a la hora de pronosticar Y_t.
- Para que la función de transferencia sea estable suponemos que

$$\sum_{j=0}^{+\infty} |\nu_j| < +\infty.$$

- Observemos que en $(\nu_0 + \nu_1 B + \nu_2 B^2 + \nu_3 B^3 + \ldots)$ hay infinitos términos.

Esta situación se solventa representando un polinomio de orden infinito como cociente de dos polinomios de orden finito. Esto es:

$$\nu(B) = \frac{\omega_s(B)}{\delta_r(B)} = \frac{\omega_0 + \omega_1 B + \omega_2 B^2 + \ldots + \omega_s B^s}{\delta_0 - \delta_1 B - \delta_2 B^2 - \ldots - \delta_r B^r}.$$

Usualmente en aplicaciones de carácter económico los valores de r y s no son superiores a 2.

$$Y_t = \frac{(\omega_0 + \omega_1 B + \omega_2 B^2 + \ldots + \omega_s B^s)}{(\delta_0 - \delta_1 B - \delta_2 B^2 - \ldots - \delta_r B^r)} X_t + \eta_t \tag{1}$$

En ocasiones el efecto de X_t no se transmite de manera instantánea a Y_t, sino que lo hace con un retardo de b periodos

$$Y_t = \frac{(\omega_0 + \omega_1 B + \omega_2 B^2 + \ldots + \omega_s B^s)}{(\delta_0 - \delta_1 B - \delta_2 B^2 - \ldots - \delta_r B^r)} X_{t-b} + \eta_t$$

$$Y_t = \frac{(\omega_0 + \omega_1 B + \omega_2 B^2 + \ldots + \omega_s B^s)}{(\delta_0 - \delta_1 B - \delta_2 B^2 - \ldots - \delta_r B^r)} B^b X_t + \eta_t$$

Suponiendo, por comodidad, que $\delta_0 = 1$,

$$\begin{aligned} Y_t &= \delta_1 Y_{t-1} + \delta_2 Y_{t-2} + \cdots + \delta_r Y_{t-r} \\ &\quad + \omega_0 X_{t-b} + \omega_1 X_{t-b-1} + \cdots + \omega_s X_{t-b-s} + perturbacion. \end{aligned}$$

Mediante esta representación se puede observar cómo la variable *output*, Y, depende de su propio pasado y de la información actual (si $b = 0$) y pasada del *input*, X.

Dado el proceso ARMA estacionario

$$(1-\phi_1 B-\cdots-\phi_p B^p)Y_t=(1+\theta_1 B+\cdots+\theta_q)\epsilon_t,$$

suponemos que se ve afectado en un instante dado, $t = t_0$, por un suceso conocido, c_0.

$$(1-\phi_1 B-\cdots-\phi_p B^p)Y_t=c_0 I_t(t_0)+(1+\theta_1 B+\cdots+\theta_q B^q)\epsilon_t, \quad I_t(t_0)=\begin{cases} 0, \text{ si } t\neq t_0 \\ 1, \text{ si } t=t_0 \end{cases},$$

donde I_t recibe el nombre de **variable impulso**.
El efecto puede darse en distintos periodos **consecutivos**. Esta situación se podría representar a partir de m variables impulso:

$$\begin{aligned}\phi_p(B)Y_t &= c_0 I_t(t_0)+c_1 I_t(t_1)+\cdots+c_m I_t(t_m)+\theta_q(B)\epsilon_t\\ &= c_0 I_t(t_0)+c_1 I_{t-1}(t_0)+\cdots+c_m I_{t-m}(t_0)+\theta_q(B)\epsilon_t\\ &= (c_0+c_1 B+\cdots+c_m B^m)I_t(t_0)+\theta_q(B)\epsilon_t\end{aligned}$$

donde $c(B)$ describe el efecto de la variable impulso (todos sus coeficientes provienen de la misma causa) y corresponde a una función de transferencia.

EJEMPLO

En una serie temporal diaria de transporte por carretera que mide el número de vehículos que circulan por una carretera ser produce un accidente que obliga a cortes parciales durante dos días hasta que se arreglan los desperfectos. Un mes después del accidente se estima el siguiente modelo de intervención donde $I_t(t_0)$ es una variable impulso que toma el valor uno en el día del accidente:

$$z_t = -(2100 + 3500B + 1000B^2 + 400B^3)I_t(t_0) + (1 - 0{,}5B)\epsilon_t$$

Interpreta los efectos de la variable impulso.

El modelo de intervención indica que el día del accidente hubo una disminución de 2100 vehículos. El segundo día la disminución es de 3500 vehículos. El tercer día, que las condiciones ya eran normales, hay todavía una disminución del tráfico de 1000 vehículos y el cuarto día de 400.

