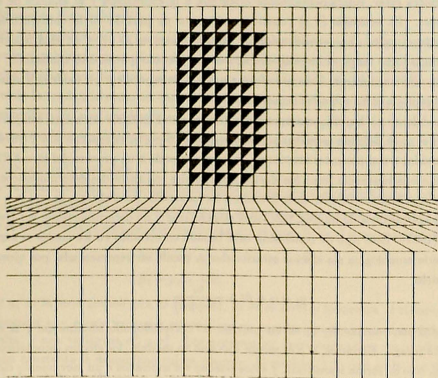


**SEXTA
JORNADA DE
MATEMATICA**

OCTUBRE 24 - 25 - 26 / 1990
TEMUCO



**MODELOS ESTRUCTURALES
PARA SERIES CRONOLOGICAS**

(CURSO B3)

PABLO MARSHALL

1 INTRODUCCION

1.1 Conceptos Básicos

Tradicionalmente se ha reconocido que, quizás después de alguna transformación, una serie cronológica y_t puede descomponerse en una tendencia μ_t , un efecto estacional γ_t y una irregularidad ϵ_t . Esto es,

$$y_t = \mu_t + \gamma_t + \epsilon_t, \quad t = 1, \dots, T. \quad (1.1)$$

La idea central de los modelos estructurales se basa en el hecho de que, en muchas aplicaciones, la tendencia y los efectos estacionales conforman los aspectos más sobresalientes en las series cronológicas. En este contexto, la construcción de un modelo está, o debiera estar, fundamentalmente orientada a la estimación y al análisis de estas componentes. La ecuación (1.1) constituye el punto de partida en la construcción de modelos estructurales. La especificación de las componentes μ_t , γ_t y ϵ_t , está basada en el conocimiento que tenga el investigador acerca del proceso que se esté analizando, y en técnicas estadísticas básicas. Esta es la forma natural de especificar un modelo, y coincide con la utilizada en otros métodos estadísticos.

El modelo (1.1) ha sido extensamente usado en el análisis de series temporales. Típicamente, μ_t es definido como un polinomio en el tiempo, γ_t es especificada a través de variables mudas o componentes armónicas, y ϵ_t corresponde a un proceso del tipo ARMA. Es decir, tanto la tendencia como los efectos estacionales son asumidos determinísticos. La evidencia empírica más reciente en la literatura ha mostrado, sin embargo, que μ_t y γ_t son, en muchas aplicaciones de interés, funciones estocásticas y no determinísticas; ver Nelson y Plosser (1982). Este tipo de conclusión está basada en estudios en los cuales se supone que una serie cronológica sin efectos estacionales y_t puede ser representada, por ejemplo, por el modelo

$$y_t = \alpha + \beta t + \frac{u_t}{(1 - \rho L)},$$

donde α y β son constantes, u_t es un proceso no correlacionado serialmente y L es el operador de rezago. Entonces, si ρ es menor que uno en modulo, los efectos estocásticos u_t afectan a y_t sólo en forma transitoria, y estos pueden ser asociados a un movimiento cíclico o irregular en la serie. Sin embargo, si ρ es igual a uno, entonces los efectos estocásticos tienen un efecto permanente sobre y_t y esto es justamente lo que se entiende por tendencia. Muchos de los estudios empíricos realizados en los últimos años, con modelos como el anterior o con obvias modificaciones para incorporar efectos estacionales, han llevado a la conclusión que, en muchas series cronológicas, la tendencia y los efectos estacionales son funciones estocásticas y no determinísticas.

La clase de modelos ARIMA, ha sido extensamente usada para el análisis de series cronológicas. Sin embargo, esta clase de modelos no está exenta de críticas. En primer lugar, y considerando lo expuesto anteriormente, la idea básica de los modelos ARIMA de diferenciar la serie para obtener estacionaridad contradice el objetivo primero del análisis por cuanto no permite la estimación de la tendencia y de los efectos estacionales. En segundo lugar, la identificación de modelos a partir de la función de autocorrelación es usualmente muy difícil. Especialmente cuando se dispone de pocas observaciones, valores

extremos, u otro tipo de irregularidades.

Otro tipo de modelos utilizados para el análisis de series cronológicas están basados en métodos automáticos; por ejemplo, autoregresiones. No pocas veces, los resultados son desastrosos; ver Jenkins (1982) y Harvey y Todd (1983).

Los modelos estructurales se desarrollan a partir de los trabajos de Muth (1960), Holt (1957) y Winters (1960). Esto es, con el suavizamiento exponencial y sus extensiones. La forma de estado y el filtro de Kalman permiten formalizar estas especificaciones desde un punto de vista estadístico, y extender fácilmente estos modelos a numerosas situaciones de interés en aplicaciones : variables exógenas, datos faltantes, datos irregularmente espaciados, etc... Referencias más recientes en el desarrollo de los modelos estructurales son Harrison y Stevens (1976), Engle (1978), Kitagawa (1981) y Harvey y Durbin (1986). Un buen resumen de todos estos trabajos es presentado en Harvey (1989).

1.2 Modelos Estructurales

En esta subsección se presentan cuatro modelos estructurales que constituyen las herramientas básicas para el modelamiento de series cronológicas. Los dos primeros modelos asumen que la serie cronológica no tiene efectos estacionales. En el primero se modela la tendencia como un camino aleatorio, mientras que en el segundo, se agrega a esta especificación una pendiente que también está representada por un camino aleatorio. Estos dos modelos son luego generalizados para incluir efectos estacionales.

Modelo Local Constante

Una forma simple de modelar una tendencia en series cronológicas en las cuales el nivel cambia lentamente a través del tiempo, es mediante un camino aleatorio. Este tipo de series temporales pueden ser representadas por el modelo local constante definido como

$$\begin{aligned} y_t &= \mu_t + \epsilon_t, & t = 1, \dots, T, \\ \mu_t &= \mu_{t-1} + \eta_t. \end{aligned} \tag{1.2}$$

Donde y_t es la serie cronológica al tiempo t , μ_t es el nivel o tendencia al tiempo t , y las componentes ϵ_t y η_t son procesos estocásticos no correlacionados entre si ni serialmente, normales, y con varianzas σ_ϵ^2 y σ_η^2 . Los procesos no correlacionados, con media cero, y varianza constante se denominan ruidos blancos. El modelo (1.2) está en lo que se conoce como la forma de estado con μ_t la variable de estado; ver Sección 1.3.

Si $\sigma_\eta^2 = 0$, entonces $u_t = u_{t-1} = u_0$, y el nivel o tendencia es fijo. Es decir, un caso especial de (1.2) es un modelo en el cual las observaciones se apartan de un nivel constante μ_0 a través de efectos aleatorios no correlacionados serialmente y representados por ϵ_t . La componente η_t , cuya varianza es en general distinta de cero, permite al nivel o tendencia evolucionar suavemente a través del tiempo.

El modelo local constante es, naturalmente, no estacionario debido a la presencia del camino aleatorio μ_t . Diferenciando las observaciones y_t se obtiene lo que se conoce como la forma estacionaria del modelo. Esta es,

$$(1 - L)y_t = \eta_t + (1 - L)\epsilon_t, \quad t = 2, \dots, T. \tag{1.3}$$

Esta forma estacionaria permite calcular la función de autocovarianzas de las primeras diferencias, $\gamma(r)$, $r > 0$. En efecto,

$$\begin{aligned} \gamma(0) &= \sigma_{\eta}^2 + 2\sigma_{\epsilon}^2 \\ \gamma(1) &= -\sigma_{\epsilon}^2 \\ \gamma(k) &= 0 \quad \text{si } k > 1. \end{aligned}$$

La función de autocovarianzas presentada define un modelo en la clase ARMA. Este modelo se conoce como la forma reducida del modelo estructural, y de los resultados anteriores es obvio que para el modelo local constante la forma reducida corresponde a un MA(1) de la forma

$$(1 - L)y_t = (1 - \theta L)a_t, \quad t = 2, \dots, T, \quad (1.4)$$

donde a_t es un ruido blanco y θ es un parametro restringido al intervalo $[0,1]$. Esta restricción aparece por cuanto la autocorrelación de orden uno es por definición menor o igual a cero. Igualando la función de autocovarianzas de (1.3) y (1.4), se establece que

$$\frac{\sigma_{\eta}^2}{\sigma_{\epsilon}^2} = \frac{(1 - \theta)^2}{\theta}.$$

Modelo Local Lineal

El modelo anterior es de limitada utilidad por cuanto muchas series temporales presentan una tendencia con pendiente. Usualmente, esta pendiente cambia lentamente a través del tiempo y puede también modelarse mediante un camino aleatorio. Series cronológicas con estas características pueden ser representadas por el modelo local lineal definido por

$$\begin{aligned} y_t &= u_t + \epsilon_t, \quad t = 1, \dots, T, \\ u_t &= u_{t+1} + \beta_{t-1} + \eta_t, \\ \beta_t &= \beta_{t-1} + \delta_t. \end{aligned} \quad (1.5)$$

Donde y_t es la observación al tiempo t , u_t es la tendencia al tiempo t y β_t es su pendiente. Por su parte, los procesos ϵ_t , η_t y δ_t son ruidos blancos no correlacionados, normales y con varianzas σ_{ϵ}^2 , σ_{η}^2 y σ_{δ}^2 . El modelo está en la forma de estado con u_t y β_t las variables de estado; ver Sección 1.3.

