

DETECCIÓN Y TRATAMIENTO DE UNA COMPONENTE ESTACIONAL ESTABLE EN UNA SERIE TEMPORAL

ALBERTO PRAT BARTRES

En el presente trabajo se analiza la detección de una estacionalidad casi determinista en una serie temporal, mediante la utilización de la función de verosimilitud exacta en la fase de estimación del modelo univariante de la serie. Las posibles formas de modelar dicha estacionalidad se discuten sobre un ejemplo concreto.

1. INTRODUCCION

Muchos modelos estadísticos se han expresado tradicionalmente en la forma:

$$Z_t = f(\underline{x}_{1t} | \beta) + N_t(\underline{x}_{2t}) \quad (1.1)$$

donde f es una cierta función de las variables explicativas \underline{x}_1 que, sobretodo en ingeniería, pueden ser objeto de un diseño experimental y por lo tanto medidas y controladas. N_t es el "ruido" que recoge el efecto de las variables \underline{x}_2 en general desconocidas y no medibles. De hecho, el objetivo de cualquier investigación científica suele tener como objeto el representado esquemáticamente por la flecha de la ecuación (1.1), es decir, extraer del ruido alguna variable adicional e incluirla dentro de las variables explicativas.

A menudo se asume que el ruido N_t es independiente de las variables \underline{x}_{1t} y que

$\{N_t\}$ ($t = 1, 2, \dots$) es una secuencia de variables aleatorias, independientes, idénticamente distribuidas cuya familia de densidades de probabilidad es la $N(0; \sigma^2)$ es decir, normal con media cero y variancia σ^2 . Una secuencia de estas características recibe el nombre de ruido blanco.

En ciertos casos, la física del sistema que se está estudiando o, por ejemplo, la teoría

económica (si estamos analizando datos tales como la evolución de los depósitos en las Cajas de Ahorro), son suficientes para determinar la forma funcional de f en (1.1).

Creo que estas situaciones son bastante frecuentes en ingeniería pero relativamente raras en el terreno económico.

Ello explicaría el auge de la formulación en el espacio de los estados de muchos problemas del área de la ingeniería de control --- / ver Kalman (1960)/ mientras que en Economía el enfoque más utilizado es el de Box y Jenkins (1970) que representa un enfoque empírico al problema en el que la obtención de un modelo del tipo (1.1) se logra mediante un ciclo iterativo de identificación-estimación-verificación.

En el presente trabajo vamos a concentrarnos en modelos del tipo (1.1) pero en los que la variable explicativa es el tiempo, t . es decir:

$$Z_t = f(t) + N(\underline{x}_{2t}) \quad (1.2)$$

Un ejemplo muy conocido de (1.2) es el modelo

$$Z_t = \beta_0 + \beta_1 t + a_t \quad \{a_t\} \text{ es ruido blanco}$$

Modelos de este tipo serán raramente adecuados

- Alberto Prat Bartés - Escuela Técnica Superior de Ingenieros Industriales de Barcelona - Av. Diagonal, 647 Barcelona, 28

- Article rebut el Maig de 1982.

dos ya que:

a) Suponen la independencia estocástica del ruido y esta hipótesis no será en general, adecuada debido a que la inercia de las variables x_2 producirá efectos de autocorrelación.

b) La tendencia, que en nuestro caso es $\beta_0 + \beta_1 t$, se supone determinista, es decir, su forma funcional se considera invariante en el tiempo.

La metodología de análisis de series temporales propuesta por Box y Jenkins (*) permite flexibilizar la clase de modelos (1.2) para evitar, entre otros, los dos problemas que acabamos de mencionar.

En efecto, consideremos el modelo: ARIMA (0,2,2):

$$(1 - B)^2 z_t = (1 - \theta_1 B - \theta_2 B^2) a_t \quad (1.3)$$

donde B es el operador de retardo: $Bf(t) = f(t-1)$ y $\{a_t\}$ es ruido blanco y donde el polinomio de medias móviles $(1 - \theta_1 B - \theta_2 B^2)$ cumple las condiciones de invertibilidad (es decir, sus ceros están fuera del círculo unidad)

El modelo (1.3) puede escribirse:

$$z_t = 2z_{t-1} - z_{t-2} + a_t + \theta_1 a_{t-1} - \theta_2 a_{t-2}$$

En condiciones muy generales, la predicción del valor de z_{t+l} ($l = 1, 2, \dots$) hecha en el instante t será:

$$\hat{z}_t(l) = E[z_{t+l} | z_t, z_{t-1}, \dots] \quad (1.4)$$

De ahí deducimos:

$$\text{Para } l = 1 \quad z_{t+1} = 2z_t - z_{t-1} + a_{t+1} - \theta_1 a_t - \theta_2 a_{t-1}$$

de donde:

$$\begin{aligned} \hat{z}_t(1) &= E(z_{t+1} | z_t, z_{t-1}, \dots) = \\ &= 2z_t - z_{t-1} - \theta_1 a_t - \theta_2 a_{t-1} \end{aligned}$$

ya que $E(z_{t-k} | z_t, z_{t-1}, \dots) = z_{t-k}$ para $k > 0$

$$E(a_{t-k} | z_t, z_{t-1}, \dots) = a_{t-k} \text{ para } k > 0$$

$$E(a_{t+k} | z_t, z_{t-1}, \dots) = 0 \text{ para } k > 0$$

Análogamente tendremos:

$$\text{Para } l = 2 \quad z_{t+2} = 2z_{t+1} - z_t + a_{t+2} - \theta_1 a_{t+1} - \theta_2 a_t$$

$$\begin{aligned} \hat{z}_t(2) &= E(z_{t+2} | z_t, z_{t-1}, \dots) = \\ &= 2\hat{z}_t(1) - z_t - \theta_2 a_t \end{aligned}$$

y finalmente, para $l > 3$

$$\begin{aligned} \hat{z}_t(l) &= E(z_{t+l} | z_t, z_{t-1}, \dots) = \\ &= 2\hat{z}_t(l-1) - \hat{z}_t(l-2) \end{aligned}$$

