



UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO

FACULTAD DE CIENCIAS

Modelos de espacio de estados para el análisis de series de
tiempo

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

Matemático

PRESENTA:

Alfie Sergio González Salcedo

TUTOR

Dr. Alberto Contreras Cristán





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.

1. Datos del alumno

González
Salcedo
Alfie Sergio
58 73 38 46
Universidad Nacional Autónoma de México
Facultad de Ciencias
Matemáticas
310578084

2. Datos del tutor

Dr.
Alberto
Contreras
Cristán

3. Datos del sinodal 1

Dr.
Luis Antonio
Rincón
Solís

4. Datos del sinodal 2

Dr.
Ricardo
Ramírez
Aldana

5. Datos del sinodal 3

Dra.
Lizbeth
Naranjo
Albarrán

6. Datos del sinodal 4

M. en C.
José Salvador
Zamora
Muñoz

7. Datos del trabajo escrito

Modelos de espacio de estados para el análisis de series de tiempo
105 p.
2018

Agradecimientos

Agradezco a la Universidad Nacional Autónoma de México por haberme dado la oportunidad de estudiar una carrera; a mi tutor, el Dr. Alberto Contreras Cristán, por su paciencia y asesoramiento para la realización de este trabajo; a las personas que se tomaron el tiempo para revisar este escrito y aportar comentarios que permitieran el mejoramiento del mismo; a mis profesores, por transmitirme sus conocimientos y motivarme a seguir aprendiendo; a mis amigos y compañeros, por su constante apoyo; a mi hermana, por ser desde siempre un ejemplo para mí. Finalmente, y sobre todo, agradezco a mis padres; a ellos les debo todo lo que tengo.

Índice general

Introducción	7
1. Clasificación de los métodos de suavizamiento exponencial	11
1.1. Tipos de tendencia	11
1.2. Pronósticos puntuales para algunos modelos conocidos	12
1.2.1. Suavizamiento exponencial simple (N,N)	12
1.2.2. Modelo lineal de Holt (A, N)	14
1.2.3. Modelo con tendencia que se desvanece (A_d, N)	14
1.2.4. Modelo con tendencia y estacionalidad aditivas (A,A) (Método de Holt-Winters aditivo)	15
1.2.5. Modelo con tendencia aditiva y estacionalidad multiplicativa (A,M) (Método de Holt-Winters multiplicativo)	15
2. Modelo lineal general de espacio de estados	17
2.1. Ecuaciones del modelo	17
2.2. Pronósticos a un paso	20
2.3. Serie de tiempo pronosticable	21
2.4. Estacionariedad débil	22
2.5. Ejemplos de modelos	23
2.5.1. Modelo de suavizamiento exponencial simple ETS(A,N,N)	24
2.5.2. Modelo lineal de Holt ETS(A,A,N)	25
2.5.3. Modelo de Holt Winters aditivo ETS(A,A,A)	26
2.5.4. Modelo de suavizamiento exponencial con tendencia	29
2.5.5. Modelo con tendencia que se termina ETS(A, A_d ,N)	31
2.5.6. Modelos compuestos	32
3. Modelo general de innovaciones de espacio de estados	33
3.1. Ecuaciones del modelo	33
3.2. Ejemplos de modelos	37
3.2.1. Modelo de nivel local con error multiplicativo ETS(M,N,N)	37
3.2.2. Modelo con error multiplicativo y tendencia aditiva ETS(M,A,N)	39
3.2.3. Modelo con error aditivo y tendencia multiplicativa ETS(A,M,N)	40
3.2.4. Modelo con error y tendencia multiplicativas ETS(M,M,N)	41
3.2.5. Modelo con error multiplicativo, tendencia aditiva y estacionalidad multiplicativa ETS(M,A,M)	41

3.2.6. Modelo con error multiplicativo y tendencia aditiva que se termina ETS(M,A _d , N)	46
3.2.7. Modelo con error aditivo y tendencia multiplicativa que se termina ETS(A,M _d ,N)	47
4. Estimación de parámetros iniciales	49
4.1. Función de máxima verosimilitud	49
4.2. Método heurístico	51
4.2.1. Promedio móvil	51
4.2.2. $2 \times m$ promedio móvil	52
4.2.3. Componente estacional inicial	52
4.2.3.1. Normalización de índices estacionales	53
4.2.4. Componente inicial del nivel	56
4.2.5. Componente inicial del factor de crecimiento	56
5. Intervalos de predicción para pronósticos puntuales	57
5.1. Media y varianza de predicción	57
5.2. Ejemplo de intervalo de predicción del modelo ETS(A,A,N)	60
5.3. Ejemplo de intervalo de predicción del modelo ETS(M,A,M)	70
6. Selección de modelos	79
6.1. Criterios de información	79
6.2. Tipos de errores para medir exactitud en un modelo	80
6.2.1. Error absoluto medio	80
6.2.2. Error cuadrático medio	81
6.2.3. Error porcentual	81
6.2.4. Error absoluto porcentual medio	81
6.2.5. Error simétrico absoluto porcentual medio	81
6.2.6. Errores escalados	82
7. Caso práctico	83
Conclusión	103
Bibliografía	105

Introducción

El objetivo de este trabajo es analizar series de tiempo mediante el uso de métodos de suavizamiento exponencial (*exponential smoothing*), introduciendo además la estructura de modelos de espacio de estados para la obtención de intervalos de predicción.

Se comenzará haciendo una revisión de algunos métodos de suavizamiento exponencial comunes. Después se procederá introduciendo el modelo lineal de espacio de estados, señalando sus propiedades y proporcionando ejemplos de algunos modelos. Posteriormente se abordará un caso más general, donde se verán modelos que pueden aplicarse incluso a series no estacionarias. Una vez hecho esto, se estudiará el proceso de estimación de parámetros iniciales, para después describir el proceso de obtención de intervalos de predicción de pronósticos puntuales. Luego se hará mención de los criterios de información para la selección de modelos y finalmente se examinará un caso práctico en el que se realizará el análisis de una serie de tiempo mediante la metodología estudiada en este trabajo.

Para comenzar, hagamos una revisión de la estructura de una serie de tiempo. Cuando se tiene una serie de tiempo, es común considerar un modelo en el cual los datos constan de ciertas componentes tales como tendencia T , estacionalidad S , una componente cíclica C y una componente de error E . Recordemos que la tendencia indica el movimiento ascendente o descendente a largo plazo en la serie. La estacionalidad se refiere a un patrón que se repite con una periodicidad conocida (por ejemplo, cada 12 meses). La componente cíclica es igualmente un patrón que se repite; sin embargo, éste tiene una periodicidad desconocida y cambiante. La componente de error representa la parte irregular (impredecible) en la serie. En este estudio consideraremos únicamente modelos que tengan a lo más tendencia, estacionalidad y una componente de error.

Estas componentes pueden combinarse de distintas maneras para establecer un modelo que explique la serie. Por ejemplo, un modelo puramente aditivo para una serie y sería:

$$y = T + S + E.$$

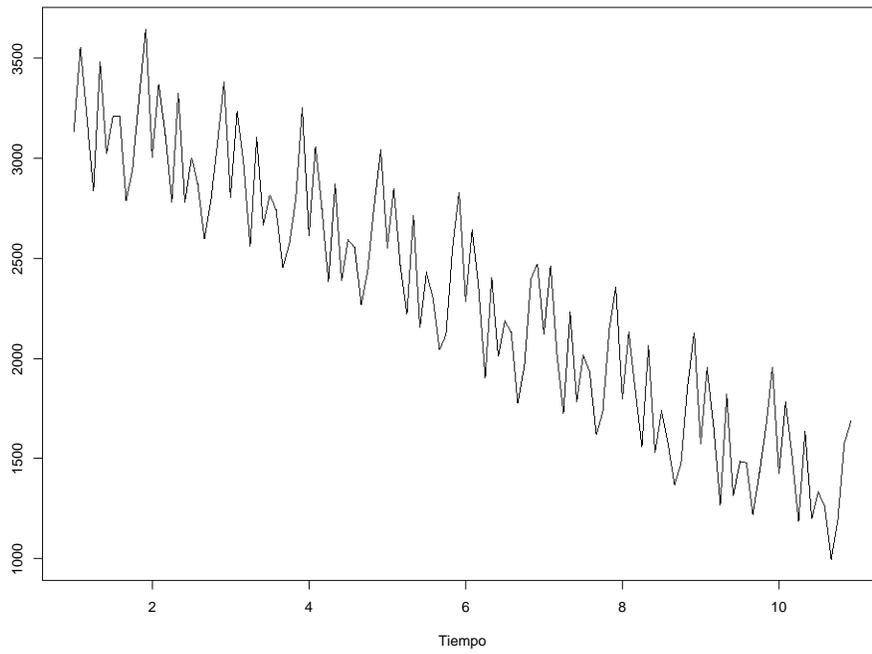


Figura 1: Simulación en R de un modelo puramente aditivo.

Otro ejemplo sería un modelo puramente multiplicativo:

$$y = T \times S \times E.$$

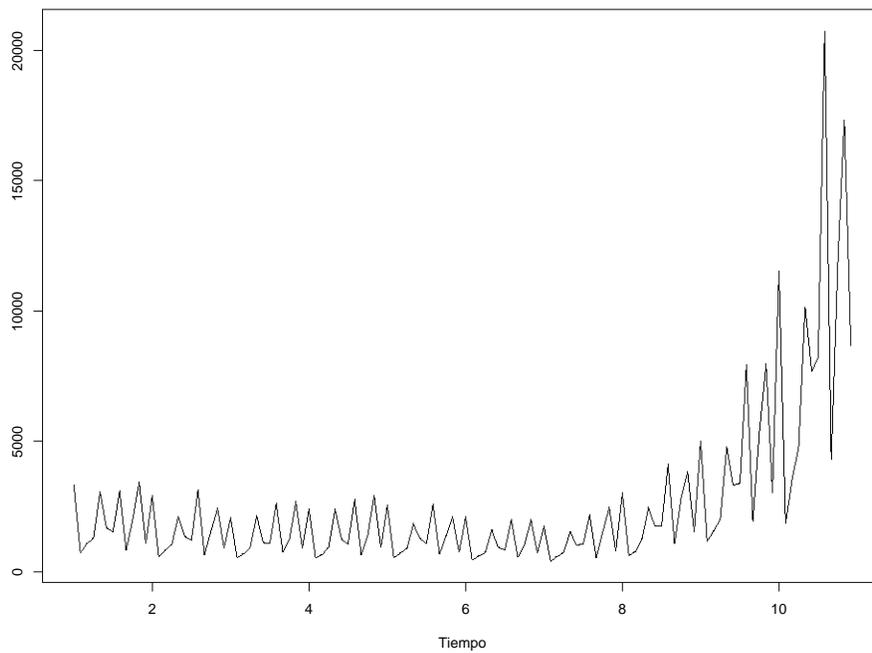


Figura 2: Simulación en R de un modelo puramente multiplicativo.

También puede haber una combinación de uno y otro:

$$y = (T + S) \times E.$$

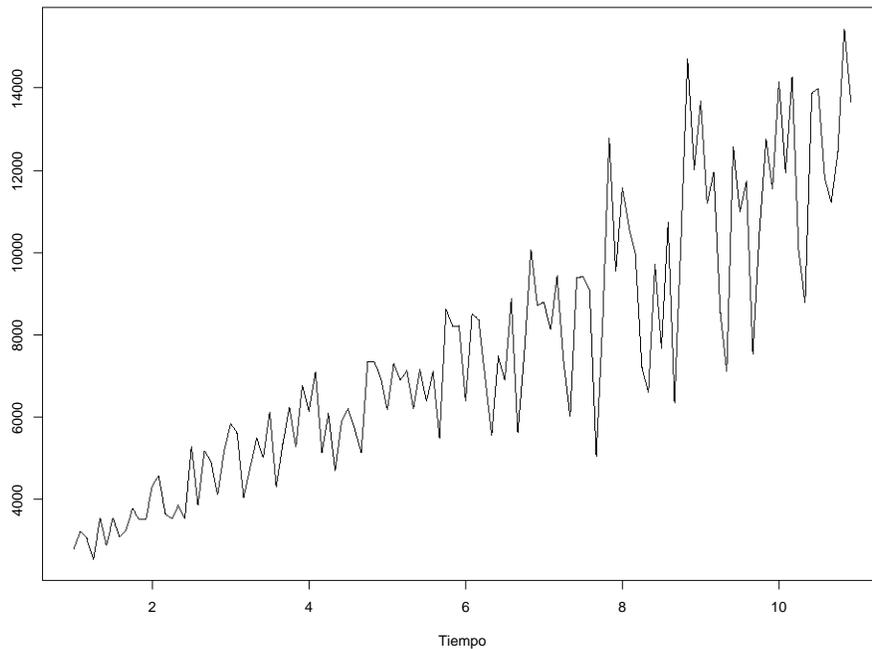


Figura 3: Simulación en R de un modelo combinado.

Estos modelos son útiles porque, como se verá a lo largo de este trabajo, emplean variables auxiliares que permiten modelar y estimar la tendencia y la estacionalidad, lo cual no hacen otros modelos como los AR, ARMA, ARIMA o SARIMA. Incluso en casos en los que se busca estimar tendencia y estacionalidad mediante la descomposición clásica, el proceso de optimización de la función de verosimilitud puede volverse complejo. Además, estos modelos no requieren de transformaciones para modelar series que no son estacionarias; se aplican directamente a la serie. De este modo, se evitan casos como el de los operadores de diferencias en los modelos ARIMA o SARIMA, en los cuales, para transformar la serie original en un proceso estacionario, se pierden observaciones.

Capítulo 1

Clasificación de los métodos de suavizamiento exponencial

1.1. Tipos de tendencia

Los modelos de suavizamiento exponencial consideran un término de tendencia que es combinación de un término de nivel ℓ y un término de crecimiento b . Sea h el número de periodos transcurridos y sea $0 < \phi < 1$. Sea T_h la tendencia a h pasos. Se tienen 5 diferentes modelos para T_h ; éstos se muestran en el cuadro 1.1.

None	$T_h = \ell$
Additive	$T_h = \ell + bh$
Additive damped	$T_h = \ell + (\phi + \phi^2 + \dots + \phi^h)b$
Multiplicative	$T_h = \ell b^h$
Multiplicative damped	$T_h = \ell b^{\phi + \phi^2 + \dots + \phi^h}$

Cuadro 1.1: Tipos de tendencia.

Los modelos de tendencia aditiva que se desvanece (*additive damped*) y de tendencia multiplicativa que se desvanece (*multiplicative damped*) se usan cuando se piensa que la tendencia está a punto de terminarse. La componente estacional puede aparecer en los modelos en forma aditiva o multiplicativa. En el caso de la componente de error, de igual manera éste puede ser del tipo aditivo o multiplicativo.

La combinación de las tendencias en la tabla, con los tipos de estacionalidad (aditiva, multiplicativa, e inexistente), dan como resultado 15 modelos, los cuales se muestran en el cuadro 1.2.

Componente de tendencia	Componente estacional		
	N (None)	A (Additive)	M (Multiplicative)
N (None)	N,N	N,A	N,M
A (Additive)	A,N	A,A	A,M
A _d (Additive damped)	A _d ,N	A _d ,A	A _d ,M
M (Multiplicative)	M,N	M,A	M,M
M _d (Multiplicative damped)	M _d ,N	M _d ,A	M _d ,M

Cuadro 1.2: Combinaciones de modelos según el tipo de tendencia y el tipo de componente estacional.

Hay que mencionar que algunos de éstos métodos tienen nombres conocidos en la práctica. El modelo N,N, es decir, aquel en el que no hay ni tendencia ni componente estacional, se conoce como suavizamiento exponencial simple (*Simple Exponential Smoothing, SES*). El modelo A,N se conoce como método lineal de Holt. El modelo A,A es el método de Holt-Winters aditivo (véase, por ejemplo, [2, sección 10.2]) y el modelo A,M es el método Holt-Winters multiplicativo. El modelo A_d, N es el modelo con tendencia que está por terminarse.

Si se considera además la parte del error, se tienen entonces 30 modelos: 15 donde el error es aditivo y 15 donde el error es multiplicativo. Se utilizará la notación ETS(, ,) para referirse a un modelo. Esto es conveniente, puesto que las letras permiten identificar fácilmente la componente a la que se refieren (E = Error, T = Trend, S = Stationality). Por ejemplo, el modelo ETS(A,A,N) tiene error de tipo aditivo y tendencia igualmente aditiva; en este caso no hay estacionalidad presente. Notemos que este modelo es el método lineal de Holt.

1.2. Pronósticos puntuales para algunos modelos conocidos

Veamos de manera introductoria algunos de los modelos más conocidos del método de suavizamiento exponencial. Vamos a considerar que se tienen una serie de datos observados y_1, y_2, \dots, y_t . Sea $\hat{y}_{t+h|t}$ el pronóstico para y_{t+h} basado en la información que se ha observado hasta tiempo t y $h = 1, 2, \dots$ es el número de pasos adelante. En el caso del pronóstico a un paso denotaremos simplemente $\hat{y}_{t+1} \equiv \hat{y}_{t+1|t}$. En las notaciones del cuadro 1.1, cuando todavía no se ha contemplado una componente estacional, se tiene que $\hat{y}_{t+h|t} \equiv T_h$. Se hará el uso de ciertos parámetros, los cuales es necesario estimar. No obstante, por ahora se supondrá que éstos ya fueron estimados.

1.2.1. Suavizamiento exponencial simple (N,N)

Si se ha observado una serie hasta tiempo t , el modelo de suavizamiento exponencial simple, según la metodología de Brown a mediados de la década de 1950, dice que pronóstico a un paso está dado por:

$$\hat{y}_{t+1} = \hat{y}_t + \alpha(y_t - \hat{y}_t), \quad (1.1)$$

donde $0 < \alpha < 1$.

Si consideramos que el error del pronóstico a tiempo t está dado por la diferencia entre el valor observado y el pronóstico, entonces para este modelo el pronóstico a un paso es la suma del pronóstico a tiempo t y una constante por el error en el paso anterior. Dependiendo del valor de α , el error en el paso anterior puede tener mayor o menor impacto en el pronóstico.

Observemos que:

$$\begin{aligned} \hat{y}_{t+1} &= \hat{y}_t + \alpha(y_t - \hat{y}_t) \\ &= \alpha y_t + (1 - \alpha)\hat{y}_t \\ &= \alpha y_t + (1 - \alpha)[\alpha y_{t-1} + (1 - \alpha)\hat{y}_{t-1}] \\ &= \alpha y_t + \alpha(1 - \alpha)y_{t-1} + (1 - \alpha)^2\hat{y}_{t-1} \\ &= \alpha y_t + \alpha(1 - \alpha)y_{t-1} + (1 - \alpha)^2[\alpha y_{t-2} + (1 - \alpha)\hat{y}_{t-2}] \\ &= \alpha y_t + \alpha(1 - \alpha)y_{t-1} + \alpha(1 - \alpha)^2 y_{t-2} + (1 - \alpha)^3 \hat{y}_{t-3} \\ &\vdots \\ &= \alpha y_t + \alpha(1 - \alpha)y_{t-1} + \alpha(1 - \alpha)^2 y_{t-2} + \cdots + \alpha(1 - \alpha)^{t-1} y_1 + (1 - \alpha)^t \hat{y}_1. \end{aligned}$$

De modo que el pronóstico a un paso está determinado al promediar de manera ponderada los valores anteriores observados. Los pesos asignados a cada observación van decreciendo de manera exponencial (de ahí el nombre de esta familia de modelos).

Para tiempos sucesivos, supondremos que la función de pronóstico es “plana”; es decir:

$$\hat{y}_{t+h|t} = \hat{y}_{t+1},$$

con $h = 2, 3, \dots$

Esto se debe a que este modelo es adecuado para series en las que no hay tendencia, estacionalidad ni otro tipo de componentes.

Ahora, podemos definir $\ell_t = \hat{y}_{t+1}$, de modo que $\hat{y}_{t+h|t} = \ell_t$. De esta manera $\ell_t = \alpha y_t + (1 - \alpha)\ell_{t-1}$. Esta variable, como anteriormente se mencionó, es el nivel de la serie a tiempo t . Como se puede observar, ℓ_t está determinada a su vez por ℓ_{t-1} ; ésta a su vez por ℓ_{t-2} , y así de manera iterativa. Es necesario especificar el valor de $\ell_0 = \hat{y}_1$, así como el valor de α .

1.2.2. Modelo lineal de Holt (A, N)

Este modelo se usa para series que tienen tendencia aditiva. Requiere de dos parámetros, α y β^* , con valores entre 0 y 1. El modelo está especificado como sigue:

$$\text{Nivel: } \ell_t = \alpha y_t + (1 - \alpha)(\ell_{t-1} + b_{t-1}), \quad (1.2a)$$

$$\text{Crecimiento: } b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)b_{t-1}, \quad (1.2b)$$

$$\text{Pronóstico: } \hat{y}_{t+h|t} = \ell_t + b_t h. \quad (1.2c)$$

La variable ℓ_t es el nivel de la serie. La variable b_t estima la tasa de crecimiento de la serie.

Para el caso particular en que $\beta^* = 0$, se tiene que:

$$\text{Crecimiento: } b_t = b_{t-1} = b_{t-2} = \dots = b_0 = b \text{ (constante).}$$

De modo que el modelo queda determinado por:

$$\text{Nivel: } \ell_t = \alpha y_t + (1 - \alpha)(\ell_{t-1} + b), \quad (1.3a)$$

$$\text{Pronóstico: } \hat{y}_{t+h|t} = \ell_t + b h. \quad (1.3b)$$

Este modelo se conoce como el modelo de suavizamiento exponencial con tendencia.

1.2.3. Modelo con tendencia que se desvanece (A_d,N)

Este modelo está especificado por:

$$\text{Nivel: } \ell_t = \alpha y_t + (1 - \alpha)(\ell_{t-1} + \phi b_{t-1}), \quad (1.4a)$$

$$\text{Crecimiento: } b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)\phi b_{t-1}, \quad (1.4b)$$

$$\text{Pronóstico: } \hat{y}_{t+h|t} = \ell_t + (\phi + \phi^2 + \dots + \phi^h)b_t. \quad (1.4c)$$

Cuando $\phi = 1$, este modelo es el modelo lineal de Holt. Cuando $0 < \phi < 1$, recordando que $\phi + \phi^2 + \dots + \phi^h = \frac{\phi - \phi^{h+1}}{1 - \phi} \xrightarrow{h \rightarrow \infty} \frac{\phi}{1 - \phi}$, se tiene que el pronóstico tiende asintóticamente a $\ell_t + \frac{\phi b_t}{1 - \phi}$. El caso $\phi < 0$ no se considera para no tener coeficientes negativos para b_t . Tampoco se va a considerar el caso $\phi > 1$, ya que de este modo b_t crecería exponencialmente.

Para modelos que consideran estacionalidad, ésta puede ser aditiva o multiplicativa. Un par de ejemplos se exhibirán a continuación.

1.2.4. Modelo con tendencia y estacionalidad aditivas (A,A) (Método de Holt-Winters aditivo)

Este modelo está determinado por las ecuaciones:

$$\text{Nivel: } \ell_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)(\ell_{t-1} + b_{t-1}), \quad (1.5a)$$

$$\text{Crecimiento: } b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)b_{t-1}, \quad (1.5b)$$

$$\text{Estacionalidad: } s_t = \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1 - \gamma)s_{t-m}, \quad (1.5c)$$

$$\text{Pronóstico: } \hat{y}_{t+h|t} = \ell_t + b_t h + s_{t-m+h_m^+}, \quad (1.5d)$$

donde se introduce la componente de estacionalidad s_t ; m es la longitud de la estacionalidad (por ejemplo, $m = 12$ para datos mensuales; $m = 4$ para datos trimestrales), y $h_m^+ = [(h - 1) \bmod m] + 1$. Los parámetros α , β^* , γ suelen considerarse entre 0 y 1.

1.2.5. Modelo con tendencia aditiva y estacionalidad multiplicativa (A,M) (Método de Holt-Winters multiplicativo)

El modelo con tendencia y estacionalidad aditivas no es muy común en la práctica; es más común considerar estacionalidad multiplicativa. Las ecuaciones del modelo son:

$$\text{Nivel: } \ell_t = \alpha \frac{y_t}{s_{t-m}} + (1 - \alpha)(\ell_{t-1} + b_{t-1}), \quad (1.6a)$$

$$\text{Crecimiento: } b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)b_{t-1}, \quad (1.6b)$$

$$\text{Estacionalidad: } s_t = \gamma y_t / (\ell_{t-1} + b_{t-1}) + (1 - \gamma)s_{t-m}, \quad (1.6c)$$

$$\text{Pronóstico: } \hat{y}_{t+h|t} = (\ell_t + b_t h) s_{t-m+h_m^+}. \quad (1.6d)$$

Los modelos anteriores establecen recursiones para encontrar pronósticos puntuales $\hat{y}_{t+h|t}$ en cada caso, dependiendo de las componentes de tendencia y estacionalidad que el usuario crea procedente modelar en los datos. No obstante, para obtener intervalos de predicción, se verá en el capítulo 2 que es necesario relacionar estos modelos con modelos de espacio de estados.

Capítulo 2

Modelo lineal general de espacio de estados

2.1. Ecuaciones del modelo

Cada uno de los modelos expuestos es un método de suavizamiento exponencial ideado para obtener pronósticos puntuales, pero también cada modelo se puede escribir como un modelo de espacio de estados (véase [1]), con el objetivo de estudiar la posibilidad de obtener intervalos de predicción para cada método. Ahora vamos a considerar un modelo general de espacio de estados que incluya todos los casos de suavizamiento exponencial. Este modelo incluye una componente de error, de la cual se va a suponer que sigue una distribución de probabilidad. Los pronósticos puntuales obtenidos son los mismos que los que se obtienen con la metodología anteriormente expuesta. Sin embargo, el modelo de espacio de estados permite, además, obtener intervalos de predicción para dichos pronósticos.

Este modelo utiliza variables auxiliares llamadas estados, que contienen información pasada de la serie, la cual se utiliza para pronosticar valores futuros de la misma. Vamos a considerar que dicha información está contenida en un vector \mathbf{x}_t , llamado *vector de estados*.

Si se tiene una serie de tiempo con valores observados y_1, y_2, \dots, y_t , el modelo de espacio de estados está dado por:

$$y_t = \mathbf{w}'\mathbf{x}_{t-1} + \varepsilon_t, \quad (2.1a)$$

$$\mathbf{x}_t = \mathbf{F}\mathbf{x}_{t-1} + \mathbf{g}\varepsilon_t, \quad (2.1b)$$

donde \mathbf{x}_t es el vector de estados de dimensión $k \times 1$; \mathbf{w} y \mathbf{g} son vectores constantes de dimensión $k \times 1$, y \mathbf{F} es una matriz de $k \times k$ con entradas constantes.

El vector \mathbf{x}_t contiene las variables de nivel ℓ_t , crecimiento b_t y estacionalidad $\{s_t, s_{t-1}, \dots, s_{t-m+1}\}$. Si se trata de un caso en que la serie tiene tendencia y estacionalidad, entonces $\mathbf{x}_t = (\ell_t, b_t, s_t, s_{t-1}, \dots, s_{t-m+1})'$. La variable aleatoria ε_t es conocido como *innovación*, y supondremos que las variables $\{\varepsilon_t\}_t$ son independientes e idénticamente distribuidas $N(0, \sigma^2)$. Denotaremos esto como $\varepsilon_t \sim NID(0, \sigma^2)$. Ésta es la componente que agrega un carácter irregular a la serie.

La ecuación 2.1a se conoce como *ecuación de medición* y la ecuación 2.1b como *ecuación de transición*; \mathbf{F} se conoce como la *matriz de transición*. El vector \mathbf{x}_0 puede ser fijo o aleatorio, pero en este trabajo consideraremos que \mathbf{x}_0 es fijo.