Dado el proceso ARMA estacionario

$$(1 - \phi_1 B - \cdots - \phi_p B^p) Y_t = (1 + \theta_1 B + \cdots + \theta_q)\epsilon_t,$$

suponemos que se ve afectado por un efecto permanente, c_0, a partir de un instante t_0:

$$(1 - \phi_1 B - \cdots - \phi_p B^p) Y_t = c_0 S_t(t_0) + (1 + \theta_1 B + \cdots + \theta_q B^q)\epsilon_t, \quad S_t(t_0) = \begin{cases} 0, & \text{si } t < t_0 \\ 1, & \text{si } t \geq t_0 \end{cases}$$

donde S_t recibe el nombre de **variable escalón**.
Si se considera que el efecto permanente se introduce gradualmente (en varios periodos **consecutivos**) hasta que se estabiliza, habría que considerar m variables escalón:

$$\begin{aligned}
\phi_p(B) Y_t &= c_0 S_t(t_0) + c_1 S_t(t_1) + \cdots + c_m S_t(t_m) + \theta_q(B)\epsilon_t \\
&= c_0 I_t(t_0) + c_1 S_{t-1}(t_0) + \cdots + c_m S_{t-m}(t_0) + \theta_q(B)\epsilon_t \\
&= (c_0 + c_1 B + \cdots + c_m B^m) S_t(t_0) + \theta_q(B)\epsilon_t
\end{aligned}$$

donde $c(B)$ describe el efecto de la variable escalón y corresponde a una función de transferencia.
El efecto total de la variable escalón, $c_0 + \cdots + c_m$, recibe el nombre de **ganancia** de la función de transferencia.

EJEMPLO

En una serie temporal que mide las ventas mensuales de un artículo el efecto de una subida de precios se representa mediante una variable escalón siguiendo este modelo de intervención:

$$z_t = (-70 - 30B + 20B^2)S_t(t_0) + (1 - 0{,}3B)\epsilon_t$$

Interpreta los efectos de la variable escalón

El modelo de intervención indica que en el mes de la subida se produjo una disminución de 70 unidades. En el mes siguiente se produjo una disminución adicional de 30 unidades y en el tercero una subida de 20. El efecto total es una bajada de 80 unidades, que es la ganancia de la función de transferencia.

Modelos de volatilidad estocástica

Tere García Muñoz (Grupos A y B, tgarciam@ugr.es)
Alessio Gaggero (Grupo C, alessiogaggero@ugr.es)

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

Tema 5– Econometría 3
Grado en Económicas

En los modelos econométricos convencionales, la varianza del término perturbación se supone constante. Sin embargo, muchas series temporales económicas, principalmente las financieras, exhiben periodos de alta volatilidad seguidos de periodos de relativa tranquilidad. En este caso, grandes choques (residuales) están seguidos por grandes choques en ambas direcciones, y pequeños shocks están seguidos por pequeñas perturbaciones. Por ejemplo, los mercados bursátiles se caracterizan por periodos *relajados* de baja volatilidad y alta volatilidad. Esto es particularmente cierto en series con altas frecuencias (diarias o semanales). Una forma de modelizar tales patrones es permitir que la varianza del término perturbación dependa de su propio pasado.

Volatilidad: variaciones bruscas que pueden encontrarse en una serie de tiempo. Aunque en el largo plazo la serie sea estacionaria y tenga, por tanto, varianza constante, puede presentar oscilaciones a corto plazo donde cada periodo esté influenciado por los inmediatamente anteriores, que es lo que recoge la varianza condicional.

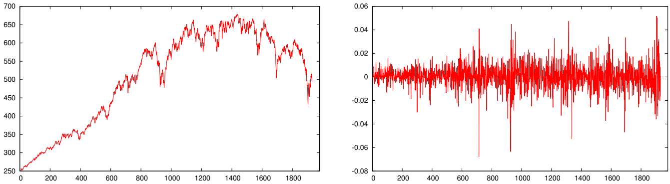


Figura: Observaciones de la bolsa de Nueva York cada 5 días (izquierda) y de las primeras diferencias regulares de su logaritmo (derecha)

Un buen modelo para abordar la volatilidad debe reflejar:

- La volatilidad tiene tendencia a aparecer agrupada por periodos, los periodos de alta volatilidad suelen mantenerse en el tiempo y lo mismo ocurre con los periodos de baja volatilidad (efecto contagio).
- A un periodo de alta/baja volatilidad, eventualmente, le sigue otro de volatilidad normal. Es decir, existe un nivel normal de volatilidad al cual ésta retorna eventualmente.

Estas características muestran la existencia de regularidades en el comportamiento que posibilitan la detección de la volatilidad y, por tanto, su modelización. Los modelos que utilizaremos reciben el nombre de **procesos con varianza condicionalmente heterocedástica** (**modelos ARCH**). En 1982, Engle introduce este tipo de modelos presuponiendo que la varianza condicional (no constante) depende de sus valores pasados siguiendo una estructura autorregresiva. Posteriormente en 1986, Bollerslev generaliza dichos modelos incorporando estructura de media móvil a la varianza condicional dando lugar a los **modelos GARCH**.

En los procesos con varianza condicionalmente heterocedástica se considera que la varianza condicionada del ruido de un modelo ARMA ($E[\varepsilon_t^2/\varepsilon_{t-1},\ldots] \equiv E_{t-1}[\varepsilon_t^2]$) está condicionada por los valores pasados. Por tanto la varianza condicionada no será constante pues depende en cada instante t de los valores pasados en $t-1, t-2, \ldots$

Una estrategia simple es modelizar el cuadrado de las perturbaciones en función de su pasado, por ejemplo:

$$\varepsilon_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + v_t,$$

donde v_t es ruido blanco. De esta manera, la varianza condicionada de las perturbaciones depende del pasado de la perturbaciones:

$$E_{t-1}[\varepsilon_t^2] = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2$$

Si α_1 es cero, la varianza condicionada será constante (no habrá volatilidad). En otro caso la varianza condicional evoluciona de acuerdo a un proceso autorregresivo.

