ESTADISTICA ESPAÑOLA
núm. 116, 1988, págs. 87 a 106

Estimación en los modelos autorregresivos y de promedios móviles

por
RAUL PEDRO MENTZ
Instituto de Investigaciones Estadísticas (INIE)
Facultad de Ciencias Económicas
Universidad Nacional de Tucumán
Casilla de Correo 209
4000 San Miguel de Tucumán
Argentina

RESUMEN

El propósito de esta presentación es seguir la trayectoria de algunas de las principales ideas aparecidas en la literatura, en el tema de la estimación de los parámetros de los modelos autorregresivos y de promedios móviles, enfatizando un enfoque en el dominio del tiempo.

En cada modelo se define la estructura probabilística y se mencionan algunas propiedades básicas. Luego se consideran sugerencias para estimar los parámetros por el método de los momentos o por mínimos cuadrados. Se consideran después los resultados de investigaciones para demostrar las propiedades asintóticas de estas propuestas y se llega al estudio de métodos de estimación por máxima verosimilitud (exacta o aproximada) en el caso normal.

Se considera brevemente el modelo mixto y algunas propuestas recientes para implementar la estimación máximo verosímil en las computadoras. Se concluye que en los últimos 65 años se han producido desarrollos muy importantes en el área.

Palabras clave: Autorregresivo, promedios móviles, estimación por mínimos cuadrados, estimación por máxima verosimilitud, propiedades asintóticas.

Clasificación AMS: 62M10

1. INTRODUCCION

El propósito de esta presentación es seguir la trayectoria de algunas de las principales ideas aparecidas en la literatura, en el tema de la estimación de los parámetros de los modelos autorregresivos y de promedios móviles, enfatizando un enfoque en el dominio del tiempo.

Se establece el origen de estos modelos en los trabajos de G. U. Yule (inglés, 1871-1951) y E. Slutzky (ruso, 1880-1948) aparecidos entre 1921 y 1937; ver Wold (1954). Antes de estos trabajos el enfoque predominante entre los analistas de datos cronológicos era la búsqueda de las "periodicidades ocultas", con herramientas como el periodograma de Schuster (1898, 1900). En la actualidad decimos que éste es un enfoque no paramétrico, y su versión moderna es el análisis espectral empírico basado en estimadores consistentes de la densidad espectral del proceso estocástico estacionario subyacente.

Un mérito importante de los trabajos de Yule y Slutzky fué insistir en la formulación de modelos estocásticos o estadísticos para el mecanismo generador de los datos. Además los modelos propuestos fueron definidos inicialmente en el dominio del tiempo, lo que sin duda contribuyó a su difusión entre los economistas y otros investigadores sociales.

Incidentalmente, en el momento actual es válido aseverar que las técnicas del análisis de series cronológicas en el dominio del tiempo, tipificadas quizás por el enfoque de Box y Jenkins (1976), y las propias del análisis en el dominio de la frecuencia, interpretadas como consecuencia de una transformada de Fourier, deben pertenecer al acervo de quienes se dedican a las series cronológicas, como investigadores metodológicos o como analistas de datos. A pesar de esta aseveración, nuestro enfoque en esta exposición estará casi exclusivamente basado en el dominio del tiempo.

2. EL MODELO AUTORREGRESIVO

2.1. Definición y Propiedades

Un proceso estocástico $\{ z_t \}$ con índice temporal discreto se dice estacionario si las distribuciones conjuntas de probabilidad asociadas con un vector $(z_{t_1}, z_{t_2}, \ldots, z_{t_k})$ son idénticas a las asociadas con el vector $(z_{t_1+h}, z_{t_2+h}, \ldots, z_{t_k+h})$ obtenido por una traslación temporal, y ésto para todo conjunto $(t_1, t_2, \ldots, t_k)$ de índices, para todo k y para todo h. Un proceso estacionario tiene todos sus momentos invariantes a cambios en el tiempo. Un proceso se dice "estacionario débil" si sus momentos de primer y segundo orden (esperanzas matemáticas, varianzas, covarianzas) son invariantes a cambios en el tiempo.

Un proceso estocástico estacionario $\{ y_t \}$ se dice que es un proceso autorregresivo (o que corresponde a un modelo autorregresivo) de orden finito p, $p \geq 1$, denotado AR(p), si tiene esperanza matemática constante (digamos, $Ey_t=\mu$ para todo t), varianza constante, positiva y finita, y si para coeficientes $\beta_0=1$, β_1, $\beta_2,\ldots,\beta_p$ satisface la ecuación estocástica en diferencias finitas.

$$\sum_{j=0}^{p} \beta_j(y_{t-j}-\mu) = (y_t-\mu) + \beta_1(y_{t-1}-\mu) +\ldots+ \beta_p(y_{t-p}-\mu) = u_t, \quad t = \ldots,-1,0,1,\ldots, \qquad (2.1.1.)$$

con u_t independiente de $y_{t-1},y_{t-2},\ldots$. Las u_t a su vez forman un proceso estocástico de variables aleatorias independientes, $Eu_t = 0, Eu_t^2= \sigma^2$, $0 < \sigma^2 < \infty$ (1). Las y_t son variables aleatorias observables, mientras que las u_t son inobservables; $\mu,\beta_1,\ldots,\beta_p$ y σ^2 son los $p+2$ parámetros del proceso.

El nombre de autorregresión y la propiedad señalada de que los u_t son independientes de los y_t pasados, se vuelven sugestivos frente a la siguiente interpretación: en la regresión lineal múltiple entre y_t y los regresores estocásticos $y_{t-1},\ldots,y_{t-p}$, que escribimos

$$y_t = \mu-\beta_1 (y_{t-1}-\mu) - \beta_2 (y_{t-2}-\mu) - \ldots -\beta_p (y_{t-p}-\mu) + u_t, \qquad (2.1.2.)$$

los "errores" o "innovaciones" contemporáneos son independientes de los regresores usados y de los pasados, es decir, de y_{t-1}, y_{t-2}, ...

La ecuación polinomial asociada,

$$w^p + \beta_1 w^{p-1} + \ldots + \beta_p = 0 \qquad (2.1.3.)$$

tiene bajo estas condiciones p raíces $w_1,\ldots,w_p$ menores que 1 en valor absoluto. A su vez, podemos "invertir" la expresión (2.1.1.), es decir, expresar a y_t como función de los u_t contemporáneo y pasados mediante sucesivos reemplazos, hasta obtener

$$y_t = \mu + \sum_{j=0}^{\infty} \gamma_j u_{t-j}, \qquad (2.1.4.)$$

para ciertos coeficientes γ_j determinados unívocamente por los β_j, donde $\gamma_0 = 1$. La representación (2.1.4.) se dice que es un promedio móvil infinito (que denotamos MA(∞), como se verá en la Sección 3); (2.1.1.) y (2.1.4.) resultan equivalentes en media cuadrática (y por lo tanto en probabilidad), en el sentido de que la serie infinita en (2.1.4.) converge en media cuadrática y $\sum_{j=0}^{\infty} \gamma_j^2 < \infty$.