Si $\sigma_{\delta}^2 = \sigma_{\eta}^2 = 0$, $\beta_t = \beta_{t-1} = \beta_0$ y $u_t = u_0 + \beta_0 t$. Es decir, la tendencia es fija. Por otra parte, si $\sigma_{\delta}^2 = 0$ pero $\sigma_{\eta}^2 > 0$, entonces $u_t = u_{t-1} + \beta_0 + \eta_t$ y se muestra que el modelo local lineal puede reescribirse como

$$\begin{aligned} y_t &= u_t^* + \beta_0 t + \epsilon_t, \\ u_t^* &= u_{t-1}^* + \eta_t, \end{aligned}$$

que corresponde al modelo local constante con una variable exógena: el tiempo; ver Sección 6.

La forma estacionaria del modelo se obtiene al tomar segundas diferencias. Así,

$$(1 - L)^2 y_t = \delta_{t-1} + (1 - L)\eta_t + (1 - L)^2 \epsilon_t, \quad t = 3, \dots, T. \quad (1.6)$$

La función de autocovarianzas de las segundas diferencias de las observaciones y_t esta dada

por

$$\begin{aligned} \gamma(0) &= \sigma_\epsilon + 2\sigma_\eta^2 + 6\sigma_\epsilon^2 \\ \gamma(1) &= -\sigma_\eta^2 - 4\sigma_\epsilon^2 \\ \gamma(2) &= \sigma_\epsilon^2 \\ \gamma(k) &= 0 \quad \text{si } k > 2. \end{aligned}$$

De aquí se desprende que la forma reducida del modelo local lineal es un MA(2) de la forma

$$(1 - L)^2 y_t = (1 - \theta_1 L - \theta_2 L^2) a_t, \tag{1.7}$$

con a_t ruido blanco y θ_1, θ_2 parámetros que satisfacen, además de las condiciones usuales de invertibilidad, las restricciones

$$\begin{aligned} \frac{\sigma_\eta^2}{\sigma_\epsilon^2} &= -4 - \frac{\theta_1(1 - \theta_2)}{\theta_2} \geq 0, \\ \frac{\sigma_\delta^2}{\sigma_\epsilon^2} &= 2 - \frac{(\theta_1 + \theta_2)^2 + (1 - 2\theta_1)}{\theta_2} \geq 0. \end{aligned}$$

Modelo Local Constante con Estacionalidad

La inclusión de efectos estacionales es una forma natural de extender los dos modelos considerados hasta ahora. Si los efectos estacionales tienen período s , el modelo local constante con estacionalidad queda definido por

$$\begin{aligned} y_t &= \mu_t + \gamma_t + \epsilon_t, \quad t = 1, \dots, T, \\ \mu_t &= \mu_{t-1} + \eta_t, \\ \sum_{i=0}^{s-1} \gamma_{t-i} &= \omega_t, \end{aligned} \tag{1.8}$$

donde y_t es la observación al tiempo t , μ_t es su nivel o tendencia, y γ_t es el efecto estacional. Los efectos aleatorios ϵ_t, η_t y ω_t son ruidos blancos no correlacionados, normales y con varianzas $\sigma_\epsilon^2, \sigma_\eta^2$ y σ_ω^2 .

La racionalidad detrás de la especificación para el efecto estacional es la siguiente. Suponga primero que $\sigma_\omega^2 = 0$. Entonces la especificación para el efecto estacional en (1.8) implica que $\gamma_t = \gamma_{t-s}$ y por tanto

$$\gamma_t = \begin{cases} \gamma_0 & t = s, 2s, 3s, \dots \\ \gamma_{-1} & t = s-1, 2s-1, 3s-1, \dots \\ \vdots & \vdots \\ \gamma_{-s+2} & t = 2, s+2, 2s+2, \dots \\ -\gamma_0 - \gamma_{-1} - \dots - \gamma_{-s+2} & t = 1, s+1, 2s+1, \dots \end{cases}$$

que corresponde a la típica especificación de variables mudas utilizada en modelos de regresión. La componente estocástica ω_t da flexibilidad a los efectos estacionales para que evolucionen lentamente a través del tiempo.

El modelo (1.8) no está escrito en la forma de estado, lo que tiene importantes desventajas para el manejo de este mediante el filtro de Kalman; ver Subsección 2.1. Sin embargo, es posible reescribir el modelo en la forma de estado. Para esto se definen componentes auxiliares $\gamma_{1t}, \gamma_{2t}, \dots, \gamma_{s-2,t}$ tales que

$$\gamma_{1t} = \gamma_{t-1},$$

y

$$\gamma_{it} = \gamma_{i-1,t-1} = \gamma_{t-i}, \quad i = 2, \dots, s-2.$$

Con estas definiciones, la ecuación para los efectos estacionales puede escribirse como

$$\begin{pmatrix} \gamma_t \\ \gamma_{1t} \\ \vdots \\ \gamma_{s-2,t} \end{pmatrix} = \begin{pmatrix} -1 & -1 & \cdots & -1 \\ 1 & 0 & \cdots & 0 \\ & \ddots & & \vdots \\ 0 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} \gamma_{t-1} \\ \gamma_{1t-1} \\ \vdots \\ \gamma_{s-2,t-1} \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \omega_t,$$

y el modelo queda expresado en la forma de estado con las variables de estado definidas por el vector $(\mu_t, \gamma_t, \gamma_{1t}, \dots, \gamma_{s-2,t})'$; ver Sección 1.3.

La forma estacionaria del modelo local constante con efectos estacionales se obtiene al aplicar, a las observaciones y_t , una diferencia estacional. Esto es,

$$(1 - L^s)y_t = \frac{(1 - L^s)\eta_t}{(1 - L)} + \frac{(1 - L^s)\omega_t}{(1 + L + \dots + L^{s-1})} + (1 - L^s)\epsilon_t \tag{1.9}$$

$$= (1 + L + \dots + L^{s-1})\eta_t + (1 - L)\omega_t + (1 - L^s)\epsilon_t.$$

A partir de esta ecuación, la función de autocovarianzas de las diferencias estacionales está dada por

$$\begin{aligned} \gamma(0) &= s\sigma_\eta^2 + 2\sigma_\omega^2 + 2\sigma_\epsilon^2 \\ \gamma(1) &= (s-1)\sigma_\eta^2 - \sigma_\omega^2 \\ \gamma(2) &= (s-2)\sigma_\eta^2 \\ &\vdots \\ \gamma(s-1) &= \sigma_\eta^2 \\ \gamma(s) &= -\sigma_\epsilon^2 \\ \gamma(k) &= 0 \quad \text{si } k > s. \end{aligned}$$

Se desprende que la forma reducida del modelo local constante con efectos estacionales corresponde a un MA(s) con restricciones en los parámetros. Mientras un modelo MA(s) sin restricciones tiene (s + 1) parámetros funcionalmente independientes, el modelo local constante con estacionalidad tiene solamente 3.

