

CAPITULO 2. La simulación y predicción en contextos con información histórica. Modelos ARIMA. Modelos de vectores autorregresivos (VAR). Modelos de vectores de corrección del error (VEC). Modelos autorregresivos y condicionales heterocedásticos (ARCH).

Bibliografía Básica

- Pulido San Román, A. (2004) Curso combinado de Predicción y Simulación. www.uam.es/predysim. Universidad Autónoma de Madrid. Capítulo 3, 4 y 5
- Gujarati, D. (2006). Econometría. 4ª Edición. Mc Graw Hill. México. Capítulos 21 y 22.
- Pindyck, R. y D. Rubinfeld (2001). McGraw Hill. 4ª Edición. Parte 4, Capítulos 15 a 19.
- Pérez López, C. (2006) Econometría. Conceptos y Problemas resueltos de Econometría. Thompson. Capítulo 4.

A. MODELOS ARIMA

- En este acápite abordaremos el estudio de la llamada metodología ARIMA o Box-Jenkins por el nombre de los dos investigadores que sintetizaron y desarrollaron estas técnicas tal y como las conocemos hoy. Se trata de una metodología que ha tenido un indudable éxito en la práctica profesional por varios motivos: En primer lugar por su rotundidad metodológica. Se constituye como una técnica avanzada que hace uso de sofisticados recursos matemático-estadísticos. En segundo lugar, existe una clara y consolidada guía de aplicación empírica de la misma que permite pasar con facilidad de las situaciones de laboratorio que crea la teoría, a la praxis profesional. En tercer lugar y no menos importante, los Modelos ARIMA han demostrado una gran utilidad en la predicción a corto plazo de series de alta frecuencia. Ese es su campo natural de aplicación.
- Desarrollaremos la metodología completa ARIMA a nivel básico haciendo especial hincapié en las posibilidades de aplicación práctica. Por ello, cada concepto se desarrolla oportunamente con un ejemplo de aplicación en EViews.
- En su conjunto, el acápite se desarrolla en 9 apartados. Cuando es oportuno, los conceptos se desarrollan en breves explicaciones adicionales. Se propone y resuelve 1 ejercicio en EViews, con sus correspondientes soluciones ilustradas. Por último, se solicita la realización de una actividad junto con el test y preguntas de autoevaluación. Este acápite le exigirá al usuario la dedicación de unas 28 horas; 13 para el aprendizaje de los conceptos teóricos, 5 para la realización de los 2 ejercicios propuestos, 4 para la realización del caso de aplicación y 6 para las actividades y test de autoevaluación.

A.1. Metodología de Box Jenkins

Resumen:

- En Econometría Avanzada se repasan, en las aplicaciones prácticas, los conceptos de modelos AR, MA y ARMA ya visto en el curso de Econometría.

- Box y Jenkins diseñaron una metodología para el tratamiento de la modelización de series temporales univariantes basada en las siguientes fases:

1) Recolección de datos. Es conveniente disponer de 50 o más datos, y en el caso de series mensuales, es habitual trabajar con entre seis y diez años completos de información.

2) Representación gráfica de la serie. Para decidir sobre la estacionariedad y estacionalidad de la serie es de gran utilidad disponer de un gráfico de la misma. Suelen utilizarse medias y desviaciones típicas por subperíodo para juzgar sobre la estacionariedad de la serie.

3) Transformación previa de la serie y eliminación de la tendencia: La transformación logarítmica es necesaria para series no estacionarias en varianza y es muy frecuente en series con dispersión relativamente constante en el tiempo. Una posibilidad práctica es ensayar siempre con la serie original y en logaritmos y comprobar los resultados. La observación del gráfico de la serie nos indicará la existencia o no de tendencia.

4) Identificación del Modelo. Consiste en determinar el tipo de modelo más adecuado para la serie, es decir, el orden de los procesos autorregresivos y de medias móviles de las componentes regular y estacional. Técnicamente esta decisión se tomará en base a las funciones de autocorrelación y autocorrelación parcial.

5) Estimación de los coeficientes del Modelo. Decidido el modelo, se precede a la estimación de sus parámetros. Dado que se trata de un procedimiento iterativo de cálculo, pueden sugerirse valores iniciales.

6) Contraste de validez conjunta del modelo. Utilizaremos diversos procedimientos para valorar el modelo o modelos inicialmente seleccionados: contraste de significación de parámetros, covarianzas entre estimadores, coeficiente de correlación, suma de cuadrados de errores, etc.

7) Análisis detallado de los errores. Las diferencias históricas entre valores reales y estimados por el modelo constituyen una fuente de especial

interés para una valoración final del modelo. Deberá comprobarse un comportamiento no sistemático de los mismos, así como analizarse la posible existencia de errores especialmente significativos.

8) Selección del modelo y predicción. En base a las etapas anteriores se selecciona el modelo y se utilizará como forma inicial de predicción.

1.Recolección de Datos.

Inicialmente se puede definir una **serie temporal** como una sucesión de valores en el tiempo. Designaremos la serie temporal por $\{y_t\}$ donde el subíndice t representa el tiempo medido en días, meses, años, etc.

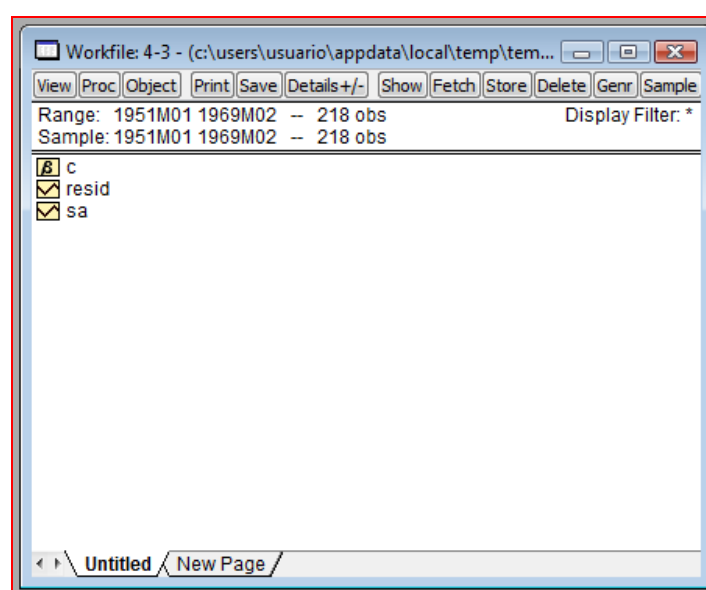
El archivo 4.3.wf1 contiene una variable de nombre SA que representa una serie de ratios mensuales correspondientes a la producción de una empresa. Con la finalidad de realizar predicciones futuras de producción se trata de ajustar un modelo ARIMA convenientemente.

En el sitio <http://www.thomsonparaninfo.com>. EN MATERIAL DE APOYO Y GUÍAS DIDÁCTICAS, ELEGIR EL TEMA ESTADÍSTICA (CLICK EN BUSCAR, LUEGO EN PROBLEMAS RESUELTOS DE ECONOMETRÍA) el usuario podrá bajar el archivo mencionado haciendo CLICK en

Conjunto de datos en formato EViews. Capítulo 4

[Descargar](#)

Guarde el archivo en un lugar adecuado. Aparecerá una carpeta en formato ZIP, ábrala y seleccione 4-3. Se encontrará con el WORKFILE 4-3 listo para trabajar.



Siga los pasos teóricos y prácticos para ir comprendiendo como trabajar con Series Temporales, especificar, estimar y ajustar un modelo ARIMA.

En nuestro caso la serie $\{y_t\}$ se denomina SA, es decir $\{y_t\} = \{SA_t\}$ cuyos valores los podemos encontrar en nuestro WORKFILE seleccionando SA y haciendo **QUICK → SHOW → OK**

Workfile: 4-3 - (c:\users\usuario\appdata\local\temp\tem...

View Proc Object Print Save Details+/- Show Fetch Store Delete Genr Sample

Range: 1951M01 1969M02 -- 218 obs Display Filter: *

Sample: 1951M01 1969M02 -- 218 obs

☐ c
☒ resid
☒ sa

Series: SA Workfile: 4-3::Untitled\

View Proc Object Properties Print Name Freeze Default Sort Edit+/- Smpl+/-

SA

Last updated: 08/03/05 - 17:17

Modified: 1951M01 1969M02 // seas(m) sa

1951M01	1.187660
1951M02	1.166821
1951M03	1.217199
1951M04	1.251932
1951M05	1.228738
1951M06	1.211244
1951M07	1.165150
1951M08	1.199093
1951M09	1.225009
1951M10	1.191793
1951M11	1.378378
1951M12	1.269002
1952M01	1.221476
1952M02	1.334204
1952M03	

Como verá se trata de una serie mensual de 218 observaciones que van desde 1951M01 a 1969M02 (es decir, Enero de 1951 a Febrero de 1969).

La teoría clásica considera una serie de tiempo formada por cuatro componentes teóricas: **tendencia**, **variaciones estacionales**, **variaciones cíclicas** y **variaciones residuales**.

La *tendencia* viene dada por el movimiento general a largo plazo de la serie. Dados los valores de la serie temporal podemos usar varios métodos para estudiar su tendencia. Entre ellos se destacan el método del ajuste analítico, el método de las medias móviles (recuerde lo ya analizado en el capítulo anterior) y el método de las diferencias.

Las *variaciones estacionales* son oscilaciones que se producen con un período igual o inferior a un año, y que se reproducen de manera reconocible en los diferentes años. El motivo principal que induce a estudiar la componente estacional es que en la inmensa mayoría de las series económicas dicha componente provoca una distorsión de su verdadero movimiento. Para eliminar estas distorsiones y captar el movimiento real de la serie es necesario aplicar criterios de desestacionalización. Es un proceso no sencillo que ha dado lugar a múltiples análisis entre los que se destacan los programas X11 y X12 del *Bureau of the Census* de Estados Unidos.

Las *variaciones cíclicas* son oscilaciones que se producen con un período superior a un año y que se deben, principalmente, a la alternancia de etapas largas (ciclos) en las que se repite el comportamiento de la serie. Es la más difícil de detectar, pues a diferencia de la tendencia, que es un movimiento a largo plazo muy general, y de las variaciones estacionales, que tienen un período fijo, estas variaciones tienen un período no fácilmente identificable y en muchos casos es variable, siendo frecuente la existencia de ciclos que se superponen, lo que hace todavía más difícil su identificación. Como vimos en otro capítulo, en la práctica suele eliminarse la tendencia y las variaciones estacionales y estudiar la parte restante de la serie. Los métodos para este estudio son varios y se destaca el análisis armónico y filtros como el de HODRICK PRESCOTT.

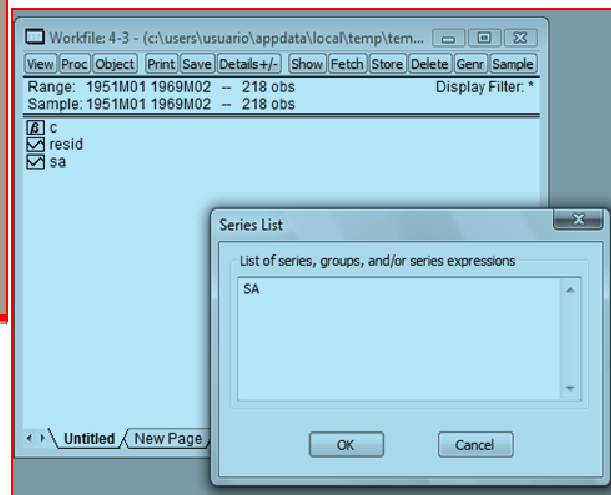
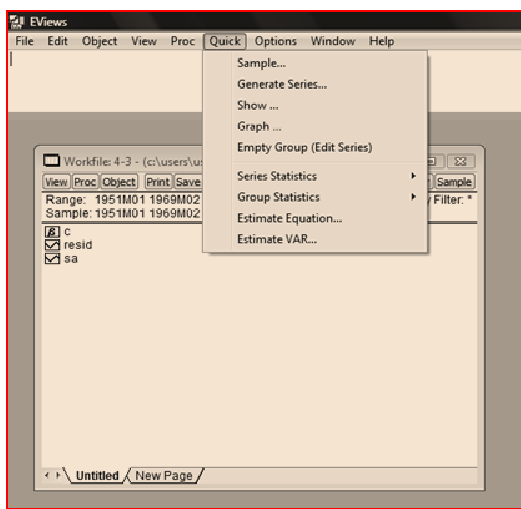
Para introducir el concepto de *variaciones residuales* debemos decir que la modelización ARIMA (AUTOREGRESIVE INTEGRATED MOVING AVERAGE) o Box Jenkins parte de considerar que el valor observado de una serie (un dato de una variable económica) en un momento determinado de tiempo t es una realización de una variable aleatoria y_t definida en dicho momento de tiempo. Por tanto, una serie de t datos $\{y_t\}$, es una muestra de un vector de t variables aleatorias ordenadas en el tiempo al que denominamos **proceso estocástico**. En ocasiones pretendemos predecir el comportamiento de una variable y_t en un momento futuro t , a partir del comportamiento que la variable tuvo en un momento pasado, por ejemplo, en el período anterior y_{t-1} . Formalmente notaríamos que $y_t = f(y_{t-1})$, es decir, que el valor de la variable en el momento t es función del valor tomado en el período $t - 1$. Puesto que en el comportamiento de una variable influyen más aspectos, debemos incluir en la relación anterior un término de error o residuo, a_t , que es una variable aleatoria a la que suponemos ciertas características estadísticas apropiadas. Por lo tanto, la variable depende de su valor anterior y de una componente residual $y_t = f(y_{t-1}, a_t)$.

