

Suavizado Exponencial con el enfoque de modelos de espacio de estados: dos aplicaciones.

Blaconá, María Teresa y Andreozzi, Lucía.

Cita:

Blaconá, María Teresa y Andreozzi, Lucía (2012). *Suavizado Exponencial con el enfoque de modelos de espacio de estados: dos aplicaciones*. X Congreso de Sociedades Latinoamericanas de Estadística. Universidad Nacional de Córdoba, Córdoba.

Dirección estable: <https://www.aacademica.org/lucia.andreozzi/35>

ARK: <https://n2t.net/ark:/13683/preH/4BC>



Esta obra está bajo una licencia de Creative Commons.
Para ver una copia de esta licencia, visite
<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.es>.

Acta Académica es un proyecto académico sin fines de lucro enmarcado en la iniciativa de acceso abierto. Acta Académica fue creado para facilitar a investigadores de todo el mundo el compartir su producción académica. Para crear un perfil gratuitamente o acceder a otros trabajos visite: <https://www.aacademica.org>.

SUAVIZADO EXPONENCIAL CON EL ENFOQUE DE MODELOS DE ESPACIO DE ESTADO: DOS APLICACIONES

BLACONÁ, M.T.¹ y ANDREOZZI, L.¹

1 Universidad Nacional de Rosario Facultad de Ciencias Económicas y Estadística - CIUNR

mblacona@fcecon.unr.edu.ar

landreozzi@fcecon.unr.edu.ar

RESUMEN

Debe ser un párrafo de a lo sumo 150 palabras. El mismo debe describir brevemente los principales contenidos del artículo y evitar las citas bibliográficas.

PALABRAS CLAVE: incluir de tres a seis palabras clave.

1. INTRODUCCIÓN

Los suavizados exponenciales surgieron con el trabajo pionero de R. G. Brown en el año 1944 y se desarrollaron durante los años 1950 a través de los trabajos de Holt (1957), Brown (1959) y Holt-Winters (1960). Estos métodos de pronóstico intuitivos y de fácil comprensión, fueron muy útiles y con muy buenos resultados especialmente en el área de negocios y empresas. No obstante sufrieron dos críticas fundamentales: 1) no existe un modelo estadístico que sustente la representación de los datos y 2) no permitían realizar intervalos de pronóstico.

Abraham y Leodolter (1983), pusieron en correspondencia los suavizados exponenciales con distintos modelos ARIMA, por ejemplo el modelo ARIMA(0,1,1) con un suavizado exponencial simple donde el parámetro θ del modelo ARIMA es igual a $(1-\alpha)$, siendo α la constante de suavizado. De esta forma se justifica la presencia de un proceso estocástico que representa los datos y permite realizar intervalos de pronósticos.

En el libro de A. Harvey (1989), se hace una presentación rigurosa de los modelos de espacio de estado, donde mediante modelos de series de tiempo estructurales, se formulan directamente los términos de los componentes de interés. En estos modelos juega un papel importante el filtro de Kalman para la estimación máximo verosímil de sus componentes. Estos componentes, como por ejemplo, el nivel, el crecimiento, la estacionalidad, etc., se

relacionan con los componentes que se suavizan en los suavizados exponenciales. Así mismo, se muestra la equivalencia entre estos modelos y los modelos ARIMA.

Más recientemente en el libro de Hyndman et al (2008), se explican sistemáticamente y desarrollan ideas de artículos de los últimos diez años, sobre los denominados modelos de innovaciones de espacio de estado, los cuales mantienen la naturaleza intuitiva del suavizado exponencial pero con el enfoque de modelos de espacio de estados, con las respectivas ecuaciones de medida y de estados. Si bien los pronósticos puntuales coinciden cuando se calculan por ambos métodos, los segundos proveen, entre otras cosas, intervalos de predicción y estimación máximo verosímil. También se pueden encontrar los modelos ARIMA equivalentes.

En este trabajo se realiza una breve introducción de las innovaciones de espacio de estado, siguiendo las ideas de Hyndman et. al. (2008) para luego mediante dos aplicaciones de series de tiempo reales, mostrar las relaciones entre los distintos tipos de modelos con sus respectivas equivalencias.

En la sección 2 se presenta una breve descripción de los distintos métodos de suavizado, los modelos de innovación de espacio de estado y su relación con los modelos ARIMA, en la sección 3 se muestran las aplicaciones de los métodos en dos series reales y en la sección 4 se realizan las conclusiones.

2. METODOLOGÍA

2.1. Métodos de suavizado exponencial

Se puede pensar a una serie de tiempo como la combinación de varios componentes tales como: tendencia (T), ciclo (C), estacionalidad (S) e irregular o error (E). Los tres componentes T, S y E se pueden pensar combinados en forma diferente, por ejemplo, un modelo puramente aditivo se expresa como:

$$y = T + S + E,$$

y uno puramente multiplicativo como:

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

existen varias otras posibles combinaciones.

Se verán algunos tipos de suavizado exponencial, que combinan estos componentes de diferentes maneras.

El componente tendencia (T) se lo piensa como una combinación del término nivel l y del término de crecimiento b , según se combinen estos dos términos, definirán cinco tipos de

tendencia, sea T_h el término de la tendencia h pasos hacia adelante, según el tipo de tendencia que se define resulta:

N Sin tendencia $T_h = l$,

A Aditiva $T_h = l + bh$,

A_d Aditiva amortiguada $T_h = l + (\phi + \phi^2 + \dots + \phi^h)$,

M Multiplicativa $T_h = lb^h$,

M_d Multiplicativa Amortiguada $T_h = b^{(\phi + \phi^2 + \dots + \phi^h)}$,

donde ϕ es un parámetro de amortiguación ($0 < \phi < 1$). Es útil utilizar un método de tendencia amortiguada cuando se piensa que la tasa de crecimiento histórica desaparecerá más allá de un cierto período de tiempo.

Si se agrega el componente estacional, sin tener en cuenta los errores, se definirán quince tipos de modelos:

Tendencia	Estacionalidad		
	Sin N	Aditiva A	Multiplicativa M
Sin N	N, N	N, A	N, M
Aditiva A	A, N	A, A	A, M
Ad. Amortiguada A_d	A_d, N	A_d, A	A_d, M
Multiplicativa M	M, N	M, A	M, M
M. Amortiguada M_d	M_d, N	M_d, A	M_d, M

Algunos de ellos describen los métodos de suavizados más conocidos, por ejemplo:

N,N : Suavizado exponencial simple,

A,N : Método lineal de Holt,

A_d,N : Método de tendencia amortiguada,

A,A : Método aditivo de Holt-Winters,

A,M : Método multiplicativo de Holt-Winters,

existen otros métodos pero no son usados tan comunmente.

Para cada uno de los quince métodos de suavizado, existen dos modelos posibles de espacio de estados, uno correspondiente al modelo con error aditivo y otro con error multiplicativo. Si se usan los parámetros equivalentes entre los modelos de espacio de estados (MEE) y su

correspondiente suavizado exponencial, los pronósticos puntuales coinciden, aunque son diferentes los intervalos de pronóstico.

Como dicen Hyndman et. al. (2008): hay que distinguir el método de suavizado exponencial del los MEE, los métodos de suavizado son algoritmos que sólo producen pronósticos puntuales, los subyacentes MEE, son estocásticos y si bien dan los mismos pronósticos puntuales, también proveen desarrollos para calcular intervalos de predicción, además de tener otras propiedades.

2.2. Pronósticos puntuales por métodos de suavizado.

Se verán algunos de los métodos de suavizado más utilizados para realizar pronósticos.

Sea la serie de tiempo observada: (y_1, y_2, \dots, y_n) . Un pronóstico de y_{t+h} basado en todos los datos hasta el tiempo t se denota como $\hat{y}_{(t+h|t)}$.

Suavizado exponencial simple

Se tienen observaciones hasta el momento $(t - 1)$ y se pronostica y_t con \hat{y}_t cuando se dispone de la observación y_t , el error de pronóstico es $(y_t - \hat{y}_t)$. El método de Brown (1959), toma el pronóstico de los períodos previos y lo ajusta usando el error de pronóstico.

$$\hat{y}_{t+1} = \hat{y}_t + \alpha(y_t - \hat{y}_t), \quad 0 < \alpha < 1 \quad (2.1)$$

si α es cercano a 1, el nuevo pronóstico incluirá un ajuste sustancial del error de los pronósticos previos. Si α es cercano a cero, el nuevo pronóstico incluirá muy poco ajuste.

