



SERIES TEMPORALES

Cecilia Esparza Catalán

ÍNDICE

| | Página |
|---|---------------|
| 1.- INTRODUCCIÓN..... | 2 |
| 2.- ANÁLISIS PRELIMINAR DE UNA SERIE..... | 3 |
| - Tendencia y nivel de la serie..... | 4 |
| - Estacionalidad..... | 9 |
| - Dependencia entre variabilidad y nivel..... | 13 |
| - Comportamientos anómalos..... | 16 |
| 3.- MÉTODOS CLÁSICOS..... | 17 |
| - Métodos de descomposición estacional y ajuste de tendencia..... | 18 |
| - Métodos de suavizado..... | 30 |
| - SERIES SIN TENDENCIA NI ESTACIONALIDAD..... | 31 |
| - Modelos “naive” o ingenuos..... | 31 |
| - Modelos de medias móviles..... | 33 |
| - Modelos de suavizado exponencial simple..... | 34 |
| - SERIES CON TENDENCIA Y SIN ESTACIONALIDAD..... | 38 |
| - SERIES CON TENDENCIA Y CON ESTACIONALIDAD..... | 41 |
| 4.- MODELOS ARIMA..... | 47 |
| - Modelos de medias móviles (MA)..... | 50 |
| - Modelos autorregresivos (AR)..... | 51 |
| - Modelos autorregresivos de medias móviles (ARMA)..... | 52 |
| - Modelos no estacionarios (ARIMA)..... | 52 |
| BIBLIOGRAFÍA..... | 53 |

SERIES TEMPORALES

1.- INTRODUCCIÓN¹

Una serie temporal es una colección de observaciones de una variable realizadas de forma secuencial en el tiempo, en las que el orden de observación es importante. Los valores de una serie temporal van ligados a instantes de tiempo, de manera que el análisis de una serie implica el manejo conjunto de dos variables; la variable en estudio propiamente dicha y la variable tiempo.

Las series pueden tener una periodicidad anual, semestral, trimestral, mensual, etc., según los periodos de tiempo en los que están recogidos los datos que la componen. Las ventas trimestrales de una empresa, el número de casos mensuales de personas afectadas por el SIDA, la cantidad de accidentes semanales de tráfico o el número de exportaciones efectuadas cada año por un determinado país, son ejemplos de series temporales con diferentes periodicidades.

El análisis de series temporales presenta un conjunto de técnicas estadísticas que permiten, además de estudiar y modelizar el comportamiento de un fenómeno que evoluciona a lo largo del tiempo, realizar previsiones de los valores que se alcanzarán en el futuro.

Con el análisis de series temporales se pretende extraer las regularidades que se observan en el comportamiento pasado de la variable, es decir, obtener el mecanismo que la genera, para tener un mejor conocimiento de la misma en el tiempo. Además, bajo el supuesto de que las condiciones estructurales que conforman la serie objeto de estudio permanecen constantes, también se trata de predecir el comportamiento futuro.

El estudio de una serie temporal puede tener distintas motivaciones. El tipo de análisis, así como los modelos en los que basemos el estudio, dependerán en gran medida del tipo de preguntas que queramos responder. Cuando las observaciones corresponden a una única variable, el análisis de series temporales suele tener como objetivo construir un modelo para explicar la estructura (descripción) y prever la evolución (predicción) de la variable de interés. Cuando se observa simultáneamente un grupo de variables, el objetivo que se persigue consiste generalmente en analizar las posibles relaciones entre las variables observadas y su evolución conjunta. Este último punto, que debe ser tratado mediante técnicas multivariantes de análisis de series temporales, no queda cubierto por el presente manual, solamente nos referiremos al análisis univariante de series temporales, al análisis de una única serie.

¹ OBSERVACIÓN: Todos los contenidos del presente manual han sido elaborados empleando la versión 14.0 del programa SPSS.

2.- ANÁLISIS PRELIMINAR DE UNA SERIE

El análisis preliminar de una serie constituye el primer paso a seguir a la hora de estudiar una serie temporal. Esta fase nos permite detectar las características más importantes de una serie, tales como su tendencia (creciente o decreciente), la existencia de ciclos, presencia de valores atípicos, etc.

La forma más sencilla de comenzar el análisis de una serie temporal es mediante su representación gráfica. El gráfico que se emplea para representar las series temporales es el gráfico de secuencia. Los gráficos de secuencia son diagramas de líneas en los cuales el tiempo se representa en el eje de abscisas (x), y la variable cuya evolución en el tiempo estudiamos en el eje de ordenadas (y).

Para obtener el gráfico de secuencia de una serie temporal empleando el programa SPSS debemos seleccionar el menú *Gráficos – secuencia...* En la casilla *Variables* del cuadro de diálogo debemos introducir la variable a estudiar y en la correspondiente a *Etiquetas del eje del tiempo* la variable que nos indica la fecha a la que corresponde cada observación.

En el caso de que en el archivo de SPSS tengamos la variable de interés pero no la correspondiente al tiempo, debemos generarla. Para ello basta con ir al menú *Datos – Definir fechas* y seleccionar la adecuada.

Veamos un ejemplo de gráfico de secuencia obtenido con los datos de la serie temporal *Airline*, serie que representa el volumen de pasajeros de una compañía aérea estadounidense.

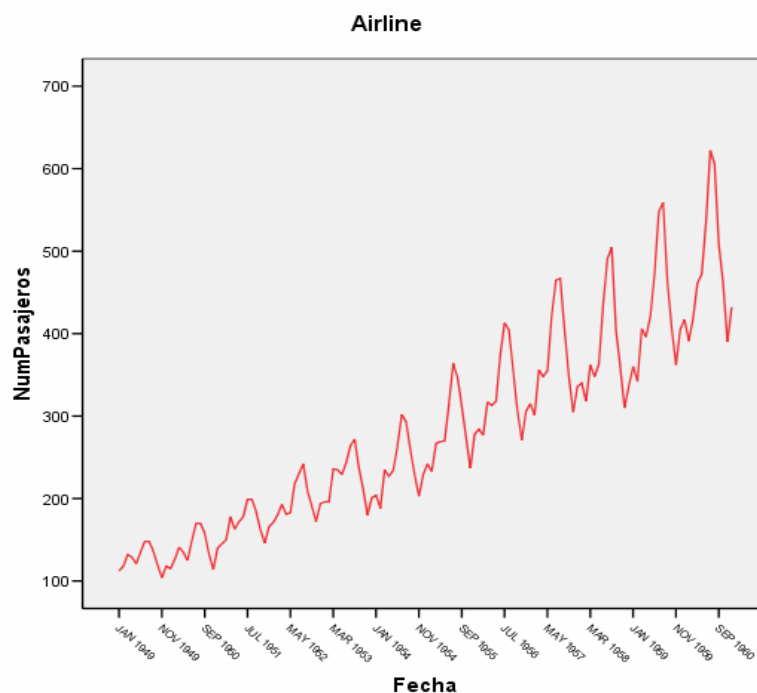


Figura 1: gráfico de secuencia de la serie Airline.

Para realizar el análisis preliminar de una serie debemos estudiar las siguientes cuestiones:

Tendencia y nivel de la serie

El nivel de una serie es una medida local de tendencia central, como por ejemplo la media, de cada periodo de tiempo que consideremos. Cuando trabajamos con un calendario (tiempo representado en días, meses o años), no es recomendable establecer periodos de tiempo antinaturales para estudiar esta característica.

Una vez calculado el nivel de la serie debemos observar su estabilidad, debemos ver si la medida de tendencia central elegida tiene valores similares en los periodos de tiempo en los que dividimos el calendario. También tendremos que observar su tendencia, es decir, si presenta una dirección constante de cambio de nivel.

Estudiamos el nivel de la serie de nuestro ejemplo, la serie Airline. Las observaciones del número de pasajeros son mensuales, luego consideraremos periodos de tiempo de un año.

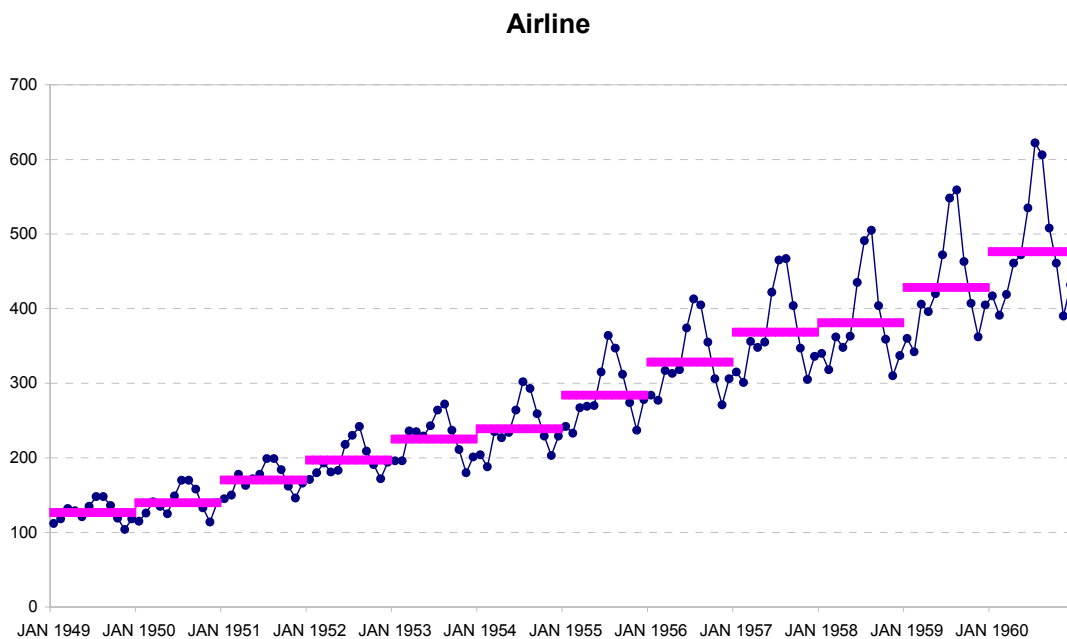


Figura 2: gráfico de secuencia de la serie Airline y su nivel.

Una vez representado el nivel de la serie lo estudiamos para determinar si es o no estable. En el caso de la serie Airline vemos que no lo es, la media del número de pasajeros contabilizados cada año (periodo elegido para dividir el calendario) va aumentando progresivamente, de manera que presenta tendencia creciente.

Otra forma de estudiar el nivel de una serie es realizando un box-plot² (diagrama de caja) de cada uno de los periodos de tiempo considerados. Para efectuarlo con el programa SPSS debemos seleccionar el menú *Gráficos – Diagramas de caja... - Simple* y en la casilla *Variable* incluimos la variable de la serie y en la correspondiente a *Eje de categorías* la referente al periodo considerado.

La línea que aparece en el interior de cada una de las cajas de un box plot representa la mediana, es decir, el nivel de la serie, puesto que es una de las posibles medidas de tendencia central. De esta manera también podemos analizar gráficamente el nivel de una serie.

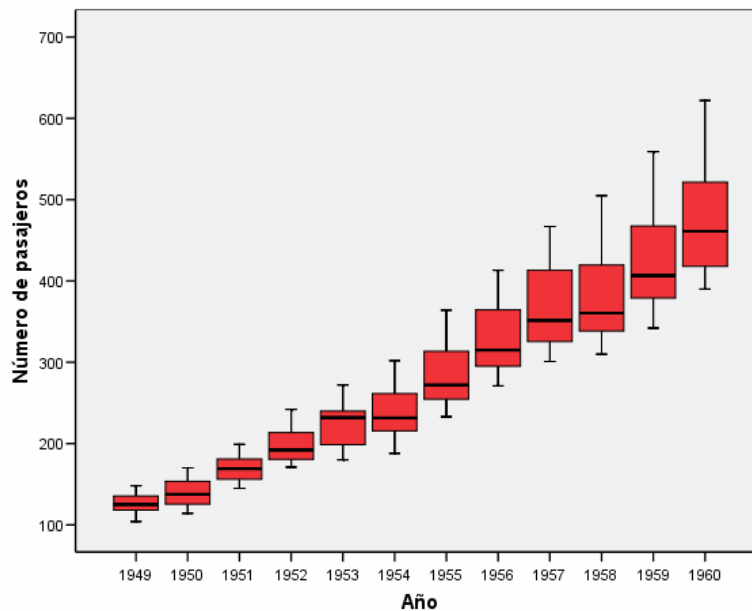
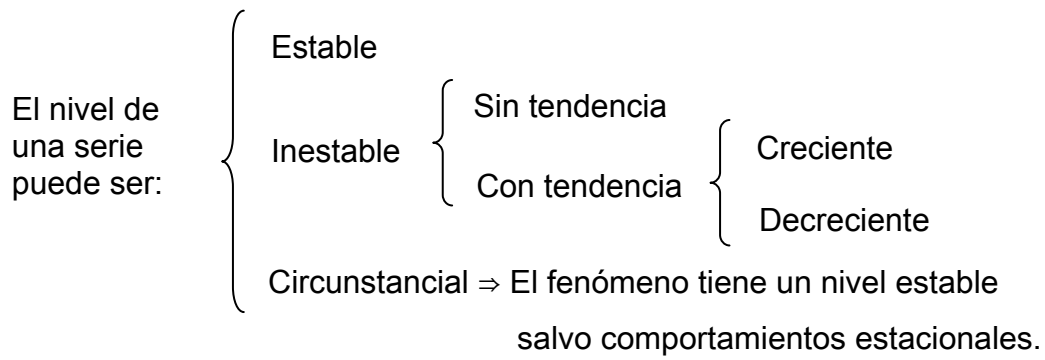


Figura 3: diagramas de cajas por año de la serie Airline.

Al igual que en el caso anterior, donde representábamos la media de la serie Airline, llegamos a la conclusión de que el nivel es inestable y presenta tendencia creciente.

² El diagrama de caja (box-plot) de una variable consiste en un rectángulo (caja), cuyo extremo inferior viene determinado por el primer cuartil de su distribución, y el superior por el tercero. Este rectángulo está dividido por un segmento que representa a la mediana (segundo cuartil) y su longitud nos indica el rango en el que está concentrado el 50% de los datos. Las dos líneas (llamadas bigotes) que parten de los extremos del rectángulo intentan alcanzar los valores mínimo y máximo, pero su longitud no puede superar 1,5 veces el rango intercuartílico. Si existe algún valor que queda fuera de los bigotes se representa a parte, como un valor extraño.



Veamos nuevos ejemplos de estudio del nivel y la tendencia de una serie.

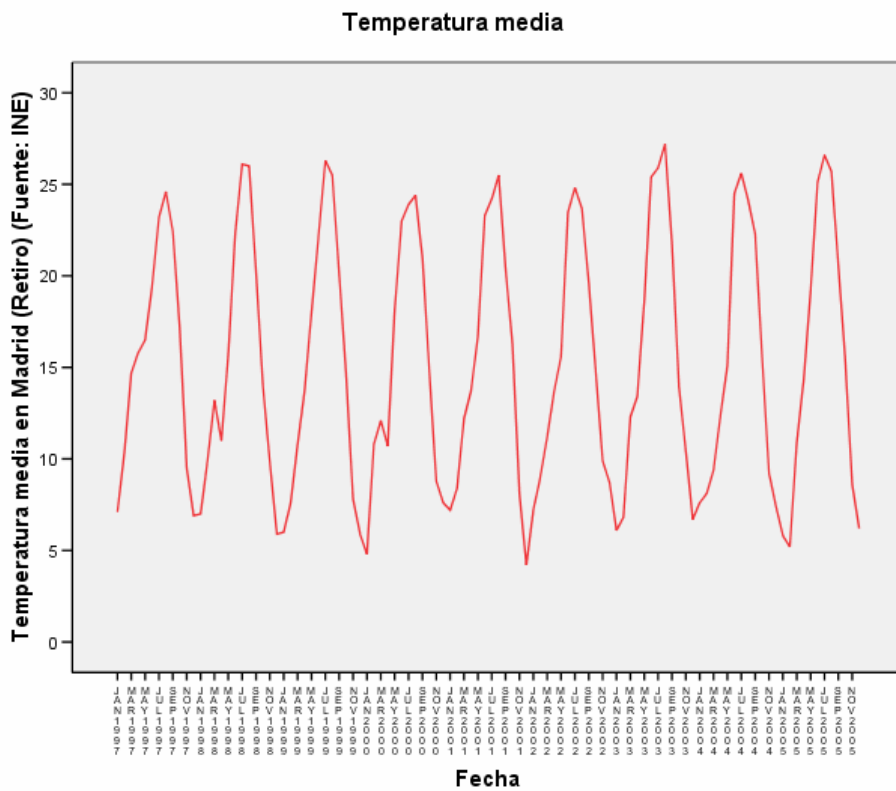


Figura 4: gráfico de secuencia de la serie Temperaturas.

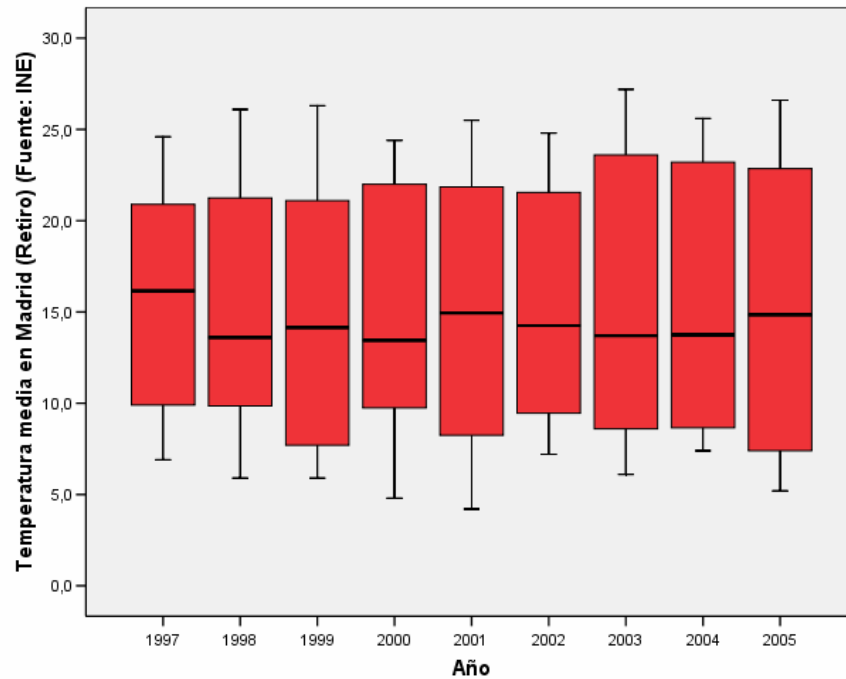


Figura 5: diagramas de cajas por año de la serie Temperaturas.

La serie Temperaturas, que representa la temperatura media registrada en Madrid mes a mes, tiene un nivel estable, es decir, es una serie estacionaria. Observando el gráfico de secuencia vemos que cada año se repite un patrón muy similar y, estudiando el diagrama de box-plot, vemos que la mediana de cada año tiene valores muy parecidos.