Modelo Estructural Básico

Finalmente, agregando un efecto estacional al modelo local lineal se obtiene el modelo estructural básico. Si la estacionalidad tiene período s, este modelo está definido por

$$\begin{aligned} y_t &= \mu_t + \gamma_t + \epsilon_t, \quad t = 1, \dots, T, \\ \mu_t &= \mu_{t-1} + \beta_{t-1} + \eta_t, \\ \beta_t &= \beta_{t-1} + \delta_t, \\ \sum_{i=0}^{s-1} \gamma_{t-i} &= \omega_t, \end{aligned} \tag{1.10}$$

donde y_t es la observación al tiempo t, μ_t es la tendencia o nivel al tiempo t, β_t es la

pendiente de esta tendencia, y γ_t es el efecto estacional al tiempo t . Los procesos ϵ_t , η_t , δ_t y ω_t son ruidos blancos no correlacionados, normales y con varianzas σ_ϵ^2 , σ_η^2 , σ_δ^2 y σ_ω^2 .

El modelo se pone en la forma de estado mediante un procedimiento análogo al usado en el modelo local constante con estacionalidad. Esto es, definiendo componentes auxiliares $\gamma_{1t}, \gamma_{2t}, \dots, \gamma_{s-2,t}$.

La forma estacionaria del modelo estructural básico se obtiene al tomar, sobre las observaciones y_t , una diferencia regular y una diferencia estacional. Así,

$$\begin{aligned}
 (1-L)(1-L^s)y_t &= \frac{(1-L^s)\delta_t}{(1-L)} + (1-L^s)\eta_t + \\
 &\frac{(1-L)(1-L^s)\omega_t}{(1+L+\dots+L^{s-1})} + (1-L)(1-L^s)\epsilon_t \\
 &= (1+L+\dots+L^{s-1})\delta_t + (1-L^s)\eta_t \\
 &+ (1-L)^2\omega_t + (1-L)(1-L^s)\epsilon_t.
 \end{aligned}
 \tag{1.11}$$

La función de autocovarianzas de las diferencias esta dada por

$$\begin{aligned}
 \gamma(0) &= s\sigma_\delta^2 + 2\sigma_\eta^2 + 6\sigma_\omega^2 + 4\sigma_\epsilon^2 \\
 \gamma(1) &= (s-1)\sigma_\delta^2 - 4\sigma_\omega^2 - 2\sigma_\epsilon^2 \\
 \gamma(2) &= (s-2)\sigma_\delta^2 + \sigma_\omega^2 \\
 \gamma(3) &= (s-3)\sigma_\delta^2 \\
 &\vdots \\
 \gamma(s-1) &= \sigma_\delta^2 + \sigma_\epsilon^2 \\
 \gamma(s) &= -\sigma_\eta^2 - 2\sigma_\epsilon^2 \\
 \gamma(s+1) &= \sigma_\epsilon^2 \\
 \gamma(k) &= 0 \quad \text{si } k > s+1.
 \end{aligned}$$

Es decir, la forma reducida del modelo estructural básico corresponde a un modelo MA($s+1$), donde existen sólo 4 parámetros funcionalmente independientes.

1.3 Forma de Estado

Los cuatro modelos presentados anteriormente pueden ponerse en la forma de estado general

$$\begin{aligned}
 y_t &= z'\theta_t + \epsilon_t, \quad t = 1, \dots, T, \\
 \theta_t &= T\theta_{t-1} + R\kappa_t,
 \end{aligned}
 \tag{1.12}$$

donde θ_t es un vector de dimensión p , κ_t es un vector de dimensión u , z es un vector de dimensión p , T es una matriz de dimensión $(p \times p)$, R es una matriz de dimensión $(p \times u)$; $V(\epsilon_t) = \sigma_\epsilon^2$ y $V(\kappa_t) = \Sigma_\kappa$. Todas estas cantidades están determinadas para cada modelo. El vector θ_t se llama vector de estado. Por ejemplo, en el modelo local constante

$z = T = R = p = u = 1$, $\theta_t = \mu_t$ y $\kappa_t = \eta_t$, $V(\kappa_t) = \sigma_{\eta}^2$. Por otra parte, en el modelo estructural básico con $s = 4$, $z = (1, 0, 1, 0, 0)$,

$$T = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & -1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}, \quad R = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

y $\kappa'_t = (\eta_t, \delta_t, \omega_t)$, $\theta'_t = (\mu_t, \beta_t, \gamma_t, \gamma_{1t}, \gamma_{2t})$, $p = 5$ y $u = 3$. Note que z , T y R no dependen de los parámetros en ninguno de los cuatro modelos presentados. Los parámetros del modelo aparecen en las varianzas de los procesos ϵ_t y κ_t .

Las componentes de los cuatro modelos presentados pueden ser generalizadas para incluir otro tipo de efectos del tipo ARMA, sin afectar la forma de estado (1.12). Es bien sabido que cualquier modelo ARMA

$$y_t = \sum_{i=1}^r \alpha_i y_{t-i} + \sum_{i=0}^s \beta_i a_{t-i}, \quad \beta_0 \equiv 1,$$

puede escribirse como $y_t = (1, 0, \dots, 0)\theta_t$, y

$$\theta_t = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_m \end{pmatrix} I_{m-1} \theta_{t-1} + \begin{pmatrix} 1 \\ \beta_1 \\ \vdots \\ \beta_{m-1} \end{pmatrix} a_t,$$

donde θ_t tiene dimensión $m = \max(r, s + 1)$.

En este sentido, los modelos estructurales presentados pueden entenderse como estructuras básicas para el modelamiento de series cronológicas. Dependiendo de las características particulares de la serie en estudio, otras componentes pueden incluirse en el modelo para representar efectos específicos a esta serie, o para generalizar las especificaciones presentadas.

2 ESTIMACION DE LAS COMPONENTES NO OBSERVADAS

En esta sección se considera la estimación de las componentes no observadas en un modelo en la forma de estado (1.12). Esto es, la atención se centra en la estimación del vector de estado θ_t condicional en las observaciones y_t . Naturalmente que el interés está fundamentalmente puesto en la aplicación a modelos estructurales donde, de la sección anterior, el vector de estado θ_t contiene los efectos de tendencia y estacionales.

En términos generales, la estimación de θ_t en (1.12) puede hacerse mediante lo que se conoce como el filtro de Kalman. Este es un algoritmo que entrega los estimadores del vector de estado, condicional en las observaciones, de manera de minimizar el error cuadrático medio. Bajo normalidad en las distintas componentes aleatorias que conforman el modelo, el filtro de Kalman entrega los mejores estimadores en el sentido del error cuadrático medio entre todos los estimadores. Es bien sabido que este estimador coincide con la esperanza de θ_t condicional en el conjunto de información usado para la estimación.

Si las componentes aleatorias no son normales, en cambio, la atención esta puesta en los estimadores lineales que minimizan el error cuadrático medio.

Para efectos de la presentación de los resultados en esta sección, se asume que los parámetros del modelo son conocidos. Vale decir, se supone que σ_ϵ^2 , σ_η^2 , σ_δ^2 , σ_ω^2 , y eventualmente algún otro parámetro, están dados. La estimación de estas varianzas es materia de la próxima sección.

La sección esta organizada de la siguiente forma. En la Subsección 2.1 se presenta el filtro de Kalman. En la Subsección 2.2 se considera la predicción del vector de estado y de las observaciones, mientras que en la Subsección 2.3 se estudia un importante problema en muchas aplicaciones como es el de observaciones faltantes.

2.1 El Filtro de Kalman

Usualmente existe interés en dos tipos de estimadores del vector de estado θ_t , $t = 1, \dots, T$, de acuerdo al conjunto de información utilizado para hacer esta estimación. En un primer tipo de estimador, el conjunto de información consiste en todas las observaciones hasta el tiempo t ; vale decir, $y_s, s = 1, \dots, t$. Este tipo de estimador se conoce como estimador en línea. En un segundo tipo de estimador, conocido como suavizamiento, el conjunto de información esta conformado por la totalidad de las observaciones en la muestra; vale decir, $y_s, s = 1, \dots, T$

Se considera primeramente la estimación en línea de θ_t condicional en y_1, \dots, y_t . Se asume además, en un comienzo, normalidad de las distintas componentes aleatorias que afectan al modelo, aunque los resultados sin el supuesto de normalidad serán también analizados. Finalmente, se llama H_0 al conjunto de información inicial con respecto al vector θ_0 y H_t al conjunto que incluye además a todas las observaciones hasta el tiempo t .