Las autocovarianzas del proceso forman una sucesión, se definen como $\sigma_s = E(y_t-\mu)(y_{t+s}-\mu)$ y son tales que $\sigma_s=\sigma_{-s}$. Multiplicando a (2.1.1.) sucesivamente por $y_t-\mu$, $y_{t-1}-\mu$, ... y tomando esperanzas matemáticas, se obtienen

$$\sigma_0 + \beta_1 \sigma_1 + \ldots + \beta_p \sigma_p = \sigma^2, \qquad (2.1.5.)$$

(1) Si fuera $Eu_t^2=0$, u_t sería la constante 0 con probabilidad 1 y la relación lineal (2.1.1) indicaría que y_t es "determinístico", lo que queremos descartar. La posibilidad de que el proceso tenga varianza infinita también es descartada, aunque es posible un análisis estadístico de este caso; ver, por ejemplo, Yohai y Maronna (1977).

y

$$\sigma_s + \beta_1 \sigma_{s-1} + \dots + \beta_p \sigma_{s-p} = 0, \quad s=1,2,\dots . \qquad (2.1.6.)$$

Estas pueden resolverse como una ecuación lineal en diferencias finitas (no estocástica), con condiciones de contorno $\sigma_s = \sigma_{-s}$ y condición de escala dada por (2.1.5.).

La transformada de Fourier de la sucesión de autocovarianzas da lugar a la aparición de la densidad espectral

$$f(\lambda) = \frac{\sigma^2}{2\pi} \left| \sum_{j=0}^{p} e^{i\lambda j} \beta_j \right|^{2}, \quad -\pi \leq \lambda \leq \pi , \qquad (2.1.7.)$$

que es tal que vale la fórmula de inversión

$$\sigma_s = \int_{-\pi}^{\pi} e^{i\lambda s} f(\lambda)\, d\lambda, \quad s=0, \pm 1, \pm 2, \dots \qquad (2.1.8.)$$

2.2. Orígenes

Wold (1954) atribuye a Yule (1927) la introducción del modelo AR(p) para el análisis empírico de las series cronológicas. Yule (1927) utilizó el modelo para el análisis de los datos de Wolfer sobre las manchas solares.

Wold (1954) propuso que el proceso (2.1.1.) fuera reconocido con el nombre de "proceso de Yule", designación que sin embargo no ha tenido aceptación generalizada en la literatura.

En el libro de Wold (1954) se aplica el modelo AR(2) para analizar empíricamente un índice de costo de la vida construído por G. Myrdal, Premio Nobel de Economía.

2.3. Primeras Sugerencias para Estimar los Parámetros

Con un registro finito $y_1, y_2, \dots, y_T$ (que denotamos con la misma letra que al proceso teórico) podemos estimar a μ mediante la media muestral

$$\bar{y} = \frac{1}{T} \sum_{t=1}^{T} y_t , \qquad (2.3.1.)$$

y a los σ_s mediante las autocovarianzas muestrales

$$c_s = \frac{1}{T} \sum_{t=s+1}^{T} (y_t - \bar{y})(y_{t-s} - \bar{y}), \quad s=0,1,\dots,T-1, \qquad (2.3.2.)$$

de manera que $c_{-s} = c_s$.

Antes de continuar, haremos una breve digresión sobre estos estimadores. Las propiedades estadísticas de estos y otros estimadores, se analizan en general a través de resultados asintóticos, cuando $T \to \infty$, por lo difícil que resultan los cálculos para muestras finitas. Es claro que $E(\bar{y}) = \mu$ y que si μ fuese conocida,

$$E\{\frac{1}{T-s}\sum_{t=s+1}^{T}(y_t-\mu)(y_{t-s}-\mu)\}=\sigma_s,\quad s=0,1,\ldots,T-1.$$

Cuando μ es desconocida, c_s y $c_s T/(T-s) = (T-s)^{-1} \sum_{t=s+1}^{T} (y_t-\bar{y})(y_{t-s}-\bar{y})$ tienen similares propiedades asintóticas, para cualquier s fijo. Por ejemplo, son asintóticamente insesgados. Se prefieren los estimadores (2.3.2.) pues matrices de covarianza muestrales formadas con ellos tienen la propiedad de ser definidas positivas; esto es importante, entre otras cosas, pues las correspondientes matrices de parámetros son también definidas (o por lo menos semi-definidas) positivas.

El uso de $\bar{y}$ en los dos factores $y_t-\bar{y}$ e $y_{t-s}-\bar{y}$ de (2.3.2.) tiene el inconveniente de que si hubiera cambios en el promedio de la serie observada (posiblemente incompatibles con el supuesto de estacionariedad, pero bastante frecuentes en la práctica) la inadecuada estimación del momento de primer orden μ, complicaría la estimación de la estructura de segundo orden: desde hace mucho tiempo se conoce la alternativa de reemplazar a estos factores por $y_t-\bar{y}^*_s$ e $y_{t-s}-\bar{y}^*_{s+}$, respectivamente, donde $\bar{y}^*_s = (T-s)^{-1} \sum_{t=1}^{T-s} y_t$ e $\bar{y}^*_{s+} = (T-s)^{-1} \sum_{t=1}^{T-s} y_{t+s}$, es decir, las medias aritméticas de los datos $y_1,\ldots,y_s$ e $y_{T-s+1},\ldots,y_T$, respectivamente. Para esta y otras alternativas, consultar, por ejemplo, Anderson (1971), Capítulo 8. Bajo el supuesto de estacionariedad, todos estos estimadores tienen las mismas propiedades asintóticas.

Para estimar a los β_j en (2.1.1.), un estimador del tipo de los momentos se obtiene al reemplazar a σ_s por c_s en las primeras p ecuaciones de (2.1.6.). El sistema resultante, escrito aquí en notación matricial,

$$\begin{bmatrix} c_0 & c_1 & \cdots & c_{p-1} \\ c_1 & c_0 & \cdots & c_{p-2} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ c_{p-1} & c_{p-2} & \cdots & c_0 \end{bmatrix} \begin{bmatrix} \beta^*_1 \\ \beta^*_2 \\ \cdot \\ \cdot \\ \cdot \\ \beta^*_p \end{bmatrix} = - \begin{bmatrix} c_1 \\ c_2 \\ \cdot \\ \cdot \\ \cdot \\ c_p \end{bmatrix} \qquad (2.3.3.)$$

se conoce con el nombre de ecuaciones muestrales de Yule y Walker, en honor a Yule y G. Walker. Por supuesto que (2.1.6.) son las ecuaciones teóricas de Yule y Walker.

Se sabe que los estimadores de los β_j provistos por este método son tales que la ecuación polinomial asociada (2.1.3.) con β_j reemplazado por su estimación, tiene todas sus raíces menores que 1 en valor absoluto; Pagano (1973), Anderson (1971a).