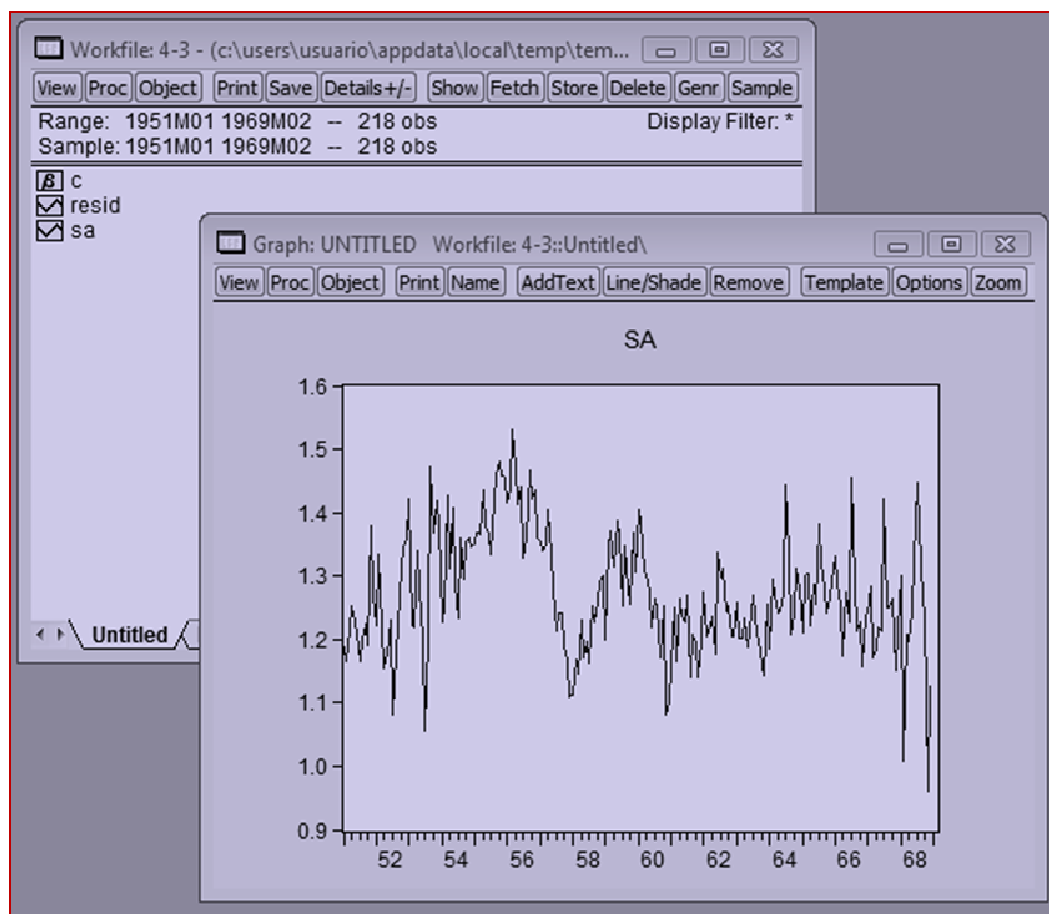
2.Representación Gráfica de la Serie

Para llegar a la fase de identificación de la serie, debemos realizar, tal cual las fases planteadas en el resumen de este acápite, dos pasos previos: la representación gráfica y la eliminación de la tendencia.

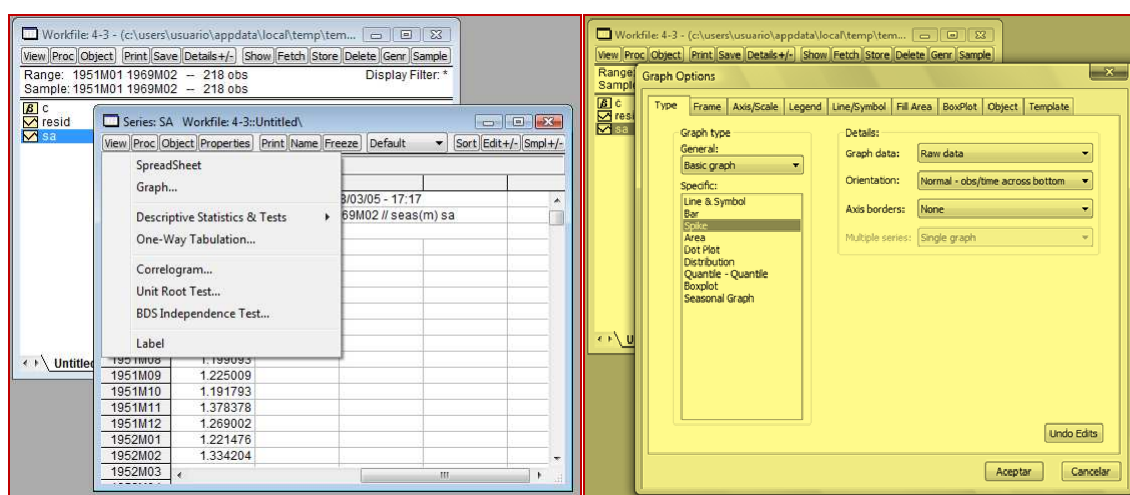
La representación gráfica de la serie es de utilidad para decidir sobre la estacionariedad y estacionalidad. Suelen utilizarse medias y desviaciones típicas por subperíodo para juzgar sobre la estacionariedad de la serie.

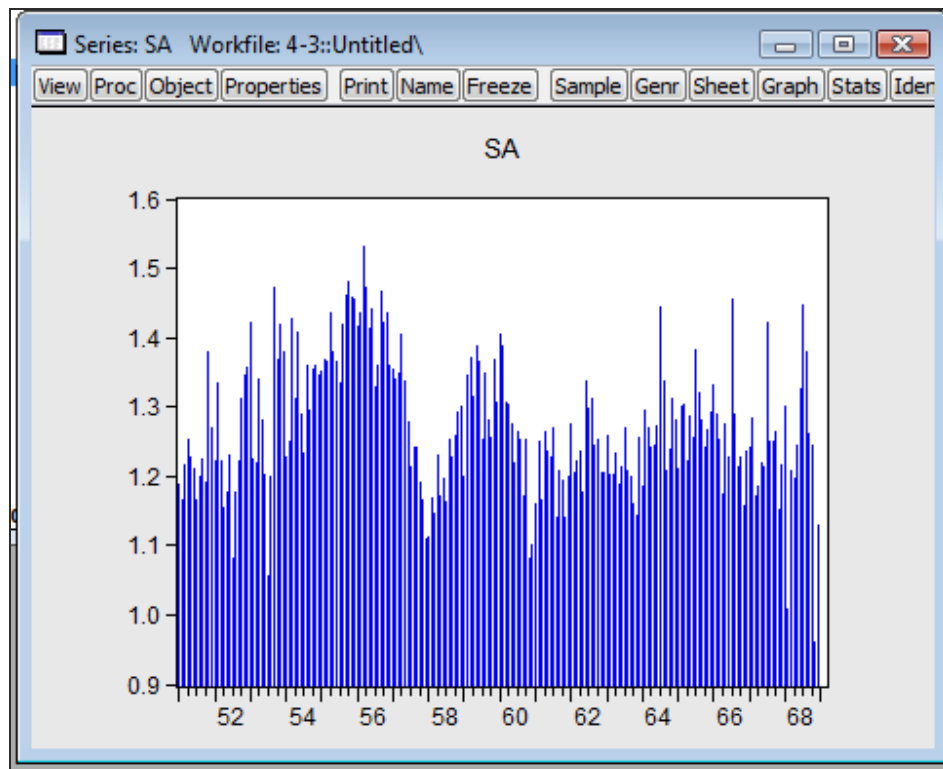
Siguiendo con nuestro ejemplo, realizamos una representación gráfica de la serie SA mediante **QUICK → GRAPH → LINE GRAPH**, indicando la serie a graficar en **SERIES LIST** para obtener la representación de la misma. Se observa a simple vista que el gráfico presenta variaciones estacionales mensuales. Sin embargo este hecho hay que comprobarlo formalmente.



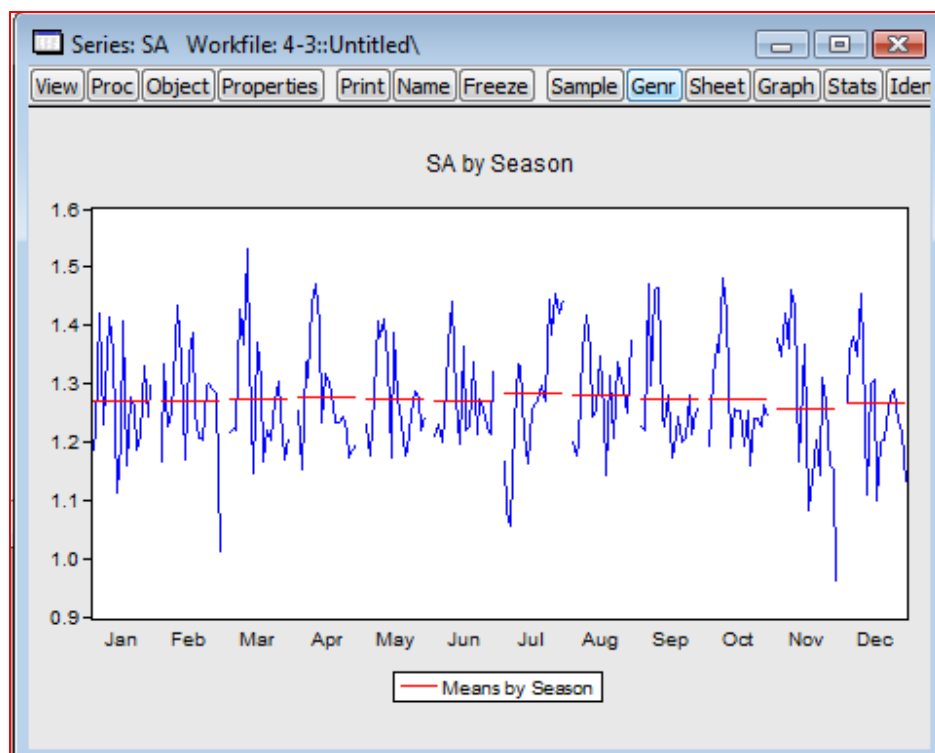


Para probar la estacionalidad podemos utilizar el gráfico vertical de la serie que se obtiene haciendo doble CLICK sobre la serie SA para ver sus valores y eligiendo **VIEW → GRAPH → SPIKE**

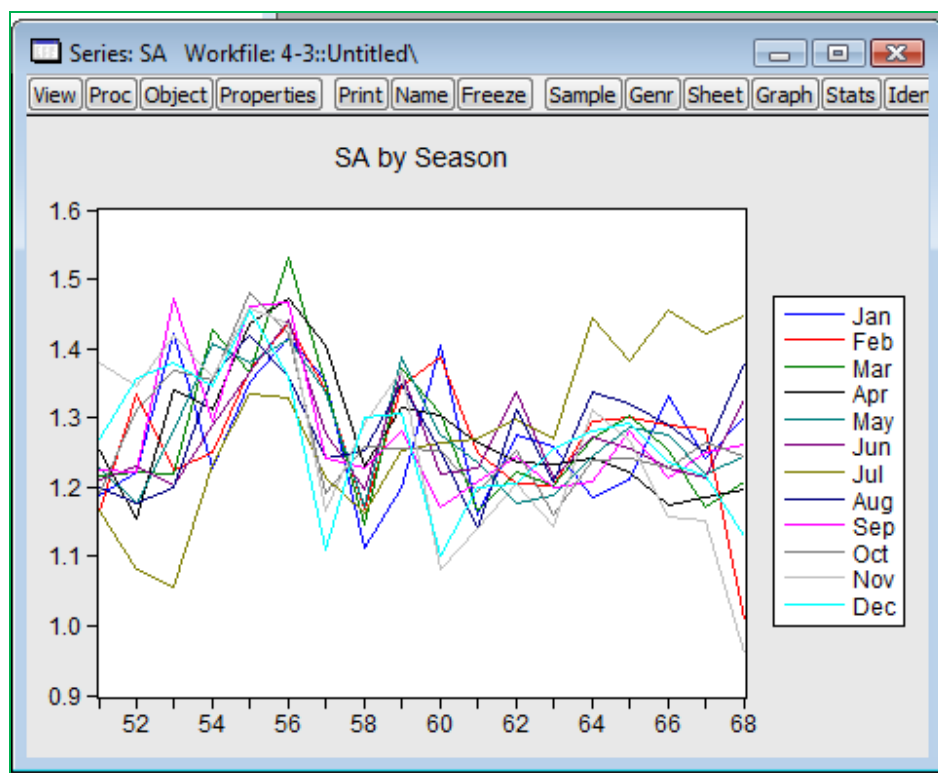




El gráfico de las subseries estacionales que se obtiene eligiendo **VIEW** → **GRAPH** → *Seasonal Stacked Line*, es el siguiente



Por último, el gráfico de las sub series anuales que se obtiene eligiendo **VIEW → GRAPH → Seasonal Split Line**, es



Si es usuario de **Eviews 6**, tenga en cuenta que para obtener estos dos últimos gráficos se hizo CLICK en GRAPH y se seleccionó la opción **SEASONAL GRAPH** y dentro de DETAILS se eligió **PANELED LINES & MEANS** para el primer caso y **MULTIPLE OVERLAYED LINES** para el segundo.

Todos los gráficos anteriores muestran la presencia de estacionalidad mensual. El gráfico de las subseries anuales presenta evoluciones paralelas de los datos en los distintos meses de todos los años. El gráfico de las subseries estacionales muestra claramente las secciones similares de las estaciones.

Pero la estacionalidad y la estacionariedad también pueden detectarse a través de las funciones de autocorrelación y autocorrelación parcial.

¿Qué es la función de autocorrelación (**fac**)?

La función de autocorrelación (**fac**) y la función de autocorrelación parcial (**facp**) miden la relación estadística entre las observaciones de una serie temporal. Por

ejemplo, el coeficiente de autocorrelación entre la variable y_t y la misma variable un período antes, y_{t-1} al que denominaremos coeficiente de autocorrelación de primer orden, se formula como:

$$\rho_1 = \frac{Cov(y_t, y_{t-1})}{\sqrt{Var(y_t) \cdot Var(y_{t-1})}}$$

Dado el supuesto de estacionariedad, se tiene que $var(y_t) = var(y_{t-1})$, por lo que

$$\rho_1 = \frac{Cov(y_t, y_{t-1})}{Var(y_t)}$$

En general, para un desfase de k períodos se tiene que:

$$\rho_k = \frac{Cov(y_t, y_{t-k})}{Var(y_t)}$$

y cuando $k = 0$,

$$\rho_0 = \frac{Cov(y_t, y_t)}{Var(y_t)} = \frac{Var(y_t)}{Var(y_t)} = 1$$

A efectos de la identificación del modelo, debemos comparar el valor que esta función presentaría para los distintos modelos teóricos, con una estimación de la misma para nuestra serie. **El estimador muestral de la (fac), para el que utilizaremos la expresión r_k , viene dado, con ciertas condiciones y aproximaciones, por:**

$$r_k = \frac{\sum_{t=1}^{N-k} (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^N (y_t - \bar{y})^2}$$

La representación gráfica de r_k se denomina **correlograma muestral** y constituye un instrumento de análisis de gran interés práctico. La función de autocorrelación muestral, en la práctica, no debe calcularse cuando $k > T/4$. Además en teoría no se puede calcular cuando $k > T + 1$.

¿Qué es la función de autocorrelación parcial (facp)?

La función de autocorrelación parcial mide la “aportación” que a las variaciones de una variable como y_t tiene otra variable, digamos y_{t-2} , aislados los efectos de las

posibles restantes variables, por ejemplo y_{t-1} . Por el contrario, la función de autocorrelación ignora el hecho de que parte de la correlación que pueda existir entre, por ejemplo y_t y y_{t-2} , se debe a que ambas están correlacionadas con y_{t-1} . Pues bien, los distintos coeficientes de autocorrelación parcial de los modelos teóricos se denotan como ϕ_{kk} , y los estimados para una muestra como $\hat{\phi}_{kk}$.