Sea la distribución de θ_0 condicional en H_0 dada por

$$\theta_0 \sim N(m_0, P_0), \tag{2.1}$$

con θ_0 no correlacionado con $\epsilon_t, \eta_t, \delta_t$ y ω_t para todo t . Las distribuciones de θ_t e y_t condicionales en H_0 son

$$\begin{aligned} \theta_1 &\sim N(Tm_0, TP_0T' + R\Sigma_\epsilon R') \\ &\sim N(Tm_0, \bar{P}_1), \end{aligned}$$

y

$$\begin{aligned} y_1 &\sim N(z'Tm_0, z'TP_0T'z + z'R\Sigma_\epsilon R'z + \sigma_\epsilon^2) \\ &\sim N(z'Tm_0, f_1). \end{aligned}$$

Siempre condicional en H_0 , la distribución conjunta de (θ_1, y_1) es normal con

$$\begin{aligned} cov(\theta_1, y_1) &= cov(\theta_1, z'\theta_1 + \epsilon_1) \\ &= \bar{P}_1 z. \end{aligned}$$

Luego, se tiene que

$$\begin{pmatrix} \theta_1 \\ y_1 \end{pmatrix} \sim N \left[\begin{pmatrix} Tm_0 \\ z'Tm_0 \end{pmatrix}, \begin{pmatrix} \bar{P}_1 & \bar{P}_1 z \\ z'\bar{P}_1 & f_1 \end{pmatrix} \right].$$

Ahora, la distribución de θ_1 condicional en $H_1 = \{H_0, y_1\}$, está dada por

$$\begin{aligned} \theta_1 &\sim [Tm_0 + \bar{P}_1 z f_1^{-1} (y_1 - z' T m_0); \bar{P}_1 - \bar{P}_1 z f_1^{-1} z' \bar{P}_1] \\ &\sim N[m_1; P_1]. \end{aligned} \tag{2.2}$$

De aquí se desprende que m_1 es el estimador de θ_1 condicional en H_1 y P_1 es la varianza asociada a este estimador. Se observa además que esta varianza no depende de las observaciones.

La idea anterior, representada por las ecuaciones (2.1) y (2.2), puede repetirse para obtener estimadores de θ_t condicionales en H_t para todo t . Esto constituye un algoritmo iterativo que se conoce como el filtro de Kalman. Sean

$$\begin{aligned} H_t &= \{H_0, y_1, \dots, y_t\} \\ m_t &= E(\theta_t / H_t) \\ \bar{P}_t &= V(\theta_t / H_{t-1}) \\ P_t &= V(\theta_t / H_t) \\ f_t &= V(y_t / H_{t-1}) \\ v_t &= y_t - E(y_t / H_{t-1}). \end{aligned}$$

Es decir, m_t es la esperanza de θ_t condicional en el conjunto de información al tiempo t , y P_t es su varianza. Las cantidades \bar{P}_t , f_t y v_t son, a estas alturas del análisis, cantidades auxiliares que facilitan la presentación del algoritmo. Más adelante se verá que para otros propósitos son de gran importancia. Con estas definiciones, se puede demostrar que el algoritmo anterior aplicado a cada tiempo t produce las siguientes ecuaciones recursivas,

$$\begin{aligned} \bar{P}_t &= T P_{t-1} T' + R \Sigma_n R' \\ f_t &= z' \bar{P}_t z + \sigma_n^2 \\ P_t &= \bar{P}_t - \bar{P}_t z f_t^{-1} z' \bar{P}_t \\ m_t &= T m_{t-1} + \bar{P}_t z' f_t^{-1} v_t \\ v_t &= y_t - z' T m_{t-1}. \end{aligned} \tag{2.3}$$

La derivación anterior del filtro de Kalman es perfectamente válida si las distintas componentes estocásticas del modelo no son normales. En ese caso, la notación $N(\mu, \sigma^2)$ debe entenderse como una distribución cualquiera con media μ y varianza σ^2 . Por otra parte, el concepto de distribución condicional utilizado en (2.2) debe ser reemplazado por el de proyección lineal.

Dos problemas adicionales quedan por resolverse para poder aplicar las ecuaciones anteriores. En primer lugar, el desarrollo anterior requiere una distribución inicial para el vector de estado. Es decir, se requiere conocer m_0 y P_0 . En segundo lugar, resulta de alguna importancia el estudiar el comportamiento de las ecuaciones (2.3) cuando t tiende a infinito.

Con respecto a la distribución de θ_0 , ésta es usualmente desconocida y por tanto, parece apropiado definir lo que se conoce como una distribución inicial difusa. Aunque esta no es la única solución para resolver el problema planteado, es la que ha tenido mayor atención en la literatura. De este modo, se define

$$\theta_0 \sim N(0, rI_p), \quad r \rightarrow \infty.$$

Con esta distribución inicial, las ecuaciones del filtro de Kalman pueden utilizarse para calcular los estimadores. Aunque esto es teóricamente razonable, la distribución inicial difusa presenta inconvenientes de tipo práctico por cuanto el valor de r debe especificarse y un valor muy alto podría generar inestabilidad numérica en los computos. Es posible mostrar que, para cualquiera de los modelos estructurales presentados en la sección anterior, ó extensiones a estos del tipo ARMA, si la distribución inicial del vector de estado es difusa, la distribución de θ_p condicional en las primeras p observaciones es propia y no difusa. En otras palabras, una distribución propia para el vector de estado se obtiene al tiempo $t = p$.

Una solución alternativa al problema de las condiciones iniciales es intentar obtener m_p y P_p usando las primeras p observaciones y luego correr el filtro de Kalman a partir de $t = p + 1$. De hecho, estas dos soluciones alternativas deben dar los mismos resultados por cuanto el estimador que minimiza el error cuadrático medio es único.

Para ejemplificar esta situación, considere el modelo local lineal (1.5) donde $p = 2$. Para $t = 1, 2$ se tiene que

$$y_2 = \mu_2 + \epsilon_2$$

$$y_1 = \mu_2 - \beta_2 - \eta_2 + \delta_2 + \epsilon_1,$$

desde donde se obtiene el estimador que minimiza el error cuadrático medio de $\theta_2 = (\mu_2, \beta_2)'$, m_2 , y su correspondiente varianza, P_2 ,

$$m_2 = \begin{pmatrix} y_2 \\ y_2 - y_1 \end{pmatrix}; \quad P_2 = \begin{pmatrix} \sigma_\epsilon^2 & \\ & \sigma_\delta^2 + \sigma_\eta^2 + \sigma_\epsilon^2 \end{pmatrix}.$$

Una solución general al problema de condiciones iniciales en modelos estructurales es presentada en Marshall (1990).

El segundo problema planteado anteriormente, que se refería al comportamiento del filtro de Kalman cuando t tiende a infinito, está estrechamente relacionado con la especificación de condiciones iniciales para el vector de estado. Para los modelos estructurales presentados en este documento se demuestra que

$$\lim_{t \rightarrow \infty} P_t = P \geq 0,$$

y que el valor de este límite es independiente de las condiciones iniciales. En otras palabras, la especificación de condiciones iniciales sólo tiene un efecto transiente en los estimadores del vector de estado. Si además, $\sigma_\eta^2 > 0$ en el modelo local constante, $\sigma_\delta^2 > 0$ en el modelo local lineal, $\sigma_\eta^2 > 0$ y $\sigma_\omega^2 > 0$ en el modelo local constante con estacionalidad, $\sigma^2 \delta > 0$ y $\sigma_\omega^2 > 0$ en el modelo estructural básico, entonces la convergencia es exponencialmente rápida.

Por ejemplo, en el modelo local constante (1.2),

$$y_1 = \mu_1 + \epsilon_1,$$

y por tanto $m_1 = y_1$ y $P_1 = \sigma_\eta^2$. Aplicando el filtro de Kalman para $t > 1$ se obtiene

$$\begin{aligned} m_t &= m_{t-1} + \bar{P}_t f_t^{-1} (y_t - m_{t-1}) \\ &= m_{t-1} + \lambda_t (y_t - m_{t-1}). \end{aligned}$$

Si $\sigma_\eta^2 > 0$ y en estado estacionario,

$$m_t = m_{t-1} + \lambda (y_t - m_{t-1}),$$

lo cual corresponde exactamente al modelo de series de tiempo conocido como suavizamiento exponencial. Se puede mostrar, usando las ecuaciones del filtro de Kalman (2.3), que en este caso la convergencia λ_t a λ es exponencialmente rápida. Por otro lado, si $\sigma_\eta^2 = 0$,

$$\begin{aligned} P_t &= P_{t-1} - \frac{P_{t-1}^2}{P_{t-1} + \sigma_\eta^2} \\ &= \frac{\sigma_\eta^2}{t}, \end{aligned}$$

y luego,

$$\begin{aligned} m_t &= m_{t-1} + (1/t)(y_t - m_{t-1}) \\ &= (1/t) \sum_{s=1}^t y_s. \end{aligned}$$

Es decir, cuando $\sigma_\eta^2 = 0$ el nivel μ_t es constante, y el filtro de Kalman entrega como estimador al promedio aritmético de las observaciones. La varianza asociada a este estimador converge a 0, pero la convergencia es lineal y no exponencial.

