

ANÁLISIS DE LA EVOLUCIÓN DE LAS TÉCNICAS DE SERIES TIEMPO. UN ENFOQUE UNIFICADO

JUAN CARLOS ABRIL

*Consejo Nacional de Investigaciones, Científicas y Técnicas (CONICET),
Facultad de Ciencias Económicas, Universidad Nacional de Tucumán
jabril@herrera.unt.edu.ar*

RESUMEN

Los sucesos variables en el tiempo reciben el nombre genérico de series de tiempo o series cronológicas. Es extremadamente difícil presentar una descripción breve del campo de las series de tiempo. La dificultad se basa en el hecho de que la materia es por sí misma muy compleja, siendo una rama de la estadística pero con su metodología y su propio vocabulario peculiar. La idea básica de una *serie de tiempo* es muy simple, consiste en el registro de cualquier cantidad fluctuante medida en diferentes puntos del tiempo. La característica común de todos los registros que pertenecen al dominio de las “series de tiempo” es que ellos están influenciados, aunque sea parcialmente, por fuentes de *variación aleatoria*. Entonces, si deseamos explicar la estructura de las fluctuaciones en una serie de tiempo debemos recurrir a lo que llamamos el *estudio de las series de tiempo*. Hay dos aspectos en el estudio de las series de tiempo: el análisis y el modelado. El objetivo del análisis es resumir las propiedades de una serie y remarcar sus características salientes. La principal razón para modelar una serie de tiempo es para permitir la predicción de sus valores futuros.

En este trabajo se presenta una breve introducción al estudio de las series de tiempo seguido de un conjunto de ejemplos que ocurren en áreas tales como medicina, astronomía, economía, etc. Luego se introduce una pequeña historia del desarrollo de esta parte de la estadística, comenzando en 1664 con los trabajos de Sir Isaac Newton, hasta llegar al gran desarrollo experimentado en los últimos sesenta años.

Posteriormente se compara el enfoque de espacio de estado (EE) para el análisis de las series de tiempo con el enfoque ARIMA de Box-Jenkins (BJ). Luego de definir lo que se entiende y cómo trabajan los enfoques de EE y BJ, nos retrotraemos a los orígenes históricos del análisis moderno y aplicado de las series de tiempo conocido como suavizado exponencial que nació en los ‘50, y tratamos de mostrar que ambos sistemas pueden ser considerados como que evolucionaron naturalmente desde esos orígenes. A continuación se realiza una comparación amplia de los méritos relativos de los dos sistemas, concluyendo en favor del enfoque de EE. Después de esto, se describen brevemente algunos trabajos recientes en donde se

aplica el enfoque de EE. Finalmente se realizan algunas consideraciones sobre el uso potencial de los métodos de EE en el trabajo sobre series de tiempo en las estadísticas oficiales y se dan algunas líneas nuevas de desarrollo del área.

Palabras clave

ARIMA; Box-Jenkins; espacio de estado; estadísticas oficiales; series de tiempo; suavizado exponencial.

ABSTRACT

Time-varying events receive the generic name of time series. It is extremely difficult to present a brief description of the field of time series. The difficulty lies in the fact that the subject is in itself complex, being a branch of statistics but with its methodology and its own peculiar vocabulary. The basic idea of a time series is very simple, consisting of the registration of any fluctuating quantity measured at different points in time. The common feature of all records that belong to the domain of "time series" is that they are influenced, even partially, by sources of random variation. So if we wish to explain the structure of the fluctuations in a time series we use what we call the study of time series. There are two aspects of the study of time series: analysis and modeling. The objective of the analysis is to summarize the properties of a series and highlight its salient features. The main reason for a time series model is to allow the prediction of their future values.

In this paper we present a brief introduction to the study of time series followed by a set of examples that occur in areas such as medicine, astronomy, economics, etc. Then we present a short history of the development of this part of statistics, beginning in 1664 with the works of Sir Isaac Newton, until the great development experienced over the last sixty years.

Later we compares the state-space approach (SS) for the analysis of time series with the approach of Box-Jenkins ARIMA (BJ). After defining what is meant and how they work both SS and BJ approaches, we go back to the historical origins of modern and applied time series known as exponential smoothing which was born in the 50s, and try to show that both systems can be considered that they naturally evolved from those origins. Following that we give a comprehensive comparison of the relative merits of the two systems, concluding in favor of the SS approach. After that we briefly describe some recent work where the SS approach is applied. Finally, some considerations are made on the potential use of SS methods in the time series work of official statistics and provides some new lines for the development of the area.

Key words

ARIMA; Box-Jenkins; exponential smoothing; official statistics; state space; time series.

1. El estudio de las series de tiempo

Los sucesos variables en el tiempo reciben el nombre genérico de series de tiempo o series cronológicas. Es extremadamente difícil presentar una descripción breve del campo de las series de tiempo. La dificultad se basa en el hecho de que la materia es por sí misma muy compleja, siendo una rama de la estadística pero con su metodología y su propio vocabulario peculiar. De cualquier manera es un área profundamente fascinante que envuelve ideas de física, matemática y estadística, cubriendo un campo inmenso de aplicaciones, desde la neurofisiología hasta la astrofísica. Las dificultades que envuelven su entendimiento son debidas parcialmente al hecho que se desarrolla a partir de dos disciplinas muy distintas entre si, una la ingeniería de las comunicaciones y la otra la estadística. La materia está, por supuesto, repleta de terminología estadística y de la ingeniería, pero las técnicas estadísticas usadas son, en algunos casos, sustancialmente diferentes de aquellas usadas en la inferencia estadística clásica. De cualquier manera, una vez que las ideas fundamentales se han entendido, la materia alcanza una elegante unidad y ofrece infinitas posibilidades para el desarrollo del ingenio tanto de aspectos teórico como prácticos.

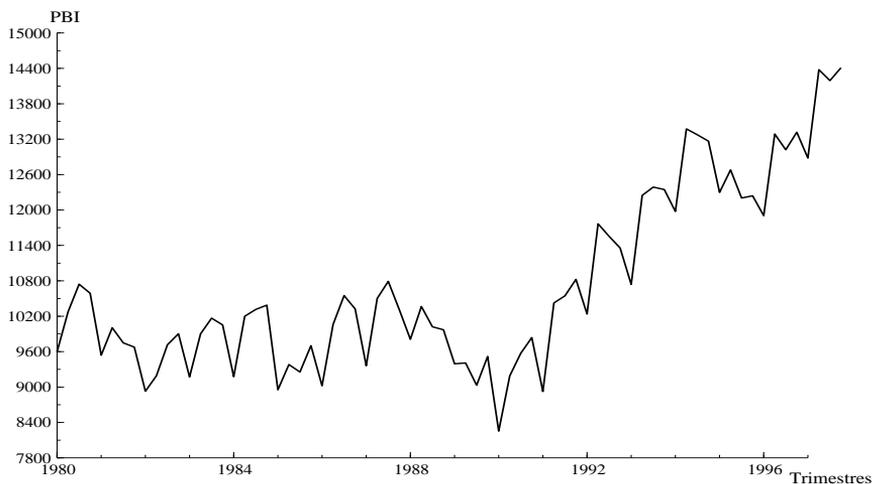
La idea básica de una *serie de tiempo* es muy simple, consiste en el registro de cualquier cantidad fluctuante medida en diferentes puntos del tiempo. Podemos tener, por ejemplo, un registro de un índice económico en un período de varios años, un registro de la presión sanguínea o el pulso de un paciente en un período de varios días, o un registro de la variación en intensidad de una señal de radio recibida en un período de varias horas. La característica común de todos los registros que pertenecen al dominio de las “series de tiempo” es que ellos están influenciados, aunque sea parcialmente, por fuentes de *variación aleatoria*. Entonces, si deseamos explicar la estructura de las fluctuaciones en una serie de tiempo debemos recurrir a lo que llamamos el *estudio de las series de tiempo*.

Hay dos aspectos en el estudio de las series de tiempo: el análisis y el modelado. El objetivo del análisis es resumir las propiedades de una serie y remarcar sus características salientes. Esto puede hacerse ya sea en el dominio del tiempo o en el dominio de las frecuencias. En el dominio del tiempo se concentra la atención en las relaciones entre las observaciones en puntos diferentes del tiempo, mientras que en el dominio de las frecuencias son los movimientos con diferentes periodicidades los que se estudian. Estas dos formas de análisis no son competitivas, muy por el contrario son complementarias. La misma información es procesada en diferentes formas, dando distintas visiones de la naturaleza de la serie de tiempo.

La principal razón para modelar una serie de tiempo es para permitir la predicción de sus valores futuros. La característica distintiva de un modelo de serie de tiempo, opuesto a un modelo econométrico, es que no se realiza ningún intento para formular una relación de comportamiento entre la serie de tiempo considerada y otras variables. Los movimientos de la serie son explicados solamente en términos de su propio pasado, o por su posición en relación al tiempo. Las predicciones se las realiza mediante extrapolación.

2. Ejemplos de series de tiempo

Figura 1. Producto Bruto Interno (PBI) de la Argentina, calculado trimestralmente a precios de mercado del año 1988. Período 1980–1997



Es frecuente cuantificar el comportamiento de fenómenos económicos mediante series de tiempo. Ejemplos de ellos son: el Producto Bruto Interno (PBI) de un país calculado en forma trimestral, el índice de precios al consumidor mes por mes, la cantidad de dinero en circulación en un cierto país mes por mes, etc. En la Figura 1 se presenta el PBI de la Argentina calculado trimestralmente a precios de mercado, en miles de pesos del año 1988. En la Figura 2 se presentan las ventas mensuales de una determinada empresa a precios constantes en cientos de miles de pesos, desde Enero de 1986 hasta Junio de 1992. Este tipo de series es muy frecuente en los negocios, siendo de gran utilidad para obtener un diagnóstico general de la empresa y su evolución en el período considerado. Es importante expresar los valores de las ventas a precios constantes (este procedimiento es conocido en economía como deflación) pues de esa manera se aísla el efecto puro del comportamiento empresarial del efecto causado por la devaluación monetaria. Box y Jenkins (1976) estudiaron la serie de pasajeros transportados por empresas aéreas internacionales. En la Figura 3 se presenta esta serie con valores mensuales expresados en miles de pasajeros transportados, desde Enero de 1949 hasta Diciembre de 1960. Debe notarse que, a diferencia de la serie anterior de ventas de una empresa, ésta es una serie de cantidades, o sea que no está afectada por las variaciones del valor del dinero, no necesitando, en consecuencia, ningún tipo de tratamiento previo como la deflación.

Muchas series de tiempo ocurren en las ciencias físicas, en especial en meteorología, ciencias marinas, geofísica, etc. Un ejemplo importante que se muestra

Figura 2. Ventas mensuales de una empresa a precios constantes. Período Enero de 1986 – Junio de 1992

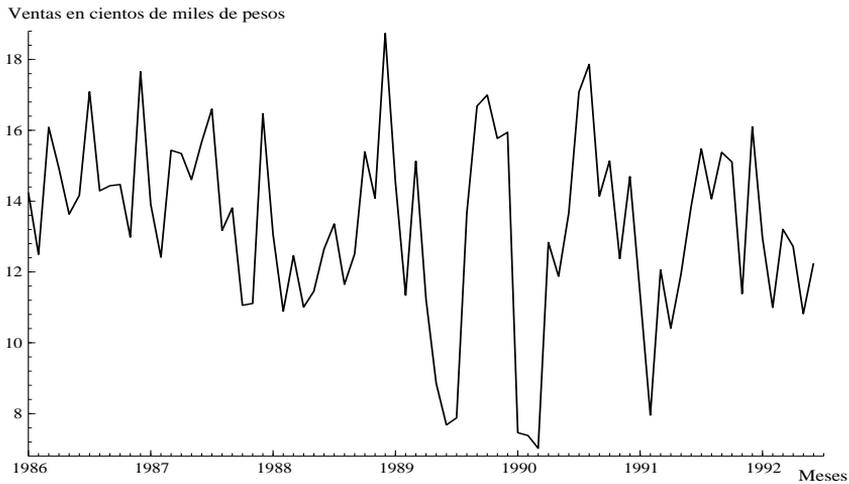
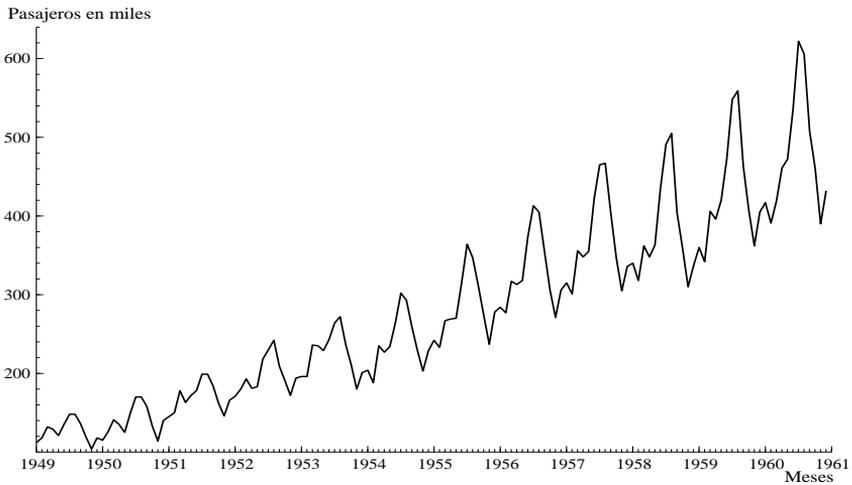


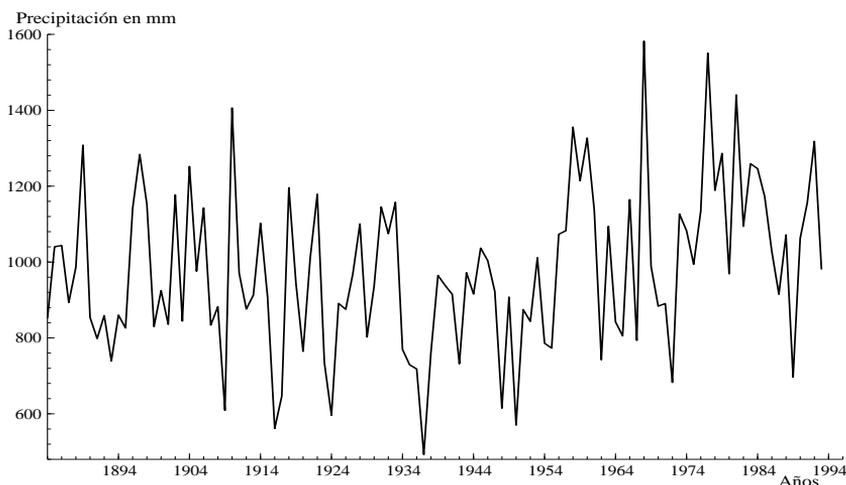
Figura 3. Pasajeros transportados mensualmente por empresas aéreas internacionales, en miles. Período Enero de 1949 – Diciembre de 1960



en la Figura 4 es la precipitación anual medida en milímetros totales caído en el año en San Miguel de Tucumán, Argentina, desde 1884 hasta 1993. Este tipo de serie fue estudiada por Pérez de del Negro (2000). En la Figura 5 se presenta el volumen anual del caudal del río Nilo en Aswan, medido en metros cúbicos, desde

1871 hasta 1970. Esta serie fue analizada por Cobb (1978), Carlstein (1988), Balke (1993) y Abril (1997).

Figura 4. Precipitación anual medida en milímetros totales caídos en San Miguel de Tucumán, Argentina, desde 1884 hasta 1993



En los estudios de población, ocurren numerosas series de tiempo. Ejemplos de ellas son los nacimientos y las defunciones mensuales en un determinado país, o región del mismo. Un uso importante de las técnicas de series de tiempo es cuando los demógrafos desean predecir los cambios en la población para períodos tan largos como diez o veinte años. Los electrocardiogramas, los electroencefalogramas y otros tipos de mediciones a través del tiempo, constituyen observaciones de series de tiempo en medicina. En los procesos de control, el problema es detectar cambios en el comportamiento de una actividad manufacturera midiendo una variable que muestre la calidad del proceso. Estas medidas pueden ser graficadas a través del tiempo como en la Figura 6. Allí, el valor objetivo es la línea horizontal que se encuentra a la altura del valor 10 del proceso y la otra línea, la fluctuante, es la que mide el valor de la variable que resume la calidad del proceso. Cuando las medidas ocurren muy alejadas de algún valor objetivo, se deben tomar acciones correctivas apropiadas de tal manera de controlar el proceso. Técnicas especiales para tratar estas series de tiempo han sido desarrolladas, y el lector puede consultar libros sobre control estadístico de calidad (por ejemplo, Wetherill, 1977).

3. Breve historia

El análisis científico de las series de tiempo tiene una larga historia. Puede pensarse que ha comenzado en 1664 cuando Sir Isaac Newton descompuso una señal luminosa (o serie de tiempo) en sus componentes a diferentes frecuencias haciendo

Figura 5. Volumen anual del caudal del río Nilo en Aswan, medido en metros cúbicos, desde 1871 hasta 1970

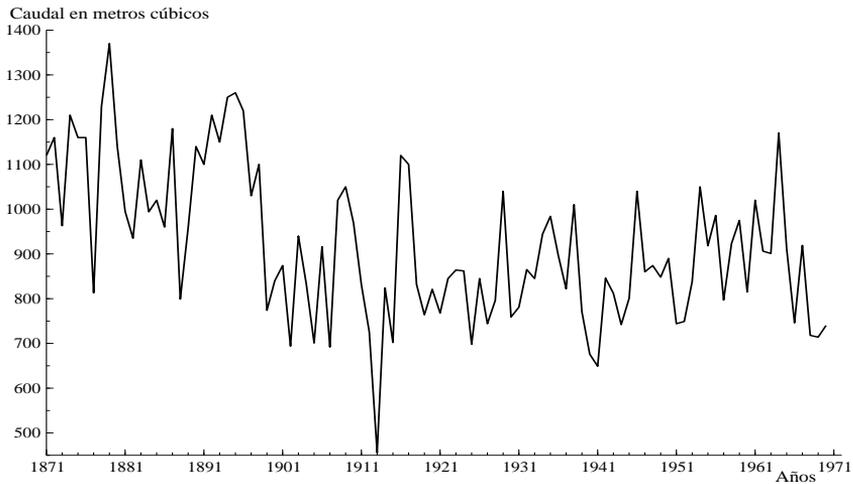
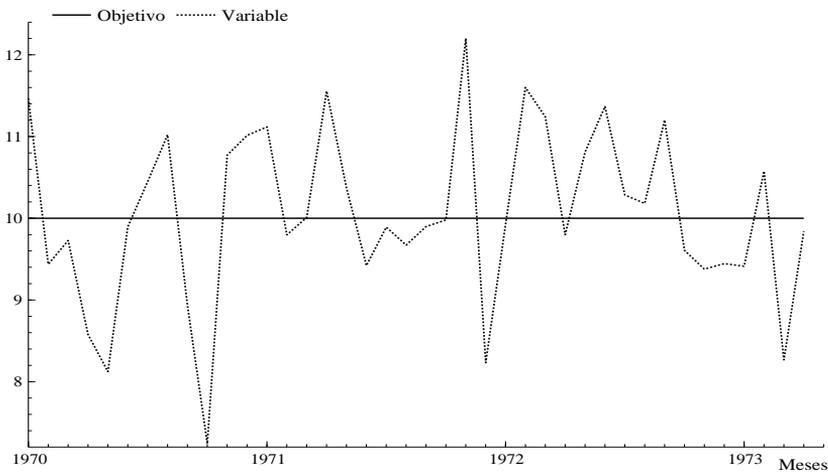


Figura 6. Proceso de control. Valor objetivo: línea llena, valor de la variable: línea de puntos



pasar la señal por un prisma de vidrio. Una imagen multicolor fue lanzada sobre una pared opuesta. El llamó a esto el espectro, y corresponde a la operación de filtrado. Newton no llevó a cabo un análisis cuantitativo de su serie de tiempo. De cualquier manera, en 1800 Sir W. Herschel lo hizo usando termómetros. Midió

la energía promedio en varias bandas de frecuencia del espectro de la luz solar emplazando termómetros a lo largo de ese espectro. Los fundamentos matemáticos del análisis de las series de tiempo se empezaron a dar a mediados del siglo XIX cuando Gouy representó a la luz blanca como una serie de Fourier. Posteriormente Lord Rayleigh reemplazó la serie por una integral. En 1881 S. P. Langley refinó el experimento de Herschel considerablemente al medir la energía de la luz con un bolómetro espectral (un aparato que él inventó y que hace uso de la corriente eléctrica generada en un cable por radiación incidental).