Esta ecuación recurrente se puede expresar:

$$\hat{z}_t(l) - 2\hat{z}_t(l-1) + \hat{z}_t(l-2) = (1-B)^2 \hat{z}_t(l) = 0$$

si se considera que B actúa sobre l; es decir:

$$B l = l - 1$$

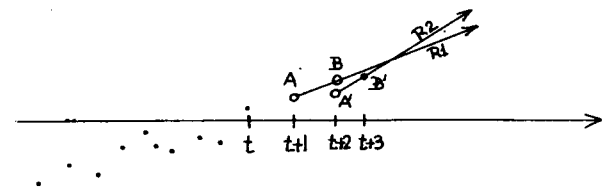
Veamos pues que para el modelo (1.3) los valores de $\hat{z}_t(1)$ y $\hat{z}_t(2)$ dependen del instante t, pero para $l > 2$ la función de previsión viene dada por la solución de la ecuación en diferencias:

$$(1 - B)^2 \hat{z}_t(l) = 0 \quad (1.5)$$

Evidentemente, la solución de (1.5) es de la forma:

$$\hat{z}_t(l) = b_0(t) + b_1(t)l \quad l > z \quad (1.6)$$

Es importante analizar este resultado. El gráfico:



nos resume el mecanismo de predicción del modelo (1.3).

En el instante t, la función de previsión vendrá dada por la recta (1.6) cuyos dos puntos iniciales A y B dependen de t. En el instante t+1, las condiciones iniciales nos determinarán los puntos A' y B' y la nueva función de precisión será la recta R2. Así pues un modelo ARIMA se caracteriza por tener una función de previsión adaptativa a la vez que permite modelar las autocorrelaciones.

El problema que se plantea es el siguiente: Si en una cierta serie temporal existe una componente determinista del tipo (1.3): ¿Será posible detectar al modelar dicha serie mediante un modelo de la clase ARIMA?. A esta pregunta intentaremos contestar en el presente trabajo.

2. DETECCION DE UNA COMPONENTE DETERMINISTA

En este apartado se demuestra que la detección de una componente determinista en un modelo ARIMA se basa en la cancelación de operadores entre la parte autoregresiva y la de medias móviles del modelo. La discusión sigue básicamente los argumentos señalados por Abraham y Box (1975).

Supongamos que una serie temporal puede ser representada por el modelo ARIMA:

$$\phi_p(B)Z_t = \theta_0 + \theta_q(B)a_t \quad (2.1)$$

en el que el polinomio autoregresivo $\phi_p(B)$ puede contener una o más diferencias, es decir, dicho polinomio podría ser de la forma:

$$\phi_p(B) = (1 - \phi_1 B)(1 - B)^2(1 - B^{12}) \quad p = 15$$

Llamemos G_i^{-1} a los ceros del polinomio $\phi_p(B)$ o sea a las soluciones de $\phi_p(B) \equiv 0$. Admitamos la posibilidad de ceros múltiples y llamemos n_i al grado de multiplicidad de G_i^{-1} , siendo $\sum n_i = p$; $i = 1, 2, \dots, r$.

Admitamos, de momento, que $\theta_q(B) = \theta_p(B)\theta_{q-p}(B)$ es tal que $\phi_p(B) \equiv \theta_p(B)$. Es evidente que el modelo (2.1) puede escribirse:

$$\phi_p(B) [Z_t - f(t)] = \theta_q(B)a_t$$

$$\text{siendo } \phi_p(B)f(t) = \theta_0 \quad (2.2)$$

Debido a la posibilidad de cancelación entre $\phi_p(B)$ y $\theta_p(B)$ obtendremos:

$$Z_t = f(t) + \theta_{q-p}(B)a_t \quad (2.3)$$

Es decir, la serie temporal podrá ser representada mediante una función determinista, -

$f(t)$, que es la solución de la ecuación en diferencias (2.2) y un ruido autocorrelacionado.

Utilizando la teoría de ecuaciones en diferencias es fácil ver que la solución para $f(t)$ sería:

$$f(t) = f_c(t) + f_p(t)$$

donde $f_c(t)$ es la solución de la ecuación en diferencias homogéneas:

$$\phi_p(B)f_c(t) = 0 \quad (2.4)$$

y $f_p(t)$ es una solución particular de (2.2)

$$f_p(t) = \frac{\theta_0}{1 - \phi_1 - \dots - \phi_p} = A_0$$

de donde:

$$f(t) = A_0 + \sum_{i=1}^r \left(\sum_{j=1}^{n_i-1} A_{j,i} t^j \right) G_i^t \quad (2.5)$$

En esta expresión las constantes $A_{j,i}$ dependen del origen de tiempos, es decir, de las condiciones iniciales. La forma concreta de esta función dependerá del carácter real o complejo de los ceros del polinomio $\phi_p(B)$. A ceros reales corresponden funciones potenciales mientras que a los ceros complejo-conjugados corresponden funciones trigonométricas.

Obviamente la componente $f(t)$ solo persistirá si alguna $|G_i| > 1$.

Hasta aquí hemos supuesto la cancelación total del operador $\phi_p(B)$ con una parte de $\theta_q(B)$. Veamos que ocurre en el caso de cancelación parcial.

Supongamos que una serie temporal puede modelarse en la forma:

$$\phi_{p+k}(B)Z_t = \theta_0 + \theta_q(B)a_t \quad p \leq q$$

en el que

$$\phi_{p+k}(B) = \phi_p(B)\phi_k(B); \theta_q(B) = \theta_p(B)\theta_{q-p}(B)$$

y supongamos que $\phi_p(B) \equiv \theta_p(B)$

En este caso, siguiendo un razonamiento análogo al anterior, se puede escribir el modelo para Z_t en la forma:

$$\phi_k^{(B)} Z_t = A_0 + \sum_{i=1}^r \left(\sum_{j=0}^{n_i-1} A_{j,i} t^j \right) G_i^t + \theta_{q-p}^{(B)} a_t \quad (2.6)$$

Estos resultados confirman que una componente determinista en un modelo podrá ser detectada por la existencia de cancelación entre factores del polinomio autoregresivo y el de medias móviles. Desde este punto de vista podemos asegurar que los modelos ARIMA engloban a los modelos con componentes deterministas.