Otra forma de escribir (2.1) es

$$\hat{y}_{t+1} = \alpha y_t + (1 - \alpha)(y_t - \hat{y}_t), \quad (2.2)$$

se puede interpretar como un promedio ponderado de los pronósticos más cercanos y las observaciones más recientes. Si se expande (2.2) se tiene

$$\hat{y}_{t+1} = \alpha y_t + (1 - \alpha)[\alpha y_{t-1} + (1 - \alpha)\hat{y}_{t-1}],$$

$$\hat{y}_{t+1} = \alpha y_t + \alpha(1 - \alpha)\hat{y}_{t-1} + \alpha(1 - \alpha)^2\hat{y}_{t-1},$$

repetiendo el proceso de sustitución se encuentra

$$\hat{y}_{t+1} = \alpha y_t + \alpha(1 - \alpha)y_{t-1} + \alpha(1 - \alpha)^2\hat{y}_{t-1} + \dots + \alpha(1 - \alpha)^{t-1}y_1 + \alpha(1 - \alpha)^t y_1,$$

\hat{y}_{t+1} representa el promedio móvil ponderado de todas las observaciones pasadas con los pesos decreciendo exponencialmente. Es importante la elección de los valores iniciales, lo cual se conoce como “problema de iniciación”.

Los pronósticos h pasos hacia adelante están dados por

$$\hat{y}_{(t+h|t)} = \hat{y}_{t+1}, \quad h = 1, 2, \dots,$$

esto resultado de valores constantes se debe a que se se tiene una serie sin tendencia ni estacionalidad.

Método lineal de Holt (A,N), (Holt, 1957)

Permite pronosticar datos con tendencia. Se necesitan dos constantes de suavizado α y β^* (con valores entre 0 y 1)

$$\text{Nivel: } l_t = \alpha y_t + (1 - \alpha)(l_{t-1} + b_{t-1}), \quad (2.4.a)$$

$$\text{Crecimiento: } b_t = \beta^*(l_{t-1} + b_{t-1}) + (1 - \beta^*)b_{t-1}, \quad (2.4.b)$$

$$\text{Pronóstico: } \hat{y}_{(t+h|t)} = l_t + b_t h, \quad (2.4.c)$$

En el caso especial que $\alpha = \beta^*$, el método de Holt es equivalente al doble suavizado exponencial de Brown (Brown 1959).

Un caso especial interesante es cuando $\beta^* = 0$, ya que se tiene

$$\text{Nivel: } l_t = \alpha y_t + (1 - \alpha)(l_{t-1} + b), \quad (2.5.a)$$

$$\text{Pronóstico: } \hat{y}_{(t+h|t)} = l_t + b h, \quad (2.5.b)$$

este método se conoce como SES con dirección, similar al método Theta de de pronósticos (Assimakopoulos y Nikolopoulos, 2000).

Método de Holt Winters con tendencia y estacionalidad, (Holt, 1957 y Winters, 1960)

El método de Holt-Winters se basa en una ecuación de suavizado para el nivel, otra para la tendencia y otra para la estacionalidad

Estacionalidad Multiplicativa (A,M)

$$\text{Nivel: } l_t = \alpha \frac{y_t}{s_{t-m}} + (1 - \alpha)(l_{t-1} + b_{t-1}), \quad (2.6.a)$$

$$\text{Crecimiento: } b_t = \beta^*(l_{t-1} + b_{t-1}) + (1 - \beta^*)b_{t-1}, \quad (2.6.b)$$

$$\text{Estacionalidad: } s_t = \gamma \frac{y_t}{s_{t-m}} + (1 - \gamma)s_{t-m}, \quad (2.6.c)$$

$$\text{Pronóstico: } \hat{y}_{(t+h|t)} = (l_t + b_t h) s_{t-m+h_m^+}, \quad (2.6.d)$$

m : longitud de la estacionalidad

$$h_m^+ = [(h - 1) \bmod m] + 1.$$

Generalmente α, β^* y γ varían entre cero y uno. Para hacer más simple la formulación del MEE, Ord et. Al. (2007) reemplazan (2.6.c) por $s_t = \gamma \frac{y_t}{l_t} + (1 - \gamma)s_{t-m}$, mostrando que prácticamente difieren levemente en los pronósticos.

Estacionalidad Aditiva (A,A)

$$\text{Nivel: } l_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)(l_{t-1} + b_{t-1}), \quad (2.7.a)$$

$$\text{Crecimiento: } b_t = \beta^*(l_{t-1} + b_{t-1}) + (1 - \beta^*)b_{t-1}, \quad (2.7.b)$$

$$\text{Estacionalidad: } s_t = \gamma(y_t - l_{t-1} - b_{t-1}) + (1 - \gamma)s_{t-m}, \quad (2.7.c)$$

$$\text{Pronóstico: } \hat{y}_{(t+h|t)} = l_t + b_t h + s_{t-m+h_m^*}, \quad (2.7.d)$$

como en el modelo multiplicativo la ecuación (2.7.c) en los libros de texto es un poco diferente, usando $s_t = \gamma^*(y_t - l_t) + (1 - \gamma^*)s_{t-m}$.

Si l_t se sustituye usando (2.7.a) se obtiene

$$s_t = \gamma^*(1 - \alpha)(y_t - l_{t-1} - b_{t-1}) + [1 - \gamma^*(1 - \alpha)]s_{t-m},$$

En este caso se obtienen los mismos pronósticos que con las fórmulas anteriores.

Los pronósticos puntuales h pasos hacia adelante de todos los métodos de suavizado exponencial, calculados mediante la fórmula recursiva se encuentran en el libro de Hyndman et. Al. (2008), Tabla 2.1 página 18.

2.3 Modelos de espacio de estado

Como se mencionó anteriormente para cada método de suavizado exponencial existen dos *MEE*, uno con errores aditivos y otro con errores multiplicativos.

Para distinguir los modelos con errores aditivos de los multiplicativos, se agrega una letra adelante de la notación del método (*ETS*), que se refiere a los tres componentes: error, tendencia y estacionalidad respectivamente. Por ejemplo, *ETS(A,A,N)*, significa error aditivo, tendencia aditiva, sin estacionalidad, correspondiente al método de Holt con error aditivo. La notación *ETS(.,.,.)* ayuda a recordar el orden en que son especificados los componentes.

Una vez que se especifica el modelo, se estudia la distribución de probabilidad de los valores futuros de la serie y se encuentra la media condicional de las observaciones futuras, dado el conocimiento del pasado, la cual está dada por $\mu_{t+h} = E[y_{t+h}/\mathbf{x}_t]$, donde \mathbf{x}_t contiene todos los componentes no observados tales como l_t , b_t y s_t , donde h representa el horizonte de pronóstico, cuando $h = 1$, $\mu_{t+1} \equiv \mu_{(t+1|t)}$.

Para muchos modelos, la media condicional será idéntica al pronóstico dado anteriormente, de modo que $\mu_{t+h} = \hat{y}_{t+h}$. Sin embargo, existen modelos (los que presentan tendencia y/o estacionalidad multiplicativa), en los que la media condicional puede diferir levemente del pronóstico puntual para $h \geq 2$.

2.3.1 El *MEE* del método lineal de Holt

Modelo con error aditivo: ETS(A, A, N)

Sea $\mu_t = \hat{y}_t = l_t + b_t$ el pronóstico un paso hacia delante suponiendo conocido los valores de todos los parámetros.

Sea $\varepsilon_t = y_t - \mu_t$ error de pronóstico un paso hacia adelante en el tiempo t, a partir de (2.4.c) se encuentra

$$y_t = l_{t-1} + b_{t-1} + \varepsilon_t, \quad (2.8)$$

de (2.4.a) y de (2.4.b) se puede escribir

$$l_t = l_{t-1} + b_{t-1} + \alpha \varepsilon_t, \quad (2.9)$$

$$b_t = b_{t-1} + \beta^*(l_t - l_{t-1} - b_{t-1}) + \alpha \varepsilon_t = b_{t-1} + \alpha \beta^* \varepsilon_t, \quad (2.10)$$

se simplifica $\beta = \alpha \beta^*$.

Las tres ecuaciones anteriores constituyen un MEE subyacente del método de Holt.

Definiendo el vector de estado $x_t = (l_t, b_t)'$ se pueden escribir las ecuaciones (2.8) a (2.10) como

$$y_t = [1 \quad 1]x_{t-1} + \varepsilon_t, \quad (2.11.a)$$

$$x_t = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} x_{t-1} + \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \varepsilon_t. \quad (2.11.b)$$

El modelo queda completamente especificado y generalmente que se establece la distribución de ε_t se supone $\varepsilon_t \sim N(0, \sigma^2)$.

