

MODELOS ARIMA

Mayo 2001

Prof. Rafael de Arce
Prof. Ramón Mahía
Dpto. Economía Aplicada
U.D.I. Econometría e Informática

INTRODUCCIÓN

En 1970, Box y Jenkins desarrollaron un cuerpo metodológico destinado a identificar, estimar y diagnosticar modelos dinámicos de series temporales en los que la variable tiempo juega un papel fundamental. Una parte importante de esta metodología está pensada para liberar al investigador econométra de la tarea de especificación de los modelos dejando que los propios datos temporales de la variable a estudiar nos indiquen las características de la estructura probabilística subyacente. En parte, los procedimientos que vamos a analizar se contraponen a la "forma tradicional" de identificar y especificar un modelo apoyándonos en las teorías subyacentes al fenómeno analizado aunque, convenientemente utilizados, los conceptos y procedimientos que examinaremos constituyen una herramienta útil para ampliar y complementar los conocimientos econométricos básicos.

Se comenzará analizando los modelos en los que una variable es explicada utilizando exclusivamente una "exógena": su propio pasado. Podemos decir que la consideración exclusiva de los valores pasados de una determinada variable para explicar su evolución presente y futura supone, al mismo tiempo, una ventaja y un inconveniente:

- la ventaja radica en el hecho de no necesitar distintas series de datos (distintas variables) referidas al mismo período de tiempo (característica común a todos los modelos univariantes) y, al mismo tiempo, ahorrarnos la identificación y especificación del modelo en el sentido de la econometría tradicional,

- el inconveniente es que, al renunciar a la inclusión de un conjunto más amplio de variables explicativas, no atendemos a las relaciones que sin duda existen entre casi todas las variables económicas perdiendo capacidad de análisis al tiempo que renunciamos, implícitamente, al estudio teórico previo del fenómeno y a su indudable utilidad.

Dentro de estos modelos **univariantes** se desarrollarán suficientemente los conocidos con el nombre de ARIMA. Posteriormente se complementará esta perspectiva univariante añadiéndose a la especificación una o más variables exógenas al modelo "tradicional" aproximándonos al estudio de los conocidos como **modelos de transferencia**.

Como es habitual en economía, definiremos una estructura que nos permita, por sus características, cumplir el fin último de predicción: proceso estocástico estacionario. Diremos cuales son las condiciones que ha de cumplir esta función para que podamos calcularla y definiremos el proceso estocástico estacionario lineal y discreto. Posteriormente, analizaremos los modelos más simples (que emplean menos retardos) conforme a una serie de funciones características (covarianza, autocorrelación total y autocorrelación parcial), describiendo sus condiciones y planteando estructuras teóricas que luego puedan ser identificables con series temporales reales.

DEFINICIÓN Y CONCEPTOS BÁSICOS DE LOS MODELOS ARIMA

Proceso estocástico y estacionariedad

Los modelos autorregresivos o de medias móviles que más tarde conceptualizaremos necesitan para su comprensión de la introducción del concepto de proceso estocástico.

Un proceso estocástico es una sucesión de variables aleatorias Y_t ordenadas, pudiendo tomar t cualquier valor entre $-\infty$ y ∞ . Por ejemplo, la siguiente sucesión de variables aleatorias puede ser considerada como proceso estocástico:

$$Y_{-5}, Y_{-4}, Y_{-3}, Y_{-2}, \dots, Y_3, Y_4$$

El subíndice t no tiene, en principio, ninguna interpretación a priori, aunque si hablamos de proceso estocástico en el contexto del análisis de series temporales este subíndice representará el paso del tiempo.

Cada una de las variables Y_t que configuran un proceso estocástico tendrán su propia función de distribución con sus correspondientes momentos. Así mismo, cada par de esas variables tendrán su correspondiente función de distribución conjunta y sus funciones de distribución marginales. Esto mismo ocurrirá, ya no para cada par de variables, sino para conjuntos más amplios de las mismas. De esta forma, para **caracterizar un proceso estocástico** deberíamos especificar las funciones de distribución conjunta de cualquier conjunto de variables:

$$(Y_{t_1}, Y_{t_2}, Y_{t_3}, \dots, Y_{t_m})$$

cualesquiera que fueran los valores de (t_1, t_2, \dots, t_m) y cualquiera que fuera el valor de m ; por ejemplo:

$$y_1, y_2, y_3 \quad (t_1 = 1 \text{ y } m = 3)$$

$$Y_3, y_4, y_5, y_6 \quad (t_1 = 3 \text{ y } m = 4)$$

Habitualmente, conocer esas funciones de distribución resulta complejo de forma que, para caracterizar un proceso estocástico, basta con especificar la media y la varianza para cada y y la covarianza para variables referidas a distintos valores de t :

$$E[Y_t] = \mathbf{m}$$

$$s_t^2 = \text{Var}(y_t) = E[y_t - \mathbf{m}]^2$$

$$\mathbf{g}_t = \text{Cov}(Y_y, Y_s) = E[(y_t - \mathbf{m})(y_s - \mathbf{m})]$$

Las distribuciones de probabilidad podrían no estar completamente caracterizadas en algunas de las variables, los momentos podrían no coincidir incluso no existir para alguna de las variables aleatorias, lo mismo puede ocurrir con las distribuciones conjuntas o marginales. Sin embargo, de todos los tipos de procesos estocásticos posibles, nos interesan especialmente dos de ellos a los que la estadística ha dado nombres precisos:

- **ruido blanco** es una sucesión de variables aleatorias (proceso estocástico) con esperanza (media) cero, varianza constante e independientes para distintos valores de t (covarianza nula).
- proceso estocástico **estacionario**.

Decimos que un proceso estocástico es estacionario si las funciones de distribución conjuntas son invariantes con respecto a un desplazamiento en el tiempo (variación de t). Es decir, considerando que $t, t+1, t+2, \dots, t+k$ reflejan períodos sucesivos:

$$F(Y_t, Y_{t+1}, \dots, Y_{t+k}) = F(Y_{t+m}, Y_{t+1+m}, \dots, Y_{t+k+m})$$

para cualquier t, k y m ; por ejemplo:

$$F(y_1, y_2, \dots, y_6) = F(y_{10}, Y_{11}, \dots, y_{15})$$

$$\text{donde } t = 1, k = 5, m = 9$$

$$F(y_3, Y_4, Y_5) = F(y_7, Y_8, Y_9)$$

$$\text{donde } t = 3, k = 2, m = 4$$

Esta definición de estacionariedad se conoce como **estacionariedad en sentido estricto o fuerte** y puede relajarse sustancialmente utilizando la denominada **estacionariedad en sentido amplio o débil**. Decimos que un proceso estocástico es débilmente estacionario si:

- Las esperanzas matemáticas de las variables aleatorias no dependen del tiempo, son constantes:

$$E[Y_t] = E[Y_{t+m}] \quad \forall m$$

- Las varianzas tampoco dependen del tiempo y son finitas:

$$\text{Var}[Y_t] = \text{Var}[Y_{t+m}] \neq \infty \quad \forall m$$

- Las covarianzas entre dos variables aleatorias del proceso correspondientes a períodos distintos de tiempo (distintos valores de t) sólo dependen del lapso de tiempo transcurrido entre ellas:

$$\text{Cov}(Y_t, Y_s) = \text{Cov}(Y_{t+m}, Y_{s+m}) \quad \forall m$$

De esta última condición se desprende que, si un fenómeno es estacionario, sus variables pueden estar relacionadas linealmente entre sí, pero de forma que la relación entre dos variables sólo depende de la distancia temporal k transcurrida entre ellas.

Lógicamente, la estacionariedad en sentido estricto garantiza la estacionariedad en sentido amplio pero no al revés.

Una vez introducido el concepto genérico de proceso estocástico puede decirse que una **serie temporal** cualquiera es, en realidad, una muestra, una realización concreta con unos valores concretos de un proceso estocástico teórico, real. El análisis de series que vamos a estudiar tratará, a partir de los datos de una serie temporal, inferir las características de la estructura probabilística subyacente, del verdadero proceso estocástico.

Modelos autorregresivos

La palabra ARIMA significa Modelos Autorregresivos Integrados de Medias Móviles.

Definimos un modelo como **autorregresivo** si la variable endógena de un período t es explicada por las observaciones de ella misma correspondientes a períodos anteriores añadiéndose, como en los modelos estructurales, un término de error. En el caso de procesos estacionarios con distribución normal, la teoría estadística de los procesos estocásticos dice que, bajo determinadas condiciones previas, toda Y_t puede expresarse como una combinación lineal de sus valores pasados (parte sistemática) más un término de error (innovación).

