



Blaconá M.T., Magnano, L. Andreozzi, L.

Instituto de Investigaciones Teóricas y Aplicadas de la Escuela de Estadística

CARACTERÍSTICAS DE LOS MODELOS DE ESPACIO DE ESTADO DE INNOVACIONES, CON APLICACIONES.

1. Introducción

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

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 empíricos 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 al 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 los MEE los y los modelos ARIMA.

La novedad en los modelos propuestos por Hyndman et al (2008) es que proponen una única fuente de error a diferencia de los clásicos MEE propuestos por Harvey (1989) que contemplan una fuente de error distinta para la ecuación de medida y para cada componente incluida en el modelo, por ello los primeros no necesitan ser estimados con el filtro de Kalman, sino que se usan los métodos iterativos tradicionales para estimación Máximo Verosímil no lineales, basado en la representación de las innovaciones de la función de densidad de probabilidad, a esto último se debe el nombre que se le da a estos modelos. En este trabajo se realiza una breve introducción de los modelos de innovaciones de espacio de estado, siguiendo las ideas de Hyndman et. al. (2008) para luego mediante aplicaciones a series de tiempo reales, mostrar las relaciones entre los distintos tipos de modelos con sus respectivas equivalencias, como así también bondades y dificultades que presentan los modelos desde un punto de vista empírico.

En la Sección 2 se presentan los tradicionales suavizados exponenciales más usados y los modelos de innovación de espacio de estado generales y los que corresponden a los suavizados exponenciales mencionados; en la Sección 3 se analizan series de tiempo Argentinas por los dos métodos enunciados anteriormente y por los tradicionales modelos ARIMA, comparando sus resultados; en la Sección 4 se realiza la discusión de los resultados encontrados.

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 el 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 $T_h = \ell$,

A Aditiva $T_h = \ell + bh$,

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

M Multiplicativa $T_h = \ell b^h$,

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

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

Si se agrega el componente estacional con la misma nomenclatura que para la tendencia, sin tener en cuenta los errores, se definirán quince tipos de modelos. 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 comúnmente.

Para cada uno de los quince métodos de suavizado, existen dos modelos posibles de espacio de estado, uno correspondiente al modelo con error aditivo y otro con error multiplicativo. Si se usan las constantes de suavizado 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 tiene un valor cercano a 1, el nuevo pronóstico introducirá un ajuste sustancial del error de los pronósticos previos. Si α , en cambio, tiene un valor 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

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

este resultado de valores constantes se debe a que 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)

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

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

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

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

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

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

este método se conoce como SES con dirección, similar al método Theta 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)

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

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

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

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

m: longitud de la estacionalidad

$$h_m^* = [(h - 1) \text{ mod } m] + 1.$$

Generalmente α , β^* y γ varían entre cero y uno. Para hacer más simple la formulación del MEE, Ord et. al. (1997) reemplazan (2.6.c) por $s_t = \gamma \frac{y_t}{\ell_t} + (1 - \gamma)s_{t-m}$, en este caso difieren levemente los pronósticos.

Estacionalidad Aditiva (A,A)

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

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

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

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

como en el modelo multiplicativo la ecuación (2.7.c) en los libros de texto es un poco diferente, se usa $s_t = \gamma^*(y_t - \ell_t) + (1 - \gamma^*)s_{t-m}$. Si ℓ_t se sustituye usando (2.7.a) se obtiene $s_t = \gamma^*(1 - \alpha)(y_t - \ell_{t-1} - b_{t-1}) + [1 - \gamma^*(1 - \alpha)]s_{t-m}$, 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).



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. Se utiliza la nomenclatura, 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 $\hat{\mu}_{t+h} = \hat{y}_{t+h}$. 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 $E(y_{t+h}/\mathbf{x}_t) = \mu_{(t+h|t)} = \hat{y}_{t+h}$, el pronóstico un paso hacia adelante suponiendo conocido los valores de todos las constantes de suavizado resulta $\mu_{(t|t-1)} = \hat{y}_{(t|t-1)} = \ell_{t-1} + b_{t-1}$.

