



**EXPLORACIÓN Y ANÁLISIS EXHAUSTIVO DE LOS MÉTODOS
ESTADÍSTICOS APLICADOS A LAS SERIES DE TIEMPO**

**EXPLORATION AND EXHAUSTIVE ANALYSIS OF THE STATISTICAL
METHODS APPLIED TO TIME SERIES**

Autor

Martino Florencia

E-mail

flormartino_42@hotmail.com

Eje Temático

Matemática en las Ciencias Económicas

Modalidad

Trabajos Científicos Académicos

Palabras claves: series de tiempo – modelado de ciclos

Resumen

Las series de tiempo han cobrado gran relevancia sobre todo en lo que refiere a las ciencias económicas, como puede ser variables de evolución de precios a lo largo del tiempo. Por ello se planteó realizar una investigación profunda sobre los métodos estadísticos aplicados a las mismas para un futuro análisis de datos.

Introducción

Una serie de tiempo es un conjunto de observaciones pasadas de una variable medida en puntos sucesivos en el tiempo o en periodos de lapsos sucesivos, es decir, si son datos históricos restringidos a valores anteriores de la variable, al método de pronóstico se lo conoce como método de series de tiempo.

El objetivo del análisis de series temporales es doble. Por un lado, se busca explicar las variaciones observadas en la sucesión en el pasado, tratando de determinar si responden a un determinado patrón de comportamiento. Por otro lado, si se logra definir ese patrón, se tratará de predecir el comportamiento futuro de la variable. Para lograr estos objetivos se admite que la serie temporal



es una función del tiempo. Así, los datos (en nuestro caso, matrícula de alumnos en educación a distancia) se define como la variable dependiente y el tiempo como la exógena, teniendo claro que el tiempo no es una variable explicativa, sino que es el escenario o contexto donde ocurre la serie temporal.

Las series de tiempo presentan diversos componentes, que definen su comportamiento o patrón. Por lo general, se supone que son 4 los componentes que se combinan para dar los valores a la misma: tendencia, cíclico, estacional o irregular. La tendencia se presenta cuando las series muestran una trayectoria gradual ya sea positiva o negativa a través de un largo plazo. El componente cíclico se debe a que, aunque la serie abarque lapsos largos de tiempo, no todos sus valores caen exactamente sobre la línea de tendencia. Puede ocurrir que ellos se encuentren alternándose por arriba y por debajo de la misma, y si, además, esto presenta una duración mayor a un año, puede decirse que la serie presenta ciclos. El elemento estacional refiere a la variabilidad en los datos debido a la influencia temporal. Además, se consideran variaciones de este tipo cuando las mismas se presentan en el lapso de un año y con regularidad durante el mismo. Por último, el componente irregular de una serie de tiempo es el factor residual o el que da cuenta de las desviaciones de los valores reales de la misma que se esperan al considerar los efectos de los restantes elementos. El irregular es ocasionado por factores de corto plazo, imprevistos y no recurrentes que afectan a la serie.

1) Materiales y métodos

Una serie de tiempo es un conjunto de observaciones pasadas de una variable medida en puntos sucesivos en el tiempo o en periodos de lapsos sucesivos, es decir, si son datos históricos restringidos a valores anteriores de la variable, al método de pronóstico se lo conoce como método de series de tiempo.

El objetivo del análisis de series temporales es doble. Por un lado, se busca explicar las variaciones observadas en la sucesión en el pasado, tratando de determinar si responden a un determinado patrón de comportamiento. Por otro lado, si se logra definir ese patrón, se tratará de predecir el comportamiento futuro de la variable. Para lograr estos objetivos se admite que la serie temporal es una función del tiempo. Así, los datos (en nuestro caso, matrícula de alumnos en educación a distancia) se define como la variable dependiente y el tiempo como la exógena, teniendo claro que el tiempo no es una variable explicativa, sino que es el escenario o contexto donde ocurre la serie temporal.



Las series de tiempo presentan diversos componentes, que definen su comportamiento o patrón. Por lo general, se supone que son 4 los componentes que se combinan para dar los valores a la misma: tendencia, cíclico, estacional o irregular. La tendencia se presenta cuando las series muestran una trayectoria gradual ya sea positiva o negativa a través de un largo plazo. El componente cíclico se debe a que, aunque la serie abarque lapsos largos de tiempo, no todos sus valores caen exactamente sobre la línea de tendencia. Puede ocurrir que ellos se encuentren alternándose por arriba y por debajo de la misma, y si, además, esto presenta una duración mayor a un año, puede decirse que la serie presenta ciclos. El elemento estacional refiere a la variabilidad en los datos debido a la influencia temporal. Además, se consideran variaciones de este tipo cuando las mismas se presentan en el lapso de un año y con regularidad durante el mismo. Por último, el componente irregular de una serie de tiempo es el factor residual o el que da cuenta de las desviaciones de los valores reales de la misma que se esperan al considerar los efectos de los restantes elementos. El irregular es ocasionado por factores de corto plazo, imprevistos y no recurrentes que afectan a la serie.

Definido lo anterior, se debe reconocer que algunas series contienen solo el componente tendencia, con las cuales podemos hacer inferencias causales con los datos de ellas. Esto es precisamente lo que se evidencia en las series de estudiantes universitarios en educación a distancia entre 2001 y 2015.

En el primer análisis, ya sea de estudiantes de grado, o de los nuevos inscriptos (a carreras de pregrado, grado o posgrado), nos centraremos en responder lo siguiente: ¿Qué clase de modelos estadísticos representan de manera adecuada el comportamiento de la tendencia?