Ahora bien, la relación lineal no tiene un tratamiento algebráico cómodo, siendo más manejable especificar el ruido de forma multiplicativa.

Se dice que $\{\varepsilon_t\}$ sigue un **proceso autorregresivo condicionalmente heterocedástico de orden 1**, **ARCH(1)**, si:

$$\varepsilon_t = v_t\sqrt{\alpha_0 + \alpha_1\varepsilon_{t-1}^2} \Leftrightarrow \varepsilon_t^2 = v_t^2\left(\alpha_0 + \alpha_1\varepsilon_{t-1}^2\right),$$

donde v_t es ruido blanco con varianza igual a 1 e independiente de ε_{t-1}.
Denotando por: $h_t = \alpha_0 + \alpha_1\varepsilon_{t-1}^2$

- La media marginal (o no condicionada) vale cero:

$$E[\varepsilon_t] \quad = \quad E[v_t\sqrt{h_t}] = \underbrace{E[v_t]}_{0}\cdot E[\sqrt{h_t}] = 0,$$

- Teniendo en cuenta que la varianza marginal (no condicionada) es constante, se tiene que:

$$E[\varepsilon_t^2] = E[v_t^2 h_t] = \underbrace{E[v_t^2]}_{1}\cdot E[h_t] = \alpha_0 + \alpha_1 E[\varepsilon_{t-1}^2] \Rightarrow E[\varepsilon_t^2] = \frac{\alpha_0}{1-\alpha_1}.$$

  Para que la varianza sea positiva se ha de verificar que $\alpha_0 > 0$ y $0 \leq \alpha_1 < 1$.

- Covarianza:

$$E[\varepsilon_t \varepsilon_{t-j}] = E[v_t \sqrt{h_t} v_{t-j} \sqrt{h_{t-j}}] = \underbrace{E[v_t]}_{0} \underbrace{E[v_{t-j}]}_{0} \cdot E[\sqrt{h_t}\sqrt{h_{t-j}}] = 0, j \neq 0$$

Aún con la modelización ARCH, ε_t sigue siendo un ruido blanco. La influencia de la modelización ARCH cae por completo en la varianza condicionada.

- La media condicionada (a los valores pasados de la serie) vale cero:

$$\begin{aligned} E_{t-1}[\varepsilon_t] &= E_{t-1}[v_t \sqrt{h_t}] = E_{t-1}[v_t] \cdot E_{t-1}[\sqrt{h_t}] \\ &= \underbrace{E[v_t]}_{0} \cdot E_{t-1}[\sqrt{h_t}] = 0, \end{aligned}$$

- La varianza condicional tiene una estructura parecida a un AR(1):

$$\begin{aligned} E_{t-1}[\varepsilon_t^2] &= E_{t-1}[v_t^2 h_t] = E_{t-1}[v_t^2] \cdot E_{t-1}[h_t] = \underbrace{E[v_t^2]}_{1} \cdot E_{t-1}[h_t] \\ &= \alpha_0 + \alpha_1 E_{t-1}[\varepsilon_{t-1}^2] = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 = h_t, \end{aligned}$$

Para que sea positiva se impone que los dos parámetros sean positivos.

De esta expresión $E_{t-1}[\varepsilon_t^2] = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 = h_t$ se deducen las características deseables de los modelos con volatilidad:

- Un valor alto/bajo de ε_{t-1}^2 conduce a un valor alto/bajo de la varianza del siguiente valor condicionada a ε_{t-1}^2, $E_{t-1}[\varepsilon_t^2]$. De esta forma, se hace más probable que el siguiente dato, ε_t^2, sea también alto/bajo. Luego, fuertes fluctuaciones inesperadas tienden a venir seguidas de períodos de iguales características.
- Por otro lado, como tanto la media marginal como la condicionada son cero, es posible que el valor de ε_t^2 sea pequeño (retorne a su media). De esta forma disminuirá la varianza condicionada de la observación siguiente, de manera que períodos de estabilidad tienden a venir seguidos por períodos igualmente estables.

Se dice que $\{\varepsilon_t\}$ sigue un **proceso autorregresivo condicionalmente heterocedástico de orden** q, **ARCH(q)**, si:

$$\varepsilon_t = v_t\sqrt{h_t},$$

donde $h_t = \alpha_0 + \sum_{k=1}^{q} \alpha_k \varepsilon_{t-k}^2$ y v_t es ruido blanco con varianza igual a 1 e independiente de $\varepsilon_{t-1}, \ldots, \varepsilon_{t-q}$.
De manera análoga a la anterior se obtiene que:

- $E[\varepsilon_t] = 0 = E_{t-1}[\varepsilon_t]$.
- $E[\varepsilon_t^2] = \frac{\alpha_0}{1-\sum_{k=1}^{q}\alpha_k}$ de donde se deduce que $\alpha_0, \alpha_1, \ldots, \alpha_q \geq 0$ y $\sum_{k=1}^{q} \alpha_k < 1$.
- $E_{t-1}[\varepsilon_t^2] = \alpha_0 + \sum_{k=1}^{q} \alpha_k \varepsilon_{t-k}^2 = h_t$ (la varianza condicionada depende de los últimos q valores de ε_{t-1}^2, luego tiene estructura autorregresiva).