Sea $\varepsilon_t = y_t - \mu_{(t|t-1)}$ error de pronóstico un paso hacia adelante en el tiempo t , a partir de (2.4.c) se encuentra

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

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

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

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

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

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



Definiendo el vector de estado $\mathbf{x}_t = (\ell_t, b_t)'$ se pueden escribir las ecuaciones (2.8) a (2.10) como

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

$$\mathbf{x}_t = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \mathbf{x}_{t-1} + \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \varepsilon_t. \quad (2.11.b)$$

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

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

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

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

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

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

donde

$$\mathbf{x}_t = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \mathbf{x}_{t-1} + [1 \quad 1] \mathbf{x}_{t-1} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \varepsilon_t,$$

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

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

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 $\mathbf{x}_t = (\ell_t, b_t, s_t, s_{t-1}, \dots, s_{t-m+1})$ y una ecuación de forma

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

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

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

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 puntal de y_t .

En el libro de Hyndman et. al. (2008) se presentan los modelos aditivos y 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 siendo $E(\varepsilon_{n+j} | \mathbf{x}_n) = 0$ 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 en los casos de tendencia y estacionalidad multiplicativa para $h \geq 2$.

Los intervalos de pronóstico de un modelo lineal, donde las predicciones se distribuyen Gausseanas, se pueden deducir de las variancias condicionales $v_{(n+h|n)} = V(y_{n+h|n})$. Este enfoque también funciona para algunos modelos no lineales.

2.4 Iniciación y estimación de los modelos de espacio de estado

Si se desea 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 las constantes de suavizado α , β , γ y ϕ .

En 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 las innovaciones del MEE (2.12), a partir de la cual se obtienen los estimadores máximo verosímiles. Esta función de verosimilitud (con los términos constantes eliminados), es condicional a las constantes de suavizado $\theta = (\alpha, \beta, \gamma, \phi)'$ y los valores iniciales del estado $\mathbf{x}_0 = (\ell_0, b_0, s_0, s_1, \dots, s_{m+1})$, donde n es en número de observaciones. Esto se calcula utilizando ecuaciones recursivas. No es necesario usar el filtro de Kalman para calcular la verosimilitud, como en los MEE clásicos, 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 , la que 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:

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

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

y_t : valor observado en el tiempo t ,

\mathbf{x}_t : vector de estado en el tiempo t ,

$\mathbf{F}\mathbf{x}_{t-1}$: efecto del pasado sobre el estado actual \mathbf{x}_t ,

\mathbf{F} : matriz de transición,

$\mathbf{g}\varepsilon_t$: 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.

Esta estructura permite explorar la necesidad de cada componente separadamente y llevar a cabo una investigación sistemática del mejor modelo para 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.

Este modelo aditivo es un caso especial del modelo más general (2.12). En el suavizado exponencial el vector de estado contiene información nueva sobre el nivel, crecimiento y patrón estacional. Por ejemplo, en un modelo con tendencia y estacionalidad $\mathbf{x}_t = (\ell_t, b_t, s_t, s_{t-1}, \dots, s_{t-m})$. Desde una perspectiva matemática, la variable de estado es esencialmente redundante. 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.

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

El valor 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 disponible, 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. Existen



alternativa diferentes del MEE y con fuentes de aleatoriedad dadas en (2.13.a) y (2.13.b), no correlacionadas, en lugar de una única fuente de aleatoriedad (las innovaciones en cada caso).

Las 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 \mathbf{x}_0 y las observaciones, se pueden calcular recursivamente a partir de los valores de la serie usando las relaciones:

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

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

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

estas transformaciones se llaman 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

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

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

Las ecuaciones (2.12), (2.14) y (2.15) demuestran lo atractivo 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

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

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

donde $\varepsilon_t \sim N(0, \sigma^2)$. En conformidad con la estructura de espacio de estado con $\mathbf{x}_t = \ell_t$, $\mathbf{w} = 1$, $\mathbf{F} = 1$ 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 regido por el tamaño de la constante de suavizado α . Los casos para $\alpha = 1$ y $\alpha = 0$ son de especial interés.

Caso $\alpha = 0$: 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

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

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

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

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

$$\ell_t = (1 - \alpha)\ell_{t-1} + \alpha y_t. \quad (2.18)$$

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

$$\hat{y}_{(t|t-1)} = (1 - \alpha)\ell_0 + \alpha \sum_{j=0}^{t-1} (1 - \alpha)^j y_{t-j}, \quad (2.19)$$

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 siendo $|1 - \alpha| < 1$) la predicción $\hat{y}_{(t|t-1)}$ es un



promedio ponderado exponencialmente de observaciones pasadas. Los coeficientes dependen del factor $1 - \alpha$:

- Si $|1 - \alpha| < 1$, los coeficientes se vuelven pequeños cuando j crece. La condición de estabilidad se cumple si y sólo si $0 < \alpha < 2$,
- Los coeficientes son positivos si y sólo si $0 < (1 - \alpha) < 1$, en este caso (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 pensar 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 frecuentemente el rango más estrecho.