Modelo con error multiplicativo: ETS(M, A, N)

Los errores se definen como $\varepsilon_t = \frac{y_t - \mu_t}{\mu_t}$ de modo que ε_t es un error relativo y se define que

$$y_t = (l_{t-1} + b_{t-1})(1 + \varepsilon_t),$$

$$l_t = (l_{t-1} + b_{t-1})(1 + \alpha \varepsilon_t),$$

$$b_t = b_{t-1} + \beta^*(l_t - l_{t-1} - b_{t-1}),$$

$$y_t = [1 \quad 1]x_{t-1}(1 + \varepsilon_t),$$

donde

$$x_t = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} x_{t-1} + [1 \quad 1]x_{t-1} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \varepsilon_t,$$

con $\varepsilon_t \sim N(0, \sigma^2)$.

Este es un MEE no lineal y generalmente es dificultoso estimar los parámetros y pronosticar, pero se pueden calcular el pronóstico en forma de innovación, como así también la verosimilitud e intervalos de pronóstico con un poco más de esfuerzo que para el modelo aditivo.

Modelo general de los métodos de suavizado exponencial

Se pueden encontrar los MEE de los treinta suavizados exponenciales. El modelo general involucra un vector de estado $\mathbf{x}_t = (l_t, b_t, s_t, s_{t-1}, \dots, s_{t-m+1})$ y una ecuación de forma

$$y_t = \mathbf{w}(\mathbf{x}_{t-1}) + \mathbf{r}(\mathbf{x}_{t-1}) \varepsilon_t, \quad (2.12.a)$$

$$\mathbf{x}_t = \mathbf{f}(\mathbf{x}_{t-1}) + \mathbf{g}(\mathbf{x}_{t-1})\varepsilon_t, \quad (2.12.b)$$

$$\varepsilon_t \sim N(\mu_t = \mathbf{w}(\mathbf{x}_{t-1}), \sigma^2)$$

El modelo con error aditivo tiene $r(\mathbf{x}_{t-1}) = 1$, entonces $y_t = \mu_t + \varepsilon_t$, mientras que en el modelo con error multiplicativo $r(\mathbf{x}_{t-1}) = \mu_t$, entonces $y_t = \mu_t(1 + \varepsilon_t)$. Donde el error relativo del modelo multiplicativo es $\varepsilon_t = \frac{y_t - \mu_t}{\mu_t}$.

Los modelos no son únicos. Cualquier valor de $r(\mathbf{x}_{t-1})$ conducirá al mismo pronóstico puntual de y_t .

En el libro de Hyndman et. al. (2008) se presenta en la tabla 2.2 pagina 21 los modelos aditivos y en la tabla 2.3 pág 22, los multiplicativos.

Los modelos multiplicativos se deberán usar con precaución porque pueden presentar dificultades numéricas debido a que alguno de los componentes de estado podrían involucrar una división por cero.

Los pronósticos puntuales de los modelos anteriores se obtienen iterando (2.12) para $t = n + 1, n + 2, \dots, n + h$ y haciendo $E(\varepsilon_{n+j} | \mathbf{x}_n) = 0$ para $j = 1, 2, \dots, h$. Los pronósticos puntuales están dados por $\mu_{(n+h|n)} = E[y_{n+h} | \mathbf{x}_n]$, la esperanza condicional del modelo correspondiente excepto los casos de tendencia y estacionalidad multiplicativa para $h \geq 2$.

Para construir los intervalos de pronóstico de un modelo lineal, donde las predicciones se distribuyen como Gausseanas, se pueden deducir de la variancias condicionales $v_{(n+h|n)} = V(y_{n+h} | \mathbf{x}_n)$. Este enfoque también funciona para muchos modelos no lineales.

2.4 Iniciación y estimación de los modelos de espacio de estados

Para utilizar los MEE para pronosticar, se necesita especificar el tipo de modelo que se va a usar (selección del modelo), el valor inicial de \mathbf{x}_0 (valor inicial o semilla) y la estimación de los parámetros α, β, γ y ϕ .

Para la iniciación, generalmente el valor de \mathbf{x}_0 se especifica usando valores ad-hoc o vía un esquema heurístico, por ejemplo el dado por Hyndman et. al. (2002).

Para la estimación se define la verosimilitud de la innovación del MEE (2.12), a partir de la cual se obtienen los estimadores máximo verosímiles. Esta función es igual a dos veces el

logaritmo negativo de la función de verosimilitud (con los términos constantes eliminados), condicional a los parámetros $\theta=(\alpha, \beta, \gamma, \phi)'$ y los valores iniciales del estado $\mathbf{x}_0 = (l_0, b_0, s_0, s_1, \dots, s_{m+1})$, donde n es el número de observaciones. Esto se calcula utilizando ecuaciones recursivas. No es necesario usar el filtro de Kalman, para calcular la verosimilitud, como en los clásicos *MEE*, por no tener múltiples fuentes de error.

2.5 Innovaciones lineales de los *MEE*

En un *MEE*, la variable de la serie de tiempo observada y_t se suplementa con una variable auxiliar no observada. Esta variable auxiliar se representa por un vector \mathbf{x}_t que se llama vector de estado. Esta es una forma parsimoniosa de resumir el comportamiento pasado de la serie de tiempo y_t , se utiliza para determinar el efecto del pasado sobre el presente y el comportamiento futuro de la serie de tiempo.

La innovación lineal del *MEE* general es:

$$\mathbf{y}_t = \mathbf{w}'\mathbf{x}_{t-1} + \varepsilon_t, \quad (2.13.a)$$

$$\mathbf{x}_t = \mathbf{F}\mathbf{x}_{t-1} + \mathbf{g}\varepsilon_t, \quad (2.13.b)$$

\mathbf{y}_t : vector observado en el tiempo t ,

\mathbf{x}_t : vector de estado,

este modelo aditivo es un caso especial del modelo más general (2.12). En el suavizado exponencial el vector de estado contiene información nueva del nivel, crecimiento y patrón estacional. Por ejemplo, en un modelo con tendencia y estacionalidad $\mathbf{x}_t = (l_t, b_t, s_t, s_{t-1}, \dots, s_{t-m})$. Desde una perspectiva matemática, la variable de estado es esencialmente redundante. Se mostrará que la variable de estado contiene un vector de estado, que se puede sustituir en la ecuación para dar una forma reducida del modelo.

¿Entonces porqué se utiliza la variable de estado?, ella permite definir en lugar de un gran modelo complejo, otro fraccionado en partes más pequeñas más manejables, reduciendo los cambios de los errores de la especificación del modelo. Los componentes del vector de estado permiten tener una mejor comprensión de la estructura de la serie.

Esta estructura permite explorar la necesidad de cada componente separadamente y llevar a cabo una investigación sistemática del mejor modelo la ecuación de transición. Esta es una relación de recurrencia de primer orden que describe como el vector de estado se desenvuelve a través del tiempo. \mathbf{F} es la matriz de transición. El término $\mathbf{F}\mathbf{x}_{t-1}$ representa el efecto del pasado sobre el estado actual \mathbf{x}_t . El término $\mathbf{g}\varepsilon_t$ muestra lo impredecible de los cambios en \mathbf{x}_t . El vector \mathbf{g} determina la extensión de los efectos de las innovaciones sobre el estado, se lo

llama vector de persistencia . La ecuación de transición es el mecanismo para crear la dependencia a través del tiempo entre las observaciones de una serie de tiempo.

Los k vectores \mathbf{w} y \mathbf{g} son fijos, y \mathbf{F} es una matriz $k \times k$ fija. Estos componentes fijos generalmente contienen parámetros que se necesitan estimar.

El valor de la semilla \mathbf{x}_0 para la ecuación de transición puede ser fijo o aleatorio. El proceso que genera la serie de tiempo puede comenzar antes del período 1, pero los períodos anteriores no están disponibles. En esta situación, el momento del comienzo del proceso se toma como $-\infty$ y \mathbf{x}_0 puede ser aleatorio. Se dice que se aplica el supuesto de comienzo infinito. Alternativamente, el proceso que genera una serie de tiempo puede haber comenzado en el período 1 y \mathbf{x}_0 ser fijo, en este caso se aplica el supuesto de comienzo finito.