Si el valor q del *ARCH(q)* es grande, se tiene que estimar un gran número de parámetros, $\alpha_1, \ldots, \alpha_q$, con los problemas que esto puede suponer. Para evitarlos se permite que la varianza condicionada tenga estructura ARMA:
El caso más sencillo responde a la expresión:

$$h_t = E_{t-1}[\varepsilon_t^2] = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1},$$

Se dice que $\{\varepsilon_t\}$ sigue un **proceso autorregresivo condicionalmente hetero-cedástico generalizado de orden (1,1)**, **GARCH(1,1)**, si:

$$\varepsilon_t = v_t \sqrt{h_t},$$

donde $h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}$ y v_t es ruido blanco con varianza igual a 1 e independiente de ε_{t-1}.

- La media marginal y condicionada son cero: $E[\varepsilon_t] = 0 = E_{t-1}[\varepsilon_t]$.
- Teniendo en cuenta que la varianza marginal es estacionaria y la ley de esperanzas iteradas:

$$
\begin{aligned}
E[\varepsilon_t^2] &= E[v_t^2 h_t] = \underbrace{E[v_t^2]}_{1} \cdot E[h_t] = \alpha_0 + \alpha_1 E[\varepsilon_{t-1}^2] + \beta_1 E[h_{t-1}] \\
&\Downarrow \\
&\Downarrow \quad E[h_{t-1}] = E[E_{t-2}[\varepsilon_{t-1}^2]] = E[E[\varepsilon_{t-1}^2/\varepsilon_{t-2}, \ldots]] = E[\varepsilon_{t-1}^2] \\
&\Downarrow \\
&= \alpha_0 + \alpha_1 E[\varepsilon_{t-1}^2] + \beta_1 E[\varepsilon_{t-1}^2] \Rightarrow E[\varepsilon_t^2] = \frac{\alpha_0}{1 - \alpha_1 - \beta_1}.
\end{aligned}
$$

Para que la varianza sea positiva se ha de verificar que $\alpha_0, \alpha_1, \beta_1 \geq 0$ y $\alpha_1 + \beta_1 < 1$.

- La varianza condicionada corresponde a una estructura ARMA(1,1):

$$\begin{aligned} E_{t-1}[\varepsilon_t^2] &= E_{t-1}[v_t^2 h_t] = E_{t-1}[v_t^2] \cdot E_{t-1}[h_t] = \alpha_0 + \alpha_1 E_{t-1}[\varepsilon_{t-1}^2] \\ & +\beta_1 E_{t-1}[h_{t-1}] = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1} = h_t, \end{aligned}$$

  donde se ha usado que:

$$E_{t-1}[h_{t-1}] = E_{t-1}[E_{t-2}[\varepsilon_{t-1}^2]] = E_{t-2}[E_{t-1}[\varepsilon_{t-1}^2]] = E_{t-2}[\varepsilon_{t-1}^2] = h_{t-1}.$$

La novedad en este caso radica en la aparición de un nuevo término que hace que la varianza condicionada, h_t, dependa de un retardo suyo, h_{t-1}. Así, es de esperar que haya rachas de mayor variabilidad que con los modelos ARCH(1) debido al efecto directo de las varianzas condicionadas altas/bajas del instante anterior.

Se dice que $\{\varepsilon_t\}$ sigue un **proceso autorregresivo condicionalmente heterocedástico generalizado de orden (p,q)**, **GARCH(p,q)**, si:

$$\varepsilon_t = v_t\sqrt{h_t},$$

donde

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j}$$

y v_t es ruido blanco con varianza igual a 1 e independiente de $\varepsilon_{t-1}, \ldots, \varepsilon_{t-q}$. De manera análoga a la anterior se obtiene que:

- $E[\varepsilon_t] = 0 = E_{t-1}[\varepsilon_t]$.
- $E[\varepsilon_t^2] = \frac{\alpha_0}{1-\sum_{i=1}^{q}\alpha_i - \sum_{j=1}^{p}\beta_j}$ de donde se deduce que $\alpha_0, \alpha_1, \ldots, \alpha_q \geq 0$, $\beta_1, \ldots, \beta_q > 0$ y $\sum_{i=1}^{q}\alpha_i + \sum_{j=1}^{p}\beta_j < 1$.
- $E_{t-1}[\varepsilon_t^2] = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j} = h_t$ (la varianza condicionada tiene estructura ARMA(p,q)).

La modelización de modelos GARCH reside en el comportamiento del cuadrado del ruido. Por tanto, para la identificación de los mismos habrá que basarse en los residuos al cuadrado del modelo de series temporales analizado.
Los pasos a seguir son los siguientes:

- Estimar el proceso estocástico $\{Y_t\}$ con el mejor modelo ARMA posible. En tal caso, dicho proceso es estacionario (media y varianza constantes) y sus residuos deben responder a las características de un ruido blanco.
- Calcular el cuadrado de los errores y representar sus funciones de autocorrelación simple y parcial. La identificación de un patrón ARMA(p,q) para los cuadrados de los residuos sería indicativo de la adecuación de un modelo GARCH(p,q).

