

75
ZEJ



UNIVERSIDAD NACIONAL AUTONOMA
DE MEXICO

FACULTAD DE CIENCIAS

"ELEMENTOS DE PRONOSTICACION
BAYESIANA"

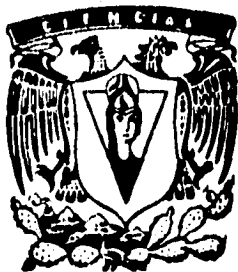
T E S I S

QUE PARA OBTENER EL TITULO DE

A C T U A R I O

P R E S E N T A :

PRUD'HOMME NIEVES EDUARDO FERNANDO



MEXICO, D. F.



1995

FALLA DE ORIGEN

FACULTAD DE CIENCIAS
SECCION ESCOLAR

**TESIS CON
FALLA DE ORIGEN**



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

TESIS CON FALLA DE ORIGEN



UNIVERSIDAD NACIONAL
AUTÓNOMA DE
MÉXICO

M. EN C. VIRGINIA ABRIN BATULE

Jefe de la División de Estudios Profesionales

Facultad de Ciencias

Presente

Los abajo firmantes, comunicamos a Usted, que habiendo revisado el trabajo de Tesis que realiz(ó)ron EL pasante(s) _____

PRUD' HOMME NIEVES EDUARDO FERNANDO

con número de cuenta 8732465-5 con el Título: _____

"ELEMENTOS DE PRONOSTICACION BAYESIANA"

Otorgamos nuestro Voto Aprobatorio y consideramos que a la brevedad deberá presentar su Examen Profesional para obtener el título de ACTUARIO

GRADO	NOMBRE(S)	APELLIDOS COMPLETOS	FIRMA
MAT. Director de Tesis	JORGE FRANCISCO DE LA VEGA	GONGORA	<i>[Signature]</i>
MAT.	MARGARITA ELVIRA CHAVEZ	CANO	<i>[Signature]</i>
M. EN C.	DINO MON VAZQUEZ		<i>[Signature]</i>
MAT. Suplente	HUGO VILLASENOR	HERNANDEZ	<i>[Signature]</i>
M. EN C. Suplente	JOSE GABRIEL HUERTA	GOMEZ	<i>[Signature]</i>

A mis dos ángeles, Ofelia y Columba. A ellas y sólo a ellas puedo dedicar este trabajo.

Gracias...

Al Mat. Jorge Francisco de la Vega Góngora por brindarme su apoyo de una manera franca e incondicional. Porque su invitación a ir más allá en el estudio de la estadística fue el génesis de esta tesis. Por ser valiente y aceptar ser director de este proyecto. Por saber detectar y criticar las fallas, por crear muchas de las ideas y patrocinar muchos de los medios. Pero sobre todo, por dejarme robar su tiempo y con parte de ese tiempo ganarme su amistad.

A mis sinodales, la Mat. Margarita Elvira Chavez Cano, el Mat. Hugo Villaseñor Hernandez, el M. en C. Dino Mon Vazquez y el M. en C. Jose Gabriel Huerta Gomez, por haber aceptado revisar este modesto trabajo y haber aportado tiempo y comentarios, sumamente valiosos para su consumación.

A la Facultad de Ciencias por ser mi espacio y mi templo.

A la Universidad Nacional Autónoma de México por ser la fuente y el oasis donde he encontrado miles de respuestas.

A mis dos ángeles, por estar siempre tan cerca en todos los tiempos. Por reservarme de sus días mucho más que el pensamiento y el corazón. Por enseñarme a ser lo que soy y darme la certeza de que siempre habrá una luz prendida para cuando yo regrese.

A Chucho y a Peter por estar desde el principio.

A mis hermanos, los otros Seis, por alentarme con dosis de desvelo y complicidad.

Salomon saith: *There is no new thing upon the earth. So that as Plato had an imagination, that all knowledge was but remembrance; so Salomon giveth his sentences, that all novelty is but oblivion.*

Francis Bacon: *Essays LVIII*

INDICE

INTRODUCCION	i
CAPITULO I	
El Enfoque Bayesiano	
1.1 Introducción	1
1.2 Enfoque Bayesiano	2
1.3 Distribución Inicial	7
CAPITULO II	
Pronosticación	
2.1 Sistemas de Pronosticación	19
2.1.1 Características de los sistemas de Pronosticación	20
2.2 Métodos de Pronosticación	22
2.2.1 Modelos de Regresión	23
2.2.2 Modelos de Series de tiempo	25
2.2.3 Limitaciones de los Modelos de Regresión y de Series de tiempo	31
2.3 Características de los Modelo Bayesianos de Pronosticación	31
CAPITULO III	
Modelo Polinomial de Primer Orden	
3.1 Introducción	36
3.2 Distribuciones de Pronóstico	45
3.3 El Modelo Constante	49
3.4 Especificación de la Varianza de Transición $\{\sigma_w^2\}$	55
3.4.1 Elección por factor de descuento	58
EJEMPLO: Tipo de Cambio	61
3.5 Estimación de la Varianza Observacional $\{\sigma_v^2\}$	78
EJEMPLO: Tipo de Cambio	87
CAPITULO IV	
El Modelo Dinámico Lineal General	
4.1 Introducción	104
4.1.1 El MDL Univariado	107

4.2	Distribución de Pronosticación	114
4.3	Varianza Observacional Desconocida	122
4.3.1	Distribuciones de Pronóstico con Varianza Observacional Desconocida	127
4.4	Distribuciones de los errores	129
4.5	Filtrado	132
	EJEMPLO: Demanda Real de Dinero	138

CONCLUSIONES

APENDICE

BIBLIOGRAFIA	179
---------------------	------------

INTRODUCCION

En la sociedad moderna es indiscutible la importancia de la pronosticación de diversas series de tiempo, particularmente las series económicas, para tomar decisiones a todos los niveles. En lo anterior radica el interés por el desarrollo y la difusión de los métodos de pronosticación.

Existe una gran variedad de métodos para analizar una serie de tiempo, abarcando diversos grados de precisión. Hay métodos constructivos, como los propuestos por Box y Jenkins (1970) y los hay también de carácter más bien descriptivo y basados en técnicas empíricas.

Sin embargo, pocos métodos de pronosticación contemplan la implementación de una metodología que incorpore apreciaciones subjetivas sin modificar la estructura matemática del modelo. Uno de los primeros avances en este sentido lo constituyen los filtros de Kalman (1960), que consisten en la combinación de dos estimadores independientes para formar un estimado ponderado o predicción. Un estimado puede ser una predicción a priori o un estimado basado en un conocimiento inicial, y el otro una predicción basada en la nueva información disponible. El propósito de los filtros de Kalman es combinar estas dos piezas de información para obtener un estimador mejorado. Esto es exactamente el proceso utilizado en la metodología bayesiana. Basado en estos filtros, Harrison y Stevens (1971) comenzaron a desarrollar la teoría que describe esta tesis. Ellos mismos han mostrado que todos los métodos de pronosticación son casos particulares de los filtros de Kalman.

Los acontecimientos políticos y sociales influyen de diversas formas en el comportamiento de las variables económicas y alteran sus perspectivas. Así el índice de

precios y cotizaciones de la Bolsa Mexicana de Valores continuamente presenta fluctuaciones causadas por eventos de orden político, como la posible candidatura a la presidencia de Camacho Solís, de orden social, como son la realización de múltiples marchas y hasta de orden deportivo, como si juega o no la selección nacional de fútbol. Los métodos bayesianos aunados con una interpretación subjetiva de la probabilidad permiten asignar probabilidades a eventos como los anteriores que de otro modo no podrían ser considerados siquiera eventos aleatorios.

Ningún método de pronóstico es perfecto, porque el futuro es incierto. Lo que se pretende en esta tesis es presentar una técnica de pronóstico desarrollada por Harrison y West y aplicarla a dos problemas económicos prácticos: la pronóstico del tipo de cambio y la demanda de dinero. En la tesis se enfocarán ambos problemas más como ejercicios matemáticos que como problemas económicos. La intención es simplemente mostrar como se aplica la teoría expuesta, sin pretender tener un conocimiento exhaustivo de los problemas.

La tesis está basada principalmente en los primeros 5 capítulos del libro de Harrison y West (1979), y se buscó hacer una exposición más clara y sistemática de los conceptos y teoremas presentados en aquel libro, presentando la estadística y metodología bayesiana como un proceso de aprendizaje. Es oportuno mencionar que bajo el nuevo enfoque económico de las "expectativas racionales", esta metodología de pronóstico es una herramienta fundamental, ya que muchas ecuaciones económicas dependen de una variable X_t por medio de su valor esperado en el tiempo $t + 1$.

En el primer capítulo de la tesis se describen los principales conceptos de la estadística bayesiana. Se introducen los conceptos de distribución inicial, distribución posterior, verosimilitud, y se resalta el proceso de incorporación de la información nueva como un proceso de aprendizaje.

En el capítulo siguiente se establecen las características de los sistemas de pronóstico. Se da una visión panorámica acerca de las series de tiempo y de los métodos de pronóstico. Al final del capítulo son presentados los modelos bayesianos de pronóstico.

En el tercer capítulo se introduce el modelo dinámico lineal en su forma más simple y se muestra su mecánica. Estableciendo ciertos supuestos se obtienen resultados importantes

que son útiles para comprender el funcionamiento de modelos más generales y sus posibles aplicaciones. Se muestra al problema de pronóstico del tipo de cambio como un ejemplo de aplicación práctica del modelo lineal.

En el capítulo cuatro se introduce el mismo modelo pero en su forma más general demostrando los resultados más importantes y por último se propone un ejemplo de aplicación a la pronóstico de la demanda por dinero, según un modelo macroeconómico.

El objetivo básico de la presente tesis es mostrar a los sinodales que el tesista tiene los elementos necesarios para afrontar un problema, primero en forma deductiva, analizándolo, entendiendo los resultados y ubicándolo en un contexto general; una vez que esto se ha logrado y se ha llegado al manejo de los conceptos, se debe ser capaz de aplicarlo a la solución de problemas concretos.

CAPITULO I

EL ENFOQUE BAYESIANO

1.1 Introducción

Comunmente un problema de inferencia estadística comienza cuando se tiene una muestra aleatoria $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)$ con su correspondiente distribución de probabilidad $F_{\mathbf{Y}}(\mathbf{y}|\Theta)$, cuyo vector de parámetros $\Theta = (\theta_1, \theta_2, \dots, \theta_p)$ es desconocido y se quiere utilizar la información proporcionada por los datos observados \mathbf{y}^1 para determinar las características de tal distribución y poder hacer conclusiones sobre la población de la que se tomó la muestra. Pero existen alternativas para hacer inferencia estadística: se puede optar por un tratamiento clásico o Frecuentista, o bien se puede escoger un enfoque Bayesiano.

La inferencia estadística bajo el enfoque clásico tiene las siguientes características:

- (1) Los estimadores y procedimientos de prueba son evaluados en términos de sus propiedades en un gran número de observaciones.
- (2) La probabilidad de un evento es definido en términos del límite de su frecuencia relativa.
- (3) No existe una forma de integrar formalmente información no procedente de la muestra.

¹ Generalmente una variable aleatoria es denotada tanto por $X(\cdot)$ como por X . Aunque la primer forma es más complicada, enfatiza el hecho de que una variable aleatoria es una función. En este trabajo se tratará de usar letras mayúsculas para denotar variables aleatorias y se usarán las correspondientes letras minúsculas para denotar los valores que tomen tales variables aleatorias.

Cuando estimamos un parámetro bajo el enfoque clásico, es deseable tener un estimador que posea las propiedades de *insesgamiento*, *eficiencia* y *consistencia* (Pindyck y Rubinfeld [1]). Dentro del conjunto de estimadores insesgados se tiene preferencia por un estimador de *varianza mínima*. En general el enfoque frecuentista busca técnicas que tengan una alta probabilidad de obtener el resultado correcto y la probabilidad está definida en términos del límite de una frecuencia relativa.

En el contexto Bayesiano destacan los siguientes aspectos:

- (1) La probabilidad es *subjetiva* y la asignación basada en frecuencias relativas es una asignación más, sin ser la única.
- (2) Los *parámetros* Θ son interpretados como *variables aleatorias* en el sentido de que se tiene incertidumbre acerca de su verdadero valor.
- (3) En este sentido, se les puede asociar una *distribución de probabilidad*² que represente el *grado de conocimiento* que se posee sobre tales parámetros y tal distribución es la fuente para toda inferencia que se haga sobre Θ . Esta distribución puede incluir información extra no contenida en la muestra.

Ambos enfoques permiten resolver problemas de inferencia con cierta validez desde el punto de vista formal. La diferencia puede consistir en la interpretación. Sin embargo en muchos problemas las soluciones son similares. Este trabajo se desenvuelve bajo el enfoque bayesiano, presentándolo como una alternativa pero no como la solución.

1.2 Enfoque Bayesiano

En esta sección se darán las principales ideas de la inferencia bayesiana.

² La estructura del espacio de probabilidad básico (Ω, \mathcal{A}, P) presenta las siguientes modificaciones. \mathcal{A} no tiene que ser una σ -álgebra de eventos de Ω , sólo se pide que sea una álgebra. La medida P entonces ya no está definida en forma única; hay una diversidad de medidas válidas (resultado de ULAM) y por último no tiene que estar definida en cada evento $E \in \mathcal{A}$, basta con definirla en un subconjunto de \mathcal{A} para que tenga sentido.

Definición 1.1

La distribución de probabilidad que representa el grado de conocimiento que se tiene inicialmente sobre el parámetro Θ es llamada *distribución a priori* o *distribución inicial*.

La distribución inicial puede basarse en datos de muestras pasadas, en la observación casual o en consideraciones teóricas. El enfoque Bayesiano debe su nombre al teorema de Bayes que enlaza la información inicial con la información muestral, contenida en la función de verosimilitud, generando la *distribución a posteriori* o *distribución final*.

Teorema 1.1 (de Bayes)

Sea $\mathbf{y}=(y_1, y_2, \dots, y_n)$ un vector de n observaciones obtenidos de una muestra aleatoria cuya función de densidad de probabilidad $f_{\mathbf{y}}(\mathbf{y}|\Theta)$ depende del parámetro $\Theta=(\theta_1, \theta_2, \dots, \theta_m)$. Supóngase que Θ tiene una distribución inicial cuya función de densidad de probabilidad es $f_{\Theta}(\theta)$, entonces

$$f_{\mathbf{y},\Theta}(\mathbf{y},\theta) = f_{\Theta|\mathbf{y}}(\theta|\mathbf{y})f_{\mathbf{y}}(\mathbf{y}) = f_{\mathbf{y}|\Theta}(\mathbf{y}|\theta)f_{\Theta}(\theta), \quad (1.1)$$

por lo tanto, la densidad condicional de Θ dado $\mathbf{Y}=\mathbf{y}$, llamada también función de densidad a posteriori o final es

$$f_{\Theta|\mathbf{y}}(\theta|\mathbf{y}) = \frac{f_{\mathbf{y}|\Theta}(\mathbf{y}|\theta)f_{\Theta}(\theta)}{f_{\mathbf{y}}(\mathbf{y})} \quad (1.2)$$

con $f_{\mathbf{y}}(\mathbf{y}) > 0$.

Si se piensa que $f_{\mathbf{y}}(\mathbf{y})$ es una constante c como función de θ , lo anterior es equivalente a tener

$$f_{\Theta|\mathbf{y}}(\theta|\mathbf{y}) = \left(\frac{1}{c}\right)f_{\mathbf{y}|\Theta}(\mathbf{y}|\theta)f_{\Theta}(\theta). \quad (1.2.a)$$

Si se sustituye la igualdad por una expresión de proporcionalidad entonces

$$f_{\Theta|\mathbf{y}}(\theta|\mathbf{y}) \propto f_{\mathbf{y}|\Theta}(\mathbf{y}|\theta)f_{\Theta}(\theta). \quad (1.2.b)$$

$f_{Y|\theta}(y|\theta)$ considerada como función de θ se le llama *función de verosimilitud* y se denota con $L(\theta|y)$ para remarcar su dependencia de θ . La función de verosimilitud es a través de la cual los datos modifican la información inicial sobre θ ; la constante de normalización c es tal que $\int_{-\infty}^{\infty} dF_{\theta|Y}(\theta|y) = 1$.

De acuerdo a (1.2.b) se puede ver que la información final esta integrada tanto por la información inicial como por la información proporcionada por la muestra y es proporcional al producto de ambas, esto es

Dist. Posterior \propto Dist. Inicial \times Verosimilitud.

Así, puede verse que el enfoque Bayesiano proporciona un proceso de aprendizaje a través de la experiencia. De manera recursiva se considera a la distribución posterior como distribución a priori y se le combina con la información de una nueva muestra obteniéndose una nueva distribución posterior. Entonces, si se toma un conjunto de muestras aleatorias Y_1, \dots, Y_k se tendrá que la distribución posterior dado el primer vector de observaciones es

$$f_{\theta|Y_1}(\theta|y_1) \propto f_{\theta}(\theta) f_{Y_1|\theta}(y_1|\theta),$$

similarmente

$$f_{\theta|Y_1, Y_2}(\theta|y_1, y_2) \propto f_{\theta|Y_1}(\theta|y_1) f_{Y_2|\theta}(y_2|\theta).$$

En consecuencia

$$f_{\theta|Y_1, Y_2}(\theta|y_1, y_2) \propto f_{\theta}(\theta) f_{Y_1|\theta}(y_1|\theta) f_{Y_2|\theta}(y_2|\theta).$$

Siguiendo del mismo modo se obtiene que

$$f_{\theta|Y_1, \dots, Y_k}(\theta|y_1, \dots, y_k) \propto f_{\theta}(\theta) f_{Y_1|\theta}(y_1|\theta) \dots f_{Y_k|\theta}(y_k|\theta). \quad (1.2.c)$$

Hay que resaltar que el teorema de Bayes es aplicable bajo las distintas interpretaciones de probabilidad porque es una afirmación cuya validez se deduce de los axiomas de esas interpretaciones. Los problemas subyacentes están relacionados con el significado de la probabilidad y con la elección necesaria de una distribución inicial. El problema del

³ En el caso en que θ es un vector aleatorio discreto se debe cumplir que $\sum_{\theta} f_{\theta|Y}(\theta|y) = 1$.

significado de la probabilidad es tratado a detalle por Bruno de Finetti [2], Dennis V. Lindley [4] y Leonard J. Savage [5], entre otros.

A continuación se dará un ejemplo (Box y Tiao [3]) en el que la forma de asignar probabilidades a los eventos es objetiva, de acuerdo a leyes bien determinadas (Leyes de Mendel), y en el que la distribución inicial es conocida.

EJEMPLO: El experimento de Mendel

Se realizan ensayos con ratones de color negro y de color café. Se sabe que el genotipo de los ratones de color negro puede ser BB, (homocigotos), o bien puede ser Bb, (heterocigoto). Por otra parte los ratones cafés sólo poseen genotipo bb. De acuerdo a las leyes de la genética, las probabilidades de distintas mezclas está dada por la siguiente tabla

Ascendencia	Descendencia		
	BB	Bb	bb
BB, bb	0	1	0
Bb, bb	0	½	½
Bb, Bb	¼	½	¼

Tabla 1.1

Se toma un ratón negro r cuyo genotipo se desconoce pero que se sabe descende de dos ratones Bb. Θ es la variable aleatoria que toma el valor de 0 si el genotipo del ratón muestra es BB, toma el valor de 1 si es Bb y 2 si es bb. De las probabilidades del último renglón de la tabla se obtiene que la distribución inicial para Θ es

$$\begin{aligned}
 f_{\omega}(0) &= P(\Theta = 0 | r \text{ es negro, padres Bb}) \\
 &= P(\text{BB} | r \text{ es negro, padres Bb}) \\
 &= \frac{P(\text{BB}, r \text{ es negro} | \text{padres Bb})}{P(r \text{ es negro} | \text{padres Bb})} \\
 &= \frac{P(\text{BB} | \text{padres Bb})}{P(\text{BB} | \text{padres Bb}) + P(\text{Bb} | \text{padres Bb})} = \frac{\frac{1}{4}}{\frac{1}{4} + \frac{1}{2}} = \frac{1}{3}
 \end{aligned}$$

$$\begin{aligned}
f_{\theta}(1) &= P(\Theta = 1 | r \text{ es negro, padres Bb}) \\
&= P(\text{Bb} | r \text{ es negro, padres Bb}) \\
&= \frac{P(\text{Bb}, r \text{ es negro} | \text{padres Bb})}{P(r \text{ es negro} | \text{padres Bb})} \\
&= \frac{P(\text{Bb} | \text{padres Bb})}{P(\text{BB} | \text{padres Bb}) + P(\text{Bb} | \text{padres Bb})} = \frac{\frac{1}{2}}{\frac{1}{4} + \frac{1}{2}} = \frac{2}{3}.
\end{aligned}$$

$$\begin{aligned}
f_{\theta}(2) &= P(\Theta = 2 | r \text{ es negro, padres Bb}) \\
&= P(\text{bb} | r \text{ es negro, padres Bb}) = 0.
\end{aligned}$$

Posteriormente ese ratón es unido con un ratón café, el cuál se sabe sólo puede ser de genotipo bb, obteniéndose 7 descendientes negros. Si $\mathbf{Y} = (Y_1, \dots, Y_7)$ denota la descendencia, cuyas componentes toman el valor de 1 si el descendiente es negro y el valor de 0 si es café, entonces la función de verosimilitud $L(\theta | \mathbf{y})$, toma los valores

$$\begin{aligned}
L(\theta | y_1 = 1, \dots, y_7 = 1) &= f_{\mathbf{Y}}(\mathbf{y} = (1, \dots, 1) | \Theta = \theta) \\
&= \begin{cases} f_{\mathbf{Y}}(\mathbf{y} = (1, \dots, 1) | \Theta = 0) = 1 & \theta = 0, \\ f_{\mathbf{Y}}(\mathbf{y} = (1, \dots, 1) | \Theta = 1) = \prod_{i=1}^7 \frac{1}{2} = (\frac{1}{2})^7 & \theta = 1, \\ f_{\mathbf{Y}}(\mathbf{y} = (1, \dots, 1) | \Theta = 2) = 0 & \theta = 2 \end{cases}
\end{aligned}$$

Entonces utilizando el teorema de Bayes se tiene que

$$\begin{aligned}
f_{\Theta | \mathbf{Y}}(0 | \mathbf{y} = (1, \dots, 1)) &\propto \frac{1}{3} \times 1 = \frac{1}{3}, \\
f_{\Theta | \mathbf{Y}}(1 | \mathbf{y} = (1, \dots, 1)) &\propto \frac{1}{3} \times (\frac{1}{2})^7 = \frac{1}{3} \times (\frac{1}{2})^6, \\
f_{\Theta | \mathbf{Y}}(2 | \mathbf{y} = (1, \dots, 1)) &\propto 0 \times 0 = 0.
\end{aligned}$$

Puesto que $\sum_{\theta=0}^2 f_{\Theta | \mathbf{Y}}(\theta | \mathbf{y} = (1, \dots, 1)) = 1$, se obtiene la constante de normalización, así

$$f_{\text{av}}(\theta | y = (1, \dots, 1)) = \begin{cases} \frac{2^6}{2^6 + 1} & \theta = 0, \\ \frac{1}{2^6 + 1} & \theta = 1, \\ 0 & \theta = 2. \end{cases}$$

De acuerdo a estos resultados es más probable, y por mucho, que el genotipo del ratón escogido sea BB. Aquí se hace patente que por medio de la experiencia se obtuvieron nuevos elementos que modificaron el conocimiento que sobre el genotipo del ratón se tenían, esto es el proceso de aprendizaje a través de la experiencia.

■

1.3 Distribución inicial

Hay casos en que la distribución inicial contiene información previa sobre Θ , pero en otros se desea utilizar el enfoque bayesiano a pesar de que no se disponga de tal información. Berger [6] da distintos métodos para encontrar una distribución inicial cuando hay información disponible. A tales distribuciones se les llama *informativas*. En otro caso se llaman *distribuciones iniciales no informativas*. Un papel importante en la selección de una distribución a priori está relacionada con la forma de la distribución posterior.

Definición 1.2

Una distribución inicial, que después de combinarla con la función de verosimilitud a través del teorema de Bayes, produce una distribución posterior de la misma forma es conocida como *distribución inicial conjugada*. Si una distribución inicial conjugada es escogida dentro de la clase de distribuciones con la misma forma funcional que la verosimilitud entonces se le llama *distribución a priori conjugada natural*.

Las distribuciones iniciales conjugadas son populares por su manejo algebraico conveniente y porque pueden representar una gran diversidad de opiniones a priori.

Un ejemplo de una familia conjugada natural es la distribución normal, como se muestra en el siguiente

Lema 1.1

Sea $Y = (Y_1, \dots, Y_n)$ una muestra aleatoria tal que $Y_i \sim N[\mu, \sigma^2]$ $i=1, \dots, n$ con μ desconocida y σ^2 conocida. Supongase que se escoge una distribución a priori informativa para M , $M \sim N[m, s^2]$

$$f_M(\mu) = \frac{1}{\sqrt{2\pi s^2}} \exp\left\{-\frac{(\mu - m)^2}{2s^2}\right\}$$

Entonces la función de densidad a posteriori para M es

$$f_{M|Y}(\mu|y) = \frac{1}{\sqrt{2\pi \bar{s}^2}} \exp\left\{-\frac{(\mu - \bar{m})^2}{2\bar{s}^2}\right\}$$

con

$$\bar{m} = \frac{s^2 n \hat{\mu} + \sigma^2 m}{s^2 n + \sigma^2} \quad \text{y} \quad \bar{s}^2 = \frac{s^2 \sigma^2}{s^2 n + \sigma^2},$$

$$\text{donde } \hat{\mu} = \frac{1}{n} \sum_{i=1}^n y_i.$$

Demostración.

La función de verosimilitud esta dada por

$$\begin{aligned} L(\mu|y) &= f_{Y|M}(y|\mu) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y_i - \mu)^2\right\} \\ &= (2\pi\sigma^2)^{-n} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2\right\} \end{aligned} \tag{1.3}$$

pero

$$\begin{aligned}
\sum_{i=1}^n (y_i - \mu)^2 &= \sum_{i=1}^n [(y_i - \hat{\mu}) - (\mu - \hat{\mu})]^2 \\
&= \sum_{i=1}^n (y_i - \hat{\mu})^2 - 2 \sum_{i=1}^n [(y_i - \hat{\mu})(\mu - \hat{\mu})] + \sum_{i=1}^n (\mu - \hat{\mu})^2 \\
&= \sum_{i=1}^n (y_i - \hat{\mu})^2 - 2[n\hat{\mu}\mu - n\hat{\mu}^2 - n\hat{\mu}\mu + n\hat{\mu}^2] + n(\mu - \hat{\mu})^2
\end{aligned}$$

con $\hat{\mu} = \frac{1}{n} \sum_{i=1}^n y_i$. Por lo tanto

$$\begin{aligned}
L(\mu|\mathbf{y}) &\propto \exp\left\{-\frac{1}{2\sigma^2} \left[n(\mu - \hat{\mu})^2 + \sum_{i=1}^n (y_i - \hat{\mu})^2 \right]\right\} \\
&\propto \exp\left\{-\frac{n}{2\sigma^2} [\mu - \hat{\mu}]^2\right\}
\end{aligned} \tag{1.3.a}$$

Utilizando (1.2.b) en el *teorema 1.1* para multiplicar $L(\mu|\mathbf{y})$ en (1.3.a) por $f_M(\mu)$ e ignorando las constantes se obtiene que la función de densidad a posteriori para M está dada por la expresión

$$\begin{aligned}
f_{M|Y}(\mu|\mathbf{y}) &\propto f_M(\mu) f_{Y|M}(\mathbf{y}|\mu) \\
&\propto \exp\left\{-\frac{n}{2\sigma^2} (\mu - \hat{\mu})^2\right\} \times \exp\left\{-\frac{1}{2s^2} (\mu - m)^2\right\} \\
&\propto \exp\left\{-\frac{1}{2} \left[\frac{n(\mu - \hat{\mu})^2}{\sigma^2} + \frac{(\mu - m)^2}{s^2} \right]\right\}
\end{aligned} \tag{1.4}$$

pero

$$\begin{aligned}
\frac{n(\mu - \hat{\mu})^2}{\sigma^2} + \frac{(\mu - m)^2}{s^2} &= \frac{s^2 n(\mu - \hat{\mu})^2 + \sigma^2(\mu - m)^2}{\sigma^2 s^2} \\
&= \frac{s^2 n(\mu^2 - 2\mu\hat{\mu} + \hat{\mu}^2) + \sigma^2(\mu^2 - 2\mu m + m^2)}{\sigma^2 s^2} \\
&= \frac{\mu^2(s^2 n + \sigma^2) - 2\mu(s^2 n\hat{\mu} + \sigma^2 m) + s^2 n\hat{\mu}^2 + \sigma^2 m^2}{\sigma^2 s^2} \\
&= \left[\frac{s^2 n + \sigma^2}{\sigma^2 s^2} \right] \left[\mu^2 - 2\mu \frac{(s^2 n\hat{\mu} + \sigma^2 m)}{s^2 n + \sigma^2} + \frac{s^2 n\hat{\mu}^2 + \sigma^2 m^2}{s^2 n + \sigma^2} \right] \\
&= \left[\frac{s^2 n + \sigma^2}{\sigma^2 s^2} \right] \left[\mu^2 - 2\mu \frac{(s^2 n\hat{\mu} + \sigma^2 m)}{s^2 n + \sigma^2} + \left(\frac{s^2 n\hat{\mu}^2 + \sigma^2 m^2}{s^2 n + \sigma^2} \right)^2 \right. \\
&\quad \left. - \left(\frac{s^2 n\hat{\mu}^2 + \sigma^2 m^2}{s^2 n + \sigma^2} \right)^2 + \frac{s^2 n\hat{\mu}^2 + \sigma^2 m^2}{s^2 n + \sigma^2} \right] \\
&= \left[\frac{s^2 n + \sigma^2}{\sigma^2 s^2} \right] \left[\left[\mu - \left(\frac{s^2 n\hat{\mu} + \sigma^2 m}{s^2 n + \sigma^2} \right) \right]^2 \right. \\
&\quad \left. - \left(\frac{s^2 n\hat{\mu}^2 + \sigma^2 m^2}{s^2 n + \sigma^2} \right)^2 + \frac{s^2 n\hat{\mu}^2 + \sigma^2 m^2}{s^2 n + \sigma^2} \right]
\end{aligned}$$

De sustituir lo anterior en (1.4) la densidad posterior se obtiene de

$$f_{\text{MY}}(\mu|y) \propto \exp \left\{ -\frac{1}{2 \left[\frac{\sigma^2 s^2}{s^2 n + \sigma^2} \right]} \left(\mu - \left[\frac{s^2 n\hat{\mu} + \sigma^2 m}{s^2 n + \sigma^2} \right] \right)^2 \right\}$$

Otro ejemplo importante de familia conjugada corresponde al siguiente ■

Lema 1.2

Sea X una variable aleatoria tal que $X \sim \text{Bin}[n, \theta]$ con n dada y θ desconocida. Supongase que se escoge una distribución a priori informativa para Θ , $\Theta \sim \text{Beta}[\alpha, \beta]$

$$f_{\theta}(\theta) = \frac{1}{B(\alpha, \beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1} I_{(0,1)}(\theta)^4.$$

Entonces la función de densidad a posteriori para Θ es

$$f_{\theta|x}(\theta|x) = \frac{1}{B(\bar{\alpha}, \bar{\beta})} \theta^{\bar{\alpha}-1} (1-\theta)^{\bar{\beta}-1} I_{(0,1)}(\theta)$$

con

$$\bar{\alpha} = \alpha + x \text{ y } \bar{\beta} = \beta + n - x.$$

Demostración.

La función de densidad a priori para Θ esta dada por

$$\begin{aligned} f_{\theta}(\theta) &= \frac{\Gamma(\alpha + \beta) \theta^{\alpha-1} (1-\theta)^{\beta-1}}{\Gamma(\alpha) \Gamma(\beta)} I_{(0,1)}(\theta) \\ &\propto \theta^{\alpha-1} (1-\theta)^{\beta-1} I_{(0,1)}(\theta) \end{aligned}$$

La función de verosimilitud esta dada por

$$\begin{aligned} L(\theta|x, n) &= \binom{n}{x} \theta^x (1-\theta)^{n-x} I_{\{0, \dots, n\}}(x) \\ &\propto \theta^x (1-\theta)^{n-x} I_{\{0, \dots, n\}}(x) \end{aligned}$$

Utilizando (1.2.b) en el teorema 1.1 para multiplicar $L(\theta|x, n)$ por $f_{\theta}(\theta)$ e ignorando las constantes se obtiene que la f.d.p. a posteriori para θ está dada por la siguiente expresión

$$\begin{aligned} f_{\theta|x}(\theta|x) &\propto \theta^{\alpha-1} (1-\theta)^{\beta-1} \theta^x (1-\theta)^{n-x} \\ &\propto \theta^{\alpha+x-1} (1-\theta)^{\beta+n-x-1} \end{aligned}$$

⁴ Sea Ω un espacio con puntos ω y $A \subset \Omega$. La función indicadora de A , denotada por $I_A(\cdot)$, es la función con dominio Ω y contradominio $\{0, 1\}$ definida por

$$I_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{si } \omega \notin A \end{cases}$$

De aquí se deduce que

$$(\theta | X = x) \sim \text{Beta}[\bar{\alpha}, \bar{\beta}],$$

con

$$\begin{aligned}\bar{\alpha} &= x + \alpha, \\ \bar{\beta} &= \beta + n - x.\end{aligned}$$

El concepto de distribución no informativa puede aclararse a través del siguiente análisis.

Definición 1.3

La *precisión* de una fuente de información se define como el inverso de la varianza de la distribución correspondiente.

Se dice que la información proveniente de cierta fuente es *vaga* o *difusa* si su precisión es pequeña, lo que indica que la varianza es grande. Por el contrario una varianza pequeña implica que la información es precisa.

En el caso tratado en el *lema 1.1*, la precisión de la información a priori es $h_0 = \frac{1}{s^2}$ y la precisión de la información muestral es $h_1 = \frac{n}{\sigma^2}$; la media y varianza de la distribución posterior se pueden reexpresar utilizando las precisiones h_0 y h_1 , siendo

$$\bar{m} = \frac{\frac{s^2 n \hat{\mu} + \sigma^2 m}{s^2 \sigma^2}}{\frac{s^2 n + \sigma^2}{s^2 \sigma^2}} = \frac{h_1 \hat{\mu} + h_0 m}{h_1 + h_0} \quad (1.5)$$

$$\bar{s}^2 = \frac{s^2 \sigma^2}{s^2 n + \sigma^2} = \left(\frac{s^2 n + \sigma^2}{s^2 \sigma^2} \right)^{-1} = (h_1 + h_0)^{-1} = h^{-1}. \quad (1.6)$$

Por lo anterior se puede notar que \bar{m} es un promedio ponderado de m y $\hat{\mu}$ cuyos pesos están dados por funciones de las precisiones de cada fuente de información. En consecuencia, el peso de la información inicial en el valor de la media posterior depende de la precisión h_0 , i.e. la información inicial tendrá más peso en \bar{m} de acuerdo a qué tan

pequeña sea la varianza s^2 . Claramente la precisión de la información posterior h_1 es la suma de las precisiones inicial h_0 y muestral h_S , por lo que la precisión de la distribución posterior siempre será mayor que ambas.

En la información inicial, la ignorancia sobre M se refleja con valores de s^2 muy grandes, y al tomar el caso límite $s^2 \rightarrow \infty$ se tiene $h_0 \rightarrow 0$ y dados (1.5) y (1.6) sucede que

$$\begin{aligned}\bar{m} &= \hat{\mu}, \\ \bar{s}^2 &= \frac{\sigma^2}{n}.\end{aligned}$$

Por lo que

$$f_{MV}(\mu|y) \propto \exp\left\{-\frac{n}{2\sigma^2}(\mu - \hat{\mu})^2\right\}, \quad (1.7)$$

indicando que en este caso límite la media \bar{m} y la varianza \bar{s}^2 sólo dependen de la muestra.

Definición 1.4

Una distribución a priori que no contiene información acerca del parámetro debido a que el conocimiento que sobre él se tiene es vago o difuso se le llama *distribución inicial no-informativa*.

A partir de una distribución a priori no-informativa se obtiene una distribución posterior dominada por la información muestral. Existen situaciones en las que es conveniente usar distribuciones de este tipo.

- (1) Cuando se tiene muy poca información sobre el parámetro y se desea una distribución que refleje esa carencia de información.
- (2) Cuando, a pesar de contar con información sustancial sobre Θ , se considera más objetivo presentar una distribución posterior que sólo contenga la información proporcionada por la muestra y la verosimilitud en vez de desviar los resultados con puntos de vista subjetivos.
- (3) Otro caso se presenta cuando es muy difícil formular una distribución inicial informativa apropiada y una forma de superar esa dificultad es tomar una no-informativa.

Se podría pensar que la distribución no-informativa para M puede ser determinada al hacer tender la varianza s^2 a ∞ , pero conforme lo anterior sucede, $f_M(\mu) \rightarrow 0$. Sin embargo, se nota que la verosimilitud escrita como en (1.3.a) es idéntica a la expresión (1.7), correspondiente a la densidad posterior obtenida a partir del caso límite $h_0 \rightarrow 0$, por lo que si se establece que la función de densidad a priori es igual a una constante

$$f_M(\mu) = k \quad (1.8)$$

la multiplicación de la verosimilitud con esta densidad a priori conduce a una función de densidad a posteriori similar a la expresada en (1.7).

El hecho de que la función de densidad a priori sea una constante se puede interpretar como una densidad uniforme sobre toda la recta real $-\infty < \mu < \infty$ y refleja la ignorancia que sobre la variable aleatoria M se tiene. Sin embargo esta densidad se caracteriza por tener que

$$\int_{-\infty}^{\infty} f_M(\mu) d\mu = \infty \quad (1.9)$$

es decir, no integra a la unidad y es una densidad impropia. En este caso, este hecho no es de grandes consecuencias, porque se vio que conduce a densidades posteriores que sí son propias.

La elección de una distribución inicial informativa puede parecer un punto cuestionable de la subjetividad del enfoque bayesiano, pudiéndose argumentar que si dos investigadores al estudiar un mismo fenómeno escogen diferentes distribuciones iniciales para un mismo parámetro, las distribuciones posteriores obtenidas por cada uno serán diferentes, conduciéndolos a conclusiones contrarias. Sin embargo se debe recordar que el enfoque Bayesiano proporciona un proceso de aprendizaje, así aún cuando se parta de dos distribuciones diferentes, los mismos datos van corrigiendo las desviaciones que pudieran haber sido causadas por la falta de información adecuada, y finalmente existe una convergencia de resultados; esta idea se enuncia en el siguiente

Lema 1.3

Se toman k muestras de tamaño n , $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{in})$, $i = 1, \dots, k$, de una población tal que $Y \sim N[\mu, \sigma^2]$. Inicialmente se elige una distribución inicial para $M \sim N[m_0, s_0^2]$. Dado el vector de observaciones $\mathbf{y}_1 = (y_{11}, y_{12}, \dots, y_{1n})$ se obtiene la distribución posterior $(M|\mathbf{Y}_1 = \mathbf{y}_1) \sim N[m_1, s_1^2]$ de acuerdo al lema 1.1. Tomando esta distribución como inicial, agregando la información proporcionada por un nuevo vector de observaciones, aplicando el lema 1.1 y repitiendo este procedimiento, en el paso k se obtiene que

$$(M|\mathbf{Y}_1 = \mathbf{y}_1, \dots, \mathbf{Y}_k = \mathbf{y}_k) \sim N[m_k, s_k^2]$$

con

$$m_k = \hat{\mathbf{M}}_k \left[1 - \frac{1}{kn \frac{s_0^2}{\sigma^2} + 1} \right] + \frac{m_0}{kn \frac{s_0^2}{\sigma^2} + 1} \quad \forall k \geq 1$$

y

$$s_k^2 = \frac{\sigma^2 s_0^2}{kns_0^2 + \sigma^2} \quad \forall k \geq 1,$$

donde

$$\hat{\mathbf{M}}_k = \frac{1}{k} \sum_{i=1}^k \hat{\mu}_i$$

es el promedio de las k medias muestrales.

Demostración.

Se hace por inducción.

Para $i = 1$ y de acuerdo al lema 1.1 se cumple que $(M|\mathbf{Y}_1 = \mathbf{y}_1) \sim N[m_1, s_1^2]$ con

$$\begin{aligned}
 m_1 &= \frac{n\hat{\mu}_1 s_0^2}{ns_0^2 + \sigma^2} + \frac{m_0 \sigma^2}{ns_0^2 + \sigma^2} \\
 &= \hat{\mu}_1 \left[1 - \frac{1}{n \frac{s_0^2}{\sigma^2} + 1} \right] + \frac{m_0}{n \frac{s_0^2}{\sigma^2} + 1}
 \end{aligned}$$

y

$$s_1^2 = \frac{\sigma^2 s_0^2}{ns_0^2 + \sigma^2}.$$

Suponemos que se cumple para $k-1$, por demostrar que se cumple para k . Por hipótesis, $(M|Y_1 = y_1, \dots, Y_{k-1} = y_{k-1}) \sim N[m_{k-1}, s_{k-1}^2]$ con

$$\begin{aligned}
 m_{k-1} &= \hat{M}_{k-1} \left[1 - \frac{1}{(k-1)n \frac{s_0^2}{\sigma^2} + 1} \right] + \frac{m_0}{(k-1)n \frac{s_0^2}{\sigma^2} + 1} \\
 &= \frac{1}{k-1} \sum_{i=1}^{k-1} \hat{\mu}_i \left[1 - \frac{1}{(k-1)n \frac{s_0^2}{\sigma^2} + 1} \right] + \frac{m_0}{(k-1)n \frac{s_0^2}{\sigma^2} + 1} \\
 &= \sum_{i=1}^{k-1} \frac{n\hat{\mu}_i s_0^2}{(k-1)ns_0^2 + \sigma^2} + \frac{m_0 \sigma^2}{(k-1)ns_0^2 + \sigma^2}
 \end{aligned}$$

y

$$s_{k-1}^2 = \frac{\sigma^2 s_0^2}{(k-1)ns_0^2 + \sigma^2}.$$

De acuerdo al lema 1.1 se tiene que $(M|Y_1 = y_1, \dots, Y_k = y_k) \sim N[m_k, s_k^2]$ con

$$m_k = \frac{s_{k-1}^2 n \hat{\mu}_k + \sigma^2 m_{k-1}}{s_{k-1}^2 n + \sigma^2} \text{ y } s_k^2 = \frac{s_{k-1}^2 \sigma^2}{s_{k-1}^2 n + \sigma^2}$$

Sustituyendo se tiene

$$\begin{aligned}
m_k &= \frac{\left[\frac{\sigma^2 s_0^2}{(k-1)ns_0^2 + \sigma^2} \right] n\hat{\mu}_k + \sigma^2 \left[\sum_{i=1}^{k-1} \frac{n\hat{\mu}_i s_0^2}{(k-1)ns_0^2 + \sigma^2} + \frac{m_0 \sigma^2}{(k-1)ns_0^2 + \sigma^2} \right]}{\left[\frac{\sigma^2 s_0^2}{(k-1)ns_0^2 + \sigma^2} \right] n + \sigma^2} \\
&= \frac{n\hat{\mu}_k \sigma^2 s_0^2 + \sigma^2 \left[\sum_{i=1}^{k-1} n\hat{\mu}_i s_0^2 + m_0 \sigma^2 \right]}{(k-1)ns_0^2 + \sigma^2} \\
&= \frac{n\sigma^2 s_0^2 + \sigma^2 ((k-1)ns_0^2 + \sigma^2)}{(k-1)ns_0^2 + \sigma^2} \\
&= \frac{n\hat{\mu}_k \sigma^2 s_0^2 + \sigma^2 \left[\sum_{i=1}^{k-1} n\hat{\mu}_i s_0^2 + m_0 \sigma^2 \right]}{n\sigma^2 s_0^2 + \sigma^2 ((k-1)ns_0^2 + \sigma^2)} \\
&= \frac{\sum_{i=1}^k n\hat{\mu}_i s_0^2 + m_0 \sigma^2}{kns_0^2 + \sigma^2}
\end{aligned}$$

y

$$\begin{aligned}
s_k^2 &= \frac{\left[\frac{s_0^2 \sigma^2}{(k-1)ns_0^2 + \sigma^2} \right] \sigma^2}{\left[\frac{s_0^2 \sigma^2}{(k-1)ns_0^2 + \sigma^2} \right] n + \sigma^2} \\
&= \sigma^2 \left[\frac{s_0^2 \sigma^2}{s_0^2 \sigma^2 n + \sigma^2 ((k-1)ns_0^2 + \sigma^2)} \right] \\
&= \frac{s_0^2 \sigma^2}{kns_0^2 + \sigma^2}
\end{aligned}$$

De acuerdo al lema anterior se observa que conforme $k \rightarrow \infty$ la precisión de la distribución posterior obtenida aumenta, o en otras palabras la varianza se hace más pequeña

$$\lim_{k \rightarrow \infty} s_k^2 = \lim_{k \rightarrow \infty} \frac{\sigma^2 s_0^2}{kns_0^2 + \sigma^2} = 0 \tag{1.10}$$

y por otra parte la media tiende a tomar el valor promedio de las k medias muestrales

$$\begin{aligned} \lim_{k \rightarrow \infty} (m_k - \hat{M}_k) &= \hat{M}_k \lim_{k \rightarrow \infty} \left(1 - \frac{1}{kn \frac{s_0^2}{\sigma^2} + 1} \right) + m_0 \lim_{k \rightarrow \infty} \frac{1}{kn \frac{s_0^2}{\sigma^2} + 1} - \hat{M}_k \\ &= \hat{M}_k - \hat{M}_k = 0 \end{aligned} \quad (1.11)$$

Claramente ambos límites no dependen de la media y varianza escogidas inicialmente, por lo tanto dos investigadores utilizando los mismos datos obtendrán a la larga los mismos resultados a pesar de partir de distintas distribuciones originales.

CAPITULO II

PRONOSTICACION

2.1 Sistemas de Pronosticación

Un *pronóstico* es una hipótesis, conjetura o visión especulativa sobre algo futuro. En situaciones en que es necesario tomar una decisión que depende de eventos futuros, la pronosticación es una importante ayuda para planear adecuadamente y buscar la reducción del riesgo de tomar una decisión en situaciones de incertidumbre. Un *Sistema de Pronosticación* lo componen tanto las personas que realizan el pronóstico a través de un *método*, como las que utilizan el pronóstico para tomar una decisión.

Ejemplo: Operaciones de la Banca Central

- (A) Entre las operaciones que realiza el Banco de México se cuenta la compra de divisas. Cada divisa d tiene un tipo de cambio que fluctúa a lo largo del tiempo, $S_d(t)$. En el tiempo $t = 0$, el Banco dispone de una cantidad C_0 en dólares para comprar divisas que debe transformar nuevamente en dólares en algún momento en el futuro. El problema de decisión consiste en determinar qué divisas se deben comprar y qué cantidad de cada una, buscando que, al momento de reconvertir a dólares las divisas se obtenga una cantidad $C_t > C_0$. Con este fin es necesario tener expectativas acerca del comportamiento que tendrán los tipos de cambio en el futuro.
- (B) Otro de los problemas a los que se enfrenta el Banco consiste en determinar diariamente la cantidad de billetes y monedas que hay en circulación, para saber si es necesario recoger o sacar dinero a través del Mercado de Valores, y mantener así un equilibrio en las tasas de interés y en el tipo de cambio.

2.1.1 Características de los sistemas de Pronosticación

Un Sistema de Pronosticación se caracteriza por los *elementos temporales*, (periodo, horizonte e intervalo de pronóstico), y la *forma* del pronóstico, (valor, intervalo, distribución).

Los elementos temporales están establecidos de acuerdo a la siguiente definición.

Definición 2.1

Los elementos temporales de un pronóstico son el *periodo* que es la unidad básica de tiempo en la que se realizan los pronósticos, el *horizonte* o número de periodos en el futuro cubiertos por el pronóstico y el *intervalo* o frecuencia con la cual se preparan los pronósticos.

Los elementos temporales los determina el problema de decisión. En el ejemplo (A) lo que interesa saber es cuál será el tipo de cambio en el momento de la revalorización, que puede ser en un mes. En el ejemplo (B), diario tiene que hacerse un pronóstico de la demanda de billetes y monedas para determinar las operaciones de Mercado Abierto, (OMA). Así, en el ejemplo (A), el periodo del pronóstico es de un mes, y en el (B) de un día. En (A) si se tiene proyectado revalorizar en 5 meses, se requieren pronósticos para los siguientes 5 meses, el horizonte del pronóstico es de 5 periodos. El intervalo puede o no ser de la misma longitud que el periodo, revisandose los pronósticos semana a semana utilizando los datos más recientes.