Estos estimadores están estrechamente vinculados a los por mínimos cuadrados simples que corresponden al modelo (2.1.1.). En efecto, las ecuaciones normales provenientes de minimizar $\sum_{t=p+1}^{T} \{ (y_t-\mu) + \beta_1(y_{t-1}-\mu) +\ldots+\beta_p (y_{t-p}-\mu) \}^2$ con respecto a $\mu, \beta_1,\ldots,\beta_p$ son

$$\sum_{t=p+1}^{T} \{ (y_t-\tilde{\mu}) +\tilde{\beta}_1 (y_{t-1}-\tilde{\mu}) +\ldots+\tilde{\beta}_p (y_{t-p}-\tilde{\mu}) \} = 0 \ , \qquad (2.3.4.)$$

y

$$\sum_{t=p+1}^{T} (y_t-\tilde{\mu})(y_{t-s}-\tilde{\mu}) = -\tilde{\beta}_1 \sum_{t=p+1}^{T} (y_{t-1}-\tilde{\mu})(y_{t-s}-\tilde{\mu})$$

$$-\ldots-\tilde{\beta}_p \sum_{t=p+1}^{T} (y_{t-p}-\tilde{\mu})(y_{t-s}-\tilde{\mu}), \quad s=1,\ldots,p. \qquad (2.3.5.)$$

Este sistema de $p+1$ ecuaciones en $p+1$ incógnitas es no lineal en los parámetros, problema que puede obviarse mediante una aproximación proveniente del siguiente razonamiento: la ecuación (2.3.4.) conduce a la estimación de μ mediante

$$\tilde{\mu} = \frac{\bar{y}_1 + \tilde{\beta}_1 \bar{y}_2 + \ldots + \tilde{\beta}_p \bar{y}_p}{1 + \tilde{\beta}_1 + \ldots + \tilde{\beta}_p} ; \qquad (2.3.6.)$$

si T es grande con respecto a p, podemos reemplazar a $\tilde{\mu}$ por $\bar{y} = \frac{1}{T} \Sigma_{t=1}^{T} y_t$. Reemplazando este valor en el sistema (2.3.5.) llegamos a las ecuaciones

$$\sum_{t=p+1}^{T} (y_t-\bar{y})(y_{t-s}-\bar{y}) = -\hat{\beta}_1 \sum_{t=p+1}^{T} (y_{t-1}-\bar{y})(y_{t-s}-\bar{y})$$

$$-\ldots-\hat{\beta}_p \sum_{t=p+1}^{T} (y_{t-p}-\bar{y})(y_{t-s}-\bar{y}), \quad s=1,\ldots,p, \qquad (2.3.7.)$$

que difieren de las ecuaciones muestrales de Yule y Walker (2.3.3.) en que todas las sumas van de $p+1$ a T, en lugar de contener cantidades variables de sumandos como en (2.3.2.).

El reemplazo de $\tilde{\mu}$ por $\bar{y}$, e incluso el del parámetro desconocido μ por uno de sus estimadores, $\tilde{\mu}$ ó $\bar{y}$, no afectará (en general) los desarrollos asintóticos ni las correspondientes propiedades, que son las que se analizan en estos casos; en efecto, tanto $\tilde{\mu}$ como $\bar{y}$ son estimadores consistentes de μ y en el análisis de las distribuciones límites (en general) es válido reemplazar a un estimador consistente por otro, o a un parámetro por uno de sus estimadores consistentes. Esto es particularmente cierto en el caso del Teorema Central del Límite, que es una herramienta típica de análisis asintótico en estos casos.

Los estimadores obtenidos de (2.3.7.) son los que se obtienen de un programa regular de regresión lineal múltiple, cuando se calcula la regresión de y_t en $y_{t-1},\ldots,y_{t-p}$ con T-p conjuntos completos de datos. Sin embargo, las propiedades tradicionales del método de mínimos cuadrados, por ejemplo las que provienen del Teorema de Gauss y Marcov, no son válidas en este caso puesto que los regresores no son fijos sino estocásticos.

Una vez disponibles los estimadores de los β_j podemos formar un estimador de σ^2 usando los "residuos"; por ejemplo, para (2.3.7.) tendremos

$$\hat{\sigma}^2 = \frac{1}{T} \sum_{t=p+1}^{T} [\, (y_t-\bar{y}) + \hat{\beta}_1 (y_{t-1}-\bar{y}) + \ldots + \hat{\beta}_p (y_{t-p}-\bar{y}) \,]^2 . \qquad (2.3.8.)$$

2.4. El Análisis de las Propiedades

Un análisis formal de las propuestas de la sección precedente implica considerar explícitamente a y_t como un proceso estocástico estacionario y a (2.1.1.) como una ecuación estocástica en diferencias finitas. Esto fué formulado a partir del trabajo de Mann y Wald (1943). Ellos comenzaron considerando la posibilidad de estimar a los β_j y σ^2 por máxima verosimilitud cuando el proceso es Gaussiano y además se supone que los valores iniciales $y_{-p+1},...,y_0$ son fijos.

Con este enfoque resultó que las propuestas de la sección precedente son asintóticamente equivalentes a la del método de máxima verosimilitud y que los resultados propios del análisis de un modelo de regresión lineal múltiple en variables explicativas no aleatorias son válidos en el sentido asintótico, cuando $T \rightarrow \infty$ mientras p permanece fijo.

2.5. Estimación por Máxima Verosimilitud

De lo dicho en la sección precedente se desprende que cuando $T \rightarrow \infty$ es razonable usar a (2.3.3.) o (2.3.7.) como equivalentes al método de máxima verosimilitud para estimar a los β_j. Un análisis detallado de estas propuestas y de sus propiedades aparece en Anderson (1971), Capítulo 5. Con la aparición del libro de Box y Jenkins (primera edición de 1970, impresión revisada en 1976) se enfatizó la posibilidad práctica de lograr estimaciones exactas σ aproximadas de los parámetros del modelo (2.1.1.) por el método de máxima verosimilitud, cuando se supone normalidad del proceso y_t.

Deben distinguirse dos casos. En uno se considera que el proceso se inició en y_{-p+1}, tuvo un estado (posiblemente) evolutivo hasta y_0, y de allí en adelante $y_1,...,y_T$ provienen de un proceso estacionario. La distribución (conjunta) condicional de $y_1,...,y_T$ dados $y_{-p+1},...,y_0$ es

$$(2\pi\sigma^2)^{-T/2}\, exp\, \{ -\frac{1}{2\sigma^2} \sum_{t=1}^{T} [\, (y_t-\mu) + \beta_1\, (y_{t-1}-\mu) + ... + \beta_p\, (y_{t-p}-\mu)\,]^2 \}. \qquad (2.5.1.)$$

Llamaremos a ésta la función de verosimilitud condicional y a los valores que la maximizan estimadores por máxima verosimilitud condicionados, pues dependen de los valores iniciales $y_{-p+1},...,y_0$.

Las ecuaciones provenientes de minimizar a (2.5.1.) con respecto a μ, $\beta_1,...,\beta_p$ son nuevamente no lineales en los parámetros, como lo eran (2.3.4.) y (2.3.5.). Consideramos entonces la simplificación consistente en estimar previamente a μ por $\bar{y}$, para luego obtener las ecuaciones normales

$$\sum_{t=1}^{T} (y_t-\bar{y})\, (y_{t-s}-\bar{y}) = -\hat{\beta}_1 \sum_{t=1}^{T} (y_{t-1}-\bar{y})\, (y_{t-s}-\bar{y}) - ... - \hat{\beta}_p \sum_{t=1}^{T} (y_{t-p}-\bar{y})\, (y_{t-s}-\bar{y}), \quad s=1,...,p\,. \qquad (2.5.2.)$$

Estas son similares a las ecuaciones (2.3.5.), excepto que las sumas van en todos los casos desde 1 hasta T. Si al resolver este sistema hacemos $y_{-p+1} = ... = y_0 = \bar{y}$ obtendremos soluciones idénticas a las de (2.3.7.).