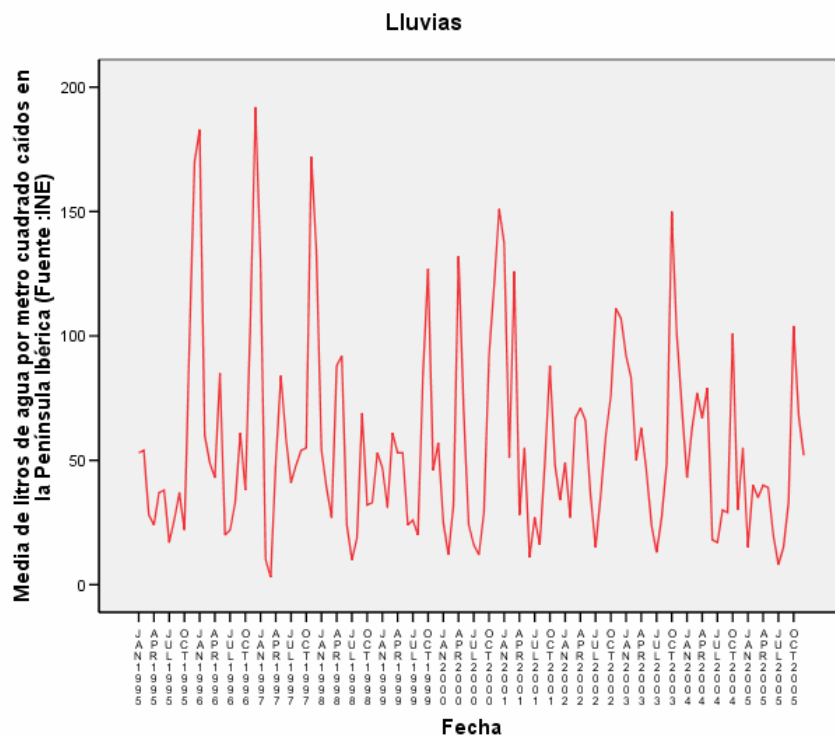


Figura 6: gráfico de secuencia de la serie Lluvias.

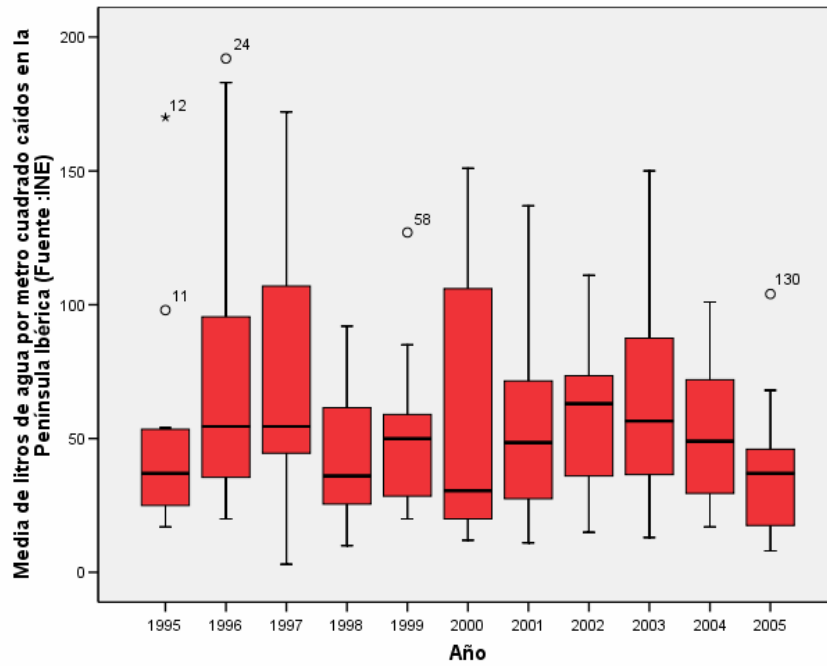


Figura 7: diagramas de cajas por año de la serie Lluvias.

En el caso de la serie Lluvias, que está formada por los valores correspondientes la media mensual del número de litros de agua caídos por metro cuadrado en la Península Ibérica, encontramos un nivel inestable y sin tendencia.

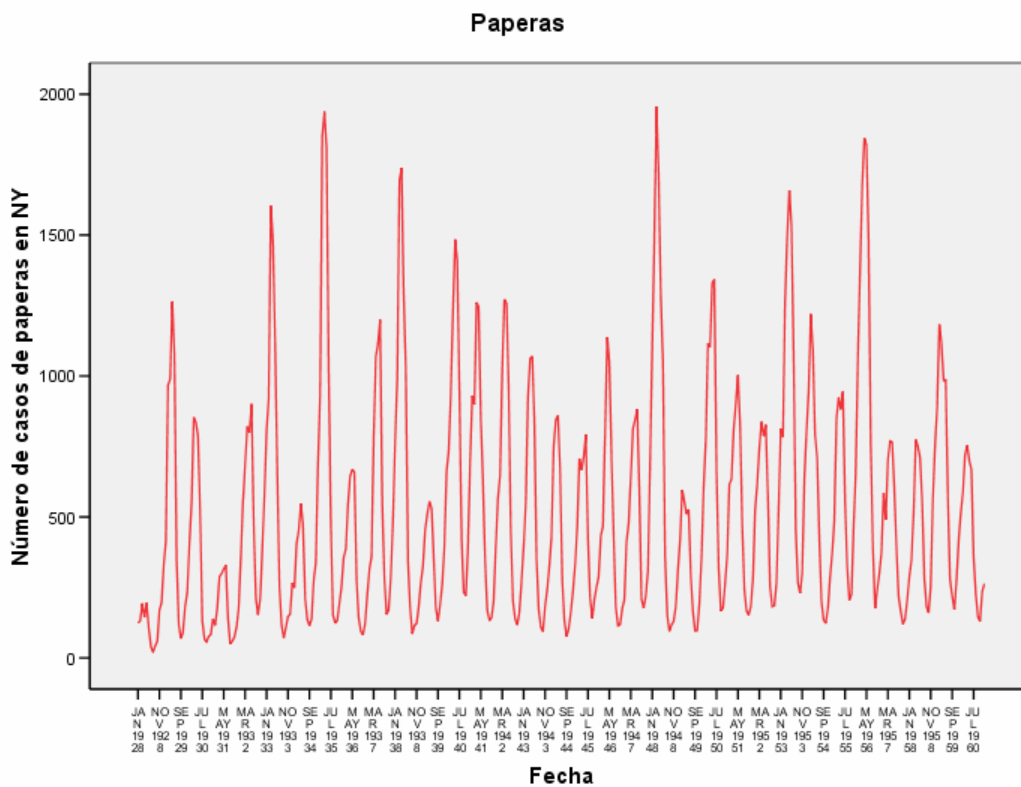


Figura 8: gráfico de secuencia de la serie Paperas.

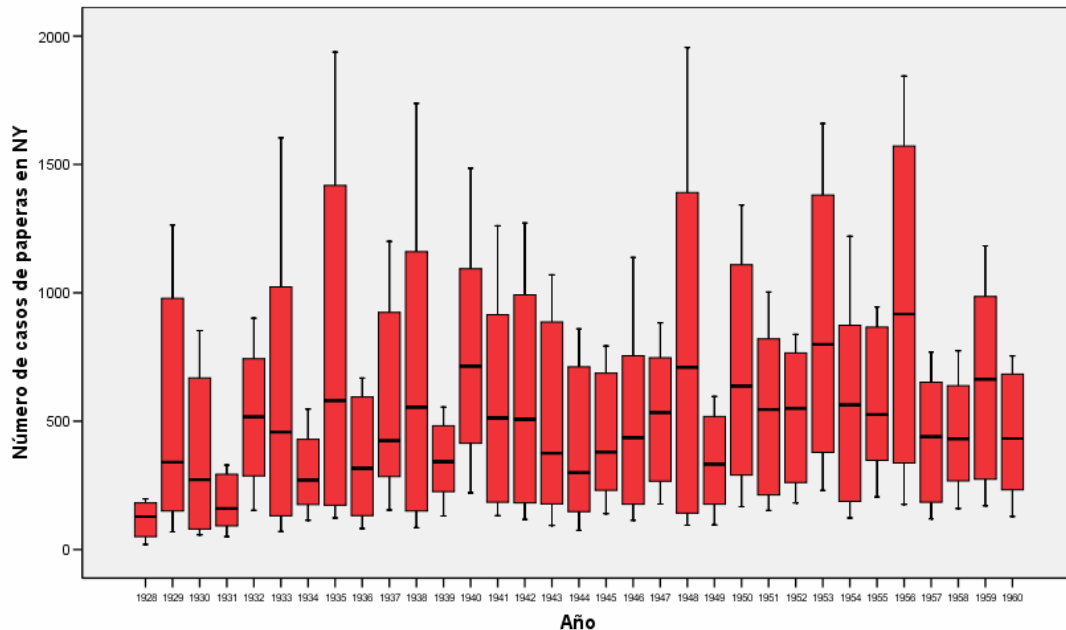


Figura 9: diagramas de cajas por año de la serie Paperas.

La serie paperas es un ejemplo de serie temporal con nivel circunstancial. Esto significa que el nivel de la serie sería estable de no ser por algunos de los picos que se observan en ella. Estos picos no son aleatorios (ruido), sino que son estacionales, se repiten periódicamente.

Estacionalidad

Una serie es estacional cuando podemos observar en ella un patrón sistemático que se repite periódicamente (cada año, cada mes, etc., dependiendo de las unidades de tiempo en que vengamos recogidos los datos).

Existen muchos ejemplos de series con comportamiento estacional. El hecho de que las vacaciones laborales se concentren en los meses de verano condiciona los valores de muchas series. Un claro ejemplo es el de las series relacionadas con el turismo, tales como número mensual de pernoctaciones hoteleras, número de viajeros en avión registrado por meses, etc. La pauta sistemática que caracteriza la estacionalidad no tiene por qué ser anual, algunas series tienen una estacionalidad cuyo periodo es de un mes, una semana o incluso un día. Un ejemplo de este último caso podría ser el de las ventas por horas de un determinado producto en un centro comercial. Es muy probable que cada día se repita un patrón de comportamiento muy similar; habrá horas punta con gran número de ventas y otras en las que apenas hay gente y por lo tanto se registran muchas menos ventas.

Si una serie tiene un comportamiento estacional muy claro se suele detectar fácilmente en el gráfico de secuencia, tal y como ocurre con la serie Airline (Figura 1). En esta serie se observa claramente que hay un patrón similar que se repite año tras año, con evidentes aumentos en el número de pasajeros en los meses de verano e incrementos más moderados en los meses correspondientes a las vacaciones de Navidad. Sin embargo, para estudiar el comportamiento estacional de una serie es más adecuado representar el diagrama de box-plot de cada una de las unidades de tiempo que componen los periodos en los que se agrupan los patrones de repetición. Es decir, en el caso de las series con observaciones mensuales representaremos el diagrama de box-plot de cada uno de los meses del año.

Para representar el gráfico con los diagramas de caja por meses con SPSS debemos seleccionar el menú *Gráficos – Diagramas de caja... - Simple* y en la casilla *Variable* incluimos la variable de la serie y en la correspondiente a *Eje de categorías* la correspondiente al mes.

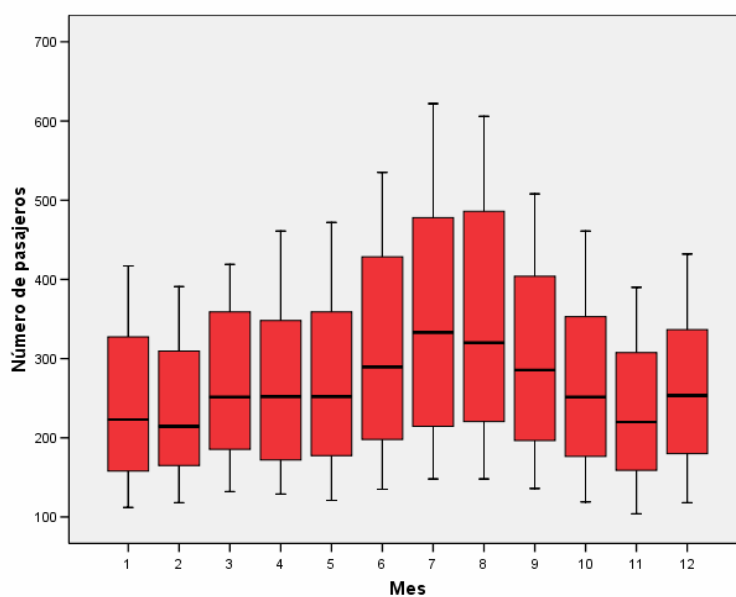


Figura 10: diagramas de cajas por mes de la serie Airline..

En la serie Airline podemos observar, tal como nos pareció en el gráfico de secuencia, la presencia de un pico en verano y otro menos pronunciado en Navidad, épocas en las que la gente más viaja en Estados Unidos. También se puede ver que existe mucha variabilidad (las cajas de los diagramas box-plot son grandes), lo que provoca que se observe peor qué es lo que ocurre. Esta variabilidad en parte es debida a la tendencia, puesto que en el diagrama de cajas de, por ejemplo, el mes de enero, se recogen tanto los datos del número de pasajeros en enero de 1949, entorno a 100, como los de 1960, que rondan los 300.

Veamos otro ejemplo:

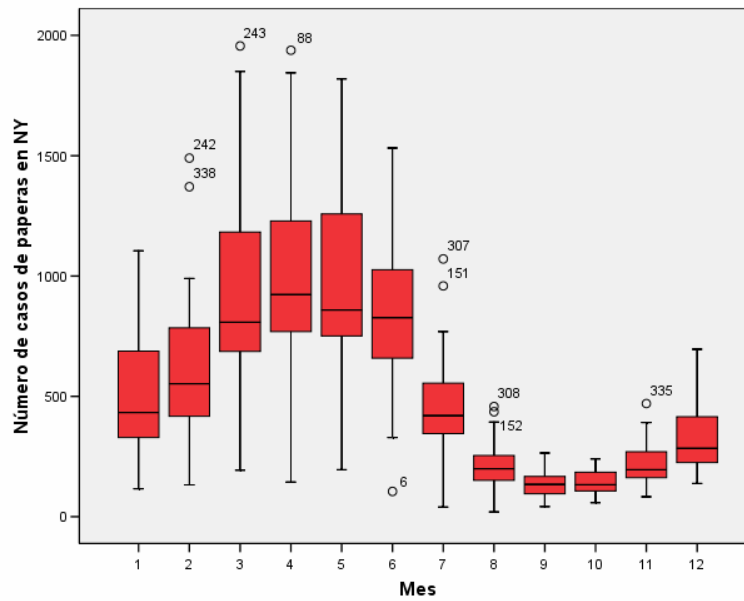


Figura 11: diagramas de cajas por mes de la serie Paperas.

En el caso de la serie paperas (Figura 8), que no presenta tendencia, la estacionalidad se ve mucho más claramente con el diagrama de box-plot por mes. Tal y como podemos observar, la incidencia de paperas en la ciudad de Nueva York es elevada en primavera pero va disminuyendo hasta alcanzar un mínimo en los meses de verano.

La estacionalidad todavía está más clara en el caso de la serie Temperaturas.

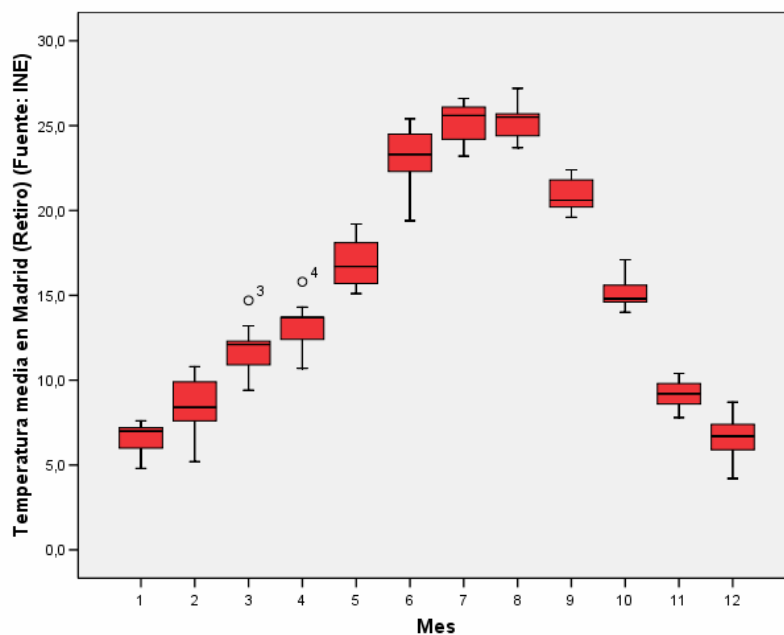


Figura 12: diagramas de cajas por mes de la serie Temperaturas.

Tal y como cabe esperar la serie correspondiente a las temperaturas medias mensuales en Madrid es claramente estacional. La media mensual de temperatura va aumentando progresivamente hasta alcanzar el máximo en los meses de verano.

A continuación, analicemos la estacionalidad de la serie ParoFem. Esta serie contiene el número de miles de mujeres entre 16 y 19 años que se encuentran en paro en EE.UU.

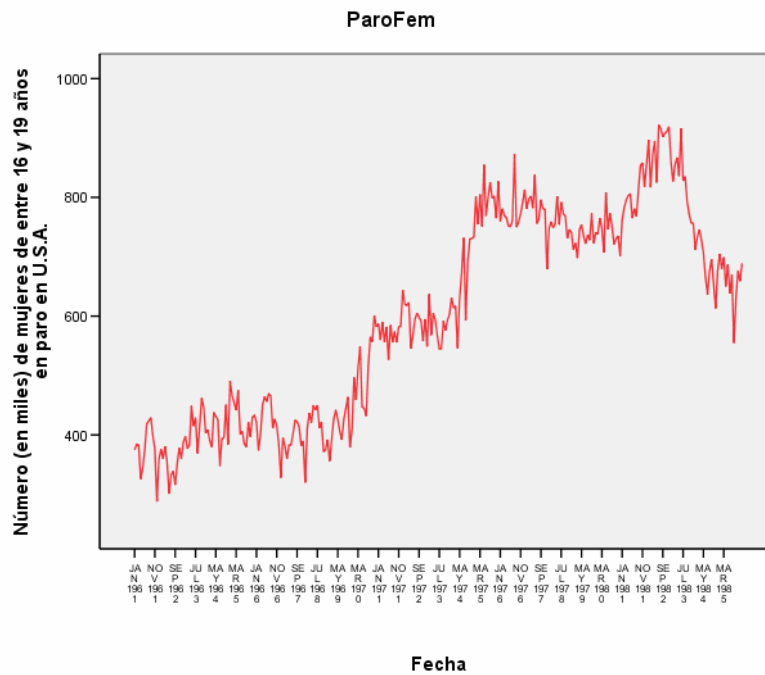


Figura 13: gráfico de secuencia de la serie ParoFem.