Los modelos autorregresivos se abrevian con la palabra **AR** tras la que se indica el **orden** del modelo: AR(1), AR(2),...etc. El orden del modelo expresa el número de observaciones retrasadas de la serie temporal analizada que intervienen en la ecuación. Así, por ejemplo, un modelo AR(1) tendría la siguiente expresión:

$$Y_t = \mathbf{f}_0 + \mathbf{f}_1 Y_{t-1} + a_t$$

El término de error de los modelos de este tipo se denomina generalmente **ruido blanco** cuando cumple las tres hipótesis básicas tradicionales mencionadas al principio del texto:

- media nula
- varianza constante
- covarianza nula entre errores correspondientes a observaciones diferentes

La expresión genérica de un modelo autorregresivo, no ya de un AR(1) sino de un AR(p) sería la siguiente:

$$Y_t = \mathbf{f}_0 + \mathbf{f}_1 Y_{t-1} + \mathbf{f}_2 Y_{t-2} + \dots + \mathbf{f}_p Y_{t-p} + a_t$$

pudiéndose escribir de forma abreviada como:

$$\mathbf{f}_p(L)Y_t = \mathbf{f}_0 + a_t$$

donde $\phi_p(L)$ es lo que se conoce como operador polinomial de retardos:

$$\mathbf{f}_p(L) = 1 - \mathbf{f}_1 L - \mathbf{f}_2 L^2 - \dots - \mathbf{f}_p L^p$$

y donde, a su vez, el término L es lo que se conoce como operador retardo tal que, aplicado al valor de una variable en t , dé como resultado el valor de esa misma variable en $t-1$:

$$LY_t = Y_{t-1}$$

y aplicado sucesivamente p veces retarda el valor en p períodos

$$L^p Y_t = Y_{t-p}$$

Normalmente, se suele trabajar con modelos autorregresivos de órdenes bajos: AR(1) o AR(2), o bien con órdenes coincidentes con la periodicidad de los datos de la serie analizada (si es trimestral AR(4), si es mensual AR(12)...).

Modelo de medias móviles

Un modelo de los denominados de medias móviles es aquel que explica el valor de una determinada variable en un período t en función de un término independiente y una sucesión de errores correspondientes a períodos precedentes, ponderados convenientemente. Estos modelos se denotan normalmente con las siglas **MA**, seguidos, como en el caso de los modelos autorregresivos, del orden entre paréntesis. Así, un modelo con q términos de error MA(q) respondería a la siguiente expresión:

$$Y_t = m + a_t + q_1 a_{t-1} + q_2 a_{t-2} + \dots + q_q a_{t-q}$$

que de nuevo puede abreviarse utilizando el polinomio de retardos (como en el caso de los modelos AR):

$$Y_t = q_q(L) a_t + m$$

Al igual que en el caso de los modelos autorregresivos, el orden de los modelos de medias móviles suele ser bajo MA(1), MA(2) o corresponderse con la periodicidad de los datos analizados MA(4), para series trimestrales, o MA(12) para series mensuales.

Interpretación de un modelo de medias móviles

Así como un modelo autorregresivo es intuitivamente sencillo de comprender, la formulación de un modelo de medias móviles resulta sorprendente para el no iniciado. ¿Qué significa que una variable aleatoria se explique en función de los errores cometidos en períodos precedentes?, ¿De dónde proceden esos errores?, ¿Cuál es la justificación de un modelo de este tipo?

En realidad, un modelo de medias móviles puede obtenerse a partir de un modelo autorregresivo sin más que realizar sucesivas sustituciones.

Efectivamente, un modelo AR(1), sin término independiente, tiene la expresión:

$$Y_t = \mathbf{f}Y_{t-1} + a_t$$

si consideramos $t-1$ en lugar de t el modelo sería en este caso:

$$Y_{t-1} = \mathbf{f}Y_{t-2} + a_{t-1}$$

y sustituyendo queda:

$$Y_t = a_t + \mathbf{f}a_{t-1} + \mathbf{f}^2 Y_{t-2}$$

si ahora sustituimos y_{t-2} por su expresión autorregresiva y así sucesivamente llegamos a un modelo del tipo:

$$Y_t = a_t + \mathbf{f}a_{t-1} + \mathbf{f}^2 a_{t-2} + \mathbf{f}^3 a_{t-3} + \dots + \mathbf{f}^j a_{t-j} + \dots$$

que es la expresión, sin término independiente, de un modelo de medias móviles como el planteado anteriormente. En realidad, de forma estricta, el paso de un modelo a otro debería realizarse al contrario (de un MA a un AR) utilizando el teorema general de descomposición de Wold.

Condiciones y raíces unitarias para los modelos AR y MA

Hemos dicho anteriormente que, bajo condiciones generales, todo proceso estocástico estacionario se prestaba a una especificación tipo AR(p) y en consecuencia podía expresarse también como un MA(q). Es ahora el momento de especificar lo que antes hemos llamado "condiciones generales" y examinar en que casos es posible la realización de un proceso AR ó MA para representar un proceso estocástico estacionario.

Para que un proceso estocástico estacionario admita una formulación del tipo que aquí estudiaremos han de cumplirse dos condiciones accesorias:

- el proceso no debe ser anticipante (*hipótesis de recursividad temporal*); lo que quiere decir que los valores de una variable en un momento t no dependerán de los que esta misma tome en $t+j$, siendo j cualquier valor superior a cero.
- el proceso ha de ser invertible; lo que supone que la correlación entre una variable y su pasado va reduciéndose a medida que nos alejamos más en el tiempo del momento para el que estamos considerando dicha correlación (*proceso ergódico*). La explicación intuitiva de esta situación derivaría de que si el especificáramos una variable en función de ciertos coeficientes que nos determinen su correlación con los valores pasados de ella misma, los valores de dichos coeficientes deberían ser necesariamente inferiores a uno, porque sino el proceso de infinitos números sería "explosivo".

La estacionariedad de las series temporales en la realidad

Queda clara que la aproximación a los procesos estocásticos con modelos AR o MA está restringida, en términos generales, a aquellos procesos estocásticos que cumplan, al menos de forma débil, la restricción de estacionariedad. Cuando, en la realidad, queremos inferir a partir de una serie temporal (muestra) la estructura del proceso estocástico mediante modelos AR ó MA, debemos cubrir dos etapas:

- asegurarnos de que la serie temporal, como muestra del proceso estocástico, es estacionaria y, si no lo es,
- transformar la serie temporal original de forma que la nueva serie transformada si lo sea.

a) ¿cómo verificamos si la serie a analizar es estacionaria en media? ¿cómo lograr que lo sea? Filtrado de la serie original

Para resolver la primera cuestión existen diversos métodos de aproximación y, de entre ellos, destacamos Podríamos subdividir la serie temporal en varios períodos de, aproximadamente, la misma longitud, y calcular su media. El proceso sería estacionario en el caso en que dichos estadísticos fueran prácticamente iguales para todos los subperíodos analizados. En la mayoría de los casos, un simple gráfico sirve para observar si existe o no una clara tendencia y, por tanto, si la serie es estacionaria o no.

Habitualmente, cuando una serie muestra tendencia, se subdivide dicha serie en dos componentes: una primera, la estimación de dicha tendencia, y, la segunda, el residuo o error que se comete cuando se utiliza dicha tendencia como valor estimado de la serie original.

$$y_t = T_t + r_t$$

Una vez estimada la tendencia, aproximada con una regresión lineal, parabólica, exponencial ... que sea más conveniente; trabajaremos con la serie del residuo, que entonces no mostrara tendencia y podremos decir que es estacionaria en media. Es sobre este residuo sobre el que llevaremos a cabo todo el proceso descrito como metodología de identificación ARIMA, sumando finalmente el valor de la tendencia estimada si queremos dar resultados de estimación de la serie original. Es decir:

- 1.- la identificación del proceso ARIMA se hará sobre esta serie del residuo $\hat{r}_t = y_t - \hat{T}_t$, estimada previamente la tendencia del modo más adecuado.
- 2.- Para obtener valores estimados de la serie original, sumaremos el componente tendencial al valor estimado del residuo mediante el modelo ARIMA $\hat{y}_t = \hat{T}_t + \hat{r}_t$.

A este procedimiento se le conoce con el nombre de filtrado de la tendencia de la serie. Por supuesto, existen muy variadas formas de aplicar un filtro, siendo la que aquí hemos enunciado la más sencilla.

b) ¿Cómo se comprueba si una serie es estacionaria en varianza? Orden de integración

Sin duda alguna, el test más habitual a la hora de determinar la estacionariedad de una serie temporal, consiste en la aplicación del conocido como test de Dickey–Fuller (Test DF) o Dickey-Fuller

Ampliado (Test ADF). Éste es un contraste de “No estacionariedad” ya que la hipótesis nula es precisamente la presencia de una raíz unitaria en el proceso generador de datos de la serie analizada.