Una formulación posible es escribir la serie $\{y_t\}$ como $y_t = \alpha_0 + \alpha_1 t + e_t$, en donde, en el caso más sencillo, $\{e_t\}$ es una secuencia independiente e idénticamente distribuida (i.i.d.) con $E(e_t)=0$ y $\text{Var}(e_t)=\sigma_e^2$. Se evidencia cómo el parámetro 1 multiplica al tiempo, t , lo que da como resultado una tendencia lineal. La interpretación de la ecuación es sencilla: si se mantienen fijos todos los demás factores, la ecuación mide el cambio en y_t de un periodo al siguiente debido al transcurso del tiempo: cuando $e_t=0$, $\Delta y_t/y_t=y_{t+1}=1$. Otra manera de considerar una secuencia que tiene una tendencia lineal es que su valor promedio es una función lineal del tiempo: $E(y_t)=\alpha_1 t$. Si $\alpha_1 > 0$, entonces, en promedio, y_t está creciendo en el tiempo y por consiguiente tiene una tendencia creciente. Si $\alpha_1 < 0$, entonces y_t tiene una tendencia decreciente. A diferencia de la media, la varianza



de y_t es constante al transcurrir los periodos: $\text{Var}(y_t) = \text{Var}(e_t) = \sigma^2_e$. Si $\{e_t\}$ es una secuencia i.i.d., entonces $\{y_t\}$ es una secuencia independiente, pero no es idénticamente distribuida.

Una caracterización más realista de la tendencia en las series de tiempo permite que $\{e_t\}$ esté correlacionada a través del tiempo, pero esto no modifica la cualidad de una tendencia lineal. De hecho, lo que es importante para el análisis de regresión, bajo los supuestos del modelo lineal clásico, es que $E(y_t)$ es lineal en t .

Además, se estudiarán modelos para el análisis de series de tiempo, se verá que tienen relación con lo mencionado anteriormente sobre la hipótesis de que la variable dependiente presenta componentes como tendencia, ciclos, estacionalidad e irregular.

2) 1) Métodos de suavizamiento

En primer lugar, se nombrarán tres métodos de pronóstico: promedios móviles, promedios móviles ponderados y suavizamiento exponencial. Estos métodos tienen por objeto amortiguar las fluctuaciones aleatorias ocasionadas por el componente irregular de la serie de tiempo, razón por la que se los conoce como métodos de suavizamiento. Éstos son adecuados para series de tiempo estables; es decir, para aquellas series que no muestran efectos importantes de tendencia, cíclicos o estacionales porque se adaptan muy bien a los cambios en el nivel de la serie de tiempo. Pero, sin alguna modificación, no funcionan muy bien cuando hay variaciones importantes de tendencia, cíclicas o estacionales.

Los métodos de suavizamiento son fáciles de utilizar y, por lo general, se obtiene una buena exactitud en pronósticos a corto plazo, como, por ejemplo, pronósticos para el periodo siguiente. Uno de estos métodos, el suavizamiento exponencial, tiene requerimientos mínimos de datos por lo que es un método adecuado cuando se requiere de pronósticos para un gran número de artículos.

3) 1) 1) Promedios móviles

En el método de los promedios móviles, para pronosticar el periodo siguiente, se emplea el promedio de los valores de los n datos más recientes de la serie de tiempo. El cálculo de un promedio móvil se hace como sigue: Promedio móvil = $\frac{\sum_{i=1}^n (\text{los } n \text{ datos mas recientes})}{n}$. Al elegir el método de pronóstico es importante considerar la exactitud del mismo, ya que se desea que el error de pronóstico



sea pequeño. Además de éste, el cuadrado de los errores de pronóstico, junto con el promedio de la suma de los errores al cuadrado, al cual se lo conoce como cuadrado medio debido al error (CME), suelen ser buenas medidas para la exactitud de predicción.

Se usará el CME en los análisis. Para utilizar este método, hay que decidir, primero, cuántos datos se emplearán para calcular los promedios móviles. Dado la cantidad que se poseen, que son 15, se decide utilizarlos a todos.

3) 1) 2) Promedios móviles ponderados

A cada uno de los valores de los datos se le da una ponderación diferente y, después, se calcula el promedio ponderado de los valores de los n datos más recientes para obtener el pronóstico. En la mayoría de los casos, a la observación más reciente se le da el mayor peso y éstos disminuyen conforme los datos son más antiguos.

En cuanto a la precisión del pronóstico, primero se debe establecer el número de datos a usar para calcular los promedios móviles ponderados y después elegir el peso o ponderación para cada uno de los datos. En general, si se cree que el pasado reciente sea un mejor predictor del futuro que el pasado distante, habrá que dar pesos mayores a las observaciones más recientes. Sin embargo, si la serie de tiempo es muy variable, puede ser mejor elegir pesos aproximadamente iguales para todos los datos. Es necesario notar que el único requerimiento es que la suma de los pesos sea igual a 1. Además, para estimar si con una determinada combinación de número de datos y pesos, se obtiene un pronóstico más exacto que con otra combinación, se seguirá usando el criterio de CME como medida de la exactitud del pronóstico. Es decir, si se supone que la combinación que es mejor para el pasado también será la mejor para el futuro, entonces para pronosticar el valor siguiente de la serie de tiempo se empleará la combinación de número de datos y pesos, que minimice el CME de la serie de tiempo histórica.

3) 1) 3) Suavizado exponencial

En el suavizado exponencial se utiliza un promedio ponderado de los valores pasados de la serie de tiempo. Este método es un caso especial del método de promedios ponderados móviles, sólo que en éste último hay que elegir un peso o proporción, para la observación más reciente. Las proporciones para los demás



datos se calculan automáticamente y son más pequeños a medida que los mismos son más antiguos.

A continuación se presenta el modelo de suavizamiento exponencial básico, el cual es sencillo de realizar y tiene pocos requerimientos de datos.

Modelo de suavizamiento exponencial donde: $F_{t+1} = \alpha Y_t + (1 - \alpha)F_t$

La ecuación muestra que el pronóstico para el periodo es un promedio ponderado del valor real en el periodo t y del valor pronosticado para el periodo t ; observe, en particular, que el peso dado al valor real del periodo t es α y el peso dado al valor pronosticado para el periodo t es $1-\alpha$.