Observemos que:

$$\begin{aligned} \mathbf{x}_t &= \mathbf{F}\mathbf{x}_{t-1} + \mathbf{g}\varepsilon_t \\ &= \mathbf{F}\mathbf{x}_{t-1} + \mathbf{g}(y_t - \mathbf{w}'\mathbf{x}_{t-1}) \\ &= \mathbf{F}\mathbf{x}_{t-1} - \mathbf{g}\mathbf{w}'\mathbf{x}_{t-1} + \mathbf{g}y_t \\ &= (\mathbf{F} - \mathbf{g}\mathbf{w}')\mathbf{x}_{t-1} + \mathbf{g}y_t. \end{aligned}$$

Sea $\mathbf{D} = \mathbf{F} - \mathbf{g}\mathbf{w}'$. Entonces:

$$\begin{aligned} \mathbf{x}_t &= \mathbf{D}\mathbf{x}_{t-1} + \mathbf{g}y_t \\ &= \mathbf{D}(\mathbf{D}\mathbf{x}_{t-2} + \mathbf{g}y_{t-1}) + \mathbf{g}y_t \\ &= \mathbf{D}^2\mathbf{x}_{t-2} + \mathbf{D}\mathbf{g}y_{t-1} + \mathbf{g}y_t \\ &= \mathbf{D}^2(\mathbf{D}\mathbf{x}_{t-3} + \mathbf{g}y_{t-2}) + \mathbf{D}\mathbf{g}y_{t-1} + \mathbf{g}y_t \\ &= \mathbf{D}^3\mathbf{x}_{t-3} + \mathbf{D}^2\mathbf{g}y_{t-2} + \mathbf{D}\mathbf{g}y_{t-1} + \mathbf{g}y_t \\ &= \\ &\vdots \\ &= \mathbf{D}^t\mathbf{x}_0 + \sum_{j=0}^{t-1} \mathbf{D}^j\mathbf{g}y_{t-j}. \end{aligned}$$

En particular:

$$\begin{aligned} \mathbf{x}_{t-1} &= \mathbf{D}^{t-1}\mathbf{x}_0 + \sum_{j=0}^{t-2} \mathbf{D}^j\mathbf{g}y_{t-1-j} \\ &= \mathbf{D}^{t-1}\mathbf{x}_0 + \sum_{j=1}^{t-1} \mathbf{D}^{j-1}\mathbf{g}y_{t-j}. \end{aligned} \tag{2.2}$$

Si $\mathbf{y} = [y_1, y_2, \dots, y_n]$, entonces la función de densidad de \mathbf{y} está dada por:

$$\begin{aligned}
p(\mathbf{y} \mid \mathbf{x}_0) &= p(y_n, y_{n-1}, \dots, y_1 \mid \mathbf{x}_0) \\
&= p(y_n, y_{n-1}, \dots, y_1, \mathbf{x}_0) \frac{1}{p(\mathbf{x}_0)} \\
&= p(y_n \mid y_{n-1}, \dots, y_1, \mathbf{x}_0) p(y_{n-1}, \dots, y_1, \mathbf{x}_0) \frac{1}{p(\mathbf{x}_0)} \\
&= p(y_n \mid y_{n-1}, \dots, y_1, \mathbf{x}_0) p(y_{n-1} \mid y_{n-2}, \dots, y_1, \mathbf{x}_0) p(y_{n-2}, \dots, y_1, \mathbf{x}_0) \frac{1}{p(\mathbf{x}_0)} \\
&= \\
&\vdots \\
&= \prod_{t=1}^n p(y_t \mid y_{t-1}, \dots, y_1, \mathbf{x}_0).
\end{aligned} \tag{2.3}$$

Anteriormente se vio que si $\mathbf{D} = \mathbf{F} - \mathbf{g}\mathbf{w}'$. Entonces $\mathbf{x}_t = \mathbf{D}\mathbf{x}_{t-1} + \mathbf{g}y_t$, de modo que:

$$\begin{aligned}
y_t &= \mathbf{w}'\mathbf{x}_{t-1} + \varepsilon_t \\
&= \mathbf{w}'(\mathbf{D}\mathbf{x}_{t-2} + \mathbf{g}y_{t-1}) + \varepsilon_t \\
&= \mathbf{w}'\mathbf{D}\mathbf{x}_{t-2} + \mathbf{w}'\mathbf{g}y_{t-1} + \varepsilon_t \\
&= \mathbf{w}'\mathbf{D}(\mathbf{D}\mathbf{x}_{t-3} + \mathbf{g}y_{t-2}) + \mathbf{w}'\mathbf{g}y_{t-1} + \varepsilon_t \\
&= \\
&\vdots \\
&= \mathbf{w}'\mathbf{D}^{t-1}\mathbf{x}_0 + \mathbf{w}' \left[\sum_{j=1}^{t-1} \mathbf{D}^{j-1} \mathbf{g}y_{t-j} \right] + \varepsilon_t.
\end{aligned} \tag{2.4}$$

De la relación 2.4, condicionalmente a y_1, y_2, \dots, y_{t-1} y x_0 , y_t tiene una distribución normal con media $\mu = \mathbf{w}'\mathbf{D}^{t-1}\mathbf{x}_0 + \mathbf{w}' \left[\sum_{j=1}^{t-1} \mathbf{D}^{j-1} \mathbf{g}y_{t-j} \right]$ y varianza σ^2 . Por otra parte, de la

relación 2.2, $\mathbf{x}_{t-1} = \mathbf{D}^{t-1}\mathbf{x}_0 + \sum_{j=1}^{t-1} \mathbf{D}^{j-1}\mathbf{g}y_{t-j}$. Es decir, condicionalmente a y_1, y_2, \dots, y_{t-1}

y x_0 , y_t tiene distribución $N(\mathbf{w}'\mathbf{x}_{t-1}, \sigma^2)$, pero como $y_t = \mathbf{w}'\mathbf{x}_{t-1} + \varepsilon_t$, donde ε_t se distribuye $N(0, \sigma^2)$, entonces, condicionalmente a \mathbf{x}_{t-1} , y_t tiene distribución normal con media

$\mu = \mathbf{w}'\mathbf{x}_{t-1} = \mathbf{w}'\mathbf{D}^{t-1}\mathbf{x}_0 + \mathbf{w}' \sum_{j=1}^{t-1} \mathbf{D}^{j-1} \mathbf{g}y_{t-j}$ y varianza σ^2 , de modo que $y_t \mid y_{t-1}, \dots, y_1, \mathbf{x}_0$

y $y_t \mid \mathbf{x}_{t-1}$ tienen la misma función de densidad. Es decir:

$$\begin{aligned}
 p(y_t | y_{t-1}, \dots, y_1, \mathbf{x}_0) &= N(\mathbf{w}'\mathbf{x}_{t-1}, \sigma^2) \\
 &= p(y_t | \mathbf{x}_{t-1}).
 \end{aligned}
 \tag{2.5}$$

De lo anterior se tiene que:

$$\begin{aligned}
 p(\mathbf{y} | \mathbf{x}_0) &= \prod_{t=1}^n p(y_t | y_{t-1}, \dots, y_2, y_1, \mathbf{x}_0) \\
 &= \prod_{t=1}^n p(y_t | \mathbf{x}_{t-1}) \\
 &= \prod_{t=1}^n N(\mathbf{w}'\mathbf{x}_{t-1}, \sigma^2) \\
 &= \prod_{t=1}^n \left\{ \frac{1}{2\pi\sigma^2} \right\}^{1/2} \exp\left(-\frac{1}{2\sigma^2}(y_t - \mathbf{w}'\mathbf{x}_{t-1})^2\right) \\
 &= \left\{ \frac{1}{2\pi\sigma^2} \right\}^{n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{t=1}^n (y_t - \mathbf{w}'\mathbf{x}_{t-1})^2\right).
 \end{aligned}$$

Puesto que $\varepsilon_t = y_t - \mathbf{w}'\mathbf{x}_{t-1}$, se tiene que:

$$p(\mathbf{y} | \mathbf{x}_0) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2} \sum_{t=1}^n \varepsilon_t^2 / \sigma^2\right).
 \tag{2.6}$$

2.2. Pronósticos a un paso

El pronóstico a un paso $\hat{y}_{t|t-1}$ es:

$$\begin{aligned}
 \hat{y}_{t|t-1} &= E[y_t | y_{t-1}, \dots, y_2, y_1, \mathbf{x}_0] \\
 &= E[y_t | \mathbf{x}_{t-1}] \\
 &= \mathbf{w}'\mathbf{x}_{t-1},
 \end{aligned}
 \tag{2.7}$$

lo cual se sigue de la relación 2.5.

Por lo tanto, obtenemos las siguientes ecuaciones:

$$\begin{aligned}\hat{y}_{t|t-1} &= \mathbf{w}'\mathbf{x}_{t-1}, \\ \varepsilon_t &= y_t - \hat{y}_{t|t-1}, \\ \mathbf{x}_t &= \mathbf{F}\mathbf{x}_{t-1} + \mathbf{g}\varepsilon_t.\end{aligned}$$

De esto se sigue que:

$$\begin{aligned}\hat{y}_{t|t-1} &= \mathbf{w}'\mathbf{x}_{t-1} \\ &= \mathbf{w}'(\mathbf{D}^{t-1}\mathbf{x}_0 + \sum_{j=0}^{t-2} \mathbf{D}^j \mathbf{g}y_{t-1-j}) \\ &= \mathbf{w}'(\mathbf{D}^{t-1}\mathbf{x}_0 + \sum_{j=1}^{t-1} \mathbf{D}^{j-1} \mathbf{g}y_{t-j}) \\ &= \mathbf{w}'\mathbf{D}^{t-1}\mathbf{x}_0 + \sum_{j=1}^{t-1} \mathbf{w}'\mathbf{D}^{j-1} \mathbf{g}y_{t-j}.\end{aligned}\tag{2.8}$$

Si $a_t = \mathbf{w}'\mathbf{D}^{t-1}\mathbf{x}_0$ y $c_j = \mathbf{w}'\mathbf{D}^{j-1}\mathbf{g}$, entonces:

$$\hat{y}_{t|t-1} = a_t + \sum_{j=1}^{t-1} c_j y_{t-j}.\tag{2.9}$$

Observamos que el pronóstico depende del valor inicial \mathbf{x}_0 y una combinación lineal de los valores observados anteriores de la serie.

2.3. Serie de tiempo pronosticable

Decimos que una serie de tiempo con $\hat{y}_{t|t-1} = a_t + \sum_{j=1}^{t-1} c_j y_{t-j}$ es *pronosticable* si se cumple que:

$$\sum_{j=1}^{\infty} |c_j| < \infty \quad \text{y} \quad \lim_{t \rightarrow \infty} a_t = a,\tag{2.10}$$

donde a es constante.

Esta condición implica que observaciones de tiempos distantes no tengan algún efecto en el pronóstico. En la mayoría de los casos, $a = 0$, aunque esto no necesariamente sucede siempre.

Se puede ver que una condición suficiente, pero no necesaria, para que un proceso sea pronosticable es que los valores propios de \mathbf{D} se encuentren dentro del círculo unitario. Esto implica que \mathbf{D}^j tiende a la matriz $\mathbf{0}$ cuando j tiende a infinito. A esta propiedad se le llama *condición de estabilidad*. A un proceso que cumple tal propiedad se le llama *proceso estable*. En un proceso estable, los coeficientes de las observaciones pasadas de la serie decaen exponencialmente. Esta es una propiedad estrechamente asociada con los modelos de suavizamiento exponencial.

En ocasiones a_t converge a una constante y la sucesión $\{c_j\}$ converge a cero incluso aun si \mathbf{D} tiene un valor propio sobre el círculo unitario. En este caso, los coeficientes de las observaciones de tiempos distantes en el pasado no afectan el pronóstico. Esto implica que si un proceso es estable, entonces es pronosticable. Se puede ver que el inverso no siempre se cumple; hay procesos que son pronosticables, pero no son estables.

2.4. Estacionariedad débil

Sea $t \in \{0, \pm 1, \pm 2, \dots\}$. Decimos que un proceso $\{Y_t\}_{t=0}^{\infty}$ con media μ_t y varianza σ_t^2 es *débilmente estacionario* si:

- (I) $\mu_t = m$, donde m es constante,
- (II) $\text{Cov}(X_{t+h}, X_t)$ es independiente de t para toda h .

Esto implica que μ_t y σ^2 son finitos y no dependen de t (véase [5, p. 30]).

Observemos que si se satisfacen 2.1a y 2.1b, entonces:

$$\begin{aligned}
 \mathbf{x}_t &= \mathbf{F}\mathbf{x}_{t-1} + \mathbf{g}\varepsilon_t \\
 &= \mathbf{F}(\mathbf{F}\mathbf{x}_{t-2} + \mathbf{g}\varepsilon_{t-1}) + \mathbf{g}\varepsilon_t \\
 &= \mathbf{F}^2\mathbf{x}_{t-2} + \mathbf{F}\mathbf{g}\varepsilon_{t-1} + \mathbf{g}\varepsilon_t \\
 &= \mathbf{F}^2(\mathbf{F}\mathbf{x}_{t-3} + \mathbf{g}\varepsilon_{t-2}) + \mathbf{F}\mathbf{g}\varepsilon_{t-1} + \mathbf{g}\varepsilon_t \\
 &= \mathbf{F}^3\mathbf{x}_{t-3} + \mathbf{F}^2\mathbf{g}\varepsilon_{t-2} + \mathbf{F}\mathbf{g}\varepsilon_{t-1} + \mathbf{g}\varepsilon_t \\
 &= \\
 &\vdots \\
 &= \mathbf{F}^t\mathbf{x}_0 + \sum_{j=0}^{t-1} \mathbf{F}^j\mathbf{g}\varepsilon_{t-j}.
 \end{aligned}$$

Por lo tanto:

$$\begin{aligned}
y_t &= \mathbf{w}' \mathbf{x}_{t-1} + \varepsilon_t \\
&= \mathbf{w}' (\mathbf{F}^{t-1} \mathbf{x}_0 + \sum_{j=0}^{t-2} \mathbf{F}^j \mathbf{g} \varepsilon_{t-1-j}) + \varepsilon_t \\
&= \mathbf{w}' (\mathbf{F}^{t-1} \mathbf{x}_0 + \sum_{j=1}^{t-1} \mathbf{F}^{j-1} \mathbf{g} \varepsilon_{t-j}) + \varepsilon_t \\
&= \mathbf{w}' \mathbf{F}^{t-1} \mathbf{x}_0 + \left[\sum_{j=1}^{t-1} \mathbf{w}' \mathbf{F}^{j-1} \mathbf{g} \varepsilon_{t-j} \right] + \varepsilon_t \\
&= d_t + \sum_{j=0}^{t-1} k_j \varepsilon_{t-j},
\end{aligned}$$

donde $d_t = \mathbf{w}' \mathbf{F}^{t-1} \mathbf{x}_0$, $k_0 = 1$ y $k_j = \mathbf{w}' \mathbf{F}^{j-1} \mathbf{g}$, para $j = 1, 2, \dots$

Esto implica que la serie está determinada por el vector inicial \mathbf{x}_0 y los errores de tiempos anteriores.

Ahora, de forma análoga a cuando un proceso es pronosticable, diremos que un proceso que se puede expresar como $y_t = d_t + \sum_{j=0}^{t-1} k_j \varepsilon_{t-j}$, con $\varepsilon_t \sim NID(0, \sigma^2)$, es *estacionario* si:

$$\sum_{j=0}^{\infty} |k_j| < \infty \quad \text{y} \quad \lim_{t \rightarrow \infty} d_t = d, \tag{2.11}$$

donde d es constante.

Esto se debe a un resultado que dice que si se tienen dos procesos $\{Y_t\}_{t=0}^{\infty}$, $\{X_t\}_{t=0}^{\infty}$ tales que $\{X_t\}_{t=0}^{\infty}$ es estacionario y además $Y_t = c + \sum_{j=0}^{\infty} k_j X_{t-j}$, con c constante, entonces el proceso $\{Y_t\}_{t=0}^{\infty}$ es estacionario (véase [2, p. 84]).

2.5. Ejemplos de modelos

Ahora vamos a retomar los modelos de suavizamiento exponencial anteriormente mencionados, vistos como modelos de espacio de estados.

2.5.1. Modelo de suavizamiento exponencial simple ETS(A,N,N)

Este modelo solamente va a considerar una variable de nivel ℓ_t más una componente aditiva de error. En este caso, la serie está descrita por:

$$y_t = \ell_{t-1} + \varepsilon_t, \quad (2.12a)$$

$$\ell_t = \ell_{t-1} + \alpha\varepsilon_t, \quad (2.12b)$$

donde $\varepsilon_t \sim NID(0, \sigma^2)$. Si $\mathbf{x}_t = \ell_t$, $\mathbf{w} = 1$, $\mathbf{F} = 1$, $\mathbf{g} = \alpha$, observamos que el modelo tiene la estructura de un modelo de espacio de estados como en 2.1. Observamos que si $\alpha = 0$, entonces el nivel es constante, teniéndose así que las variables y_t son independientes e idénticamente distribuidas. Si $\alpha = 1$, entonces:

$$\begin{aligned} y_t &= \ell_{t-1} + \varepsilon_t \\ &= \ell_t - \varepsilon_t + \varepsilon_t \\ &= \ell_t. \end{aligned}$$

Por lo tanto $y_t = \ell_{t-1} + \varepsilon_t = y_{t-1} + \varepsilon_t$; es decir, una caminata aleatoria. En este caso cada observación depende de la anterior.

Si α está en el intervalo $(0, 1)$, para los pronósticos se tiene:

$$\begin{aligned} \hat{y}_{t|t-1} &= \mathbf{w}' \mathbf{x}_{t-1} \\ &= 1 \cdot \ell_{t-1} \\ &= \ell_{t-1}, \end{aligned}$$

$$\begin{aligned} \varepsilon_t &= y_t - \hat{y}_{t|t-1} \\ &= y_t - \ell_{t-1}, \end{aligned}$$

$$\begin{aligned} \ell_t &= \ell_{t-1} + \alpha\varepsilon_t \\ &= \ell_{t-1} + \alpha(y_t - \ell_{t-1}) \\ &= \ell_{t-1} + \alpha y_t - \alpha\ell_{t-1} \\ &= (1 - \alpha)\ell_{t-1} + \alpha y_t. \end{aligned}$$

De modo que:

$$\begin{aligned}
\hat{y}_{t|t-1} &= \mathbf{w}' \mathbf{D}^{t-1} \mathbf{x}_0 + \sum_{j=1}^{t-1} \mathbf{w}' \mathbf{D}^{j-1} \mathbf{g} y_{t-j} \quad (\text{Véase la ecuación 2.8.}) \\
&= \mathbf{w}' (\mathbf{F} - \mathbf{g} \mathbf{w}')^{t-1} \mathbf{x}_0 + \sum_{j=1}^{t-1} \mathbf{w}' (\mathbf{F} - \mathbf{g} \mathbf{w}')^{j-1} \mathbf{g} y_{t-j} \\
&= (1 - \alpha)^{t-1} \ell_0 + \alpha \sum_{j=1}^{t-1} (1 - \alpha)^{j-1} y_{t-j}.
\end{aligned} \tag{2.13}$$

La serie $\sum_{j=1}^{\infty} |1 - \alpha|^{j-1}$ converge si y sólo si $|1 - \alpha| < 1$, lo cual sucede si y sólo si $0 < \alpha < 2$.

Esto implica que el proceso es estable cuando $0 < \alpha < 2$, pues $\mathbf{D} = \mathbf{F} - \mathbf{g} \mathbf{w}' = 1 - \alpha$, de manera que $\mathbf{D}^j = (1 - \alpha)^j$ tiende a cero si $0 < \alpha < 2$.

2.5.2. Modelo lineal de Holt ETS(A,A,N)

Recordemos que este modelo, además de una variable de nivel ℓ_t , también contiene una variable de crecimiento b_t . Tenemos que:

$$y_t = \ell_{t-1} + b_{t-1} + \varepsilon_t, \tag{2.14a}$$

$$\ell_t = \ell_{t-1} + b_{t-1} + \alpha \varepsilon_t, \tag{2.14b}$$

$$b_t = b_{t-1} + \beta \varepsilon_t. \tag{2.14c}$$

El modelo puede verse como un modelo de espacio de estados si:

$$\mathbf{x}_t = [\ell_t \ b_t]', \quad \mathbf{w} = [1 \ 1]', \quad \mathbf{F} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{g} = [\alpha \ \beta]'$$

Si $\beta = 0$, entonces b_t es constante. Si además se tiene que $\alpha = 0$, entonces ℓ_t tiene una razón de crecimiento constante. A esto se le conoce como *tendencia global*. La razón de crecimiento suele interpretarse como la razón de crecimiento a largo plazo. Para otros valores de β distintos de cero, por ejemplo $\beta = 1$, se tiene que b_t es una caminata aleatoria.

Para el pronóstico se tiene que:

$$\begin{aligned}
\hat{y}_{t|t-1} &= \mathbf{w}' \mathbf{x}_{t-1} \\
&= [1 \ 1] [\ell_{t-1} \ b_{t-1}]' \\
&= \ell_{t-1} + b_{t-1},
\end{aligned}$$

$$\begin{aligned}
\varepsilon_t &= y_t - \hat{y}_{t|t-1} \\
&= y_t - (\ell_{t-1} + b_{t-1}) \\
&= y_t - \ell_{t-1} - b_{t-1},
\end{aligned}$$

$$\begin{aligned}
\ell_t &= \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t \\
&= \ell_{t-1} + b_{t-1} + \alpha(y_t - \ell_{t-1} - b_{t-1}) \\
&= \ell_{t-1} + b_{t-1} + \alpha y_t - \alpha\ell_{t-1} - \alpha b_{t-1} \\
&= \alpha y_t + (1 - \alpha)(\ell_{t-1} + b_{t-1}),
\end{aligned}$$

$$\begin{aligned}
b_t &= b_{t-1} + \beta\varepsilon_t \\
&= b_{t-1} + \beta(\ell_t - \ell_{t-1} - b_{t-1})/\alpha \\
&= b_{t-1}(1 - \beta/\alpha) + (\beta/\alpha)(\ell_t - \ell_{t-1}) \\
&= \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)b_{t-1},
\end{aligned}$$

donde $\beta^* = \beta/\alpha$.

Las ecuaciones $\ell_t = \alpha y_t + (1 - \alpha)(\ell_{t-1} + b_{t-1})$ y $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)b_{t-1}$ se pueden interpretar como promedios ponderados si $0 < \alpha < 1$ y $0 < \beta^* < 1$, o bien, si $0 < \alpha < 1$ y $0 < \beta < \alpha$. En la práctica suelen usarse estas restricciones. Se puede verificar (véase [6]) que el proceso es estable si $0 < \alpha$, $0 < \beta$ y $2\alpha + \beta < 4$.

2.5.3. Modelo de Holt Winters aditivo ETS(A,A,A)

A este modelo se añade una componente estacional s_t . Las ecuaciones que lo describen modelo son:

$$y_t = \ell_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t, \quad (2.15a)$$

$$\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t, \quad (2.15b)$$

$$b_t = b_{t-1} + \beta\varepsilon_t, \quad (2.15c)$$

$$s_t = s_{t-m} + \gamma\varepsilon_t, \quad (2.15d)$$

donde m indica la longitud del ciclo que define la estacionalidad.

Notemos que si $\mathbf{w}' = [1 \ 1 \ 0 \ \cdots \ 0 \ 1]$, $\mathbf{x}_t = \begin{bmatrix} \ell_t \\ b_t \\ s_t \\ s_{t-1} \\ \vdots \\ s_{t-m+1} \end{bmatrix}$, $\mathbf{F} = \begin{bmatrix} 1 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix}$ y

$\mathbf{g} = \begin{bmatrix} \alpha \\ \beta \\ \gamma \\ 0 \\ \vdots \\ 0 \end{bmatrix}$, el modelo se puede ver como un modelo de espacio de estados.

Notemos que si en las ecuaciones que definen al modelo sumamos un número arbitrario δ a la componente estacional s_t y lo restamos de la variable de nivel ℓ_t , obtenemos las ecuaciones equivalentes:

$$\begin{aligned} y_t &= (\ell_{t-1} - \delta) + b_{t-1} + (s_{t-m} + \delta) + \varepsilon_t, \\ \ell_t - \delta &= \ell_{t-1} - \delta + b_{t-1} + \alpha\varepsilon_t, \\ b_t &= b_{t-1} + \beta\varepsilon_t, \\ (s_t + \delta) &= (s_{t-m} + \delta) + \gamma\varepsilon_t. \end{aligned}$$

Esto implica que podemos obtener una infinidad de variables de nivel ℓ_t y de parámetros de estacionalidad s_t , obteniendo una infinidad de ecuaciones que determinan el modelo ETS(A,A,A). Para evitar este problema, vamos a imponer una restricción respecto a la variable de estacionalidad s_t . Se va a requerir que las variables iniciales ($t = 0$) de estacionalidad $s_{-m+1}, \dots, s_{-1}, s_0$ estén normalizadas; esto es, que la suma de éstas sea igual a cero.

Para el pronóstico se tiene que:

$$\begin{aligned} \hat{y}_{t|t-1} &= \mathbf{w}'\mathbf{x}_{t-1} \\ &= [1 \ 1 \ 0 \ \cdots \ 0 \ 1] \begin{bmatrix} \ell_t \\ b_t \\ s_t \\ s_{t-1} \\ \vdots \\ s_{t-m+1} \end{bmatrix} \\ &= \ell_{t-1} + b_{t-1} + s_{t-m}, \end{aligned}$$

$$\begin{aligned}\varepsilon_t &= y_t - \hat{y}_{t|t-1} \\ &= y_t - \ell_{t-1} - b_{t-1} - s_{t-m},\end{aligned}$$

$$\begin{aligned}\ell_t &= \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t \\ &= \ell_{t-1} + b_{t-1} + \alpha(y_t - \ell_{t-1} - b_{t-1} - s_{t-m}) \\ &= \alpha(y_t - s_{t-m}) + (1 - \alpha)(\ell_{t-1} + b_{t-1}),\end{aligned}$$

$$\begin{aligned}b_t &= b_{t-1} + \beta\varepsilon_t \\ &= b_{t-1} + \beta(\ell_t - \ell_{t-1} - b_{t-1})/\alpha \\ &= (\beta/\alpha)(\ell_t - \ell_{t-1}) + (1 - \beta/\alpha)b_{t-1} \\ &= \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)b_{t-1},\end{aligned}$$

donde $\beta^* = \beta/\alpha$.

Notemos que:

$$\begin{aligned}y_t - \ell_t - s_{t-m} &= y_t - \alpha(y_t - s_{t-m}) - (1 - \alpha)(\ell_{t-1} + b_{t-1}) - s_{t-m} \\ &= (1 - \alpha)(y_t - \ell_{t-1} - b_{t-1} - s_{t-m}) \\ &= (1 - \alpha)\varepsilon_t,\end{aligned}$$

de donde:

$$\begin{aligned}s_t &= s_{t-m} + \gamma\varepsilon_t \\ &= s_{t-m} + \gamma(y_t - \ell_t - s_{t-m})/(1 - \alpha) \\ &= s_{t-m} + [\gamma/(1 - \alpha)](y_t - \ell_t - s_{t-m}) \\ &= \gamma^*(y_t - \ell_t) + (1 - \gamma^*)s_{t-m},\end{aligned}$$

donde $\gamma^* = \gamma/(1 - \alpha)$.

Las ecuaciones para ℓ_t , b_t y s_t se pueden ver como promedios ponderados siempre que α , β^* y γ^* estén en el intervalo $(0, 1)$; esto es, $0 < \alpha < 1$, $0 < \beta < \alpha$ y $0 < \gamma < 1 - \alpha$.

2.5.4. Modelo de suavizamiento exponencial con tendencia

Las ecuaciones que describen este modelo son las siguientes:

$$y_t = \ell_{t-1} + b + \varepsilon_t, \quad (2.16a)$$

$$\ell_t = b + \ell_{t-1} + \alpha \varepsilon_t. \quad (2.16b)$$

El modelo se puede ver como un modelo de espacio de estados si $\mathbf{x}_t = [\ell_t \ b]'$, $\mathbf{w} = [1 \ 1]'$, $\mathbf{F} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ y $\mathbf{g} = [\alpha \ 0]'$.

Tenemos que:

$$\begin{aligned} \mathbf{D} &= \mathbf{F} - \mathbf{g}\mathbf{w}' \\ &= \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} \alpha \\ 0 \end{bmatrix} [1 \ 1] \\ &= \begin{bmatrix} 1 - \alpha & 1 - \alpha \\ 0 & 1 \end{bmatrix}. \end{aligned}$$

Sin embargo, los valores propios de \mathbf{D} son 1 y $1 - \alpha$, de modo que \mathbf{D}^j no converge a la matriz $\mathbf{0}$. Sin embargo, se puede ver que el modelo es pronosticable si $0 < \alpha < 2$.