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

El modelo de nivel local se va incrementando a una tasa de crecimiento b_t , dando

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

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

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

Hay dos constantes 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 estado

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

Las constantes 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 las constantes de suavizado, la tasa de crecimiento sigue un camino aleatorio en el tiempo. Cuando las constantes de suavizado crecen, la tendencia de la serie se hace más suave.

En este modelo la transformación (2.14) de los valores de la serie, genera innovaciones de la forma

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

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

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

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

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

Un sistema de ecuaciones equivalente es:



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

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

$$\ell_t = \alpha y_t + (1 - \alpha)(\ell_{t-1} + b_{t-1}), \quad (2.21.c)$$

$$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)b_{t-1}, \quad (2.21.d)$$

donde $\beta^* = \frac{\beta}{\alpha}$. A menudo al término $\ell_t - \ell_{t-1}$ se lo interpreta 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 D^j converge a 0 cuando j crece) cuando $\alpha > 0$, $\beta > 0$ y $2\alpha + \beta < 4$. Esto conduce a una región paramétrica 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 planteada permite cambios en el patrón estacional. El modelo es:

$$y_t = \ell_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t, \quad (2.22.a)$$

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

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

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

este modelo corresponde al MEE de primer orden donde

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

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

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

y

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



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

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

$$\ell_t - \delta = \ell_{t-1} - \delta + b_{t-1} + \alpha\varepsilon_t,$$

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

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

las cuales son equivalentes a (2.22). Para eliminar este 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 menos tenga media cero). Cuando se cumple esta condición del componente estacional se dice que está 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 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

$$\hat{y}_t = \ell_{t-1} + b_{t-1} + s_{t-m},$$

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

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

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

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

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

$$\hat{y}_t = \ell_{t-1} + b_{t-1} + s_{t-m},$$

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

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

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

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

donde el valor de la serie se desestacionaliza en la ecuación de tendencia y se considera sin tendencia en la ecuación de estacionalidad, $\beta^* = \beta/\alpha$ y $\gamma^* = \gamma/(1 - \alpha)$. Las ecuaciones anteriores pueden interpretarse como promedios ponderados. En el caso que las restricciones paramétricas sean que α , β^* y γ^* caigan entre 0 y 1, según sean las consideraciones sobre la matriz de descuento D conducirán a diferentes regiones a las constantes de suavizado.



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

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

$$(1 - L)l_t = \alpha\varepsilon_t, \quad (2.23.b)$$

Para encontrar la forma reducida del ETS(A, N, N) al modelo ARIMA se le aplica el operador de diferencia $(1 - L)$ a ambos lados de la ecuación de medida (2.23.b). 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. Al igual que producirán también la misma variancia de pronósticos e intervalos de pronóstico. El espacio paramétrico del ETS es $\alpha \in (0,2)$ correspondiente al espacio del parámetro del ARIMA con $|\theta_1| < 1$. Sin embargo, el supuesto de inicio finito permite el esquema ETS con $\alpha = 0$, correspondiente a una media constante, en cambio, el modelo ARIMA no incluye este caso.

Modelo ETS(A, A, N)

En forma similar, se puede escribir el modelo local ETS(A, A, N), el cual es un MEE de innovaciones correspondiente al 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)



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

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

Las condiciones de invertibilidad son: $\theta_1 + \theta_2 < 1$, $\theta_2 - \theta_1 < 1$ y $|\theta_2| < 1$, que en función de los coeficientes del ETS son $\beta > 0$, $4 - 2\alpha - \beta > 0$ y $0 < \alpha < 2$.