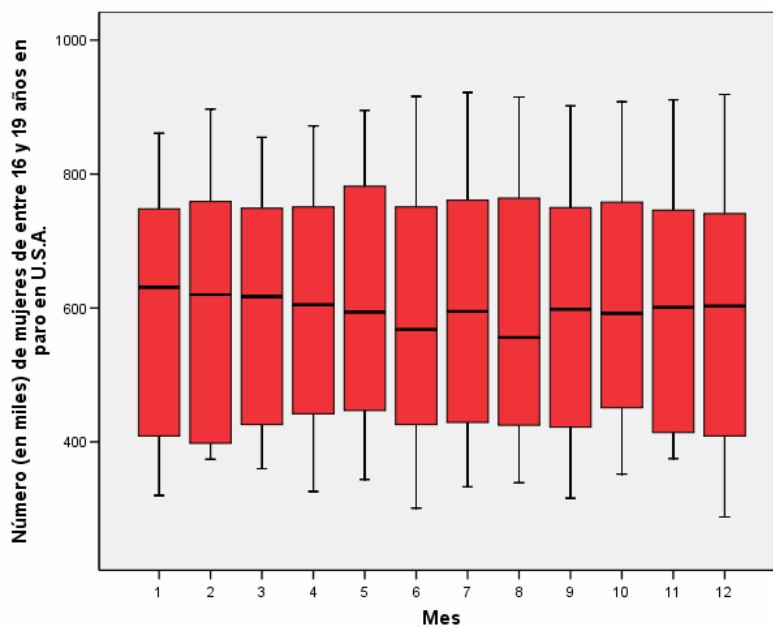


Figura 14: diagramas de cajas por mes de la serie ParoFem.

En este caso nos encontramos ante una serie que no parece tener estacionalidad. En el gráfico de secuencia no se observa un patrón de repetición cíclico y los diagramas de cajas mensuales son muy similares entre sí.

Dependencia entre variabilidad y nivel

Otra de las cuestiones a determinar a la hora de realizar el análisis preliminar de una serie es si existe dependencia entre su variabilidad y el nivel.

Si la variabilidad de una serie no depende del nivel significa que los componentes de la serie se combinan de forma aditiva, es decir, el incremento debido a la estacionalidad siempre es el mismo aunque exista tendencia creciente o decreciente. Si la variabilidad y el nivel dependen entre sí los elementos de la serie se combinan de forma multiplicativa. Esto quiere decir que el incremento debido a la estacionalidad aumenta o disminuye conforme la tendencia crece o decrece.

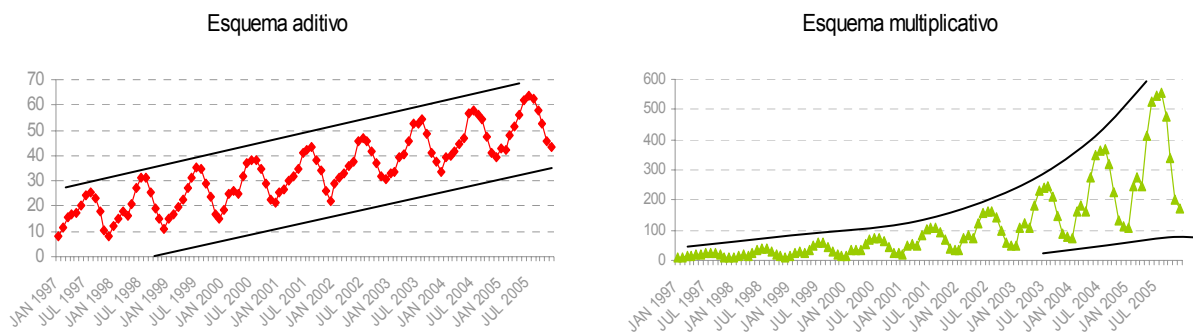


Figura 15: Representación gráfica de dos series con distinto esquema de combinación entre sus componentes de tendencia y estacionalidad.

Para estudiar la dependencia existente entre variabilidad y nivel se emplea el gráfico de dispersión por nivel. Este gráfico es un diagrama de dispersión en el que se representa el logaritmo neperiano de la mediana (medida de tendencia central) frente al logaritmo neperiano de la diferencia entre los percentiles 80 y 20 (medida de la dispersión) de cada uno de los periodos considerados en la serie.

Para obtener con el programa SPSS el gráfico de dispersión por nivel, también llamado gráfico de rango-media, debemos seleccionar el menú *Analizar - Estadísticos descriptivos - Explorar...* y en la casilla *Dependientes* introducir la variable correspondiente a los datos de la serie, y en *Factores* la que determina los periodos elegidos. A continuación pinchamos sobre el botón *Gráficos...* y en *Dispersión por nivel con prueba de Levene* seleccionamos *Estimación de potencia*.

En el caso de la serie Airline el diagrama de dispersión por nivel que obtenemos es el siguiente:

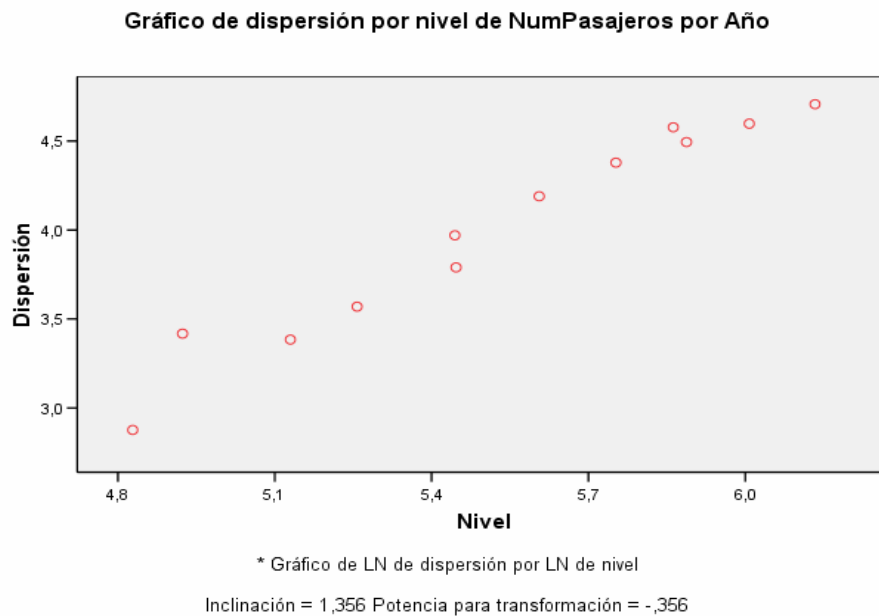


Figura 16: Gráfico de dispersión por nivel de la serie Airline.

En el gráfico podemos ver que existe una dependencia clara entre variabilidad y nivel, ya que los puntos del diagrama pueden ajustarse bastante bien a una recta. También podemos observar que la variabilidad crece con el nivel, es decir, que cuanto mayor es el nivel correspondiente un año de la serie mayor es su variabilidad. Esto también se puede observar en el diagrama de box-plots por año (Figura); cuanto mayor es el nivel (línea central que representa la mediana) mayor es la caja.

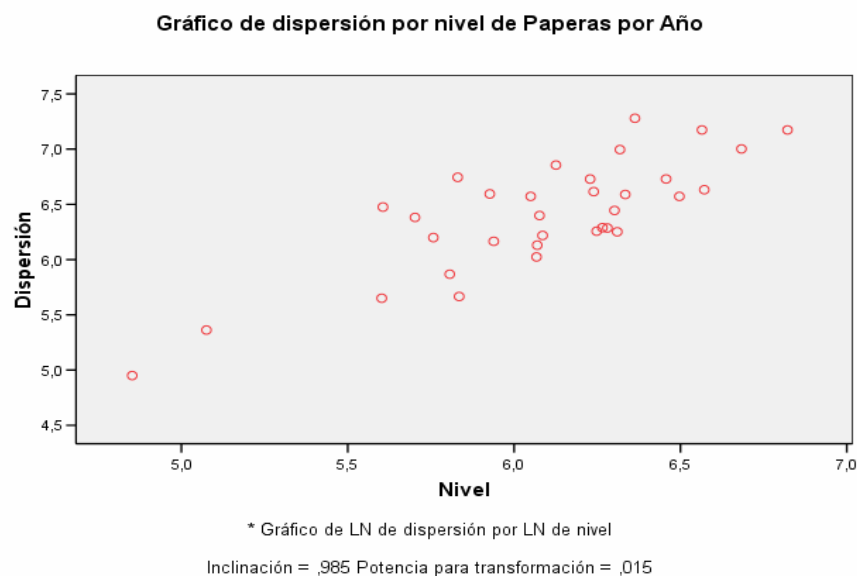


Figura 17: Gráfico de dispersión por nivel de la serie Paperas.

Comportamientos anómalos

Otro aspecto a estudiar en el análisis preliminar de una serie temporal es el de los comportamientos extraños.

Si una serie temporal tiene valores perdidos (en un determinado momento no se han recogido datos) o valores raros, no podemos ignorarlos. Los comportamientos anómalos pueden ser de tres tipos: cambios de tendencia, subidas bruscas de nivel o aparición de valores extraños.

Veamos unos ejemplos:

En este primer caso tenemos una serie con tres cambios de tendencia.

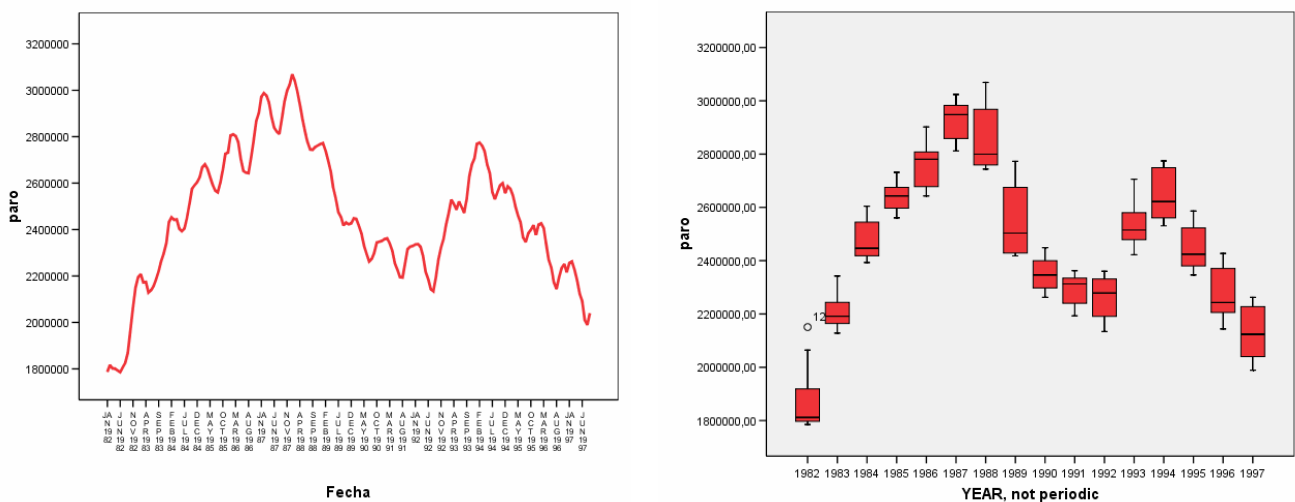


Figura 20: Gráfico de secuencia y diagrama de cajas por año de la serie Paro.

En la serie ParoFem detectamos varios cambios bruscos de nivel.

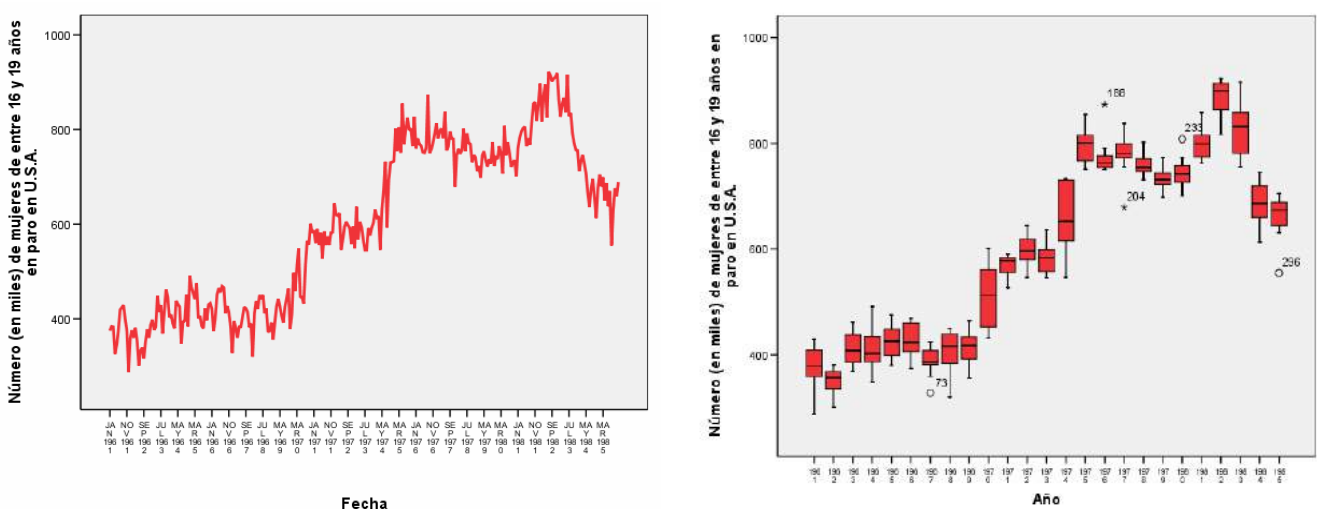


Figura 21: Gráfico de secuencia y diagrama de cajas por año de la serie ParoFem.

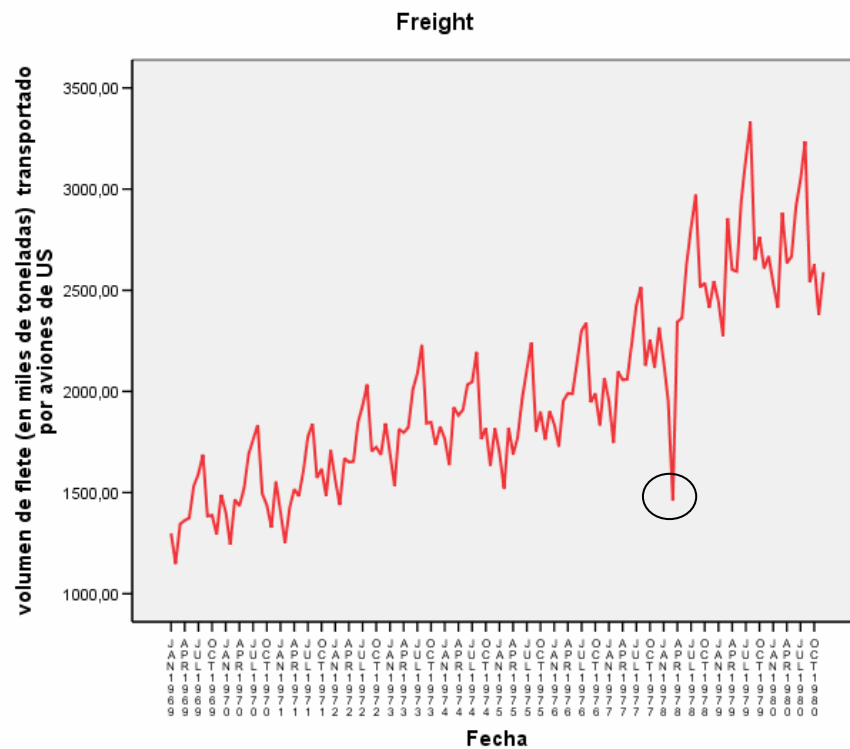


Figura 22: gráfico de secuencia de la serie Freight.

Estudiando el gráfico de secuencia de la serie Freight encontramos un outlier (valor muy diferente a los demás), en este caso un valor anormalmente bajo.

Ante un outlier en primer lugar debemos estudiar si se trata de un error en la recogida de los datos, o de causas externas que han afectado a la variable. En el caso de comportamientos anómalos hay que tratar de analizar las causas del mismo, puesto que en ocasiones puede estar provocado por cuestiones como cambios metodológicos en la recogida de datos, etc.

3.- MÉTODOS CLÁSICOS

Desde una perspectiva teórica el enfoque clásico de análisis de series temporales considera que el comportamiento de una variable en el tiempo es el resultado de la integración de cuatro componentes fundamentales (aunque no siempre aparecen todos): tendencia (T_t), ciclo (C_t), componente estacional (S_t) y componente irregular o ruido (E_t). De esta forma con los métodos clásicos una serie temporal X_t es una función de estos cuatro componentes.

$$X_t = f(C_t, T_t, S_t, E_t)$$

Se considera tendencia (T_t) al movimiento suave y regular de la serie a largo plazo. Refleja la dirección del movimiento de una determinada variable; creciente, decreciente o estable.

El componente cíclico (C_t) consiste en variaciones superiores al año que no son estrictamente periódicas. Se trata de un factor de tipo oscilante caracterizado por movimientos recurrentes en torno a la tendencia, y suele aparecer fundamentalmente en series de tipo económico.

En muchas series temporales podemos observar un patrón sistemático que se repite todos los años, es decir, todos los años aparece un cambio de valor en un determinado mes. La estacionalidad (S_t) de una serie son los movimientos regulares de la misma que tienen una periodicidad inferior al año. Recoge las oscilaciones que año a año se repiten en una serie de forma periódica.

El componente irregular o ruido (E_t) incluye las variaciones de la serie cuyas leyes nos resultan desconocidas. Se caracteriza porque no responde a un comportamiento sistemático o regular y, en consecuencia, no es posible su predicción. El enfoque clásico atribuye esta irregularidad al azar. De esta forma el ruido lo compone todo lo que no queda explicado por la tendencia, el ciclo y la estacionalidad.

Los métodos clásicos de análisis de series temporales tienen la ventaja de no ser excesivamente complejos, aunque como contrapartida responden a preguntas menos ambiciosas. Se pueden emplear para realizar predicciones a corto plazo, pero no a medio o largo plazo. Por ejemplo, en series mensuales se utilizan para predecir uno o dos meses, no un año completo³.

Métodos de descomposición estacional y ajuste de tendencia

Los métodos de descomposición estacional son eminentemente descriptivos. Tratan de separar la serie en subseries correspondientes a la tendencia-ciclo⁴, la estacionalidad y el ruido (componente aleatorio).

³ En algunos de los ejemplos que veremos a continuación se emplean los métodos clásicos para predecir un año completo, pero se hace únicamente para poder ver gráficamente cómo pronostican.

⁴ Separar la tendencia y el ciclo es muy difícil, habitualmente van a quedar confundidos, de manera que consideraremos una única componente tendencia-ciclo.

