

SERIES TEMPORALES, MODELO ARIMA METODOLOGÍA DE BOX - JENKINS



George Edward
Pelham Box



Gwilym Meirion
Jenkins

MODELO ARIMA(p, d, q) (P, D, Q)_s

Se han analizado las series temporales desde un punto de vista determinista o clásico. A partir de ahora se estudian desde un punto de vista estocástico o moderno, que utiliza métodos más complejos y su aplicación requiere series más largas.

Box y Jenkins han desarrollado modelos estadísticos para series temporales que tienen en cuenta la dependencia existente entre los datos, esto es, cada observación en un momento dado es modelada en función de los valores anteriores. Los análisis se basan en un modelo explícito. Los modelos se conocen con el nombre genérico de **ARIMA** (*AutoRegressive Integrated Moving Average*), que deriva de sus tres componentes **AR** (Autoregresivo), **I**(Integrado) y **MA** (Medias Móviles).

El modelo **ARIMA** permite describir un valor como una función lineal de datos anteriores y errores debidos al azar, además, puede incluir un componente cíclico o estacional. Es decir, debe contener todos los elementos necesarios para describir el fenómeno. Box y Jenkins recomiendan como mínimo 50 observaciones en la serie temporal.

La metodología de Box y Jenkins se resume en cuatro fases:

- La **primera fase** consiste en identificar el posible modelo **ARIMA** que sigue la serie, lo que requiere:
 - Decidir qué transformaciones aplicar para convertir la serie observada en una serie estacionaria.
 - Determinar un modelo **ARMA** para la serie estacionaria, es decir, los órdenes p y q de su estructura autorregresiva y de media móvil.
- La **segunda fase**: Seleccionado provisionalmente un modelo para la serie estacionaria, se pasa a la segunda etapa de estimación, donde los parámetros AR y MA del modelo se estiman por máxima verosimilitud y se obtienen sus errores estándar y los residuos del modelo.
- La **tercera fase** es el diagnóstico, donde se comprueba que los residuos no tienen estructura de dependencia y siguen un proceso de ruido blanco. Si los residuos muestran estructura se modifica el modelo para incorporarla y se repiten las etapas anteriores hasta obtener un modelo adecuado.
- La **cuarta fase** es la predicción, una vez que se ha obtenido un modelo adecuado se realizan predicciones con el mismo.

PASOS A SEGUIR PARA EL ANÁLISIS DE DATOS

1. **Recogida de datos:** Es conveniente disponer de 50 o más datos, y en el caso de series mensuales, trabajar entre seis y diez años completos.

- 2. Representación gráfica:** Es de gran utilidad disponer de un gráfico de la serie para decidir sobre la estacionariedad. En ocasiones, se utilizan medias y desviaciones típicas por subperiodo para juzgar sobre la estacionariedad de la serie.
- 3. Transformación previa de la serie:** Cuando la serie no es estacionaria en varianza se requiere una transformación logarítmica. No obstante, la transformación logarítmica es muy frecuente incluso en series con dispersión relativamente constante en el tiempo. Una práctica habitual es ensayar con la serie original y en logaritmos y comprobar resultados.
- 4. Eliminación de la tendencia:** La observación del gráfico de la serie indica la existencia o no de tendencia. Una tendencia lineal será corregida tomando primeras diferencias, que será el caso más frecuente. Una tendencia no lineal suele llevar en la práctica al uso de dos diferencias como mucho.
- 5. Identificación del modelo:** Consiste en determinar el tipo de modelo más adecuado, esto es, el orden de los procesos autorregresivos y de medias móviles de las componentes regular y estacional. Técnicamente esta decisión se toma en base a las funciones de autocorrelación (FAC) y autocorrelación parcial (FAC parcial), tanto en la parte regular como estacional. Es habitual terminar eligiendo entre los procesos más simples AR(1), AR(2), MA(1), MA(2) y ARMA(1,1), tanto en la parte regular como estacional. En caso de duda pueden seleccionarse varios modelos alternativos que serán estimados y contrastados posteriormente, para definir finalmente el modelo adoptado.
- 6. Estimación de los coeficientes del modelo:** Decidido el modelo, se procede a la estimación de sus parámetros, dado que se trata de un procedimiento iterativo de cálculo, pueden sugerirse valores iniciales.
- 7. Contraste de validez del modelo:** Se utilizan distintos 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.
- 8. Análisis detallado de los errores:** Se tendrán en cuenta las diferencias históricas entre valores reales y estimados por el modelo para su valoración final. Hay que verificar un comportamiento no sistemático de los mismos, así como analizar la posible existencia de errores especialmente significativos.
- 9. Selección del modelo:** En base a los resultados de pasos anteriores, se decide sobre el modelo adoptado.
- 10. Predicción:** El modelo seleccionado se utilizará como fórmula inicial de predicción.

IDENTIFICACIÓN PRÁCTICA DEL MODELO

Identificar un modelo significa utilizar los datos recogidos, así como cualquier información de cómo se general la serie temporal objeto de estudio, para sugerir un conjunto reducido de posibles modelos, que tengan muchas posibilidades de ajustarse a los datos. Ante una serie temporal empírica, se deben encontrar los valores (p, d, q) más apropiados.

- Si la serie temporal presenta una tendencia, lo primero que debe de hacerse es convertirla en estacionaria mediante una diferenciación de orden d . Una vez diferenciada la serie, una buena estrategia consiste en comparar los correlogramas de la función de autocorrelación (ACF) y la función de autocorrelación parcial (ACFP), proceso que suele ofrecer una orientación para la formulación del modelo orientativo.
- ♣ Los procesos autorregresivos presentan función de autocorrelación parcial (ACFP) con un número finito de valores distinto de cero. *Un proceso AR(p) tiene los primeros p términos de la función de autocorrelación parcial distintos de cero y los demás son nulos.*

Esta afirmación es muy fuerte, y en la práctica se considera que una muestra dada proviene de un proceso autorregresivo de orden p si los términos de la función de autocorrelación parcial son casi cero a partir del que ocupa el lugar p .

Un valor se considera casi cero cuando su módulo es inferior a $2/\sqrt{T}$. Los programas de ordenador constituyen la franja $(-2/\sqrt{T}, 2/\sqrt{T})$ y detectan los valores de la ACFP que caen fuera de ella.

- ♣ Los procesos de medias móviles presentan función de autocorrelación con un número finito de valores distintos de cero. *Un proceso MA(q) tiene los primeros q términos de la función de autocorrelación distintos de cero y los demás son nulos.*

Las dos propiedades descritas son muy importantes con vistas a la identificación de un proceso mediante el análisis de las funciones de autocorrelación y autocorrelación parcial.

El resumen de los pasos de identificación de un modelo de series temporales:

- 1) Decidir si X_t necesita ser transformada para eliminar la no estacionariedad en media p en la no estacionariedad en varianza (heteroscedasticidad). Puede ser conveniente utilizar logaritmos de la serie o aplicar la transformación de Box-Cox.
- 2) Determinación del grado d de diferenciación adecuado.
En general, la falta de estacionariedad se manifiesta en que los coeficientes de la función de autocorrelación estimada tienden a decrecer muy lentamente.
La pregunta es, ¿cuán lentamente ha de ser el decrecimiento de los coeficientes de la función de autocorrelación parcial (ACFP) para que el proceso sea estacionario?.
En general, solo ocasionalmente los datos económicos del correlograma dejarán de decrecer tras las primeras diferencias, y en este caso serían necesarias segundas diferencias. Una diferenciación superflua solo sirve para alterar el esquema de autocorrelación evidente en una serie estacionaria y complicarlo innecesariamente.
- 3) Decidir los valores de (p, q) , y si existe una componente estacional, decidir los órdenes de los operadores estacionales (P, Q). Para este apartado se utilizan las funciones de autocorrelación (ACF) y autocorrelación parcial (ACFP) según el siguiente cuadro:

Proceso	Función de autocorrelación (ACF)	Función de autocorrelación parcial (ACFP)
MA(q)	Solo los q primeros coeficientes son significativos. El resto se anulan bruscamente (coef. 0 para retardo >q)	Decrecimiento rápido exponencial atenuado u ondas sinusoidales.
AR(p)	Decrecimiento rápido exponencial atenuado u ondas sinusoidales.	Solo los p primeros coeficientes son significativos. El resto se anulan bruscamente (coef. 0 para retardo >q)
ARIMA(p, d, q)	Comportamiento irregular en los retardos (1, ..., q) con q picos. Decrecimiento para retardos posteriores a q .	Decrece (aproximadamente con exponenciales atenuados y ondas sinusoidales). No cero pronto.

DETENCIÓN PRÁCTICA DE LA ESTACIONARIEDAD

- Para detectar rápidamente la estacionariedad (**Analizar/Estadísticos descriptivos/ Explorar**) se pueden calcular la sucesión de medias y varianzas por años, si se obtienen variaciones significativas crecientes y decrecientes a lo largo de los años, indica que no hay estacionariedad. Este resultado conduce a tomar logaritmos y diferenciar la serie original con el objetivo de atenuar la falta de estacionariedad en media y varianza.
- Otro método (**Analizar/Series temporales/Autocorrelaciones**), si los coeficientes de la ACF no decaen rápidamente hay un indicio claro de falta de estacionariedad en media, lo que llevaría a tomar primeras diferencias en la serie original.

Si hay duda sobre diferenciar o no, o sobre cuántas veces hay que diferenciar, se calcula la varianza de la serie original y de la serie sometida a diferentes diferenciaciones, tomando como diferenciación adecuada aquella para la que la varianza es mínima. El método es tanto más adecuado cuanto mayor se la diferencia entre las varianzas anteriores. La *sobrediferenciación* suele evitarse observando si en la parte de medias móviles alguna raíz es próxima a la unidad.

DETENCIÓN PRÁCTICA DE LA ESTACIONALIDAD: PERIODGRAMA

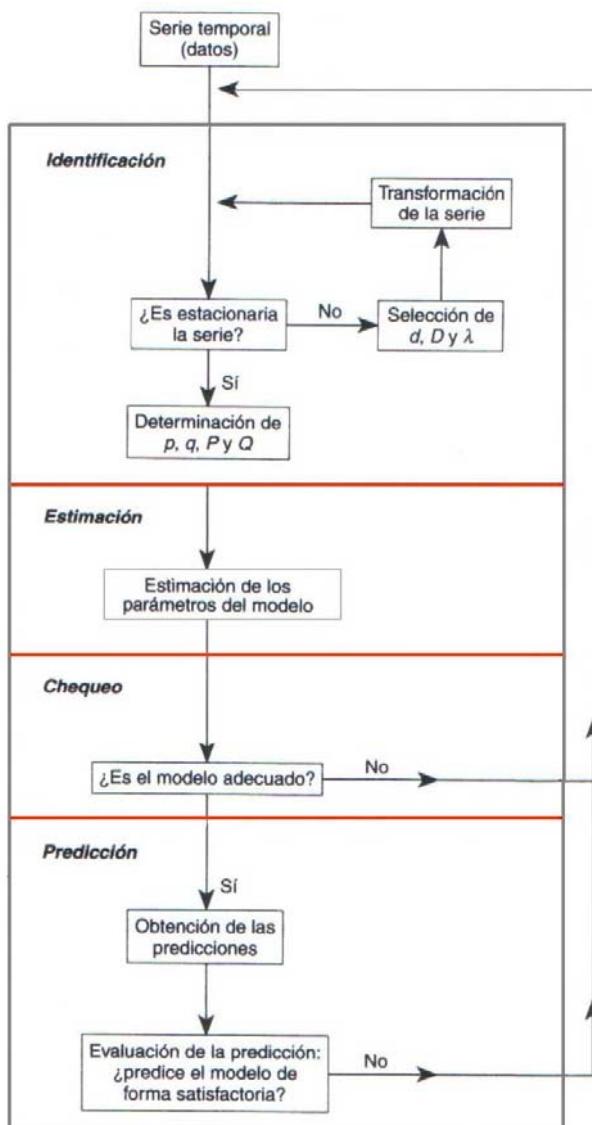
Para detectar rápidamente la estacionalidad se puede utilizar directamente el gráfico de la serie (**Analizar/Series temporales/Análisis espectral**), así se obtiene el PERIODGRAMA, que es una figura que transforma la serie temporal de su dominio natural (el tiempo) al dominio de las frecuencias (a los valores de la serie se les aplican transformaciones de Fourier), en el eje X se presentan frecuencias y en el eje Y las amplitudes. Respecto al PERIODGRAMA hay que establecer las siguientes consideraciones:

- No hay estacionalidad si no hay picos destacables.
- Cada pico destacable identifica un período que incluso puede ser un ciclo.
- A cada amplitud destacable le corresponde una frecuencia cuya inversa es el período estacional o el ciclo, con lo que *el periodograma identifica la longitud del período estacional y en su caso el ciclo*.