En 1872 Lord Kelvin construyó un analizador armónico y un sintetizador armónico para ser usados en el análisis y predicción de las series de la altura de las mareas en un lugar y tiempo determinado. Este aparato era mecánico y estaba basado en poleas. Durante el mismo período de tiempo otros investigadores (por ejemplo G. Stokes) estaban realizando análisis de Fourier numérico usando métodos de computación. En particular, en 1891 S. C. Chandler realizó un análisis de la variación de latitud con tiempo. Su análisis lo condujo a sugerir que el movimiento del polo de rotación de la tierra era compuesto, conteniendo componentes de período 12 y de aproximadamente 14 meses.

Un avance sustancial en el análisis de las series de tiempo correspondientes a las señales luminosas ocurrió en 1891 cuando A. A. Michelson inventó el interferómetro. Ese aparato permitía la medición del valor promedio

$$\frac{\int_0^n [y(t) + y(t + u)]^2 dt}{n}$$

para valores grandes de n y valores determinados del rezago (lag) u , donde n es la cantidad de observaciones disponibles de la señal luminosa o serie de tiempo bajo estudio $y(t)$, la cual se la considera como una función continua del tiempo t . Esto permitía la estimación de la función de autocovarianzas de la señal. En 1898 Michelson y Stratton describieron un analizador armónico basado en resortes, y lo usaron para obtener la transformada de Fourier de la fórmula dada anteriormente. Esta transformada de Fourier proveía una estimación del espectro de la señal. Michelson visualizaba la señal como una suma de cosenos y vio al espectro estimado como un estadístico descriptivo de la fuente emisora de luz.

En 1894 M. I. Pupin inventó el filtro de onda eléctrica. Este aparato ampliaba considerablemente el dominio de las frecuencias sobre las cuales la serie de tiempo podía ser analizada. La potencia de una señal eléctrica podía ser entonces medida en un intervalo de bandas de frecuencias.

En una serie de trabajos escritos durante los años 1894-1898, A. Schuster propuso el periodograma y discutió su utilización. Su motivación fue investigar las periodicidades escondidas. En los años sucesivos, el periodograma y sus equivalentes fueron computados por muchos investigadores para una variedad de fenómenos. El análisis del periodograma es una técnica de mucho valor cuando se la aplica a una serie con una estructura apropiada. En efecto, el mismo está motivado

por las ideas clásicas de la matemática aplicada que consisten esencialmente en suponer que la serie es la suma de un componente periódico determinístico más un término de error que representa la contribución estocástica.

En 1927 Yule estudió una serie de tiempo que, desde entonces, se convirtió en un clásico de la literatura. Ella es la serie de manchas solares de Wolfer. Esta serie oscila de una forma más o menos regular, pero Yule notó que las alturas de los máximos y las distancias entre sucesivos máximos y mínimos varían en una forma muy irregular. Una mejor inspección de la serie reveló que los ciclos tienen una forma asimétrica (esto se refiere a que el tiempo que toma en ascender desde un mínimo a un máximo es diferente al tiempo que toma en descender desde un máximo a un mínimo). Por lo tanto, observó que sería poco realista si tratara de ajustar a esta serie un modelo consistente en uno o varios términos estrictamente periódicos y simétricos, tales como funciones seno y coseno. Yule (1927) propuso, en consecuencia, un tipo de modelo completamente diferente basado en una analogía con sistemas mecánicos oscilantes. Él consideró un péndulo moviéndose en un medio con resistencia. Observó que si se le aplica un solo impulso, la resistencia llevará eventualmente al sistema al reposo, pero si se le aplica una sucesión de impulsos, entonces el movimiento se mantendrá. Si esos impulsos forman una sucesión estrictamente periódica, el movimiento resultante será estrictamente periódico. El ingenioso enfoque de Yule fue permitir que esos impulsos ocurran en instantes aleatorios de tiempo. El sistema continuaría oscilando, pero exhibiendo amplitudes e intervalos irregulares, muy parecidos a los observados en la serie de manchas solares. Este sistema pudo ser descrito matemáticamente, con un buen grado de aproximación, como una ecuación diferencial estocástica de segundo orden. El análogo para tiempo discreto de esta ecuación diferencial estocástica de segundo orden es la ecuación en diferencias estocástica de segundo orden, lo cual hoy en día es conocido como *modelo autorregresivo de segundo orden* (denotado como AR(2)).

El modelo AR(2) todavía constituye uno de los modelos básicos usados para describir series de tiempo “pseudo” periódicas. Una vez que estas ideas esenciales fueron dadas a conocer a la comunidad científica, no hubo dificultades en extenderlas a formas más complicadas, definiendo, en consecuencia, los *modelos autorregresivos de orden p* como una ecuación en diferencias estocástica de orden p (denotados como AR(p)). Así, sea $\{y_t\}$ la serie de tiempo bajo estudio, el modelo AR(p) se representa como

$$y_t + \alpha_1 y_{t-1} + \dots + \alpha_k y_{t-p} = \varepsilon_t, \quad t = 0, \pm 1, \pm 2, \dots, \quad (1)$$

donde $\{\varepsilon_t\}$ es un proceso tal que $E(\varepsilon_t) = 0$, $E(\varepsilon_t^2) = \text{var}(\varepsilon_t) = \sigma^2$ y $E(\varepsilon_t \varepsilon_s) = 0$ si $t \neq s$. Usualmente se denomina a $\{\varepsilon_t\}$ un proceso ortogonal o ruido blanco.

Una clase alternativa, y en muchas formas complementaria, de modelos fueron los llamados *modelos de promedios móviles de orden q* (denotados como MA(q)), cuya representación es

$$y_t = \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \dots + \beta_q \varepsilon_{t-q}, \quad t = 0, \pm 1, \pm 2, \dots, \quad (2)$$

donde $\{\varepsilon_t\}$ es un proceso ortogonal.

Los modelos de promedios móviles tienen propiedades bastantes diferentes a las de los modelos autorregresivos, pero todavía pueden mostrarnos una forma de comportamiento “pseudo” periódico. Un estudio sistemático de las propiedades estadísticas de los modelos de promedios móviles fue realizado por Wold (1949).

La estimación de los parámetros de los modelos AR, MA y ARMA es un problema que puede resultar bastante difícil, pero afortunadamente hay un número razonable de paquetes estándares de computación los cuales efectúan la estimación por máxima verosimilitud, bajo el supuesto de normalidad, usando algoritmos numéricos. Existen varios criterios para la determinación del orden de los modelos, los cuales nos permiten seleccionar los valores más adecuados para el orden de los modelos AR o MA a ajustar. Los más conocidos de estos criterios son el criterio AIC de Akaike (1974) y el criterio CAT de Parzen (1974). Los modelos AR parecen, a primera vista, modelos ordinarios de regresión, pero el hecho de que la variable bajo estudio juega el doble papel de variable dependiente y variable independiente, complica las propiedades de los estimadores por mínimos cuadrados de los parámetros. Un estudio pionero de la estimación de los parámetros de los modelos AR fue realizado por Mann y Wald (1943).

La idea subyacente del modelo de Yule era generar una serie de tiempo que, aunque no sea estrictamente periódica, exhibiera un tipo de comportamiento “estable” en algún sentido. Ese comportamiento estable se lo describe con más precisión introduciendo la idea de *estacionariedad*. Existe un importante teorema, el Teorema de Wold, que dice que cualquier serie de tiempo “bien comportada” y estacionaria en cierto sentido, que no contiene componentes estrictamente periódicos, puede ser representada ya sea como una autorregresión o un promedio móvil (ambos posiblemente de orden infinito). Por lo tanto, se puede modelar casi cualquier serie de ese tipo ya sea mediante una AR o un MA de alto orden con el grado de exactitud que uno requiera.

Cuando una serie de tiempo representa algunas características físicas puede ser interesante estudiar sus propiedades directamente, en lugar de construir un modelo para ella. Una técnica importante usada en el análisis de un proceso físico es descomponer el mismo en sus componentes de diferentes frecuencias. Esta operación se la describe como *análisis espectral*. A partir de 1930, y en base a una modificación de los estudios matemáticos de Wiener sobre al análisis armónico generalizado, se logró encontrar la teoría adecuada que proveía exactamente el tipo correcto de descomposición en frecuencias para una serie de tiempo estacionaria. Con ello se definió la *función de densidad espectral* también conocida simplemente como densidad espectral o espectro de la serie de tiempo. La evaluación de esta función de densidad espectral directamente de la definición constituye un problema formidable, pero existe un resultado muy importante, conocido como el Teorema de Wiener-Kintchine, que provee una relación simple entre la función de densidad espectral y la función de *autocovarianzas* de la serie de tiempo.

Durante la década de 1950 y el inicio de la de 1960, el problema de estimar el espectro a partir de los datos de una serie de tiempo atrajo gran atención. Esto mostró ser algo más dificultoso de lo que se anticipaba originariamente. En efecto, el periodograma como estimador obvio del espectro, resultó no ser lo suficientemente satisfactorio como se lo pensaba. Esta situación condujo a un número de autores a sugerir otros estimadores basados en un suavizado del periodograma. Estos estimadores se los logra mediante un promedio localmente ponderado del periodograma en un vecindario de la frecuencia considerada y se los suele denominar estimadores de ventana. Los trabajos pioneros fueron los de Daniell (1946), Tukey (1949), Bartlett (1950), Grenander y Roseblatt (1957) y Parzen (1957).

Posteriormente, diferentes métodos de estimación espectral han sido propuestos basados en el ajuste de un modelo autorregresivo finito a los datos, y luego usando la forma teórica de la densidad espectral del modelo ajustado (ver Akaike 1969; Parzen 1974). Este método, conocido como estimación autorregresiva del espectro, puede producir resultados superiores cuando los datos conforman bien a un adecuado modelo autorregresivo, pero los estimadores de ventana tienen la gran ventaja de ser estimadores no-paramétricos, o sea que no requieren que los datos satisfagan un tipo particular de modelo.

La estimación espectral por el método de máxima entropía fue introducido por Burg (1967, 1972), y a pesar que la motivación de este método es totalmente diferente a la de la estimación autorregresiva del espectro, resulta que los dos enfoques conducen a idénticas formas de estimadores. En efecto, se puede mostrar que la función que maximiza la entropía sujeta a la condición que sus primeros k coeficientes de Fourier sean exactamente iguales a las k autocovarianzas muestrales es la misma que la que se obtiene al ajustar por mínimos cuadrados un modelo autorregresivo de orden k y luego usar las estimaciones para realizar una estimación autorregresiva del espectro (ver Priestley, 1981, p. 604).

El método de máxima entropía fue propuesto primeramente por Burg en 1967 pero el primer trabajo publicado en donde se lo describía fue de Lacoss (1971). En ese trabajo, Lacoss describe otra forma “no-cuadrática” de estimación espectral llamada estimación por máxima verosimilitud, desarrollada originalmente por Capon (1969). La idea básica de la estimación por máxima verosimilitud es construir un filtro que deja pasar un componente de una frecuencia particular de una forma no distorsionada, rechazando los otros componentes, haciendo todo esto de una forma “óptima”. Burg (1972) estableció una relación entre la estimación por máxima verosimilitud y la estimación por máxima entropía mostrando que el recíproco del estimador por máxima entropía basado en las primeras autocovarianzas es igual al promedio del recíproco del estimador por máxima verosimilitud basado en las mismas primeras autocovarianzas

Hemos observado que cualquier serie de tiempo “razonable” puede ser descripta por un modelo AR o MA (posiblemente de orden infinito). Ahora bien, si tratamos de ajustar un modelo AR a datos generados por un modelo MA, requeriremos

un modelo de orden extremadamente alto para obtener un ajuste adecuado, y viceversa. Entonces existe una ventaja obvia en combinar los dos tipos de modelos AR y MA en un *modelo mixto autorregresivo de promedios móviles* (denotado como ARMA), cuya representación está dada por

$$y_t + \alpha_1 y_{t-1} + \cdots + \alpha_p y_{t-p} = \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \cdots + \beta_q \varepsilon_{t-q}, \quad (3)$$

para $t = 0, \pm 1, \dots$, donde $\{\varepsilon_t\}$ es un proceso ortogonal con varianza σ^2 . En ese caso decimos que $\{y_t\}$ es generado por el modelo mixto, autorregresivo y de promedio móvil de ordenes p, q , al cual denotamos como $\text{ARMA}(p, q)$. Al modelo (3) se lo conoce también como proceso ARMA(p, q). Modelos de este tipo fueron vigorosamente propuestos por Box y Jenkins (1976) como la clase “estándar” de modelos de series de tiempo. Es efectivamente cierto que el esquema ARMA puede ser usado para modelar una gran variedad de estructuras de series de tiempo. La ventaja de los modelos ARMA es que en el ajuste se requerirán menos parámetros que en el ajuste de un modelo puro AR o MA.

Box y Jenkins (1976) consideraron una extensión de los modelos ARMA para ocuparse de ciertos tipos especiales de series no-estacionarias. Ellos consideraron una serie de tiempo que es no-estacionaria, pero que puede llegar a ser estacionaria si se la diferencia un número suficiente de veces. Esto condujo a los llamados modelos ARIMA.

Box y Jenkins presentaron técnicas claramente simples para identificar los ordenes adecuados de los modelos, y para la evaluación numérica de los estimadores de máxima verosimilitud aproximada de los parámetros del modelo. Es cierto que, cuando se los aplica correctamente, las clases ARMA y ARIMA pueden proveer descripciones muy útiles de una gran variedad de datos de series de tiempo, explicando porqué el “enfoque de Box-Jenkins” (como se lo conoce) ha sido usado extensamente por una gran variedad de practicantes del análisis de series de tiempo en una gran diversidad de campos. Debe notarse, no obstante, que el paso denominado “identificación del modelo” en el análisis (o sea, la selección de una forma apropiada de modelos ARMA o ARIMA) requiere cuidados y experiencia considerable. El éxito de este enfoque depende mayormente en la habilidad del analista para seleccionar un modelo apropiado que en los aspectos más técnicos de la estimación de los parámetros por máxima verosimilitud. Más adelante daremos algunos detalles del “enfoque de Box-Jenkins”.

Uno de los más importantes problemas en el análisis de las series de tiempo es el de *predicción* de valores futuros de la serie, dado algunos datos sobre los valores pasados. Cuando la serie representa alguna variable económica, la importancia práctica de la predicción es inmediatamente evidente. No obstante, los primeros estudios sistemáticos de este problema fueron desarrollados por Wiener (1949) y Kolmogorov (1941) en contextos completamente diferentes. Wiener se interesó en el problema de predecir la trayectoria del vuelo de un avión. Supongamos que tenemos datos que se extienden en la historia pasada y remota de la serie, o sea que tenemos todas las observaciones hasta el momento actual, y deseamos predecir el valor de algún punto futuro. Podemos considerar una predicción basada en

una combinación lineal de todos los valores pasados. El problema así presentado implica lograr una predicción “óptima” en algún sentido. Suponiendo que las propiedades estadísticas de la serie son totalmente conocidas, esto se transforma en un problema puramente matemático, y su solución fue dada en forma independiente por Wiener y Kolmogorov usando el error medio cuadrático como el criterio de optimalidad. El método de Wiener está basado en un enfoque en el *dominio de las frecuencias* e involucra técnicas matemáticas complicadas conocidas como “factorización espectral”. Este método es muy difícil de aplicar a menos que la serie tenga una densidad espectral muy simple. Por otra parte el enfoque de Wiener se ajusta más a las series de tiempo de la ingeniería que a las de la economía.

Box y Jenkins (1976) propusieron un método mucho más simple basado en ajustar un modelo ARMA, y luego calcular las predicciones directamente del modelo ajustado. Las predicciones son simplemente calculadas usando un algoritmo recursivo. El método es extremadamente fácil de aplicar, pero requiere, por supuesto, que la serie sea bien ajustada por un modelo ARMA. El método de Wiener-Kolmogorov no requiere que la serie sea conformable con un modelo con un número finito de parámetros, pero es mucho más difícil de aplicar a los datos.

Wiener (1949) extendió su método para tratar el problema asociado del *filtrado* donde no se puede observar directamente la serie sino que, en su lugar, observamos una serie que consiste en la serie de interés más un ruido que corrompe a la primera. El problema aquí es construir predicciones de la serie de interés basándose en la observación de la serie corrompida. La solución de Wiener al problema de filtrado envuelve esencialmente las mismas herramientas matemáticas usadas en su solución al problema de predicción.

Una nueva y poderosa solución al problema de filtrado fue ideado por Kalman (1960) y Kalman y Bucy (1961), usando la llamada *representación de espacio de estado* de una serie de tiempo. Esto provee una descripción muy compacta del modelo y está basado en el resultado conocido que dice que cualquier ecuación en diferencias (o diferencial) lineal de orden finito puede ser escrita como una ecuación vectorial en diferencias (o diferencial) lineal de primer orden. La ventaja de esta última representación es que involucra solamente dependencia de un paso, o sea que posee la propiedad de Markov, lo cual conduce a un algoritmo simple y elegante para calcular las predicciones de valores futuros de la serie conocido como el *algoritmo del filtro y suavizador de Kalman*.

Una gran parte del análisis convencional de las series de tiempo está basado en el supuesto que la serie es estacionaria (al menos hasta el segundo orden). En la práctica este supuesto es frecuentemente muy incierto, y no obstante cierto tipo de series no-estacionarias pueden ser reducidas a una forma estacionaria mediante transformaciones tales como la diferenciación, hay muchas series cuyas propiedades estadísticas están “continuamente” cambiando en el tiempo, no pudiendo por lo tanto, ser reducidas a estacionarias mediante este tipo de transformaciones. El problema de analizar esas series es muy difícil; si las propiedades

estadísticas cambian muy rápidamente en el tiempo, entonces claramente será imposible desarrollar cualquier análisis estadístico de utilidad. No obstante, si sus propiedades estadísticas cambian “suavemente” en el tiempo, entonces podemos considerar la posibilidad de ajustar modelos “locales” a conjuntos “cortos” de los datos, permitiendo que los parámetros cambien sus valores a medida que nos movemos de un conjunto a otro de datos. La idea conduce, por un lado, a un modelo global del tipo ARMA en el cual los parámetros dependen del tiempo. Modelos de esta forma han sido considerados por varios autores, particularmente Subba Rao (1970) ha estudiado el problema de estimar el conjunto de parámetros asociados dependientes del tiempo.

Por otro lado, esa idea conduce a la representación de espacio de estado en donde las matrices de sistema que contienen los parámetros dependen del tiempo. Luego, mediante una aplicación adecuada del algoritmo del filtro y suavizador de Kalman, se obtienen las estimaciones y las predicciones de la serie. Un trabajo pionero dentro de esta área es el de Harvey y Durbin (1986). Un estudio profundo de este enfoque es el trabajo enciclopédico de Harvey (1989). Un tratamiento moderno del tema puede verse en Abril (1997, 1999).