En una serie de tiempo, como ejemplo, de tres datos: Y_1 , Y_2 y Y_3 se demostrará que el pronóstico obtenido mediante suavizamiento exponencial para cualquier periodo es un promedio ponderado de todos los valores reales anteriores de la serie de tiempo. Para empezar, sea F_1 igual al valor real de la serie de tiempo en el periodo 1; es decir $F_1 = Y_1$. Por tanto, el pronóstico para el periodo 2 es $F_2 = \alpha Y_1 + (1 - \alpha)Y_1$, con $F_2 = Y_1$

De tal manera que el pronóstico obtenido mediante suavizamiento exponencial para el periodo 2 es igual al valor real de la serie de tiempo para el periodo 1. El pronóstico para el periodo 3 es $F_3 = \alpha Y_2 + (1 - \alpha)F_2 = \alpha Y_2 + (1 - \alpha)Y_1$. Y así sucesivamente. Por tanto, F_3 es un promedio ponderado de los dos primeros valores de la serie de tiempo. La suma de los coeficientes o pesos de Y_1 y Y_2 es igual a uno. Mediante un argumento similar se puede demostrar que, en general, cualquier pronóstico F_{t+1} es un promedio ponderado de los valores previos de la serie de tiempo.

A pesar de que con el suavizamiento exponencial se obtiene un pronóstico que es el promedio ponderado de todas las observaciones pasadas, no es necesario conservar todos los datos pasados para calcular el pronóstico para el periodo siguiente. En efecto, una vez elegida la constante de suavizamiento α , sólo se necesitan dos informaciones para calcular el pronóstico, en el caso de F_{t+1} solo son necesarios conocer el valor real (Y_t) y el pronosticado (F_t).

Exactitud del pronóstico: para α se puede usar cualquier valor entre 0 y 1, con algunos valores se obtendrá un mejor pronóstico que con otros. Una idea de cómo elegir el mejor valor para α se obtiene al revisar el modelo básico de suavizamiento exponencial. De manera que el nuevo pronóstico F_{t+1} es igual al pronóstico anterior F_t más un ajuste, el cual es igual a α multiplicado por el error



de pronóstico más reciente. Es decir, el pronóstico para el periodo $t+1$ se obtiene al ajustar el pronóstico para el periodo t mediante una fracción del error de pronóstico. Si en la serie de tiempo existe una gran variabilidad aleatoria, se prefiere un valor pequeño para la constante de suavizamiento. La razón es que como gran parte del error de pronóstico se debe a la variabilidad aleatoria, no se quiere reaccionar de manera exagerada y ajustarlo muy rápidamente. En una serie de tiempo con variabilidad aleatoria relativamente pequeña, valores mayores para la constante de suavizado permiten ajustar rápidamente los pronósticos cuando ocurren errores en los mismos, esto permite adaptarlos, en forma rápida, a las condiciones cambiantes. El criterio que se usará para determinar el valor adecuado para la constante de α es elegir el valor α que minimice el cuadrado medio debido al error (CME).

Para definir si habrá un valor de α que dé mejores resultados en términos de un valor menor para CME, la forma más sencilla es probar otros valores para α , y después comparar los cuadrados medios de los errores con ese valor. Al probar con otros valores de α se puede hallar un “buen” valor para la constante de suavizamiento. Este valor puede emplearse en el modelo de suavizado exponencial para obtener pronósticos para el futuro.

3) 3) 1) 3) Métodos de suavizamiento exponencial

El modelo simple es adecuado para series en las que no existe tendencia ni estacionalidad, dado que el pronóstico no tiene en cuenta estos efectos. Su estructura matemática es la siguiente: $F_{t+1} = F_t + \alpha(Y_t - F_t)$

Su único parámetro de suavizado es α , que controla el peso relativo dado a las observaciones más recientes, en contraposición a la media global de la serie. Cuando α toma el valor 1, se utiliza exclusivamente la única observación más reciente: si $\alpha = 1 \Rightarrow F_{t+1} = Y_t, \text{ con } F_t = Y_{t-1}$; y así sucesivamente, hasta $t=1$ donde, $F_2 = Y_1$, el primer valor observado.

Cuando alfa toma el valor 0, las observaciones antiguas cuentan con tanto peso como las recientes: si $\alpha = 0 \Rightarrow F_{t+1} = F_t, \text{ con } F_t = F_{t-1}$; y así sucesivamente, hasta $t=1$ donde $F_2 = F_1$. Este modelo es el más similar a un modelo Autorregresivo Integrado de Media Móvil, más comúnmente como ARIMA, de órdenes $(p,d,q) = (0,1,1)$ sin constantes, que se presentará posteriormente.

Los valores pronosticados se calculan como: $\hat{y}_t(k) = F_t$ con $t < k$; siendo F_t como lo establece esta ecuación (vista anteriormente): $F_{t+1} = F_t + \alpha(Y_t - F_t)$.



Una variante del modelo simple cuando la serie presenta estacionalidad moderada, es el siguiente: $F_t = \alpha(Y_t - E_{t-s}) + (1 - \alpha)F_{t-1}$. Siendo: $E_t = \gamma(Y_t - F_t) + (1 - \gamma)E_{t-s}$; con $0 \leq \alpha \leq 1$; $0 \leq \gamma \leq 1$.

La primer ecuación corrige el valor Y_t observado en el tiempo t por la componente estacional E_{t-s} , correspondiente al período de retardo $t-s$ (ecuación siguiente). El parámetro α es como antes, es decir, el parámetro general; el parámetro γ es el parámetro estacional. Para el período k -ésimo, en general, el valor pronosticado será: $\hat{y}_t(k) = F_t + E_{t+k-s}$, con $t < k$, siendo F_t y E_t como en las anteriores ecuaciones. Este modelo considera la hipótesis aditiva para su construcción y es similar a un ARIMA $(0, 1, (1, p, p+1)) (0, 1, 0)$. Aquí p indica el número de períodos incluidos en el intervalo estacional.