Sea $U = \begin{bmatrix} u & u \\ 0 & 1 \end{bmatrix}$. Entonces:

$$\begin{aligned} U^2 &= \begin{bmatrix} u & u \\ 0 & 1 \end{bmatrix} \begin{bmatrix} u & u \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} u^2 & u^2 + u \\ 0 & 1 \end{bmatrix}, \end{aligned}$$

$$\begin{aligned} U^3 &= U^2U \\ &= \begin{bmatrix} u^2 & u^2 + u \\ 0 & 1 \end{bmatrix} \begin{bmatrix} u & u \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} u^3 & u^3 + u^2 + u \\ 0 & 1 \end{bmatrix}, \end{aligned}$$

⋮

$$U^{j-1} = \begin{bmatrix} u^{j-1} & \sum_{i=1}^{j-1} u^i \\ 0 & 1 \end{bmatrix}. \quad (2.17)$$

De 2.17 tenemos que $D^{j-1} = \begin{bmatrix} (1-\alpha)^{j-1} & \sum_{i=1}^{j-1} (1-\alpha)^i \\ 0 & 1 \end{bmatrix}$. Luego, por 2.9, se tiene que:

$$\begin{aligned} a_t &= \mathbf{w}' \mathbf{D}^{t-1} \mathbf{x}_0 \\ &= [1 \ 1] \begin{bmatrix} (1-\alpha)^{t-1} & \sum_{i=1}^{t-1} (1-\alpha)^i \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \ell_0 \\ 0 \end{bmatrix} \\ &= [1 \ 1] \begin{bmatrix} \ell_0(1-\alpha)^{t-1} \\ 0 \end{bmatrix} \\ &= \ell_0(1-\alpha)^{t-1}, \end{aligned} \quad (2.18)$$

$$\begin{aligned} c_j &= \mathbf{w}' \mathbf{D}^{j-1} \mathbf{g} \\ &= [1 \ 1] \begin{bmatrix} (1-\alpha)^{j-1} & \sum_{i=1}^{j-1} (1-\alpha)^i \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha \\ 0 \end{bmatrix} \\ &= [1 \ 1] \begin{bmatrix} \alpha(1-\alpha)^{j-1} \\ 0 \end{bmatrix} \\ &= \alpha(1-\alpha)^{j-1}. \end{aligned} \quad (2.19)$$

Por lo tanto, si $|1-\alpha| < 1$, por 2.19, $\sum_{j=1}^{\infty} |c_j|$ converge. Además, por 2.18, $\lim_{t \rightarrow \infty} a_t$ existe y es constante. De este modo, el modelo es pronosticable si $|1-\alpha| < 1$, si i sólo si $0 < \alpha < 2$.

Observamos que si $\alpha = 0$, entonces $y_t = \ell_0 + bt + \varepsilon_t$; esto es, un modelo de regresión lineal.

Para el pronóstico se tiene que:

$$\begin{aligned} \hat{y}_{t|t-1} &= \mathbf{w}' \mathbf{x}_{t-1} \\ &= [1 \ 1] \begin{bmatrix} \ell_{t-1} \\ b \end{bmatrix} \\ &= \ell_{t-1} + b, \end{aligned}$$

$$\begin{aligned}\varepsilon_t &= y_t - \hat{y}_{t|t-1} \\ &= y_t - \ell_{t-1} - b,\end{aligned}$$

$$\begin{aligned}\ell_t &= b + \ell_{t-1} + \alpha\varepsilon_t \\ &= b + \ell_{t-1} + \alpha(y_t - \ell_{t-1} - b) \\ &= \alpha y_t + (1 - \alpha)(\ell_{t-1} + b).\end{aligned}$$

2.5.5. Modelo con tendencia que se termina ETS(A,A_d,N)

Este modelo considera un factor ϕ en $(0, 1)$, el cual modifica la tendencia. Las ecuaciones del modelo son:

$$y_t = \ell_{t-1} + \phi b_{t-1} + \varepsilon_t, \quad (2.20a)$$

$$\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t, \quad (2.20b)$$

$$b_t = \phi b_{t-1} + \beta\varepsilon_t. \quad (2.20c)$$

Si $\mathbf{x}_t = [\ell_t \ b_t]'$, $\mathbf{w} = [1 \ \phi]'$, $\mathbf{F} = \begin{bmatrix} 1 & \phi \\ 0 & \phi \end{bmatrix}$, $\mathbf{g} = [\alpha \ \beta]'$, el modelo se puede ver como un modelo de espacio de estados.

Para el pronóstico se tiene que:

$$\begin{aligned}\hat{y}_{t|t-1} &= \mathbf{w}'\mathbf{x}_{t-1} \\ &= [1 \ \phi] \begin{bmatrix} \ell_{t-1} \\ b_{t-1} \end{bmatrix} \\ &= \ell_{t-1} + \phi b_{t-1},\end{aligned}$$

$$\begin{aligned}\varepsilon_t &= y_t - \hat{y}_{t|t-1} \\ &= y_t - \ell_{t-1} - \phi b_{t-1},\end{aligned}$$

$$\begin{aligned}\ell_t &= \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t \\ &= \ell_{t-1} + \phi b_{t-1} + \alpha(y_t - \ell_{t-1} - \phi b_{t-1}) \\ &= \alpha y_t + (1 - \alpha)(\ell_{t-1} + \phi b_{t-1}),\end{aligned}$$

$$\begin{aligned}
b_t &= \phi b_{t-1} + \beta \varepsilon_t \\
&= \phi b_{t-1} + \beta(\ell_t - \ell_{t-1} - \phi b_{t-1})/\alpha \\
&= (\beta/\alpha)(\ell_t - \ell_{t-1}) + (1 - \beta/\alpha)\phi b_{t-1} \\
&= \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)\phi b_{t-1},
\end{aligned}$$

donde $\beta^* = \beta/\alpha$.

Cuando $\phi = 1$, este modelo es el modelo ETS(A,A,N).

2.5.6. Modelos compuestos

Podemos combinar dos modelos de espacio de estados. Si se tienen dos modelos de espacio de estados, entonces, para $i = 1, 2$, se tiene que:

$$y_t = \mathbf{w}'_i \mathbf{x}_{i,t-1} + \varepsilon_{i,t}, \quad (2.21a)$$

$$\mathbf{x}_{i,t} = \mathbf{F}_i \mathbf{x}_{i,t-1} + \mathbf{g}_i \varepsilon_{i,t}, \quad (2.21b)$$

donde $\varepsilon_{i,t} \sim NID(0, v_i)$.

Al combinarlos, se tiene un nuevo modelo dado por:

$$y_t = \mathbf{w}'_1 \mathbf{x}_{1,t-1} + \mathbf{w}'_2 \mathbf{x}_{2,t-1} + \varepsilon_t, \quad (2.22a)$$

$$\begin{bmatrix} \mathbf{x}_{1,t} \\ \mathbf{x}_{2,t} \end{bmatrix} = \begin{bmatrix} \mathbf{F}_1 & 0 \\ 0 & \mathbf{F}_2 \end{bmatrix} \begin{bmatrix} \mathbf{x}_{1,t-1} \\ \mathbf{x}_{2,t-1} \end{bmatrix} + \begin{bmatrix} \mathbf{g}_1 \\ \mathbf{g}_2 \end{bmatrix} \varepsilon_t. \quad (2.22b)$$

Capítulo 3

Modelo general de innovaciones de espacio de estados

3.1. Ecuaciones del modelo

Vamos ahora a introducir una forma más general del modelo de espacio de estados, la cual va a describir otros modelos. Las ecuaciones de este modelo son las siguientes:

$$y_t = w(\mathbf{x}_{t-1}) + r(\mathbf{x}_{t-1})\varepsilon_t, \quad (3.1a)$$

$$\mathbf{x}_t = f(\mathbf{x}_{t-1}) + g(\mathbf{x}_{t-1})\varepsilon_t, \quad (3.1b)$$

donde $\mathbf{x}_t = (\ell_t, b_t, s_t, s_{t-1}, \dots, s_{t-m+1})$ es el vector de estados; w y r son funciones escalares, y f y g son funciones vectoriales. Las funciones w , r , f y g no dependen de t . La componente $\{\varepsilon_t\}_t$ es un ruido blanco con varianza σ^2 . Recordamos que un ruido blanco es un proceso de variables no correlacionadas e idénticamente distribuidas con media 0 y varianza σ^2 .

Tenemos que $y_t = w(\mathbf{x}_{t-1}) + r(\mathbf{x}_{t-1})\varepsilon_t$; entonces $\varepsilon_t = \frac{[y_t - w(\mathbf{x}_{t-1})]}{r(\mathbf{x}_{t-1})}$.

De modo que:

$$\begin{aligned} \mathbf{x}_t &= f(\mathbf{x}_{t-1}) + g(\mathbf{x}_{t-1})\varepsilon_t \\ &= f(\mathbf{x}_{t-1}) + g(\mathbf{x}_{t-1})\frac{[y_t - w(\mathbf{x}_{t-1})]}{r(\mathbf{x}_{t-1})}. \end{aligned}$$

Definiendo $D(\mathbf{x}_t) = f(\mathbf{x}_t) - \frac{[g(\mathbf{x}_t)w(\mathbf{x}_t)]}{r(\mathbf{x}_t)}$, se tiene que:

$$\begin{aligned}\mathbf{x}_t &= f(\mathbf{x}_{t-1}) + g(\mathbf{x}_{t-1}) \frac{[y_t - w(\mathbf{x}_{t-1})]}{r(\mathbf{x}_{t-1})} \\ &= f(\mathbf{x}_{t-1}) - \frac{[g(\mathbf{x}_{t-1})w(\mathbf{x}_{t-1})]}{r(\mathbf{x}_{t-1})} + g(\mathbf{x}_{t-1}) \frac{y_t}{r(\mathbf{x}_{t-1})},\end{aligned}$$

de donde:

$$\mathbf{x}_t = D(\mathbf{x}_{t-1}) + g(\mathbf{x}_{t-1}) \frac{y_t}{r(\mathbf{x}_{t-1})}. \quad (3.2)$$

Cuando $f(\mathbf{x}_t)$ y $g(\mathbf{x}_t)$ son lineales para \mathbf{x}_t , y $w(\mathbf{x}_t) = r(\mathbf{x}_t)$, se tiene que $D(\mathbf{x}_t)$ es lineal. Si $g(\mathbf{x}_t)/r(\mathbf{x}_t)$ no depende de las variables de estado, entonces podemos escribir a \mathbf{x}_t como:

$$\mathbf{x}_t = D\mathbf{x}_{t-1} + gy_t.$$

Éste es el modelo que se había analizado antes.

Anteriormente se había demostrado en 2.3 que para $\mathbf{y} = [y_1, y_2, \dots, y_n]$, en el modelo lineal se tiene que:

$$p(\mathbf{y} \mid \mathbf{x}_0) = \prod_{t=1}^n p(y_t \mid y_{t-1}, \dots, y_2, y_1, \mathbf{x}_0).$$

Por otra parte, se tiene que:

$$\begin{aligned}
y_t &= w(\mathbf{x}_{t-1}) + r(\mathbf{x}_{t-1})\varepsilon_t \\
&= w\left(D(\mathbf{x}_{t-2}) + g(\mathbf{x}_{t-2})\frac{y_{t-1}}{r(\mathbf{x}_{t-2})}\right) + r\left(D(\mathbf{x}_{t-2}) + g(\mathbf{x}_{t-2})\frac{y_{t-1}}{r(\mathbf{x}_{t-2})}\right)\varepsilon_t \\
&= w\left(D\left(D(\mathbf{x}_{t-3}) + g(\mathbf{x}_{t-3})\frac{y_{t-2}}{r(\mathbf{x}_{t-3})}\right)\right. \\
&\quad \left.+ g\left(D(\mathbf{x}_{t-3}) + g(\mathbf{x}_{t-3})\frac{y_{t-2}}{r(\mathbf{x}_{t-3})}\right)\frac{y_{t-1}}{r\left(D(\mathbf{x}_{t-3}) + g(\mathbf{x}_{t-3})\frac{y_{t-2}}{r(\mathbf{x}_{t-3})}\right)}\right) \\
&\quad \left.+ r\left(D\left(D(\mathbf{x}_{t-3}) + g(\mathbf{x}_{t-3})\frac{y_{t-2}}{r(\mathbf{x}_{t-3})}\right)\right)\right. \\
&\quad \left.+ g\left(D(\mathbf{x}_{t-3}) + g(\mathbf{x}_{t-3})\frac{y_{t-2}}{r(\mathbf{x}_{t-3})}\right)\frac{y_{t-1}}{r\left(D(\mathbf{x}_{t-3}) + g(\mathbf{x}_{t-3})\frac{y_{t-2}}{r(\mathbf{x}_{t-3})}\right)}\right)\varepsilon_t \\
&= \\
&\quad \vdots \\
&= h(y_1, y_2, \dots, y_{t-1}, \mathbf{x}_0) + k(y_1, y_2, \dots, y_{t-1}, \mathbf{x}_0)\varepsilon_t.
\end{aligned}$$

Por lo tanto, tenemos que:

$$y_t = w(\mathbf{x}_{t-1}) + r(\mathbf{x}_{t-1})\varepsilon_t = h(y_1, y_2, \dots, y_{t-1}, \mathbf{x}_0) + k(y_1, y_2, \dots, y_{t-1}, \mathbf{x}_0)\varepsilon_t,$$

donde $w(\mathbf{x}_{t-1}) = h(y_1, y_2, \dots, y_{t-1}, \mathbf{x}_0)$ y $r(\mathbf{x}_{t-1}) = k(y_1, y_2, \dots, y_{t-1}, \mathbf{x}_0)$. De este modo, condicionando a $y_1, y_2, \dots, y_{t-1}, \mathbf{x}_0$, tenemos que y_t sigue la misma distribución que condicionando a \mathbf{x}_{t-1} . Entonces $p(y_t | y_{t-1}, \dots, y_2, y_1, \mathbf{x}_0) = p(y_t | \mathbf{x}_{t-1})$.

Así, asumiendo que $\varepsilon_t \sim N(0, \sigma^2)$, obtenemos:

$$\begin{aligned}
 p(\mathbf{y} \mid \mathbf{x}_0) &= \prod_{t=1}^n p(y_t \mid y_{t-1}, \dots, y_2, y_1, \mathbf{x}_0) \\
 &= \prod_{t=1}^n p(y_t \mid \mathbf{x}_{t-1}) \\
 &= \prod_{t=1}^n N(w(\mathbf{x}_{t-1}), r^2(\mathbf{x}_{t-1})\sigma^2) \\
 &= \prod_{t=1}^n \left\{ \frac{1}{2\pi r^2(\mathbf{x}_{t-1})\sigma^2} \right\}^{1/2} \exp\left(-\frac{1}{2r^2(\mathbf{x}_{t-1})\sigma^2} [y_t - w(\mathbf{x}_{t-1})]^2\right) \\
 &= \left\{ \frac{1}{2\pi\sigma^2} \right\}^{n/2} \left| \prod_{t=1}^n r(\mathbf{x}_{t-1}) \right|^{-1} \exp\left(-\frac{1}{2} \sum_{t=1}^n \frac{[y_t - w(\mathbf{x}_{t-1})]^2}{r^2(\mathbf{x}_{t-1})\sigma^2}\right).
 \end{aligned}$$

Puesto que $\varepsilon_t = \frac{[y_t - w(\mathbf{x}_{t-1})]}{r(\mathbf{x}_{t-1})}$, se tiene que:

$$p(\mathbf{y} \mid \mathbf{x}_0) = \left\{ \frac{1}{2\pi\sigma^2} \right\}^{n/2} \left| \prod_{t=1}^n r(\mathbf{x}_{t-1}) \right|^{-1} \exp\left(-\frac{1}{2} \sum_{t=1}^n \varepsilon_t^2 / \sigma^2\right). \quad (3.3)$$

Luego, para el pronóstico se tiene que:

$$\begin{aligned}
 \hat{y}_{t|t-1} &= E[y_t \mid y_{t-1}, \dots, y_2, y_1, \mathbf{x}_0] \\
 &= E[y_t \mid \mathbf{x}_{t-1}] \quad (\text{Puesto que } p(y_t \mid y_{t-1}, \dots, y_2, y_1, \mathbf{x}_0) = p(y_t \mid \mathbf{x}_{t-1})) \\
 &= E[w(\mathbf{x}_{t-1}) + r(\mathbf{x}_{t-1})\varepsilon_t \mid \mathbf{x}_{t-1}] \\
 &= E[w(\mathbf{x}_{t-1}) \mid \mathbf{x}_{t-1}] + E[r(\mathbf{x}_{t-1})\varepsilon_t \mid \mathbf{x}_{t-1}] \\
 &= w(\mathbf{x}_{t-1}) + r(\mathbf{x}_{t-1})E[\varepsilon_t \mid \mathbf{x}_{t-1}] \\
 &= w(\mathbf{x}_{t-1}) + r(\mathbf{x}_{t-1})E[\varepsilon_t]^* \\
 &= w(\mathbf{x}_{t-1}) + r(\mathbf{x}_{t-1}) \cdot 0 \\
 &= w(\mathbf{x}_{t-1}) + 0 \\
 &= w(\mathbf{x}_{t-1}).
 \end{aligned} \quad (3.4)$$

Nota*: La igualdad $E[\varepsilon_t] = E[\varepsilon_t \mid \mathbf{x}_{t-1}]$ se sigue de que \mathbf{x}_{t-1} es función de $\varepsilon_{t-1}, \varepsilon_{t-2}, \dots$, y las variables aleatorias $\{\varepsilon_t\}_t$ son independientes.

Por otra parte:

$$\varepsilon_t = (y_t - \hat{y}_{t|t-1})/r(\mathbf{x}_{t-1}), \quad (3.5)$$

$$\mathbf{x}_{t-1} = f(\mathbf{x}_{t-1}) + g(\mathbf{x}_{t-1})\varepsilon_t. \quad (3.6)$$

Estas relaciones definen el modelo y nos sirven para calcular y actualizar predicciones de forma recursiva.

3.2. Ejemplos de modelos

Vamos ahora a introducir algunos modelos en los cuales la estructura de las ecuaciones descritas en 3.1 está presente.

3.2.1. Modelo de nivel local con error multiplicativo ETS(M,N,N)

Este modelo está determinado por:

$$y_t = \ell_{t-1}(1 + \varepsilon_t), \quad (3.7a)$$

$$\ell_t = \ell_{t-1}(1 + \alpha\varepsilon_t), \quad (3.7b)$$

donde $\mathbf{x}_t = \ell_t$, $w(\mathbf{x}_{t-1}) = r(\mathbf{x}_{t-1}) = f(\mathbf{x}_{t-1}) = \ell_{t-1}$ y $g(\mathbf{x}_{t-1}) = \alpha\ell_{t-1}$.

Si $\alpha = 0$, entonces $\ell_t = \ell_{t-1} = \dots = \ell_0 = \ell$ (constante). Así, $y_t = \ell(1 + \varepsilon_t)$. Si $\alpha = 1$, entonces $\ell_t = \ell_{t-1}(1 + \varepsilon_t) = y_t$, de modo que la serie se expresa como $y_t = y_{t-1}(1 + \varepsilon_t)$.

Notemos que:

$$\begin{aligned} y_t &= \ell_{t-1}(1 + \varepsilon_t) \\ &= \ell_{t-2}(1 + \alpha\varepsilon_{t-1})(1 + \varepsilon_t) \\ &= (y_{t-1}/(1 + \varepsilon_{t-1}))(1 + \alpha\varepsilon_{t-1})(1 + \varepsilon_t) \quad (\text{Ya que } y_{t-1} = \ell_{t-2}(1 + \varepsilon_{t-1}).) \\ &= y_{t-1}(1 + \varepsilon_t)(1 + \alpha\varepsilon_{t-1})/(1 + \varepsilon_{t-1}). \end{aligned}$$

Por otra parte:

$$\begin{aligned}
y_t &= \ell_{t-1}(1 + \varepsilon_t) \\
&= \ell_{t-2}(1 + \alpha\varepsilon_{t-1})(1 + \varepsilon_t) \\
&= \ell_{t-3}(1 + \alpha\varepsilon_{t-2})(1 + \alpha\varepsilon_{t-1})(1 + \varepsilon_t) \\
&= \\
&\vdots \\
&= \ell_0(1 + \varepsilon_t) \prod_{j=1}^{t-1} (1 + \alpha\varepsilon_j),
\end{aligned}$$

$$\begin{aligned}
\ell_t &= \ell_{t-1}(1 + \alpha\varepsilon_t) \\
&= \ell_{t-1}(1 + \alpha\varepsilon_t) + \alpha\ell_{t-1} - \alpha\ell_{t-1} \\
&= \alpha\ell_{t-1}(1 + \varepsilon_t) + (1 - \alpha)\ell_{t-1} \\
&= \alpha y_t + (1 - \alpha)\ell_{t-1}.
\end{aligned}$$

Observamos que ℓ_t se determina de la misma manera que para el modelo ETS(A,N,N).

Para el pronóstico se tiene que:

$$\begin{aligned}
\hat{y}_{t|t-1} &= w(\mathbf{x}_{t-1}) \\
&= \ell_{t-1} \\
&= (1 - \alpha)\ell_{t-2} + \alpha y_{t-1} \\
&= (1 - \alpha)[(1 - \alpha)\ell_{t-3} + \alpha y_{t-2}] + \alpha y_{t-1} \\
&= (1 - \alpha)^2 \ell_{t-3} + \alpha(1 - \alpha)y_{t-2} + \alpha y_{t-1} \\
&= \\
&\vdots \\
&= (1 - \alpha)^{t-1} \ell_0 + \alpha \sum_{j=1}^{t-1} (1 - \alpha)^{j-1} y_{t-j}.
\end{aligned} \tag{3.8}$$

Por lo tanto, el pronóstico a un paso es el mismo que en el modelo ETS(A,N,N).

Como $\sum_{j=1}^{\infty} |1 - \alpha|^{j-1}$ converge cuando $|1 - \alpha| < 1$, entonces el proceso es estable si $-1 < 1 - \alpha < 1$, si y sólo si $0 < \alpha < 2$.

3.2.2. Modelo con error multiplicativo y tendencia aditiva ETS(M,A,N)

Este modelo contiene una variable de nivel ℓ_t y una variable de crecimiento b_t para la tendencia, además de los parámetros α y β . Las ecuaciones que lo describen son las siguientes:

$$y_t = (\ell_{t-1} + b_{t-1})(1 + \varepsilon_t), \quad (3.9a)$$

$$\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t), \quad (3.9b)$$

$$b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t. \quad (3.9c)$$

Si $\mathbf{x}_{t-1} = [\ell_t \ b_t]'$, $w(\mathbf{x}_{t-1}) = r(\mathbf{x}_{t-1}) = \ell_{t-1} + b_{t-1}$, $f(\mathbf{x}_{t-1}) = [\ell_{t-1} + b_{t-1}, b_{t-1}]'$ y $g(\mathbf{x}_{t-1}) = [\alpha(\ell_{t-1} + b_{t-1}), \beta(\ell_{t-1} + b_{t-1})]'$, el modelo se puede ver como un modelo de espacio de estados.

Notemos que si $\beta = 0$, entonces $b_t = \text{cte}$, de modo que el factor de crecimiento es constante.

Si $\beta = 0$ y $\alpha = 1$, entonces $b_t = \text{cte}$ y $y_t = (\ell_{t-1} + b_{t-1})(1 + \varepsilon_t) = \ell_t$. Por lo tanto, $y_t = (y_{t-1} + b)(1 + \varepsilon_t)$, lo cual se conoce como una *caminata aleatoria con tendencia*.

Si $\beta = 0$ y $\alpha = 0$, entonces $b_t = \text{cte}$ y $\ell_t = \ell_{t-1} + b$, teniéndose así que $\ell_t = \ell_{t-1} + b = \ell_{t-2} + 2b = \ell_{t-3} + 3b = \dots = \ell_0 + bt$, por lo que $y_t = (\ell_0 + bt)(1 + \varepsilon_t)$.

Para el pronóstico se tiene que:

$$\begin{aligned} \hat{y}_{t|t-1} &= w(\mathbf{x}_{t-1}) \\ &= \ell_{t-1} + b_{t-1}, \end{aligned}$$

$$\begin{aligned} \varepsilon_t &= (y_t - \hat{y}_{t|t-1})/r(\mathbf{x}_{t-1}) \\ &= (y_t - \ell_{t-1} - b_{t-1})/(\ell_{t-1} + b_{t-1}), \end{aligned}$$

$$\begin{aligned} \ell_t &= (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t) \\ &= (\ell_{t-1} + b_{t-1})[1 + \alpha(y_t - \ell_{t-1} - b_{t-1})/(\ell_{t-1} + b_{t-1})] \\ &= \ell_{t-1} + b_{t-1} + \alpha(y_t - \ell_{t-1} - b_{t-1}) \\ &= \alpha y_t + (1 - \alpha)(\ell_{t-1} + b_{t-1}), \end{aligned}$$

$$\begin{aligned}
b_t &= b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t \\
&= b_{t-1} + (\beta/\alpha)(\ell_{t-1} + b_{t-1})\alpha\varepsilon_t \\
&= b_{t-1} + (\beta/\alpha)(\ell_t - \ell_{t-1} - b_{t-1}) \\
&= (\beta/\alpha)(\ell_t - \ell_{t-1}) + (1 - (\beta/\alpha))b_{t-1} \\
&= \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)b_{t-1},
\end{aligned}$$

donde $\beta^* = \beta/\alpha$.

También podemos escribir a b_t de la siguiente manera:

$$\begin{aligned}
b_t &= b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t \\
&= b_{t-1} + \beta(y_t - \ell_{t-1} - b_{t-1}) \\
&= \beta(y_t - \ell_{t-1}) + (1 - \beta)b_{t-1}.
\end{aligned}$$

En [3, sección 10.2] se establece que el proceso es estable si $\alpha > 0$, $\beta > 0$ y $2\alpha + \beta < 4$.

3.2.3. Modelo con error aditivo y tendencia multiplicativa ETS(A,M,N)

Este modelo está determinado por las ecuaciones siguientes:

$$y_t = \ell_{t-1}b_{t-1} + \varepsilon_t, \quad (3.10a)$$

$$\ell_t = \ell_{t-1}b_{t-1} + \alpha\varepsilon_t, \quad (3.10b)$$

$$b_t = b_{t-1} + \beta\varepsilon_t/\ell_{t-1}, \quad (3.10c)$$

donde $\beta = \alpha\beta^*$.

De modo que:

$$\begin{aligned}
\ell_t &= \ell_{t-1}b_{t-1} + \alpha\varepsilon_t \\
&= \ell_{t-1}b_{t-1} + \alpha(y_t - \ell_{t-1}b_{t-1}) \\
&= (1 - \alpha)\ell_{t-1}b_{t-1} + \alpha y_t,
\end{aligned}$$

$$\begin{aligned}
b_t &= b_{t-1} + \beta\varepsilon_t/\ell_{t-1} \\
&= b_{t-1} + \beta(y_t - \ell_{t-1}b_{t-1})/\ell_{t-1} \\
&= b_{t-1} + \beta(y_t/\ell_{t-1} - b_{t-1}) \\
&= (1 - \beta)b_{t-1} + \beta y_t/\ell_{t-1}.
\end{aligned}$$

Para este modelo no se pueden garantizar las condiciones de estabilidad (véase [3, sección 10.2]).

3.2.4. Modelo con error y tendencia multiplicativas ETS(M,M,N)

Modificando el modelo anterior se tienen las siguientes ecuaciones:

$$y_t = \ell_{t-1} b_{t-1} (1 + \varepsilon_t), \quad (3.11a)$$

$$\ell_t = \ell_{t-1} b_{t-1} (1 + \alpha \varepsilon_t), \quad (3.11b)$$

$$b_t = b_{t-1} (1 + \beta \varepsilon_t). \quad (3.11c)$$

Por lo tanto, si $1 + \varepsilon_t > 0$ y si $0 < \beta < 1$, entonces, por la relación 3.11c, se tiene que $b_t > 0$. Si además $0 < \alpha < 1$, entonces, por la relación 3.11b, tenemos que $\ell_t > 0$. De estas observaciones se sigue que $y_t > 0$ si se tienen las condiciones $0 < \alpha < 1$, $0 < \beta < 1$ y $1 + \varepsilon_t > 0$.

En el caso en que $\beta = 0$, se tiene que $b_t = b_{t-1} = \dots = b_0 = b$, de modo que el factor de crecimiento es constante. Si $b < 1$, se tiene que la tendencia va desapareciendo y si $b > 1$, se tiene que la tendencia aumenta cada vez más, lo cual no permitiría la condición de estabilidad en el proceso.

Veamos un ejemplo de un modelo en el que se incluye estacionalidad multiplicativa.