De lo expuesto se deduce también la gran importancia que tiene el introducir en los programas de estimación de series temporales -- existentes en el mercado, la posibilidad de estimar conjuntamente componentes deterministas y parámetros del modelo estocástico. A ello dedica sus esfuerzos en la actualidad -- un profesor ayudante de la Cátedra bajo nuestra supervisión.

3. SOBREDIFERENCIA DE UNA SERIE DEBIDA A UNA COMPONENTE DETERMINISTA NO DETECTADA.

En uno de los estudios que hemos realizado -- para la empresa Gas y Electricidad, S.A. de Palma de Mallorca y como paso previo a la modelización de la relación dinámica entre la temperatura media mensual y la punta mensual media de demanda de energía eléctrica, tuvimos que buscar un modelo ARIMA adecuado para la serie de temperaturas medias mensuales.

Por razones solidamente fundadas en la física de una esfera girando alrededor de un punto emisor de calor, se puede intuir que la evolución de la temperatura media mensual a lo largo de varios años a tener una componente trigonométrica (senoides y cosenoides) relativamente estable, a la que se va a superponer un cierto proceso estocástico, ya que en primera aproximación la tierra puede considerarse esférica y el sol puntual en cuyo caso la temperatura en un punto de la tierra seguirá el modelo indicado.

Desde esta óptica parecería lógico ajustar -- un modelo del tipo:

$$Z_t = A_0 + A_1 \cos \left(\frac{2\pi t}{12} \right) + A_2 \sin \left(\frac{2\pi t}{12} \right) + \psi(B) a_t \quad (3.1)$$

$$Z_t = f(t) + \psi(B) a_t$$

donde $f(t)$ representa la componente determinista antes mencionada y $\psi(B)$ es el filtro lineal correspondiente al ruido de la serie. La utilidad de modelos de este tipo parece -- también justificada en modelos hidrológicos para precisión del caudal de ciertos ríos -- (Ver Kashyap y Rao (1976)).

Del contenido del apartado 2 del presente -- trabajo, se deduce que un análisis cuidadoso de la serie de temperaturas mediante la metodología de Box y Jenkins debería detectar la existencia de esta componente determinista -- por cancelación de operadores y veremos en -- el apartado 5 que ello es así, si se utiliza el algoritmo de estimación basado en la maximización de la función de verosimilitud exacta.

Por desgracia, hasta hace muy poco tiempo, -- dicho algoritmo no estaba implementado en -- los paquetes de programas más utilizados en la práctica y ello ha conducido a una sobrediferenciación de bastantes series temporales existentes en la literatura.

Como es sabido, Box y Jenkins (1970), recomiendan la utilización del operador diferencia $(1-B)$ u otros operadores tales como $(1-B^{12})^D$, como paso previo necesario para -- convertir en estacionaria una serie que no -- lo sea inicialmente. En cualquier caso, dichos autores advierten sobre los efectos de una sobrediferenciación.

Los efectos de aplicar el operador $(1-B)^d$ -- a una serie temporal Z_t son, entre otros:

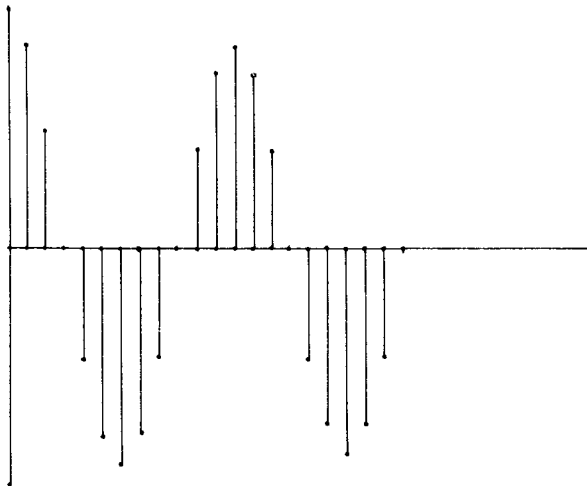
- a) -- Eliminar de la serie una tendencia polinómica de grado k ($k = 0, 1, \dots, d$)
- b) -- Eliminar la no estacionariedad causada por $d-k$ ($k = 0, 1, \dots, d$) raíces unitarias en el operador autoregresivo.

Veamos a continuación cual es el efecto de -- una sobrediferenciación:

Supongamos que el verdadero modelo para una serie temporal fuese:

$$(1 - \sqrt{3} B + B^2) z_t = a_t \quad (3.2)$$

La función de autocorrelación del modelo --- (3.2) es:



La falta de amortiguamiento en esta función de autocorrelación es debida a que la serie no es estacionaria. En efecto, las soluciones de la ecuación: $1 - \sqrt{3} B + B^2 = 0$

$$\text{son } G_1^{-1} = \frac{\sqrt{3}}{2} + \frac{1}{2} i \quad G_2^{-1} = \frac{3}{2} - \frac{1}{2} i ,$$

ambas en el círculo unidad. El carácter de complejo-conjugadas de estas dos soluciones es la justificación teórica del comportamiento trigonométrico de la función de autocorrelación antes citada, ya que, según lo expuesto en el apartado 2, y teniendo en cuenta -- que el modelo (3.2) es autoregresivo puro, -- su función de autocorrelación se obtendría -- resolviendo las ecuaciones de Yule-Walker:

$$(1 - \sqrt{3} B + B^2) p_k = 0 \quad (3.3)$$

donde ahora el operador B actúa sobre k.

La solución de la ecuación (3.3) es:

$$p_k = A_1 \cos \frac{2\pi k}{12} + A_2 \sin \frac{2\pi k}{12}$$

que justifica la forma observada en la función de autocorrelación.