El modelo de Holt con tendencia lineal es adecuado para las series en las que existe dicha tendencia pero no hay estacionalidad. El algoritmo asociado actualiza en cada salto el nivel de la serie: $\{F_t\}$ (altura que ha alcanzado) y la pendiente de la recta de tendencia: $\{b_{1t}\}$. Su estructura algorítmica es la siguiente: $F_t = \alpha Y_t + (1 - \alpha)(F_{t-1} + b_{1t-1})$, para $F_1 = Y_1 - \frac{1}{2} b_{11}$; $b_{1t} = \beta(F_t - F_{t-1}) + (1 - \beta)b_{1t-1}$, para $b_{11} = \frac{Y_N - Y_1}{N-1}$; $0 < \alpha < 1$, $0 < \beta < 1$

Sus parámetros de suavizado son α y β (este último controla el peso relativo dado a las observaciones recientes al estimar la tendencia de la serie en el presente y toma valores entre 0 y 1; los valores mayores indican un mayor peso para los valores recientes), y los valores de los mismos no se encuentran restringidos mutuamente.

El pronóstico será: $\hat{y}_t(k) = F_t + k\beta_{1,t}$, $t < k$. Siendo b_{1t} y F_t como en las anteriores ecuaciones.

Este modelo es muy similar a un modelo ARIMA(0,2,2).

El modelo de Brown de un parámetro con tendencia lineal (o modelo de ajuste doble) es adecuado para series en las que existe la misma y no hay estacionalidad. En este método, se calcula primero una suavización exponencial simple para cada valor de la serie y luego se vuelve a calcular otra sobre los datos resultantes de la primera, mediante el doble ajuste exponencial.

En todo el método se usan las siguientes ecuaciones: $F_t = \alpha Y_t + (1 - \alpha)F_{t-1}$; $F_t^T = \alpha(F_t - F_{t-1}) + (1 - \alpha)F_{t-1}^T$; donde F_t^T es la estimación de la tendencia en el período t .



El valor pronosticado es: $\hat{y}_t(k) = F_t + ((K-1) + \alpha^{-1})F_t^T$, $t < k$.

Otro modelo es el de Winters, el cual puede ser aditivo o multiplicativo. El primero (aditivo) es adecuado para las series con tendencia lineal y un patrón estacional que no depende del nivel de la serie; sus parámetros de suavizado son α , β y γ (el parámetro gamma controla el peso relativo dado a las observaciones recientes al estimar la estacionalidad del presente y toma valores de 0 a 1; los valores próximos a 1 corresponden a un mayor peso para las observaciones recientes). El método está formado por tres ecuaciones: una que ajusta el nivel de la serie, otra que ajusta la pendiente de la recta tendencia y otra que ajusta el valor de los factores estacionales.

El esquema algorítmico del modelo aditivo es el siguiente: $F_t = \alpha(Y_t - E_{t-k}) + (1 - \alpha)(F_{t-1} + b_{1,t-1})$; $b_{1,t} = \beta(F_t - F_{t-1}) + (1 - \beta)b_{1,t-1}$; $E_t = \gamma(Y_t - F_t) + (1 - \gamma)E_{t-p}$.

Siendo p el periodo estacional, E_t el componente estacional t -ésimo, $t=1,2,3,\dots,k$

El valor pronosticado será: $\hat{y}_t = F_t + kb_{1,t} + E_{t+k-p}$.

El modelo multiplicativo es adecuado para series con tendencia lineal y un efecto estacional que depende del nivel de la serie; sus parámetros de suavizado son los mismos que en el modelo aditivo. No es similar a ningún modelo ARIMA. Las ecuaciones de estimación de la tendencia, la estacionalidad y el ajuste del modelo son: $F_t = \alpha \frac{Y_t}{E_{t-k}} + (1 - \alpha)(F_{t-1} + b_{1,t-1})$; $b_{1,t} = \beta(F_t - F_{t-1}) + (1 - \beta)b_{1,t-1}$, $E_t = \gamma \frac{Y_t}{F_t} + (1 - \gamma)E_{t-p}$; siendo E_t los componentes estacionales, con $t = 1, \dots, n$.

El pronóstico resulta: $\hat{Y}_t(k) = (F_t + kb_{1t})E_{t+k-p}$.

El modelo propuesto por Holt y ampliado por Winters, que es conocido como alisado exponencial lineal con doble parámetro o técnica de Holt-Winters, cuyo objetivo es eliminar el sesgo en la predicción de series con tendencia aproximadamente lineal (mismo objetivo que el de Brown). Ahora, la propia media móvil incorpora ya un componente de tendencia: $F_t = \alpha y_t + (1 - \alpha)(F_{t-1} + \beta_{t-1})$. Donde β es un factor de variación definido a partir de otra nueva constante de alisado para la tendencia: $\beta_t = b(F_t - F_{t-1}) + (1 - b)\beta_{t-1}$.

La ecuación de predicción será: $\hat{Y}_{t+h} = F_t + \beta_t h$.



Otro modelo es el Damped, que se utiliza para amortiguar la serie con tendencia lineal (y con estacionalidad). Éste surge de pensar modificaciones del modelo de Brown. Algunos autores recomiendan la tendencia amortiguada, como un método de pronóstico que debe mejorar la precisión en las aplicaciones prácticas. Gardner Jr. (2006) propone un parámetro (Φ) de amortización para modificar la tendencia en el método de tendencia lineal de Holt. El resultado es un método fijo en primeras diferencias, en lugar de segundas diferencias como en el de Holt.

Así, y siempre considerando la hipótesis inicial en estos modelos que es la existencia de ruido blanco, resulta: $F_t = \alpha Y_t + (1 - \alpha)(F_{t-1} + \phi b_{t-1})$, y además, $b_t = \beta(F_t - F_{t-1}) + (1 - \beta)\phi b_{t-1}$.

Aquí $0 \leq \phi \leq 1$, es el parámetro de amortiguamiento de la tendencia. Si $\phi = 1$, se considera que la tendencia es altamente persistente, si $\phi = 0$, se implicaría la ausencia de tendencia.