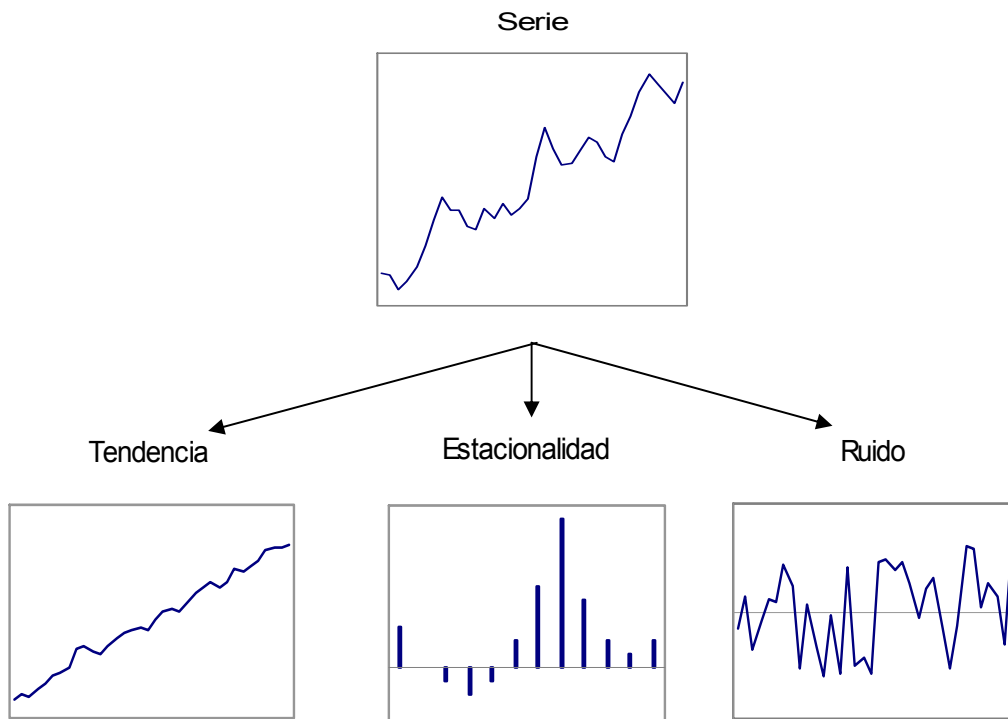


Figura 23: Descomposición estacional.

En ocasiones nos interesa desestacionalizar una serie, eliminar la influencia estacional. Supongamos que tratamos de analizar si, a lo largo del año, los niveles de ozono aumentan, se mantienen o disminuyen. El nivel de ozono sube en verano, pero lo que tratamos de determinar es si existe una subida más allá del incremento propio del verano. De esta forma lo que nos interesa es estudiar la tendencia de la serie independientemente de la subida que se produce cada estío. Otro ejemplo en el que nos puede interesar eliminar la influencia estacional de una serie es el caso en el que necesitamos decidir la capacidad que debe tener un almacén. Para ello analizamos los datos correspondientes al volumen de producto a almacenar cada mes. Trataremos de tomar la decisión abstractándonos de los picos de la estacionalidad. No nos interesa montar una nueva planta de almacén que solo se va a ocupar en un pico como, por ejemplo, navidad.

En ocasiones tendencia y estacionalidad se enmascaran, a veces una tendencia marcada puede no dejarnos ver la estacionalidad, y viceversa. Los métodos de descomposición estacional separan tendencia, estacionalidad y ruido, pero no predicen. Para predecir es necesario combinarlos con métodos de ajuste de tendencia.

A la hora de predecir consideramos la estacionalidad constante periodo a periodo y el ruido cero. El ruido es aleatorio, impredecible, y tiene media cero, de manera que la mejor previsión que podemos hacer de él es cero. De esta forma

para pronosticar realizamos un ajuste de tendencia con el fin de obtener un modelo extrapolable, y le añadimos la estacionalidad.

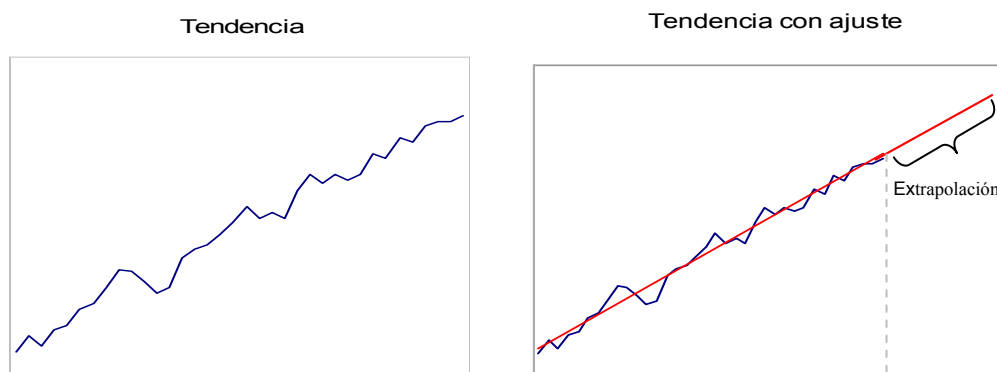


Figura 24: Tendencia de una serie (datos concretos, no podemos extrapolar) y tendencia con ajuste (podemos extrapolar la recta ajustada a la tendencia)

El primer paso a seguir a la hora de descomponer una serie es determinar cómo se combinan sus componentes. Las combinaciones aditiva y multiplicativa son las más habituales⁵.

Tal y como hemos visto en la introducción a los métodos clásicos, según éstos una serie temporal X_t es una función que depende de cuatro componentes.

$$X_t = f(C_t, T_t, S_t, E_t)$$

Si dichos componentes se combinan de forma aditiva entonces:

$$X_t = C_t + T_t + S_t + E_t$$

y si lo hacen de forma multiplicativa:

$$X_t = C_t \times T_t \times S_t \times E_t$$

Desarrollemos el proceso para el caso aditivo.

En primer lugar eliminamos el ruido y la estacionalidad. El ruido se elimina sustituyendo cada observación por una media de lo ocurrido anteriormente (media móvil anterior) y la estacionalidad realizando un proceso de media móvil centrada. Este último procedimiento suaviza cada observación tomando la media de igual número de valores anteriores y posteriores a la misma. El orden de la media móvil centrada, es decir, el número total de observaciones que generará cada media móvil centrada, habitualmente es igual al periodo de la serie. En cualquier caso debe ser tal que no incluya más observaciones de una unidad de periodo que de las demás. Es decir, si la serie tiene periodicidad semanal, en la media móvil centrada no deben estar incluidos los valores correspondien-

⁵ Ver el apartado relativo a la dependencia entre la variabilidad y el nivel (Análisis preliminar).

tes a dos lunes si no se incluyen también dos veces los del resto de los días de la semana.

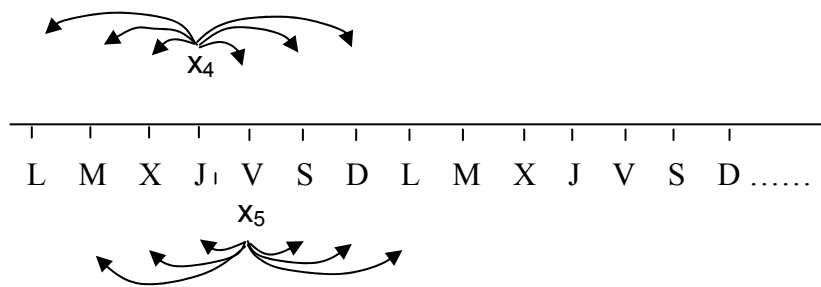


Figura 25: Media móvil centrada de una serie con información diaria (periodo 7).

En el caso de que el periodo de la serie sea par, por ejemplo los trimestres del año, la media móvil centrada se calcula tomando la media de las dos posibles medias móviles a considerar.

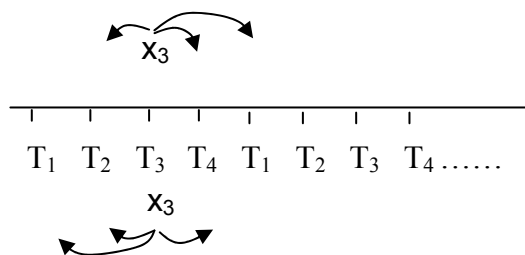


Figura 26: Posibles medias móviles centradas a considerar en una serie con información trimestral (periodo 4).

Una vez eliminados la estacionalidad y el ruido obtenemos una serie que únicamente está formada por la tendencia y el ciclo: $M_t = C_t + T_t$.

A continuación, eliminando la tendencia y el ciclo de la serie de partida, conseguimos una serie integrada solo por la estacionalidad y el ruido: $X_t - M_t = E_t + S_t$. Para estimar el factor estacional a partir de esta última serie se realiza la media de todas las observaciones disponibles de cada unidad de periodo (por ejemplo, cada mes de un año). De esta forma la serie del factor estacional estará formada por los valores de estas medias, es decir, será una serie que repetirá constantemente los mismos valores en cada unidad de periodo. Haciendo $X_t - M_t - S_t = E_t$ obtenemos una serie con el error, con el ruido.

En el caso de que los componentes de la serie se combinen de forma multiplicativa el proceso es equivalente. En primer mediante medias móviles obtene-

mos una serie formada únicamente por la tendencia y el ciclo: $M_t = C_t \times T_t$. A continuación, haciendo $\frac{X_t}{M_t} = E_t \times S_t$, conseguimos una serie formada solo por la estacionalidad y el ruido. Finalmente, tras estimar la componente estacional, haciendo $\frac{X_t}{M_t \times S_t} = E_t$ obtenemos una serie con el error.

Realicemos un ejemplo de análisis de una serie mediante este método clásico empleando el programa SPSS.

Sea la serie Cava, correspondiente a datos de venta de cava. En primer lugar debemos realizar un análisis preliminar de la serie.

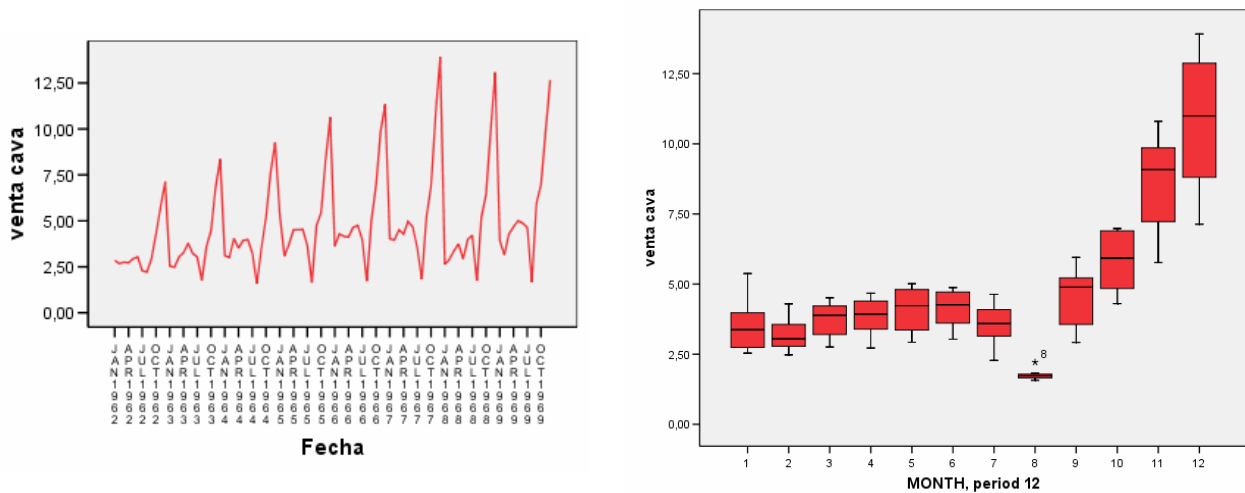


Figura 27: Gráfico de secuencia y diagrama de box-plots por mes de la serie cava.

Para poder aplicar este método la serie debe tener un comportamiento estacional. Al observar el gráfico de secuencia de la serie vemos de forma bastante clara que existe estacionalidad, hecho que se confirma al analizar el gráfico de box-plot por mes. En este último advertimos que la venta de cava en los primeros meses del año es más o menos parecida hasta agosto, donde encontramos un mínimo, a partir del cual las ventas van creciendo hasta alcanzar un máximo en diciembre. Seguramente este crecimiento estará causado por la proximidad de la Navidad, periodo de mayor consumo de cava del año, y la consecuente necesidad de abastecimiento de las tiendas en los meses previos a las fiestas.

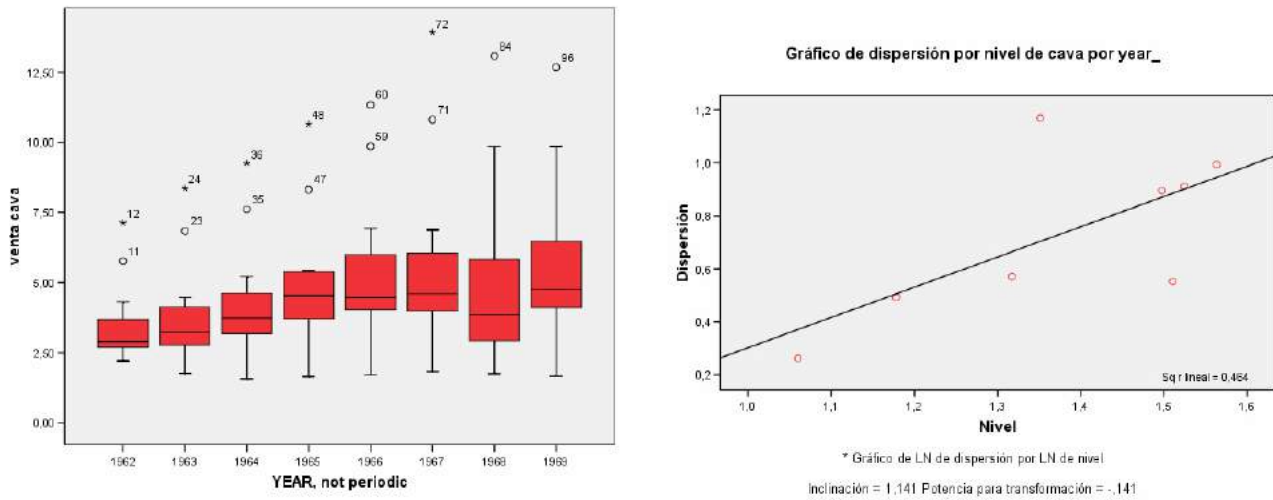


Figura 28: Diagrama de box-plots por año y gráfico de dispersión por nivel de la serie cava.

Observando el gráfico de dispersión por nivel de la serie, vemos que los puntos del mismo parecen ajustarse a una recta, lo que indicaría dependencia. Sin embargo existe un punto que puede estar falseando la tendencia. Además, el valor del R^2 no es alto (0,494), a pesar de que esto es usual en este tipo de análisis, puesto que habitualmente se dispone de pocos casos. Si observamos el diagrama de cajas por año la variabilidad no parece depender del nivel, realmente lo parece ocurrir es que el penúltimo año tiene mayor variabilidad. De esta manera nos decantaremos por considerar que no existe dependencia entre variabilidad y nivel, es decir, por el modelo aditivo.

Para realizar la descomposición estacional de la serie con SPSS debemos seleccionar *Analizar – Series temporales – Descomposición estacional...* y, tras introducir la variable con los datos de la serie en la casilla *Variables*, activar las opciones *modelo aditivo* y *Ponderación de la media móvil Puntos finales ponderados por ,5* (seleccionamos esta opción siempre que la longitud del periodo sea par, y en este caso lo es, consta de 12 meses). El resultado que nos proporciona SPSS es el siguiente:

Descripción del modelo

| | |
|---|---|
| Nombre del modelo | MOD_1 |
| Tipo de modelo | Aditivo |
| Nombre de la serie | 1 cava venta cava |
| Longitud del período estacional | 12 |
| Método de computación de medias móviles | Amplitud igual a la periodicidad y todos los puntos ponderados igualmente |

Aplicando las especificaciones del modelo de MOD_1

Factores estacionales

Nombre de la serie: cava venta cava

| Período | Factor estacional |
|---------|-------------------|
| 1 | -1,19183 |
| 2 | -1,55867 |
| 3 | -,94680 |
| 4 | -,84001 |
| 5 | -,63348 |
| 6 | -,64499 |
| 7 | -1,24087 |
| 8 | -2,92404 |
| 9 | -,38616 |
| 10 | ,93247 |
| 11 | 3,68062 |
| 12 | 5,75377 |

El primer recuadro es un resumen del tipo de modelo empleado para realizar la descomposición estacional, y el siguiente nos da los valores concretos de los factores estacionales. Estas cifras nos indican qué ocurre con los valores de la serie cada uno de los meses del año. Por ejemplo, en los meses de enero las ventas se reducen en un 1,19, mientras que meses como octubre suponen un aumento de 0,93.

Además de estos resultados, el programa SPSS crea cuatro nuevas variables:

- ERR_1 Componente aleatorio, ruido (E_t)
- SAS_1 Serie desestacionalizada ($X_t - S_t$)
- SAF_1 Componente estacional (S_t)
- STC_1 Componente de tendencia-ciclo ($C_t + T_t$)

Se puede comprobar fácilmente que la suma SAS_1+STC_1+ERR_1 da como resultado la serie original.

Gráficamente:

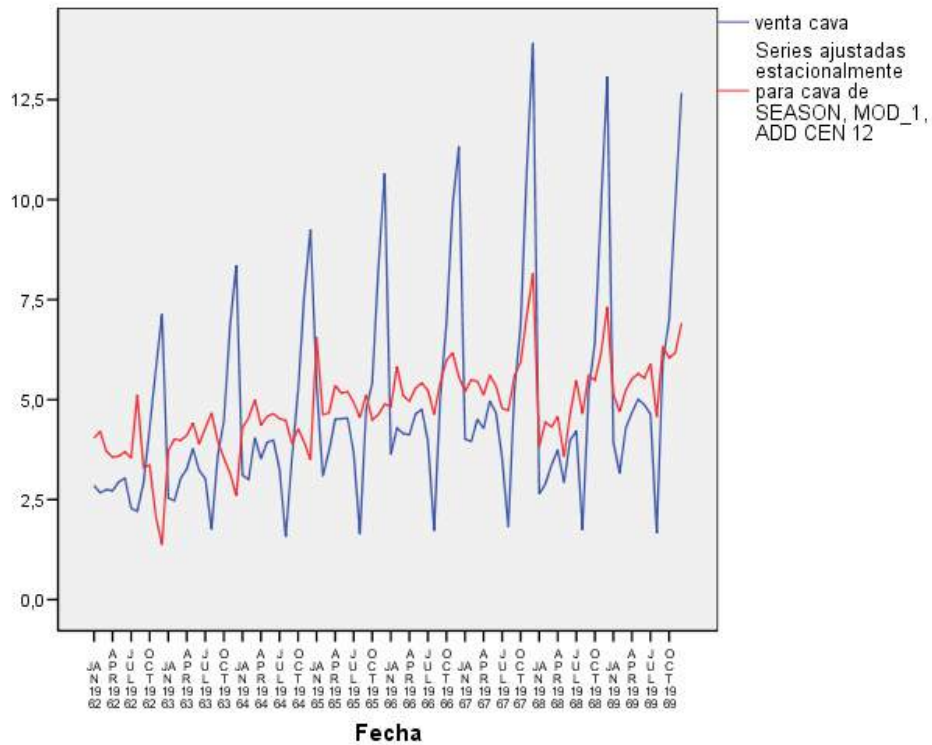


Figura 29: Gráfico de secuencia de la serie cava original y la serie desestacionalizada.