3.2.5. Modelo con error multiplicativo, tendencia aditiva y estacionalidad multiplicativa ETS(M,A,M)

Este modelo agrega una componente estacional s_t de tipo multiplicativo. Las ecuaciones que lo describen están dadas por:

$$y_t = (\ell_{t-1} + b_{t-1}) s_{t-m} (1 + \varepsilon_t), \quad (3.12a)$$

$$\ell_t = (\ell_{t-1} + b_{t-1}) (1 + \alpha \varepsilon_t), \quad (3.12b)$$

$$b_t = b_{t-1} + \beta (\ell_{t-1} + b_{t-1}) \varepsilon_t, \quad (3.12c)$$

$$s_t = s_{t-m} (1 + \gamma \varepsilon_t). \quad (3.12d)$$

El modelo se puede ver como un modelo de espacio de estados si $\mathbf{x}_t = \begin{bmatrix} \ell_t \\ b_t \\ s_t \\ s_{t-1} \\ \vdots \\ s_{t-m+1} \end{bmatrix}$,

$$w(\mathbf{x}_{t-1}) = r(\mathbf{x}_{t-1}) = (\ell_{t-1} + b_{t-1})s_{t-m}, \quad g(\mathbf{x}_{t-1}) = \begin{bmatrix} \alpha(\ell_{t-1} + b_{t-1}) \\ \beta(\ell_{t-1} + b_{t-1}) \\ \gamma s_{t-m} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad \text{y } f(\mathbf{x}_{t-1}) = \mathbf{F}\mathbf{x}_{t-1},$$

$$\text{con } \mathbf{F} = \begin{bmatrix} 1 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix}.$$

Notemos que podemos multiplicar y dividir por un número arbitrario $\delta \neq 0$ y obtenemos las ecuaciones equivalentes:

$$\begin{aligned} y_t &= (1/\delta)(\ell_{t-1} + b_{t-1}) \cdot \delta s_{t-m}(1 + \varepsilon_t), \\ (1/\delta)\ell_t &= (1/\delta)(\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t), \\ (1/\delta)b_t &= (1/\delta)[b_{t-1} + \beta(\ell_{t-1} + b_{t-1})]\varepsilon_t, \\ \delta s_t &= \delta s_{t-m}(1 + \gamma\varepsilon_t). \end{aligned}$$

De modo que, así como en el modelo ETS(A,A,A), hay una infinidad de maneras de determinar el modelo. Para resolver este problema, es necesaria la condición de que las variables iniciales de estacionalidad $s_{-m+1}, \dots, s_{-1}, s_0$ estén normalizadas; en este caso esto significa que la suma de éstas se iguala a m .

Por otra parte, tenemos que:

$$\begin{aligned} \mathbf{x}_t &= \left[[f(\mathbf{x}_{t-1}) - g(\mathbf{x}_{t-1}) \frac{w(\mathbf{x}_{t-1})}{r(\mathbf{x}_{t-1})}] + \frac{g(\mathbf{x}_{t-1})}{r(\mathbf{x}_{t-1})} y_t \right] \\ &= [\mathbf{F}\mathbf{x}_{t-1} - g(\mathbf{x}_{t-1})] + \frac{g(\mathbf{x}_{t-1})}{r(\mathbf{x}_{t-1})} y_t. \end{aligned}$$

Al sustituir, se obtiene:

$$\begin{aligned}
\mathbf{x}_t &= \begin{bmatrix} \ell_t \\ b_t \\ s_t \\ s_{t-1} \\ \vdots \\ s_{t-m+1} \end{bmatrix} \\
&= \begin{bmatrix} (1-\alpha)(\ell_{t-1} + b_{t-1}) \\ b_{t-1} - \beta(\ell_{t-1} + b_{t-1}) \\ (1-\gamma)s_{t-m} \\ s_{t-1} \\ \vdots \\ s_{t-m+1} \end{bmatrix} + \begin{bmatrix} \alpha/s_{t-m} \\ \beta/s_{t-m} \\ \gamma/(\ell_{t-1} + b_{t-1}) \\ 0 \\ \vdots \\ 0 \end{bmatrix} y_t.
\end{aligned}$$

Haciendo $\beta = \alpha\beta^*$, se obtiene:

$$\ell_t = (1-\alpha)(\ell_{t-1} + b_{t-1}) + \alpha y_t / s_{t-m},$$

$$\begin{aligned}
b_t &= b_{t-1} - \beta(\ell_{t-1} + b_{t-1}) + \beta y_t / s_{t-m} \\
&= b_{t-1} - (\alpha\beta^*)(\ell_{t-1} + b_{t-1}) + (\alpha\beta^*)y_t / s_{t-m} \\
&= b_{t-1} + \beta^* \alpha (y_t / s_{t-m} - \ell_{t-1} - b_{t-1}) \\
&= b_{t-1} + \beta^* (\ell_t - \ell_{t-1} - b_{t-1}) \\
&= \beta^* (\ell_t - \ell_{t-1}) + (1 - \beta^*) b_{t-1},
\end{aligned}$$

$$s_t = (1-\gamma)s_{t-m} + \gamma y_t / (\ell_{t-1} + b_{t-1}).$$

A h pasos, tenemos que:

$$y_{t+h} = (\ell_{t+h-1} + b_{t+h-1})s_{t+h-m}(1 + \varepsilon_{t+h}).$$

Observamos que:

$$\begin{aligned}
\ell_{t+h-1} + b_{t+h-1} &= (\ell_{t+h-2} + b_{t+h-2})(1 + \alpha\varepsilon_{t+h-1}) \\
&\quad + b_{t+h-2} + \beta(\ell_{t+h-2} + b_{t+h-2})\varepsilon_{t+h-1} \\
&= \ell_{t+h-2} + b_{t+h-2} + f_1(\ell_{t+h-2}, b_{t+h-2})\varepsilon_{t+h-1} \\
&\quad + b_{t+h-2} + g_1(\ell_{t+h-2}, b_{t+h-2})\varepsilon_{t+h-1} \\
&= \ell_{t+h-2} + 2b_{t+h-2} + k_1(\ell_{t+h-2}, b_{t+h-2})\varepsilon_{t+h-1} \quad (\text{Con } k_1 = f_1 + g_1.) \\
&= \ell_{t+h-3} + b_{t+h-3}(1 + \alpha\varepsilon_{t+h-2}) \\
&\quad + 2(b_{t+h-3} + \beta(\ell_{t+h-3} + b_{t+h-3})\varepsilon_{t+h-2}) + k_1(\ell_{t+h-2}, b_{t+h-2})\varepsilon_{t+h-1} \\
&= \ell_{t+h-3} + b_{t+h-3} + f_2(\ell_{t+h-3}, b_{t+h-3})\varepsilon_{t+h-2} \\
&\quad + 2b_{t+h-3} + g_2(\ell_{t+h-3}, b_{t+h-3})\varepsilon_{t+h-2} + k_1(\ell_{t+h-2}, b_{t+h-2})\varepsilon_{t+h-1} \\
&= \ell_{t+h-3} + 3b_{t+h-3} + k_2(\ell_{t+h-3}, b_{t+h-3})\varepsilon_{t+h-2} \\
&\quad + k_1(\ell_{t+h-2}, b_{t+h-2})\varepsilon_{t+h-1} \quad (\text{Con } k_2 = f_2 + g_2.) \\
&= \\
&\quad \vdots \quad (\text{Se requieren } h - 1 \text{ iteraciones.}) \\
&= \ell_{t-1} + hb_{t-1} + k_{h-1}(\ell_{t-1}, b_{t-1})\varepsilon_t + \cdots + k_2(\ell_{t+h-3}, b_{t+h-3})\varepsilon_{t+h-2} \\
&\quad + k_1(\ell_{t+h-2}, b_{t+h-2})\varepsilon_{t+h-1}, \text{ con } k_i = f_i + g_i \text{ para } i = 1, 2, \dots, h - 1.
\end{aligned}$$

Las funciones k_1, k_2, \dots, k_{h-1} pueden a su vez iterarse hasta llegar al tiempo $t - 1$. En cada iteración se obtienen parámetros $\ell_{\text{índice anterior}-1}$, $b_{\text{índice anterior}-1}$ y $\varepsilon_{\text{índice anterior}}$. Al hacer todas las iteraciones necesarias, tenemos que:

$$\begin{aligned}
\ell_{t+h-1} + b_{t+h-1} &= \ell_{t-1} + hb_{t-1} + \hat{k}_{h-1}(\ell_{t-1}, b_{t-1})\varepsilon_t + \cdots + \hat{k}_2(\ell_{t-1}, b_{t-1})\varepsilon_{t+h-2} \\
&\quad + \hat{k}_1(\ell_{t-1}, b_{t-1})\varepsilon_{t+h-1},
\end{aligned}$$

donde las funciones $\hat{k}_i(\ell_{t-1}, b_{t-1})$ son el resultado de iterar las funciones k_i .

Además, si $h^* = h \bmod m$, se tiene que:

$$\begin{aligned}
s_{t+h-m} &= s_{t+h-2m}(1 + \gamma\varepsilon_{t+h-m}) \\
&= s_{t+h-2m} + \lambda_1(s_{t+h-2m})\varepsilon_{t+h-m} \\
&= s_{t+h-3m} + (1 + \gamma\varepsilon_{t+h-2m}) + \lambda_1(s_{t+h-2m})\varepsilon_{t+h-m} \\
&= s_{t+h-3m} + \lambda_2(s_{t+h-3m})\varepsilon_{t+h-2m} + \lambda_1(s_{t+h-2m})\varepsilon_{t+h-m} \\
&= \\
&\vdots \quad (\text{Se requieren } h + m - h^* \text{ iteraciones.}) \\
&= s_{t+h-(h+m+1-h^*)} + \lambda_{h+m-h^*}(s_{t+h-(h+m+1-h^*)})\varepsilon_{t+h-(h+m-h^*)} \\
&\quad + \cdots + \lambda_1(s_{t+h-2m})\varepsilon_{t+h-m} \\
&= s_{t-m-1+h^*} + \lambda_{h+m-h^*}(s_{t-m-1+h^*})\varepsilon_{t-m+h^*} \\
&\quad + \cdots + \lambda_1(s_{t+h-2m})\varepsilon_{t+h-m}.
\end{aligned}$$

De igual manera podemos iterar las funciones λ_i , obteniendo:

$$\begin{aligned}
s_{t+h-m} &= s_{t-m-1+h^*} + \hat{\lambda}_{h+t-h^*}(s_{t-m-1+h^*})\varepsilon_{t-m+h^*} \\
&\quad + \cdots + \hat{\lambda}_1(s_{t-m-1+h^*})\varepsilon_{t+h-m}.
\end{aligned}$$

Ahora vemos que:

$$\begin{aligned}
y_{t+h-1} &= (\ell_{t+h-1} + b_{t+h-1})s_{t+h-1}(1 + \varepsilon_{t+h}) \\
&= [\ell_{t-1} + hb_{t-1} + \hat{k}_{h-1}(\ell_{t-1}, b_{t-1})\varepsilon_t + \cdots + \hat{k}_2(\ell_{t-1}, b_{t-1})\varepsilon_{t+h-2} \\
&\quad + \hat{k}_1(\ell_{t-1}, b_{t-1})\varepsilon_{t+h-1}][s_{t-m-1+h^*} + \lambda_{h+m-h^*}(s_{t-m-1+h^*})\varepsilon_{t-m+h^*} \\
&\quad + \cdots + \lambda_1(s_{t+h-2m})\varepsilon_{t+h-m}] \\
&= (\ell_{t-1} + hb_{t-1})s_{t-m-1+h^*} + \sum_{i,j} c_{i,j}\varepsilon_i\varepsilon_j,
\end{aligned}$$

donde $c_{i,j}$ es un coeficiente que depende de variables de nivel ℓ_t y variables de crecimiento b_t , para tiempos anteriores o iguales a $t - 1$.

El pronóstico a h pasos está dado por:

$$\begin{aligned}
\hat{y}_{t+h-1|t-1} &= \mathbf{E}[y_{t+h-1} \mid \mathbf{x}_{t-1}] \\
&= \mathbf{E}[(\ell_{t-1} + hb_{t-1})s_{t-m-1+h^*} + \sum_{i,j} c_{i,j}\varepsilon_i\varepsilon_j \mid \mathbf{x}_{t-1}] \\
&= \mathbf{E}[(\ell_{t-1} + hb_{t-1})s_{t-m-1+h^*} \mid \mathbf{x}_{t-1}] + \sum_{i,j} c_{i,j}\mathbf{E}[\varepsilon_i\varepsilon_j \mid \mathbf{x}_{t-1}] \\
&= \mathbf{E}[(\ell_{t-1} + hb_{t-1})s_{t-m-1+h^*}] + \sum_{i,j} c_{i,j}\mathbf{E}[\varepsilon_i\varepsilon_j] \\
&= (\ell_{t-1} + hb_{t-1})s_{t-m-1+h^*} + 0^* \\
&= (\ell_{t-1} + hb_{t-1})s_{t-m-1+h^*}.
\end{aligned}$$

Nota*: Condicionando a \mathbf{x}_{t-1} , se tiene que ℓ_{t-1} , b_{t-1} y $s_{t-m-1+h^*}$ son constantes, y por otra parte, como los $c_{i,j}$ corresponden a tiempos a lo más $t-1$, también son constantes. En el caso de $\varepsilon_i\varepsilon_j$, si $i, j > t-1$, entonces, como la sucesión $\{\varepsilon_t\}_t$ es de variables no correlacionadas y $\mathbf{E}[\varepsilon_n] = 0$ para toda n , se tiene que $\mathbf{E}[\varepsilon_i\varepsilon_j \mid \mathbf{x}_{t-1}] = 0$.

En la práctica comúnmente se eligen $0 < \alpha, \beta, \gamma < 1$.

Finalmente exhibiremos un par de modelos con tendencia que se termina.

3.2.6. Modelo con error multiplicativo y tendencia aditiva que se termina ETS(M, A_d, N)

A tiempo t , este modelo tiene una razón de crecimiento modificada ϕb_{t-1} , con $0 \leq \phi < 1$, lo cual disminuye la variable de nivel ℓ_t . Esto es útil para casos en los que se piensa que la tendencia no habrá de continuar.

Este modelo está descrito por:

$$y_t = (\ell_{t-1} + \phi b_{t-1})(1 + \varepsilon_t), \quad (3.13a)$$

$$\ell_t = (\ell_{t-1} + \phi b_{t-1})(1 + \alpha \varepsilon_t), \quad (3.13b)$$

$$b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1})\varepsilon_t, \quad (3.13c)$$

con $0 \leq \phi < 1$.

**3.2.7. Modelo con error aditivo y tendencia multiplicativa que se termina
ETS(A,M_d,N)**

Las ecuaciones de este modelo son:

$$y_t = \ell_{t-1} b_{t-1}^\phi + \varepsilon_t, \quad (3.14a)$$

$$\ell_t = \ell_{t-1} b_{t-1}^\phi + \alpha \varepsilon_t, \quad (3.14b)$$

$$b_t = b_{t-1}^\phi + \beta \varepsilon_t / \ell_{t-1}, \quad (3.14c)$$

con $0 \leq \phi < 1$.

Capítulo 4

Estimación de parámetros iniciales

4.1. Función de máxima verosimilitud

Una manera de estimar los parámetros es mediante la función de máxima verosimilitud. En este caso, anteriormente se obtuvo que:

$$p(\mathbf{y} \mid \boldsymbol{\theta}, \mathbf{x}_0, \sigma^2) = \prod_{t=1}^n p(y_t \mid \boldsymbol{\theta}, \mathbf{x}_{t-1}, \sigma^2), \quad (4.1)$$

donde $\boldsymbol{\theta}$ es un vector de parámetros a estimar.

Dado que $y_t = w(\mathbf{x}_{t-1}) + r(\mathbf{x}_{t-1})\varepsilon_t$, condicionando a \mathbf{x}_{t-1} , se tiene que $w(\mathbf{x}_{t-1})$ y $r(\mathbf{x}_{t-1})$ son constantes, de modo que $p(y_t \mid \boldsymbol{\theta}, \mathbf{x}_{t-1}, \sigma^2) = p(\varepsilon_t) / |r(\mathbf{x}_{t-1})|$. Para esto recordemos un resultado que dice que si $Y = a + bX$, con a y b constantes, entonces $p(Y) = p(X) / |b|$.

De esta manera, tenemos que:

$$p(\mathbf{y} \mid \boldsymbol{\theta}, \mathbf{x}_0, \sigma^2) = \prod_{t=1}^n p(\varepsilon_t) / |r(\mathbf{x}_{t-1})|. \quad (4.2)$$

En el caso en que la distribución de ε_t es normal, se obtuvo que la función de verosimilitud está dada por:

$$L(\boldsymbol{\theta}, \mathbf{x}_0, \sigma^2 \mid \mathbf{y}) = (2\pi\sigma^2)^{-n/2} \left| \prod_{t=1}^n r(\mathbf{x}_{t-1}) \right|^{-1} \exp \left(-\frac{1}{2} \sum_{t=1}^n \varepsilon_t^2 / \sigma^2 \right). \quad (4.3)$$

Por lo tanto, la log-verosimilitud es:

$$\ln(L) = -(n/2)\ln(2\pi\sigma^2) - \sum_{t=1}^n \ln(|r(\mathbf{x}_{t-1})|) - \frac{1}{2} \sum_{t=1}^n \varepsilon_t^2 / \sigma^2. \quad (4.4)$$

Si bien σ^2 forma parte de los parámetros a estimar, vamos a considerarlo a parte. Al calcular la derivada parcial respecto a σ^2 , se tiene que:

$$\frac{\partial \ln(L)}{\partial \sigma^2} = -(n/2)\sigma^2 + \frac{1}{2} \sum_{t=1}^n \varepsilon_t^2 / (\sigma^2)^2.$$

Igualando a 0, se tiene que:

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{t=1}^n \varepsilon_t^2. \quad (4.5)$$

Sustituyendo $\hat{\sigma}^2$ en la ecuación 4.3, se obtiene que:

$$\begin{aligned} L(\boldsymbol{\theta}, \mathbf{x}_0 | \mathbf{y}) &= (2\pi\hat{\sigma}^2)^{-n/2} \left| \prod_{t=1}^n r(\mathbf{x}_{t-1}) \right|^{-1} e^{-n/2} \\ &= (2\pi e \hat{\sigma}^2)^{-n/2} \left| \prod_{t=1}^n r(\mathbf{x}_{t-1}) \right|^{-1}. \end{aligned}$$

De esto se tiene que:

$$\begin{aligned} -2\ln(L) &= -2\ln \left((2\pi e \hat{\sigma}^2)^{-n/2} \left| \prod_{t=1}^n r(\mathbf{x}_{t-1}) \right|^{-1} \right) \\ &= -2 \left[-(n/2)\ln(2\pi e \hat{\sigma}^2) - \ln \left(\prod_{t=1}^n r(\mathbf{x}_{t-1}) \right) \right] \\ &= n\ln(2\pi e \hat{\sigma}^2) + 2 \sum_{t=1}^n \ln(|r(\mathbf{x}_{t-1})|) \\ &= n\ln \left(\frac{2\pi e}{n} \right) + n\ln(n\hat{\sigma}^2) + 2 \sum_{t=1}^n \ln(|r(\mathbf{x}_{t-1})|) \\ &= c_n + n\ln \left(\sum_{t=1}^n \varepsilon_t^2 \right) + 2 \sum_{t=1}^n \ln(|r(\mathbf{x}_{t-1})|), \end{aligned}$$

donde $c_n = n\ln \left(\frac{2\pi e}{n} \right)$ (no depende ni de $\boldsymbol{\theta}$ ni de \mathbf{x}_0).

Por lo tanto, para estimar los parámetros, se tiene que minimizar la función:

$$L^*(\boldsymbol{\theta}, \mathbf{x}_0) = n \ln \left(\sum_{t=1}^n \varepsilon_t^2 \right) + 2 \sum_{t=1}^n \ln(|r(\mathbf{x}_{t-1})|). \quad (4.6)$$

De manera equivalente, se pueden estimar los parámetros al minimizarse la función:

$$S(\boldsymbol{\theta}, \mathbf{x}_0) = [\exp(L^*(\boldsymbol{\theta}, \mathbf{x}_0))]^{1/n} = \left| \prod_{t=1}^n r(\mathbf{x}_{t-1}) \right|^{2/n} \sum_{t=1}^n \varepsilon_t^2. \quad (4.7)$$

Este criterio es conocido en la literatura como Mínimos Cuadrados Aumentados (*Augmented Least Squares, ALS*).

4.2. Método heurístico

Vamos a considerar el siguiente método propuesto en [3, p. 23-24]. Éste sirve para obtener valores iniciales de las variables de estado $\ell_0, b_0, s_{-m+1}, \dots, s_0$, los cuales serán utilizados en el proceso de optimización de la función de verosimilitud. Estos valores iniciales ayudan a que el proceso converja más rápido. Para esto vamos a mencionar conceptos que serán de utilidad para tal método.

4.2.1. Promedio móvil

Si se tiene un proceso $\{Y_t\}_{t=0}^{\infty}$, para $m = 2k + 1$, con k un número natural, podemos definir un nuevo proceso $\{W_t\}_{t=0}^{\infty}$, llamado *m-promedio móvil*, el cual está definido por:

$$W_t = \frac{1}{m} \sum_{j=-k}^k Y_{t+j}, \quad (4.8)$$

con $k + 1 \leq t \leq n - k$.

Si m es un número par ($m = 2k$, con k un número natural), el *m-promedio móvil* está definido por:

$$W_t = \frac{1}{m} \sum_{j=-k+1}^k Y_{t+j}, \quad (4.9)$$

con $k \leq t \leq n - k$.

4.2.2. $2 \times m$ promedio móvil

Si se tiene un proceso $\{Y_t\}_{t=0}^{\infty}$, y $\{W_t\}_{t=0}^{\infty}$ es un m -promedio móvil de éste, con $m = 2k$, donde k es un número natural, entonces definimos un $2 \times m$ -promedio móvil como un proceso $\{\hat{T}_t\}_t$ tal que:

$$\hat{T}_t = \frac{1}{2}(W_{t-1} + W_t), \quad (4.10)$$

con $k + 1 \leq t \leq n - k$.

Observemos que:

$$\begin{aligned} \hat{T}_t &= \frac{1}{2}(W_{t-1} + W_t) \\ &= \frac{1}{2} \left(\frac{1}{m} \sum_{j=-k+1}^k Y_{t-1+j} + \frac{1}{m} \sum_{j=-k+1}^k Y_{t+j} \right) \\ &= \frac{1}{2m} Y_{t-k} + \frac{1}{2m} \sum_{j=-k+2}^k Y_{t-1+j} + \frac{1}{2m} \sum_{j=-k+1}^{k-1} Y_{t+j} + \frac{1}{2m} Y_{t+k} \\ &= \frac{1}{2m} Y_{t-k} + \frac{1}{2m} \sum_{j=-k+1}^{k-1} Y_{t+j} + \frac{1}{2m} \sum_{j=-k+1}^{k-1} Y_{t+j} + \frac{1}{2m} Y_{t+k} \\ &= \frac{1}{2m} Y_{t-k} + \frac{1}{m} \sum_{j=-k+1}^{k-1} Y_{t+j} + \frac{1}{2m} Y_{t+k} \\ &= \sum_{j=-k}^k \alpha_j Y_{t+j}, \end{aligned}$$

donde $\alpha_j = \frac{1}{m}$ para $-k + 1 \leq j \leq k - 1$, y $\alpha_j = \frac{1}{2m}$ para $j = \pm k$.

De este modo, un $2 \times m$ -promedio móvil se puede ver como un promedio ponderado de los valores Y_t tales que los valores extremos tienen un peso igual a $\frac{1}{2m}$ y los demás valores tienen un peso igual a $\frac{1}{m}$ (véase [4]).

4.2.3. Componente estacional inicial

Si m es el valor de la longitud de la estacionalidad, primeramente, si m es par, hay que calcular un $2 \times m$ -promedio móvil para los datos de los primeros años (2 para el modelo de Holt-Winters y 3 para los demás modelos ETS). Si m es impar, se calcula un m -promedio móvil. Los valores de los promedios móviles se denotarán $\{f_t\}_t$, con $t = m/2 + 1, m/2 + 2,$

... si m es par, y $t = (m - 1)/2 + 1, (m - 1)/2 + 2, \dots$ si m es impar.

Si la estacionalidad es aditiva, para eliminar la tendencia hay que definir el proceso $y_t - f_t$. Si la estacionalidad es multiplicativa, definimos el proceso y_t/f_t . Luego hay que calcular el promedio de los datos sin tendencia para cada estación. Por ejemplo, si se tienen datos mensuales y la serie empieza en enero, entonces el 2×12 promedio móvil $\{f_t\}_t$ se calcula para los tiempos $t = 7, 8, 9, \dots$. En este caso, $t = 7$ corresponde al mes de julio. Entonces, promediando los valores de $y_t - f_t$ (o de y_t/f_t , según sea la componente estacional, aditiva o multiplicativa) para los meses de julio, se obtiene el índice estacional para dicho mes. El índice estacional de enero sería s_{-11} , el de febrero s_{-10} , ... hasta el de diciembre, que sería s_0 . En general así se obtienen s_{-m+1}, \dots, s_0 . Finalmente se normalizan los valores obtenidos; esto es, que la suma de éstos sea 0 si la estacionalidad es aditiva, o que la suma sea igual a m si la estacionalidad es multiplicativa.

4.2.3.1. Normalización de índices estacionales

Sea $i = 0, \dots, m - 1$. Denotaremos $s_t^{(i)}$ la componente estacional a tiempo t que corresponde a la estación en el periodo de tiempo $t - i$. Podemos observar que $s_{t-1}^{(i-1)}$ es la componente estacional a tiempo $t - 1$ que corresponde a la estación en el periodo de tiempo $t - 1 - (i - 1) = t - i$, de forma que $s_t^{(i)}$ y $s_{t-1}^{(i-1)}$ son componentes estacionales de la misma estación, pero para tiempos distintos. De manera general, en todos los modelos que tienen componente estacional del tipo aditivo, podemos expresar las componentes estacionales por medio de las siguientes ecuaciones:

$$s_t^{(0)} = s_{t-1}^{(m-1)} + \gamma q(\mathbf{x}_{t-1})\varepsilon_t, \quad (4.11a)$$

$$s_t^{(i)} = s_{t-1}^{(i-1)}, \text{ para } i = 1, \dots, m - 1. \quad (4.11b)$$

Si el error del modelo es de tipo aditivo, se tiene que $q(\mathbf{x}_{t-1}) = 1$, y si el error es de tipo multiplicativo, entonces $q(\mathbf{x}_{t-1}) = \hat{y}_{t|t-1}$.

Para normalizar los datos con estacionalidad aditiva, buscamos que $\sum_{i=0}^{m-1} s_{t-i} = 0$. Para esto, para cada i modificaremos la componente $s_t^{(i)}$ por una nueva componente $\tilde{s}_t^{(i)}$ definida por:

$$\tilde{s}_t^{(0)} = \tilde{s}_{t-1}^{(m-1)} + \gamma q(\tilde{\mathbf{x}}_{t-1})\varepsilon_t - a_t, \quad (4.12a)$$

$$\tilde{s}_t^{(i)} = \tilde{s}_{t-1}^{(i-1)} - a_t, \text{ para } i = 1, \dots, m - 1, \quad (4.12b)$$

con a_t el factor de normalización dado por $a_t = (\gamma/m)q(\tilde{\mathbf{x}}_{t-1})\varepsilon_t$.

Notemos que:

$$\begin{aligned}
\sum_{i=0}^{m-1} \tilde{s}_t^{(i)} &= \tilde{s}_t^{(0)} + \sum_{i=1}^{m-1} \tilde{s}_t^{(i)} \\
&= \tilde{s}_t^{(0)} + \sum_{i=1}^{m-1} [\tilde{s}_{t-1}^{(i-1)} - a_t] \quad (\text{Usando 4.12b.}) \\
&= \tilde{s}_{t-1}^{(m-i)} + \gamma q(\tilde{\mathbf{x}}_{t-1})\varepsilon_t - a_t + \sum_{i=1}^{m-1} [\tilde{s}_{t-1}^{(i-1)} - a_t] \quad (\text{Usando 4.12a.}) \\
&= \tilde{s}_{t-1}^{(m-1)} + \gamma q(\tilde{\mathbf{x}}_{t-1})\varepsilon_t - a_t + \left[\sum_{i=1}^{m-1} \tilde{s}_{t-1}^{(i-1)} \right] - (m-1)a_t \\
&= \tilde{s}_{t-1}^{(m-1)} + \gamma q(\tilde{\mathbf{x}}_{t-1})\varepsilon_t - ma_t + \sum_{i=1}^{m-1} \tilde{s}_{t-1}^{(i-1)} \tag{4.13} \\
&= \tilde{s}_{t-1}^{(m-1)} + \gamma q(\tilde{\mathbf{x}}_{t-1})\varepsilon_t - m(\gamma/m)q(\tilde{\mathbf{x}}_{t-1})\varepsilon_t + \sum_{i=1}^{m-1} \tilde{s}_{t-1}^{(i-1)} \\
&= \tilde{s}_{t-1}^{(m-1)} + \sum_{i=1}^{m-1} \tilde{s}_{t-1}^{(i-1)} \\
&= \sum_{i=1}^m \tilde{s}_{t-1}^{(i-1)} \\
&= \sum_{i=0}^{m-1} \tilde{s}_{t-1}^{(i)}.
\end{aligned}$$

Por lo tanto, si las componentes estacionales para el tiempo $t-1$ están normalizadas, también los estarán para el tiempo t . De este modo, si suponemos que $\sum_{i=0}^{m-1} \tilde{s}_{-i} = \sum_{i=0}^{m-1} \tilde{s}_0^{(-i)} = 0$, entonces, para cada tiempo en el futuro de la serie, las componentes estacionales estarán normalizadas.