El análisis espectral de series no-estacionarias contiene algunas dificultades fundamentales. Cuando la serie es no-estacionaria, la “representación espectral” de la misma no es posible de expresar como “sumas” de funciones senos y cosenos. Esto significa que todo el concepto de frecuencia requiere una nueva y cuidadosa evaluación si se desea preservar la interpretación física del espectro. Priestley (1965) mostró que es posible definir al espectro para una clase general de series no-estacionarias. El punto crucial es que para esas series no-estacionarias el espectro es dependiente del tiempo y describe las propiedades locales en un vecindario de cada punto del tiempo. Es posible en estos casos estimar al espectro y construir una teoría de predicción lineal y filtrado para series no-estacionarias, la cual sería muy próxima y paralela a la teoría de Wiener-Kolmogorov para series estacionarias.

Desarrollos recientes en el análisis de las series de tiempo tienen que ver con el estudio de modelos no-lineales. Los esquemas AR, MA y ARMA son todos modelos lineales, en el sentido que están basados en ecuaciones lineales en diferencias cuyas soluciones se expresan como una combinación lineal de valores presentes y pasados de una serie de ruido blanco. Mientras que esos modelos pueden proveer descripciones acertadas y razonables para una gran variedad de datos, la noción de linealidad, como la de estacionariedad, es a lo sumo solamente una aproximación al “mundo real”. La construcción de tipos generales de modelos no-lineales de series de tiempo presenta problemas formidables, pero el estudio de ciertas formas especiales de modelos no-lineales ha mostrado que posee algunas propiedades estructurales fascinantes que no pueden ser reproducidas por los modelos lineales. Algunos de estos casos especiales son: a) los modelos bilineales estudiados e investigados principalmente por Granger y Anderson (1978), Subba Rao (1981) y Subba Rao y Gabr (1984), b) los modelos autorregresivos de umbral (“threshold autoregressive models”) propuestos por Tong y Lim (1980), c) los modelos

autorregresivos exponenciales introducidos por Haggan y Ozaki (1979) y d) los “modelos generales de espacios dependientes” introducidos por Priestley (1980).

Un precursor del estudio de las series de tiempo es James Durbin quien viene trabajando desde 1950 dentro del tema, habiendo producido grandes e impactantes aportes científicos. Detalles sobre las contribuciones de Durbin al estudio de las series de tiempo y la econometría pueden verse en Abril (2000).

4. Algunos modelos simples

Como ya se dijo, una serie de tiempo es un conjunto de observaciones ordenadas en el tiempo. El modelo básico para representar una serie de tiempo es el modelo aditivo

$$y_t = \mu_t + \gamma_t + \vartheta_t + \varepsilon_t. \quad (4)$$

Donde, μ_t es un componente que varía suavemente llamado la *tendencia*, γ_t es un componente periódico de período fijo llamado la *estacionalidad*, ϑ_t es un componente periódico con frecuencia sustancialmente menor que la estacionalidad llamado el *ciclo* y ε_t es un componente irregular llamado el *error*. En muchas aplicaciones, particularmente en economía, los componentes combinan multiplicativamente, dando

$$y_t = \mu_t \gamma_t \vartheta_t \varepsilon_t. \quad (5)$$

No obstante, tomando logaritmos, el modelo (5) reduce al modelo (4). Por lo tanto se puede usar el modelo (4) a todos los efectos prácticos.

La tendencia es un componente que cambia suavemente en el tiempo. En muchos casos se la trató como una función determinística del tiempo, pero, con mayor generalidad se puede pensar que es un proceso estocástico que varía en forma suave con el tiempo.

Muchas series están influenciadas por factores estacionales variables tales como el clima, el efecto de los cuales puede ser modelado por medio de un componente periódico con período fijo y conocido. La estacionalidad también se la trató como una función determinística del tiempo, pero con argumentos similares a los dados para la tendencia, se puede pensar que es aleatoria. Usualmente, la parte determinística de la estacionalidad puede ser expresada de dos formas: una como un conjunto de variables ficticias (“dummy”) tales que suman cero durante el período estacional, o bien como una combinación de senos y cosenos con frecuencias acordes con el período estacional.

El ciclo es un componente periódico de baja frecuencia. En realidad, es un componente cuya frecuencia es superior a la de la tendencia pero sustancialmente inferior a la de la estacionalidad. Es altamente conveniente pensar que es aleatorio, siendo su parte determinística una combinación de senos y cosenos con frecuencias acordes con el período cíclico.

En cuanto al componente irregular, el mismo ha sido tratado siempre como una variable aleatoria con ciertas características adecuadas al fenómeno que da origen a la serie de tiempo bajo estudio.

Un adecuado y moderno estudio de modelos del tipo presentado en la fórmula (4) puede verse en Abril (1997 y 1999).

5. El enfoque de Box-Jenkins

Esta sección está basada en el enfoque propuesto por Box y Jenkins (1970 y 1976) para el tratamiento de diferentes series de tiempo y que fuera muy popular en los últimos años. A fin de facilitar la lectura de la obra original de Box y Jenkins (BJ) vamos a tratar de acercarnos lo más que se pueda a la notación del libro de estos autores. En todos los casos necesarios se efectuarán las aclaraciones pertinentes a fin de compatibilizar nuestra notación con la de BJ.

El sistema de BJ siguen básicamente las siguientes secuencia de pasos: (a) identificación, (b) estimación, (c) control de diagnóstico, (d) predicción.

La *identificación* consiste en determinar el posible modelo generador de la serie observada. Una vez determinado ese modelo, se procede a la *estimación* de sus parámetros. Luego, en el *control de diagnóstico*, determinamos, mediante una serie de test si el modelo identificado y estimado es el adecuado para explicar la serie observada. Si los test de este *control de diagnóstico* nos conducen a aceptar el modelo, procedemos a continuación a la *predicción* de los valores futuros de la serie. Por el contrario, si el *control de diagnóstico* nos lleva a rechazar el modelo, debemos reiniciar la secuencia con la *identificación* de otro posible modelo, y así seguir todo el proceso hasta encontrar el adecuado a fin de realizar la *predicción*.

Por razones didácticas y de facilidad de lectura, presentaremos primero todo el proceso de predicción, para luego seguir en orden con los otros pasos a partir de la identificación.

5.1 Predicción de modelos estacionarios

Supongamos que tenemos las observaciones x_t , donde $E(x_t) = \mu$. Sea $y_t = x_t - \mu$. El modelo básico de BJ es el ARMA(p, q) para la serie y_t , o sea

$$y_t - \varphi_1 y_{t-1} - \cdots - \varphi_p y_{t-p} = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}, \quad (6)$$

donde, si se compara con (3) vemos que $\alpha_j = -\varphi_j$ y $\beta_k = -\theta_k$ para todo j y k , y los ε_t s forman un proceso ortogonal (o ruido blanco). La fórmula (6) puede escribirse como

$$\varphi(B)y_t = \theta(B)\varepsilon_t,$$

con

$$\begin{aligned}\varphi(B) &= 1 - \varphi_1 B - \dots - \varphi_p B^p, \\ \theta(B) &= 1 - \theta_1 B - \dots - \theta_q B^q,\end{aligned}$$

B es el operador de rezago, o sea que $B^r y_t = y_{t-r}$. Suponemos que $\varphi(B)$ y $\theta(B)$ tienen todas sus raíces fuera del círculo unitario, lo que implica que el modelo es estacionario y tiene representación MA(∞) y también AR(∞), esto es,

$$y_t = \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \dots = \psi(B) \varepsilon_t,$$

$$\begin{aligned}y_t - \pi_1 y_{t-1} - \pi_2 y_{t-2} - \dots &= \varepsilon_t, \\ \pi(B) y_t &= \varepsilon_t.\end{aligned}$$

Como se aprecia, todo el enfoque está basado en el hecho de que la serie observada sea estacionaria. Las predicciones de $y_{t+\ell}$, para $\ell = 1, 2, \dots$ dado $Y_t = \{y_t, y_{t-1}, y_{t-2}, \dots\}$ se las realiza mediante las recursiones

$$\begin{aligned}\widehat{y}_t(1) - \varphi_1 y_t - \dots - \varphi_p y_{t-p+1} &= -\theta_1 (y_t - \widehat{y}_{t-1}(1)) - \theta_2 (y_{t-1} - \widehat{y}_{t-2}(1)) \\ &\quad - \dots - \theta_q (y_{t-q+1} - \widehat{y}_{t-q}(1)),\end{aligned}\tag{7}$$

$$\begin{aligned}\widehat{y}_t(2) - \varphi_1 \widehat{y}_t(1) - \varphi_2 y_t - \dots - \varphi_p y_{t-p+2} &= -\theta_2 (y_t - \widehat{y}_{t-1}(1)) \\ &\quad - \dots - \theta_q (y_{t-q+2} - \widehat{y}_{t-q+1}(1)),\end{aligned}\tag{8}$$

$$\begin{aligned}\widehat{y}_t(3) - \varphi_1 \widehat{y}_t(2) - \varphi_2 \widehat{y}_t(1) - \dots - \varphi_p y_{t-p+3} &= -\theta_3 (y_t - \widehat{y}_{t-1}(1)) - \dots \\ &\quad - \theta_q (y_{t-q+3} - \widehat{y}_{t-q+2}(1)),\end{aligned}\tag{9}$$

y así sucesivamente. Este conjunto de ecuaciones nos permite calcular $\widehat{y}_t(1), \widehat{y}_t(2), \dots, \widehat{y}_t(\ell)$ recursivamente. El conjunto de valores $\widehat{y}_t(1), \widehat{y}_t(2), \dots, \widehat{y}_t(\ell)$ se denomina *perfil de predicción*. Luego, obtenemos las predicciones de x_t simplemente usando la relación $y_t = x_t - \mu$.

5.2 Predicción de modelos no estacionarios (y no estacionales)

Sea el operador $\Delta = 1 - B$, entonces $\Delta f(t) = f(t) - f(t-1)$, $\Delta^2 f(t) = (1 - B)^2 f(t) = f(t) - 2f(t-1) + f(t-2)$, y así sucesivamente.

Supongamos que la serie x_t contiene tendencia, la cual se la puede aproximar adecuadamente en forma local por un polinomio de orden d . Podemos eliminarla tomado diferencias de orden d de las observaciones x_t , esto es

$$y_t = \Delta^d x_t.$$

Luego suponemos que y_t puede ser representado por un modelo ARMA(p, q) estacionario de la forma

$$\varphi(B)y_t = \theta(B)\varepsilon_t.$$

Esto implica que el modelo para la serie no estacionaria x_t es

$$\begin{aligned}\varphi(B)\Delta^d x_t &= \theta(B)\varepsilon_t \\ \varphi(B)(1-B)^d x_t &= \theta(B)\varepsilon_t.\end{aligned}$$

Debemos observar que el operador autorregresivo $\varphi(B)(1-B)^d$ tiene ahora algunas de sus raíces sobre el círculo unitario, lo que implica que la serie x_t es no estacionaria.

BJ llamaron a esto un modelo ARIMA(p, d, q) o modelo autorregresivo integrado de promedio móvil de orden p, d, q . “Integrado” viene de suma, por ejemplo

$$y_t = \Delta x_t = x_t - x_{t-1},$$

entonces

$$\begin{aligned}x_t &= y_t + x_{t-1} \\ &= y_t + y_{t-1} + x_{t-2} \\ &\quad \vdots \\ &= y_t + y_{t-1} + y_{t-2} + \cdots,\end{aligned}$$

lo que implica que x_t es igual a la suma o integración de variables aleatorias estacionarias.

Las predicciones de BJ son equivalentes a nuestro tratamiento de predicción del modelo ARMA presentado en (7), (8), (9) y demás necesarias. Así, predecir $x_{t+\ell}$ mediante $\hat{x}_t(\ell) = E(x_{t+\ell} | X_t)$, donde X_t es la historia x_t, x_{t-1}, \dots , se basa en

$$\varphi(B)\Delta^d x_t = \theta(B)\varepsilon_t, \tag{10}$$

lo que es equivalente a

$$\varphi^*(B)x_t = \theta(B)\varepsilon_t, \tag{11}$$

donde $\varphi^*(B) = \varphi(B)\Delta^d$. En el momento $t + j$ tenemos

$$x_{t+j} - \varphi_1^* x_{t+j-1} - \cdots - \varphi_p^* x_{t+j-p} = \varepsilon_{t+j} - \theta_1 \varepsilon_{t+j-1} - \cdots - \theta_q \varepsilon_{t+j-q}, \tag{12}$$

donde $p^* = p + d$. Tomando esperanza condicional, $E(\cdot | X_t)$, en (12) obtenemos

$$\hat{x}_{t+j} - \varphi_1^* \hat{x}_{t+j-1} - \cdots - \varphi_p^* \hat{x}_{t+j-p} = \hat{\varepsilon}_{t+j} - \theta_1 \hat{\varepsilon}_{t+j-1} - \cdots - \theta_q \hat{\varepsilon}_{t+j-q}, \tag{13}$$

para $j = 1, 2, \dots, \ell$, y donde

$$\widehat{x}_{t+i} = \begin{cases} x_{t+i}, & i \leq 0, \\ \widehat{x}_t(i), & i > 0, \end{cases} \quad (14)$$

$$\widehat{\varepsilon}_{t+i} = \begin{cases} x_{t+i} - \widehat{x}_{t+i-1}(1), & i \leq 0, \\ 0, & i > 0. \end{cases} \quad (15)$$

O sea que son las mismas recursiones que para los modelos estacionarios, presentadas en (7), (8), (9) y demás necesarias. Para $\ell > q$ la recursión es

$$\widehat{x}_{t+\ell} - \varphi_1^* \widehat{x}_{t+\ell-1} - \dots - \varphi_{p^*}^* \widehat{x}_{t+\ell-p^*} = 0, \quad (16)$$

la que es una ecuación en diferencias ordinaria con solución

$$\widehat{x}_{t+\ell} = b_0^{(t)} f_0(\ell) + b_1^{(t)} f_1(\ell) + \dots + b_{p^*}^{(t)} f_{p^*}(\ell), \quad (17)$$

donde las funciones f_j 's son polinomios, exponenciales, senos, cosenos, etc.. Esta última fórmula se denomina *función de predicción eventual*. La fórmula (17) es diferente para cada valor de t y a su vez está determinada enteramente (su estructura, no sus coeficientes) por el operador $\varphi^*(B)$.

5.3 Predicción de modelos no estacionarios (y estacionales)

Sea el operador $\Delta_s = 1 - B^s$, entonces $\Delta_s f(t) = f(t) - f(t-s)$, $\Delta_s^2 f(t) = (1 - B^s)^2 f(t) = f(t) - 2f(t-s) + f(t-2s)$, y así sucesivamente.

Supongamos que la serie x_t contiene un movimiento periódico con periodicidad s (como podría ser el caso de la estacionalidad), y que se puede aproximar adecuadamente en forma local por un polinomio de orden D . Podemos eliminarlo tomado diferencias estacionales de orden D de las observaciones x_t , esto es

$$y_t = \Delta_s^D x_t = (1 - B^s)^D x_t.$$

Luego suponemos que y_t puede ser representado por un modelo ARMA(p, q) estacionario de la forma

$$\varphi(B)y_t = \theta(B)\varepsilon_t.$$

Esto implica que el modelo para la serie no estacionaria x_t es

$$\begin{aligned} \varphi(B)\Delta_s^D x_t &= \theta(B)\varepsilon_t \\ \varphi(B)(1 - B^s)^D x_t &= \theta(B)\varepsilon_t. \end{aligned}$$

Debemos observar que el operador autorregresivo $\varphi(B)(1 - B^s)^D$ tiene ahora algunas de sus raíces sobre el círculo unitario, esto implica que la serie x_t es no

estacionaria. A fin de realizar las predicciones se procede de forma similar al caso en que teníamos tendencia solamente.

Debe notarse que se pueden combinar los casos en que la serie observada tenga tendencia y estacionalidad, siendo su solución similar a la que ya vimos.

Es importante poner atención en el hecho de que con el procedimiento de BJ lo que se hace es eliminar la tendencia y la estacionalidad a fin de lograr estacionariedad, para luego trabajar con la serie transformada. Posteriormente se deshacen las transformaciones para efectuar las predicciones de la serie original. De esta forma nunca se conocen y estiman los componentes tendencia y estacionalidad de forma separada, simplemente se los elimina, lo cual en muchas aplicaciones (por ejemplo economía, y otras ciencias sociales) es un problema pues es importante tenerlos a esos componentes estimados de forma explícita.

5.4 Identificación del modelo

Dada las observaciones de una serie de tiempo x_t , se presupone que, luego de efectuar las transformaciones pertinentes y tomar las diferencias adecuadas (tanto para eliminar la tendencia como para eliminar la estacionalidad u otros movimientos periódicos) la serie resultante seguirá apropiadamente un modelo ARMA(p, q) estacionario. En consecuencia, queremos determinar las transformaciones necesarias y los valores de p , d y q (también D en su caso) para lograr ese modelo. Este proceso es el que se conoce como *identificación*, el cual no es lo mismo que la identificación en econometría. A continuación enunciamos las diferentes formas de proceder ante los distintos casos que suelen presentarse.

Se logra una idea preliminar del valor de d al graficar la serie $\Delta^r x_t$, para $r = 0, 1, 2, \dots$, hasta que tenga la apariencia de una serie estacionaria, básicamente sin tendencia. También calculamos las autocorrelaciones muestrales (útil para determinar q) y las autocorrelaciones parciales muestrales (útil para determinar p) de cada una de las $\Delta^r x_t$, para $r = 0, 1, 2, \dots$. De forma similar, se logra una idea preliminar del valor de D al graficar la serie $\Delta_s^r x_t$, para $r = 0, 1, 2, \dots$, hasta que tenga la apariencia de una serie estacionaria, básicamente sin estacionalidad u otros movimientos periódicos.

Para la determinación de los valores de p y q debemos remitirnos a las propiedades de los respectivos procesos estacionarios. Así, si las observaciones provienen de una AR(p), esto implica que $q = 0$, que el correlograma de la serie tiende a cero de una manera rápida y que el correlograma parcial es idénticamente igual a cero para ordenes superiores a p . Por otra parte, si las observaciones provienen de un MA(q), esto implica que $p = 0$, que el correlograma de la serie es idénticamente igual a cero para ordenes superiores a q y que el correlograma parcial tiende a cero de una manera rápida. Además, si p y q son ambos distintos de cero, entonces el proceso generador de las observaciones es un ARMA y tanto el correlograma como el correlograma parcial tienden a cero de una manera rápida. Estas características teóricas son muy útiles, pero en la práctica no conocemos los valores verdaderos

de las autocorrelaciones ni de las autocorrelaciones parciales del proceso generador de las observaciones. Lo único que podemos hacer es estimarlas, tomando el conjunto de las observaciones disponibles, y usamos el hecho de que ambas, tanto las autocorrelaciones muestrales como las autocorrelaciones parciales muestrales, son asintóticamente normales con media igual al parámetro a estimar y varianza igual a $\frac{1}{n}$, donde n es el número de observaciones. Luego graficamos esas estimaciones y una banda de confianza de la forma $(-2/\sqrt{n}, 2/\sqrt{n})$, que es aproximadamente la correspondiente al 95%, y esperamos que los estimadores se comporten de forma similar a los parámetros que queremos estimar. En particular, si la estimación está dentro de la banda de confianza, podemos aceptar que el respectivo parámetro es igual a cero.