Supongamos que el analista no fuese consciente de la discusión precedente y considerase el no amortiguamiento rápido de la f.a.c. (**) como una no estacionariedad en la serie que se puede eliminar por diferenciación de la -- misma. Dado el carácter cíclico de la serie con un período de 12 meses, es posible que -- ajustase un modelo: ARMA a la serie $(1-B^{12})z_t$. Si tenemos en cuenta que:

$$(1-B^{12}) = (1-B)(1+B)(1+\sqrt{3}B+B^2)(1+B+B^2) \\ (1+B^2)(1-B+B^2)(1-\sqrt{3}B+B^2)$$

y la estructura real del modelo (3.2), resultaría que:

$$(1-B^{12})z_t = \frac{1-B^{12}}{1-\sqrt{3}B+B^2} a_t = \theta_{10}(B)a_t \quad (3.4)$$

con lo que se habría introducido en el modelo un polinomio de medias móviles con diez -- parámetros innecesarios y además no invertible.

Este hecho no sería especialmente grave si -- en la fase de estimación del modelo (3.4) se detectasen realmente las raíces unitarias en $\theta_{10}(B)$. En la práctica se utiliza la estimación mínimo-cuadrática condicional que tiende a infravalorar los parámetros cuando hay ceros en el círculo unidad con lo que es posible que la sobrediferenciación pasase desapercibida. Este aspecto del problema se discute en el apartado que sigue:

4. ESTIMACION MAXIMO-VEROSIMIL Y SUS APROXIMACIONES

La función de máxima verosimilitud de un modelo ARIMA cualquiera ha sido objeto de numerosos estudios en los últimos años. Podríamos citar, entre otros, los trabajos de Dent (1975), Newbold (1974), Ansley (1979) y ---- Ljung (1976). Es sobre este último artículo en el que basaremos la exposición que sigue:

Supongamos que una serie temporal z_t está generada por el modelo ARIMA (p,d,q):

$$\phi_p(B)w_t = \theta_q(B)a_t \\ w_t = (1-B)^d z_t \quad (4.1)$$

y que disponemos de $n + d$ observaciones z_1, z_2, \dots, z_{n+d} . Con ellas podemos obtener:

$$\underline{w} = [w_1, w_2, \dots, w_n]^t \quad (4.2)$$

sin pérdida de generalidad $E[\underline{w}] = \underline{0}$

Considerando \underline{w} como una observación de una ley normal multivariante $N(\underline{0}; \sigma^2 \underline{\Sigma})$ donde $\underline{\Sigma}$ es la matriz de correlaciones de \underline{w} tendríamos que la función de verosimilitud sería:

$$L(\underline{\phi}, \underline{\theta}, \sigma | \underline{w}) = (2\pi\sigma^2)^{-n/2} |\underline{\Sigma}|^{-1/2} \exp - \frac{1}{2\sigma^2} \underline{w}' \underline{\Sigma}^{-1} \underline{w} \quad (4.3)$$

$$\text{Llamemos } S(\underline{\phi}, \underline{\theta}) = \underline{w}' \underline{\Sigma}^{-1} \underline{w}$$

La estimación máximo-verosímil de σ^2 se obtendrá maximizando el logaritmo neperiano de (4.3) y es inmediato comprobar que:

$$\hat{\sigma}_2 = \frac{s(\underline{\phi}, \underline{\theta})}{n}$$

con lo cual:

$$L(\underline{\phi}, \underline{\theta} | \underline{w}, \hat{\sigma}_2) = k (\underline{w}' \underline{\Sigma}^{-1} \underline{w})^{-n/2} |\underline{\Sigma}|^{-1/2} \quad (4.4)$$

Los estimadores máximo-verosímiles de $\underline{\phi}$ y $\underline{\theta}$ se obtendrán pues maximizando (4.4) o, lo que es equivalente, minimizando

$$(\underline{w}' \underline{\Sigma}^{-1} \underline{w})^{n/2} |\underline{\Sigma}|^{1/2}$$

es decir, MINIMIZANDO:

$$(\underline{w}' \underline{\Sigma}^{-1} \underline{w}) |\underline{\Sigma}|^{1/n} \quad (4.5)$$

Así pues, el cálculo exacto de los estimadores de máxima verosimilitud requeriría la inversión de la matriz de correlaciones, $\underline{\Sigma}$, que es de orden $n \times n$. Un procedimiento general para invertir dicha matriz fué propuesto por Cleveland (1972). El método propuesto por Ljung y Box (1976) permite reducir considerablemente las dimensiones de la matriz a invertir.

Para poder iniciar la solución de la ecuación en diferencias que es el modelo (4.1) se deben conocer o estimar los valores iniciales

$$\underline{u}_* = [w_{1-p}, \dots, w_0, a_{1-q}, \dots, a_0]^t$$

Si llamamos $\underline{a} = [a_1, a_2, \dots, a_n]^t$, el modelo (4.1) puede escribirse:

$$\underline{L}_1 \underline{w} = \underline{L}_2 \underline{a} + \underline{V} \underline{u}_* \quad (4.6)$$

donde \underline{L}_1 y \underline{L}_2 , son matrices de orden $n \times n$; \underline{V} es de orden $n \times (p+q)$ y todas ellas perfectamente definidas en función de los parámetros del modelo.