- Las amplitudes más fuertes (correspondientes a valores más bajos de las frecuencias) suelen corresponder a ciclos y las menos fuertes (correspondientes a valores no tan bajos de las frecuencias) suelen corresponder a estaciones. En caso de duda entre ciclos y estaciones se puede recurrir a las funciones de autocorrelación para discriminar.
- PERIODGRAMA ACUMULATIVO, representa en el eje de abscisas las frecuencias y en el eje de ordenadas las amplitudes acumuladas. En esta línea, hay que hacer las consideraciones:
 - Para una serie aleatoria coincide con la diagonal del primer cuadrante.
 - Desvíos bruscos de la diagonal provocan presencia de ciclos o estaciones para las respectivas frecuencias, que serán ciclos cuando las frecuencias sean bajas.

La estacionalidad, así como la estacionariedad, también puede detectarse a través de las funciones de autocorrelación (ACF) y autocorrelación parcial estimadas (ACFP) - **Analizar/Series temporales/Autocorrelaciones** (en Opciones para representar ACF con un tramo significativo se eligen 36 retardos) -

METODOLOGÍA BOX-JENKINS



MODELOS AUTORREGRESIVOS AR(p)

Un modelo autorregresivo **AR** describe una clase particular de proceso en que las observaciones en un momento dado son predecibles a partir de las observaciones previas del proceso más un término de error. El caso más simple es el ARIMA(1,0,0) o AR(1) o de primer orden, cuya expresión matemática es:

$$AR(1) \equiv X_t = \phi_1 X_{t-1} + a_t$$

El proceso autorregresivo de orden p, representado por ARIMA(p,0,0) o simplemente por AR(p):

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

que puede ponerse, mediante el operador de cambio retroactivo B, en la forma:

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) X_t = a_t \quad B^k (X_t) = X_{t-k}$$

- ♣ Un proceso autorregresivo AR(p) es **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 ecuación: $x^p - \phi_1 x^{p-1} - \phi_2 x^{p-2} - \dots - \phi_{p-1} x - \phi_p = 0$ sean todas inferiores a uno en módulo.
- ♣ Un proceso autorregresivo siempre es **invertible**.

MODELO DE MEDIAS MÓVILES Ma(q)

Un modelo de medias móviles **MA** describe una serie temporal estacionaria. En este modelo el valor actual puede predecirse a partir de la componente aleatoria de este momento y, en menor medida, de los impulsos aleatorios anteriores. El modelo ARIMA(0,0,1), también denotado por MA(1), viene dado por la expresión:

$$X_t = a_t - v_1 a_{t-1}$$

El proceso de medias móviles de orden q, representado por ARIMA(0,0,q) o también por Ma(q), viene dado por la expresión:

$$X_t = a_t - v_1 a_{t-1} - v_2 a_{t-2} - \dots - v_q a_{t-q}$$

que puede ponerse, mediante el operador de cambio retroactivo B, en la forma:

$$X_t = (1 - v_1 B - v_2 B^2 - \dots - v_q B^q) a_t$$

- ♣ Un proceso de medias móviles es siempre **estacionario**.
- ♣ Un proceso de medias móviles Ma(q) es **invertible** si las raíces del polinomio en B definido por $(1 - v_1 B - v_2 B^2 - \dots - v_q B^q)$ caen fuera del círculo unidad. Esta condición es equivalente a que las raíces de la ecuación $x^q - \phi_1 x^{q-1} - \phi_2 x^{q-2} - \dots - \phi_{q-1} x - \phi_q = 0$ sean todas inferiores a uno en módulo.

MODELOS ARMA (p, q)

Una extensión natural de los modelos **AR(p)** y **MA(q)** es un tipo de modelos que incluyen tanto términos autorregresivos como de medias móviles y se definen como **ARIMA(p, 0, q)**. Se representan por la ecuación:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + a_t - v_1 a_{t-1} - v_2 a_{t-2} - \cdots - v_q a_{t-q}$$

que puede ponerse de la forma:

$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \cdots - \phi_p X_{t-p} = a_t - v_1 a_{t-1} - v_2 a_{t-2} - \cdots - v_q a_{t-q}$$

$$\text{es decir, } X_t(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p) = a_t(1 - v_1 B - v_2 B^2 - \cdots - v_q B^q)$$

El proceso **ARMA(p, q)** es **estacionario** si lo es su componente autorregresiva, y es **invertible** si lo es su componente de medias móviles.

- Un modelo **ARMA(p, q)** es **estacionario** si las raíces del polinomio definido por $(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p)$ caen fuera del círculo unidad. Esta condición es equivalente a que las raíces de la ecuación:

$$x^p - \phi_1 x^{p-1} - \phi_2 x^{p-2} - \cdots - \phi_{p-1} x - \phi_p = 0 \text{ sean todas inferiores a uno en módulo.}$$

- Un modelo **ARMA(p, q)** es **invertible** si las raíces del polinomio en B definido mediante $(1 - v_1 B - v_2 B^2 - \cdots - v_q B^q)$ caen fuera del círculo unidad. Esta condición es equivalente a que las raíces de la ecuación:

$$x^q - \phi_1 x^{q-1} - \phi_2 x^{q-2} - \cdots - \phi_{q-1} x - \phi_q = 0 \text{ sean todas inferiores a uno en módulo.}$$

MODELOS ARIMA (p, d, q)

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

El modelo (0, d, 0) se expresa mediante: $(1 - B)^d X_t = a_t$

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

$$(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p)(1 - B)^d X_t = (1 - v_1 B - v_2 B^2 - \cdots - v_q B^q)a_t$$

Un modelo ARIMA(p, d, q) 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 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 cílicas, también denominadas variaciones estacionales. Las variaciones cílicas a veces se superponen a una tendencia secular.

Las series con tendencia secular y variaciones cílicas pueden representarse mediante los modelos **ARIMA(p, d, q)(P, D, Q)**. El primer paréntesis (**p, d, q**) se refiere a la **tendencia secular o parte regular** de la serie y el segundo paréntesis (**P, D, Q**) se refiere a las **variaciones estacionales, o parte cílica** de la serie temporal.

En este sentido, se adjuntan algunas expresiones del modelo:

- **ARIMA(0, 1, 1)(0,0,1)₁₂**: $(1 - B) X_t = (1 - \nu_1 B^{12}) (1 - \delta_{12} B^{12})$
- **ARIMA(0, 1, 1)(0,1,1)₁₂**: $(1 - B) (1 - B^{12}) X_t = (1 - \nu_1 B^{12}) (1 - \delta_{12} B^{12})$
- **ARIMA(2, 1, 0)(1,0,0)₁₂**: $(1 - \phi_1 B - \phi_2 B^{12}) (1 - \Omega_1 B^{12}) (1 - B) X_t = a_t$
- **ARIMA(1, 1, 1)(2,1,1)₁₂**: $(1 - \phi_1 B) (1 - \Omega_1 B^{12} - \Omega_2 B^{24}) (1 - B^{12}) (1 - B) X_t = (1 - \nu_1 B) (1 - \delta_{12} B^{12}) a_t$

IDENTIFICACIÓN DE MODELOS ARIMA(p, d, q)

Un **proceso estocástico** $(X_t)_{t=1,2,3,\dots}$ se define como una colección de variables aleatorias X_t ordenadas de acuerdo con el parámetro t tiempo.

Los modelos estocásticos de series temporales contemplan una serie temporal X_t como una colección de observaciones muéstrales, cada una correspondiente a una variable del proceso.

Las leyes de probabilidad que rigen cualquier proceso estocástico se describen exhaustivamente mediante las funciones de distribución de probabilidad conjunta de todos y cada uno de los vectores de variables aleatorias que se pueden formar con las variables que constituyen el proceso. No obstante, con finalidad práctica, los procesos estocásticos se suelen describir mediante sus momentos.

La media del proceso estocástico se define por $u_t = E(X_t)$ y generalmente es una función del tiempo. La función de autocovarianza se define como:

$$g(t, t+k) = \text{Cov}(X_t, X_{t+k}) = E((X_t - E(X_t))(X_{t+k} - E(X_{t+k}))) \quad t = \dots -3, -2, -1, 0, 1, 2, 3, \dots$$

A partir de la función de autocovarianza se obtienen dos resultados útiles:

- Función de varianza del proceso: $g(t, t) = \text{Var} X_t$
- Función de autocorrelación: $h(t, t+k) = \frac{g(t, t+k)}{\sqrt{g(t, t) g(t+k, t+k)}}$

ESTACIONARIEDAD

Un proceso estocástico es **estacionario en sentido estricto** si los vectores $[X_{t_1}, X_{t_2}, \dots, X_{t_n}]$ y $[X_{t_1+s}, X_{t_2+s}, \dots, X_{t_n+s}]$ tienen la misma función de distribución de probabilidad, independientemente de s , para cualquier n dado. Esta definición de estacionariedad implica que las características del proceso estocástico no sufren alteración en tiempos históricamente diferentes, condición quizá demasiado fuerte para imponer en la práctica.

Un proceso es **estacionario en sentido amplio** (o estacionario de segundo orden, o de covarianza estacionaria, o débilmente estacionario) cuando se verifica que $u_t = u < \infty$ y $g(t, t+k) = g_k < \infty$, es decir, la media del proceso es constante (no depende del tiempo) y la autocovarianza es solo función del lapso temporal considerado, y no del tiempo histórico. Los momentos de orden superior pueden variar con el tiempo.

En el caso de procesos con función de distribución de probabilidad normal, la estacionariedad en sentido amplio implica la estacionariedad en sentido estricto.

FUNCIÓN DE AUTOCORRELACIÓN

En procesos estacionarios, la función de autocorrelación es:

$$h_k = \frac{g_k}{g_0} = \frac{\text{Cov}(X_t, X_{t+k})}{\text{Var}(X_t)} \quad k = \dots, -2, -1, 0, 1, 2, \dots$$

Para procesos reales se verifica además que $g_0 > 0$, $g_k = g_{-k}$, $h_k = h_{-k}$, $h_0 = 1$ y $|h_k| \leq 1$

Se denomina **correlograma** del proceso a la representación gráfica con h_k en ordenadas y k en abscisas.

La función de autocorrelación de las series estacionarias disminuye sensiblemente a medida que aumenta el desfase temporal k . Esta característica no suele suceder en las series no estacionarias.