Hasta ahora, la atención ha estado puesta en el estimador en línea del vector de estado. A menudo, en aplicaciones, interesa también calcular el estimador suavizado, que considera como conjunto de información a toda la muestra (y_1, \dots, y_T) . Es posible derivar ecuaciones recursivas similares a (2.3) para este tipo de estimador y su correspondiente varianza. Formulas generales son presentadas en Harvey (1981) y Anderson y Moore (1979).

2.2 Predicción

Las ecuaciones del filtro de Kalman pueden adaptarse para obtener predicciones del vector de estado o predicciones de observaciones futuras. En términos generales interesa estimar θ_{t+l} con información hasta el tiempo t . Considere primero la predicción de θ_{t+l} condicional en H_t . De la forma de estado

$$\theta_{t+l} = T^l \theta_t + \sum_{j=0}^{l-1} T^j R \kappa_{t+j+1}, \quad l > 0.$$

Luego, tomando esperanza condicional, el estimador de θ_{t+l} condicional en H_t y la varianza del error de predicción son

$$m_t(l) = T^l m_t, \quad l > 0,$$

y

$$P_t(l) = T^l P_t T^{l'} + \sum_{j=0}^{l-1} T^j R \Sigma_* R' T^{j'}$$
, $l > 0$,

donde, de acuerdo con la notación utilizada en la subsección previa, $P_t(1) = \bar{P}_{t+1}$. En cuanto a la predicción de y_{t+l} condicional en H_t , se tiene

$$y_{t+l} = z^l \theta_{t+l} + \epsilon_{t+l}$$
,

y luego la predicción de y_{t+l} condicional en el conjunto de información H_t , $\tilde{y}_t(l)$, y la varianza del error de predicción, $f_t(l)$, están dadas por

$$\tilde{y}_t(l) = z^l m_t(l)$$
,

y

$$f_t(l) = z^l P_t(l) z + \sigma_\epsilon^2$$
,

para $l > 0$ y, de acuerdo a la notación de la subsección anterior, $f_t(1) = f_{t+1}$.

2.3 Observaciones Faltantes

Muchas aplicaciones están basadas en series con observaciones faltantes. Afortunadamente, y como se muestra en las siguientes líneas, esto no representa ningún inconveniente en la metodología de los modelos estructurales. Suponga, para simplificar la exposición, que no se dispone de una de las observaciones y que esta es y_s . Condicional en el conjunto de información H_{s-1}

$$\theta_s \sim N(m_{s-1}(1), P_{s-1}(1))$$
.

Sin embargo, el estimador de θ_s no puede adaptarse por cuanto y_s es desconocido. Es decir, el algoritmo del filtro de Kalman no puede usarse al tiempo $t = s$. Ahora, siempre condicional en el conjunto de información al tiempo $t = s - 1$,

$$\theta_{s+1} \sim N(m_{s-1}(2), P_{s-1}(2))$$
,

y el estimador de θ_{s+1} si puede adaptarse usando esta última como la distribución inicial del vector de estado θ_{s+1} . En consecuencia, observaciones faltantes no presentan un problema por cuanto si y_s es una observación faltante, las cantidades m_s y P_s son reemplazados por $m_{s-1}(1)$ y $P_{s-1}(1)$. Es decir por las predicciones a un paso. Resultados semejantes son obtenidos para los estimadores suavizados. Algo que tendrá alguna importancia en la próxima sección, en la cual se considera la estimación de las varianzas del modelo, es que v_s pierde sentido si y_s es una observación faltante.

3 ESTIMACION DE LOS PARAMETROS

Esta sección considera la estimación de los parámetros de un modelo estructural. Estos corresponden a las varianzas de los distintos efectos aleatorios que generan las observaciones, las componentes de tendencia y los efectos estacionales: σ_ϵ^2 , σ_η^2 , σ_δ^2 y σ_ω^2 .

El procedimiento de estimación que se considera está basado en el principio de máxima verosimilitud, o en aproximaciones asintóticas, y en el supuesto de que las observaciones son normales. En la Subsección 3.1 se entregan resultados relativos a la verosimilitud exacta de un modelo en la forma de estado obtenida a partir del filtro de Kalman y la descomposición del error de predicción. Los resultados que ahí se presentan son más generales que lo que se requiere para estimar los parámetros de los modelos estructurales de la Sección 1. El enfoque incluye también, por ejemplo, a los modelos ARMA.

En algunas situaciones, la maximización de la verosimilitud es engorrosa y llama a algún tipo de simplificación. Con este propósito, la Subsección 3.2 considera resultados relativos a aproximaciones de la verosimilitud para el caso de modelos estructurales. Aunque estos procedimientos pueden también usarse en otro tipo de modelos, su simplicidad se hace más relevante en el caso de los modelos estructurales.

3.1 Verosimilitud Exacta

En esta subsección se considera la estimación de los parámetros de un modelo en la forma de estado (1.12). Usando la descomposición del error de predicción se tiene que para cualquier valor de $q \geq 1$,

$$l(y_1, \dots, y_T) = l(y_1, \dots, y_q) + \sum_{t=q+1}^T l(y_t/H_{t-1}), \quad (3.1)$$

donde $l(x)$ representa al logaritmo de la función verosimilitud de x y H_t es el conjunto de información hasta el tiempo t definido como en la Sección 2. En el caso en que las observaciones tengan distribución normal, las ecuaciones del filtro de Kalman entregan

$$l(y_t/H_{t-1}) = -(1/2) \log 2\pi - (1/2) \log f_t - (1/2) \frac{v_t^2}{f_t}.$$

Lo que implica que la suma en (3.1) puede evaluarse usando el filtro de Kalman. Con respecto al primer término en (3.1), $l(y_1, \dots, y_q)$, se tienen varias posibilidades dependiendo del modelo y de la especificación de condiciones iniciales en el vector de estado. Estas posibilidades son:

(i) Si la distribución de θ_0 es propia, vale decir conocida y no difusa, se puede formar $l(y_1, \dots, y_T/H_0)$ tomando $q = 0$. Sin embargo esto entrega una verosimilitud condicional en H_0 , que puede no ser el objetivo. Si se desea construir la verosimilitud incondicional, entonces, la verosimilitud se forma a partir de

$$l(y_1, \dots, y_T) = l(\theta_0) + l(y_1, \dots, y_T/\theta_0) - l(\theta_0/y_1, \dots, y_T).$$

El primer término en esta identidad es evaluado directamente con la distribución inicial del vector de estado. El segundo término se obtiene a partir del filtro de Kalman, mientras que la tercera expresión también puede obtenerse en términos de las cantidades entregadas por el filtro de Kalman; ver de Jong (1988) y Marshall (1990).

(ii) Si la distribución de θ_0 es difusa, $l(y_1, \dots, y_p)$, donde p es la dimensión del vector de estado θ_t , no está definida y la función objetivo a maximizar es

$$Q = -(1/2) \sum_{t=p+1}^T \left[\log f_t + \frac{v_t^2}{f_t} \right]. \quad (3.2)$$

(iii) En un proceso estacionario se puede tomar $\varphi = 1$ y luego $l(y_1)$ se obtiene a partir de la distribución incondicional del proceso. Por ejemplo, en un modelo AR(1) con media 0,

$$l(y_1) = -(1/2) \log 2\pi - (1/2) \log \frac{\sigma^2}{1 - \phi^2} - (1/2) \frac{y_1^2(1 - \phi^2)}{\sigma^2},$$

$$l(y_t/H_{t-1}) = -(1/2) \log 2\pi - (1/2) \log \sigma^2 - (1/2) \frac{(y_t - \phi y_{t-1})^2}{\sigma^2}, \quad t = 2, \dots, T.$$

Al igual que en la sección anterior, para modelos estructurales se supone que la distribución de θ_0 es difusa. Los estimadores máximo verosímiles hacen máximo a Q , y esta maximización debe hacerse numéricamente. Para esto, es posible calcular, a partir de las ecuaciones del filtro de Kalman, las derivadas de v_t , f_t y Q con respecto a los parámetros.