Una serie de tiempo es un conjunto de observaciones x_t , cada una registrada en un tiempo t específico⁵. Para el análisis, se establece que una serie de tiempo $\{x_t, t \in T\}$ es una realización de la familia de variables aleatorias $\{X_t, t \in T\}$. Esta consideración sugiere modelar los datos como una realización (o parte de una realización), de un *proceso estocástico*.

Definición 2.2

⁵Una serie de tiempo discreta x_t es aquella en la que el conjunto ordenado $T = \{t: \text{se presenta una observación en } t\}$ es un conjunto discreto. Una serie de tiempo continua $x(t)$ se presenta cuando las observaciones se registran continuamente sobre un intervalo de tiempo $T = [t_0, t_f]$.

Sea (Ω, \mathcal{A}, P) un espacio de probabilidad, \mathfrak{R} el campo de los números reales y sea T un conjunto ordenado. Un proceso estocástico es una función $X(t, \omega): T \times \Omega \rightarrow \mathfrak{R}$, tal que para cada t fija, $X_t(\omega) = X(t, \omega)$ es una variable aleatoria en (Ω, \mathcal{A}, P) .

Generalmente en el análisis de series de tiempo discretas, el conjunto ordenado T se toma igual al conjunto de los números *naturales* o al de los *enteros*.

Definición 2.3

Para cada $\omega \in \Omega$ fijo, a la función $X_t(\omega) = X(\cdot, \omega): T \rightarrow \mathfrak{R}$ se le llama *realización* o *senda muestral* del proceso $\{X_t, t \in T\}$.

Si bien el término "serie de tiempo" se utiliza para nombrar tanto a los datos registrados como al proceso que los genera, es importante darse cuenta que cuando se ve una tabla de registros de una serie de tiempo, por ejemplo la correspondiente a la tasa anual de crecimiento de una economía, conceptualmente se está viendo una tabla de $X(t, \omega)$ con ω fijo, es decir, una realización. Al conjunto de todas las posibles realizaciones se le llama *ensamble de realizaciones*.

En el esquema de series de tiempo, la pronosticación consiste en hacer conjeturas sobre el desarrollo futuro de un proceso $\{X_t, t \in T\}$ correspondiente a una variable de interés, es decir, especular sobre el valor que toman variables aleatorias en el futuro. Por ello, la forma en que un pronóstico es proporcionado, puede ser una de las tres siguientes

- (1) Estimación del *Valor Esperado* de la variable.
- (2) *Intervalo* de confianza.
- (3) *Distribución* de Probabilidad.

En el ejemplo, si se requiere saber el valor esperado de C_t , se utiliza la función de pronóstico

$$f_t(k) = E[S_d(k)] \quad \forall d.$$

Si $\{X_t\}$ es la serie de tiempo de la variable de interés, al pronóstico sobre X_{t+k} con $k > 0$ realizado en el tiempo t se le llama *pronóstico de k periodos en el futuro* y se le denota por

$$f_t(k).$$

Si los valores que tomará la serie en los k periodos siguientes al tiempo t no serán observables y sólo se puede especular sobre el valor acumulado $\sum_{i=1}^k X_{t+i}$ o sobre $\sum_{i=j}^k X_{t+i}$ con $k \geq j \geq 1$, a los pronósticos realizados se les llama *pronóstico guía de horizonte k* en el primer caso

$$f_t^H(k)$$

y *pronóstico guía de horizonte k diferido $j-1$ periodos*

$$f_t^H(j,k)$$

en el segundo caso.

2.2 Métodos de Pronosticación

Los *Métodos de Pronosticación* pueden clasificarse en *Cuantitativos*, si involucran el uso de métodos matemáticos y estadísticos, y *Cualitativos*, si la información utilizada para realizar los pronósticos es procesada en forma subjetiva.

Los métodos cuantitativos se caracterizan por definir explícitamente cómo es determinado el pronóstico, pueden ser muy informales e intuitivos, o bien estar completamente basados en principios estadísticos. Involucran el análisis de los datos históricos en busca del proceso estocástico que genera la variable en estudio y asumiendo que el modelo es estable, se utiliza el conocimiento obtenido para extrapolar hacia el futuro. Dos tipos básicos de modelos son utilizados: los modelos de *serie de tiempo* y los modelos *causales*.

2.2.1 Modelos de Regresión

Los *Modelos Causales* centran su atención en la relación entre la serie de tiempo de interés y otras series. Si estas series están correlacionadas con la variable de interés y si parece haber una causa para esta correlación, se puede construir un modelo estadístico que describa tal relación. El modelo más importante de este tipo es el modelo de *Regresión*.

El *Análisis de Regresión* es una técnica estadística que busca describir la relación entre ciertas variables independientes llamadas explicativas o regresores y la variable de interés.

El modelo de *Regresión Lineal* está especificado por la siguiente ecuación

$$\mathbf{Y} = \mathbf{X}\Theta + \varepsilon$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_T \end{bmatrix} = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1K} \\ X_{21} & X_{22} & \cdots & X_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ X_{T1} & X_{T2} & \cdots & X_{TK} \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_K \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_T \end{bmatrix}$$

donde

\mathbf{Y} es el $(T \times 1)$ vector aleatorio de *observaciones* para $t=1,2,\dots,T$. de la variable de interés.

\mathbf{X} es la $(T \times K)$ matriz de *diseño* o de *regresores*.

Θ es el $(K \times 1)$ vector de *parámetros* desconocidos que se quieren estimar.

ε es el $(T \times 1)$ vector aleatorio de *errores* no observables.

Los supuestos del modelo son los siguientes.

1. El modelo está perfectamente especificado en el sentido siguiente:
 - i. Los parámetros son constantes.
 - ii. El modelo es lineal en los parámetros.
2. \mathbf{X} no es estocástica.
3. $E[\varepsilon] = 0$.
4. $AR[\varepsilon] = \Sigma_{\varepsilon} = \sigma^2 \mathbf{1}_T$, (homoscedasticidad y no autocorrelación).
5. Las columnas de \mathbf{X} son linealmente independientes y en consecuencia el rango de \mathbf{X} es K , (No hay Multicolinealidad).

Pronosticación

Si T es el tiempo actual, sea $\mathbf{Y}_0' = (Y_{T+1} \ Y_{T+2} \ \cdots \ Y_{T+N})$ el vector de N observaciones futuras, para $t = T+1, \dots, T+N$, y

$$\mathbf{X}_0 = \begin{bmatrix} X_{T+1,1} & X_{T+1,2} & \cdots & X_{T+1,K} \\ X_{T+2,1} & X_{T+2,2} & \cdots & X_{T+2,K} \\ \vdots & \vdots & \ddots & \vdots \\ X_{T+N,1} & X_{T+N,2} & \cdots & X_{T+N,K} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_{T+1} \\ \mathbf{X}_{T+2} \\ \vdots \\ \mathbf{X}_{T+N} \end{bmatrix}$$

la matriz de valores futuros de las variables explicativas.

Para realizar pronósticos sobre tales valores futuros se supone que el proceso que genera al vector \mathbf{Y}_0 es el mismo que generó al vector \mathbf{Y} , y por lo tanto

$$\hat{\mathbf{Y}}_0 = \mathbf{X}_0 \hat{\boldsymbol{\theta}}$$

donde $\hat{\boldsymbol{\theta}}$ es el estimador del vector parametral. Si este estimador es el obtenido por el método de mínimos cuadrados, se tiene que

$$\hat{\mathbf{Y}}_0 = \mathbf{X}_0 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}$$

Si al modelo de Regresión Lineal se le agrega el supuesto de *normalidad*, esto es

$$\boldsymbol{\varepsilon} \sim N[\mathbf{0}, \sigma^2 \mathbf{I}_T]$$

se puede construir un *intervalo de predicción* (Judge, Hill, Griffiths, Lütkepohl, Lee [7]) para cada observación futura. Así, se pronostica que con probabilidad $(1 - \alpha)$ el valor de la variable Y_t , $t = T+1, \dots, T+N$ está contenido en el intervalo

$$\left[\mathbf{X}'_0 \hat{\boldsymbol{\theta}} - t_{(T-K, \alpha/2)} \tilde{\sigma} \sqrt{\mathbf{X}'_0 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}_0 + 1}, \mathbf{X}'_0 \hat{\boldsymbol{\theta}} + t_{(T-K, \alpha/2)} \tilde{\sigma} \sqrt{\mathbf{X}'_0 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}_0 + 1} \right],$$

con

$$\tilde{\sigma}^2 = \frac{\mathbf{e}'\mathbf{e}}{T-K} \quad \text{y} \quad \mathbf{e} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\theta}}$$

2.2.2 Modelos de Series de tiempo

⁶ El número $t_{T-K, \alpha/2}$ es el cuantil $\alpha/2$ de una distribución *t de Student* con $(T-K)$ grados de libertad.

$$\text{esto es } P\left(Z \leq t_{T-K, \alpha/2}\right) = \frac{\Gamma\left(\frac{T-K+1}{2}\right)}{\sqrt{\pi(T-K)} \Gamma\left(\frac{T-K}{2}\right)} \int_{-\infty}^{t_{T-K, \alpha/2}} \left[1 + \left(\frac{t^2}{T-K}\right)\right]^{-\frac{(T-K+1)}{2}} dt = \frac{\alpha}{2}$$

En los modelos de *Series de tiempo* sólo se analizan las observaciones pasadas de la variable de interés. El objetivo de los métodos basados en series de tiempo es descubrir el patrón en las series de datos históricos y extrapolar tal patrón al futuro. Sin embargo, en ocasiones para poder modelar y realizar pronósticos es necesario que los procesos que generan a las series sean estacionarios.

Definición 2.4

Si $\{X_t, t \in T\}$ es un proceso estocástico tal que $AR[X_t] < \infty$ para cada $t \in T$, entonces la *función de autocovarianza* $\gamma_X(r, s)$ de $\{X_t\}$ está definida por

$$\gamma_X(r, s) = COV[X_r, X_s] = E[(X_r - E[X_r])(X_s - E[X_s])], \quad r, s \in T$$

La función de autocovarianza es una extensión del concepto de la matriz de covarianza y es útil para descubrir la forma de la dependencia entre las diferentes variables aleatorias que componen un proceso estocástico.

Definición 2.5

La serie de tiempo $\{X_t, t \in T\}$ con $T = \{0, \pm 1, \pm 2, \dots\}$, se dice que es *estacionaria* si

$$\begin{aligned} E[|X_t|^2] &< \infty \quad \forall t \in T \\ E[X_t] &= \mu \quad \forall t \in T \\ \gamma_X(r, s) &= \gamma_X(r+t, r+s) \quad \forall r, s, t \in T \end{aligned}$$

Lo que se estableció como *estacionariedad* en la *definición 2.5*, es también conocido frecuentemente como *estacionariedad débil*, *estacionariedad en el sentido amplio* o *estacionariedad de segundo orden*, para distinguirlo de lo que se conoce como *estacionariedad estricta*.

Definición 2.6

La serie de tiempo $\{X_t, t \in T\}$ con $T = \{0, \pm 1, \pm 2, \dots\}$, se dice que es *estrictamente estacionaria* si cumple que

$$F_{X_{t_1}, \dots, X_{t_h}}(x_{t_1}, \dots, x_{t_h}) = F_{X_{t_1+h}, \dots, X_{t_h+h}}(x_{t_1}, \dots, x_{t_h})$$

para todo $t_1, \dots, t_h, h \in T$ y para todo vector $(x_{t_1}, \dots, x_{t_h}) \in \mathbb{R}^h$.

para todo $t_1, \dots, t_k, h \in T$ y para todo vector $(x_1, \dots, x_k) \in \mathfrak{R}^k$.

Es claro que un proceso estrictamente estacionario con segundos momentos finitos es débilmente estacionario. Es interesante hacer notar que los índices t_1, \dots, t_k , no son necesariamente consecutivos. Si una serie de tiempo es estrictamente estacionaria, la función de distribución de la variable aleatoria $X_t(\omega)$ es la misma para toda $t \in T$. La distribución conjunta depende de la distancia entre los elementos del conjunto índice T , y no de su valor cardinal. Esto no significa que una realización en particular será igual a otra realización, sino que, en intervalos de tiempo de la misma longitud, la serie de tiempo debe mostrar características estadísticas similares.

Los procesos estacionarios juegan un papel crucial en el análisis de series de tiempo, pero muchas de las series observadas en la práctica son no-estacionarias en apariencia, como los datos del Producto Interno Bruto de un país o el Índice de Precios al Consumidor. Frecuentemente tales conjuntos de datos pueden ser transformados, mediante diversas técnicas, a series que pueden ser razonablemente modeladas como realizaciones de un proceso estacionario (Brockwell, Davis [8]).

Una clase muy amplia de procesos estacionarios pueden ser generados usando ruidos blancos⁷ como términos de ajuste en un sistema de ecuaciones en diferencia lineales, dando origen a lo que se conoce como *Proceso Autoregresivo Promedios Móviles (ARMA)*.

Definición 2.7

Un proceso $\{X_t, t \in T\}$ con $T = \{0, \pm 1, \pm 2, \dots\}$ se dice que es un proceso *ARMA*(p, q), si $\{X_t\}$ es estacionario y para todo $t \in T$ es solución de la ecuación en diferencias

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q} \quad (2.1)$$

donde $\{Z_t\}$ es un proceso ruido blanco. Se dice que $\{X_t, t \in T\}$ es un *ARMA*(p, q) con media μ si $\{X_t - \mu\}$ es un *ARMA*(p, q).

Otra forma de escribir la ecuación (2.1) es

⁷Un proceso $\{Z_t, t \in T\}$ que cumple con $E[Z_t] = 0 \quad \forall t \in T$, $AR[Z_t] = \sigma_z^2 \quad \forall t \in T$ y $\gamma_z(t, t+h) = 0 \quad \forall t \in T, h \neq 0$, se le llama ruido blanco.

$$\phi(L)X_t = \theta(L)Z_t \quad (2.1.a)$$

donde $\phi(\cdot), \theta(\cdot)$ son los polinomios de grado p y q tales que

$$\phi(s) = 1 - \sum_{i=1}^p \phi_i s^i$$

$$\theta(s) = 1 + \sum_{i=1}^q \theta_i s^i$$

y L es el operador rezago^a.

Ejemplo: El proceso MA(q)

Si $\phi(s) \equiv 1$ entonces

$$X_t = \theta(L)Z_t \quad (2.2)$$

y se dice que el proceso es de *promedios móviles (moving average)*, de orden q y se denota por MA(q). En este caso la ecuación en diferencias (2.1) tiene la solución única (2.2). De hecho la solución $\{X_t\}$ es un proceso estacionario al cumplir que

$$E[X_t] = E[Z_t] + \sum_{j=1}^q \theta_j E[Z_{t-j}] = 0$$

y

$$\gamma_X(t+h, t) = \begin{cases} \sigma^2 \sum_{j=0}^{q-|h|} \theta_j \theta_{j+|h|} & |h| \leq q \\ 0 & |h| > q \end{cases}$$

La siguiente gráfica corresponde a una realización de $\{X_t\}_{t=1}^{100}$, con $q=1$, $\theta = -0.75$ y $Z_t \sim N[0,1]$, para toda t .

^a El operador rezago L opera de la siguiente forma: $L^n X_t = L^{n-1}(LX_t) = X_{t-n}$ y $L^0 X_t = X_t$.

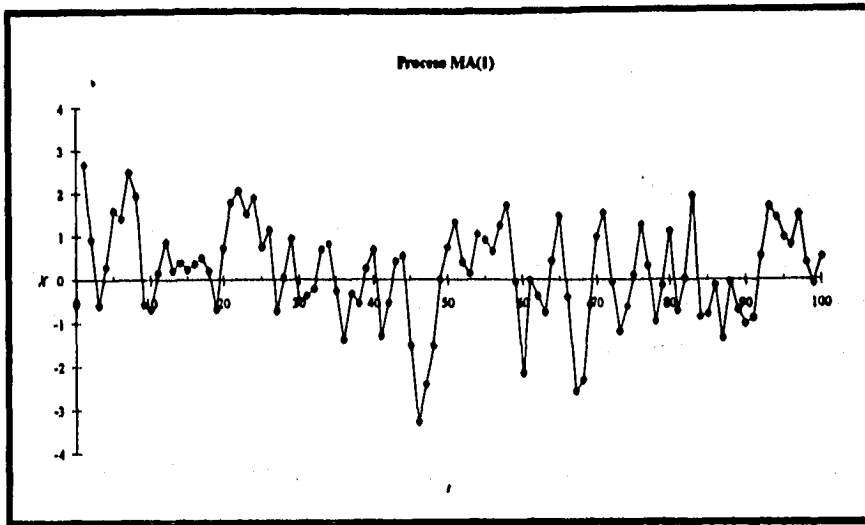


Figura 2.1. Realización de un proceso MA(1).

Ejemplo: El proceso AR(p)

Si $\theta(s) \equiv$ entonces

$$\phi(L)X_t = Z_t \quad (2.3)$$

A tal proceso se le llama *autoregresivo de orden p*, denotándose **AR(p)**. En este caso la existencia y unicidad de una solución estacionaria de (2.3) sólo se garantiza bajo condiciones especiales, especificadas en el siguiente

Teorema 2.1

Si $\phi(s) \neq 0$ para toda s en el campo de los números complejos y que cumpla con $|s| =$, entonces la ecuación en diferencia $\phi(L)X_t = \theta(L)Z_t$ tiene la única solución estacionaria,

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}$$

donde los coeficientes ψ_j están determinados por

$$\alpha(s)\{\phi(s)\}^{-1} = \sum_{j=-\infty}^{\infty} \psi_j s^j = \psi(s), \quad \frac{1}{r} < |s| < r$$

con $r > 1$.

En la práctica es difícil establecer los órdenes p y q de los polinomios de la ecuación (2.1) que describan la serie de tiempo en estudio. Una forma de determinarlos es utilizando la función de autocovarianza. Diferentes métodos para calcular tal función son tratados a detalle por Brockwell y Davis [11]. Sin embargo si la autocovarianza decae muy lentamente, el proceso puede no ser estacionario. Supongase que se inicia un proceso

$$Y_t = Y_{t-1} + X_t \quad (2.4)$$

donde X_t es un proceso estacionario con media $\mu \neq 0$ y donde $Y_0 = 0$. Desarrollando (2.4) se tiene

$$\begin{aligned} Y_t &= (Y_{t-2} + X_{t-1}) + X_t \\ &= ((Y_{t-3} + X_{t-2}) + X_{t-1}) + X_t \\ &= \dots = X_1 + X_2 + \dots + X_t \end{aligned}$$

y por lo tanto $E[Y_t] = t\mu$ por lo que se puede afirmar que la serie no es estacionaria. Sin embargo la diferencia

$$X_t = Y_t - Y_{t-1} = (1-L)Y_t \quad (2.4.a)$$

es una serie estacionaria. Considerar diferencias como (2.4.a) permite el análisis de procesos no estacionarios, muy frecuentes en la vida real, estudiando una transformación de tales procesos que cumplan con la estacionalidad. Si X_t en (2.4) es un proceso $ARMA(p,q)$ entonces Y_t es un proceso de *Auto-Regresivo Promedios Móviles Integrado*, denotado por $ARIMA(p,1,q)$. Si

$$X_t = (1-L)^d Y_t$$

es un proceso $ARMA(p,q)$ entonces Y_t es un proceso $ARIMA(p,d,q)$.

Pronosticación

Una vez que se ha ajustado un modelo de Series de tiempo apropiado, se busca predecir el valor de la variable aleatoria Y_{T+h} basandose en la información reunida hasta al actual tiempo T . $f_T(h)$ es una variable aleatoria que depende de las variables Y_T, Y_{T-1}, \dots .

La función de pronóstico es generada tomando la siguiente esperanza

$$f_T(h) = E[Y_{T+h} | Y_T, Y_{T-1}, \dots]$$

Generalmente, el pronóstico para el tiempo $T+h$ es construido sucesivamente a partir de los pronósticos para los tiempos $T+1, T+2, \dots, T+h-1$. Con este procedimiento las variables aún no ocurridas, Y_{T+j} , son remplazados por los pronósticos

$$f_T(j), \quad j = 1, \dots, h-1.$$

Los errores aún no ocurridos se toman igual a cero, $\epsilon_{T+j} = 0$, y los ya ocurridos son reemplazados por los residuos, $e_{T-j} = Y_{T-j} - f_{T-j-1}(1)$.

Para obtener un intervalo de predicción se requiere suponer una distribución para el proceso $\{Y_t\}$.

2.2.1 Limitaciones de los Modelos de Regresión y de Series de tiempo

Una de las limitaciones más importantes de los modelos antes mencionados es que no existe una forma de incorporar información externa. Los datos históricos no son la única fuente de información, así, en el ejemplo (b), es claro que ciertos eventos políticos internos o externos al país, influyen en el ambiente económico y provocan tendencias sobre el comportamiento de la demanda de dinero, pero si se utiliza alguno de los modelos anteriores aún cuando el pronosticador perciba cambios en las condiciones del mercado de dinero, no puede incorporar ese conocimiento en tales modelos sin modificarlos estructuralmente.

Por otro lado, y relacionado con este mismo problema, los modelos son *estáticos* en el sentido de que los parámetros del modelo son fijos, así como sus varianzas. Los modelos estáticos con parámetros constantes requieren que la media y la varianza sean estacionarias a

través del rango completo de las observaciones. En otro caso, los resultados son inútiles desde el punto de vista estadístico.

2.3 Características de los Modelos Bayesianos de Pronosticación

En el mundo real, ningún sistema está totalmente aislado y siempre está expuesto a interacciones con el exterior, por ello, se debe tener presente que el proceso que genera a la variable de interés tampoco está ajeno a cambios y perturbaciones, por lo que tomar un modelo estático para describir tal proceso puede no ser adecuado.

De manera alternativa, supongase que el proceso que genera a la variable Y se puede representar con un modelo lineal parametrizado de la siguiente forma

$$Y = F\theta + \varepsilon \quad (2.5)$$

donde F representa a las variables explicativas, ε es un error aleatorio y θ es un vector parámetro desconocido. Bajo el enfoque bayesiano, se puede considerar que el vector parámetro es un vector aleatorio Θ , y que, el conocimiento que sobre él se tiene puede ser expresado a través de una distribución de probabilidad F_{Θ} .

Las condiciones ambientales que determinan el comportamiento de la variable de interés cambian con el tiempo, es decir, el sistema en estudio es dinámico y por lo tanto se debe reconocer un grado de incertidumbre adicional debido al paso del tiempo. Por lo anterior se debe suponer que el parámetro Θ evoluciona a lo largo del tiempo, Θ_t . De esta forma, la distribución de probabilidad cambiará con el tiempo y entonces el modelo que describe a la serie $\{Y_t\}$ debe considerarse como una sucesión de modelos, cada uno con una incertidumbre medida por la distribución F_{Θ_t} .

Ejemplo: Demanda Real de Dinero

La gente demanda dinero por motivos de transacción, de precaución y de especulación, por ello la demanda de dinero en el tiempo t depende del nivel de ingreso en t Y_t y de la tasa de interés en ese tiempo r_t ,

$$L_t(r_t, Y_t),$$

con $\frac{\partial L_t}{\partial r_t} < 0$ y $\frac{\partial L_t}{\partial Y_t} > 0$.

Si P_t es el nivel de precios al tiempo t existente en la economía, la demanda de dinero en términos nominales crece proporcionalmente con los precios

$$M_t^d = P_t L_t \left(r_t, \frac{Y_t}{P_t} \right)$$

donde $\frac{Y_t}{P_t}$ es el ingreso real. De esta manera la demanda de dinero en términos reales es

$$\frac{M_t^d}{P_t} = L_t \left(r_t, \frac{Y_t}{P_t} \right).$$

Bajo este marco teórico se realiza un estudio estadístico para verificar si en la realidad el ingreso y la tasa de interés influyen en la demanda de saldos reales y la forma en que evoluciona la serie de demanda de dinero en el tiempo. Se toma un modelo paramétrico para describir el sistema, y se utilizan, por conveniencia, los logaritmos de las variables involucradas. Para corregir un comportamiento estacionalidad⁹ se agrega un variable dummy al modelo, estableciéndose la ecuación

$$(m-p)_t = \sum_{j=1}^{12} \gamma_j d_j + \alpha_1 y_t + \alpha_2 i_t + \alpha_3 (m-p)_{t-1}$$

donde

d_j = dummy estacional del mes j

$y_t = \ln Y_t$.

$i_t = \ln r_t$.

$$(m-p)_t = \ln \left(\frac{M_t}{P_t} \right).$$

y $\theta = (\gamma_1, \dots, \gamma_{12}, \alpha_1, \alpha_2, \alpha_3)$ es el vector de parámetros a estimar.

⁹ La estacionalidad ocurre cuando las variables explicativas tienen diferente efecto en la variable de interés dependiendo de la "estación" del año.

Si se estiman los parámetros considerando que el modelo es estático y utilizando los datos de 3 meses, de enero a marzo de 1988, por ejemplo, se obtiene un vector estimado

$$\hat{\theta}^1 = (\hat{\gamma}_1^1, \dots, \hat{\gamma}_{12}^1, \hat{\alpha}_1^1, \hat{\alpha}_2^1, \hat{\alpha}_3^1)$$

Si se vuelve a realizar una estimación, ahora con los datos de febrero a abril de 1988, se obtiene un nuevo vector estimado

$$\hat{\theta}^2 = (\hat{\gamma}_1^2, \dots, \hat{\gamma}_{12}^2, \hat{\alpha}_1^2, \hat{\alpha}_2^2, \hat{\alpha}_3^2)$$

que no necesariamente es igual al vector $\hat{\theta}^1$.

Es claro que parte de la diferencia entre estas dos estimaciones se debe a que las condiciones del sistema económico fueron modificadas por sucesos de diversa índole. Si se escogió un modelo paramétrico y si el sistema en estudio es muy cambiante, no es conveniente pensar que los parámetros no sufrirán cambio alguno. Se debe dar cierta flexibilidad al modelo y permitir que los parámetros se modifiquen con el tiempo. Estas ideas dan origen a los *Modelos Dinámicos*.

Definición 2.8

Un *modelo dinámico* es una sucesión $\{M_t\}$ de modelos paramétricos $M_t = M(\Theta_t)$ teniendo Θ_t una distribución de probabilidad F_{Θ_t} . Si se utiliza un modelo dinámico para describir un proceso $\{Y_t\}$, el modelo está estructurado para

- (a) Describir la evolución del proceso entre los tiempos t y $t+1$.
- (b) Modificar el conocimiento sobre Θ_t , a través de la distribución $F_{\Theta_t|Y_t}(\Theta_t, Y_t)$.

Al realizar un pronóstico para Y_t en el esquema del modelo dinámico $\{M(\Theta_t)\}$, cada valor Θ_t , que puede tomar el vector Θ_t , proporciona un pronóstico condicional en términos de la densidad condicional $f_{Y_t|\Theta_t}(y_t|\Theta_t)$, así el pronóstico puede ser obtenido de la densidad marginal

$$f_t(y_t) = \int_{\Theta_t} f_{Y_t|\Theta_t}(y_t|\Theta_t) d\Theta_t. \quad (2.6)$$

La metodología Bayesiana permite interpretar a los modelos Dinámicos como *rutinas de aprendizaje*. Si $\{M_t\}$ es un modelo dinámico, antes de presentarse una observación Y_t el pronosticador establece una distribución a priori para el parámetro, cuya densidad es $f_{\Theta_t}(\Theta_t)$. Como el comportamiento de la variable en el siguiente periodo depende del valor

de Θ_t , se puede obtener un pronóstico de la densidad condicional $f_{y_t|\Theta_t}(y_t|\theta_t)$. Una vez que se presenta la observación y_t , la información que se tiene sobre el parámetro Θ_t , se modifica, teniéndose la densidad posterior

$$f_{\Theta_t|y_t}(\theta_t|y_t) \propto f_{y_t|\Theta_t}(y_t|\theta_t)f_{\Theta_t}(\theta_t)$$

Cuando se considera que las condiciones ambientales del proceso han cambiado sustancialmente, la metodología Bayesiana permite que el pronosticador pueda *intervenir*, por medio de la integración de información *subjetiva*. Sin embargo, hasta que no ocurra una intervención, se considera que un *Modelo Dinámico* es cerrado a la información externa para $t \geq 1$. Así si D_0 representa a la información en el tiempo $t = 0$, la información disponible en el tiempo t es simplemente

$$D_t = \{y_t, D_{t-1}\}$$

En resumen, los **Modelos Dinámicos**, al ser un ejemplo del Enfoque Bayesiano aplicado a la Pronosticación, presentan las siguientes características:

- (1) Estructuración *paramétrica*, que permite la comunicación entre el pronosticador y Método de Pronosticación en ambas direcciones.
- (2) Tienen una definición *secuencial* que describe como los parámetros cambian en el tiempo, tanto sistemáticamente como por perturbaciones aleatorias.
- (3) Se tiene una representación *probabilística* de la información acerca de los parámetros en cualquier tiempo t . Esta representación permite a un pronosticador:
 - (i) integrar información *subjetiva*.
 - (ii) realizar pronósticos en situaciones de escasa información histórica.
 - (iii) cuando las condiciones ambientales del proceso cambian sustancialmente, el pronosticador puede *intervenir*, haciendo que el modelo responda efectivamente a tales cambios.
- (4) Los pronósticos son en forma de *distribuciones de probabilidad*.
- (5) Los modelos son *rutinas de aprendizaje*.

Dentro de la clase de Modelos Dinámicos existe una subclase caracterizada por su linealidad en los parámetros y este grupo es el de los denominados *Modelo Dinámicos Lineales MDL*, que son los modelos que se desarrollarán en esta tesis.

CAPITULO III

MODELO POLINOMIAL DE PRIMER ORDEN

3.1 Introducción

Para entender el funcionamiento de los Modelos Dinámicos Lineales, se comenzará por estudiar uno de los casos más sencillos y también ampliamente utilizado en la práctica: el *Modelo Polinomial de Primer Orden* o también denominado *Modelo Lineal ML*.

Supongase que $\{Y_t\}$ es una serie de tiempo con observaciones y_t de una variable en estudio.

Definición 3.1

El *Modelo Polinomial de primer orden* consiste en un modelo de series de tiempo dado por las ecuaciones

$$Y_t = \Theta_t + \varepsilon_{V_t} \quad \varepsilon_{V_t} \sim N[\mu_{V_t}, \sigma_{V_t}^2] \quad (3.1)$$

$$\Theta_t = \Theta_{t-1} + \varepsilon_{W_t} \quad \varepsilon_{W_t} \sim N[\mu_{W_t}, \sigma_{W_t}^2] \quad (3.2)$$

y la *información inicial*

$$(\Theta_0 | D_0) \sim N(m_0, S_0^2) \quad (3.3)$$

donde al parámetro Θ_t se le llama *nivel de la serie al tiempo t*, y $\{\varepsilon_{v_t}\}, \{\varepsilon_{w_t}\}$ ¹⁰ son independientes. m_0 es una estimación sobre el nivel de la serie y S_0^2 una medida de la incertidumbre, ambas establecidas de acuerdo al grado de creencia del modelador. El conjunto de información inicial D_0 contiene los valores de los hiperparámetros¹¹ $\mu_{v_t}, \mu_{w_t}, \sigma_{v_t}^2$ y $\sigma_{w_t}^2$, para toda t y $D_t = (Y_t, D_{t-1})$ es el conjunto de información disponible en el tiempo t .

Una forma de denotar a este modelo es $\{1, 1, \mu_{v_t}, \sigma_{v_t}^2, \mu_{w_t}, \sigma_{w_t}^2\}$, donde los dos primeros 1 corresponden a los coeficientes de Θ_t y Θ_{t-1} en las ecuaciones (3.1) y (3.2) respectivamente.

La denominación de "polinomial de primer orden" surge de la siguiente consideración: si al proceso $\{\Theta_t\}$ se le toma como un proceso continuo en el tiempo $\mathcal{Q}(t)$, y además se supone que es diferenciable, entonces utilizando el desarrollo en series de Taylor de grado n en el punto t asociado a dicha función se tiene la siguiente expresión

$$\begin{aligned} \mathcal{Q}(t + \delta) &= \sum_{i=0}^n \frac{\mathcal{Q}^{(i)}(t)\delta^i}{i!} + R_{n,t} \\ &= \mathcal{Q}(t) + \sum_{i=1}^n \frac{\mathcal{Q}^{(i)}(t)\delta^i}{i!} + R_{n,t} \\ &= \mathcal{Q}(t) + e(t, \delta, n) \end{aligned}$$

con lo que el modelo describe a los términos restantes como ruido blanco ε_{w_t} (llamado *error de transición*). De este modo se tiene una aproximación lineal local para el parámetro Θ_t .

Aún cuando en el largo plazo se espera que ocurran cambios significativos en la serie, si el modelador no desea anticipar la forma de la evolución porque no existe alguna razón para esperar que el nivel de la serie aumente o disminuya, se puede agregar el supuesto de que $\mu_{w_t} = 0$ para toda t . Bajo esta condición el modelo presenta las siguientes particularidades.

¹⁰ A partir de este momento a los términos de error del modelo lineal se les denotará con la letra griega epsilon minúscula, y no con una letra mayúscula como a las demás variables aleatorias presentadas en este trabajo. Como tales errores no son observables no debe existir confusión en la notación de los errores aleatorios y los valores que toman.

¹¹ A los términos $\mu_{v_t}, \mu_{w_t}, \sigma_{v_t}^2, \sigma_{w_t}^2$, se les llama *hiperparámetros* para distinguirlos de los parámetros que entran en el modelo a través de las ecuaciones de la definición 3.1.

Constancia local.

En la siguiente figura se muestra la realización de una serie $\{Y_t\}$, cuyos datos fueron obtenidos a partir de una simulación del modelo lineal, suponiendo que el nivel de la serie en el tiempo 0 cumple que $\theta_0 = 0$. Además se supuso que $\mu_{v_t} = \mu_{w_t} = 0$ para toda t , y que las varianzas son constantes en todo tiempo, tomando los valores $\sigma_{v_t}^2 = \sigma_{v_t}^2 =$ y $\sigma_{w_t}^2 = \sigma_{w_t}^2 = 0.05$.

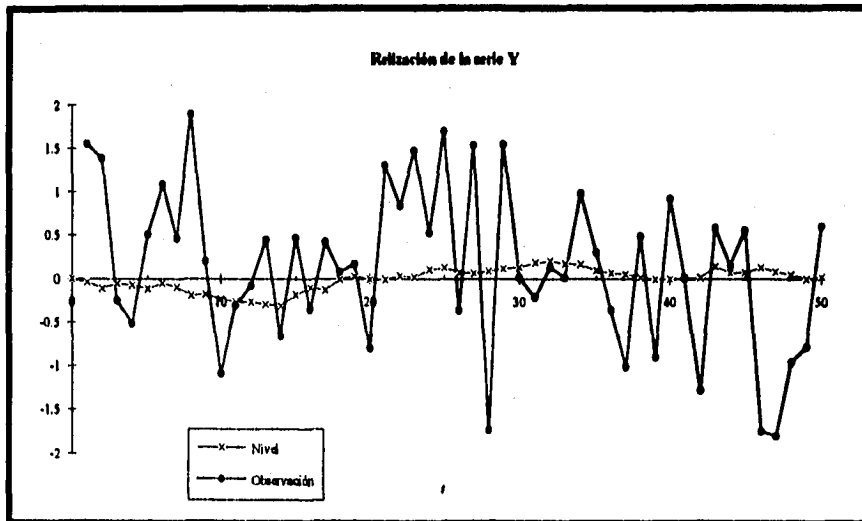


Figura 3.1

En la gráfica se puede observar que para un tiempo dado, el nivel de la serie no presenta cambios bruscos en los periodos alrededor de ese tiempo, es decir, el nivel es "constante localmente". Lo anterior ocurre porque existe una alta probabilidad de que el nivel de la serie θ_t entre los tiempos t y $t+1$, no presente cambios bruscos, como se ilustra en la figura 3.2.

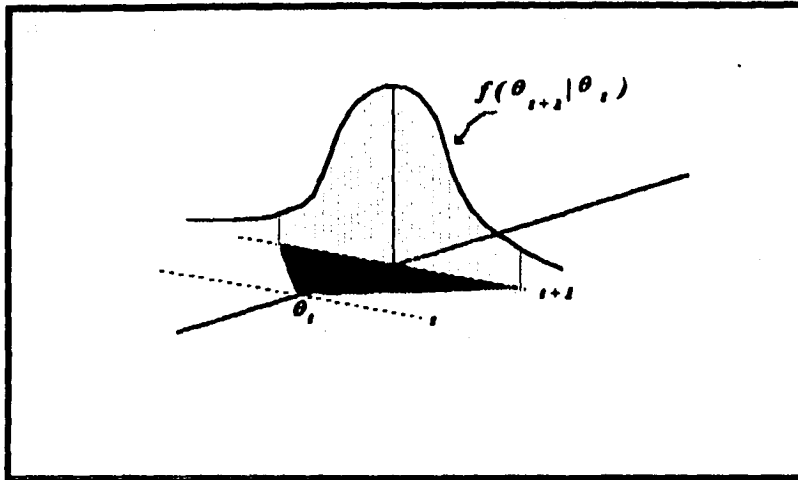


Figura 3.2

Valor esperado en el futuro constante

Otra particularidad es que, el valor esperado de una observación Y en el futuro condicionado al nivel actual es constante

$$E[Y_{i+k} | \theta_i] = E[\theta_{i+k} | \theta_i] = \theta_i, \quad \forall k > 0, \quad (3.4)$$

ya que

$$\begin{aligned} E[Y_{i+k} | \theta_i] &= E[\theta_{i+k} + \varepsilon_{Y,i+k} | \theta_i] = E[\theta_{i+k} | \theta_i] \\ &= E\left[\theta_i + \sum_{i=1}^k \varepsilon_{w,i} | \theta_i\right] = \theta_i \end{aligned}$$

El involucrar el nivel de la serie para modelar a la serie $\{Y_t\}$ es de una gran utilidad para problemas de pronosticación con fines de planeación de producción y control de inventarios. El nivel de la serie representa la demanda promedio o demanda real por un producto. La demanda observada está compuesta por esa demanda real más una perturbación debida a sorpresas que se presentan en el mercado.

Para seguir la transición del nivel de la serie a lo largo del tiempo, las distribuciones más importantes para actualizar el conocimiento del proceso se establecen en el siguiente teorema¹².

Teorema 3.1

Si $(\Theta_0|D_0) \sim N[m_0, S_0^2]$ entonces para $t \geq 1$ se tienen las distribuciones

(a) A priori para Θ_t ,

$$(\Theta_t|D_{t-1}) \sim N[\bar{m}_t, \bar{S}_t^2],$$

donde

$$\bar{m}_t = m_{t-1} + \mu_w, \quad \bar{S}_t^2 = S_{t-1}^2 + \sigma_w^2,$$

(b) Pronóstico de un periodo

$$f_{t-1}(1) \sim N[\hat{Y}_t, \hat{\sigma}_t^2],$$

donde

$$\hat{Y}_t = \bar{m}_t + \mu_v, \quad \hat{\sigma}_t^2 = \bar{m}_t + \sigma_v^2,$$

(c) A posteriori para Θ_t ,

$$(\Theta_t|D_t) \sim N[m_t, S_t^2],$$

con

$$\begin{aligned} m_t &= \bar{m}_t + A_t e_t \\ &= \bar{m}_t + \frac{\bar{S}_t^2}{\hat{\sigma}_t^2} (y_t - \hat{Y}_t) \end{aligned}$$

$$\begin{aligned} S_t^2 &= A_t \sigma_v^2 \\ &= \bar{S}_t^2 - \frac{(\bar{S}_t^2)^2}{\hat{\sigma}_t^2} \end{aligned}$$

¹² Los teoremas de las secciones 3.1 y 3.2 se pueden encontrar en el libro de Harrison y West, pero ahí sólo se considera el caso en que $\mu_v = \mu_w = 0$, para toda t . En este trabajo se pretende extender los teoremas al caso más general.

donde

$$A_t = \frac{\bar{S}_t^2}{\hat{\sigma}_t^2}$$

es el coeficiente adaptativo y

$$e_t = y_t - \hat{Y}_t$$

es el error de pronóstico de un periodo.

Demostración.

Demostración por inducción.

Supóngase que se cumple la distribución (c) para $t-1$. Θ_t condicionado sobre D_{t-1} es la suma de dos variables aleatorias independientes con distribución normal, $\Theta_{t-1}, \varepsilon_{w_t}$, por lo tanto $(\Theta_t | D_{t-1})$ tiene también distribución normal. La media y la varianza son obtenidas a partir de la suma de las medias y las varianzas de $\Theta_{t-1}, \varepsilon_{w_t}$,

$$\begin{aligned} E[\Theta_t | D_{t-1}] &= E[\Theta_{t-1} + \varepsilon_{w_t} | D_{t-1}] \\ &= m_{t-1} + \mu_{w_t} \end{aligned}$$

$$\begin{aligned} AR[\Theta_t | D_{t-1}] &= VAR[\Theta_{t-1} + \varepsilon_{w_t} | D_{t-1}] \\ &= S_{t-1}^2 + \sigma_{w_t}^2 \end{aligned}$$

Por lo tanto

$$(\Theta_t | D_{t-1}) \sim N[m_{t-1} + \mu_{w_t}, S_{t-1}^2 + \sigma_{w_t}^2]$$

quedando demostrado (a).

Similarmente, y de nuevo condicionada sobre D_{t-1} , Y_t es la suma de variables aleatorias independientes con distribución normal, $\Theta_t, \varepsilon_{v_t}$, por lo que $(Y_t | D_{t-1})$ tiene distribución normal con esperanza y varianza obtenidas de

$$\begin{aligned} E[Y_t | D_{t-1}] &= E[\Theta_t + \varepsilon_{v_t} | D_{t-1}] \\ &= \bar{m}_t + \mu_{v_t} \end{aligned}$$

$$\begin{aligned} AR[Y_t | D_{t-1}] &= VAR[\Theta_t + \varepsilon_{v_t} | D_{t-1}] \\ &= \bar{S}_{t-1}^2 + \sigma_{v_t}^2 \end{aligned}$$

Por lo tanto

$$(Y_i | D_{i-1}) \sim N[\bar{m}_i + \mu_{Y_i}, \bar{S}_i^2 + \sigma_{Y_i}^2]$$

quedando demostrado (b).

Existen dos maneras de demostrar (c).

(i) Via Teorema de Bayes.

(ii) Via propiedades de la distribución Normal.

(i) De acuerdo a la función de densidad a priori para Θ , es

$$f_{\Theta_i | D_{i-1}}(\theta_i) = \frac{1}{\sqrt{2\pi\bar{S}_i^2}} \exp\left\{-\frac{(\theta_i - \bar{m}_i)^2}{2\bar{S}_i^2}\right\}$$

y de acuerdo a (3.1), la función de densidad para Y_i está dada por

$$\begin{aligned} f_{Y_i | \Theta_i, D_{i-1}}(y_i | \theta_i) &= L(\theta_i | y_i, D_{i-1}) \\ &= \frac{1}{\sqrt{2\pi\sigma_{Y_i}^2}} \exp\left\{-\frac{(y_i - (\theta_i + \mu_{Y_i}))^2}{2\sigma_{Y_i}^2}\right\} \end{aligned}$$

Sea $\Theta_i^* = \Theta_i + \mu_{Y_i}$ entonces

$$f_{\Theta_i | D_{i-1}}(\theta_i) = \frac{1}{\sqrt{2\pi\bar{S}_i^2}} \exp\left\{-\frac{(\theta_i - \hat{Y}_i)^2}{2\bar{S}_i^2}\right\}$$

y

$$\begin{aligned} f_{Y_i | \Theta_i^*, D_{i-1}}(y_i | \Theta_i^*) &= L(\Theta_i^* | y_i, D_{i-1}) \\ &= \frac{1}{\sqrt{2\pi\sigma_{Y_i}^2}} \exp\left\{-\frac{(y_i - \Theta_i^*)^2}{2\sigma_{Y_i}^2}\right\} \end{aligned}$$

De acuerdo al lema 1.1 se tiene que

$$\begin{aligned} f_{\Theta_i^* | Y_i, D_{i-1}}(\Theta_i^* | y_i) &= f_{\Theta_i | D_{i-1}}(\theta_i) \\ &\propto \exp\left\{-\frac{1}{2\bar{S}_i^2}(\theta_i - \bar{m}_i)^2\right\} \end{aligned}$$

$$\begin{aligned}
 m_i^* &= \frac{\bar{S}_i^2 y_i + \sigma_{v_i}^2 \hat{Y}_i}{\bar{S}_i^2 + \sigma_{v_i}^2} = \frac{\bar{S}_i^2 y_i + \sigma_{v_i}^2 \hat{Y}_i}{\hat{\sigma}_i^2} \\
 &= A_i y_i + \hat{Y}_i \left(\frac{\hat{\sigma}_i^2 - \bar{S}_i^2}{\hat{\sigma}_i^2} \right) = A_i y_i + (1 - A_i) \hat{Y}_i
 \end{aligned}$$

Y

$$S_i^2 = \frac{\bar{S}_i^2 \sigma_{v_i}^2}{\bar{S}_i^2 + \sigma_{v_i}^2}$$

Por lo tanto

$$(\theta_i^* | D_i) \sim N[m_i^*, S_i^2]$$

Entonces

$$(\theta_i | D_i) \sim N[m_i, S_i^2]$$

con

$$\begin{aligned}
 m_i &= m_i^* - \mu_{v_i} \\
 &= A_i (y_i - \hat{Y}_i) + \hat{Y}_i - \mu_{v_i} \\
 &= A_i (y_i - \hat{Y}_i) + \bar{m}_i
 \end{aligned}$$

(ii) Debido a que cualquier función lineal de Y_i y θ_i es una combinación lineal de variables normales e independientes $\varepsilon_{v_i}, \varepsilon_{u_i}, \theta_{i-1}$, tal función lineal condicionada sobre D_{i-1} también se distribuye de forma normal y por definición la distribución conjunta $(Y_i, \theta_i | D_{i-1})$ es normal bivariada. La covarianza entre Y_i y θ_i está dada por:

$$\begin{aligned}
 COV[Y_i, \theta_i | D_{i-1}] &= COV[\theta_i + \varepsilon_{v_i}, \theta_i | D_{i-1}] \\
 &= COV[\varepsilon_{v_i}, \theta_i | D_{i-1}] + VAR[\theta_i | D_{i-1}] \\
 &= \bar{S}_i^2
 \end{aligned}$$

Por lo tanto

$$\begin{pmatrix} Y_i \\ \Theta_i \end{pmatrix} | D_{i-1} \sim N \left[\begin{pmatrix} \hat{Y}_i \\ \bar{m}_i \end{pmatrix}, \begin{pmatrix} \hat{\sigma}_i^2 & \bar{S}_i^2 \\ \bar{S}_i^2 & \bar{S}_i^2 \end{pmatrix} \right]$$

De acuerdo a la sección A.1.2 del Apéndice se tiene la distribución condicional

$$(\Theta_i | Y_i = y_i, D_{i-1}) \sim N[m_i(y_i), S_i^2(y_i)]$$

con

$$m_i(y_i) = \bar{m}_i + \frac{\bar{S}_i^2}{\hat{\sigma}_i^2} (y_i - \hat{Y}_i)$$

$$S_i^2(y_i) = \bar{S}_i^2 - \frac{\bar{S}_i^2}{\hat{\sigma}_i^2}$$

donde

$$A_i = \frac{\bar{S}_i^2}{\hat{\sigma}_i^2}$$

es el coeficiente de regresión de Θ_i sobre Y_i .

Se observa que la correlación entre Y_i y Θ_i es $\rho_i = \frac{\bar{S}_i^2}{\sqrt{\bar{S}_i^2 \hat{\sigma}_i^2}}$ por lo tanto $\rho_i^2 = A_i$.

Finalmente se completa la demostración a partir de que (a) se cumple para $t = 1$.

■

En la demostración del teorema se observa que A_i es el *coeficiente a priori de regresión* de Θ_i sobre Y_i , siendo también, en este caso particular, el *cuadrado de su coeficiente de correlación*; cumpliéndose que $0 \leq A_i \leq 1$. Al coeficiente A_i se le llama *coeficiente adaptativo* debido a la representación alternativa para m_i :

$$\begin{aligned} m_i &= \bar{m}_i + A_i (y_i - \hat{Y}_i) \\ &= A_i y_i - A_i (\hat{Y}_i - \bar{m}_i) + (1 - A_i) \bar{m}_i \end{aligned} \tag{3.5}$$

m_i es un promedio ponderado del *nivel estimado a priori* \bar{m}_i , la observación y_i y el valor esperado del error de observación μ_{ν_i} , siendo A_i lo que determina los pesos.