Para el caso en el que la estacionalidad es multiplicativa, buscamos que se cumpla $\sum_{i=0}^{m-1} s_{t-i} = m$.

Usando la misma notación del caso aditivo, modificaremos las componentes de la siguiente

manera:

$$\tilde{s}_t^{(0)} = [\tilde{s}_{t-1}^{(m-1)} + \gamma q(\tilde{\mathbf{x}}_{t-1})\varepsilon_t]/r_t, \quad (4.14a)$$

$$\tilde{s}_t^{(i)} = \tilde{s}_{t-1}^{(i-1)}/r_t, \text{ para } i = 1, \dots, m-1, \quad (4.14b)$$

con r_t el factor de normalización dado por $r_t = 1 + (\gamma/m)q(\tilde{\mathbf{x}}_{t-1})\varepsilon_t$, donde $q(\tilde{\mathbf{x}}_{t-1})$ toma valores según los siguientes casos:

a) Modelo con error de tipo multiplicativo: $q(\tilde{\mathbf{x}}_{t-1}) = \tilde{s}_{t-1}^{(m-1)}$.

b) Modelo sin tendencia y con error aditivo: $q(\tilde{\mathbf{x}}_{t-1}) = 1/\tilde{\ell}_{t-1}$.

c) Modelo con tendencia aditiva y error de tipo aditivo: $q(\tilde{\mathbf{x}}_{t-1}) = 1/(\tilde{\ell}_{t-1} + \tilde{b}_{t-1})$.

d) Modelo con tendencia aditiva que se desvanece y error de tipo aditivo: $q(\tilde{\mathbf{x}}_{t-1}) = 1/(\tilde{\ell}_{t-1} + \phi\tilde{b}_{t-1})$.

e) Modelo con tendencia multiplicativa y error aditivo: $q(\tilde{\mathbf{x}}_{t-1}) = 1/(\tilde{\ell}_{t-1}\tilde{b}_{t-1})$.

f) Modelo con tendencia multiplicativa que se desvanece y error aditivo: $q(\tilde{\mathbf{x}}_{t-1}) = 1/(\tilde{\ell}_{t-1}\tilde{b}_{t-1}^\phi)$.

Si para el tiempo $t-1$ las componentes estacionales están normalizadas; esto es, $\sum_{i=0}^{m-1} \tilde{s}_{t-1}^{(i)} = m$.

Entonces, se tiene que:

$$\begin{aligned}
\sum_{i=0}^{m-1} \tilde{s}_t^{(i)} &= \tilde{s}_t^{(0)} + \sum_{i=1}^{m-1} \tilde{s}_t^{(i)} \\
&= [\tilde{s}_{t-1}^{(m-1)} + \gamma q(\tilde{\mathbf{x}}_{t-1})\varepsilon_t]/r_t + \sum_{i=1}^{m-1} \tilde{s}_{t-1}^{(i-1)}/r_t \\
&= \left[\sum_{i=1}^m \tilde{s}_{t-1}^{(i-1)}/r_t \right] + \gamma q(\tilde{\mathbf{x}}_{t-1})\varepsilon_t/r_t \\
&= \left[\sum_{i=0}^{m-1} \tilde{s}_{t-1}^{(i)}/r_t \right] + \gamma q(\tilde{\mathbf{x}}_{t-1})\varepsilon_t/r_t \\
&= m/r_t + \gamma q(\tilde{\mathbf{x}}_{t-1})\varepsilon_t/r_t \\
&= \frac{m + \gamma q(\tilde{\mathbf{x}}_{t-1})\varepsilon_t}{1 + (\gamma/m)q(\tilde{\mathbf{x}}_{t-1})\varepsilon_t} \\
&= \frac{m[1 + (\gamma/m)q(\tilde{\mathbf{x}}_{t-1})\varepsilon_t]}{1 + (\gamma/m)q(\tilde{\mathbf{x}}_{t-1})\varepsilon_t} \\
&= m.
\end{aligned}$$

Por lo tanto, las componentes estacionales a tiempo t también están normalizadas. De esta manera, si suponemos que $\sum_{i=0}^{m-1} \tilde{s}_{-i} = \sum_{i=0}^{m-1} \tilde{s}_0^{(-i)} = m$, se tiene que para el resto de la serie las componentes estacionales estarán normalizadas.

4.2.4. Componente inicial del nivel

Para estimar el componente inicial del nivel ℓ_0 , si se tienen datos estacionales, se debe ajustar una regresión lineal simple* respecto al tiempo t , $t = 1, \dots, 10$ para los primeros 10 valores de los índices estacionales obtenidos en el paso anterior. Si los datos no son estacionales, hay que ajustar un modelo de regresión lineal simple respecto al tiempo t , $t = 1, \dots, 10$ para los primeros 10 valores de la serie. En cualquiera de los dos casos se ajusta un modelo del tipo $a + bt$. En este caso, el valor inicial para el nivel será el intercepto de la regresión. Por lo tanto, $\ell_0 = a$.

Nota*: Para ver cómo se ajusta un modelo de regresión lineal simple, consultar [7, sección 11.3].

4.2.5. Componente inicial del factor de crecimiento

Para estimar el componente inicial del factor de crecimiento b_0 , si se trata de un modelo de tendencia aditiva, entonces se elige $b_0 = b$, con b el valor de la pendiente de la recta ajustada en el procedimiento anterior. Si la tendencia es multiplicativa, se elige $b_0 = 1 + b/a$.

Capítulo 5

Intervalos de predicción para pronósticos puntuales

5.1. Media y varianza de predicción

Con los modelos de espacio de estados podemos obtener pronósticos a h pasos a partir de observaciones hasta un tiempo $t = n$. Si se asumen conocidos el modelo, los parámetros y el valor de \mathbf{x}_0 , entonces se puede conocer el valor del vector de estados a tiempo n . Por ejemplo, dados $\mathbf{x}_0, y_1, \dots, y_n$, podemos usar las ecuaciones $\varepsilon_t = (y_t - \hat{y}_{t|t-1})/r(\mathbf{x}_{t-1})$ y $\mathbf{x}_t = f(\mathbf{x}_{t-1}) + g(\mathbf{x}_{t-1})\varepsilon_t$ para calcular de forma iterativa $\varepsilon_1, \mathbf{x}_1, \varepsilon_2, \mathbf{x}_2, \dots$. Vamos a obtener entonces pronósticos $\hat{y}_{n+h|n}$; esto es, condicionando hasta \mathbf{x}_n . Para obtener intervalos de predicción para los pronósticos, una manera es hacer simulaciones de diferentes posibles trayectorias a partir de las ecuaciones que describen el modelo y usar los cuantiles de las trayectorias. Para esto, se hacen simulaciones sobre el error ε_t para obtener las trayectorias. Esto tiene la ventaja de que incluso si la distribución del error no es normal, se pueden hacer simulaciones de otra distribución. Para todos los modelos ETS se pueden obtener intervalos por medio de este método.

Otra manera de obtener los intervalos de predicción es obteniendo una expresión analítica para la media $\mu_{n+h|n}$ y la varianza $v_{n+h|n}$ de $y_{n+h|n}$. Para realizar esto, se tiene que:

$$y_{n+h} = w(\mathbf{x}_{n+h-1}) + r(\mathbf{x}_{n+h-1})\varepsilon_{n+h}, \quad (5.1a)$$

$$\mathbf{x}_{n+h} = f(\mathbf{x}_{n+h-1}) + g(\mathbf{x}_{n+h-1})\varepsilon_{n+h}. \quad (5.1b)$$

Para el caso de los modelos ETS lineales y homocedásticos podemos escribir:

$$\begin{aligned} y_{n+h} &= \mathbf{w}'\mathbf{x}_{n+h-1} + \varepsilon_{n+h}, \\ \mathbf{x}_{n+h} &= \mathbf{F}\mathbf{x}_{n+h-1} + \mathbf{g}\varepsilon_{n+h}. \end{aligned}$$

Si $h = 1$, entonces:

$$y_{n+1} = \mathbf{w}'\mathbf{x}_n + \varepsilon_{n+1}.$$

Usando el mismo argumento que en 2.7, para la media de predicción tenemos que:

$$\begin{aligned}\mu_{n+1|n} &= \hat{y}_{n+1|n} \\ &= \mathbb{E}[y_{n+1|n} \mid \mathbf{x}_n] \\ &= \mathbf{w}'\mathbf{x}_n.\end{aligned}\tag{5.2}$$

Para la varianza de predicción se tiene que:

$$\begin{aligned}v_{n+1|n} &= \text{Var}[y_{n+1} \mid \mathbf{x}_n] \\ &= \text{Var}[\mathbf{w}'\mathbf{x}_n + \varepsilon_{n+1} \mid \mathbf{x}_n] \\ &= \text{Var}[\varepsilon_{n+1} \mid \mathbf{x}_n] \\ &= \text{Var}[\varepsilon_{n+1}]^* \\ &= \sigma^2.\end{aligned}\tag{5.3}$$

Nota*: Ya que en el caso lineal, \mathbf{x}_n depende de $\varepsilon_n, \varepsilon_{n-1}, \dots, \varepsilon_1$ y \mathbf{x}_0 , pues

$$\mathbf{x}_t = \mathbf{F}^t\mathbf{x}_0 + \sum_{j=0}^{t-1} \mathbf{F}^j\mathbf{g}\varepsilon_{t-j}.$$

Si $h \geq 2$, entonces:

$$\begin{aligned}y_{n+h} &= \mathbf{w}'\mathbf{x}_{n+h-1} + \varepsilon_{n+h} \\ &= \mathbf{w}'(\mathbf{F}\mathbf{x}_{n+h-2} + \mathbf{g}\varepsilon_{n+h-1}) + \varepsilon_{n+h} \\ &= \mathbf{w}'\mathbf{F}\mathbf{x}_{n+h-2} + \mathbf{w}'\mathbf{g}\varepsilon_{n+h-1} + \varepsilon_{n+h} \\ &= \mathbf{w}'\mathbf{F}(\mathbf{F}\mathbf{x}_{n+h-3} + \mathbf{g}\varepsilon_{n+h-2}) + \mathbf{w}'\mathbf{g}\varepsilon_{n+h-1} + \varepsilon_{n+h} \\ &= \mathbf{w}'\mathbf{F}^2\mathbf{x}_{n+h-3} + \mathbf{w}'\mathbf{F}\mathbf{g}\varepsilon_{n+h-2} + \mathbf{w}'\mathbf{g}\varepsilon_{n+h-1} + \varepsilon_{n+h} \\ &= \\ &\vdots \\ &= \mathbf{w}'\mathbf{F}^{h-1}\mathbf{x}_n + \left[\sum_{j=1}^{h-1} \mathbf{w}'\mathbf{F}^{j-1}\mathbf{g}\varepsilon_{n+h-j} \right] + \varepsilon_{n+h}.\end{aligned}$$

Por lo tanto:

$$\begin{aligned}
\mu_{n+h|n} &= E[y_{n+h} \mid \mathbf{x}_n] \\
&= E[\mathbf{w}'\mathbf{F}^{h-1}\mathbf{x}_n + \left[\sum_{j=1}^{h-1} \mathbf{w}'\mathbf{F}^{j-1}\mathbf{g}\varepsilon_{n+h-j} \right] + \varepsilon_{n+h} \mid \mathbf{x}_n] \\
&= \mathbf{w}'\mathbf{F}^{h-1}\mathbf{x}_n + \left[\sum_{j=1}^{h-1} \mathbf{w}'\mathbf{F}^{j-1}\mathbf{g}E[\varepsilon_{n+h-j} \mid \mathbf{x}_n] \right] + E[\varepsilon_{n+h} \mid \mathbf{x}_n] \\
&= \mathbf{w}'\mathbf{F}^{h-1}\mathbf{x}_n, *
\end{aligned} \tag{5.4}$$

$$\begin{aligned}
v_{n+h|n} &= \text{Var}[y_{n+h} \mid \mathbf{x}_n] \\
&= \text{Var}[\mathbf{w}'\mathbf{F}^{h-1}\mathbf{x}_n + \left[\sum_{j=1}^{h-1} \mathbf{w}'\mathbf{F}^{j-1}\mathbf{g}\varepsilon_{n+h-j} \right] + \varepsilon_{n+h} \mid \mathbf{x}_n] \\
&= \left[\sum_{j=1}^{h-1} (\mathbf{w}'\mathbf{F}^{j-1}\mathbf{g})\text{Var}[\varepsilon_{n+h-j} \mid \mathbf{x}_n](\mathbf{w}'\mathbf{F}^{j-1}\mathbf{g})' \right] + \text{Var}[\varepsilon_{n+h} \mid \mathbf{x}_n] \\
&= \sigma^2 \left[\sum_{j=1}^{h-1} (\mathbf{w}'\mathbf{F}^{j-1}\mathbf{g})(\mathbf{w}'\mathbf{F}^{j-1}\mathbf{g})' \right] + \sigma^{2*} \\
&= \sigma^2 \left[1 + \sum_{j=1}^{h-1} (\mathbf{w}'\mathbf{F}^{j-1}\mathbf{g})(\mathbf{w}'\mathbf{F}^{j-1}\mathbf{g})' \right] \\
&= \sigma^2 \left[1 + \sum_{j=1}^{h-1} c_j^2 \right],
\end{aligned} \tag{5.5}$$

donde $c_j = \mathbf{w}'\mathbf{F}^{j-1}\mathbf{g}$.

Nota*: Asumiendo que $\{\varepsilon_t\}_t$ son variables independientes e idénticamente distribuidas.

Como $y_{n+h} = \mathbf{w}'\mathbf{x}_{n+h-1} + \varepsilon_{n+h}$ y además se tiene el supuesto de que ε_t tiene distribución normal, entonces, condicionalmente a \mathbf{x}_n , se tiene que y_{n+h} también tiene distribución normal con media $\mu_{n+h|n}$ y varianza $v_{n+h|n}$. Por lo tanto, el $100(1 - \alpha)$ intervalo de confianza para los pronósticos a h pasos es $\text{Int}_{100(1-\alpha)} = \mu_{n+h|n} \pm z_{1-\alpha/2} \sqrt{v_{n+h|n}}$, donde $z_{1-\alpha/2}$ es el cuantil $1 - \alpha/2$ de una distribución $N(0, 1)$.

En general, el procedimiento consiste en calcular $\mu_{n+h|n}$ y $v_{n+h|n}$, y con el supuesto de normalidad de ε_t , el $100(1 - \alpha)$ intervalo de confianza es el mismo que el mencionado arriba. Lo único que varía de un modelo a otro es la complejidad de los cálculos necesarios para obtener

$\mu_{n+h|n}$ y $v_{n+h|n}$, lo cual no siempre es posible.

En [3, p. 76] podemos encontrar el cuadro 5.1, que hace una clasificación de los modelos en 5 clases, según la derivación de la fórmula analítica para $\mu_{n+h|n}$ y $v_{n+h|n}$. La clase 1 consiste de modelos en los que el error es aditivo y las componentes de tendencia y estacionalidad pueden ser aditivas o inexistentes. La clase 2 se compone de modelos en los que el error es multiplicativo y las componentes de tendencia y estacionalidad pueden ser aditivas o inexistentes. La clase 3 consiste de modelos en los que el error es multiplicativo, la componente de tendencia puede ser aditiva o no existir y la componente estacional es multiplicativa. La clase 4 contiene modelos en los que el error es multiplicativo, la componente de tendencia es multiplicativa y la componente estacional puede ser multiplicativa o inexistente. Finalmente, la clase 5 se conforma de modelos en los que el error es principalmente aditivo (con algunos casos de tipo multiplicativo), la componente de tendencia es aditiva o multiplicativa y la componente estacional de todo tipo.

Clase 1 →	A,N,N	A,N,A			
	A,A,N	A,A,A			
	A,A _d ,N	A,A _d ,A			
Clase 2 →	M,N,N	M,N,A	M,N,M		
	M,A,N	M,A,A	M,A,M	← Clase 3	
	M,A _d ,N	M,A _d ,A	M,A _d ,M		
Clase 4 →	M,M,N		M,M,M		
	M,M _d ,N		M,M _d ,M		
Clase 5 →		M,M,A	A,N,M	A,M,N	A,M _d ,N
		M,M _d ,A	A,A,M	A,M,A	A,M _d ,A
			A,A _d ,M	A,M,M	A,M _d ,M

Cuadro 5.1: Clases de los modelos ETS.

Para las primeras tres clases es posible obtener una fórmula analítica de la media y la varianza, de modo que con éstas se pueden obtener los intervalos de predicción. Sin embargo, para la clase 4 no se puede obtener una expresión analítica de la media o la varianza, de modo que se realiza el método de simulación para obtener los intervalos de predicción de los pronósticos. Para la clase 5, se realiza de igual manera el método de simulación. Véase [3, p. 95-104], donde se muestra cómo se derivan las fórmulas de $\mu_{n+h|n}$ y $v_{n+h|n}$ para cada clase. En la página 81 se muestra una tabla con los valores de $\mu_{n+h|n}$ y c_j para los modelos de las clases 1 y 2, y en la página 84 los valores de $\mu_{n+h|n}$ y c_j para los modelos de la clase 3.

5.2. Ejemplo de intervalo de predicción del modelo ETS(A,A,N)

Se tienen datos mensuales del índice bursátil S&P 500, de enero de 2010 a septiembre de 2016 (81 observaciones). Se va a utilizar un modelo ETS(A,A,N) (método lineal de Holt) y se va a considerar que ya se conocen los valores de los parámetros del modelo, así como de las

variables de estado iniciales. Se tiene que $\alpha = 0.5$, $\beta = 0.4$, $\ell_0 = 1087.9488$ y $b_0 = 34.6055$. Se obtendrán los pronósticos e intervalos de predicción con un nivel de significancia del 95 % para los siguientes 6 tiempos; es decir, los últimos 3 meses de 2016 y los primeros 3 meses de 2017.

Se utilizará la librería `quantmod` de R para obtener los datos:

```
library(quantmod)
```

La función `getSymbols()` incluida en la librería `quantmod` descarga los datos del símbolo seleccionado. La clave del S&P 500 es `^GSPC`. Los parámetros `from` y `to` especifican el intervalo de tiempo para el cual se obtienen los datos. Los datos consisten de un objeto que tiene como columnas la fecha, valor de apertura (Open), valor máximo (High), valor mínimo (Low), valor de cierre (Close), Volumen (Volume) y valor de cierre ajustado (Adjusted Close). El objeto se almacena en una variable de nombre `GSPC`:

```
getSymbols('^GSPC', from = "2010-01-01", to = "2016-10-01",
           adjusted = TRUE)
```

La función `getSymbols()` obtiene valores diarios, por lo que para obtener los valores mensuales se usa la función `to.monthly()`. Guardamos los valores mensuales del índice en la variable `datos1`:

```
datos1 <- to.monthly(GSPC)
```

Necesitamos obtener los valores de cierre, por lo que se usa la función `Cl()`. Creamos la variable `serie1`, la cual le da a los valores una estructura de serie de tiempo. El parámetro `frequency = 12` indica que los valores son mensuales. El parámetro `start` especifica el año y mes desde donde comienza la serie:

```
serie1 <- ts(Cl(datos1), frequency = 12, start = c(2010, 01))
```

Visualizamos el gráfico de la variable `serie1` con la función `plot()`:

```
plot(serie1, main = 'S&P 500', xlab = 'Tiempo', ylab = 'Puntos')
```

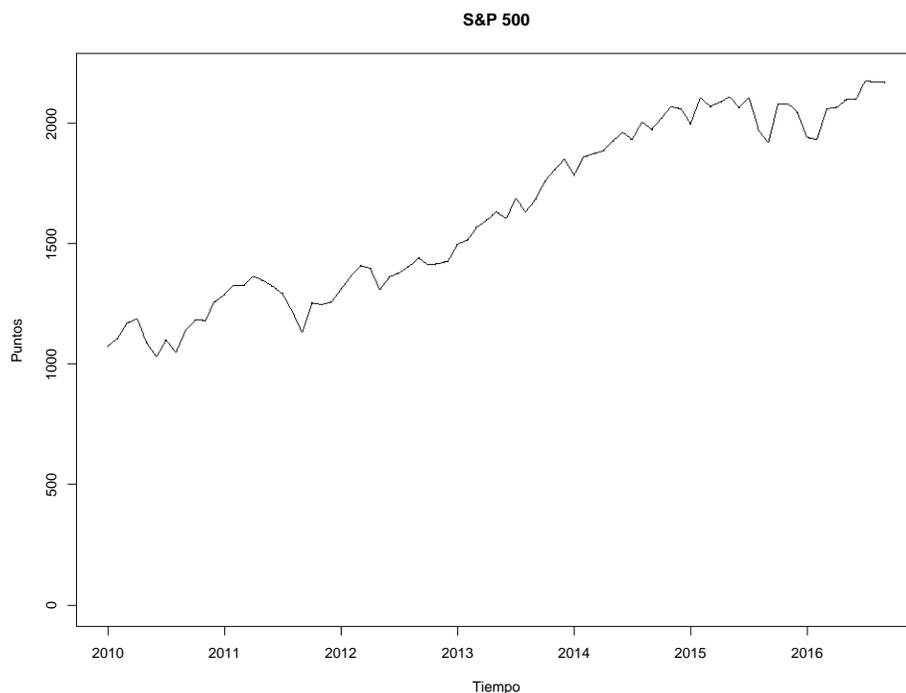


Figura 5.1: Gráfico de puntos mensuales del S&P 500.

Especificamos los parámetros del modelo:

```
alpha <- 0.5
beta <- 0.4
```

Creamos las variables de estado iniciales:

```
l_0 <- 1087.9488
b_0 <- 34.6055

l_t <- c(l_0)
b_t <- c(b_0)
```

l_t y b_t son inicialmente escalares cuyo único valor es l_0 y b_0 , respectivamente. A continuación se realiza un ciclo `for` para ir agregando, en la iteración i , los valores de l_i y b_i de l_0 y b_0 . De esta manera, en la iteración i , $l_t = (l_0, l_1, l_2, \dots, l_i)$; $b_t = (b_0, b_1, \dots, b_i)$.

Creamos el vector de estados x_t con las variables de estado l_t y b_t para cada tiempo:

```
for(i in seq(1, length(serie1))){
  if(i == 1){
    l_t = c(l_t, alpha*serie1[i]+(1-alpha)*(l_0 + b_0))
    b_t = c(b_t, (beta/alpha)*(l_t[2]-l_0) + (1-(beta/alpha))*b_0)
  } else{
    l_t = c(l_t, alpha*serie1[i]+(1-alpha)*(l_t[i]+b_t[i]))
```

```

    b_t = c(b_t, (beta/alpha)*(l_t[i+1]-l_t[i])
            + (1-(beta/alpha))*b_t[i])
  }
}

```

Ahora vamos a obtener los valores ajustados del modelo para los tiempos t , de $t = 1$ a $t = 81$.

Se crea un vector sin elementos de nombre `fitted` y con un ciclo `for` se agregan los valores ajustados:

```

fitted <- c()

for(i in seq(1, length(serie1))){
  if(i == 1){
    fitted = c(fitted, l_0 + b)
  } else{
    fitted = c(fitted, l_t[i] + b_t[i])
  }
}

```

Se convierten los valores en un objeto del tipo de serie de tiempo:

```
fitted <- ts(fitted, frequency = 12, start = c(2010, 01))
```

Vamos a realizar el pronóstico de los 6 periodos siguientes. Se crea la variable `serie2` para replicar la serie original y añadir los valores pronosticados. Para esto se crean también las variables `l1_t` y `b1_t` para replicar las variables `l_t` y `b_t`, respectivamente, y añadir los nuevos valores:

```

serie2 <- serie1
l1_t <- l_t
b1_t <- b_t

```

Realizamos los pronósticos:

```

for(i in seq(length(serie1)+1, length(serie1) + 6)){
  serie2 = c(serie2, l1_t[i] + b1_t[i])
  l1_t = c(l1_t, alpha*serie2[i]+(1-alpha)*(l1_t[i]+b1_t[i]))
  b1_t = c(b1_t, (beta/alpha)*(l1_t[i+1]-l1_t[i])
            + (1-(beta/alpha))*b1_t[i])
}

```

Almacenamos los pronósticos en la variable `prediction`, la cual es del tipo de serie de tiempo:

```
prediction <- ts(tail(serie2, 6), frequency = 12, start = c(2016, 10))
```

Estimaremos $\hat{\sigma}^2$. Para este modelo se tiene que $\varepsilon_t = y_t - \hat{y}_t$ y $\hat{\sigma}^2 = \frac{1}{n} \sum_{t=1}^n \varepsilon_t^2$.

Se crea la variable `residuals`, la cual es un vector de ceros. Con un ciclo `for` se sustituye cada entrada con el valor de ε_t :

64CAPÍTULO 5. INTERVALOS DE PREDICCIÓN PARA PRONÓSTICOS PUNTUALES

```
residuals <- rep(0, length(serie1))

for(i in seq(1, length(serie1))) {
  residuals[i] = serie1[i]-fitted[i]
}

sigma2 <- mean(residuals^2)
```

Obtendremos el intervalo de predicción mediante la fórmula analítica. Para este modelo se tiene que $\mu_{n+h|n} = \ell_n + hb_n$ y que $c_j = \alpha + \beta_j$:

```
lower <- c()
upper <- c()
c_j <- alpha+beta*seq(1,5)
v <- 0

for (i in seq(1, 6)){
  if(i == 1){lower = c(lower,
    qnorm(0.025, mean = l_t[length(l_t)] + i*b_t[length(b_t)],
    sd = sqrt(sigma2)))
  upper = c(upper,
    qnorm(0.975, mean = l_t[length(l_t)] + i*b_t[length(b_t)],
    sd = sqrt(sigma2)))
  } else{

  v = sigma2*(1 + sum(c_j[1:i-1]^2))

  lower = c(lower,
    qnorm(0.025, mean = l_t[length(l_t)] + i*b_t[length(b_t)],
    sd = sqrt(v)))
  upper = c(upper,
    qnorm(0.975, mean = l_t[length(l_t)] + i*b_t[length(b_t)],
    sd = sqrt(v)))
  }
}

forecast_analytic <- cbind(prediction, lower, upper)
colnames(forecast_analytic) <- c('Forecast', 'Lower', 'Upper')
```

Se obtuvo que:

```
forecast_analytic
      Forecast Lower   Upper
Oct 2016 2203.761 2074.492 2333.029
Nov 2016 2213.460 2039.547 2387.374
Dec 2016 2223.160 1981.320 2465.000
Jan 2017 2232.860 1906.088 2559.632
Feb 2017 2242.559 1817.738 2667.381
Mar 2017 2252.259 1718.486 2786.032
```

Veamos gráficamente los resultados:

```
plot(serie1, main = 'S&P 500', xlab = 'Tiempo', ylab = 'Puntos',
     xlim = c(2010, 2018), ylim = c(0, 2800))
lines(fitted, col = 'red')
lines(forecast_analytic[,1], col = 'blue', type = 'o', pch = 20, cex = 1)
lines(forecast_analytic[,2], col = 'blue', type = 'o', pch = 20, cex = 1)
lines(forecast_analytic[,3], col = 'blue', type = 'o', pch = 20, cex = 1)
legend(2010, 2600, c('A,A,N', 'Valores ajustados', 'Serie original'),
      lty = c(1, 1, 1), col = c('blue', 'red', 'black'))
```

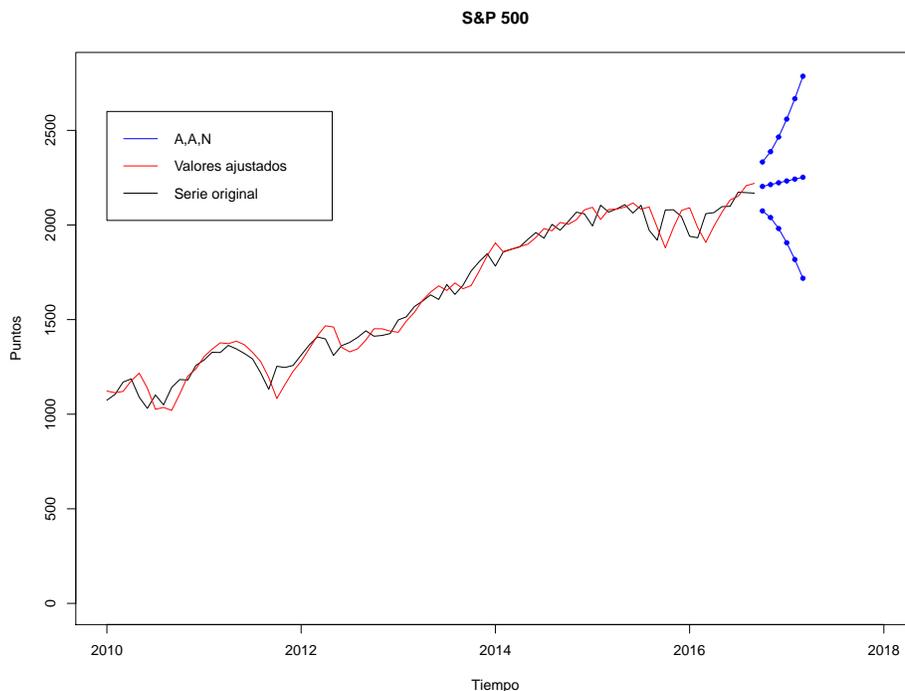


Figura 5.2: Método de fórmula analítica.