Si además de eliminar la influencia estacional nos interesa predecir, en primer lugar tenemos que ajustar la tendencia de la serie.

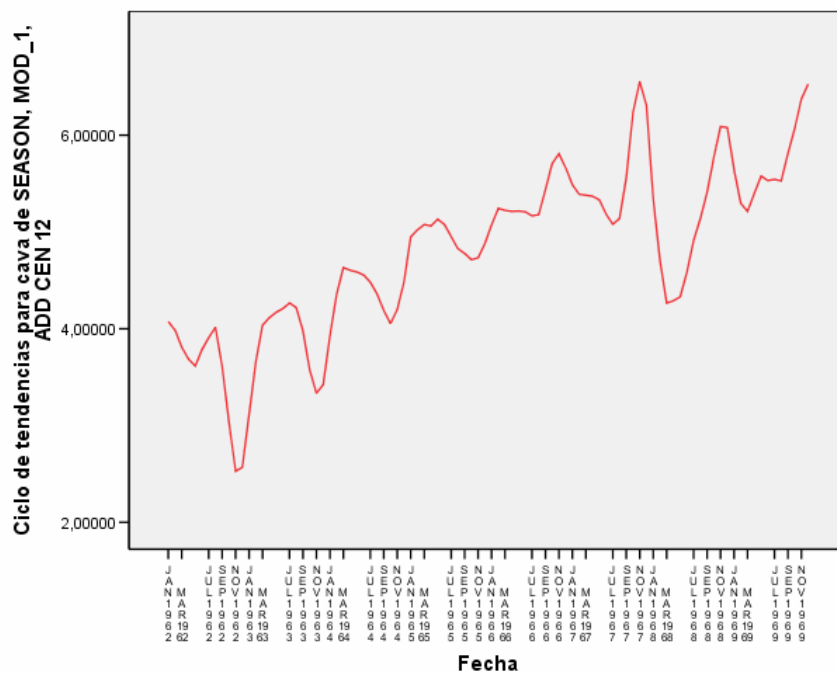


Figura 30: Gráfico de secuencia de la subserie tendencia-ciclo.

Para ajustar la tendencia con el programa SPSS debemos seleccionar *Analizar – Regresión – Estimación curvilínea...* y en la casilla variable dependiente incluimos la variable correspondiente a la componente de tendencia-ciclo (STC) y como variable independiente activamos *Tiempo*. Dentro del apartado *Modelos* disponemos de diferentes tipos de ajustes posibles, elegimos los que nos interesen y activamos la casilla *Mostrar tabla de ANOVA*.

| | | |
|--|---|--|
| Lineal $y = b_0 + b_1 \cdot t$ | Cuadrático $y = b_0 + b_1 \cdot t + b_2 \cdot t^2$ | Inverso $y = b_0 + \frac{b_1}{t}$ |
| Logarítmico $y = b_0 + b_1 \cdot \ln t$ | Cúbico $y = b_0 + b_1 \cdot t + b_2 \cdot t^2 + b_3 \cdot t^3$ | Compuesto $y = b_0 \cdot b_1^t$ |
| Potencia $y = b_0 \cdot t^{b_1}$ | S $y = e^{b_0 + \frac{b_1}{t}}$ | Exponencial $y = b_0 \cdot e^{b_1 \cdot t}$ |
| Crecimiento $y = e^{b_0 + b_1 \cdot t}$ | Logístico $y = \frac{1}{\frac{1}{u} + b_0 \cdot b_1^t}$ | |

En nuestro ejemplo analizaremos el ajuste lineal, el S y el logarítmico.

El resultado que nos proporciona SPSS es:

Descripción del modelo

| | | |
|--|---|--|
| Nombre del modelo | | MOD_4 |
| Variable dependiente | 1 | STC_1 Ciclo de tendencias para cava de SEASON, MOD_1, ADD CEN 12 |
| Ecuación | 1 | Lineal |
| | 2 | Logarítmica |
| | 3 | S ^a |
| Variable independiente | | Secuencia de casos |
| Constante | | Incluidos |
| Variable cuyos valores etiquetan las observaciones en los gráficos | | Sin especificar |

a. El modelo necesita todos los valores no perdidos para ser positivo.

Resumen del procesamiento de los casos

| | N |
|------------------------------|----|
| Total de casos | 96 |
| Casos excluidos ^a | 0 |
| Casos pronosticados | 0 |
| Casos creados nuevos | 0 |

a. Los casos con un valor perdido en cualquier variable se excluyen del análisis.

Resumen del procesamiento de las variables

| | Variables | |
|-----------------------------|--|----|
| | Dependiente | |
| | STC_1 Ciclo de tendencias para cava de SEASON, MOD_1, ADD CEN 12 | |
| Número de valores positivos | | 96 |
| Número de ceros | | 0 |
| Número de valores negativos | | 0 |
| Número de valores perdidos | Perdidos definidos por el usuario | 0 |
| | Perdidos del sistema | 0 |

STC_1 Ciclo de tendencias para cava de SEASON, MOD_1, ADD CEN 12**Lineal****Resumen del modelo**

| R | R cuadrado | R cuadrado corregida | Error típico de la estimación |
|------|------------|----------------------|-------------------------------|
| ,817 | ,667 | ,663 | ,498 |

ANOVA

| | Suma de cuadrados | gl | Media cuadrática | F | Sig. |
|-----------|-------------------|----|------------------|---------|------|
| Regresión | 46,660 | 1 | 46,660 | 188,269 | ,000 |
| Residual | 23,297 | 94 | ,248 | | |
| Total | 69,957 | 95 | | | |

Coefficientes

| | Coeficientes no estandarizados | | Coeficientes estandarizados | t | Sig. |
|--------------------------------|--------------------------------|--------------|-----------------------------|--------|------|
| | B | Error típico | Beta | | |
| Secuencia de casos (Constante) | ,025 | ,002 | ,817 | 13,721 | ,000 |
| | 3,581 | ,102 | | 34,967 | ,000 |

Logarítmica

Resumen del modelo

| R | R cuadrado | R cuadrado corregida | Error típico de la estimación |
|------|------------|----------------------|-------------------------------|
| ,747 | ,558 | ,554 | ,573 |

ANOVA

| | Suma de cuadrados | gl | Media cuadrática | F | Sig. |
|-----------|-------------------|----|------------------|---------|------|
| Regresión | 39,070 | 1 | 39,070 | 118,900 | ,000 |
| Residual | 30,888 | 94 | ,329 | | |
| Total | 69,957 | 95 | | | |

Coefficientes

| | Coeficientes no estandarizados | | Coeficientes estandarizados | t | Sig. |
|-------------------|--------------------------------|--------------|-----------------------------|--------|------|
| | B | Error típico | Beta | | |
| ln(Case Sequence) | ,692 | ,064 | ,747 | 10,904 | ,000 |
| (Constante) | 2,310 | ,236 | | 9,796 | ,000 |

S

Resumen del modelo

| R | R cuadrado | R cuadrado corregida | Error típico de la estimación |
|------|------------|----------------------|-------------------------------|
| ,346 | ,120 | ,110 | ,181 |

ANOVA

| | Suma de cuadrados | gl | Media cuadrática | F | Sig. |
|-----------|-------------------|----|------------------|--------|------|
| Regresión | ,419 | 1 | ,419 | 12,778 | ,001 |
| Residual | 3,079 | 94 | ,033 | | |
| Total | 3,497 | 95 | | | |

Coefficientes

| | Coeficientes no estandarizados | | Coeficientes estandarizados | t | Sig. |
|-------------------|--------------------------------|--------------|-----------------------------|--------|------|
| | B | Error típico | Beta | | |
| 1 / Case Sequence | -,555 | ,155 | -,346 | -3,575 | ,001 |
| (Constante) | 1,581 | ,020 | | 78,058 | ,000 |

La variable dependiente es ln(STC_1 Ciclo de tendencias para cava de SEASON, MOD_1, ADD CEN 12).

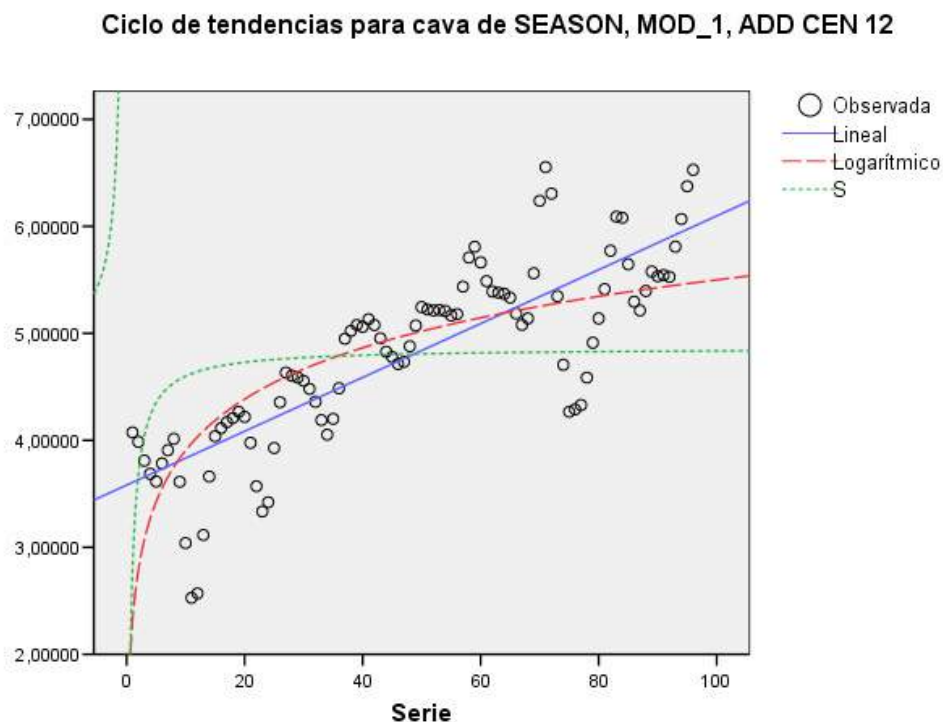


Figura 31: Gráfico de la subserie tendencia-ciclo con los ajustes lineal, logarítmico y S.

Tal y como podemos observar en las tablas ANOVA, los tres ajustes son significativos (Sig.<0,05), lo que quiere decir que podemos considerar los tres modelos. También son significativos los coeficientes ajustados en los tres casos, es decir, que ninguno de ellos es considerado nulo. Sin embargo, para decidirnos entre uno de los tres ajustes debemos analizar el gráfico con los tres ajustes sobre los datos originales, así como el R cuadrado corregido que aparece en el resumen del modelo. Este valor nos da una medida del porcentaje de variabilidad de los datos explicado por el modelo, de manera que cuanto más próximo a 1 sea su valor, mejor representados estarán nuestros datos.

Observando el gráfico vemos que el modelo S no ajusta bien la serie correspondiente a la tendencia, se aleja de los datos de la misma. Sin embargo, al estudiar el ajuste lineal y el logarítmico vemos que éstos se ajustan mejor. Si ahora nos remitimos a los valores del R^2 corregido, vemos que el ajuste S es el que presenta un menor valor del mismo (0,110), como cabía esperar tras fijarse en el gráfico. En el caso del modelo lineal y logarítmico los valores son 0,663 y 0,554 respectivamente, de manera que nos inclinamos por el ajuste lineal, puesto que tiene el valor más elevado del R^2 corregido.

A continuación, para realizar predicciones de nuestra serie con SPSS, debemos generar las previsiones de la componente de tendencia-ciclo. Para ello debemos acudir de nuevo al menú *Analizar – Regresión – Estimación curvilínea...* y, al ajustar el modelo lineal a la variable STC_1, pinchar en el botón

Guardar y tras activar la casilla *Valores pronosticados* en *Pronosticar casos* seleccionamos *Predecir hasta* e indicamos la fecha hasta la que queremos obtener predicciones. Tras hacer esto se genera una nueva variable que contiene el ajuste de todos los puntos de la subserie tendencia-ciclo para los que hay datos, así como las previsiones que hemos solicitado. Para obtener las predicciones de la serie original, debemos sumar los valores pronosticados (variable FIT) y los valores de SAF_1 (componente estacional). La previsión que hacemos de la componente de ruido será cero, de manera que sumando estas dos variables (*Transformar – Calcular...*), obtenemos las predicciones de venta de cava.

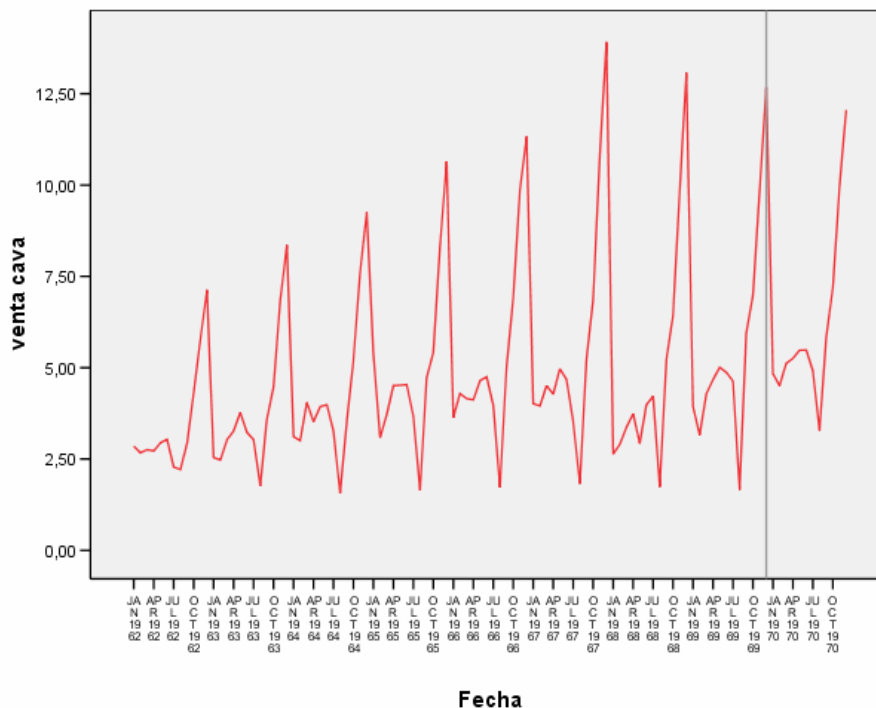


Figura 32: Gráfico de secuencia de la serie cava original y las predicciones para un año.

Métodos de suavizado

Los métodos de suavizado o alisado son técnicas de tipo predictivo más que descriptivo. Resultan más adecuados para pronosticar, y proporcionan previsiones razonables para horizontes de predicción inmediatos. Además los resultados que se obtienen con ellos son satisfactorios, incluso cuando no se dispone de un gran número de datos históricos.

A diferencia de los métodos de descomposición estacional, para aplicar los de suavizado no es necesario que la serie presente comportamiento estacional. Dentro de estos últimos existen modelos para series no afectadas por tendencia ni estacionalidad, para series con tendencia y para series con tendencia y estacionalidad.

MODELOS APLICABLES A SERIES SIN TENDENCIA NI ESTACIONALIDAD

Este tipo de series tiene un comportamiento más o menos estable que sigue un patrón subyacente (μ) salvo fluctuaciones aleatorias (e_t), de manera que se modelizan de la siguiente forma:

$$X_t = \mu + e_t$$

Todos los modelos que se van a exponer a continuación proporcionan una misma previsión para todo el horizonte de previsión. Por esta razón se exige que se apliquen a series sin tendencia ni estacionalidad. Se emplean para realizar predicciones a corto plazo, puesto que a largo plazo lo que se haría es actualizar la serie con la nueva información y efectuar de nuevo una predicción a corto plazo.

Supongamos que tenemos una serie de este tipo con n observaciones. Para predecir el valor $n+1$ de la serie existen diversas opciones:

Modelos “naive” o ingenuos

- Se otorga la misma importancia ($1/n$) a todas las observaciones a la hora de predecir. De esta forma la previsión vendrá dada por la media de las observaciones.

$$\hat{X}_{n+1} = \bar{x}$$

Consideremos la serie Combustible, que contiene datos mensuales sobre la producción de gasolina para la automoción en España.

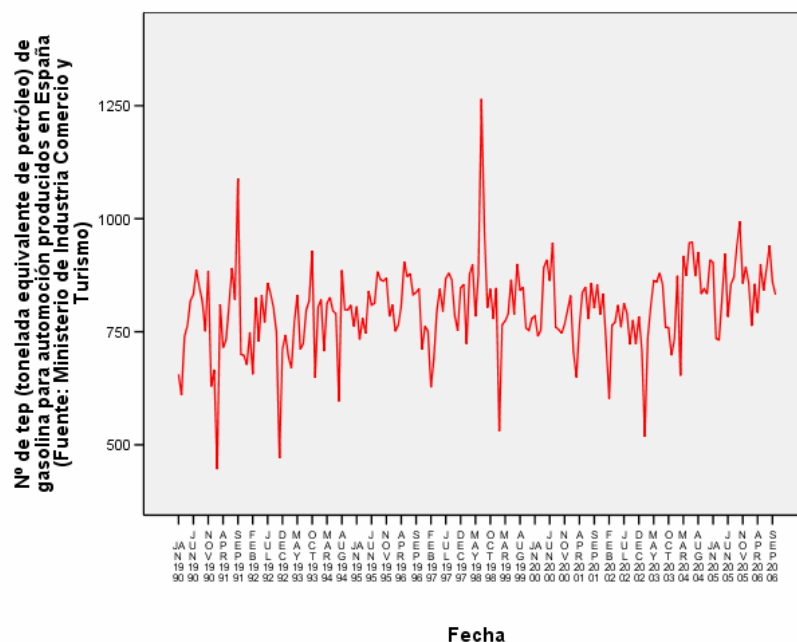


Figura 33: Gráfico de secuencia de la serie Combustible.

Se trata de una serie que no presenta tendencia ni estacionalidad, de manera que podemos emplearla para aplicar los distintos métodos que estamos viendo.

Veamos el ajuste de la serie Combustible que resulta dando la misma importancia a todas las observaciones disponibles, es decir, ajustando la serie a través de su media.