El valor pronosticado será: $\hat{Y}_t(k) = F_t + \sum_{i=1}^k \phi^i b_t$

Este modelo es similar funcionalmente a un ARIMA (1,1,2).

3)2) Modelado de ciclos: modelos de promedios móviles (MA), modelos autorregresivos (AR), modelos de promedios móviles autorregresivos (ARMA), ARIMA y SARIMA

Al formar modelos de pronósticos se pretende que sea acercarse a una realidad más compleja. Por ello, el modelado y pronóstico de series temporales es la aproximación parsimoniosa y a la vez exacta de la representación de Wold.

Teorema (Wold): Cualquier proceso estacionario $\{y_t\}$ con media cero se puede representar de la forma $y_t = B(L)\varepsilon_t = \sum b_i \varepsilon_{t-i}$, con $\varepsilon_t \sim WN(0, \sigma^2)$, en donde $b_0 = 1$ y $\sum b_i^2 < \infty$.

En resumen, el modelo correcto para cualquier serie de covarianza estacionaria es un rezago distribuido de ruido blanco, llamado Representación de Wold. Con frecuencia, a las ε_t se las denomina innovaciones porque corresponden a los errores de correlación a una etapa, que cometeríamos si usáramos un pronóstico bueno. Esto es, las ε_t representan la parte de la evolución de y , linealmente impredecible con base en el pasado de la variable endógena. También las ε_t , aunque no están correlacionadas, no necesariamente son independientes. De nuevo, la falta de correlación solo implica independencia cuando las variables



son aleatorias distribuidas independiente e idénticamente como normales, y las innovaciones no son necesariamente así.

En el análisis, supusimos promedio cero, dado que ello facilita la lectura por el ahorro de notación y además, no implica pérdida de generalidad.

Otro término mencionado es el de Ruido Blanco, el cual es un modelo especial de proceso estocástico (sucesión de variables aleatorias). Éste es el componente impredecible del mismo. El proceso $\{t\}$ es ruido blanco y representa el error resultante de predecir y_t con una función lineal de los retardos de y_t : $t = y_t - E(y_t | y_{t-1}, y_{t-2}, \dots)$

Además para definir el ruido blanco, las propiedades de ciertos procesos de serie de tiempo resultan importantes para pronosticar. Antes de estimarlos necesitamos comprender las propiedades de las poblaciones, suponiendo que el modelo postulado es válido. Para representar la serie observada de interés usaremos y .

Supongamos que: $y_t = \varepsilon_t$, con $\varepsilon_t \sim (0, \sigma^2)$, en la que el “choque” ε_t no está correlacionado en el tiempo. Se dice que ε_t y entonces también, y_t es no correlacionado en serie, o seriadamente no correlacionado. Ese proceso, con promedio cero varianza constante y sin correlación en serie, se llama ruido blanco de (o con) promedio cero, o simplemente ruido blanco. A veces, se escribe: $\varepsilon_t \sim WN(0, \sigma^2)$, y entonces $y_t \sim WN(0, \sigma^2)$.

Aclarados ya los conceptos mencionados, procederemos a explicar los métodos de modelado de ciclos que luego se utilizarán para un caso concreto.

En análisis de serie del tiempo, el modelo de medias móviles (MA) es una aproximación común para series de tiempo univariadas. El MA especifica que la variable de salida depende linealmente del valor actual y varios de los anteriores de un término estocástico.

Junto con el modelo autorregresivo (AR), MA es un caso especial y componente clave de los modelos de serie de tiempo más generales ARMA y ARIMA, los cuales tienen una estructura estocástica más compleja.

Contrariamente a los modelos autorregresivos, el MA es siempre estacionario.

3) 2) 1) Modelos de promedios móviles (MA)



El proceso de promedios móviles es un tipo de proceso estocástico, que permite modelar las series de tiempo que presentan rezagos distribuidos de choques actuales y pasados.

La característica que define el proceso MA en general y el promedio móvil de primer orden en particular, es que el valor actual de la serie observada se expresa como función de choques actuales y rezagados inobservables, como si fuera un modelo de regresión solo con perturbaciones actual y retrasadas en el lado derecho.

El promedio móvil de primer orden, o proceso MA(1) es: $y_t = \varepsilon_t + \theta\varepsilon_{t-1} = (1 + \theta L)\varepsilon_t$, siendo θ un factor de ponderación, que se encuentra entre 0 y 1, donde $\varepsilon_t \sim WN(0, \sigma^2)$, es decir, ε_t es el ruido blanco.

Esta función presenta un solo valor no nulo, el correspondiente al primer retardo, mientras que su función de autocorrelación parcial decrece exponencialmente a cero, es decir, solo aparece la primer demora del choque, y tiene un choque actual y choques inobservables rezagados hacia la derecha.

Por otro lado, podemos ver que un proceso de medias móviles de orden 2, MA(2), es un proceso estocástico que sigue la función: $y_t = \delta + \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2}$.

En este caso, la función de autocorrelación simple (o solamente función de autocorrelación) es cero para valores $k > 2$, mientras que su función de autocorrelación parcial disminuye exponencialmente a cero.

Por lo tanto es fácil afirmar que la identificación de los modelos se realiza principalmente analizando sus funciones de autocorrelación y de autocorrelación parcial. Podemos ver que la identificación de modelo MA (2) mediante su función de autocorrelación parcial (FAP) es difícil, lo mismo ocurre con la simple (FAS), pero es muy sencillo hacerlo con la función AR(2) por medio de la FAP o de MA(2) por medio de la FAS. En la MA(2) si la FAS es cero para $k > 2$ al igual que en la AR(2). Por otra parte, estos resultados son perfectamente generalizables para modelos de orden superior, AR(p) y MA(q), con p, q enteros positivos cualquiera.