Reemplazando estos estimadores en (2.5.1.) y maximizando la expresión resultante con respecto a σ^2, obtendremos el correspondiente estimador

$$\hat{\hat{\sigma}}^2 = \frac{1}{T} \sum_{t=1}^{T} [(y_t - \bar{y}) + \hat{\hat{\beta}}_1 (y_{t-1} - \bar{y}) + \ldots + \hat{\hat{\beta}}_p (y_{t-p} - \bar{y})]^2 . \tag{2.5.3.}$$

Un segundo enfoque consiste en suponer que el proceso estocástico $\{ y_t \}$ que genera los datos es Gaussiano, estacionario y estable, de manera que las observaciones $y_1,\ldots,y_T$ tienen una densidad normal conjunta dada por

$$(2\pi)^{-T/2} \mid \underset{\sim}{\Sigma} \mid^{-1/2} exp \{ - \frac{1}{2} (\underset{\sim}{y}_T - \underset{\sim}{\mu})' \underset{\sim}{\Sigma}^{-1} (\underset{\sim}{y}_T - \underset{\sim}{\mu}) \} , \tag{2.5.4.}$$

donde $\underset{\sim}{y}_T = (y_1,\ldots,y_T)'$ es el vector de observaciones, $\underset{\sim}{\mu} = E\, \underset{\sim}{y}_T = (\mu,\ldots,\mu)'$ el vector de esperanzas matemáticas, y $\underset{\sim}{\Sigma} = E\, (\underset{\sim}{y}_T - \underset{\sim}{\mu})\, (\underset{\sim}{y}_T - \underset{\sim}{\mu})'$ es la matriz de varianzas y covarianzas. Por comparación con el caso anterior, llamamos a (2.5.4.) como función de los parámetros, la función de verosimilitud incondicional y a los valores que la maximizan estimadores por máxima verosimilitud incondicionados.

Si bien los elementos σ_s de la matriz $\underset{\sim}{\Sigma}$, es decir, las autocovarianzas del proceso que satisfacen (2.1.5.) y (2.1.6.), no tienen formas explícitas fáciles de escribir en el caso general, los componentes de $\underset{\sim}{\Sigma}^{-1}$ fueron dados explícitamente por Siddiqui (1958) en función de los parámetros σ^2 y β_j, mientras que $|\underset{\sim}{\Sigma}|$ está dado explícitamente en Anderson y Mentz (1980) como función de las raíces w_j de (2.1.3.) y de σ^2.

La maximización de (2.5.4.) (independientemente del tratamiento que se de a μ) conduce a problemas no lineales, algunos de los cuales serán considerados en la Sección 4.

Desde el punto de vista de las propiedades estadísticas de los estimadores de los parámetros $\mu,\beta_1,\ldots,\beta_p$ y σ^2, asintóticamente todos los procedimientos sugeridos en esta sección para el método de máxima verosimilitud tienen las mismas propiedades: los estimadores son consistentes y se pueden realizar las inferencias como en el caso de regresores no estocásticos. A su vez, estos resultados asintóticos coinciden con los de los estimadores considerados en la Sección 2.3.

El trabajo Anderson y Mentz (1980) contiene un resumen de las propiedades del modelo AR(p) y un análisis del problema de existencia de los estimadores correspondientes a la función (2.5.4.).

2.6. Ejemplo: El Modelo AR(1)

Para aclarar algunos de los detalles de la presentación, consideraremos el caso más sencillo del modelo AR(1). Para simplificar la escritura consideraremos para la esperanza matemática que, o bien es conocida, en cuyo caso podemos tomarla como μ=0, o si es desconocida estimarla por $\bar{y}$ y en todas las expresiones donde corresponde, reemplazar a y_t por $y_t-\bar{y}$. Esta es a menudo la manera como se opera en la práctica, por ejemplo (internamente) en muchos de los programas de cómputo disponibles.

El modelo autorregresivo de primer orden, AR(1), postula que

$$y_t = -\beta\, y_{t-1} + u_t\,, \qquad t=\ldots, -1,0,1,\ldots\,. \tag{2.6.1.}$$

La ecuación polinomial asociada es $w+\beta=0$, de donde se deduce que $|\beta| < 1$ es la condición para que y_t sea estacionario e independiente de $u_{t+1}, u_{t+2},\ldots$, e invertible en $y_t = \sum_{j=0}^{\infty} (-\beta)^j\, u_{t-j}$. Las funciones de autocovarianzas y de autocorrelaciones son,

$$\sigma_s = (-\beta)^s \frac{\sigma^2}{1-\beta^2}\,, \qquad \rho_s = (-\beta)^s\,, \tag{2.6.2.}$$

respectivamente, y la densidad espectral es

$$f(\lambda) = \frac{\sigma^2}{2\pi}\,\frac{1}{1+\beta^2+2\beta\cos\lambda}\,, \qquad -\pi \leq \lambda \leq \pi\,. \tag{2.6.3.}$$

Para derivar la función de verosimilitud en el caso normal, consideramos a $u_1,\ldots,u_T$ normales independientes, con densidad

$$(\sigma^2 2\pi)^{-T/2} \exp\left[-\frac{1}{2\sigma^2}\sum_{t=1}^{T} u_t^2\right]. \tag{2.6.4.}$$

Suponemos que $y_1=u_1-\beta y_0$, $y_t+\beta y_{t-1}=u_t$, $t=2,\ldots,T$. En resumen

$$\begin{bmatrix} 1 & 0 & 0 & \ldots & 0 & 0 \\ \beta & 1 & 0 & \ldots & 0 & 0 \\ 0 & \beta & 1 & \ldots & 0 & 0 \\ . & . & . & & . & . \\ . & . & . & & . & . \\ . & . & . & & . & . \\ 0 & 0 & 0 & \ldots & \beta & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ . \\ . \\ . \\ y_T \end{bmatrix} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ . \\ . \\ . \\ u_T \end{bmatrix} + \begin{bmatrix} -\beta y_0 \\ 0 \\ 0 \\ . \\ . \\ . \\ 0 \end{bmatrix}. \tag{2.6.5.}$$

De esta transformación se desprende que $y_1,\ldots,y_T$ tienen densidad

$$(\sigma^2 2\pi)^{-T/2} \exp\left[-\frac{1}{2\sigma^2}\sum_{t=1}^{T} (y_t+\beta y_{t-1})^2\right], \tag{2.6.6.}$$

condicional en y_0. El máximo de (2.6.6.) con respecto a β conduce a

$$\hat{\beta} = -\frac{\sum_{t=1}^{T} y_t y_{t-1}}{\sum_{t=1}^{T} y_{t-1}^2}\,. \tag{2.6.7.}$$

Poniendo y_0=0 construímos el estimador por mínimos cuadrados simples,

$$\tilde{\beta} = - \frac{\sum_{t=2}^{T} y_t y_{t-1}}{\sum_{t=2}^{T} y_{t-1}^2} = - \frac{\sum_{t=1}^{T-1} y_t y_{t+1}}{\sum_{t=1}^{T-1} y_t^2} . \qquad (2.6.8.)$$

El criterio de Yule-Walker conduce a estimar a β por

$$\beta^* = - \frac{\sum_{t=1}^{T-1} y_t y_{t+1}}{\sum_{t=1}^{T} y_t^2} = - \frac{\sum_{t=2}^{T} y_t y_{t-1}}{\sum_{t=1}^{T} y_t^2} , \qquad (2.6.9.)$$

que utiliza a todas las observaciones disponibles en el denominador.

En el caso en que suponemos a $y_1,...,y_T$ normal multivariante, puede demostrarse que la función de verosimilitud es

$$(\sigma^2 2\pi)^{-T/2} \sqrt{1-\beta^2}\, exp \{ - \frac{1}{2\sigma^2} [(1-\beta^2)\, y_1^2 + \sum_{t=2}^{T} (y_t+\beta y_{t-1})^2] \} . \qquad (2.6.10.)$$

Los valores de β de σ^2 que hacen máxima esta expresión fueron escritos en forma explícita por Hasza (1980).