Con una función de R que realiza los pronósticos de manera automática, se puede hacer este mismo cálculo:

```
library(expsmooth)

modelo <- ets(serie1, model = "AAN", alpha = 0.5, beta = 0.4,
             gamma = NULL, damped = FALSE)
pred_analytic <- forecast(modelo, h = 6, level = 95)

pred_analytic
  Point Forecast      Lo 95      Hi 95
Oct 2016      2203.761 2074.492 2333.029
Nov 2016      2213.460 2039.547 2387.374
Dec 2016      2223.160 1981.320 2465.000
Jan 2017      2232.860 1906.088 2559.632
```

Feb 2017	2242.559	1817.738	2667.381
Mar 2017	2252.259	1718.486	2786.032

Los resultados son los mismos. Para otro modelo solamente habría que cambiar la expresión para $\mu_{n+h|n}$ y C_j .

Ahora realizaremos los intervalos de predicción para el mismo modelo por medio del método de simulación.

Para obtener los intervalos de predicción vamos a generar un data frame para los posibles trayectorias y sus posibles vectores de estados. Se crea la variable `path`, cuyas columnas almacenan 5000 posibles pasos para los siguientes 6 periodos. La variable `lpred` y `bpred` almacenan en cada una de sus columnas los 5000 posibles valores para las variables `l_t` y `b_t`, respectivamente:

```
path <- as.data.frame(matrix(0, 5000, 6))
lpred <- as.data.frame(matrix(0, 5000, 6))
bpred <- as.data.frame(matrix(0, 5000, 6))
```

Se crea la variable `epsilon` y se realiza un ciclo `for` para crear en cada paso un vector de 5000 números aleatorios con distribución $N(0, \hat{\sigma}^2)$ por medio de la función `rnorm()`. En cada paso se van almacenando los valores de las trayectorias y las variables de estados:

```
epsilon = c()
for (i in seq(1,6)){
  epsilon = rnorm(5000, sd = sqrt(sigma2))
  if(i == 1){
    path[,i] = l_t[length(serie1)+1]+b_t[length(serie1)+1]+epsilon
    lpred[,i] = l_t[length(serie1)+1]+b_t[length(serie1)+1]+alpha*epsilon
    bpred[,i] = b_t[length(serie1)+1] + beta*epsilon
  } else{
    path[,i] = lpred[i-1]+bpred[i-1]+epsilon
    lpred[,i] = lpred[i-1]+bpred[i-1]+alpha*epsilon
    bpred[,i] = bpred[i-1] + beta*epsilon
  }
}
```

Se crea la variable `confint`, que es un data frame de 6 filas y 2 columnas donde se van a almacenar el intervalo inferior y superior:

```
confint <- as.data.frame(matrix(0, 6, 2))
```

Con un ciclo doble se almacenan los intervalos de las trayectorias en cada tiempo. Para esto se usan los cuantiles 0.025 y 0.975 de los vectores de trayectorias en cada paso. Esto se realiza con la función `quantile()`:

```
for(j in seq(1,2)){
  for(i in seq(1,6)){
    if(j == 1){
```

```

        confint[i,j] = quantile(path[,i], probs = c(0.025))
    }
    else{
        confint[i,j] = quantile(path[,i], probs = c(0.975))
    }
}
}
}

```

La variable `forecast` almacena los pronósticos y los intervalos de predicción:

```

forecast <- cbind(prediction, confint[,1], confint[,2])
colnames(forecast) <- c('Forecast', 'Lower', 'Upper')

```

Se obtuvo:

```

forecast
      Forecast Lower   Upper
Oct 2016 2203.761 2076.165 2335.427
Nov 2016 2213.460 2038.568 2388.100
Dec 2016 2223.160 1976.781 2468.534
Jan 2017 2232.860 1899.038 2564.384
Feb 2017 2242.559 1816.364 2670.723
Mar 2017 2252.259 1727.710 2789.073

```

Graficamos los resultados:

```

plot(serie1, main = 'S&P 500', xlab = 'Tiempo', ylab = 'Puntos',
     xlim = c(2010, 2018), ylim = c(0, 2800))
lines(fitted, col = 'red')
lines(forecast[,1], col = 'blue', type = 'o', pch = 20, cex = 1)
lines(forecast[,2], col = 'blue', type = 'o', pch = 20, cex = 1)
lines(forecast[,3], col = 'blue', type = 'o', pch = 20, cex = 1)
legend(2010, 2600, c('A,A,N', 'Valores ajustados', 'Serie original'),
     lty = c(1, 1, 1), col = c('blue', 'red', 'black'))

```

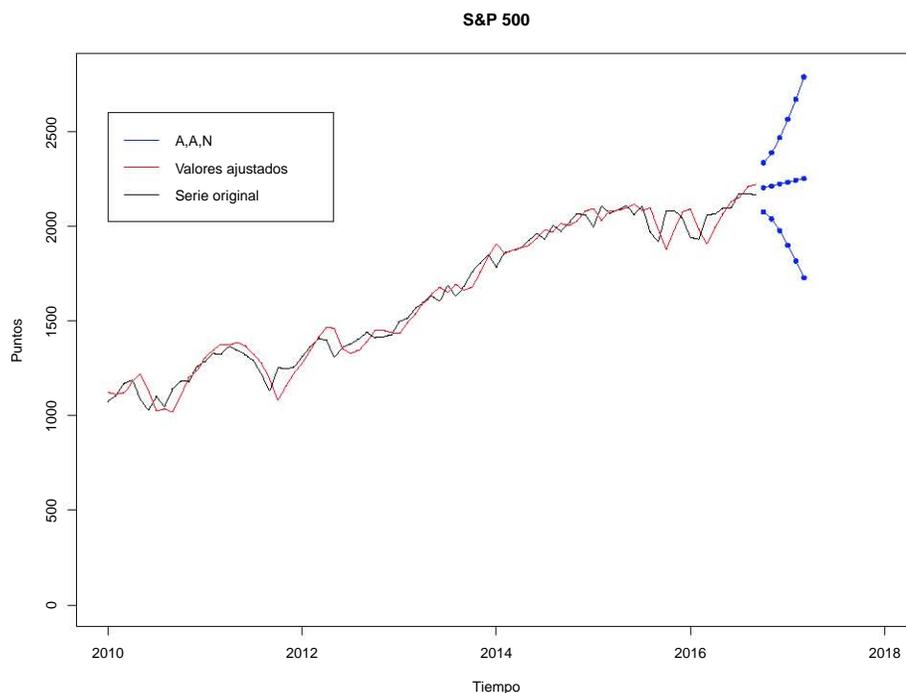


Figura 5.3: Método de simulación.

Con la función de R se obtuvo lo siguiente:

```
library(expsmooth)

modelo <- ets(serie1, model = "AAN", alpha = 0.5, beta = 0.4,
             gamma = NULL, damped = FALSE)
pred_analytic <- forecast(modelo, h = 6, level = 95, simulate = TRUE,
                        npaths = 5000)

predsim
  Point Forecast      Lo 95      Hi 95
Oct 2016      2203.761 2075.771 2335.260
Nov 2016      2213.460 2042.484 2387.173
Dec 2016      2223.160 1983.179 2473.192
Jan 2017      2232.860 1907.871 2561.356
Feb 2017      2242.559 1817.050 2669.593
Mar 2017      2252.259 1715.175 2798.040
```

Como se puede observar, los resultados son muy similares para los intervalos de predicción. Difieren ligeramente debido a que para cada proceso los números aleatorios generados no son los mismos. Para otros modelos, sería necesario modificar las ecuaciones que describen la serie y las variables de estado. Respecto a la componente de error, en caso de no tener normalidad, sería necesario generar números aleatorios con la distribución correspondiente. En caso de que el error tenga una distribución normal, la desviación estándar será la raíz cuadrada del

valor estimado de σ^2 , el cual se obtiene en la fase de estimación de los parámetros iniciales ya antes mencionada.

Comparemos ahora los resultados obtenidos por los dos modelos:

```
plot(serie1, main = 'S&P 500', xlab = 'Tiempo', ylab = 'Puntos',
     xlim = c(2010, 2018), ylim = c(0, 2800))

lines(fitted, col = 'red')
lines(forecast_analytic[,1], col = 'blue', type = 'o', pch = 20, cex = 1)
lines(forecast_analytic[,2], col = 'red', type = 'p', pch = 3, cex = 1)
lines(forecast_analytic[,3], col = 'red', type = 'p', pch = 3, cex = 1)

lines(forecast[,2], col = 'blue', type = 'p', pch = 0, cex = 1)
lines(forecast[,3], col = 'blue', type = 'p', pch = 0, cex = 1)
legend(2010, 2800,
      c('Pronósticos', 'Simulación', 'Fórmula analítica',
        'Valores ajustados', 'Serie original'),
      pch = c(20, 0, 3, NA, NA),
      lty = c(NA, NA, NA, 1, 1),
      col = c('blue', 'blue', 'red', 'red', 'black'))
```

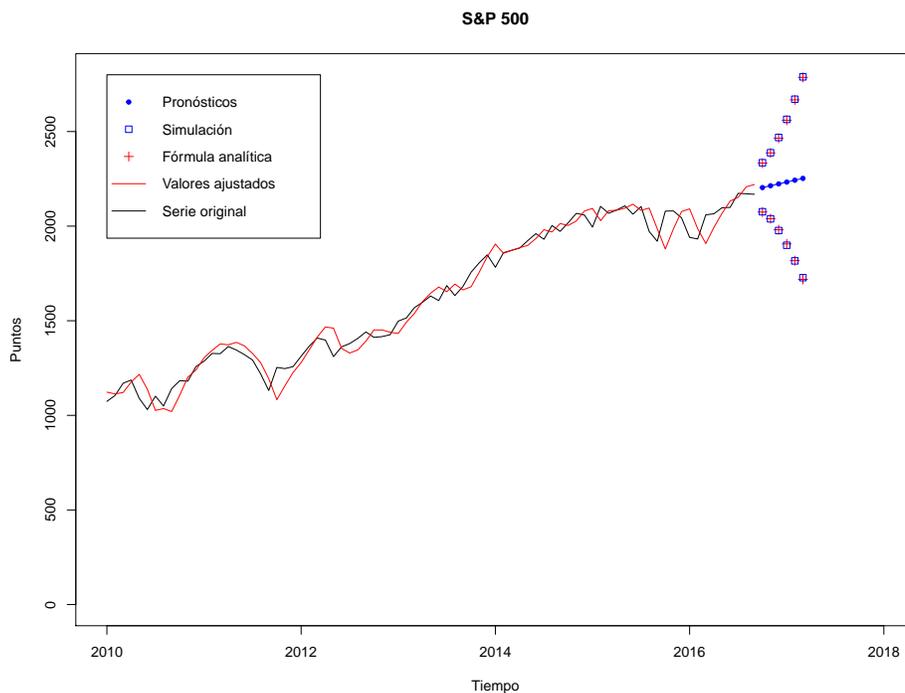


Figura 5.4: Comparación de los resultados obtenidos por los modelos.

Podemos observar que los intervalos de predicción obtenidos por cada modelo son muy cercanos.

5.3. Ejemplo de intervalo de predicción del modelo ETS(M,A,M)

Veamos otro ejemplo del método de simulación. Se tienen datos trimestrales del S&P 500, de enero de 2010 hasta septiembre de 2016. Utilizaremos un modelo ETS(M,A,M) y supondremos que los valores de los parámetros y de las variables iniciales ya están especificados. Tenemos que $\alpha = 0.02$, $\beta = 0.01$, $\gamma = 0.01$, $\ell_0 = 996.7884$ y $b_0 = 45.7497$, $s_0 = 1.0049$, $s_{-1} = 0.9682$, $s_{-2} = 0.9925$ y $s_{-3} = 1.0343$. Se obtendrán los pronósticos e intervalos de predicción con un nivel de significancia del 95 % para los siguientes 6 tiempos.

Nuevamente utilizamos la librería `quantmod` y la función `getSymbols()` para obtener los datos:

```
library(quantmod)

getSymbols('^GSPC', from = "2010-01-01", to = "2016-10-01",
          adjusted = TRUE)
```

Recordemos que la función `getSymbols()` obtiene valores diarios, por lo que para obtener los valores trimestrales se usa la función `to.quarterly()`. Guardamos los valores mensuales del índice en la variable `datos1`:

```
datos1 <- to.quarterly(GSPC)
```

Con la función `Cl()` obtenemos los valores de cierre y los guardamos en la variable `serie1`. La función `ts()` le da a los datos una estructura de una serie de tiempo. En este caso, el parámetro `frequency = 4` se refiere a que son datos trimestrales:

```
serie1 <- ts(Cl(datos1), frequency = 4, start = c(2010, 01))
```

Veamos el gráfico de la serie:

```
plot(serie1, main = 'S&P 500', xlab = 'Tiempo', ylab = 'Puntos')
```

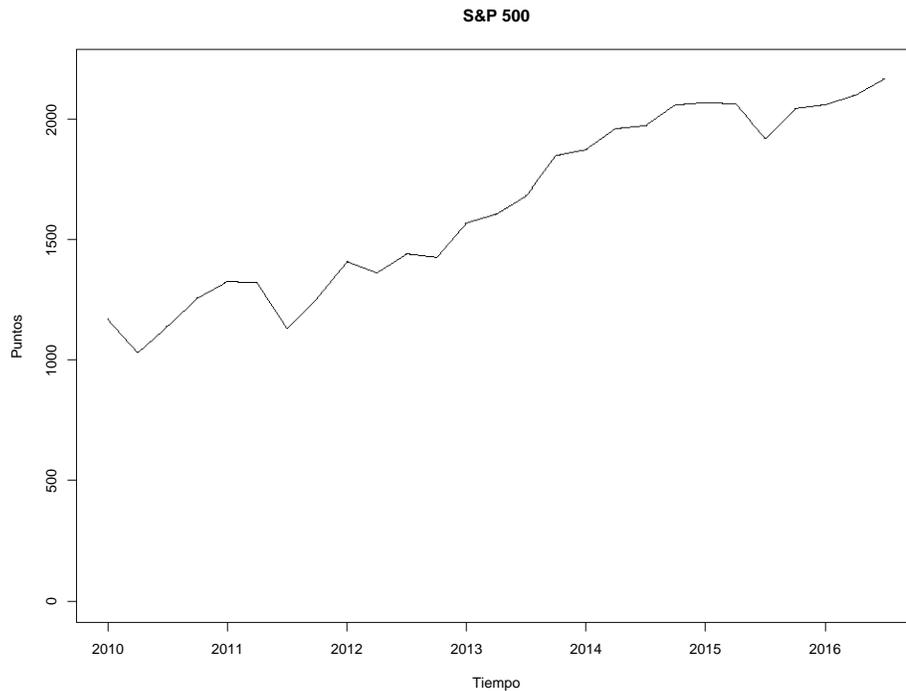


Figura 5.5: Gráfico de puntos trimestrales del S&P 500.

Definimos los parámetros iniciales:

```
alpha <- 0.02
beta <- 0.01
gamma <- 0.01
l_0 <- 996.7884
b_0 <- 45.7497
s1 <- 1.0049
s2 <- 0.9682
s3 <- 0.9925
s4 <- 1.0343
l_t <- c(l_0)
b_t <- c(b_0)
s_t <- c(s4, s3, s2, s1)
```

Hay que mencionar que s_4 corresponde a la variable de estado (teórica) s_{-3} ; s_3 corresponde a s_{-2} ; s_2 a s_{-1} , y s_1 a s_0 . Es importante que al definir la variable s_t se acomoden las variables s_4 , s_3 , s_2 y s_1 en el orden especificado arriba.

Creamos el vector x_t . Para esto calculamos las variables de estado l_t , b_t y s_t para cada tiempo:

```
for(i in seq(1, length(serie1))) {
  if(i == 1) {
    l_t = c(l_t, alpha*(serie1[1]/s_t[1]) + (1-alpha)*(l_0+b_0))
    b_t = c(b_t, beta/alpha*(l_t[2]-l_0)+(1-beta/alpha)*b_0)
    s_t = c(s_t, gamma*(serie1[1]/(l_0+b_0)) + (1-gamma)*s_t[1])
  }
}
```

```

} else if(i == 2){
  l_t = c(l_t, alpha*(serie1[2]/s_t[2]) + (1-alpha)*(l_t[2]+b_t[2]))
  b_t = c(b_t, beta/alpha*(l_t[3]-l_t[2])+(1-beta/alpha)*b_t[2])
  s_t = c(s_t, gamma*(serie1[2]/(l_t[2]+b_t[2])) + (1-gamma)*s_t[2])
} else if(i == 3){
  l_t = c(l_t, alpha*(serie1[3]/s_t[3]) + (1-alpha)*(l_t[3]+b_t[3]))
  b_t = c(b_t, beta/alpha*(l_t[4]-l_t[3])+(1-beta/alpha)*b_t[3])
  s_t = c(s_t, gamma*(serie1[3]/(l_t[3]+b_t[3])) + (1-gamma)*s_t[3])
} else if(i == 4){
  l_t = c(l_t, alpha*(serie1[4]/s_t[4]) + (1-alpha)*(l_t[4]+b_t[4]))
  b_t = c(b_t, beta/alpha*(l_t[5]-l_t[4])+(1-beta/alpha)*b_t[4])
  s_t = c(s_t, gamma*(serie1[4]/(l_t[4]+b_t[4])) + (1-gamma)*s_t[4])
} else{
  l_t = c(l_t, alpha*(serie1[i]/s_t[i]) + (1-alpha)*(l_t[i]+b_t[i]))
  b_t = c(b_t, beta/alpha*(l_t[i+1]-l_t[i])+(1-beta/alpha)*b_t[i])
  s_t = c(s_t, gamma*(serie1[i]/(l_t[i]+b_t[i])) + (1-gamma)*s_t[i])
}
}
}

```

Realizaremos el pronóstico de los 6 periodos siguientes. Se crea la variable `serie2` para replicar la serie original y añadir los valores pronosticados. Para esto se crean también las variables `l1_t`, `b1_t` y `s1_t` para replicar las variables `l_t` y `b_t` y `s_t`, respectivamente, y añadimos los nuevos valores:

```

serie2 <- serie1
l1_t <- l_t
b1_t <- b_t
s1_t <- s_t

for(i in seq(length(serie1)+1, length(serie1) + 6)){
  serie2 = c(serie2, (l1_t[i] + b1_t[i])*s1_t[i])
  l1_t = c(l1_t, alpha*(serie2[i]/s1_t[i]) + (1-alpha)*(l1_t[i]+b1_t[i]))
  b1_t = c(b1_t, beta/alpha*(l1_t[i+1]-l1_t[i])+(1-beta/alpha)*b1_t[i])
  s1_t = c(s1_t, gamma*(serie2[i]/(l1_t[i]+b1_t[i])) + (1-gamma)*s1_t[i])
}

```

Calcularemos los valores ajustados y los almacenaremos en la variable `fitted`. Además, le daremos a ésta una estructura de serie de tiempo con la función `ts()`:

```

fitted <- c()

for(i in seq(1, length(serie1))){
  if(i == 1){
    fitted = c(fitted, (l_0 + b_0)*s_t[1])
  }
  else if(i == 2){
    fitted = c(fitted, (l_t[2] + b_t[2])*s_t[2])
  }
}

```

```

else if(i == 3){
  fitted = c(fitted, (l_t[3] + b_t[3])*s_t[3])
}
else if(i == 4){
  fitted = c(fitted, (l_t[4] + b_t[4])*s_t[4])
}
else{
  fitted = c(fitted, (l_t[i] + b_t[i])*s_t[i])
}
}

fitted <- ts(fitted, frequency = 4, start = c(2010, 01))

```

Vamos a almacenar los pronósticos en la variable `prediction`, dándole además una estructura de serie de tiempo con la función `ts()`:

```
prediction <- ts(tail(serie2, 6), frequency = 4, start = c(2016, 4))
```

Estimaremos $\hat{\sigma}^2$ obteniendo los residuos del modelo. En este caso:

$$\varepsilon_t = (y_t - \hat{y}_t)/r(\mathbf{x}_{t-1}) = (y_t - \hat{y}_t)/((\ell_{t-1} + b_{t-1})s_{t-m}).$$

Después utilizamos que $\hat{\sigma}^2 = \frac{1}{n} \sum_{t=1}^n \varepsilon_t^2$. El resultado se guarda en la variable `sigma2`:

```

residuals <- rep(0, length(serie1))

for(i in seq(1, length(serie1))) {
  if(i == 1){
    residuals[i] = (serie1[i]-fitted[i])/((l_0+b_0)*s_t[1])

  } else if(i == 2){
    residuals[i] = (serie1[i]-fitted[i])/((l_t[2]+b_t[2])*s_t[2])

  } else if(i == 3){
    residuals[i] = (serie1[i]-fitted[i])/((l_t[3]+b_t[3])*s_t[3])

  } else if(i == 4){
    residuals[i] = (serie1[i]-fitted[i])/((l_t[4]+b_t[4])*s_t[4])

  } else{
    residuals[i] = (serie1[i]-fitted[i])/((l_t[i]+b_t[i])*s_t[i])
  }
}

sigma2 <- mean(residuals^2)

```

Calcularemos el intervalo de predicción simulando 5000 trayectorias para y_t almacenadas en la variable `path`, así como 5000 simulaciones para los parámetros ℓ_t , b_t y s_t , guardados en las variables `lpred`, `bpred` y `spred`, respectivamente. Para esto, en cada paso se simulan números aleatorios con distribución $N(0, \hat{\sigma}^2)$:

74CAPÍTULO 5. INTERVALOS DE PREDICCIÓN PARA PRONÓSTICOS PUNTUALES

```
path <- as.data.frame(matrix(0, 5000, 6))
lpred <- as.data.frame(matrix(0, 5000, 6))
bpred <- as.data.frame(matrix(0, 5000, 6))
spred <- as.data.frame(matrix(0, 5000, 6))
epsilon = c()
for (i in seq(1,6)){
  epsilon = rnorm(5000, sd = sqrt(sigma2))
  if(i == 1){
    path[,i] = (l1_t[length(serie1)+1]
    +b1_t[length(serie1)+1])*s1_t[length(serie1)+1]*(1+epsilon)

    lpred[,i] = (l1_t[length(serie1)+1]
    +b1_t[length(serie1)+1])*(1+alpha*epsilon)

    bpred[,i] = b1_t[length(serie1)+1]
    +beta*(l1_t[length(serie1)+1]+b1_t[length(serie1)+1])*epsilon

    spred[,i] = s1_t[length(serie1)+1]*(1+gamma*epsilon)
  } else if(i == 2){
    path[,i] = (l1_t[length(serie1)+2]
    +b1_t[length(serie1)+2])*s1_t[length(serie1)+2]*(1+epsilon)

    lpred[,i] = (l1_t[length(serie1)+2]
    +b1_t[length(serie1)+2])*(1+alpha*epsilon)

    bpred[,i] = b1_t[length(serie1)+2]
    +beta*(l1_t[length(serie1)+2]+b1_t[length(serie1)+1])*epsilon

    spred[,i] = s1_t[length(serie1)+2]*(1+gamma*epsilon)
  } else if(i == 3){
    path[,i] = (l1_t[length(serie1)+3]
    +b1_t[length(serie1)+3])*s1_t[length(serie1)+3]*(1+epsilon)

    lpred[,i] = (l1_t[length(serie1)+3]
    +b1_t[length(serie1)+3])*(1+alpha*epsilon)

    bpred[,i] = b1_t[length(serie1)+3]
    +beta*(l1_t[length(serie1)+3]+b1_t[length(serie1)+3])*epsilon

    spred[,i] = s1_t[length(serie1)+3]*(1+gamma*epsilon)
  } else if(i == 4){
    path[,i] = (l1_t[length(serie1)+4]
    +b1_t[length(serie1)+4])*s1_t[length(serie1)+4]*(1+epsilon)
```

```

lpred[,i] = (l1_t[length(serie1)+4]
+b1_t[length(serie1)+4])*(1+alpha*epsilon)

bpred[,i] = b1_t[length(serie1)+4]
+beta*(l1_t[length(serie1)+4]+b1_t[length(serie1)+4])*epsilon

spred[,i] = s1_t[length(serie1)+4]*(1+gamma*epsilon)

} else{
  path[,i] = (lpred[i-1]+bpred[i-1])*spred[i-4]*(1+epsilon)
  lpred[,i] = (lpred[i-1]+bpred[i-1])*(1+alpha*epsilon)
  bpred[,i] = bpred[i-1] + beta*(lpred[i-1]+bpred[i-1])*epsilon
  spred[,i] = spred[i-4]*(1+gamma*epsilon)
}
}

```

La variable `confint` es un data frame que almacena el intervalo de predicción para cada pronóstico:

```

confint <- as.data.frame(matrix(0, 6, 2))

for(j in seq(1,2)){
  for(i in seq(1,6)){
    if(j == 1){
      confint[i,j] = quantile(path[,i], probs = c(0.025))
    }
    else{
      confint[i,j] = quantile(path[,i], probs = c(0.975))
    }
  }
}

```

La variable `forecast` contiene los pronósticos a cada paso, así como el intervalo de predicción correspondiente:

```

forecast <- cbind(prediction, confint[,1], confint[,2])
colnames(forecast) <- c('Forecast', 'Lower', 'Upper')

```

Se obtuvo:

```

forecast
      Forecast      Lower      Upper
2016 Q4 2301.125 2035.229 2563.954
2017 Q1 2416.702 2122.608 2702.739
2017 Q2 2365.238 2090.700 2652.064
2017 Q3 2352.038 2065.273 2634.920
2017 Q4 2487.665 2188.731 2783.452
2018 Q1 2608.719 2306.568 2920.822

```

Graficamos los resultados:

```
plot(serie1, main = 'S&P 500', xlab = 'Tiempo', ylab = 'Puntos',
     xlim = c(2010, 2019), ylim = c(0, 3000))
lines(fitted, col = 'red')
lines(prediction, col = 'blue', type = 'o', pch = 20, cex = 1)
lines(forecast_lower, col = 'blue', type = 'o', pch = 20, cex = 1)
lines(forecast_upper, col = 'blue', type = 'o', pch = 20, cex = 1)
legend(2010, 2600, c('M,A,M', 'Valores ajustados', 'Serie original'),
      lty = c(1, 1, 1), col = c('blue', 'red', 'black'))
```

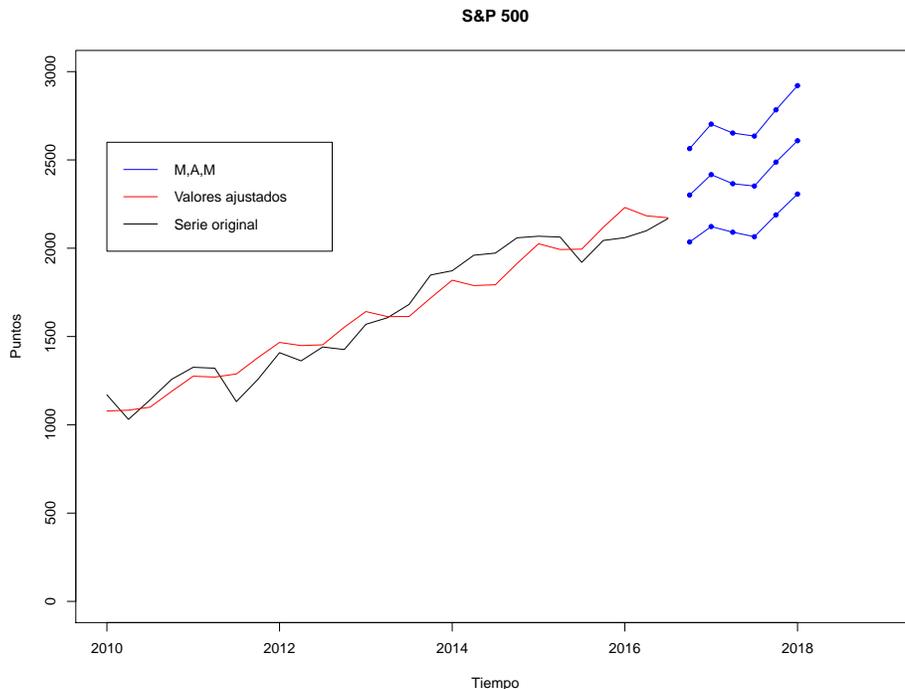


Figura 5.6: Método de simulación.