La función de autocorrelación simple (FAS) de un proceso estocástico, es una función que para cada instante t y cada entero k toma un valor $p_k(t)$ igual a la correlación entre y_t e y_{t-k} . La función de autocorrelación parcial (FAP) de un proceso estocástico es una función que para cada instante t y cada entero k toma



un valor igual a la correlación entre y_t e y_{t-k} , ajustada por el efecto de los retardos intermedios $y_{t-1}, y_{t-2}, \dots, y_{t-k-1}$. El concepto de estacionalidad permite deducir algunas propiedades de estas funciones (un proceso estacionario es aquel en que la esperanza, la varianza, las covarianzas y las correlaciones entre diversos retardos de la variable bajo análisis son invariantes en el tiempo).

Lo más común es que en el análisis empírico de series temporales se encuentren representaciones que tienen un componente autorregresivo así como un componente de medias móviles. Estos modelos se denotan como modelos ARMA(p,q), donde p y q son los órdenes de los componentes autorregresivo y de medias móviles, respectivamente. Dicho modelo será explicado ulteriormente.

3) 2) Modelo autorregresivo o AR

El proceso autorregresivo también es una aproximación natural a la representación de Wold. Por otro lado, al igual que el MA, tiene autocorrelación simple y autocorrelación parcial, solamente que en el AR tenemos una ecuación en diferencias estocástica. Un proceso estocástico es un conjunto ordenado (sucesión) de variables aleatorias que depende de un indicador. En el caso de las series temporales el indicador es el tiempo t. También se puede decir que un proceso estocástico es un modelo matemático de valor actual de una serie que se encuentra linealmente relacionado con sus valores en otro modelo que es estocástico aditivo.

El proceso autorregresivo de primer orden, AR(1), es: $y_t = \phi y_{t-1} + \varepsilon_t$, donde $\varepsilon_t \sim WN(0, \sigma^2)$

En su forma de operador de rezago se escribe: $(1 - \phi L)y_t = \varepsilon_t$. Operando algebraicamente, llegamos a que $y_t = \phi y_{t-1} + \varepsilon_t$. Comprobando así que tiene sentido la representación en el operador de rezago.

Se debe destacar que la disminución gradual de la autocorrelación, característica de proceso autorregresivos, tienden a cero pero solo en el límite cuando el desplazamiento tiende a infinito. En particular, no se cortan y bajan a cero, como en el caso de los procesos de promedios móviles. Si ϕ es positiva, el decaimiento de la autocorrelación es unilateral (aumenta o disminuye continuamente); si es negativa, el decaimiento implica oscilaciones.



Por otro lado, es necesario mencionar que el modelo de medias móviles es esencialmente un filtro de respuesta de impulso finito aplicado a ruido blanco, al que se ha añadido una interpretación adicional.

El papel de la función de los shocks aleatorios en el modelo MA difiere de su función en el autorregresivo (AR) de dos maneras: están propagados a valores futuros de la serie de tiempo directamente, por ejemplo, ε_{t-1} aparece directamente en el lado correcto de la ecuación para X_{t-1} ; y, por otro lado, en el modelo MA, un shock afecta a valores X solamente para el período actual y q a períodos futuros; en contraste, en el modelo AR, un shock afecta valores futuros infinitamente lejanos, porque ε_t afecta a X_t , el cual afecta a X_{t+1} , X_{t+2} , y así sucesivamente.

3) 2) 3) Promedios móviles autorregresivos (ARMA)

El proceso ARMA tiene una motivación directa: si el choque aleatorio que impulsa un proceso autorregresivo es a su vez un proceso de promedios móviles, se puede demostrar que se obtienen dichos procesos. Además, éste se puede originar por agregación. Por último, el proceso AR observado, sujetos a error de medición, también es ARMA. Estos modelos a menudo se combinan con frecuencia para tratar de obtener mejores aproximaciones y más parsimoniosas a la representación de Wold, y producen el proceso de promedios móviles autorregresivos o ARMA (p,q) .

El proceso ARMA más simple, que no es una autorregresión pura ni un proceso puro de promedios móviles, es el proceso ARMA $(1,1)$, definido como: $y_t = \varphi y_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}$, con $\varepsilon_t \sim (0, \sigma^2)$; o, como operador de demora: $(1 - \varphi L)y_t = (1 + \theta L)\varepsilon_t$.

En donde se requiere que para que haya estabilidad y Θ menor a uno en valor absoluto para la existencia de invertibilidad (tener un choque actual y uno rezagado, e invertirlo para obtener valores rezagados de la serie y otro choque actual).

El proceso para ARMA (p,q) es una generalización del anterior (ARMA $(1,1)$), que permite rezagos de promedios móviles múltiples y autorregresivos.

Se escribe: $y_t = \varphi_1 y_{t-1} + \dots + \varphi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$ con: $\varepsilon_t \sim WN(0, \sigma^2)$; o bien, $\varphi(L) y_t = \Theta(L)\varepsilon_t$.

La función de autocorrelación de este proceso se comporta, a partir de $k=1$, como la función de autocorrelación de un proceso AR(1).



Esta propiedad puede generalizarse: la función de autocorrelación (FAC) de un proceso ARMA (p,q) se comporta como una FAC de un proceso AR(p), a partir de $k > q$. Pero la identificación de estos modelos por inspección de la serie temporal correspondiente resulta menos sencilla que en casos de modelos AR(p) y MA(q) puros, dado que normalmente no se ajusta a unas normas tan bien definidas. Ello se debe a que tanto la función de autocorrelación como la de autocorrelación parcial (FACP) de los modelos ARMA heredan características de ambos componentes.

De todas maneras, como se llega al modelo que más se ajusta a la serie de tiempo bajo análisis luego de varios intentos, no hace falta obtener la mejor identificación del modelo en un primer momento. Es decir, lo más probable es que se comience probando con un modelo que, al analizar las funciones de autocorrelación y autocorrelación parcial de la serie, se considera el adecuado. Pero, luego de realizar el chequeo de los residuos, cabe la posibilidad de que notemos que las funciones de autocorrelación y autocorrelación parcial correspondientes tengan un comportamiento conocido, por lo cual deberá realizarse una corrección del modelo utilizado.