Ejemplo.La serie correspondiente al PIB trimestral desde 1995 a 2015 es transformada de la siguiente manera: se calcula su logaritmo neperiano y se hacen primeras diferencias regulares y estacionales. El correlograma de la serie transformada es:

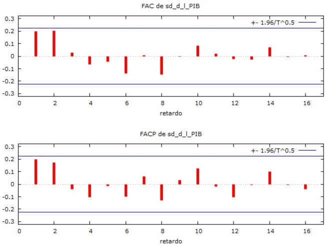


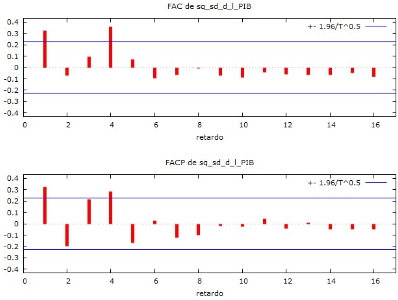
Figura: Correlogramas del logaritmo neperiano, primeras diferencias regulares y estacionales del PIB trimestral

Escribir la ecuación del modelo.

Puesto que los correlogramas se corresponden con un ruido blanco, la ecuación del modelo sería:

$$\nabla_4 \cdot \nabla \cdot \ln PIB_t = \varepsilon_t,$$

donde ε_t es ruido blanco.
El correlograma de los residuos al cuadrado de este modelo es:



En estos correlogramas se observa la existencia de estructura, por lo que en el modelo ajustado hay volatilidad y tendríamos que identificar un modelo GARCH.

Un procedimiento más formal para detectar la presencia de estructura ARCH en los errores es el siguiente:

1. Estimar el proceso estocástico $\{Y_t\}$ con el mejor modelo ARMA posible y obtener el cuadrado de los errores.
2. Realizar una regresión de dichos residuos al cuadrado sobre una constante y q retardos de los mismos:
$$e_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \cdots + \alpha_q e_{t-q}^2 + v_t$$
.
3. Si no hay un patrón ARCH, los valores estimados de α_i, con $i = 1, \ldots, q$, deberían ser estadísticamente cero y el coeficiente de determinación de esta regresión auxiliar, R^2_{aux}, muy bajo.
Se rechazaría la hipótesis nula
$$H_0 : \alpha_1 = \cdots = \alpha_q = 0$$
si $\chi^2_{exp} = T \cdot R^2_{aux} > \chi^2_q(1 - \alpha)$

En definitiva, si se rechaza la hipótesis nula anterior se podría concluir (al nivel de significación usado) que hay estructura ARCH en los errores.

Ejemplo. En el ejemplo anterior, se ha realizado el contraste para detectar la presencia de estructura ARCH en los errores del modelo ajustado que tiene por hipótesis nula

$$H_0 : \alpha_1 = 0$$

y se obtiene un coeficiente de determinación para la regresión auxiliar igual a 0.048.
Teniendo en cuenta que $q = 1$ y que se disponen de 167 observaciones se tiene que se rechaza la hipótesis nula ya que

$$\chi^2_{exp} = 167 \cdot 0{,}048 = 8{,}016 > 3{,}81 = \chi^2_1(0{,}95)$$

.
Por tanto, en los errores hay estructura ARCH.

Ejemplo. A partir de los residuos, e, del modelo

$$(1 + 0{,}2 \cdot B) \cdot Y_t = (1 - 0{,}7 \cdot B) \cdot \varepsilon_t$$

se desea comprobar si hay volatilidad.
Con tal objetivo se plantea la regresión auxiliar

$$e_t^2 = \alpha_0 + \alpha_1 \cdot e_{t-1}^2 + \alpha_2 \cdot e_{t-2}^2 + \alpha_3 \cdot e_{t-3}^2 + v_t$$

obteniéndose un coeficiente de determinación $R_{aux}^2 = 0{,}018$ y utilizando 100 observaciones.
Contraste
$H_0 : \alpha_1 = \alpha_2 = \alpha_3 = 0$ (no existe estructura ARCH)
$H_1 : \exists i$ tal que $\alpha_i \neq 0$ (existe estructura ARCH)
NO se rechaza la hipótesis nula ya que:

$$\chi_{exp}^2 = 100 \cdot 0{,}018 = 1{,}8 \not> 7{,}815 = \chi_3^2(0{,}95).$$

Por tanto, no existe estructura ARCH en el modelo.

Modelos vectoriales autorregresivos

Tere García Muñoz (Grupos A y B, tgarciam@ugr.es)
Alessio Gaggero (Grupo C, alessiogaggero@ugr.es)

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

Tema 6– Econometría 3
Grado en Económicas

Los modelos VAR:

- analizan conjuntamente dos o más series temporales
- son muy útiles cuando existe evidencia de causalidad simultánea entre un grupo de variables y sus relaciones se transmiten a lo largo de un determinado número de períodos.