3.2 Distribuciones de pronóstico

En esta sección se enunciará la manera en que se distribuyen los pronósticos mencionados en la sección 2.1.1.

Teorema 3.2

En el tiempo t se tiene las siguientes distribuciones de pronóstico para $k \geq j \geq 1$:

(a) Pronóstico de k periodos en el futuro

$$f_t(k) \sim N[\hat{Y}_t(k), \hat{\sigma}_t^2(k)],$$

con

$$\hat{Y}_t(k) = m_t + \sum_{i=1}^k \mu_{w_{t+i}} + \mu_{v_{t+k}}$$

$$\hat{\sigma}_t^2(k) = S_t^2 + \sum_{j=1}^k \sigma_{w_{t+j}}^2 + \sigma_{v_{t+k}}^2$$

(b) Pronóstico guía de horizonte k

$$f_t^H(k) \sim N[m_{H_t}(k), S_{H_t}^2(k)],$$

con

$$m_{H_t}(k) = km_t + \sum_{i=1}^k i \mu_{w_{t+i-1}} + \sum_{i=1}^k \mu_{v_{t+i}}$$

$$S_{H_t}^2(k) = k^2 S_t^2 + \sum_{i=1}^k i^2 \sigma_{w_{t+i-1}}^2 + \sum_{i=1}^k \sigma_{v_{t+i}}^2$$

(c) Pronóstico guía de horizonte k diferido $j-1$ periodos

$$f_t^H(j, k) \sim N[m_{H_t}(j, k), S_{H_t}^2(j, k)],$$

con

$$m_{H_t}(j, k) = \sum_{r=j}^k \left[m_t + \sum_{i=1}^r \mu_{w_{t+i}} + \mu_{v_{t+r}} \right]$$

$$S_{H_t}^2(j, k) = (k-j+1) \left[(k-j+1) S_t^2 + \sum_{i=1}^j \sigma_{w_{t+i}}^2 \right]$$

$$+ \sum_{i=0}^{k-j-1} (k-j-i) \sigma_{w_{t+j+i}}^2 + \sum_{r=1}^k \sigma_{v_{t+r}}^2$$

Demostración.

(a) De acuerdo a (3.5), para $k > 0$

$$\Theta_{i,k} = \Theta_i + \sum_{l=1}^k \varepsilon_{W_{i,l}},$$

y usando (3.1) es claro que

$$Y_{i,k} = \Theta_i + \sum_{l=1}^k \varepsilon_{W_{i,l}} + \varepsilon_{V_{i,k}} \quad (3.6)$$

es una combinación lineal de cantidades aleatorias independientes con distribución normal. Entonces $Y_{i,k}$ tiene distribución normal con momentos obtenidos a partir de (3.6)

$$\begin{aligned} E[f_i(k)] &= E[Y_{i,k} | D_i] \\ &= E[\Theta_i | D_i] + \sum_{l=1}^k E[\varepsilon_{W_{i,l}} | D_i] + E[\varepsilon_{V_{i,k}} | D_i] \\ &= m_i + \sum_{l=1}^k \mu_{W_{i,l}} + \mu_{V_{i,k}} \\ &= \hat{Y}_i(k) \end{aligned}$$

y

$$\begin{aligned} AR[f_i(k)] &= VAR[Y_{i,k} | D_i] \\ &= VAR[\Theta_i | D_i] + \sum_{l=1}^k VAR[\varepsilon_{W_{i,l}} | D_i] + VAR[\varepsilon_{V_{i,k}} | D_i] \\ &= S_i^2 + \sum_{l=1}^k \sigma_{W_{i,l}}^2 + \sigma_{V_{i,k}}^2 \\ &= \hat{\sigma}_i^2(k) \end{aligned}$$

quedando (a) demostrado.

(b) A partir de la *definición 3.1* y de acuerdo a (3.1) y (3.2) se tiene que

$$\begin{aligned}
\sum_{i=1}^k Y_{t,i} &= \sum_{i=1}^k [\theta_{t,i} + \varepsilon_{v_{t,i}}] \\
&= \sum_{i=1}^k \left[\theta_t + \sum_{j=1}^i \varepsilon_{w_{t,j}} \right] + \sum_{i=1}^k \varepsilon_{v_{t,i}} \\
&= k\theta_t + \sum_{i=1}^k \sum_{j=1}^i \varepsilon_{w_{t,j}} + \sum_{i=1}^k \varepsilon_{v_{t,i}} \\
&= k\theta_t + \sum_{i=1}^k i\varepsilon_{t,k-i+1} + \sum_{i=1}^k \varepsilon_{v_{t,i}}
\end{aligned}$$

es una función lineal de términos independientes con distribución normal

$$\begin{aligned}
E[f_t''(k)] &= kE[\theta_t | D_t] + \sum_{i=1}^k i E[\varepsilon_{w_{t,k-i+1}} | D_t] + \sum_{i=1}^k E[\varepsilon_{v_{t,i}} | D_t] \\
&= km_t + \sum_{i=1}^k i \mu_{w_{t,k-i+1}} + \sum_{i=1}^k \mu_{v_{t,i}} \\
&= m_{H_t}(k)
\end{aligned}$$

$$\begin{aligned}
AR[f_t''(k)] &= k^2 VAR[\theta_t | D_t] + \sum_{i=1}^k i^2 VAR[\varepsilon_{w_{t,k-i+1}} | D_t] + \sum_{i=1}^k VAR[\varepsilon_{v_{t,i}} | D_t] \\
&= k^2 S_t^2 + \sum_{i=1}^k i^2 \sigma_{w_{t,k-i+1}}^2 + \sum_{i=1}^k \sigma_{v_{t,i}}^2 \\
&= S_{H_t}^2(k)
\end{aligned}$$

quedando (b) demostrado.

(c) De igual manera, a partir de la *definición 3.1* y de acuerdo a (3.1) y (3.2) se tiene que

$$\sum_{r=j}^k Y_{t,r} = \sum_{r=j}^k \left[\theta_t + \sum_{i=1}^r \varepsilon_{w_{t,i}} + \varepsilon_{v_{t,r}} \right]$$

es una función lineal de términos independientes y de forma normal y en consecuencia

$$\begin{aligned}
E[f_t''(j,k)] &= \sum_{r=j}^k \left[E[\theta_t | D_t] + \sum_{i=1}^r E[\varepsilon_{w_{t,i}} | D_t] + E[\varepsilon_{v_{t,r}} | D_t] \right] \\
&= \sum_{r=j}^k \left[m_t + \sum_{i=1}^r \mu_{w_{t,i}} + \mu_{v_{t,r}} \right]
\end{aligned}$$

y

$$\begin{aligned}
AR[f_t^h(j,k)] &= VAR[(k-j+1)\theta_t | D_t] \\
&+ \sum_{r=j}^k \left[\sum_{i=1}^j \sigma_{w_{i,t}}^2 + \sum_{h=0}^{r-j} \sigma_{w_{i,j+h}}^2 + \sigma_{v_{i,t}}^2 \right] \\
&= (k-j+1)^2 S_t^2 \\
&+ (k-j+1) \sum_{i=1}^j \sigma_{w_{i,t}}^2 + \sum_{h=1}^{k-j} (k-j-h+1) \sigma_{w_{i,j+h}}^2 + \sum_{r=j}^k \sigma_{v_{i,t}}^2
\end{aligned}$$

■

En las siguientes secciones se verán distintos casos particulares de los hiperparámetros para el modelo general descrito hasta aquí. Específicamente, se considera:

- (i) El caso $\sigma_{v_t}^2 = \sigma_v^2$, $\sigma_{w_t}^2 = \sigma_w^2$ para toda t , ambas conocidas.
- (ii) Elección de σ_w^2 .
- (iii) El caso de $\sigma_{v_t}^2 = \sigma_v^2$ con σ_v^2 desconocida.

3.3 El Modelo Constante

Definición 3.2

El caso particular del *Modelo Lineal (ML)* en el que, $\sigma_{v_t}^2 = \sigma_v^2$ y $\sigma_{w_t}^2 = \sigma_w^2$ para toda t , con σ_v^2, σ_w^2 constantes positivas y acotadas es llamado *constante (MLC)* y se denota por $\{1, 1, \mu_v, \sigma_v^2, \mu_w, \sigma_w^2\}$.

Aún cuando el modelo constante es bastante restringido, es útil para comprender las características de ciertos modelos más generales. Un rasgo particular del modelo constante es que el coeficiente adaptativo A_t , converge a un valor constante como se ve en el siguiente teorema.

Teorema 3.3

En el MLC las series A_t y S_t^2 cumplen con lo siguiente

$$\lim_{t \rightarrow \infty} A_t = A$$

y

$$\lim_{t \rightarrow \infty} S_t^2 = S^2 = A \sigma_v^2$$

donde

$$A = \frac{\sigma_w^2 \left(\sqrt{1 + \frac{4\sigma_v^2}{\sigma_w^2}} - 1 \right)}{2\sigma_v^2}$$

Demostración.

Por el teorema 3.1 se tiene

$$S_t^2 = \sigma_v^2 \frac{\bar{S}_t^2}{\hat{\sigma}_t^2} = \sigma_v^2 \left(\frac{\bar{S}_t^2}{\bar{S}_t^2 + \sigma_v^2} \right) = \frac{1}{\frac{1}{\sigma_v^2} + \frac{1}{\bar{S}_t^2}}$$

Por lo que

$$(S_t^2)^{-1} = (\sigma_v^2)^{-1} + (\bar{S}_t^2)^{-1} \tag{3.7}$$

tomando la diferencia

$$\begin{aligned} (S_t^2)^{-1} - (S_{t-1}^2)^{-1} &= \left[(\bar{S}_t^2)^{-1} + V^{-1} \right] - \left[(\bar{S}_{t-1}^2)^{-1} + V^{-1} \right] \\ &= (\bar{S}_t^2)^{-1} - (\bar{S}_{t-1}^2)^{-1} = \frac{\bar{S}_{t-1}^2 - \bar{S}_t^2}{\bar{S}_t^2 \bar{S}_{t-1}^2} \end{aligned}$$

usando el hecho que $\bar{S}_t^2 = S_{t-1}^2 + \sigma_w^2$ se observa que para toda $t > 2$

$$\begin{aligned} (S_t^2)^{-1} - (S_{t-1}^2)^{-1} &= \frac{[S_{t-2}^2 + \sigma_w^2] - [S_{t-1}^2 + \sigma_w^2]}{\bar{S}_t^2 \bar{S}_{t-1}^2} \\ &= \frac{S_{t-2}^2 - S_{t-1}^2}{\bar{S}_t^2 \bar{S}_{t-1}^2} = \frac{S_{t-2}^2 S_{t-1}^2 [(S_{t-1}^2)^{-1} - (S_{t-2}^2)^{-1}]}{\bar{S}_t^2 \bar{S}_{t-1}^2} \end{aligned}$$

puesto que $\frac{S_{t-2}^2 S_{t-1}^2}{\bar{S}_t^2 \bar{S}_{t-1}^2} > 0$ para toda $t > 2$, entonces

$$\frac{(S_t^2)^{-1} - (S_{t-1}^2)^{-1}}{(S_{t-1}^2)^{-1} - (S_{t-2}^2)^{-1}} > 0 \quad \forall t > 2$$

Es decir, la sucesión $\{(S_t^2)^{-1}\}$ es monótona.

Por otra parte

$$S_t^2 = A_t \sigma_v^2$$

y

$$0 < A_t < 1 \quad \forall t > 0,$$

por lo que

$$0 < S_t^2 < \sigma_v^2 \quad \forall t > 0$$

lo que muestra que la sucesión $\{(S_t^2)^{-1}\}$ es acotada y como también es monótona, entonces converge a un límite. Como $\bar{S}_t^2 = S_{t-1}^2 + \sigma_w^2$ y $A_t = \frac{S_t^2}{\sigma_v^2}$, las sucesiones $\{S_t^2, \bar{S}_t^2, A_t\}$ también

convergen a los límites S^2, \bar{S}^2 y A , respectivamente, cumpliéndose además que $\bar{S}^2 = S^2 + \sigma_w^2$ y $A = \frac{S^2}{\sigma_v^2}$. De acuerdo al *teorema 3.1* se ve que

$$S_t^2 = A_t \sigma_v^2 = \frac{\bar{S}_t^2 \sigma_v^2}{\bar{S}_t^2 + \sigma_v^2}$$

entonces

$$\begin{aligned} \lim_{t \rightarrow \infty} S_t^2 &= \lim_{t \rightarrow \infty} \frac{\bar{S}_t^2 \sigma_v^2}{\bar{S}_t^2 + \sigma_v^2} \\ &= \frac{\bar{S}^2 \sigma_v^2}{\bar{S}^2 + \sigma_v^2} = \frac{[S^2 + \sigma_w^2] \sigma_v^2}{[S^2 + \sigma_w^2] + \sigma_v^2} \\ &= S^2 \end{aligned}$$

de donde se obtiene la ecuación de segundo grado

$$(S^2)^2 + \sigma_w^2 S^2 - \sigma_w^2 \sigma_v^2 = 0$$

Tomando la raíz positiva se obtiene el límite buscado

$$\begin{aligned} S^2 &= \frac{-\sigma_w^2 + \sqrt{(\sigma_w^2)^2 + 4\sigma_w^2 \sigma_v^2}}{2} \\ &= \sigma_w^2 \left[\frac{\sqrt{1 + \frac{4\sigma_v^2}{\sigma_w^2}} - 1}{2} \right] \end{aligned}$$

y

$$A = \frac{S^2}{\sigma_v^2} = \frac{\sigma_w^2}{2\sigma_v^2} \left[\sqrt{1 + \frac{4\sigma_v^2}{\sigma_w^2}} - 1 \right]$$

Para tener una idea de la relación entre los valores que toma A y el cociente $r = \frac{\sigma_w^2}{\sigma_v^2}$, se da la siguiente gráfica

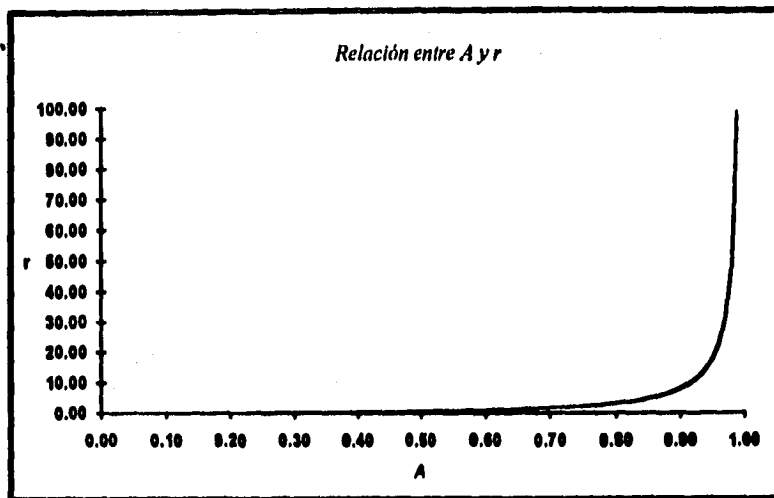


Figura 3.3

De acuerdo a (3.5) se obtiene la siguiente aproximación

$$m_t \approx A(Y_t - \mu_{Y_t}) + (1 - A)(m_{t-1} + \mu_{w_t}) \quad (3.8)$$

En la gráfica anterior se observa que cuando $\frac{\sigma_w^2}{\sigma_Y^2} \rightarrow \infty$ $A \rightarrow 1$, en consecuencia y utilizando (3.8), se puede afirmar que cuando $\sigma_Y^2 \approx 0$, cualquier cambio en Y_t se refleja en el predictor puntual m_t , esto es

$$m_t \approx Y_t - \mu_{Y_t} \quad (3.8.a)$$

Por el contrario, si $A \rightarrow 0$ entonces

$$m_t \approx m_{t-1} + \mu_{w_t} \quad (3.8.b)$$

y el predictor m_t es poco sensible a los cambios en la serie porque los datos más recientes tienen poco peso en él.

Corolario 3.1

En el MLC las sucesiones $\bar{S}_t^2, \hat{\sigma}_t^2$ cumplen con lo siguiente

$$\lim_{t \rightarrow \infty} \bar{S}_t^2 = \bar{S}^2 = \frac{S^2}{1-A} = \frac{A\sigma_v^2}{1-A},$$

$$\lim_{t \rightarrow \infty} \hat{\sigma}_t^2 = \hat{\sigma}^2 = \frac{\sigma_v^2}{1-A},$$

$$\sigma_w^2 = A^2 \hat{\sigma}^2$$

Demostración.

Del teorema 3.3

$$\begin{aligned} \lim_{t \rightarrow \infty} (S_t^2)^{-1} &= \lim_{t \rightarrow \infty} \left[(S_t^2)^{-1} + (\sigma_v^2)^{-1} \right] \\ &= (\sigma_v^2)^{-1} + \lim_{t \rightarrow \infty} (\bar{S}_t^2)^{-1} = (\sigma_v^2)^{-1} + (\bar{S}^2)^{-1} \\ &= \frac{1}{A\sigma_v^2} \end{aligned}$$

de donde se obtiene que

$$(\bar{S}^2)^{-1} = \frac{1-A}{A\sigma_v^2}$$

en consecuencia y por el mismo teorema

$$\bar{S}^2 = \frac{A\sigma_v^2}{1-A} = \frac{S^2}{1-A}.$$

Como $S_t^2 = A_t \sigma_v^2$ ocurre que $\hat{\sigma}_t^2 = \frac{\bar{S}_t^2}{S_t^2} \sigma_v^2$, obteniéndose el límite

$$\begin{aligned} \lim_{t \rightarrow \infty} \hat{\sigma}_t^2 &= \lim_{t \rightarrow \infty} \frac{\bar{S}_t^2}{S_t^2} \sigma_v^2 = \frac{\bar{S}^2}{S^2} \sigma_v^2 \\ &= \frac{S^2 \sigma_v^2}{S^2(1-A)} = \frac{\sigma_v^2}{(1-A)} \end{aligned}$$

Finalmente

$$\begin{aligned} \sigma_w^2 &= \bar{S}^2 - S^2 = \frac{A\sigma_v^2}{1-A} - A\sigma_v^2 \\ &= \frac{A^2\sigma_v^2}{1-A} = A^2 \hat{\sigma}^2 \end{aligned}$$

La experiencia muestra que el modelo constante se aplica adecuadamente para pronósticos a corto plazo y que se cumpla que $5 \leq \frac{\sigma_v^2}{\sigma_w^2} \leq 1000$, esto último con el fin de que el nivel de la serie tienda a parecer una curva suave. La cuestión es ¿qué se entiende por corto plazo? Una respuesta muy general puede ser que si h denota el número de periodos del horizonte de corto plazo, se debe cumplir que $2h \frac{\sigma_w^2}{\sigma_v^2} < 1$, por lo que generalmente el rango $1 \leq h \leq 6$ es una buena idea de lo que se entiende por corto plazo (Harrison y West [9]).

3.4 Especificación de la Varianza de Transición $\{\sigma_w^2\}$

El buen desempeño del *ML* como rutina de pronosticación depende en gran medida de que se escojan los valores apropiados de las varianzas σ_w^2 y σ_v^2 .

Considerese que un pronosticador aplica un modelo constante $\{1, 1, \mu_v, \sigma_v^2, \mu_w, \sigma_w^2\}$ y asigna a las varianzas los valores σ_v^2 y σ_w^2 . El error observacional es $e_i = y_i - \hat{Y}_i$, así que antes de que ocurra la observación se puede considerar a este error como una variable aleatoria

$$\varepsilon_i = Y_i - \hat{Y}_i$$

Se toma la diferencia

$$\begin{aligned} y_i - y_{i-1} &= [\hat{Y}_i + \varepsilon_i] - [\hat{Y}_{i-1} + \varepsilon_{i-1}] \\ &= m_{i-1} + \mu_w + \mu_v + \varepsilon_i - \hat{Y}_{i-1} - \varepsilon_{i-1} \end{aligned}$$

pero como

$$\bar{m}_i = m_i - A_i \varepsilon_i = \hat{Y}_i - \mu_v$$

entonces

$$\hat{Y}_{i-1} = m_{i-1} + \mu_{v,i-1} - A_{i-1} \varepsilon_{i-1}$$

Por lo tanto

$$\begin{aligned} y_i - y_{i-1} &= m_{i-1} + \mu_w + \mu_v + \varepsilon_i - m_{i-1} - \mu_{v,i-1} + A_{i-1} \varepsilon_{i-1} - \varepsilon_{i-1} \\ &= \mu_w + \mu_v + \varepsilon_i - \mu_{v,i-1} + A_{i-1} \varepsilon_{i-1} - \varepsilon_{i-1} \\ &= \mu_w + \mu_v - \mu_{v,i-1} + \varepsilon_i - \varepsilon_{i-1} (1 - A_{i-1}) \end{aligned}$$

Supongase además que la convergencia de la sucesión $\{A_i\}$ es rápida por lo que se obtiene la siguiente expresión

$$\begin{aligned} y_i - y_{i-1} &\approx \varepsilon_i + (\mu_w + \mu_v) - \varepsilon_{i-1} (1 - A) - \mu_{v,i-1} \\ &= \varepsilon_i + (\mu_w + \mu_v) - \varepsilon_{i-1} \delta - \mu_{v,i-1} \end{aligned} \quad (3.9)$$

con $\delta = (1 - A)$. Por otra parte como $e_t = y_t - \hat{Y}_t$ y de acuerdo al *corolario 3.1* entonces

$$(\varepsilon_t | D_0) \sim N[0, \hat{\sigma}_t^2].$$

$$\text{con } \hat{\sigma}_t^2 = \frac{\sigma_v^2}{\delta}.$$

Aún cuando ningún modelo representa totalmente a la realidad supongase que las observaciones son generadas realmente por un modelo constante con varianzas σ_v^2 y σ_w^2 . Si se aplica el proceso de aprendizaje con estas varianzas se obtienen otros pronósticos con medias \hat{Y}_t y en consecuencia se observan diferentes errores de pronóstico $e_{t*} = y_t - \hat{Y}_{t*}$ y se obtiene un diferente límite para el coeficiente adaptativo A , teniéndose la siguiente diferencia

$$\begin{aligned} y_t - y_{t-1} &= m_{t-1} + \mu_{w_t} + \mu_{v_t} + e_{t*} - m_{t-1} - \mu_{v_{t-1}} + A_{t-1} e_{t-1*} - e_{t-1*} \\ &= \mu_{w_t} + \mu_{v_t} + e_{t*} - \mu_{v_{t-1}} + A_{t-1} e_{t-1*} - e_{t-1*} \\ &= \mu_{w_t} + \mu_{v_t} - \mu_{v_{t-1}} + e_{t*} - e_{t-1*} (1 - A_{t-1}) \end{aligned} \quad (3.9.a)$$

y por el *corolario 3.1*

$$(\varepsilon_{t*} | D_0) \sim N[0, \hat{\sigma}_{t*}^2],$$

$$\text{con } \hat{\sigma}_{t*}^2 = \frac{\sigma_{v*}^2}{\delta}.$$

Si igualamos (3.9) con (3.9.a) se tiene que

$$e_t = e_{t*} + (\delta - \delta_*) \sum_{i=0}^{t-1} \delta^i e_{t-i-1*} \quad (3.10)$$

Dada la independencia de la sucesión $\{\varepsilon_{t*}\}$ se tiene que los momentos a priori iniciales para los errores ε_t son

$$E(\varepsilon_t | D_0) = E(\varepsilon_{t*} | D_0) + (\delta - \delta_*) \sum_{i=0}^{t-1} \delta^i E(\varepsilon_{t-i-1*} | D_0)$$

y

$$\begin{aligned}
 AR(\varepsilon_t|D_0) &= VAR(\varepsilon_t|D_0) + (\delta - \delta_0)^2 \sum_{i=0}^{t-1} \delta^{2i} VAR(\varepsilon_{t-i-1}) \\
 &= \hat{\sigma}_{\varepsilon}^2 \left[1 + \frac{(\delta - \delta_0)^2 (1 - \delta^{2t+2})}{(1 - \delta)^2} \right]
 \end{aligned}$$

Por lo tanto

$$\lim_{t \rightarrow \infty} VAR(\varepsilon_t|D_0) = \hat{\sigma}_{\varepsilon}^2 \left[1 + \frac{(\delta - \delta_0)^2}{(1 - \delta)^2} \right] \quad (3.11)$$

De esta ecuación se puede concluir que cuando el cociente $r = \frac{\sigma_w^2}{\sigma_v^2}$ tiende a cero, esto es, cuando $\delta \rightarrow \delta_0$, ocurre que $\hat{\sigma}_{\varepsilon}^2 \left[1 + \frac{(\delta - \delta_0)^2}{(1 - \delta)^2} \right] \rightarrow \infty$. El valor de la varianza de $(\varepsilon_t|D_0)$ diverge cuando $t \rightarrow \infty$ a menos que $\delta = \delta_0$, es decir, que se elijan las varianzas que generan el proceso en realidad. Para elegir la varianza σ_v^2 se debe considerar esta inconveniencia que se agrega al hecho de que cuando $r \rightarrow 0$, el modelo resulta ser poco sensible a los cambios en la serie, como ya se vio en la sección 3.3.

Por el contrario si $r \rightarrow 0$, δ toma valores cercanos a cero, por lo que

$$\lim_{t \rightarrow \infty} VAR(\varepsilon_t|D_0) = \sigma_v^2 (1 + \delta^2) \quad (3.11.a)$$

El escoger valores de r que no sean cercanos a cero, y hacer el modelo sensible a los datos más recientes es preferible en el sentido de que se limita la varianza del error de pronóstico.

3.4.1 Elección por factor de descuento

En el modelo constante se tiene que $\bar{S}_t^2 = S_{t-1}^2 + \sigma_w^2$, al tomar el caso límite se tiene que

$$\lim_{t \rightarrow \infty} \bar{S}_t^2 = \lim_{t \rightarrow \infty} S_{t-1}^2 + \sigma_w^2 = S^2 + \sigma_w^2 \quad (3.12)$$

y de acuerdo al *Corolario 3.1* se tiene entonces que

$$\begin{aligned} \sigma_w^2 &= \frac{S^2}{1-A} - S^2 \\ &= \frac{AS^2}{1-A} \end{aligned} \quad (3.13)$$

por lo que σ_w^2 puede ser vista como una proporción fija de S^2 , esto sugiere que en el caso límite, al pasar del tiempo t al siguiente, el error ε_{w_t} conduce a un incremento de la incertidumbre, y el incremento es en una proporción $\frac{A}{(1-A)}$ de la incertidumbre inicial S^2 .

Esto sugiere una manera de escoger la varianza σ_w^2 en relación a la tasa límite de adaptación de datos A . Si se supone que el comportamiento límite es rápidamente alcanzado en este modelo, es conveniente adoptar una tasa constante de incremento en la incertidumbre, para toda t y no sólo en el caso límite. *El factor de descuento* se define como $\delta = 1 - A$. De acuerdo a esta idea el error de transición cumple con lo siguiente

$$\varepsilon_{w_t} \sim N(\mu_{w_t}, \sigma_{w_t}^2), \quad (3.14)$$

donde

$$\sigma_{w_t}^2 = \frac{S_{t-1}^2(1-\delta)}{\delta} \quad (3.14.a)$$

por lo tanto

$$\bar{S}_t^2 = \frac{S_{t-1}^2}{\delta} \quad (3.14.b)$$

tal varianza no corresponde a un modelo constante, pero su comportamiento límite corresponde al de un modelo constante con $A = 1 - \delta$, como se ve en el siguiente lema.

Lema 3.1

En el modelo polinomial de primer orden con varianza observacional σ_v^2 y varianza de transición $\sigma_w^2 = \frac{S_{t-1}^2(1-\delta)}{\delta}$ siendo $\delta = (1-A)$ el factor de descuento se tiene que

$$\lim_{t \rightarrow \infty} S_t^2 = S^2 = A \sigma_v^2$$

con

$$A = 1 - \delta.$$

Demostración.

Usando (3.8) y (3.14.b) se ve que

$$\begin{aligned} (S_t^2)^{-1} &= (\sigma_v^2)^{-1} + \delta(S_{t-1}^2)^{-1} \\ &= (\sigma_v^2)^{-1} + \delta[(\sigma_v^2)^{-1} + \delta(S_{t-2}^2)^{-1}] \\ &= (\sigma_v^2)^{-1} + \delta(\sigma_v^2)^{-1} + \delta^2[(\sigma_v^2)^{-1} + \delta(S_{t-3}^2)^{-1}] \\ &= \dots \\ &= (\sigma_v^2)^{-1} + \delta(\sigma_v^2)^{-1} + \delta^2(\sigma_v^2)^{-1} + \dots + \delta^i(S_{t-i}^2)^{-1} \\ &= (\sigma_v^2)^{-1} [1 + \delta + \delta^2 + \dots + \delta^{i-1} + \delta^i(S_0^2)^{-1}] \\ &= (\sigma_v^2)^{-1} \left[\frac{1 - \delta^i}{1 - \delta} + \delta^i(S_0^2)^{-1} \right] \end{aligned}$$

entonces

$$\begin{aligned} \lim_{t \rightarrow \infty} (S_t^2)^{-1} &= (\sigma_v^2)^{-1} \lim_{t \rightarrow \infty} \left[\frac{1 - \delta^i}{1 - \delta} \right] + (\sigma_v^2)^{-1} \lim_{t \rightarrow \infty} \delta^i (S_0^2)^{-1} \\ &= \frac{1}{\sigma_v^2(1-\delta)} = \frac{1}{\sigma_v^2 A} \end{aligned}$$

por lo tanto

$$\lim_{t \rightarrow \infty} S_t^2 = A \sigma_v^2 = S^2$$

La idea de factor de descuento permite asignar valores a la secuencia $\sigma_{w_t}^2$ de manera simple y natural, y es aplicable al caso general del modelo polinomial de primer orden, porque tal asignación depende sólo de los valores de S_{t-1}^2 y δ , los cuales son conocidos al tiempo $t-1$, y aún más, si se supone que D_0 contiene los valores de la sucesión $\sigma_{v_t}^2$, entonces la sucesión $\sigma_{w_t}^2$ también es conocida al tiempo $t=0$, porque puede ser calculada a partir de $D_0, S_0^2, \sigma_{v_{t-1}}^2, \dots, \sigma_{v_t}^2$

Para comparar modelos con distintos factores de descuento δ y con el fin de determinar su conveniencia, se definen las siguientes medidas que pueden servir de guía.

Definición 3.3

La *desviación media absoluta DMA* se define como

$$\sum_{i=1}^n \frac{|e_i|}{n},$$

el *error medio al cuadrado EMC* se define como

$$\sum_{i=1}^n \frac{e_i^2}{n},$$

y la *función de verosimilitud dada la información inicial* es

$$\begin{aligned} L(\delta | y_1, \dots, y_n) &= f_{y_1, \dots, y_n | D_0}(y_1, \dots, y_n) \\ &= \prod_{i=1}^n f_{y_i | D_{i-1}}(y_i) \end{aligned}$$

con $\delta \in D_0$.

A continuación se dará un ejemplo en el que se aplica un modelo lineal en el que se escoge la sucesión $\{\sigma_{w_t}^2\}$ a través de factores de descuento.

Ejemplo: Pronóstico del tipo de cambio \$/£

Sea S_t el tipo de cambio dólar-libra esterlina al cierre del mes t , la tasa de crecimiento¹³ de la serie $\{S_t\}$ es

$$\begin{aligned} Y_t &= \ln\left(\frac{S_t}{S_{t-1}}\right) \\ &= \ln(S_t) - \ln(S_{t-1}) \\ &= s_t - s_{t-1} \end{aligned} \quad (3.15)$$

con

$$s_t = \ln S_t$$

En la tabla 3.1 están contenidos los datos correspondientes a las tasas de crecimiento registradas entre enero de 1980 y marzo de 1994. Se escoge un *ML* con varianza $\sigma_v^2 = \sigma_e^2$ para toda t , para describir el comportamiento de la serie $\{Y_t\}$ en el intervalo de tiempo antes mencionado, plantéandose dos problemas:

- (i) ¿Cuál es la información inicial que se debe utilizar?
- (ii) ¿Cómo se eligen las varianzas?

Para (i), existen diversas soluciones, pero por simplicidad y dado que se cuenta con toda la información de la serie entre enero de 1980 y marzo de 1994 que es el intervalo de tiempo en estudio, una solución posible es asignar a m_0 el valor de la media muestral,

$$\bar{y} = \sum_{t=1}^{171} \frac{y_t}{171} = 1.6817 \times 10^{-2}$$

y a S_0^2 un porcentaje del valor de σ_v^2 , (en este ejemplo se toma el 10 %).

Para (ii), una solución posible es considerar que el valor de la varianza observacional σ_v^2 es el mismo de la varianza muestral

¹³ En términos continuos se tiene que, $d \ln S = \frac{dS}{S}$. De acuerdo a lo anterior, en términos discretos

una aproximación de la tasa de crecimiento del tipo de cambio es la siguiente diferencia

$$\ln S_t - \ln S_{t-1} = s_t - s_{t-1} = \Delta \ln S \approx \frac{\Delta S}{S}$$

$$\hat{\sigma}_y^2 = \sum_{i=1}^{171} \frac{(y_i - \bar{y})^2}{170} = 0.0766 \times 10^{-4}$$

	1980	1983	1986	1989	1992
ENE	2.2667	1.5202	1.4120	1.7520	1.7875
FEB	2.2730	1.5115	1.4460	1.7434	1.7577
MAR	2.1600	1.4825	1.4735	1.6855	1.7363
ABR	2.2585	1.5573	1.5510	1.6900	1.7769
MAY	2.3425	1.6005	1.4723	1.5700	1.8297
JUN	2.3555	1.5330	1.5320	1.5500	1.9050
JUL	2.3425	1.5200	1.4915	1.6660	1.9275
AGO	2.3960	1.4940	1.4880	1.5680	1.9884
SEP	2.3870	1.4977	1.4468	1.6205	1.7773
OCT	2.4365	1.4952	1.4055	1.5810	1.5985
NOV	2.3575	1.4630	1.4335	1.5705	1.5171
DEC	2.3890	1.4515	1.4825	1.6125	1.5105
	1981	1984	1987	1990	1993
ENE	2.3655	1.4012	1.5140	1.6780	1.4861
FEB	2.2040	1.4905	1.5460	1.6775	1.4255
MAR	2.2325	1.4375	1.6070	1.6480	1.5145
ABR	2.1395	1.3980	1.6605	1.6420	1.5740
MAY	2.0700	1.3855	1.6295	1.6780	1.5630
JUN	1.9300	1.3575	1.6128	1.7490	1.4912
JUL	1.8400	1.3073	1.5910	1.8635	1.4825
AGO	1.8450	1.3075	1.6328	1.8950	1.4925
SEP	1.8030	1.2340	1.6248	1.8788	1.4960
OCT	1.8620	1.2185	1.7215	1.9435	1.4800
NOV	1.9550	1.1970	1.8256	1.9435	1.4858
DEC	1.9150	1.1580	1.8860	1.9300	1.4775
	1982	1985	1988	1991	1994
ENE	1.8790	1.1305	1.7691	1.9670	1.5069
FEB	1.8200	1.0790	1.7735	1.9130	1.4855
MAR	1.7810	1.2360	1.8865	1.7510	1.4840
ABR	1.7945	1.2410	1.8785	1.7260	
MAY	1.7910	1.2860	1.8382	1.6980	
JUN	1.7355	1.3095	1.7085	1.6220	
JUL	1.7380	1.4110	1.7100	1.6870	
AGO	1.7205	1.3910	1.6825	1.6815	
SEP	1.6945	1.4085	1.6910	1.7510	
OCT	1.6770	1.4405	1.7690	1.7392	
NOV	1.6320	1.4890	1.8493	1.7679	
DEC	1.6180	1.4450	1.8110	1.8707	

Tabla 3.1. Tasas de crecimiento $\times 100$ del tipo de cambio Dólar-Libra Esterlina.

Fuente Bloomberg Market.

A $\sigma_{w_t}^2$ se le determina por medio del factor de descuento δ , pudiendo tomar diferentes valores para este último parámetro con el fin de comparar diferentes grados de sensibilidad a los datos más recientes.

Por otra parte, como no existe ninguna razón para pensar que la tasa de crecimiento presente una clara tendencia hacia arriba o a la baja, se decide tomar a $\mu_v = 0$ y $\mu_w = 0$ para toda t .

A continuación se mostrará el desarrollo de la rutina de aprendizaje entre los tiempos $t=0$ y $t=$, proporcionada por un modelo lineal en el que se toma como factor de descuento a $\delta = 0.8$ ¹⁴.

Información Inicial

$$D_0 = \left\{ \begin{array}{ll} \mu_v = 0 & \sigma_v^2 = 0.0766 \times 10^{-4} \\ \mu_w = 0 & \sigma_w^2 = S_{t-1}^2 \left(\frac{1-\delta}{\delta} \right), \delta = 0.8 \\ m_0 = 1.6817 \times 10^{-2} & S_0^2 = 0.0076 \times 10^{-4} \end{array} \right\}$$

De acuerdo a la información contenida en D_0 , se tiene la siguiente distribución inicial

$$(\Theta_0 | D_0) \sim N[1.6817 \times 10^{-2}, 0.0076 \times 10^{-4}]$$

Transición

Como la elección de $\{\sigma_w^2\}$ es por factor de descuento entonces

$$\begin{aligned} \sigma_w^2 &= S_0^2 \left(\frac{1-0.8}{0.8} \right) \\ &= (0.0076 \times 10^{-4}) \times (0.25) \\ &= 0.0019 \times 10^{-4} \end{aligned}$$

Los momentos de la distribución a priori de Θ_1 se obtienen de la siguiente forma

$$\begin{aligned} \bar{m}_1 &= m_0 + \mu_v \\ &= 1.6817 \times 10^{-2} + 0 = 1.6817 \times 10^{-2}, \end{aligned}$$

¹⁴ Las operaciones efectuadas en este desarrollo para obtener los datos contenidos en la tabla 3.2 fueron realizadas con un hoja de cálculo electrónica (Excel 4.0).

$$\begin{aligned}\bar{S}_1^2 &= S_0^2 + \sigma_{\mu_1}^2 \\ &= 0.0076 \times 10^{-4} + 0.0019 \times 10^{-4} \\ &= 0.0095 \times 10^{-4}\end{aligned}$$

Por lo tanto se tiene que la distribución a priori para Θ_1 es

$$(\Theta_1 | D_0) \sim N[1.6817 \times 10^{-2}, 0.0095 \times 10^{-4}]$$

Pronóstico

Los momentos de la distribución del pronóstico $f_0(1)$, son obtenidos realizando las siguientes operaciones

$$\begin{aligned}\hat{Y}_1 &= \bar{m}_1 + \mu_{\nu_1} \\ &= 1.6817 \times 10^{-2} + 0 = 1.6817 \times 10^{-2},\end{aligned}$$

$$\begin{aligned}\hat{\sigma}_1^2 &= \bar{S}_1^2 + \sigma_{\nu_1}^2 \\ &= 0.0095 \times 10^{-4} + 0.0766 \times 10^{-4} = 0.0862 \times 10^{-4}.\end{aligned}$$

Entonces la distribución de pronóstico es

$$(Y_1 | D_0) \sim N[1.6817 \times 10^{-2}, 0.0862 \times 10^{-4}]$$

Actualización

Una vez presentada la observación $y_1 = 2.2667 \times 10^{-2}$, este nuevo dato se debe de incorporar al conjunto de información disponible

$$D_1 = \{y_1 = 2.2667 \times 10^{-2}, D_0\}$$

Se calcula el error de pronóstico y el coeficiente adaptativo

$$\begin{aligned}e_1 &= y_1 - \hat{Y}_1 \\ &= (2.2667 \times 10^{-2} - 1.6817 \times 10^{-2}) \\ &= 0.5850 \times 10^{-2},\end{aligned}$$

$$A_1 = \frac{\bar{S}_1^2}{\hat{\sigma}_1^2} = \frac{0.0095 \times 10^{-4}}{0.0862 \times 10^{-4}} = 0.1111$$

Finalmente se procede a actualizar el conocimiento que se tiene sobre Θ_t

$$\begin{aligned} m_t &= \bar{m}_t + A_t e_t \\ &= 1.6817 \times 10^{-2} + (0.1111) \times (0.5850 \times 10^{-2}) \\ &= 1.7467 \times 10^{-2} \end{aligned}$$

$$\begin{aligned} S_t^2 &= A_t \sigma_{\epsilon_t}^2 \\ &= (0.1111) \times (0.0766 \times 10^{-4}) \\ &= 0.0085 \times 10^{-4} \end{aligned}$$

obteniéndose la distribución posterior para Θ_t

$$(\Theta_t | D_t) \sim N[1.7467 \times 10^{-2}, 0.0085 \times 10^{-4}]$$

Continuando con este procedimiento se obtienen los datos contenidos en la tabla 3.2. En ella se presentan las medias \hat{Y}_t y varianzas $\hat{\sigma}_t^2$ de las distribuciones de pronóstico $f_{t-1}(1)$ para cada periodo comprendido entre enero de 1980 y marzo de 1994. Se utilizaron tres diferentes factores de descuento, $\{0.8, 0.9, 1.0\}$, con el fin de comparar diferentes grados de adaptabilidad a la información más reciente.

Cabe hacer notar que, como se supuso que las medias μ_w y μ_v son iguales a cero para toda t , entonces se cumple que

$$\hat{Y}_t = \bar{m}_t = m_{t-1}$$

por lo que en la tabla 3.2 se pueden encontrar los valores de los tres diferentes predictores puntuales.