3) 2) 4) Modelos ARIMA (autorregresivo integrado de media móvil)

Cuando se mencionó anteriormente, un modelo ARMA(p,q) para una variable diferenciada se dice que se tiene un modelo ARIMA(p,d,q). El parámetro d indica la cantidad de veces que se ha diferenciado la variable bajo análisis, entonces ARIMA es un caso especial de ARMA diferenciado d veces.

Por otro lado, el ARIMA permite describir un valor como una función lineal de datos anteriores y errores debidos al azar, y además, es posible incluir un componente cíclico o estacional. La estacionariedad exige que se cumplan dos requisitos: una media aproximadamente constante en el tiempo y una varianza o dispersión también constante (propiedad de homocedasticidad). Esto implica que si dividimos una serie en subperíodos arbitrarios, la media y la varianza son aproximadamente iguales en cada uno de ellos.

Al desarrollar la descomposición de las series temporales, se evidencia que las series pueden presentar características estacionales cuando éstas tienen una frecuencia inferior a la anual. La especificación de las características estacionales de una serie puede ser del tipo: $y_t = \Phi_{12}y_{t-12} + e_t$.



La determinación de la estructura estacional de una serie temporal se lleva a cabo de modo similar al descrito para la identificación de modelo ARMA, con la salvedad que para ello se examinan los valores estacionales de las funciones de autocorrelación, correspondientes a los retardos 4, 8, 12... si los datos son trimestrales, ó 12, 24, 36... si son mensuales.

3) Construcción del modelo ARIMA

Habiendo explicado los modelos, se procederá a describir la aplicación práctica del modelo ARIMA. Para construir un modelo, primero se debe obtener, utilizar y analizar alguna información de los datos en cuanto a cómo fue generada la serie, para así identificar el modelo a utilizar, estimar los resultados del mismo y validar dichos resultados.

3) 3) 1) La primera etapa en la construcción de un modelo ARIMA corresponde al análisis del gráfico de la serie.

La función de autocorrelación de los procesos AR tiene el inconveniente de que, si bien decrece a partir de determinado orden, nunca se hace cero. De ahí que sea difícil muchas veces identificar el orden de un proceso AR a partir de la FAC. Los coeficientes de autocorrelación parcial, y la función de autocorrelación parcial (FACP) están diseñados especialmente para facilitar la identificación de procesos AR.

Para el cálculo de la FACP estimada, el coeficiente de autocorrelación parcial de primer orden coincide con el coeficiente de autocorrelación "natural" de primer orden.

El coeficiente de autocorrelación parcial de segundo orden se calcula como el coeficiente de autocorrelación entre y_t y y_{t-2} , cuando se ha eliminado de ambas variables el efecto de y_{t-1} .

En general, el coeficiente de autocorrelación parcial de orden "k" se calcula como el coeficiente de autocorrelación entre y_t y y_{t-k} , cuando se ha eliminado de ambas variables el efecto de $y_{t-1}, y_{t-2}, \dots, Y_{t-k+1}$.

En términos prácticos, para calcular el coeficiente de segundo orden se realiza una regresión entre y_t y y_{t-1} . Y luego se calcula el coeficiente de correlación entre los residuos de dicha regresión y y_{t-2} .

El coeficiente de tercer orden surge de la correlación entre los residuos de la regresión de y_t versus y_{t-1} y y_{t-2} , por un lado, y y_{t-3} , por otro.

Por otro lado, a los efectos de llevar adelante la estimación del modelo, debe decidirse sobre la inclusión de un término independiente. Es decir, si la serie una vez transformada en estacionaria tiene media cero o distinta de cero. Incluir o no



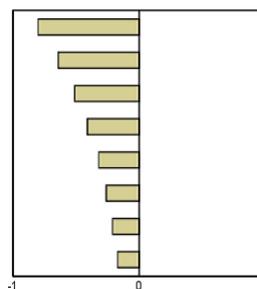
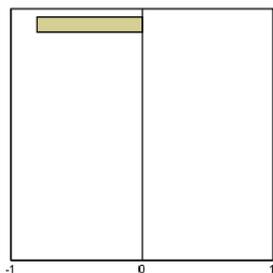
una constante puede resultar un punto crucial a los resultados del modelo, dado que cuando la serie se diferencia, añadirla significa que la serie original sigue un patrón con tendencia.

El problema que esto ocasiona es menor, ya que siempre puede especificarse el modelo con término independiente y luego realizar pruebas de significación de este parámetro, para decidir su inclusión o no.

En el análisis que se realizará, se utilizarán los siguientes gráficos para la identificación, de las funciones de autocorrelación y autocorrelación parcial. Para interpretar los gráficos de las funciones mencionadas se tendrán en cuenta los esquemas de procesos ARIMA puros o teóricos para ambas funciones.

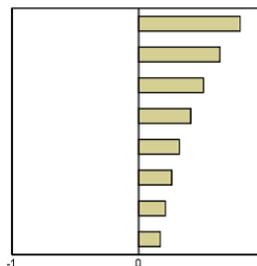
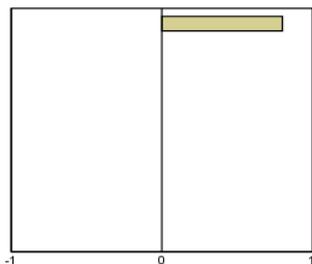
ARIMA(0,0,1), $\theta > 0$

FAS FAP



ARIMA(0,0,1), $\theta < 0$

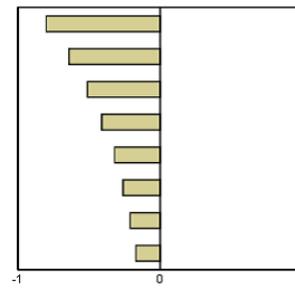
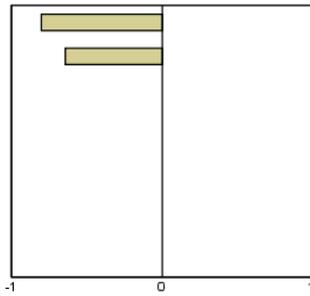
FAS FAP