Modelo VAR con p retardos para dos variables:

$$\begin{aligned} X_t &= \delta_1 + \alpha_{11} X_{t-1} + \cdots + \alpha_{1p} X_{t-p} + \beta_{11} Y_{t-1} + \cdots + \beta_{1p} Y_{t-p} + a_{1t}, \\ Y_t &= \delta_2 + \alpha_{21} X_{t-1} + \cdots + \alpha_{2p} X_{t-p} + \beta_{21} Y_{t-1} + \cdots + \beta_{2p} Y_{t-p} + a_{2t}, \end{aligned}$$

donde

- δ_i representa el término independiente de la ecuación i, con $i = 1, 2$
- α_{ij} y β_{ij} representan los coeficientes de los procesos X e Y, respectivamente, en la ecuación i y el retardo j, con $j = 1, \ldots, p$
- a_{1t} y a_{2t} son dos series independientes e idénticamente distribuidas con media cero.

Así, por ejemplo, si $p = 1$ se tiene:

$$\begin{aligned} X_t &= \delta_1 + \alpha_{11} X_{t-1} + \beta_{11} Y_{t-1} + a_{1t}, \\ Y_t &= \delta_2 + \alpha_{21} X_{t-1} + \beta_{21} Y_{t-1} + a_{2t}. \end{aligned}$$

Expresado matricialmente:

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{pmatrix} \delta_1 \\ \delta_2 \end{pmatrix} + \begin{pmatrix} \alpha_{11} & \beta_{11} \\ \alpha_{21} & \beta_{21} \end{pmatrix} \cdot \begin{pmatrix} X_{t-1} \\ Y_{t-1} \end{pmatrix} + \begin{pmatrix} a_{1t} \\ a_{2t} \end{pmatrix}$$

- α_{11}, muestra la dependencia lineal de X_t con X_{t-1} en presencia de Y_{t-1}.
- β_{11}, muestra la dependencia lineal de X_t con Y_{t-1} en presencia de X_{t-1}.
- α_{21}, mide la relación lineal entre Y_t y X_{t-1} en presencia de Y_{t-1}.
- β_{21}, mide la relación lineal entre Y_t y Y_{t-1} en presencia de X_{t-1}.

Caso 1. $\beta_{11} = \alpha_{21} = 0$.

$$\begin{aligned} X_t &= \delta_1 + \alpha_{11} X_{t-1} + a_{1t}, \\ Y_t &= \delta_2 + \beta_{21} Y_{t-1} + a_{2t}. \end{aligned}$$

- X_t e Y_t no están dinámicamente relacionados.
- Cada serie sigue un modelo univariante *AR(1)* y se pueden tratar como tales de acuerdo a lo estudiado en los tres primeros capítulos.
- Se dice que las dos series están desacopladas.

Caso 2. $\beta_{11} = 0$ pero $\alpha_{21} \neq 0$

$$\begin{aligned} X_t &= \delta_1 + \alpha_{11} X_{t-1} + a_{1t}, \\ Y_t &= \delta_2 + \alpha_{21} X_{t-1} + \beta_{21} Y_{t-1} + a_{2t}. \end{aligned}$$

- X_t no depende del valor pasado Y_{t-1}, pero Y_t si depende del valor pasado X_{t-1}. Relación unidireccional con X_t actuando como una variable *input* y Y_t como una variable *output*. X_t y Y_t tienen una relación de función de transferencia.
- El modelo implica la existencia de causalidad en sentido de Granger entre las dos series con X_t causando Y_t, pero no X_t siendo causada por Y_t.
- La información pasada X_{t-1} mejora la predicción de Y_t, reduciendo la varianza del error de predicción.

De forma análoga, si $\alpha_{21} = 0$ pero $\beta_{11} \neq 0$, entonces Y_t causa X_t pero X_t no causa Y_t.

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{pmatrix} \delta_1 \\ \delta_2 \end{pmatrix} + \begin{pmatrix} \alpha_{11} & \beta_{11} \\ \alpha_{21} & \beta_{21} \end{pmatrix} \cdot \begin{pmatrix} X_{t-1} \\ Y_{t-1} \end{pmatrix} + \begin{pmatrix} a_{1t} \\ a_{2t} \end{pmatrix}$$

Condición de estacionariedad: Las series de un modelo VAR(1) son estacionarias si todos los valores propios de la matriz $\begin{pmatrix} \alpha_{11} & \beta_{11} \\ \alpha_{21} & \beta_{21} \end{pmatrix}$ son menores que uno en valor absoluto.

$$\begin{aligned} X_t &= \delta_1 + \alpha_{11} X_{t-1} + \alpha_{12} X_{t-2} + \beta_{11} Y_{t-1} + \beta_{12} Y_{t-2} + a_{1t}, \\ Y_t &= \delta_2 + \alpha_{21} X_{t-1} + \alpha_{22} X_{t-2} + \beta_{21} Y_{t-1} + \beta_{22} Y_{t-2} + a_{2t}. \end{aligned}$$

$$\begin{aligned} \begin{pmatrix} X_t \\ Y_t \end{pmatrix} &= \begin{pmatrix} \delta_1 \\ \delta_2 \end{pmatrix} + \begin{pmatrix} \alpha_{11} & \beta_{11} \\ \alpha_{21} & \beta_{21} \end{pmatrix} \cdot \begin{pmatrix} X_{t-1} \\ Y_{t-1} \end{pmatrix} \\ &+ \begin{pmatrix} \alpha_{12} & \beta_{12} \\ \alpha_{22} & \beta_{22} \end{pmatrix} \cdot \begin{pmatrix} X_{t-2} \\ Y_{t-2} \end{pmatrix} + \begin{pmatrix} a_{1t} \\ a_{2t} \end{pmatrix} \end{aligned}$$