FUNCIÓN DE AUTOCORRELACIÓN ESTIMADA

En aplicaciones prácticas, en las que se dispone de ciertas observaciones, $(X_t)_{t=1,2,3,\dots,T}$, relativas a un proceso estocástico que se supone estacionario, la media del proceso se estima mediante: $\bar{X} = \sum_{t=1}^T \frac{X_t}{T}$

Análogamente, la función de autocorrelación h_k se estima mediante la *función de autocorrelación*

$$\text{muestral, que se define por } r_k = \frac{\sum_{t=1}^T (X_t - \bar{X})(X_{t-k} - \bar{X})}{\sum_{t=1}^T (X_t - \bar{X})^2}$$

Se denomina **correlograma muestral** a la representación gráfica de r_k , instrumento de gran interés práctico del análisis de series temporales.

Para obtener correlogramas debe partirse en la práctica de muestras de tamaño suficientemente grande (al menos 50 observaciones).

La función de autocorrelación muestral no se puede calcular cuando $(k > T + 1)$, y en la práctica no debe calcularse para $(T > T / 4)$

RUIDO BLANCO

Es un proceso puramente aleatorio, se define por las condiciones:

$$u = E(X_t) = 0, g_0^2 = \text{var}(X_t), g_k = \text{cov}(X_t, X_{t+k}) = 0 \quad k = \dots, -2, -1, 0, 1, 2, \dots$$

En este tipo de procesos puramente aleatorios el **correlograma** se reduce a un segmento de longitud unitaria sobre el eje de ordenadas.

ESTACIONARIEDAD Y ELIMINACIÓN DE LA TENDENCIA

Muy pocas series temporales reales del mundo económico son **estacionarias**. La mayoría presentan tendencia, varianza no constante y variaciones estacionales.

La presencia de variaciones estacionales se traduce en una variabilidad de la media del proceso, lo que es contrario a la hipótesis de estacionariedad.

Normalmente, es posible transformar muchas series económicas reales no estacionarias en otras aproximadamente estacionarias, sometiéndolas a operaciones algebraicas adecuadas.

A las series no estacionarias que presentan una tendencia lineal se les somete a la transformación $Z_t = X_t - X_{t-1}$ para convertirlas en estacionarias. Si X_t muestra una tendencia lineal, la primera diferencia de la serie Z_t ya no tendrá esa tendencia. En este caso se dice que X_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 puede conseguirse mediante una doble diferenciación. Esta operación se realiza en dos etapas, primero se obtiene $W_t = X_t - X_{t-1}$ 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 X_t es una *serie temporal homogénea 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 I(p) de orden p*.

TRANSFORMACIÓN LOGARÍTMICA

En general, se denomina *proceso homogéneo de orden h*, o *integrado de orden h*, denotado por $I(h)$, a un proceso no estacionario que se convierte en estacionario después de h operaciones de diferencias y no antes.

Si X_t muestra una tendencia exponencial, puede eliminarse la tendencia hallando primero el logaritmo de la serie, y luego la diferencia primera de la nueva serie así calculada. La serie $Z_t = \ln X_t - \ln X_{t-1}$ puede tener la tendencia eliminada.

TRANSFORMACIÓN DE Box-Cox

Permite estabilizar la varianza de una serie temporal (*serie estacionaria en varianza*) y aproximar su distribución a una normal.

Si X_t es la serie temporal inicial, la transformación viene dada por:

$$\begin{cases} Z_t = \frac{(X_t + l_2)^{l_1-1}}{l_1 g^{l_1-1}} & \text{si } l_1 \neq 0 \text{ y } X_t > -l_2 \\ Z_t = g \ln(X_t + l_2) & \text{si } l_1 = 0 \text{ y } l_2 < 0 \end{cases}$$

donde g es la media geométrica simple de $X_t + l_2$, el primer parámetro l_1 gobierna la fuerza de la transformación. Para $l_1 = 1$ se tiene la serie original X_t y l_2 se elige de forma que $X_t + l_2$ sea siempre positiva. En consecuencia, l_2 será cero si se trabaja con datos positivos e igual en valor absoluto al valor más negativo observado, en otro caso.

La transformación de Box-Cox es una familia de transformaciones dependiente del parámetro l_1 , que incluye como casos particulares la transformación logarítmica, la raíz cuadrada y la inversa.

La *eliminación de las variaciones estacionales, para inducir la estacionariedad*, suele hacerse casi siempre, mediante la diferenciación estacional.

Si los datos son mensuales, la diferenciación estacional de la serie temporal X_t consiste en calcular $Z_t = X_t - X_{t-12}$. Con datos trimestrales se calcula $Z_t = X_t - X_{t-4}$. 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 sucesivamente.

FUNCIÓN DE AUTOCORRELACIÓN PARCIAL ESTIMADA

Un concepto muy útil en el análisis de series temporales es la función de autocorrelación parcial.

El primer término de la función de autocorrelación parcial se denota por ϕ_{11} , puede estimarse transformando la serie X_t en desviaciones respecto a su media muestral $Y_t = X_t - \bar{X}$ y a continuación estimando una regresión de Y_t sobre Y_{t-1} .

El modelo de regresión: $Y_t = \phi_{11} Y_{t-1} + u_t$, la pendiente estimada de esta regresión es $\hat{\phi}_{11}$.

Además, el primer valor de la función de autocorrelación parcial ϕ_{11} es igual al primer valor de la función de autocorrelación, propiedad de las funciones de autocorrelación de todo proceso estocástico estacionario.

- El segundo valor de la función de autocorrelación parcial ϕ_{22} se estima mediante una regresión de Y_t sobre Y_{t-1} e Y_{t-2} . El modelo de regresión: $Y_t = \phi_{21} Y_{t-1} + \phi_{22} Y_{t-2} + u_t$.
- El tercer valor de la función de autocorrelación parcial ϕ_{33} se estima mediante una regresión de Y_t sobre Y_{t-1} , Y_{t-2} e Y_{t-3} . El modelo de regresión: $Y_t = \phi_{31} Y_{t-1} + \phi_{32} Y_{t-2} + \phi_{33} Y_{t-3} + u_t$.

La función de autocorrelación parcial puede estimarse mediante una serie de regresiones, cada una de las cuales contiene como variable explicativa un retardo más que la anterior, y en cada caso se eligen los coeficientes estimados en los retardos más altos ($\phi_{11}, \phi_{22}, \phi_{33}, \dots$), que son así los valores estimados de la función de autocorrelación parcial.

Otra manera de obtener la función de autocorrelación parcial estimada es mediante fórmulas recursivas, utilizando la función de autocorrelación previamente estimada y utilizando las ecuaciones de Yule-Walker. A veces se suele denominar **correlograma** a la representación gráfica de las funciones de autocorrelación y autocorrelación parcial.

IDENTIFICACIÓN DEL TÉRMINO INDEPENDIENTE

Para ajustar la serie temporal a veces conviene introducir un término independiente en el modelo ARIMA.

Para contrastar la hipótesis nula de que el modelo se ajusta con una constante, se utiliza el estadístico:

- Si X_t es ruido blanco (proceso aleatorio), el estadístico: $t_{N-1} = \frac{\bar{X}}{\sqrt{S_x^2/(N-1)}}$ (t-Student)
- Si X_t está autocorrelacionada, siendo significativos los primeros k coeficientes de autocorrelación (r_1, r_2, \dots, r_k) , el estadístico a utilizar es: $t_{N-1} = \frac{\bar{X}}{\sqrt{S_x^2(1+2r_1+2r_2+\dots+2r_k)/N}}$

ESTIMACIÓN DE MODELOS ARIMA(p, d, q)

Los parámetros se suelen obtener de manera que la suma cuadrática de los errores sea la menor posible. Representando el proceso ARIMA(p, d, q) de la forma $\phi(B) X_t = v(B) a_t$, los errores del modelo pueden expresarse de la forma $a_t = \phi^{-1}(B) \phi(B) a_t$.

El objetivo es encontrar el vector de parámetros $\phi = \phi(\phi_1, \dots, \phi_p)$ y $v = (v_1, \dots, v_p)$ que minimice la suma de cuadrados de los errores $\sum_t a_t^2 = S(\phi, v)$.

La estimación es complicada ya que la ecuación es no lineal en los parámetros. Se debe recurrir a un método iterativo de estimación no lineal (Marquardt). Para comenzar el algoritmo se necesitan estimaciones preliminares de los parámetros, que se obtienen mediante el método de los momentos.

DIAGNÓSTICO, VALIDACIÓN O CONTRSTE DE MODELOS ARIMA(p, d, q)

Box y Jenkins sugirieron un número considerable de tests para verificar si el modelo elegido se ajusta correctamente al conjunto de datos dado. Uno de ellos, conocido como **sobreparametrización**, consiste en ajustar un modelo de orden superior al elegido y comprobar si los parámetros son significativamente distintos de cero.

- De otro lado, si el modelo se aproxima satisfactoriamente a la serie observada, los residuos deben tender a comportarse como ruido blanco, lo que se comprobaría mediante las funciones de autocorrelación de los residuos (ACF, ACFP). Dichas funciones de autocorrelación deben de ser nulas en todo su recorrido, excepto en el cero.
- Si el modelo no se aproxima satisfactoriamente a la serie observada, los residuos se comportarían como un ruido autocorrelado. Por ello, deben emplearse contrastes como el de Durbin-Watson (para la autocorrelación de primer orden) o el de Wallis (para la de cuarto orden).

Otros tests aplicados a los residuos van encaminados a comprobar si los residuos obtenidos son consistentes con el supuesto de ruido blanco (aleatorios):

- ◆ Box y Pierce proponen el estadístico $Q = \sum_{k=1}^m r_k^2$ donde $r_k = \frac{\sum_{t=k+1}^n a_t a_{t-k}}{\sum_{t=1}^n a_t^2}$, siendo a_t ≡ residuos

estimados y n el número de observaciones. Bajo el supuesto de que m es suficientemente grande, Box y Pierce demuestran que el estadístico Q se distribuye como una Chi-cuadrado con $(m-p-q)$ grados de libertad. Rechazándose la hipótesis de que los residuos son un ruido blanco para valores de Q muy altos. Más concretamente, se halla la región crítica para un nivel de significación α , calculando un valor I que cumpla $P(Q > I) = \alpha$. Cuando el valor de Q cae dentro de la región crítica se rechaza la hipótesis nula de que los residuos son un ruido blanco. Si cae fuera de la región crítica se acepta la hipótesis nula. El valor de m es arbitrario, aunque conviene tomarlo lo más elevado posible.

Para valores de m no muy grandes, Ljung y Box proponen un estadístico alternativo:

$$Q' = \frac{n(n+2) \sum_{k=1}^m r_k^2}{(n-k)} \approx \chi^2_{m-p-q}$$

Se halla la región crítica para un nivel de significación α , calculando un valor I que cumpla $P(Q' > I) = \alpha$. Cuando el valor de Q' cae dentro de la región crítica se rechaza la hipótesis nula de que los residuos son un ruido blanco. Si cae fuera de la región crítica se acepta la hipótesis nula.

- ◆ Un diagnóstico completo surge de la **inspección del gráfico de los residuos**. Si los residuos provienen de un proceso de ruido blanco, deben de ser incorrelacionados entre sí, lo que les hará alternar en signo, sin ningún criterio obvio. Por el contrario, rachas de residuos consecutivos de un mismo signo son, en general, un indicativo de mala especificación del modelo, bien por ser una indicación de autocorrelación de los residuos o por indicar no estacionariedad en los mismos. Si el gráfico (t, a_t) tiene una tendencia conocida, puede haber heteroscedasticidad de los residuos.

PREDICCIÓN EN MODELOS ARIMA

Los modelos ARIMA proporcionan, además de una predicción puntual, la distribución de probabilidad completa para los futuros valores de la serie.

Considerando como predicción óptima la que tiene un error cuadrático medio de predicción mínimo, se trata de elegir una predicción a horizonte l , $Z_t(l)$, tal que $E[e_t^2(l)] = E[X_{t+1} - Z_t(l)]^2$ fuese mínimo.