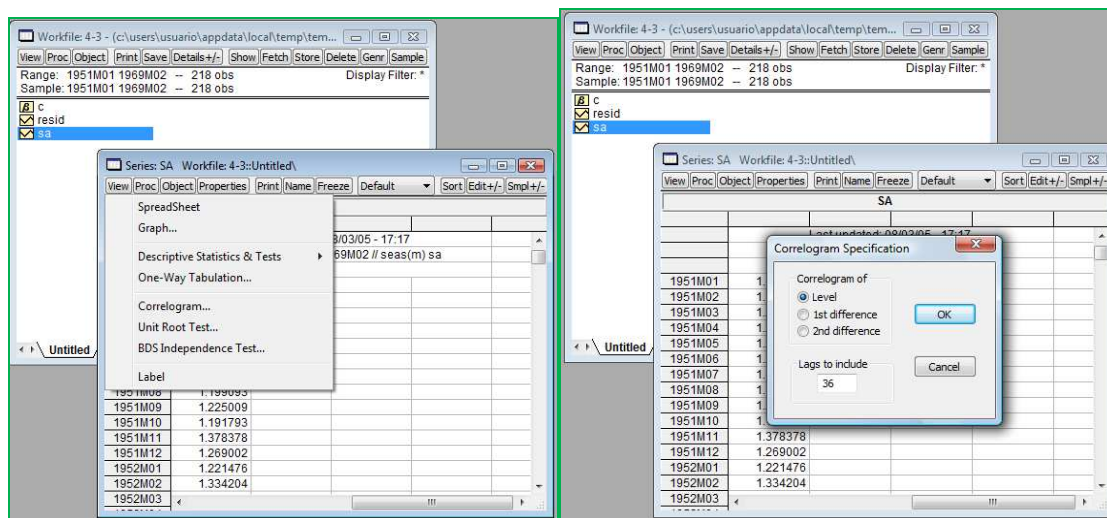
La utilidad de los mismos se deriva de que en determinadas ocasiones el simple conocimiento de la *(fac)* muestral no sería suficiente para la determinación del verdadero proceso generador de la serie.

El primer término de la función de autocorrelación parcial, que vamos a simbolizar por ϕ_{11} puede estimarse transformando la serie $\{y_t\}$ en desviaciones a su media muestral $Y_t = y_t - \bar{y}$ y a continuación estimando una regresión de Y_t sobre Y_{t-1} . La pendiente estimada de esta regresión es $\hat{\phi}_{11}$. El modelo de regresión es $Y_t = \phi_{11}Y_{t-1} + \mu_t$. Además, el primer valor de la función de autocorrelación parcial ϕ_{11} es precisamente igual al primer valor de la función de autocorrelación.

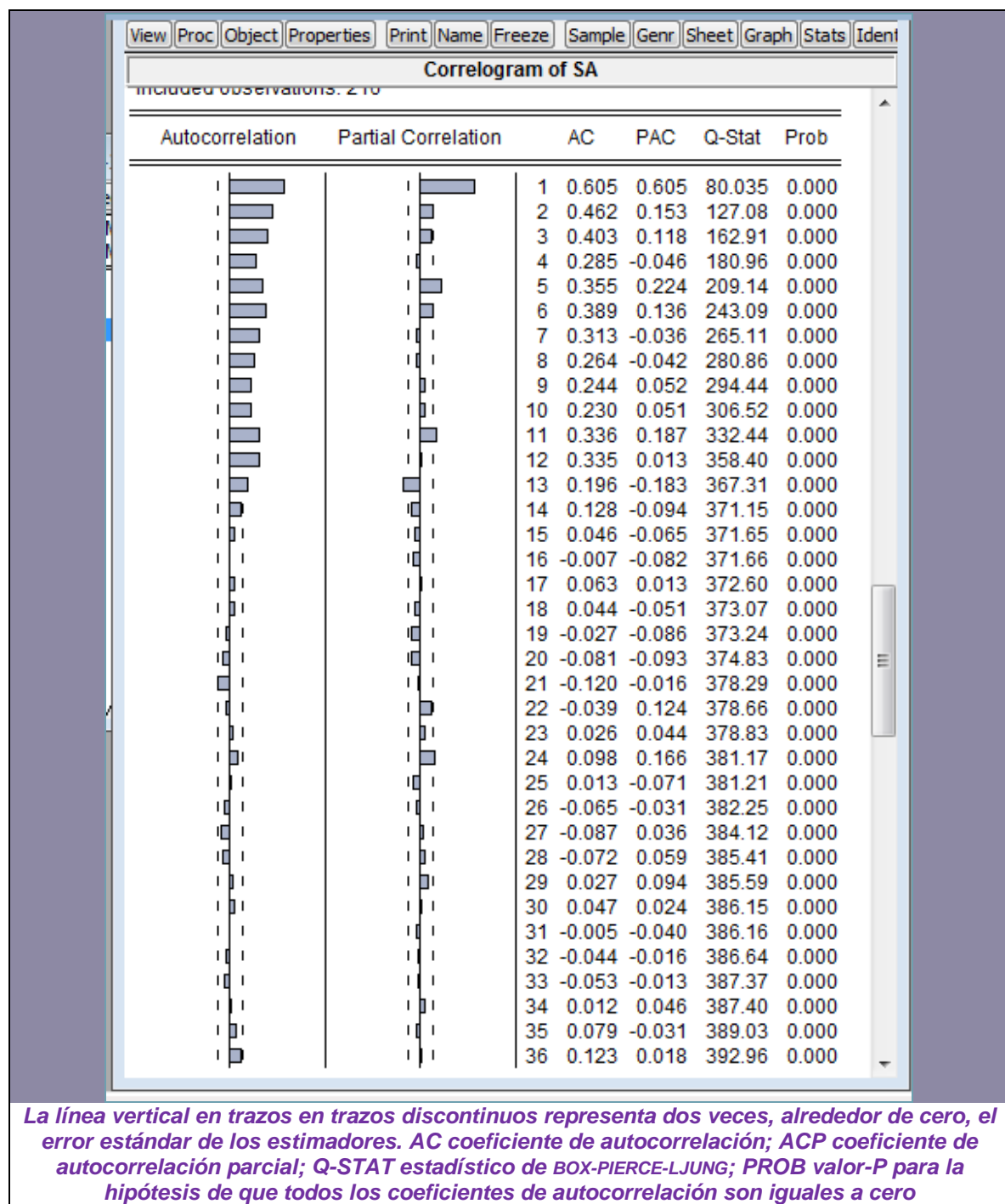
De la misma manera podemos encontrar los sucesivos valores de los siguientes términos, esto es haciendo la regresión en desviaciones con respecto a la media de Y_t sobre Y_{t-1} e Y_{t-2} . El modelo de regresión es $Y_t = \phi_{21}Y_{t-1} + \phi_{22}Y_{t-2} + \mu_t$. El tercero se obtendrá de hacer la regresión: $Y_t = \phi_{31}Y_{t-1} + \phi_{32}Y_{t-2} + \phi_{33}Y_{t-3} + \mu_t$ y así siguiendo...

Se denomina **correlograma** a la representación conjunta de las funciones de autocorrelación y autocorrelación parcial. Como veremos para el rezago de orden 1 estas dos funciones son coincidentes.

Para nuestro ejemplo, con la finalidad de estimar estas funciones y estudiar la estacionalidad y estacionariedad, elegimos en el **WORKFILE View** → **Correlogram** y elegimos la serie en niveles con 36 retardos:



Se obtienen las funciones de autocorrelación (fac) y de autocorrelación parcial (fap) estimadas. La (fac) muestra valores altos en los retardos múltiplos del período estacional 12, 24 y 36.



Se observa que las funciones de autocorrelación y autocorrelación parcial estimadas también validan los períodos estacionales porque los coeficientes de la (*fac*) para retardos múltiplos del período estacional de la serie son significativamente distintos de cero.

La significancia estadística de cualquier r_k puede ser evaluada por su error estándar. Como vemos en la nota al pie del correlograma la línea en trazos discontinuos representa aproximadamente el intervalo de confianza del 95% (aproximadamente 2 veces el desvío estándar) para contrastar la hipótesis de que $\rho_k = 0$.

Formalmente, BARLETT ha demostrado en "ON THE TEORETICAL SPECIFICATION OF SAMPLING PROPERTIES OF AUTOCORRELATED TIME SERIES" (JOURNAL OF THE ROYAL SATATISTICAL SOCIETY 1946) que si una serie de tiempo es puramente aleatoria, es decir presenta ruido blanco (el término de error a_t tiene media es nula, varianza constante y no está autocorrelacionado), los coeficientes de autocorrelación muestral están distribuidos en forma *aproximadamente* normal con media cero y varianza $1/n$, donde n es el tamaño de la muestra.

Para la serie SA $n = 218$, lo que implica un error estándar de $1/\sqrt{218} = 0,0677$. Entonces siguiendo las propiedades de la distribución normal estándar, el intervalo de confianza al 95% para cualquier ρ_k será $\pm 1,96 \cdot (1/\sqrt{n}) = \pm 0,1327$ a cualquier lado del cero. Así si un r_k se encuentra dentro del intervalo $(-0.1327; 0.1327)$ se acepta la hipótesis de que cualquier ρ_k es cero. Pero, si se encuentra por fuera de dicho intervalo se rechaza tal hipótesis. El lector puede ver que todos los coeficientes estimados hasta el rezago 14 son estadísticamente significativos de manera individual, es decir significativamente distintos de cero.

Para probar la **hipótesis conjunta** de que todos los coeficientes de autocorrelación son ρ_k son simultáneamente iguales a cero, se puede utilizar el estadístico Q desarrollado por BOX-PIERCE-LJUNG EN "ON A MEASURE OF LACK OF FIT IN TIME SERIES MODELS" (BIOMETRIKA, 1978) que está definido como:

$$Q = n(n+2) \sum_{k=1}^m \left(\frac{r_k^2}{n-k} \right) \sim \chi_m^2$$

Donde, m es la longitud del rezago.

Como puede verse en el correlograma este estadístico es igual a 392.96 para 36 rezagos, siendo altamente significativos; los valores p de obtener tales valores ji-cuadrado son cero. Es de destacar que los valores p son cero para cualquier rezago considerado. Por ejemplo para el tercer rezago tendríamos:

$$Q = 218(218 + 2) \cdot \left(\frac{0.605^2}{217} + \frac{0.462^2}{216} + \frac{0.403^2}{215} \right) \cong 162.91 > \chi^2_{m=3; \alpha=0.05} = 7.81473$$

Por lo que con base al correlograma y los test estadísticos hay evidencia empírica suficiente como para abonar la conclusión general de que nuestra serie es no estacionaria.

Además, para una cantidad grande de retardos la (fac) se configura en forma de abanico que completa su ciclo girando sobre el eje de las abscisas para una cantidad de retardos igual al período estacional. Por otro lado, la $(facp)$ presenta estructura de coeficientes significativos para retardos periódicos (largos). La (fac) y la $(facp)$ deben considerarse a la vez, pues a veces intercambian sus papeles en el comportamiento estacional (como aquí sucede con el rezago 24). Asimismo, los coeficientes de la (fac) no decaen rápidamente, lo que indica falta de estacionariedad en media.

Por otra parte, si el lector calcula la serie de medias y varianzas por meses a lo largo de toda la serie observará variaciones significativas crecientes y decrecientes a lo largo de los años, lo que indica que no hay estacionariedad ni en media ni en varianza en la serie original. Invitamos a que lo compruebe.

3.Transformación de la Serie

Cuando una serie no es estacionaria hay que proceder a su transformación.

Para detectar la no estacionariedad deberemos recordar lo siguiente.

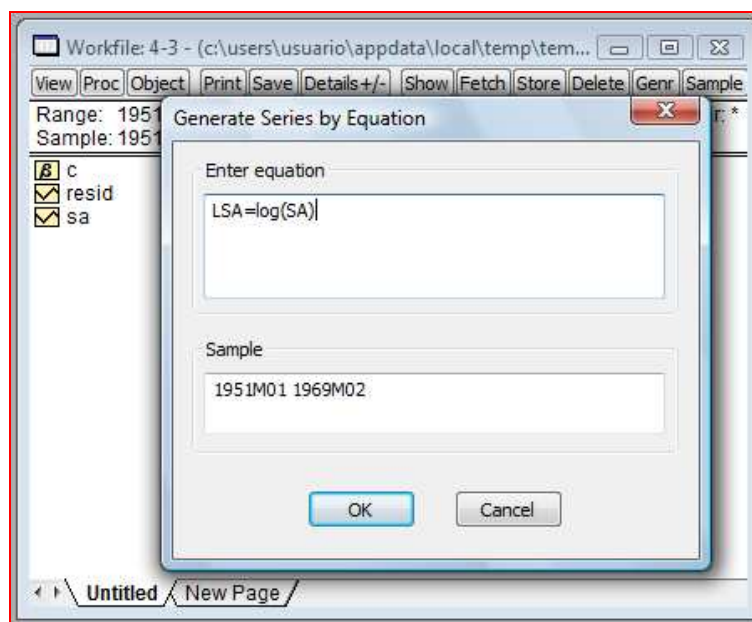
- a) Para detectar rápidamente la estacionariedad se puede utilizar el **gráfico de la serie**, como se ha visto en la aplicación práctica. Para ello se divide el campo de variación total de la serie en varios intervalos calculándose para cada uno de ellos la media y la varianza. Si existe estacionalidad se toma como longitud del intervalo la del período estacional. Para ver si la serie es estacionaria en media y varianza hay que comprobar que las mismas no fluctúen mucho, que sean más bien estables.
- b) Otro criterio para detectar la estacionariedad en varianza es el **gráfico de Box-Cox**, consistente en representar los puntos (media, rango) para todos los intervalos en que se ha dividido la serie. Si los puntos del gráfico son ajustables a una recta con pendiente positiva no hay estacionariedad en varianza (será necesario tomar logaritmo de la serie original). Si el gráfico

no tiene tendencia definida o es ajustable a una recta paralela al eje de las abscisas hay estacionariedad en varianza.