- α_{11}, muestra la dependencia lineal de X_t con X_{t-1} en presencia de X_{t-2},Y_{t-1} e Y_{t-2}.
- β_{21}, mide la relación lineal entre Y_t y Y_{t-1} en presencia de X_{t-1}, X_{t-2} e Y_{t-2} .

$$\begin{aligned} X_t &= \delta_1 + \alpha_{11} X_{t-1} + \alpha_{12} X_{t-2} + \beta_{11} Y_{t-1} + \beta_{12} Y_{t-2} + a_{1t}, \\ Y_t &= \delta_2 + \alpha_{21} X_{t-1} + \alpha_{22} X_{t-2} + \beta_{21} Y_{t-1} + \beta_{22} Y_{t-2} + a_{2t}. \end{aligned}$$

$$\left(\begin{array}{c} X_t \\ Y_t \\ X_{t-1} \\ Y_{t-1} \end{array} \right) = \left(\begin{array}{c} \delta_1 \\ \delta_2 \\ 0 \\ 0 \end{array} \right) + \left(\begin{array}{cccc} \alpha_{11} & \beta_{11} & \alpha_{12} & \beta_{12} \\ \alpha_{21} & \beta_{21} & \alpha_{22} & \beta_{22} \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{array} \right) \cdot \left(\begin{array}{c} X_{t-1} \\ Y_{t-1} \\ X_{t-2} \\ Y_{t-2} \end{array} \right) + \left(\begin{array}{c} a_{1t} \\ a_{2t} \\ 0 \\ 0 \end{array} \right)$$

$$Z_t = \delta_t + \Phi Z_{t-1} + a_t$$

Dado el contraste:

$$\left. \begin{array}{l} H_0 : \text{el orden del VAR es } p_1 \\ H_1 : \text{el orden del VAR es } p_2 \end{array} \right\},$$

donde $p_1 < p_2$, se tiene que se rechaza la hipótesis nula si el valor experimental:

$$(T - m) \cdot \left(\ln \left(|\widehat{\Sigma}_{p_1}| \right) - \ln \left(|\widehat{\Sigma}_{p_2}| \right) \right),$$

es mayor que el valor teórico de una χ^2_r siendo r el número de parámetros de menos que hay que estimar en el modelo de la hipótesis nula frente al de la hipótesis alternativa, y donde:

- T corresponde al número de observaciones disponible,
- m es el número de parámetros a estimar bajo la hipótesis alternativa
- $|\widehat{\Sigma}_{p_1}|$ y $|\widehat{\Sigma}_{p_2}|$ son los determinantes de las estimaciones de las matrices de varianzas-covarianzas de los residuos de cada modelo.

La forma idónea de actuar es considerar inicialmente que $p_1 = 1$ y $p_2 = 2$ e ir aumentando gradualmente estos valores hasta que no se rechaze la hipótesis nula. En tal caso se tendría el orden idóneo para el modelo.

Para seleccionar el orden adecuado para el VAR es usar los distintos criterios de información: Akaike, Schwarz o Hannan-Quinn. Así, dado un criterio, se elegirá como orden idóneo del VAR el de aquel modelo que presente un menor valor.

Ejemplo. Determinar el orden del modelo VAR teniendo en cuenta:

Orden	AIC	BIC	HQC
1	2.0749	2.3043	2.1622
2	1.7702	2.1526	1.9158
3	1.8080	2.3434	2.0119
4	1.9331	2.6214	2.1952

Orden	AIC	BIC	HQC
1	2.0031	2.1621	2.0627
2	1.5355	1.8536	1.6547
3	1.5251	2.0021	1.7038
4	1.6646	2.3006	1.9028

El objetivo del **análisis de causalidad de Granger** es determinar si los retardos de una variable influyen en otra y, por tanto, son útiles para explicarla.

$$
\begin{aligned}
X_t &= \delta_1 + \alpha_{11} X_{t-1} + \cdots + \alpha_{1p} X_{t-p} + \beta_{11} Y_{t-1} + \cdots + \beta_{1p} Y_{t-p} + a_{1t}, \\
Y_t &= \delta_2 + \alpha_{21} X_{t-1} + \cdots + \alpha_{2p} X_{t-p} + \beta_{21} Y_{t-1} + \cdots + \beta_{2p} Y_{t-p} + a_{2t},
\end{aligned}
$$

se tiene que:

- Si los coeficientes β_{1j} (primera ecuación) son significativamente distintos de cero, mientras que los α_{2j} (segunda ecuación) no lo son, $j = 1, \ldots, p$, se dice que hay causalidad en el sentido de Granger de Y hacia X.
- Si los coeficientes α_{2j} (segunda ecuación) son significativamente distintos de cero, mientras que los β_{1j} (primera ecuación) no lo son, $j = 1, \ldots, p$, se dice que hay causalidad en el sentido de Granger de X hacia Y.
- Si ambos conjuntos de parámetros son estadísticamente significativos (distintos de cero), entonces se dice que hay causalidad en el sentido de Granger bidireccional.
- Si ninguno de los dos conjuntos es estadísticamente significativo (distinto de cero), entonces se dice que no hay ninguna relación de causalidad. En este caso, se deberían analizar las series de forma individual siguiendo un planteamiento unidimensional.