Esta forma reducida supone implícitamente 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 $Z_t = \delta(L)y_t$ resulta el modelo ARIMA

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

El MEE de innovaciones lineal general $y_t = \mathbf{w}'(\mathbf{x}_{t-1}) + \varepsilon_t$ y $\mathbf{x}_t = \mathbf{F}(\mathbf{x}_{t-1}) + \mathbf{g}\varepsilon_t$ se puede reducir a un modelo ARIMA utilizando el operador de rezago. Primero la ecuación de transición se puede escribir como

$$(\mathbf{I} - \mathbf{FL})\mathbf{x}_t = \mathbf{g}\varepsilon_t, \quad (2.25)$$

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

$$\det(\mathbf{I} - \mathbf{FL})\mathbf{x}_t = \text{adj}(\mathbf{I} - \mathbf{FL})\mathbf{g}\varepsilon_t. \quad (2.26)$$

Aplicando el operador $\det(\mathbf{I} - \mathbf{FL})$ a ambos lados de la ecuación de medida se encuentra

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

A partir (2.26) sustituyendo el $\det(\mathbf{I} - \mathbf{FL})\mathbf{x}_t$ se obtiene el siguiente modelo ARIMA

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

En este modelo ARIMA $\eta(L) = \det(\mathbf{I} - \mathbf{FL})$ y $\theta(L) = \mathbf{w}'\text{adj}(\mathbf{I} - \mathbf{FL})\mathbf{g} + \det(\mathbf{I} - \mathbf{FL})$.

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. Esta forma reducida corresponderá a un modelo ARIMA después de eliminar algunos factores comunes a $\eta(L)$ y $\theta(L)$. Técnicamente la eliminación de las 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

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

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

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

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

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

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

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

y

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

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

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

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

$$= L(1 - L)(1 - L^m)\alpha + L(1 - L^m)\beta + L^m(1 - L)^2\gamma + (1 - L)^2(1 - L^m).$$

En ambos polinomios está presente el factor común $(1 - L)$, indicando que el MEE tiene sobre diferenciación. La eliminación de la raíz unitaria común de ambos lados lleva a

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

$$\theta(L) = (1 - L)^2(1 - L^m) + L(1 - L^m)\alpha + (L + \dots + L^m)\beta + L^m(1 - L)\gamma.$$



El modelo contiene $(m + 1)$ términos promedios móviles pero sólo tres parámetros, por lo tanto difiere de un proceso usual ARIMA estacional. Cuando $\beta = 0$ este modelo corresponde al modelo de la aerolínea, difiriendo sólo en el factor α que acompaña al 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: 1) índice anual de mortalidad total de Lee-Carter para la provincia de Santa Fe (1980 – 2010) y 2) litros de leche mensuales con destino a la fabricación yogur en Argentina (enero 1991 a diciembre 2009). La primera ejemplifica las series que sólo poseen tendencia, mientras que la segunda presenta componentes de tendencia y estacionalidad.

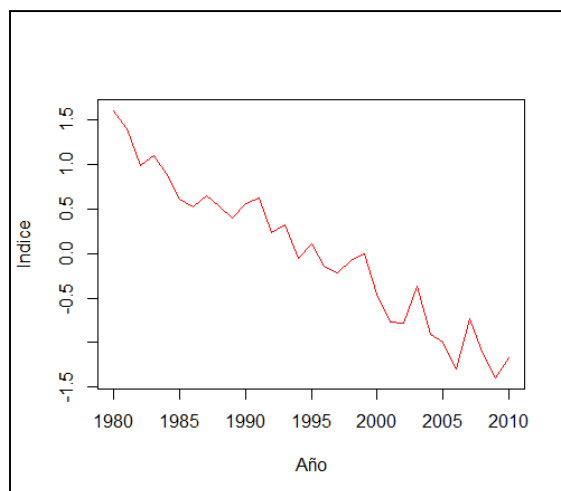
Los resultados numéricos se calculan con el programa R, utilizando el paquete forecast, versión 3.22., elaborado por R. J. Hyndman (2012).

1) Índice general de mortalidad

La serie representa el índice general de mortalidad total de Lee-Carter en el período 1980-2010, la que fue 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.

El modelo ETS que mejor ajusta los datos, según el criterio de Akaike, es un ETS(A, A, N), el cuál posee error y tendencia aditivos.

Gráfico 1. Índice general de mortalidad total de Lee-Carter para la provincia de Santa Fe (1980-2010)





Fuente: Elaboración propia en base a datos del MSAL.