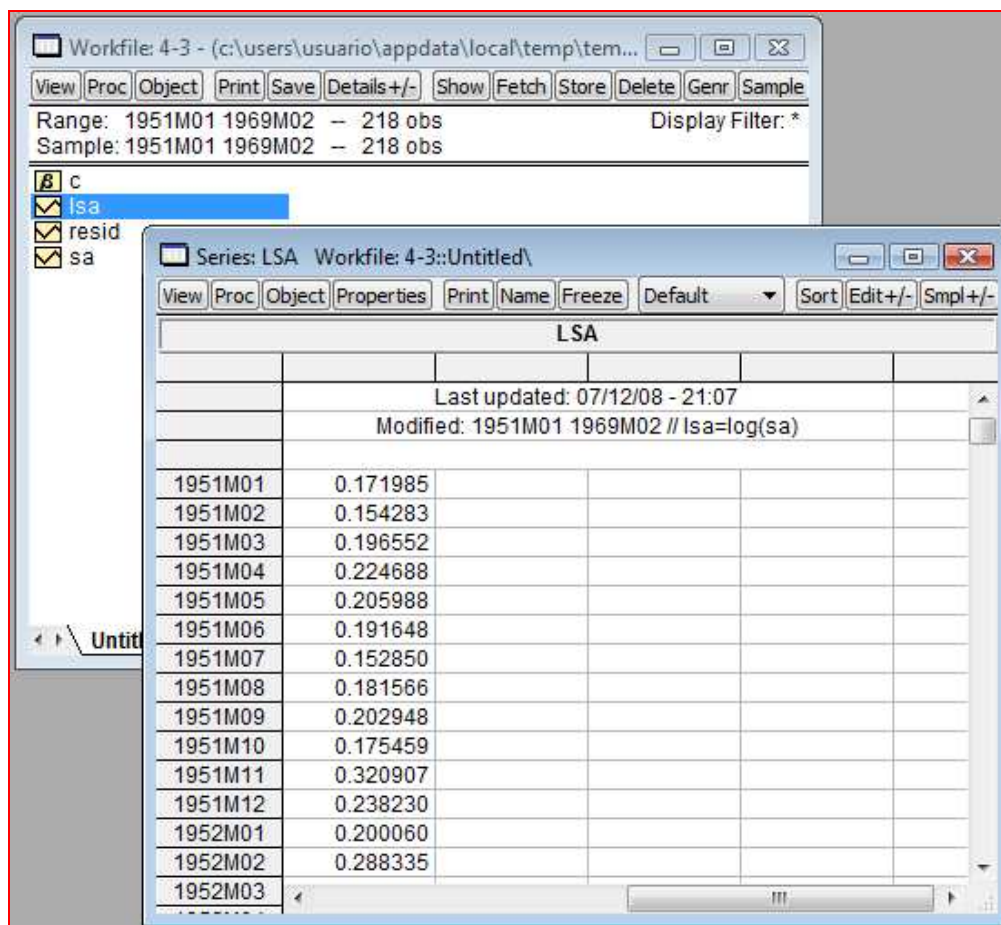
- c) También se puede detectar la estacionariedad en varianza mediante los **contrastes de varianzas** aplicados a los diferentes intervalos en que se divide la serie
- d) Otro criterio para detectar la estacionariedad en media es el criterio de la función de autocorrelación estimada. Si los coeficientes de las **(*fac*)** no decaen rápidamente hay un indicio claro de falta de estacionariedad en media, lo que nos llevaría a tomar primeras diferencias en la serie original o también mediante los **contrastes de raíces unitarias**. Ambos criterios se verán más adelante en este mismo acápite.

Muy pocas series temporales reales del mundo económico son estacionarias. La mayoría suelen presentar tendencia, suelen tener varianza no constante y también suelen presentar variaciones estacionales. La presencia de estas variaciones se traduce en una variabilidad de la media del proceso estocástico, lo que es contrario a la hipótesis de estacionariedad. Pero normalmente es posible transformar muchas series económicas reales no estacionarias en otras aproximadamente estacionarias, sometiéndolas a transformaciones algebraicas adecuadas.

En el ejemplo, la transformación nos lleva a tomar logaritmos sobre la serie original. Esto en esencia significa generar una nueva variable, la que denominaremos $LSA = \log(SA)$. En *Eviews* esto se logra por medio del comando GENR,

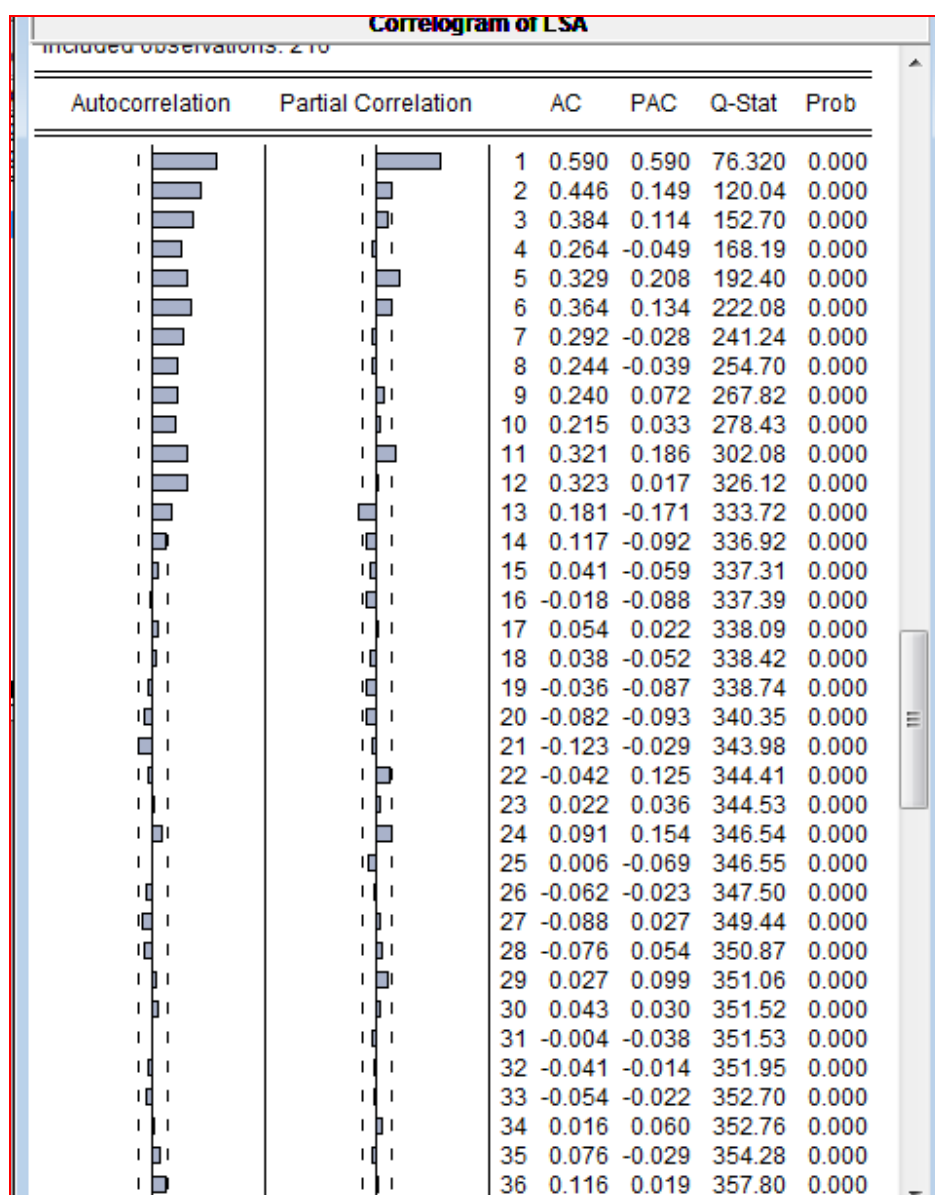


Nuestro archivo de trabajo y la nueva serie en logaritmos son los que se muestran en la siguiente pantalla de *Eviews*



Calculando el correlograma de la serie transformada en logaritmo vemos que el problema aún no se ha solucionado ya que las funciones de autocorrelación y autocorrelación parcial siguen mostrando un comportamiento similar al detectado, con esas mismas funciones, en la serie original.

Se puede decir que con la transformación logarítmica se ha solucionado la estacionariedad en varianza, pero la serie sigue mostrando una tendencia lineal.



¿Qué significa diferenciar una serie?

El proceso para atenuar la estacionariedad se completa tomando primeras o más diferencias de la serie original o transformada en logaritmo. Para transformar una serie en primeras diferencias se hace el siguiente cálculo: $Z_t = (y_t - y_{t-1})$. De esta forma si y_t muestra una tendencia lineal, la primera diferencia de la serie Z_t ya no tendrá esa tendencia. En este caso se dice que y_t es una serie temporal homogénea de primer orden o integrada de primer orden y se denota por $I(1)$.

La eliminación de una tendencia cuadrática, de la serie original o de su transformación logarítmica, puede conseguirse mediante doble diferenciación. Esta operación se realiza en dos etapas, primero se obtiene $W_t = (y_t - y_{t-1})$ y, si sigue existiendo tendencia, se obtiene $Z_t = (W_t - W_{t-1})$. Si Z_t ya no incorpora tendencia (es estacionaria), se dice que y_t es una serie temporal homogénea de segundo orden o integrada de segundo orden $I(2)$. Análogamente una tendencia de orden p puede eliminarse llevando a cabo una diferenciación de orden p dando lugar a una serie homogénea o integrada de orden p , esto es $I(p)$.

En la práctica una serie económica se transforma en estacionaria en media con una o a lo sumo dos diferenciaciones. Y como dijimos anteriormente, se transforma en estacionaria en varianza, generalmente, a partir de la transformación logarítmica.

Habíamos dicho también que una serie puede presentar variaciones estacionales. Estas series presentan oscilaciones que se producen con un período igual o inferior a un año, y que se reproducen de manera reconocible en los diferentes años. El motivo principal que induce a estudiar la componente estacional es que en la inmensa mayoría de las series económicas dicha componente provoca una distorsión de su verdadero movimiento. Para eliminar estas distorsiones y captar el movimiento real de la serie, es necesario eliminar las oscilaciones estacionales desestacionalizando la serie.

Para detectar la estacionalidad, en la práctica pueden utilizarse los siguientes caminos:

- a) El gráfico de la serie da una idea de los posibles períodos estacionales.
- b) El gráfico de las subseries estacionales identifica gráficamente los períodos estacionales presentando secciones sucesivas de los mismos.
- c) El gráfico de las subseries anuales valida gráficamente los períodos estacionales presentando comportamientos paralelos para cada estación.

d) Las (fac) y las $(facp)$ estimadas también validan los períodos estacionales de acuerdo a las siguientes consideraciones:

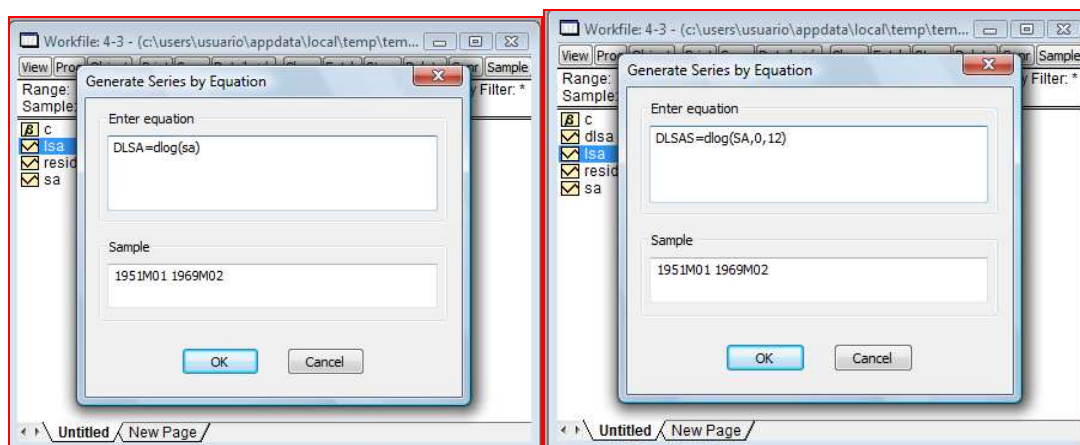
- Los coeficientes de la (fac) para retardos múltiplos del período estacional de la serie deben ser significativamente distintos de cero.
- Para una cantidad grande de retardos la (fac) se configura en forma de abanico que completa su ciclo girando sobre el eje de las abscisas para una cantidad de retardos igual al período estacional. La $(facp)$ debe presentar estructura de coeficientes significativos para retardos periódicos (largos).
- Las (fac) y las $(facp)$ deben considerarse a la vez, pues a veces intercambian sus papeles en el comportamiento estacional.

La desestacionalización es una tarea no trivial que ha dado lugar a multitud de estudios y algoritmos, entre los que destacan los programas X11 y X12 del Bureau of the Census de Estados Unidos. A nivel trivial existen varios métodos de desestacionalización, entre ellos el más usado es el **método de las diferencias estacionales**. El mismo permite eliminar la mayor parte del efecto estacional de una serie y consiste en obtener la serie de diferencias de orden m (período estacional), definida como $Z_t = (y_t - y_{t-m})$. De todos modos es conveniente recordar que en cada diferenciación de orden m perdemos m observaciones de la serie original o de su transformada logarítmica.

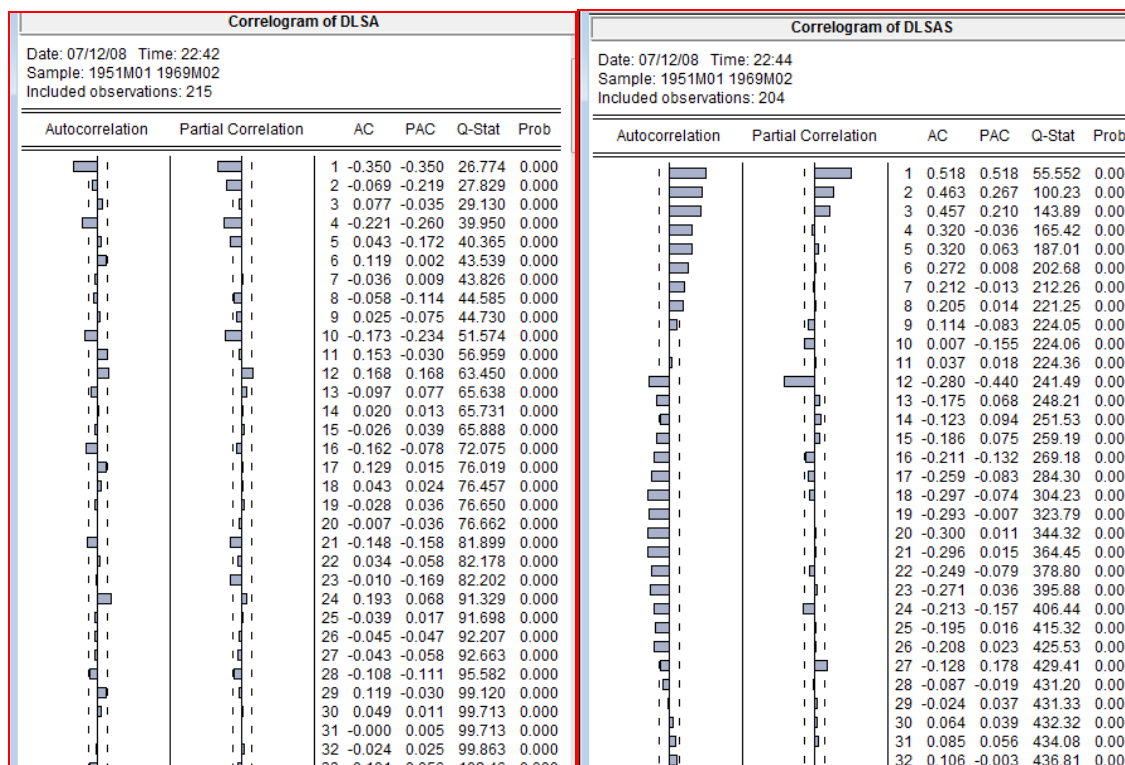
La decisión de diferenciar estacionalmente la serie se basa en la (fac) con el mismo criterio que para la diferenciación estacionaria pero considerando sólo los retardos referidos a períodos estacionales (m y sus múltiplos). Si los coeficientes de la función de autocorrelación no decaen rápidamente en los retardos múltiplos del período estacional m hay que diferenciar estacionalmente la serie.