No solamente la presencia de tendencia, estacionalidad u otros movimientos periódicos pueden ser fuentes de falta de estacionariedad de la serie a estudiar. Otra fuente es el cambio con el tiempo de la varianza. En particular, cuando la variabilidad de la serie aumenta o disminuye con el nivel. Para detectarlo se grafica la serie y se espera que las oscilaciones sean todas aproximadamente de la misma amplitud. En caso de que esto no suceda se puede aplicar una transformación estabilizadora de la varianza. La más usada es la de Box y Cox (1964), la cual, para una serie de observaciones x_t , produce la serie y_t de la siguiente manera

$$y_t = \begin{cases} \frac{x_t^\lambda - 1}{\lambda}, & \lambda \neq 0, \\ \log x_t, & \lambda = 0. \end{cases} \quad (18)$$

Mediante una elección adecuada de λ , se puede hacer a la variabilidad aproximadamente constante. En particular, para datos positivos cuyo desvío estándar aumenta linealmente con el nivel, se puede estabilizar a la variabilidad tomando $\lambda = 0$. Algunas veces, no es necesario realizar este tipo de transformaciones ya que la serie tiene una variabilidad estable.

Este es un proceso ad hoc, el que nos da valores preliminares de p , d , D , q y λ . Debe tomarse en cuenta que en situaciones prácticas la serie observada puede tener simultáneamente variabilidad cambiante con el tiempo, tendencia, estacionalidad u otros movimientos periódicos, por lo tanto se deberán aplicar todos, o buena parte, de los procedimientos presentados en esta subsección para lograr la identificación. Si estamos en duda, tomemos diferentes conjuntos de valores de p , d , D , q y λ , y decidamos luego después del ajuste usando los test de significación que correspondan.

Nótese que la identificación esta basada en gran medida en las autocorrelaciones muestrales y las autocorrelaciones parciales muestrales, las que, de acuerdo a lo visto, están altamente autocorrelacionadas y tienen una alta variabilidad muestral, lo que las hace inestables para su uso en la práctica. Se pueden encontrar ejemplos en los cuales los datos aparecen explicados igualmente bien por modelos cuyas especificaciones son muy diferentes. Esto es un gran problema del enfoque de BJ.

Ahora consideremos algunas otras desventajas del enfoque de BJ. La eliminación de la tendencia y la estacionalidad u otros movimientos periódicos mediante diferenciación puede no ser un inconveniente si la predicción es el único objetivo del análisis, pero en muchos contextos, particularmente en estadísticas oficiales, el conocimiento de esos componentes tiene importancia intrínseca. Es cierto también que se pueden “recuperar” las estimaciones de la tendencia, de la estacionalidad o de otros movimientos periódicos a partir de la serie diferenciada maximizando la media cuadrática residual, pero este parece ser un procedimiento artificial que no es tan atractivo como modelar los componentes directamente.

El requerimiento que la serie diferenciada debe ser estacionaria es una debilidad de la teoría. En los campos de la economía y las ciencias sociales, las series nunca son estacionarias, no obstante toda la diferenciación que se haga. El investigador debe enfrentar la cuestión siguiente: ¿cuán cerca de la estacionariedad es suficientemente cerca? Esta pregunta es muy difícil de responder. Hay otro punto que está íntimamente relacionado con este. Al comienzo se supone que las observaciones están formadas por los términos tendencia, estacionalidad u otros movimientos periódicos e irregular. El término irregular se supone que se comporta de acuerdo a algún modelo estocástico. Encuentro una coincidencia a remarcar que la cantidad de diferenciación requerida para eliminar los términos de tendencia y estacionalidad (u otros movimientos periódicos) es exactamente la misma que se necesita para lograr un adecuado modelo ARMA estacionario para el término irregular diferenciado.

En el sistema de BJ es relativamente difícil de manejar cuestiones tales como observaciones perdidas, el agregado de variables explicativas, ajustes calendarios y cambios en el comportamiento a través del tiempo.

Una interesante discusión sobre los problemas del enfoque de BJ y las alternativas disponibles puede verse en Abril (2001).

5.5 Estimación del modelo

Supongamos que tenemos observaciones de la serie no estacionaria x_t , la que puede ser la serie realmente observada o bien una serie resultante luego de haberle estabilizado la varianza usando la transformación (18). Entonces, siguiendo el enfoque de BJ, hacemos $y_t = \Delta_s^D \Delta^d x_t$ para valores adecuados de D y d logrados durante el proceso de identificación. También, durante ese proceso identificamos al modelo ARMA (p, q) de la forma dada en (6) que entendemos que satisface la serie y_t . Esto último implica lograr los valores de p y q pertinentes. Con ello debemos proceder a estimar los parámetros respectivos, procedimiento que en esencia sigue los lineamientos que a continuación resumimos.

Suponemos inicialmente que $\varepsilon_0 = \varepsilon_{-1} = \dots = 0$ y que $y_0, y_{-1}, \dots, y_{-p+1}$ están

dados. También suponemos que $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ son independientes $N(0, \sigma_\varepsilon^2)$. Luego

$$\log L = K^* - \frac{n}{2} \log \sigma_\varepsilon^2 - \frac{1}{2\sigma_\varepsilon^2} S(\varphi, \theta), \quad (19)$$

donde K^* es una constante, $S(\varphi, \theta) = \sum_{t=1}^n \varepsilon_t^2$, y los ε_t 's están expresados en términos de y_1, \dots, y_n , y de los φ_j 's y los θ_j 's de acuerdo a la expresión (6). De acuerdo a (19), Máxima Verosimilitud (MV) $\equiv \min S(\varphi, \theta)$ con respecto a los φ_j 's y los θ_j 's.

Si no queremos suponer que $\varepsilon_0 = \varepsilon_{-1} = \dots = 0$, se los puede estimar mediante "back"-forecasting o "backcasting" como lo mostrado por Abril (2004) en el Capítulo 5. Con ello obtenemos las estimaciones $\hat{y}_0, \hat{y}_{-1}, \dots$, de y_0, y_{-1}, \dots y así, utilizándolas podemos lograr la estimación por "MV exacta" de los φ_j 's y los θ_j 's. También se obtiene el estimador por MV de σ_ε^2 .

5.6 Control de diagnóstico

Esto es equivalente a realizar tests del ajuste del modelo a las observaciones.

Sea $\{y_t\}$ la serie corregida por media, transformada y diferenciada de acuerdo a lo visto anteriormente. El problema es ahora, decidir si el modelo ARMA(p, q) estimado es el adecuado para representar a $\{y_t\}$. Si p y q hubieran sido conocidos de antemano esto hubiera constituido una aplicación directa de las técnicas de estimación vistas previamente, y allí el proceso hubiera terminado para luego continuar con la predicción. Ahora bien, esto usualmente no es el caso, por lo tanto resulta necesario determinar si los valores de p y q identificados y con los cuales se realizó la estimación son los apropiados.

Puede parecer a primera vista que cuando más grandes sean los valores de p y q mejor será el ajuste. No obstante debemos tener cuidado de los peligros del *sobreajuste*, esto es, de "hacer a medida" el ajuste de tal manera de que esté muy cercano a los valores particulares observados.

Se han desarrollado criterios, en particular el criterio de información AIC de Akaike (1974b) y el criterio *CAT* de Parzen (1974), los cuales intentan prevenir el sobreajuste asignándole efectivamente un costo a la introducción de cada parámetro adicional. El estadístico AIC para un modelo ARMA(p, q) con vectores de coeficientes $\varphi = (\varphi_1, \dots, \varphi_p)'$ y $\theta = (\theta_1, \dots, \theta_q)'$ es definido como

$$\text{AIC}(\varphi, \theta) = -2 \log L(\varphi, \theta, S(\varphi, \theta)/n) + 2(p + q + 1), \quad (20)$$

donde $L(\varphi, \theta, \sigma^2)$ es la verosimilitud de los datos bajo un modelo ARMA Gaussiano con parámetros $(\varphi, \theta, \sigma^2)$ y $S(\varphi, \theta)$ es la suma de cuadrados residual. El modelo seleccionado es aquel que minimiza el valor de AIC definido en (20).

Ante dos modelos igualmente aceptables, debemos elegir aquel que sea más parsimonioso.

Definición. *Parsimonia* es el requerimiento de que ante modelos igualmente “buenos”, nos debemos quedar con el que tenga menor número de parámetros.

Estudios de tipo Monte Carlo (ver Jones, 1975; Shibata, 1976) sugieren que el estadístico AIC definido en (20) tiene una tendencia a sobreestimar p (y también q). Ante ello, Akaike (1978) propuso el estadístico BIC el cual intenta corregir la propensión al sobreajuste de AIC y trabaja de la misma manera que (20). La definición de BIC es

$$\begin{aligned} \text{BIC} = & (n - p - q) \log [n\hat{\sigma}^2/(n - p - q)] + n \left(1 + \log \sqrt{2n}\right) \\ & + (p + q) \log \left[\left(\sum_{t=1}^n y_t^2 - n\hat{\sigma}^2 \right) / (p + q) \right], \end{aligned} \quad (21)$$

donde $\hat{\sigma}^2$ es la estimación por MV de la varianza del ruido blanco.

La versión corregida por sesgo de AIC, conocida como AICC y que fue sugerida por Hurvich y Tsai (1989), es la que más se usa en la práctica. La definición de AICC es

$$\text{AICC}(\varphi, \theta) = -2 \log L(\varphi, \theta, S(\varphi, \theta)/n) + 2(p + q + 1)n/(n - p - q - 2), \quad (22)$$

donde $L(\varphi, \theta, \sigma^2)$ es la verosimilitud de los datos bajo un modelo ARMA Gaussiano con parámetros $(\varphi, \theta, \sigma^2)$ y $S(\varphi, \theta)$ es la suma de cuadrados residual. Intuitivamente, podemos pensar que $2(p + q + 1)n/(n - p - q - 2)$ en (22) es una penalización para desalentar la sobre parametrización. Nuevamente, debemos elegir aquel modelo que minimice AICC. La introducción del estadístico AICC (o sus análogos) reduce la cuestión de elegir los valores de p y q a un problema bien definido. No obstante, la búsqueda de un modelo que minimice el AICC puede ser muy lenta si no se tiene una idea de la clase de modelos que se deben explorar. Diversas técnicas pueden usarse para acelerar la búsqueda, pero las más adecuadas son aquellas basadas en el estudio de las autocorrelaciones y las autocorrelaciones parciales muestrales de $\{y_t\}$, de acuerdo a lo enunciado en la subsección denominada *identificación del modelo*.

La siguiente es una alternativa para decidir si el orden del modelo ARMA ajustado a la serie $\{y_t\}$ es el adecuado: Sean $\hat{S}_1 = S(\hat{\varphi}, \hat{\theta})$ suma de cuadrados residual luego de ajustar p_1 y q_1 parámetros, y $\hat{S}_2 =$ suma de cuadrados residual luego de ajustar p_2 y q_2 parámetros, con $p_2 \geq p_1, q_2 \geq q_1$ y $p_2 + q_2 > p_1 + q_1$, entonces el estadístico del test a usar es

$$\frac{\frac{1}{p_2 + q_2 - p_1 - q_1}(\hat{S}_1 - \hat{S}_2)}{\frac{1}{n - p_2 - q_2}\hat{S}_2}. \quad (23)$$

El test definido en (23) es un test de cocientes de verosimilitudes, en consecuencia tiene una distribución χ^2 con $(p_2 + q_2 - p_1 - q_1)$ grados de libertad bajo la hipótesis nula que el modelo más chico (el más parsimonioso, el basado en \hat{S}_1) ajusta tan bien como el otro. Por lo tanto, si se acepta la hipótesis nula debemos tomar como modelo adecuado a aquel basado en \hat{S}_1 .

Una vez que se ha encontrado un modelo que minimice el valor de AICC, debemos controlar la bondad del ajuste, esencialmente verificando que los residuos se comportan como si fueran un ruido blanco.

Sabemos que las autocorrelaciones muestrales de una sucesión z_1, \dots, z_n de variables aleatorias independientes e idénticamente distribuidas (iid) con $E(z_t^2) < \infty$ se distribuyen aproximadamente como $N(0, 1/n)$ para n grande. Entonces, suponiendo que hemos ajustado el modelo ARMA apropiado a nuestros datos y que ese modelo ARMA es generado por una sucesión de ruidos blancos iid, la misma aproximación debería ser válida para las autocorrelaciones muestrales $r_k(\hat{\varepsilon}_t)$, $k = 1, 2, \dots$, de los residuos $\hat{\varepsilon}_t$. Los valores de $\hat{\varepsilon}_t$ son estimados mediante la aplicación del análogo muestral de la fórmula $\varepsilon_t = y_t - \hat{y}_{t-1}(1)$, esto es con los coeficientes reemplazados por sus estimaciones. Por lo tanto, si los residuos $\hat{\varepsilon}_t$ estiman un ruido blanco iid, cada una de sus autocorrelaciones muestrales deberían estar dentro del intervalo $(-2/\sqrt{n}, 2/\sqrt{n})$, lo que significa que estamos realizando el respectivo test con un nivel de significación aproximado del 5%. En la práctica se grafica la función de autocorrelación muestral de los residuos y la anterior banda de confianza, viéndose de esa forma si se acepta o no la hipótesis nula que los ε_t generadores del modelo ARMA son un conjunto de ruidos blancos iid.

En lugar de comprobar si cada $r_k(\hat{\varepsilon}_t)$ cae dentro de la banda de confianza $\pm \frac{2}{\sqrt{n}}$, es posible considerar en cambio un solo estadístico el que depende de $r_k(\hat{\varepsilon}_t)$ para $k = 1, \dots, K$. A lo largo de esta discusión se supone que K depende del tamaño muestral n de tal manera que (i) $K \rightarrow \infty$ cuando $n \rightarrow \infty$, y (ii) las condiciones de Box y Pierce (1970) son satisfechas, las cuales son

1. $\psi_j = O(n^{-1/2})$ para $j \geq K$, donde ψ_j , $j = 0, 1, \dots$, son los coeficientes en la expansión $y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$, y
2. $K = O(n^{1/2})$.

Entonces, de acuerdo a lo dicho anteriormente, el estadístico propuesto por Box y Pierce (1970) es

$$Q = n \sum_{k=1}^K r_k^2(\hat{\varepsilon}_t), \tag{24}$$

el que se distribuye aproximadamente como χ^2 con $(K - p - q)$ grados de libertad.

El modelo estimado es rechazado al nivel α si

$$Q > \chi_{1-\alpha}^2(K - p - q), \quad (25)$$

donde $\chi_{1-\alpha}^2(K - p - q)$ es el valor obtenido de una tabla de la distribución χ^2 con $(K - p - q)$ grados de libertad que acumula el $100(1 - \alpha)\%$ del área. De acuerdo a la condición 2 de Box y Pierce, K es del orden de \sqrt{n} , pero frecuentemente en la práctica se lo toma $15 \leq K \leq 30$.

Desafortunadamente la aproximación χ^2 del estadístico Q definido en (24) no suele ser muy buena cuando $n < 100$. Varias alternativas han sido propuestas. La más usada es la de Ljung y Box (1978) quienes propusieron el estadístico

$$Q^* = n(n + 2) \sum_{k=1}^K \frac{r_k^2(\hat{\varepsilon}_t)}{n - k}, \quad (26)$$

el que también se distribuye aproximadamente como χ^2 con $(K - p - q)$ grados de libertad.

A los tests basados en los estadísticos (24) y (26) se los suele denominar tests “*portmanteau*”. Una ventaja de los tests portmanteau es que ellos juntan información de las autocorrelaciones muestrales $r_k(\hat{\varepsilon}_t)$ para diferentes rezagos $k = 1, \dots, K$. Una clara desventaja es que frecuentemente dejan de rechazar modelos con ajustes pobres, o sea que tienen poco poder. En la práctica, los tests portmanteau son usados para descalificar modelos insatisfactorios y no para seleccionar el modelo de mejor ajuste entre competidores muy próximos.

Existen muchos otros tests que pueden ser usados en este proceso de control de diagnóstico. Aquí solamente hemos presentado aquellos que son usados con mayor frecuencia en situaciones prácticas. Mayores detalles sobre otros tests pueden verse en Brockwell y Davis (1991).

Con respecto a la tarea de computo necesaria para realizar el análisis práctico de las series de tiempo, la misma puede ejecutarse con la ayuda de una gran cantidad de programas generales de estadística tales como SPSS, SAS, etc. Ahora bien, hay un paquete específico de series de tiempo, que realiza todas las tareas de forma muy efectiva, fue desarrollado por Brockwell y Davis (2000) y su nombre es ITSM 2000.

6. El enfoque de espacio de estado

En esta sección se presenta un enfoque unificado para el análisis de series de tiempo. El tratamiento técnico está basado en los métodos de espacio de estado (EE). Estos métodos pueden ser aplicados a cualquier modelo lineal, incluyendo

aquellos dentro de la clase de los autorregresivos integrados de promedios móviles (ARIMA).

Todos los modelos lineales de series de tiempo tienen una representación en la forma de espacio de estado. Esta representación relaciona el vector de errores o disturbios $\{\varepsilon_t\}$ con el vector de observaciones $\{\mathbf{y}_t\}$ vía un proceso de Markov $\{\alpha_t\}$. Una expresión conveniente de la forma de espacio de estado es

$$\begin{aligned} \mathbf{y}_t &= \mathbf{Z}_t \alpha_t + \varepsilon_t, & \varepsilon_t &\sim N(\mathbf{0}, \mathbf{H}_t), \\ \alpha_t &= \mathbf{T}_t \alpha_{t-1} + \mathbf{R}_t \eta_t, & \eta_t &\sim N(\mathbf{0}, \mathbf{Q}_t), \quad t = 1, \dots, n, \end{aligned} \quad (27)$$

donde \mathbf{y}_t es un vector de orden $p \times 1$ de observaciones y α_t es un vector de orden $m \times 1$ inobservable, llamado el *vector de estado*.

La idea subyacente en el modelo es que el desarrollo del sistema en el tiempo está determinado por α_t de acuerdo con la segunda ecuación de (27); pero debido a que α_t no puede ser observado directamente, debemos basar nuestro análisis en las observaciones \mathbf{y}_t . Las matrices \mathbf{Z}_t , \mathbf{T}_t , \mathbf{R}_t , \mathbf{H}_t y \mathbf{Q}_t se suponen inicialmente conocidas y los términos de error ε_t y η_t se suponen que son serialmente independientes e independientes entre sí en todo momento de tiempo. Las matrices \mathbf{Z}_t y \mathbf{T}_t pueden depender de $\mathbf{y}_1, \dots, \mathbf{y}_{t-1}$. El estado inicial α_0 se supone que es $N(\mathbf{a}_0, \mathbf{P}_0)$ e independiente de $\varepsilon_1, \dots, \varepsilon_n$ y de η_1, \dots, η_n , donde en principio se supone que \mathbf{a}_0 y \mathbf{P}_0 son conocidos. Posteriormente consideraremos cómo proceder ante la ausencia de conocimiento sobre \mathbf{a}_0 y \mathbf{P}_0 , y esto se conoce como *iniciación*.

Al sistema (27) se lo denomina usualmente *modelo básico de espacio de estado* (MBEE), aunque también se lo conoce como *modelo lineal Gaussiano de espacio de estado*. A la primera ecuación de (27) se la conoce usualmente como la *ecuación de medida* o *ecuación de observación* y a la segunda, la *ecuación de transición* o *relación de transición* o *ecuación de estado*. Es interesante destacar que la ecuación de medida es equivalente a un modelo de regresión con coeficientes α_t estocásticos que satisfacen la ecuación de transición.