Los valores iniciales del vector de estado son $\ell_0 = 1,367$ y $b_0 = -0,0774$. Los valores estimados de las constantes de suavizado por máxima verosimilitud son: $\hat{\alpha} = 0,017$ y $\hat{\beta} = 0,017$ y $\hat{\sigma}_\varepsilon^2 = 0,2005$.

Según las equivalencias expresadas en la sección 2.4.1, entre los MEE de innovación y los suavizados exponenciales, el suavizado que corresponde al ETS(A,A,N) es el de Holt, 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, en forma tradicional (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 obtenidos para el modelo ETS(A,A,N) dado que los equivalentes por ETS son $\hat{\alpha} = 0,0017$ 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 2.5, el modelo ARIMA equivalente es un ARIMA(0,2,2) en consecuencia $(1 - L)^2 y_t = (1 - L\theta_1 - L^2\theta_2)\varepsilon_t$. Los valores estimados de los parámetros para este modelo se presentan en la Tabla 1.

Tabla 1. Parámetros estimados para el modelo ARIMA(0,2,2), serie Índice general de mortalidad total de Lee-Carter para la provincia de Santa Fe.

	Estimación	DS	t	p - valor
θ_1	-1,9936	0,2749	-7,25	0,0000
θ_2	0,9999	0,2752	3,63	

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

Para este caso según el criterio de Akaike el modelo apropiado es un ARIMA(0,1,1) con constante, lo que indicaría un crecimiento fijo, no coincidente con el modelo ETS donde el crecimiento varía levemente en la misma forma que el nivel. El modelo ARIMA es $(1 - L)y_t = \mu_t + (1 - L\theta_1)\varepsilon_t$ donde los parámetros estimados son $\hat{\mu} = 0,7704$ ($p = 0,0294$) y $\hat{\theta}_1 = -0,1278$ ($p = 0,0537$).



Tabla 2. Pronósticos cinco pasos hacia adelante por diferentes métodos de la serie Índice general de mortalidad total de Lee-Carter para la provincia de Santa Fe.

h	ETS(A, A, N)	HOLT Tradicional	ARIMA(0,2,2)
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.

Se realizan los pronósticos de 5 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 tradicional y la del mejor modelo ARIMA. Tabla 2.

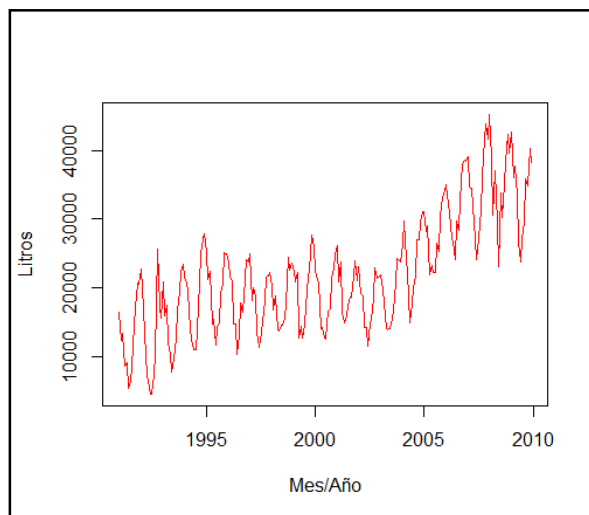
Según estos pronósticos el método con menor incertidumbre en sus predicciones (intervalos más estrechos), es el ETS, seguido por el modelo ARIMA y por último el modelo de Holt tradicional.

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

La serie litros de leche mensuales con destino a fabricación de yogur en Argentina, brindada por el Sistema Integrado de Información Agropecuaria, abarca el período enero 1991 a diciembre de 2009. 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 mensual litros de leche pasteurizada destinadas a la elaboración de yogur. (enero 1991 a diciembre 2009).