<i>t</i>	$y_i \times 100$	$\delta = 1.0$		$\delta = 0.9$		$\delta = 0.8$	
		$\hat{Y}_i \times 100$	$\hat{\sigma}_i^2 \times 10,000$	$\hat{Y}_i \times 100$	$\hat{\sigma}_i^2 \times 10,000$	$\hat{Y}_i \times 100$	$\hat{\sigma}_i^2 \times 10,000$
80							
<i>E</i>	2.2667	1.6817	0.0843	1.6817	0.0851	1.6817	0.0862
<i>F</i>	2.2730	1.7349	0.0836	1.7402	0.0851	1.7467	0.0872
<i>M</i>	2.1600	1.7797	0.0830	1.7935	0.0851	1.8109	0.0883
<i>A</i>	2.2585	1.8090	0.0825	1.8301	0.0851	1.8571	0.0893
<i>M</i>	2.3425	1.8411	0.0821	1.8730	0.0851	1.9140	0.0902
<i>J</i>	2.3555	1.8745	0.0817	1.9199	0.0851	1.9786	0.0910
<i>J</i>	2.3425	1.9046	0.0814	1.9635	0.0851	2.0383	0.0918
<i>A</i>	2.3960	1.9303	0.0811	2.0014	0.0851	2.0886	0.0924
<i>S</i>	2.3870	1.9562	0.0809	2.0409	0.0851	2.1412	0.0930
<i>O</i>	2.4365	1.9789	0.0806	2.0755	0.0851	2.1846	0.0935
<i>N</i>	2.3575	2.0018	0.0804	2.1116	0.0851	2.2301	0.0939
<i>D</i>	2.3890	2.0187	0.0803	2.1362	0.0851	2.2535	0.0942
81							
<i>E</i>	2.3655	2.0355	0.0801	2.1614	0.0851	2.2789	0.0945
<i>F</i>	2.2040	2.0499	0.0799	2.1819	0.0851	2.2953	0.0948
<i>M</i>	2.2325	2.0563	0.0798	2.1841	0.0851	2.2778	0.0950
<i>A</i>	2.1395	2.0634	0.0797	2.1889	0.0851	2.2691	0.0951
<i>M</i>	2.0700	2.0663	0.0796	2.1840	0.0851	2.2439	0.0952
<i>J</i>	1.9300	2.0664	0.0794	2.1726	0.0851	2.2099	0.0953
<i>J</i>	1.8400	2.0616	0.0793	2.1483	0.0851	2.1549	0.0954
<i>A</i>	1.8450	2.0539	0.0792	2.1175	0.0851	2.0928	0.0955
<i>S</i>	1.8030	2.0469	0.0792	2.0902	0.0851	2.0438	0.0955
<i>O</i>	1.8620	2.0391	0.0791	2.0615	0.0851	1.9961	0.0956
<i>N</i>	1.9550	2.0335	0.0790	2.0416	0.0851	1.9695	0.0956
<i>D</i>	1.9150	2.0312	0.0789	2.0329	0.0851	1.9666	0.0956
82							
<i>E</i>	1.8790	2.0277	0.0789	2.0211	0.0851	1.9563	0.0957
<i>F</i>	1.8200	2.0235	0.0788	2.0069	0.0851	1.9409	0.0957
<i>M</i>	1.7810	2.0178	0.0787	1.9882	0.0851	1.9168	0.0957
<i>A</i>	1.7945	2.0114	0.0787	1.9675	0.0851	1.8897	0.0957
<i>M</i>	1.7910	2.0057	0.0786	1.9502	0.0851	1.8707	0.0957
<i>J</i>	1.7355	2.0002	0.0786	1.9343	0.0851	1.8548	0.0957
<i>J</i>	1.7380	1.9936	0.0785	1.9144	0.0851	1.8310	0.0957
<i>A</i>	1.7205	1.9874	0.0785	1.8968	0.0851	1.8124	0.0957
<i>S</i>	1.6945	1.9810	0.0784	1.8791	0.0851	1.7940	0.0957
<i>O</i>	1.6770	1.9744	0.0784	1.8607	0.0851	1.7741	0.0957
<i>N</i>	1.6320	1.9676	0.0783	1.8423	0.0851	1.7547	0.0957
<i>D</i>	1.6180	1.9601	0.0783	1.8213	0.0851	1.7302	0.0958

Tabla 3.2 Sección I

<i>t</i>	$y_i \times 100$	$\delta = 1.0$		$\delta = 0.9$		$\delta = 0.8$	
		$\hat{y}_i \times 100$	$\hat{\sigma}_i^2 \times 10,000$	$\hat{y}_i \times 100$	$\hat{\sigma}_i^2 \times 10,000$	$\hat{y}_i \times 100$	$\hat{\sigma}_i^2 \times 10,000$
83							
<i>E</i>	1.5202	1.9527	0.0783	1.8009	0.0851	1.7078	0.0958
<i>F</i>	1.5115	1.9435	0.0782	1.7729	0.0851	1.6703	0.0958
<i>M</i>	1.4825	1.9345	0.0782	1.7467	0.0851	1.6385	0.0958
<i>A</i>	1.5573	1.9253	0.0782	1.7203	0.0851	1.6073	0.0958
<i>M</i>	1.6005	1.9179	0.0781	1.7040	0.0851	1.5973	0.0958
<i>J</i>	1.5330	1.9117	0.0781	1.6937	0.0851	1.5979	0.0958
<i>J</i>	1.5200	1.9044	0.0781	1.6776	0.0851	1.5850	0.0958
<i>A</i>	1.4940	1.8972	0.0781	1.6618	0.0851	1.5720	0.0958
<i>S</i>	1.4977	1.8897	0.0780	1.6450	0.0851	1.5564	0.0958
<i>O</i>	1.4952	1.8826	0.0780	1.6303	0.0851	1.5446	0.0958
<i>N</i>	1.4630	1.8756	0.0780	1.6168	0.0851	1.5348	0.0958
<i>D</i>	1.4515	1.8684	0.0780	1.6014	0.0851	1.5204	0.0958
84							
<i>E</i>	1.4012	1.8612	0.0779	1.5864	0.0851	1.5066	0.0958
<i>F</i>	1.4905	1.8534	0.0779	1.5679	0.0851	1.4855	0.0958
<i>M</i>	1.4375	1.8474	0.0779	1.5602	0.0851	1.4865	0.0958
<i>A</i>	1.3980	1.8407	0.0779	1.5479	0.0851	1.4767	0.0958
<i>M</i>	1.3855	1.8335	0.0778	1.5329	0.0851	1.4610	0.0958
<i>J</i>	1.3575	1.8264	0.0778	1.5182	0.0851	1.4459	0.0958
<i>J</i>	1.3073	1.8191	0.0778	1.5021	0.0851	1.4282	0.0958
<i>A</i>	1.3075	1.8112	0.0778	1.4826	0.0851	1.4040	0.0958
<i>S</i>	1.2340	1.8036	0.0778	1.4651	0.0851	1.3847	0.0958
<i>O</i>	1.2185	1.7951	0.0778	1.4420	0.0851	1.3546	0.0958
<i>N</i>	1.1970	1.7866	0.0777	1.4196	0.0851	1.3274	0.0958
<i>D</i>	1.1580	1.7780	0.0777	1.3974	0.0851	1.3013	0.0958
85							
<i>E</i>	1.1305	1.7692	0.0777	1.3734	0.0851	1.2726	0.0958
<i>F</i>	1.0790	1.7602	0.0777	1.3492	0.0851	1.2442	0.0958
<i>M</i>	1.2360	1.7507	0.0777	1.3221	0.0851	1.2112	0.0958
<i>A</i>	1.2410	1.7437	0.0777	1.3135	0.0851	1.2161	0.0958
<i>M</i>	1.2860	1.7369	0.0776	1.3063	0.0851	1.2211	0.0958
<i>J</i>	1.3095	1.7309	0.0776	1.3042	0.0851	1.2341	0.0958
<i>J</i>	1.4110	1.7253	0.0776	1.3048	0.0851	1.2492	0.0958
<i>A</i>	1.3910	1.7213	0.0776	1.3154	0.0851	1.2815	0.0958
<i>S</i>	1.4085	1.7170	0.0776	1.3230	0.0851	1.3034	0.0958
<i>O</i>	1.4405	1.7131	0.0776	1.3315	0.0851	1.3244	0.0958
<i>N</i>	1.4890	1.7097	0.0776	1.3424	0.0851	1.3477	0.0958
<i>D</i>	1.4450	1.7070	0.0776	1.3571	0.0851	1.3759	0.0958

Tabla 3.2 Sección II

<i>t</i>	$y_t \times 100$	$\delta = 1.0$		$\delta = 0.9$		$\delta = 0.8$	
		$\hat{y}_t \times 100$	$\hat{\sigma}_t^2 \times 10,000$	$\hat{y}_t \times 100$	$\hat{\sigma}_t^2 \times 10,000$	$\hat{y}_t \times 100$	$\hat{\sigma}_t^2 \times 10,000$
86							
<i>E</i>	1.4120	1.7038	0.0775	1.3659	0.0851	1.3897	0.0958
<i>F</i>	1.4460	1.7003	0.0775	1.3705	0.0851	1.3942	0.0958
<i>M</i>	1.4735	1.6972	0.0775	1.3780	0.0851	1.4046	0.0958
<i>A</i>	1.5510	1.6946	0.0775	1.3876	0.0851	1.4183	0.0958
<i>M</i>	1.4723	1.6929	0.0775	1.4039	0.0851	1.4449	0.0958
<i>J</i>	1.5320	1.6904	0.0775	1.4108	0.0851	1.4504	0.0958
<i>J</i>	1.4915	1.6886	0.0775	1.4229	0.0851	1.4667	0.0958
<i>A</i>	1.4880	1.6864	0.0775	1.4297	0.0851	1.4716	0.0958
<i>S</i>	1.4468	1.6842	0.0775	1.4356	0.0851	1.4749	0.0958
<i>O</i>	1.4055	1.6816	0.0774	1.4367	0.0851	1.4693	0.0958
<i>N</i>	1.4335	1.6786	0.0774	1.4336	0.0851	1.4565	0.0958
<i>D</i>	1.4825	1.6759	0.0774	1.4336	0.0851	1.4519	0.0958
87							
<i>E</i>	1.5140	1.6739	0.0774	1.4385	0.0851	1.4580	0.0958
<i>F</i>	1.5460	1.6722	0.0774	1.4460	0.0851	1.4692	0.0958
<i>M</i>	1.6070	1.6709	0.0774	1.4560	0.0851	1.4846	0.0958
<i>A</i>	1.6605	1.6702	0.0774	1.4711	0.0851	1.5091	0.0958
<i>M</i>	1.6295	1.6701	0.0774	1.4900	0.0851	1.5394	0.0958
<i>J</i>	1.6128	1.6697	0.0774	1.5040	0.0851	1.5574	0.0958
<i>J</i>	1.5910	1.6691	0.0774	1.5149	0.0851	1.5685	0.0958
<i>A</i>	1.6328	1.6684	0.0774	1.5225	0.0851	1.5730	0.0958
<i>S</i>	1.6248	1.6680	0.0774	1.5335	0.0851	1.5849	0.0958
<i>O</i>	1.7215	1.6676	0.0774	1.5426	0.0851	1.5929	0.0958
<i>N</i>	1.8256	1.6681	0.0773	1.5605	0.0851	1.6186	0.0958
<i>D</i>	1.8860	1.6696	0.0773	1.5870	0.0851	1.6600	0.0958
88							
<i>E</i>	1.7691	1.6717	0.0773	1.6169	0.0851	1.7052	0.0958
<i>F</i>	1.7735	1.6726	0.0773	1.6322	0.0851	1.7180	0.0958
<i>M</i>	1.8865	1.6735	0.0773	1.6463	0.0851	1.7291	0.0958
<i>A</i>	1.8785	1.6755	0.0773	1.6703	0.0851	1.7606	0.0958
<i>M</i>	1.8382	1.6773	0.0773	1.6911	0.0851	1.7842	0.0958
<i>J</i>	1.7085	1.6788	0.0773	1.7058	0.0851	1.7950	0.0958
<i>J</i>	1.7100	1.6790	0.0773	1.7061	0.0851	1.7777	0.0958
<i>A</i>	1.6825	1.6793	0.0773	1.7065	0.0851	1.7641	0.0958
<i>S</i>	1.6910	1.6793	0.0773	1.7041	0.0851	1.7478	0.0958
<i>O</i>	1.7690	1.6794	0.0773	1.7028	0.0851	1.7364	0.0958
<i>N</i>	1.8493	1.6802	0.0773	1.7094	0.0851	1.7430	0.0958
<i>D</i>	1.8110	1.6816	0.0773	1.7234	0.0851	1.7642	0.0958

Tabla 3.2 Sección III

<i>t</i>	$y_t \times 100$	$\delta = 1.0$		$\delta = 0.9$		$\delta = 0.8$	
		$\hat{y}_t \times 100$	$\hat{\sigma}_t^2 \times 10,000$	$\hat{y}_t \times 100$	$\hat{\sigma}_t^2 \times 10,000$	$\hat{y}_t \times 100$	$\hat{\sigma}_t^2 \times 10,000$
89							
<i>E</i>	1.7520	1.6827	0.0773	1.7322	0.0851	1.7736	0.0958
<i>F</i>	1.7434	1.6833	0.0773	1.7341	0.0851	1.7693	0.0958
<i>M</i>	1.6835	1.6838	0.0772	1.7351	0.0851	1.7641	0.0958
<i>A</i>	1.6900	1.6838	0.0772	1.7301	0.0851	1.7484	0.0958
<i>M</i>	1.5700	1.6839	0.0772	1.7261	0.0851	1.7367	0.0958
<i>J</i>	1.5500	1.6830	0.0772	1.7105	0.0851	1.7034	0.0958
<i>J</i>	1.6660	1.6819	0.0772	1.6944	0.0851	1.6727	0.0958
<i>A</i>	1.5680	1.6818	0.0772	1.6916	0.0851	1.6714	0.0958
<i>S</i>	1.6205	1.6809	0.0772	1.6792	0.0851	1.6507	0.0958
<i>O</i>	1.5810	1.6804	0.0772	1.6734	0.0851	1.6446	0.0958
<i>N</i>	1.5705	1.6796	0.0772	1.6641	0.0851	1.6319	0.0958
<i>D</i>	1.6125	1.6788	0.0772	1.6548	0.0851	1.6196	0.0958
90							
<i>E</i>	1.6780	1.6783	0.0772	1.6505	0.0851	1.6182	0.0958
<i>F</i>	1.6775	1.6783	0.0772	1.6533	0.0851	1.6302	0.0958
<i>M</i>	1.6480	1.6782	0.0772	1.6557	0.0851	1.6396	0.0958
<i>A</i>	1.6420	1.6780	0.0772	1.6549	0.0851	1.6413	0.0958
<i>M</i>	1.6780	1.6777	0.0772	1.6536	0.0851	1.6414	0.0958
<i>J</i>	1.7490	1.6778	0.0772	1.6561	0.0851	1.6488	0.0958
<i>J</i>	1.8635	1.6783	0.0772	1.6654	0.0851	1.6688	0.0958
<i>A</i>	1.8950	1.6796	0.0772	1.6852	0.0851	1.7077	0.0958
<i>S</i>	1.8788	1.6812	0.0772	1.7062	0.0851	1.7452	0.0958
<i>O</i>	1.9435	1.6826	0.0772	1.7234	0.0851	1.7719	0.0958
<i>N</i>	1.9435	1.6845	0.0772	1.7454	0.0851	1.8062	0.0958
<i>D</i>	1.9300	1.6863	0.0772	1.7652	0.0851	1.8337	0.0958
91							
<i>E</i>	1.9670	1.6880	0.0771	1.7817	0.0851	1.8529	0.0958
<i>F</i>	1.9130	1.6900	0.0771	1.8002	0.0851	1.8758	0.0958
<i>M</i>	1.7510	1.6915	0.0771	1.8115	0.0851	1.8832	0.0958
<i>A</i>	1.7260	1.6919	0.0771	1.8055	0.0851	1.8568	0.0958
<i>M</i>	1.6980	1.6922	0.0771	1.7975	0.0851	1.8306	0.0958
<i>J</i>	1.6220	1.6922	0.0771	1.7876	0.0851	1.8041	0.0958
<i>J</i>	1.6870	1.6917	0.0771	1.7710	0.0851	1.7677	0.0958
<i>A</i>	1.6815	1.6917	0.0771	1.7626	0.0851	1.7515	0.0958
<i>S</i>	1.7510	1.6916	0.0771	1.7545	0.0851	1.7375	0.0958
<i>O</i>	1.7392	1.6920	0.0771	1.7542	0.0851	1.7402	0.0958
<i>N</i>	1.7679	1.6923	0.0771	1.7527	0.0851	1.7400	0.0958
<i>D</i>	1.8707	1.6928	0.0771	1.7542	0.0851	1.7456	0.0958

Tabla 3.2 Sección IV

t	y _t × 100	δ = 1.0		δ = 0.9		δ = 0.8	
		Ŷ _t × 100	σ̂ _t ² × 10,000	Ŷ _t × 100	σ̂ _t ² × 10,000	Ŷ _t × 100	σ̂ _t ² × 10,000
92							
E	1.7875	1.6940	0.0771	1.7658	0.0851	1.7706	0.0958
F	1.7577	1.6946	0.0771	1.7680	0.0851	1.7740	0.0958
M	1.7363	1.6950	0.0771	1.7670	0.0851	1.7707	0.0958
A	1.7769	1.6953	0.0771	1.7639	0.0851	1.7638	0.0958
M	1.8297	1.6958	0.0771	1.7652	0.0851	1.7665	0.0958
J	1.9050	1.6966	0.0771	1.7717	0.0851	1.7791	0.0958
J	1.9275	1.6979	0.0771	1.7850	0.0851	1.8043	0.0958
A	1.9884	1.6993	0.0771	1.7992	0.0851	1.8289	0.0958
S	1.7773	1.7011	0.0771	1.8182	0.0851	1.8608	0.0958
O	1.5585	1.7016	0.0771	1.8141	0.0851	1.8441	0.0958
N	1.5171	1.7007	0.0771	1.7885	0.0851	1.7870	0.0958
D	1.5105	1.6996	0.0771	1.7614	0.0851	1.7330	0.0958
93							
E	1.4861	1.6985	0.0771	1.7363	0.0851	1.6885	0.0958
F	1.4255	1.6972	0.0771	1.7113	0.0851	1.6480	0.0958
M	1.5145	1.6956	0.0771	1.6827	0.0851	1.6035	0.0958
A	1.5740	1.6945	0.0771	1.6659	0.0851	1.5857	0.0958
M	1.5630	1.6938	0.0771	1.6567	0.0851	1.5834	0.0958
J	1.4912	1.6930	0.0771	1.6473	0.0851	1.5793	0.0958
J	1.4825	1.6919	0.0771	1.6317	0.0851	1.5617	0.0958
A	1.4925	1.6907	0.0770	1.6168	0.0851	1.5458	0.0958
S	1.4960	1.6895	0.0770	1.6044	0.0851	1.5352	0.0958
O	1.4800	1.6884	0.0770	1.5935	0.0851	1.5273	0.0958
N	1.4858	1.6872	0.0770	1.5822	0.0851	1.5179	0.0958
D	1.4775	1.6861	0.0770	1.5725	0.0851	1.5115	0.0958
94							
E	1.5069	1.6849	0.0770	1.5630	0.0851	1.5047	0.0958
F	1.4855	1.6838	0.0770	1.5574	0.0851	1.5051	0.0958
M	1.4840	1.6828	0.0770	1.5502	0.0851	1.5012	0.0958

Tabla 3.2 sección V

Es interesante observar como los datos de la tabla 3.2 confirman lo expuesto en el lema 3.1: la convergencia de la sucesión $\{S_t^2\}$ al límite $S^2 = (1 - \delta) \times (0.0766 \times 10^{-4})$ y con ello la convergencia de $\{\hat{\sigma}_t^2\}$ al límite $\hat{\sigma}^2 = \frac{0.0766 \times 10^{-4}}{\delta}$.

En el caso en que $\delta = 1.0$ la sucesión $\{\hat{\sigma}_t^2\}$, converge más lentamente que en los otros dos casos. El modelo aplicado con este factor δ es rígido, se adapta lentamente a los nuevos datos, por lo tanto, no es el mejor para describir el comportamiento de una serie de tiempo en estudio. El siguiente análisis gráfico apoya esta afirmación.

En las figuras 3.4, 3.5 y 3.6 se grafican, para los diferentes factores de descuento, tanto las observaciones y_t , formando una línea continua, como las medias \hat{Y}_t , formando estas una línea punteada.

Para el caso $\delta = 1.0$, se puede observar que ambas líneas presentan amplias diferencias en ciertas regiones. La línea de datos tiene un gran número de quiebres y en cambio la línea de pronóstico es bastante suave, no presenta cambios bruscos de un tiempo t al siguiente. De hecho a partir de 1988 la curva de pronósticos parece una función constante, los valores que toman los predictores \hat{Y}_t están acotados entre $.67 \times 10^{-2}$ y $.71 \times 10^{-2}$, aún cuando las observaciones constantemente salen de ese rango a partir del mismo año. Este análisis gráfico señala que con $\delta = 1.0$, \hat{Y}_t es poco sensible a los datos recientes y parece que no se ajusta a los cambios que se presentan en la serie.

En contraste al utilizar el factor de descuento $\delta = 0.8$, se observa que ambas líneas muestran patrones de comportamiento muy similares. Se podría decir que la línea punteada va siguiendo a la línea continua, el predictor puntual \hat{Y}_t rápidamente se ajusta a los datos más recientes. El modelo es muy sensible a los cambios en la serie ajustándose de acuerdo a la información más reciente.

En el caso en que se escoge un factor $\delta = 0.9$, se tiene un comportamiento intermedio de los dos anteriores.

De acuerdo a estas primeras gráficas al parecer el factor que parece ser el más adecuado para describir el comportamiento de la serie es $\delta = 0.8$. Para apoyar esta idea, en las figuras 3.4.a, 3.5.a y 3.6.a se grafican los correspondientes residuales para cada periodo.

Al analizar gráficamente a los errores de pronóstico o residuales, nuevamente el primer caso es el que parece menos apropiado para realizar pronósticos: la magnitud de los errores es muy grande y sólo cuando la variabilidad de la serie disminuyó los pronósticos fueron más acertados. Esto confirma lo establecido en la sección 3.4: si $\delta \rightarrow 1$, la varianza del error de pronóstico tiende a ser muy grande.

En cambio cuando se tomó $\delta = 0.8$, los errores de pronóstico no son tan grandes como en los otros casos, de hecho se puede decir que los pronósticos son bastante buenos. Los

errores más importantes se registraron cuando la serie presentaba cambios muy bruscos de un periodo a otro, como el registrado de agosto a septiembre de 1992.

Se debe observar que en los tres casos, en los primeros periodos, los errores de pronóstico son relativamente más grandes respecto a los presentados hacia la mitad del intervalo de tiempo en estudio y parecen tener un patrón muy similar. Este fenómeno es atribuible a la información inicial que los tres comparten, al parecer esta no fue muy apropiada. De alguna manera el valor que se asignó a m_0 desvió el desarrollo de los primeros pronósticos puntuales $\{\hat{y}_t\}$. Sin embargo en los casos en que los predictores puntuales son sensibles a los datos recientes, la incorporación de nuevas observaciones y la pérdida de peso de la información inicial en los pronóstico, corrigen las fallas ocasionadas por D_0 .

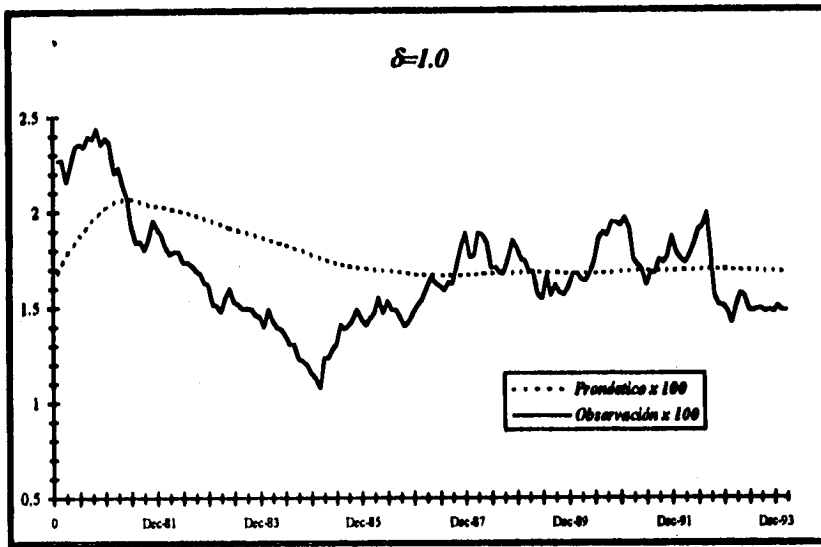


Figura 3.4

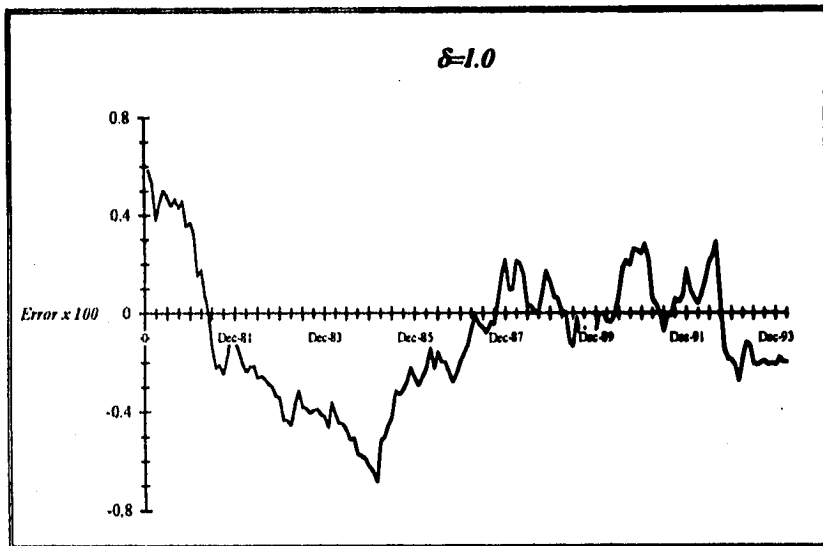


Figura 3.4.a

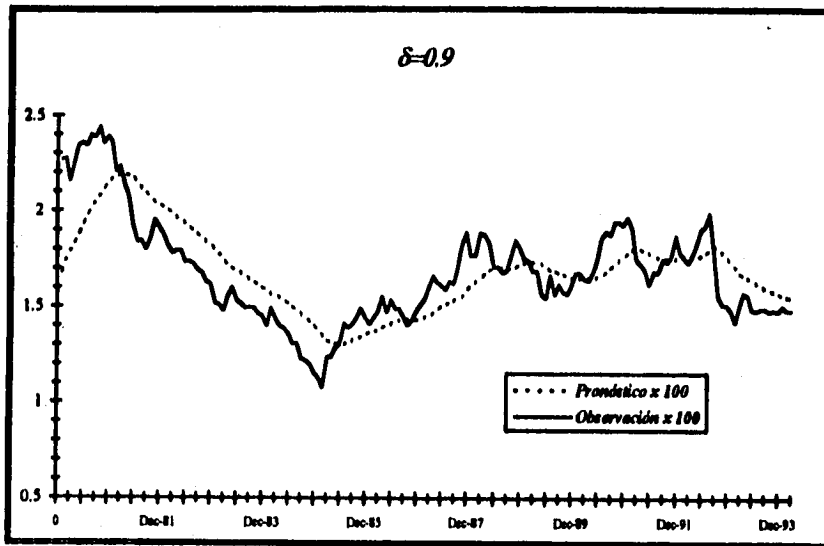


Figura 3.5

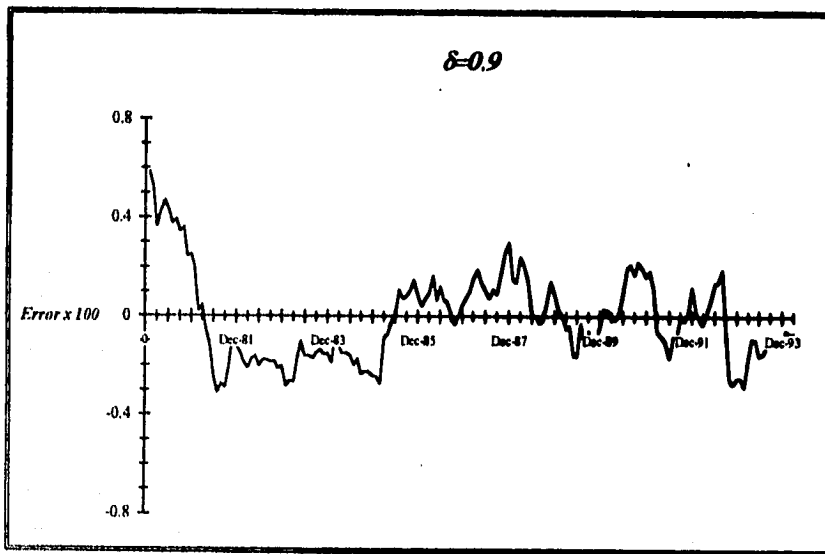


Figura 3.5.a

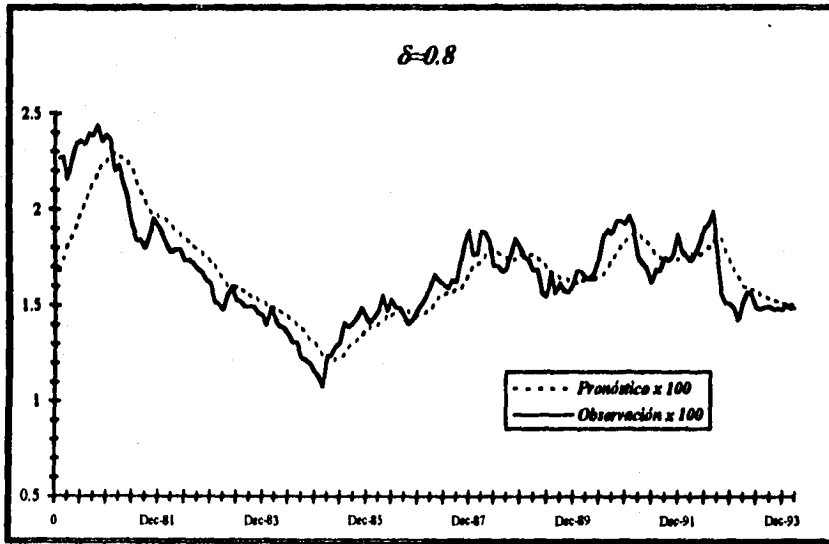


Figura 3.6

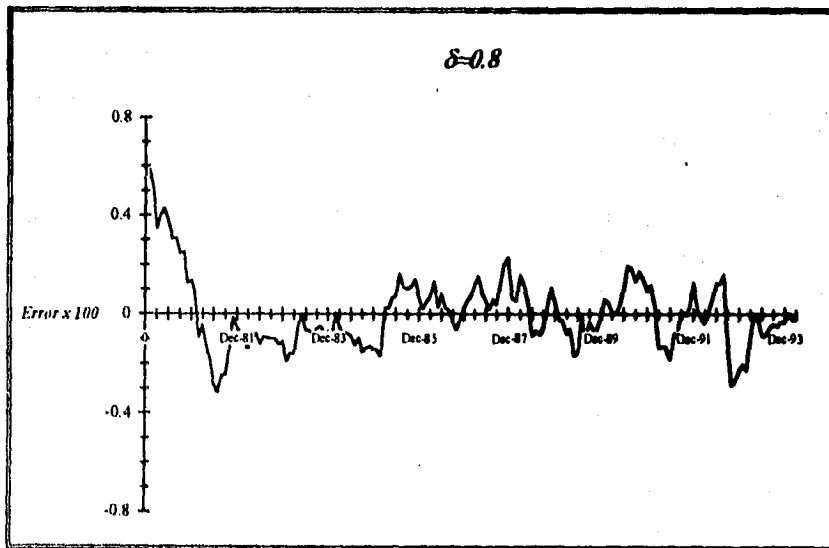


Figura 3.6.a

Para un análisis cuantitativo del rendimiento del modelo como método de pronosticación para diferentes factores de descuento, a continuación se presenta una tabla comparativa que muestra para las δ utilizadas la desviación media absoluta (DMA) y el error medio al cuadrado (EMC). Por otra parte en vez de utilizar a la función de verosimilitud $L(d|y_1, \dots, y_{171})$, con el fin de tener una cantidad con la cual comparar, se utilizará el cociente $\frac{\ln L(d|y_1, \dots, y_{171})}{\ln L(1.0|y_1, \dots, y_{171})}$, cuyo valor indica el grado de soporte de los datos.

δ	DMA	EMC	$\frac{\ln L(d y_1, \dots, y_{171})}{\ln L(1.0 y_1, \dots, y_{171})}$
1.0	0.2285	0.0781	0.0000
0.9	0.1458	0.0324	44.8780
0.8	0.1087	0.0208	48.5150

Tabla 3.3

Este último análisis confirma que el factor de descuento más apropiado es $\delta = 0.8$ al presentar la menor desviación media absoluta, el menor error medio al cuadrado y ser el factor con el que los datos son más verosímiles.

Con este ejemplo se muestra que el modelo lineal es un método cuya mecánica es muy sencilla de aplicar: una hoja de cálculo electrónica permite hacer los cálculos de una manera rápida y sencilla. Sin embargo, se debe prestar especial atención a la información inicial que se va a utilizar. Esto último es con el fin de evitar que el modelo sea poco descriptivo y no corresponda en nada al proceso en estudio, teniendo como consecuencia que los pronósticos no sean lo suficientemente efectivos.

La elección del factor de descuento, contenido también en la información inicial, es determinante en la flexibilidad o rigidez de la rutina de aprendizaje. Con la práctica se descubre cuál es el factor más adecuado para cada problema en particular. En el ejemplo, entre más flexible es el modelo, mejores pronósticos se obtienen. Esto puede deberse a que las fluctuaciones del tipo de cambio, en mercados no inestables, presentan cierta inercia, si se presenta una alza significativa es muy probable que se mantenga una tendencia a la alza

durante algún tiempo: la magnitud de la tasa de crecimiento del tipo de cambio de hoy es más parecida a la de ayer que a la de hace un mes.

Para completar este ejemplo, considerando un horizonte de 6 periodos a partir de $t = \text{marzo } 94$, se proporcionan los siguientes pronósticos.

t		$\delta = 1.0$		$\delta = 0.9$		$\delta = 0.8$	
		$\hat{Y}_t \times 100$	$\hat{\sigma}_t^2 \times 10,000$	$\hat{Y}_t \times 10$	$\hat{\sigma}_t^2 \times 10,000$	$\hat{Y}_t \times 100$	$\hat{\sigma}_t^2 \times 10,000$
94							
A	$f_t(1)$	1.6817	0.0770	1.5436	0.0851	1.4978	0.0958
M	$f_t(2)$	1.6817	0.0770	1.5436	0.0860	1.4978	0.0996
J	$f_t(3)$	1.6817	0.0770	1.5436	0.0868	1.4978	0.1034
J	$f_t(4)$	1.6817	0.0770	1.5436	0.0877	1.4978	0.1073
A	$f_t(5)$	1.6817	0.0770	1.5436	0.0885	1.4978	0.1111
S	$f_t(6)$	1.6817	0.0770	1.5436	0.0894	1.4978	0.1149

Tabla 3.4

Los predictores puntuales no cambian debido a que se supuso que $\mu_v = \mu_w = 0$, para toda t . Sin embargo se debe observar que conforme se va más hacia adelante en el horizonte, la varianza de la distribución de pronóstico se va incrementando, excepto en el caso en que $\delta = 1.0$ al ocurrir que

$$\begin{aligned} AR[f_t(k)] &= S_t^2 + \sum_{i=1}^k 0 + \sigma_v^2 \\ &= S_t^2 + \sigma_v^2 \end{aligned}$$

Este hecho ratifica la inconveniencia de realizar pronósticos con un modelo rígido.¹⁵

¹⁵ Desafortunadamente al momento de terminar este trabajo no se contaba con los datos publicados para el periodo entre abril y septiembre de 1994, por lo que no fue posible confrontarlos con los pronósticos.

3.5 Estimación de la Varianza Observacional $\{\sigma_{v_t}^2\}$

En la práctica el pronosticador puede no establecer valores precisos para la sucesión de varianzas observacionales $\{\sigma_{v_t}^2\}$, pero puede tener una opinión acerca de su forma. Así

pueden considerarse los siguientes tres casos, como los más prácticos:

- (i) $\sigma_{v_t}^2 = \sigma_v^2$ para toda t , con σ_v^2 desconocida.
- (ii) $\sigma_{v_t}^2 = \sigma_v^2 k_t(\theta_t)$, donde $k_t(\theta_t)$ es una función de la media.
- (iii) $\sigma_{v_t}^2 = \sigma_v^2 k_t$, con la sucesión de pesos $\{k_t\}$ desconocida.

A continuación se definirá el concepto de precisión, que puede hacer más claro el desarrollo de las ideas; a partir de este concepto se discutirá el caso (i), por ser el más sencillo.

Definición 3.6

En el modelo polinomial de primer orden la *precisión* ϕ_t , se define como el recíproco de la varianza observacional: $\phi_t \equiv \frac{1}{\sigma_{v_t}^2}$.

Cuando la varianza σ_v^2 es un parámetro desconocido, para que la rutina de pronóstico proporcionada por el *teorema 3.1* permita realizar pronósticos bajo esta condición se deben hacer las siguientes modificaciones al modelo de la *definición 3.1*:

- (I) Se considera que $\sigma_{v_t}^2 = \sigma_v^2$ y la precisión $\phi_t = \phi$ se considera como una cantidad aleatoria Φ , y como tal se le asocia una distribución de probabilidad $F_\phi(\phi)$. Se escoge una distribución *Gamma* ya que por permitir simular otras distribuciones, haciendo combinaciones en sus parámetros. Como la distribución *Gamma* es a priori conjugada, entonces el conocimiento sobre Φ puede actualizarse. En consecuencia la información inicial debe contener una distribución a priori inicial

$$(\Phi|D_0) \sim G\left[\frac{n_0}{2}, \frac{d_0}{2}\right]$$

para algunos $n_0, d_0 > 0$. Sea $S_{v_0}^2 \equiv \frac{d_0}{n_0}$, de las propiedades de la distribución *Gamma*

se sabe que

$$E[\Phi|D_0] = \frac{n_0}{d_0} = \frac{1}{S_{v_0}^2}$$

entonces S_{ν}^2 es el valor estimado a priori para σ_{ν}^2 o equivalentemente $\frac{1}{S_{\nu}^2}$ es el valor

esperado para Φ .

- (2) Para facilitar los cálculos, se considera que para toda t , la varianza de transición $\sigma_{w_t}^2$ es un múltiplo de σ_{ν}^2

$$\sigma_{w_t}^2 = \sigma_{\nu}^2 \sigma_{w_t}^{2*}$$

y similarmente

$$S_0^2 = \sigma_{\nu}^2 S_0^{2*}$$

siendo S_0^{2*} y la sucesión $\{\sigma_{w_t}^{2*}\}$ conocidas e independientes de la escala de las observaciones determinada por σ_{ν}^2 y son tales que al multiplicarlas por σ_{ν}^2 se obtienen las verdaderas varianzas S_0^2 y $\{\sigma_{w_t}^2\}$ cuando σ_{ν}^2 es conocida.

- (3) Se supone que las sucesiones de errores $\{\varepsilon_{\nu_t}\}$ y $\{\varepsilon_{w_t}\}$ son independientes y mutuamente independientes, pero *condicionadas* sobre σ_{ν}^2 .

Teorema 3.4

Con el modelo modificado, si $(\Theta_0 | D_0) \sim N[m_0, S_0^2]$ y $(\Phi | D_0) \sim G[n_0/2, d_0/2]$, para cada

$t \geq 1$ se tienen las distribuciones

(I) Condicionales sobre $\Phi = \frac{1}{\sigma_{\nu}^2}$:

(a) A priori para Θ_t

$$\left(\Theta_t | D_{t-1}, \Phi = \frac{1}{\sigma_{\nu}^2} \right) \sim N[\bar{m}_t, \sigma_{\nu}^2 \bar{S}_t^{2*}]$$

donde

$$\bar{m}_t = m_{t-1} + \mu_{w_t}, \quad \bar{S}_t^{2*} = S_{t-1}^{2*} + \sigma_{w_t}^{2*}$$

(b) Pronóstico de un periodo

$$\left(f_{t-1}(1) \middle| \Phi = \frac{1}{\sigma_{\nu}^2} \right) \sim N[\hat{Y}_t, \sigma_{\nu}^2 \hat{\sigma}_t^{2*}]$$

donde

$$\hat{y}_i = \bar{m}_i + \mu_{\nu_i}, \quad \hat{\sigma}_i^2 = 1 + \bar{S}_i^2$$

(c) A posteriori para Θ_i

$$\left(\Theta_i \mid D_i, \Phi = \frac{1}{\sigma_{\nu_i}^2} \right) \sim N \left[m_i, \sigma_{\nu_i}^2 S_i^2 \right],$$

con

$$m_i = \bar{m}_i + A_i e_i, \quad S_i^2 = \bar{S}_i^2 - A_i^2 \hat{\sigma}_i^2 = A_i$$

Donde

$$A_i = \frac{\bar{S}_i^2}{\hat{\sigma}_i^2}, \quad e_i = y_i - \hat{y}_i$$

(II) Para la precisión Φ :

(a) Distribución a priori

$$(\Phi \mid D_{i-1}) \sim G \left[\frac{n_{i-1}}{2}, \frac{d_{i-1}}{2} \right]$$

(b) Distribución a posteriori

$$(\Phi \mid D_i) \sim G \left[\frac{n_i}{2}, \frac{d_i}{2} \right],$$

con

$$n_i = n_{i-1} + 1, \quad d_i = d_{i-1} + \frac{e_i^2}{\hat{\sigma}_i^2}$$

(III) Incondicionales sobre $\Phi = \frac{1}{\sigma_{\nu_i}^2}$:

(a) $(\Theta_i \mid D_{i-1}) \sim T_{n_{i-1}} \left[\bar{m}_i, \bar{S}_i^2 \right]$.

(b) $f_{i-1}(1) \sim T_{n_{i-1}} \left[\hat{y}_i, \hat{\sigma}_i^2 \right]$.

(c) $(\Theta_i \mid D_i) \sim T_{n_i} \left[m_i, S_i^2 \right]$.

Con

$$\begin{aligned} S_{i-1}^2 &= S_{\hat{y}_{i-1}}^2 S_{i-1}^{2*}, & \bar{S}_i^2 &= S_{\hat{y}_{i-1}}^2 \bar{S}_i^{2*}, \\ \hat{\sigma}_i^2 &= S_{\hat{y}_{i-1}}^2 \hat{\sigma}_i^{2*}, & S_i^2 &= S_{\hat{y}_{i-1}}^2 S_i^{2*}. \end{aligned}$$

donde

$$S_{\hat{y}_{i-1}}^2 = \frac{d_i}{n_i}$$

Demostración.

La sección (I) es inmediata del teorema 3.1.

La sección (II) se demuestra por inducción.

Supóngase que se cumple (a), es decir

$$(\phi | D_{i-1}) \sim G\left[\frac{n_{i-1}}{2}, \frac{d_{i-1}}{2}\right],$$

siendo la densidad a priori

$$f_\phi(\phi) = \frac{\left(\frac{d_{i-1}}{2}\right)^{\frac{n_{i-1}}{2}} \phi^{\frac{n_{i-1}}{2}-1} \exp\left\{-\phi \frac{d_{i-1}}{2}\right\}}{\Gamma\left(\frac{n_{i-1}}{2}\right)} \propto \phi^{\frac{n_{i-1}}{2}-1} \exp\left\{-\phi \frac{d_{i-1}}{2}\right\}$$

De acuerdo a la sección (I)

$$(f_{i-1}(1) | \sigma_{\hat{y}_i}^2) \sim N\left[\hat{y}_i, \sigma_{\hat{y}_i}^2 \hat{\sigma}_i^{2*}\right],$$

por lo tanto

$$(f_i(1) | \phi = \phi) \sim N\left[\hat{y}_i, \frac{\hat{\sigma}_i^{2*}}{\phi}\right],$$

cuya función de densidad es

$$f_{y_i | \phi, D_{i-1}}(y_i | \phi) \propto \phi^{\frac{1}{2}} \exp\left\{-\frac{\phi (y_i - \hat{y}_i)^2}{2 \hat{\sigma}_i^{2*}}\right\}$$

Por el teorema de Bayes

$$f_{\phi, \nu, D_{t-1}}(\phi | y_t) \propto \phi^t \exp\left\{-\frac{\phi(y_t - \hat{Y}_t)^2}{\hat{\sigma}_t^2}\right\} \phi^{\frac{n_{t-1}}{2}-1} \exp\left\{-\phi \frac{d_{t-1}}{2}\right\}$$

$$\propto \phi^{\frac{n_{t-1}+1}{2}-1} \exp\left\{-\frac{\phi}{2}\left[d_{t-1} + \frac{e_t^2}{\hat{\sigma}_t^2}\right]\right\}$$

de donde se deduce que

$$(\phi | D_t) \sim G\left[\frac{n_t}{2}, \frac{d_t}{2}\right]$$

con

$$n_t = n_{t-1} + 1 \quad \text{y} \quad d_t = d_{t-1} + \frac{e_t^2}{\hat{\sigma}_t^2}$$

Como (a) también se cumple para $t = 1$, queda demostrada la sección (II).

La sección (III) se demuestra directamente de las propiedades de las distribuciones *Normal-Gamma* y *t-Student* que son enunciadas en la sección A.4 del Apéndice.

En el tiempo t la media

$$E[\phi | D_{t-1}] = \frac{n_{t-1}}{d_{t-1}} = \frac{1}{S_{\nu_{t-1}}^2}$$

proporciona una estimación puntual a priori de la precisión $(\sigma_\nu^2)^{-1}$, y de manera similar

$$E[\phi | D_t] = \frac{n_t}{d_t} = \frac{1}{S_\nu^2}$$

es una estimación puntual posterior para $(\sigma_\nu^2)^{-1}$.

De acuerdo al teorema anterior se pueden resumir las siguientes ecuaciones recurrentes de actualización del modelo:

$$m_t = \bar{m}_t + A_t e_t$$

$$S_t^2 = \left(\frac{S_{\nu_t}^{2*}}{S_{\nu_{t-1}}^2} \right) (\bar{S}_t^2 - A_t^2 \hat{\sigma}_t^2) = A_t S_{\nu_t}^2$$

$$S_{\nu_t}^2 = \frac{d_t}{n_t}, \text{ con } \begin{cases} n_t = n_{t-1} + 1, & d_t = d_{t-1} + \frac{S_{\nu_{t-1}}^2 e_t^2}{\hat{\sigma}_t^2}, \\ \hat{\sigma}_t^2 = S_{\nu_{t-1}}^2 + \bar{S}_t^2, & A_t = \frac{\bar{S}_t^2}{\hat{\sigma}_t^2}. \end{cases}$$

Es importante hacer notar que en el Modelo Lineal con varianza σ_v^2 conocida los pronósticos se basan en distribuciones *normales*, mientras que cuando σ_v^2 es desconocida tales distribuciones son reemplazadas por distribuciones *t de Student*¹⁶. En ambos casos las ecuaciones de actualización son esencialmente las mismas, sustituyendo a σ_v^2 en el segundo caso por la estimación $S_{\nu_{t-1}}^2$. Las únicas diferencias aparecen al actualizar a S_t^2, d_t , debido a la actualización de la estimación de σ_v^2 .

Similarmente, para las distribuciones de pronóstico se tiene el siguiente

Teorema 3.5

Con el modelo modificado y dado que $(\Phi|D_t) \sim G\left[\frac{n_t}{2}, \frac{d_t}{2}\right]$, en el tiempo t y para $k \geq j \geq 1$ se tienen las siguientes distribuciones de pronóstico

(I) Condicionales sobre $\Phi = \frac{1}{\sigma_v^2}$:

(a) Pronóstico de k periodos en el futuro

$$\left(f_t(k) \middle| \Phi = \frac{1}{\sigma_v^2} \right) \sim N\left[\hat{Y}_t(k), \sigma_v^2 (\hat{\sigma}_t^*(k)) \right]$$

con

$$\hat{Y}_t(k) = m_t + \sum_{i=1}^k \mu_{w_{t,i}} + \mu_{v_{t,k}}$$

$$\hat{\sigma}_t^*(k) = S_t^{2*} + \sum_{j=1}^k \sigma_{w_{t,j}}^{2*} +$$

¹⁶ En la sección A.3 del Apéndice se muestra la relación entre el comportamiento límite de la distribución *t de Student* y la distribución *Normal*.

(b) Pronóstico guía de horizonte k

$$\left(f_i^H(k) \middle| \Phi = \frac{1}{\sigma_v^2} \right) \sim N \left[m_{H_i}(k), \sigma_v^2 S_{H_i}^{2^*}(k) \right]$$

con

$$m_{H_i}(k) = km_i + \sum_{i=1}^k i \mu_{w_{i,i+1}} + \sum_{i=1}^k \mu_{v_{i,i}}$$

$$S_{H_i}^{2^*}(k) = k^2 S_i^{2^*} + \sum_{i=1}^k i^2 \sigma_{w_{i,i+1}}^{2^*} + k$$

(c) Pronóstico guía de horizonte k diferido $j-1$ periodos

$$\left(f_i^H(j,k) \middle| \Phi = \frac{1}{\sigma_v^2} \right) \sim N \left[m_{H_i}(j,k), \sigma_v^2 S_{H_i}^{2^*}(j,k) \right]$$

con

$$m_{H_i}(j,k) = \sum_{r=1}^k \left[m_i + \sum_{i=1}^r \mu_{w_{i,i+1}} + \mu_{v_{i,i}} \right]$$

$$S_{H_i}^{2^*}(j,k) = (k-j+1) \left[(k-j+1) S_i^{2^*} + \sum_{i=1}^j \sigma_{w_{i,i}}^{2^*} + 1 \right] + \sum_{i=0}^{k-j-1} (k-j-i) \sigma_{w_{i,j+i}}^{2^*}$$

con m_i y $S_i^{2^*}$ como en el teorema 3.4.

(II) Incondicionales sobre $\Phi = \frac{1}{\sigma_v^2}$:

(a) $f_i(k) \sim T_n \left[\hat{Y}_i(k), \hat{\sigma}_i^2(k) \right]$,

(b) $f_i^H(k) \sim T_n \left[m_{H_i}(k), S_{H_i}^2(k) \right]$,

(c) $f_i^H(j,k) \sim T_n \left[m_{H_i}(j,k), S_{H_i}^2(j,k) \right]$,

con

$$\hat{\sigma}_i^2(k) = S_{H_i}^2 \hat{\sigma}_i^2(k),$$

$$S_{H_i}^2(k) = S_{V_i}^2 S_{H_i}^{2*}(k),$$

$$S_{H_i}^2(j, k) = S_{V_i}^2 S_{H_i}^{2*}(j, k)$$

y $S_{V_i}^2$ como en el teorema 3.4.

Demostración.

La sección (I) es inmediata del teorema 3.2.

La sección (II) se demuestra directamente de las propiedades de las distribuciones *Normal-Gamma* y *t-Student* que son enunciadas en la sección A.4 del Apéndice.

Para finalizar el capítulo se dará un ejemplo en el que se aplica un modelo lineal que comprenda los métodos hasta aquí estudiados.

Ejemplo: Pronóstico del tipo de cambio \$/DM

Se considera un problema de pronóstico sobre la tasa de crecimiento del tipo de cambio. En esta ocasión S_t representa el tipo de cambio dólar-marco alemán al cierre del mes t , y Y_t está dada como en (3.15). En la tabla 3.5 están contenidos los datos correspondientes a las tasas de crecimiento registradas entre enero de 1980 y marzo de 1994. Se escoge un ML para describir el comportamiento de la serie $\{Y_t\}$ en dicho intervalo de tiempo.

Se considera que $\sigma_{y_t}^2$ es constante pero desconocida y que la sucesión $\{\sigma_{y_t}^2\}$ se determina por medio de factores de descuento δ , utilizando para este ejemplo $\delta = 1.0, 0.9$ y 0.8 .

De manera similar al ejemplo realizado anteriormente, una solución posible para determinar la información inicial respecto al parámetro Θ es asignar a m_0 el valor de la media muestral,

$$\bar{y} = \sum_{t=1}^{171} \frac{y_t}{171} = 2.0593 \times 10^{-2}$$

y a S_0^2 un porcentaje del valor de $S_{y_0}^2$, (en este ejemplo se toma el 10 %).

A diferencia del ejemplo anterior, en este caso, se debe proveer información inicial acerca de la precisión. Por simplicidad y dado que se cuenta con el total de los datos del periodo en estudio, una solución posible es asignarle a $S_{y_0}^2$ el valor de la varianza muestral

$$\begin{aligned} S_{y_0}^2 &= \hat{\sigma}_y^2 \\ &= \sum_{t=1}^{171} \frac{(y_t - \bar{y})^2}{170} \\ &= 0.2155 \times 10^{-4} \end{aligned}$$

También por simplicidad, el hiperparámetro μ_0 se toma igual a 1 de tal forma que $d_0 = S_{y_0}^2$.

	1980	1983	1986	1989	1992
ENE	1.7438	2.4650	2.3880	1.8785	1.6075
FEB	1.7770	2.4340	2.2308	1.8230	1.6373
MAR	1.9515	2.4310	2.3450	1.8965	1.6465
ABR	1.8000	2.4665	2.1630	1.8802	1.6525
MAY	1.7785	2.5245	2.3250	1.9760	1.6065
JUN	1.7641	2.5410	2.2010	1.9545	1.5240
JUL	1.7805	2.6485	2.0935	1.8658	1.4747
AGO	1.7920	2.6970	2.0355	1.9582	1.4030
SEP	1.8127	2.6315	2.0272	1.8725	1.4121
OCT	1.9058	2.6315	2.0610	1.8413	1.5406
NOV	1.9269	2.6990	1.9730	1.7800	1.5903
DEC	1.9735	2.7250	1.9235	1.6895	1.6212
	1981	1984	1987	1990	1993
ENE	2.1320	2.8165	1.8330	1.6880	1.6109
FEB	2.1310	2.5983	1.8275	1.6965	1.6457
MAR	2.1105	2.5940	1.8020	1.6883	1.6070
ABR	2.2105	2.7215	1.7985	1.6790	1.5855
MAY	2.3300	2.7318	1.8240	1.6990	1.5902
JUN	2.3940	2.7815	1.8262	1.6658	1.7085
JUL	2.4710	2.9063	1.8590	1.5875	1.7420
AGO	2.4475	2.8895	1.8108	1.5760	1.6763
SEP	2.3250	3.0720	1.8428	1.5665	1.6345
OCT	2.2435	3.0305	1.7283	1.5157	1.6880
NOV	2.2138	3.1100	1.6405	1.5000	1.7165
DEC	2.2380	3.1550	1.5708	1.4970	1.7387
	1982	1985	1988	1991	1994
ENE	2.3165	3.1630	1.6785	1.4785	1.7345
FEB	2.3880	3.3440	1.6883	1.5260	1.7045
MAR	2.4175	3.0830	1.6558	1.6975	1.6742
ABR	2.3310	3.0980	1.6746	1.7185	
MAY	2.3475	3.0590	1.7290	1.7397	
JUN	2.4595	3.0290	1.8170	1.8138	
JUL	2.4660	2.7995	1.8750	1.7465	
AGO	2.4970	2.8110	1.8787	1.7485	
SEP	2.5290	2.6790	1.8730	1.6625	
OCT	2.5600	2.6190	1.7820	1.6680	
NOV	2.4650	2.5125	1.7360	1.6270	
DEC	2.3780	2.4470	1.7735	1.5200	

Tabla 3.5. Tasas de crecimiento $\times 100$ del tipo de cambio Dólar-Marco Alemán.