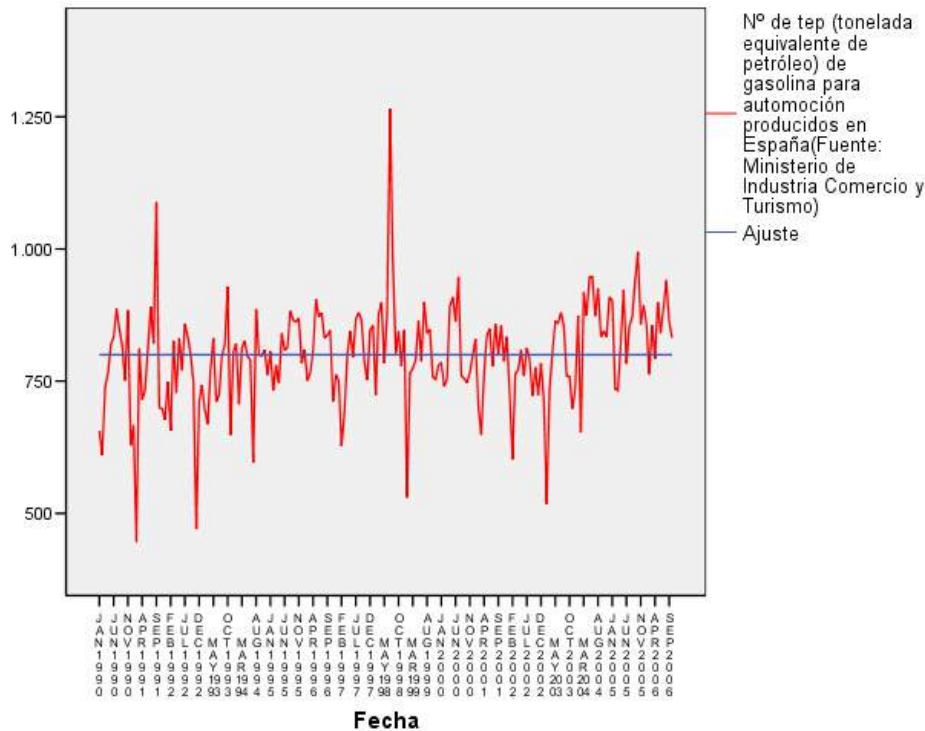


Figura 34: Gráfico de secuencia de la serie Combustible con el ajuste efectuado empleando la media de la serie.

La previsión del siguiente mes para el que no se dispone de datos con este método será la propia media.

- Se da importancia únicamente al último de los datos de que disponemos, ignorando el resto.

$$\hat{X}_{n+1} = X_n$$

Para realizar este ajuste con el programa SPSS debemos crear una nueva variable con el comando *Calcular* empleando la función *Lag(1)*. Para ello seleccionamos el menú *Transformar – Calcular...* y a continuación en la casilla *Variable de destino* escribimos el nombre de la nueva variable que contendrá la serie ajustada. En la casilla *Expresión numérica* debemos incluir la variable *Lag(1)* que encontramos dentro de la agrupación *Otras* del listado *Grupo de funciones*. Finalmente debemos incluir la variable que contiene los valores de nuestra serie como argumento de la función *Lag*.

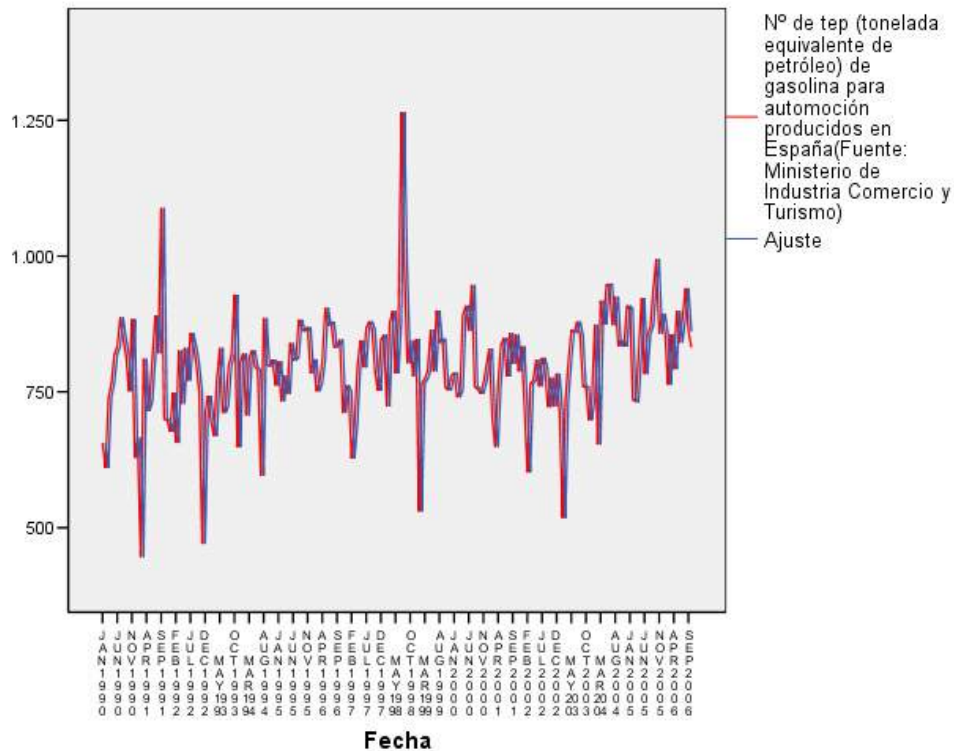


Figura 35: Gráfico de secuencia de la serie Combustible con el ajuste efectuado considerando únicamente el dato anterior.

Tal y como podemos observar en el gráfico de secuencia, el ajuste de la serie es su “sombra”, es la misma serie pero retardada en una unidad de periodo, en este caso un mes.

La previsión del siguiente mes para el que no se dispone de datos con este método será el último valor de la serie.

Modelos de medias móviles

Se basan en considerar únicamente las últimas k observaciones. De esta manera se da el mismo peso a los últimos k datos ($1/k$) y cero al resto, mediante un procedimiento de medias móviles. Este procedimiento no es tan extremo como los anteriores, y al sustituir cada dato por una media de los k últimos la serie se suaviza y se elimina ruido, obteniendo el patrón subyacente de la misma. Cuantas más observaciones relevantes (k) tomemos al aplicar este tipo de ajuste más se suavizará la serie.

$$\hat{X}_{n+1} = \frac{\sum_{i=1}^k X_{n-i+1}}{k}$$

Para obtener el ajuste de una serie mediante este procedimiento empleando el programa SPSS, en primer lugar tenemos que seleccionar el menú *Transformar – Crear serie temporal...* A continuación debemos introducir la variable que

contiene los valores de la serie en la casilla *Nuevas variables* y en *Función* seleccionar *Media móvil anterior*. Al seleccionar esta función se activará la casilla *Amplitud*, donde debemos indicar el número de observaciones anteriores que queremos que se tengan en cuenta para calcular la media móvil anterior. Al ejecutar este menú SPSS crea una nueva variable con la serie ajustada. Hay que tener en cuenta que esta nueva variable no dispone de valores en los k primeros registros, puesto que en esos casos no existen k observaciones anteriores para calcular la media.

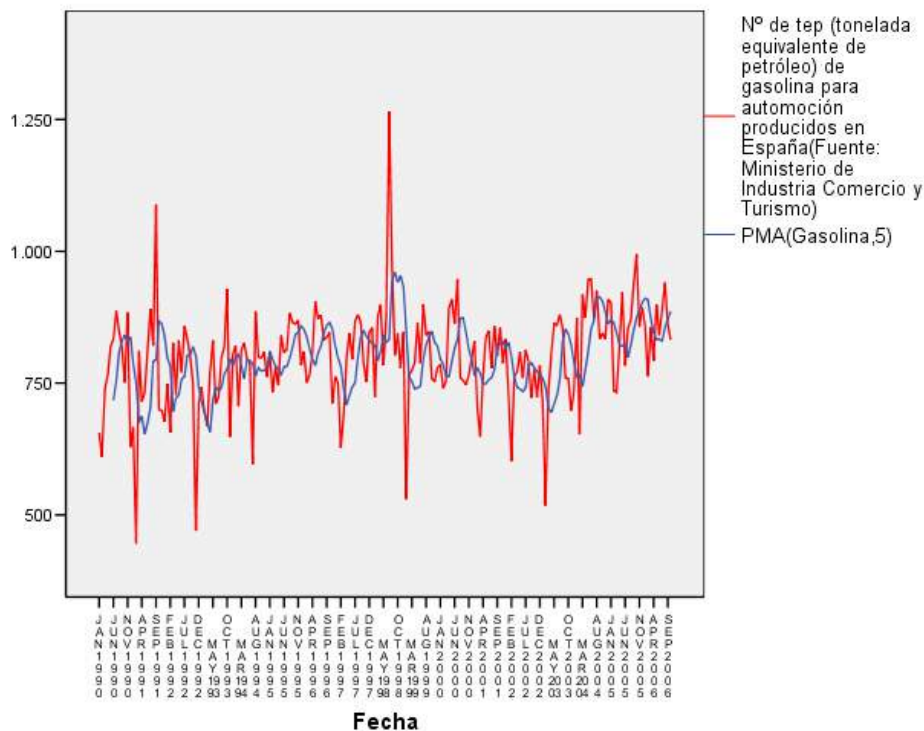


Figura 36: Gráfico de secuencia de la serie Combustible con el ajuste efectuado considerando un modelo de medias móviles anteriores de amplitud 5.

La previsión del siguiente mes para el que no se dispone de datos con este método será la media de los últimos k valores de la serie.

Modelos de suavizado exponencial simple

Consisten en dar importancia a todos los datos anteriores, pero concediéndoles diferentes pesos. Los datos más relevantes a la hora de efectuar una previsión son los últimos de los que se dispone, de forma que este método considera que la importancia disminuye conforme nos alejamos de ellos. De esta manera surgen los métodos de suavizado exponencial, que sustituyen cada dato de la serie por una media ponderada de las observaciones anteriores, considerando que los pesos de las mismas decaen de forma exponencial conforme éstas se alejan en el tiempo.

$$\hat{X}_{n+1} = \alpha \cdot X_n + (1 - \alpha) \cdot \hat{X}_n \quad \text{donde} \quad 0 \leq \alpha \leq 1$$

La fórmula del ajuste es recursiva (se alimenta de sí misma), de manera que es necesario determinar la estimación del primer valor de la serie. También hay que establecer el valor de α (peso), que establece la diferencia de importancia que se va a dar a la observación inmediatamente anterior de la serie y a toda la información anterior recogida en el ajuste de la misma. El parámetro α puede tomar valores entre 0 y 1, de manera que cuanto menor sea α más suavizaremos la serie. Si $\alpha = 0$ ó $\alpha = 1$ nos encontramos con los modelos naive. El valor de α que se suele tomar es aquel que minimiza una función de pérdida establecida.

Para efectuar el suavizado exponencial simple de una serie con el programa SPSS el primer paso a realizar es seleccionar el menú *Analizar – Series temporales – Suavizado exponencial...* A continuación debemos incluir la variable que contiene los datos de la serie en la casilla *Variables* y manteniendo activado el modelo simple pinchar sobre el botón parámetros. Al hacerlo aparece un cuadro de diálogo en el que podemos determinar los valores de los parámetros α y \hat{X}_1 . En lo que se refiere a este último, el ajuste del primer valor de la serie, podemos dejar activada la opción *Automático de Valores iniciales*, en cuyo caso este valor de partida será la media de la serie, o activar la opción *Personalizado* e incluir un valor concreto que creamos más apropiado.

A la hora de determinar el α a emplear también existen dos posibilidades: proponer al programa un valor determinado del parámetro que queremos que utilice, o indicar al programa que realice una búsqueda en rejilla. En este último caso SPSS determina el valor de α minimizando el ECM (error cuadrático medio), función de pérdida que emplea el programa, entre todos los valores *Iniciar* y *Parar* que se generan empleando el paso indicado en *Por*.

Al ejecutar el menú correspondiente al suavizado exponencial simple, el programa SPSS crea dos nuevas variables llamadas FIT y ERR que corresponden al ajuste efectuado y al error cometido respectivamente. Los resultados generados al efectuar el suavizado de la serie Combustible efectuando una búsqueda en rejilla entre 0,01 y 0,99 por 0,05 (no permitimos los valores extremos 0 y 1) y dejando en *Automática* la selección del valor inicial son:

Descripción del modelo

| | | |
|-------------------|----------------|--|
| Nombre del modelo | | MOD_16 |
| Serie | 1 | Gasolina Nº de tep (tonelada equivalente de petróleo) de gasolina para automoción producidos en España(Fuente: Ministerio de Industria Comercio y Turismo) |
| Modelo simple | Tendencia | Ninguno |
| | Estacionalidad | Ninguno |

Aplicando las especificaciones del modelo de MOD 16

Estado de suavizado inicial

| | |
|-------|-----------|
| | Gasolina |
| Nivel | 799,89604 |

Sumas menores de los errores cuadráticos

| Serie | Rango del modelo | Alpha (Nivel) | Sumas de los errores cuadráticos |
|----------|------------------|---------------|----------------------------------|
| Gasolina | 1 | ,11000 | 1723447,19 |
| | 2 | ,06000 | 1724757,02 |
| | 3 | ,16000 | 1738049,66 |
| | 4 | ,21000 | 1753475,83 |
| | 5 | ,26000 | 1766888,59 |
| | 6 | ,31000 | 1778693,23 |
| | 7 | ,01000 | 1788982,41 |
| | 8 | ,36000 | 1790294,87 |
| | 9 | ,41000 | 1803331,42 |
| | 10 | ,46000 | 1819315,21 |

Parámetros del suavizado

| Serie | Alpha (Nivel) | Sumas de los errores cuadráticos | gl error |
|----------|---------------|----------------------------------|----------|
| Gasolina | ,11000 | 1723447,19 | 201 |

A continuación, se muestran los parámetros con las sumas menores de errores cuadráticos. Estos parámetros se utilizan para pronosticar.

La tabla correspondiente a la *Descripción del modelo* es un resumen de las características del modelo aplicado, en nuestro caso sin tendencia ni estacionalidad puesto que se trata de un modelo de suavizado exponencial simple.

La tabla *Estado de suavizado inicial* nos muestra el valor de la estimación (ajuste) del primer valor de la serie. En este caso es la media de la serie, puesto que en *Valores iniciales* se ha dejado activada la opción *Automático*.

La tabla *Sumas menores de los errores cuadráticos* nos muestra los diez valores más pequeños del ECM encontrados y sus correspondientes α entre todos los indicados en la búsqueda en rejilla. Si desactivamos la casilla *Mostrar solo los 10 mejores modelos de la búsqueda en rejilla*, también obtendríamos la siguiente tabla con el valor del ECM para todos los posibles valores de α indica-

dos en la búsqueda en rejilla, no solo los diez de ellos con valores más pequeños.

Búsqueda en rejilla de parámetros de suavizado

| Serie | Alpha (Nivel) | Sumas de los errores cuadráticos |
|----------|---------------|----------------------------------|
| Gasolina | ,01000 | 1788982,41 |
| | ,06000 | 1724757,02 |
| | ,11000 | 1723447,19 |
| | ,16000 | 1738049,66 |
| | ,21000 | 1753475,83 |
| | ,26000 | 1766888,59 |
| | ,31000 | 1778693,23 |
| | ,36000 | 1790294,87 |
| | ,41000 | 1803331,42 |
| | ,46000 | 1819315,21 |
| | ,51000 | 1839479,25 |
| | ,56000 | 1864753,68 |
| | ,61000 | 1895816,03 |
| | ,66000 | 1933172,04 |
| | ,71000 | 1977240,74 |
| | ,76000 | 2028431,67 |
| | ,81000 | 2087211,21 |
| | ,86000 | 2154160,08 |
| | ,91000 | 2230025,75 |
| | ,96000 | 2315774,42 |

La tabla parámetros de suavizado nos muestra el valor óptimo de α entre todos los examinados, así como el ECM que lleva asociado.

Para efectuar la predicción del siguiente mes para el que no disponemos de datos basta con aplicar la fórmula recursiva indicada anteriormente. Utilizando el programa SPSS al efectuar el suavizado exponencial en el botón *Guardar* debemos indicar el año y el mes hasta el que deseamos realizar pronósticos y obtendremos las correspondientes previsiones incluidas en la variable FIT.

Al igual que en los modelos ingenuos o en los de medias móviles, en los de suavizado exponencial simple las previsiones son constantes, es decir, se pronostica siempre con el mismo valor sea cual sea el horizonte de predicción.

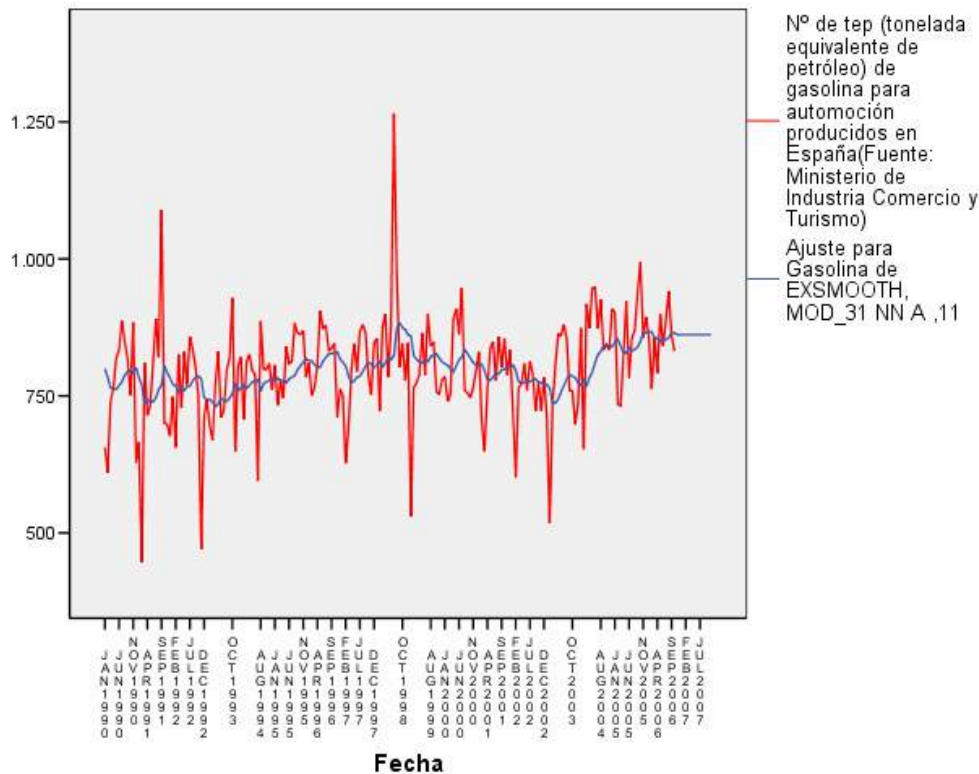


Figura 37: Gráfico de secuencia de la serie Combustible con el ajuste efectuado considerando un modelo exponencial simple (serie suavizada) y la previsión a varios meses vista.

MODELOS APLICABLES A SERIES NO ESTACIONALES CON TENDENCIA

En el caso de series temporales con tendencia lineal (creciente o decreciente) pero sin comportamiento estacional, el modelo clásico que más se suele aplicar es el de Holt. Es un modelo de suavizado exponencial doble y, al igual que en el caso de los modelos de suavizado exponencial simple, su fórmula de ajuste es recursiva:

$$\hat{X}_{n+1} = a_n + b_n$$

siendo:

$$\begin{cases} a_n = \alpha \cdot X_n + (1-\alpha) \cdot (a_{n-1} + b_{n-1}) & \alpha \in (0,1) \\ b_n = \gamma \cdot (a_n - a_{n-1}) + (1-\gamma) \cdot b_{n-1} & \gamma \in (0,1) \end{cases}$$

donde α es la constante de suavizado del nivel y γ la del escalón.