3. EL MODELO DE PROMEDIOS MOVILES

3.1. Definición y Propiedades

El modelo de promedios móviles de orden $q \geq 1$, denotado MA(q), se define por

$$y_t - \mu = u_t + \alpha_1 u_{t-1} + ... + \alpha_q u_{t-q} , \qquad (3.1.1.)$$

donde los y_t son observables y los u_t son como en el modelo AR(p) ya considerado. Los q+2 parámetros del modelo son $\mu, \alpha_1, ..., \alpha_q$ y $\sigma^2 = Eu_t^2 (0 < \sigma^2 < \infty)$. Aquí q es finito.

Para t=...,-1,0,1,... (3.1.1.) define un proceso estacionario débil para cualquier valor de los α_j. Si además la ecuación polinomial asociada

$$z^q + \alpha_1 z^{q-1} + ... + \alpha_q = 0 \qquad (3.1.2.)$$

tiene raíces $z_1,...,z_q$ menores que 1 en valor absoluto, el proceso es invertible,

$$u_t = \sum_{j=0}^{\infty} \delta_j (y_{t-j} - \mu) , \qquad (3.1.3.)$$

donde $\delta_0=1$. Las representaciones (3.1.1.) y (3.1.3.) son equivalentes (en media cuadrática y en probabilidad) y permiten escribir MA(q) = AR(∞), advirtiendo que en (3.1.3.) $\delta_j=\delta_j(\alpha_1,\dots,\alpha_q)$, funciones de los α_r .

El proceso tiene autocovarianzas

$$\sigma_s=\sigma^2\sum_{j=0}^{q-|s|}\alpha_j\,\alpha_{j+|s|}\,,\qquad s=0,\pm 1,\pm 2,\dots,\pm q\,, \tag{3.1.4.}$$
$$=0\,,\qquad |s|>q\,.$$

La densidad espectral es

$$f(\lambda)=\frac{\sigma^2}{2\pi}\,\Big|\sum_{j=0}^{q}\alpha_j\,e^{i\lambda j}\Big|^2\,,\qquad -\pi\le\lambda\le\pi\,, \tag{3.1.5.}$$

y es tal que vale la fórmula de inversión (2.1.7.).

3.2. Orígenes

Yule (1921, 1926) y Slutzky (1927, 1937) consideraron esquemas de este tipo. Slutzky (1937) consideró combinaciones lineales de términos puramente aleatorios (nuestros u_t) como fuentes de procesos cíclicos.

Wold (1954) propuso que el proceso (3.1.1.) fuera reconocido con el nombre de "proceso de Slutzky", designación que sin embargo no ha tenido aceptación generalizada en la literatura.

En el libro de Wold (1954) se aplican los modelos MA(1), MA(2) y MA(4) a los índices de precios del trigo en Europa Occidental, 1518-1869, recopilados por W. Beveridge.

3.3. Primeras Sugerencias para Estimar los Parámetros

Wold (1954) aparece como el primer autor que estimó concretamente los parámetros de un modelo de promedios móviles con datos observados. Su estimación es por lo que llamamos actualmente el método de los momentos. Por ejemplo, si $q=1$, tenemos que $\sigma_0=\sigma^2(1+\alpha^2)$, $\sigma_1=\sigma^2\alpha$ y por lo tanto el coeficiente de autocorrelación (teórico) de primer orden es $\rho_1=\alpha\,/\,(1+\alpha^2)$. Formando el análogo muestral

$$r_1=\frac{\sum_{t=1}^{T-1}(y_t-\bar{y})\,(y_{t+1}-\bar{y})}{\sum_{t=1}^{T}(y_t-\bar{y})^2}\,, \tag{3.3.1.}$$

y resolviendo para α la ecuación $r_1=\alpha/(1+\alpha^2)$, obtenemos como estimador admisible (menor que 1 en valor absoluto cuando $|r_1|<\frac{1}{2}$)

$$\alpha^* = \frac{1-\sqrt{1-4r_1^2}}{2r_1}. \tag{3.3.2.}$$

De manera similar se puede proceder para modelos de orden $q > 1$.

3.4. El Análisis de las Propiedades

Correspondió a Whittle (1953), un discípulo de Wold, analizar las propiedades asintóticas de los estimadores considerados en la sección precedente. Whittle argumentó que α^* definido en (3.3.2.) es consistente para $\alpha(plim_{T\to\infty}\ \alpha^* = \alpha)$, pero que es asintóticamente ineficiente comparado con el estimador por máxima verosimilitud en el caso normal.

Whittle encontró que la varianza de la distribución normal límite de α^* definido en (3.3.2.) es

$$\frac{1+\alpha^2+4\alpha^4+\alpha^6+\alpha^8}{T(1-\alpha^2)^2} = \frac{1-\alpha^2}{T} + \alpha^2\,\frac{4+\alpha^2\,(1+\alpha^2)^2}{T(1-\alpha^2)^2}, \tag{3.4.1.}$$

mientras que para el estimador por máxima verosimilitud la varianza de la distribución normal límite es sólo el primer sumando del miembro de (3.4.1.), es decir,

$$\frac{1-\alpha^2}{T}. \tag{3.4.2.}$$

Los argumentos de Whittle fueron en general de tipo informal, pero luego se proveyeron demostraciones rigurosas de las aseveraciones citadas.

Como conclusión de este análisis surgió la necesidad de mejorar la estimación sugerida por el método de los momentos. Una posibilidad consistía (en el caso q=1) en mejorar la estimación de r_1 que entra en (3.3.2.); otra consistía en explotar la relación MA(q)=AR(∞), pues hemos visto que para el modelo AR(p) las ideas intuitivas del tipo de las ecuaciones de Yule-Walker o mínimos cuadrados simples, condujeron a estimadores asintóticamente equivalentes a los por máxima verosimilitud.

3.5. Aproximaciones a la Estimación por Máxima Verosimilitud

Whittle (1951) sugirió un enfoque iterativo para resolver un sistema de ecuaciones que eran asintóticamente equivalentes al método de máxima verosimilitud. Este autor utilizó sus ideas para estimar los parámetros con datos simulados. Esta sugerencia no tuvo un gran impacto práctico hasta la difusión del trabajo de Durbin que comentamos a continuación.

Durbin (1959, 1961) logró mejorar la estimación en el modelo MA(q) explotando la relación MA(q)=AR(∞). Su razonamiento es que si aproximamos al esquema MA(q) por uno AR(k), para un orden k suficientemente grande, y estimamos los parámetros del modelo AR(k) por máxima verosimilitud o por mínimos cuadrados (sin restricciones), esos estimadores pueden combinarse y proporcionar estimadores de los parámetros del modelo MA(q) que "heredarán" entonces (aproximadamente) las propiedades asintóticamente óptimas.

En el caso concreto del modelo AR(1), si $\hat{\beta}_1,...,\hat{\beta}_k$ con los estimadores de los parámetros del modelo AR(k), mostró que

$$\tilde{\alpha} = - \frac{\sum_{i=0}^{k-1} \hat{\beta}_i \hat{\beta}_{i+1}}{\sum_{i=0}^{k-1} \hat{\beta}_i^2} \tag{3.5.1.}$$

es el estimador deseado, donde $\hat{\beta}_0=1$. Las propiedades de este estimador fueron analizadas por Durbin (1959) y más formalmente por Mentz (1977). Varios autores realizaron comparaciones para muestras finitas en pruebas Monte Carlo.