La eliminación de las variaciones estacionales, para inducir estacionariedad, suele hacerse casi siempre, mediante la diferenciación estacional. Si los datos son mensuales, la diferenciación estacional de la serie temporal y_t consiste en calcular $Z_t = (y_t - y_{t-12})$. Con datos trimestrales calcularíamos $Z_t = (y_t - y_{t-4})$. Si después de efectuar esta transformación la serie sigue presentando evidencias de variaciones estacionales, es posible aplicar de nuevo el procedimiento, es decir, calcular las diferencias de segundo orden y así siguiendo...

En el ejemplo, una vez aplicado logaritmos, como la serie es estacional, el problema es identificar si diferenciamos la parte regular de la serie en logaritmos o en la parte estacional. Para ello generamos con el botón GENR los dos casos, donde d simboliza la diferenciación.



A continuación representamos las funciones de autocorrelación estimada y de autocorrelación parcial estimada bajo los supuestos de diferenciación en la parte regular de la serie en logaritmos o en la parte estacional. Esto lo hacemos solicitando los respectivos correlogramas,



Se observa que al diferenciar sólo la parte regular de la serie en logaritmos, las funciones de autocorrelación y autocorrelación parcial estimadas no superan el problema de la falta de estacionariedad ya que la *(fac)* no decae rápidamente. Pero al diferenciar sólo una vez la parte estacional de la serie en logaritmos, las funciones de autocorrelación y de autocorrelación parcial estimadas ya superan el problema de la no estacionariedad. Asimismo, estas dos funciones cumplen las condiciones para que haya estacionalidad porque los coeficientes de la *(fac)* para retardos múltiplos de período estacional de la serie son significativamente distintos de cero. Además, para una cantidad grande de retardos, la *(fac)* se configura en forma de abanico que completa su ciclo girando sobre el eje de abscisas para una cantidad de retardos igual al período estacional.

Luego el problema de la estacionalidad y estacionariedad en media y varianza se ha arreglado aplicando logaritmos, diferenciando una vez la parte estacional y no diferenciando la parte regular. Luego la parte regular de la serie en logaritmos es integrada de orden cero $I(0)$ y la parte estacional es integrada de orden uno $I(1)$.

4. Identificación del Modelo

Ahora debemos elegir una forma funcional concreta para la serie de datos, que como vimos depende de sus valores anteriores. Por ejemplo, una forma lineal como $y_t = \phi_0 + \phi_1 y_{t-1} + a_t$, donde ϕ_0 es un término independiente y ϕ_1 es un parámetro que multiplica al valor de la variable y en el período $t - 1$. Utilizando métodos estadísticos adecuados podemos estimar los parámetros ϕ_0 y ϕ_1 de forma que estos cumplan propiedades estadísticas razonables y sean una buena (la mejor posible) estimación. Con ello obtendríamos una expresión como $\hat{y}_t = \hat{\phi}_0 + \hat{\phi}_1 y_{t-1}$ que utilizaríamos a efectos de predicción.

Aunque no entraremos en el detalle, en el modelo teórico, el coeficiente de autocorrelación parcial coincide con el último coeficiente autorregresivo de un modelo **AR**. En otros términos, el coeficiente de correlación parcial de orden 1 será el valor de ϕ_1 en un **AR(1)**; el de orden 2 coincidirá con ϕ_2 en un **AR(2)**, y así sucesivamente. En la práctica, los coeficientes de autocorrelación parcial calculados no son buenos estimadores de los parámetros correspondientes, aunque pueden servir como valores iniciales para el proceso iterativo de cómputo que ha de seguirse.

¿Qué son los modelos AR?

Lo que acabamos de ver es la esencia de los *modelos autorregresivos* (o modelos AR). Se realiza una regresión de la variable y_t sobre sí misma (auto-regresión) o, mejor dicho, sobre los valores que la variable tomó en el período anterior.

Un aspecto importante es el orden del modelo AR. Por ejemplo, el modelo

$$y_t = \phi_0 + \phi_1 y_{t-1} + a_t$$

es de orden 1, y se denota como **AR(1)**. Si tomamos en el modelo como explicativas los valores de la variable y_t en los 2 períodos anteriores, es decir

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + a_t$$

entonces hemos especificado un **AR(2)**. De igual forma un **AR(3)** vendría dado por

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + a_t$$

En general, un **AR(p)** viene dado por

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + a_t$$

Que puede ponerse mediante el operador de cambio retroactivo **B** como:

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) y_t = a_t, \text{ donde } (B)^k y_t = y_{t-k}$$

Es frecuente encontrarnos con Modelos AR con un bajo orden (1 o 2). En series con componente estacional es habitual que el desfase sea coincidente con la periodicidad de los datos. En ese caso hablamos de *modelos SAR*.

¿Qué son los modelos SAR?

Los modelos AR relacionan el valor de la variable y_t con sus retardos inmediatamente anteriores. Por ejemplo un **AR(2)**, relacionaría el valor de la variable y (digamos la tasa anual de inflación) en el año 2007, con el valor de la variable (inflación) en el año 2006 y 2005. Ahora bien, cuando modelizamos una serie con estacionalidad (por ejemplo, la tasa de variación mensual de inflación, con 12 datos al año), la comparación adecuada no solo debe ser, por ejemplo, de la inflación de junio de 2007 con mayo y abril de 2007, sino con el mismo mes (junio) de los años anteriores, en nuestro ejemplo 2006 y 2005. Ello da lugar a los modelos SAR.

La formulación de un modelo **SAR(1)**, viene dado por:

$$y_t = \phi_0 + \phi_1 y_{t-s} + a_t$$

donde $s=4$, si la serie a modelizar es de frecuencia trimestral, o $s=12$, si la serie es mensual. Un modelo **SAR(2)** se especificaría como:

$$y_t = \phi_0 + \phi_1 y_{t-s} + \phi_2 y_{t-2s} + a_t$$

¿Qué son los modelos MA?

Un modelo ARIMA está conformado por un proceso *autorregresivo* (AR) *integrado* (I) de *medias móviles* (MA). Una alternativa de modelización pasa por tratar de explicar el comportamiento de una variable y , no en función de los valores que tomó en el pasado (modelos AR) sino a través de los *errores* al estimar el valor de la variable en los períodos anteriores. Ello da lugar a los *modelos de medias móviles* (o Modelos MA, por sus siglas en inglés).

Por ejemplo, un modelo **MA(1)** viene dado por la expresión

$$y_t = \mu + a_t + \theta_1 a_{t-1}$$

donde μ es el valor constante alrededor del cual se mueve la variable, y ha de ser estimado igualmente con los coeficientes θ . En general un modelo **MA(q)** viene dado por la expresión:

$$y_t = \mu + a_t + \theta_1 a_{t-1} + \dots + \theta_q a_{t-q}$$

Que puede expresarse mediante el *operador de cambio retroactivo* como:

$$y_t = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) a_t$$

Al igual que ocurría con los modelos AR, en series con componente estacional es frecuente que el retardo coincida con la periodicidad de los datos, dando lugar a los modelos SMA.

¿Qué son los modelos SMA?

Al igual que ocurre con los modelos AR, en series con componente estacional (periodicidad inferior a la anual) es frecuente que en los modelos MA los retardos se establezcan no con los períodos inmediatamente anteriores, sino que sean coincidentes con la periodicidad de los datos.

Así, un modelo **SMA(1)** vendría dado por:

$$y_t = \mu + a_t + \theta_1 a_{t-s}$$

donde $s=4$, si la serie a modelizar es de frecuencia trimestral, o $s=12$, si la serie es mensual. Un modelo **SMA(2)** se especificaría como:

$$y_t = \mu + a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2s}$$

Relación entre modelos AR y MA

Entre los modelos AR y los Modelos MA existe una relación, bajo ciertas condiciones, que es útil conocer. Los modelos ARMA integran a los Modelos AR y a los Modelos MA en una única expresión. Por tanto, la variable y queda explicada en función de los valores tomados por la variable en períodos anteriores, y los errores cometidos en la estimación. Una expresión general de un modelo **ARMA(p,q)** viene dado por

$$y_t = \mu + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + a_t + \theta_1 a_{t-1} + \dots + \theta_q a_{t-q}$$

que, es la unión de un modelo **AR(p)** y un modelo **MA(q)**.

Obviamente, los modelos **AR(p)** se corresponden con modelo **ARMA(p,0)**, mientras que los modelos **MA(q)** se corresponden con **ARMA(0,q)**.

Bajo ciertas condiciones los modelos **AR** y los **MA** pueden relacionarse. Estas condiciones se denominan de **invertibilidad** y de **estacionariedad**. La explicación de tales características requiere de un instrumental matemático específico, que puede ser encontrado en la lectura adicional 1.

Intuitivamente, por ejemplo, supongamos un modelo **AR(1)** sin término independiente como:

$$y_t = \phi_1 y_{t-1} + a_t$$

Puesto que

$$y_{t-1} = \phi_1 y_{t-2} + a_{t-1}$$

podríamos llegar, por sustituciones sucesivas a una expresión como:

$$y_t = a_t + \phi_1 a_{t-1} + \phi_1^2 a_{t-2} + \phi_1^3 a_{t-3} + \dots$$

Es decir, un proceso autorregresivo de primer orden es equivalente a una media móvil de infinitos términos con una ponderación decreciente en forma exponencial, cuando

$$0 < |\phi_1| < 1$$

Este resultado es generalizable, y puede demostrarse que bajo las condiciones de **estacionariedad** un modelo **AR** de orden reducido puede transformarse en un modelo **MA** de orden elevado o, teóricamente, infinito.

De igual forma, bajo las condiciones de **invertibilidad**, modelos **MA** de orden reducido pueden aproximarse por modelos AR de un número suficientemente elevado de términos.

La utilidad de esta relación estriba en que, por aplicación de un **principio de parsimonia** (y sin olvidar que, en la práctica, puede favorecer la estimación del modelo) es preferible un modelo sencillo, con el menor número posible de términos y, por lo tanto, de parámetros a estimar, frente a un modelo con un gran número de coeficientes, siempre que, por supuesto, nos conduzca a resultados similares.

En general, un modelo **AR(p)** es siempre *invertible* y será *estacionario* si las raíces del polinomio en **B** dado por: $(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)$ caen fuera del círculo unidad. **Esta condición es equivalente a que las raíces de la siguiente ecuación sean todas inferiores a 1 en módulo:**

$$y^p - \phi_1 y^{p-1} - \phi_2 y^{p-2} - \dots - \phi_{p-1} y - \phi_p = 0$$

Por otro lado, un proceso de medias móviles **MA(q)** es siempre *estacionario* y será *invertible* si las raíces del polinomio en **B** dado por: $(1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q)$ caen fuera del círculo unidad. Esta condición es equivalente a decir que las raíces de la siguiente ecuación sean todas inferiores a uno en módulo:

$$y^q - \phi_1 y^{q-1} - \phi_2 y^{q-2} - \dots - \phi_{q-1} y - \phi_q = 0$$

Modelos ARIMA

Un modelo **ARIMA(0,d,0)** es una serie temporal que se convierte en un ruido blanco (proceso puramente aleatorio) después de ser diferenciada **d** veces. El modelo **ARIMA(0,d,0)** se expresa mediante:

$$(1 - B)dy_t = a_t$$

El modelo general **ARIMA(p,d,q)** denominado proceso autorregresivo integrado de medias móviles de orden **p,d,q**, toma la siguiente expresión:

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)(1 - B)^d y_t = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) a_t$$

Este modelo permite describir una serie de observaciones después de que hayan sido diferenciadas **d** veces, a fin de extraer las posibles fuentes de no estacionariedad. Esta fórmula general se puede aplicar a cualquier modelo. Si hay alguna componente **p,d,q** igual a cero, se elimina el término correspondiente de la fórmula general.

Los modelos cíclicos o estacionales son aquellos que se caracterizan por oscilaciones cíclicas, también denominadas variaciones estacionales. Las variaciones cíclicas a veces se superponen a una tendencia secular. Las series con tendencia secular y variaciones cíclicas pueden representarse mediante los modelos **ARIMA(p,d,q)(P,D,Q)**. El primer paréntesis se refiere a la tendencia secular o parte regular de la serie y el segundo paréntesis se refiere a las variaciones estacionales, o parte cíclica de la serie temporal.

La expresión de un modelo **ARIMA(0,1,1)(0,0,1)₁₂** será:

$$(1 - B)y_t = (1 - \theta_1 B^{12})(1 - \delta_{12} B^{12}) a_t$$

O sea, un modelo de medias móviles de orden 1, que ha sido diferenciado una vez y que presenta una estacionalidad mensual en el proceso de medias móviles.

La expresión de un modelo **ARIMA(1,1,1)(2,1,1)₁₂** será:

$$(1 - \phi_1 B)(1 - \omega_1 B^{12} - \omega_2 B^{24})(1 - B^{12})(1 - B)y_t = (1 - \theta_1 B)(1 - \delta_{12} B^{12}) a_t$$

O sea, un modelo autorregresivo y de medias móviles de orden 1, que presenta una estacionalidad mensual de segundo orden en su parte autorregresiva y de orden uno en su parte de medias móviles, luego de haber sido diferenciado una vez.

En resumen, **con la serie tratada para convertirla en estacionaria** es posible estimar un Modelo ARIMA. Pues bien, un Modelo Autorregresivo Integrado de Medias Móviles de orden p , d , q , o abreviadamente **ARIMA(p, d, q)**, no es más que un modelo **ARMA(p, q)** aplicado a una serie integrada de orden d , **$I(d)$** es decir, a la que ha sido necesario diferenciar d veces para eliminar la tendencia.

Por lo tanto, el problema principal parte de identificar el modelo que mejor describe el fenómeno pasa por determinar el más adecuado de los órdenes autorregresivos, de la media móvil y el orden de integrabilidad o diferenciación.

Aparte de la simple representación gráfica a partir del correlograma, para la determinación del orden de integrabilidad, esto es, para determinar el número de veces que será necesario diferenciar la serie para hacerla estacionaria en media, existen dos procedimientos fundamentales de *detección del número de raíces unitarias*, como son el **Test de Dickey – Fuller** (DF), el **Test de Dickey – Fuller Aumentado** (DFA) y el **Test de Phillips – Perron** (PP).