Fuente Bloomberg Market.

De igual manera que en el primer ejemplo de este capítulo, se toma $\mu_{v_t} = 0$ y $\mu_{w_t} = 0$ para toda t .

A continuación se mostrará el desarrollo de la rutina de aprendizaje proporcionada por el modelo lineal con varianza constante pero desconocida entre los tiempos $t = 0$ y $t = \dots$. Nuevamente la sucesión $\{\sigma_{w_t}^2\}$ se escoge con el factor de descuento $\delta = 0.8$.

Información Inicial

$$D_0 = \left\{ \begin{array}{ll} \mu_v = 0 & S_{v_0}^2 = 0.2155 \times 10^{-4} \\ \mu_w = 0 & \sigma_{w_t}^2 = S_{t-1}^2 \left(\frac{1-\delta}{\delta} \right), \quad \delta = 0.8 \\ m_0 = 2.0593 \times 10^{-2} & S_0^2 = 0.0215 \times 10^{-4} \\ n_0 = 1 & d_0 = 0.2155 \times 10^{-4} \end{array} \right.$$

De acuerdo a la información contenida en D_0 , se tienen las siguientes distribuciones iniciales

$$(\theta_0 | D_0) \sim T_1 [2.0593 \times 10^{-2}, 0.0215 \times 10^{-4}],$$

$$(\phi_0 | D_0) \sim G \left[\frac{1}{2}, \frac{0.2155 \times 10^{-4}}{2} \right]$$

Transición

Como la elección de $\{\sigma_{w_t}^2\}$ es por factor de descuento entonces

$$\begin{aligned} \sigma_{w_t}^2 &= S_0^2 \left(\frac{1-0.8}{0.8} \right) \\ &= (0.0215 \times 10^{-4}) \times (0.25) \\ &= 0.0054 \times 10^{-4} \end{aligned}$$

Los momentos de la distribución a priori de θ , se obtienen a través de las siguientes operaciones

$$\begin{aligned} \bar{m}_t &= m_0 + \mu_{v_0} \\ &= 2.0593 \times 10^{-2} + 0 = 2.0593 \times 10^{-2}, \end{aligned}$$

	1980	1983	1986	1989	1992
ENE	1.7438	2.4650	2.3880	1.8785	1.6075
FEB	1.7770	2.4340	2.2308	1.8230	1.6373
MAR	1.9515	2.4310	2.3450	1.8965	1.6465
ABR	1.8000	2.4665	2.1650	1.8802	1.6525
MAY	1.7785	2.5245	2.3250	1.9760	1.6065
JUN	1.7641	2.5410	2.2010	1.9545	1.5240
JUL	1.7805	2.6485	2.0935	1.8658	1.4747
AGO	1.7920	2.6970	2.0355	1.9582	1.4030
SEP	1.8127	2.6315	2.0272	1.8725	1.4121
OCT	1.9058	2.6315	2.0610	1.8413	1.5406
NOV	1.9269	2.6990	1.9730	1.7800	1.5903
DEC	1.9735	2.7250	1.9235	1.6895	1.6212
	1981	1984	1987	1990	1993
ENE	2.1320	2.8165	1.8330	1.6880	1.6109
FEB	2.1310	2.5983	1.8275	1.6965	1.6457
MAR	2.1105	2.5940	1.8020	1.6883	1.6070
ABR	2.2105	2.7215	1.7985	1.6790	1.5855
MAY	2.3300	2.7318	1.8240	1.6990	1.5902
JUN	2.3940	2.7815	1.8262	1.6658	1.7085
JUL	2.4710	2.9063	1.8590	1.5875	1.7420
AGO	2.4475	2.8895	1.8108	1.5760	1.6763
SEP	2.3250	3.0720	1.8428	1.5665	1.6345
OCT	2.2435	3.0305	1.7283	1.5157	1.6880
NOV	2.2138	3.1100	1.6405	1.5000	1.7165
DEC	2.2380	3.1550	1.5708	1.4970	1.7387
	1982	1985	1988	1991	1994
ENE	2.3165	3.1630	1.6785	1.4785	1.7345
FEB	2.3880	3.3440	1.6883	1.5260	1.7045
MAR	2.4175	3.0830	1.6358	1.6975	1.6742
ABR	2.3310	3.0980	1.6746	1.7185	
MAY	2.3475	3.0590	1.7290	1.7397	
JUN	2.4595	3.0290	1.8170	1.8138	
JUL	2.4660	2.7995	1.8750	1.7465	
AGO	2.4970	2.8110	1.8787	1.7485	
SEP	2.5290	2.6790	1.8730	1.6625	
OCT	2.5600	2.6190	1.7820	1.6680	
NOV	2.4650	2.5125	1.7360	1.6270	
DEC	2.3780	2.4470	1.7735	1.5200	

Tabla 3.5. Tasas de crecimiento $\times 100$ del tipo de cambio Dólar-Marco Alemán.

Fuente Bloomberg Market.

De igual manera que en el primer ejemplo de este capítulo, se toma $\mu_v = 0$ y $\mu_w = 0$ para toda t .

A continuación se mostrará el desarrollo de la rutina de aprendizaje proporcionada por el modelo lineal con varianza constante pero desconocida entre los tiempos $t=0$ y $t=$. Nuevamente la sucesión $\{\sigma_w^2\}$ se escoge con el factor de descuento $\delta = 0.8$.

Información Inicial

$$D_0 = \left\{ \begin{array}{ll} \mu_v = 0 & S_v^2 = 0.2155 \times 10^{-4} \\ \mu_w = 0 & \sigma_w^2 = S_{t-1}^2 \left(\frac{1-\delta}{\delta} \right), \delta = 0.8 \\ m_0 = 2.0593 \times 10^{-2} & S_0^2 = 0.0215 \times 10^{-4} \\ n_0 = 1 & d_0 = 0.2155 \times 10^{-4} \end{array} \right\}$$

De acuerdo a la información contenida en D_0 , se tienen las siguientes distribuciones iniciales

$$(\Theta_0 | D_0) \sim T_1 [2.0593 \times 10^{-2}, 0.0215 \times 10^{-4}]$$

$$(\Phi_0 | D_0) \sim G \left[\frac{1}{2}, \frac{0.2155 \times 10^{-4}}{2} \right]$$

Transición

Como la elección de $\{\sigma_w^2\}$ es por factor de descuento entonces

$$\begin{aligned} \sigma_w^2 &= S_0^2 \left(\frac{1-0.8}{0.8} \right) \\ &= (0.0215 \times 10^{-4}) \times (0.25) \\ &= 0.0054 \times 10^{-4} \end{aligned}$$

Los momentos de la distribución a priori de Θ_1 se obtienen a través de las siguientes operaciones

$$\begin{aligned} \bar{m}_1 &= m_0 + \mu_v \\ &= 2.0593 \times 10^{-2} + 0 = 2.0593 \times 10^{-2}, \end{aligned}$$

$$\begin{aligned}\bar{S}_1^2 &= S_0^2 + \sigma_{w_1}^2 \\ &= 0.0215 \times 10^{-4} + 0.0054 \times 10^{-4} \\ &= 0.0269 \times 10^{-4}\end{aligned}$$

Por lo tanto la distribución a priori para Θ_1 es

$$(\Theta_1 | D_0) \sim T_1[2.0593 \times 10^{-2}, 0.0269 \times 10^{-4}]$$

Pronóstico

Los momentos de la distribución del pronóstico $f_0(1)$, se obtienen de la siguiente manera

$$\begin{aligned}\hat{Y}_1 &= \bar{m}_1 + \mu_{v_1} \\ &= 2.0593 \times 10^{-2} + 0 \\ &= 2.0593 \times 10^{-2}\end{aligned}$$

$$\begin{aligned}\hat{\sigma}_1^2 &= \bar{S}_1^2 + S_{v_0}^2 \\ &= 0.0269 \times 10^{-4} + 0.2155 \times 10^{-4} \\ &= 0.2424 \times 10^{-4}\end{aligned}$$

Entonces la distribución de pronóstico es

$$(Y_1 | D_0) \sim T_1[2.0593 \times 10^{-2}, 0.2424 \times 10^{-4}]$$

Actualización

Una vez que se observa $y_1 = 1.7438 \times 10^{-2}$, este dato se incorpora al conjunto de información disponible

$$D_1 = \{y_1 = 1.7438 \times 10^{-2}, D_0\}$$

El error de pronóstico y el coeficiente adaptativo están dados por

$$\begin{aligned}e_1 &= y_1 - \hat{Y}_1 \\ &= (1.7438 \times 10^{-2} - 2.0593 \times 10^{-2}), \\ &= -0.3155 \times 10^{-2}\end{aligned}$$

$$A_1 = \frac{\bar{S}_1^2}{\hat{\sigma}_1^2} = \frac{0.0269 \times 10^{-4}}{0.2424 \times 10^{-4}} = 0.1111$$

Utilizando este nuevo dato se actualiza la información acerca de la precisión Φ

$$n_1 = n_0 + 1 = 1 + 1 = 2$$

$$\begin{aligned} d_1 &= d_0 + \frac{S_{\nu_0}^2 e_1^2}{\hat{\sigma}_1^2} \\ &= 0.2155 \times 10^{-4} + \frac{(0.2155 \times 10^{-4}) \times (0.0995 \times 10^{-4})}{0.2424 \times 10^{-4}} \\ &= 0.3040 \times 10^{-4} \end{aligned}$$

Por lo tanto la estimación puntual para la varianza observacional es

$$S_{\nu_1}^2 = \frac{d_1}{n_1} = \frac{0.3040 \times 10^{-4}}{2} = 0.1520 \times 10^{-4}$$

Finalmente se procede a actualizar el conocimiento que se tiene sobre Θ_1

$$\begin{aligned} m_1 &= \bar{m}_1 + A_1 e_1 \\ &= 2.0593 \times 10^{-2} + (0.1111) \times (-0.3155 \times 10^{-2}) \\ &= 2.0243 \times 10^{-2} \end{aligned}$$

$$\begin{aligned} S_1^2 &= A_1 S_{\nu_1}^2 \\ &= (0.1111) \times (0.1520 \times 10^{-4}) \\ &= 0.0169 \times 10^{-4} \end{aligned}$$

De esta manera las distribuciones posteriores para Θ_1 y Φ son

$$(\Theta_1 | D_1) \sim T_2 [2.0243 \times 10^{-2}, 0.0169 \times 10^{-4}]$$

$$(\Phi | D_1) \sim G \left[\frac{2}{2}, \frac{0.3040 \times 10^{-4}}{2} \right]$$

Este procedimiento es muy similar al expuesto en el ejemplo anterior, sólo que se agregan ciertas operaciones para ajustar el valor de la estimación de la varianza observacional $S_{\nu_1}^2$. Las operaciones relativas a las modificaciones del conocimiento sobre Θ_1 son esencialmente las mismas que en el ejemplo anterior, pero utilizando $S_{\nu_1}^2$ en lugar de $\sigma_{\nu_1}^2$.

Repitiendo esta rutina se obtienen los datos contenidos en la tabla 3.6. Ahí se muestran las medias y varianzas de las distribuciones de pronóstico $f_{t-1}(1)$ para todos los meses entre enero de 1980 y marzo de 1994. Nuevamente se utilizaron tres diferentes factores de descuento 0.8, 0.9 y 1.0.

Como se conservo el supuesto de que μ_w y μ_v son iguales a cero para toda t , entonces se cumple que

$$\hat{Y}_t = \bar{m}_t = m_{t-1}$$

por lo que en la tabla 3.6 se pueden encontrar los valores de los tres predictores puntuales.

<i>t</i>	$y_t \times 100$	$\delta = 1.0$		$\delta = 0.9$		$\delta = 0.8$	
		$\hat{y}_t \times 100$	$\hat{\sigma}_t^2 \times 10,000$	$\hat{y}_t \times 100$	$\hat{\sigma}_t^2 \times 10,000$	$\hat{y}_t \times 100$	$\hat{\sigma}_t^2 \times 10,000$
80							
<i>E</i>	1.7438	2.0593	0.2370	2.0593	0.2394	2.0593	0.2424
<i>F</i>	1.7770	2.0306	0.1669	2.0278	0.1695	2.0243	0.1731
<i>M</i>	1.9515	2.0095	0.1318	2.0027	0.1340	1.9941	0.1374
<i>A</i>	1.8000	2.0050	0.0991	1.9976	0.1011	1.9885	0.1047
<i>M</i>	1.7785	1.9904	0.0872	1.9778	0.0887	1.9617	0.0918
<i>J</i>	1.7641	1.9763	0.0798	1.9579	0.0805	1.9341	0.0828
<i>J</i>	1.7805	1.9630	0.0746	1.9385	0.0744	1.9072	0.0757
<i>A</i>	1.7920	1.9523	0.0692	1.9227	0.0682	1.8862	0.0688
<i>S</i>	1.8127	1.9434	0.0641	1.9096	0.0625	1.8701	0.0625
<i>O</i>	1.9058	1.9365	0.0593	1.8999	0.0572	1.8600	0.0569
<i>N</i>	1.9269	1.9349	0.0538	1.9005	0.0520	1.8683	0.0521
<i>D</i>	1.9735	1.9346	0.0492	1.9032	0.0477	1.8791	0.0482
81							
<i>E</i>	2.1320	1.9363	0.0455	1.9102	0.0445	1.8967	0.0454
<i>F</i>	2.1310	1.9448	0.0449	1.9324	0.0448	1.9413	0.0462
<i>M</i>	2.1105	1.9526	0.0441	1.9522	0.0444	1.9777	0.0456
<i>A</i>	2.2105	1.9589	0.0428	1.9681	0.0432	2.0033	0.0439
<i>M</i>	2.3300	1.9686	0.0440	1.9923	0.0441	2.0436	0.0439
<i>J</i>	2.3940	1.9820	0.0487	2.0261	0.0480	2.0996	0.0461
<i>J</i>	2.4710	1.9967	0.0550	2.0629	0.0526	2.1575	0.0483
<i>A</i>	2.4475	2.0130	0.0634	2.1037	0.0583	2.2193	0.0508
<i>S</i>	2.3250	2.0275	0.0693	2.1381	0.0612	2.2644	0.0509
<i>O</i>	2.2435	2.0371	0.0701	2.1568	0.0600	2.2764	0.0488
<i>N</i>	2.2138	2.0436	0.0689	2.1654	0.0577	2.2699	0.0467
<i>D</i>	2.2380	2.0487	0.0671	2.1703	0.0554	2.2587	0.0449
82							
<i>E</i>	2.3165	2.0543	0.0658	2.1770	0.0534	2.2546	0.0431
<i>F</i>	2.3880	2.0618	0.0659	2.1910	0.0521	2.2669	0.0416
<i>M</i>	2.4175	2.0709	0.0673	2.2107	0.0516	2.2911	0.0406
<i>A</i>	2.3310	2.0802	0.0692	2.2314	0.0512	2.3163	0.0398
<i>M</i>	2.3475	2.0868	0.0689	2.2413	0.0498	2.3192	0.0384
<i>J</i>	2.4595	2.0935	0.0688	2.2519	0.0485	2.3249	0.0372
<i>J</i>	2.4660	2.1027	0.0709	2.2727	0.0484	2.3518	0.0365
<i>A</i>	2.4970	2.1115	0.0728	2.2920	0.0480	2.3746	0.0358
<i>S</i>	2.5290	2.1207	0.0750	2.3125	0.0478	2.3991	0.0352
<i>O</i>	2.5600	2.1302	0.0777	2.3342	0.0478	2.4250	0.0346
<i>N</i>	2.4650	2.1400	0.0807	2.3568	0.0479	2.4520	0.0342
<i>D</i>	2.3780	2.1472	0.0813	2.3676	0.0469	2.4546	0.0332

Tabla 3.6 Sección I

<i>t</i>	$y_t \times 100$	$\delta = 1.0$		$\delta = 0.9$		$\delta = 0.8$	
		$\hat{Y}_t \times 100$	$\hat{\sigma}_t^2$ <small>$\times 10,000$</small>	$\hat{Y}_t \times 100$	$\hat{\sigma}_t^2$ <small>$\times 10,000$</small>	$\hat{Y}_t \times 100$	$\hat{\sigma}_t^2$ <small>$\times 10,000$</small>
83							
<i>E</i>	2.4650	2.1522	0.0805	2.3686	0.0456	2.4393	0.0325
<i>F</i>	2.4340	2.1589	0.0810	2.3783	0.0447	2.4444	0.0317
<i>M</i>	2.4310	2.1646	0.0808	2.3838	0.0436	2.4423	0.0308
<i>A</i>	2.4665	2.1700	0.0805	2.3886	0.0426	2.4401	0.0301
<i>M</i>	2.5245	2.1760	0.0807	2.3963	0.0417	2.4454	0.0294
<i>J</i>	2.5410	2.1828	0.0816	2.4092	0.0411	2.4612	0.0288
<i>J</i>	2.6485	2.1897	0.0827	2.4223	0.0405	2.4771	0.0283
<i>A</i>	2.6970	2.1983	0.0855	2.4450	0.0408	2.5114	0.0283
<i>S</i>	2.6315	2.2076	0.0891	2.4702	0.0413	2.5485	0.0284
<i>O</i>	2.6315	2.2153	0.0911	2.4863	0.0409	2.5651	0.0280
<i>N</i>	2.6990	2.2227	0.0928	2.5008	0.0405	2.5784	0.0275
<i>D</i>	2.7250	2.2311	0.0956	2.5206	0.0405	2.6025	0.0272
84							
<i>E</i>	2.8165	2.2396	0.0986	2.5411	0.0405	2.6270	0.0270
<i>F</i>	2.5983	2.2494	0.1032	2.5686	0.0412	2.6649	0.0271
<i>M</i>	2.5940	2.2552	0.1035	2.5716	0.0404	2.6516	0.0267
<i>A</i>	2.7215	2.2607	0.1037	2.5738	0.0397	2.6401	0.0262
<i>M</i>	2.7318	2.2682	0.1058	2.5886	0.0393	2.6564	0.0259
<i>J</i>	2.7815	2.2755	0.1077	2.6029	0.0390	2.6714	0.0255
<i>J</i>	2.9063	2.2834	0.1104	2.6208	0.0389	2.6935	0.0253
<i>A</i>	2.8895	2.2930	0.1153	2.6493	0.0396	2.7360	0.0256
<i>S</i>	3.0720	2.3020	0.1195	2.6733	0.0399	2.7667	0.0256
<i>O</i>	3.0305	2.3135	0.1277	2.7132	0.0420	2.8278	0.0267
<i>N</i>	3.1100	2.3241	0.1342	2.7449	0.0430	2.8683	0.0270
<i>D</i>	3.1550	2.3355	0.1422	2.7814	0.0445	2.9167	0.0275
85							
<i>E</i>	3.1630	2.3472	0.1509	2.8188	0.0460	2.9643	0.0280
<i>F</i>	3.3440	2.3587	0.1591	2.8532	0.0472	3.0041	0.0282
<i>M</i>	3.0830	2.3724	0.1720	2.9023	0.0503	3.0720	0.0296
<i>A</i>	3.0980	2.3821	0.1772	2.9204	0.0500	3.0742	0.0291
<i>M</i>	3.0590	2.3918	0.1823	2.9381	0.0497	3.0790	0.0287
<i>J</i>	3.0290	2.4007	0.1862	2.9502	0.0492	3.0750	0.0282
<i>J</i>	2.7995	2.4089	0.1893	2.9581	0.0486	3.0658	0.0278
<i>A</i>	2.8110	2.4140	0.1887	2.9422	0.0482	3.0125	0.0285
<i>S</i>	2.6790	2.4191	0.1883	2.9291	0.0478	2.9722	0.0287
<i>O</i>	2.6190	2.4224	0.1865	2.9041	0.0480	2.9136	0.0295
<i>N</i>	2.5125	2.4248	0.1844	2.8756	0.0484	2.8547	0.0303
<i>D</i>	2.4470	2.4259	0.1819	2.8393	0.0496	2.7862	0.0315

Tabla 3.6 Sección II

<i>t</i>	$y_t \times 100$	$\delta = 1.0$		$\delta = 0.9$		$\delta = 0.8$	
		$\hat{y}_t \times 100$	$\hat{\sigma}_t^2 \times 10,000$	$\hat{y}_t \times 100$	$\hat{\sigma}_t^2 \times 10,000$	$\hat{y}_t \times 100$	$\hat{\sigma}_t^2 \times 10,000$
86							
<i>E</i>	2.3880	2.4262	0.1794	2.8001	0.0510	2.7184	0.0326
<i>F</i>	2.2308	2.4257	0.1770	2.7588	0.0526	2.6523	0.0337
<i>M</i>	2.3450	2.4234	0.1751	2.7060	0.0556	2.5680	0.0356
<i>A</i>	2.1650	2.4225	0.1728	2.6699	0.0566	2.5234	0.0358
<i>M</i>	2.3250	2.4195	0.1714	2.6194	0.0592	2.4517	0.0370
<i>J</i>	2.2010	2.4184	0.1693	2.5900	0.0596	2.4264	0.0367
<i>J</i>	2.0935	2.4159	0.1678	2.5511	0.0607	2.3813	0.0369
<i>A</i>	2.0355	2.4123	0.1669	2.5053	0.0626	2.3237	0.0375
<i>S</i>	2.0272	2.4081	0.1666	2.4584	0.0645	2.2661	0.0380
<i>O</i>	2.0610	2.4039	0.1663	2.4152	0.0660	2.2183	0.0383
<i>N</i>	1.9730	2.4002	0.1657	2.3798	0.0667	2.1869	0.0381
<i>D</i>	1.9235	2.3956	0.1659	2.3391	0.0679	2.1441	0.0382
87							
<i>E</i>	1.8330	2.3906	0.1666	2.2976	0.0691	2.1000	0.0383
<i>F</i>	1.8275	2.3847	0.1682	2.2511	0.0708	2.0466	0.0387
<i>M</i>	1.8020	2.3789	0.1698	2.2088	0.0721	2.0028	0.0388
<i>A</i>	1.7985	2.3730	0.1717	2.1681	0.0732	1.9626	0.0388
<i>M</i>	1.8240	2.3671	0.1734	2.1311	0.0739	1.9298	0.0387
<i>J</i>	1.8262	2.3616	0.1748	2.1004	0.0741	1.9086	0.0384
<i>J</i>	1.8590	2.3563	0.1760	2.0730	0.0741	1.8921	0.0380
<i>A</i>	1.8108	2.3513	0.1767	2.0516	0.0738	1.8855	0.0376
<i>S</i>	1.8428	2.3460	0.1780	2.0275	0.0736	1.8706	0.0373
<i>O</i>	1.7283	2.3412	0.1787	2.0090	0.0732	1.8650	0.0369
<i>N</i>	1.6405	2.3353	0.1808	1.9810	0.0733	1.8377	0.0367
<i>D</i>	1.5708	2.3286	0.1839	1.9469	0.0737	1.7982	0.0367
88							
<i>E</i>	1.6785	2.3215	0.1879	1.9093	0.0744	1.7528	0.0369
<i>F</i>	1.6883	2.3155	0.1902	1.8862	0.0742	1.7379	0.0366
<i>M</i>	1.6558	2.3097	0.1923	1.8664	0.0738	1.7280	0.0362
<i>A</i>	1.6746	2.3037	0.1946	1.8454	0.0735	1.7135	0.0359
<i>M</i>	1.7290	2.2980	0.1966	1.8283	0.0731	1.7058	0.0356
<i>J</i>	1.8170	2.2928	0.1978	1.8184	0.0725	1.7104	0.0352
<i>J</i>	1.8750	2.2886	0.1981	1.8182	0.0718	1.7317	0.0350
<i>A</i>	1.8787	2.2849	0.1978	1.8239	0.0711	1.7604	0.0349
<i>S</i>	1.8730	2.2814	0.1975	1.8294	0.0705	1.7840	0.0347
<i>O</i>	1.7820	2.2778	0.1972	1.8337	0.0698	1.8018	0.0344
<i>N</i>	1.7360	2.2735	0.1976	1.8286	0.0692	1.7979	0.0341
<i>D</i>	1.7735	2.2689	0.1984	1.8193	0.0686	1.7855	0.0338

Tabla 3.6 Sección III

<i>t</i>	$y_t \times 100$	$\delta = 1.0$		$\delta = 0.9$		$\delta = 0.8$	
		$\hat{y}_t \times 100$	$\hat{\sigma}_t^2 \times 10,000$	$\hat{y}_t \times 100$	$\hat{\sigma}_t^2 \times 10,000$	$\hat{y}_t \times 100$	$\hat{\sigma}_t^2 \times 10,000$
89							
<i>E</i>	1.8785	2.2647	0.1988	1.8147	0.0680	1.7831	0.0335
<i>F</i>	1.8230	2.2615	0.1984	1.8211	0.0674	1.8022	0.0333
<i>M</i>	1.8965	2.2578	0.1983	1.8213	0.0668	1.8063	0.0330
<i>A</i>	1.8802	2.2549	0.1977	1.8288	0.0663	1.8244	0.0328
<i>M</i>	1.9760	2.2518	0.1972	1.8340	0.0657	1.8355	0.0325
<i>J</i>	1.9545	2.2495	0.1961	1.8482	0.0653	1.8636	0.0324
<i>J</i>	1.8658	2.2472	0.1951	1.8588	0.0649	1.8818	0.0322
<i>A</i>	1.9582	2.2441	0.1947	1.8595	0.0643	1.8786	0.0319
<i>S</i>	1.8725	2.2418	0.1937	1.8694	0.0638	1.8945	0.0317
<i>O</i>	1.8413	2.2389	0.1932	1.8697	0.0633	1.8901	0.0314
<i>N</i>	1.7800	2.2358	0.1929	1.8668	0.0628	1.8804	0.0312
<i>D</i>	1.6895	2.2323	0.1930	1.8582	0.0623	1.8603	0.0310
90							
<i>E</i>	1.6880	2.2281	0.1938	1.8413	0.0620	1.8261	0.0310
<i>F</i>	1.6965	2.2240	0.1946	1.8260	0.0617	1.7985	0.0309
<i>M</i>	1.6883	2.2200	0.1953	1.8130	0.0613	1.7781	0.0307
<i>A</i>	1.6790	2.2160	0.1960	1.8005	0.0610	1.7601	0.0305
<i>M</i>	1.6990	2.2120	0.1967	1.7884	0.0606	1.7439	0.0303
<i>J</i>	1.6658	2.2082	0.1972	1.7795	0.0602	1.7349	0.0301
<i>J</i>	1.5875	2.2042	0.1980	1.7681	0.0598	1.7211	0.0299
<i>A</i>	1.5760	2.1997	0.1994	1.7500	0.0596	1.6944	0.0298
<i>S</i>	1.5665	2.1952	0.2009	1.7326	0.0594	1.6707	0.0297
<i>O</i>	1.5157	2.1907	0.2024	1.7160	0.0591	1.6499	0.0296
<i>N</i>	1.5000	2.1858	0.2043	1.6960	0.0590	1.6230	0.0295
<i>D</i>	1.4970	2.1810	0.2063	1.6764	0.0588	1.5984	0.0294
91							
<i>E</i>	1.4785	2.1762	0.2082	1.6584	0.0586	1.5781	0.0292
<i>F</i>	1.5260	2.1713	0.2103	1.6405	0.0584	1.5582	0.0291
<i>M</i>	1.6975	2.1668	0.2118	1.6290	0.0581	1.5518	0.0289
<i>A</i>	1.7185	2.1636	0.2119	1.6359	0.0577	1.5809	0.0288
<i>M</i>	1.7397	2.1605	0.2118	1.6441	0.0573	1.6084	0.0287
<i>J</i>	1.8138	2.1577	0.2115	1.6537	0.0570	1.6347	0.0287
<i>J</i>	1.7465	2.1553	0.2108	1.6697	0.0568	1.6705	0.0287
<i>A</i>	1.7485	2.1526	0.2105	1.6774	0.0564	1.6857	0.0285
<i>S</i>	1.6625	2.1499	0.2102	1.6845	0.0560	1.6983	0.0283
<i>O</i>	1.6680	2.1467	0.2103	1.6823	0.0556	1.6911	0.0281
<i>N</i>	1.6270	2.1435	0.2105	1.6809	0.0553	1.6865	0.0280
<i>D</i>	1.5200	2.1401	0.2108	1.6755	0.0549	1.6746	0.0278

Tabla 3.6 Sección IV

<i>t</i>	$y_t \cdot 100$	$\delta = 1.0$		$\delta = 0.9$		$\delta = 0.8$	
		$\hat{y}_t \cdot 100$	$\hat{\sigma}_t^2$ ×10,000	$\hat{y}_t \cdot 100$	$\hat{\sigma}_t^2$ ×10,000	$\hat{y}_t \cdot 100$	$\hat{\sigma}_t^2$ ×10,000
92							
<i>E</i>	1.6075	2.1361	0.2120	1.6599	0.0547	1.6437	0.0278
<i>F</i>	1.6373	2.1327	0.2125	1.6547	0.0543	1.6364	0.0276
<i>M</i>	1.6465	2.1295	0.2127	1.6529	0.0540	1.6366	0.0274
<i>A</i>	1.6525	2.1265	0.2128	1.6523	0.0536	1.6386	0.0272
<i>M</i>	1.6065	2.1235	0.2129	1.6523	0.0532	1.6414	0.0270
<i>J</i>	1.5240	2.1202	0.2133	1.6477	0.0529	1.6344	0.0269
<i>J</i>	1.4747	2.1165	0.2142	1.6354	0.0526	1.6123	0.0268
<i>A</i>	1.4030	2.1125	0.2155	1.6193	0.0525	1.5848	0.0267
<i>S</i>	1.4121	2.1081	0.2174	1.5977	0.0524	1.5484	0.0267
<i>O</i>	1.5406	2.1038	0.2191	1.5791	0.0523	1.5212	0.0267
<i>N</i>	1.5903	2.1004	0.2197	1.5753	0.0520	1.5251	0.0265
<i>D</i>	1.6212	2.0973	0.2200	1.5768	0.0517	1.5381	0.0264
93							
<i>E</i>	1.6109	2.0944	0.2200	1.5812	0.0513	1.5547	0.0263
<i>F</i>	1.6457	2.0915	0.2201	1.5842	0.0510	1.5660	0.0261
<i>M</i>	1.6070	2.0889	0.2199	1.5903	0.0507	1.5819	0.0260
<i>A</i>	1.5855	2.0860	0.2200	1.5920	0.0504	1.5869	0.0258
<i>M</i>	1.5902	2.0831	0.2202	1.5913	0.0501	1.5866	0.0257
<i>J</i>	1.7085	2.0802	0.2203	1.5912	0.0498	1.5874	0.0255
<i>J</i>	1.7420	2.0781	0.2198	1.6030	0.0496	1.6116	0.0254
<i>A</i>	1.6763	2.0761	0.2191	1.6169	0.0494	1.6377	0.0254
<i>S</i>	1.6345	2.0738	0.2188	1.6228	0.0491	1.6454	0.0252
<i>O</i>	1.6880	2.0713	0.2186	1.6240	0.0488	1.6432	0.0251
<i>N</i>	1.7165	2.0691	0.2182	1.6304	0.0485	1.6522	0.0250
<i>D</i>	1.7387	2.0671	0.2176	1.6390	0.0483	1.6650	0.0248
94							
<i>E</i>	1.7345	2.0653	0.2170	1.6490	0.0481	1.6798	0.0247
<i>F</i>	1.7045	2.0634	0.2163	1.6575	0.0478	1.6907	0.0246
<i>M</i>	1.6742	2.0614	0.2158	1.6622	0.0476	1.6935	0.0244

Tabla 3.6 Sección V

En este ejemplo, aún cuando se escogió la varianza $\{\sigma_{w_t}^2\}$ por medio de factores de descuento no se observa una convergencia de la sucesión $\{\hat{\sigma}_t^2\}$. Esto no contradice al *lema* 3.1. Lo que sucede es que la estimación puntual de la varianza observacional $S_{v_t}^2$ cambia con el tiempo y no existe ningún motivo por el cual la sucesión $\{S_{v_t}^2\}$ tienda a un límite en particular.

En las figuras 3.7, 3.8 y 3.9 se muestran gráficamente el desarrollo de la serie y_t y de los predictores puntuales \hat{y}_t .

Los resultados son similares que en el ejemplo anterior. En el caso en que se toma $\delta=1.0$, la brecha de diferencia entre las dos líneas es bastante considerable. El modelo reacciona de manera poco efectiva a los cambios en la serie.

En cambio, el factor de descuento $\delta=0.8$ reacciona casi de manera inmediata a los cambios de dirección de la serie. La distancia entre ambas líneas siempre es pequeña. Este valor de δ parece ser el más adecuado para describir el comportamiento de la serie.

En las figuras 3.7.a, 3.8.a y 3.6.a están graficados los errores de pronóstico para cada mes. Nuevamente los resultados confirman que el factor $\delta=1.0$ no es muy apropiado para realizar pronósticos. La magnitud de los errores es consistentemente grande y esto no es atribuible a cambios bruscos de dirección de la serie sino a la poca eficiencia del modelo para describir el proceso. Una evidencia es que la longitud de las rachas de los signos de los residuales es demasiado grande.

El factor $\delta=0.8$ es el que proporciona los mejores pronósticos, confirmando así, ser la mejor elección para describir el proceso $\{Y_t\}$. La rutina de aprendizaje con este factor es lo suficientemente sensible para adaptarse a los cambios en la tendencia de la serie y al mismo tiempo retener información de las observaciones no tan recientes. Los residuales más grandes se observan cuando la serie presenta caídas o alzas muy fuertes.

A diferencia del primer ejemplo, al parecer la información inicial fue adecuada al no desviar el desarrollo del predictor \hat{Y}_t y no provocar que los errores de pronóstico al comienzo del proceso fueran significativamente más grandes que en el resto del proceso.

Los resultados son similares que en el ejemplo anterior. En el caso en que se toma $\delta=1.0$, la brecha de diferencia entre las dos líneas es bastante considerable. El modelo reacciona de manera poco efectiva a los cambios en la serie.

En cambio, el factor de descuento $\delta=0.8$ reacciona casi de manera inmediata a los cambios de dirección de la serie. La distancia entre ambas líneas siempre es pequeña. Este valor de δ parece ser el más adecuado para describir el comportamiento de la serie.

En las figuras 3.7.a, 3.8.a y 3.6.a están graficados los errores de pronóstico para cada mes. Nuevamente los resultados confirman que el factor $\delta=1.0$ no es muy apropiado para realizar pronósticos. La magnitud de los errores es consistentemente grande y esto no es atribuible a cambios bruscos de dirección de la serie sino a la poca eficiencia del modelo para describir el proceso. Una evidencia es que la longitud de las rachas de los signos de los residuales es demasiado grande.

El factor $\delta=0.8$ es el que proporciona los mejores pronósticos, confirmando así, ser la mejor elección para describir el proceso $\{Y_t\}$. La rutina de aprendizaje con este factor es lo suficientemente sensible para adaptarse a los cambios en la tendencia de la serie y al mismo tiempo retener información de las observaciones no tan recientes. Los residuales más grandes se observan cuando la serie presenta caídas o alzas muy fuertes.

A diferencia del primer ejemplo, al parecer la información inicial fue adecuada al no desviar el desarrollo del predictor \hat{Y}_t y no provocar que los errores de pronóstico al comienzo del proceso fueran significativamente más grandes que en el resto del proceso.

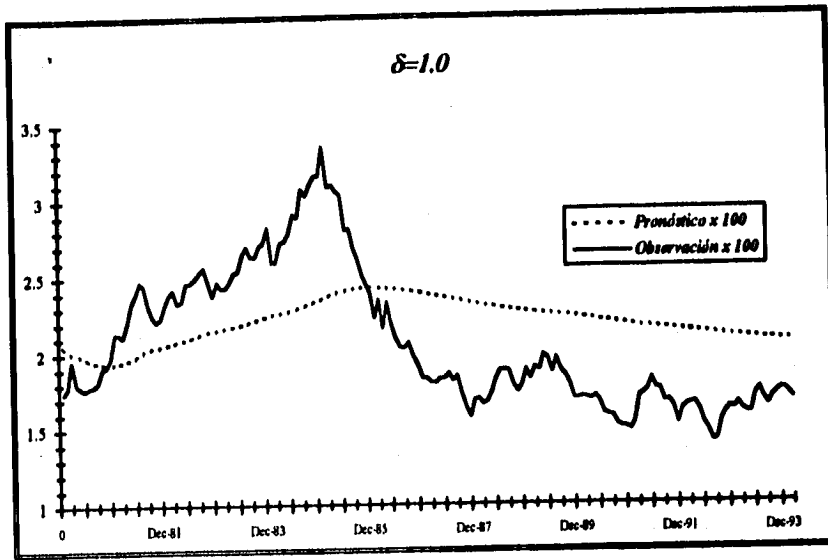


Figura 3.7

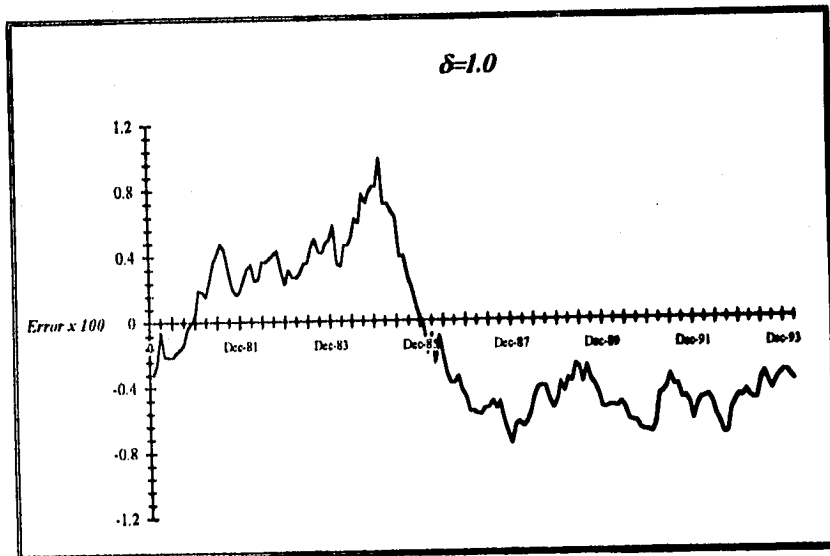


Figura 3.7.a

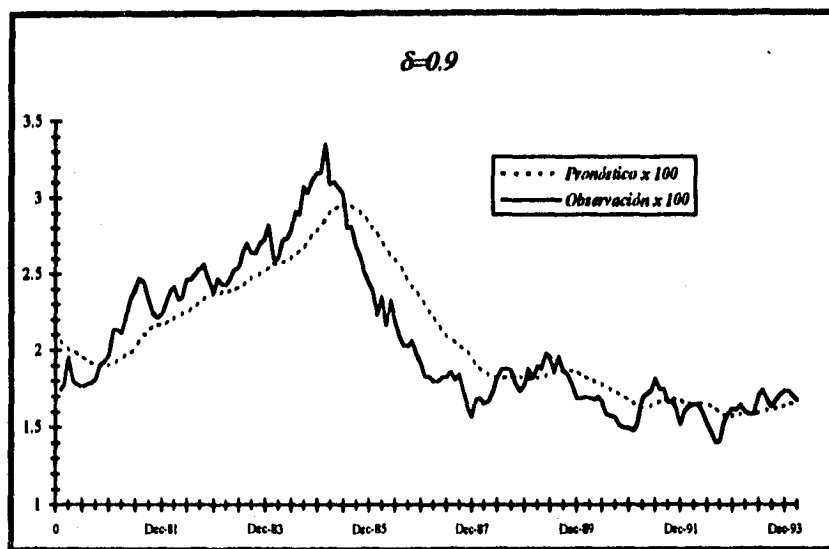


Figura 3.8

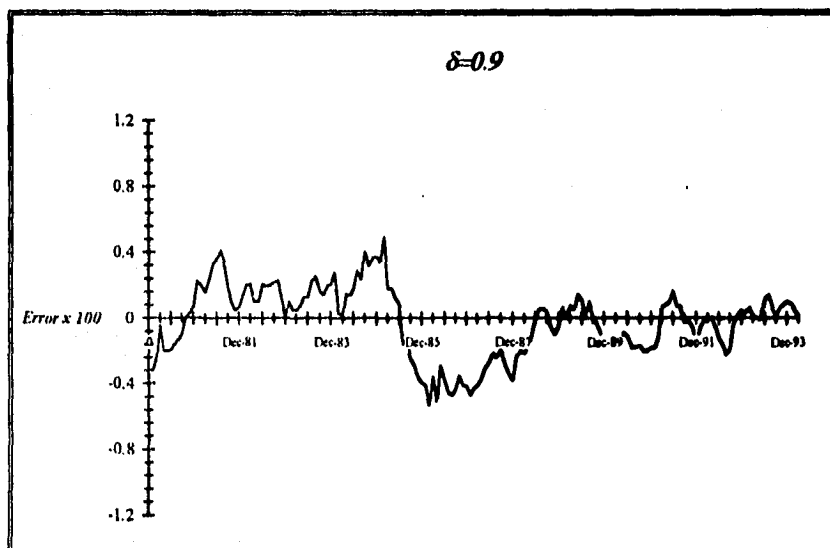


Figura 3.8.a

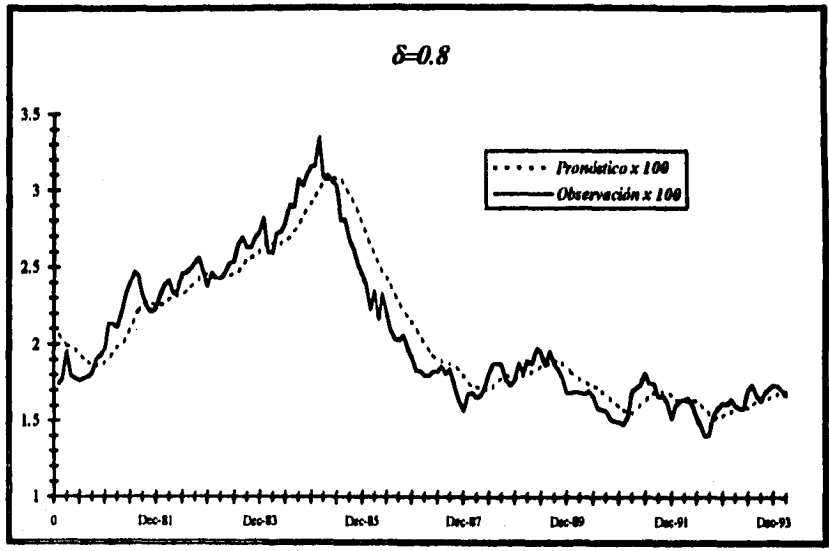


Figura 3.9

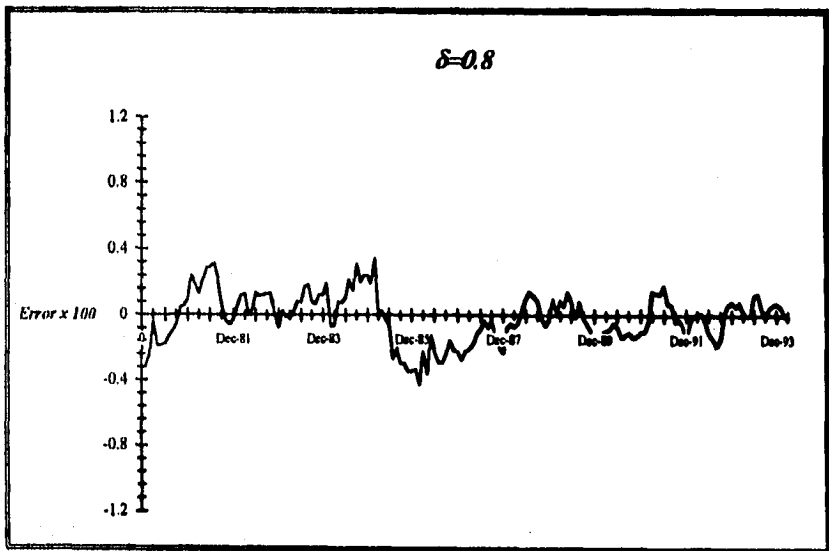


Figura 3.9.a

A continuación se muestra el análisis cuantitativo del desenvolvimiento del modelo para diferentes factores de descuento. La tabla 3.7 presenta los valores de la desviación media absoluta y el error medio cuadrado. Para medir el grado de soporte de los datos se utiliza el cociente $\frac{\ln L(d|y_1, \dots, y_{171})}{\ln L(1.0|y_1, \dots, y_{171})}$.

δ	DMA	EMC	$\frac{\ln L(d y_1, \dots, y_{171})}{\ln L(1.0 y_1, \dots, y_{171})}$
1.0	0.4299	0.2168	0.0000
0.9	0.1720	0.0462	131.6643
0.8	0.1203	0.0227	189.4476

Tabla 3.7.

Comparando las cifras de la tabla se confirma que $\delta = 0.8$ es el factor más apropiado de entre los tres factores utilizados. Escogiendo ese factor se presenta la menor desviación media absoluta, el menor error medio al cuadrado y es el factor con el que los datos son más verosímiles.

Los resultados en ambos ejemplos muestran que un modelo rígido es ineficiente en la descripción de una serie de tiempo y los errores de pronóstico tienden a ser muy grandes. La experiencia indica que factores de descuento alrededor de 0.8 son los que mejor rendimiento tienen para realizar pronósticos. Sin embargo la práctica va formando en el pronosticador un sentido intuitivo que le permite seleccionar la δ adecuada para cada problema.

La intención de este segundo ejemplo es mostrar que integrar un mecanismo en el modelo lineal para modificar el conocimiento sobre σ_v^2 , no implica mayores complicaciones. Al igual que en el ejemplo anterior los cálculos se pueden realizar en una hoja de cálculo electrónica de manera sencilla y rápida.

Finalmente se proporciona la siguiente tabla con los momentos de las distribuciones de pronóstico para diferentes horizontes.

<i>t</i>		$\delta = 1.0$		$\delta = 0.9$		$\delta = 0.8$	
		$\hat{Y}_i(k)$ ×100	$\hat{\sigma}_i^2(k)$ ×10,000	$\hat{Y}_i(k)$ ×10	$\hat{\sigma}_i^2(k)$ ×10,000	$\hat{Y}_i(k)$ ×10	$\hat{\sigma}_i^2(k)$ ×10,000
94							
A	$f_i(1)$	2.0593	0.2154	1.6634	0.0473	1.6896	0.0243
M	$f_i(2)$	2.0593	0.2154	1.6634	0.0475	1.6896	0.0253
J	$f_i(3)$	2.0593	0.2154	1.6634	0.0477	1.6896	0.0262
J	$f_i(4)$	2.0593	0.2154	1.6634	0.0479	1.6896	0.0272
A	$f_i(5)$	2.0593	0.2154	1.6634	0.0481	1.6896	0.0282
S	$f_i(6)$	2.0593	0.2154	1.6634	0.0484	1.6896	0.0292

Tabla 3.8

Las esperanzas de las distribuciones de pronóstico para una δ en particular, es el mismo para los seis diferentes horizontes. Esto es consecuencia de que se supuso que el error observacional ε_{V_t} y el error de transición ε_{W_t} tienen media cero para toda t .

Desafortunadamente, al terminar este trabajo aún no se contaba con los datos necesarios para verificar qué tan exactos son los pronósticos realizados. Sin embargo, es de esperarse que los pronósticos realizados con $\delta = 1.0$ sean poco apropiados debido a que las distribuciones de pronóstico son idénticas para cada horizonte. Esto se debe a que $\sigma_{W_{11},1}^2 = 0$ para $i = 1, \dots, 6$.

CAPITULO IV

EL MODELO DINAMICO LINEAL GENERAL

4.1 Introducción

El *Modelo Dinámico Lineal* es un sistema de ecuaciones en el que se especifica:

- (i) la forma en que las observaciones dependen estocásticamente de los parámetros actuales del proceso.
- (ii) la forma en que los parámetros evolucionan en el tiempo tanto por el carácter dinámico del proceso como por perturbaciones aleatorias.