Nótese que en (3.5.1.) el valor de k debe determinarse explícitamente para una aplicación práctica de la propuesta.

Walker (1961) siguió un enfoque similar, pero basado en la estimación de los coeficientes de autocorrelación. En consecuencia, su propuesta puede considerarse como un mejoramiento de (3.3.2.) basado en un mejoramiento de r_1 como estimador de ρ_1 (para q=1), o como una formulación alternativa de la propuesta de Durbin (1959). Si $r_1,...,r_k$ son estimadores de las primeras k autocorrelaciones del proceso, definidos por

$$r_s = \frac{\sum_{t=1}^{T-s} (y_t-\bar{y})(y_{t+s}-\bar{y})}{\sum_{t=1}^{T} (y_t-\bar{y})^2} \quad , \quad s = 1,..., k, \tag{3.5.2.}$$

el estimador de ρ_1 propuesto por Walker (1961) es

$$\hat{\rho}_1 = r_1 + \sum_{j=2}^{k} m_j r_j \, , \tag{3.5.3.}$$

donde los m_j son también funciones de r_1 . Al igual que en la propuesta de Durbin (1959), el valor de k debe establecerse en cada aplicación práctica de la propuesta.

Walker (1961) analizó algunas propiedades de su propuesta y Mentz (1977a) formalizó un análisis cuando T $\rightarrow \infty$ y k=k(T) $\rightarrow \infty$

Las propuestas de Durbin y de Walker originaron a partir de 1960 un gran caudal de estudios. Se analizaron matemáticamente las propiedades asintóticas y se propusieron modificaciones a los métodos, y se realizaron además comparaciones Monte Carlo entre éstas y otras propuestas.

De estos estudios surgió que la varianza asintótica (3.4.2.) es alcanzada por estas propuestas (y por otras derivadas de ellas), cuando no sólo consideramos que $T \to \infty$, sino que la cantidad k de valores muestrales es función de T, $k=k(T) \to \infty$, pero a una velocidad menor que T; por ejemplo, en algunos casos se requiere que $k(T)^2 / T \to 0$ cuando $T \to \infty$.

3.6. Estimación por Máxima Verosimilitud

El libro de Box y Jenkins (1976) enfatizó para este modelo la posibilidad práctica de lograr estimaciones exactas o aproximadas de los parámetros por el método de máxima verosimilitud bajo el supuesto de normalidad. Un excelente resúmen es el trabajo de Godolphin y de Gooijer (1982).

Nuevamente la función de verosimilitud para el vector $\underset{\sim}{y}_T=(y_1,...,y_T)'$ de observaciones del modelo es

$$(2\pi)^{-T/2} \mid \underset{\sim}{V} \mid^{-\frac{1}{2}} exp \left[-\frac{1}{2}(\underset{\sim}{y}_T-\underset{\sim}{\nu})' \; \underset{\sim}{V}^{-1} \; (\underset{\sim}{y}_T-\underset{\sim}{\nu}) \right] , \tag{3.6.1.}$$

donde $\underset{\sim}{\nu} = E \; \underset{\sim}{y}_T$, $\underset{\sim}{V} = E \; (\underset{\sim}{y}_T-\underset{\sim}{\nu}) \; (\underset{\sim}{y}_T-\underset{\sim}{\nu})'$. Aquí se desconoce la forma general de $\underset{\sim}{V}^{-1}$ y de $\mid \underset{\sim}{V} \mid$, por lo que las posibilidades de operar con la función de verosimilitud exacta son escasas. En el caso $q=1$ Godolphin y de Gooijer derivaron un procedimiento iterativo para obtener los estimadores (exactos) de α y σ^2 usando expresiones conocidas de los componentes de $\underset{\sim}{V}^{-1}$, y $\mid \underset{\sim}{V} \mid = (1-\alpha^{2T+2}) / (1-\alpha^2)$.

Un resúmen de una gran cantidad de propuestas para estimar el parámetro α de un modelo MA(1) aparece en el trabajo de Mentz y Antelo (1977). En el trabajo Anderson y Mentz (1980) se demuestra que si $\underset{\sim}{y}_T \neq \underset{\sim}{0} = (0,...,0)'$ los estimadores correspondientes a (3.6.1.) existen.

4. EL MODELO MIXTO

Las ideas de una autorregresión que liga a las variables aleatorias observables, definida por el miembro de la izquierda de (2.1.1.), y la de una combinación lineal de variables aleatorias inobservables, definida por el miembro de la derecha de (3.1.1.), se pueden combinar para definir un proceso lineal con un número finito de parámetros que generaliza simultáneamente al AR(p) y al MA(q). Se trata del modelo mixto o ARMA(p,q) (autorregresivo con errores promedios móviles, de orden (p,q)), que con las mismas letras definimos por

$$\sum_{j=0}^{p} \beta_j (y_{t-j}-\mu) = \sum_{k=0}^{q} \alpha_k u_{t-k} . \qquad (4.1.)$$

Para la estacionariedad se requiere que los β_j hagan que las raíces de (2.1.3.) sean todas menores que 1 en valor absoluto; en cambio las α_k pueden tomar valores reales sin restricciones. Si las raíces de (2.1.3.) son todas menores que 1 en valor absoluto, (4.1.) se puede invertir para obtener un promedio móvil infinito: simbólicamente, ARMA(p,q)=MA(∞); si las raíces de (3.1.2.) son todas menores que 1 en valor absoluto, (4.1.) se puede invertir para obtener una autorregresión infinita, ARMA(p,q)=AR(∞). En cada caso tendremos que deducir las relaciones que ligan a los coeficientes de las representaciones.

La posibilidad de usar este tipo de modelos en la práctica fué enfatizada por Box y Jenkins en su libro publicado inicialmente en 1970 (1976) y desde entonces atrajo considerable interés teórico y aplicado.

No es fácil en este caso operar con la sucesión de autocovarianzas, definida nuevamente por $E(y_t-\mu)(y_{t+s}-\mu)$. Resulta interesante destacar que la densidad espectral es

$$\frac{\sigma^2}{2\pi} \frac{\left|\sum_{k=0}^{q} \alpha_k e^{i\lambda k}\right|^2}{\left|\sum_{j=0}^{p} \beta_j e^{i\lambda j}\right|^2}, \qquad (4.2.)$$

o sea una función racional que generaliza a (2.1.7.) y (3.1.5.).

Un tratamiento de este nuevo modelo en sus aspectos probabilísticos y estadísticos nos llevaría mas allá del propósito de esta presentación. En esta sección nos limitaremos a dar una descripción concisa de algunos aspectos de trabajos contemporáneos destinados a la estimación de los parámetros de (4.1.) por máxima verosimilitud, exacta o aproximada, en el caso Gaussiano. Como ya lo comentamos antes (Sección 2.5.) existe mucho interés en este tipo de estimación, con preferencia a otros tipos de aproximaciones que se pueden considerar en el caso del modelo mixto.

Aparte de limitar la longitud del trabajo, esta discusión servirá para el siguiente propósito: Está claro que la familia de modelos ARMA(p,q) para $p \geq 0$ y $q \geq 0$ (finitos) incluye a los modelos que tratamos en las secciones 2 y 3, pues AR(p) = ARMA(p,0) y MA(q) = ARMA(0,q). Una de las consecuencias del interés práctico en estos modelos, ha sido el desarrollo de enfoques, algoritmos y programas de cómputo eficaces. Nuestras referencias a estos importantes problemas fueron limitadas y en esta sección presentaremos una discusión, que al considerar el caso de los modelos ARMA, incluye a los "puros" AR y MA ya estudiados.