Vamos a suponer inicialmente, como modelo de partida para el análisis de una determinada serie y_t , el de un proceso estacionario autorregresivo de orden uno:

$$y_t = a_1 y_{t-1} + \mathbf{e}_t \quad (\text{Ec. 1})$$

frente a este modelo se plantea, como hipótesis nula H_0 , el modelo alternativo de un paseo aleatorio no estacionario del tipo¹:

$$y_t = y_{t-1} + \mathbf{e}_t$$

se trata por tanto de contrastar si el coeficiente a_1 es igual a la unidad o distinto de ella.

Sin embargo, para contrastar la nulidad del coeficiente a_1 , no podemos utilizar el contraste “t” habitual sobre una estimación por MCO del primer modelo. La razón es que la hipótesis nula que habitualmente se contrasta y, a partir de la cual se deriva la expresión y propiedades del test “t”, es la de nulidad del parámetro ($a_1=0$) de la (Ec.2), sin embargo, en nuestro caso, necesitaríamos contrastar $H_0: a_1=1$. Si la hipótesis nula fuera cierta, la varianza de y no sería estacionaria sino que crecería con los valores de “t” según la expresión de la varianza de un paseo aleatorio con deriva:

$$\text{Var}(y_t) = t\mathbf{S}_e^2$$

La estimación de a_1 en $y_t = a_1 y_{t-1} + \mathbf{e}_t$ será siempre consistente sin embargo, su distribución variará según los valores que tome la estimación. Utilizando las palabras de Novales (1993), la distribución de probabilidad asintótica del estimador de MCO del modelo AR(1) presenta una “discontinuidad” cuando $a_1=1$ y, como sustituto, deberán utilizarse las distribuciones derivadas de forma empírica mediante un procedimiento de Montecarlo realizado por Dickey (1976). Más recientemente, MacKinnon (1991) realizó un número mayor de simulaciones que las tabuladas por Dickey y Fuller. Además, MacKinnon estimó la superficie de respuesta usando los resultados de la simulación, lo que permite calcular los valores críticos del test DF para cualquier tamaño muestral y cualquier número de variables en el lado derecho de la ecuación.

En la práctica, por cuestiones de sencillez operativa, el modelo utilizado para el contraste DF no es el expuesto al comienzo del epígrafe sino otro, equivalente al anterior, que se obtiene restando a uno y otro lado el término y_{t-1} :

¹ $y_t = y_{t-1} + \mathbf{e}_t \rightarrow \Delta y_t = \mathbf{e}_t$ en un proceso de este tipo, sustituyendo recursivamente se obtiene:

$$y_t = y_0 + \sum_{i=1}^t \mathbf{e}_i, \text{ proceso con media constante } E[y_t] = E\left[y_0 + \sum_{i=1}^t \mathbf{e}_i\right] = E[y_0] = y_0 \text{ y varianza}$$

$$\begin{aligned} \text{estocástica: } V[y_t] &= E[y_t - E[y_t]]^2 = E\left[y_0 + \sum_{i=1}^t \mathbf{e}_i - y_0\right]^2 = E\left[\sum_{i=1}^t \mathbf{e}_i\right]^2 = \\ &= E[\mathbf{e}_1^2 + \mathbf{e}_2^2 + \dots + \mathbf{e}_1 \mathbf{e}_2 + \mathbf{e}_1 \mathbf{e}_3 + \dots] = E[\mathbf{e}_1^2 + \mathbf{e}_2^2 + \dots + \mathbf{e}_t^2] = t\mathbf{S}_e^2 \end{aligned}$$

$$\begin{aligned}
 y_t - y_{t-1} &= a_0 + a_1 y_{t-1} - y_{t-1} + \mathbf{e}_t \\
 \Delta y_t &= a_0 + (a_1 - 1) y_{t-1} + \mathbf{e}_t \quad (\text{Ec. 2}) \\
 \Delta y_t &= a_0 + \mathbf{g} \cdot y_{t-1} + \mathbf{e}_t
 \end{aligned}$$

por tanto, la hipótesis nula inicial para la (Ec. 2), se transforma ahora en $H_0: \gamma=0$ frente a $H_1: \gamma < 0$. Decir que γ es nulo es lo mismo que decir que $a_1=1$, o sea, que existe una raíz unitaria, decir que es menor que cero equivale a decir que a_1 es menor que la unidad (proceso autorregresivo estacionario)².

El procedimiento básico para la aplicación simple del test DF es, a partir de aquí, aparentemente sencillo. Se estima el modelo propuesto y se calcula el valor estimado de la “t” del parámetro analizado. Una vez calculado se compara con el valor empírico de referencia obtenido con las tablas de Dickey y Fuller o de MacKinnon. Si el valor estimado para γ es inferior en valor absoluto al tabulado dado un determinado nivel de confianza, admitiremos la hipótesis nula, o sea, la presencia de una raíz unitaria.

El modelo expuesto hasta el momento es el más simple posible, pero cabe que el modelo más adecuado a la realidad incluya otros términos, como una constante y/o una tendencia. Dolado *et al.* (1990) y Perron (1990) propusieron, entre otros autores, seguir un proceso en etapas a fin de garantizar el éxito en la elección del modelo de referencia en el mayor número de ocasiones:

- En primer lugar se estimaría el modelo menos restringido (con término constante y tendencia determinista).
- Dado que el principal error de esta táctica inicial consistiría en la escasa potencia del contraste para el rechazo de la hipótesis nula por inclusión de variables irrelevantes, si los valores críticos indican rechazo (ausencia de raíz unitaria), terminaríamos el procedimiento.
- En el caso de no rechazarse la hipótesis nula de presencia de una raíz unitaria, es decir, en el caso en que admitamos la presencia de una raíz unitaria ($H_0: \gamma=0$) pasaríamos ahora a examinar la significatividad del parámetro tendencial determinista a_2 . Dado que, en este punto, estaríamos bajo la hipótesis ya admitida de que $\gamma=0$, utilizaríamos el valor de referencia de $\tau_{\beta\tau}$ e incluso, para mayor seguridad, también el contraste conjunto ϕ_3 ($a_2=\gamma=0$).
- Si el término tendencial resulta significativo ($a_2 \neq 0$) contrastaremos de nuevo la presencia de una raíz unitaria ($H_0: \gamma=0$) pero utilizando entonces las tablas de una normal estandarizada. Sea cual sea el resultado del test con las nuevas tablas finalizaríamos aquí el contraste admitiendo o rechazando la presencia de una raíz unitaria.
- Si el término tendencial es no significativo, deberá replantearse el modelo inicialmente estimado pasándose a examinar otro con término constante pero sin esta tendencia determinista. Con este modelo se vuelve a analizar la presencia de una raíz unitaria ($\gamma=0$).
- En el caso en que, nuevamente, se sostenga la presencia de una raíz unitaria, se contrastará entonces la adecuación del término independiente a_0 bien con el contraste $\tau_{\alpha 1}$, bien con ϕ_1 . Si el término independiente resulta significativo usamos de nuevo las tablas de

² No se considera el caso de procesos autorregresivos explosivos en que $a_1 > 1$

una normal para contrastar la presencia de la raíz unitaria, concluyendo de nuevo aquí el contraste.

- Sólo si entonces la constante a_0 es no significativa se utiliza el modelo más simple como modelo de referencia contrastándose, de nuevo, la presencia de raíz unitaria. En este caso, no tiene cabida el uso de la distribución normal estandarizada.