Sin perjuicio de lo anterior, uno de los parámetros en un modelo estructural, por ejemplo σ_v^2 , puede concentrarse fuera de la verosimilitud. Efectivamente, a partir de las recursiones del filtro de Kalman se demuestra que para todo t ,

$$f_t = f_t^* \sigma_v^2,$$

donde f_t^* se calcula igual que f_t en (2.3) pero con los parámetros $(1, \sigma_n^2, \sigma_b^2, \sigma_s^2/\sigma_v^2, \sigma_\omega^2/\sigma_v^2)$ en lugar de $(\sigma_v^2, \sigma_n^2, \sigma_b^2, \sigma_s^2, \sigma_\omega^2)$. Más aún, se demuestra que v_t no depende del valor absoluto de estas varianzas sino de las razones. Entonces,

$$Q = -(1/2) \sum_t \log(\sigma_v^2 f_t^*) - (1/2) \sum_t \frac{v_t^2}{\sigma_v^2 f_t^*},$$

y por tanto el estimador máximo verosímil de σ_v^2 es

$$\hat{\sigma}_v^2 = \frac{1}{T-p} \sum_t \frac{v_t^2}{f_t^*};$$

dejando la verosimilitud concentrada

$$Q_c = -((T-p)/2) \log \hat{\sigma}_v^2 - ((T-p)/2) - (1/2) \sum_t \log f_t^*,$$

para ser maximizada con respecto a los restantes parámetros del modelo. El análisis anterior es igualmente válido cualquiera sea la varianza que se concentre fuera de la verosimilitud.

En consecuencia, en el modelo estructural básico que tiene cuatro parámetros, uno de estos es concentrado fuera de la verosimilitud y los restantes tres son obtenidos mediante

una rutina de optimización numérica. En el modelo local constante, en cambio, que sólo tiene dos parámetros, la optimización numérica es llevada a cabo sobre uno de los parámetros.

Bajo condiciones de regularidad y si todas las varianzas del modelo son mayores que cero, los estimadores que entrega el método de máxima verosimilitud son consistentes, normales, asintóticamente insesgados, y la matriz

$$\left[\frac{\partial^2 Q}{\partial \psi \partial \psi'} \right]_{\psi = \hat{\psi}}, \quad \psi' = (\sigma_\epsilon^2, \dots, \sigma_\omega^2),$$

representa una aproximación de la varianza de los estimadores $\hat{\psi}$; ver Pagan (1980). Este resultado no depende del supuesto de normalidad aunque si las observaciones no son normales, la maximización de Q no corresponde exactamente al método de máxima verosimilitud.

3.2 Verosimilitud Espectral

El método de verosimilitud exacta presentado en la subsección anterior requiere una pasada del filtro de Kalman en cada oportunidad en que la función verosimilitud es evaluada por el algoritmo numérico de optimización. Esto, a menudo, requiere de una cantidad considerable de recursos computacionales. En esta subsección se consideran aproximaciones asintóticas al método de máxima verosimilitud.

La versión estacionaria de los modelos estructurales presentados en la Sección 1 tienen la forma general

$$p(L)y_t = p_\epsilon(L)\epsilon_t + p_\eta(L)\eta_t + p_\delta(L)\delta_t + p_\omega(L)\omega_t, \quad t = p + 1, \dots, T,$$

donde $p(L)$, $p_\epsilon(L)$, $p_\eta(L)$, $p_\delta(L)$ y $p_\omega(L)$ son polinomios en el operador de rezago L . Sea y un vector de dimensión $(T - p \times 1)$ con t -ésima componente $p(L)y_t$. Sea la varianza de este vector $V(y) = \Sigma$ y suponga que $(T - p)$ es impar. Sea además $m = [(T - p - 1)/2]$. Entonces, para un valor de T suficientemente grande,

$$\Sigma = HDH'$$

donde H es la matriz de Fourier de dimensión $(T - p \times T - p)$ con elemento típico h_{ij} dado por

$$h_{ij} = \begin{cases} (1/T - p)^{1/2} & \text{si } j = 1, \\ (2/T - p)^{1/2} \cos(\pi ij / (T - p)) & \text{si } j = 2, 4, \dots, (T - p - 1), \\ (2/T - p)^{1/2} \sin(\pi i(j - 1) / (T - p)) & \text{si } j = 3, 5, \dots, (T - p). \end{cases}$$

Se puede mostrar fácilmente que H satisface $H'H = I_{T-p}$. Por su parte, D es una matriz diagonal de la forma

$$\text{diag}(D) = (d_0, d_1, d_1, \dots, d_m, d_m).$$

con elemento típico definido como

$$d_j = \gamma_0 + 2 \sum_{k=1}^{\infty} \gamma_k \cos(2\pi j k / (T - p)), \quad j = 0, \dots, m,$$

y con γ_k la autocovarianza de orden k . El elemento d_j es 2π veces la densidad espectral evaluada en la frecuencia $(2\pi j)/(T - p)$. El caso $(T - p)$ par requiere de algunos cambios en las definiciones de H y D ; para mayores detalles ver Fuller (1976).

Usando estos resultados, es posible obtener una expresión asintóticamente equivalente a la función verosimilitud. En efecto, el logaritmo de la verosimilitud exacta del vector y es

$$\begin{aligned} l(y) &= -((T - p)/2) \log(2\pi) - (1/2) \log |\Sigma| - (1/2) y' \Sigma^{-1} y \\ &= -((T - p)/2) \log(2\pi) - 1/2 \sum_{j=0}^m \omega_j \log d_j - 1/2 \sum_{j=0}^m \frac{p_j}{d_j}, \end{aligned}$$

donde $\omega_j = 1$ si $j = 0$ y $\omega_j = 2$ si $j > 0$; la componente p_j corresponde al periodograma de y ,

$$p_j = \begin{cases} \left(\frac{2}{T-p} \right) \left[\left(\sum_t y_t \cos \left(\frac{2\pi j t}{T-p} \right) \right)^2 + \left(\sum_t y_t \sin \left(\frac{2\pi j t}{T-p} \right) \right)^2 \right] & \text{si } j \neq 0 \\ \left(\frac{1}{T-p} \right) (\sum_t y_t)^2 & \text{si } j = 0; \end{cases}$$

y d_j se calcula para cada modelo a partir de la función de autocovarianzas. Definiendo la constante $c_{rj} = 2 - 2 \cos(2\pi jr/(T - p))$, se tiene que, para el modelo local constante,

$$d_j = \sigma_\eta^2 + c_{1j} \sigma_\epsilon^2, \quad j = 0, \dots, m;$$

mientras que para el modelo local lineal,

$$d_j = \sigma_\delta^2 + c_{1j} \sigma_\eta^2 + c_{2j}^2 \sigma_\epsilon^2, \quad j = 0, \dots, m.$$

Por su parte, en el modelo local constante con estacionalidad,

$$d_j = \begin{cases} s^2 \sigma_\eta^2 & \text{si } j = 0 \\ \frac{c_{1j}}{c_{1j}} \sigma_\eta^2 + c_{1j} \sigma_\omega^2 + c_{2j} \sigma_\epsilon^2 & \text{si } j \neq 0, \end{cases}$$

y, finalmente, en el modelo estructural básico,

$$d_j = \begin{cases} s^2 \sigma_\delta^2 & \text{si } j = 0 \\ \frac{c_{1j}}{c_{1j}} \sigma_\delta^2 + c_{2j} \sigma_\eta^2 + c_{1j} \sigma_\omega^2 + c_{1j} c_{2j} \sigma_\epsilon^2 & \text{si } j \neq 0. \end{cases}$$

Con estas definiciones, la función objetivo a maximizar puede escribirse como

$$Q^* = -(1/2) \sum_{j=0}^m \left[\omega_j \log d_j + \frac{p_j}{d_j} \right], \quad (3.3)$$

con

$$d_j = c_{\epsilon_j} \sigma_{\epsilon}^2 + c_{\eta_j} \sigma_{\eta}^2 + c_{\delta_j} \sigma_{\delta}^2 + c_{\omega_j} \sigma_{\omega}^2,$$

y donde c_{ϵ_j} , c_{η_j} , c_{δ_j} y c_{ω_j} son constantes conocidas.