Si se calcula la densidad de probabilidad conjunta:

$$p(\underline{w}, \underline{u}_* | \underline{\theta}, \underline{\phi}, \hat{\sigma}) = p(\underline{w} | \underline{u}_*, \underline{\phi}, \underline{\theta}, \hat{\sigma}) \cdot p(\underline{u}_* | \underline{\phi}, \underline{\theta}, \hat{\sigma}) \quad (4.7)$$

en la que $p(\underline{u}_* | \underline{\phi}, \underline{\theta}, \sigma)$ es la densidad de una ley normal multivariante $N(\underline{0}, \sigma^2 \underline{\Sigma}_*)$ siendo $\underline{\Sigma}_*$ la matriz de correlaciones de los valores iniciales, de orden $(p+q) \times (p+q)$ y $p(\underline{w} | \underline{u}_*, \underline{\phi}, \underline{\theta}, \sigma)$ se obtendrá a partir de $p(\underline{a} | \underline{u}_*, \underline{\phi}, \underline{\theta}, \sigma)$ mediante la transformación lineal (4.6) es fácil encontrar la expresión de la verosimilitud. En efecto, la densidad conjunta (4.7) admite la descomposición:

$$p(\underline{w}, \underline{u}_* | \underline{\phi}, \underline{\theta}, \sigma) = p(\underline{w} | \underline{\phi}, \underline{\theta}, \sigma) p(\underline{u}_* | \underline{w}, \underline{\phi}, \underline{\theta}, \sigma)$$

y, evidentemente, la función de verosimilitud se obtendrá teniendo en cuenta que:

$$L(\underline{\phi}, \underline{\theta}, \sigma | \underline{w}) = k p(\underline{w} | \underline{\phi}, \underline{\theta}, \sigma)$$

En el trabajo antes citado, Ljung y Box, llegan a las siguientes expresiones para el cálculo de la expresión (4.5):

$$|\underline{\Sigma}| = |\underline{I}_{p+q} + \underline{\Sigma}_*^{-1} \underline{V}' \underline{L}_2^{-1} \underline{L}_2^{-1} \underline{V}| \quad (4.8)$$

$$\underline{w}' \underline{\Sigma}^{-1} \underline{w} = \hat{\underline{u}}_*' \underline{\Sigma}_*^{-1} \hat{\underline{u}}_* + \hat{\underline{a}}' \hat{\underline{a}} \quad (4.9)$$

donde $\hat{\underline{u}}_* = E[\underline{u}_* | \underline{w}, \underline{\phi}, \underline{\theta}, \sigma]$ y $\hat{\underline{a}}$ se obtiene de (4.6) sustituyendo \underline{u}_* por $\hat{\underline{u}}_*$

Si aplicamos estos resultados al caso de un modelo de medias móviles MA (9) tendremos:

$$\underline{u}_* = \underline{a}_* = [a_{1-q}, \dots, a_0]^t$$

$$\underline{L}_1 = \underline{I}_n$$

$$\underline{\Sigma}_* = \underline{I}_q$$

de donde:

$$|\underline{\underline{\Sigma}}| = |\underline{\underline{I}}_q + \underline{\underline{V}}^{-1} \underline{\underline{L}}_2^{1-1} \underline{\underline{L}}_2^{-1} \underline{\underline{V}}| = |\underline{\underline{X}}' \underline{\underline{X}}| \quad (4.10)$$

siendo $\underline{\underline{X}} = \underline{\underline{L}}_2^{-1} \underline{\underline{V}}, y$

$$\underline{\underline{w}}' \underline{\underline{\Sigma}}^{-1} \underline{\underline{w}} = \hat{\underline{\underline{a}}}_* \hat{\underline{\underline{a}}}_* + \hat{\underline{\underline{a}}}' \hat{\underline{\underline{a}}} = \sum_{1-q}^0 \hat{\underline{\underline{a}}}_t^2 + \sum_1^n \hat{\underline{\underline{a}}}_t^2 \quad (4.11)$$

La estimación máximo-verosímil consistiría - pues en minimizar la expresión (4.6) que, para el caso particular que nos ocupa sería:

$$\left[\sum_{1-q}^n \hat{\underline{\underline{a}}}_t^2 + \sum_1^n \hat{\underline{\underline{a}}}_t^2 \right] |\underline{\underline{X}}' \underline{\underline{X}}|^{1/n} \quad (4.12)$$

Box y Jenkins (1970) indican que, excepto si los parámetros del modelo están próximos a la región de no invertibilidad y para tamaños de muestra no muy pequeños, la expresión (4.12) viene denominada por la suma de cuadrados:

$$\underline{\underline{w}}' \underline{\underline{\Sigma}}^{-1} \underline{\underline{w}} \approx \sum_{1-q}^0 \hat{\underline{\underline{a}}}_t^2 + \sum_1^n \hat{\underline{\underline{a}}}_t^2 \quad (4.13)$$

Indican además que si la muestra es suficientemente grande, la minimización de (4.13) se conseguirá con suficiente aproximación si se minimiza la suma de cuadrados condicional a $\underline{\underline{a}}_* = 0$, es decir, buscando el mínimo de la suma $\sum_{t=1}^n \hat{\underline{\underline{a}}}_t^2$. Este procedimiento es el que hemos llamado mínimos cuadrados condicionales. Una mejor aproximación se logra minimizando directamente (4.12) en donde $\hat{\underline{\underline{a}}}_*$ se calculan mediante retrovisión (backforecasting). De todos modos, la estimación máximo-verosímil exacta debe hacerse teniendo en cuenta el determinante $|\underline{\underline{X}}' \underline{\underline{X}}|^{1/n}$ y minimizando (4.12).

Si llamamos $S(\underline{\underline{\theta}})$ al valor de $\underline{\underline{w}}' \underline{\underline{\Sigma}}^{-1} \underline{\underline{w}}$; $S_0(\underline{\underline{\theta}})$ el valor de (4.13) condicional a $\underline{\underline{a}}_* = \underline{\underline{\theta}}$ y $L_0(\underline{\underline{\theta}})$ al valor exacto de (4.12) y citándonos al caso de un modelo MA (1), es fácil probar que:

$$S(\underline{\underline{\theta}}) = S_0(\underline{\underline{\theta}}) - \hat{\underline{\underline{a}}}_0^2 \frac{1 - \theta^{2(n+1)}}{1 - \theta^2} \quad (4.14)$$

$$L_0(\underline{\underline{\theta}}) = S(\underline{\underline{\theta}}) \left| \underline{\underline{\Sigma}} \right|^{1/n} = S(\underline{\underline{\theta}}) \left\{ \frac{1 - \theta^{2(n+1)}}{1 - \theta^2} \right\}^{1/n} \quad (4.15)$$