La fórmula de previsión de este modelo es:

$$f_{n,h} = a_n + h \cdot b_n$$

donde h es el momento hasta el que queremos predecir.

Para aplicar a una serie el método de suavizado exponencial doble de Holt empleando el programa SPSS, en primer lugar debemos seleccionar *Analizar – Series temporales - Suavizado exponencial...* A continuación debemos incluir la variable que contiene los datos de la serie en la casilla *Variables* y, manteniendo activado el modelo *Holt*, pinchar sobre el botón *Parámetros...* Al hacerlo aparece un cuadro de diálogo en el que podemos determinar los valores de los parámetros α , γ y los valores iniciales a_0 y b_0 . En lo que se refiere a estos últimos, podemos dejar activada la opción *Automático de Valores iniciales*, en cuyo caso estos valores de partida serán determinados por el SPSS, o activar la opción *Personalizado* e incluir valores concretos que creamos más apropiados.

Al igual que en el caso del suavizado exponencial simple SPSS nos proporciona dos opciones para determinar las constantes de suavizado: proporcionar al programa valores concretos o realizar una búsqueda en rejilla para localizar los valores de los coeficientes α y γ que minimicen el ECM. Si nos decantamos por esta última opción no conviene establecer rejillas demasiado finas, porque sino el programa debe examinar un número excesivo de modelos (demasiadas combinaciones de α y γ).

Al ejecutar el menú correspondiente al método de Holt, el programa crea dos nuevas variables llamadas FIT y ERR que corresponden al ajuste efectuado y al error cometido por el mismo respectivamente.

Consideremos la serie SuavizadoHolt, serie con información mensual no estacional y con tendencia creciente. Los resultados generados por el SPSS al ajustar esta serie efectuando una búsqueda en rejilla entre 0,01 y 0,99 por 0,1 tanto para determinar α como para establecer el valor de γ , y dejando en *Automática* la selección de los valores iniciales son:

Descripción del modelo

| | | |
|-------------------|----------------|---------|
| Nombre del modelo | | MOD_13 |
| Serie | 1 | Serie |
| Modelo de Holt | Tendencia | Lineal |
| | Estacionalidad | Ninguno |

Aplicando las especificaciones del modelo de MOD_13

Estado de suavizado inicial

| | |
|-----------|----------|
| | Serie |
| Nivel | 2571,104 |
| Tendencia | 17,41751 |

Sumas menores de los errores cuadráticos

| Serie | Rango del modelo | Alpha (Nivel) | Gamma (Tendencia) | Sumas de los errores cuadráticos |
|-------|------------------|---------------|-------------------|----------------------------------|
| Serie | 1 | ,21000 | ,01000 | 6850632,76 |
| | 2 | ,31000 | ,01000 | 6890834,73 |
| | 3 | ,41000 | ,01000 | 6956397,38 |
| | 4 | ,11000 | ,01000 | 6966523,47 |
| | 5 | ,51000 | ,01000 | 7080451,93 |
| | 6 | ,11000 | ,11000 | 7195698,29 |
| | 7 | ,61000 | ,01000 | 7299329,02 |
| | 8 | ,21000 | ,11000 | 7349480,20 |
| | 9 | ,31000 | ,11000 | 7502083,64 |
| | 10 | ,41000 | ,11000 | 7586003,07 |

Parámetros del suavizado

| Serie | Alpha (Nivel) | Gamma (Tendencia) | Sumas de los errores cuadráticos | gl error |
|-------|---------------|-------------------|----------------------------------|----------|
| Serie | ,21000 | ,01000 | 6850632,76 | 141 |

A continuación, se muestran los parámetros con las sumas menores de errores cuadráticos. Estos parámetros se utilizan para pronosticar.

Estas tablas de resultados son análogas a las obtenidas en el caso del suavizado exponencial simple, con la diferencia de que en lugar de un valor inicial aparecen dos y el resultado nos proporciona dos parámetros de suavizado.

Al igual que en el caso de suavizado exponencial simple si desactivamos la casilla Mostrar solo los 10 mejores modelos de la búsqueda en rejilla de la casilla Parámetros obtendremos una tabla con los valores del ECM para todas las posibles combinaciones de valores de α y γ en función de la rejilla indicada.

El botón Guardar es común a todos los análisis de series temporales que realiza el SPSS, y en él podemos elegir entre ajustar la serie (activando la opción *Desde el período de estimación hasta el último caso*) o predecir casos hasta la fecha que indiquemos al programa (eligiendo la opción *Pronosticar casos hasta*:). Por defecto SPSS no pronostica, solo ajusta.

Al contrario que en los modelos de suavizado exponencial simple, en el de Holt las previsiones conforman una recta.

Representando en un gráfico de secuencia la serie original, el ajuste y las predicciones logradas mediante el método de Holt obtenemos:

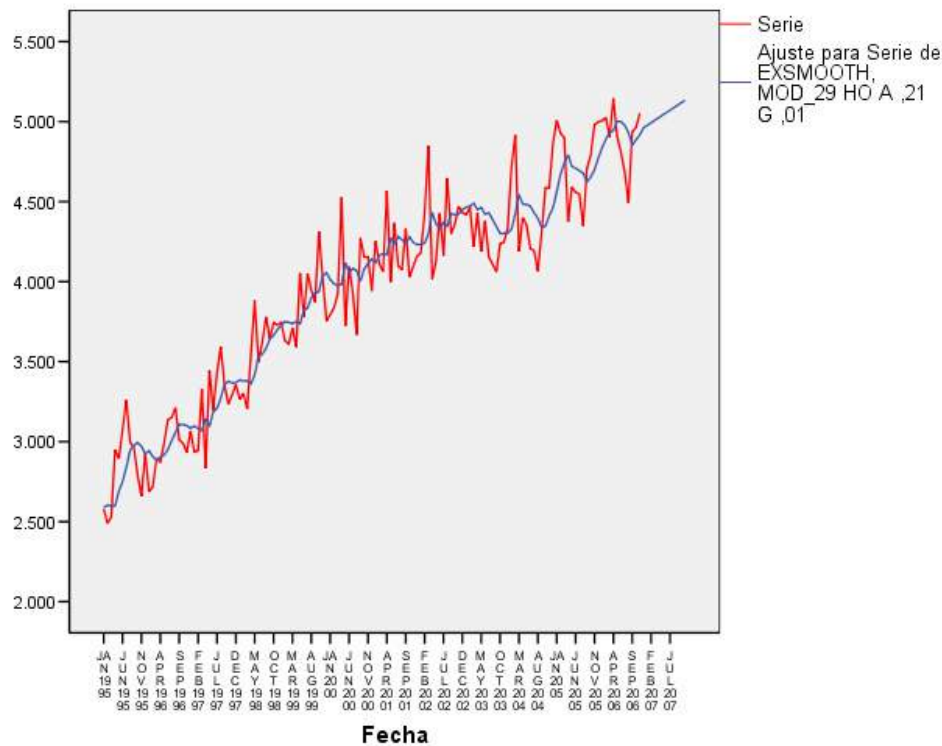


Figura 38: Gráfico de secuencia de la serie SuavizadoHolt con el ajuste efectuado considerando un modelo exponencial doble (serie suavizada) y la previsión a varios meses vista.

MODELOS APLICABLES A SERIES CON TENDENCIA Y ESTACIONALIDAD

En el caso de series temporales con tendencia lineal (creciente o decreciente) y comportamiento estacional, el modelo clásico que se aplica es el de Holt-Winters. Es una extensión del modelo de Holt, visto en el apartado anterior, que considera estacionalidad. La tendencia y la estacionalidad se pueden combinar de diferentes maneras, pero las que se consideran más frecuentemente son la combinación aditiva y la multiplicativa⁶.

Igual que en el caso de todos los modelos de suavizado exponencial vistos hasta ahora, la fórmula de ajuste del modelo de Holt-Winters es recursiva:

- Caso aditivo:

$$\hat{X}_{n+1} = a_n + b_n + S_{n-p}$$

siendo:

$$\begin{cases} a_n = \alpha \cdot (X_n - S_{n-p}) + (1 - \alpha) \cdot (a_{n-1} + b_{n-1}) & \alpha \in (0,1) \\ b_n = \gamma \cdot (a_n - a_{n-1}) + (1 - \gamma) \cdot b_{n-1} & \gamma \in (0,1) \\ S_t = \delta \cdot (X_n - a_n) + (1 - \delta) \cdot S_{n-p} & \delta \in (0,1) \end{cases}$$

⁶ Ver el apartado Análisis preliminar.

donde S_{n-p} representa al factor estacional para la misma estación pero un año antes, p es el periodo de la serie, α es la constante de suavizado del nivel, γ la del escalón y δ la de la estacionalidad.

La fórmula de previsión de este modelo es:

$$f_{n,h} = (a_n + h \cdot b_n) + S_{n+h-p}$$

donde h es el momento hasta el que queremos predecir.

- Caso multiplicativo:

$$\hat{X}_{n+1} = (a_n + b_n) \cdot S_{n-p}$$

siendo:

$$\begin{cases} a_n = \alpha \cdot \left(\frac{X_n}{S_{n-p}}\right) + (1 - \alpha) \cdot (a_{n-1} + b_{n-1}) & \alpha \in (0,1) \\ b_n = \gamma \cdot (a_n - a_{n-1}) + (1 - \gamma) \cdot b_{n-1} & \gamma \in (0,1) \\ S_t = \delta \cdot \left(\frac{X_n}{a_n}\right) + (1 - \delta) \cdot S_{t-p} & \delta \in (0,1) \end{cases}$$

donde S_{n-p} representa al factor estacional para la misma estación pero un año antes, p es el periodo de la serie, α es la constante de suavizado del nivel, γ la del escalón y δ la de la estacionalidad.

La fórmula de previsión de este modelo es:

$$f_{n,h} = (a_n + h \cdot b_n) \cdot S_{n+h-p}$$

donde h es el momento hasta el que queremos predecir.

Para aplicar a una serie el método de suavizado exponencial de Holt-Winters empleando SPSS, en primer lugar debemos seleccionar el menú *Analizar – Series temporales – Suavizado exponencial...* A continuación tenemos que incluir la variable que contiene los datos de la serie en la casilla *Variables* e indicar el modelo a considerar. Si es multiplicativo, en el recuadro *Modelo* seleccionamos *Winters*, y si es aditivo activamos la opción *Personalizado*, y en el botón *Personalizar* escogemos *Componente de tendencia lineal* y *Componente estacional aditivo*. Una vez seleccionado el modelo a emplear, pinchamos sobre el botón *Parámetros...* Al hacerlo aparece un cuadro de diálogo en el que podemos determinar los valores de los parámetros α , γ , δ y los valores iniciales a_0 , b_0 y S_0^1, \dots, S_0^p (valores iniciales del factor estacional para cada una de las estaciones, periodos, de la serie). En lo que se refiere a estos últimos, podemos dejar activada la opción *Automático de Valores iniciales*, en cuyo caso estos valores de partida serán determinados por el SPSS, o activar la opción *Per-*

sonalizado e incluir valores concretos que creamos más apropiados para a_0 (nivel) y b_0 (tendencia). En el caso de los S_0 debemos crear en el archivo de datos una variable que los contenga. Esta variable debe tener tantos valores como factores estacionales tengamos en la serie (p). A continuación debemos introducir esta variable en la casilla *Factores estacionales* del cuadro de diálogo principal para indicar al programa que tome estos valores como valores de partida de S .

El programa SPSS no nos permite aplicar el modelo de Holt-Winters a una serie para la que no se ha definido una variable periódica (correspondiente a meses, trimestres, cuatrimestres...). En estos casos aparece desactivada la opción *Winters* en el apartado *Modelo* del cuadro de diálogo del suavizado exponencial. SPSS tampoco nos deja aplicar el modelo de Holt-Winters a una serie cuando no hay, como mínimo, cuatro periodos completos (cuatro años si los datos son mensuales, cuatro semanas si son diarios, etc.).

Al ejecutar el menú correspondiente al método de Holt-Winters, tanto en el caso aditivo como en el multiplicativo, el programa crea dos nuevas variables llamadas FIT y ERR que corresponden al ajuste efectuado y al error cometido por el mismo respectivamente.

Consideremos la serie Papel, serie con información trimestral sobre producción de papel, estacional y con tendencia creciente. Realizando un análisis preliminar de la serie concluimos que no parece existir dependencia entre variabilidad y nivel, de manera que consideraremos que tendencia y estacionalidad se combinan de forma aditiva. Los resultados generados por el SPSS al ajustar esta serie considerando un modelo aditivo y efectuando una búsqueda en rejilla entre 0,01 y 0,99 por 0,1 tanto para determinar α como para establecer el valor de γ y δ , y dejando en *Automática* la selección de los valores iniciales, son análogos a los obtenidos al ajustar otros modelos de suavizado exponencial:

Descripción del modelo

| | | |
|---------------------------------|-----------------------------|--------------------------------------|
| Nombre del modelo | | MOD_2 |
| Serie | 1 | Papel Producción trimestral de papel |
| Modelo aditivo de Winters | Tendencia Estacionalidad | Lineal Aditivo |
| Longitud del periodo estacional | | 4 |

Aplicando las especificaciones del modelo de MOD_2

Estado de suavizado inicial

| | | Papel |
|--------------|---|-----------|
| Índices | 1 | 442,18480 |
| estacionales | 2 | 341,71173 |
| | 3 | -896,248 |
| | 4 | 112,35141 |
| Nivel | | 2662,228 |
| Tendencia | | 39,55301 |

Sumas menores de los errores cuadráticos

| Serie | Rango del modelo | Alpha (Nivel) | Gamma (Tendencia) | Delta (Estación) | Sumas de los errores cuadráticos |
|-------|------------------|---------------|-------------------|------------------|----------------------------------|
| Papel | 1 | ,21000 | ,01000 | ,91000 | 1606687,12 |
| | 2 | ,31000 | ,01000 | ,91000 | 1611388,28 |
| | 3 | ,21000 | ,01000 | ,81000 | 1620179,67 |
| | 4 | ,31000 | ,01000 | ,81000 | 1630482,97 |
| | 5 | ,11000 | ,01000 | ,91000 | 1642031,84 |
| | 6 | ,21000 | ,01000 | ,71000 | 1642967,56 |
| | 7 | ,11000 | ,01000 | ,81000 | 1647363,64 |
| | 8 | ,41000 | ,01000 | ,91000 | 1650512,46 |
| | 9 | ,31000 | ,01000 | ,71000 | 1659694,37 |
| | 10 | ,11000 | ,01000 | ,71000 | 1660222,69 |

Parámetros del suavizado

| Serie | Alpha (Nivel) | Gamma (Tendencia) | Delta (Estación) | Sumas de los errores cuadráticos | gl error |
|-------|---------------|-------------------|------------------|----------------------------------|----------|
| Papel | ,21000 | ,01000 | ,91000 | 1606687,12 | 43 |

A continuación, se muestran los parámetros con las sumas menores de errores cuadráticos. Estos parámetros se utilizan para pronosticar.

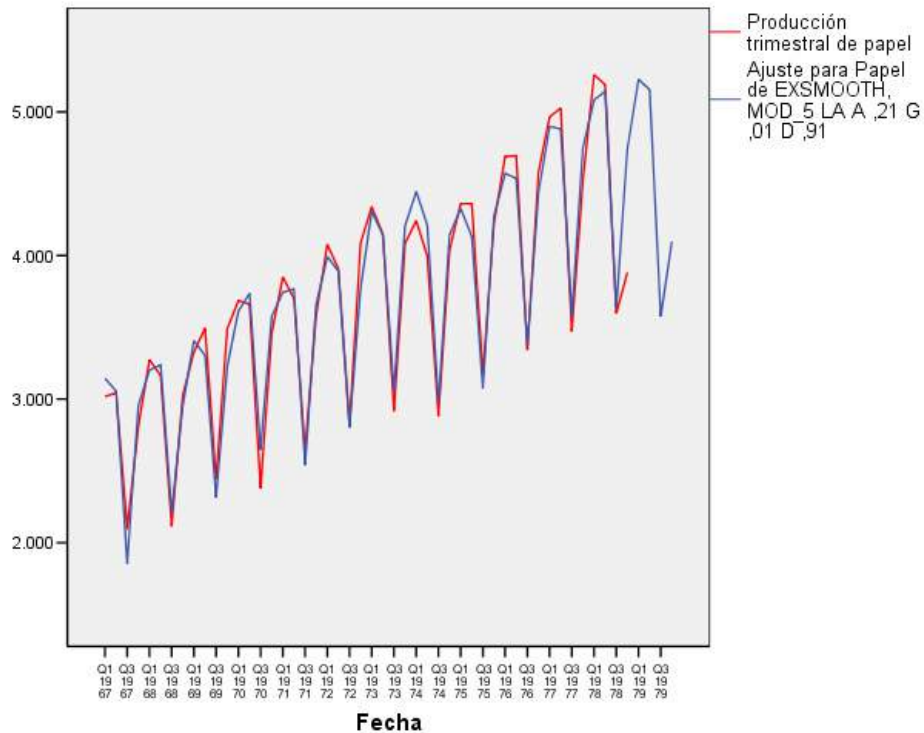


Figura 39: Gráfico de secuencia de la serie Papel con el ajuste efectuado considerando un modelo de Holt-Winters aditivo y la previsión a un año vista.

Consideremos ahora la serie Airline. Tal y como vimos en el apartado de análisis preliminar esta serie es claramente estacional y de tendencia creciente. Tanto observando su gráfico de secuencia como su correspondiente gráfico de dispersión por nivel, queda bastante claro que la estacionalidad y la tendencia se combinan de forma multiplicativa.

Los resultados generados por el SPSS al ajustar esta serie considerando un modelo aditivo y efectuando una búsqueda en rejilla entre 0,01 y 0,99 por 0,1 tanto para determinar α como para establecer el valor de γ y δ , y dejando en *Automática* la selección de los valores iniciales, son:

| Descripción del modelo | | |
|----------------------------------|-----------------------------|----------------------------------|
| Nombre del modelo | | MOD_29 |
| Serie | 1 | NumPasajeros Número de pasajeros |
| Modelo multiplicativo de Winters | Tendencia Estacionalidad | Lineal Multiplicativo |
| Longitud del periodo estacional | | 12 |

Aplicando las especificaciones del modelo de MOD_29

Estado de suavizado inicial

| | | Num Pasajeros |
|-------------------------|----|------------------|
| Índices estacionales | 1 | 91,18565 |
| | 2 | 88,22166 |
| | 3 | 100,80678 |
| | 4 | 97,29508 |
| | 5 | 98,12681 |
| | 6 | 111,33470 |
| | 7 | 123,13503 |
| | 8 | 121,46574 |
| | 9 | 105,79046 |
| | 10 | 92,16911 |
| | 11 | 80,28828 |
| | 12 | 90,18069 |
| Nivel | | 110,78030 |
| Tendencia | | 2,64773 |

Sumas menores de los errores cuadráticos

| Serie | Rango del modelo | Alpha (Nivel) | Gamma (Tendencia) | Delta (Estación) | Sumas de los errores cuadráticos |
|--------------|------------------|---------------|----------------------|---------------------|--|
| NumPasajeros | 1 | ,81000 | ,01000 | ,01000 | 16171,54550 |
| | 2 | ,71000 | ,01000 | ,01000 | 16199,74848 |
| | 3 | ,91000 | ,01000 | ,01000 | 16469,89221 |
| | 4 | ,81000 | ,01000 | ,11000 | 16549,96064 |
| | 5 | ,61000 | ,01000 | ,01000 | 16581,46507 |
| | 6 | ,91000 | ,01000 | ,11000 | 16616,58962 |
| | 7 | ,91000 | ,01000 | ,21000 | 16786,28523 |
| | 8 | ,71000 | ,01000 | ,11000 | 16927,10486 |
| | 9 | ,91000 | ,01000 | ,31000 | 16973,47493 |
| | 10 | ,81000 | ,01000 | ,21000 | 17016,56711 |

Parámetros del suavizado

| Serie | Alpha (Nivel) | Gamma (Tendencia) | Delta (Estación) | Sumas de los errores cuadráticos | gl error |
|--------------|---------------|----------------------|---------------------|--|----------|
| NumPasajeros | ,81000 | ,01000 | ,01000 | 16171,54550 | 131 |

A continuación, se muestran los parámetros con las sumas menores de errores cuadráticos. Estos parámetros se utilizan para pronosticar.