El **test DF simple**, parte de considerar que el proceso estocástico que subyace a la serie objeto de estudio es un **AR(1)**, y contrasta la significatividad del parámetro asociado a la variable Y_{t-1} , que en la versión más general adoptaría la expresión:

$$(1) y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 t + \varepsilon_t$$

o, análogamente:

$$(2) \Delta y_t = \alpha_0 + \gamma y_{t-1} + \alpha_2 t + \varepsilon_t$$

Donde t es una variable de tendencia. Además, Δ es el operador de primeras diferencias, tal que $\Delta y_t = y_t - y_{t-1}$.

De esta forma, si en la expresión (1) se encuentra que $\alpha_1 = 1$ entonces se dice que la variable Y_t tiene una **raíz unitaria**. En econometría una serie de tiempo que tiene raíz unitaria se conoce como una *caminata aleatoria*. Una caminata aleatoria es un ejemplo de una serie de tiempo **no estacionaria**.

La expresión (1) con frecuencia se expresa de la forma (2) donde se supone que $\gamma = (\alpha_1 - 1)$ y donde Δ es el operador de primeras diferencias. Haciendo uso de estas definiciones el lector puede ver que las expresiones son iguales. Sin embargo, ahora la hipótesis nula es que $\gamma = 0$. Por lo tanto, si el parámetro *gamma* es estadísticamente distinto de cero, entonces la serie es estacionaria en

media. Si no podemos rechazar la hipótesis nula de que el parámetro es igual a cero, entonces podemos concluir que la serie presenta, al menos, una raíz unitaria.

El **Test ADF** incorpora una corrección paramétrica útil cuando presuponemos que el proceso estocástico subyacente a la serie no sigue un esquema **AR(1)**. Si esto es así, entonces la estimación de la regresión auxiliar del test DF no arrojaría un residuo ruido blanco.

El **TEST PHILLIPS PERRON** es un método no paramétrico de detección de raíces unitarias que, para corregir la correlación serial, en lugar de añadir más retardos en la regresión, lo que realiza es una corrección en el **estadístico t** asociado al coeficiente *gamma* en una regresión, que supone que el proceso estocástico subyacente sigue un **AR(1)**, como

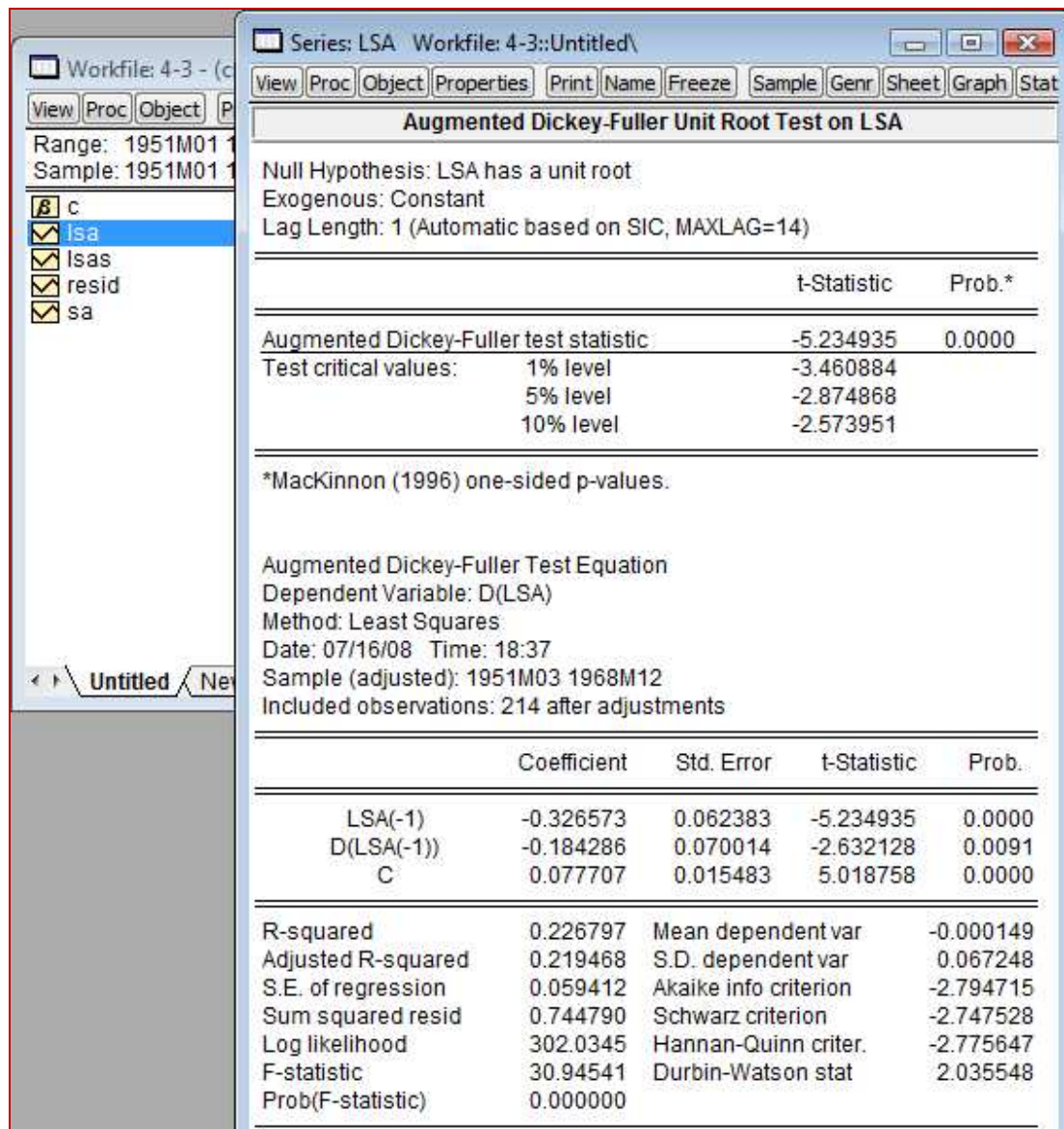
$$\Delta y_t = \alpha_0 + \gamma y_{t-1} + \varepsilon_t$$

Se trata, pues de una alternativa al ADF, en la medida en que realiza distinto tratamiento de la correlación.

Ambos procedimientos presentan alternativas propias en la estimación de las regresiones auxiliares, inclusión o no de término independiente, de una variable de tendencia, número de retardos óptimos... Por ello, la utilización correcta de estos procedimientos pasa por una metodología ordenada y rigurosa.

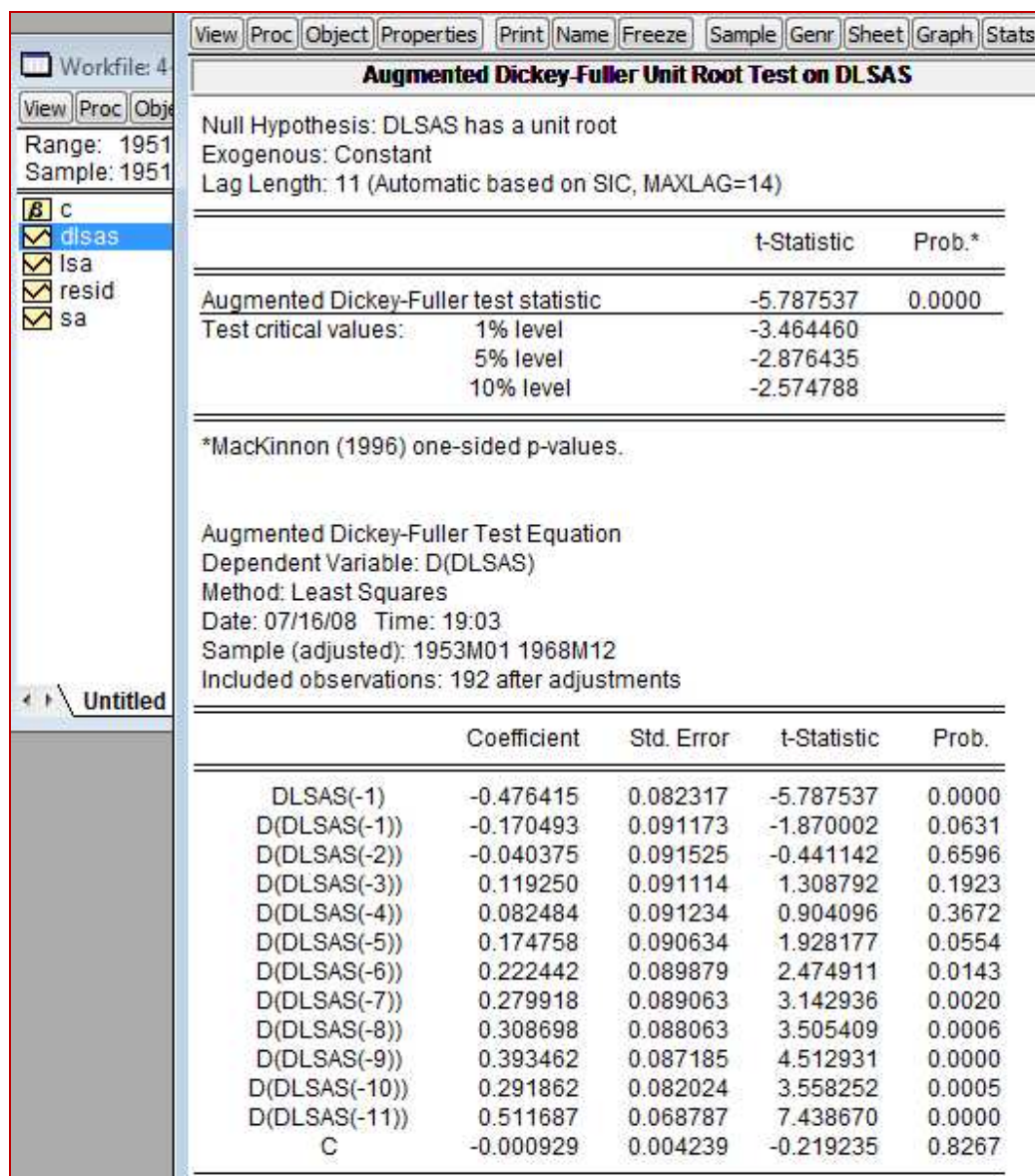
En la serie que estamos tratando habíamos identificado previamente, a partir de análisis de las funciones de autocorrelación y de autocorrelación parcial y de sus respectivos correlogramas, luego de la transformación logarítmica que el proceso estocástico que la generaba era integrado de orden uno en la parte estacional y no integrado en la parte regular de la serie.

Para comprobar esta situación utilizemos el contraste de raíz unitaria, para ello, con los datos de la variable en pantalla se elige **View → Unit Root Test → Ok**. Se puede ver que el p-valor de la **t de Student** en el **Test de Dickey – Fuller Aumentado** (0.000) es menor que 0.05, lo que nos lleva a aceptar la estacionariedad de LSA (hecho que ya habíamos demostrado a partir de las funciones de autocorrelación y autocorrelación parcial).



Si repetimos estos pasos para DLSAS, o sea al diferenciar una vez la parte estacional de la serie en logaritmos se observa que la misma también presenta estacionariedad. Por lo tanto se comprueba que la serie regular en logaritmo es integrada de orden cero y la serie estacional es integrada de orden uno, ya que es necesario tomar las primeras diferencias para que la misma sea estacionaria.

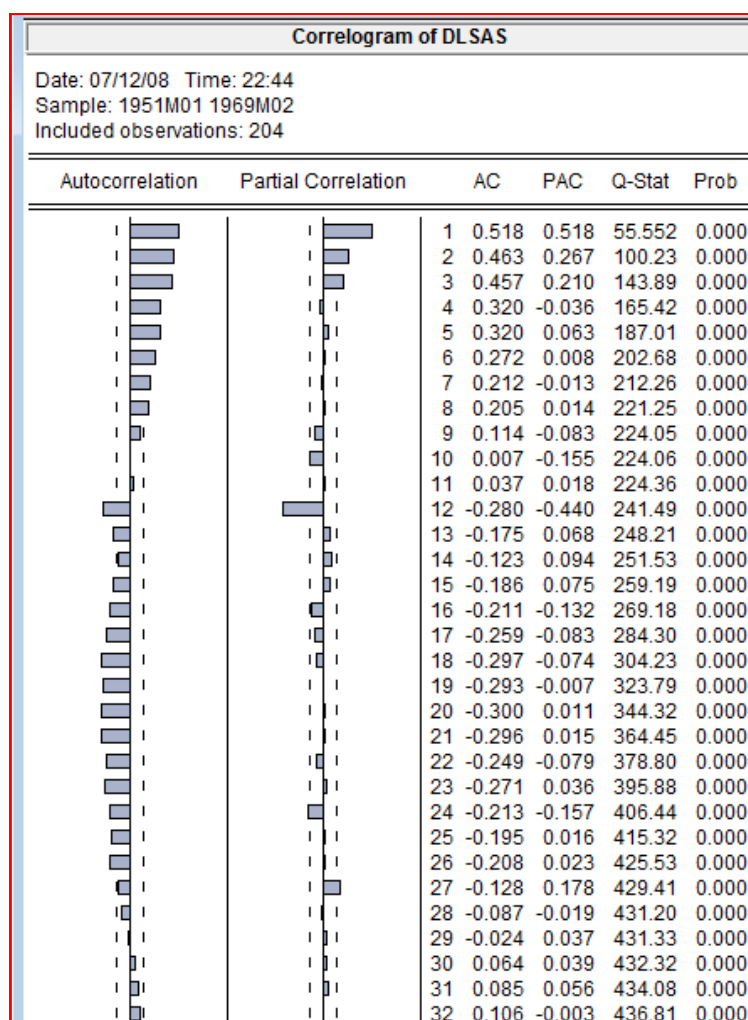
Se comprueba, entonces, todo lo realizado a partir de las funciones de autocorrelación y autocorrelación parcial y del análisis del respectivo correlograma.



Lo mismo se puede comprobar con el **Test de Phillips - Perron**. Para ello elegimos **Quick → Series Statistics → Unit Root Test**, rellenamos la pantalla **Series Name** con la variable **LSA** y al pulsar **OK** se obtiene la pantalla **Unit Root Test** en cuyo campo **Test Type** elegimos **Phillips Perron** y en **Test for unit root in** elegimos **Level** ya que estamos probando la estacionariedad sin diferenciar la serie. Al pulsar **Ok** se obtiene un p-valor menor que 0,05 en los resultados del contraste de **Phillips - Perron** lo que indica estacionariedad en el nivel de la serie SA. Invitamos al lector a realizar estos pasos.

Resta por identificar el orden de los procesos AR y MA. Para la obtención del orden (p, q) se realiza una comparación entre las características que 2 importantes funciones estadísticas presentan para los distintos modelos ARIMA teóricos y las características que tales funciones, pero muestrales, presentan en la serie objeto de estudio.