De estas expresiones se deduce que:

- a) $S_0(\underline{\underline{\theta}})$ es siempre mayor que $S(\underline{\underline{\theta}})$
- b) Si $|\theta|$ no es próximo a la unidad y cumple la condición de invertibilidad $|\theta| < 1$ entonces $S(\underline{\underline{\theta}})$ y $S_0(\underline{\underline{\theta}})$ no diferirán mucho, - con lo cual, las estimaciones obtenidas mediante la aproximación mínimo-cuadrática condicional y mediante la utilización de backforecasting coincidirán prácticamente.
- c) Si $|\theta| \approx 1$ entonces $\frac{1 - \theta^{2(n+1)}}{1 - \theta^2} \rightarrow n+1$ y por tanto, $S(\underline{\underline{\theta}})$ y $S_0(\underline{\underline{\theta}})$ pueden diferir notablemente.
- d) $L_0(\underline{\underline{\theta}})$ es siempre mayor que $S(\underline{\underline{\theta}})$ en la región de invertibilidad. Ello nos permite deducir que para tamaños de muestra no muy pequeños y $|\theta|$ no próximos a la unidad, las esperanzas matemáticas de $S_0(\underline{\underline{\theta}})$, $S(\underline{\underline{\theta}})$ y $L_0(\underline{\underline{\theta}})$ serían del tipo:

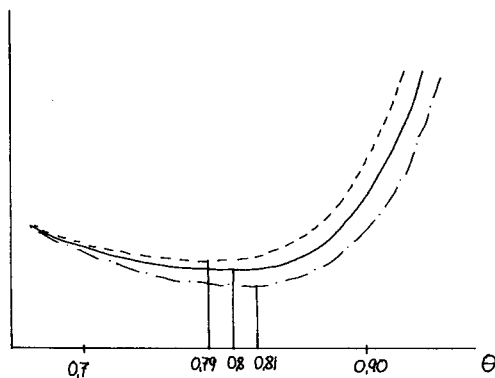


Fig. 4.1

E	$\left[L_0(\underline{\underline{\theta}}) \right]$	_____	$n = 100$
$\frac{E}{\omega}$	$\left[S(\underline{\underline{\theta}}) \right]$	-----	$\theta = 0,8$
$\frac{E}{\omega}$	$\left[S_0(\underline{\underline{\theta}}) \right]$	- - - - -	

En cambio, si $\theta = 0,95$ entonces se tendría - que:

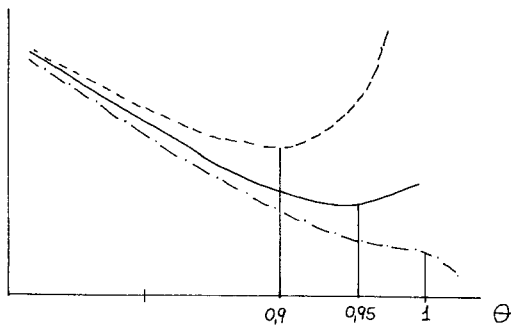


Fig. 4.2

$n = 100$
 $\theta = 0,95$

en la que se observa claramente la infravaloración del parámetro al minimizar $S_0(\theta)$ y su sobrevaloración al utilizar $S(\theta)$ que puede ser causa incluso de la no convergencia del algoritmo de minimización, hecho frecuentemente observado en la práctica.

Las consecuencias de lo expuesto hasta aquí son obvias. Para detectar por ejemplo la sobrediferenciación en el modelo (3.4) deberían detectarse raíces en el círculo unidad en el polinomio de medias móviles. Si se utiliza la estimación mínimo-cuadrática condicional es posible que el sesgo hacia 0 de estos estimadores no nos permitiese detectar el valor unitario de los ceros del polinomio.

Es claro también que para detectar cancelación entre factores de los polinomios autorepresivos y las de medias móviles (hecho que permitiría ajustar una tendencia determinista al modelo) debemos ser cuidadosos en el proceso de estimación. Es por ello que recomendamos en Prat et al. (1980) el uso de la función de verosimilitud exacta como herramienta básica en el proceso de inferencia.

5. EJEMPLOS

Para ilustrar los conceptos anteriores, vamos a utilizar un ejemplo concreto. La serie temporal que vamos a estudiar es la serie de temperaturas medias mensuales de Palma de Mallorca.

La función de autocorrelación estimada se encuentra en la fig. 5.1. Del carácter cíclico

de esta función se puede deducir la necesidad de aplicar el operador $(1 - \sqrt{3B + B^2})$. De todos modos, supongamos que el analista no es consciente de este hecho y, basándose en la estacionalidad de período 5 = 12 de dicha serie aplica el operador $\nabla_{12} = 1 - B^{12}$. La función de autocorrelación de la serie $\nabla_{12}X_t$ es la de la figura (5.2).

Si en la fase de identificación tentativa, el modelo retenido hubiese sido:

$$\nabla_{12}X_t = (1 - \theta_1 B)(1 - \theta_2 B^{12})a_t \quad (5.1)$$

entonces los resultados de la estimación mínimo-cuadrática condicional serían:

$$\begin{aligned} \hat{\theta}_1 &= -0,29 & [-0,46; -0,11] \\ \hat{\theta}_2 &= 0,61 & [0,46; 0,76] \end{aligned}$$

donde los números entre corchetes son los extremos del intervalo de confianza del 95% de dichos parámetros.

Los residuos del modelo no presentan más anomalía que un valor de 2,86 para el cociente entre la media de los residuos y su desviación tipo. Ello indica la existencia de una componente determinista como ya era de esperar.

Pero el punto que pretendíamos ilustrar en este apartado es que el valor de $\hat{\theta}_2$ y su intervalo de confianza del 95% no permiten detectar la existencia de raíces unitarias en el polinomio.