Fuente: Sistema integrado de Información Agropecuaria

El modelo de innovaciones de espacio de estado más conveniente según el AIC es ETS(A,A,A). Las estimaciones máximo-verosímiles de los parámetros de suavizado son $\hat{\alpha} = 0,3351$, $\hat{\beta} = 0,0016$, $\hat{\gamma} = 0,0004$ y $\hat{\sigma}_{\varepsilon}^2 = 2131,512$. Según la relación de equivalencia el método de suavizado correspondiente es el de Holt-Winters. Para este método las constantes de suavizado equivalentes son $\hat{\alpha} = 0,3351$, $\hat{\beta}^* = \frac{\hat{\beta}}{\hat{\alpha}} = 0,0048$, $\hat{\gamma}^* = \hat{\gamma}/(1 - \hat{\alpha}) = 0,0006$, muy distantes de las constantes de suavizado encontradas por el método tradicional: $\hat{\alpha} = 0,2473$, $\hat{\beta} = 0,0134$, $\hat{\gamma} = 0,3881$. Las estimaciones máximo verosímiles del modelo ETS proporcionan valores de $\hat{\alpha}$ y $\hat{\beta}$ muy cercanos a cero. Incluso tienen interpretaciones distintas, mientras el coeficiente γ estimado por ETS es prácticamente nulo por lo que la estacionalidad puede considerarse constante; por HW presenta un valor que representa un cambio de este componente a través del tiempo. El cambio en la estimación de α y β son más coincidentes por ambos métodos, mostrando a través de las estimaciones de α un cambio en el nivel a través del tiempo y a través de las estimaciones de β un crecimiento constante.

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. En este caso como la estimación de β es un valor cercano a cero, se lo considera nulo para obtener un modelo 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 se observan en la tabla 3.

Tabla 3 Estimaciones de los parámetros del modelo SARIMA(0,1,1)(0,1,1)₁₂, serie litros de leche mensuales con destino a fabricación de yogur en Argentina.

	Estimación	DS	t	p – valor
θ_1	-0,6685	0,0523	-12,78	0,0000
θ_1	-0,7792	0,0699	-11,14	0,0000

Tabla 4. Pronósticos seis pasos hacia adelante por diferentes métodos de la serie litros de leche mensuales con destino a fabricación de yogur en Argentina.

h	ETS(A, A, A)	HW tradicional	SARIMA(0,1,1)(0,1,1) ₁₂	Valor real
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

Los valores que toman las estimaciones de los parámetros si se utilizan las relaciones de equivalencia dadas en la sección 2.5 son $\hat{\theta}_1 = 1 - \hat{\alpha} = 0,6649$ y $\hat{\Theta}_1 = 1 - \hat{\gamma} = 0,9996$. La mayor diferencia se presenta en la estimación de θ_1 .

Se realizan los pronósticos 6 pasos hacia adelante de los modelos ETS(A, A, A), suavizado aditivo de HW tradicional y SARIMA(0,1,1)(0,1,1)₁₂ con sus respectivos intervalos de pronóstico, tabla 4.

Los intervalos de pronóstico de los tres métodos contienen el valor real, siendo el que presenta menos incertidumbre el ETS, segundo del SARIMA y por último HW, aunque las diferencias no son muy pronunciadas. Se calcula el MAPE de los tres modelos para $h = 1, 2, \dots, 6$, Tabla 5. Los pronósticos más exactos son los del suavizado de HW tradicional seguido por el SARIMA y finalmente el ETS.



Tabla 5. MAPE calculado para diferentes métodos de la serie litros de leche mensuales con destino a fabricación de yogur en Argentina.

h	ETS(A, A, A)	HW Tradicional	SARIMA(0,1,1)(0,1,1) ₁₂
1	3,3	1,1	1,0
2	5,0	4,9	3,8
3	6,4	4,6	4,8
4	5,8	5,3	5,5
5	7,8	5,7	6,8
6	8,7	5,1	6,8

Los MAPE se calculan para cada h como $\frac{1}{h} \sum_{i=1}^h \left| \frac{y_{t+i} - \hat{y}_{t+i}}{y_{t+i}} \right| \times 100$

3. Consideraciones Finales

En este trabajo se realiza una breve introducción de los modelos de innovaciones de espacio de estado y su relación con los métodos de suavizado exponencial. Si bien los pronósticos puntuales de ambos métodos coinciden cuando se definen las constantes de suavizado equivalentes, los MEE permiten entre otras cosas encontrar intervalos de predicción y estimación máximo verosímil de dichas constantes.

Se presentan los métodos de suavizado generalmente más usados y se definen los correspondiente MEE de innovaciones. Se encuentra la forma reducida de los MEE, las cuales coinciden con distintos modelos ARIMA.

Para la estimación máximo verosímil de MEE de innovaciones, se utilizan ecuaciones recursivas, no siendo necesario usar el filtro de Kalman, para calcular la verosimilitud, como en los clásicos MEE, por no tener múltiples fuentes de error.