Ahora, usando la librería `expsmooth` de R, procedemos a calcular de nuevo los pronósticos y los intervalos de predicción para los pronósticos:

```
library(expsmooth)

modelo <- ets(serie1, model = "MAM", alpha = 0.02, beta = 0.01,
             gamma = 0.01, damped = FALSE)
predsim <- forecast(modelo, h = 6, level = 95, simulate = TRUE,
                  npaths = 5000)

predsim
  Point Forecast      Lo 95      Hi 95
2016 Q4      2301.086 2034.745 2581.476
2017 Q1      2416.716 2124.956 2697.428
2017 Q2      2365.168 2082.565 2642.166
```

2017 Q3	2352.063	2073.347	2632.383
2017 Q4	2487.617	2194.542	2786.748
2018 Q1	2608.729	2293.678	2910.976

Nuevamente observamos que los intervalos de predicción son muy similares. En el caso de los pronósticos, los decimales difieren ligeramente debido a que en el proceso de cálculo los valores de la variable de estado s_t fueron distintos solamente por decimales.

De esta manera se procede en el caso de un modelo que tiene una componente estacional s_t . Notemos que el procedimiento es casi el mismo que el que se realizó para el modelo A,A,N en los datos del S&P 500. Así concluimos con la parte del cálculo de intervalos de predicción para los pronósticos.

Capítulo 6

Selección de modelos

6.1. Criterios de información

Una vez que se tienen los elementos para ajustar modelos ETS hay que determinar qué modelo habría que escoger en caso de tener varios modelos adecuados para los datos. Si bien agregar parámetros a un modelo puede ayudar para explicar diferentes aspectos que intervienen en algún fenómeno de estudio, en ocasiones una cantidad excesiva de parámetros puede hacer que el modelo obtenga pronósticos menos precisos. Recordemos que la finalidad de estos modelos es obtener pronósticos lo más precisos posibles. Para esto, introduciremos los criterios de información, lo cuales penalizan la sobreparametrización en los modelos. De manera general, el criterio de información IC está dado por la siguiente ecuación:

$$\text{IC} = -2\ln(L(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{x}}_0 | \boldsymbol{y})) + q\zeta(n), \quad (6.1)$$

donde $L(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{x}}_0 | \boldsymbol{y})$ es la función de verosimilitud maximizada, q es el número de parámetros en $\hat{\boldsymbol{\theta}}$, más el número de estados libres; esto es, las variables que hay que estimar en $\hat{\boldsymbol{x}}_0$. La función $\zeta(n)$ depende del tamaño de muestra n y varía según el criterio de información. Presentamos algunos de ellos:

AIC: $\zeta(n) = 2$, de modo que:

$$\text{IC} = -2\ln(L(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{x}}_0 | \boldsymbol{y})) + 2q. \quad (6.2)$$

Este es el criterio de Akaike (1974).

BIC: $\zeta(n) = \ln(n)$, de modo que:

$$\text{IC} = -2\ln(L(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{x}}_0 | \boldsymbol{y})) + q\ln(n). \quad (6.3)$$

Este es el criterio de información bayesiano.

AICc: $\zeta(n) = 2n/(n - q - 1)$, de modo que:

$$\text{IC} = -2\ln(L(\hat{\boldsymbol{\theta}}, \hat{\mathbf{x}}_0 | \mathbf{y})) + 2qn/(n - q - 1). \quad (6.4)$$

Este es criterio de Akaike corregido.

El elemento $q\zeta(n)$ es la penalización particular de cada modelo.

En la práctica, al calcular el criterio de información de cada modelo, se debería elegir el modelo cuyo valor del criterio de información sea menor. En realidad, no se puede asegurar que un modelo sea mejor que otro, pero el criterio de información indica qué modelos están más sobreparametrizados que otros. Por lo tanto, éste es una referencia para elegir el modelo “más adecuado” de entre los que se tienen.

En el caso en que la distribución del error es normal, se tiene que:

$$\begin{aligned} \text{IC} &= -2\ln(L(\hat{\boldsymbol{\theta}}, \hat{\mathbf{x}}_0 | \mathbf{y})) + 2q \\ &= n\ln\left(\sum_{t=1}^n \varepsilon_t^2\right) + 2\sum_{t=1}^n \ln(|r(\mathbf{x}_{t-1})| + q\zeta(n)). \end{aligned} \quad (6.5)$$

6.2. Tipos de errores para medir exactitud en un modelo

A continuación veremos algunas medidas comunes para determinar la exactitud de un modelo al momento de realizar pronósticos puntuales.

6.2.1. Error absoluto medio

Sea $e_t = y_t - \hat{y}_t$. A h pasos se define $e_{t+h|t} = y_{t+h} - \hat{y}_{t+h|t}$. El error absoluto medio está determinado por:

$$\text{MAE} = \frac{1}{t} \sum_{i=1}^t |e_i|. \quad (6.6)$$

6.2.2. Error cuadrático medio

Definiendo e_t de la misma manera que en el caso anterior, el error cuadrático medio está dado por:

$$\text{MSE} = \frac{1}{t} \sum_{i=1}^t e_i^2. \quad (6.7)$$

6.2.3. Error porcentual

Sea $e_t = y_t - \hat{y}_t$, o bien $e_{t+h|t} = y_{t+h} - \hat{y}_{t+h|t}$. Definimos el error porcentual p_t de la siguiente manera:

$$p_t = 100e_t/y_t. \quad (6.8)$$

Este error mide el porcentaje, en cada paso, de la diferencia entre la observación y_t y el pronóstico \hat{y}_t respecto al valor observado. Este tipo de error puede no ser adecuado cuando se tienen magnitudes para las cuales un porcentaje no tendría sentido: por ejemplo, escalas de temperatura.

6.2.4. Error absoluto porcentual medio

Más comúnmente que el error porcentual, se utiliza el error absoluto porcentual medio, el cual está determinado por:

$$\text{MAPE} = \frac{1}{t} \sum_{i=1}^t |p_i|. \quad (6.9)$$

6.2.5. Error simétrico absoluto porcentual medio

Este tipo de error está definido por:

$$\text{SMAPE} = \frac{1}{t} \sum_{i=1}^t 200 |y_i - \hat{y}_i| / (y_i + \hat{y}_i). \quad (6.10)$$

En este caso se tiene el inconveniente de que puede haber una división por cero, además de que padece de las mismas limitaciones presentes en el error porcentual.

6.2.6. Errores escalados

Sea $e_t = y_t - \hat{y}_t$, o bien $e_{t+h|t} = y_{t+h} - \hat{y}_{t+h|t}$. Definimos el error escalado q_t como:

$$q_t = \frac{e_t}{\frac{1}{n-1} \sum_{i=2}^n |y_i - y_{i-1}|}. \quad (6.11)$$

Este tipo de error es independiente de la escala de las unidades de medición. A partir de éste podemos definir el error escalado absoluto medio de la siguiente manera:

$$\text{MASE} = \frac{1}{t} \sum_{i=1}^t |q_i|. \quad (6.12)$$

Capítulo 7

Caso práctico

Ahora veremos un ejemplo de cómo se hace un análisis de una serie de tiempo utilizando los modelos de espacio de estados. Para esto utilizaremos las librerías de R `expsmooth` y `forecast`. La librería `expsmooth` contiene las funciones necesarias para ajustar modelos de espacio de estados a una serie de tiempo, mientras que la librería `forecast` permite obtener los pronósticos puntuales así como los intervalos de predicción.

Con la librería `quantmod` y la función `getSymbols()` obtenemos datos de los precios de acciones de la compañía Amazon Inc, los cuales se guardan en un objeto de nombre `AMZN`. Con la función `to.monthly()` extraemos los precios de cierre mensuales, comenzando en octubre de 2010 y terminando en noviembre de 2016 (74 observaciones), y los almacenamos en la variable `datos`.

Posteriormente creamos la variable `serie`, la cual da a la variable `datos` una estructura de serie de tiempo. Hacemos entonces lo siguiente:

```
library(quantmod)

getSymbols('AMZN', from = "2010-10-01", to = "2016-11-30",
          adjusted = TRUE)

datos <- to.monthly(AMZN)

serie <- ts(Cl(datos), frequency = 12, start = c(2010, 10))
```

Ahora introduciremos la función `ets()` de la librería `expsmooth`. Esta función es la que ajusta los modelos de espacio de estados. La función se utiliza de la siguiente manera:

```
library(expsmooth)

ets(serie, model = 'ANN')
```

La función `ets()` requiere de un objeto de serie de tiempo. El parámetro `model` permite especificar el modelo de espacio de estados. En este ejemplo se ajusta el modelo A,N,N (Suavizamiento Exponencial Simple). La primera letra indica el tipo de error, el cual puede

tener cualquiera de las siguientes 3 opciones: A = error aditivo, M = error multiplicativo y Z = seleccionar automáticamente. De manera análoga, la segunda letra se refiere al tipo de tendencia; se tienen 4 opciones: N = sin tendencia, A = tendencia aditiva, M = tendencia multiplicativa y Z = seleccionar automáticamente. La tercera letra especifica el tipo de estacionalidad: N = sin componente estacional, A = estacionalidad aditiva, M = estacionalidad multiplicativa y Z = seleccionar automáticamente. Si quisiéramos ajustar el modelo AAA (Holt-Winters), se usaría lo siguiente:

```
ets(serie, model = 'AAA')
```

Hay que mencionar que por defecto los parámetros α , β , γ del modelo se estiman automáticamente, ya que la función `ets()` optimiza el modelo basándose en los parámetros que obtienen el menor valor para el criterio de Akaike. De igual manera, la decisión sobre si usar tendencia normal (sea ésta aditiva o multiplicativa) o si usar tendencia que se desvanece, se realiza de manera automática, según la opción que minimice el criterio de información. A su vez, en caso de que se use tendencia que se desvanece, el valor de ϕ se estima también de manera automática. Para especificar que se quiere usar tendencia normal, se hace lo siguiente:

```
ets(serie, model = 'AAA', damped = FALSE)
```

Si queremos usar específicamente tendencia que se desvanece, se utiliza:

```
ets(serie, model = 'AAA', damped = TRUE)
```

Por otra parte, los parámetros pueden especificarse manualmente. Si queremos usar el modelo AA_dA (siendo A_d tendencia que se desvanece) y queremos elegir $\alpha = 0.9999$, $\beta = 0.0261$, $\gamma = 1 \times 10^{-4}$ (1e-4 en R), $\phi = 0.98$, utilizamos:

```
ets(serie, model = 'AAA', alpha = 0.9999, beta = 0.0261, gamma = 1e-04,
    phi = 0.98)
```

Con esto ya podemos especificar cualquier modelo que queramos. Hay que resaltar que no todas las combinaciones de valores de parámetros son aceptadas por la función `ets()`. Finalmente, mencionamos que el parámetro `model` tampoco es necesario. Si queremos que la función `ets()` decida el modelo y los parámetros que optimizan el valor del criterio de información, escribimos simplemente:

```
ets(serie)
```

Vamos ahora a analizar la serie. Recordemos que la serie termina en el mes de noviembre de 2016. Nuestro objetivo es obtener pronósticos para los siguientes 4 meses, con sus respectivos intervalos de predicción.

Para esto, como se tienen los datos correspondientes a diciembre de 2016, enero, febrero y marzo de 2017, vamos a leer éstos de manera que podamos comparar los pronósticos del modelo con los valores que realmente tomó la serie en los tiempos posteriores. Utilizaremos nuevamente la función `getSymbols()` para obtener los datos, los convertiremos a datos mensuales con la función `to.monthly()`, los almacenaremos en la variable `datos_ref` y se creará la variable `serie_ref`, la cual ya tiene la estructura de una serie de tiempo. De esta manera la serie tiene

los mismos datos que la serie original más los valores para diciembre de 2016, enero, febrero y marzo de 2017:

```
getSymbols('AMZN', from = "2010-10-01", to = "2017-04-01",
           adjusted = TRUE)

datos_ref <- to.monthly(AMZN)

serie_ref <- ts(Cl(datos_ref), frequency = 12, start = c(2010, 10))
```

Visualizamos el gráfico de la variable serie con la función plot():

```
plot(serie, main = 'Precios mensuales Amazon Inc.', xlab = 'Tiempo',
     ylab = 'Precio')
```

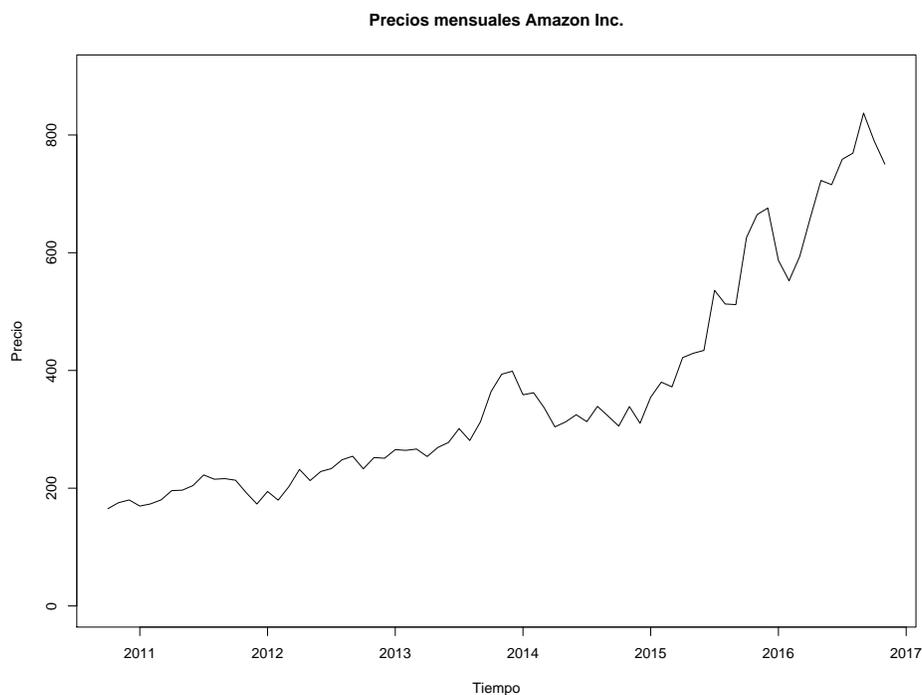


Figura 7.1: Gráfico de precios.

Podemos observar que los valores de los precios siguen una tendencia ascendente. No parece haber ciclos estacionales, de modo que consideraremos un modelo que contenga solamente una componente de error y una de tendencia. No parece que la tendencia esté por terminarse (véase figura 7.1), de modo que ésta se considerará como una tendencia normal (sin *dampening*). El modelo que elegiremos será el modelo A,A,N; esto es, el modelo de error aditivo y tendencia aditiva.

Vamos a ajustar a la variable `serie` el modelo con la función `ets()`. Hacemos lo siguiente:

```
modelo <- ets(serie, model = 'AAN', damped = FALSE)
```

Ahora veamos los parámetros optimizados estimados por la función `ets()`:

```
modelo
ETS(A,A,N)
```

Call:

```
ets(y = serie, model = "AAN", damped = FALSE)
```

Smoothing parameters:

```
alpha = 0.9999
```

```
beta = 1e-04
```

Initial states:

```
l = 158.4272
```

```
b = 8.0153
```

```
sigma: 31.6903
```

AIC	AICc	BIC
839.9904	840.8727	851.5107

El valor de α es de 0.9999 y el de β es de 1×10^{-4} . Los parámetros iniciales son: $\ell_0 = 158.4272$ y $b_0 = 8.0153$. El valor estimado de $\hat{\sigma}$ es de 31.6903. La función `ets()` también nos proporciona los valores del criterio de Akaike, el criterio de Akaike corregido y el del criterio de información bayesiano.

Para obtener los pronósticos utilizaremos la función `forecast()` de la librería `forecast`. Ésta se aplica a la variable `modelo`, la cual contiene el modelo ajustado a la serie. Por defecto, si existe una expresión analítica de la media y varianza de predicción, entonces el intervalo de predicción se calcula a partir de la fórmula analítica. En caso contrario, la función `forecast()` utiliza el método de simulación. Por otra parte, si desde un principio se desea usar el método de simulación, se agrega el parámetro `simulate = TRUE`.

Vamos a obtener los pronósticos e intervalos de predicción para los siguientes 4 periodos usando un nivel de confianza del 95 por ciento. La función muestra tres columnas, siendo la primera la de los pronósticos puntuales, la segunda la del intervalo inferior y la tercera la del intervalo superior. Almacenamos los resultados en la variable `prediction`:

```
library(forecast)
```

```
prediction <- forecast(modelo, h = 4, level = 95)
```

Si quisiéramos utilizar el método de simulación para obtener el intervalo de predicción, haríamos:

```
prediction <- forecast(modelo, h = 4, level = 95, simulate = TRUE)
```

El parámetro `h` se refiere al número de periodos para calcular los pronósticos y el parámetro `level` indica el nivel de confianza. Veamos los resultados obtenidos al usar la fórmula analítica:

```
prediction
  Point Forecast    Lo 95    Hi 95
Dec 2016      758.5901 696.4782 820.7019
Jan 2017      766.6054 678.7660 854.4448
Feb 2017      774.6207 667.0363 882.2051
Mar 2017      782.6360 658.4030 906.8689
```

Ahora veamos los valores reales almacenados en la variable `serie_ref`. Para esto usamos la función `tail()`, la cual utiliza como objeto la variable `serie_ref` y además usamos el parámetro `n = 4` para indicar que se muestren los 4 últimos valores:

```
tail(serie_ref, n = 4)
  Jan    Feb    Mar Apr May Jun Jul Aug Sep Oct Nov Dec
2016
2017 823.48 845.04 886.54                                749.87
```

Podemos observar que los valores de los pronósticos no coinciden con los valores reales; sin embargo, estos últimos se encuentran dentro de los intervalos de predicción, lo cual es un buen resultado del modelo.

Esto se puede ver más claramente en el siguiente gráfico. Hacemos lo siguiente:

```
plot(prediction, main = 'Pronósticos de ETS(A,A,N) vs datos reales',
      xlab = 'Tiempo', ylab = 'Precio', ylim = c(0, 900), shaded = FALSE,
      flty = 0, pch.lty = 0)

lines(fitted(modelo), col = 'red')

lines(prediction$mean, col = 'blue', type = 'o', pch = 20, cex = 1)

lines(prediction$lower, col = 'blue', type = 'o', pch = 20, cex = 1)

lines(prediction$upper, col = 'blue', type = 'o', pch = 20, cex = 1)

lines(tail(serie_ref, 4), col = 'black', type = 'o', pch = 20, cex = 1)

legend(2011, 900, c('A,A,N', 'Valores ajustados', 'Serie original'),
      lty = c(1, 1, 1), col = c('blue', 'red', 'black'))
```

Con la función `plot()` graficamos la serie original junto con los pronósticos puntuales y la región de los intervalos de predicción. La función `lines()` agrega los valores ajustados del modelo obtenidos con la función `fitted()` (de la librería `forecast`) al gráfico original. De igual manera, agregamos los valores futuros reales, los cuales son los últimos 4 valores de la variable `serie_ref`:

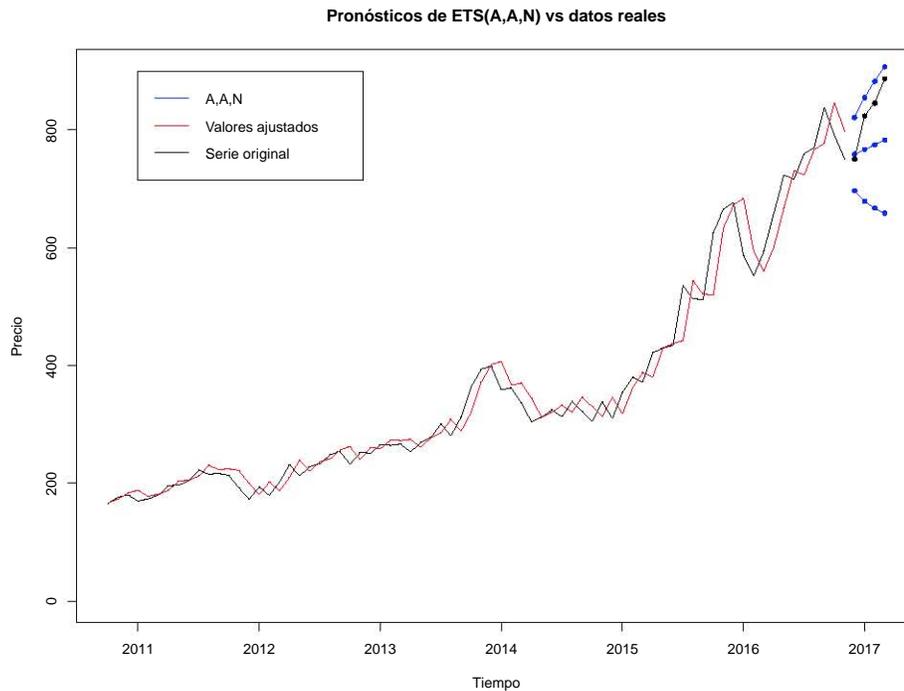


Figura 7.2: Comparación de pronósticos con datos reales.

Las líneas azules de los extremos corresponden a los intervalos de predicción. La línea negra que se muestra antes de estos últimos corresponde a la serie y_t hasta noviembre de 2016. La línea roja muestra los valores ajustados \hat{y}_t por el modelo hasta noviembre de 2016. La línea azul que está entre los intervalos de predicción muestra los pronósticos puntuales para diciembre de 2016, enero, febrero y marzo de 2017. La línea negra que se encuentra entre los intervalos muestra los valores reales de la serie para los 4 meses mencionados.

De esta manera es más fácil notar qué tanto coinciden los valores de los pronósticos con los valores reales de la serie y además si estos últimos se encuentran dentro de la región de los intervalos de predicción, lo cual se cumple.

Ahora vamos a analizar los residuos para verificar el supuesto de que son un ruido blanco con distribución normal. Para esto, primeramente obtenemos los residuos y los guardamos en la variable `residuos`. Hacemos lo siguiente:

```
residuos <- modelo$residuals
```

Ahora vamos a utilizar la función `tsdiag()` de la librería `forecast`, la cual nos muestra el gráfico de los residuos, su función de autocorrelación y el gráfico de los p-values de la prueba de Ljung-Box:

```
tsdiag(modelo)
```

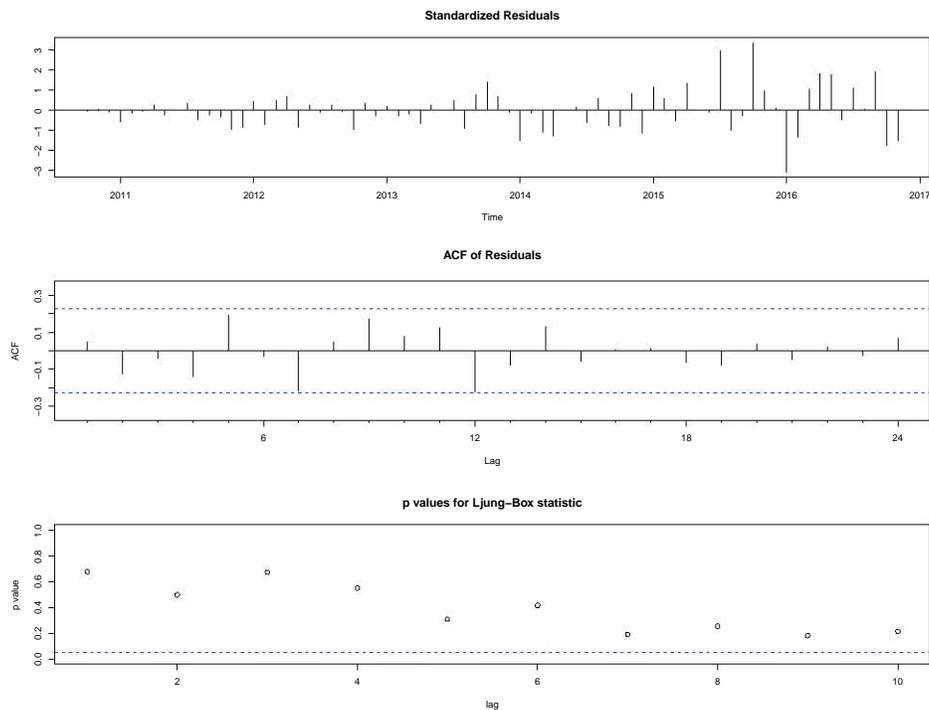


Figura 7.3: Gráfico y función de autocorrelación de los residuos y prueba de Ljung-Box.

Podemos ver que los residuos tienen algunos valores que son considerablemente mayores que los otros, de modo que la varianza no es constante. En la función de autocorrelación ninguno de los valores se sale de las bandas de predicción (color azul), que es lo que se esperaría de un ruido blanco. Todos los p-valores superan la línea de valor 0.05 (dependiendo este número del nivel de significancia de la prueba), por lo tanto, no se rechaza la hipótesis nula de la prueba de Ljung-Box que dice que los valores de las autocorrelaciones son cero.

Veamos ahora si éstos siguen una distribución normal. Usaremos las funciones `qqnorm()` y `qqline()`:

```
qqnorm(residuos, xlab = "Residuos", ylab = "Valores Esperados",
       main = "Probabilidad Normal", pch = 1, col = "blue")
```

```
qqline(residuos, col = 2)
```

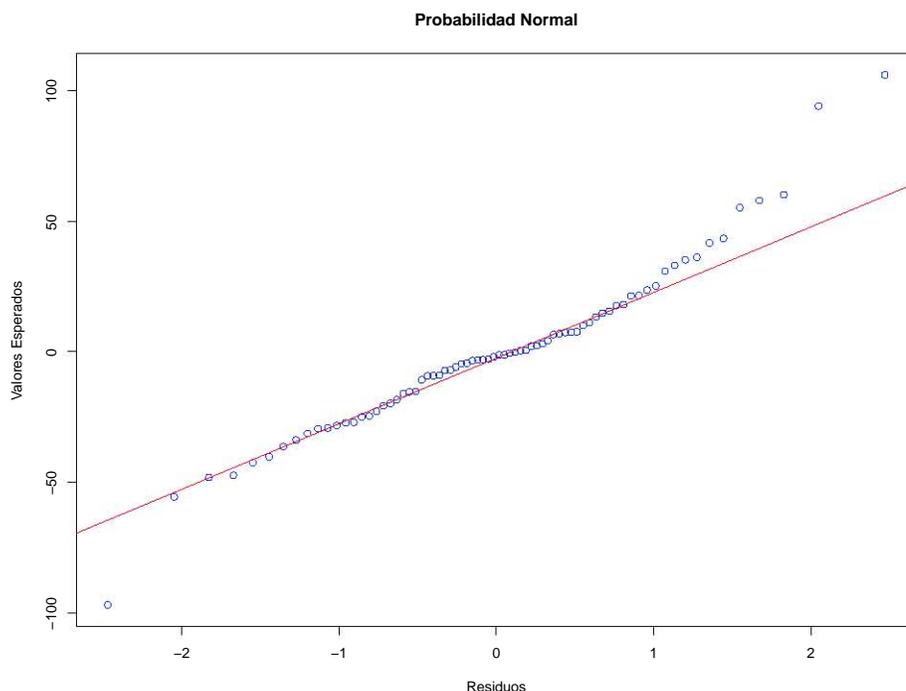


Figura 7.4: Prueba de normalidad.

La función `qqnorm()` y `qqline()` comparan los valores estandarizados de los residuos con lo que serían los valores esperados de una distribución normal estándar. Entre más parecidos sean los valores de los residuos (color azul) a la línea roja, más indicios hay de que éstos siguen una distribución normal. En nuestro caso, en la región del centro los valores se ajustan a la línea, pero en los valores de los extremos se alejan de la línea. Esto sucede debido que los residuos tienen la cola derecha pesada.