Al modelo (2.13) a menudo se lo llama el *MEE* de innovaciones Gausseanas, porque se define en términos de innovaciones que siguen la distribución Gausseana. Se puede tener también la alternativa de *MEE* diferentes y con fuentes de aleatoriedad en (2.13.a) y (2.13.b), no correlacionadas en lugar de una única fuente de aleatoriedad (las innovaciones en cada caso).

La función de densidad de probabilidad para $\mathbf{y} = [y_1, \dots, y_n]$ es una función de las innovaciones y tiene la forma relativamente simple

$$p(\mathbf{y}|\mathbf{x}_0) = \prod_{t=1}^n p(y_t|y_1, \dots, y_n, \mathbf{x}_0) = \prod_{t=1}^n p(y_t|\mathbf{x}_{t-1}) = \prod_{t=1}^n p(\varepsilon_t),$$

Si se supone que la distribución es Gausseana, esta expresión se vuelve

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

una expresión que depende de las innovaciones $\{\varepsilon_t\}$ las cuales se calculan fácilmente. Si el valor de \mathbf{x}_0 se conoce, las innovaciones ε_t son el error de predicción una etapa hacia adelante, aplicando (2.13.a) y (2.13.b) se obtiene

$$E(y_t|y_1, \dots, y_n, \mathbf{x}_0) = E(y_t|\mathbf{x}_{t-1}) = E(\mathbf{w}'\mathbf{x}_{t-1}|\mathbf{x}_{t-1}) + E(\varepsilon_t|\mathbf{x}_{t-1}) = \mathbf{w}'\mathbf{x}_{t-1}.$$

Las predicciones de y_t , dados el valor inicial de \mathbf{x}_0 y las observaciones, se pueden calcular recursivamente de los valores de la serie usando las relaciones:

$$\hat{y}_{(t|t+1)} = \mathbf{w}'\mathbf{x}_{t-1}, \tag{2.14.a}$$

$$\varepsilon_t = y_t - \hat{y}_{(t|t+1)}, \tag{2.14.b}$$

$$\mathbf{x}_t = \mathbf{F}\mathbf{x}_{t-1} + \mathbf{g}\varepsilon_t, \tag{2.14.c}$$

Esta transformación se llamará suavizado exponencial general. Fue dada por primera vez por Box y Jenkins (Box et. al. 1994).

Los pronósticos que se obtienen con estas transformaciones son funciones lineales de observaciones pasadas. Se sustituye (2.14.a) y (2.14.b) en (2.14.c) para encontrar

$$\mathbf{x}_t = \mathbf{D}\mathbf{x}_{t-1} + \mathbf{g}y_t,$$

donde $\mathbf{D} = \mathbf{F} - \mathbf{g}\mathbf{w}'$. Resolviendo hacia atrás la relación de recurrencia (2.14) se tiene

$$\mathbf{x}_t = \mathbf{D}^t \mathbf{x}_0 + \sum_{j=0}^{t-1} \mathbf{D}^j \mathbf{g} y_{t-j}, \quad (2.15)$$

Esto indica que el estado actual \mathbf{x}_t es una función lineal de los valores pasados y presentes de la serie de tiempo. Finalmente sustituyendo (2.15) rezagado un período en (2.14.a) se tiene

$$\hat{y}_{(t|t+1)} = a_t + \sum_{j=1}^{t-1} \mathbf{c}_j y_{t-j}, \quad (2.16)$$

donde $a_t = \mathbf{w}\mathbf{D}^{t-1}\mathbf{x}_0$ y $\mathbf{c}_j = \mathbf{w}'\mathbf{D}^{j-1}\mathbf{g}$. El pronóstico es una función lineal de las observaciones pasadas y el vector semilla.

Las ecuaciones (2.12), (2.14) y (2.15) demuestran la belleza del enfoque de las innovaciones. Se puede comenzar a partir del *MEE* (2.13) y generar los pronósticos una etapa adelante usando (2.15). Cuando se dispone de una nueva observación, el vector de estado se adapta usando (2.14) y se dispone inmediatamente de un nuevo pronóstico una etapa adelante. Otros enfoques consiguen la adaptación y la transición del modelo a la función de pronóstico con menos transparencia y considerablemente más esfuerzo.

2.4.1 Casos especiales de innovaciones lineales de los *MEE*

Modelo de nivel local ETS(A, N, N)

La forma más simple de transmitir la historia de un proceso es a través de un único estado l_t , llamado nivel. El *MEE* resultante se define por las ecuaciones

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

$$l_t = l_{t-1} + \alpha \varepsilon_t, \quad (2.16.b)$$

donde $\varepsilon_t \sim N(0, \sigma^2)$. En conformidad con la estructura de espacio de estado con $\mathbf{x}_t = l_t$, $\mathbf{w} = 1$, $\mathbf{F} = 1$ y $\mathbf{g} = \alpha$. Los valores que son generados por este proceso estocástico se mueven aleatoriamente alrededor del nivel (local) como lo describe (2.16.a).

El grado de cambio de sucesivos niveles está gobernado por el tamaño del parámetro de suavizado α . Los casos para $\alpha = 1$ y $\alpha = 0$ son de especial interés.

Caso $\alpha = 0$: En este caso el nivel local no cambia, y se lo denomina nivel global. Valores sucesivos de la serie y_t son independientes e idénticamente distribuidos. Sus momentos no cambian en el tiempo.

Caso $\alpha = 1$: el modelo se convierte en un camino aleatorio $y_t = y_{t-1} + \varepsilon_t$. Sucesivos valores de la serie de tiempo son dependientes.

El caso especial de la transformación (2.14), del modelo (2.16) es

$$\hat{y}_{(t|t-1)} = l_{t-1}, \quad (2.17.a)$$

$$\varepsilon_t = y_t - l_{t-1}, \quad (2.17.b)$$

$$l_t = l_{t-1} + \alpha \varepsilon_t, \quad (2.17.c)$$

Esto corresponde a un suavizado exponencial simple (Brown, 1959), uno de los métodos más ampliamente usados para pronósticos en aplicaciones comerciales. Es un esquema recursivo simple para calcular las innovaciones de los datos brutos. La ecuación (2.17) se reduce a

$$l_t = (1 - \alpha)l_{t-1} + \alpha y_t, \quad (2.18)$$

la predicción una etapa adelante se obtiene de este esquema que es linealmente independiente de los valores anteriores. La ecuación (2.16) indica que

$$\hat{y}_{(t|t+1)} = (1 - \alpha)l_0 + \alpha \sum_{j=0}^{t-1} (1 - \alpha)^j y_{t-j}, \quad (2.19)$$

es una función lineal de los datos y del nivel semilla. Ignorando el primer término (el cuál es despreciable para valores grandes de t y $|1 - \alpha| < 1$) la predicción $\hat{y}_{(t|t+1)}$ es un promedio ponderado exponencialmente de observaciones pasadas. Los coeficientes dependen del factor discontinuo $1 - \alpha$:

- Si $|1 - \alpha| < 1$, los coeficientes se vuelven pequeños cuando j crece. Luego la condición de estabilidad se cumple si y sólo si $0 < \alpha < 2$,

- Los coeficientes son positivos si y sólo si $0 < (1 - \alpha) < 1$ y (2.19) se puede interpretar como un promedio ponderado de los niveles pasados y el valor actual y_t . Así la predicción sólo se puede interpretar como un promedio ponderado si $0 < \alpha < 1$.

Consecuentemente hay dos posibles rangos de α que se pueden proponer: $0 < \alpha < 2$ sobre la base de un argumento de estabilidad y $0 < \alpha < 1$ sobre la base de interpretarlo como promedios ponderados. En la práctica se usa más el rango más estrecho.

Modelo de tendencia local ETS(A, A, N)

El modelo de nivel local se puede ir aumentando por una tasa de crecimiento b_t dando

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

$$l_t = l_{t-1} + b_{t-1} + \alpha \varepsilon_t, \quad (2.20.b)$$

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

Donde hay dos parámetros de suavizado α y β . La tasa de crecimiento o pendiente b_t puede ser positiva, cero o negativa. El modelo (2.20) tiene una estructura de espacio de estados

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

El impacto de los parámetros de suavizado reflejan el impacto de las innovaciones sobre el nivel y la tasa de crecimiento.

Cuando $\beta = 0$ la tasa de crecimiento es constante en el tiempo. Si además $\alpha = 0$ el nivel cambia a una tasa constante en el tiempo, es decir, no hay cambios aleatorios en el nivel o crecimiento a largo plazo. Para otros valores de los parámetros de suavizado, la tasa de crecimiento sigue un camino aleatorio en el tiempo. Cuando los parámetros de suavizado crecen, la tendencia de la serie se hace más suave.