El problema de obtener estimadores de los parámetros aplicando exactamente el método de máxima verosimilitud es complicado, pues las matrices de covarianzas, o sus determinantes o inversas, no son conocidas en todos los casos, y cuando lo son, incluyen formas paramétricas complicadas. Tres enfoques posibles son la obtención de fórmulas explícitas para los estimadores, procedimientos basados en el cálculo de valores de la función de verosimilitud y los métodos iterativos.

4.1. Fórmulas Explícitas

Aparentemente el único resultado conocido es el provisto por Hasza (1980), para la función de verosimilitud incondicional del modelo AR(1); ver Sección 2.6.

4.2. Cálculo de Valores de la Función de Verosimilitud

Un método usado frecuentemente consiste en deducir algún procedimiento para computar los valores numéricos de la función de verosimilitud, o de su logaritmo, a medida que se dan valores a los parámetros, para luego usar un algoritmo de optimización en el computador que requiera solamente los valores de la función de verosimilitud para proporcionar los estimadores deseados.

Por ejemplo, el paquete de subrutinas Fortran "IMSL Library" (1979) utiliza un algoritmo de camino más escarpado ("steepest descent") modificado para llegar a los estimadores de los parámetros de modelos ARMA y sus versiones no estacionarias. El paquete de programas "BMDP" (1985) usa un método del tipo Gauss-Marquardt para realizar estimaciones en casos lineales y no lineales.

Una idea en este terreno fué la de Box y Jenkins (1976) que utilizaron el procedimiento conocido como "backcasting" para evaluar la función de verosimilitud aproximada de varios modelos de series de tiempo. Otras ideas útiles han sido el uso de la descomposición de Cholesky de la matriz de covarianzas o de su inversa, y la "fórmula de Woodbury", como en Phadke y Kedem (1978).

Ansley (1979) estudió varios enfoques y propuso un nuevo algoritmo; él analizó las contribuciones de Newbold (1974), Dent (1977), Ali (1977), Osborn (1977) y Hillmer y Tiao (1979), mostró la equivalencia de las dos primeras y las relacionó con la contribución de Ali (1977). Otro trabajo en este grupo es el de Nicholls y Hall (1979).

4.3. Procedimientos Iterativos

Muchos de los procedimientos iterativos usan las ecuaciones de verosimilitud (obtenidas al igualar a 0 las derivadas de la función de verosimilitud) para obtener procedimientos de la forma $\hat{\theta}^{(i)} = g(\hat{\theta}^{(i-1)})$, para alguna función g que depende de los datos $y_1,...,y_T$, donde $\hat{\theta}^{(i)}$ para $i=1,2,...$ son sucesivos valores numéricos de las estimaciones, mientras que $\hat{\theta}^{(0)}$ es un valor inicial. Anderson (1975, 1977) consideró varios procedimientos para la estimación máximo verosímil exacta y algunas aproximaciones. Godolphin y de Gooijer (1982) presentaron un procedimiento para el modelo MA(1), como comentamos en la Sección 3.6.

En general, derivar procedimientos iterativos tiende a implicar mas elaboraciones matemáticas con la función de verosimilitud y mas conocimiento de ella, con relación a los enfoques considerados en la sección precedente.

Anderson y Mentz (1987) analizaron en detalle una serie de procedimientos iterativos para el modelo MA(1). Algunos de estos procedimientos se pueden relacionar con uno conocido en la literatura como el "algoritmo EM" (Dempster *et al*, 1977) que fué propuesto inicialmente para resolver problemas de datos faltantes. La conexión la proporciona el hecho de que podemos considerar a los valores iniciales como datos faltantes, en cuyo caso el algoritmo EM sugerirá una sucesión de pasos alternando el cálculo de esperanzas condicionales (E) y maximizaciones (M), lo que conducirá a un procedimiento iterativo.

Por otra parte, varios autores han relacionado estos problemas con el "filtro de Kalman", que es otro procedimiento iterativo que ha encontrado múltiples aplicaciones en el campo de las series cronológicas. Ver, por ejemplo, Harvey (1981).

5. CONCLUSIONES

El progreso alcanzado en el campo de la estimación de los parámetros de los modelos lineales con un número finito de parámetros (AR(p), MA(q) y ARMA(p,q)), desde su difusión en el campo de la estadística hasta el presente, es muy importante. En los 65 años que van desde 1921 hasta el presente, se ha logrado progresar en el conocimiento teórico de las propiedades de los procesos, en el análisis inferencial basado en teoría asintótica y en algunos resultados para muestras finitas, en comparaciones Monte Carlo para tamaños muestrales pequeños, en el desarrollo de algoritmos computacionales y en el uso concreto de los modelos para análisis, predicción y control.

Los primeros trabajos de estimación enfocaron el problema a través del método de los momentos o intentaron la estimación por mínimos cuadrados. El método de máxima verosimilitud fué considerado en forma teórica o a través de aproximaciones, pues los problemas de cómputos eran difíciles o imposibles de solucionar. Al fin de esta etapa los trabajos de Durbin (1959) y Walker (1961) intentaron aproximar la estimación máximo verosímil. A fines de la década de 1960 y comienzos de la siguiente, aparecieron una serie de trabajos importantes, entre otros los de Hannan (1969), Parzen (1971) y Anderson (1975), los que fueron reconsiderados y comparados desde un punto de vista teórico por Anderson (1975a, 1977); la colección de trabajos editada por Brillinger y Tiao (1980) es importante.

La aparición de los libros de Box y Jenkins (1976) y Anderson (1971) a comienzos de la década de 1970 coincidió con un gran incremento en el interés en los temas de series cronológicas en general, y en la estimación de parámetros en particular. Algunos de los desarrollos recientes fueron considerados en nuestra Sección 4. A comienzos de la década de 1980 debe destacarse la aparición del texto de Priestley (1981), entre otros de lo que es ahora una bibliografía muy extensa en el campo de las series cronológicas.

El área de la inferencia en procesos estocásticos es comparativamente difícil y quedan muchas cosas para investigar y descubrir. La cantidad de especialistas dedicados al tema ha aumentado de unos pocos alrededor de 1970, a una cantidad importante en 1980. Es de confiar que con el transcurso de pocos años más, los problemas importantes reconocidos en la actualidad sean resueltos y se desarrollen nuevas y atrayentes áreas de aplicación.

BIBLIOGRAFIA

ALI, M. M. (1977), Analysis of autoregressive-moving average models: Estimation and prediction, *Biometrika*, 64, 535-45.

ANDERSON, T. W. (1971), *The Statistical Analysis of Time Series*, New York: John Wiley.

ANDERSON, T. W. (1971a), "The Stationarity of an Estimated Autoregressive Process", Technical Report N.º 7, Department of Statistics, Stanford University.

ANDERSON, T. W. (1975), "Estimation by Maximum Likelihood in Autoregressive Moving Average Models in the Time and Frequency Domains", Technical Report N.º 20, Department of Statistics, Stanford University.

ANDERSON, T. W. (1975a), "Maximum Likelihood Estimation of Parameters of an Autoregressive Process with Moving Average Residuals and Other Covariance Matrices with Linear Structure", *Annals of Statistics*, 3, 1283-04.