Está claro que lo expuesto hasta este momento permite contrastar la presencia de una o más raíces unitarias en una determinada serie temporal para la que se supone un proceso AR(1). Sin embargo, muchas serie temporales se ajustan más adecuadamente a procesos autorregresivos de orden superior AR(2) o AR(3). No parece, por tanto, muy correcto, contrastar la presencia de una o más raíces unitarias utilizando siempre la estructura de un modelo AR(1) ya que las raíces unitarias pueden aparecer también en estructuras más complejas. Este problema da lugar a lo que se conoce como test de raíces unitarias de Dickey-Fuller Ampliado (DFA): si se quiere contrastar la presencia de una raíz unitaria en una serie que sigue un proceso AR(p), deberá aplicarse el procedimiento expuesto para el caso simple AR(1), pero suponiendo ahora del modelo:

$$\Delta y_t = a_0 + \mathbf{g} \cdot y_{t-1} + \sum_{i=2}^p \mathbf{b}_i \Delta y_{t-i+1} + \mathbf{e}_t$$

donde:

$$\mathbf{g} = - \left(1 - \sum_{i=1}^p a_i \right)$$

y:

$$\mathbf{b}_i = \sum_{j=1}^p a_j$$

MODELO ARIMA(p,d,q) SARIMA(P,D,Q)

En su forma más general el modelo ARIMA(p,d,q) ARIMA(P,D,Q)_s podría escribirse como:

$$Y_T = \mathbf{j}_1 Y_{T-1} + \mathbf{j}_2 Y_{T-2} + \dots + \mathbf{j}_{p_s+p+D_s+d} Y_{T-p_s-p-sD-d} + \\ + \mathbf{d} + U_T + \mathbf{q}_1 U_{T-1} + \dots + \mathbf{q}_{Q_s+q} U_{T-sQ-q}$$

Entendiendo que puede haber más de un proceso generador de la serie (en la parte regular y en la estacional) y escribiendo una combinación de los modelos MA(q) y AR(p) que han precisado de una serie de diferenciaciones "d" en la parte regular o "D" en la parte estacional para que fueran estacionarios.

FUNCIONES DE UN PROCESO ESTOCÁSTICO ESTACIONARIO

Definido un proceso estocástico como estacionario (ya sea de forma débil o fuerte), ya se ha comentado que si cumple las condiciones en sentido estricto, también cumple las condiciones en sentido débil. Siendo así, el proceso estaría perfectamente definido si conociéramos su media constante (μ), su varianza constante (σ) y la covarianza entre cada par de momentos diferentes en el

tiempo. Dicho esto:

- La **función de autocovarianza** vendrá definida por los distintos valores que tomaría dicha covarianza cuando cambiamos el lapso temporal entre las observaciones de la serie que manejamos. Analíticamente, se podría expresar como:

$$\mathbf{g}_j = cov(y_t, y_{t-j}) = E[(y_t - \mathbf{m})(y_{t-j} - \mathbf{m})]$$

donde, evidentemente, cuando el valor de "j" es cero, tendríamos la varianza de la función:

$$\mathbf{g}_0 = cov(y_t, y_{t-0}) = E(y_t - \mathbf{m})^2 = \mathbf{s}^2$$

- La **función de autocorrelación** se define igualmente como:

$$\mathbf{r}_j = \frac{cov(y_t, y_{t-j})}{\sqrt{var(y_t)} \sqrt{var(y_{t-j})}}$$

como nos encontramos ante un proceso definido como estacionario, la varianza es constante, por lo que podemos escribir:

$$\mathbf{r}_j = \frac{\mathbf{g}_j}{\mathbf{g}_0}$$

La función de autocorrelación de un proceso estocástico estacionario manifiesta las siguientes propiedades:

1.- $\mathbf{r}_0 = 1$

2.- $|\mathbf{r}_j| \leq 1$, ya que $|\gamma_j| \leq \gamma_0$

Esto asegura que el comportamiento de la función no sea explosivo.

3.- $\mathbf{r}_j = \mathbf{r}_{-j}$ (simetría)

dato que si un proceso es estacionario, la covarianza de dos variables aleatorias separadas por el mismo lapso de tiempo es la misma.

4.- $\lim_{j \rightarrow \infty} \mathbf{r}_j = 0$ (proceso ergódico).

Esta última propiedad, que define al proceso como "ergódico", es la que posibilita inferir valores de una serie en función de la información que sobre ella nos da su propio pasado, logrando estimadores consistentes. Sólo si se da esta propiedad, la pérdida de información al no considerar la influencia de los infinitos valores obtenidos en el pasado es cada vez más escasa e, incluirlos, añadiría poca información para la definición del proceso generador de datos que se intenta reproducir para aplicar al futuro.

- Además de las dos funciones anteriores, se puede definir una tercera conocida como **función de autocorrelación parcial**, con el fin de tener en cuenta los valores de correlación entre dos variables aleatorias separadas entre si "j" períodos y en función de los valores intermedios entre ellas. Es decir:

$$\Pi_j = \text{corr}(y_t, y_{t-j} / y_{t-1}, y_{t-2}, \dots, y_{t-j+1})$$

Si planteamos las mejores predicciones de y_t e y_{t-j} como los resultantes de plantear Mínimos Cuadrados Ordinarios (MCO) siendo el primero de ellos del siguiente modo:

$$\hat{y}_t = \mathbf{a}_1 y_{t-1} + \mathbf{a}_2 y_{t-2} + \dots + \mathbf{a}_{j-1} y_{t-j+1}$$

Se puede escribir la función de autocorrelación parcial, si la media es nula, como:

$$\Pi_j = \frac{\text{cov}[(y_t - \hat{y}_t)(y_{t-j} - \hat{y}_{t-j})]}{\sqrt{\text{var}(y_t - \hat{y}_t)}\sqrt{\text{var}(y_{t-j} - \hat{y}_{t-j})}}$$

Pudiendo demostrarse que:

$$\Pi_j = \frac{\mathbf{r}_j - \mathbf{a}_1 \mathbf{r}_{j-1} - \mathbf{a}_2 \mathbf{r}_{j-2} - \dots - \mathbf{a}_{j-1} \mathbf{r}_1}{1 - \mathbf{a}_1 \mathbf{r}_1 - \mathbf{a}_2 \mathbf{r}_2 - \dots - \mathbf{a}_{j-1} \mathbf{r}_{j-1}}$$

APLICACIÓN DE ESTAS FUNCIONES A MUESTRAS CONCRETAS

En este apartado se pretender especificar estimaciones de los valores que caracterizan el proceso estacionario del tipo que se está describiendo ya no para el proceso estocástico general, sino para una manifestación concreta de éste traslucida en una serie temporal. Habrá que estimar la media (μ), para lo que usaremos la media muestral; la varianza (γ_0) y la función de autocovarianza (γ_j), para lo que emplearemos la fórmula de la covarianza muestral) y la función de autocorrelación (ρ_j).

- **Media muestral.**

Como ya se ha dicho, el estimador $\mu = E(y_t)$ será la media muestral:

$$\bar{y} = \frac{1}{T} \sum_{t=1}^T y_t$$

dicho estimador cumplirá dos propiedades:

- a) Insensatez.- la esperanza de la media de la serie será igual a μ

Esto se demuestra ya que:

$$E(\bar{y}) = \frac{1}{T} \sum_{t=1}^T E(y_t) = \frac{1}{T} \sum_{t=1}^T \mu = \mu$$

- b) Consistencia.- es decir, la varianza se anula cuando ampliamos la muestra a la población

siendo el estimador insesgado, propiedad que se cumplirá siempre que se dé la siguiente condición, que no desarrollamos:

- Función de **autocovarianza muestral**.

El estimador de γ_j se obtendrá según la siguiente expresión:

$$C_j = \frac{1}{T-j} \sum_{t=j+1}^T (y_t - \bar{y})(y_{t-j} - \bar{y})$$

que, a pesar de ser sesgado, dicho sesgo será determinable y cada vez más reducido según se aumente la muestra.

- Función de **autocorrelación muestral**.

Para su cálculo se recurrirá al cociente de funciones de autocovarianza del siguiente modo:

$$r_j = \frac{g_j}{g_0} = \frac{C_j}{C_0}$$

- El estimador para la función de **autocorrelación parcial** a emplear se calculará según el método recursivo de Durbin del siguiente modo:

$$\hat{f}_{11} = r_1$$

$$\hat{f}_{j+1,j+1} = \frac{r_{j+1} - \sum_{i=1}^j \hat{f}_{ji} r_{j+1-i}}{1 - \sum_{i=1}^j \hat{f}_{ji} r_i}$$

$$\hat{f}_{ji}, i = \hat{f}_{ji} - \hat{f}_{j+1,j+1} \hat{f}_{j,j+1-i} \quad / i = 1 \dots j$$

Empleando estas fórmulas para los primeros casos, podremos escribir:

$$\hat{f}_{11} = r_1$$

$$\hat{f}_{22} = \frac{r_2 - \hat{f}_{11} r_1}{1 - \hat{f}_{11} r_1}$$

$$f_{21} = \hat{f}_{11} - \hat{f}_{22} r_1$$

$$\hat{f}_{33} = \frac{r_3 - \hat{f}_{21} r_2 - \hat{f}_{22} r_1}{1 - \hat{f}_{21} r_1 - \hat{f}_{22} r_2}$$

$$\hat{f}_{31} = \hat{f}_{21} - \hat{f}_{33} \hat{f}_{22}$$

$$\hat{f}_{32} = \hat{f}_{22} - \hat{f}_{33} \hat{f}_{21}$$

PROCESO ESTOCÁSTICO ESTACIONARIO LINEAL DISCRETO

Vamos a definir un nuevo caso especial de un proceso estocástico que nos permita luego intentar encontrar algo parecido en la realidad y que sea fácilmente identificable.