Para este modelo la transformación (2.14) de los valores de la serie, se transforma en innovaciones de la forma

$$\hat{y}_{(t|t+1)} = l_{t-1} + b_{t-1},$$

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

$$l_t = l_{t-1} + b_{t-1} + \alpha\varepsilon_t,$$

$$b_t = b_{t-1} + \beta\varepsilon_t,$$

Este modelo corresponde al suavizado exponencial lineal de Holt (1957).

Un sistema equivalente de ecuaciones es:

$$\hat{y}_{(t|t+1)} = l_{t-1} + b_{t-1}, \tag{2.21.a}$$

$$\varepsilon_t = y_t - \hat{y}_{(t|t+1)}, \tag{2.21.b}$$

$$l_t = \alpha y_t + (1 - \alpha)(l_{t-1} + b_{t-1}), \tag{2.21.c}$$

$$b_t = \beta^*(l_t - l_{t-1}) + (1 - \beta^*)b_{t-1}, \tag{2.21.d}$$

donde $\beta^* = \frac{\beta}{\alpha}$. El término $l_t - l_{t-1}$ se lo interpreta a menudo como el crecimiento real, es una forma distinta del crecimiento predicho b_{t-1} .

Las ecuaciones (2.21.c) y (2.21.d) se pueden interpretar como promedios ponderados si $0 < \alpha < 1$ y $0 < \beta^* < 1$, o equivalentemente si $0 < \alpha < 1$ y $0 < \beta^* < \alpha$. Estas restricciones se utilizan comúnmente en la práctica. También se puede demostrar que el modelo es estable (la matriz de descuento \mathbf{D}^j converge a $\mathbf{0}$ cuando j crece) cuando $\alpha > 0$, $\beta > 0$ y $2\alpha + \beta < 4$. Esto da una región de parámetros mucho más grande que la que se permite generalmente.

Modelo estacional aditivo local ETS(A, A, A)

Para series con patrón estacional, al modelo con tendencia local se le agrega el efecto estacional, llamado s_t . A menudo la estructura del patrón estacional va cambiando levemente a través del tiempo, en respuesta a cambios en gustos y tecnología.

La fórmula usada representa el efecto estacional se permitiría la posibilidad de cambios del patrón estacional. El modelo es:

$$y_t = l_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t, \tag{2.22.a}$$

$$l_t = l_{t-1} + b_{t-1} + \alpha \varepsilon_t, \quad (2.22.b)$$

$$b_t = b_{t-1} + \beta \varepsilon_t, \quad (2.22.c)$$

$$s_t = s_{t-m} + \gamma \varepsilon_t, \quad (2.22.d)$$

Este modelo corresponde al MEE de primer orden donde

$$\mathbf{w}' = [1 \quad 1 \quad 0 \quad \dots \quad 0 \quad 1],$$

$$\mathbf{x}_t' = [l_t \quad b_t \quad s_t \quad s_{t-1} \quad \dots \quad s_{t-m+1}],$$

$$\mathbf{F} = \begin{bmatrix} 1 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 & 0 \end{bmatrix},$$

$$\mathbf{y} \quad \mathbf{g}' = [\alpha \quad \beta \quad \gamma \quad 0 \quad \dots \quad 0].$$

La inspección cuidadosa del modelo (2.22) muestra que los términos nivel y estacionalidad son confundidos. Si se suma una cantidad arbitraria S al elemento estacional y se sustrae la misma cantidad al nivel, se obtienen las siguientes ecuaciones

$$y_t = (l_{t-1} - \delta) + b_{t-1} + (s_{t-m} + \delta) + \varepsilon_t,$$

$$l_t - \delta = l_{t-1} - \delta + b_{t-1} + \alpha \varepsilon_t,$$

$$b_t = b_{t-1} + \beta \varepsilon_t,$$

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

la cual es equivalente a (2.22). Para eliminar el problema es deseable restringir el componente estacional a que cualquier secuencia $\{s_t, s_{t+1}, \dots, s_{t+m-1}\}$ sume cero (o al mneos tenga media cero). Cuando es cierta la condición para el componente estacional se dice que esta normalizado. La normalización del factor estacional involucra una modificación “sutil” del modelo. Se puede imponer la restricción que el factor estacional en el estado inicial \mathbf{x}_0 debe ser cero. Esto significa que el componente estacional comienza siendo normalizado, aunque no es una restricción, entonces se mueve alrededor de cero en el tiempo.

Las transformaciones de los valores de la serie para predecir errores se puede mostrar que son

$$\hat{y}_t = l_{t-1} + b_{t-1} + s_{t-m_t},$$

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

$$l_t = l_{t-1} + b_{t-1} + \alpha \varepsilon_t,$$

$$b_t = b_{t-1} + \beta \varepsilon_t,$$

$$s_t = s_{t-m} + \gamma \varepsilon_t,$$

esto corresponde a la versión aditiva usada comúnmente del suavizado exponencial estacional (Winters, 1960). Una forma equivalente de las ecuaciones de transición es

$$\hat{y}_t = l_{t-1} + b_{t-1} + s_{t-m_t},$$

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

$$l_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)(l_{t-1} + b_{t-1}),$$

$$b_t = (1 - \beta^*)b_{t-1} + \beta^*(l_t - l_{t-1}),$$

$$s_t = \gamma^*(y_t - l_{t-1} - b_{t-1}) + (1 - \alpha)s_{t-m},$$

Donde el valor de la serie es desestacionalizado en la ecuación de tendencia y sin tendencia en la ecuación de estacionalidad, $\beta^* = \beta/\alpha$ y $\gamma^* = \gamma/(1 - \alpha)$. Las ecuaciones (...?) pueden interpretarse como promedios ponderados, en el caso que las restricciones paramétricas sean que α , β^* y γ^* caigan entre 0 y 1, consideraciones sobre la matriz de descuento \mathbf{D} conducen a diferentes regiones de los parámetros.

2.5 Relación entre los *MEE* de innovación y modelos *ARIMA*

El enfoque que se sigue es el de eliminar las variables de estado del modelo de estado, para llegar a un proceso *ARIMA*, que se lo llama la forma reducida del modelo.

Como el modelo de innovaciones opera con un punto de comienzo finito, podemos restringir nuestra atención a aquellos *MEE* de innovación que son estables de manera que el efecto de las condiciones de inicio se pueda ignorar para series suficientemente largas.

Suavizado exponencial simple ETS(A, N, N)

Ya fue definido el modelo *ETS(A, A, N)* el cuál es un modelo de estado de espacio de innovaciones para suavizado exponencial simple. Usando el operador de rezago L , este modelo se puede escribir en forma levemente diferente

$$y_t = l_{t-1} + \varepsilon_t, \tag{2.23.a}$$

$$(1 - L)l_t = \alpha\varepsilon_t, \tag{2.23.b}$$

Para encontrar la forma reducida del *ETS(A, A, N)* al modelo *ARIMA* se aplica el operador de diferencia $(1 - L)$ a ambos lados de la ecuación de medida (2.23.a). El resultado es un modelo *ARIMA(0,1,1)*

$$(1 - L)y_t = (1 - L)l_{t-1} + (1 - L)\varepsilon_t = \alpha\varepsilon_{t-1} + \varepsilon_t - \varepsilon_{t-1} =$$

$$(-1 + \alpha)\varepsilon_{t-1} + \varepsilon_t = (1 - \theta_1 L)\varepsilon_t$$

Donde $\theta_1 = (1 - \alpha)$. Es bien conocido que el modelo *ARIMA(0,1,1)* provee el mismo pronóstico puntual que un suavizado exponencial simple. El modelo *ETS(A, A, N)* y el

modelo $ARIMA(0,1,1)$ producirán también la misma variancia de pronósticos e intervalos de pronóstico. El espacio del parámetro del ETS $\alpha \in (0,2)$ correspondiente exactamente al espacio del parámetro del $ARIMA$ $|\theta_1| < 1$. Sin embargo observamos que el supuesto de inicio finito permite el esquema ETS con $\alpha = 0$, correspondiente a una constante media, el modelo $ARIMA$ no incluye en este caso.