Tales funciones estadísticas son la (fac) y la $(facp)$, que son los dos instrumentos básicos en la fase de identificación del ARIMA, al permitirnos inferir el verdadero mecanismo subyacente que ha generado nuestros datos. Para ello volvemos a traer aquí el correlograma analizado oportunamente, el de la serie DLSAS:



Al observar estas dos funciones vemos que sus coeficientes no se anulan bruscamente con periodicidades y que sus estructuras se ajustan claramente a un $ARMA(1,1)(0,1)_{12}$.

La parte $AR(1)$ de la parte regular proviene del decrecimiento rápido inicial y las ondas sinusoidales de la (fac) añadido a que la $(facp)$ presenta sólo un coeficiente significativo en la mayoría de los períodos (salvo en el primero), anulándose bruscamente el resto de los coeficientes. Asimismo, la parte $MA(1)$ de la parte regular proviene de que la (fac) presenta un solo retardo significativo en la mayoría de los períodos (salvo en el primero).

En series con estacionalidad la identificación adquiere matices. En este caso, junto a la identificación del orden del autorregresivo y de la media móvil de la componente regular (ya comentado) debemos identificar los órdenes de la componente estacional. Para ello las reglas de identificación son similares a las comentadas para la parte regular, pero adaptadas a la frecuencia de la serie. Es decir, en una serie mensual, debemos prestar atención a los valores de las funciones para los retardos 12, 24, 36,... En una serie trimestral fijaremos la atención para los retardos 4, 8, 12,...

Para tales valores utilizaremos las mismas reglas: decrecimiento exponencial hacia cero de los valores de la función (fac) o $(facp)$ y número de coeficientes significativamente distintos de cero. Los resultados extraídos nos darán los órdenes de la componente estacional del modelo.

En conjunto, para modelos con componente regular y estacional podemos señalar cuatro reglas para la identificación de la parte estacional:

- 1) Un coeficiente significativo para un retardo no típico (es decir, que no corresponda a los primeros valores o a múltiplos de la estacionalidad) no deberá tenerse, en principio, en cuenta.
- 2) Coeficientes significativos cercanos a los retardos de estacionalidad (por ejemplo, en el once o trece, para datos mensuales) pueden ser asimilados como posible "contagio" de los coeficientes limítrofes. Se trata de los llamados *satélites*.
- 3) Coeficientes teóricos relativamente pequeños, como corresponde a los retardos, por ejemplo, de orden 4 en adelante en una función decreciente en forma exponencial o sinusoidal, no es fácil detectarlos como significativamente distintos de cero.
- 4) Precisamente, será la existencia de coeficientes significativos, en forma permanente, para retardos de orden 4 en adelante un aviso de posible fallo en la

eliminación de la tendencia. Un proceso no estacionario se caracteriza porque los coeficientes de autocorrelación van disminuyendo muy lentamente, en lugar de rápidamente, que se corresponde con situaciones correctamente estacionarias.

Visto lo anterior, la parte estacional de nuestra serie responde a un modelo $ARMA(1,1)(0,1)_{12}$ como lo habíamos establecido. La única duda posible sería considerar también $AR(1)$ la parte estacional.

Una vez identificado el modelo que subyace a nuestra serie, ya es posible la estimación de los parámetros del modelo.

5. Estimación de los coeficientes del Modelo

La fase de la estimación consiste en la obtención de unos valores numéricos para los parámetros θ_i y ϕ_i del modelo previamente identificado, de forma que contengan buenas propiedades estadísticas. Para ello, la metodología ARIMA realiza un proceso de búsqueda iterativo de tales valores, de prueba y error, hasta dar con los valores óptimos, puesto que las ecuaciones de resolución no son lineales.

La estimación presupone el conocimiento de valores no disponibles de la variable a analizar, así como de la variable de error, para lo cual los programas de estimación plantean distintas alternativas.

Además, en la búsqueda iterativa de las estimaciones, se debe partir de un valor inicial. Todos los programas lo realizan de forma automática pero existen ciertas relaciones teóricas que es importante conocer. Por ejemplo, en el caso más sencillo de un $AR(1)$ puede demostrarse que si

$$z_t = \phi z_{t-1} + a_t$$

el coeficiente de autocorrelación de orden 1 puede servir como estimación inicial de ϕ . Por ello, el proceso iterativo de búsqueda de la estimación óptima, puede iniciarse con $\hat{\phi} = r_1$.

Para un $AR(2)$ podría partirse de:

$$\hat{\phi}_1 = \frac{r_1(1-r_2)}{1-r_1^2}$$

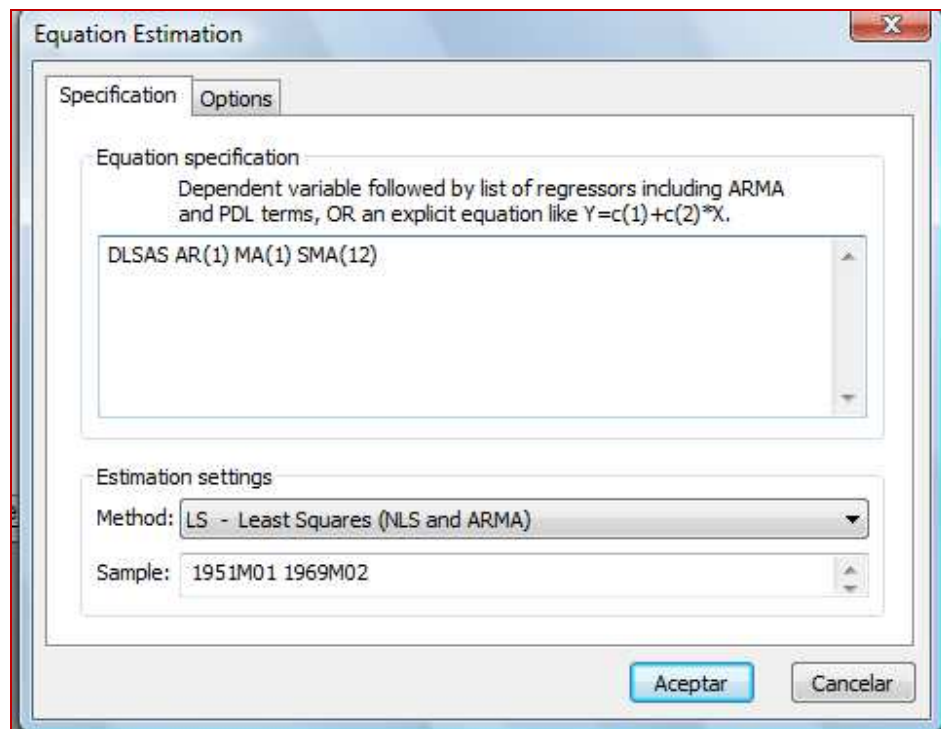
$$\hat{\phi}_2 = \frac{r_2-r_1^2}{1-r_1^2}$$

Para un **ARMA(1,1)** se tiene:

$$\hat{\phi} = \frac{r_2}{r_1}$$

$$\hat{\theta} = \frac{b \pm \sqrt{b^2 - 4}}{2} \quad \text{con } b = \frac{1 - 2r_2 + \hat{\phi}^2}{r_1 - \hat{\phi}}$$

Ya tenemos identificada completamente la serie original como un modelo **ARMA(1,0,1)(0,1,1)₁₂**. Es decir, ya hemos realizado el trabajo más importante en la modelización de una serie temporal mediante la metodología Box – Jenkins. Ahora, realizaremos su estimación. Para ello se elige **Quick → Estimate Equation**, se escribe la ecuación del modelo a ajustar en el campo **Equation Specification** de la solapa **Specification** teniendo en cuenta la estructura ARIMA previamente identificada, se elige **LS – Least Squares (NLS and ARIMA)** en el campo **Method** y se hace **CLICK** en **Aceptar**. Se obtienen los siguientes resultados:



Equation: UNTITLED Workfile: 4-3::Untitled\				
View Proc Object Print Name Freeze Estimate Forecast Stats Resids				
Dependent Variable: DLSAS				
Method: Least Squares				
Date: 07/16/08 Time: 20:39				
Sample (adjusted): 1952M02 1968M12				
Included observations: 203 after adjustments				
Convergence achieved after 10 iterations				
MA Backcast: 1951M01 1952M01				
	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.895466	0.044583	20.08537	0.0000
MA(1)	-0.513893	0.086498	-5.941087	0.0000
SMA(12)	-0.683951	0.059664	-11.46330	0.0000
R-squared	0.545357	Mean dependent var		-0.000419
Adjusted R-squared	0.540810	S.D. dependent var		0.083414
S.E. of regression	0.056525	Akaike info criterion		-2.893614
Sum squared resid	0.639006	Schwarz criterion		-2.844650
Log likelihood	296.7018	Hannan-Quinn criter.		-2.873805
Durbin-Watson stat	1.987643			
Inverted AR Roots	.90			
Inverted MA Roots	.97	.84+.48i	.84-.48i	.51
	.48+.84i	.48-.84i	.00+.97i	-.00-.97i
	-.48+.84i	-.48-.84i	-.84-.48i	-.84+.48i
	-.97			

Una vez estimado el modelo podemos interpretar los parámetros del mismo.

Por ejemplo, un **ARIMA (0,1,1)**, como dijimos, se puede expresar sintéticamente como:

$$\nabla y_t = \theta_1(B)a_t$$

y desarrollando esta expresión tenemos

$$(1-B)y_t = (1+\theta_1 B)a_t$$

$$y_t - y_{t-1} = a_t + \theta_1 a_{t-1}$$

Supongamos que la estimación nos arroja un valor de $\theta_1 = -0,75$. Entonces la predicción, puesto que a_t se hace cero, se realizaría como

$$y_t = y_{t-1} - 0,75a_{t-1}$$

Es decir, que el valor de la variable y en el período t será el valor de la variable en el período anterior menos el 75% del error de estimación cometido en aquel período.

Un **ARIMA (1,1,1)**, vendría dado por

$$\phi_1(B)\nabla y_t = \theta_1(B)a_t$$

$$(1 - \phi_1 B)(1 - B)y_t = (1 + \theta_1 B)a_t$$

$$(y_t - y_{t-1}) - \phi_1(y_{t-1} - y_{t-2}) = a_t + \theta_1 a_{t-1}$$

Si, por ejemplo, los valores estimados fueran $\phi_1 = 0,5$ y $\theta_1 = 0,75$ la predicción se realizaría como:

$$y_t = 1,5y_{t-1} - 0,5y_{t-2} - 0,75a_{t-1}$$

Esto es, en el período t la variable y se calcula como 1,5 por el valor de la variable en el período anterior menos la mitad del valor de la variable hace dos períodos y menos el 75% del error cometido en la estimación del período anterior.

6. Contraste de validez conjunta del modelo

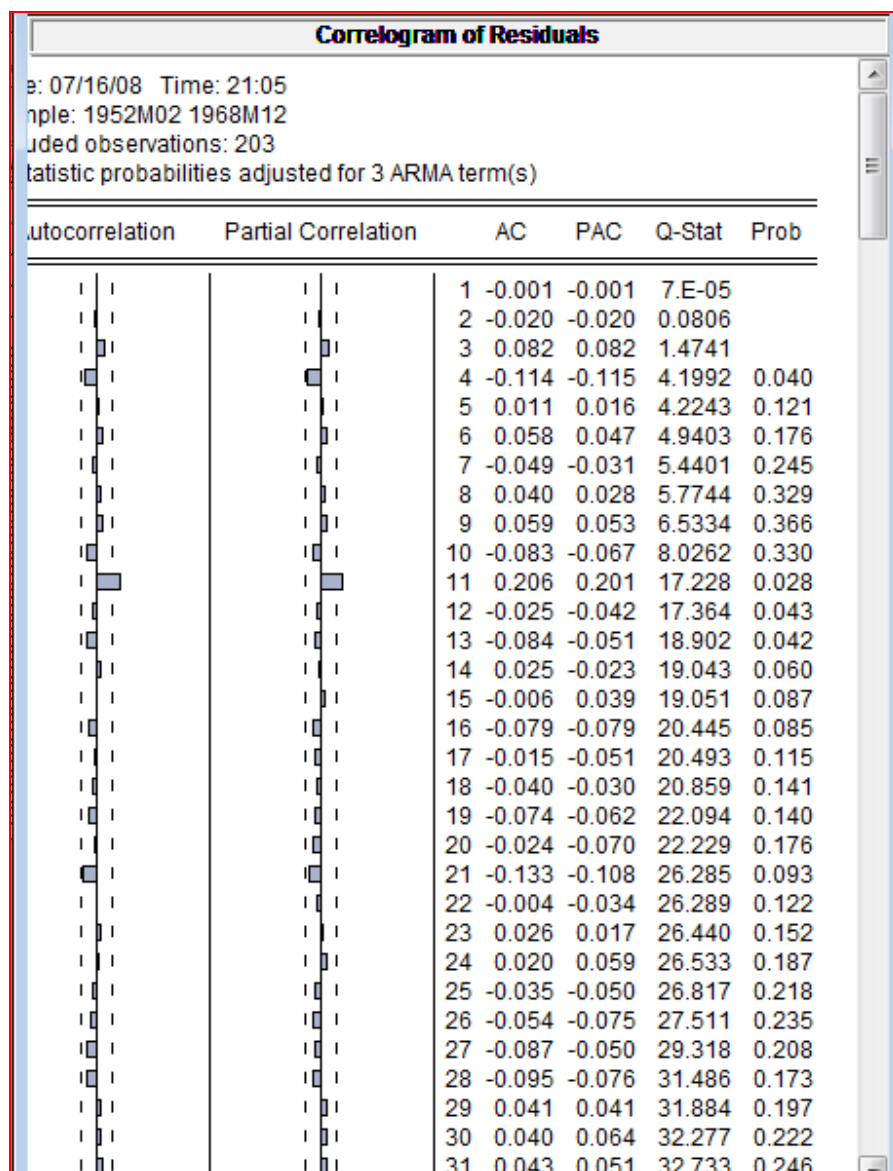
A la hora de realizar la etapa de validación (o contraste) del modelo podemos recurrir a diversos criterios y contrastes estadísticos. Todos los programas de ordenador recogen estos test.

El modelo para la serie SA presenta buena significatividad individual y conjunta de los parámetros estimados, altos coeficientes de determinación y un estadístico de Durbin y Watson casi igual a 2. Luego la diagnosis del ajuste es correcta.