La función Q^* debe maximizarse numéricamente aunque, al igual que en la expresión (3.2), uno de los parámetros puede concentrarse fuera de la verosimilitud. La gran ventaja de Q^* sobre Q definida en la sección anterior, es que la evaluación de Q^* es considerablemente más simple por cuanto los parámetros aparecen explícitamente en la función y no es necesario hacer algún tipo de cálculo equivalente a una pasada del filtro de Kalman cada vez que se evalúe la verosimilitud. Si bien es cierto que la expresión (3.3) requiere el cálculo del periodograma, esto se realiza sólo una vez por cuanto el periodograma depende exclusivamente de las observaciones y no de los parámetros del modelo.

Las primeras dos derivadas de Q^* con respecto al vector de parámetros ψ son

$$\frac{\partial Q^*}{\partial \psi} = -1/2 \sum_j \left[\frac{\omega_j}{d_j} - \frac{p_j}{d_j^2} \right] x_j, \quad (3.4)$$

donde, para el modelo estructural básico, $x_j = (c_{\epsilon_j}, c_{\eta_j}, c_{\delta_j}, c_{\omega_j})$, y

$$\frac{\partial^2 Q^*}{\partial \psi \partial \psi'} = -1/2 \sum_j \left[-\frac{\omega_j}{d_j^2} + \frac{2p_j}{d_j^3} \right] x_j x_j'.$$

Además, como aparte de una constante multiplicativa el periodograma es asintóticamente insesgado de la densidad espectral, la matriz

$$I(\psi) = (1/2) \sum_j \frac{\omega_j}{d_j^2} x_j x_j'$$

es asintóticamente equivalente a la matriz de información.

Bajo ciertas condiciones de regularidad, y si todas las varianzas de un modelo estructural son mayores que cero, los estimadores que maximizan Q^* , $\hat{\psi}$, son consistentes, insesgados y asintóticamente normales. La matriz $I^{-1}(\hat{\psi})$ es una aproximación finita de la varianza de los estimadores. Este resultado no depende del supuesto de normalidad; ver Dunsmuir (1979).

El resultado anterior justifica la formulación de test estadísticos clásicos como Wald, Razón de Verosimilitud y Multiplicadores de Lagrange.

Una perspectiva distinta del estimador que maximiza la verosimilitud espectral puede obtenerse al igualar a cero el vector de primeras derivadas en (3.4). Esto implica que

$$\sum_j \frac{\omega_j x_j x_j' \hat{\psi}}{d_j^2} = \sum_j \frac{p_j x_j}{d_j^2},$$

y luego

$$\hat{\psi} = \left[\sum_j \frac{\omega_j x_j x_j'}{d_j^2} \right]^{-1} \left[\sum_j \frac{p_j x_j}{d_j^2} \right]. \quad (3.5)$$

Es decir, $\hat{\psi}$ tiene la forma de un estimador mínimo cuadrático. El problema ciertamente es que la matriz de varianzas y covarianzas depende justamente del vector de parámetros. Un estimador consistente de ψ se puede obtener al hacer $d_j = 1 \forall j$ en (3.5). Más aún, el método de puntajes considera un estimador obtenido iterativamente mediante

$$\hat{\psi}_i = \hat{\psi}_{i-1} + I^{-1}(\hat{\psi}_{i-1}) \left[\frac{\partial Q}{\partial \psi} \right]_{\psi = \hat{\psi}_{i-1}},$$

donde $\hat{\psi}_0$ es un estimador inicial consistente. El estimador $\hat{\psi}_1$ es, asintóticamente, tan eficiente como el estimador de máxima verosimilitud. Se demuestra además que $\hat{\psi}_i$ converge al estimador de máxima verosimilitud.

4 DIAGNOSTICOS

Una forma importante de contrastar el modelo especificado con las observaciones es a través de un estudio de las características de los residuos del modelo. Lo que usualmente se conoce como diagnósticos. Para modelos en la forma de estado (1.12), estos diagnósticos están basados en las innovaciones o errores de predicción estandarizados. Esto es,

$$\hat{v}_t = \frac{v_t}{f_t^{1/2}}, \quad t = p+1, \dots, T.$$

Si los parámetros del modelo son conocidos, y el modelo está bien especificado, estos residuos son no correlacionados, y bajo normalidad,

$$\hat{v}_t \sim N(0, 1), \quad t = p+1, \dots, T.$$

En lo que sigue de esta sección se presentan tests de hipótesis típicamente usados en modelos de series cronológicas. Estos son, un test de igualdad de las varianzas de los residuos en el período muestral, un test de correlación serial ó autocorrelación, y un test de normalidad.

Heterocedasticidad

Para docimar que la varianza de los residuos es constante en la muestra, se puede utilizar el estadígrafo

$$H = \frac{\sum_{t=r_2}^{T-p} \hat{v}_t^2 / (T-p-r_2+1)}{\sum_{t=p+1}^{r_1} \hat{v}_t^2 / r_1},$$

que tiene distribución $F(T-p-r_2+1, r_1)$ bajo la hipótesis de homocedasticidad. Es decir bajo la hipótesis que no hay cambios en la varianza de los residuos v_t entre el período $(p+1, r_1)$ y el período $(r_2, T-p)$. Típicamente r_1 y r_2 son elegidos de manera de comparar el tercer y el primer tercio del período muestral.

Una especificación ligeramente diferente de H permite formular un test para la bondad de ajuste del modelo. Esto se realiza comparando la varianza de los residuos en la muestra con la de las predicciones fuera de la muestra. Suponga que el modelo es ajustado con observaciones desde $t = 1$ y hasta $t = T$, pero se dispone de observaciones hasta $T + 1$. Entonces,

$$H(l) = \frac{\sum_{t=T+1}^{T+l} \hat{v}_t^2 / l}{\sum_{t=p+1}^T \hat{v}_t^2 / T - p}$$

tiene distribución $F(l, T - p)$ y asintóticamente $lH(l)$ se distribuye $\chi^2(l)$. El problema de este test es que un valor pequeño de $H(l)$ no implica necesariamente un buen ajuste. Podría estar indicando carencia de ajuste tanto dentro como fuera de la muestra.

Correlación Serial

Uno de los diagnósticos más importantes, y que usualmente recibe la mayor atención es el test de correlación serial o autocorrelación. El estadígrafo

$$B(l) = (T - p)(T - p + 2) \sum_{k=1}^l (T - k)^{-1} r_k^2,$$

donde r_k es la autocorrelación de los residuos \hat{v}_t de orden k , tiene, cuando T y l tienden a infinito la distribución χ^2 con grados de libertad igual $(l + 1)$ menos el número de parámetros, bajo la hipótesis de que los residuos no están serialmente correlacionados; ver Ljung y Box (1978).

Normalidad

Un tercer test de hipótesis corresponde al test de normalidad. El estadígrafo

$$N = \frac{T - p}{6} \left(\sum_t \frac{\hat{v}_t^3}{s^3(T - p)} \right) + \frac{T - p}{24} \left(\sum_t \frac{\hat{v}_t^4}{s^4(T - p)} - 3 \right)^2,$$

donde $s^2 = \sum_t \hat{v}_t^2 / (T - p)$, tiene en muestras grandes la distribución $\chi^2(2)$ bajo la hipótesis de normalidad. Este test es particularmente útil para detectar valores extremos; ver Bowman y Shenton (1975).

5 BONDAD DE AJUSTE

El coeficiente de determinación R^2 es usualmente utilizado, en modelos de regresión, para comparar la suma de cuadrados de los residuos del modelo definida como

$$SCR = \sum_t \hat{v}_t^2,$$

con la suma de cuadrados $\sum (y_t - \bar{y})^2$, que supone que el modelo alternativo al ajustado es

$$y_t = u + \epsilon_t, \quad t = 1, \dots, T.$$

En series de tiempo no estacionarias, este modelo alternativo es demasiado inocente y la suma de cuadrados de los residuos del modelo debe compararse con una suma de cuadrados más realista. Dos medidas alternativas al coeficiente R^2 han sido propuestas en la literatura, dependiendo de si la serie cronológica contiene o no efectos estacionales.