Al utilizar la estimación con retrovisión (backforecasting) y de acuerdo con lo previsto por la teoría resumida en la figura 4.2, el valor del mínimo se encuentra en la región de no invertibilidad. En efecto, después de la primera iteración los valores estimados fueron:

$$\hat{\theta}_1^{(1)} = -0,3246 \quad \hat{\theta}_2^{(1)} = 0,9589$$

En la segunda iteración:

$$\hat{\theta}_1^{(2)} = -0,3006 \quad \hat{\theta}_2^{(2)} = 1,08$$

con lo que el procedimiento deja de iterar por entrar en la zona de no invertibilidad. También de acuerdo con la teoría expuesta en

*** IDENTIFICACION TEMPERATURA GESA MENSUAL ***

GRAPH OF OBSERVED SERIES ACF

GRAPH INTERVAL IS .2000-01

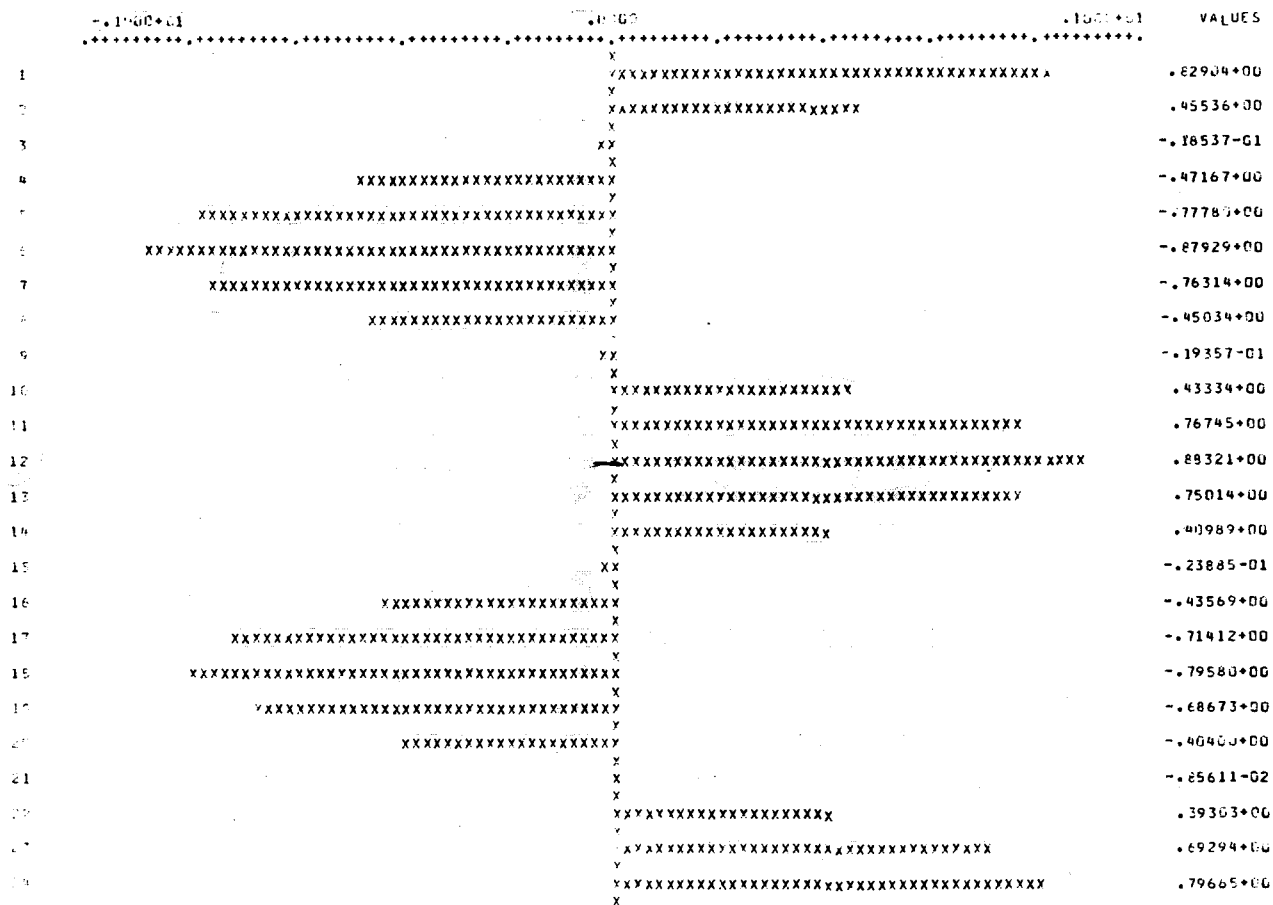


Figura 5.1
 Función de autocorrelación de la serie X_t
 (Temperaturas medias mensuales)