Definición 4.1

Sea \mathbf{Y}_t el vector *observacional* de $(r \times 1)$ en el tiempo t y sea Θ_t el vector *parámetro* o vector *de estado* de $(n \times 1)$ en el tiempo t , entonces para cada t el *Modelo General Dinámico Lineal Normal*, (MDL), establece la relación entre \mathbf{Y}_t y Θ_t , a través de

$$\mathbf{Y}_t = \mathbf{F}_t \Theta_t + \varepsilon_{V_t} \quad \varepsilon_{V_t} \sim N[\mu_{V_t}, \sigma_{V_t}^2] \quad (4.1)$$

y

$$\Theta_t = \mathbf{G}_t \Theta_{t-1} + \varepsilon_{W_t} \quad \varepsilon_{W_t} \sim N[\mu_{W_t}, \sigma_{W_t}^2] \quad (4.2)$$

donde

\mathbf{F}_t es la *matriz de regresores* de $(n \times r)$,

\mathbf{G}_t es la *matriz de transferencia* de $(n \times n)$,

ε_{V_t} es el *error observacional* de $(r \times 1)$, con media μ_{V_t} y matriz de varianzas $\sigma_{V_t}^2$.

ε_{W_t} es el *error de transición* de $(n \times 1)$, con media μ_{W_t} y matriz de varianzas $\sigma_{W_t}^2$.

Las sucesiones $\{\varepsilon_{V_t}\}$ y $\{\varepsilon_{W_t}\}$ son independiente y mutuamente independientes.¹⁷

El *MDL* se caracteriza por la sextupla

$$\{F_t, G_t, \mu_{V_t}, \sigma_{V_t}^2, \mu_{W_t}, \sigma_{W_t}^2\}.$$

Se debe de observar que el *Modelo Lineal Normal* convencional es un caso simple y estático del *MDL* en el cual $\Theta_t = \Theta$ y $\{\sigma_{V_t}^2\}$ son independientes del tiempo.

Definición 4.2

Dentro de la clase de los *MDL* existen dos subconjuntos importantes

- (i) Si la pareja $\{F_t, G_t\}$ es constante para toda t entonces el modelo se define como un *Modelo Dinámico Lineal de Serie de Tiempo* o (*MDLST*) y se caracteriza por la sextupla

$$\{F, G, \mu_{V_t}, \sigma_{V_t}^2, \mu_{W_t}, \sigma_{W_t}^2\}.$$

- (ii) Si en un *MDLST* la cuadrupla $\{\mu_{V_t}, \sigma_{V_t}^2, \mu_{W_t}, \sigma_{W_t}^2\}$ es constante para toda t entonces se dice que el modelo es homogéneo en el tiempo y se le conoce como *Modelo Dinámico Lineal Constante*, (*MDLC*), y se caracteriza por la sextupla

$$\{F, G, \mu_V, \sigma_V^2, \mu_W, \sigma_W^2\}.$$

El modelo polinomial de primer orden es un ejemplo univariado de la clase *MDLST*, al ocurrir que $F_t = 1$, $G_t = 1$ para toda t .

Una caminata aleatoria

$$\begin{aligned} Y_t &= \Theta_t + \varepsilon_{V_t} & \varepsilon_{V_t} &\sim N[0, \sigma_{V_t}^2] \\ \Theta_t &= \Theta_{t-1} + \varepsilon_{W_t} & \varepsilon_{W_t} &\sim N[0, \sigma_{W_t}^2] \end{aligned}$$

es un ejemplo univariado de un modelo dinámico lineal constante.

¹⁷ A la estructura del Modelo Dinámico Lineal General, sin el supuesto de Normalidad en los errores, se le conoce también como la Forma General Estado Espacio, (en inglés General State Space Form).

En esta tesis, el trabajo restante será describir el caso univariado, ($r = 1$), siendo el vector Y_t de dimensión 1, teniendo como ecuaciones

$$Y_t = F_t \Theta_t + \varepsilon_{v_t} \quad \varepsilon_{v_t} \sim N[\mu_{v_t}, \sigma_{v_t}^2] \quad (4.1.a)$$

y

$$\Theta_t = G_t \Theta_{t-1} + \varepsilon_{w_t} \quad \varepsilon_{w_t} \sim N[\mu_{w_t}, \sigma_{w_t}^2] \quad (4.2.a)$$

y la información inicial

$$(\Theta_0 | D_0) \sim N[m_0, S_0^2] \quad (4.3)$$

para algunos m_0, S_0^2 ; las sucesiones de *errores de transición* $\{\varepsilon_{w_t}\}$ y de *observación* son independientes, mutuamente independientes e independientes de la información inicial, $(\Theta_0 | D_0)$.

La definición del vector Θ , para cualquier modelo estadístico en particular se determina por construcción. Sus elementos pueden o no ser identificables con componentes que tengan una interpretación substancial, como una tendencia o factor de estacionalidad. Desde el punto de vista técnico, el objeto de la representación en la forma estado espacio es definir a Θ , de tal forma que, al tiempo t , contenga toda la información relevante que se tiene del sistema, por esta razón se debe buscar que el vector de estado tenga el menor número de elementos posible.

No necesariamente hay una representación única para un problema en particular, por lo que la forma del vector de estado en la *definición 4.1* puede no ser la misma. Si se toma a una matriz arbitraria no singular R de $(n \times n)$ y se considera a un nuevo vector de estado $\Theta_t^* = R\Theta_t$, premultiplicando por R la ecuación (4.2.a) se tiene

$$\Theta_t^* = G_t^* \Theta_{t-1}^* + \varepsilon_{w_t}^* \quad \varepsilon_{w_t}^* \sim N[R\mu_{w_t}, R\sigma_{w_t}^2 R'] \quad (4.2.b)$$

donde $G_t^* = R G_t R^{-1}$. La correspondiente ecuación de observación queda de la siguiente forma

$$Y_t = F_t^* \Theta_t^* + \varepsilon_{v_t} \quad \varepsilon_{v_t} \sim N[\mu_{v_t}, \sigma_{v_t}^2] \quad (4.3.b)$$

donde $F_t' = F_t' R^{-1}$.

4.1.1 El MDL Univariado

Una vez que el modelo se ha especificado de acuerdo a la *definición 4.1* con $r = 1$, se puede aplicar el procedimiento enunciado en el siguiente teorema y que es un caso particular del llamado *filtro de Kalman*.¹⁸

Teorema 4.1

En el MDL univariado, si $(\Theta_0|D_0) \sim N[m_0, S_0^2]$ entonces para $t \geq 1$ se tienen las distribuciones

(a) A priori para Θ_t ,

$$(\Theta_t|D_{t-1}) \sim N[\bar{m}_t, \bar{S}_t^2],$$

donde

$$\bar{m}_t = G_t m_{t-1} + \mu_{w_t},$$

$$\bar{S}_t^2 = G_t S_{t-1}^2 G_t' + \sigma_{w_t}^2.$$

(b) Pronóstico de un periodo

$$f_{t-1}(1) \sim N[\hat{Y}_t, \hat{\sigma}_t^2],$$

donde

$$\hat{Y}_t = F_t \bar{m}_t + \mu_{v_t},$$

$$\hat{\sigma}_t^2 = F_t' \bar{S}_t^2 F_t + \sigma_{v_t}^2.$$

(c) A posteriori para Θ_t ,

¹⁸ El filtro de Kalman es un procedimiento recursivo para calcular el estimador óptimo del vector de estado en el tiempo t , basándose en la información disponible en el tiempo t . El filtro de Kalman permite que la estimación del vector de estado sea continuamente actualizada en el tiempo conforme se presentan nuevas observaciones. [10 Harvey].

$$(\Theta_t | D_t) \sim N[\mathbf{m}_t, \mathbf{S}_t^2],$$

con

$$\begin{aligned} \mathbf{m}_t &= \bar{\mathbf{m}}_t + \mathbf{A}_t \mathbf{e}_t \\ &= \bar{\mathbf{m}}_t + (\hat{\sigma}_t^2)^{-1} \bar{\mathbf{S}}_t' \mathbf{F}_t (Y_t - \hat{Y}_t), \end{aligned}$$

$$\begin{aligned} \mathbf{S}_t^2 &= \bar{\mathbf{S}}_t^2 - (\hat{\sigma}_t^2)^{-1} \bar{\mathbf{S}}_t' \mathbf{F}_t \mathbf{F}_t' \bar{\mathbf{S}}_t \\ &= \bar{\mathbf{S}}_t^2 - \hat{\sigma}_t^2 \mathbf{A}_t \mathbf{A}_t', \end{aligned}$$

donde

$$\mathbf{A}_t = (\hat{\sigma}_t^2)^{-1} \bar{\mathbf{S}}_t' \mathbf{F}_t$$

es la matriz adaptativa

$$\mathbf{e}_t = y_t - \hat{Y}_t$$

es el error de pronóstico de un periodo.

Demostración.

Demostración por inducción.

Supóngase que se cumple la distribución en (c) para $t-1$. Θ_{t-1} , condicionada sobre D_{t-1} es la suma de dos vectores aleatorios independientes con distribución normal, Θ_{t-1} y ϵ_{w_t} , por lo tanto $(\Theta_t | D_{t-1})$ tiene también distribución normal, con media y varianza

$$\begin{aligned} E[\Theta_t | D_{t-1}] &= E[\mathbf{G}_t \Theta_{t-1} + \epsilon_{w_t} | D_{t-1}] \\ &= \mathbf{G}_t E[\Theta_{t-1} | D_{t-1}] + E[\epsilon_{w_t} | D_{t-1}] \\ &= \mathbf{G}_t \mathbf{m}_{t-1} + \mu_{w_t} = \bar{\mathbf{m}}_t \end{aligned}$$

$$\begin{aligned} \text{AR}[\Theta_t | D_{t-1}] &= \text{VAR}[\mathbf{G}_t \Theta_{t-1} + \epsilon_{w_t} | D_{t-1}] \\ &= \mathbf{G}_t \text{VAR}[\Theta_{t-1} | D_{t-1}] \mathbf{G}_t' + \text{VAR}[\epsilon_{w_t} | D_{t-1}] \\ &= \mathbf{G}_t \bar{\mathbf{S}}_{t-1}^2 \mathbf{G}_t' + \sigma_{w_t}^2 = \bar{\mathbf{S}}_t^2 \end{aligned}$$

quedando demostrado (a). Similarmente, $(Y_i|D_{i-1})$ es la suma de cantidades aleatorias independientes con distribución normal, $F_i\Theta_i$ y ε_{vi} , por lo tanto también tiene distribución normal con esperanza y varianza obtenidas de la suma de las esperanzas y varianzas de cada sumando

$$\begin{aligned} E[Y_i|D_{i-1}] &= E[F_i\Theta_i + \varepsilon_{vi}|D_{i-1}] \\ &= E[F_i\Theta_i|D_{i-1}] + E[\varepsilon_{vi}|D_{i-1}] \\ &= F_i\bar{m}_i + \mu_{vi} \end{aligned}$$

$$\begin{aligned} AR[Y_i|D_{i-1}] &= VAR[F_i\Theta_i + \varepsilon_{vi}|D_{i-1}] \\ &= F_i VAR[\Theta_i|D_{i-1}] F_i + VAR[\varepsilon_{vi}|D_{i-1}] \\ &= F_i \bar{S}_i^2 F_i + \sigma_{vi}^2 \end{aligned}$$

Entonces

$$f_{i-1}(1) \sim N[\hat{Y}_i, \hat{\sigma}_i^2],$$

con

$$\begin{aligned} \hat{Y}_i &= F_i \bar{m}_i + \mu_{vi} \\ \hat{\sigma}_i^2 &= F_i \bar{S}_i^2 F_i + \sigma_{vi}^2 \end{aligned}$$

quedando demostrado (b). Existen dos maneras de demostrar (c).

(i) Vía Teorema de Bayes.

(ii) Vía propiedades de la distribución Normal.

(i) De acuerdo a (a) la función de densidad a priori para Θ , es

$$f_{\Theta_i|D_{i-1}}(\theta_i) \propto \exp\left\{-\frac{1}{2}(\theta_i - \bar{m}_i)' (\bar{S}_i^2)^{-1} (\theta_i - \bar{m}_i)\right\} \quad (4.4)$$

y de acuerdo a (4.1) la función de densidad para Y_i está dada por

$$f_{Y_i|\Theta_i}(y_i|\theta_i) \propto \exp\left\{-\frac{1}{2}(\sigma_{vi}^2)^{-1} [y_i - (F_i\theta_i + \mu_{vi})]^2\right\} \quad (4.5)$$

Por el Teorema de Bayes se tiene que

$$\begin{aligned}
f_{\theta_i|D_i}(\theta_i) &\propto f_{\theta_i|D_{i-1}}(\theta_i) f_{y_i|\theta_i}(y_i|\theta_i) \\
&\propto \exp\left\{-\frac{1}{2}(\theta_i - \bar{m}_i)'(\bar{S}_i^2)^{-1}(\theta_i - \bar{m}_i)\right\} \exp\left\{-\frac{1}{2}(\sigma_{v_i}^2)^{-1}[y_i - (\mathbf{F}_i\theta_i + \mu_{v_i})]^2\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[(\theta_i - \bar{m}_i)'(\bar{S}_i^2)^{-1}(\theta_i - \bar{m}_i) + (\sigma_{v_i}^2)^{-1}[y_i - (\mathbf{F}_i\theta_i + \mu_{v_i})]^2\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[\theta_i'(\bar{S}_i^2)^{-1}\theta_i - \bar{m}_i'(\bar{S}_i^2)^{-1}\theta_i - \theta_i'(\bar{S}_i^2)^{-1}\bar{m}_i + \bar{m}_i'(\bar{S}_i^2)^{-1}\bar{m}_i\right.\right. \\
&\quad \left.\left.+ (\sigma_{v_i}^2)^{-1}(y_i^2 - 2y_i\mathbf{F}_i\theta_i - 2y_i\mu_{v_i} + \theta_i'\mathbf{F}_i\mathbf{F}_i\theta_i + 2\mathbf{F}_i\theta_i\mu_{v_i} + \mu_{v_i})\right]\right\}
\end{aligned}$$

Reordenando términos y excluyendo constantes que no involucren a θ , se obtiene

$$f_{\theta_i|D_i}(\theta_i) \propto \exp\left\{-\frac{1}{2}\left[\theta_i'[(\bar{S}_i^2)^{-1} + \mathbf{F}_i\mathbf{F}_i(\sigma_{v_i}^2)^{-1}]\theta_i - 2\theta_i'[(\bar{S}_i^2)^{-1}\bar{m}_i + (\sigma_{v_i}^2)^{-1}(\mathbf{F}_iy_i - \mathbf{F}_i\mu_{v_i})]\right]\right\} \quad (4.6)$$

Por otra parte, utilizando la definición de \mathbf{S}_i^2 enunciada en el teorema, se observa que

$$\begin{aligned}
((\bar{S}_i^2)^{-1} + \mathbf{F}_i\mathbf{F}_i(\sigma_{v_i}^2)^{-1})\mathbf{S}_i^2 &= ((\bar{S}_i^2)^{-1} + \mathbf{F}_i\mathbf{F}_i(\sigma_{v_i}^2)^{-1})\left(\bar{S}_i^2 - (\hat{\sigma}_i^2)^{-1}\bar{S}_i^2'\mathbf{F}_i\mathbf{F}_i\bar{S}_i^2\right) \\
&= \mathbf{I} - (\hat{\sigma}_i^2)^{-1}\mathbf{F}_i\mathbf{F}_i\bar{S}_i^2' + (\sigma_{v_i}^2)^{-1}\mathbf{F}_i\mathbf{F}_i\bar{S}_i^2' - (\sigma_{v_i}^2)^{-1}(\hat{\sigma}_i^2)^{-1}\mathbf{F}_i\mathbf{F}_i\bar{S}_i^2'\mathbf{F}_i\mathbf{F}_i\bar{S}_i^2 \\
&= \mathbf{I} - (\sigma_{v_i}^2\hat{\sigma}_i^2)^{-1}\left[(\sigma_{v_i}^2 - \hat{\sigma}_i^2)\mathbf{F}_i\mathbf{F}_i\bar{S}_i^2' + \mathbf{F}_i\mathbf{F}_i\bar{S}_i^2'\mathbf{F}_i\mathbf{F}_i\bar{S}_i^2\right]
\end{aligned}$$

pero como

$$\hat{\sigma}_i^2 - \sigma_{v_i}^2 = \mathbf{F}_i\bar{S}_i^2\mathbf{F}_i$$

entonces

$$\begin{aligned}
((\bar{S}_i^2)^{-1} + \mathbf{F}_i\mathbf{F}_i(\sigma_{v_i}^2)^{-1})\mathbf{S}_i^2 &= \mathbf{I} - (\sigma_{v_i}^2\hat{\sigma}_i^2)^{-1}\left[(\sigma_{v_i}^2 - \hat{\sigma}_i^2)\mathbf{F}_i\mathbf{F}_i\bar{S}_i^2' + \mathbf{F}_i(\hat{\sigma}_i^2 - \sigma_{v_i}^2)\mathbf{F}_i\bar{S}_i^2\right] \\
&= \mathbf{I}
\end{aligned}$$

por lo tanto

$$(\mathbf{S}_i^2)^{-1} = ((\bar{S}_i^2)^{-1} + \mathbf{F}_i\mathbf{F}_i(\sigma_{v_i}^2)^{-1}) \quad (4.7)$$

Utilizando la definición de \mathbf{m}_i , establecida en el teorema y la forma (4.7) se tiene que

$$\begin{aligned}
 (\mathbf{S}_i^2)^{-1} \mathbf{m}_i &= \left((\bar{\mathbf{S}}_i^2)^{-1} + \mathbf{F}_i \mathbf{F}_i' (\sigma_{\nu_i}^2)^{-1} \right) (\bar{\mathbf{m}}_i + \mathbf{A}_i e_i) \\
 &= (\bar{\mathbf{S}}_i^2)^{-1} \bar{\mathbf{m}}_i + (\bar{\mathbf{S}}_i^2)^{-1} \mathbf{A}_i e_i + \mathbf{F}_i \mathbf{F}_i' (\sigma_{\nu_i}^2)^{-1} \bar{\mathbf{m}}_i + \mathbf{F}_i \mathbf{F}_i' (\sigma_{\nu_i}^2)^{-1} \mathbf{A}_i e_i \\
 &= (\bar{\mathbf{S}}_i^2)^{-1} \bar{\mathbf{m}}_i + (\bar{\mathbf{S}}_i^2)^{-1} (\hat{\sigma}_i^2)^{-1} \bar{\mathbf{S}}_i^2 \mathbf{F}_i e_i + \mathbf{F}_i \mathbf{F}_i' (\sigma_{\nu_i}^2)^{-1} \bar{\mathbf{m}}_i + \mathbf{F}_i \mathbf{F}_i' (\sigma_{\nu_i}^2)^{-1} (\hat{\sigma}_i^2)^{-1} \bar{\mathbf{S}}_i^2 \mathbf{F}_i e_i \\
 &= (\bar{\mathbf{S}}_i^2)^{-1} \bar{\mathbf{m}}_i + (\hat{\sigma}_i^2 \sigma_{\nu_i}^2)^{-1} \left[\sigma_{\nu_i}^2 \mathbf{F}_i e_i + \hat{\sigma}_i^2 \mathbf{F}_i \bar{\mathbf{m}}_i + \mathbf{F}_i (\hat{\sigma}_i^2 - \sigma_{\nu_i}^2) e_i \right] \\
 &= (\bar{\mathbf{S}}_i^2)^{-1} \bar{\mathbf{m}}_i + (\sigma_{\nu_i}^2)^{-1} \left[\mathbf{F}_i \bar{\mathbf{m}}_i + \mathbf{F}_i (y_i - \hat{y}_i) \right] \\
 &= (\bar{\mathbf{S}}_i^2)^{-1} \bar{\mathbf{m}}_i + (\sigma_{\nu_i}^2)^{-1} \left[\mathbf{F}_i \bar{\mathbf{m}}_i + \mathbf{F}_i y_i - \mathbf{F}_i (\mathbf{F}_i \bar{\mathbf{m}}_i + \mu_{\nu_i}) \right] \\
 &= (\bar{\mathbf{S}}_i^2)^{-1} \bar{\mathbf{m}}_i + (\sigma_{\nu_i}^2)^{-1} \left[\mathbf{F}_i y_i - \mathbf{F}_i \mu_{\nu_i} \right]
 \end{aligned}$$

en consecuencia (4.6) puede reescribirse de la siguiente forma

$$f_{\Theta_i | D_i}(\theta_i) \propto \exp \left\{ -\frac{1}{2} \left[\theta_i' (\mathbf{S}_i^2)^{-1} \theta_i - 2 \theta_i' (\mathbf{S}_i^2)^{-1} \mathbf{m}_i \right] \right\} \quad (4.6.a)$$

que a su vez es proporcional a

$$f_{\Theta_i | D_i}(\theta_i) \propto \exp \left\{ -\frac{1}{2} \left[(\theta_i - \mathbf{m}_i)' (\mathbf{S}_i^2)^{-1} (\theta_i - \mathbf{m}_i) \right] \right\}$$

que al multiplicar por las constantes de normalización resulta ser la densidad de una distribución normal multivariada con media \mathbf{m}_i y varianza \mathbf{S}_i^2 .

(ii) Debido a que cualquier función lineal de Y_i y Θ_i es una combinación lineal de variables y vectores normales e independientes $\varepsilon_{\nu_i}, \varepsilon_{\nu_i}, \Theta_{i-1}$, tal función lineal condicionada sobre D_{i-1} también se distribuye de forma normal y por definición la distribución conjunta $(Y_i, \Theta_i | D_{i-1})$ es normal bivariada. La covarianza entre Y_i y Θ_i , está dada por

$$\begin{aligned}
 COV[Y_i, \Theta_i | D_{i-1}] &= COV[\mathbf{F}_i \Theta_i + \varepsilon_{\nu_i}, \Theta_i | D_{i-1}] \\
 &= COV[\varepsilon_{\nu_i}, \Theta_i | D_{i-1}] + \mathbf{F}_i VAR[\Theta_i | D_{i-1}] \\
 &= \mathbf{F}_i \bar{\mathbf{S}}_i^2
 \end{aligned}$$

por lo tanto

$$\begin{pmatrix} Y_t \\ \Theta_t \end{pmatrix} | D_{t-1} \sim N \left[\begin{pmatrix} \hat{Y}_t \\ \bar{m}_t \end{pmatrix}, \begin{pmatrix} \hat{\sigma}_t^2 & \mathbf{F}_t' \bar{\mathbf{S}}_t^2 \\ \bar{\mathbf{S}}_t^2 \mathbf{F}_t & \bar{\mathbf{S}}_t^2 \end{pmatrix} \right]$$

De acuerdo a la sección A.1.2 del Apéndice se tiene la distribución condicional

$$(\Theta_t | Y_t = y_t, D_{t-1}) \sim N[\mathbf{m}_t(y_t), \mathbf{S}_t^2(y_t)],$$

con

$$\begin{aligned} \mathbf{m}_t(y_t) &= \bar{m}_t + \frac{(y_t - \hat{Y}_t)}{\hat{\sigma}_t^2} \bar{\mathbf{S}}_t^2 \mathbf{F}_t \\ &= \bar{m}_t + \mathbf{A}_t e_t \end{aligned}$$

$$\begin{aligned} \mathbf{S}_t^2(y_t) &= \bar{\mathbf{S}}_t^2 - \bar{\mathbf{S}}_t^2 \mathbf{F}_t (\hat{\sigma}_t^2)^{-1} \mathbf{F}_t' \bar{\mathbf{S}}_t^2 \\ &= \bar{\mathbf{S}}_t^2 - \mathbf{A}_t \mathbf{A}_t' \hat{\sigma}_t^2 \end{aligned}$$

donde

$$\mathbf{A}_t = \bar{\mathbf{S}}_t^2 \mathbf{F}_t (\hat{\sigma}_t^2)^{-1}$$

es la matriz de regresión de Θ_t sobre Y_t .

Finalmente se completa la demostración a partir de que (a) se cumple para $t = 1$.

Es importante resaltar la importancia del *Teorema 4.1*

- (1) Proporciona una estructura racional y coherente para realizar pronósticos en situaciones de escasa información histórica.
- (2) Facilita la intervención del pronosticador.

4.2 Distribuciones de Pronóstico

Definición 4.3

Para cada t a

$$E[Y_t | \Theta_t] = F_t \Theta_t + \mu_{v_t}$$

se le llama *función de respuesta media*. A

$$\mu_t = F_t \Theta_t$$

se le define como el *nivel* de la serie en el tiempo t .

Definición 4.4

En el tiempo t y de acuerdo a la información disponible D_t , se definen el pronóstico *parametral* de k pasos en el futuro

$$(\Theta_{t+k} | D_t) \quad k > 0$$

La forma de las distribuciones de pronóstico esta dada en el siguiente

Teorema 4.2

En cada tiempo t y para $k \geq 1$ se tienen las siguientes distribuciones con momentos definidos recursivamente

(a) Pronóstico parametral de horizonte k

$$(\Theta_{t+k} | D_t) \sim N[m_{\Theta_t}(k), S_{\Theta_t}^2(k)]$$

con

$$m_{\Theta_t}(k) = G_{t+k} m_{\Theta_t}(k-1) + \mu_{w_{t+k}},$$

$$S_{\Theta_t}^2(k) = G_{t+k} S_{\Theta_t}^2(k-1) G_{t+k}' + \sigma_{w_{t+k}}^2,$$

con valores iniciales $m_{\Theta_t}(0) = m_t$ y $S_{\Theta_t}^2(0) = S_t^2$.

(b) Pronóstico de k periodos en el futuro

$$f_t(k) \sim N[\hat{Y}_t(k), \hat{\sigma}_t^2(k)]$$

con

$$\hat{Y}_i(k) = F'_{i+k} m_{\Theta_i}(k) + \mu_{V_{i+k}},$$

$$\hat{\sigma}_i^2(k) = F'_{i+k} S_{\Theta_i}^2(k) F_{i+k} + \sigma_{V_{i+k}}^2.$$

Demostración.

Se define para toda t y para $0 \leq r < t$ al conjunto de matrices $H_i(r) = G_i G_{i-1} \dots G_{i-r+1}$ siendo $H_i(0) = I$. De acuerdo a (4.1.a) se tiene que

$$\begin{aligned} \Theta_{i+k} &= H_{i+k}(k) \Theta_i + \sum_{r=1}^k H_{i+k}(k-r) \varepsilon_{W_{i+r}}, \\ &= G_{i+k} \left[H_{i+k-1}(k-1) \Theta_i + \sum_{r=1}^{k-1} H_{i+k-1}(k-1-r) \varepsilon_{W_{i+r}} \right] + \varepsilon_{W_{i+k}} \end{aligned} \quad (4.8)$$

Puesto que Θ_i y $\varepsilon_{W_{i+r}}$, $i=1, \dots, k$ condicionados sobre D_i , son independientes y se distribuyen normalmente, entonces

$$(\Theta_{i+k} | D_i) \sim N[m_{\Theta_i}(k), S_{\Theta_i}^2(k)].$$

Los momentos de tal distribución se obtienen de la siguiente manera

$$\begin{aligned} E[\Theta_{i+k} | D_i] &= H_{i+k}(k) E[\Theta_i | D_i] + \sum_{r=1}^k H_{i+k}(k-r) E[\varepsilon_{W_{i+r}} | D_i] \\ &= H_{i+k}(k) m_i + \sum_{r=1}^k H_{i+k}(k-r) \mu_{W_{i+r}} \\ &= m_{\Theta_i}(k) \end{aligned}$$

usando (4.8) se tiene que

$$m_{\Theta_i}(k) = G_{i+k} m_{\Theta_i}(k-1) + \mu_{W_{i+k}}.$$

$$\begin{aligned}
AR[\Theta_{t+k}|D_t] &= \mathbf{H}_{t+k}(k)VAR[\Theta_t|D_t]\mathbf{H}'_{t+k}(k) + \\
&+ \sum_{r=1}^k \mathbf{H}_{t+k}(k-r)VAR[\varepsilon_{v_{t+r}}|D_t]\mathbf{H}'_{t+k}(k-r) \\
&= \mathbf{H}_{t+k}(k)\mathbf{S}_t^2\mathbf{H}'_{t+k}(k) + \sum_{r=1}^k \mathbf{H}_{t+k}(k-r)\sigma_{v_{t+r}}^2\mathbf{H}'_{t+k}(k-r) \\
&= \mathbf{S}_{\Theta_t}^2(k)
\end{aligned}$$

nuevamente usando (4.8) se tiene que

$$\mathbf{S}_{\Theta_t}^2(k) = \mathbf{G}_t\mathbf{S}_{\Theta_t}^2(k-1)\mathbf{G}'_t + \sigma_{w_t}^2.$$

Para $k=0$ se tiene que $E[\Theta_{t+0}|D_t] = \mathbf{m}_t$ y $AR[\Theta_{t+0}|D_t] = \mathbf{S}_t^2$ completándose la demostración de (a).

De acuerdo a (4.1.a) se tiene que

$$Y_{t+k} = \mathbf{F}'_{t+k}\Theta_{t+k} + \mu_{v_{t+k}}$$

Como Θ_t, μ_{v_t} son independientes $\forall t$, entonces

$$\begin{aligned}
f_t(k) &\sim N[\hat{Y}_t(k), \hat{\sigma}_t^2(k)] \\
&\quad \circ \\
(Y_{t+k}|D_t) &\sim N[\hat{Y}_t(k), \hat{\sigma}_t^2(k)],
\end{aligned}$$

con

$$\begin{aligned}
\hat{Y}_t(k) &= E[Y_{t+k}|D_t] \\
&= E[\mathbf{F}'_{t+k}\Theta_{t+k} + \varepsilon_{v_{t+k}}|D_t] \\
&= \mathbf{F}'_{t+k}E[\Theta_{t+k}|D_t] + E[\varepsilon_{v_{t+k}}|D_t] \\
&= \mathbf{F}'_{t+k}\mathbf{m}_{\Theta_t}(k) + \mu_{v_{t+k}}
\end{aligned}$$

$$\begin{aligned}
\hat{\sigma}_i^2(k) &= \text{VAR}[Y_{i+k}|D_i] \\
&= \text{VAR}[\mathbf{F}'_{i+k}\boldsymbol{\Theta}_{i+k} + \varepsilon_{i+k}|D_i] \\
&= \mathbf{F}'_{i+k}\text{VAR}[\boldsymbol{\Theta}_{i+k}|D_i]\mathbf{F}_{i+k} + \text{VAR}[\varepsilon_{i+k}|D_i] \\
&= \mathbf{F}'_{i+k}\mathbf{S}_{\boldsymbol{\Theta}_i}^2(k)\mathbf{F}_{i+k} + \sigma_{\varepsilon_{i+k}}^2
\end{aligned}$$

Corolario 4.1

Para cualquier par de enteros j, k tales que $k > j \geq 1$ se tiene que

$$\text{COV}[\boldsymbol{\Theta}_{i+k}, \boldsymbol{\Theta}_{i+j}|D_i] = \mathbf{S}_{\boldsymbol{\Theta}_i}(k, j)$$

y

$$\begin{aligned}
\text{COV}[Y_{i+k}, Y_{i+j}|D_i] &= \mathbf{F}'_{i+k}\mathbf{S}_{\boldsymbol{\Theta}_i}(k, j)\mathbf{F}_{i+j} \\
&= \hat{\sigma}_i(k, j)
\end{aligned}$$

con $\mathbf{S}_{\boldsymbol{\Theta}_i}(k, j)$ definida recursivamente por

$$\mathbf{S}_{\boldsymbol{\Theta}_i}(i, j) = \mathbf{G}_{i,i}\mathbf{S}_{\boldsymbol{\Theta}_i}(i-1, j), \quad i = j+1, \dots, k$$

con valor inicial $\mathbf{S}_{\boldsymbol{\Theta}_i}(j, j) = \mathbf{S}_{\boldsymbol{\Theta}_i}^2(j) \quad \forall i$.

Demostración.

Supongase que la afirmación es válida para i

$$\begin{aligned}
\mathbf{S}_{\boldsymbol{\Theta}_i}(i+1, j) &= \text{COV}[\boldsymbol{\Theta}_{i+1}, \boldsymbol{\Theta}_{i+j}|D_i] \\
&= \text{COV}[\mathbf{G}_{i+1,i}\boldsymbol{\Theta}_{i+1} + \varepsilon_{i+1}, \boldsymbol{\Theta}_{i+j}|D_i] \\
&= \mathbf{G}_{i+1,i}\text{COV}[\boldsymbol{\Theta}_{i+1}, \boldsymbol{\Theta}_{i+j}|D_i] + \text{COV}[\varepsilon_{i+1}, \boldsymbol{\Theta}_{i+j}|D_i] \\
&= \mathbf{G}_{i+1,i}\text{COV}[\boldsymbol{\Theta}_{i+1}, \boldsymbol{\Theta}_{i+j}|D_i] = \mathbf{G}_{i+1,i}\mathbf{S}_{\boldsymbol{\Theta}_i}(i, j)
\end{aligned}$$

y puesto que

$$\text{COV}[\boldsymbol{\Theta}_{i+j}, \boldsymbol{\Theta}_{i+j}|D_i] = \text{VAR}[\boldsymbol{\Theta}_{i+j}|D_i]$$

entonces

$$S_{\theta_i}(j, j) = S_{\theta_i}^2(j)$$

Finalmente

$$\begin{aligned} \text{COV}[Y_{it+k}, Y_{it+j} | D_i] &= \text{COV}[\mathbf{F}'_{it+k} \boldsymbol{\Theta}_{it+k}, \mathbf{F}'_{it+j} \boldsymbol{\Theta}_{it+j} | D_i] \\ &= \mathbf{F}'_{it+k} \mathbf{S}_{\theta_i}(k, j) \mathbf{F}_{it+j} \\ &= \hat{\sigma}_i(k, j) \end{aligned}$$

Corolario 4.2

En cada tiempo t y para $k \geq$ se tiene la distribución de pronóstico gata de horizonte k

$$f_i^H(k) \sim N[m_{H_i}(k), S_{H_i}^2(k)]$$

con momentos definidos recursivamente

$$\begin{aligned} m_{H_i}(k) &= \sum_{i=1}^k [F'_{i,i} m_{\Theta_i}(i) + \mu_{V_{i,i}}] \\ &= \sum_{i=1}^k \hat{Y}_i(k) \end{aligned}$$

$$S_{H_i}^2(k) = \sum_{i=1}^k \hat{\sigma}_i^2(i) + \sum_{i=1}^k \sum_{j \neq i} \sigma_i(i, j) + \sum_{i=1}^k \varepsilon_{V_{i,i}}$$

Demostración.

$$\begin{aligned} \sum_{i=1}^k Y_{i,i} &= \sum_{i=1}^k [F'_{i,i} \Theta_{i,i} + \varepsilon_{V_{i,i}}] \\ &= \sum_{i=1}^k F'_{i,i} \Theta_{i,i} + \sum_{i=1}^k \varepsilon_{V_{i,i}} \quad k > 0 \end{aligned}$$

entonces

$$\begin{aligned} E\left[\sum_{i=1}^k Y_{i,i} | D_t\right] &= \sum_{i=1}^k E[F'_{i,i} \Theta_{i,i} | D_t] + \sum_{i=1}^k E[\varepsilon_{V_{i,i}} | D_t] \\ &= \sum_{i=1}^k F'_{i,i} m_{\Theta_i}(i) + \sum_{i=1}^k \mu_{V_{i,i}} \\ &= \sum_{i=1}^k \hat{Y}_i(i) \end{aligned}$$

$$\begin{aligned}
AR \left[\sum_{i=1}^k Y_{i,i} | D_i \right] &= \sum_{i=1}^k \mathbf{F}'_{i,i} VAR[\boldsymbol{\Theta}_{i,i} | D_i] \mathbf{F}_{i,i} + \sum_{i=1}^k \sum_{j \neq i}^k \mathbf{F}'_{i,i} COV[\boldsymbol{\Theta}_{i,i}, \boldsymbol{\Theta}_{i,j} | D_i] \mathbf{F}_{i,j} \\
&+ \sum_{i=1}^k VAR[\varepsilon_{v,i} | D_i] + \sum_{i=1}^k \sum_{j \neq i}^k COV[\varepsilon_{v,i}, \varepsilon_{v,j} | D_i] \\
&+ \sum_{i=1}^k \sum_{j=1}^k \mathbf{F}'_{i,i} COV[\boldsymbol{\Theta}_{i,i}, \varepsilon_{v,j} | D_i] + \sum_{i=1}^k \sum_{j=1}^k COV[\varepsilon_{v,i}, \boldsymbol{\Theta}_{i,j} | D_i] \mathbf{F}_{i,j} \\
&= \sum_{i=1}^k \mathbf{F}'_{i,i} \mathbf{S}_{\boldsymbol{\Theta}_i}^2(i) \mathbf{F}_{i,i} + \sum_{i=1}^k \sum_{j \neq i}^k \mathbf{F}'_{i,i} \mathbf{S}_{\boldsymbol{\Theta}_i}(i, j) \mathbf{F}_{i,j} + \sum_{i=1}^k \sigma_{v,i}^2 \\
&= \sum_{i=1}^k \hat{\sigma}_i^2(i) + \sum_{i=1}^k \sum_{j \neq i}^k \sigma_i(i, j) + \sum_{i=1}^k \sigma_{v,i}^2
\end{aligned}$$

Corolario 4.3

Para el caso particular en que la matriz de transferencia es constante $\mathbf{G}_t = \mathbf{G}$ para toda t entonces para $k \geq 0$ se tiene que

$$\mathbf{m}_{\boldsymbol{\Theta}_i}(k) = \mathbf{G}^k \mathbf{m}_i + \sum_{i=1}^{k-1} \mathbf{G}^{k-i} \boldsymbol{\mu}_{w_{i,i}} + \boldsymbol{\mu}_{w_{i,i}}$$

y por lo tanto

$$\hat{Y}_i(k) = \mathbf{F}'_{i,k} \left[\mathbf{G}^k \mathbf{m}_i + \sum_{i=1}^{k-1} \mathbf{G}^{k-i} \boldsymbol{\mu}_{w_{i,i}} + \boldsymbol{\mu}_{w_{i,i}} \right] + \boldsymbol{\mu}_{v_{i,i}}$$

Si además la matriz de regresores es constante, $\mathbf{F}_t = \mathbf{F}$ para toda t entonces

$$\hat{Y}_i(k) = \mathbf{F}' \left[\mathbf{G}^k \mathbf{m}_i + \sum_{i=1}^{k-1} \mathbf{G}^{k-i} \boldsymbol{\mu}_{w_{i,i}} + \boldsymbol{\mu}_{w_{i,i}} \right] + \boldsymbol{\mu}_{v_{i,i}}$$

Demostración.

En la demostración del *teorema 4.2* se observó que

$$\mathbf{m}_{\boldsymbol{\Theta}_i}(k) = \mathbf{H}_{i,k}(k) \mathbf{m}_i + \sum_{r=1}^k \mathbf{H}_{i,k}(k-r) \boldsymbol{\mu}_{w_{i,i}}$$

pero como

$$G_t = G \quad \forall t$$

entonces

$$H_t(r) = G_t G_{t-1} \dots G_{t-r+1} = G^r,$$

por lo tanto

$$m_{\theta_t}(k) = G^k m_1 + \sum_{r=1}^{k-1} G^{k-r} \mu_{w_{1,r}} + \mu_{w_{1,k}}$$

Por otra parte se tiene por el mismo teorema que

$$\hat{Y}_t(k) = F'_{t+k} m_{\theta_t}(k) + \mu_{v_{t,k}}$$

Si $F_t = F$ para toda t entonces

$$\begin{aligned} \hat{Y}_t(k) &= F' m_{\theta_t}(k) + \mu_{v_{t,k}} \\ &= F' \left[G^k m_1 + \sum_{r=1}^{k-1} G^{k-r} \mu_{w_{1,r}} + \mu_{w_{1,k}} \right] + \mu_{v_{t,k}} \end{aligned}$$

4.3 Varianza Observacional Desconocida

Tanto en la definición de el caso general de los *MDL*, como en el caso Univariado se estableció que la sextupla $\{F_t, G_t, \mu_{Y_t}, \sigma_{Y_t}^2, \mu_{W_t}, \sigma_{W_t}^2\}$ es conocida para cualquier t . Las matrices F_t , G_t , μ_{Y_t} y μ_{W_t} son diseñadas por el modelador, la sucesión de matrices $\{\sigma_{W_t}^2\}$ puede ser escogida de acuerdo a un factor de descuento δ de forma similar a como se escoge en el *Modelo Lineal*. En cambio el modelador generalmente desconoce la forma de la sucesión $\{\sigma_{Y_t}^2\}$, pero puede tener diferentes opiniones acerca de ella. Sin embargo en este trabajo se considerará el caso más sencillo, en el que la varianza observacional es constante pero desconocida

$$\sigma_{Y_t}^2 = \sigma_V^2 \text{ para toda } t.$$

Similarmente al modelo lineal y retomando el concepto de *precisión* $\phi \equiv \frac{1}{\sigma_V^2}$, la metodología Bayesiana proporciona un proceso de aprendizaje sobre la varianza observacional desconocida:

- (1) Se considera que la precisión es una cantidad aleatoria Φ , y por lo tanto, se le puede asociar una función de distribución $F_\phi(\phi)$.
- (2) Se considera que todas las varianzas en la definición del modelo son multiples de σ_V^2 , o dicho de otra forma están *escaladas* por σ_V^2 .

Definición 4.5

Para cada t el *MDL Univariado* con varianza observacional desconocida se establece la relación entre Y_t y Θ_t a través de la *ecuación observacional*

$$Y_t = F_t' \Theta_t + \varepsilon_{Y_t} \quad \varepsilon_{Y_t} \sim N[\mu_{Y_t}, \sigma_V^2] \quad (4.1.b)$$

y la *ecuación de transición*

$$\Theta_t = G_t \Theta_{t-1} + \varepsilon_{W_t} \quad \varepsilon_{W_t} \sim N[\mu_{W_t}, \sigma_V^2 \sigma_{W_t}^2] \quad (4.2.b)$$

y la *información inicial*

$$\left(\Theta_0 | D_0, \Phi = \frac{1}{\sigma_V^2} \right) \sim N[\mathbf{m}_0, \sigma_V^2 \mathbf{S}_0^2] \quad (4.3.a)$$

$$(\Phi|D_0) \sim G\left[\frac{n_0}{2}, \frac{d_0}{2}\right] \quad (4.3.b)$$

Se debe recordar que se toma una distribución *Gamma* para Φ por permitir simular otras distribuciones así como por ser una distribución a priori conjugada. También se debe recalcar los siguiente

(i) El valor puntual estimado a priori para la precisión es

$$E[\Phi|D_0] = \frac{n_0}{d_0} = \frac{1}{S_v^2}$$

(ii) Se supone que S_0^2 y $\{\sigma_w^2\}$ son conocidas e independientes de la escala de las observaciones determinada por σ_v^2 y son tales que al multiplicarlas por σ_v^2 se obtienen las verdaderas varianzas S_i^2 y $\{\sigma_w^2\}$ cuando σ_v^2 es conocida.

Teorema 4.3

Con el modelo especificado en la definición 4.6, si $(\Theta_0|D_0, \Phi = \frac{1}{\sigma_v^2}) \sim N[m_0, \sigma_v^2 S_0^2]$ y

$(\Phi|D_0) \sim G\left[\frac{n_0}{2}, \frac{d_0}{2}\right]$, para cada $t \geq 1$ se tienen las siguientes distribuciones

(I) Condicionales sobre $\Phi = \frac{1}{\sigma_v^2}$:

(a) A priori para Θ_t

$$\left(\Theta_t | D_{t-1}, \Phi = \frac{1}{\sigma_v^2}\right) \sim N[\bar{m}_t, \sigma_v^2 \bar{S}_t^2]$$

donde

$$\bar{m}_t = G_t m_{t-1} + \mu_w$$

$$\bar{S}_t^2 = G_t \bar{S}_{t-1}^2 G_t' + \sigma_w^2$$

(b) Pronostico de un periodo

$$\left(f_{t-1}(1) | \Phi = \frac{1}{\sigma_v^2}\right) \sim N[\hat{Y}_t, \sigma_v^2 \hat{\sigma}_t^2]$$

donde

$$\hat{Y}_i = F_i \bar{m}_i + \mu_{\nu_i}$$

$$\hat{\sigma}_i^2 = 1 + F_i \bar{S}_i^2 F_i$$

(c) A posteriori para Θ_i

$$\left(\Theta_i \mid D_i, \Phi = \frac{1}{\sigma_i^2} \right) \sim N \left[m_i, \sigma_i^2 S_i^2 \right]$$

con

$$m_i = \bar{m}_i + A_i e_i$$

$$S_i^2 = \bar{S}_i^2 - A_i A_i' \hat{\sigma}_i^2$$

Donde

$$A_i = \frac{\bar{S}_i^2 F_i}{\hat{\sigma}_i^2}$$

$$e_i = y_i - \hat{Y}_i$$

(II) Para la precisión Φ :

(a) Distribución a priori

$$(\phi \mid D_{i-1}) \sim G \left[n_{i-1}/2, d_{i-1}/2 \right]$$

(b) Distribución a posteriori

$$(\phi \mid D_i) \sim G \left[n_i/2, d_i/2 \right]$$

con

$$n_i = n_{i-1} + 1$$

$$d_i = d_{i-1} + \frac{e_i^2}{\hat{\sigma}_i^2}$$

(III) Incondicionales sobre $\Phi = \frac{1}{\sigma_v^2}$:

(a) $(\Theta_i | D_{i-1}) \sim T_{n_i}[\bar{m}_i, \bar{S}_i^2]$.

(b) $f_{i-1}(1) \sim T_{n_i}[\hat{\gamma}_i, \hat{\sigma}_i^2]$.

(c) $(\Theta_i | D_i) \sim T_{n_i}[m_i, S_i^2]$.

Con

$$\bar{S}_i^2 = S_{v,i-1}^2 \bar{S}_i^{2*}$$

$$\hat{\sigma}_i^2 = S_{v,i-1}^2 \hat{\sigma}_i^{2*}$$

$$S_i^2 = S_{v,i}^2 S_i^{2*}$$

donde

$$S_{v,i}^2 = d_i / n_i$$

(IV) Las ecuaciones recurrentes de actualización se definen de la siguiente forma

$$m_i = \bar{m}_i + A_i e_i$$

$$S_i^2 = \left(\frac{S_{v,i}^2}{S_{v,i-1}^2} \right) [\bar{S}_i^2 - A_i A_i' \hat{\sigma}_i^2]$$

$$S_{v,i}^2 = \frac{d_i}{n_i}$$

$$n_i = n_{i-1} + 1$$

$$d_i = d_{i-1} + \frac{S_{v,i-1}^2 e_i^2}{\hat{\sigma}_i^2}$$

donde

$$\hat{\sigma}_i^2 = S_{v,i-1}^2 + F_i' \bar{S}_i^2 F_i$$

$$A_i = \frac{\bar{S}_i^2 F_i}{\hat{\sigma}_i^2}$$

Demostración.

La sección (I) es inmediata del teorema 4.1.

La sección (II) se demuestra por inducción.

Suponemos que se cumple (a), es decir

$$(\Phi|D_{i-1}) \sim G\left[n_{i-1}/2, d_{i-1}/2\right]$$

siendo la densidad a priori

$$f_{\Phi}(\phi) = \frac{\left(\frac{d_{i-1}}{2}\right)^{\frac{n_{i-1}}{2}} \phi^{\frac{n_{i-1}}{2}-1} \exp\left\{-\phi \frac{d_{i-1}}{2}\right\}}{\Gamma\left(\frac{n_{i-1}}{2}\right)} \propto \phi^{\frac{n_{i-1}}{2}-1} \exp\left\{-\phi \frac{d_{i-1}}{2}\right\}$$

De acuerdo a la sección (I)

$$(Y_i|D_{i-1}, \Phi = \phi) \sim N\left[\hat{Y}_i, \frac{\hat{\sigma}_i^2}{\phi}\right]$$

en consecuencia

$$f_{Y_i|D_{i-1}}(y_i|\phi) \propto \phi^{\frac{1}{2}} \exp\left\{-\frac{\phi(y_i - \hat{Y}_i)^2}{2\hat{\sigma}_i^2}\right\}$$

Por el teorema de Bayes

$$\begin{aligned} f_{\Phi Y_i|D_{i-1}}(\phi|y_i) &\propto \phi^{\frac{1}{2}} \exp\left\{-\frac{\phi(y_i - \hat{Y}_i)^2}{2\hat{\sigma}_i^2}\right\} \phi^{\frac{n_{i-1}}{2}-1} \exp\left\{-\phi \frac{d_{i-1}}{2}\right\} \\ &\propto \phi^{\frac{n_{i-1}+1}{2}-1} \exp\left\{-\frac{\phi}{2}\left[d_{i-1} + \frac{e_i^2}{\hat{\sigma}_i^2}\right]\right\} \end{aligned}$$

deduciéndose que

$$(\Phi|D_i) \sim G\left[n_i/2, d_i/2\right]$$

con $n_i = n_{i-1} + 1$ y $d_i = d_{i-1} + \frac{e_i^2}{\hat{\sigma}_i^2}$.

Como (a) también se cumple para $t = 0$, queda demostrada la sección (III).