En general, se demuestra que dicha predicción viene dada por la esperanza condicionada de X_{t+1} :

$$Z_t(l) = E[X_{t+1} | X_t, X_{t-1}, \dots, X_1]$$

El cálculo real de la predicción $Z_t(l)$ puede hacerse de forma recursiva utilizando el modelo ARIMA estimado, de forma que si el modelo se expresa como:

$$d_t = \phi_1 d_{t-1} + \phi_2 d_{t-2} + \dots + \phi_p d_{t-p} + a_t - v_1 a_{t-1} - v_2 a_{t-2} - \dots - v_q a_{t-q}$$

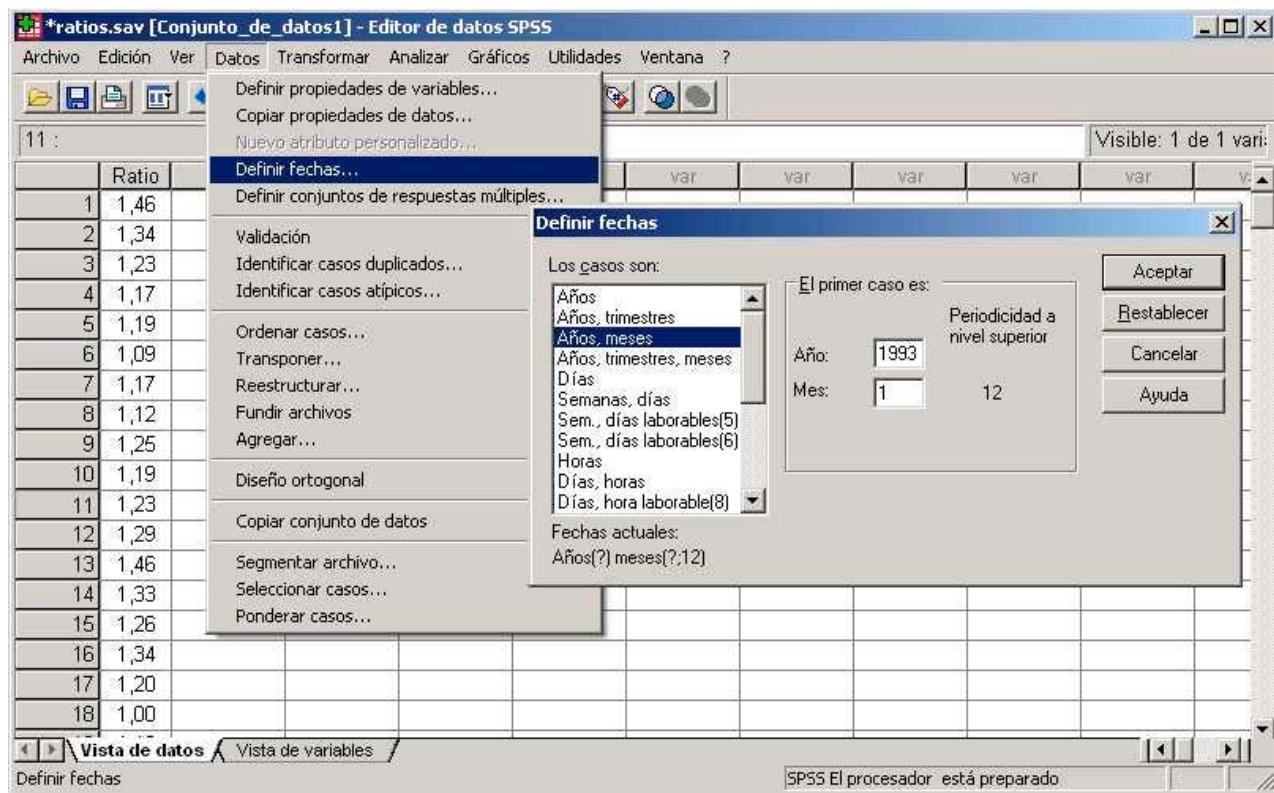
donde d_t ≡ diferencia de orden d de X_t (supuesto X_t no estacionaria y convertible en estacionaria mediante un proceso de d diferenciaciones consecutivas).

Para calcular la predicción $Z_t(l)$ se comienza calculando la estimación de $d_t(1)$ como la esperanza condicionada de d_{t+1} , y posteriormente se calcula la estimación de $d_t(2)$, y así sucesivamente hasta calcular la estimación de $d_t(l)$. Una vez que la serie d_t ha sido predicha, se puede obtener una predicción de X_t sumando d_t d -veces. Para calcular la predicción $Z_t(l)$ se utiliza la fórmula:

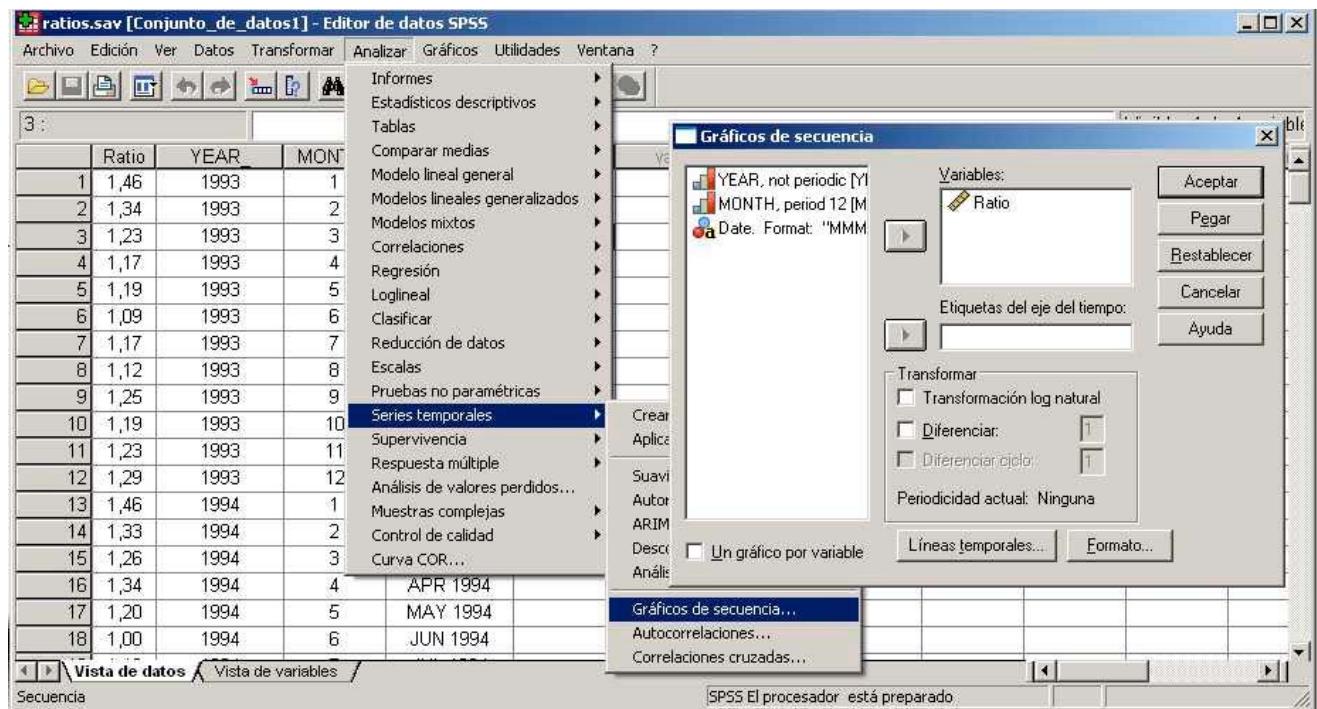
$$Z_t(l) = \phi_l d_t + \phi_{l+1} d_{t-1} + \phi_{l+2} d_{t-2} + \dots = Z_{t+l}$$

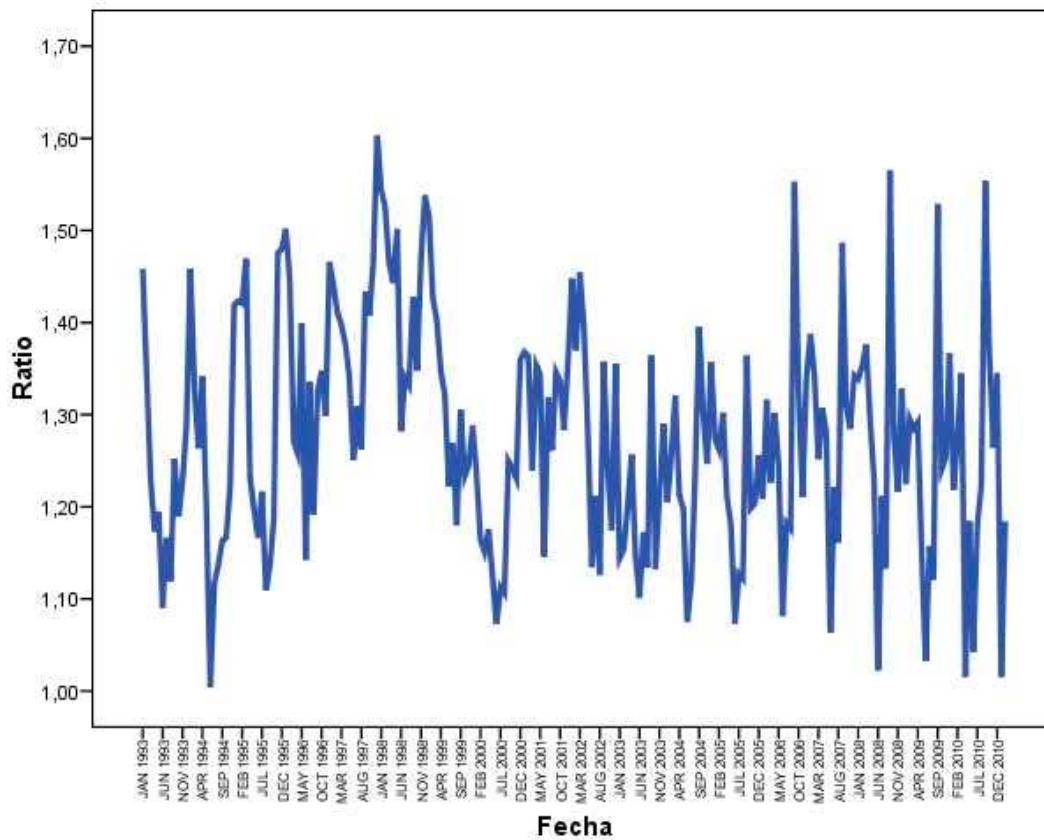
IDENTIFICACIÓN DE MODELOS ARIMA(p,d,q)(P,D,Q)_s

El archivo ***rations.sav*** muestra una serie de ratios mensuales de enero de 1993 a febrero de 2011.