En muchas aplicaciones \mathbf{R}_t es la matriz identidad. En otras, uno puede definir $\eta_t^* = \mathbf{R}_t \eta_t$ y $\mathbf{Q}_t^* = \mathbf{R}_t \mathbf{Q}_t \mathbf{R}_t'$, y proceder sin la inclusión explícita de \mathbf{R}_t , con lo cual se hace parecer al modelo mucho más simple. No obstante, si \mathbf{R}_t es de orden $m \times g$ y \mathbf{Q}_t tiene rango $g < m$, existe una ventaja obvia en trabajar con η_t no singular en lugar de η_t^* singular. En modelos univariados $p = 1$, por lo tanto \mathbf{Z}_t es un vector fila y \mathbf{H}_t es un escalar que se lo suele denotar como σ_ε^2 .

Denotemos como \mathbf{Y}_{t-1} al conjunto $\mathbf{y}_1, \dots, \mathbf{y}_{t-1}$ junto con toda la información anterior al tiempo $t = 1$. Comenzando en $t = 1$ y construyendo la distribución de α_t e \mathbf{y}_t recursivamente, se puede mostrar que

$$p(\mathbf{y}_t \mid \alpha_1, \dots, \alpha_t, \mathbf{Y}_{t-1}) = p(\mathbf{y}_t \mid \alpha_t)$$

y

$$p(\alpha_t \mid \alpha_1, \dots, \alpha_{t-1}, \mathbf{Y}_{t-1}) = p(\alpha_t \mid \alpha_{t-1}),$$

estableciendo con ello la verdadera naturaleza Markoviana del modelo.

El modelo (27) provee herramientas poderosas para el análisis de una amplia clase de problemas.

El análisis estadístico de modelos no Gaussianos es ahora posible debido a los desarrollos recientes en las técnicas de simulación, las que tienen soluciones dentro del enfoque de espacio de estado.

Por muchos años los modelos autorregresivos integrados de promedios móviles (ARIMA) han sido considerados como los que proveen la base principal para el modelado de series de tiempo. Ahora bien, dada la tecnología actual, puede haber alternativas más atractivas, particularmente cuando se tratan datos desordenados. En las próximas secciones presentaremos las ideas básicas del enfoque de espacio de estado y del modelado estructural de series de tiempo, y haremos notar la relación con los modelos autorregresivos integrados de promedios móviles. Una muy interesante discusión sobre las semejanzas y diferencias entre los modelos de espacio de estado y los otros modelos usados en la práctica del análisis de las series de tiempo puede verse en Abril (2001b).

El filtro de Kalman es descripto más adelante. Este es necesario para el manejo de modelos estructurales de series de tiempo, pero con mayor importancia, y más allá de la clase de modelos usados, el mismo es crucial para el tratamiento de datos desordenados.

En general, cuando un modelo dinámico puede ser puesto en una forma lineal Gaussiana de Markov se lo trata usualmente con los métodos de espacio de estado.

Un libro recomendado como apoyo a lo presentado aquí es el de Harvey (1989). Este provee un excelente tratamiento enciclopédico del área hasta aproximadamente 1989. Muy buenas fuentes suplementarias para aplicaciones futuras y trabajos posterior son Harvey y Shephard (1993) y Abril (1997 y 1999). Interesantes aplicaciones de estas técnicas en econometría pueden verse en Abril, Ferullo y Gaínza Córdoba (1998 y 1999) y Abril y Blanco (2002). Un paquete de computación interesante para implementar el análisis estructural de las series de tiempo, con el nombre de STAMP (*Structural Time Series Analyser, Modeller and Predictor*), ha sido desarrollado por Koopman, Harvey, Doornik y Shephard (1995 y 2000). Otro conjunto de rutinas escritas en C, que llevan el nombre de SsfPack, han sido desarrolladas para realizar el análisis estadístico de modelos univariados y multivariados puestos en la forma de espacio de estado. Las instrucciones de uso y las propiedades de estas rutinas están documentadas en Koopman, Shephard y Doornik (1999), además se puede lograr información adicional sobre ellas en la página de internet <http://www.ssfpack.com/>.

El modelo estructural de series de tiempo más simple presenta una situación en la que no hay estacionalidad ni ciclos, la tendencia está compuesta solamente por el nivel de la serie que cambia a través del tiempo. Ese nivel es modelado por

un camino aleatorio (“random walk”) sobre el cual se le sobreimpone un error aleatorio o ruido blanco. Además, todas las variables aleatorias se distribuyen normalmente. Esto da el modelo

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t, & \varepsilon_t &\sim N(0, \sigma_\varepsilon^2), \\ \mu_t &= \mu_{t-1} + \eta_t, & \eta_t &\sim N(0, \sigma_\eta^2), \end{aligned} \quad (28)$$

para $t = 1, \dots, n$, donde las ε_t 's y las η_t 's son serialmente y mutuamente independientes, y $N(a, \sigma^2)$ denota una distribución normal con media a y varianza σ^2 . A (28) se lo conoce como modelo de nivel local (NL).

Una característica práctica muy importante del modelo (28) es que el estimador del nivel, basado en la información corriente disponible, está dado por un promedio móvil exponencialmente ponderado de las observaciones pasadas, donde la constante de suavizado es una función del *cociente señal-ruido*, $q = \sigma_\eta^2 / \sigma_\varepsilon^2$ (ver Abril, 2002). La predicción de observaciones futuras, no importa cuantos pasos adelante, está dada exactamente por la misma expresión. Esto fue mostrado por Muth (1960). Para un camino aleatorio puro, q es infinito, llegándose a que la predicción es la última observación. A medida que q tiende a 0, la predicción se acerca a la media muestral.

6.1 Filtrado

El objetivo del filtrado es actualizar nuestro conocimiento del sistema cada vez que una nueva observación \mathbf{y}_t es obtenida. Una vez que el modelo ha sido puesto en la forma (27) de espacio de estado, el camino está abierto para la aplicación de un número importante de algoritmos. En el centro de ellos está al *filtro de Kalman*. Este filtro es un procedimiento recursivo para computar el estimador óptimo del vector de estado en el momento t , basado en la información disponible hasta ese tiempo t . Esa información consiste en observaciones hasta \mathbf{y}_t incluida.

En ciertas aplicaciones de ingeniería el filtro de Kalman es importante debido a la posibilidad de lograr estimaciones sobre la marcha (“on-line”). El valor actual del vector de estado es de primordial interés (por ejemplo, puede representar las coordenadas de un cohete en el espacio) y el filtro de Kalman permite que la estimación del vector de estado sea continuamente actualizada cada vez que una nueva observación está disponible. A primera vista, el valor de tal procedimiento en las aplicaciones económicas parece limitado. Nuevas observaciones tienden a aparecer a intervalos menos frecuentes y el énfasis está en hacer *predicciones* de observaciones futuras basadas en una muestra dada. El vector de estado no siempre tiene una interpretación económica, y en los casos en que la tiene, es más apropiado estimar su valor en un punto particular del tiempo usando toda la información de la muestra y no solamente parte de ella. Estos dos problemas son conocidos como *predicción* y *suavizado* respectivamente. Sucede que el filtro de Kalman nos provee las bases para las soluciones de ambos problemas.

Otra razón para el rol central del filtro de Kalman es que cuando los errores y el vector de estado inicial están normalmente distribuidos, permite que la función de verosimilitud sea calculada a través de lo que se conoce como la descomposición del error de predicción. Esto abre el camino para la estimación de cualquier parámetro desconocido en el modelo. También provee las bases para los tests estadísticos y de especificación del modelo.

La forma en que se deriva más abajo el filtro de Kalman para el modelo (27) se basa en el supuesto que el estado inicial α_0 es $N(\mathbf{a}_0, \mathbf{P}_0)$ e independiente de $\varepsilon_1, \dots, \varepsilon_n$ y de η_1, \dots, η_n , con \mathbf{a}_0 y \mathbf{P}_0 . Luego es usado un resultado estándar sobre la distribución normal multivariada para mostrar cómo es posible calcular recursivamente la distribución de α_t condicional en la información establecida en el tiempo t , para todo t desde 1 hasta n . Estas distribuciones condicionales son a su vez normales y por lo tanto están completamente especificadas por sus matrices de medias y varianzas. Son estas cantidades las que el filtro de Kalman computa. Así, supóngase que queremos obtener la distribución a posteriori de α_{t+1} dado \mathbf{Y}_t . Puesto que todas las distribuciones son normales, las distribuciones condicionales son también normales. Supongamos que α_t dado \mathbf{Y}_{t-1} es $N(\mathbf{a}_t, \mathbf{P}_t)$ y la de α_t dado \mathbf{Y}_t es $N(\mathbf{a}_{t|t}, \mathbf{P}_{t|t})$. Nuestro objetivo es calcular recursivamente $\mathbf{a}_{t|t}$, $\mathbf{P}_{t|t}$, \mathbf{a}_{t+1} y \mathbf{P}_{t+1} dado $\mathbf{a}_t, \mathbf{P}_t$.

Sea

$$\mathbf{v}_t = \mathbf{y}_t - \mathbf{Z}_t \mathbf{a}_t. \quad (29)$$

Entonces \mathbf{v}_t es el *error de predicción un paso adelante* y satisface

$$\mathbf{v}_t = \mathbf{y}_t - E(\mathbf{y}_t | \mathbf{Y}_{t-1}).$$

Denotemos a su matriz de varianzas como $\mathbf{F}_t = \text{var}(\mathbf{v}_t)$. Entonces

$$\mathbf{F}_t = \mathbf{Z}_t \mathbf{P}_t \mathbf{Z}_t' + \mathbf{H}_t, \quad t = 1, \dots, n. \quad (30)$$

Mediante la teoría elemental de regresión se puede mostrar que la regresión de α_t en \mathbf{Y}_t es igual a la regresión en \mathbf{Y}_{t-1} más la regresión en $\mathbf{y}_t - E(\mathbf{y}_t | \mathbf{Y}_{t-1})$. Consecuentemente,

$$\mathbf{a}_{t|t} = \mathbf{a}_t + \mathbf{B}_t \mathbf{v}_t, \quad (31)$$

donde

$$\begin{aligned} \mathbf{B}_t &= E(\alpha_t \mathbf{v}_t') [\text{var}(\mathbf{v}_t)]^{-1} \\ &= E[\alpha_t (\alpha_t - \mathbf{a}_t)' \mathbf{Z}_t' + \alpha_t \varepsilon_t'] \mathbf{F}_t^{-1} \\ &= \mathbf{P}_t \mathbf{Z}_t' \mathbf{F}_t^{-1}, \quad t = 1, \dots, n. \end{aligned} \quad (32)$$

Nuevamente, mediante la teoría elemental de regresión,

$$\text{var}(\mathbf{a}_{t|t}) = \text{var}(\mathbf{a}_t) - \mathbf{B}_t \text{var}(\mathbf{v}_t) \mathbf{B}_t',$$

así

$$\mathbf{P}_{t|t} = \mathbf{P}_t - \mathbf{P}_t \mathbf{Z}_t' \mathbf{F}_t^{-1} \mathbf{Z}_t \mathbf{P}_t, \quad t = 1, \dots, n. \quad (33)$$

Estos resultados también pueden obtenerse mediante la técnica de “completar el cuadrado” y basándose en la teoría de la distribución normal multivariada y sus distribuciones condicionales resultantes, pero la manipulación de matrices es mucho más intrincada; ver, por ejemplo, el apéndice del Capítulo 3 y sus aplicaciones en la sección 3.2.2 del libro de Harvey (1989). Ellos también surgen de la teoría Bayesiana estándar. Para obtener \mathbf{a}_{t+1} y \mathbf{P}_{t+1} simplemente notamos que

$$\begin{aligned}\mathbf{a}_{t+1} &= E(\alpha_{t+1} \mid \mathbf{Y}_t) \\ &= E(\mathbf{T}_{t+1}\alpha_t + \mathbf{R}_{t+1}\eta_{t+1} \mid \mathbf{Y}_t) \\ &= \mathbf{T}_{t+1}\mathbf{a}_{t|t},\end{aligned}\tag{34}$$

y

$$\begin{aligned}\mathbf{P}_{t+1} &= \text{var}(\alpha_{t+1} \mid \mathbf{Y}_t) \\ &= \text{var}(\mathbf{T}_{t+1}\alpha_t + \mathbf{R}_{t+1}\eta_{t+1} \mid \mathbf{Y}_t) \\ &= \mathbf{T}_{t+1}\mathbf{P}_{t|t}\mathbf{T}'_{t+1} + \mathbf{R}_{t+1}\mathbf{Q}_{t+1}\mathbf{R}'_{t+1},\end{aligned}\tag{35}$$

para $t = 0, \dots, n - 1$, con $\mathbf{a}_{0|0} = \mathbf{a}_0$ y $\mathbf{P}_{0|0} = \mathbf{P}_0$.

Podemos combinar las fórmulas (31) a la (35) para que den un solo par de ecuaciones de actualización

$$\mathbf{a}_{t+1} = \mathbf{T}_{t+1}\mathbf{a}_t + \mathbf{K}_t\mathbf{v}_t,\tag{36}$$

donde

$$\begin{aligned}\mathbf{K}_t &= \mathbf{T}_{t+1}\mathbf{B}_t \\ &= \mathbf{T}_{t+1}\mathbf{P}_t\mathbf{Z}'_t\mathbf{F}_t^{-1},\end{aligned}\tag{37}$$

y

$$\mathbf{P}_{t+1} = \mathbf{T}_{t+1}\mathbf{P}_t(\mathbf{T}'_{t+1} - \mathbf{Z}'_t\mathbf{K}'_t) + \mathbf{R}_{t+1}\mathbf{Q}_{t+1}\mathbf{R}'_{t+1},\tag{38}$$

para $t = 0, \dots, n - 1$, con $\mathbf{K}_0 = \mathbf{0}$.

El conjunto de fórmulas (31) a la (35), o alternativamente (36) a la (38), constituyen el filtro de Kalman para el modelo (27).

Se puede mostrar que cuando las observaciones no están normalmente distribuidas, restringiendo a estimaciones que son lineales en las \mathbf{y}_t 's y bajo supuestos apropiados, los valores de \mathbf{a}_{t+1} dados por el filtro minimizan el error cuadrático medio de estimación de cada componente de α_{t+1} . Ver, por ejemplo, Duncan y Horn (1972), Anderson y Moore (1979) y Abril (2001a, 2002) para detalles de este enfoque.

6.2 Iniciación

Ahora consideremos cómo iniciar el filtrado cuando nada se conoce sobre los parámetros \mathbf{a}_0 y \mathbf{P}_0 de la distribución de α_0 . En esta situación es razonable

representar a esa distribución como una distribución *difusa a priori*, esto es, se fija \mathbf{a}_0 en un valor arbitrario y se hacen tender los elementos diagonales de \mathbf{P}_0 a ∞ .

Detalles sobre otras alternativas de iniciación pueden verse en Abril (1999) y en las referencias allí citadas.

6.3 Suavizado

Consideraremos ahora la estimación de $\alpha_1, \dots, \alpha_n$ dada la muestra completa \mathbf{Y}_n . El estimador con *error cuadrático medio mínimo* (ECMM) de α_t es $\hat{\alpha}_t = E(\alpha_t | \mathbf{Y}_n)$. Llamamos a $\hat{\alpha}_t$ el *valor suavizado* de α_t y llamamos a la operación de calcular $\hat{\alpha}_1, \dots, \hat{\alpha}_n$ *suavizado*. Consideremos primero el caso en que \mathbf{a}_0 y \mathbf{P}_0 son conocidos y luego tratamos el caso difuso.

Puesto que $\mathbf{v}_1, \dots, \mathbf{v}_n$ son independientes y a su vez son transformaciones lineales de $\mathbf{y}_1, \dots, \mathbf{y}_n$ y puesto que $\mathbf{v}_t, \dots, \mathbf{v}_n$ son independientes de $\mathbf{y}_1, \dots, \mathbf{y}_{t-1}$ y tienen media cero, donde los \mathbf{v}_t 's fueron definidos en § 6.1, vemos que

$$\begin{aligned}\hat{\alpha}_t &= E(\alpha_t | \mathbf{Y}_{t-1}) + \sum_{j=t}^n E(\alpha_t | \mathbf{v}_j) \\ &= \mathbf{a}_t + \sum_{j=t}^n \mathbf{B}_{tj} \mathbf{v}_j, \quad t = 1, \dots, n,\end{aligned}\tag{39}$$

donde \mathbf{B}_{tj} es el coeficiente de la regresión de α_t en \mathbf{v}_t y está dado por $\mathbf{B}_{tj} = E(\alpha_t \mathbf{v}_j') \mathbf{F}_j^{-1}$. Ahora bien, $\mathbf{v}_j = \mathbf{Z}_j(\alpha_j - \mathbf{a}_j) + \varepsilon_j$, y $\alpha_s - \mathbf{a}_s = \mathbf{L}_{s-1}(\alpha_{s-1} - \mathbf{a}_{s-1}) + \mathbf{R}_s \eta_s - \mathbf{K}_{s-1} \varepsilon_{s-1}$, donde $\mathbf{L}_{s-1} = \mathbf{T}_s - \mathbf{K}_{s-1} \mathbf{Z}_{s-1}$ para $s = j, j-1, \dots$. Puesto que $E[(\alpha_t - \mathbf{a}_t)(\alpha_t - \mathbf{a}_t)'] = \mathbf{P}_t$ tenemos, en consecuencia,

$$E(\alpha_t \mathbf{v}_j') = \mathbf{P}_t \mathbf{L}'_t \mathbf{L}'_{t+1} \cdots \mathbf{L}'_{j-1} \mathbf{Z}'_j, \quad j = t+1, \dots, n.$$

Sustituyendo en (39) resulta

$$\begin{aligned}\hat{\alpha}_t &= \mathbf{a}_t + \mathbf{P}_t \mathbf{Z}'_t \mathbf{F}_t^{-1} \mathbf{v}_t + \mathbf{P}_t \mathbf{L}'_t \mathbf{Z}'_{t+1} \mathbf{F}_{t+1}^{-1} \mathbf{v}_{t+1} + \mathbf{P}_t \mathbf{L}'_t \mathbf{L}'_{t+1} \mathbf{Z}'_{t+2} \mathbf{F}_{t+2}^{-1} \mathbf{v}_{t+2} + \cdots \\ &= \mathbf{a}_t + \mathbf{P}_t \mathbf{r}_{t-1},\end{aligned}\tag{40}$$

donde \mathbf{r}_t satisface la recursión hacia atrás

$$\mathbf{r}_{t-1} = \mathbf{Z}'_t \mathbf{F}_t^{-1} \mathbf{v}_t + \mathbf{L}'_t \mathbf{r}_t, \quad t = n, n-1, \dots, 1,\tag{41}$$

con $\mathbf{r}_n = \mathbf{0}$. La varianza de \mathbf{r}_{t-1} está dada por

$$\mathbf{N}_{t-1} = \mathbf{Z}'_t \mathbf{F}_t^{-1} \mathbf{Z}_t + \mathbf{L}'_t \mathbf{N}_t \mathbf{L}_t, \quad t = n, n-1, \dots, 1,\tag{42}$$

con $\mathbf{N}_n = \mathbf{0}$. Los *residuos suavizados*, tal como los definen de Jong (1989) y de Jong y Penzer (1998), son estimados mediante la recursión

$$\mathbf{u}_t = \mathbf{F}_t^{-1} \mathbf{v}_t - \mathbf{K}_t' \mathbf{r}_t. \quad (43)$$

La varianza de \mathbf{u}_t está dada por

$$\mathbf{M}_t = \mathbf{F}_t^{-1} + \mathbf{K}_t' \mathbf{N}_t \mathbf{K}_t. \quad (44)$$

Como se puede ver, las cantidades \mathbf{u}_t y \mathbf{r}_t juegan el rol de pivotes en la construcción de test de diagnóstico para observaciones atípicas y cambios estructurales.