Se conoce por proceso estocástico estacionario lineal y discreto a aquel que puede expresarse como:

$$y_t = \mathbf{m} + a_t + \mathbf{q}_1 a_{t-1} + \mathbf{q}_2 a_{t-2} + \dots$$

donde:

- es lineal porque puede escribirse como combinación lineal de los errores.
- a_t es ruido blanco (esperanza y covarianza nulas y varianza constante).
- es discreto porque los lapsos temporales considerados son uniformes (no hay saltos temporales distintos entre las variables consideradas).

El siguiente paso será especificar las condiciones que nos aseguran que este proceso es estacionario, es decir que tiene media y varianza constantes y que su covarianza sólo varía cuando lo hace el espacio temporal que separa las observaciones empleadas para calcularla. Para ver estas condiciones, calcularemos los momentos de primer y segundo orden asegurando la estacionariedad en sentido débil.

- Media constante:

$$E(y_t) = E(\mathbf{m} + a_t + \mathbf{q}_1 a_{t-1} + \dots) = \mathbf{m} + \sum_{i=0}^{\infty} \mathbf{q}_i E(a_t)$$

$$/ \mathbf{q}_0 = 1 \text{ y } \sum_{i=1}^{\infty} \mathbf{q}_i = k < \infty$$

Es decir la media será constante en la medida en que exista la segunda parte del sumando que, al quedar multiplicada por la esperanza del "ruido blanco" será finalmente cero, y la media quedará igual a μ .

- Varianza constante:

$$\begin{aligned} \mathbf{g}_0 &= E(y_t - \mathbf{m})^2 = E(\mathbf{m} - \mathbf{m} + a_t + \mathbf{q}_1 a_{t-1} + \mathbf{q}_2 a_{t-2} + \dots)^2 = \\ &= E(a_t^2 + \mathbf{q}_1^2 a_{t-1}^2 + \mathbf{q}_2^2 a_{t-2}^2 + \dots + 2\mathbf{q}_1 a_t a_{t-1} + 2\mathbf{q}_2 a_t a_{t-2} + \dots) = \\ &= \mathbf{s}_a^2 (1 + \mathbf{q}_1^2 + \mathbf{q}_2^2 + \dots + 0 + 0) = \\ &= \mathbf{s}_a^2 \sum_{i=0}^{\infty} \mathbf{q}_i^2 \quad / \mathbf{q}_0 = 1 \end{aligned}$$

Siendo entonces condición necesaria para que la varianza exista que el último sumatorio sea calculable (converja).

- Covarianza constante:

$$\begin{aligned} \mathbf{g}_j &= E[(y_t - \mathbf{m})(y_{t-j} - \mathbf{m})] = \\ &= E[(a_t + \mathbf{q}_1 a_{t-1} + \mathbf{q}_2 a_{t-2} + \dots)(a_{t-j} + \mathbf{q}_1 a_{t-j-1} + \mathbf{q}_2 a_{t-j-2} + \dots)] = \\ &= \mathbf{q}_j E(a_{t-j}^2) + \mathbf{q}_1 \mathbf{q}_{j+1} E(a_{t-j-1}^2) + \mathbf{q}_2 \mathbf{q}_{j+2} E(a_{t-j-2}^2) + \dots = \\ &= \mathbf{s}_a^2 (\mathbf{q}_j + \mathbf{q}_1 \mathbf{q}_{j+1} + \mathbf{q}_2 \mathbf{q}_{j+2} + \dots) = \mathbf{s}_a^2 \sum_{i=0}^{\infty} \mathbf{q}_i \mathbf{q}_{i+j} \end{aligned}$$

Luego el proceso será estacionario en la medida en que se cumplan estas tres condiciones:

$$\sum_{i=0}^{\infty} q_i E(a_t) / q_0 = 1 \text{ y } \sum_{i=1}^{\infty} q_i = k < \infty$$

$$\sum_{i=0}^{\infty} q_i^2 < \infty \quad / \quad q_0 = 1$$

$$\sum_{i=0}^{\infty} q_i q_{i+j} < \infty$$

- **Ventajas:**

Una vez definido este proceso particular, vamos a ver resumidamente sus ventajas respecto a no contar con él.

En principio, si quisieramos definir un proceso estocástico en general, tendríamos, al menos, que definir sus momentos de primer y segundo orden, para lo cual sería necesario estimar T varianzas, T esperanzas y $(T^2 - T)/2$ covarianzas, lo que nos es imposible si sólo contamos con T datos.

Si el proceso fuera estacionario, ya sólo tendríamos que estimar una esperanza y una varianza (media y una varianza constantes) y (T-1) covarianzas ($\text{cov}(y_t, y_{t-1}), \text{cov}(y_t, y_{t-2}), \dots$); en total $1+1+(T-1)=T+1$ parámetros, lo que tampoco es posible. Si estamos ante un proceso estocástico estacionario lineal discreto, sólo necesitaremos contar con:

$$T > p + q + 2$$

siendo "p" y "q" los órdenes de los retardos de los modelos autorregresivos y de medias móviles que ya hemos definido anteriormente.

MODELOS MA(1)

Una vez tenemos definidas las ventajas de contar con un proceso estocástico estacionario lineal y discreto, y que podemos calcular las funciones de autocovarianza y autocorrelación, puede resultar interesante ver que valores toman éstas en aquellos casos sencillos que luego nos permitan comprobar si series económicas generales pueden tener un comportamiento similar, simplemente acudiendo a la comparación de sus correlogramas (de la función de autocorrelación total y parcial).

El primer caso a analizar será el modelo de medias móviles de orden uno, que se define como:

$$y_t = m + a_t + q_1 a_{t-1}$$

Este modelo también se puede escribir en función del operador retardo, ya comentado, del siguiente modo:

$$y_t = m + q(L) a_t$$

$$/ \quad q(L) = 1 - q_1 L$$

Se dice que un modelo MA(q) es invertible en la medida en que se pueda escribir como un proceso autorregresivo de orden infinito. Para que esta circunstancia pueda darse, será condición necesaria que las raíces de:

$$1 - q(L) = 0$$

caigan fuera del círculo unitario, lo que se cumplirá siempre que $|\theta| < 1$

Esta situación proviene de la conversión del modelo de medias móviles en modelo autorregresivo. Si escribimos el MA(1) como:

$$y_t = m + a_t + q_1 a_{t-1}$$

Podemos hacer sucesivas sustituciones hasta llegar al modelo autorregresivo:

$$a_{t-1} = y_{t-1} - m - q_1 a_{t-2}$$

$$y_t = m + q_1 (y_{t-1} - m - q_1 a_{t-2}) + a_t$$

$$y_t = m + a_t + q_1 y_{t-1} - q_1 m - q_1^2 y_{t-2} - \dots =$$

$$= q_1 y_{t-1} + q_1^2 y_{t-2} + \dots + m(1 + q_1 + q_1^2 + q_1^3 + \dots + q_1^{n-1}) + a_t - q_1^n a_{t-n}$$

si $n \rightarrow \infty$:

$$y_t = q_1 y_{t-1} + q_1^2 y_{t-2} + \dots + m \frac{1}{1 - q_1} + a_t$$

donde es necesario que $|\theta| < 1$ para que la progresión geométrica que se produce sobre los parámetros θ sea calculable y no explosiva.

El siguiente paso, una vez definida la condición de invertibilidad, es definir las funciones que se han descrito para los procesos estocásticos en general para el caso del modelo MA(1).