Modelo ETS(A,A,N)

En forma similar, podemos escribir el modelo local $ETS(A,A,N)$, el cual es MEE de innovaciones del método de Holt como

$$y_t = l_{t-1} + b_{t-1} + \varepsilon_t, \quad (2.24.a)$$

$$(1 - L)l_t = b_{t-1} + \alpha\varepsilon_t, \quad (2.24.b)$$

$$(1 - L)b_t = \beta\varepsilon_t, \quad (2.24.c)$$

Aplicando el operador de diferencia $(1 - L)^2$ en la ecuación de medida (2.24.a) y usando las ecuaciones (2.24.b) y (2.24.c) se obtiene el siguiente modelo $ARIMA(0,2,2)$

$$(1 - L)^2 y_t = (1 - \theta_1 L - \theta_2 L^2) \varepsilon_t,$$

donde $\theta_1 = 2 - \alpha - \beta$ y $\theta_2 = \alpha - 1$

Las condiciones de invertibilidad son

$$\theta_1 + \theta_2 < 1$$

$$\theta_2 - \theta_1 < 1$$

$$|\theta_2| < 1$$

que en función de los coeficientes del ETS son

$$\beta > 0$$

$$4 - 2\alpha - \beta < 0$$

$$0 < \alpha < 2$$

Notamos que esta forma reducida implícitamente supone un inicio finito, el modelo (2.24) debe ser suplementado por los valores iniciales l_0 y b_0 .

2.5.1 Forma reducida del modelo de innovación general

Un objetivo es transformar el MEE en un modelo $ARIMA$. Un modelo general $ARIMA$ se puede expresar como

$$\eta(L)y_t = \Phi(L)\delta(L)y_t = \Theta(L)\varepsilon_t$$

donde de $\eta(L)$ y $\theta(L)$ son ambos polinomios del retardo L y se pueden incluir potencias de L relacionadas con el período estacional m . Además, $\delta(L)$ contiene todas las raíces unitarias en el polinomio. Si hacemos $Z_t = \delta(L)y_t$

$$\Phi(L)Z_t = \Theta(L)\varepsilon_t,$$

Es un modelo ARIMA.

El MEE de innovaciones lineal general $y_t = w'(x_{t-1}) + \varepsilon_t$ y $x_t = F(x_{t-1}) + g\varepsilon_t$ se puede reducir a un modelo ARIMA con la ayuda del operador de rezago. Primero la ecuación de transición se puede escribir como

$$(I - FL)x_t = g\varepsilon_t, \quad (2.25)$$

Como puede suceder que $(I - FL)$ no tenga inversa, multiplicamos a ambos lados de (2.25) por su adjunto, $adj(I - FL)$, se tiene

$$det(I - FL)x_t = adj(I - FL)g\varepsilon_t. \quad (2.26)$$

Luego, aplicando el operador $det(I - FL)$ a ambos lados de la ecuación de medida se encuentra $det(I - FL)y_t = w'det(I - FL)x_{t-1} + det(I - FL)\varepsilon_t$.

Usando (2.26) sustituyendo el $det(I - FL)x_t$ se obtiene el siguiente modelo ARIMA

$$det(I - FL)y_t = w'adj(I - FL)g\varepsilon_t + det(I - FL)\varepsilon_t. \quad (2.27)$$

En este modelo ARIMA

$$\eta(L) = det(I - FL)$$

y

$$\theta(L) = w'det(I - FL)x_{t-1} + det(I - FL)\varepsilon_t.$$

Es posible que los polinomios a ambos lados de esta ecuación tengan factores comunes, en este caso deberían ser cancelados, de acuerdo con la dimensión minimal de la representación de los estados. Esto es la forma reducida corresponderá a un modelo ARIMA después de eliminar de algunos factores que son comunes de $\eta(L)$ y $\theta(L)$. Técnicamente la eliminación de la raíces unitarias deberían conducir a la introducción de una constante en el lado derecho de la ecuación. Sin embargo, tales constantes a menudo son cero cuando al menos queda una raíz unitaria en el lado izquierdo. El siguiente ejemplo ilustra el proceso.

Forma reducida del modelo ARIMA de un MEE de innovación ETS(A, A, A)

El MEE de innovación ETS(A, A, A) viene dado por

$$y_t = l_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t,$$

$$l_t = l_{t-1} + b_{t-1} + \alpha\varepsilon_t,$$

$$b_t = b_{t-1} + \beta\varepsilon_t,$$

$$s_t = s_{t-m} + \gamma\varepsilon_t,$$

$$\mathbf{w}' = [1 \quad 1 \quad 0 \quad \dots \quad 0 \quad 1],$$

$$\mathbf{x}_t' = [l_t \quad b_t \quad s_t \quad s_{t-1} \quad \dots \quad s_{t-m+1}],$$

$$\mathbf{F} = \begin{bmatrix} 1 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 & 0 \end{bmatrix},$$

$$\mathbf{y} \mathbf{g}' = [\alpha \quad \beta \quad \gamma \quad 0 \quad \dots \quad 0].$$

$$\eta(L) = \det(\mathbf{I} - \mathbf{FL}) = \begin{bmatrix} 1 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 & 0 \end{bmatrix} - \begin{bmatrix} L & L & 0 & 0 & \dots & 0 & 0 \\ 0 & L & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & L \\ 0 & 0 & L & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & L & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & L & 0 \end{bmatrix} = (\mathbf{1} - L)^2(\mathbf{1} - L^m)$$

$$\Theta(L) = \mathbf{w}' \det(\mathbf{I} - \mathbf{FL}) \mathbf{x}_{t-1} + \det(\mathbf{I} - \mathbf{FL}) \varepsilon_t =$$

$$= [1 \quad 1 \quad 1 \quad \dots \quad 0 \quad 0] \det(\mathbf{I} - \mathbf{FL}) \begin{bmatrix} L\alpha \\ L\beta \\ L\gamma \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + (\mathbf{1} - L)^2(\mathbf{1} - L^m)$$

$$= L(\mathbf{1} - L)(\mathbf{1} - L^m)\alpha + L(\mathbf{1} - L^m)\beta + L^m(\mathbf{1} - L)^2\gamma + (\mathbf{1} - L)^2(\mathbf{1} - L^m)$$

Mirando los dos polinomios revelan la presencia del factor común $(\mathbf{1} - L)$ en ambos polinomios, indicando que el MEE tiene sobre diferenciación. La eliminación de la raíz unitaria común de ambos lados lleva a

$$\eta(L) = (\mathbf{1} - L)^2(\mathbf{1} - L^m)$$

$$\Theta(L) = (\mathbf{1} - L)^2(\mathbf{1} - L^m) + L(\mathbf{1} - L^m)\alpha + (L + \dots + L^m)\beta + L^m(\mathbf{1} - L)\gamma$$

El modelo contiene $(m + 1)$ términos promedios móviles pero sólo 3 parámetros, entonces difiere de un proceso ARIMA estacional usual. Cuando $\beta = 0$ este modelo corresponde al modelo de la aerolínea, difiriendo sólo en el factor $\alpha\gamma$ en el coeficiente de L^{m+1}

3. RESULTADOS

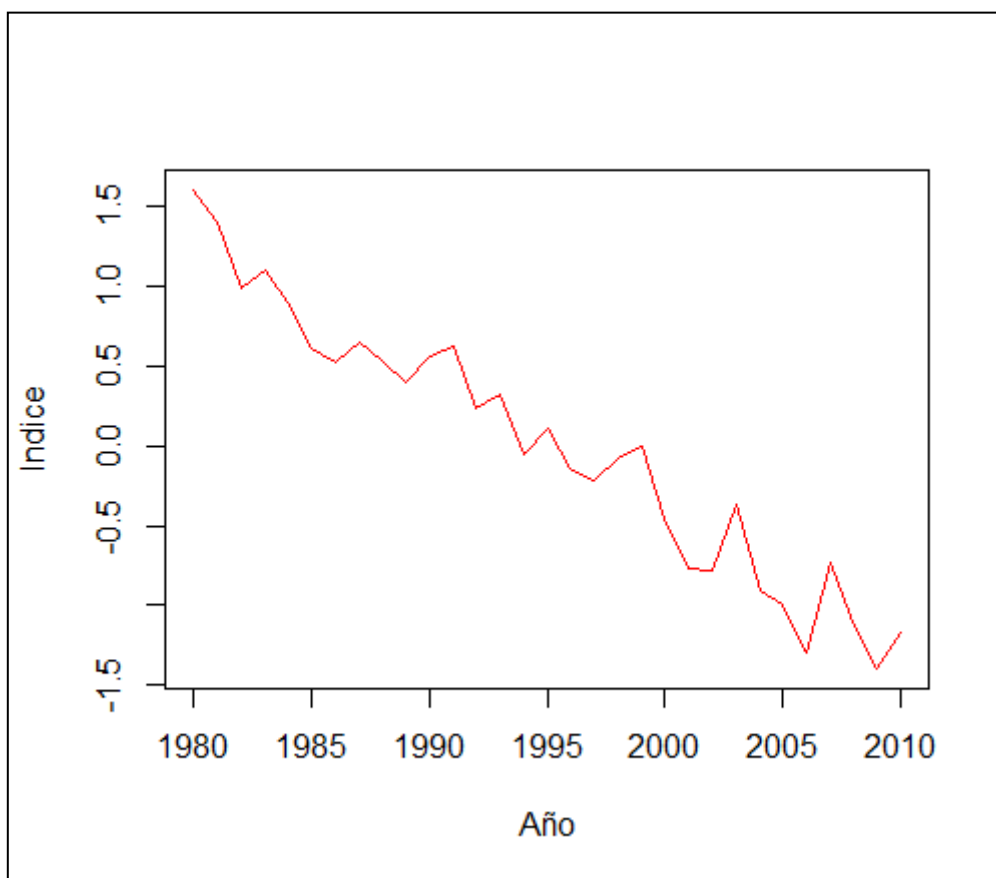
Para mostrar los resultados empíricos de los distintos modelos y las equivalencias entre ellos se trabaja con dos series: índice anual de mortalidad total de Lee-Carter para la provincia de Santa Fe (1980-2010) y litro de leche con destino a yogur mensuales. La primera ejemplifica