El suavizador, definido por (40) y (41), fue introducido por de Jong (1989) en un esquema más general que el considerado aquí. En la literatura el suavizador $E(\alpha_t | \mathbf{Y}_n)$ para t variable y n fijo se llama *suavizador de intervalo fijo*. Si t es fijo y $n = t + 1, t + 2, \dots$ se llama *suavizador de punto fijo*, y si ambos varían con $n - t$ fijo se llama *suavizador de rezago fijo*. En su trabajo, de Jong (1989), presenta algoritmos análogos para los suavizadores de punto fijo y de rezago fijo. Un suavizador anterior, al cual nos referiremos como *suavizador clásico*, está en la sección 3.6 de Harvey (1989) y es derivado en el Capítulo 7 de Anderson y Moore (1979). Este último toma la forma

$$\begin{aligned} \hat{\alpha}_t &= \mathbf{T}_t \mathbf{a}_t + \mathbf{P}_t^* (\hat{\alpha}_{t+1} - \mathbf{a}_{t+1}), \quad t = n - 1, \dots, 1, \\ \mathbf{P}_t^* &= \mathbf{P}_{t|t} \mathbf{T}_{t+1}' \mathbf{P}_t^{-1}, \end{aligned} \quad (45)$$

con $\hat{\alpha}_n = \mathbf{a}_{n|n}$. Comparando (45) con (40) y (41) se observa que el suavizador clásico requiere la inversión de $n - 1$ matrices \mathbf{P}_t , las que posiblemente pueden ser grandes, mientras que el suavizador de de Jong no requiere inversiones salvo la de \mathbf{F}_t , la cual ya fue invertida durante el proceso del filtrado de Kalman. Esto puede ser una ventaja considerable para modelos grandes.

Se puede mostrar directamente (ver Abril, 1999) que la varianza

$$\text{var}(\hat{\alpha}_t - \alpha_t) = \mathbf{P}_t - \mathbf{P}_t \mathbf{D}_{t-1} \mathbf{P}_t, \quad (46)$$

donde \mathbf{D}_{t-1} está dada por la recursión

$$\mathbf{D}_{t-1} = \mathbf{Z}_t' \mathbf{F}_t^{-1} \mathbf{Z}_t + \mathbf{L}_t' \mathbf{D}_t \mathbf{L}_t, \quad t = n, n - 1, \dots, 1, \quad (47)$$

con $\mathbf{D}_n = \mathbf{0}$, y también que la covarianza

$$\text{cov}(\hat{\alpha}_s - \alpha_s, \hat{\alpha}_t - \alpha_t) = \mathbf{P}_s \mathbf{L}_s' \cdots \mathbf{L}_{t-1}' (\mathbf{I} - \mathbf{D}_{t-1} \mathbf{P}_t), \quad s < t. \quad (48)$$

Si las matrices de varianzas y covarianzas no son requeridas, una variante del suavizador de de Jong, debida a Koopman (1993), es computacionalmente más

rápida. Sea $\hat{\eta}_t = E(\eta_t | \mathbf{Y}_n)$, $t = 1, \dots, n$. Puesto que $E(\eta_t \mathbf{v}'_j) = \mathbf{0}$ para $j < t$, tenemos

$$\hat{\eta}_t = \sum_{j=t}^n E(\eta_t \mathbf{v}'_j) \mathbf{F}_j^{-1} \mathbf{v}_j. \quad (49)$$

Como antes, $\mathbf{v}_j = \mathbf{Z}_j(\alpha_j - \mathbf{a}_j) + \varepsilon_j$ y $\alpha_s - \mathbf{a}_s = \mathbf{L}_{s-1}(\alpha_{s-1} - \mathbf{a}_{s-1}) + \mathbf{R}_s \eta_s - \mathbf{K}_{s-1} \varepsilon_{s-1}$, así para $s > t$, $E[(\eta_t(\alpha_s - \mathbf{a}_s))'] = E[(\eta_t(\alpha_{s-1} - \mathbf{a}_{s-1}))'] \mathbf{L}'_{s-1}$ mientras que $E[\eta_t(\alpha_t - \mathbf{a}_t)'] = \mathbf{Q}_t \mathbf{R}'_t$. De esta manera

$$E(\eta_t \mathbf{v}'_j) = \mathbf{Q}_t \mathbf{R}'_t \mathbf{L}'_t \mathbf{L}'_{t+1} \cdots \mathbf{L}'_{j-1} \mathbf{Z}'_j, \quad j = t+1, \dots, n. \quad (50)$$

Sustituyendo en (49) nos da

$$\hat{\eta}_t = \mathbf{Q}_t \mathbf{R}'_t \mathbf{r}_{t-1}, \quad (51)$$

donde \mathbf{r}_t satisface la recursión (41). Los $\hat{\alpha}_t$ son entonces calculados mediante la recursión hacia adelante

$$\hat{\alpha}_t = \mathbf{T}_t \hat{\alpha}_{t-1} + \mathbf{R}_t \hat{\eta}_t, \quad t = 1, \dots, n-1, \quad (52)$$

junto con $\hat{\alpha}_n = \mathbf{a}_{n|n}$. El valor inicial es $\hat{\alpha}_0 = \mathbf{a}_0 + \hat{\eta}_0$ con $\hat{\eta}_0 = \mathbf{P}_0 \mathbf{r}_0$.

La ventaja de este suavizador sobre el de de Jong es que es significativamente menos costoso en términos computacionales, y requiere menos espacio de almacenamiento. Los costos relativos de computación son examinados con todo detalle por Koopman en su trabajo. Por otra parte, el suavizador de Koopman tiene la desventaja de que las matrices de varianzas y covarianzas de los errores de estimación no pueden calcularse a partir de él; así, si ellas son necesarias, se deberá usar el suavizador de de Jong o el clásico.

6.4 Estimación de los Hiperparámetros

Los parámetros en los modelos de espacio de estado usualmente se denominan *hiperparámetros*, presumiblemente para distinguirlos de los elementos del vector de estado los cuales pueden pensarse como parámetros aleatorios. Además, recordemos que la magnitud por la cual los parámetros aleatorios pueden variar está gobernada por esos hiperparámetros.

Ahora debemos estimar por máxima verosimilitud los hiperparámetros del modelo de espacio de estado.

Comenzamos construyendo la verosimilitud. Suponiendo que \mathbf{a}_0 y \mathbf{P}_0 son conocidos, la densidad conjunta de $\mathbf{y}_1, \dots, \mathbf{y}_n$ es

$$p(\mathbf{Y}_n) = \prod_{t=1}^n p(\mathbf{y}_t | \mathbf{Y}_{t-1}), \quad (53)$$

donde $p(\mathbf{y}_t | \mathbf{Y}_{t-1}) = N(\mathbf{Z}_t \mathbf{a}_t, \mathbf{F}_t)$. Por lo tanto, tomando logaritmos obtenemos

$$\log L = -\frac{np}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^n \log |\mathbf{F}_t| - \frac{1}{2} \sum_{t=1}^n \mathbf{v}_t' \mathbf{F}_t^{-1} \mathbf{v}_t, \quad (54)$$

donde $\mathbf{v}_t = \mathbf{y}_t - \mathbf{Z}_t \mathbf{a}_t$ y \mathbf{F}_t está dada en (30). Se llama a (54) la *descomposición del error de predicción* del logaritmo de la verosimilitud.

Esta última fórmula debe maximizarse con respecto a los elementos del vector ψ de hiperparámetros desconocidos. Poniendo $\mathbf{H}_t = \sigma^2 \mathbf{H}_t^*$, $\mathbf{Q}_t = \sigma^2 \mathbf{Q}_t^*$, encontramos que $\mathbf{F}_t = \sigma^2 \mathbf{F}_t^*$ donde \mathbf{F}_t^* depende solamente de \mathbf{H}_t^* y de \mathbf{Q}_t^* junto con cualquier hiperparámetro desconocido que se encuentre en las restantes matrices del modelo. El parámetro de escala σ^2 puede concentrarse, reduciendo así la dimensión de la búsqueda numérica en una unidad.

En el caso difuso, cuando \mathbf{a}_0 y \mathbf{P}_0 son desconocidos, supongamos que τ es el menor valor de t para el cual $p(\alpha_t | \mathbf{Y}_t)$ existe. Entonces tomamos la verosimilitud condicional con \mathbf{Y}_τ fijo, dando

$$\log L = -\frac{(n-\tau)p}{2} \log(2\pi) - \frac{1}{2} \sum_{t=\tau+1}^n \log |\mathbf{F}_t| - \frac{1}{2} \sum_{t=\tau+1}^n \mathbf{v}_t' \mathbf{F}_t^{-1} \mathbf{v}_t. \quad (55)$$

Las expresiones (54) o (55) pueden luego maximizarse numéricamente, posiblemente después de haber sido concentradas.

6.5 Predicción

Supongamos que tenemos observaciones $\mathbf{y}_1, \dots, \mathbf{y}_n$ que satisfacen el modelo (27) de espacio de estado y queremos predecir $\mathbf{y}_{n+\ell}$, $\ell = 1, 2, \dots, J$. Siguiendo con el argumento dado en § 5 anterior, queremos que la predicción sea con *mínimo error medio cuadrático de predicción* (MEMCP) dada la muestra completa \mathbf{Y}_n . O sea que buscamos aquella estimación $\hat{\mathbf{y}}_n(\ell)$ de $\mathbf{y}_{n+\ell}$ que minimice

$$\bar{\mathbf{F}}_{n+\ell} = E[(\hat{\mathbf{y}}_n(\ell) - \mathbf{y}_{n+\ell})(\hat{\mathbf{y}}_n(\ell) - \mathbf{y}_{n+\ell})' | \mathbf{Y}_n] \quad (56)$$

en el sentido matricial, para todas las estimaciones de $\mathbf{y}_{n+\ell}$. De acuerdo a lo mostrado en §6.2, la predicción con MEMCP de $\mathbf{y}_{n+\ell}$ dada la muestra completa \mathbf{Y}_n es la media condicional

$$\hat{\mathbf{y}}_n(\ell) = E(\mathbf{y}_{n+\ell} | \mathbf{Y}_n). \quad (57)$$

Para $\ell = 1$ la predicción es directa. De (27) tenemos que

$$\mathbf{y}_{n+1} = \mathbf{Z}_{n+1} \alpha_{n+1} + \varepsilon_{n+1},$$

por lo tanto

$$\begin{aligned}\hat{\mathbf{y}}_n(1) &= \mathbf{Z}_{n+1}E(\alpha_{n+1} | \mathbf{Y}_n) \\ &= \mathbf{Z}_{n+1}\mathbf{a}_{n+1},\end{aligned}\quad (58)$$

donde \mathbf{a}_{n+1} es la estimación de α_{n+1} producida por el filtro de Kalman en (36). La matriz de varianza del error o matriz de error medio cuadrático

$$\begin{aligned}\bar{\mathbf{F}}_{n+1} &= E[(\hat{\mathbf{y}}_n(1) - \mathbf{y}_{n+1})(\hat{\mathbf{y}}_n(1) - \mathbf{y}_{n+1})' | \mathbf{Y}_n] \\ &= \mathbf{Z}_{n+1}\mathbf{P}_{n+1}\mathbf{Z}'_{n+1} + \mathbf{H}_{n+1},\end{aligned}\quad (59)$$

es calculada por la relación (30) del filtro de Kalman.

Para $\ell = 2, \dots, J$, definamos

$$\bar{\mathbf{a}}_{n+\ell} = E(\alpha_{n+\ell} | \mathbf{Y}_n),\quad (60)$$

y

$$\bar{\mathbf{P}}_{n+\ell} = E[(\bar{\mathbf{a}}_{n+\ell} - \alpha_{n+\ell})(\bar{\mathbf{a}}_{n+\ell} - \alpha_{n+\ell})' | \mathbf{Y}_n].\quad (61)$$

Puesto que

$$\mathbf{y}_{n+\ell} = \mathbf{Z}_{n+\ell}\alpha_{n+\ell} + \varepsilon_{n+\ell},$$

tenemos

$$\begin{aligned}\hat{\mathbf{y}}_n(\ell) &= \mathbf{Z}_{n+\ell}E(\alpha_{n+\ell} | \mathbf{Y}_n) \\ &= \mathbf{Z}_{n+\ell}\bar{\mathbf{a}}_{n+\ell},\end{aligned}\quad (62)$$

con matriz de error medio cuadrático

$$\begin{aligned}\bar{\mathbf{F}}_{n+\ell} &= E\{\mathbf{Z}_{n+\ell}(\bar{\mathbf{a}}_{n+\ell} - \alpha_{n+\ell}) - \varepsilon_{n+\ell}\}[\mathbf{Z}_{n+\ell}(\bar{\mathbf{a}}_{n+\ell} - \alpha_{n+\ell}) - \varepsilon_{n+\ell}]' | \mathbf{Y}_n\} \\ &= \mathbf{Z}_{n+\ell}\mathbf{P}_{n+\ell}\mathbf{Z}'_{n+\ell} + \mathbf{H}_{n+\ell}.\end{aligned}\quad (63)$$

Ahora derivamos las recursiones para calcular $\bar{\mathbf{a}}_{n+\ell}$ y $\bar{\mathbf{P}}_{n+\ell}$. De (27) tenemos que

$$\alpha_{n+\ell+1} = \mathbf{T}_{n+\ell+1}\alpha_{n+\ell} + \mathbf{R}_{n+\ell+1}\eta_{n+\ell+1},$$

entonces

$$\begin{aligned}\bar{\mathbf{a}}_{n+\ell+1} &= \mathbf{T}_{n+\ell+1}E(\alpha_{n+\ell} | \mathbf{Y}_n) \\ &= \mathbf{T}_{n+\ell+1}\bar{\mathbf{a}}_{n+\ell},\end{aligned}\quad (64)$$

para $\ell = 1, \dots, J-1$ y con $\bar{\mathbf{a}}_{n+1} = \mathbf{a}_{n+1}$. También

$$\begin{aligned}\bar{\mathbf{P}}_{n+\ell+1} &= E[(\bar{\mathbf{a}}_{n+\ell+1} - \alpha_{n+\ell+1})(\bar{\mathbf{a}}_{n+\ell+1} - \alpha_{n+\ell+1})' | \mathbf{Y}_n] \\ &= \mathbf{T}_{n+\ell+1}E[(\bar{\mathbf{a}}_{n+\ell} - \alpha_{n+\ell})(\bar{\mathbf{a}}_{n+\ell} - \alpha_{n+\ell})' | \mathbf{Y}_n]\mathbf{T}'_{n+\ell+1} \\ &\quad + \mathbf{R}_{n+\ell+1}E(\eta_{n+\ell+1}\eta'_{n+\ell+1})\mathbf{R}'_{n+\ell+1} \\ &= \mathbf{T}_{n+\ell+1}\bar{\mathbf{P}}_{n+\ell}\mathbf{T}'_{n+\ell+1} + \mathbf{R}_{n+\ell+1}\mathbf{Q}_{n+\ell+1}\mathbf{R}'_{n+\ell+1}.\end{aligned}\quad (65)$$

Observamos que las recursiones (64) y (65) para $\bar{\mathbf{a}}_{n+\ell}$ y $\bar{\mathbf{P}}_{n+\ell}$ respectivamente, son las mismas que las recursiones para $\mathbf{a}_{n+\ell}$ y $\mathbf{P}_{n+\ell}$ del filtro de Kalman dadas en (36) y (38) respectivamente, siempre que tomemos en estas últimas $\mathbf{v}_{n+\ell} = \mathbf{0}$ y $\mathbf{K}_{n+\ell} = \mathbf{0}$ para $\ell = 1, \dots, J - 1$. Pero estas son precisamente las condiciones que nos permiten tratar a las observaciones faltantes mediante la aplicación rutinaria del filtro de Kalman.

En consecuencia, hemos demostrado que las predicciones de $\mathbf{y}_{n+1}, \dots, \mathbf{y}_{n+J}$ junto con sus matrices de varianzas de los errores de predicción pueden ser obtenidas simplemente tratando a \mathbf{y}_t para $t > n$ como observaciones faltantes. En algún sentido esta conclusión podría ser considerada intuitivamente obvia, no obstante pensamos que es importante demostrarlo algebraicamente. Para resumir, las predicciones y sus matrices de varianzas de los errores asociadas pueden obtenerse rutinariamente en el análisis de series de tiempo basado en los modelos Gaussianos de espacio de estado mediante la continuación del filtrado de Kalman más allá de $t = n$ con $\mathbf{v}_t = \mathbf{0}$ y $\mathbf{K}_t = \mathbf{0}$ para $t > n$. Estos resultados para predicción constituyen una característica elegante de los métodos de espacio de estado para el análisis de las series de tiempo.

7. Orígenes del análisis moderno de series de tiempo

La década del '50 puede ser considerada como el inicio del análisis moderno de las series de tiempo debido a la introducción del suavizado exponencial. Efectivamente, el promedio móvil exponencialmente ponderado (PMEP) fue introducido en los '50 para lograr la predicción un paso adelante $\hat{y}_t(1)$ de y_{t+1} dada \mathbf{Y}_t , o sea toda la información hasta el momento t . Este tiene la forma

$$\hat{y}_t(1) = (1 - \lambda) \sum_{j=0}^{\infty} \lambda^j y_{t-j}, \quad 0 < \lambda < 1. \quad (66)$$

De (66), deducimos inmediatamente la recursión

$$\hat{y}_t(1) = (1 - \lambda)y_t + \lambda\hat{y}_{t-1}(1), \quad (67)$$

la que es usada en lugar de (66) para los cálculos prácticos. Esta es una estructura simple y requiere poco almacenamiento, por lo tanto fue muy conveniente para las computadoras primitivas disponibles en los '50. Como resultado, el PMEP se transformó en una técnica de predicción muy popular en la industria, particularmente para predecir ventas.

Una contribución importante fue realizada por Muth (1960) quien mostró que las predicciones de PMEP realizadas mediante la recursión (67) son de mínimo error medio cuadrático de predicción (MEMCP) en el sentido que minimizan $E[\hat{y}_t(1) - y_{t+1}]^2$ dada \mathbf{Y}_t para series generadas por el modelo

$$y_t = \mu_t + \varepsilon_t, \quad \mu_t = \mu_{t-1} + \eta_t, \quad (68)$$

donde ε_t y η_t son ruidos blancos independientes, esto es, son variables aleatorias serialmente independientes, independientes entre sí, con media cero y varianza constante. Este es, en efecto, un modelo simple de espacio de estado similar al definido en (28).

Tomando primeras diferencias de las observaciones generadas por (68) logramos

$$\Delta y_t = y_t - y_{t-1} = \varepsilon_t - \varepsilon_{t-1} + \eta_t.$$

Puesto que ε_t y η_t son serialmente no correlacionadas, la autocorrelación de orden 1 de es distinta de cero, pero todas las otras autocorrelaciones de ordenes mayores son cero. Esta es la función de autocorrelación de un modelo MA(1), esto es, de un modelo de promedios móviles de orden 1 dado por

$$\Delta y_t = \nu_t - \theta \nu_{t-1},$$

el cual es el modelo básico de BJ dado en (6) con $p = 0$ y $q = 1$, y en donde se reemplazó a ε_t por el proceso ortogonal ν_t .

Observamos que estas dos formas simples de modelos de EE y de BJ producen las mismas predicciones un paso adelante y que estas pueden ser calculadas mediante los PMEP dados en (67) y que estos probaron tener buenos valores prácticos. Podemos escribir a (67) de la siguiente forma

$$\hat{y}_t(1) = \hat{y}_{t-1}(1) + (1 - \lambda) \{y_t - \hat{y}_{t-1}(1)\},$$

a que es la forma del filtro de Kalman para predicciones generales de EE especializadas para este caso simple.