- Esperanza:

$$E(y_t) = E(\mathbf{m} + a_t + \mathbf{q}_1 a_{t-1}) = E(\mathbf{m}) + E(a_t) + \mathbf{q}_1 E(a_{t-1}) = \mathbf{m}$$

- Varianza:

$$\text{var}(y_t) = E(y_t - \mathbf{m})^2 = E(\mathbf{m} + a_t + \mathbf{q}_1 a_{t-1} - \mathbf{m})^2$$

$$\begin{aligned} E(a_t^2 + \mathbf{q}_1^2 a_{t-1}^2 + 2\mathbf{q}_1 a_t a_{t-1}) &= \\ &= \mathbf{s}_a^2 + \mathbf{q}_1^2 \mathbf{s}_a^2 + 0 = \end{aligned}$$

- Función de autocovarianza:

$$\mathbf{g}_1 = \text{cov}(y_t, y_{t-1}) = E((y_t - \mathbf{m})(y_{t-1} - \mathbf{m})) =$$

$$E((a_t + \mathbf{q}_1 a_{t-1})(a_{t-1} + \mathbf{q}_1 a_{t-2})) =$$

$$E(a_t a_{t-1} + \mathbf{q}_1 a_t a_{t-2} + \mathbf{q}_1 a_{t-1} a_{t-1} + \mathbf{q}_1^2 a_{t-1} a_{t-2}) =$$

$$\mathbf{q}_1 E(a_{t-1})^2 = \mathbf{q}_1 \mathbf{s}_a^2$$

$$\mathbf{g}_j = \text{cov}(y_t, y_{t-j}) = E((y_t - \mathbf{m})(y_{t-j} - \mathbf{m})) =$$

$$E((a_t + \mathbf{q}_1 a_{t-1})(a_{t-j} + \mathbf{q}_1 a_{t-j-1})) =$$

$$= E(a_t a_{t-j} + \mathbf{q}_1 a_{t-1} a_{t-j} + \mathbf{q}_1 a_t a_{t-j-1} + \mathbf{q}_1^2 a_{t-1} a_{t-j-1})$$

$$E(a_t a_{t-j} + \mathbf{q}_1 a_{t-1} a_{t-j} + \mathbf{q}_1 a_t a_{t-j-1} + \mathbf{q}_1^2 a_{t-1} a_{t-j-1}) = 0 = \mathbf{g}_j$$

Luego la función de autocovarianza se anula para valores de "j" mayores que uno y es una fracción de la varianza del error para el valor de j=1.

- **Función de autocorrelación:**

Calculada como cociente entre la función de autocovarianza y la varianza, tal y como ya se vio antes, tendríamos:

$$r_j = \frac{g_j}{g_0}$$

$$r_1 = \rho_1 s_a^2$$

$$r_j = 0 \quad /j > 1$$

- La **función de autocorrelación parcial** se calcularía siguiendo la siguiente expresión, que no demostramos:

$$\hat{f}_{jj} = \frac{\rho_1^j (1 - \rho_1^2)}{1 - \rho_1^{2(j+1)}}$$

Definidas las funciones características de los procesos estocásticos para el caso concreto del MA(1), podemos enunciar las siguientes **particularidades** de este tipo de proceso:

- 1.- Siempre es estacionario.
- 2.- Para ser invertible, es necesario que $|\theta| < 1$.
- 3.- La ρ_j sólo tiene un punto significativo. El modelo "olvida" la correlación con períodos distintos al inmediatamente anterior y el correlograma sólo tendrá un punto significativo.
- 4.- La función de autocorrelación parcial no se anula, pero tendrá un comportamiento amortiguado hacia cero.

MODELOS AR(1)

Definido el modelo AR(1) como:

$$y_t = \mathbf{f}_0 + \mathbf{f}_1 y_{t-1} + a_t$$

que también puede escribirse como:

$$\mathbf{f}(L) y_t = \mathbf{f}_0 + a_t$$

$$\mathbf{f}(L) = 1 - \mathbf{f}_1 L$$

a diferencia de los modelos de medias móviles, los autorregresivos no son estacionarios por definición y, para que lo sean, ha de cumplirse que las raíces de la siguiente ecuación sean mayores que uno:

$$\mathbf{f}(L) = 1 - \mathbf{f}_1 L = 0$$

lo que nos permitiría escribirlo como un modelo de medias móviles y, en definitiva, esto supone que los coeficientes $|\phi|$ han de ser menores que 1.

Pasamos a describir las funciones definidas:

- Esperanza matemática:

$$\begin{aligned} E(y_t) &= E(\mathbf{f}_1 y_{t-1} + \mathbf{f}_0 + a_t) = \\ &= \mathbf{f}_1 E(y_{t-1}) + \mathbf{f}_0 + 0 \end{aligned}$$

donde, como el proceso es estacionario, las esperanzas $E(y_t) = E(y_{t-1}) = \dots = E(y_{t-j}) = \mu$ y puedo escribir:

$$E(y_t) = \mathbf{m} = \mathbf{f}_1 \mathbf{m} + \mathbf{f}_0$$

$$\mathbf{m} = \frac{\mathbf{f}_0}{1 - \mathbf{f}_1}$$

- Varianza:

Para hacer los cálculos con mayor facilidad es conveniente poner el modelo autorregresivo en desviaciones a la media, sin que ello suponga ningún cambio en éste (se puede hacer la prueba escribiendo el modelo en desviaciones y llegando al modelo normal).

El modelo en desviaciones a la media lo definiremos como:

$$\tilde{y}_t = \mathbf{f}_1 \tilde{y}_{t-1} + a_t$$

donde:

$$\tilde{y}_{t-j} = y_{t-j} - \mathbf{m} = y_{t-j} - \frac{\mathbf{f}_0}{1 - \mathbf{f}_1}$$

Para calcular la varianza, escribimos el momento de segundo orden:

$$\begin{aligned} \text{var}(y_t) &= \mathbf{g}_0 = E(\mathbf{f}_1 \tilde{y}_{t-1} + a_t)^2 \\ &= E(\mathbf{f}_1^2 y_{t-1}^2 + a_t^2 + 2\mathbf{f}_1 \tilde{y}_{t-1} a_t) = \\ &= \mathbf{f}_1^2 E(\tilde{y}_{t-1})^2 + E(a_t)^2 + 2\mathbf{f}_1 E(\tilde{y}_{t-1} a_t) \end{aligned}$$

dado que el proceso es estacionario:

$$E(\tilde{y}_t)^2 = E(\tilde{y}_{t-1})^2 = \mathbf{g}_0$$

$$E(a_t)^2 = \mathbf{s}_a^2$$

como el proceso en desviaciones a la media se puede escribir como un proceso de medias móviles y por lo que hemos visto anteriormente, podríamos escribir:

$$E(a_t \tilde{y}_{t-h}) = \mathbf{s}_a^2 / h = 0 \quad \text{y } 0 \text{ si } h > 0$$

Por todo lo cual:

$$\mathbf{g}_0 = \mathbf{f}_1^2 \mathbf{g}_0 + \mathbf{s}_a^2$$

$$\mathbf{g}_0 = \frac{\mathbf{s}_a^2}{1 - \mathbf{f}_1^2}$$

- Función de autocovarianza:

$$\begin{aligned}\mathbf{g}_1 &= cov(\tilde{y}_t, \tilde{y}_{t-1}) = E(\tilde{y}_{t-1}(\mathbf{f}_1 \tilde{y}_{t-1} + a_t)) = \\ &= \mathbf{f}_1 \mathbf{g}_0 = \mathbf{f}_1 \frac{\mathbf{s}_a^2}{1 - \mathbf{f}_1^2}\end{aligned}$$

$$\begin{aligned}\mathbf{g}_2 &= cov(\tilde{y}_t, \tilde{y}_{t-2}) = E(\tilde{y}_{t-2}(\mathbf{f}_1 \tilde{y}_{t-1} + a_t)) = \\ &\mathbf{f}_1 \mathbf{g}_1 = \mathbf{f}_1^2 \mathbf{g}_0 = \mathbf{f}_1^2 \frac{\mathbf{s}_a^2}{1 - \mathbf{f}_1^2}\end{aligned}$$

Lo que generalizando, se podría escribir como:

$$\mathbf{g}_j = \frac{\mathbf{s}_a^2}{1 - \mathbf{f}_1^2} \mathbf{f}_1^j$$

- Función de autocorrelación total

$$\mathbf{r}_j = \mathbf{f}_1^j \quad / j \geq 1$$

- Función de autocorrelación parcial

$$\mathbf{f}_{jj} = \mathbf{r}_1 \quad / j = 1 \quad \text{y} \quad 0 \quad \text{si} \quad j > 1$$

Definidas todas estas funciones, podemos caracterizar el proceso autorregresivo del siguiente modo:

- 1.- Siempre es invertible (está directamente invertido).
- 2.- Para ser estacionario, ha de cumplirse que $|\phi| < 1$.
- 3.- La función de autocorrelación total no se anula, pero se va amortiguando hacia cero.
- 4.- La función de autocorrelación parcial se anula para retardos superiores a uno.