las series que sólo poseen tendencia, mientras que la segunda presenta componentes de tendencia y estacionalidad.

1) *Índice general de mortalidad*

La serie representa el índice general de mortalidad total de Lee-Carter en el período 1980-2010 la serie es elaborada por las autoras en base a datos proporcionados por el ministerio de Salud de la Nación. La misma se representa en el gráfico 1.

Gráfico 1. Índice general de mortalidad total de Lee-Carter



Fuente: Elaboración propia en base a datos del MSAL.

El modelo *ETS* que mejor ajusta los datos es un *ETS(A,A,N)*, es decir error aditivo y tendencia aditiva, por lo cual se puede escribir

$$y_t = l_{t-1} + b_{t-1} + \varepsilon_t$$

$$l_t = l_{t-1} + b_{t-1} + \alpha \varepsilon_t$$

$$b_t = b_{t-1} + \beta \varepsilon_t$$

Los valores iniciales del vector de estado son $l_0 = 1,367$ y $b_0 = -0,0774$. Los valores estimados por máxima verosimilitud de los parámetros de suavizado son: $\alpha = 0,017$ y $\beta = 0,017$ y $\sigma_\varepsilon^2 = 0,2005$.

Según las equivalencias expresadas en la sección ..., entre los *MEE* de innovación y los suavizados exponenciales, el suavizado que corresponde al *ETS(A, A, N)* es el de Holt, y toma la forma

$$\hat{y}_{(t|t+1)} = l_{t-1} + b_{t-1},$$

$$l_t = \alpha y_t + (1 - \alpha)(l_{t-1} + b_{t-1}),$$

$$b_t = \beta^*(l_t - l_{t-1}) + (1 - \beta^*)b_{t-1},$$

donde α coincide con la constante α del *ETS(A, A, N)* y $\beta^* = \beta/\alpha$ del mismo *ETS*. Si se estiman las constantes de Holt, haciendo mínima la suma de cuadrados de error se obtiene $\hat{\alpha} = 0,3794$ y $\hat{\beta} = 0,3276$. Estas constantes difieren notoriamente de los valores obtenido para el modelo *ETS(A, A, N)* dado que $\hat{\alpha} = 0,0017$ y $\hat{\beta} = 1$, incluso este último resultado mostraría que en lugar de utilizar el suavizado de Holt, se debería utilizar un suavizado exponencial simple.

Si se tiene en cuenta la equivalencia entre el modelo *ETS(A, A, N)* y los modelos *ARIMA* presentada en la sección El modelo *ARIMA* equivalente es el *ARIMA(0,0,2)* es decir $(1 - L)^2 y_t = (1 - L\theta_1 - L^2\theta_2)\varepsilon_t$. Los valores estimados de los parámetros para este modelo son:

	<i>Estimación</i>	<i>DS</i>	<i>t</i>	<i>p - valor</i>
θ_1	-1,9936	0,2749	-7,25	
θ_2	0,9999	0,2752	3,63	

Los valores equivalentes al *ETS* donde $\hat{\theta}_1 = (2 - \hat{\alpha} - \hat{\beta}) = 2 - 0,017 - 0,017 = 1,966$ y $\hat{\theta}_2 = (\hat{\alpha} - 1) = 0,017 - 1 = 0,993$. Si bien los parámetros equivalentes son bastante similares, se puede apreciar que el modelo *ARIMA* que se obtiene no cumple la condición de invertibilidad.

En este caso parece más apropiado proponer un *ARIMA(0,1,1) con constante*, lo que hablaría de un crecimiento fijo, no coincidente con el modelo *ETS* donde el crecimiento varía levemente en la misma forma que el nivel. El modelo *ARIMA* sería $(1 - L)y_t = \mu_t + (1 - L\theta_1)\varepsilon_t$ donde los parámetros estimados son $\hat{\mu} = 0,7704$ ($t = 1,61$) y $\hat{\theta}_1 = -0,1278$ (1,89).

Se realizan los pronósticos de ... pasos hacia adelante con sus correspondientes intervalos de predicción para el modelo *ETS*, para el suavizado de Holt equivalente no se exponen por ser los pronósticos puntuales coincidentes, en cambio se presentan los del método de Holt que dio como mejor según la estimación tradicional y la del modelo *ARIMA* equivalente y la del mejor modelo *ARIMA*. Tabla ... Gráfico...

X CONGRESO LATINOAMERICANO DE SOCIEDADES DE ESTADÍSTICA
CÓRDOBA, ARGENTINA. 16 A 19 DE OCTUBRE 2012

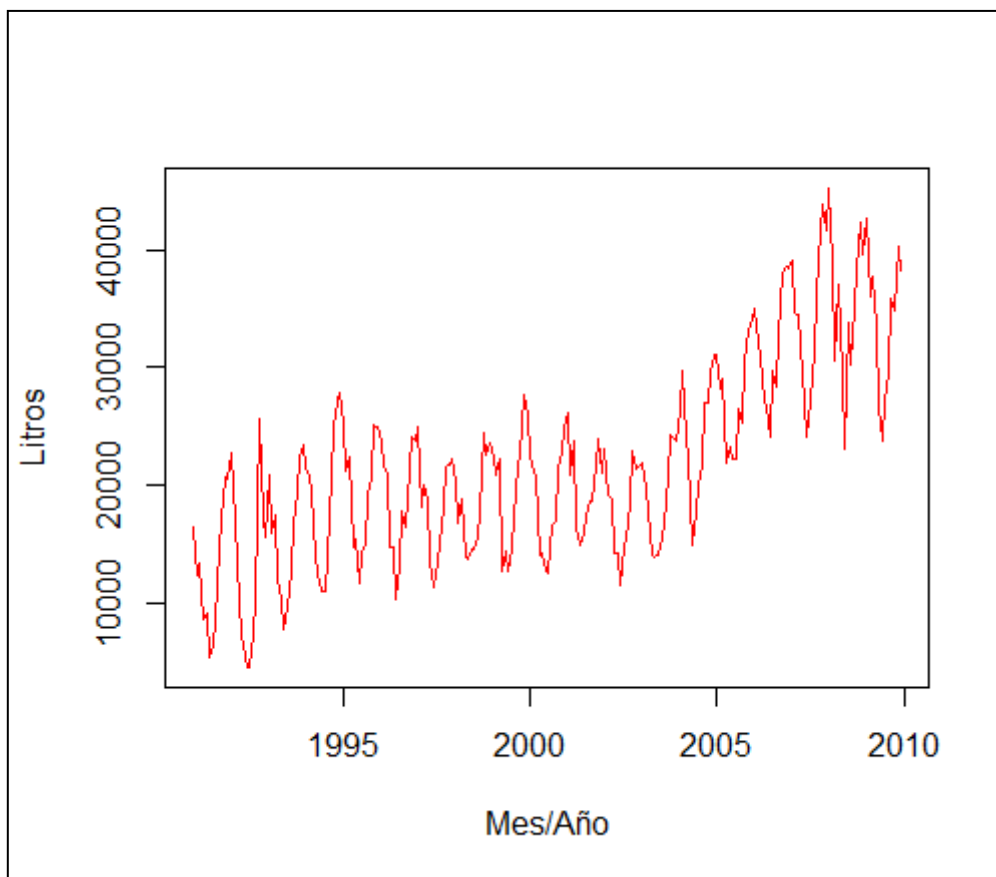
h	$ETS(A, A, N)$	$HOLT Tradicional$	$ARIMA(0,1,1) con constante$
1	-1,421 (-1,814; -1,028)	-1,369 (-1,852; -0,886)	-1,391 (-1,794; -0,989)
2	-1,514 (-1,908; -1,121)	-1,438 (-1,978; -0,898)	-1,478 (-1,882; -1,075)
3	-1,607 (-2,002; -1,213)	-1,508 (-2,127; -0,888)	-1,565 (-1,970; -1,162)
4	-1,701 (-2,096; -1,305)	-1,577 (-2,295; -0,858)	-1,623 (-2,057; -1,248)
5	-1,794 (-1,192; -1,396)	-1,646 (-2,480; -0,812)	-1,740 (-2,145; -1,334)

Los valores entre paréntesis corresponden a los intervalos de pronóstico del 95% de confianza.