El modelo de PMEP fue extendido por Holt (1957) y Winters (1960) (HW) a series que contienen tendencia y efectos estacionales. La extensión para la tendencia en el caso aditivo es

$$\hat{y}_t(1) = m_t + b_t,$$

donde m_t y b_t son el nivel y la pendiente generados por las recursiones del tipo PMEP

$$m_t = (1 - \lambda_1)y_t + \lambda_1(m_{t-1} + b_{t-1}), \quad b_t = (1 - \lambda_2)(m_t - m_{t-1}) + \lambda_2 b_{t-1}.$$

En una extensión interesante de los resultados de Muth (1960), Theil y Wage (1964) mostraron que las predicciones logradas mediante las recursiones de HW son de MEMCP para el modelo de EE

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t, & \varepsilon_t &\sim N(0, \sigma_\varepsilon^2), \\ \mu_t &= \mu_{t-1} + \beta_{t-1} + \eta_t, & \eta_t &\sim N(0, \sigma_\eta^2), \\ \beta_t &= \beta_{t-1} + \zeta_t, & \zeta_t &\sim N(0, \sigma_\zeta^2). \end{aligned} \quad (69)$$

Esto extiende al modelo de camino aleatorio más ruido (68) agregándole un término de pendiente β_t el que permite al modelo seguir una tendencia variable en el tiempo la que es localmente lineal.

Tomando segundas diferencias de y_t generada por (69), obtenemos

$$\Delta^2 y_t = \zeta_t + \eta_t - \eta_{t-1} + \varepsilon_t - 2\varepsilon_{t-1} + \varepsilon_{t-2}.$$

Esta es una serie estacionaria con autocorrelaciones de ordenes 1 y 2 distintas de cero pero todas las otras autocorrelaciones son cero. Por lo tanto sigue el modelo MA(2)

$$\Delta^2 y_t = \nu_t - \theta_1 \nu_{t-1} - \theta_2 \nu_{t-2},$$

el que es una forma simple del modelo de BJ.

Ahora consideremos qué pasa cuando la estacionalidad es agregada al cuadro. Comenzando con el modelo (68) de camino aleatorio más ruido, le agregamos un término de estacionalidad γ_t , dando

$$y_t = \mu_t + \gamma_t + \varepsilon_t.$$

Si la estructura estacional fuera constante en el tiempo, los γ_t 's satisfarían la condición $\gamma_t + \gamma_{t-1} + \dots + \gamma_{t-s+1} = 0$, donde s es el número de “meses” por “año”. Para permitir que la estructura estacional cambie en el tiempo, agregamos un término de ruido blanco y obtenemos el modelo estructural

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t, & \varepsilon_t &\sim N(0, \sigma_\varepsilon^2), \\ \mu_t &= \mu_{t-1} + \eta_t, & \eta_t &\sim N(0, \sigma_\eta^2), \\ \gamma_t &= -\gamma_{t-1} - \dots - \gamma_{t-s+1} + \omega_t, & \omega_t &\sim N(0, \sigma_\omega^2). \end{aligned} \quad (70)$$

Esto satisface la forma general de EE dada en (27).

Tomando primeras diferencias y primeras diferencias estacionales de (70), encontramos

$$\Delta \Delta_s y_t = \eta_t - \eta_{t-s} + \omega_t - 2\omega_{t-1} + \omega_{t-2} + \varepsilon_t - \varepsilon_{t-1} - \varepsilon_{t-s} + \varepsilon_{t-s+1},$$

la que constituye efectivamente una serie de tiempo estacionaria con correlaciones de ordenes 1, 2, $s-1$, s y $s+1$ distintas de cero. Consideremos el modelo de BJ que se lo puede expresar en la forma

$$\Delta \Delta_s y_t = \nu_t - \theta_1 \nu_{t-1} - \theta_s \nu_{t-s} + \theta_1 \theta_s \nu_{t-s-1}.$$

Este es el famoso “modelo de aerolínea” de BJ el cual ha resultado ser bueno para ajustar numerosas series de tiempo económicas que contienen tendencia y efectos estacionales. Tiene autocorrelaciones de ordenes 1, $s-1$, s y $s+1$ diferentes de cero. La autocorrelación de orden 2 del modelo (70) sucede sólo a través de $\text{var}(\omega_t)$ la que en muchos casos prácticos es pequeña. De esa manera, cuando agregamos efectos estacionales encontramos una correspondencia cercana entre los modelos de EE y de BJ. Un término de pendiente β_t puede agregarse a (70) como se hizo en (69) sin que afecte las conclusiones.

Un patrón comienza a emerger. Partiendo de las predicciones de PMEP, que mostraron que en circunstancias apropiadas trabajan bien en la práctica, encontramos que hay dos tipos de modelos distintos, los modelos de EE y los modelos de BJ, los que aparentan ser conceptualmente muy diferentes pero ambos dan MEMCP a partir de las recursiones PMEP. La explicación es que cuando la serie de tiempo tiene una estructura subyacente que es suficientemente simple, los modelos apropiados de EE y de BJ son esencialmente equivalentes. Es realmente cuando nos movemos a estructuras más complejas cuando las diferencias emergen. En la sección siguiente se comparan los enfoques de EE y de BJ ante un conjunto de problemas del análisis de series de tiempo más amplios.

8. Comparación de los enfoques de espacio de estado y de Box-Jenkins para el análisis de series de tiempo

Los desarrollos iniciales de la metodología de EE tuvieron lugar en el campo de la ingeniería y no en la estadística, comenzando con el importante trabajo de Kalman (1960). Kalman realizó dos cosas crucialmente importantes en su trabajo. Mostró que una clase muy amplia de problemas puede ser encapsulado en un modelo lineal simple, esencialmente el modelo de EE dado en (27). Luego mostró cómo, debido a la naturaleza Markoviana del modelo, los cálculos necesarios para la aplicación práctica pueden ser desarrollados de una manera recursiva muy conveniente para la computación. Una gran cantidad de trabajo se realizó a partir de allí para desarrollar esas ideas en el campo de la ingeniería.

Durante los años '60 y hasta los '80, las contribuciones de los estadísticos a la metodología de EE fueron aisladas y esporádicas. No es mi interés realizar aquí una revisión histórica de esas contribuciones. Los nombres de Jones, Akaike, Duncan, Harrison, Stevens, Kitagawa, Gersch y Harvey son importantes. Andrew Harvey merece una mención particular, no sólo por sus contribuciones en las investigaciones dentro de este campo sino también por su libro (Harvey, 1989) y su liderazgo en el proyecto de software conocido como STAMP (Structural Time series Analyser, Modeller and Predictor). Harvey (1989) describe como fue el estado del arte alrededor de 1988 en el análisis de series de tiempo mediante el enfoque de EE, y también provee referencias históricas que voy a omitir.

La ventaja clave del enfoque de EE es que está basado en el análisis estructural del problema. Los diferentes componentes que hacen a la serie, tales como tendencia, estacionalidad, ciclo y variaciones calendarias, junto con los efectos de variables explicativas e intervenciones, son modelados de forma separada antes de ponerlos conjuntamente en el modelo de EE. Es de la responsabilidad del investigador identificar y modelar cualquier rasgo que requiera especial tratamiento en situaciones particulares. En contraste, el enfoque de BJ es una "caja negra" en la cual el modelo adoptado depende puramente de los datos sin ningún análisis previo de la estructura del sistema que generó los datos.

Una segunda ventaja de los modelos de EE es que son muy flexibles. Debido a su naturaleza recursiva y a las técnicas computacionales usadas para analizarlos,

es inmediata la posibilidad de incorporar dentro de la estructura del sistema los cambios conocidos que suceden en el tiempo. Por el contrario, los modelos de BJ son homogéneos a través del tiempo porque están basados en el supuesto que la serie diferenciada es estacionaria.

Los modelos de EE son muy generales. Cubren una amplia gama, incluyendo a todos los modelos de BJ. Las observaciones multivariadas pueden ser manejadas mediante una extensión directa de la teoría univariada, lo que no sucede con los modelos de BJ.

Es muy fácil incorporar ciclos y observaciones perdidas a los modelos de EE; además las variables explicativas pueden ser agregadas a ellos sin ninguna dificultad. Más aún, los coeficientes de regresión asociados pueden variar estocásticamente en el tiempo si ello fuera necesario en las aplicaciones. Ajustes por días laborables y otras variaciones calendarias pueden fácilmente ser incorporadas.

Debido a la naturaleza Markoviana de los modelos de EE, los cálculos necesarios para su implementación pueden ser puestos de una manera recursiva. Esto permite que grandes modelos puedan ser efectivamente manejados sin incrementos desproporcionados en la carga computacional.

No se necesita teoría adicional para las predicciones. Todo lo que se requiere es proyectar el filtro de Kalman hacia adelante en el futuro. Esto da las predicciones pertinentes y sus errores estándares estimados mediante la utilización de fórmulas ya usadas con anterioridad para la serie.

Si estas son las ventajas del modelado de EE, uno se pregunta cuáles son sus desventajas relativas al enfoque de BJ. En mi opinión, la única desventaja es la falta de información, conocimiento y software relativos a estos modelos dentro de la comunidad estadística. El modelado de BJ forma parte integrante de la mayoría de los cursos típicos de series de tiempo ofrecidos en las universidades, existiendo numerosos libros de textos sobre el mismo. El software está disponible en la mayoría de los paquetes generales tales como el SAS, BMDP y MINITAB y en los paquetes especializados en series de tiempo, tal como el ITSM. Por el contrario, el modelado de EE se enseña en muy pocas universidades en el mundo, y desde el punto de vista estadístico, en contraste con el punto de vista de la ingeniería, hay solamente cuatro libros generales sobre el tema que son Harvey (1989), West y Harrison (1989), Abril (1999) y Durbin y Koopman (2001). Ningún paquete estadístico general tiene software sobre EE, y hay muy pocos paquetes especializados tales como STAMP (Structural Time series Analyser, Modeller and Predictor) de Koopman, Harvey, Doornik y Shephard (1995 y 2000).

Ahora consideremos algunas de las desventajas del enfoque de BJ. La eliminación de la tendencia y la estacionalidad mediante diferenciación puede no ser un inconveniente si la predicción es el único objetivo del análisis, pero en muchos contextos, particularmente en estadísticas oficiales, el conocimiento de esos componentes tiene importancia intrínseca. Es cierto también que se pueden “recuperar” las estimaciones de la tendencia y de la estacionalidad a partir de la serie diferenciada

maximizando la media cuadrática residual, pero este parece ser un procedimiento artificial que no es tan atractivo como modelar los componentes directamente.

El requerimiento que la serie diferenciada debe ser estacionaria es una debilidad de la teoría. En los campos de la economía y las ciencias sociales, las series nunca son estacionarias, no obstante toda la diferenciación que se haga. El investigador debe enfrentar la cuestión siguiente: ¿cuán cerca de la estacionariedad es suficientemente cerca? Esta pregunta es muy difícil de responder. Hay otro punto que está íntimamente relacionado con este. Al comienzo se supone que las observaciones están formadas por los términos tendencia, estacionalidad e irregular. El término irregular se supone que se comporta de acuerdo a algún modelo estocástico. Encuentro una coincidencia a remarcar que la cantidad de diferenciación requerida para eliminar los términos de tendencia y estacionalidad es exactamente la misma que se necesita para lograr un adecuado modelo ARMA estacionario para el término irregular diferenciado.

En el sistema de BJ es relativamente difícil de manejar cuestiones tales como observaciones perdidas, el agregado de variables explicativas, ajustes calendarios y cambios en el comportamiento a través del tiempo. Estas son cosas que se pueden tratar de una manera directa en el sistema de EE.

En la práctica se encontró que el modelo de aerolínea y otros modelos similares de BJ ajustan bastante bien a muchos conjuntos de datos, pero como se dijo anteriormente, ellos son aproximadamente equivalentes a modelos plausibles de EE. Este punto es discutido con amplitud en Harvey (1989, pp. 72, 73). A medida que nos apartamos de los modelos de tipo aerolínea, el proceso de identificación se hace cada vez más difícil de aplicar en el sistema de BJ. La principal herramienta es la función de autocorrelación muestral la que es notoriamente imprecisa debido a su alta variabilidad muestral. Se pueden encontrar ejemplos en los cuales los datos aparecen explicados igualmente bien por modelos cuyas especificaciones son muy diferentes.

Un punto final a favor de los modelos estructurales es su transparencia. Uno puede examinar los gráficos de la tendencia, la estacionalidad y de los otros componentes para comprobar si se comportan de acuerdo a lo esperado. En caso de encontrar algún punto extraño, por ejemplo en el gráfico de la estacionalidad, se podría ir a los datos originales a fin de rastrear y quizás hacer los ajustes correspondientes al modelo.

Para resumir, los modelos de EE están basados en el modelado de la estructura observada de los datos. Son más generales, flexibles y transparentes que los modelos de BJ. Pueden manejar situaciones presentes en los datos que son muy difíciles de manejar con los sistemas de BJ. El software está disponible en el paquete STAMP para el análisis de EE de las series de tiempo.

Una presentación completa, detallada y actual de los métodos y modelos para el análisis de las series de tiempo, tanto bajo el enfoque de BJ como de EE, y

también en el dominio del tiempo como en las frecuencias, puede encontrarse en Abril (2004).

9. Algunos trabajos recientes en modelado de EE de series de tiempo

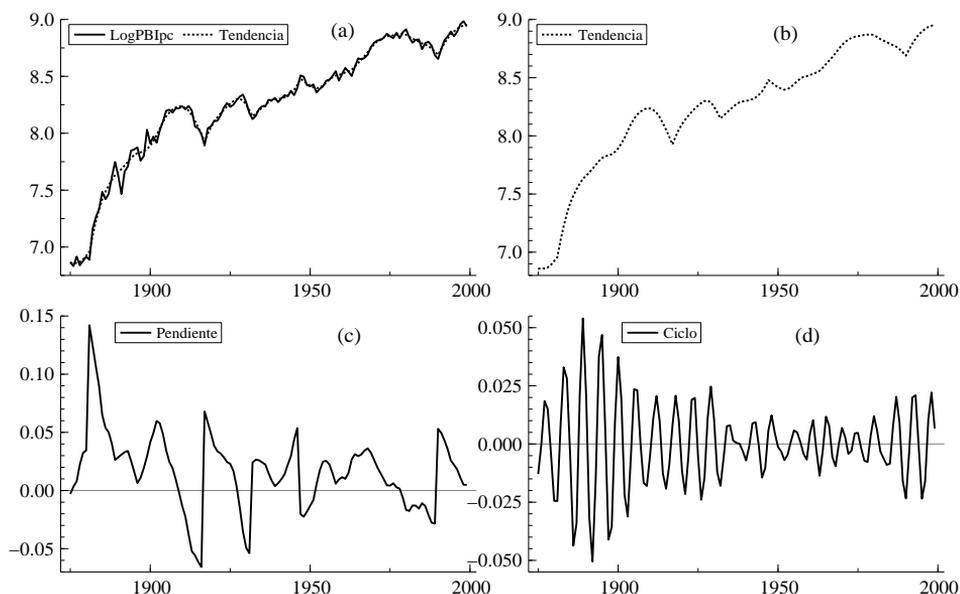
En esta sección se presentaran algunos trabajos recientes sobre el modelado de EE de series de tiempo.

En un trabajo, Abril y Blanco (2002) estudian los “hechos estilizados” del Producto Bruto Interno total (PBI_{total}) y del Producto Bruto Interno per cápita (PBI_{pc}) de la Argentina para el período 1875–1999. Establecer los “hechos estilizados” asociados con un conjunto de series de tiempo es considerado un paso crucial en investigación económica. Para que esos hechos sean útiles deben (i) ser consistentes con las propiedades estocásticas de los datos, y (ii) presentar información significativa. Ahora bien, muchos hechos estilizados citados en la literatura no satisfacen esos criterios. Particularmente, si el trabajo está basado en series a las que se les eliminó la tendencia de una manera mecánica puede producir erróneamente la impresión de tener un comportamiento cíclico falso. El análisis basado en los sistemas de BJ puede ser confuso si los modelos respectivos son elegidos principalmente sobre la base de la parsimonia.

Se sabe que los modelos estructurales de series de tiempo proveen los marcos más útiles dentro de los cuales se pueden presentar los hechos estilizados de las series de tiempo. Estos modelos estructurales se basan explícitamente en las propiedades estocásticas de los datos. Ellos brindan información significativa sobre los componentes de la serie y ayudan a la comprensión de los mismos.

En el análisis de los hechos estilizados del PBI_{total} y del PBI_{pc}, Abril y Blanco (2002) identificaron y estimaron un modelo con nivel fijo, pendiente estocástica y ciclo, con algunos cambios estructurales en la pendiente y observaciones atípicas (“outliers”). Los cambios estructurales y los outliers fueron identificados en coincidencia con hechos y acontecimientos concretos de la vida económica de la Argentina, y para ambas series el ciclo encontrado tiene un período de 5 años y 7 meses. En la Figura 7 se presentan los componentes de la serie PBI_{pc}. Es claro, a partir de la observación de esa figura, que el método de EE es adecuado para el estudio de los hechos estilizados de las series de tiempo. Por otra parte, Abril, Ferullo y Gaínza Córdoba (1999) estudian la relación entre los cambios en la tasa de desempleo y el crecimiento económico de Argentina en el período 1980–1996. Para ello plantean, identifican y estiman un modelo en donde los cambios en la tasa de desempleo son explicados mediante una tendencia estocástica, movimientos estacionales estocásticos, cambios estructurales en la tendencia, outliers y el crecimiento como variable explicativa. Al crecimiento se lo define como la tasa de cambio en el Producto Bruto Interno. En la Figura 8 se presentan los componentes de la serie de primeras diferencias de la tasa de desempleo. Es importante destacar que el sistema de EE resultó ser el más adecuado para realizar este análisis, el cual en términos económicos, es una generalización de la llamada

Figura 7. PBIpc de Argentina 1875–1999 con tendencia [(a) y (b)], la pendiente (c) y el ciclo (d)

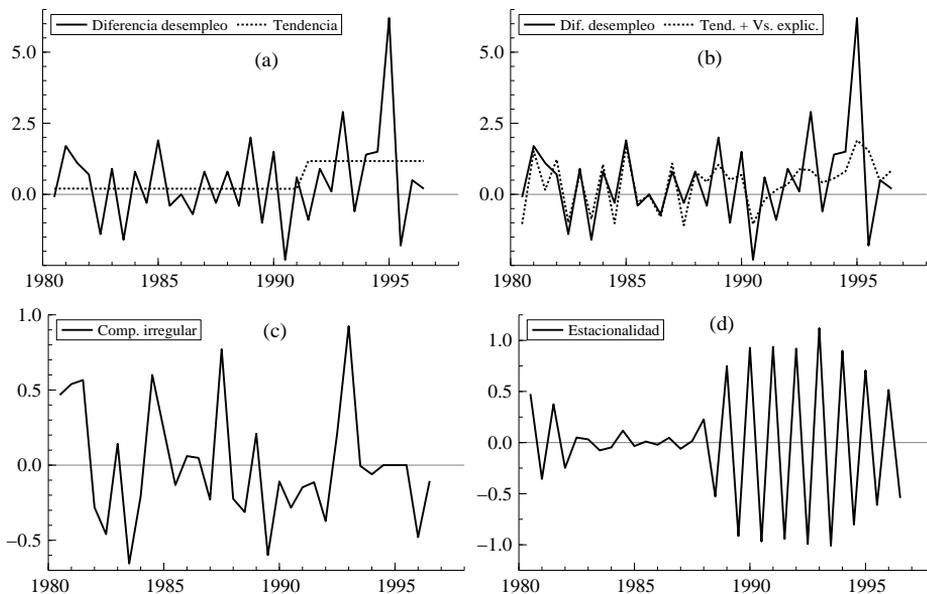


“Ley de Okun”. En un trabajo de 1997, Durbin y Quenneville (1997) analizan el problema del “benchmarking” para las ventas minoristas de Canadá. Benchmarking es muy importante en estadísticas oficiales y consiste en lo siguiente: se tiene una serie mensual o trimestral de observaciones (por simplicidad, supongamos que es mensual) obtenida a partir de una muestra. Esas observaciones están sujetas a sesgo y errores mensuales. Por otra parte, se tiene una serie de valores anuales obtenida a partir de datos administrativos, censos o encuestas anuales, la que se considera más precisa que la correspondiente serie mensual. Los valores anuales pueden ser totales del año o valores de un mes particular. El problema es ajustar las observaciones de tal manera que los valores mensuales y anuales sean consistentes.