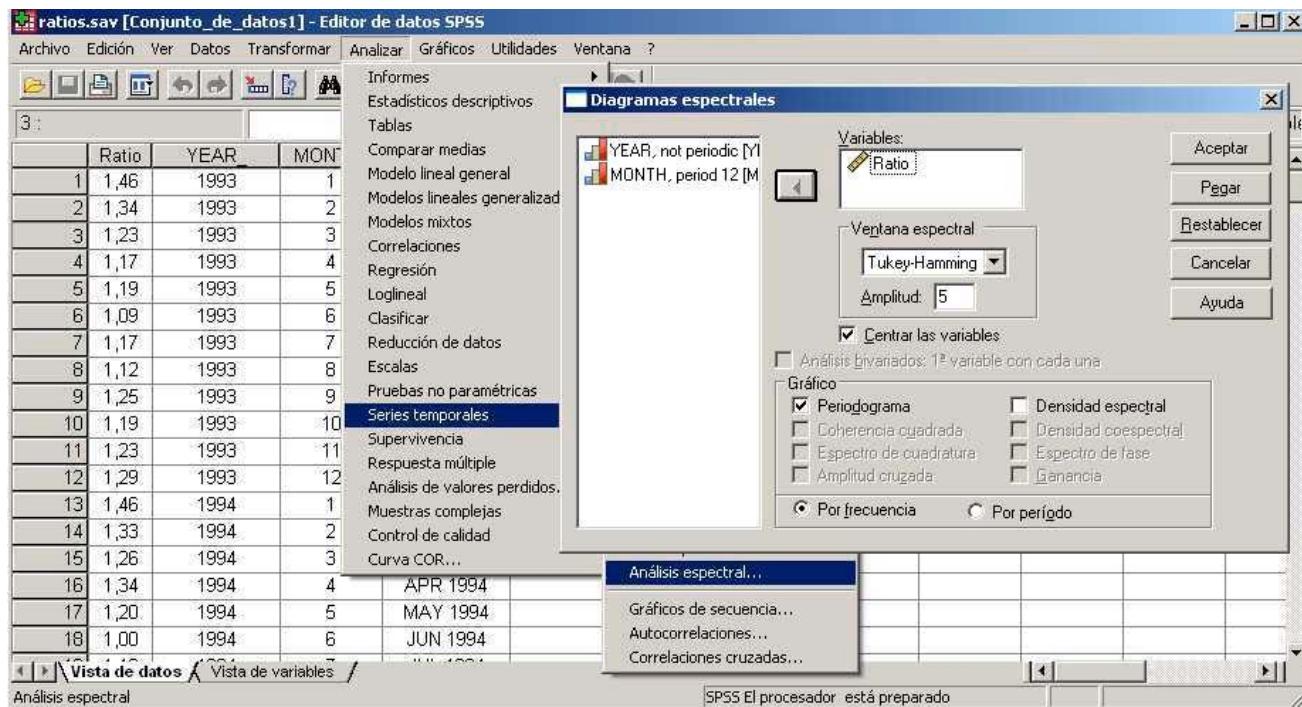
La fase de identificación comienza realizando una representación gráfica de la variable [Analizar/Series temporales/Gráficos de secuencia] con objeto de observar la estacionalidad.



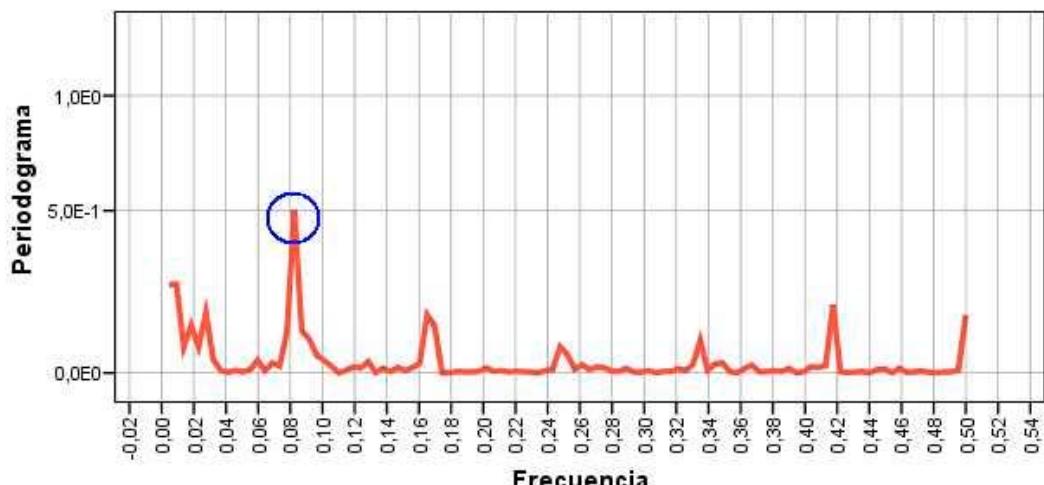


La serie muestra muchos picos, muchos de los cuales parecen estar espaciados uniformemente, sugiriendo la presencia de un componente periódico en la serie.

Para observar mejor la estacionalidad, se representa el PERIODGRAMA por frecuencia de la serie mediante [Analizar/Series temporales/Análisis espectral].

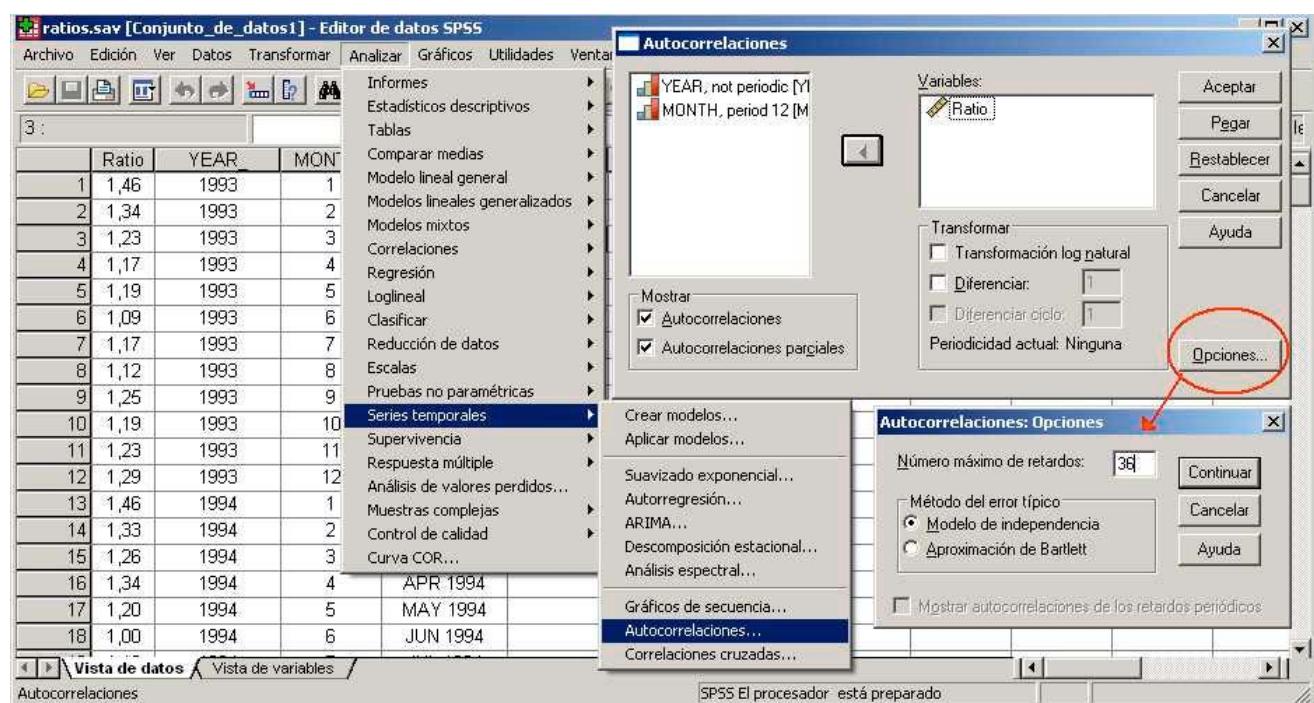


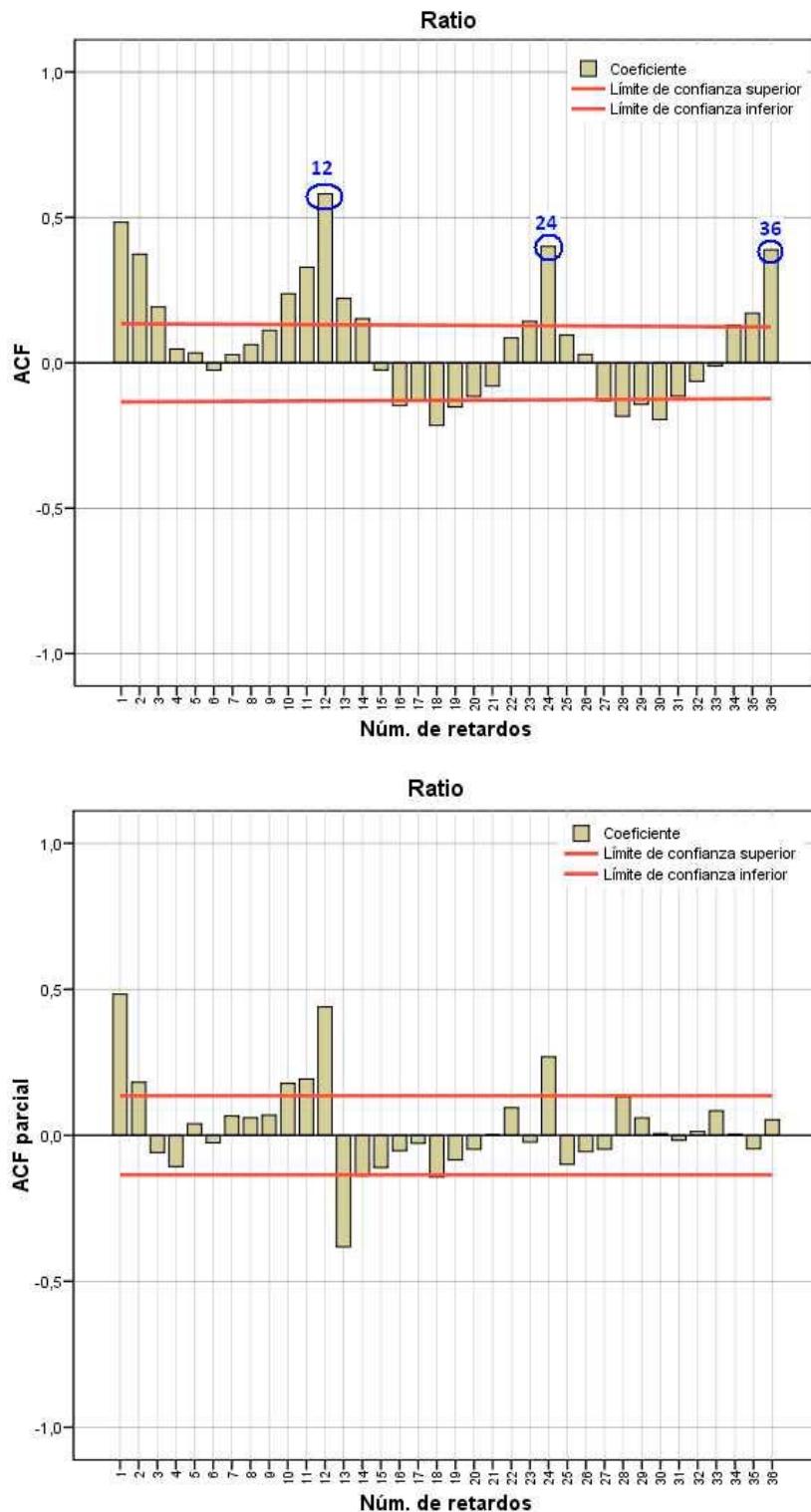
Periodograma de Ratio por frecuencia



El pico señalado corresponde a la frecuencia 0,08, es decir, la estación es $1/0,08 = 12$ meses.