2) *Serie litros de leche pasteurizada destinadas a la elaboración de yogur*

La serie litros de leche con destino a yogur mensuales, brindada por el Sistema Integrado de Información Agropecuaria. Abarca el período ene91' al dic09', se dejan los valores del primer semestre del 2010 para comparar pronósticos. Esta serie presenta una leve tendencia creciente y una marcada estacionalidad, gráfico 2.

Gráfico 2. Serie litros de leche pasteurizada destinadas a la elaboración de yogur



Fuente: Sistema integrado de Información Agropecuaria

El modelo de innovaciones de espacio de estado que corresponde es $ETS(A,A,A)$ que se escribe como.

$$y_t = l_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t,$$

$$l_t = l_{t-1} + b_{t-1} + \alpha \varepsilon_t,$$

$$b_t = b_{t-1} + \beta \varepsilon_t,$$

$$s_t = s_{t-m} + \gamma \varepsilon_t.$$

Las estimaciones máximo-verosímiles de los parámetros de suavizado son $\alpha = 0,3351$, $\beta = 0,0016$, $\gamma = 0,0004$ y $\sigma_\varepsilon^2 = 0,2005$.

Según la relación de equivalencia el método de suavizado correspondiente es el de Holt Winters, es decir

$$l_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)(l_{t-1} + b_{t-1}),$$

$$b_t = (1 - \beta^*)b_{t-1} + \beta^*(l_t - l_{t-1}),$$

$$s_t = \gamma^*(y_t - l_{t-1} - b_{t-1}) + (1 - \alpha)s_{t-m},$$

por lo tanto para este método las constantes de suavizado equivalentes son $\alpha = 0,3351$, $\beta^* = \frac{\beta}{\alpha} = 0,0048$, $\gamma^* = \gamma/(1 - \alpha) = 0,0006$, muy distantes de las constantes de suavizado encontradas por el método tradicional: $\alpha = 0,2473$, $\beta = 0,0134$, $\gamma = 0,3881$. Las

estimaciones máximo verosímiles del modelo *ETS* proporcionan valores de α y β muy cercanos a cero.

La considerable diferencia entre las estimaciones de las constantes de suavizado para el modelo *ETS* el respectivo método de Holt Winters fue advertido por otras personas que aplicaron estos métodos en otros países, a lo que Hyndman (2011) respondió que existen varios temas involucrados en estos resultados:

1) Holt Winters y *ETS* utilizan diferentes criterios de optimización, mientras Holt Winters utiliza valores heurísticos para el estado inicial y luego estima los parámetros de optimización por el error medio mínimo cuadrático (MSE), con *ETS* tanto el estado inicial como los parámetros de suavizado la optimización se realiza mediante de la función de verosimilitud (lo cuál es equivalente sólo para modelos aditivos lineales).

2) Las dos funciones utilizan diferentes rutinas de optimización y diferentes valores iniciales,

3) *ETS* investiga sobre un espacio de parámetros restringido para asegurar que el modelo resultante sea pronosticable, mientras que Holt Winters ignora este tema.

Cuando hay diferencias entre *ETS* y Holt Winters, los resultados de *ETS* son generalmente más confiables, Hyndman(2011).

El modelo *ARIMA* equivalente al *ETS(A,A,A)* suponiendo que el $\beta = 0$, es el llamado modelo de la aerolínea, *SARIMA(0,1,1)(0,1,1)₁₂* y en este caso como la estimación de β fue un valor cercano a cero, se los supone igual a cero para obtener el modelo más parsimonioso y frecuentemente usado en las aplicaciones. Por lo tanto, el modelo propuesto es:

$$(1 - L^{12})(1 - L)y_t = (1 - \theta_1 L)(1 - \theta_1 L^{12})\varepsilon_t$$

Los parámetros estimados por Máxima verosimilitud son:

	<i>Estimación</i>	<i>DS</i>	<i>t</i>	<i>p - valor</i>
θ_1	-0,6685	0,0523	-12,78	0,0000
θ_1	-0,7792	0,0699	-11,14	0,0000

Los valores que toman las estimaciones de los parámetros si se utilizan las relaciones de equivalencia dadas en la sección ... son $\hat{\theta}_1 = 1 - \hat{\alpha} = 0,6649$ y $\hat{\theta}_1 = 1 - \hat{\gamma} = 0,9996$, donde se observa que la mayor diferencia se presenta en la estimación de θ_1 . Se realizan los pronósticos ... hacia delante de los modelos *ETS(A,A,A)*, suavizado aditivo de holt Winters tradicional y *SARIMA(0,1,1)(0,1,1)₁₂* con sus respectivos intervalos de pronóstico.

Tabla Pronósticos

h	$ETS(A, A, A)$	$HW Tradicional$	$SARIMA(0,1,1)(0,1,1)_{12}$	<i>Valor real</i>
1	39.252 (35.074; 43.429)	40.133 (35.427; 44.838)	40.171 (35.710; 44.632)	40.571
2	35.903 (31.493; 40.314)	35.082 (30.232; 39.933)	35.937 (31.237; 40.636)	38.459
3	35.373 (30.740; 40.006)	33.655 (28.659; 38.651)	34.519 (29.592; 39.446)	32.349
4	30.856 (26.008; 35.704)	31.866 (26.726; 37.006)	32.044 (26.900; 37.187)	29.692
5	29.002 (23.946; 34.057)	26.912 (21.627; 32.196)	28.059 (22.706; 33.411)	25.090
6	27.174 (21.917; 32.431)	24.388 (18.960; 29.816)	25.511 (19.958; 31.064)	23.922

Tabla MAPE

h	$ETS(A, A, A)$	$HW Tradicional$	$SARIMA(0,1,1)(0,1,1)_{12}$
1	3,25	1,08	0,99
2	6,65	8,78	6,56
3	9,35	4,04	6,71
4	3,92	7,32	7,92
5	15,59	7,26	11,83
6	13,59	1,95	6,64
	8,73	5,07	6,77

4. DISCUSIÓN

5. CONCLUSIONES

6. REFERENCIAS

FIGURAS, TABLAS Y ECUACIONES

Las tablas, figuras y gráficos deberán tener un título y estar numeradas correlativamente. Las ecuaciones deberán estar numeradas a la derecha.

CITAS DE REFERENCIAS EN EL TEXTO

Para citar un artículo en el texto, se indicará autor y año de publicación, como en los siguientes ejemplos:

..... the model proposed by Barnett (1969)

The theoretical treatment provided by Fuller (1987, cap.4)

Bold et al. (1995) also find....

REFERENCIAS

Las referencias deberán disponerse en orden alfabético del nombre del autor y, para un mismo autor, en orden cronológico.

Las partes que debe contener una referencia son las siguientes: Autor(es), año de publicación, título, información sobre la publicación.

Las referencias deberán presentarse en el siguiente estilo:

THEOBALD, C. M. and MALLISON, J. R. (1978). Comparative Calibration, Linear Structural Relationship and Congeneric Measurements. *Biometrics*, **34**, 39-45.

FULLER, W. A. (1987). *Measurement Error Models*. Wiley. New York.

LINDLEY, D. V. and SMITH, A. F. M. (1972) Bayes Estimates for the Linear Model (with discussion). *Journal of the Royal Statistical Society, Series B*, **34**, 1-41.

AGRADECIMIENTOS

La palabra AGRADECIMIENTOS, justificada a izquierda.

La extensión máxima total del trabajo completo (desde el título hasta el final de los agradecimientos) será de veinticinco páginas.