Hillmer y Trabelsi (1987) desarrollan un procedimiento en dos etapas; en la primera ajustan un modelo de BJ a los datos muestrales y en la segunda lo adicionan en los datos anuales.

Durbin y Quenneville (1997) realizan un tratamiento de EE para el problema. Usando el enfoque de EE es posible dar solución a diferentes problemas accesorios. Por ejemplo, se pueden incorporar los efectos de los días laborables y de los años bisiestos de una manera inmediata. Un segundo problema es que los datos

Figura 8. Primeras diferencias de la tasa de desempleo. Serie original [línea llena de (a) y (b)], tendencia estimada [línea de puntos de (a)], tendencia más variables explicativas [línea de puntos de (b)], componente irregular (c) y estacionalidad (d)



muestrales están sesgados por “no respuestas” y otros factores. Suponiendo que los valores anuales son insesgados, se incorpora un coeficiente en el modelo de tal manera que permite estimar el sesgo muestral. Un tercer problema es que los datos económicos se comportan, frecuentemente, de forma multiplicativa. Una solución usual es tomar logaritmos antes de analizar los datos. Por otra parte, la suma de valores mensuales para obtener totales anuales es una operación lineal. Entonces, la combinación de datos multiplicativos y totales anuales en el mismo modelo produce un sistema que es no-lineal. Estimando el modo del vector de estado dadas todas las observaciones, mensuales y anuales, le permitió a Durbin y Quenneville (1997) ajustar el modelo, asegurando que los valores mensuales satisfacen las restricciones anuales.

La mayoría de las técnicas de EE están basadas en el supuesto que el modelo sea Gaussiano y lineal. En el mundo real, muchas observaciones están muy lejos de la normalidad y la falta de linealidad es muy frecuente. Hay por lo tanto una necesidad por desarrollar técnicas para el tratamiento de estos casos. En Abril (1999, Cap. 4) y en Durbin y Koopman (1997, 2000) se presentan métodos de EE para el análisis de observaciones no-Gaussianas y para el tratamiento de modelos no-lineales.

10. Uso potencial en estadísticas oficiales

En esta sección se considera la potencialidad que los métodos de EE pueden tener para el trabajo de series de tiempo en las estadísticas oficiales. Para el trabajo cotidiano e individual en el análisis de series de tiempo, se mostró en §8 que los modelos de EE ofrecen mayor alcance, flexibilidad y relevancia para las necesidades prácticas que los modelos de BJ.

En §9 se presentaron ejemplos de aplicación de los sistemas de EE para el tratamiento de observaciones que típicamente corresponden al ámbito de las estadísticas oficiales. Se mostró también allí la existencia de elementos dentro de los modelos de EE para el tratamiento de datos no-Gaussianos y no-lineales.

El uso más importante del análisis de series de tiempo en estadísticas oficiales es, por lejos, en ajuste estacional. En la mayoría de las agencias gubernamentales (Institutos o Departamentos de Estadísticas Oficiales), este procedimiento es realizado mediante alguna variante del procedimiento X-11, X-12 o por alguna técnica de suavizado similar de base intuitiva. La mayoría de los especialistas académicos en series de tiempo se sorprendieron cuando vieron por primera vez estos métodos. La opinión fue que alguien debía producir una técnica basada en modelos y en ideas estadísticas modernas que superara cualquier procedimiento de base intuitiva. No obstante, la metodología básica del X-11 o similares fue desarrollada hace más de 40 años y no ha sido cambiada en su esencia hasta el momento. Debemos admirar el éxito de aquellos estadísticos que han creado un método que ha sido tan popular por tanto tiempo y que ha sobrevivido toda una revolución en la computación.

Parte de la explicación es que ni la tendencia ni la estacionalidad pueden ser objetivamente definidas. ¿Cuán suave debe ser una tendencia? Ningún criterio objetivo puede decidir esto; en última instancia la respuesta debe estar determinada por el consumidor. ¿Cuán rápido debe la estructura estacional cambiar en el tiempo? La respuesta en este caso es la misma. Parece que quienes diseñaron el X-11 deben haber logrado las respuestas casi adecuadas, presumiblemente mediante algún procedimiento de prueba y error unido al ejercicio del discernimiento.

De cualquier manera, el X-11 y sus derivados tales como el X-12 no son, de ninguna manera, perfectos. En particular, no hacen un uso eficiente de los datos más recientes que se encuentran al final de las series. En este punto es importante citar algunos resultados logrados por Maravall (1996), los cuales son: (i) el X-11 puede extraer componentes estacionales falsos, incluso a series de observaciones que son ruido blanco; (ii) el X-11 subajustará cuando el ancho del pico estacional en el espectro de la serie es mayor que el capturado por el filtro del método X-11, y sobreajustará cuando el ancho del pico sea menor que el del filtro; (iii) el X-11 no posee un adecuado marco de detección de casos en los cuales su uso es inapropiado; (iv) aún cuando sea apropiado, el X-11 no contiene las bases para una adecuada inferencia; (v) el X-11 no permite calcular predicciones óptimas de los componentes; (vi) a pesar que el X-11 calcula estimaciones separadas de los

componentes tendencia, estacionalidad e irregular, sus propiedades estadísticas son desconocidas.

Con el desarrollo de los modelos de EE creo que se encuentra abierto el camino para la aplicación de métodos basados en modelos que hagan un uso eficiente de los datos. Se ha discutido anteriormente, en este trabajo, la flexibilidad y generalidad de estos modelos. Se mostró cómo se manejaron la falta de normalidad y la no-linealidad en muchas aplicaciones. Variables explicativas tales como los días laborales, variaciones calendarias y variables climáticas pueden ser acomodadas de forma relativamente fácil. El uso de estos modelos proveerá una mejor base para el análisis subsecuente debido a su sustento analítico.

La tarea de desarrollar modelos de EE para ajuste estacional será larga debido a que los modelos relativamente simples de tendencia y estacionalidad existentes en el modelado estructural de series de tiempo tienen poca posibilidad de ser aceptados de forma inmediata y sin una investigación adicional por los estadísticos oficiales. Un abanico de modelos deberá ser probado, de tal manera de asegurar que el suavizado adecuado de la tendencia y los cambios justos de la estacionalidad en el tiempo son alcanzados. Pero a pesar que la tarea es grande, se espera que alguien se vea persuadido de realizarla.

Nota

Este artículo está basado en una conferencia pronunciada en el XXIX Coloquio Argentino de Estadística, Santa Fe, Provincia de Santa Fe, Argentina, 12 al 14 de Octubre de 2011.

Referencias

ABRIL, J. C. (1987). "The approximate densities of some quadratic forms of stationary random variables." *Journal of Time Series Analysis*. **8**: 249–59.

ABRIL, J. C. (1997). "Series de tiempo irregulares: un enfoque unificado". Conferencia invitada pronunciada durante el XXV Coloquio Argentino de Estadística. Sociedad Argentina de Estadística. Noviembre de 1997.

ABRIL, J. C. (1999). *Análisis de Series de Tiempo Basado en Modelos de Espacio de Estado*. EUDEBA, Buenos Aires.

ABRIL, J. C. (2000). "James Durbin. Más allá de los tests de correlación serial". Reporte técnico. Facultad de Ciencias Económicas de la Universidad Nacional de Tucumán.

ABRIL, J. C. (2001a). "On time series of observations from exponential family distributions." *Pakistan Journal of Statistics*. **17**: 235–48.

- ABRIL, J. C. (2001b). “El enfoque de espacio de estado para el análisis de las series de tiempo - Comparaciones.” *Revista de la Sociedad Argentina de Estadística*. **5**: 1–16.
- ABRIL, J. C. (2002). “Outliers, structural shifts and heavy-tailed distributions in state space time series models.” *Pakistan Journal of Statistics*. **18**: 25–43.
- ABRIL, J. C. (2004). *Modelos para el Análisis de las Series de Tiempo*. Ediciones Cooperativas, Buenos Aires.
- ABRIL, J. C. y BLANCO, M. B. (2002). “Stylized Facts of the Gross National Product of Argentina: 1875–1999.” *Anales de la XXXVII Reunión Anual de la Asociación Argentina de Economía Política*.
- ABRIL, J. C., FERULLO, H. D. y CÓRDOBA, A. G. (1998). “Estimación de la relación de Okun: Argentina 1980–1996.” *Anales de la XXXIII Reunión Anual de la Asociación Argentina de Economía Política*.
- ABRIL, J. C., FERULLO, H. D. y CÓRDOBA, A. G. (1999). “Tendencias estocásticas en modelos de regresión dinámica: una aplicación a la relación desempleo-crecimiento en la Argentina.” *Revista de la Sociedad Argentina de Estadística*. **3**, 9–35.
- AKAIKE, H. (1969). “Power spectrum estimation through autoregressive model fitting.” *Annals of the Institute of Statistical Mathematics*. **21**: 407–19.
- AKAIKE, H. (1974). A new look at the statistical model identification.” *IEEE Transactions on Automatic Control*. **AC-19**: 716–22.
- ANDERSON, B. D. O. y MOORE, J. B. (1979).” *Optimal Filtering*. Prentice-Hall, Englewood Cliffs, NJ.
- BALKE, N. S. (1993). “Detecting level shifts in time series.” *Journal of Business and Economics Statistics*. **11**: 81–92.
- BARTLETT, M. S. (1950). “Periodogram analysis and continuous spectra.” *Biometrika*. **37**: 1–16.
- BOX, G. E. P. y JENKINS, G. M. (1976). *Time Series Analysis: Forecasting and Control* (Revised edition). Holden-Day, Inc, San Francisco, CA.
- BOX, G. E. P. y PIERCE, D. A. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models.” *Journal of the American Statistical Association*. **65**: 1509–26.
- BROCKWELL, P. J. y DAVIS, R. A. (1991). *Time Series: Theory and Methods*, second edition. Springer-Verlag, New York.

- BROCKWELL, P. J. y DAVIS, R. A. (2000). *ITSM 2000: Interactive Time Series Modelling Package for the PC*, version 6.0. Springer-Verlag, New York.
- BURG, J. P. (1967). "Maximum entropy spectral analysis." Paper presented at the 37th Annual International SEG Meeting, Oklahoma City, OK.
- BURG, J. P. (1972). "The relationship between maximum entropy spectra and maximum likelihood spectra." *Geophysics*. **37**: 375–6.
- CAPON, J. (1969). "High resolution frequency-wave number spectral analysis." *Proceedings of the IEEE*. **57**: 1408–18.
- CARLSTEIN, E. (1988). "Non-parametric change point estimation." *Biometrika*. **75**: 188–97.
- COBB, G. W. (1978). "The problem of the Nile: conditional solution to a change point problem." *Biometrika*. **65**: 243–51.
- DANIELL, P. J. (1946). "Discussion on 'Symposium on autocorrelation in time series'." *Journal of the Royal Statistical Society, Supplement*. **8**: 88–90.
- DE JONG, P. (1989). "Smoothing and interpolation with the state-space model." *Journal of the American Statistical Association*. **84**: 1085–88.
- DE JONG, P. y PENZER, J. R. (1998). "Diagnosing shocks in time series." *Journal of the American Statistical Association*. **93**: 796–806.
- DUNCAN, D. B. y HORN, S. D. (1972). "Linear dynamic recursive estimation from the viewpoint of regression analysis." *Journal of the American Statistical Association*. **67**: 815–21.
- DURBIN, J. y KOOPMAN, S. J. (1997). "Monte Carlo maximum likelihood estimation for non-Gaussian state space models." *Biometrika*. **84**: 669–84.
- DURBIN, J. y KOOPMAN, S. J. (2000). "Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives (with discussion)." *Journal of the Royal Statistical Society, B*. **62**: 3–56.
- DURBIN, J. y KOOPMAN, S. J. (2001). *Time Series Analysis by State Space Methods*. Oxford University Press, Oxford.
- DURBIN, J. y QUENNEVILLE, B. (1997). "Benchmarking by state space models." *International Statistical Review*. **65**: 23–98.
- GRANGER, C. W. J. y ANDERSON, A. P. (1978). *An Introduction to Bilinear Time Series Models*. Vandenhoeck and Ruprecht, Göttingen.

GRENANDER, U. y ROSEMBLATT, M. (1957). *Statistical Analysis of Stationary Time Series*. Wiley, New York.

HAGGAN, V. y OZAKI, T. (1979). "Amplitude-dependent AR model fitting for non-linear random vibrations." Paper presented at the International Time Series Meeting, University of Nottingham, UK, March 1979.

HARVEY, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press, Cambridge.

HARVEY, A. C. y DURBIN, J. (1986). "The effects of seat belt legislation on British road casualties: a case study in structural time series modelling (with discussion)." *Journal of the Royal Statistical Society, A*. **149**: 187–227.

HARVEY, A. C. y SHEPHARD, N. (1993). "Structural time series models" en *Handbook of Statistics, Vol. 11: Econometrics* (G. S. Maddala, C. R. Rao y H. D. Vinod, Eds.), 261–302.

HILLMER, C. H. y TRABELSI, A. (1987). "Benchmarking of economic time series." *Journal of the American Statistical Association*. **82**: 1064–71.

HOLT, C. C. (1957). "Forecasting seasonals and trends by exponentially weighted moving average." ONR Research Memorandum **52**. Carnegie Institute of Technology.

HURVICH, C. M. y TSAI, C. L. (1989). "Regression and time series model selection in small samples." *Biometrika*. **76**: 297–307.

JONES, R. H. (1975). "Fitting autoregressions." *Journal of the American Statistical Association*., **70**: 590–2.

JONES, R. H. (1993). *Longitudinal Data with Serial Correlation: A State-Space Approach*, Chapman and Hall, London.

KALMAN, R. E. (1960). "A new approach to linear filtering and prediction problems." *Transactions of the ASME. Journal of Basic Engineering*. **82D**: 35–45.

KALMAN, R. E. y BUCY, R. S. (1961). "New results in linear filtering and prediction problems." *Transactions of the ASME. Journal of Basic Engineering*. **83D**: 95–108.

KOLMOGOROV, A. (1941). "Interpolation und extrapolation von stationären Zufälligen Folgen." *Bulletin of the Russian Academy of Sciences: Mathematics*. **5**: 3–14.

KOOPMAN, S. J. (1993). "Disturbance smoother for state space models." *Biometrika*. **80**: 117–26.

- KOOPMAN, S. J., HARVEY, A. C., DOORNIK, J. A. y SHEPHARD, N. (1995). *STAMP 5.0: Structural Time Series Analyser, Modeller and Predictor*, Chapman and Hall, London.
- KOOPMAN, S. J., HARVEY, A. C., DOORNIK, J. A. y SHEPHARD, N. (2000). *STAMP 6.0: Structural Time Series Analyser, Modeller and Predictor*, Timberlake Consultants, London.
- KOOPMAN, S. J., SHEPHARD, N. y DOORNIK, J. A. (1999). "Statistical algorithms for models in state space form using SsfPack 2.2." *Econometrics Journal*. **2**: 113–66. <http://www.ssfpack.com/>.
- LACOSS, R. T. (1971). "Data adaptive spectral analysis methods." *Geophysics*. **36**: 661–75.
- LJUNG, G. M. y BOX, G. E. P. (1978). "On a measure of lack of fit in time series models." *Biometrika*. **65**: 297–303.
- MANN, H. B. y WALD, A. (1943). "On the statistical treatment of linear stochastic difference equations." *Econometrica*. **11**: 173–220.
- MARAVALL, A. (1996). "Short-term analysis of macroeconomic time series. Documento de Trabajo N° 9607, Servicio de Estudios, Banco de España.
- MUTH, J. F. (1960). "Optimal properties of exponentially weighted forecasts." *Journal of the American Statistical Association*. **55**: 299–305.
- PARZEN, E. (1957). "On consistent estimates of the spectrum of a stationary time series." *Annals of Mathematical Statistics*. **28**: 329–48.
- PARZEN, E. (1974). "Some recent advances in time series modelling." *IEEE Transactions on Automatic Control*. **AC-19**: 723–9.
- PÉREZ DE DEL NEGRO, M. A. (2000). "Modelos estructurales de series de tiempo. Precipitaciones en San Miguel de Tucumán (1884–1993)." Trabajo presentado en el XXVIII Coloquio Argentino de Estadística, Posadas, Misiones, Argentina, Agosto de 2000.
- PRIESTLEY, M. B. (1965). "Evolutionary spectra and non-stationary processes." *Journal of the Royal Statistical Society, B*. **27**: 204–37.
- PRIESTLEY, M. B. (1980). "State dependent models: a general approach to non-linear time series analysis." *Journal of Time Series Analysis*. **1**: 41–71.
- PRIESTLEY, M. B. (1981). *Spectral Analysis and Time Series*, vols I, II. Academic Press, London.

- SCHUSTER, A. (1898). "On the investigation of hidden periodicities with application to a supposed 26-day period of meteorological phenomena." *Terrestrial Magnetism and Atmospheric Electricity*, **3**: 13–41.
- SUBBA RAO, T. (1970). "The fitting of non-stationary time series with time-dependent parameters." *Journal of the Royal Statistical Society, B*, **32**: 312–22.
- SUBBA RAO, T. (1981). "On the theory of bilinear time series models." *Journal of the Royal Statistical Society, B*, **43**: 244–55.
- SUBBA RAO, T. y GABR, M. M. (1984). *An Introduction to Bispectral Analysis and Bilinear Time Series Models*. Springer-Verlag, Berlin.
- THEIL, H. y WAGE, S. (1964). "Some observations on adaptive forecasting." *Management Science*, **10**: 198–206.
- TONG, H. y LIM, K. S. (1980). "Threshold autoregression, limit cycles, and cyclical data." *Journal of the Royal Statistical Society, B*, **42**: 245–92.
- TUKEY, J. W. (1949). "The sampling theory of power spectrum estimates." *Proc. Symp. on Applications of Autocorrelation Analysis to Physical Problems*, NAVEXOS-P-735, 46–67. Office of Naval Research, Department of the Navy, Washington, DC.
- WEST, M. y HARRISON, P. J. (1989). *Bayesian Forecasting and Dynamic Models*. Springer-Verlag, New York.
- WETHERILL, G. B. (1977). *Sampling Inspection and Quality Control* (2nd. edn.). Chapman and Hall, London.
- WIENER, N. (1949). *The Extrapolation, Interpolation and Smoothing of Stationary Time Series with Engineering Applications*. Wiley, New York.
- WINTERS, P. R. (1960). "Forecasting sales by exponentially weighted moving average." *Management Science*, **6**: 324–42.
- WOLD, H. O. A. (1949). "A large sample test of moving averages." *Journal of the Royal Statistical Society, B*, **11**: 297.
- YULE, G. U. (1927). "On a method of investigating periodicities in disturbed series with special reference to Wolfer's sunspot numbers." *Philosophical Transactions of the Royal Society of London, A*, **226**: 267–98.

Invited Paper

Received November 2011

Revised March 2012