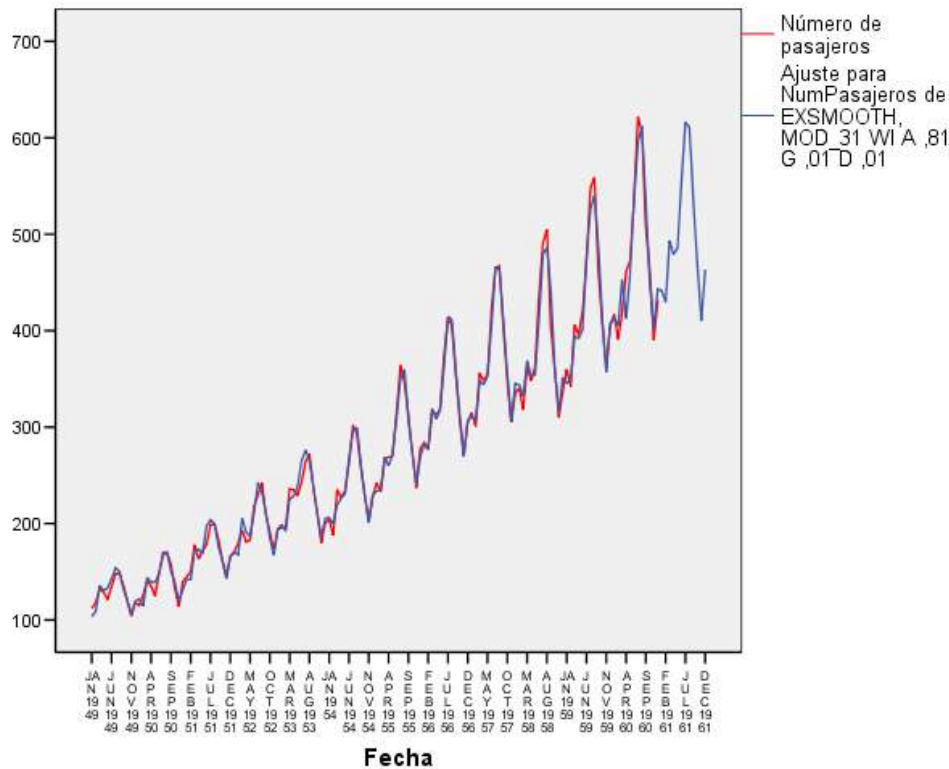


Figura 40: Gráfico de secuencia de la serie Airline con el ajuste efectuado considerando un modelo de Holt-Winters multiplicativo y la previsión a un año vista.

Las tablas de resultados obtenidas en ambos casos son análogas a las resultantes en el caso de suavizado exponencial doble, con la diferencia de que, además de los valores iniciales de nivel y tendencia, aparecen p (periodo de la serie) valores iniciales para la componente estacional. Además en este caso obtenemos el valor de tres parámetros de suavizado.

4.- MODELOS ARIMA

Al estudiar los métodos clásicos de análisis de series temporales no hemos tenido en cuenta de dónde salen los datos de las mismas, cuál es el mecanismo que las genera. A la hora de estudiar una serie temporal empleando los modelos ARIMA suponemos que la serie puede estar generada por un proceso estocástico.

Un proceso estocástico $\langle X_t \rangle_{t \in \mathbb{Z}}$ se define como una familia de variables aleatorias⁷ que corresponden a momentos sucesivos en el tiempo. De esta forma una

⁷ Una variable es aleatoria si toma diferentes valores como resultado de un experimento aleatorio. Se puede pensar en una variable aleatoria como un valor o una magnitud que cambia de una presentación a otra, sin seguir una secuencia predecible. Los valores de una variable aleatoria son los valores numéricos correspondientes a cada posible resultado de un experimento aleatorio (realización).

serie temporal con n observaciones estaría generada por X_1, \dots, X_n variables aleatorias, y sus valores concretos x_1, \dots, x_n serían una realización del proceso estocástico.

Un proceso estocástico queda determinado si conocemos la función de distribución de las variables aleatorias que lo componen y todas las posibles funciones de distribución conjuntas de dichas variables.

$$\begin{aligned} & - F(X_t) \quad \forall t \in \mathbb{Z} \\ & - F(X_{t_1}, \dots, X_{t_n}) \quad \forall t_1, \dots, t_n \in \mathbb{Z} \quad \forall n \in \mathbb{N} \end{aligned}$$

La determinación de las características de un proceso estocástico a partir de las funciones de distribución es, en general, un procedimiento complicado, por lo que se acostumbra a caracterizarlos a partir de los momentos de primer y segundo orden (media y covarianza).

$$\text{Media o momento de primer orden:} \quad \mu_t = E(X_t) \quad t \in \mathbb{Z}$$

$$\text{Covarianzas o momentos de segundo orden}^8: \quad \gamma_{t,s} = \text{cov}(X_t, X_s) \quad t, s \in \mathbb{Z}$$

Con el número de datos obtenidos en una sola realización del proceso estocástico se debe estimar un número superior de parámetros. De esta forma, si se dispone de n datos, con ellos se deberían estimar n medias, n varianzas y todas las covarianzas. En otras palabras, el problema no tiene ninguna solución. Para poder, a partir de una sola realización, efectuar inferencias sobre los parámetros de un proceso estocástico, es preciso imponer restricciones a este último. Las restricciones que se imponen habitualmente son que sea estacionario y ergódico.

Un proceso estocástico $\langle X_t \rangle_{t \in \mathbb{Z}}$ es estacionario en sentido estricto si se mantienen estables todas las posibles relaciones entre las variables aleatorias que lo componen, de manera que solo dependen de la distancia que las separa:

$$\begin{aligned} & - F(X_t) = F(X) \quad \forall t \in \mathbb{Z} \\ & - F(X_{t_1}, \dots, X_{t_n}) = F(X_{t_1+k}, \dots, X_{t_n+k}) \quad \forall t_1, \dots, t_n \in \mathbb{Z}, \forall k \in \mathbb{Z}, \forall n \in \mathbb{N} \end{aligned}$$

Restringirnos a los procesos estacionarios estrictos es excesivo, de manera que se opta por una solución intermedia, con los llamados procesos estocásticos estacionarios en sentido débil o, simplemente, procesos estacionarios.

Un proceso es estacionario en sentido débil cuando su media y su varianza permanecen constantes a lo largo del tiempo y las relaciones lineales entre las variables solo dependen de la distancia que las separa:

$$\begin{aligned} & - E(X_t) = \mu \quad \forall t \in \mathbb{Z} \\ & - \text{Var}(X_t) = \sigma^2 \quad \forall t \in \mathbb{Z} \\ & - \text{Cov}(X_{t_1}, X_{t_2}) = \text{Cov}(X_{t_1+k}, X_{t_2+k}) \quad t_1, t_2 \in \mathbb{Z}, \forall k \in \mathbb{Z} \end{aligned}$$

⁸ OBSERVACIÓN: si $t=s$ la covarianza pasa a ser la varianza. $\text{Cov}(X_t, X_t) = \text{Var}(X_t)$

Antes de introducir la definición de proceso estocástico ergódico necesitamos definir la función de autocorrelación.

Se denomina autocorrelación de orden k , ρ_k , a la correlación de cualesquiera dos variables aleatorias del proceso estocástico, distanciadas k instantes de tiempo:

$$\rho_k = \frac{\text{Cov}(X_t, X_{t+k})}{\sqrt{\text{Var}(X_t)} \cdot \sqrt{\text{Var}(X_{t+k})}} = \frac{\gamma_k}{\gamma_0} \quad \forall k \in \mathbb{Z}, \quad \forall t \in \mathbb{Z}$$

Propiedades de la autocorrelación:

- $\rho_i = \rho_{-i}$
- $\rho_0 = 1$
- $|\rho_k| \leq 1$

La función de autocorrelación simple es la representación de ρ_k frente a k . Como la función de autocorrelación simple es una simétrica debido la primera de sus propiedades, únicamente se suele representar la parte positiva.

Se denomina autocorrelación parcial de orden k , Φ_{kk} , a la correlación entre dos variables aleatorias del proceso estocástico distanciadas k instantes de tiempo, pero sin considerar la existencia de las variables aleatorias intermedias. Es decir, para calcular la autocorrelación parcial entre X_t y X_{t+k} eliminamos la influencia que sobre ambas tienen $X_{t+1}, X_{t+2}, \dots, X_{t+k-1}$.

$$\phi_{kk} = \text{corr}(X_t, X_{t+k} \mid X_{t+1}, X_{t+2}, \dots, X_{t+k-1})$$

La función de autocorrelación parcial es la representación de Φ_{kk} frente a k .

Las funciones de autocorrelación simple y parcial, tal y como veremos más adelante, constituyen uno de los instrumentos clave para ajustar el modelo que genera una serie temporal.

Un proceso es ergódico cuando conforme k se hace más grande la autocorrelación, ρ_k , se hace más pequeña. Es decir, que lo que ocurre hoy, conforme va pasando el tiempo va teniendo menos importancia.

$$\rho_k \xrightarrow{k \rightarrow +\infty} 0$$

Cuando valores de la serie temporal alejados en el tiempo están muy correlacionados, es decir, cuando ρ_k se mantiene en unas cotas elevadas para un k grande, sucede que al aumentar el tamaño de la muestra se añade poca información nueva. La consecuencia de este hecho en el plano estadístico es que los estimadores obtenidos no son consistentes, ya que el aumento del tamaño de la muestra no tendrá una especial utilidad, puesto que se deberá calcular un mayor número de covarianzas para caracterizar el proceso.

De esta forma, cuando estamos ante un proceso estacionario y ergódico, todo el problema de inferencia se simplifica de forma considerable.

Nuestro objetivo al analizar una serie temporal es estimar el proceso estocástico que la genera y para ello, según hemos visto hasta ahora, debemos partir del supuesto de que dicho proceso estocástico es estacionario y ergódico.

Modelos de medias móviles (MA)

$$X_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}$$

Un proceso de medias móviles de orden q es un proceso en el que la variable X_t se obtiene como un promedio de variables de ruido blanco (a_i), siendo los θ_i sus coeficientes de ponderación.

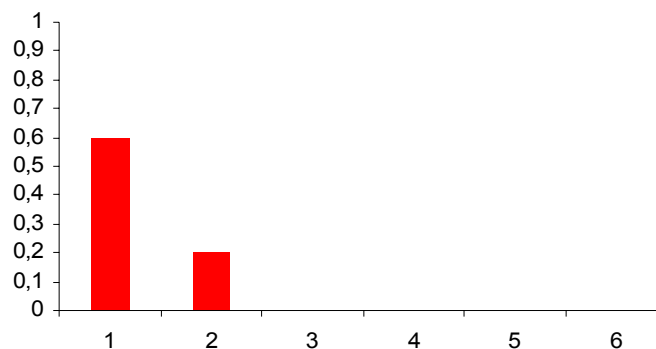
Todos los procesos de medias móviles son procesos estacionarios.

Existen $2q$ procesos de medias móviles de orden q que poseen la misma función de autocorrelación, pero solo uno de ellos es invertible. De esta manera si solo consideramos procesos invertibles la función de autocorrelación determina unívocamente un proceso.

- La función de autocorrelación simple de un modelo de medias móviles se corta (se hace cero) en el orden del modelo (q) y caracteriza los procesos de medias móviles.
- La función de autocorrelación parcial de un modelo de medias móviles no se corta, tiende a cero rápidamente, exponencialmente.

Veamos un ejemplo de funciones de autocorrelación de un modelo de medias móviles de orden dos:

Función de autocorrelación simple (FAS)



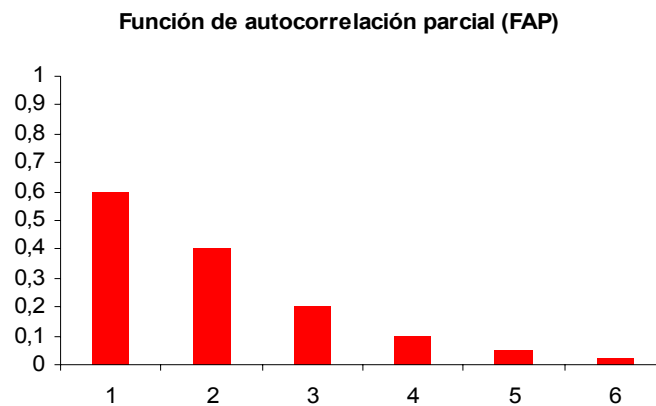


Figura: Funciones de autocorrelación simple y parcial de un proceso de medias móviles de orden 2.

Hay que tener en cuenta que los coeficientes de correlación de ambas funciones pueden tomar valores positivos o negativos.

Modelos autorregresivos (AR)

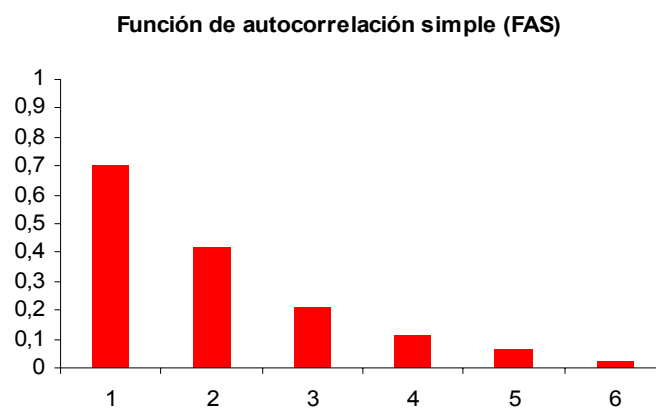
$$X_t = a_t + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p}$$

Un proceso autorregresivo de orden p es un proceso en el que la variable X_t se obtiene efectuando una regresión sobre valores pasados de la misma.

Todos los procesos autorregresivos son procesos invertibles.

- La función de autocorrelación simple de un modelo autorregresivo no se corta, tiende a cero rápidamente, exponencialmente.
- La función de autocorrelación parcial de un modelo autorregresivo se corta (se hace cero) en el orden del modelo (p).

Veamos un ejemplo de funciones de autocorrelación de un modelo autorregresivo de orden dos:



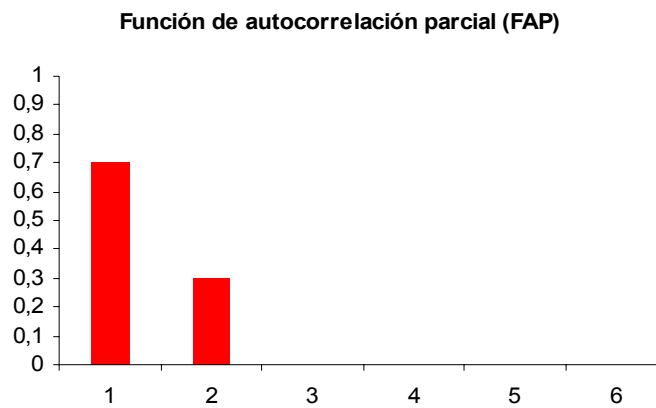


Figura: Funciones de autocorrelación simple y parcial de un proceso autorregresivo de orden 2. Hay que tener en cuenta que los coeficientes de correlación de ambas funciones pueden tomar valores positivos o negativos.

Modelos autorregresivos de medias móviles ARMA

$$X_t = a_t + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}$$

Los modelos autorregresivos de medias móviles, ARMA(p,q), son la suma de un proceso autorregresivo de orden p y uno de medias móviles de orden q.

Las funciones de autocorrelación simple y parcial de un proceso ARMA también son la suma de las correspondientes a un proceso MA y a otro AR. De esta forma ninguna de las dos se corta.

- La función de autocorrelación simple de un proceso ARMA en cuanto supera el orden de la parte MA, q, se comporta como si solo hubiera parte AR.
- La función de autocorrelación parcial se comporta como si solo hubiera parte MA en cuanto superamos el orden de la parte AR, p.

Estas características de las funciones de autocorrelación dificultan la identificación del proceso a través de las mismas, de manera que no es posible identificar p y q al mismo tiempo. En un primer paso debemos identificar un proceso AR (o un MA) y, tras ajustar la serie, identificar en los residuos un MA (o un AR).

Modelos no estacionarios (ARIMA)

Existe gran cantidad de series temporales que no son estacionarias, de manera que no podemos analizarlas ajustando los modelos vistos hasta ahora, puesto que exigen estacionariedad (nivel de la serie y variabilidad constantes a lo largo del tiempo).

Sin embargo es posible transformar las series no estacionarias para que verifiquen los supuestos que necesitamos. En primer lugar debemos analizar la dependencia entre variabilidad y nivel. Si existe dependencia de este tipo la serie no es estacionaria en varianza, de manera que debemos transformarla. En estos casos, cuando la variabilidad depende del nivel, se realizan transformaciones de la familia de box-cox. Cuando estamos ante una serie no estacionaria en nivel (serie con tendencia) se consigue que la serie sea estacionaria diferenciándola. Diferenciar una serie X_t (integrarla) consiste en aplicarle el operador diferencia simple (∇) que hace lo siguiente:

$$\nabla X_t = X_t - X_{t-1}$$

Es posible que, tras diferenciar una serie, esta siga siendo no estacionaria, en cuyo caso es posible que necesite una nueva diferencia. De esta manera la notación que se emplea con estos modelos es ARIMA(p,d,q), donde d es el número de diferencias que se efectúan sobre la serie original antes de ajustar un ARMA.

BIBLIOGRAFÍA

- Carmen Rodríguez Morilla. (2000). *Análisis de series temporales*. La muralla.
- Ezequiel Uriel. (1985). *Análisis de series temporales. Modelos ARIMA*. Paraninfo.
- Antonio Aznar, Francisco Javier Trivez. (1993). *Métodos de predicción en economía II. Análisis de series temporales*. Ariel Economía.
- Quintín Martín Martín, María Teresa Cabero Morán, Ramón Ardanuy Albajar. *Paquetes estadísticos SPSS 8.0*. Hespérides.