La sección (III) se demuestra directamente de las propiedades de las distribuciones *Normal-Gamma* y *t-Student* que son enunciadas en la sección A.4 del Apéndice.

Las ecuaciones de la sección (IV) son directas de la sección (III) con la diferencia que las varianzas y covarianzas en las distribuciones *t-Student* incluyen la estimación de σ_v^2 .

■

4.3.1 Distribuciones de Pronóstico con Varianza Observacional Desconocida

Similarmente para la forma de las distribuciones de pronóstico se da el siguiente

Teorema 4.4

Con el modelo especificado en la definición 4.6, si $(\Phi | D_t) \sim G\left[\frac{n_t}{2}, \frac{d_t}{2}\right]$ como en el teorema 4.3 en cada t y para $k \geq 1$ se tienen las siguientes distribuciones con momentos definidos recursivamente

(I) Condicionales sobre σ_v^2 :

(a) Pronóstico parametral de horizonte k

$$\left(\Theta_{t+k} \left| D_t, \Phi = \frac{I}{\sigma_v^2} \right. \right) \sim N\left[\mathbf{m}_{\Theta_t}(k), \sigma_v^2 \mathbf{S}_{\Theta_t}^*(k) \right]$$

con

$$\mathbf{m}_{\Theta_t}(k) = \mathbf{G}_{t+k} \mathbf{m}_{\Theta_t}(k-1) + \mu_{w_{t+k}},$$

$$\mathbf{S}_{\Theta_t}^*(k) = \mathbf{G}_{t+k} \mathbf{S}_{\Theta_t}^*(k-1) \mathbf{G}_{t+k}' + \sigma_{w_{t+k}}^2.$$

Con valores iniciales $\mathbf{m}_{\Theta_t}(0) = \mathbf{m}_t$, $\mathbf{S}_{\Theta_t}^*(0) = \mathbf{S}_t^*$.

(b) Pronóstico de k periodos en el futuro

$$(f_t(k) | \sigma_v^2) \sim N\left[\hat{Y}_t(k), \sigma_v^2 \hat{\sigma}_t^2(k) \right],$$

con

$$\hat{Y}_t(k) = \mathbf{F}_{t+k}' \mathbf{m}_{\Theta_t}(k) + \mu_{v_{t+k}},$$

$$\hat{\sigma}_i^2(k) = \mathbf{F}'_{i,t} \mathbf{S}_{\theta_i}^2(k) \mathbf{F}_{i,t} +$$

(II) Incondicionales sobre $\Phi = \frac{1}{\sigma_v^2}$:

(a) $(\Theta_{i,t} | D_i) \sim T_n[\mathbf{m}_{\theta_i}(k), \mathbf{S}_{\theta_i}^2(k)].$

(b) $f_i(k) \sim T_n[\hat{Y}_i(k), \hat{\sigma}_i^2(k)].$

con

$$\mathbf{S}_{\theta_i}^2(k) = S_{v_i}^2 \mathbf{S}_{\theta_i}^2(k),$$

$$\hat{\sigma}_i^2(k) = S_{v_i}^2 \hat{\sigma}_i^2(k)$$

con $S_{v_i}^2$ como en el teorema 4.3.

Demostración.

La sección (I) es inmediata del teorema 4.2. La sección (II) se demuestra directamente de propiedades de las distribuciones *Normal-Gamma* y *t-Student* que son enunciadas en sección A.4 del Apéndice. ■

4.4 Distribuciones de los errores

Con el fin de mejorar el conocimiento sobre las sucesiones $\{\varepsilon_{v_i}\}$ y $\{\varepsilon_{w_i}\}$ es importante identificar las distribuciones posteriores de los errores observacional y de transición, para lo cual se establece el siguiente

Teorema 4.5

Si para cada t se tiene que $\varepsilon_{v_i} \sim N[\mu_{v_i}, \sigma_{v_i}^2]$ y $\varepsilon_{w_i} \sim N[\mu_{w_i}, \sigma_{w_i}^2]$ entonces se tienen las siguientes distribuciones

(a) A posteriori para ε_{v_i}

$$(\varepsilon_{v_i} | D_i) \sim N[\bar{\mu}_{v_i}, \bar{\sigma}_{v_i}^2]$$

donde

$$\bar{\mu}_{v_i} = \mu_{v_i} + \left(\frac{\sigma_{v_i}^2}{\hat{\sigma}_i^2} \right) e_i,$$

$$\bar{\sigma}_{v_i}^2 = \sigma_{v_i}^2 - \frac{(\hat{\sigma}_{v_i}^2)^2}{\hat{\sigma}_i^2}$$

(b) *A posteriori para ε_w*

$$(\varepsilon_w | D_t) \sim N[\bar{\mu}_w, \bar{\sigma}_w^2],$$

donde

$$\bar{\mu}_w = \mu_w + \frac{\sigma_w^2 \mathbf{F}_t' e_t}{\hat{\sigma}_i^2},$$

$$\bar{\sigma}_w^2 = \sigma_w^2 - \frac{\sigma_w^2 \mathbf{F}_t' \mathbf{F}_t \sigma_w^2}{\hat{\sigma}_i^2}$$

Demostración.

Debido a que cualquier función lineal de $Y_t, \varepsilon_{v_t}, \varepsilon_w$, define una combinación lineal de variables normales, tal función lineal condicionada sobre D_{t-1} también se distribuye de forma normal y por definición las distribuciones conjuntas $(Y_t, \varepsilon_{v_t} | D_{t-1})$ y $(Y_t, \varepsilon_w | D_{t-1})$ son normales bivariadas.

(a) De acuerdo a lo establecido en el modelo

$$\begin{aligned} \text{COV}[Y_t, \varepsilon_{v_t} | D_{t-1}] &= \text{COV}[\mathbf{F}_t' \Theta_t + \varepsilon_{v_t}, \varepsilon_{v_t} | D_{t-1}] \\ &= \mathbf{F}_t' \text{COV}[\Theta_t, \varepsilon_{v_t} | D_{t-1}] + \text{VAR}[\varepsilon_{v_t} | D_{t-1}] \\ &= \sigma_{v_t}^2 \end{aligned}$$

por lo tanto

$$\begin{pmatrix} Y_t \\ \varepsilon_{v_t} \end{pmatrix} | D_{t-1} \sim N \left[\begin{pmatrix} \hat{Y}_t \\ \mu_{v_t} \end{pmatrix}, \begin{pmatrix} \hat{\sigma}_i^2 & \hat{\sigma}_{v_t}^2 \\ \hat{\sigma}_{v_t}^2 & \hat{\sigma}_{v_t}^2 \end{pmatrix} \right]$$

(b) Similarmente

$$\begin{aligned}
\text{COV}[Y_t, \boldsymbol{\varepsilon}_{w_t} | D_{t-1}] &= \text{COV}[\mathbf{F}_t'(\boldsymbol{\Theta}_{t-1} + \boldsymbol{\varepsilon}_{w_t}) + \varepsilon_{v_t}, \boldsymbol{\varepsilon}_{w_t} | D_{t-1}] \\
&= \mathbf{F}_t' \text{COV}[\boldsymbol{\Theta}_{t-1} + \boldsymbol{\varepsilon}_{w_t}, \boldsymbol{\varepsilon}_{w_t} | D_{t-1}] + \text{COV}[\varepsilon_{v_t}, \boldsymbol{\varepsilon}_{w_t} | D_{t-1}] \\
&= \mathbf{F}_t' (\text{COV}[\boldsymbol{\Theta}_{t-1}, \boldsymbol{\varepsilon}_{w_t} | D_{t-1}] + \text{VAR}[\boldsymbol{\varepsilon}_{w_t} | D_{t-1}]) = \mathbf{F}_t' \boldsymbol{\sigma}_{w_t}^2
\end{aligned}$$

por lo tanto

$$\begin{pmatrix} Y_t \\ \boldsymbol{\varepsilon}_{w_t} \end{pmatrix} | D_{t-1} \sim N \left[\begin{pmatrix} \hat{Y}_t \\ \boldsymbol{\mu}_{w_t} \end{pmatrix}, \begin{pmatrix} \hat{\sigma}_t^2 & \mathbf{F}_t' \boldsymbol{\sigma}_{w_t}^2 \\ \boldsymbol{\sigma}_{w_t}^2 \mathbf{F}_t & \boldsymbol{\sigma}_{w_t}^2 \end{pmatrix} \right]$$

De acuerdo a la sección A.1.2 del Apéndice se tiene las distribuciones condicionales

(a) $(\varepsilon_{v_t} | Y_t = y_t, D_{t-1}) \sim N[\bar{\mu}_{v_t}, \bar{\sigma}_{v_t}^2]$

con

$$\bar{\mu}_{v_t}(y_t) = \mu_{v_t} + \left(\frac{\sigma_{v_t}^2}{\hat{\sigma}_t^2} \right) (y_t - \hat{Y}_t)$$

$$\bar{\sigma}_{v_t}^2(y_t) = \sigma_{v_t}^2 - \frac{(\sigma_{v_t}^2)^2}{\hat{\sigma}_t^2}$$

(b) $(\boldsymbol{\varepsilon}_{w_t} | Y_t = y_t, D_{t-1}) \sim N[\bar{\boldsymbol{\mu}}_{w_t}, \bar{\boldsymbol{\sigma}}_{w_t}^2]$

con

$$\bar{\boldsymbol{\mu}}_{w_t}(y_t) = \boldsymbol{\mu}_{w_t} + \frac{(y_t - \hat{Y}_t)}{\hat{\sigma}_t^2} \boldsymbol{\sigma}_{w_t}^2 \mathbf{F}_t$$

$$\bar{\boldsymbol{\sigma}}_{w_t}^2(y_t) = \boldsymbol{\sigma}_{w_t}^2 - \frac{\boldsymbol{\sigma}_{w_t}^2 \mathbf{F}_t \mathbf{F}_t' \boldsymbol{\sigma}_{w_t}^2}{\hat{\sigma}_t^2}$$

4.5 Filtrado

Con fines de control y monitoreo, a veces se requiere de hacer inferencias sobre el estado del proceso en el pasado utilizando la información reciente y a esta acción se le denomina *filtrado*.

Teorema 4.6

En el MDL univariado $\{F, G, \mu_v, \sigma_v^2, \mu_w, \sigma_w^2\}$ para cada t se define

$$B_t = S_t^2 G'_{t+1} (\bar{S}_{t+1}^2)^{-1}$$

Entonces para toda k tal que $t > k \geq 1$ las distribuciones filtro están dadas por

$$(\Theta_{t-k} | D_t) \sim N[m_{\Theta_t}(-k), S_{\Theta_t}^2(-k)],$$

donde

$$m_{\Theta_t}(-k) = m_{t-k} + B_{t-k} [m_{\Theta_t}(-k+1) - \bar{m}_{t-k+1}]$$

con valor inicial

$$m_{\Theta_t}(0) = m_t,$$

y

$$S_{\Theta_t}^2(-k) = S_{t-k}^2 - B_{t-k} [\bar{S}_{t-k+1}^2 - S_{\Theta_t}^2(-k+1)] B'_{t-k},$$

con

$$S_{\Theta_t}^2(0) = S_t^2.$$

Así mismo se tiene que

$$\bar{m}_{t-k+1} = m_{\Theta_{t-k}}(1) \quad \text{y} \quad \bar{S}_{t-k+1}^2 = S_{\Theta_{t-k}}^2(1).$$

Demostración.

Las distribuciones filtro están definidas recursivamente por

FALTA PAGINA

No. 1 3 2

$$f_{\Theta_{t-k}}(\Theta_{t-k}|D_t) = \int f_{\Theta_{t-k}|\Theta_{t-k+1}, D_t}(\Theta_{t-k}|\Theta_{t-k+1}) f_{\Theta_{t-k+1}|D_t}(\Theta_{t-k+1}) d\Theta_{t-k+1}, \quad (4.9)$$

sugiriendo que el teorema se demuestra por inducción. Se supone que el teorema es válido para $k - 1$, esto es

$$(\Theta_{t-k+1}|D_t) \sim N[\mathbf{m}_{\Theta_{t-k+1}}(-k+1), \mathbf{S}_{\Theta_{t-k+1}}^2(-k+1)].$$

Por el teorema de Bayes y puesto que $D_t = \{D_{t-k}, y_{t-k+1}, \dots, y_t\}$, el primer término en el integrando es

$$f_{\Theta_{t-k}|\Theta_{t-k+1}, D_t}(\Theta_{t-k}|\Theta_{t-k+1}) = \frac{f_{\Theta_{t-k}|\Theta_{t-k+1}, D_{t-k}}(\Theta_{t-k}|\Theta_{t-k+1}) f_{y_{t-k+1}, \dots, y_t|\Theta_{t-k}, \Theta_{t-k+1}, D_{t-k}}(y_{t-k+1}, \dots, y_t|\Theta_{t-k}, \Theta_{t-k+1})}{f_{y_{t-k+1}, \dots, y_t|\Theta_{t-k+1}, D_{t-k}}(y_{t-k+1}, \dots, y_t|\Theta_{t-k+1})}$$

Dado $\Theta_{t-k+1} = \Theta_{t-k+1}$, el vector (y_{t-k+1}, \dots, y_t) es independiente de Θ_{t-k} por lo que

$$f_{\Theta_{t-k}|\Theta_{t-k+1}, D_t}(\Theta_{t-k}|\Theta_{t-k+1}) = f_{\Theta_{t-k}|\Theta_{t-k+1}, D_{t-k}}(\Theta_{t-k}|\Theta_{t-k+1}).$$

Nuevamente por el teorema de Bayes se tiene

$$f_{\Theta_{t-k}|\Theta_{t-k+1}, D_{t-k}}(\Theta_{t-k}|\Theta_{t-k+1}) \propto f_{\Theta_{t-k}|D_{t-k}}(\Theta_{t-k}) f_{\Theta_{t-k+1}|\Theta_{t-k}, D_{t-k}}(\Theta_{t-k+1}|\Theta_{t-k}) \quad (4.10)$$

pero

$$(\Theta_{t-k+1}|\Theta_{t-k} = \Theta_{t-k}, D_{t-k}) \sim N[\mathbf{G}_{t-k+1}\Theta_{t-k} + \boldsymbol{\mu}_{W_{t-k+1}}, \boldsymbol{\Sigma}_{W_{t-k+1}}^2]$$

y

$$(\Theta_{t-k}|D_{t-k}) \sim N[\mathbf{m}_{t-k}, \mathbf{S}_{t-k}^2].$$

Como

$$\begin{aligned} COV[\Theta_{t-k}, \Theta_{t-k+1}|D_{t-k}] &= COV[\Theta_{t-k}, \mathbf{G}_{t-k+1}\Theta_{t-k} + \boldsymbol{\varepsilon}_{W_{t-k+1}}|D_{t-k}] \\ &= \mathbf{G}_{t-k+1} VAR[\Theta_{t-k}|D_{t-k}] + COV[\Theta_{t-k}, \boldsymbol{\varepsilon}_{W_{t-k+1}}|D_{t-k}] \\ &= \mathbf{G}_{t-k+1} \mathbf{S}_{t-k}^2 \end{aligned}$$

la distribución conjunta bivariada para $\Theta_{t-k}, \Theta_{t-k+1}$ dado D_{t-k} está dada por

$$\begin{pmatrix} \Theta_{t-k} \\ \Theta_{t-k+1} \end{pmatrix} | D_{t-k} \sim N \left[\begin{pmatrix} \mathbf{m}_{t-k} \\ \mathbf{G}_{t-k+1} \mathbf{m}_{t-k} + \boldsymbol{\mu}_{W_{t-k+1}} \end{pmatrix}, \begin{pmatrix} \mathbf{S}_{t-k}^2 & \mathbf{G}_{t-k+1} \mathbf{S}_{t-k}^2 \\ \mathbf{S}_{t-k}^2 \mathbf{G}'_{t-k+1} & \bar{\mathbf{S}}_{t-k+1}^2 \end{pmatrix} \right]$$

por lo que utilizando los resultados de la sección A.1.2 del Apéndice se tiene la distribución condicional

$$(\Theta_{t-k} | \Theta_{t-k+1} = \theta_{t-k+1}, D_{t-k}) \sim N[\mathbf{h}_t(k), \mathbf{H}_t(k)] \quad (4.11)$$

con

$$\begin{aligned} \mathbf{h}_t(k) &= \mathbf{m}_{t-k} + \mathbf{S}_{t-k}^2 \mathbf{G}'_{t-k+1} (\bar{\mathbf{S}}_{t-k+1}^2)^{-1} (\theta_{t-k+1} - (\mathbf{G}_{t-k+1} \mathbf{m}_{t-k} + \boldsymbol{\mu}_{\eta_{t-k+1}})) \\ &= \mathbf{m}_{t-k} + \mathbf{S}_{t-k}^2 \mathbf{G}'_{t-k+1} (\bar{\mathbf{S}}_{t-k+1}^2)^{-1} (\theta_{t-k+1} - \bar{\mathbf{m}}_{t-k+1}) \end{aligned}$$

y

$$\mathbf{H}_t(k) = \mathbf{S}_{t-k}^2 - \mathbf{S}_{t-k}^2 \mathbf{G}'_{t-k+1} (\bar{\mathbf{S}}_{t-k+1}^2)^{-1} \mathbf{G}_{t-k+1} \mathbf{S}_{t-k}^2$$

Con la definición de \mathbf{B}_t establecida en el teorema, se tiene

$$\mathbf{h}_t(k) = \mathbf{m}_{t-k} + \mathbf{B}_{t-k} (\theta_{t-k+1} - \bar{\mathbf{m}}_{t-k+1}),$$

$$\mathbf{H}_t(k) = \mathbf{S}_{t-k}^2 - \mathbf{B}_{t-k} \bar{\mathbf{S}}_{t-k+1}^2 \mathbf{B}'_{t-k}.$$

Regresando a (4.9), la densidad $f_{\Theta_{t-k}|D_t}(\theta_{t-k})$ es el valor esperado de (4.11) con respecto a $(\Theta_{t-k+1}|D_t)$, cuya forma, por hipótesis, cumple con lo establecido en el teorema para $k - 1$, por lo que

$$(\Theta_{t-k}|D_t) \sim N[\mathbf{m}_{\Theta_t}(-k), \mathbf{S}_{\Theta_t}(-k)]$$

donde

$$\begin{aligned} \mathbf{m}_{\Theta_t}(-k) &= E[\mathbf{h}_t(k)|D_t] \\ &= \mathbf{m}_{t-k} + \mathbf{B}_{t-k} [\mathbf{m}_{\Theta_t}(-k+1) - \bar{\mathbf{m}}_{t-k+1}] \end{aligned}$$

y

$$\begin{aligned} \mathbf{S}_{\Theta_t}(-k) &= E[\mathbf{H}_t(k)|D_t] + \text{VAR}[\mathbf{h}_t(k)|D_t] \\ &= \mathbf{S}_{t-k}^2 - \mathbf{B}_{t-k} [\bar{\mathbf{S}}_{t-k+1}^2 \mathbf{S}_{\Theta_t}^2(-k+1)] \mathbf{B}'_{t-k} \end{aligned}$$

cumpliendo el teorema para k . Para completar la demostración se debe probar que el teorema es válido para $k = 1$, pero en tal caso $t - k + 1 = t$ y el desarrollo anterior se aplica con $\mathbf{m}_{\theta, (-k+1)} = \mathbf{m}_{\theta, (0)} = \mathbf{m}$, y $\mathbf{S}_{\theta, (-k+1)}^2 = \mathbf{S}_{\theta, (0)}^2 = \mathbf{S}_t^2$.

■

Para el caso en que la varianza σ_v^2 es constante para toda t , pero desconocida, si se utiliza el análisis de la sección 4.3 se pueden construir distribuciones filtro utilizando el siguiente

Corolario 4.4

Si $\sigma_v^2 = \sigma_v^2 \equiv \frac{1}{\phi}$ es desconocida y si es aplicada la rutina de aprendizaje proporcionada por el teorema 4.3, entonces se tienen las distribuciones filtro

(I) Condicional sobre $\Phi = \frac{1}{\sigma_v^2}$

$$\left(\Theta_{t-k} \mid \Phi = \frac{1}{\sigma_v^2}, D_t \right) \sim N \left[\mathbf{m}_{\theta, (-k)}, \sigma_v^2 \mathbf{S}_{\theta, (-k)}^2 \right],$$

con

$$\mathbf{S}_{\theta, (-k)}^2 = \mathbf{S}_{t-k}^2 - \mathbf{B}_{t-k} \left[\bar{\mathbf{S}}_{t-k+1}^2 - \mathbf{S}_{\theta, (-k+1)}^2 \right] \mathbf{B}_{t-k}'$$

$$\mathbf{S}_{\theta, (0)}^2 = \mathbf{S}_t^2.$$

(II) Incondicional sobre $\Phi = \frac{1}{\sigma_v^2}$

$$\left(\Theta_{t-k} \mid D_t \right) \sim T_{\eta} \left[\mathbf{m}_{\theta, (-k)}, \left(\frac{S_{v_t}^2}{S_{v_{t-k}}^2} \right) \mathbf{S}_{\theta, (-k)}^2 \right]$$

Demostración.

La sección (I) es inmediata del teorema 4.6.

La sección (II) se demuestra a partir de las propiedades de las distribuciones *Normal-Gamma* y *t-Student* que son enunciadas en la sección A.4 del Apéndice y el factor $\left(\frac{S_{V_t}^2}{S_{V_{t-k}}^2}\right)$ es para corregir la escala.



Un concepto relacionado con el filtrado es el de *suavizamiento* de la serie, que consiste en la estimación retrospectiva del nivel de la serie μ , utilizando las distribuciones filtro.

Corolario 4.5

La distribución *suavizada* para el nivel de la serie condicional sobre $\Phi = \frac{1}{\sigma_V^2}$ está dada por

$$\left(\mu_{t-k} \mid \Phi = \frac{1}{\sigma_V^2}, D_t\right) \sim N\left[\hat{\mu}_t(-k), \mathbf{F}'_{t-k} \mathbf{S}_{\Theta_t}^2(-k) \mathbf{F}_{t-k}\right],$$

e incondicional sobre $\Phi = \frac{1}{\sigma_V^2}$

$$(\mu_{t-k} \mid D_t) \sim T_{n_t}\left[\hat{\mu}_t(-k), \left(\frac{S_{V_t}^2}{S_{V_{t-k}}^2}\right) \mathbf{F}'_{t-k} \mathbf{S}_{\Theta_t}^2(-k) \mathbf{F}_{t-k}\right],$$

con

$$\hat{\mu}_t(-k) = \mathbf{F}'_{t-k} \mathbf{m}_{\Theta_t}(-k).$$

Demostración.

De acuerdo a la *definición 4.3* se tiene

$$\begin{aligned} E\left[\mu_{t-k} \mid \Phi = \frac{1}{\sigma_V^2}, D_t\right] &= E\left[\mathbf{F}'_{t-k} \Theta_{t-k} \mid \Phi = \frac{1}{\sigma_V^2}, D_t\right] \\ &= \mathbf{F}'_{t-k} E\left[\Theta_{t-k} \mid \Phi = \frac{1}{\sigma_V^2}, D_t\right] = \mathbf{F}'_{t-k} \mathbf{m}_{\Theta_t}(-k) \\ &= \hat{\mu}_t(-k), \end{aligned}$$

$$\begin{aligned}
 AR \left[\mu_{t-k} \mid \Phi = \frac{1}{\sigma_v^2}, D_t \right] &= VAR \left[\mathbf{F}'_{t-k} \Theta_{t-k} \mid \Phi = \frac{1}{\sigma_v^2}, D_t \right] \\
 &= \mathbf{F}'_{t-k} VAR \left[\Theta_{t-k} \mid \Phi = \frac{1}{\sigma_v^2}, D_t \right] \mathbf{F}_{t+k} \\
 &= \mathbf{F}'_{t-k} \mathbf{S}_{\Theta_t}^2 (-k) \mathbf{F}_{t-k}
 \end{aligned}$$

Por lo tanto

$$\left(\mu_{t-k} \mid \Phi = \frac{1}{\sigma_v^2}, D_t \right) \sim N \left[\hat{\mu}_t(-k), \mathbf{F}'_{t-k} \mathbf{S}_{\Theta_t}^2 (-k) \mathbf{F}_{t-k} \right]$$

De acuerdo a las propiedades de las distribuciones *Normal-Gamma* y *t-Student* que son enunciadas en la sección A.4 del Apéndice se tiene que

$$\left(\mu_{t-k} \mid D_t \right) \sim T_n \left[\hat{\mu}_t(-k), \left(\frac{S_{v_t}^2}{S_{v_{t-k}}^2} \right) \mathbf{F}'_{t-k} \mathbf{S}_{\Theta_t}^2 (-k) \mathbf{F}_{t-k} \right]$$

y el factor $\left(\frac{S_{v_t}^2}{S_{v_{t-k}}^2} \right)$ es para corregir la escala.

Ejemplo

En este ejemplo se retoma el problema de la demanda real por dinero, planteado en la sección 2.3 del segundo capítulo. La ecuación paramétrica que representa la demanda real por dinero en el mes t es la siguiente

$$(m-p)_t = \sum_{i=1}^{12} \gamma_i d_i + \alpha_1 y_t + \alpha_2 i_t + \alpha_3 m p_{t-1} \quad (4.12)$$

donde

$$(m-p)_t = \ln\left(\frac{M_t}{P_t}\right) = \ln(M_t) - \ln(P_t)$$

$$y_t = \ln(Y_t)$$

$$i_t = \ln\left(1 - \frac{I_t}{12}\right)$$

d_i = dummy estacional del mes i .

Se escoge un *MDL* para describir el comportamiento de la serie $\{(m-p)_t\}$ y realizar pronósticos. En este punto se presenta el primer problema al que un analista se enfrenta: ¿Cuáles datos están disponibles? ¿Qué variables pueden representar a la demanda de saldos reales en la economía o al ingreso nacional? Una solución es tomar variables proxy. En este problema se tomaron los siguientes datos

M_t = Billetes y Monedas totales (en millones de NS).

P_t = Índice Nacional de Precios al Consumidor (Base 1979 = 100).

I_t = Tasa de Cetes a 28 días.

Y_t = Índice de la Producción Industrial (Base 1980 = 100).

Bajo la estructura de los modelos dinámicos el vector de regresores y el vector paramétrico al tiempo t están determinados por

$$F_t = \begin{pmatrix} d_1 \\ \vdots \\ d_{12} \\ y_t \\ i_t \\ mp_{t-1} \end{pmatrix} \quad \Theta_t = \begin{pmatrix} \theta_{11} = \gamma_{11} \\ \vdots \\ \theta_{12t} = \gamma_{12t} \\ \theta_{13t} = \alpha_{11} \\ \theta_{14t} = \alpha_{21} \\ \theta_{15t} = \alpha_{31} \end{pmatrix}$$

En las siguientes tablas se presentan los valores de los regresores correspondientes al periodo comprendido entre enero de 1987 y marzo de 1994 y en la tabla 4.4, se indica el mes correspondiente a cada variable dummy

	$(m - p)_t$			
	1987	1988	1989	1990
ENE	-0.34198	-0.47909	-0.27190	-0.18548
FEB	-0.37703	-0.52308	-0.28956	-0.21150
MAR	-0.42550	-0.42222	-0.26112	-0.20197
ABR	-0.40941	-0.44397	-0.25928	-0.19634
MAY	-0.41161	-0.41417	-0.26425	-0.18619
JUN	-0.46019	-0.33680	-0.25070	-0.18419
JUL	-0.44291	-0.32078	-0.24463	-0.20855
AGO	-0.47603	-0.36014	-0.27716	-0.22446
SEP	-0.53905	-0.37641	-0.27313	-0.23104
OCT	-0.47486	-0.34806	-0.22854	-0.17369
NOV	-0.38609	-0.21464	-0.17935	-0.12710
DEC	-0.27773	-0.10607	0.01705	0.07533
	1991	1992	1993	1994
ENE	-0.07937	0.05266	0.08468	0.13662
FEB	-0.10114	0.02270	0.04851	0.10170
MAR	-0.04034	-0.02546	0.01133	0.15416
ABR	-0.11148	0.01655	0.01757	
MAY	-0.06467	0.04133	0.04012	
JUN	-0.05093	0.02036	0.03321	
JUL	-0.06433	0.03674	0.04891	
AGO	-0.06687	0.00083	-0.00328	
SEP	-0.08546	-0.04761	-0.01037	
OCT	-0.05107	-0.01327	0.04135	
NOV	-0.00531	0.02740	0.06605	
DEC	0.19268	0.22966	0.26702	

Tabla 4.1. Logaritmo del cociente de los Billetes y Monedas en circulación entre el INPC.

Fuente: Banco de México.

	i_t			
	1987	1988	1989	1990
<i>ENE</i>	0.07735	0.12342	0.04148	0.03377
<i>FEB</i>	0.07749	0.12042	0.04016	0.03692
<i>MAR</i>	0.07619	0.07773	0.03906	0.03819
<i>ABR</i>	0.07467	0.05167	0.04091	0.03637
<i>MAY</i>	0.07346	0.04400	0.04232	0.03021
<i>JUN</i>	0.07361	0.04400	0.04615	0.02663
<i>JUL</i>	0.07319	0.03305	0.03904	0.02490
<i>AGO</i>	0.07257	0.03397	0.02864	0.02448
<i>SEP</i>	0.07225	0.03425	0.02818	0.02481
<i>OCT</i>	0.07217	0.03656	0.03107	0.02456
<i>NOV</i>	0.08289	0.04082	0.03181	0.02074
<i>DEC</i>	0.09716	0.04267	0.03314	0.02139
	1991	1992	1993	1994
<i>ENE</i>	0.01950	0.01271	0.01384	0.00863
<i>FEB</i>	0.01915	-0.01200	0.01468	0.00789
<i>MAR</i>	0.01819	0.00977	0.01447	0.00776
<i>ABR</i>	0.01764	0.01034	0.01337	
<i>MAY</i>	0.01643	0.01124	0.01244	
<i>JUN</i>	0.01463	0.01250	0.01283	
<i>JUL</i>	0.01528	0.01339	0.01143	
<i>AGO</i>	0.01380	0.01399	0.01132	
<i>SEP</i>	0.01448	0.01444	0.01135	
<i>OCT</i>	0.01477	0.01602	0.01085	
<i>NOV</i>	0.01375	0.01500	0.01185	
<i>DEC</i>	0.01377	0.01394	0.00977	

Tabla 4.2. Logaritmo del monto obtenido de invertir 1 peso a la tasa de cetes a 28 días.

Fuente Banco de México.

	Y_t			
	1987	1988	1989	1990
ENE	4.61245	4.68014	4.73166	4.76532
FEB	4.60013	4.67240	4.70737	4.73877
MAR	4.68965	4.70146	4.72555	4.81706
ABR	4.64907	4.68450	4.76515	4.75458
MAY	4.69123	4.68910	4.75959	4.80650
JUN	4.67373	4.68951	4.75151	4.79725
JUL	4.66246	4.63635	4.72041	4.78093
AGO	4.67975	4.69987	4.76009	4.81319
SEP	4.67494	4.66581	4.71459	4.77029
OCT	4.70900	4.70236	4.75320	4.83964
NOV	4.71390	4.74783	4.79162	4.84884
DEC	4.73714	4.73051	4.74283	4.81382
	1991	1992	1993	1994
ENE	4.83227	4.84519	4.86330	4.88701
FEB	4.78769	4.83356	4.82857	4.85227
MAR	4.80238	4.89238	4.91846	4.94216
ABR	4.86060	4.83793	4.86053	
MAY	4.84677	4.87236	4.87061	
JUN	4.82954	4.87192	4.87975	
JUL	4.82013	4.86708	4.81705	
AGO	4.83190	4.85719	4.82112	
SEP	4.80909	4.86340	4.87384	
OCT	4.87662	4.88335	4.90008	
NOV	4.88650	4.88521	4.90231	
DEC	4.84855	4.89676	4.91339	

Tabla 4.3. Logaritmo del Índice de la Producción Industrial.

Fuente: Banco de México.

i	$d_i =$	$d_i = 0$
1	Mes de Semana Santa	otro caso
2	Febrero	"
3	Marzo	"
4	Abril	"
5	Mayo	"
6	Junio	"
7	Julio	"
8	Agosto	"
9	Septiembre	"
10	Octubre	"
11	Noviembre	"
12	Diciembre	"

Tabla 4.4. Variables Dummy.

Para completar la especificación del modelo se deben resolver los siguientes problemas:

- (i) ¿Cuál es la forma de la sucesión $\{G_t\}$?
- (ii) ¿Cómo determinar la forma de las varianzas $\{\sigma_{w_t}^2\}$ y $\{\sigma_{v_t}^2\}$ y los valores de $\{\mu_{w_t}\}$ y $\{\mu_{v_t}\}$?
- (iii) ¿Qué información inicial utilizar?

Dinámica

Los anteriores planteamientos están ligados directamente a la forma en que evolucionan los parámetros, la elección de tales elementos determina la dinámica del modelo. Una solución posible es utilizar la información del pasado disponible para generar un dinamismo en los parámetros, realizando regresiones en las que los intervalos de observación se desplacen con el tiempo. De esta manera cada regresión proporcionará una estimación diferente del vector parametral. Bajo el contexto estático de un modelo de regresión, la forma en que cambian las estimaciones puede sugerir la forma en que evoluciona la sucesión de parámetros en el tiempo bajo el marco de un modelo dinámico.

A partir de la ecuación (4.12) se realizan las regresiones considerando primero el periodo comprendido entre febrero de 1987 y marzo de 1994, (el intervalo no comienza en enero para poder tener la observación rezagada $(m-p)_{t-1}$), en segundo lugar al periodo entre marzo de 1987 y marzo de 1994 y así sucesivamente hasta tomar finalmente el intervalo de septiembre de 1988 a marzo de 1994, completándose 20 regresiones de la manera ilustrada en la figura 4.1 y los valores de los parámetros de cada regresión se registran en la tabla 4.5.

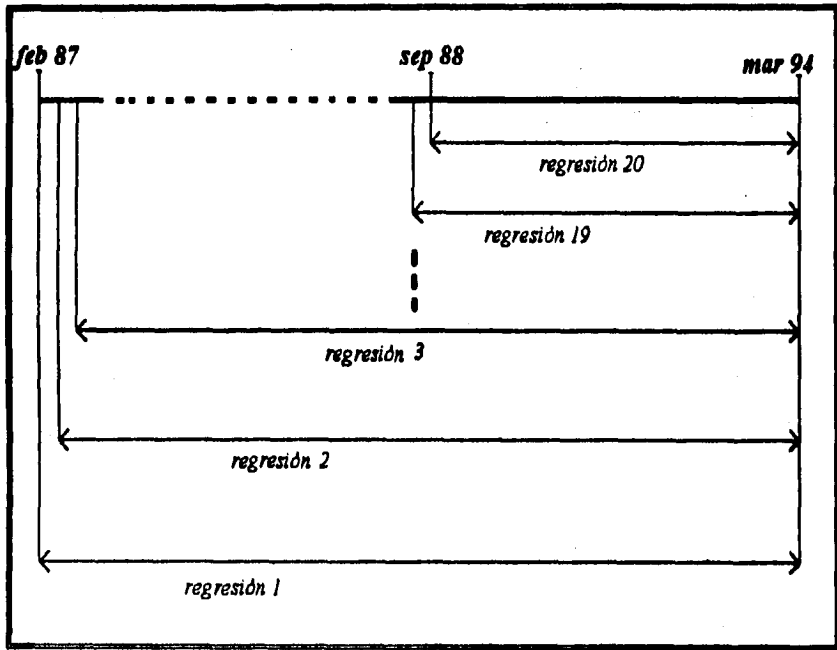


Figura 4.1

Regresión	α_i				
	$i=1$	$i=2$	$i=3$	$i=4$	$i=5$
1	0.8893	-0.9724	-0.0252	0.3143	0.1932
2	0.8890	-0.9775	-0.0251	0.3142	0.1931
3	0.8916	-0.9339	-0.0255	0.3149	0.1940
4	0.8913	-0.9394	-0.0255	0.3148	0.1939
5	0.8922	-0.9260	-0.0256	0.3150	0.1942
6	0.8957	-0.8607	-0.0261	0.3159	0.1954
7	0.8959	-0.8789	-0.0260	0.3159	0.1954
8	0.8957	-0.8861	-0.0259	0.3158	0.1953
9	0.8965	-0.8396	-0.0263	0.3161	0.1957
10	0.8983	-0.8538	-0.0262	0.3164	0.1961
11	0.8943	-0.9457	-0.0255	0.3154	0.1876
12	0.9094	-0.7576	-0.0271	0.3258	0.1931
13	0.9172	-0.6644	-0.0274	0.3261	0.1938
14	0.8863	-1.0302	-0.0250	0.3207	0.1866
15	0.8128	-2.1422	-0.0184	0.3081	0.1696
16	0.8133	-2.1456	-0.0184	0.3083	0.1697
17	0.8076	-2.1890	-0.0181	0.3072	0.1684
18	0.8422	-1.9392	-0.0200	0.3134	0.1766
19	0.8397	-1.9616	-0.0199	0.3130	0.1760
20	0.8297	-2.0335	-0.0193	0.3112	0.1737
PROMEDIO	0.8744	-1.2449	-0.0238	0.3151	0.1871
VAR.	0.0012	0.3154	0.0000	0.0000	0.0001

Tabla 4.5 Sección I. Estimaciones de los parámetros obtenidas de las regresiones.

Regresión	α_i				
	$i=6$	$i=7$	$i=8$	$i=9$	$i=10$
1	0.1700	0.1033	0.0989	0.1344	0.1369
2	0.1699	0.1032	0.0988	0.1343	0.1368
3	0.1710	0.1042	0.0998	0.1352	0.1376
4	0.1708	0.1041	0.0996	0.1351	0.1375
5	0.1712	0.1045	0.1000	0.1354	0.1378
6	0.1727	0.1059	0.1013	0.1367	0.1464
7	0.1726	0.1058	0.1012	0.1336	0.1462
8	0.1725	0.1057	0.1002	0.1334	0.1461
9	0.1730	0.1121	0.1011	0.1342	0.1467
10	0.1700	0.1122	0.1012	0.1343	0.1470
11	0.1678	0.1100	0.0991	0.1324	0.1452
12	0.1738	0.1160	0.1047	0.1376	0.1501
13	0.1749	0.1170	0.1055	0.1383	0.1506
14	0.1664	0.1087	0.0980	0.1315	0.1445
15	0.1464	0.0890	0.0801	0.1156	0.1311
16	0.1466	0.0892	0.0803	0.1157	0.1312
17	0.1450	0.0876	0.0789	0.1144	0.1300
18	0.1547	0.0969	0.0872	0.1223	0.1269
19	0.1540	0.0963	0.0866	0.1223	0.1264
20	0.1512	0.0936	0.0879	0.1203	0.1245
PROMEDIO	0.1647	0.1033	0.0955	0.1299	0.1390
VAR.	0.0001	0.0001	0.0001	0.0001	0.0001

Tabla 4.5 Sección II. Estimaciones de los parámetros obtenidas de las regresiones.

Regresión	α_i		γ_j		
	$i=11$	$i=12$	$j=1$	$j=2$	$j=3$
1	0.1488	0.0953	0.1292	0.1140	0.0601
2	0.1487	0.0952	0.1291	0.1133	0.0601
3	0.1496	0.0981	0.1355	0.1138	0.0565
4	0.1495	0.0976	0.1356	0.1137	0.0559
5	0.1511	0.0979	0.1359	0.1139	0.0559
6	0.1526	0.0993	0.1372	0.1145	0.0557
7	0.1525	0.0991	0.1370	0.1145	0.0558
8	0.1524	0.0990	0.1369	0.1145	0.0559
9	0.1530	0.0997	0.1375	0.1146	0.0556
10	0.1533	0.0999	0.1376	0.1149	0.0558
11	0.1515	0.0981	0.1361	0.1143	0.0562
12	0.1567	0.1029	0.1404	0.1167	0.0560
13	0.1573	0.1033	0.1406	0.1160	0.0559
14	0.1506	0.0975	0.1337	0.1099	0.0563
15	0.1355	0.0929	0.1186	0.0989	0.0413
16	0.1356	0.0919	0.1184	0.0990	0.0421
17	0.1360	0.0909	0.1174	0.0982	0.0418
18	0.1434	0.0969	0.1234	0.1035	0.0434
19	0.1429	0.0965	0.1230	0.1031	0.0433
20	0.1408	0.0947	0.1213	0.1016	0.0429
PROMEDIO	0.1481	0.0973	0.1313	0.1102	0.0523
VAR.	0.0000	0.0000	0.0001	0.0000	0.0000

Tabla 4.5 Sección III. Estimaciones de los parámetros obtenidas de las regresiones.

Una solución sencilla y natural para (i), es considerar que la transición de cada parámetro puede ser descrita como un proceso autoregresivo de primer orden, (2.3), expresado de la siguiente forma

$$\Theta_{it} = \rho_i \Theta_{i,t-1} + \varepsilon_{w_{it}} \quad i = 1, 2, \dots, 15 \quad (4.13)$$

o

$$\begin{aligned} \gamma_{it} &= \rho_i \gamma_{i,t-1} + \varepsilon_{w_{it}} & i &= 1, \dots, 12 \\ \alpha_{jt} &= \rho_{j+12} \alpha_{j,t-1} + \varepsilon_{w_{j+12,t}} & j &= 1, 2, 3 \end{aligned} \quad (4.13.a)$$

sin que el valor actual de un parámetro dependa del valor anterior de otro parámetro que no sea él mismo. Esta consideración tiene como objetivo que la matriz de transferencia sea diagonal.

Si se supone que (4.13.a) se cumple para los parámetros obtenidos en las regresiones realizadas, con el subíndice t indicando el número de la regresión en lugar del tiempo, se puede utilizar la información contenida en la tabla 4.5 para estimar los valores de los coeficientes ρ_i . Con este fin se realiza un proceso similar al ilustrado en la figura 4.1: a partir de (4.13) se realizan regresiones, en este caso auto-regresiones, considerando primero los datos de la regresión 2 a la regresión 20 (nuevamente no se comienza desde la regresión 1 para poder tener el dato rezagado), en segundo lugar se toman los datos de la regresión 2 hasta la regresión 20 y así sucesivamente hasta finalizar con la autoregresión de los datos de las regresiones 11 a la 20, completándose 10 autoregresiones.

En la tabla 4.6 están contenidos los valores estimados para los coeficientes ρ_i , $i = 1, 2, \dots, 15$ correspondientes a cada autoregresión.

Auto-Regresión	ρ_i				
	$i=1$	$i=2$	$i=3$	$i=4$	$i=5$
1	0.9962	1.0261	0.9857	0.9994	0.9943
2	0.9960	1.0267	0.9850	0.9994	0.9940
3	0.9956	1.0289	0.9831	0.9992	0.9933
4	0.9953	1.0296	0.9820	0.9992	0.9929
5	0.9950	1.0309	0.9803	0.9991	0.9923
6	0.9943	1.0340	0.9770	0.9988	0.9912
7	0.9938	1.0344	0.9753	0.9987	0.9904
8	0.9933	1.0351	0.9730	0.9987	0.9896
9	0.9925	1.0378	0.9684	0.9985	0.9883
10	0.9915	1.0383	0.9646	0.9982	0.9867
Promedia	0.9944	1.0322	0.9774	0.9989	0.9913
Desv. Estándar	0.0016	0.0044	0.0071	0.0004	0.0025
Variación	0.16%	0.42%	0.73%	0.04%	0.25%

Tabla 4.6 Sección I. Estimaciones del coeficiente ρ obtenidas de las autoregresiones.

Auto-Regresión	ρ_i				
	$i=6$	$i=7$	$i=8$	$i=9$	$i=10$
1	0.9936	0.9938	0.9928	0.9938	0.9952
2	0.9933	0.9935	0.9924	0.9935	0.9950
3	0.9925	0.9926	0.9913	0.9927	0.9944
4	0.9920	0.9922	0.9908	0.9922	0.9941
5	0.9913	0.9914	0.9899	0.9915	0.9936
6	0.9899	0.9899	0.9881	0.9901	0.9889
7	0.9891	0.9891	0.9871	0.9912	0.9880
8	0.9881	0.9883	0.9868	0.9905	0.9870
9	0.9866	0.9815	0.9846	0.9890	0.9853
10	0.9871	0.9791	0.9826	0.9877	0.9835
Promedio	0.9903	0.9891	0.9887	0.9912	0.9905
Desv. Estándar	0.0026	0.0050	0.0034	0.0020	0.0044
Variación	0.26%	0.51%	0.35%	0.20%	0.45%

Tabla 4.6 Sección II. Estimaciones del coeficiente ρ obtenidas de las autoregresiones.

Auto-Regresión	ρ_i				
	$i=11$	$i=12$	$i=13$	$i=14$	$i=15$
1	0.9968	0.9993	0.9964	0.9938	0.9812
2	0.9966	0.9994	0.9962	0.9938	0.9799
3	0.9961	0.9977	0.9933	0.9932	0.9830
4	0.9959	0.9978	0.9928	0.9928	0.9825
5	0.9949	0.9974	0.9921	0.9922	0.9812
6	0.9938	0.9962	0.9908	0.9911	0.9799
7	0.9933	0.9961	0.9901	0.9904	0.9779
8	0.9928	0.9958	0.9893	0.9895	0.9756
9	0.9917	0.9948	0.9877	0.9883	0.9734
10	0.9906	0.9941	0.9862	0.9867	0.9695
Promedio	0.9942	0.9969	0.9915	0.9912	0.9784
Desv. Estándar	0.0022	0.0018	0.0033	0.0024	0.0044
Variación	0.22%	0.18%	0.34%	0.25%	0.45%

Tabla 4.6 Sección III. Estimaciones del coeficiente ρ obtenidas de las autoregresiones.

De los datos de la tabla 4.6 se puede observar que la variación de los coeficientes ρ_i es muy pequeña, hecho que sugiere que una solución para (i) es suponer que

$$G_i = G \quad \forall i$$

siendo la matriz G una matriz diagonal cuyas entradas en la diagonal están contenidas en la tabla 4.7 y que corresponden a los valores promedio presentados en la tabla 4.6.

i	G_{ii}	i	G_{ii}
1	0.9944	9	0.9912
2	1.0322	10	0.9784
3	0.9774	11	0.9942
4	0.9989	12	0.9969
5	0.9913	13	0.9915
6	0.9942	14	0.9912
7	0.9969	15	0.9784
8	0.9913		

Tabla 4.7. Entradas de la matriz de transferencia.

Para solucionar (ii), se puede proceder de manera similar a los ejemplos del capítulo anterior, determinando la sucesión $\{\sigma_{it}^2\}$ por medio de un factor de descuento,

$$\sigma_{w_t}^2 = \left(\frac{1-\delta}{\delta} \right) S_{t-1}^2$$

Así mismo, como no existe ninguna razón para pensar lo contrario se supone que $\mu_{w_t} = 0$ y $\mu_{v_t} = 0$ para toda t .

También se supone que la varianza observacional es constante en el tiempo, $\sigma_{v_t}^2 = \sigma_v^2$ para toda t , pero desconocida y por tanto se le asocia una distribución de probabilidad, aplicándose el análisis desarrollado en la sección 4.3.

Información Inicial

Para resolver el problema (iii), primero se debe establecer el tiempo a partir del cual se realizarán los pronósticos, i.e. determinar el momento inicial. Puesto que el intervalo de observación de la Regresión 20 es el periodo entre septiembre de 1988 y marzo de 1994, se escoge al mes de septiembre de 1988 como $t=0$, y el primer pronóstico $f_0(i) = ((M-P)_i | D_0)$, será el correspondiente al mes de octubre de 1991.

El valor de $S_{v_0}^2$ es determinado por la varianza muestral

$$\begin{aligned} S_{v_0}^2 &= \hat{\sigma}_{mp}^2 \\ &= \sum_{t=\text{octubre1988}}^{\text{marzo1994}} \frac{\left[(m-p)_t - \sum_{i=\text{octubre1988}}^{\text{marzo1994}} \frac{(m-p)_i}{67} \right]^2}{66} \\ &= 2.004664 \times 10^{-2} \end{aligned}$$

Por simplicidad, se escoge que $n_0 =$ para que $d_0 = S_{v_0}^2$.