Para series sin efectos estacionales, se propone

$$R_d^2 = 1 - \frac{SCR}{\sum_t (y_t^* - \bar{y}^*)^2}, \quad (5.1)$$

donde y_t^* corresponde a la serie de las primeras diferencias de y_t , y \bar{y}^* al promedio de estas diferencias. La medida R_d^2 supone que el modelo alternativo al ajustado es

$$y_t = y_{t-1} + \beta + \epsilon_t, \quad t = 2, \dots, T.$$

Para series con efectos estacionales de período s , se propone

$$R_s^2 = 1 - \frac{SCR}{\sum_t (y_t^* - \bar{\gamma}_j)^2}, \quad (5.2)$$

donde y_t^* está definido como en (5.1) y $\bar{\gamma}_j$ es el promedio de los y_t^* correspondientes a los meses o trimestres j . El modelo alternativo en este caso está dado por

$$y_t = y_{t-1} + \sum_{j=1}^s \gamma_j z_{tj} + \epsilon_t, \quad t = 2, \dots, T,$$

donde z_{tj} es igual a 0 ó 1 dependiendo de si t corresponde o no al mes o trimestre j , y γ_j es el efecto estacional del mes o trimestre j .

6 EXTENSIONES

En esta sección se consideran algunas extensiones a los modelos estructurales presentados en la Sección 1. La primera de ellas es la inclusión de una nueva componente estocástica correspondiente a un efecto cíclico. La segunda extensión corresponde a la inclusión de variables exógenas.

6.1 Efectos Cíclicos

Un modelo que ha sido tradicionalmente popular en el análisis de series cronológicas es el modelo armónico

$$y_t = \alpha \cos \lambda t + \beta \sin \lambda t + \epsilon_t, \quad t = 1, \dots, T, \quad (6.1)$$

donde $0 \leq \lambda \leq \pi$ corresponde a la frecuencia asociada al período $2\pi/\lambda$, y α y β determinan la amplitud de la componente. Usualmente α , β y λ son constantes desconocidas. Las

aplicaciones del modelo armónico han mostrado que éste es extremadamente rígido y no permite cambios en la forma como la componente es generada en el tiempo. Esta es la misma crítica que se le hace, por ejemplo, a las tendencias de tipo determinístico. En las siguientes líneas se muestra que (6.1) puede escribirse en la forma de estado, y se sugieren extensiones a éste de manera de hacerlo más flexible. Escribiendo,

$$[\beta \cos \lambda t + \alpha \sin \lambda t] = \cos \lambda [\alpha \cos \lambda(t-1) + \beta \sin \lambda(t-1)] \\ + \sin \lambda [\beta \cos \lambda(t-1) - \alpha \sin \lambda(t-1)],$$

y

$$[\beta \cos \lambda t - \alpha \sin \lambda t] = -\sin \lambda [\alpha \cos \lambda(t-1) + \beta \sin \lambda(t-1)] \\ + \cos \lambda [\beta \cos \lambda(t-1) - \alpha \sin \lambda(t-1)],$$

y agregando efectos aleatorios, y un factor de estacionaridad, (6.1) puede generalizarse a

$$\begin{pmatrix} \psi_t \\ \psi_t^* \end{pmatrix} = \rho \begin{pmatrix} \cos \lambda & \sin \lambda \\ -\sin \lambda & \cos \lambda \end{pmatrix} \begin{pmatrix} \psi_{t-1} \\ \psi_{t-1}^* \end{pmatrix} + \begin{pmatrix} \kappa_t \\ \kappa_t^* \end{pmatrix},$$

donde $0 \leq \rho \leq 1$ es una constante de estacionaridad y κ_t y κ_t^* son ruidos blancos no correlacionados.

Componentes armónicas pueden incorporarse en los modelos local constante y local lineal presentados en la Sección 1 para representar efectos estacionales. Así, asumiendo que el período estacional s es par,

$$\gamma_t = \sum_{j=1}^{s/2} \gamma_{jt}, \quad t = 1, \dots, T,$$

donde

$$\gamma_{jt} = \begin{cases} \alpha_j \cos \frac{2\pi jt}{s} + \beta_j \sin \frac{2\pi jt}{s}, & j < s/2, \\ \alpha_j \cos(\pi t), & j = s/2. \end{cases}$$

Por ejemplo, si $s = 4$,

$$\begin{pmatrix} \gamma_{1t} \\ \gamma_{1t}^* \\ \gamma_{2t} \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} \gamma_{1t-1} \\ \gamma_{1t-1}^* \\ \gamma_{2t-1} \end{pmatrix} + \begin{pmatrix} \kappa_{1t} \\ \kappa_{1t}^* \\ \kappa_{2t} \end{pmatrix}.$$

6.2 Variables Exógenas

La forma de estado presentada en la Sección 1 puede extenderse para incluir variables exógenas. Suponga

$$y_t = z_t' \theta_t + x_t' \delta + \varepsilon_t, \quad t = 1, \dots, T,$$

donde x_t' es un vector fila de variables exógenas y δ es un vector de coeficientes. Una importante aplicación de esta extensión es el caso en que $x_t' \delta$ representa saltos o valores extremos en la serie. Es decir x_t' corresponde a una variable del tipo 0 o 1.

La estimación del vector de coeficientes δ tiene varias posibilidades. En primer lugar,

el filtro de Kalman puede correrse sobre $y_t - z_t' \delta$ y δ se estima con el procedimiento numérico que maximiza Q . Por otro lado, δ puede incluirse en un vector de estado ampliado. Sin embargo, la más eficiente forma de tratar a δ fue propuesta por Kohn y Ansley (1985). Estos autores demostraron que las innovaciones v_t pueden escribirse como

$$v_t = v_{y_t} - V_{z_t}' \delta, \quad t = 1, \dots, T,$$

donde v_{y_t} y V_{z_t}' son pseudo innovaciones obtenidas al correr el filtro de Kalman sobre y_t y cada elemento de z_t . Se desprende que el estimador máximo verosímil de δ es

$$\hat{\delta} = \left[\sum_t \frac{V_{z_t} V_{z_t}'}{f_t} \right]^{-1} \left[\sum_t \frac{V_{z_t} v_{y_t}}{f_t} \right]$$

Un resultado análogo a este se obtiene con la verosimilitud espectral.

7 REFERENCIAS

Anderson, B.D.O. y J.B. Moore (1979), , Prentice-Hall.

Bowman, K.O. y L.R. Shenton (1975) Omnibus test contours for departures from normality based on k_1 and k_2 , , 243-250.

de Jong, P. (1988) The Likelihood for a State Space Model, , 165-169.

Duncan, D.B. y S.D. Horn (1972) Linear Dynamic Regression from the Viewpoint of Regression Analysis, , 815-821.

Dunsmuir, W. (1979) A Central Limit Theorem for Parameter Estimation in Stationary Vector Time Series and its Application to Models for a Signal Observed with Noise, , 490-506.

Engle, R.F. (1978) Estimating Structural Models of Seasonality, In , A. Zellner editions, 281-308.

Fuller, W.A. (1976) , John Wiley.

Harrison, P.J. y C.F. Stevens (1976) Bayesian Forecasting, , 205-247 (with discussion).

Harvey, A.C. (1981) , Philip Allan.

- Harvey, A.C. (1989)
Cambridge University Press.
- Harvey, A.C. y J. Durbin (1986) *The Effects of Seat Belt Legislation on British Road Casualties: A Case Study in Structural Time Series Modelling*,
....., 187-227. (with discussion).
- Harvey, A.C. y P.H.J. Todd (1983) *Forecasting Economic Time Series with Structural and Box-Jenkins Models: A Case Study*, vol. 1
N° 4, 299-315 (with discussion).
- Holt, C.C. (1957)
....., Carnegie Institute of Technology Pittsburgh, Pennsylvania, ONR Research Memorandum N° 52.
- Jenkins, G.M. (1982) *Some Practical Aspects of Forecasting in Organizations*,
....., 3-21.
- Kitagawa, G. (1981) *A Nonstationary Time Series Model and its Fitting by a Recursive Filter*,
....., 103-116.
- Kohn, R. y C. Ansley (1985) *Efficient Estimation and Prediction in Time Series Regression Models*,
....., 694-697.
- Marshall, P. (1990)
....., Unpublished Thesis, London School of Economics.
- Muth, J.F. (1960) *Optimal Properties of Exponentially Weighted Forecasts*,
....., 299-305.
- Nelson, C.R. y C.I. Plosser (1982) *Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications*,
....., 130-162.
- Pagan, A. (1980) *Some Identification and Estimation Results for Regression Models with Stochastically Varying Coefficients*,
....., 341-363.
- Winters, P.R. (1960) *Forecasting Sales by Exponentially Weighted Moving Averages*,
....., 324-342.