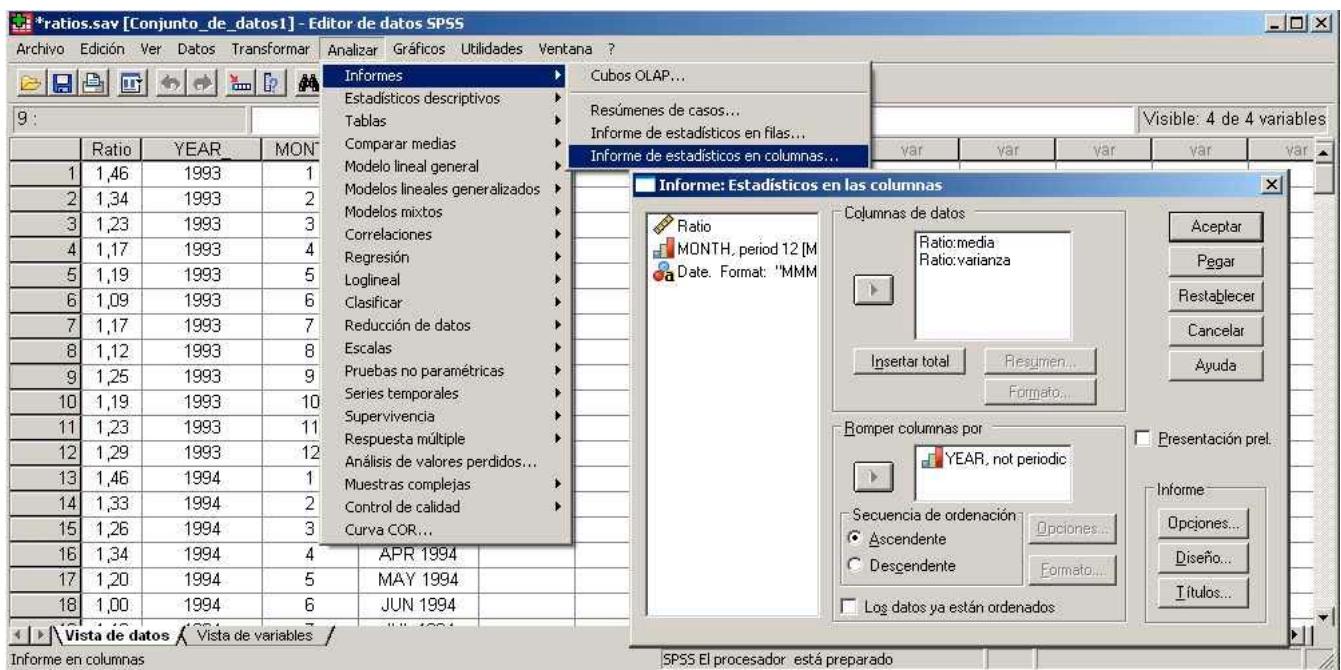
La estacionalidad, así como la estacionariedad, pueden detectarse también mediante las funciones de **autocorrelación** y **autocorrelación parcial estimadas (ACF y ACFP respectivamente)**. Para ello se elige **Analizar/Series temporales/Autocorrelaciones**. En el botón **Opciones** se escribe **36**, número máximo de retardos para representar a la función de autocorrelación **ACF** con un tramo significativo.





- Se observa que las funciones de autocorrelación (ACF) y autocorrelación parcial (ACFP) estimadas también validan los períodos estacionales porque los coeficientes de la ACF para retardos múltiplos del período estacional de la serie son significativamente distintos de cero.
- Además, para una cantidad grande de retardos la ACF 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. Por otro lado, la ACFP presenta estructura de coeficientes significativos para retardos periódicos largos. La ACF y ACFP deben considerarse a la vez, puesto que a veces intercambian sus papeles en el comportamiento estacional.
- Los coeficientes de la ACF no decaen rápidamente, indicando falta de estacionariedad en media.

Para certificar la falta de estacionariedad en varianza se recurre a **Analizar/Informes/Informes estadísticos en columnas**



Años	Media	Varianza
1993	1,23	0,01
1994	1,24	0,02
1995	1,29	0,02
1996	1,33	0,01
1997	1,39	0,01
1998	1,43	0,01
1999	1,31	0,01
2000	1,19	0,01
2001	1,31	0
2002	1,3	0,01
2003	1,19	0,01
2004	1,24	0,01
2005	1,21	0,01
2006	1,26	0,01
2007	1,29	0,01
2008	1,28	0,02
2009	1,25	0,02
2010	1,25	0,02
2011	1,1	0,01

Se obtiene una sucesión de medias y varianzas por años con variaciones significativas crecientes y decrecientes a lo largo de los años, indicando que **no hay estacionariedad** ni en media ni en varianza en la serie original.

*Para atenuar la falta de estacionariedad en media y en varianza se toman **logaritmos** y se **diferencia** la serie original.*

Para ello, **Analizar/Series temporales/Autocorrelaciones**

ratios.sav [Conjunto_de_datos1] - Editor de datos SPSS

Archivo Edición Ver Datos Transformar Analizar Gráficos Utilidades Ventana ?

1 : Ratio 1,45856

Informes
Estadísticos descriptivos
Tablas
Comparar medias
Modelo lineal general
Modelos lineales generalizados
Modelos mixtos
Correlaciones
Regresión
Loglineal
Clasificar
Reducción de datos
Escalas
Pruebas no paramétricas
Series temporales
Supervivencia
Respuesta múltiple
Análisis de valores perdidos...
Muestras complejas
Control de calidad
Curva COR...
Crear modelos...
Aplicar modelos...
Suavizado exponencial...
Autorregresión...
ARIMA...
Descomposición estacional...
Análisis espectral...
Gráficos de secuencia...
Autocorrelaciones...
Correlaciones cruzadas...

Vista de datos Vista de variables /

Autocorrelaciones SPSS El procesador está preparado

Aplicando logaritmos, como la serie es estacional, el problema consiste en identificar si se diferencia la parte regular o la parte estacional de la serie. Para ello se representan las funciones de autocorrelación estimada y autocorrelación parcial estimada bajo los supuestos de diferenciación en la parte regular o en la parte estacional (Diferenciar ciclo: 1).

ratios.sav [Conjunto_de_datos1] - Editor de datos SPSS

Archivo Edición Ver Datos Transformar Analizar Gráficos Utilidades Ventana ?

Autocorrelaciones

Variables: Ratio

Mostrar Autocorrelaciones Autocorrelaciones parciales

Aceptar Pegar Restablecer Cancelar Ayuda Opciones...

Transformar Transformación log natural Diferenciar: 1 **PARTE REGULAR** Diferenciar ciclo: 1 Periodicidad actual: 12

Autocorrelaciones

Variables: Ratio

Mostrar Autocorrelaciones Autocorrelaciones parciales

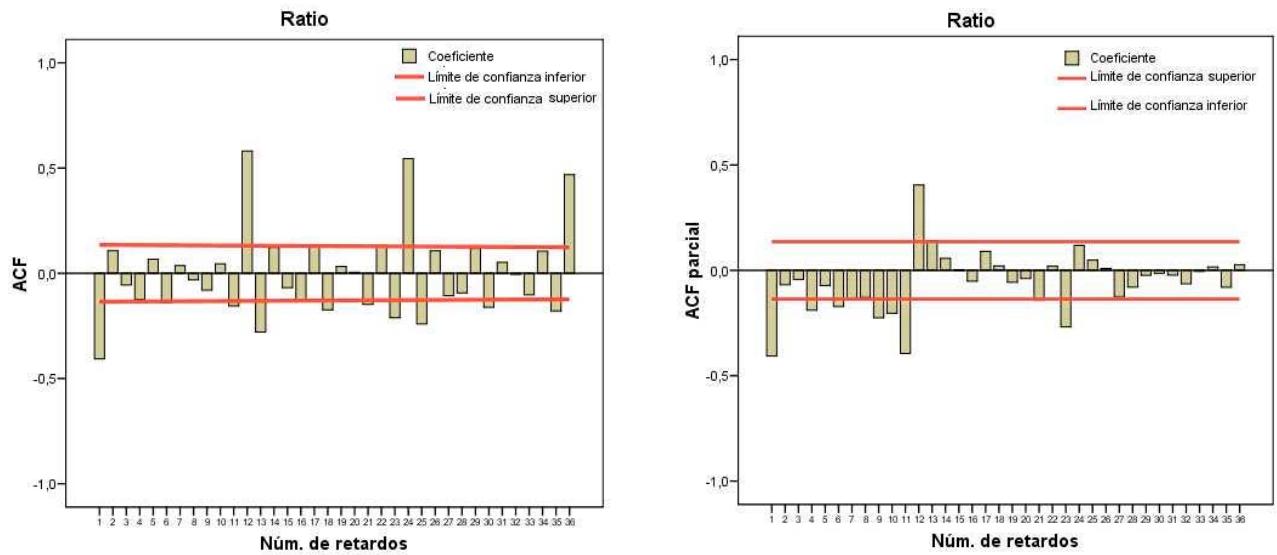
Aceptar Pegar Restablecer Cancelar Ayuda Opciones...

Transformar Transformación log natural Diferenciar: 1 **PARTE ESTACIONAL** Diferenciar ciclo: 1 Periodicidad actual: 12

Vista de datos Vista de variables /

Autocorrelaciones SPSS El procesador está preparado

DIFERENCIACIÓN PARTE REGULAR

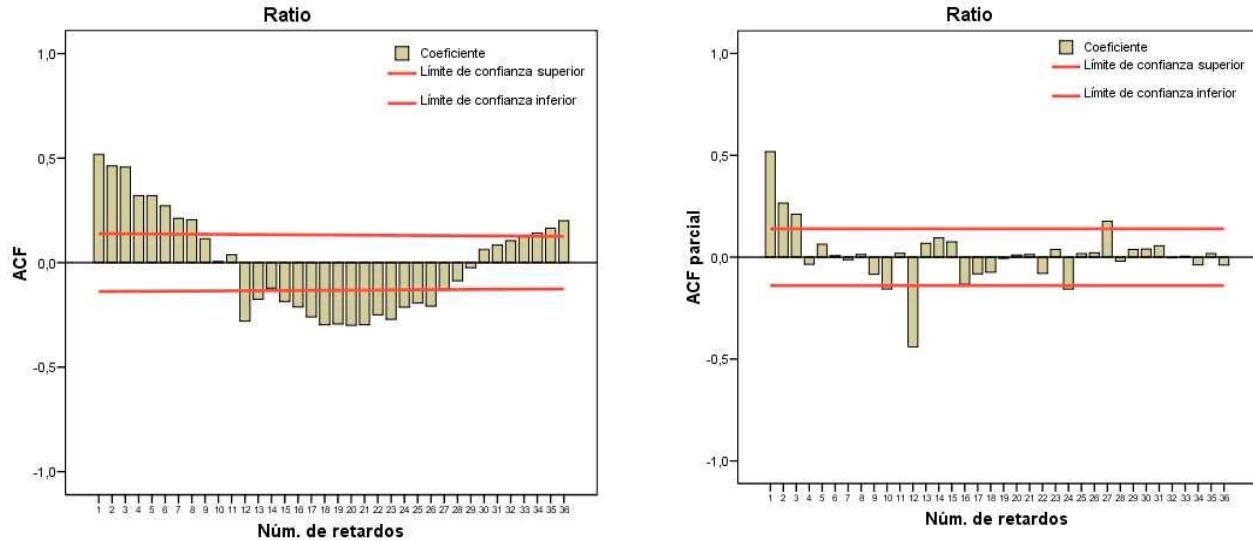


Al diferenciar solo la **parte regular**, las funciones de autocorrelación (ACF) y autocorrelación parcial estimadas (ACFP) **no superan** el problema de la falta de estacionariedad, pues la ACF no decae rápidamente.