	FAC	FAP
MA(q)	Se anula para retardos superiores a q	Decrecimiento rápido sin llegar a anularse
AR(p)	Decrecimiento rápido sin llegar a anularse	Se anula para retardos superiores a p
ARMA(p,q)	Decrecimiento rápido sin llegar a anularse	Decrecimiento rápido sin llegar a anularse

IDENTIFICACIÓN DEL MODELO

Aunque podríamos seguir definiendo las características de otros procesos ARIMA de órdenes mayores, no tiene mayor interés una vez se ha entendido el procedimiento y sí interesa precisar la forma que adoptarían los correlogramas de estas funciones porque, fruto de su comparación con los que obtendremos con nuestras series de interés, podremos asociar a nuestra serie temporal de estudio un proceso ARIMA que identifique su proceso generador de datos, tanto a pasado como a futuro.

Los correlogramas o funciones de autocorrelación total y parcial están disponibles en el libro de Pulido,A(1989): "Predicción Económica y Empresarial" Editorial Pirámide.

De forma muy poco académica, el proceso de identificación consistirá en calcular las funciones de autocorrelación total y parcial de nuestra serie (una vez estamos seguros de que ésta cumple las condiciones que definen un proceso estocástico estacionario) y comparar sus correlogramas con los correspondientes a los modelos teóricos AR(p),MA(q) o ARMA(p,q).

En principio, si el proceso está bien identificado, procederemos a su estimación y, si analizamos los correlogramas de los residuos obtenidos en la estimación, serán "ruido blanco". Si esto no es así, habrá que realizar una nueva estimación incorporando la estructura más parecida al modelo teórico que podamos intuir con la comparación con los modelos teóricos.

Para saber cuando estamos ante un "ruido blanco", se pueden hacer las siguientes comprobaciones:

- Media nula

Puede observarse en el gráfico de residuos si el error se mueve en torno al valor cero o bien calcularse el cociente entre la media y la varianza muestral de los residuos. Si ese ratio es inferior a 2, podemos concluir (con un $\epsilon=0,05$) que la media no es significativamente distinta de cero.

- Varianza constante

Observando el gráfico de los residuos puede analizarse la constancia de la varianza del error. En caso de heterocedasticidad y es recomendable una transformación logarítmica en la serie original.

- Incorrelación

Deben observarse los coeficientes de autocorrelación muestral de los residuos y comprobar que ninguno de ellos supera el valor de las bandas de significatividad al 5% ($\pm 1,96(1/T^{1/2})$). El valor $T^{1/2}$ es una aproximación de la varianza asintótica pero resulta sólo adecuada para valores grandes de "j". Se aconseja, por tanto, utilizar distinta amplitud de bandas como por ejemplo $\pm(1/T^{1/3})$ para los términos más cercanos a cero.

El estudio de las funciones de autocorrelación muestral y autocorrelación parcial muestral de los residuos, pueden servirnos fácilmente para el replanteamiento del modelo inicial.

El contraste de la "Q" de Box-Pierce analiza la hipótesis nula de que:

$$H_0: \rho_1(a) = \rho_2(a) = \rho_3(a) = \dots = \rho_M(a) = 0$$

suponiendo que la expresión:

$$Q = T \sum_{j=1}^M r_j^2(\hat{a})_t$$

o la alternativa propuesta por Ljung-Box:

$$Q^* = T(T+2) \sum_{j=1}^M (T-j)^{-1} r_j^2(\hat{a})_t$$

se distribuye como una chi-cuadrado con M-k grados de libertad. Otros contrastes alternativos son los de Vandaele (1983), que analiza las autocorrelaciones muestrales de los residuos transformados mediante una diferencia regular y el de Peña (1983).

BREVE RESEÑA SOBRE LA ESTIMACIÓN DE LOS MODELOS ARIMA

Se analizará a continuación el proceso de estimación de los modelos ARMA(p,q) x ARMA(P,Q), centrando los desarrollos en el caso específico de un modelo sin componente estacional, es decir, un ARMA(p,q) ó ARIMA (p,d,q):

$$(1 - \mathbf{f}_1 L - \mathbf{f}_2 L^2 - \dots - \mathbf{f}_p L^p) w_t = \mathbf{d} + (1 - \mathbf{q}_1 L - \mathbf{q}_2 L^2 - \dots - \mathbf{q}_q L^q) a_t$$

donde se entiende por W_t la serie ya en diferencias y donde, como siempre, a_t se supone un ruido blanco con media cero y varianza constante (σ_u^2).

Definida esta función genérica, el objetivo principal es estimar el vector formado por los parámetros correspondientes a la parte autorregresiva ϕ_j y de medias móviles θ_j (incluido, si fuera necesario, el término independiente) así como la varianza residual.

Problemas iniciales: los valores iniciales y la no linealidad.

La naturaleza del modelo implica que la variable a explicar se hace depender de valores pasados de la misma y errores cometidos en la estimación de dichos valores pasados. De esta forma el planteamiento de minimización de los errores como procedimiento de estimación lleva necesariamente aparejada **la necesidad de conocer valores pasados de la variable endógena y**

de los errores ya que la expresión del error, por ejemplo para un período "t" sería:

$$a_t = w_t - \mathbf{f}_1 W_{t-1} - \dots - \mathbf{f}_p W_{t-p} - \mathbf{d} + \mathbf{q}_1 u_{t-1} - \dots - \mathbf{q}_q a_{t-q}$$

Al conjunto de valores iniciales requeridos de la variable endógena desde "t-1" a "t-p" y de los errores desde "t-1" a "t-q" los notaremos por los vectores:

$$W^{o'} = (W_{t-1}, W_{t-2}, \dots, W_{t-p})$$

$$a^{o'} = (a_{t-1}, a_{t-2}, \dots, a_{t-q})$$

El procedimiento de estimación que lleva implícita la especificación a priori de unos valores iniciales se denomina "enfoque condicional", mientras que aquel en el que se estiman simultáneamente los valores iniciales y los parámetros se denomina "enfoque exacto".

Además de este primer problema, se señala el de la **no linealidad** del modelo cuando este incluye medias móviles, lo que puede comprobarse fácilmente a partir de una transformación de la especificación de un modelo sencillo [por ejemplo MA(1)]:

$$Y_t = a_t - \mathbf{q}_1 a_{t-1} \rightarrow a_t = Y_t + \mathbf{q}_1 a_{t-1}$$

↓

$$\text{para } t = 1 \quad a_1 = Y_1 + \mathbf{q}_1 a_0$$

$$\text{para } t = 2 \quad a_2 = Y_2 + \mathbf{q}_1 a_1 =$$

$$= Y_2 + \mathbf{q}_1 (Y_1 + \mathbf{q}_1 a_0) =$$

$$= Y_2 + \mathbf{q}_1 Y_1 + \mathbf{q}_1^2 a_0$$

Partiendo de esta expresión, se observa como al minimizar:

$$\sum_{t=1}^T \hat{a}_t^2$$

(3)

³ El término " T " hace referencia a los valores disponibles de la serie una vez tomadas diferencias tanto en la componente regular (d) como en la estacional (D):

en el proceso de estimación, esta expresión no será lineal. Por ello, una primera conclusión es que sea cual sea el método de estimación utilizado (Mínimos cuadrados o Máxima verosimilitud) deberán aplicarse algoritmos de resolución no lineales.

PREDICTOR ÓPTIMO.

"La predicción es el fin último y primordial del análisis univariante de series temporales"⁴

Una vez identificado y estimado el modelo ARIMA, se plantea su utilización para conseguir la mejor predicción de los valores a futuro de una serie a partir de su propia historia. El primer interrogante que surge se referirá a la determinación del **PREDICTOR ÓPTIMO** para este fin.

Intuitivamente, el mejor predictor posible será "el que menos se equivoca" o, en términos estadísticos, aquel que minimiza el error cuadrático medio respecto a otro potencial predictor alternativo. Esto se puede expresar:

$$E[(Y_{T+l} - \hat{Y}_T(l))^2 | I_T] \leq E[(Y_{T+l} - \hat{Y}_T^*(l))^2 | I_T]$$

donde $\hat{Y}_T(l)$ sería el valor de predicción de la serie para el período (T+l), condicionado a los valores históricos de Y_T ($Y_T = Y_{T-1}, Y_{T-2}, \dots$).

Se demuestra que el predictor elegido es óptimo cuando su valor esperado es igual al valor real de predicción condicionado a la información existente en el período T respecto a la serie que nos ocupa; es decir:

$$\hat{Y}_T(l) = E[Y_{T+l} | I_T]$$

El error cuadrático medio de un predictor arbitrario siempre es mayor que aquel cuyo valor coincide con la esperanza del valor real en el período que estemos considerando. Esta propiedad será fundamental para el posterior desarrollo de la predicción puntual.