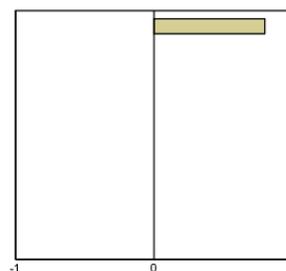
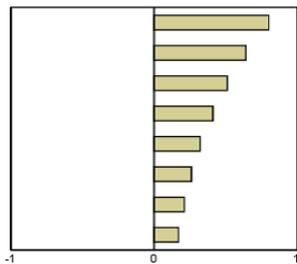
ARIMA(0,0,2), $\theta_1 \theta_2 > 0$

FAS FAP



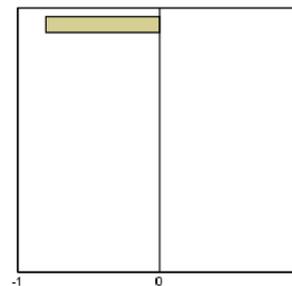
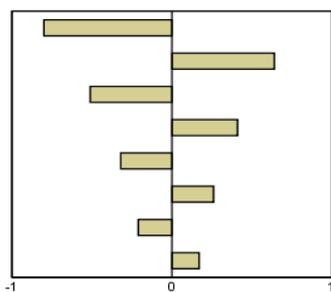
ARIMA(1,0,0), $\phi > 0$

FAS FAP



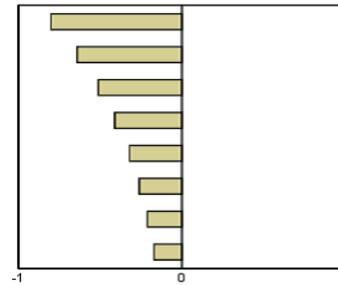
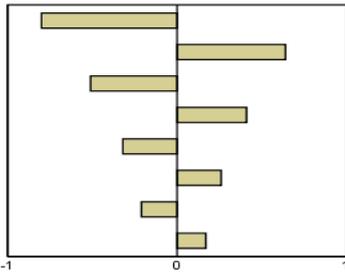
ARIMA(1,0,0), $\phi < 0$

FAS FAP



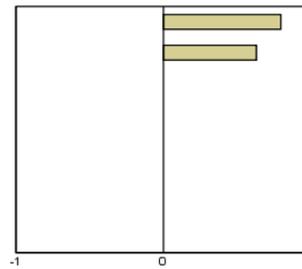
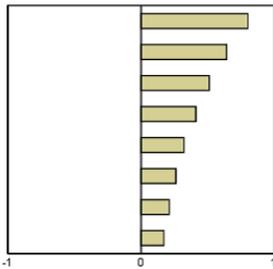
ARIMA(1,0,1), $\phi < 0, \theta > 0$

FAS FAP



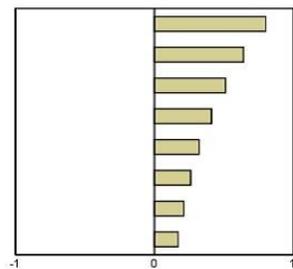
ARIMA(2,0,0), $\phi_1 \phi_2 > 0$

FAS FAP



ARIMA(0,1,0) (serie integrada)

FAS



Fuente: Manual SPSS Trends™ 14.0.

El segundo paso de la estimación consiste en el uso eficiente de los datos para realizar inferencias acerca de los parámetros del modelo adecuado.



Y, por último, la validación permite verificar si el modelo construido es el que mejor describe el comportamiento de los datos o si es inadecuado, en cuyo caso es preciso mejorarlo, o de no ser posible, descartarlo por completo.

3) 4) Estadístico de Box- Ljung

El estadístico de Box- Ljung se utiliza con el objetivo de validar lo observado gráficamente, se somete a prueba de aleatoriedad los rezagos de autocorrelación observados mediante la siguiente prueba (basados en una distribución chi-cuadrado).

Este test tiene dos particularidades: a) se basa en el conjunto de autocorrelación, y b) provee una prueba general (y no en particular de pares de lags o retardos). Su formulación es la siguiente:

H₀: los retardos son aleatorios; **H₁**: los retardos no son aleatorios.

El estadístico del test es: $Q_{BL} = (n(n+2)) \sum_{i=1}^k \frac{r_i^2}{n-i}$. Donde: n es el tamaño muestral, r_j^2 es la autocorrelación del j-ésimo retardo, y k es el número de rezagos testados. Regla de Decisión: Rechace H_0 si $Q_{BL} > \chi^2_{(1-\alpha), k}$

4) Referencias

- Box & Jenkins(1976). Time Series Analysis.Forecasting and Control. Edición revisada, Holden Day, San Francisco, Estados Unidos.
- Croxton y Cowden(1942). Estadística General Aplicada. Fondo de Cultura Económica, México.
- Diebold, F. (1999). Elementos de pronósticos. S.A. Ediciones Paraninfo. México.
- Jaramillo Ayerbea, M., González Gómez, D. E., Nuñez Cabrera, M.E., PORTILLA, G.E. y García, J. H. L. (2007). Análisis de series de tiempo univariante aplicando metodología de Box-Jenkins para la predicción de ozono en la ciudad de Calí, Colombia. Revista Facultad de Ingeniería Universidad de Antioquia, Medellín, Colombia.
- Kolb, D. (1984). Experiential Learning.Experience as the source of learning and Development. New Jersey: Prentice Hall P T R, Englewood Cliffs.



- Levin, R. y Rubin, D. (2004). Estadística para administración y economía. Editorial Prentice Hall. Séptima edición. México.
- Wooldridge, J. (2001). Introducción a la Econometría. Ed. Thomson, México.