Autocorrelaciones						Autocorrelaciones parciales					
Serie: Ratio						Serie: Ratio					
Retardo	Autocorrelación	Error típico ^a	Estadístico de Box-Ljung			Retardo	Autocorrelación parcial	Error típico			
			Valor	gl	Sig. ^b						
1	-.406	.067	36,343	1	,000	1	-,406	,068			
2	,108	.067	38,937	2	,000	2	-,068	,068			
3	-,056	.067	39,624	3	,000	3	-,044	,068			
4	-,124	.067	43,081	4	,000	4	-,189	,068			
5	,066	.067	44,072	5	,000	5	-,073	,068			
6	-,140	.067	48,514	6	,000	6	-,172	,068			
7	,036	.066	48,812	7	,000	7	-,138	,068			
8	-,031	.066	49,029	8	,000	8	-,128	,068			
9	-,080	.066	50,492	9	,000	9	-,225	,068			
10	,045	.066	50,953	10	,000	10	-,203	,068			
11	-,155	.066	56,491	11	,000	11	-,395	,068			
12	,581	.066	134,696	12	,000	12	,405	,068			
13	-,279	.066	152,834	13	,000	13	,140	,068			
14	,123	.065	156,381	14	,000	14	,057	,068			
15	-,069	.065	157,497	15	,000	15	,001	,068			
16	-,133	.065	161,673	16	,000	16	-,051	,068			
17	,130	.065	165,658	17	,000	17	,090	,068			
18	-,174	.065	172,868	18	,000	18	,021	,068			
19	,033	.065	173,130	19	,000	19	-,056	,068			
20	,005	.064	173,135	20	,000	20	-,038	,068			
21	-,148	.064	178,468	21	,000	21	-,142	,068			
22	,130	.064	182,588	22	,000	22	,019	,068			
23	-,212	.064	193,547	23	,000	23	-,268	,068			
24	,545	.064	266,628	24	,000	24	,118	,068			
25	-,241	.064	281,037	25	,000	25	,048	,068			
26	,107	.063	283,911	26	,000	26	,009	,068			
27	-,106	.063	286,742	27	,000	27	-,127	,068			
28	-,093	.063	288,923	28	,000	28	-,080	,068			
29	,118	.063	292,416	29	,000	29	-,023	,068			
30	-,162	.063	299,089	30	,000	30	-,015	,068			
31	,052	.063	299,781	31	,000	31	-,023	,068			
32	-,005	.062	299,788	32	,000	32	-,064	,068			
33	-,103	.062	302,522	33	,000	33	-,004	,068			
34	,105	.062	305,359	34	,000	34	,016	,068			
35	-,179	.062	313,700	35	,000	35	-,081	,068			
36	,469	.062	371,400	36	,000	36	,027	,068			

Al diferenciar solo una vez la **parte estacional**, las funciones de autocorrelación (ACF) y autocorrelación parcial estimadas (ACFP) superan el problema de la no estacionariedad.

PARTE ESTACIONAL



Serie: Ratio

Autocorrelaciones

Retardo	Autocorrelación	Error típico ^a	Estadístico de Box-Ljung		
			Valor	gl	Sig. ^b
1	,518	,069	56,046	1	,000
2	,463	,069	100,994	2	,000
3	,457	,069	145,061	3	,000
4	,321	,069	166,862	4	,000
5	,320	,068	188,687	5	,000
6	,272	,068	204,495	6	,000
7	,212	,068	214,203	7	,000
8	,205	,068	223,284	8	,000
9	,114	,068	226,108	9	,000
10	,006	,068	226,117	10	,000
11	,037	,067	226,424	11	,000
12	-,280	,067	243,754	12	,000
13	-,176	,067	250,608	13	,000
14	-,122	,067	253,936	14	,000
15	-,186	,067	261,660	15	,000
16	-,212	,067	271,820	16	,000
17	-,260	,066	287,139	17	,000
18	-,297	,066	307,221	18	,000
19	-,293	,066	326,931	19	,000
20	-,300	,066	347,656	20	,000
21	-,297	,066	368,111	21	,000
22	-,250	,066	382,615	22	,000
23	-,272	,065	399,877	23	,000
24	-,213	,065	410,538	24	,000
25	-,194	,065	419,413	25	,000
26	-,208	,065	429,718	26	,000
27	-,129	,065	433,685	27	,000
28	-,087	,064	435,489	28	,000
29	-,024	,064	435,630	29	,000
30	,063	,064	436,607	30	,000
31	,084	,064	438,349	31	,000
32	,105	,064	441,051	32	,000
33	,123	,064	444,808	33	,000
34	,141	,063	449,730	34	,000
35	,165	,063	456,514	35	,000
36	,201	,063	466,661	36	,000

Autocorrelaciones parciales

Retardo	Autocorrelación parcial	Error típico
1	,518	,070
2	,266	,070
3	,211	,070
4	-,035	,070
5	,063	,070
6	,007	,070
7	-,012	,070
8	,013	,070
9	-,084	,070
10	-,156	,070
11	,019	,070
12	-,441	,070
13	,068	,070
14	,094	,070
15	,076	,070
16	-,133	,070
17	-,082	,070
18	-,074	,070
19	-,006	,070
20	,010	,070
21	,014	,070
22	-,080	,070
23	,038	,070
24	-,157	,070
25	,017	,070
26	,022	,070
27	,176	,070
28	-,020	,070
29	,038	,070
30	,040	,070
31	,056	,070
32	-,003	,070
33	,004	,070
34	-,038	,070
35	,017	,070
36	-,039	,070

Las dos funciones (ACF, ACFP) cumplen las condiciones para que exista estacionalidad, dado que los coeficientes de las ACF para retardos múltiples del período estacional (12, 24, 36) de la serie son significativamente distintos de cero.

Además, para una cantidad de retardos igual al periodo estacional, la ACF se configura en forma de abanico que completa su ciclo girando sobre el eje de abscisas.

El problema de estacionalidad y estacionariedad en media y en varianza se resuelve tomando logaritmos, **diferenciando una vez la parte estacional y no diferenciando la parte regular**.

En consecuencia, 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)**.

Para identificar la parte autorregresiva **AR** y la parte de medias móviles **MA** se utiliza la ACF y ACFP con lo que se ha obtenido la estacionariedad y la estacionalidad.

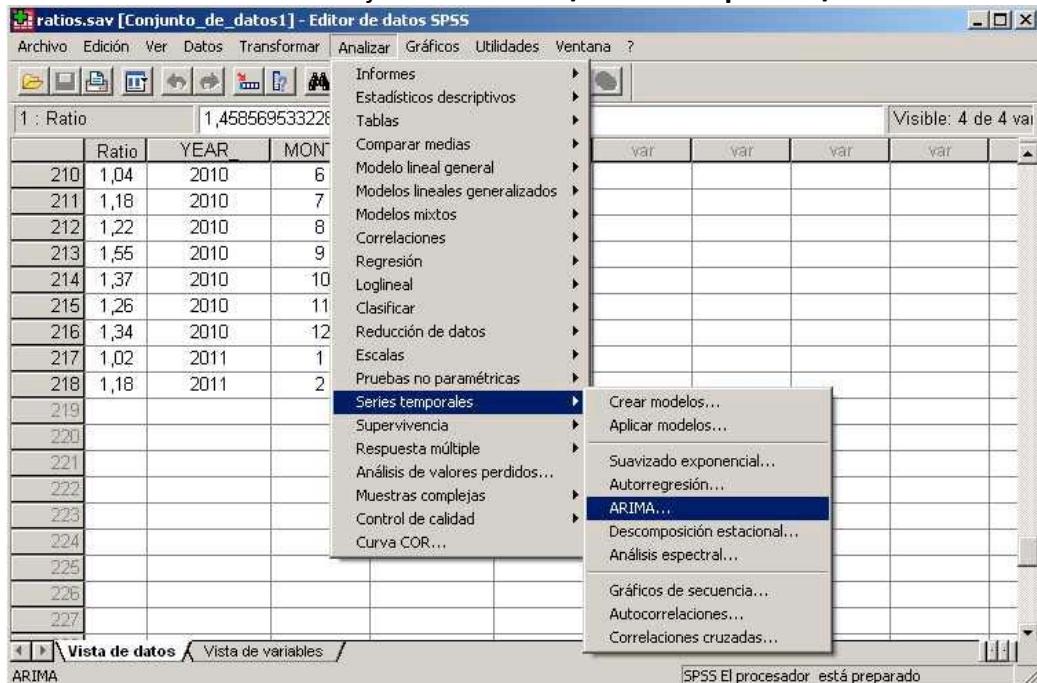
Observando estas dos funciones se distingue como sus coeficientes no se anulan bruscamente con periodicidades y que sus estructuras se ajustan a un modelo **ARMA(1,1)(0,1)₁₂**

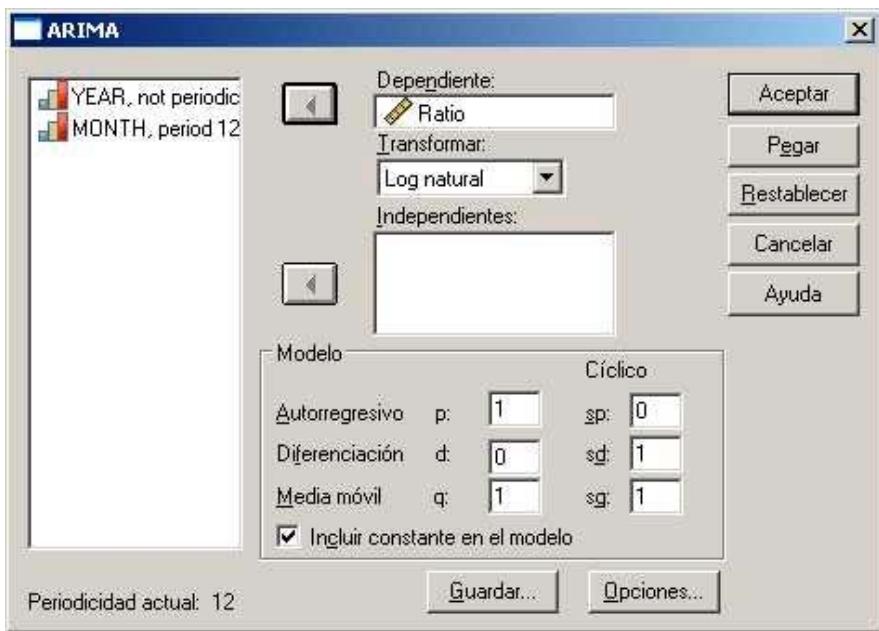
- La parte **AR(1) de la parte regular** proviene del decrecimiento rápido inicial y las ondas sinusoidales de la ACF añadido a que la ACFP presenta solo un coeficiente significativo en la mayoría de los periodos (salvo en el primero), anulándose bruscamente el resto de los coeficientes.
- La parte **MA(1) de la parte regular** proviene de que la ACF presenta un solo retardo significativo en la mayoría de los periodos (salvo en el primero).

La única duda posible sería considerar también **AR(1) la parte estacional**.

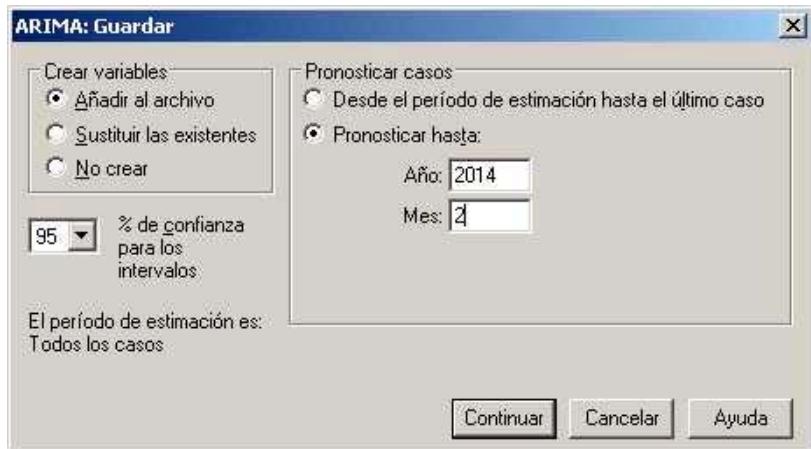
Identificada la serie inicial como un modelo **ARIMA(1,0,1)(0,1,1)₁₂** queda realizado el trabajo más importante en la modelización de la serie temporal mediante la metodología de Box-Jenkins.