PREDICCIÓN PUNTUAL

Partiendo de un modelo ARIMA sobre el que se han realizado una serie de diferenciaciones para lograr una serie estacionaria, el planteamiento de la predicción se hace sobre los valores reales de la serie, por entender que es de éstos de los que se quiere obtener valores a futuro. Esta circunstancia deberá ser tomada en cuenta a la hora de interpretar los subíndices que acompañan a las fórmulas de este apartado.

La aparición de órdenes autorregresivos superiores a "p" será debida a dos circunstancias: traslación del modelo a períodos fuera del espacio temporal conocido y aplicación de las expresiones

$$T = N - d - sD$$

⁴TRIVEZ, F.J. y AZNAR, A.: "Métodos de predicción en economía" Volumen II, Análisis de series temporales" Pgs.240-267. Editorial Ariel, 1993.

obtenidas para $W_T = (1 - L)^d Y_T$ directamente a los valores de la serie sin transformar (sustituir W_T por su valor en cada caso).

En su forma más general el modelo ARIMA(p,d,q) ARIMA(P,D,Q)_S podría escribirse como:

$$Y_T = \mathbf{j}_1 Y_{T-1} + \mathbf{j}_2 Y_{T-2} + \dots + \mathbf{j}_{p_s+p+D_s+d} Y_{T-p_s-p-sD-d} + \\ + \mathbf{d} + a_T + \mathbf{q}_1 a_{T-1} + \dots + \mathbf{q}_{Q_s+q} a_{T-sQ-q}$$

El desarrollo teórico se hace para un modelo sin parte estacional, sin que ello suponga pérdida de generalidad en los resultados. El modelo podría escribirse entonces como:

$$\hat{\mathbf{f}}_p(L)W_t = \mathbf{d} + \mathbf{q}_q^*(L)a_t$$

como sabemos que $W_T = (1 - L)^d Y_T$, la expresión anterior se puede reescribir como:

$$\mathbf{j}_p^*(L)Y_t = \mathbf{d} + \mathbf{q}_q^*(L)a_t$$

donde:

$$\mathbf{j}(L)Y_T = 1 - \mathbf{j}_1 L - \mathbf{j}_2 L^2 + \dots + \mathbf{j}_{p+d} L^{p+d} +$$

$$\mathbf{q}(L) = 1 - \mathbf{q}_1 L + \dots + \mathbf{q}_q L^q$$

El modelo ARIMA correspondiente sería:

$$Y_T = \mathbf{j}_1 Y_{T-1} + \mathbf{j}_2 Y_{T-2} + \dots + \mathbf{j}_{p+d} Y_{T-p-d} + \\ + \mathbf{d} + a_T + \mathbf{q}_1 a_{T-1} + \dots + \mathbf{q}_q a_{T-q}$$

En la medida en que aparece la perturbación aleatoria en la definición de cada valor de predicción de la serie Y_T , el modelo es susceptible de ser escrito como función infinita de los valores de la perturbación aleatoria del período considerado y de los anteriores (no habría más que despejar el valor de la perturbación aleatoria de la expresión general de un modelo ARIMA).

Para hacer predicción hay que tener en cuenta dos supuestos iniciales:

- 1.- Los parámetros de las funciones presentadas son conocidos.
- 2.- Las perturbaciones aleatorias se conocen en el período muestral y tiene carácter de

ruido blanco para los períodos de predicción.

Teniendo en cuenta ambos supuestos, se puede especificar un modelo ARIMA para definir el valor de predicción en función de la serie en "p" períodos anteriores (por la parte autorregresiva) y de los "q" errores cometidos al estimar la serie en los "q" períodos previos (por la parte de las medias móviles). Dado que son los valores reales aquellos sobre los que es verdaderamente interesante hacer predicción, se plantea la ecuación de predicción de un período (T+1) como:

$$\hat{Y}(l) = \mathbf{j}_1 \hat{Y}_T(l-1) + \mathbf{j}_2 \hat{Y}_T + \mathbf{j}_{l-1} Y_T(l) + \dots + \\ + \mathbf{j}_{l+1} Y_{T-1} + \dots + \mathbf{j}_{p+d} Y_{T+l-p-d} + \mathbf{d} - \mathbf{q}_1 a_T - \mathbf{q}_{l+1} a_{T-1} - \dots - \mathbf{q}_q a_{T+l-q}$$

Es conveniente remarcar tres aspectos de la predicción siguiendo esta formulación:

- Los valores de predicción se calculan de forma secuencial. Para hacer posible la aplicación de la parte autorregresiva del modelo en períodos distintos al primero de predicción, se toma el valor predicho en el inmediatamente anterior ($\hat{Y}_T(l-1)$).

- Por la condición necesaria para estar ante el predictor óptimo (el valor esperado de la serie en el período de predicción es igual al predicho si es óptimo), se demuestra que las perturbaciones aleatorias empleadas en la predicción son sólo las de períodos anteriores, es decir:

$$a_{T+j} \quad \forall \quad j \leq 0$$

ya que se ha supuesto que en el período de predicción, las perturbaciones aleatorias tienen esperanza nula.

Para realizar las sucesivas predicciones de $\hat{Y}_T(l)$ tenemos que contar con unos valores para $a_T, a_{T-1}, \dots, a_{T+l-q}$. Como tales se tomarán:

$$a_{T+j} = Y_{T+j} - \hat{Y}_{T+j-l}(l) \quad \forall \quad j \leq 0$$

- Las predicciones ARIMA son adaptativas y los resultados obtenidos para (T+1), con la información disponible hasta el período T, son los mismos que las que obtendríamos para el mismo período tomando como base informativa hasta T-1, y añadiendo un término de error.

CARACTERÍSTICAS DE LAS PREDICCIONES REALIZADAS CON MODELOS ARIMA

- Modelos AR(p): la predicción tiende a μ (media del proceso) a medida que aumenta el horizonte temporal de la predicción.

- Modelos MA(q): dada la memoria limitada que caracteriza a estos procesos, la predicción es

igual a μ (media del proceso) cuando el horizonte temporal de la predicción es mayor que el orden del proceso (q).

- **Modelos ARMA(p,q)**: a partir de "q" períodos futuros la predicción tiende a μ (media del proceso) a medida que aumenta el horizonte temporal de la predicción.

- **Modelos ARI(p,d) e IMA(d,q)**: la predicción ya no tiende a μ sino que será una línea recta que parte de $\hat{Y}(1)$ con pendiente igual a la media del proceso w_T (serie resultante de las transformaciones necesarias para hacerla estacionaria).

SELECCIÓN DE MODELOS

Si entendemos que una predicción es mejor que otra cuando comete menor error, los criterios de selección de modelos serían el **error cuadrático medio** (ECM), **error absoluto medio** (EAM) y **error absoluto porcentual medio** (EAPM). Estos indicadores se calcularían a período histórico, es decir, se calcularían los valores que el modelo ofrece para las H últimas observaciones y se compararían con el valor real, del siguiente modo:

$$ECM(H) = \frac{1}{H} \sum_{l=1}^H e_{T-H}^2(l) = \frac{1}{H} \sum_{l=1}^H [y_{T-H+l} - \hat{y}_{T-H}(l)]^2$$

$$EAM(H) = \frac{1}{H} \sum_{l=1}^H |e_{T-H}(l)| = \frac{1}{H} \sum_{l=1}^H |y_{T-H+l} - \hat{y}_{T-H}(l)|$$

$$EAPM(H) = \frac{1}{H} \sum_{l=1}^H \frac{|e_{T-H}(l)|}{y_{T-H+l}} * 100 = \frac{1}{H} \sum_{l=1}^H \frac{|y_{T-H+l} - \hat{y}_{T-H}(l)|}{y_{T-H+l}} * 100$$

El problema es que estos indicadores no tienen en cuenta la estructura estocástica del modelo, no informan sobre alguna característica estocástica supuesta sobre el período extramuestral.

ALGUNAS REFERENCIAS

AZNAR, A. Y TRIVEZ, F.J.(1993): Métodos de Predicción en Economía II. Análisis de series temporales
Editorial Ariel Economía, Barcelona 1993.

ESPASA, A. Y CANCELO, J.R. (1993): Métodos cuantitativos para el análisis de la coyuntura Económica
Alianza Editorial, Madrid 1993.

PULIDO, A. Y PÉREZ GARCÍA, J. (2001): Modelos Econométricos
Editorial Pirámide, Madrid 2001.

Enders, W. (1995). Applied Econometric Times Series.
John Wiley & Sons, Inc. United States.

Mahía Casado, R (1999): Procedimientos de análisis de la estacionariedad: Integración y raíces unitarias. Documento de trabajo Instituto de Predicción Económica LR Klein 99/1 (disponible en <http://www.uam.es/klein>, documentos de trabajo.