Se realizan dos aplicaciones a series reales, una con tendencia y otra con tendencia y estacionalidad. Los resultados muestran que en estos casos las constantes de suavizado de los métodos de suavizado exponencial equivalentes a las del modelo MEE de innovaciones difieren notablemente en sus valores. Según Hyndman (2011), estas diferencias se deben a que: a) se utilizan diferentes criterios de optimización, b) se usan diferentes rutinas de optimización y valores iniciales y c) ETS investiga sobre un espacio paramétrico restringido para asegurar que el modelo resultante sea pronosticable, aspecto que ignoran los suavizados exponenciales

La forma reducida del MEE de innovaciones, coincidentes con los modelos ARIMA, no resultan las más parsimoniosos e incluso presentan problemas de invertibilidad, por ello se comparan los modelos ETS con el mejor modelo ARIMA según AIC.



Para el segundo ejemplo los intervalos de pronóstico de los tres métodos contienen el valor real, siendo el que presenta menor incertidumbre (intervalos de pronóstico más estrechos), el ETS, seguido del SARIMA y por último HW, aunque las diferencias son pequeñas.

Con respecto a la exactitud de los pronósticos, la serie 2) presenta valores de MAPE mayores en el modelo ETS, seguido por el modelo ARIMA y con los valores más bajos para Holt Winters.

La considerable diferencia entre las estimaciones de las constantes de suavizado para el modelo ETS con respecto a las equivalentes del método de Holt Winters fue advertido también en otras aplicaciones de estos métodos en otros países, a lo que Hyndman (2011) responde que existen varios temas involucrados en estos resultados:

- 1) Holt Winters y ETS utilizan diferentes criterios de optimización, mientras Holt Winters emplea valores heurísticos para el estado inicial y luego estima los parámetros de optimización por el error medio mínimo cuadrático (MSE), mientras que en los modelos ETS el estado inicial y la optimización de los parámetros de suavizado se realiza mediante de la función de verosimilitud;
- 2) Las dos funciones utilizan diferentes rutinas de optimización y distintos valores iniciales,
- 3) ETS investiga sobre un espacio paramétrico restringido para asegurar que el modelo resultante sea pronosticable, mientras que Holt Winters ignora este aspecto.

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

Estos resultados muestran que no hay una supremacía absoluta de un método sobre los otros, sin embargo, los modelos ETS permiten una interpretación más directa de los componentes que intervienen en la serie.

Dado los múltiples interrogantes que se plantean en las aplicaciones de este trabajo, en futuras investigaciones se seguirá profundizando sobre las bondades y debilidades del modelo de espacio de estado de innovaciones.



4. Referencias

- ABRAHAM, B. and LEDOLTER, J. (1983). Statistical Methods for Forecasting. Wiley. New York.
- BOX, G. E. P., JENKINS, G. M. and REINSEL, G. C. (1994). Time Series Analysis: Forecasting and Control. Prentice-Hall, Englewood Cliffs, New Jersey.
- BROWN, R. G. (1959). Statistical Forecasting for Inventory Control. McGraw-Hill. New York.
- HARVEY, A. C., (1989). Forecasting structural time Series Models and the Kalman Filter. Cambridge University Press, Cambridge.
- HYNDMAN, R. J., KOEHLER, A.B., SNEIDER, R. D. and GROSE, S (2002). A State Space Framework for Automatic Forecasting using Exponential Smoothing Methods. International Journal of Forecasting, 18, (3), 439-454
- HYNDMAN, R. J., KOEHLER, A.B., ORD, J.K., and SNEIDER, R. D. (2008). Forecasting with Exponential Smoothing – The State Space Approach. Springer-Verlag. Berlin Heidelberg.
- HYNDMAN, R. J., (2011) <http://robjhyndman.com/researchtips/estimation2>.
- HOLT, C. C., (1957). Forecasting trend and seasonals by Exponential Weighted Average, O.N.R. Memorandum 52/1957. Carnegie Institute of Technology.
- ORD, J.K., KOEHLER, A.B., and SNEIDER, R. D. (1997). Estimation and prediction for a class of dynamic nonlinear statistical models. Journal of American statistical association 92 1621-1629.
- WINTERS, P. R., (1960). Forecasting Sales by Exponentially Weighted Moving Average. Management Sciences, 6 324-342.