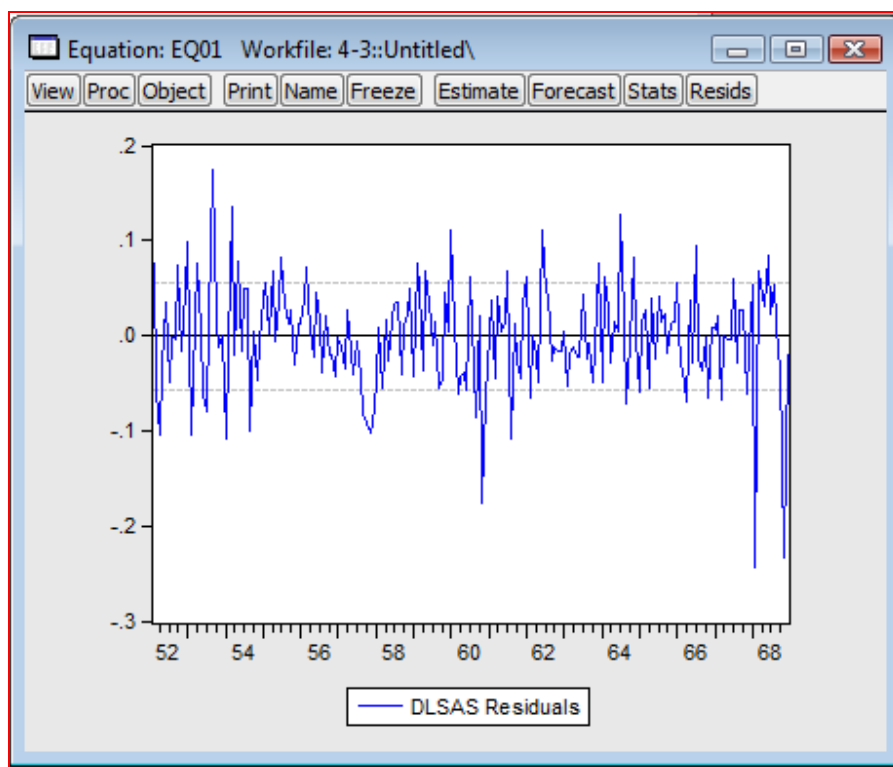
7. Análisis detallado de los errores

También es un buen instrumento de diagnosis el correlograma residual obtenido mediante *View → Residual Tests → Correlogram Q – Statistics*. Se observa que tanto la función de autocorrelación y la función de autocorrelación parcial no tienen retardos claramente significativos y además las probabilidades asociadas al estadístico Q son casi todas

mayores que 0.05, lo que indica que los residuos del modelo estimado se comportan como un ruido blanco:



Además, para validar el modelo podemos estudiar el gráfico de los residuos, que nos proporciona una visión de conjunto de la cuantía de los errores, sesgos sistemáticos y puntos de errores excepcionales (llamados *outliers*) que deben ser analizados con especial atención.



8. Selección del modelo y predicción

Seleccionado el modelo, puede pasarse a la etapa de la predicción. La verdadera predicción será la realizada a partir del último dato del período muestral, aunque también puede resultar útil analizar cómo se habría comportado el modelo si hubiera tenido que hacer una predicción dentro del período histórico ya conocido, y que ha servido de base a su estimación y contraste. Para ello disponemos de dos alternativas: la estática y la dinámica.

Pero antes de pasar a dichas alternativas es preciso expresar como ha quedado el modelo estimado en este ejemplo:

No olvidemos que $\{y_t\} = \{SA_t\}$ y entonces, $DLSAS = \ln(y_t) - \ln(y_{t-12})$ con lo que podemos escribir,

$$\ln(y_t) - \ln(y_{t-12}) = 0.89[\ln(y_t) - \ln(y_{t-12}) - \ln(y_{t-1}) - \ln(y_{t-12})] + e_t + 0.51e_{t-1} + 0.68e_{t-12} + 0.346e_{t-13}$$

$$0.11\ln(y_t) - 0.11\ln(y_{t-12}) + 0.89\ln(y_{t-1}) - 0.89\ln(y_{t-13}) = e_t + 0.51e_{t-1} + 0.68e_{t-12} + 0.346e_{t-13}$$

En la predicción estática se utilizan los valores verdaderos de las variables desplazadas. Desde el punto de vista de EViews, en primer lugar estimamos el modelo correspondiente, con **QUICK → ESTIMATE EQUATION**. Para predecir

utilizamos la instrucción **FORECAST** del menú de la ventana de ecuación, inmediatamente después de la estimación, porque ésta quedará residente en memoria sólo hasta la siguiente estimación que realicemos. Por ejemplo, si hemos estimado un modelo **ARMA(1,1)** tal como

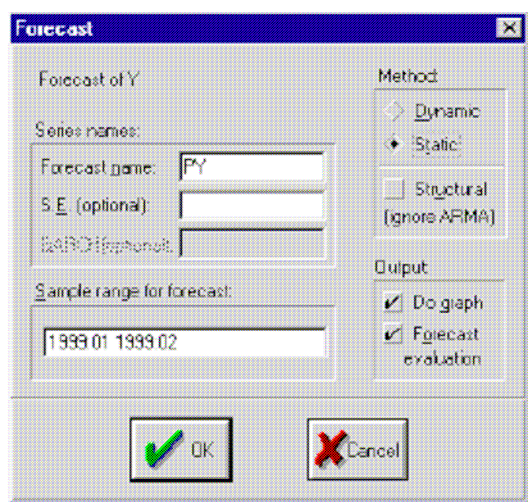
$$y_t = 0.7y_{t-1} + a_t + 0.3a_{t-1}$$

Que también se puede re-expresar como

$$y_t - 0.7y_{t-1} = a_t + 0.3a_{t-1}$$

La predicción estática para datos históricos se realizaría indicando, por ejemplo, que el período de predicción sea el comprendido entre los meses de enero y febrero de 1999.

Indicamos en el cuadro de diálogo que el método de estimación sea estático (**static**) y que las predicciones se guarden en una nueva variable denominada **PY**.



De esta forma, las predicciones se calcularán como

$$PY_{99:01} = 0.7 Y_{98:12} + 0.3 a_{98:12}$$

$$PY_{99:02} = 0,7 Y_{99:01} + 0,3 a_{99:01}$$

Por tanto, la predicción para febrero de 1999 se realizará con datos reales conocidos del mes de enero. Por el contrario, en una predicción dinámica, se indicaría como método, precisamente, el dinámico (*dynamic*):

Ahora las predicciones, contenidas en la variable **PY**, se calcularían como

$$PY_{99:01} = 0,7 Y_{98:12} + 0,3 a_{98:12}$$

$$PY_{99:02} = 0,7 PY_{99:01} + 0,3 a_{99:01}$$

Coinciden, por tanto, ambas predicciones (estática y dinámica) para el primer periodo, pero, a partir del **segundo**, la predicción dinámica utiliza el valor estimado y no el valor real del periodo precedente.

La predicción estática nos informa, pues, de los errores que hubiéramos cometido de utilizar el modelo para predecir sólo un periodo por delante. La predicción dinámica deja al modelo que vaya realimentando sus propias predicciones.

Naturalmente, a efectos de una auténtica predicción hacia futuro, sólo será posible el segundo tipo de predicción y, además, terminará perdiendo la corrección por los componentes de error. Así, si el modelo fue estimado con datos hasta finales de 1998, una predicción para dos periodos hacia delante daría:

$$PY_{99:01} = 0,7 Y_{98:12} + 0,3 a_{98:12}$$

$$PY_{99:02} = 0,7 PY_{99:01}$$

Cuando se disponga de nueva información real sobre los periodos de predicción, puede comprobarse el grado de error de nuestras estimaciones (un criterio a posteriori básico para enjuiciar la validez del modelo). Aunque puede re-estimarse el modelo con cada nueva estimación disponible, la práctica habitual es continuar con los mismos parámetros estimados durante varios periodos, aunque alimentando el modelo con los valores conocidos de los errores realmente cometidos.

9. Análisis de intervención.

Una limitación de la modelización ARIMA univariante está en que, provocados por distintos motivos, aparecen con frecuencia puntos atípicos (outliers), que dan lugar a errores excepcionalmente elevados. Se trata de circunstancias excepcionales que perturban la dinámica general de la serie que modelizamos.

Para incorporar tales fenómenos al modelo contamos con el llamado *análisis de intervención* con dos posibilidades principales: variables de impulso y variables de escalón.

¿Qué provoca puntos atípicos?

- Errores en la cuantificación de algún dato de la serie o cambio en el criterio de cálculo (cambios metodológicos no corregidos, errores de transcripción,...).
- Acontecimientos extraordinarios que afectan puntualmente al fenómeno en estudio (huelga, cambio de gobierno, devaluación, etc.)
- Variaciones en el comportamiento estacional (oscilación del período de semana santa en marzo o abril, cambios climatológicos para un mismo período en diversos años, etc.)
- Acciones especiales o intervenciones propiamente dichas (promociones especiales, aumento de tarifas, reforma fiscal, lanzamiento de un nuevo producto, etc.)

¿En qué consiste una variable de impulso?

Cuando para el análisis de intervención utilizamos las variables de impulso, consideramos que el fenómeno en cuestión sólo actúa en un punto aislado, o en diversos puntos pero separados entre sí. Por ejemplo, cuando modelizamos las ventas semanales de una empresa, debemos recoger en el modelo la posibilidad de que en determinados momentos se realizan promociones, o liquidaciones especiales.

Aspectos como estos (posibilidad de que en una semana haya o no ofertas, o huelgas,...) son aspectos cualitativos, definidos sobre la base de existencia o no de un fenómeno. Pues bien, el procedimiento ordinario para incluir aspectos cualitativos en modelos como los ARIMA pasa por añadir a nuestra especificación del modelo una (o varias) *variables ficticias (dummy)*. Esta variable adopta el valor 1 para el período en el que se produce el fenómeno (por ejemplo, en la cuarta semana de 2002 hubo una oferta especial) y 0 en el resto de períodos. En ese caso el modelo adoptaría la siguiente forma resumida:

$$y_t = \omega_0 x_t + v_t$$

donde v_t representa al modelo $ARMA(p, q)(P, Q)_s$. Por tanto, al modelo ARIMA univariante original le hemos añadido una nueva variable, que nos mide el efecto sobre la variable dependiente a través de un nuevo parámetro ω_0 , a estimar.

¿Qué es una variable de escalón?

Al contrario que para las variables de impulso, la utilización de una variable de escalón obedece a la idea de que el fenómeno que provoca la intervención afecta a todos los datos, en igual magnitud, durante un período de tiempo. Por ejemplo, supongamos que a partir de un determinado momento, los precios se elevasen en una determinada cuantía, y tal elevación se mantuviese, no en un solo período o varios alternos (una semana), sino que se mantuviese a partir de tal momento (durante una serie de semanas consecutivas, por ejemplo). En ese caso utilizaríamos una variable de escalón.

La aplicación de una variable de escalón es similar a la de la variable de impulso: se introduce una variable ficticia que toma el valor de 1 a partir del momento en que se produjo tal fenómeno, y 0 para los períodos anteriores.

APLICACIÓN

1. Se recomienda a los usuarios del curso que una vez seguido el caso de aplicación descrito en el acápite repliquen el mismo con una serie económica de su elección.
2. Se solicita trabajar con el caso de la **serie SA** y realizar las predicciones estáticas y dinámicas de la misma para el período **1969M01 1969M02**

TEST DE AUTOEVALUACIÓN

Marque la casilla de la respuesta que considere correcta:

- 1) En la modelización ARIMA se supone que una serie de n datos es...
 - a) n muestras de una variable aleatoria definida en cada momento del tiempo.
 - b) n muestras de otras tantas variables aleatorias.
 - c) Una muestra de tamaño 1 de una variable aleatoria.
 - d) Una muestra de tamaño 1 de un conjunto de n variables aleatorias.
- 2) Los modelos SARMA se diferencian de los ARMA en que...
 - a) Los ARMA se aplican a series estacionarias, mientras que los SARMA admiten series no estacionarias.
 - b) En los SARMA el desfase es coincidente con la frecuencia de los datos.
 - c) Los SARMA se aplican a series con estacionalidad, mientras que los ARMA no se pueden aplicar a este tipo de series.
 - d) En los SARMA además de estacionariedad las series son estacionales.
- 3) Los modelos ARMA se diferencian de los ARIMA en que...
 - a) Los ARIMA integran la metodología ARMA también para series no estacionarias.
 - b) Los ARIMA abarcan también series con estacionalidad, además de estacionarias.
 - c) En los ARIMA las series son estacionarias, mientras que en los ARMA las series no presentan esta cualidad.
 - d) La única diferencia es que para pasar de un modelo ARMA a uno ARIMA hay que integrar la serie.
- 4) Se entiende que una serie es estacionaria en sentido débil...
 - a) Cuando sus propiedades estadísticas permanecen invariables ante cambios de origen temporal.
 - b) Cuando se ha diferenciado y tomado logaritmos.
 - c) Cuando la media y la varianza son aproximadamente constantes para todo el período muestral.
 - d) Cuando la pendiente de la serie muestra una dispersión semejante para todo el periodo.
- 5) Si los test de detección estudiados nos muestran la existencia de 2 raíces unitarias en una serie entonces debemos...
 - a) Tomar una diferencia en la parte regular, y otra en la parte estacional.
 - b) Tomar dos diferencias en la parte estacional, siempre que la serie presente ese componente estacional.
 - c) Tomar dos diferencias en la parte regular, y comprobar la existencia de tendencia en la parte estacional.
 - d) Tomar dos diferencias en la parte regular y otras dos en la parte estacional.

- 6) En una serie en la cual el correlograma muestra una *fac* que decrece lentamente hacia cero, y una *facp* con un coeficiente significativo, se puede identificar...
- a) Un ARMA(1,0)
 - b) Un ARMA(0,1)
 - c) Un ARMA (1,0)*SARMA(1,0)
 - d) No se puede identificar un modelo. La serie no es estacionaria.
- 7) Observamos un correlograma en el que la *facp* decrece sinusoidalmente hacia cero, y una *fac* con los 3 primeros, y el quinto coeficientes significativos. Lo más razonable es que se trate de un modelo...
- a) ARMA(0,3)
 - b) ARMA(1, 1)
 - c) ARIMA(1,0,1)
 - d) ARIMA(1,1,1)
- 8) La expresión $\phi_2(B)\nabla y_t = \theta_1(B)a_t$ se corresponde con un modelo...
- a) ARIMA (3,1,1)
 - b) ARIMA (1,2,1)
 - c) ARIMA (2,1,1)
 - d) ARIMA (2,1,2)
- 9) señale la afirmación INCORRECTA referida a las etapas de una aplicación ARIMA:
- a) Una vez transformada la serie sólo debemos continuar por el resto de fases con un único modelo: el que se deriva de la fase de identificación.
 - b) La etapa de contraste nos sirve para comparar entre modelos alternativos.
 - c) Las fases de aplicación de modelos ARIMA son circulares: cuentan con un proceso de revisión permanente.
 - d) El análisis de los errores no sólo sirve para validar un modelo, también para observar si puede ser necesario realizar análisis de intervención.
- 10) Cuando en un modelo identificamos varios errores atípicos debemos...
- a) Tratar de identificar las razones económicas de la desviación en tales períodos de tiempo, y tratar de modelizarlos con análisis de intervención.
 - b) Revisar los datos por si se tratase de errores de transcripción.
 - c) Replantear el proceso, o incluso cambiar de técnica.
 - d) Todo lo anterior es posible, y aconsejable uno u otro en función del resultado.