Otra manera de verificar la normalidad es mediante un histograma. Utilizaremos la librería MASS y la función `truehist()` y `curve()` para mostrar el histograma de los residuos y agregar la curva de una distribución normal con la media y desviación estándar de éstos. Hacemos lo siguiente:

```
library(MASS)

truehist(residuos, main = "Histograma de los residuos", prob = TRUE)

curve(dnorm(x, mean(residuos), sd(residuos)), add = T, col = "purple")
```

La función `truehist()` grafica el histograma de los residuos. El parámetro `prob = TRUE` hace que los valores del histograma se muestren como probabilidades y no como frecuencias. La función `curve()` muestra la curva de una función de densidad de una normal. En este caso indicamos que la media y la desviación estándar son las de los residuos. El histograma se ve así:

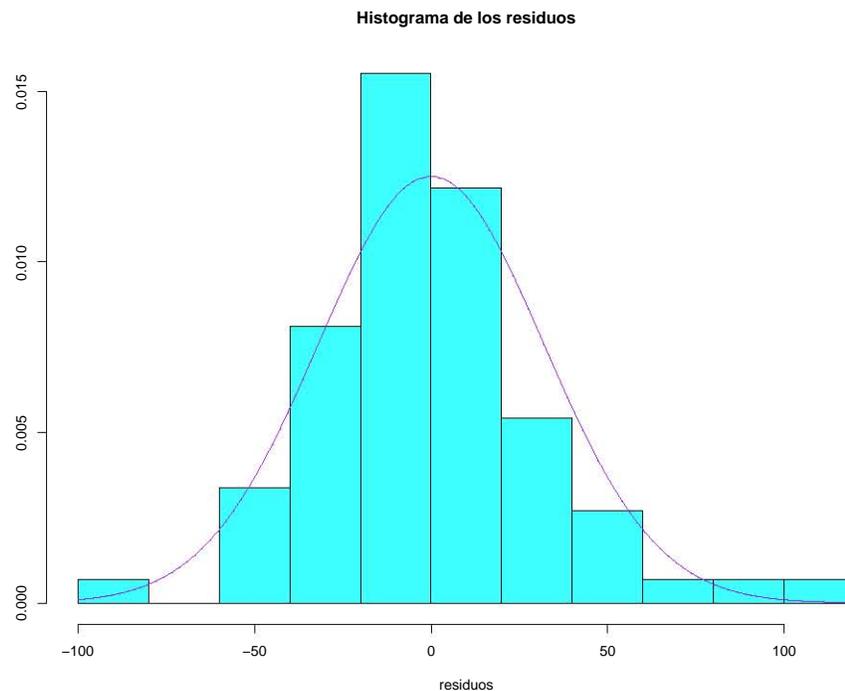


Figura 7.5: Histograma.

Si bien algunas de las barras se salen de la curva, en general las demás tienen la estructura de la curva de la función de densidad normal. Esta prueba apoya más la hipótesis de normalidad que la prueba realizada con la función `qqnorm()`.

Una prueba formal es la prueba de Anderson-Darling, cuya hipótesis nula es que los residuos siguen una distribución normal. Esta prueba puede realizarse con la función `ad.test()` de la librería `nortest`. Hacemos lo siguiente:

```
library(nortest)

ad.test(residuos)

Anderson-Darling normality test

data:  residuos
A = 0.95136, p-value = 0.01533
```

Usando un nivel de significancia de 0.05, el p-value es igual a 0.01533, de modo que es menor que el nivel de significancia, por lo que se rechaza la hipótesis nula. Por lo tanto, los residuos no tienen una distribución normal.

Esta prueba es la más estricta de las pruebas mencionadas. Si bien se puede concluir que la distribución de los residuos no es normal, en la práctica éste es un supuesto que no se cumple

siempre.

El hecho de que los residuos no tienen una distribución normal y que la varianza no es constante indican que este modelo no es adecuado para el análisis de la serie. A pesar de esto, para tener una idea de la precisión de los pronósticos, vamos a calcular algunos de los tipos de errores de exactitud.

Creamos la variable `e_t`, la cual corresponde a $e_t = y_t - \hat{y}_t$:

```
e_t <- serie-fitted(modelo)
```

Error absoluto medio:

```
MAE <- sum(abs(e_t))
```

Error cuadrático medio:

```
MSE <- mean(e_t^2)
```

o bien,

```
MSE <- modelo$mse
```

Error porcentual:

```
p_t <- 100*e_t/serie
```

Error porcentual absoluto medio:

```
MAPE <- mean(abs(p_t))
```

Error porcentual absoluto medio simétrico:

```
sMAPE <- mean(200*abs(e_t)/(serie+fitted(modelo)))
```

Error escalado:

```
q_t <- e_t/(1/(length(serie)-1)*sum(abs(diff(serie, differences = 1))))
```

Error escalado absoluto medio:

```
MASE <- mean(abs(q_t))
```

Creamos un data frame de nombre `errors` para mostrar los valores obtenidos de los errores:

```
errors <- data.frame(c('MAE', 'MSE', 'MAPE', 'sMAPE', 'MASE'),
                    c(MAE, MSE, MAPE, sMAPE, MASE))
```

```
colnames(errors) <- c('Tipo de error', 'Valor')
```

```

errors
  Tipo de error      Valor
1          MAE      22.5695301
2          MSE    1004.2748129
3          MAPE      6.0957125
4          sMAPE     6.0494185
5          MASE      0.9520477

```

El error absoluto medio nos dice que en promedio el valor absoluto de la diferencia entre el valor real y el pronóstico del modelo es de 22.57. Observamos que los errores de porcentaje indican que en promedio los pronósticos difieren en un 6% de los valores de la serie.

Vamos ahora a hacer el mismo análisis, pero utilizando el modelo que la función `ets()` selecciona de manera automática. Hacemos lo siguiente:

```

modelo.auto <- ets(serie)

modelo.auto
ETS(M,A,N)

Call:
ets(y = serie)

Smoothing parameters:
  alpha = 0.9281
  beta  = 1e-04

Initial states:
  l = 161.506
  b = 5.6969

sigma: 0.0762

      AIC      AICc      BIC
802.9436 803.8260 814.4639

```

El modelo seleccionado es el M,A,N. Notemos que el valor del criterio de Akaike es $AIC = 802.9436$, menor que 839.9904, que es el valor del criterio de Akaike para el modelo A,A,N. Recordemos que el modelo seleccionado por la función `ets()` es el que tiene el valor óptimo para el criterio de información.

En cuanto a los parámetros, tenemos $\alpha = 0.9281$, $\beta = 1 \times 10^{-4}$, siendo el valor de α diferente del obtenido para el modelo A,A,N ($\alpha = 0.9999$). El valor de β coincide en ambos modelos. Por otra parte, los parámetros iniciales son $\ell_0 = 161.506$, $b_0 = 5.6969$, los cuales no difieren mucho de los obtenidos para el modelo A,A,N ($\ell_0 = 158.4272$ y $b_0 = 8.0153$). El valor de $\hat{\sigma}$ es 0.0762, considerablemente distinto del valor obtenido para el otro modelo (31.6903).

Vamos ahora a calcular los pronósticos y los intervalos de predicción. Almacenamos los resultados en la variable `prediction.auto`:

```
prediction.auto <- forecast(modelo.auto, h = 4, level = 95)
```

```
prediction.auto
      Point Forecast      Lo 95      Hi 95
Dec 2016      759.7728 646.3566 873.1890
Jan 2017      765.4882 609.9070 921.0693
Feb 2017      771.2035 582.0590 960.3481
Mar 2017      776.9189 558.7992 995.0387
```

Los valores reales fueron:

```
tail(serie_ref, n = 4)
      Jan      Feb      Mar Apr May Jun Jul Aug Sep Oct Nov Dec
2016                                     749.87
2017 823.48 845.04 886.54
```

Podemos notar que los pronósticos puntuales nuevamente fueron distintos; sin embargo, como en el otro modelo, todos los valores reales están dentro de los intervalos de predicción.

Veamos una comparación de los resultados de cada modelo:

Modelo A,A,N

```
prediction
      Point Forecast      Lo 95      Hi 95
Dec 2016      758.5901 696.4782 820.7019
Jan 2017      766.6054 678.7660 854.4448
Feb 2017      774.6207 667.0363 882.2051
Mar 2017      782.6360 658.4030 906.8689
```

Modelo M,A,N

```
prediction.auto
      Point Forecast      Lo 95      Hi 95
Dec 2016      759.7728 646.3566 873.1890
Jan 2017      765.4882 609.9070 921.0693
Feb 2017      771.2035 582.0590 960.3481
Mar 2017      776.9189 558.7992 995.0387
```

Los valores de los pronósticos son muy parecidos. Recordemos que cuando los modelos tienen los mismos parámetros, los pronósticos puntuales son los mismos sin importar si el error es aditivo o multiplicativo, pero ese no es el caso aquí, puesto que difieren en el parámetro α . También observamos que los intervalos de predicción para el modelo M,A,N son más grandes.

Veamos en un gráfico cómo difieren los pronósticos puntuales del modelo M,A,N con los valores reales:

```
plot(prediction.auto, main = 'Pronósticos de ETS(M,A,N) vs datos reales',
```

```

xlab = 'Tiempo', ylab = 'Precio', ylim = c(0, 1000), shaded = FALSE,
flty = 0, pi.lty = 0)

lines(fitted(modelo.auto), col = 'red')

lines(prediction.auto$mean, col = 'blue', type = 'o', pch = 20, cex = 1)
lines(prediction.auto$lower, col = 'blue', type = 'o', pch = 20, cex = 1)
lines(prediction.auto$upper, col = 'blue', type = 'o', pch = 20, cex = 1)
lines(tail(serie_ref, 4), col = 'black', type = 'o', pch = 20, cex = 1)

legend(2011, 1000, c('M,A,N', 'Valores ajustados', 'Serie original'),
      lty = c(1, 1, 1), col = c('blue', 'red', 'black'))

```

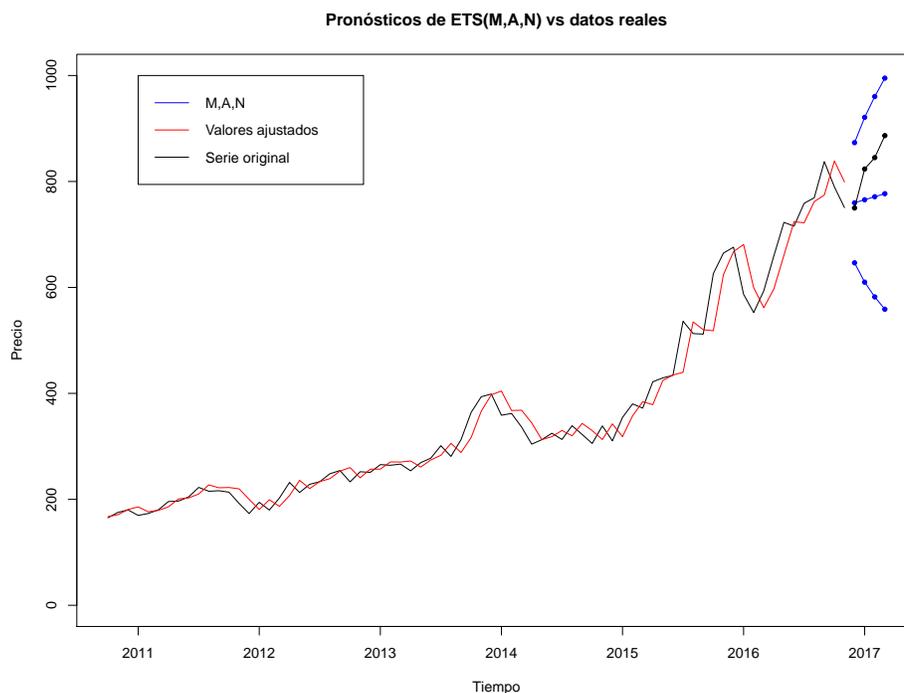


Figura 7.6: Comparación de pronósticos con datos reales.

Observamos que la región de los intervalos de predicción es más amplia que la del modelo anterior.

Veamos ahora los resultados de los dos modelos juntos:

```

plot(prediction.auto, main = 'Pronósticos de ETS(M,A,N) y
  ETS(A,A,N) vs datos reales',
      xlab = 'Tiempo', ylab = 'Precio', ylim = c(0, 1000), shaded = FALSE,

```

```

    flty = 0, pi.lty = 0)
lines(fitted(modelo), col = 'red')
lines(prediction$mean, col = 'red', type = 'o', pch = 20, cex = 1)
lines(prediction$lower, col = 'red', type = 'o', pch = 20, cex = 1)
lines(prediction$upper, col = 'red', type = 'o', pch = 20, cex = 1)

lines(fitted(modelo.auto), col = 'blue')
lines(prediction.auto$mean, col = 'blue', type = 'o', pch = 20, cex = 1)
lines(prediction.auto$lower, col = 'blue', type = 'o', pch = 20, cex = 1)
lines(prediction.auto$upper, col = 'blue', type = 'o', pch = 20, cex = 1)
lines(tail(serie_ref, 4), col = 'black', type = 'o', pch = 20, cex = 1)
legend(2011, 1000, c('M,A,N', 'A,A,N', 'Serie original'),
      lty = c(1, 1, 1), col = c('blue', 'red', 'black'))

```

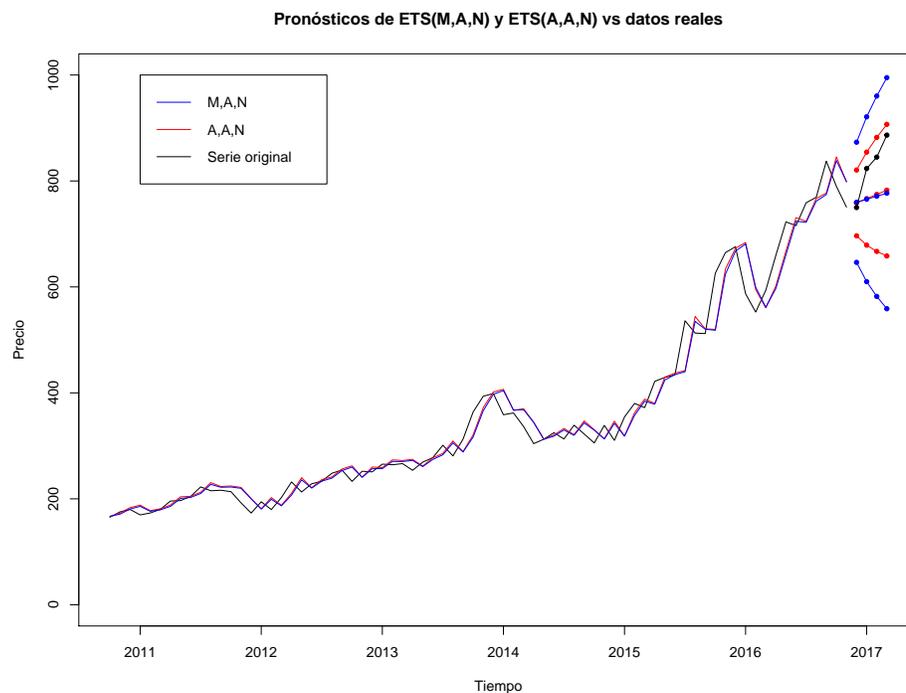


Figura 7.7: Comparación de modelos.

Podemos observar que los valores ajustados de ambos modelos son muy cercanos entre sí. Los valores de los pronósticos puntuales de igual manera son muy parecidos. Por otra parte, los intervalos de predicción del modelo M,A,N (color azul) son más grandes que los del modelo A,A,N (color rojo), como ya se había mencionado. En ambos casos, los valores futuros reales de la serie (color negro) están contenidos en los intervalos.

Ahora vamos a verificar los supuestos para el modelo M,A,N. Primeramente extraemos los residuos del modelo y los guardamos en la variable `residuos.auto`:

```
residuos.auto <- modelo.auto$residuals
```

Ahora utilizamos la función `tsdiag()`:

```
tsdiag(modelo.auto)
```

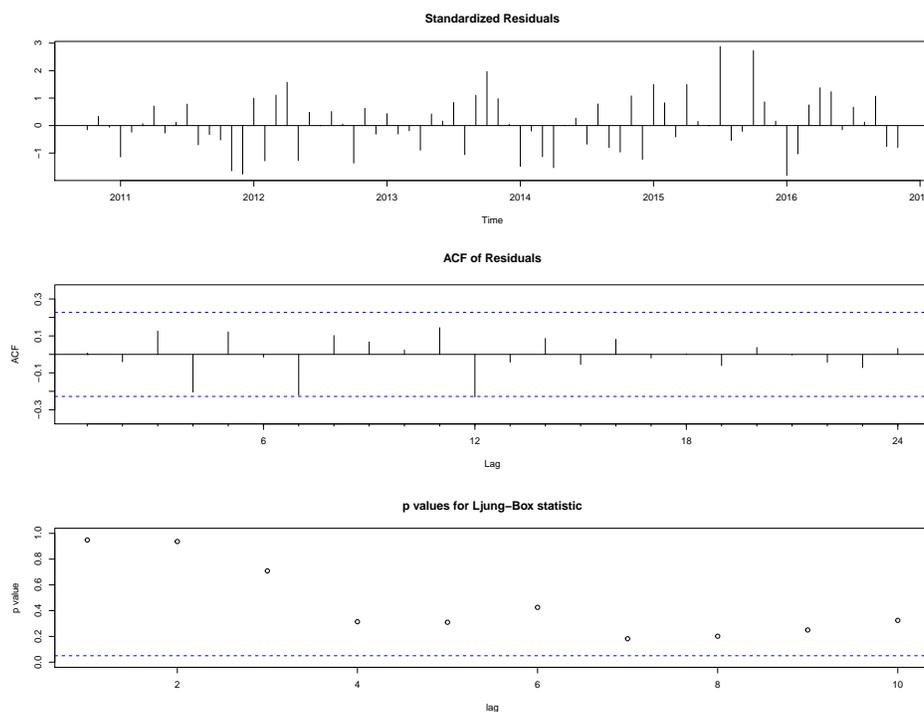


Figura 7.8: Gráfico y función de autocorrelación de los residuos y prueba de Ljung-Box.

Notemos que en el gráfico de los residuos la varianza es constante. En el caso de la función de autocorrelación, los valores de dicha función no se salen de las bandas de confianza. Además, todos los p-values de la prueba de Ljung-Box son mayores al nivel de significancia $\alpha = 0.05$, por lo que podemos concluir que los valores de las autocorrelaciones entre los p-values son cero.

Veamos ahora la prueba de normalidad. Primeramente, utilizamos las funciones `qqnorm()` y `qqline()`:

```
qqnorm(residuos.auto, xlab = "Residuos", ylab = "Valores Esperados",
        main = "Probabilidad Normal", pch = 1, col = "blue")

qqline(residuos.auto,col = 2)
```

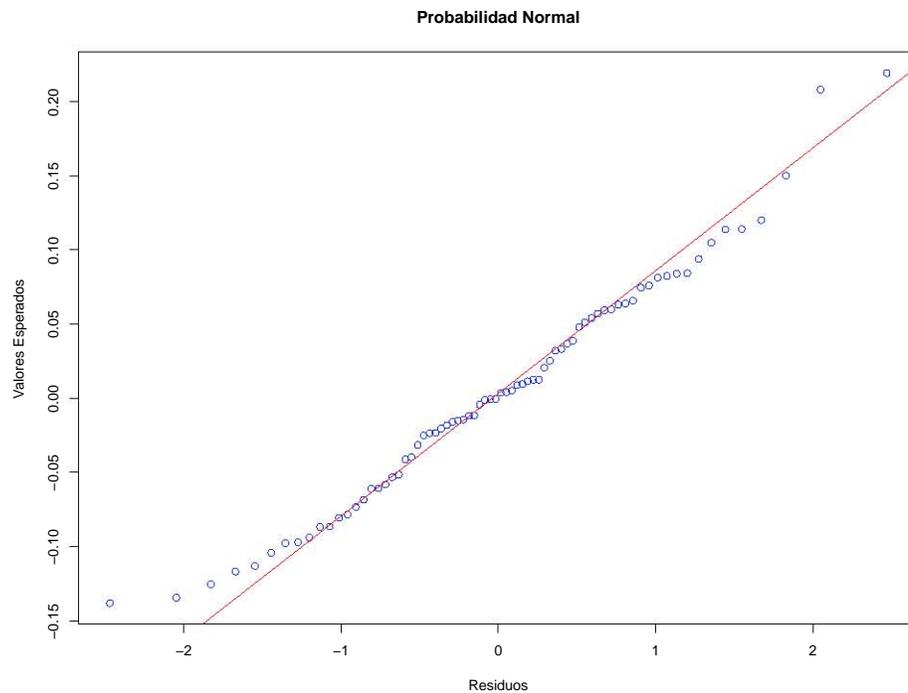


Figura 7.9: Prueba de normalidad.

Si bien algunos puntos de los extremos se alejan de ésta, la mayoría se ajusta a la línea normal.

Ahora veamos el histograma de los residuos:

```
truehist(residuos.auto, main = "Histograma de los residuos", prob = TRUE,
          xlim = c(-0.4, 0.4), ylim = c(0,6))

curve(dnorm(x, mean(residuos.auto), sd(residuos.auto)), add = T,
      col = "purple")
```

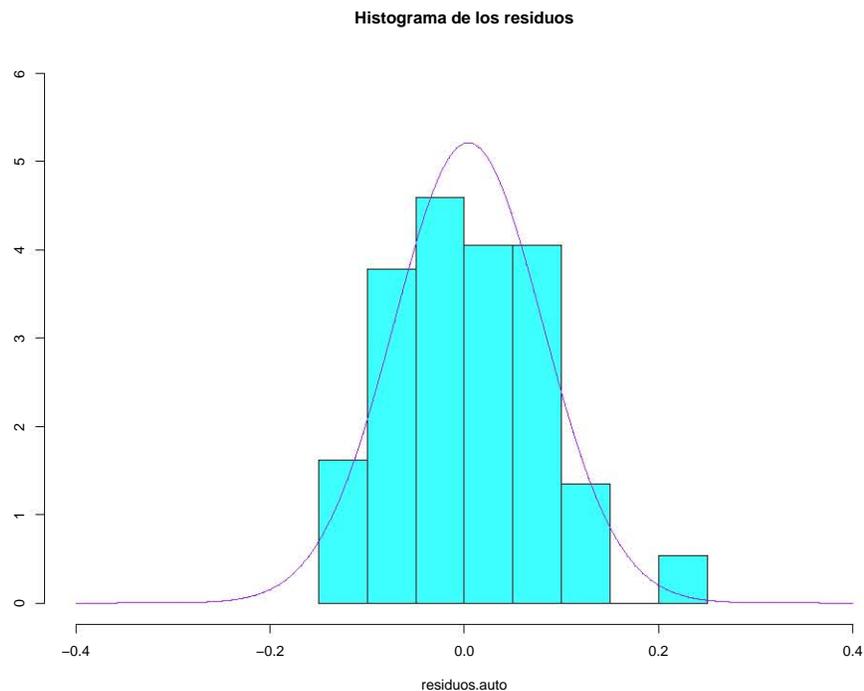


Figura 7.10: Histograma.

El histograma se ajusta aproximadamente a la curva de la función de densidad de una distribución normal.

Finalmente veamos la prueba de Anderson-Darling:

```
ad.test(residuos.auto)
```

```
Anderson-Darling normality test
```

```
data:  residuos.auto
A = 0.24375, p-value = 0.7566
```

El p-value es igual a 0.7566, de modo que es mayor que el nivel de significancia α de 0.05, por lo que no se rechaza la hipótesis nula de que los residuos siguen una distribución normal.

Con esto podemos concluir que el modelo M,A,N es un modelo adecuado. Puesto que cumple más supuestos que el modelo A,A,N, lo consideramos más apropiado para el análisis de nuestra serie. Es necesario mencionar que no siempre el modelo seleccionado al optimizar el criterio de información es el más adecuado. Siempre es necesario verificar que los supuestos se cumplen para aceptar o no un modelo. Cuando dos o más modelos cumplen los supuestos, entonces en ese caso sí elegimos el modelo que tenga el menor valor del criterio de información.

Calculamos ahora los tipos de error:

Definimos nuevamente la variable `e_t`:

```
e_t <- serie - fitted(modelo.auto)
```

Error absoluto medio:

```
MAE <- sum(abs(e_t))
```

Error cuadrático medio:

```
MSE <- mean(e_t^2)
```

o bien,

```
MSE <- modelo.auto$mse
```

Error porcentual:

```
p_t <- 100*e_t/serie
```

Error porcentual absoluto medio:

```
MAPE <- mean(abs(p_t))
```

Error porcentual absoluto medio simétrico:

```
sMAPE <- mean(200*abs(e_t)/(serie+fitted(modelo.auto)))
```

Error escalado:

```
q_t <- e_t/(1/(length(serie)-1)*sum(abs(diff(serie, differences = 1))))
```

Error escalado absoluto medio:

```
MASE <- mean(abs(q_t))
```

Creamos un data frame de nombre `errors.auto` para mostrar los valores obtenidos de los errores:

```
errors.auto <- data.frame(c('MAE', 'MSE', 'MAPE', 'sMAPE', 'MASE'),
                          c(MAE, MSE, MAPE, sMAPE, MASE))
```

```
colnames(errors.auto) <- c('Tipo de error', 'Valor')
```

```
errors.auto
  Tipo de error      Valor
1          MAE    22.5806941
2          MSE 1021.2744008
3          MAPE     5.9967792
4         sMAPE     5.9939417
5          MASE     0.9525186
```

Comparemos los valores con los del modelo anterior:

Modelo A,A,N			Modelo M,A,N		
errors			errors.auto		
	Tipo de error	Valor		Tipo de error	Valor
1	MAE	22.5695301	1	MAE	22.5806941
2	MSE	1004.2748129	2	MSE	1021.2744008
3	MAPE	6.0957125	3	MAPE	5.9967792
4	sMAPE	6.0494185	4	sMAPE	5.9939417
5	MASE	0.9520477	5	MASE	0.9525186

Notemos que el error absoluto es ligeramente mayor, así como el error cuadrático medio. No obstante, los errores de tipo porcentual son menores.

Conclusión

En este trabajo se hizo una revisión de la estructura y propiedades de los modelos de suavizamiento exponencial para el análisis de series de tiempo, auxiliándose además del concepto de modelos de espacio de estados, los cuales permiten la obtención de intervalos de predicción para pronósticos puntuales. Se vio que estos modelos pueden utilizarse tanto en procesos estacionarios como no estacionarios sin tener que recurrir a algún tipo de transformación y además pueden estimar componentes de tendencia y estacionalidad.

No obstante, así como otros modelos, los modelos de espacio de estados se realizan bajo ciertos supuestos, de modo que si éstos no se cumplen, los modelos no serán adecuados para el análisis de datos, aunque, como se vio, los supuestos requeridos son similares a los supuestos realizados por otros modelos como los de la metodología de Box-Jenkins (por ejemplo, modelos ARIMA).

Puesto que hay muchas combinaciones de estos modelos, se tiene una gran cantidad de opciones para elegir el modelo que más se ajuste a las características de los datos en cuestión. Sin embargo, en ciertas áreas hay modelos de otras metodologías que en la práctica se han mostrado bastante adecuados, por lo que los modelos de espacio de estados tal vez no sean los más convenientes en ciertos casos. Un ejemplo de esto es el modelo GARCH, el cual es un modelo muy utilizado para el análisis de rendimientos de acciones, datos para los cuales posiblemente este modelo sea más recomendable que los modelos de espacio de estados.

Finalmente, cabe mencionar que el estudio de esta metodología puede extenderse a, por ejemplo, casos en los que hay más de un patrón de estacionalidad, análisis de series multivariadas, entre otras cosas.

Bibliografía

El contenido de este trabajo está basado en las siguientes referencias:

- [1] Durbin, J., Koopman, S.J., *Time Series Analysis by State Space Methods*, Second edition, Oxford University Press, 2012.
- [2] Brockwell, P.J., Davis, R.A., *Introduction to Time Series and Forecasting*, Third edition, Springer-Verlag, 2016.
- [3] Hyndman, R., Koehler, A.B., Ord, J.K., Snyder, R.D., *Forecasting with Exponential Smoothing: The State Space Approach*, Springer-Verlag, 2008.
- [4] Hyndman, R., Athanasopoulos, G., *Forecasting: Principles and Practice*, Sección 6.2, <https://www.otexts.org/fpp2/moving-averages.html>, 2012.
- [5] Tsay, R.S., *Analysis of Financial Time Series*, Third edition, Wiley, 2010.
- [6] McClain, J.O., Thomas L. J., *Response-variance tradeoffs in adaptive forecasting*, Operations Research, 21, 554–568, 1973.
- [7] Casella, G., Berger, R.L., *Statistical Inference*, Second edition, Duxbury, 2001.