Para determinar la distribución del parámetro Θ_0 , se puede utilizar la información generada por las regresiones, tomando los valores promedio, mostrados en la tabla 4.5, como las entradas del vector m_0 . Utilizando las varianzas, también mostradas en la tabla 4.5, y suponiendo que

$$COV[\theta_0, \theta_{j_0} | D_0] = 0 \quad i \neq j$$

se puede construir la matriz S_0^2 . Así

$$[\Theta_0 | D_0] \sim T_1[m_0, S_0^2]$$

con m_0 y S_0^2 mostrados en la tabla 4.8.

i	m_{i0}	C_{i0}
1	0.8744	1.24E-03
2	-1.2449	3.15E-01
3	-0.0238	1.10E-05
4	0.3151	2.38E-05
5	0.1871	1.07E-04
6	0.1647	1.11E-04
7	0.1033	7.56E-05
8	0.0955	7.27E-05
9	0.1299	6.47E-05
10	0.1390	7.08E-05
11	0.1481	4.55E-05
12	0.0973	1.04E-05
13	0.1313	6.34E-05
14	0.1102	4.32E-05
15	0.0523	4.55E-05
$i \neq j$		$C_{i0} = 0$

Tabla 4.8. Entradas de los vectores correspondientes a los momentos de la distribución inicial.

Los anteriores elementos constituyen el conjunto D_0 y de acuerdo a esta información se establecen las siguientes distribuciones

$$\begin{aligned}
 (\Theta_{10} | D_0) &\sim T_1[0.8744, 1.24 \times 10^{-3}], & (\Theta_{20} | D_0) &\sim T_1[-1.2449, 3.15 \times 10^{-1}], \\
 (\Theta_{30} | D_0) &\sim T_1[-0.0238, 1.10 \times 10^{-5}], & (\Theta_{40} | D_0) &\sim T_1[0.3151, 2.38 \times 10^{-5}], \\
 (\Theta_{50} | D_0) &\sim T_1[0.1871, 1.07 \times 10^{-4}], & (\Theta_{60} | D_0) &\sim T_1[0.1647, 1.11 \times 10^{-4}], \\
 (\Theta_{70} | D_0) &\sim T_1[0.1033, 7.56 \times 10^{-5}], & (\Theta_{80} | D_0) &\sim T_1[0.0955, 7.27 \times 10^{-5}], \\
 (\Theta_{90} | D_0) &\sim T_1[0.1299, 6.47 \times 10^{-5}], & (\Theta_{100} | D_0) &\sim T_1[0.1390, 7.08 \times 10^{-5}], \\
 (\Theta_{110} | D_0) &\sim T_1[0.1481, 4.55 \times 10^{-5}], & (\Theta_{120} | D_0) &\sim T_1[0.0973, 1.04 \times 10^{-5}], \\
 (\Theta_{130} | D_0) &\sim T_1[0.1313, 6.34 \times 10^{-5}], & (\Theta_{140} | D_0) &\sim T_1[0.1102, 4.32 \times 10^{-5}], \\
 (\Theta_{150} | D_0) &\sim T_1[0.0523, 4.55 \times 10^{-5}], & & \\
 (\phi_0 | D_0) &\sim G\left[\frac{1}{2}, \frac{2.0046 \times 10^{-2}}{2}\right].
 \end{aligned}$$

El establecimiento de una distribución inicial para Θ_0 tiene sentido si existe una interpretación para este vector: si se considera que las observaciones $(m-p)_1, (m-p)_2, \dots$

son la continuación de una serie observada previamente y que el origen $t = 0$ es sólo una referencia arbitraria, Θ_0 representa el vector parametral final resultante de los datos históricos. Si por el contrario no existe ningún antecedente de la serie, Θ_0 carece de interpretación y el modelo se inicializa con una distribución a priori para Θ_1 .

Transición

Como la elección de $\{\sigma_{w_t}^2\}$ es por factor de descuento entonces

$$\sigma_{w_t}^2 = \left(\frac{1-0.8}{0.8} \right) S_0^2$$

$$= \begin{pmatrix} 310 \cdot 10^{-4} & 0 \\ 0 & 788 \cdot 10^{-2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 275 \cdot 10^{-6} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 593 \cdot 10^{-6} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 267 \cdot 10^{-3} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 277 \cdot 10^{-5} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 189 \cdot 10^{-5} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 181 \cdot 10^{-5} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 161 \cdot 10^{-5} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 177 \cdot 10^{-5} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 113 \cdot 10^{-5} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 24 \cdot 10^{-6} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 158 \cdot 10^{-5} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 108 \cdot 10^{-5} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 113 \cdot 10^{-5} & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Los momentos de la distribución a priori de Θ_1 se obtienen de las siguientes operaciones de matrices

$$\bar{m}_1 = Gm_0 + \mu_{V_0}$$

$$= \begin{pmatrix} 0.8695 \\ -1.2849 \\ -0.0233 \\ 0.3148 \\ 0.1854 \\ 0.1631 \\ 0.1021 \\ 0.0944 \\ 0.1287 \\ 0.1377 \\ 0.1472 \\ 0.0970 \\ 0.1302 \\ 0.1092 \\ 0.0512 \end{pmatrix} + 0 = \begin{pmatrix} 0.8695 \\ -1.2849 \\ -0.0233 \\ 0.3148 \\ 0.1854 \\ 0.1631 \\ 0.1021 \\ 0.0944 \\ 0.1287 \\ 0.1377 \\ 0.1472 \\ 0.0970 \\ 0.1302 \\ 0.1092 \\ 0.0512 \end{pmatrix}$$

mes comprendido entre octubre de 1988 y marzo de 1994. Para escoger las matrices σ_w^2 , se utilizaron los factores $\delta = 0.8, 0.9, 1.0$.

<i>t</i>	<i>mp_t</i>	$\delta = 1.0$		$\delta = 0.9$		$\delta = 0.8$	
		\hat{Y}_t	$\hat{\sigma}_t^2$	\hat{Y}_t	$\hat{\sigma}_t^2$	\hat{Y}_t	$\hat{\sigma}_t^2$
88							
<i>O</i>	-0.3481	0.7346	0.0215	0.7346	0.0217	0.7346	0.0219
<i>N</i>	-0.2146	0.6766	0.5937	0.6693	0.5978	0.6603	0.6040
<i>D</i>	-0.1061	0.5739	0.6561	0.5559	0.6579	0.5329	0.6629
89							
<i>E</i>	-0.2719	0.4413	0.6049	0.4126	0.6032	0.3750	0.6053
<i>F</i>	-0.2896	-1.0680	12.5748	-1.1103	19.4857	-1.1670	31.8999
<i>M</i>	-0.2611	1.2142	0.5178	1.1722	0.5367	1.1158	0.5892
<i>A</i>	-0.2593	0.6375	0.7143	0.5555	0.6873	0.4412	0.6629
<i>M</i>	-0.2642	0.4559	0.7258	0.3475	0.6897	0.1949	0.6577
<i>J</i>	-0.2507	0.3995	0.7009	0.2721	0.6554	0.1004	0.6152
<i>J</i>	-0.2446	0.3117	0.6702	0.1691	0.6148	-0.0127	0.5651
<i>A</i>	-0.2772	0.2819	0.6364	0.1260	0.5756	-0.0596	0.5248
<i>S</i>	-0.2731	0.2897	0.6077	0.1230	0.5405	-0.0622	0.4873
<i>O</i>	-0.2285	0.2753	0.5846	0.0953	0.5123	-0.0923	0.4603
<i>N</i>	-0.1794	0.2758	0.5600	0.0849	0.4829	-0.0991	0.4291
<i>D</i>	0.0170	0.2185	0.5346	0.0228	0.4522	-0.1499	0.3908
90							
<i>E</i>	-0.1855	0.1273	0.5030	-0.0620	0.4236	-0.2029	0.3661
<i>F</i>	-0.2115	-0.5505	1.4212	-0.4278	2.8724	-0.3397	7.6699
<i>M</i>	-0.2020	0.0916	0.4542	-0.0983	0.3819	-0.2244	0.3446
<i>A</i>	-0.1963	1.1307	0.4532	0.9134	0.4684	0.7480	1.0535
<i>M</i>	-0.1862	0.2222	0.4980	-0.0059	0.4024	-0.0954	0.4095
<i>J</i>	-0.1842	0.1892	0.4817	-0.0430	0.3854	-0.1320	0.4075
<i>J</i>	-0.2085	0.1274	0.4651	-0.1057	0.3653	-0.1891	0.3725
<i>A</i>	-0.2245	0.1113	0.4494	-0.1234	0.3504	-0.2016	0.3679
<i>S</i>	-0.2310	0.1345	0.4348	-0.0983	0.3361	-0.1713	0.3674
<i>O</i>	-0.1737	0.1299	0.4227	-0.1076	0.3249	-0.1785	0.3806
<i>N</i>	-0.1271	0.1441	0.4095	-0.0916	0.3114	-0.1543	0.3612
<i>D</i>	0.0753	0.1032	0.3962	-0.1266	0.2948	-0.1781	0.2829
91							
<i>E</i>	-0.0794	0.0194	0.3818	-0.1905	0.2842	-0.2029	0.2466
<i>F</i>	-0.1011	-0.4501	0.8874	-0.2253	1.7868	-0.1953	4.6692
<i>M</i>	-0.0403	0.6307	0.3707	0.2299	0.4276	-0.2376	1.5877
<i>A</i>	-0.1115	0.3040	0.3606	0.1063	0.2688	0.1012	0.4116
<i>M</i>	-0.0647	0.1307	0.3551	-0.0778	0.2733	-0.0974	0.6909
<i>J</i>	-0.0509	0.1074	0.3453	-0.0940	0.2660	-0.1108	0.7386
<i>J</i>	-0.0643	0.0582	0.3353	-0.1339	0.2525	-0.1421	0.6442
<i>A</i>	-0.0669	0.0488	0.3260	-0.1342	0.2456	-0.1465	0.6720
<i>S</i>	-0.0855	0.0777	0.3171	-0.0966	0.2404	-0.1239	0.7426
<i>O</i>	-0.0511	0.0769	0.3092	-0.0934	0.2360	-0.1084	0.8078
<i>N</i>	-0.0053	0.0982	0.3014	-0.0647	0.2287	-0.0718	0.8132
<i>D</i>	0.1927	0.0668	0.2935	-0.0841	0.2132	-0.0423	0.5216

Tabla 4.9 sección I

<i>t</i>	<i>mp_t</i>	$\delta = 1.0$		$\delta = 0.9$		$\delta = 0.8$	
		\hat{Y}_t	$\hat{\sigma}_t^2$	\hat{Y}_t	$\hat{\sigma}_t^2$	\hat{Y}_t	$\hat{\sigma}_t^2$
92							
E	0.0527	-0.0128	0.2863	-0.1381	0.2058	-0.0789	0.3130
F	0.0227	-0.3265	0.6287	-0.0964	1.2845	-0.1039	3.3738
M	-0.0255	-0.0278	0.2739	-0.1334	0.1996	-0.0311	0.3859
A	0.0165	0.8482	0.2755	0.4330	0.4384	0.0705	2.6479
M	0.0413	0.0983	0.2771	-0.0009	0.2288	-0.0229	1.2051
J	0.0204	0.0781	0.2709	-0.0142	0.2253	-0.0216	1.1905
J	0.0367	0.0345	0.2648	-0.0495	0.2097	-0.0382	1.0966
A	0.0008	0.0286	0.2590	-0.0474	0.2054	-0.0397	1.0951
S	-0.0476	0.0568	0.2536	-0.0193	0.2062	-0.0563	1.1534
O	-0.0133	0.0562	0.2486	-0.0183	0.2048	-0.0343	1.1392
N	0.0274	0.0810	0.2436	0.0107	0.2012	0.0051	1.1974
D	0.2297	0.0547	0.2387	-0.0033	0.1778	0.1533	1.0498
93							
E	0.0847	-0.0225	0.2345	-0.0609	0.1651	0.0166	0.6282
F	0.0485	-0.1680	0.5050	0.0328	1.0142	0.0083	2.6414
M	0.0113	-0.0330	0.2264	-0.0544	0.1622	0.0106	0.6315
A	0.0176	0.7764	0.2279	0.1946	0.4002	0.0264	1.3023
M	0.0401	0.0814	0.2286	0.0522	0.2229	0.0335	1.0978
J	0.0332	0.0622	0.2246	0.0342	0.2201	0.0156	1.0617
J	0.0489	0.0236	0.2205	0.0099	0.1980	0.0290	1.0167
A	-0.0033	0.0182	0.2167	0.0029	0.1946	0.0005	0.9930
S	-0.0104	0.0445	0.2131	0.0124	0.2028	-0.0430	1.0282
O	0.0414	0.0445	0.2096	0.0178	0.2033	-0.0066	1.0244
N	0.0660	0.0721	0.2062	0.0509	0.2060	0.0324	1.1212
D	0.2670	0.0502	0.2028	0.0618	0.1680	0.2196	1.1111
94							
E	0.1366	-0.0250	0.2002	-0.0062	0.1418	0.0694	0.8864
F	0.1017	-0.0534	0.4284	0.0553	0.8359	0.0324	2.4759
M	0.1542	-0.0324	0.1947	0.0066	0.1407	0.0341	0.6947

Tabla 4.9 sección II

Los resultados muestran nuevamente la inconveniencia de utilizar el factor de descuento $\delta = 1.0$. La línea de observaciones presenta un comportamiento mucho más estable que la línea de pronóstico. El predictor puntual parece nunca ajustarse a los datos.

En las gráficas se puede observar que en los tres casos estudiados, ocurre un desfase en la estacionalidad de ambas series. Mientras la línea de observaciones presenta picos o cimas en los meses de diciembre, las líneas de pronóstico muestran picos y valles en diferentes meses del año. La causa parece ser la información inicial. En el caso más sensible a la información

reciente, $\delta = 0.8$, a pesar de que al principio del proceso se presentan diferencias muy importante entre el predictor puntual y la observación y de hecho parecen tener patrones de comportamiento muy dispares, conforme pasa el tiempo y se integran nuevos datos, la brecha entre las dos líneas se hace muy pequeña. Parece que el predictor \hat{Y}_t aprende de los datos y esto se demuestra con lo siguiente

(1) Después de 3 años comienza a predecir los picos de diciembre que presenta la serie de interés.

(2) Los errores de pronóstico disminuyen su magnitud conforme avanza el tiempo.

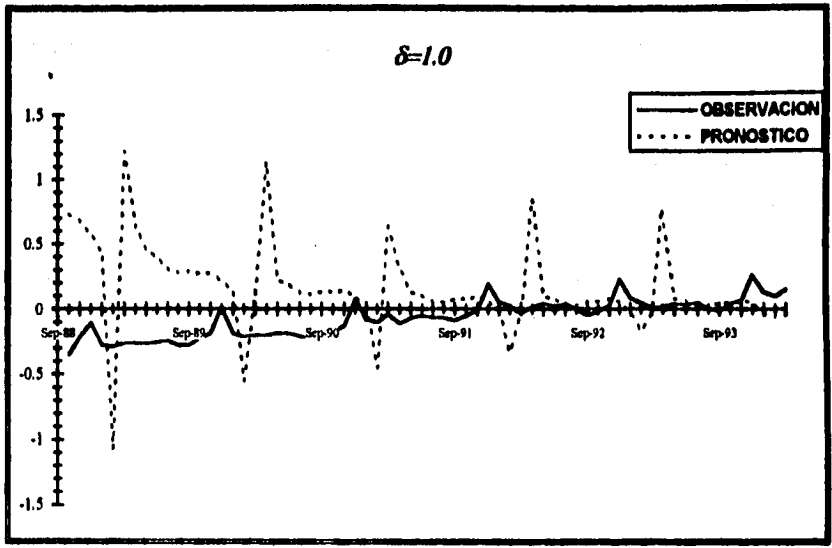


Figura 4.2

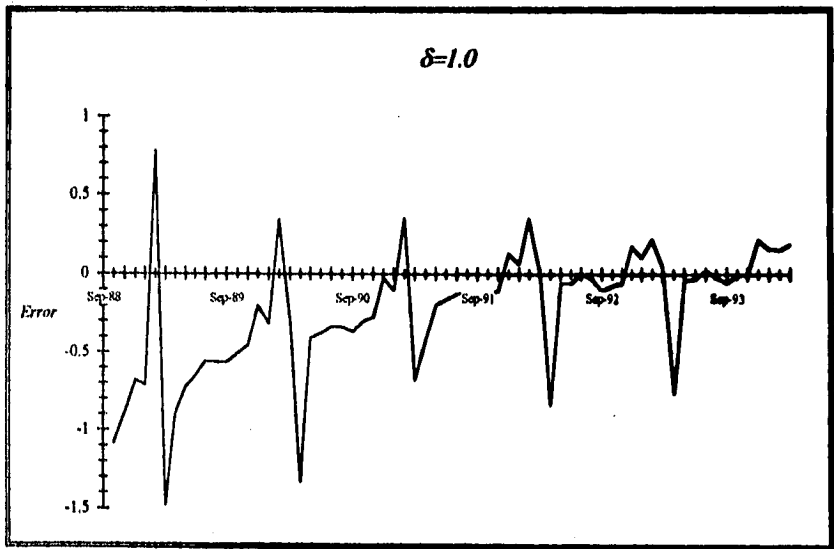


Figura 4.2.a

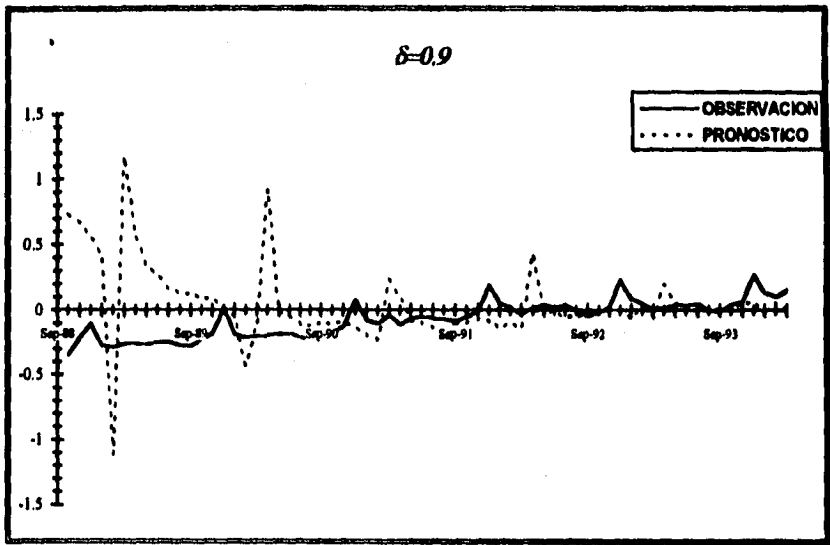


Figura 4.3

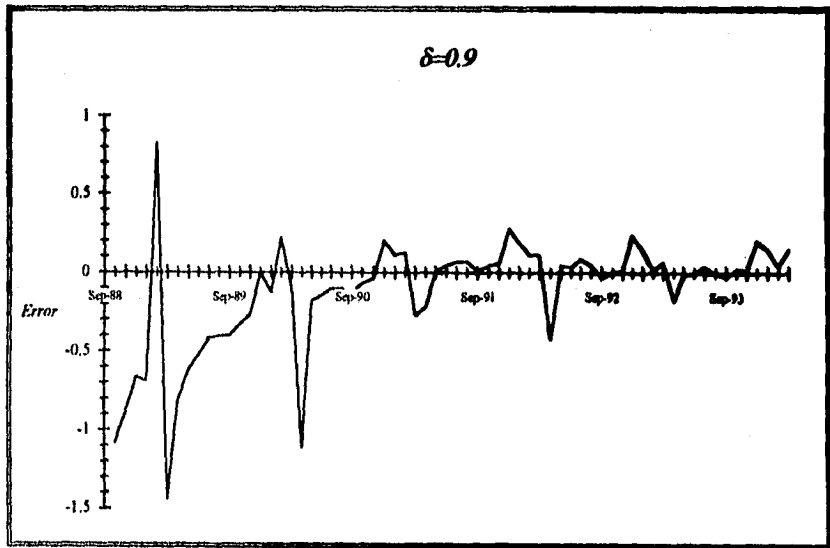


Figura 4.3.a

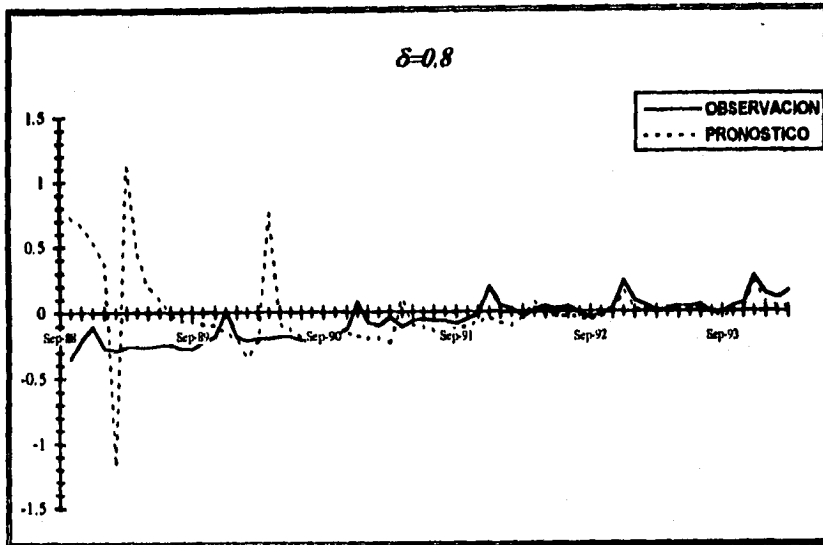


Figura 4.4

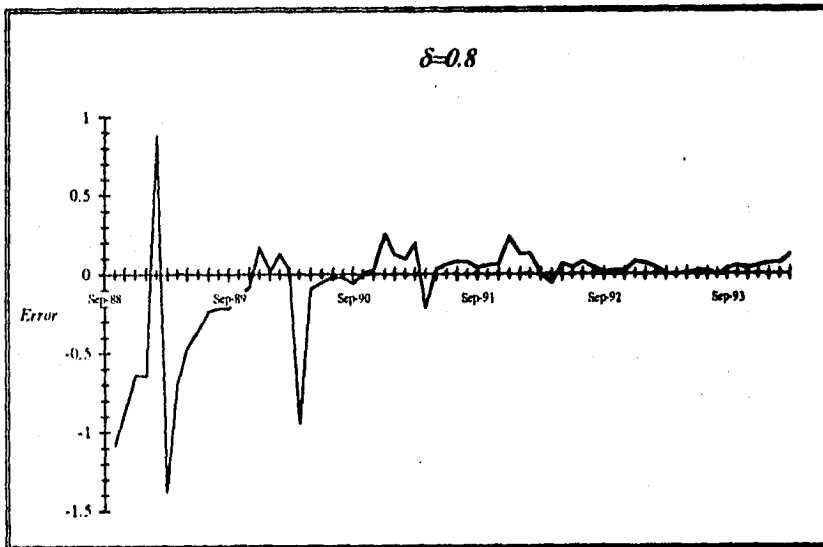


Figura 4.4.a

Para completar este ejemplo, a continuación se presenta una tabla comparativa de la desviación media absoluta y el error medio cuadrático y el cociente de verosimilitudes $\frac{\ln L(d|y_1, \dots, y_{171})}{\ln L(1.0|y_1, \dots, y_{171})}$ para diferentes valores de δ utilizados.

δ	<i>DMA</i>	<i>EMC</i>	$\frac{\ln L(d y_1, \dots, y_{171})}{\ln L(1.0 y_1, \dots, y_{171})}$
1.0	21.8915	14.3628	0.0000
0.9	15.5383	9.6985	8.4066
0.8	12.2220	7.7265	14.5921

Tabla 4.10

En esta tabla se confirma que el factor de descuento $\delta = 0.8$ es el de mayor calidad de entre los tres seleccionadas.

La intención de este ejemplo es mostrar que el modelo dinámico lineal se puede aplicar de manera no muy complicada. Para este trabajo se elaboró una macro en una hoja de cálculo y la aplicación del procedimiento de actualización no presenta dificultades de cálculo.

CONCLUSIONES

Del estudio y análisis de la metodología Bayesiana, los modelos desarrollados y los ejemplos expuestos en este trabajo, se establecen las siguientes conclusiones:

1. En el problema de inferencia estadística bajo el contexto bayesiano destacan los siguientes aspectos:

- a) La probabilidad es subjetiva.
- b) Los parámetros son interpretados como variables aleatorias.
- c) A los parámetros se les asocia una distribución de probabilidad.

2. El problema de Pronosticación es un problema que está presente en muchas de las actividades humanas. Los pronósticos buscan reducir el riesgo de tomar una decisión en situaciones de incertidumbre, por lo que ayudan a hacer una planeación adecuada.

3. Entre los métodos de Pronosticación se cuentan a los Modelos de Series de tiempo y al Modelo de Regresión Lineal. Estos modelos presentan importantes limitaciones al no permitir la incorporación de información externa y al ser estáticos.

4. A pesar de que se cuestiona su aplicabilidad a problemas de inferencia científica, el enfoque Bayesiano ofrece un proceso de aprendizaje a través de la experiencia, abriendo un amplio campo para el desarrollo de métodos de Pronosticación.

5. A partir de la consideración de que ningún sistema es ajeno a las perturbaciones surgen los Modelos Dinámicos. La metodología Bayesiana permite interpretar a los Modelos Dinámicos como rutinas de aprendizaje, constituyendo una alternativa para la realización de pronósticos. Los Modelo Dinámicos se caracterizan por:

- a) Tener una estructura paramétrica.
- b) Tener una definición secuencial.
- c) Tener una representación probabilística de la información acerca de los parámetros, permitiendo integrar información subjetiva, realizar pronósticos en situaciones de escasa información y realizar intervenciones.
- d) Los pronósticos son en forma de distribuciones de probabilidad.

6. El Modelo Polinomial de primer orden o simplemente Modelo Lineal es el caso más simple dentro de la clase de los Modelos Dinámicos Lineales y su estudio permite la comprensión de modelos más complejos. El modelo es aplicable para la descripción de series de tiempo cuyo nivel no presenta cambios demasiado bruscos.

7. El Modelo Dinámico Lineal General es un sistema de ecuaciones que especifica la forma en que las observaciones dependen estocásticamente de los parámetros actuales del proceso y la forma en que los parámetros evolucionan en el tiempo. Constituye una poderosa herramienta para la realización de pronósticos al poseer una gran maleabilidad y permitir describir diversos procesos. Entre las principales dificultades que se presentan durante la utilización de un modelo Dinámico se cuentan:

- a) La forma en que se escogen los diferentes elementos del modelo: vectores esperanza, matrices de varianzas y de transferencia.
- b) La forma de la información inicial.

8. De los ejemplos se puede observar que en los Modelos Dinámicos Lineales, el uso de diferentes factores de descuento provoca diferentes desenvolvimientos de la rutina de pronosticación, los cuales son comparables bajo los criterios de la Desviación Media Absoluta, el Error Medio Cuadrado y la función de Verosimilitud. Se observa también que conforme se va integrando la información de las nuevas observaciones, se observa que los errores tienden a hacerse más pequeños, sin embargo si se escoge un factor de descuento δ muy cercano a 1, el modelo es demasiado rígido y la varianza del error de pronóstico se hace cada vez más grande.

APENDICE

A.1 Distribución Normal

Distribución Normal

Una variable aleatoria X es definida como *normalmente* distribuida si su función de densidad está dada por

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \quad (A1)$$

donde los parámetros μ, σ^2 satisfacen $-\infty < \mu < \infty$ y $\sigma > 0$. Cualquier distribución definida por una función de densidad dada por (A1) es llamada una distribución *normal*.

Momentos

Si X es una variable aleatoria con distribución normal entonces

$$E[X] = \mu$$

$$VAR[X] = \sigma^2$$

Distribución Normal Multivariada

Un n -vector aleatorio $\mathbf{X} = (X_1, X_2, \dots, X_n)'$ tiene una *distribución normal multivariada* de dimensión n si y sólo si $Y = \sum_{i=1}^n a_i X_i = \mathbf{a}\mathbf{X}$ tiene distribución *normal* para todo vector constante $\mathbf{a} = (a_1, a_2, \dots, a_n)$ distinto del vector cero $\mathbf{0}$.

Momentos

Si \mathbf{X} se distribuye como una normal multivariada entonces $E[\mathbf{X}] = \boldsymbol{\mu}$, $VAR[\mathbf{X}] = \boldsymbol{\sigma}^2$ con

$$\mu = \begin{pmatrix} E[X_1] \\ E[X_2] \\ \vdots \\ E[X_n] \end{pmatrix}$$

y

$$\sigma^2 = \begin{pmatrix} VAR[X_1] & COV[X_1, X_2] & \dots & COV[X_1, X_n] \\ COV[X_2, X_1] & VAR[X_2] & \dots & COV[X_2, X_n] \\ \vdots & \vdots & \ddots & \vdots \\ COV[X_n, X_1] & COV[X_n, X_2] & \dots & VAR[X_n] \end{pmatrix}$$

Si $\mathbf{X} \sim N[\mu, \sigma^2]$ la función de densidad de \mathbf{X} es

$$f_{\mathbf{X}}(\mathbf{x}) = \left[(2\pi)^n |\sigma^2| \right]^{-1/2} \exp \left[\frac{-(\mathbf{x} - \mu)' (\sigma^2)^{-1} (\mathbf{x} - \mu)}{2} \right]$$

A.1.1 Distribuciones Marginales

Supongase que se particiona el vector \mathbf{X} en dos subvectores

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix},$$

teniendo que

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad \sigma^2 = \begin{pmatrix} \sigma_{11}^2 & \sigma_{12}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 \end{pmatrix}$$

entonces \mathbf{X}_1 tiene una distribución *marginal* $\mathbf{X}_1 \sim N[\mu_1, \sigma_{11}^2]$.

Cualquier par de elementos X_i, X_j , de un vector \mathbf{X} con distribución normal multivariada se distribuyen en forma normal bivariada.

Distribución Normal Bivariada

Sean X_1, X_2 variables aleatorias con función de densidad conjunta

$$f_{X_i, X_j}(x_i, x_j) = \left[(2\pi) \sqrt{s_{ii}s_{jj}(1-\rho_{ij}^2)} \right]^{-1} \exp \left[-\frac{Q(x_i, x_j)}{2} \right]$$

donde

$$Q(x_i, x_j) = \frac{(x_i - m_i)^2}{s_{ii}^2} + \frac{(x_j - m_j)^2}{s_{jj}^2} - 2\rho_{ij} \frac{(x_i - m_i)(x_j - m_j)}{s_{ii}s_{jj}},$$

y

$$\rho_{ij} = \text{corr}(X_i, X_j) = \frac{\sigma_{ij}}{\sigma_{ii}\sigma_{jj}}$$

es el coeficiente de correlación entre X_i, X_j .

A.1.2 Distribuciones Condicionales

Para la partición del vector \mathbf{X} en \mathbf{X}_1 y \mathbf{X}_2 se tienen la distribución condicional

$$(\mathbf{X}_1 | \mathbf{X}_2 = \mathbf{x}_2) \sim N[\mu_1(\mathbf{x}_2), \sigma_1^2(\mathbf{x}_2)]$$

donde

$$\mu_1(\mathbf{x}_2) = \mu_1 + \sigma_{12}^2 (\sigma_{22}^2)^{-1} (\mathbf{x}_2 - \mu_2),$$

$$\sigma_1^2(\mathbf{X}_2) = \sigma_{11}^2 - \sigma_{12}^2 (\sigma_{22}^2)^{-1} \sigma_{21}^2.$$

La matriz $\mathbf{A}_1 = \sigma_{12}^2 (\sigma_{22}^2)^{-1}$ es llamada *matriz de regresión de \mathbf{X}_1 sobre \mathbf{X}_2* . Los momentos condicionales están dados en términos de la matriz de regresión

$$\mu_1(\mathbf{x}_2) = \mu_1 + \mathbf{A}_1(\mathbf{x}_2 - \mu_2),$$

$$\sigma_1^2(\mathbf{X}_2) = \sigma_{11}^2 - \mathbf{A}_1 \sigma_{22}^2 \mathbf{A}_1'.$$

En el caso especial de la normal bivariada, los momentos son escalares, la correlación entre X_1 y X_2 es

$$\rho = r_{12} = \frac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}}$$

El coeficiente de regresión está dado por

$$A_1 = \rho \sqrt{\frac{\sigma_{11}}{\sigma_{22}}}$$

además

$$VAR[X_1 | X_2 = x_2] = (1 - r^2) \sigma_{11}^2$$

A.1.4 Regresión Lineal

Supóngase que el vector $\mathbf{Y}' = (Y_1, Y_2, \dots, Y_p)$ y el vector $\Theta = (\theta_1, \theta_2, \dots, \theta_n)$ están relacionados a través de:

$$\mathbf{Y} = \mathbf{F}\Theta + \varepsilon \quad \varepsilon \sim N[0, \sigma^2],$$

donde la matriz de $(n \times p)$ \mathbf{F} y la matriz simétrica σ^2 de $(p \times p)$ son constantes.

Supóngase además que

$$\Theta = \mathbf{a} + \omega$$

con

$$\omega \sim N[0, \Sigma^2]$$

independiente de ε . Entonces el vector (\mathbf{Y}, Θ) tiene una distribución conjunta normal multivariada con momentos

$$E[\Theta] = \mathbf{a}$$

$$E[\mathbf{Y}] = E[\mathbf{F}'\Theta + \varepsilon] = \mathbf{F}'\mathbf{a}$$

$$AR[\Theta] = \Sigma^2$$

$$AR[\mathbf{Y}] = VAR[\mathbf{F}'\boldsymbol{\Theta} + \varepsilon] = \mathbf{F}'\boldsymbol{\Sigma}^2\mathbf{F} + \sigma^2$$

$$COV[\mathbf{Y}, \boldsymbol{\Theta}] = COV[\mathbf{F}'\boldsymbol{\Theta} + \varepsilon, \boldsymbol{\Theta}] = \mathbf{F}'\boldsymbol{\Sigma}^2$$

A.2 Distribución Gamma

Distribución Gamma

Si una variable aleatoria Φ tiene función de densidad dada por

$$f_F(f) = \frac{d^n f^{n-1} \exp(-df)}{\Gamma(n)} I_{[0, \infty)}(f)$$

donde $n > 0$ y $d > 0$ y Γ es la función gamma

$$\Gamma(n) = \int_0^{\infty} x^{n-1} e^{-x} dx, \quad n > 0, \quad (\text{A2})$$

entonces Φ así definida tiene una distribución Gamma, $\Phi \sim G[n, d]$.

Momentos

Si $\Phi \sim G[n, d]$ entonces

$$E[\Phi] = \frac{n}{d}$$

$$AR[\Phi] = \frac{E[\Phi]^2}{n} = \frac{n}{d^2}$$

Existen dos casos particulares dentro de esta clase de distribuciones la *exponencial* y la *ji-cuadrada*.

Distribución exponencial

Si $n = 1$, se dice que Φ tiene una distribución exponencial (negativa)

$$f_F(f) = d \exp(-df) I_{[0, \infty)}(f), \quad (\text{A3})$$

con

$$E[\Phi] = \frac{1}{d}$$

Distribución ji-cuadrada

Si X es una variable aleatoria con densidad

$$f_x(x) = \frac{1}{\Gamma\left(\frac{k}{2}\right)} \left(\frac{1}{2}\right)^{\frac{k}{2}-1} \exp\left\{-\frac{1}{2}x\right\} I_{(0,\infty)}(x) \quad (\text{A4})$$

entonces X así definida tiene una distribución *ji-cuadrada* con k grados de libertad, $\chi^2_{(k)}$, donde el parámetro k es un entero positivo. La densidad de una *ji-cuadrada* con k grados de libertad es un caso particular de la densidad de una *Gamma* con parámetros $n = \frac{k}{2}$ y $d = \frac{1}{2}$.

De esta forma si

$$\Phi \sim G\left[\frac{n}{2}, \frac{d}{2}\right]$$

con n un entero positivo entonces la variable aleatoria $Y \equiv d\Phi$ tiene una distribución *ji-cuadrada* con n grados de libertad,

$$Y \sim \chi^2_{(k)}$$

A.3 Distribución t de Student

Sean $W \sim N[0,1]$ y $V \sim \chi^2_{(k)}$ variables aleatorias independientes. La densidad conjunta de W y V es

$$f_{w,v}(w, v) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{w^2}{2}\right) \frac{1}{\Gamma\left(\frac{k}{2}\right) 2^{\frac{k}{2}}} v^{\frac{k}{2}-1} \exp\left(-\frac{v}{2}\right) I_{(0,\infty)}(v)$$

Si se realiza la transformación $X = \frac{W}{\sqrt{\frac{V}{k}}}$ y $Y = V$, se obtiene que la densidad conjunta

$$f_{x,y}(x,y) = \frac{1}{\sqrt{2\pi}\Gamma\left(\frac{k}{2}\right)} y^{\frac{k}{2}-1} \exp\left[-\frac{y}{2}\left(1+\frac{x^2}{k}\right)\right] \frac{\sqrt{y}}{\sqrt{k}} I_{\left(\frac{k}{2}\right)}(y).$$

Entonces la densidad marginal de X es

$$f_X(x) = \frac{\Gamma\left(\frac{(k+1)}{2}\right)}{\Gamma\left(\frac{k}{2}\right)} \frac{1}{\sqrt{k\pi}} \left(1+\frac{x^2}{k}\right)^{-\frac{(k+1)}{2}} \quad (\text{A5})$$

y se dice que X tiene una distribución *t de Student* estándar con k grados de libertad.

Una variable aleatoria Y tiene distribución *t-Student* con k grados de libertad, moda μ y escala σ^2 si y sólo si

$$f_Y(y) \propto \left[k + \frac{(y-\mu)^2}{\sigma^2}\right]^{-\frac{(k+1)}{2}} \quad (\text{A6})$$

Sea Z una variable aleatoria tal que $Z = \frac{Y-\mu}{\sigma}$, entonces la función de densidad de Z está dada por (A5).

Momentos

Si X tiene una distribución *t de Student* estándar con k grados de libertad entonces

$$E[X] = 0 \text{ para } k > 1,$$

$$AR[X] = \frac{k}{k-2} \text{ para } k > 2.$$

La ecuación (A5) puede escribirse de la siguiente forma para remarcar su dependencia del parámetro k

$$f_X^{(k)}(x) = B_k \left(1 + \frac{x^2}{k}\right)^{-\frac{(k+1)}{2}} \quad (\text{A5.a})$$

Se puede mostrar que

$$B_k = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{k\pi}\Gamma\left(\frac{k}{2}\right)} = \exp\left(-\frac{1}{2}\right) (\pi k)^{-\frac{1}{2}} \left(\frac{k-1}{k-2}\right)^{\frac{(k-2)}{2}} \left(\frac{k-1}{k-2}\right)^{\frac{1}{2}} \left(\frac{k-1}{2}\right)^{\frac{1}{2}}$$

y

$$\lim_{k \rightarrow \infty} B_k = (2\pi)^{-\frac{1}{2}}.$$

Por otra parte

$$\begin{aligned} \lim_{k \rightarrow \infty} \left(1 + \frac{x^2}{k}\right)^{-\frac{(k+1)}{2}} &= \lim_{k \rightarrow \infty} \left(1 + \frac{x^2}{k}\right)^{-\frac{1}{2}} \left(1 + \frac{x^2}{k}\right)^{-\frac{k}{2}} \\ &= \exp\left(-\frac{x^2}{2}\right) \end{aligned}$$

Por lo anterior se puede concluir que conforme $k \rightarrow \infty$ la distribución *t de Student* estándar converge en distribución a una *Normal* estándar, (Kenney y Keeping [10])

$$\lim_{k \rightarrow \infty} f_X^{(k)} = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\} \quad (\text{A7})$$

A.4 Distribución Normal-Gamma

Distribución Normal-Gamma Univariada

Sea Φ una v.a. tal que $\Phi \sim G\left[\frac{n}{2}, \frac{d}{2}\right]$, con $n > 0$ y $d > 0$. Supongase que $(X|\Phi = \phi) \sim N\left[m, \frac{S^2}{\phi}\right]$, para alguna m y S^2 . (Nótese que $E[X] = E[X|\Phi = \phi] = m$, no depende de ϕ). Entonces la distribución conjunta de X y Φ es llamada *distribución Normal-Gamma*, y su función de densidad es

$$\begin{aligned} f_{X,\Phi}(x, \phi) &= \left(\frac{\phi}{2\pi S^2}\right)^{1/2} \exp\left[-\frac{\phi(x-m)^2}{2S^2}\right] \frac{d^{n/2}}{2^{n/2} \Gamma\left(\frac{n}{2}\right)} \phi^{\frac{n}{2}-1} \exp\left[-\frac{d\phi}{2}\right] I_{(0,\infty)}(\phi) \\ &\propto \phi^{\left(\frac{n+1}{2}\right)-1} \exp\left\{-\frac{\phi}{2}\left[\frac{(x-m)^2}{S^2} + d\right]\right\} I_{(0,\infty)}(\phi) \end{aligned} \quad (\text{A8})$$

Distribución Condicional

A partir de (A8) se obtiene

$$f_{\phi|X}(\phi|x) \propto \phi^{\frac{(n+1)}{2}-1} \exp \left[-\frac{\phi}{2} \left\{ \frac{(x-m)^2}{S^2} + d \right\} \right] \quad (\text{A9})$$

por lo que

$$(\phi|X=x) \sim G \left[n^*/2, d^*/2 \right]$$

donde

$$n^* = n+1 \text{ y } d^* = d + \frac{(x-m)^2}{S^2}.$$

Distribución Marginal

Combinando la densidad conjunta (A8) con la densidad condicional (A9) se obtiene

$$f_X(x) = \frac{f_{X,\phi}(x,\phi)}{f_{\phi|X}(\phi|x)} \quad -\infty < x < \infty.$$
$$\propto \left[n + \frac{(x-m)^2}{\tilde{S}^2} \right]^{-\frac{(n+1)}{2}}$$

donde

$$\tilde{S}^2 = S^2 \left(\frac{d}{n} \right) = \frac{S^2}{E[\phi]}.$$

A partir de la forma de $f_X(x)$ se deduce que X tiene distribución *t de Student* con n grados de libertad, moda m y escala \tilde{S}^2

$$X \sim T_n[m, \tilde{S}^2].$$

Estandarizando se tiene que

$$\frac{X-m}{\sqrt{\tilde{S}^2}} \sim T_n[0,1].$$

Distribución Normal-Gamma Generalizada

Sea Φ una v.a. tal que $\Phi \sim G\left[\frac{n}{2}, \frac{d}{2}\right]$ y sea $\mathbf{X} = (X_1, \dots, X_p)$ un p -vector aleatorio con distribución normal condicional sobre Φ

$$\mathbf{X} \sim N\left[\mathbf{m}, \frac{1}{\Phi} \mathbf{S}^2\right]$$

con \mathbf{m} un vector conocido de dimensión p y \mathbf{S}^2 una matriz de $p \times p$, conocida, simétrica y positiva definida. Entonces \mathbf{X} y Φ tienen distribución conjunta *Normal-Gamma* y en consecuencia

$$(\Phi | \mathbf{X} = \mathbf{x}) \sim G\left[\frac{n^*}{2}, \frac{d^*}{2}\right]$$

donde

$$n^* = n + p \text{ y } d^* = d + (\mathbf{x} - \mathbf{m})' (\mathbf{S}^2)^{-1} (\mathbf{x} - \mathbf{m}).$$

La distribución marginal de \mathbf{X} es *t de Student multivariada*, con moda \mathbf{m} y matriz de escala $\tilde{\mathbf{S}}^2 = \left(\frac{d}{n}\right) \mathbf{S}^2 = \frac{\mathbf{S}^2}{E(\Phi)}$,

$$\mathbf{X} \sim T_n[\mathbf{m}, \tilde{\mathbf{S}}^2],$$

con densidad

$$f(\mathbf{X}) \propto \left[n + (\mathbf{x} - \mathbf{m})' (\tilde{\mathbf{S}}^2)^{-1} (\mathbf{x} - \mathbf{m}) \right]^{-\frac{(n+p)}{2}}.$$

En particular, si X_i es el i -ésimo elemento de \mathbf{X} ,

$$X_i \sim T_n[m_i, \tilde{S}_i^2],$$

donde

$$\tilde{S}_i^2 = \frac{d}{n} S_{ii}^2.$$

FALTA PAGINA

No. 176 a la 178

BIBLIOGRAFIA

- **BERGER, J. O.**, [1985]. *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag, New York.
- **BOX, G. E. P. y TIAO, G. C.**, [1973]. *Bayesian Inference in Statistical Analysis*. Addison-Wesley, Massachusetts.
- **BROCKWELL, P. J. y DAVIS, R. A.**, *Time Series: Theory and Methods*.
- **DE FINEETTI, B.**, [1974,1975]. *Theory of Probability (vols. 1 y 2)*. Wiley, Chichester.
- **HARRISON, P. J. y STEVENS, C. F.**, [1976]. Bayesian Forecasting. *J. Royal Statist. Soc. (serie B)*. **38**, 205-247.
- **HARVEY, ANDREW C.** [1989]. *Forecasting, structural time series models an the Kalman filter*. Cambridge University Press, Cambridge.
- **JUDGE, G. G. et al**, [1985]. *Introduction to the Theory and Practice of Econometrics*. John Wiley & sons, New York.
- **KENNEY, J. F. y KEEPING, E.S.**, *Mathematics of Statistics (parte 2)*. D. Van Nostrand Company.
- **LINDLEY, D. V.**, [1965]. *Introduction to Probability and Statistics from a bayesian viewpoint (partes 1 y 2)*. Cambridge University Press, Cambridge.
- **MAKRIDAKIS, S., WHEELWRIGHT, S. C. y McGEE, V.**, [1983]. *Forecasting. Methods and Applications*. John Wiley & sons, New York.
- **MOOD, A. M. y GRAYBILL, F. A.**, [1974]. *Introduction to the Theory of Statistics*. McGraw-Hill, New York.

- PINDYCK, R. S. y RUBINFELD, D. L., [1981]. *Econometric Models and Economic Forecasts*. New York , McGraw-Hill.
- SAVAGE, L. J., [1971]. *The foundations of Statistics*. Dover, New York.
- SMITH, J. Q., [1979]. A Generalization of the Bayesian Steady Forecasting Model. *J. Royal. Statist. Soc. (serie B)*, **41**, 375-387.
- WEST, M. y HARRISON, P. J., [1989]. *Bayesian forecasting and dynamical models*. Springer-Verlag, New York.
- WEST, M. y HARRISON, P.J., [1986]. Monitoring and adaptation in Bayesian Forecasting Models. *J. American Statist. Asso.* **81**, 741-750.
- WEST, M., HARRISON, J. y MIGON, H. S., [1985]. Dynamic Generalized Models and Bayesian Forecasting. *J. American Statist. Asso.* **80**, 73-85.

BIBLIOGRAFIA

- BERGER, J. O., [1985]. *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag, New York.
- BOX, G. E. P. y TIAO, G. C., [1973]. *Bayesian Inference in Statistical Analysis*. Addison-Wesley, Massachusetts.
- BROCKWELL, P. J. y DAVIS, R. A., *Time Series: Theory and Methods*.
- DE FINEETTI, B., [1974,1975]. *Theory of Probability (vols. 1 y 2)*. Wiley, Chichester.
- HARRISON, P. J. y STEVENS, C. F., [1976]. Bayesian Forecasting. *J. Royal Statist. Soc. (serie B)*. **38**, 205-247.
- HARVEY, ANDREW C. [1989]. *Forecasting, structural time series models an the Kalman filter*. Cambridge University Press, Cambridge.
- JUDGE, G. G. et al, [1985]. *Introduction to the Theory and Practice of Econometrics*. John Wiley & sons, New York.
- KENNEY, J. F. y KEEPING, E.S., *Mathematics of Statistics (parte 2)*. D. Van Nostrand Company.
- LINDLEY, D. V., [1965]. *Introduction to Probability and Statistics from a bayesian viewpoint (partes 1 y 2)*. Cambridge University Press, Cambridge.
- MAKRIDAKIS, S., WHEELWRIGHT, S. C. y MCGEE, V., [1983]. *Forecasting. Methods and Applications*. John Wiley & sons, New York.
- MOOD, A. M. y GRAYBILL, F. A., [1974]. *Introduction to the Theory of Statistics*. McGraw-Hill, New York.

- PINDYCK, R. S. y RUBINFELD, D. L., [1981]. *Econometric Models and Economic Forecasts*. New York , McGraw-Hill.
- SAVAGE, L. J., [1971]. *The foundations of Statistics*. Dover, New York.
- SMITH, J. Q., [1979]. A Generalization of the Bayesian Steady Forecasting Model. *J. Royal. Statist. Soc. (serie B)*. **41**, 375-387.
- WEST, M. y HARRISON, P. J., [1989]. *Bayesian forecasting and dynamical models*. Springer-Verlag, New York.
- WEST, M. y HARRISON, P.J., [1986]. Monitoring and adaptation in Bayesian Forecasting Models. *J. American Statist. Asso.* **81**, 741-750.
- WEST, M., HARRISON, J. y MIGON, H. S., [1985]. Dynamic Generalized Models and Bayesian Forecasting. *J. American Statist. Asso.* **80**, 73-85.