GENIFICACION TEMPERATURA GESA MENSUAL

GRAPH OF DIFFERENCE 1 ACF

GRAPH INTERVAL IS .2000-01

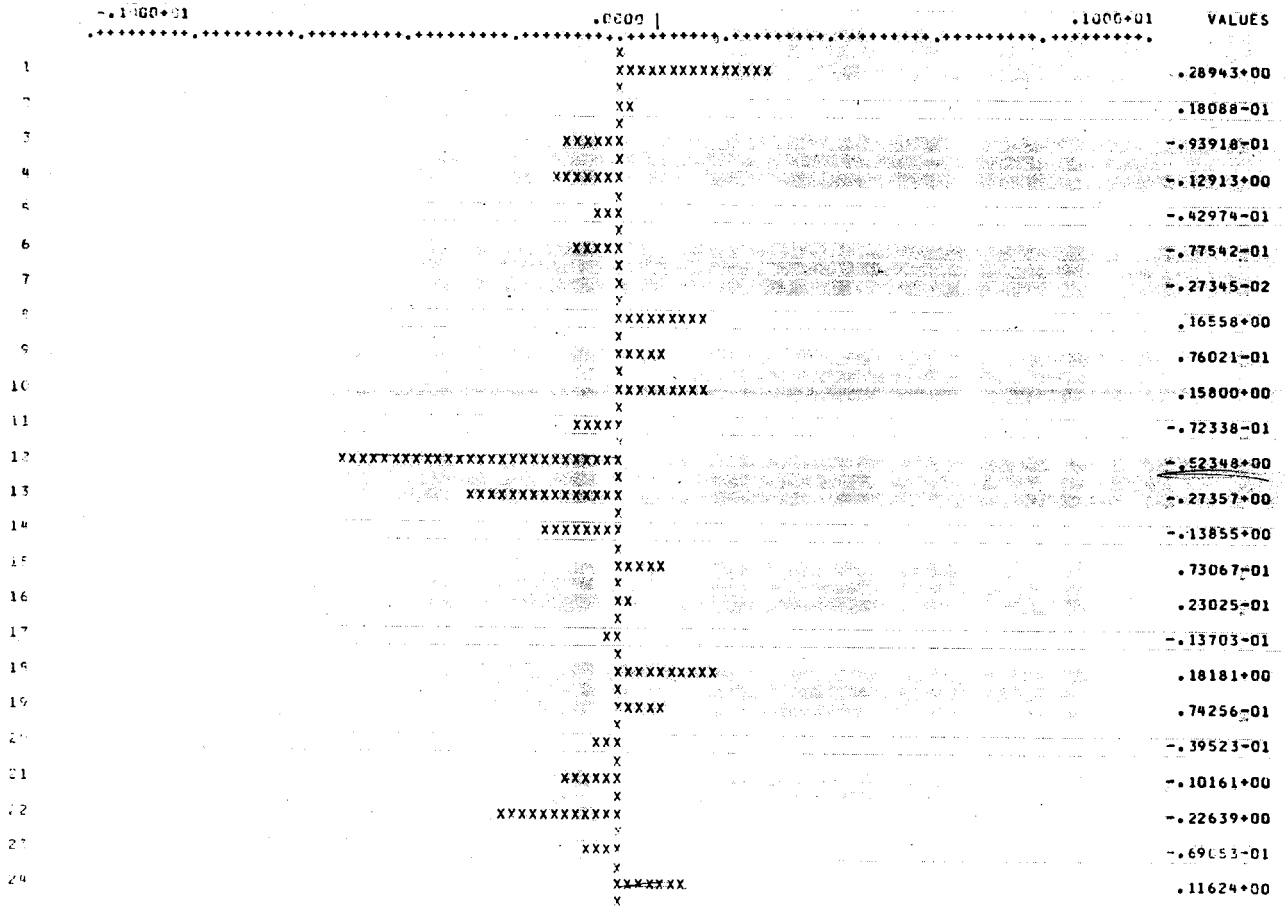


Figura 5.2

Función de autocorrelación de la serie $V_{12}X_t$

el apartado 4, la suma de cuadrados con retrovisión $S(\theta) = 0,147 \cdot 10^3$ es menor que la obtenida con mínimos cuadrados condicionales $S_0(\theta) = 0,175 \cdot 10^3$.

Finalmente, la estimación máximo-verosímil - dió como resultados:

$$\hat{\theta}_1 = -0,3054 \quad \hat{\theta}_2 = 0,987$$

con lo que es fácil darse cuenta de la sobre diferenciación de la serie.

6. CONCLUSIONES

Hemos visto en primer lugar que la clase de modelos ARIMA propuestos por Box y Jenkins - (1970) incluyen como caso particular a modelos con tendencias y / o estacionalidades de tipo casi determinista.

En segundo lugar hemos comprobado que la detección de una tendencia y / o estacionalidad de tipo determinista se basa en la cancelación de operadores entre los polinomios -- autoregresivos y de medias móviles del modelo.

El problema de la sobrediferenciación de una serie se expone en el apartado 3.

Tanto para detectar la cancelación de operadores como la posible sobrediferenciación en una serie temporal se ha visto, en el presente trabajo, la importancia de utilizar la estimación máximo-verosímil de los parámetros, en lugar de los mínimos-cuadrados condicionales.

Finalmente estos conceptos han sido estudiados mediante una serie temporal concreta.

7. BIBLIOGRAFIA

- /1/ ABRAHAM, B., BOX G.E.P. "Linear models, Time Series and Outliers. 3". Technical Report No 438. Department of Statistics. University of Wisconsin-Madison. (1975).
- /2/ ANSLEY, C.F. "An Algorithm for the exact likelihood of a mixed autoregressive-moving average process". *Biometrika*, 66, 59-65 (1979)

- /3/ BOX, G.E.P., JENKINS G.M. "Time Series Analysis: Forecasting and Control. Holden Day." (1970).
- /4/ CLEVELAND, W.P. "Analysis and Forecasting of Seasonal Time Series". Ph.D. Thesis. Department of Statistics. University of Wisconsin-Madison. (1972).
- /5/ DENT, W.T. "Computation of the exact likelihood function for an ARIMA process", *J. Statist. Comp X Simul*, 5, 193-206. (1977).
- /6/ KALMAN, R.E. "A new approach to linear filtering and prediction problems", *J. Basic Eng (ASME Trans)* 82, D. 33-45 (1960)
- /7/ KASHYAP, R.L., RAO, A.R. "Dynamic Stochastic Models from Empirical Data". Academic Press. (1976).
- /8/ LJUNG, G .M., BOX, G.E.P. "Studies in -- the modelling of Discrete Time Series, 2" Technical Report No 476. Department of Statistics. University of Wisconsin-Madison. (1976).
- /9/ NEWBOLD, P. "The exact likelihood function for a mixed autoregressive-moving average process". *Biometrika*, 61, 423-26 (1974).
- /10/ PRAT-BARTES, A., PEÑA D., MARTI, M. "La Teoría de Series Temporales. Una -- evaluación crítica de los desarrollos -- más recientes". Ponencia invitada en la reunión anual de la E.E.I.I.O. Setiembre 1980, JACA.

8. NOTAS

- (*) Para una exposición completa de la metodología, nomenclatura, etc., remitimos al lector de la citada obra de Box y -- Jenkins (1970)
- (**) Designaremos en adelante función de -- autocorrelación por f.a.c. y función de autocorrelación parcial por f.a.c.p.