Contrastes F-Snedecor:

$H_0 : \beta_{11} = \beta_{12} = \cdots = \beta_{1p} = 0$
H_1 : Alguno de esos parámetros es distinto de cero

$H_0 : \alpha_{21} = \alpha_{22} = \cdots = \alpha_{2p} = 0$
H_1 : Alguno de esos parámetros es distinto de cero

Ejemplo. Para un modelo VAR(2) se ha obtenido la siguiente información sobre los contrastes de restricciones cero. El número de observaciones utilizadas son 54. ¿Existe relación de causalidad en el sentido de Granger entre ambas variables?:

	Ecuación 1ª variable	Ecuación 2ª variable
Todos los retardos de 1ª variable tienen coeficientes nulos de forma simultánea	$F_{2,49} = 31.295$ p-valor <0.001	$F_{2,49} = 3.7821$ p-valor$=0.03$
Todos los retardos de 2ª variable tienen coeficientes nulos de forma simultánea	$F_{2,49} = 0.086$ p-valor$=0.9175$	$F_{2,49} = 48.405$ p-valor < 0.001

Puesto que los coeficientes del VAR son díficiles de interpretar individualmente, para analizar la respuesta de una variable dependiente se estudia la función de respuesta al impulso.
El objetivo del análisis de la respuesta a un impulso radica en estudiar cómo afecta al análisis realizado una distorsión en los datos. Es decir, calcular los efectos (incremento o disminución) a lo largo del tiempo ocasionados por una distorsión en las perturbaciones del sistema. La distorsión corresponde a un aumento de una desviación estándar en una de las perturbaciones del modelo. Es decir, la distorsión en la ecuación i consiste en la raíz cuadrada del elemento (i, i) de la diagonal principal de $\mathbf{\Sigma_a}$.

Ejemplo. Para el modelo *VAR(1)* para dos variables sin término independiente:

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{pmatrix} 1{,}168 & -38{,}031 \\ 0{,}0017 & 0{,}7158 \end{pmatrix} \cdot \begin{pmatrix} X_{t-1} \\ Y_{t-1} \end{pmatrix} + \begin{pmatrix} \widehat{a}_{1t} \\ \widehat{a}_{2t} \end{pmatrix},$$

con:

$$\widehat{\Sigma} = \begin{pmatrix} 63{,}947 & 0{,}2528 \\ 0{,}2528 & 0{,}001324 \end{pmatrix},$$

Se pide analizar el efecto que tiene una distorsión de una desviación estándar en la perturbación asociada a la primera variable X sobre X e Y en los tres primeros periodos:

Teniendo en cuenta que $\sqrt{63{,}947} = 7{,}9967$:

$$\begin{pmatrix} X_1 \\ Y_1 \end{pmatrix} = \begin{pmatrix} 1{,}168 & -38{,}031 \\ 0{,}0017 & 0{,}7158 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 7{,}9967 \\ 0 \end{pmatrix} = \begin{pmatrix} 7{,}9967 \\ 0 \end{pmatrix},$$

$$\begin{pmatrix} X_2 \\ Y_2 \end{pmatrix} = \begin{pmatrix} 1{,}168 & -38{,}031 \\ 0{,}0017 & 0{,}7158 \end{pmatrix} \cdot \begin{pmatrix} 7{,}9967 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 9{,}3401 \\ 0{,}0136 \end{pmatrix},$$

$$\begin{pmatrix} X_3 \\ Y_3 \end{pmatrix} = \begin{pmatrix} 1{,}168 & -38{,}031 \\ 0{,}0017 & 0{,}7158 \end{pmatrix} \cdot \begin{pmatrix} 9{,}3401 \\ 0{,}0136 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 10{,}3923 \\ 0{,}0256 \end{pmatrix}.$$

Se puede observar que un incremento en X implica aumentos en X e Y.

Una de las ventajas de los modelos VAR sobre otro tipo de modelizaciones es su capacidad predictiva. A partir de la expresión :

$$\mathbf{Z}_t = \mathbf{\Phi}_0 + \mathbf{\Phi}_1 \mathbf{Z}_{t-1} + \mathbf{\Phi}_2 \mathbf{Z}_{t-2} + \cdots + \mathbf{\Phi}_p \mathbf{Z}_{t-p} + \mathbf{a}_t,$$

la predicción óptima de $\mathbf{Z}_t$ para k periodos futuros a partir de T observaciones vendrá dada por:

$$\hat{\mathbf{Z}}_T(k) = E[\mathbf{Z}_{T+k} | \mathbf{Z}_T, \ldots, \mathbf{Z}_1] \equiv E_T[\mathbf{Z}_{T+k}].$$

Recordad que:

$$E_T[\mathbf{a}_{T+s}] = 0, \; s > 0,$$

$$E_T[\mathbf{Z}_{T+s}] = \begin{cases} \hat{\mathbf{Z}}_T(s) & s > 0 \\ \mathbf{Z}_{T+s} & s \leq 0 \end{cases} .$$