ANDERSON, T. W. (1977), "Estimation for Autoregressive Moving Average Models in the Time and Frequency Domains", *Annals of Statistics*, 5, 842-865.

ANDERSON, T. W. and R. P. MENTZ (1980), "On the Structure of the Likelihood Function of Autoregressive and Moving Average Models", *Journal of Time Series Analysis*, 1, 83-94.

ANDERSON, T. W. and R. P. MENTZ (1987), Some results on exact maximum likelihood estimation in time series models, *Bulletin of the ISI*, 19-20.

ANSLEY, C. F. (1979), "An Algorithm for the Exact Likelihood of a Mixed Autoregressive Moving Average Process", *Biometrika*, 66, 59-65.

ANSLEY, C. F. and P. NEWBOLD (1980), "Finite Properties of Estimators for Autoregressive Moving Average Models", *Journal of Econometrics*, 13, 159-183.

BMDP (1985), BMDP Statistical Software, Los Angeles, California.

BOX, G. E. P. and G. M. JENKINS (1976), *Time Series Analysis, Forecasting and Control*, Revised Edition, San Francisco: Holden-Day.

BRILLINGER, D. R. and G. C. TIAO (editors) (1980), *Directions in Time Series*, Institute of Mathematical Statistics.

DEMPSTER, A. P., N. M. LAIRD and D. B. RUBIN (1977), Maximum Likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society*, Series B, 39, 1-38.

DENT, W. (1977), Computation of the exact Likelihood function of an ARIMA process, *J. Statist. Comp. and Simul.* 5, 193-206.

DURBIN, J. (1959), "Efficient Estimation of Parameters in Moving Average Models", *Biometrika*, 46, 306-316.

DURBIN, J. (1961), "Efficient Fitting of Linear Models for Continuous Stationary Time Series from Discrete Data", *Bulletin of the International Statistical Institute*, 38, 273-282.

GODOLPHIN, E. J. and J. G. de GOOIJER (1982), "On the Maximum Likelihood Estimation of the Parameters of Gaussian Moving Average process", *Biometrika*, 69, 443-451.

HANNAN, E. J. (1969), "The Estimation of Mixed Moving Average Autoregressive Systems", *Biometrika*, 56, 579-593.

HARVEY, A. C. (1981), *Time Series Models*, Philip Allan Pub. Ltd., Oxford.

HASZA, D. F. (1980), "A Note on Maximum Estimation for the First-Order Autoregressive Process", *Communications in Statistics*, A9, 1411-1415.

HILLMER, S. G. and G. C. TIAO (1979), Likelihood function of stationary multiple autoregressive moving average models, *J. Amer Statist. Assoc.*, 74, 652-61.

IMSL (1979), The IMSL Library, Houston, Texas.

LJUNG, G. M. and G. E. P. BOX (1979), "The Likelihood Function of Stationary Autoregressive Moving Average Models", *Biometrika*, 66, 265-270.

MANN, H. B. and A. WALD (1943), "On the Statistical Treatment of Linear Stochastic Difference Equations", *Econometrica*, 11, 173-220.

MENTZ, R. P. (1977), "Estimation in the First Order Moving Average Model Through the finite Autoregressive Approximation", *J. of Econometrics*, 6, 225-236.

MENTZ, R. P. (1977a), "Estimation in the First Order Moving Average Model Based on Sample Autocorrelations", *Annals of Statistics*, 5, 1250-57.

MENTZ, R. P. and R. ANTELO (1977), "Estimation Procedures for the First-Order Moving Average Model", *Proceedings of the Business and Economic Statistics Section*, Part 1, 329-334, American Statistical Association.

NEWBOLD, P. (1974), The exact likelihood function for a mixed autoregressive-moving average process, *Biometrika*, 61, 423-26.

NICHOLLS, D. F. and A. D. HALL (1979), The exact likelihood function of multivariate autoregressive moving average models, *Biometrika*, 66, 259-64.

OSBORN, D. R. (1977), Exact and approximate maximum likelihood estimators for vector moving average processes, *J. R. Statist. Soc. B*, 39, 114-18.

PAGANO, M. (1973), When is an Autoregressive Scheme Sationary?, *Communications in Statistics*, 1, 533-544.

PARZEN, E. (1971). "Efficient Estimation of Stationary Time Series Mixed Schemes", *Bulletin of the International Statistical Institute*, 44, 315-319.

PHADKE M. S. and G. KEDEM (1978), Computation of the exact likelihood function of multivariate moving average models, *Biometrika*, 65, 511-20.

PRIESTLEY, M. R. (1981), *Spectral Analysis and Time Series*, Vol. 1, London: Academic Press.

SCHUSTER, A. (1898), "On the Investigation of Hidden Periodicities with Application to a Supposed 26 Day Period of Meteorological Phenomena", *Terrestrial Magnetism*, March, 13-41.

SCHUSTER A. (1900), "The Periodogram of Magnetic Declination as obtained from the record of the Greenwich Observatory during the years 1871-1895". *Transactions of the Cambridge Philosophical Society*, 18, 107-135.

SIDDIQUI, M. (1958), "On the Inversion of the Sample Covariance Matrix in a Stationary Autoregressive Process", *The Annals of Mathematical Statistics*, 29, 585-588.

SLUTZKY, E. (1927), "The Summation of Random Causes as the Source of Cyclic Processes". (Russian with an English summary). *Problems of Economics Conditions*, 3, N.º 1.

SLUTZKY, E. (1937), "The Summation of Random Causes as the Source of Cyclic Processes", *Econometrica*, 5, 105.

WALKER, A. M. (1961), "Large Sample Estimation of Parameters for Moving Average Models", *Biometrika*, 48, 343-357.

WHITTLE, P. (1951), *Hypothesis Testing in Time Series Analysis*, Uppsala: Almqvist and Wicksells.

WHITTLE, P. (1953), "Estimation and Information in Stationary Time Series", *Arkiv für Mathematik*, 2, 423-424.

WOLD, H. (1954), *A Study in the Analysis of Stationary Time Series, 2nd. edition* (first edition of 1938), Stockholm: Almqvist and Wicksells.

YOHAI, V. J. and R. A. MARONNA (1977), Asymptotic behavior of least squares estimates for autoregressive processes with infinite variances, *Annals of Statistics*, Vol 5, 554.

YULE, G. U. (1921), "On the Time-Correlation Problem, with Special Reference to the variate-Difference Correlation Method", *Journal of the Royal Statistical Society*, 84, 497-526.

YULE, G. U. (1926), "Why do we Sometimes get Nonsense-Correlations Between Time-Series? - A Study in Sampling and the Nature of Time-Series", *Journal of the Royal Statistical Society*, 89, 1-64.

YULE, G. U. (1927), "On a Method of Investigating Periodicities in Disturbed Series, with special reference to Wolfer's Sunspot Numbers", *Philosophical Transactions*, A 226, 267-298.

ESTIMATION OF AUTOREGRESSIVE MOVING AVERAGE MODELS

SUMMARY

This work summarizes the present status of the estimation of autoregressive moving average models in the time domain.

For every model studied, the probability distributions are defined and the most important properties are mentioned. Then the suggestions for moment and least square estimation are analyzed as well as maximum likelihood for normal errors.

Finally, some recent ideas for maximum likelihood estimation of mixed models are revised. A conclusion of this survey is that in the last 65 years many important developments have taken place in this area.