Para estimar el modelo se ejecuta: **Analizar/Series temporales/ARIMA**

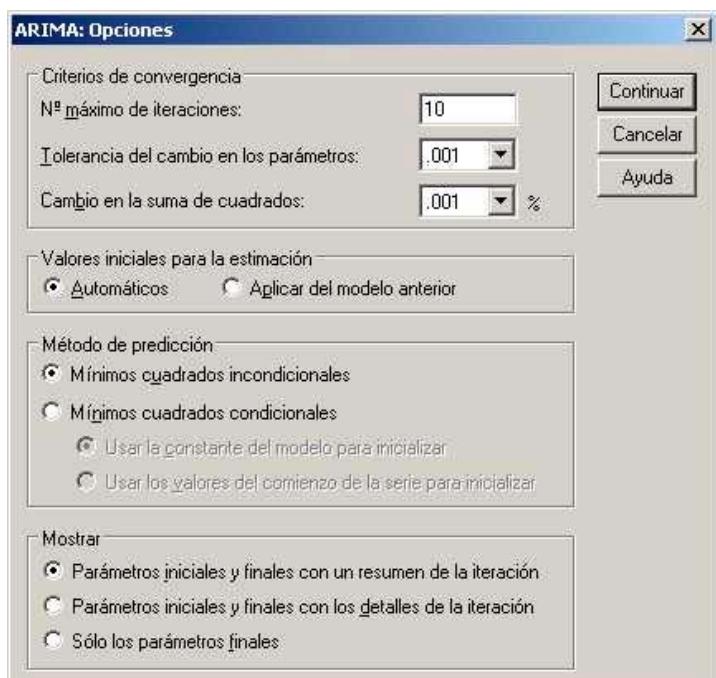




Desactivar <Incluir constante en el modelo> si no se desea incluir constante.



Se elige **Guardar** para crear nuevas variables que contengan valores pronosticados (**FIT_1**), residuos (**ERR_1**), intervalos de confianza (**LCL_1, UCL_1**), errores estándar para las predicciones (**SEP_1**). Todas las variables se añaden al Editor de datos como nuevas columnas.



En **Opciones** se seleccionan los criterios de convergencia, establecer valores iniciales para el modelo y elegir cómo mostrar los parámetros en los resultados.

ratiostest.sav [Conjunto_de_datos1] - Editor de datos SPSS

Archivo Edición Ver Datos Transformar Analizar Gráficos Utilidades Ventana ?

1 : Ratio 1,4585695332288 Visible: 9 de 9 variables:

	Ratio	YEAR	MONTH	DATE	FIT_1	ERR_1	LCL_1	UCL_1	SEP_1
1	1,46	1993	1	JAN 1993					
2	1,34	1993	2	FEB 1993					
3	1,23	1993	3	MAR 1993					
4	1,17	1993	4	APR 1993					
5	1,19	1993	5	MAY 1993					
6	1,09	1993	6	JUN 1993					
7	1,17	1993	7	JUL 1993					
8	1,12	1993	8	AUG 1993					
9	1,25	1993	9	SEP 1993					
10	1,19	1993	10	OCT 1993					
11	1,23	1993	11	NOV 1993					
12	1,29	1993	12	DEC 1993					
13	1,46	1994	1	JAN 1994	1,45519	,00197	1,23878	1,70940	,11883
14	1,33	1994	2	FEB 1994	1,33988	-,00759	1,16768	1,53748	,09348
15	1,26	1994	3	MAR 1994	1,22308	,03250	1,07023	1,39776	,08281
16	1,34	1994	4	APR 1994	1,18249	,12632	1,03555	1,35028	,07957
17	1,20	1994	5	MAY 1994	1,25830	-,04773	1,10206	1,43668	,08460
18	1,00	1994	6	JUN 1994	1,12208	-,11065	,98275	1,28115	,07545
19	1,12	1994	7	JUL 1994	1,14846	-,02802	1,00589	1,31124	,07720
20	1,14	1994	8	AUG 1994	1,09189	,04026	,95642	1,24654	,07335
21	1,16	1994	9	SEP 1994	1,24166	-,06536	1,08777	1,41732	,08332
22	1,17	1994	10	OCT 1994	1,15247	,01228	1,00985	1,31524	,07722
				NOV 1994	1,18840	,02426	1,05840	1,32722	,08244

Vista de datos Vista de variables SPSS El procesador está preparado

Al pulsar **Aceptar** se obtiene el ajuste a un modelo de **Box-Jenkins ARIMA(1,0,1)(0,1,1)₁₂**:

Criterios de finalización de las iteraciones

Cambio máximo en los parámetros menor que	,001
Máxima constante de Marquardt mayor que	1E+009
Cambio en el porcentaje de la suma de cuadrados menor que	,001%
Número de iteraciones igual a	10

Diagnóstico residual

Número de residuos	206
Número de parámetros	3
GL residuales	202
Suma de cuadrados residual corregida	,681
Suma de cuadrados residual	,898
Varianza residual	,003
Error típico del modelo	,057
Log-verosimilitud	296,004
Criterio de información de Akaike (AIC)	-584,009
Criterio bayesiano de Schwarz (BIC)	-570,697

Estimaciones de los parámetros

		Estimaciones	Error típico	t	Sig. aprox.
Retardos no estacionales	AR1	,903	,043	21,007	,000
	MA1	,529	,083	6,343	,000
Retardos estacionales	Seasonal MA1	,630	,069	9,190	,000
Constante		-,002	,008	-,304	,762

Se ha utilizado el algoritmo de Melard para la estimación.

Matriz de covarianza

		Retardos no estacionales		Retardos estacionales	Constante
		AR1	MA1	Seasonal MA1	
Retardos no estacionales	AR1	,002	,002	,000	0 ^a
	MA1	,002	,007	,000	0 ^a
Retardos estacionales	Seasonal MA1	,000	,000	,005	0 ^a
Constante		0 ^a	0 ^a	0 ^a	,000

Se ha utilizado el algoritmo de Melard para la estimación.

- a. La estimación del parámetro ARMA y la estimación de los parámetros de la regresión no están correlacionados asintóticamente.

Matriz de correlaciones

		Retardos no estacionales		Retardos estacionales	Constante
		AR1	MA1	Seasonal MA1	
Retardos no estacionales	AR1	1,000	,693	,118	0 ^a
	MA1	,693	1,000	-,009	0 ^a
Retardos estacionales	Seasonal MA1	,118	-,009	1,000	0 ^a
Constante		0 ^a	0 ^a	0 ^a	1,000

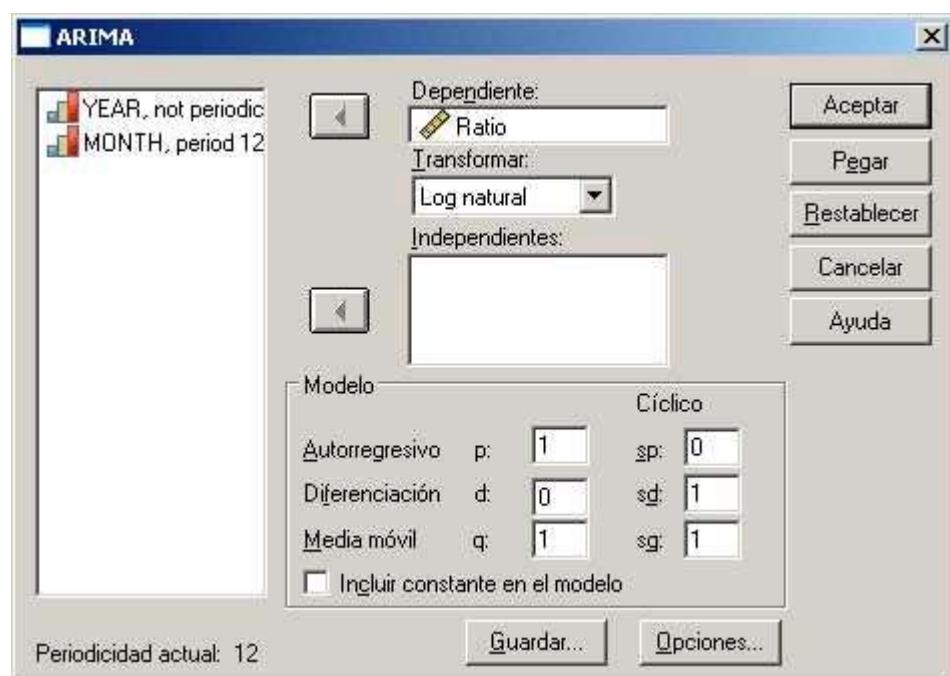
Se ha utilizado el algoritmo de Melard para la estimación.

- a. La estimación del parámetro ARMA y la estimación de los parámetros de la regresión no están correlacionados asintóticamente.

Matriz covarianzas:
$$\begin{pmatrix} & \text{AR1} & \text{MA1} & \text{SMA1} \\ \text{AR1} & 0,02 & 0,02 & 0 \\ \text{MA1} & 0,02 & 0,07 & 0 \\ \text{SMA1} & 0 & 0 & 0,05 \end{pmatrix}$$

Matriz correlaciones:
$$\begin{pmatrix} & \text{AR1} & \text{MA1} & \text{SMA1} \\ \text{AR1} & 1 & 0,693 & 0,118 \\ \text{MA1} & 0,693 & 1 & -0,09 \\ \text{SMA1} & 0,118 & -0,09 & 1 \end{pmatrix}$$

Excepto la constante, el resto de los parámetros son significativos, lo que conduce a estimar el modelo sin constante.



El nuevo ajuste sin contraste:

Historial de iteraciones

	Retardos no estacionales		Retardos estacionales Estacional MM1	Suma de cuadrados corregida	Constante de Marquardt
	AR1	MA1			
0	,893	,241	,304	,830	,001
1	,870	,447	,628	,685	,001
2	,908	,537	,634	,682	,000
3	,904	,530	,628	,682 ^a	,000

Se ha utilizado el algoritmo de Melard para la estimación.

- a. La estimación finalizó en esta iteración, ya que la suma de los cuadrados disminuyó en menos de ,001%.

Diagnóstico residual

Número de residuos	206
Número de parámetros	3
GL residuales	203
Suma de cuadrados residual corregida	,682
Suma de cuadrados residual	,830
Varianza residual	,003
Error típico del modelo	,057
Log-verosimilitud	295,950
Criterio de información de Akaike (AIC)	-585,900
Criterio bayesiano de Schwarz (BIC)	-575,916

Estimaciones de los parámetros

		Estimaciones	Error típico	t	Sig. aprox.
Retardos no estacionales	AR1	,905	,043	21,254	,000
	MA1	,531	,083	6,401	,000
Retardos estacionales	Seasonal MA1	,632	,068	9,256	,000

Se ha utilizado el algoritmo de Melard para la estimación.

El nuevo ajuste es bueno con una significación muy alta de sus parámetros (p-valor nulos para sus parámetros).

Matriz de covarianza

		Retardos no estacionales		Retardos estacionales
		AR1	MA1	Seasonal MA1
Retardos no estacionales	AR1	,002	,002	,000
	MA1	,002	,007	,000
Retardos estacionales	Seasonal MA1	,000	,000	,005

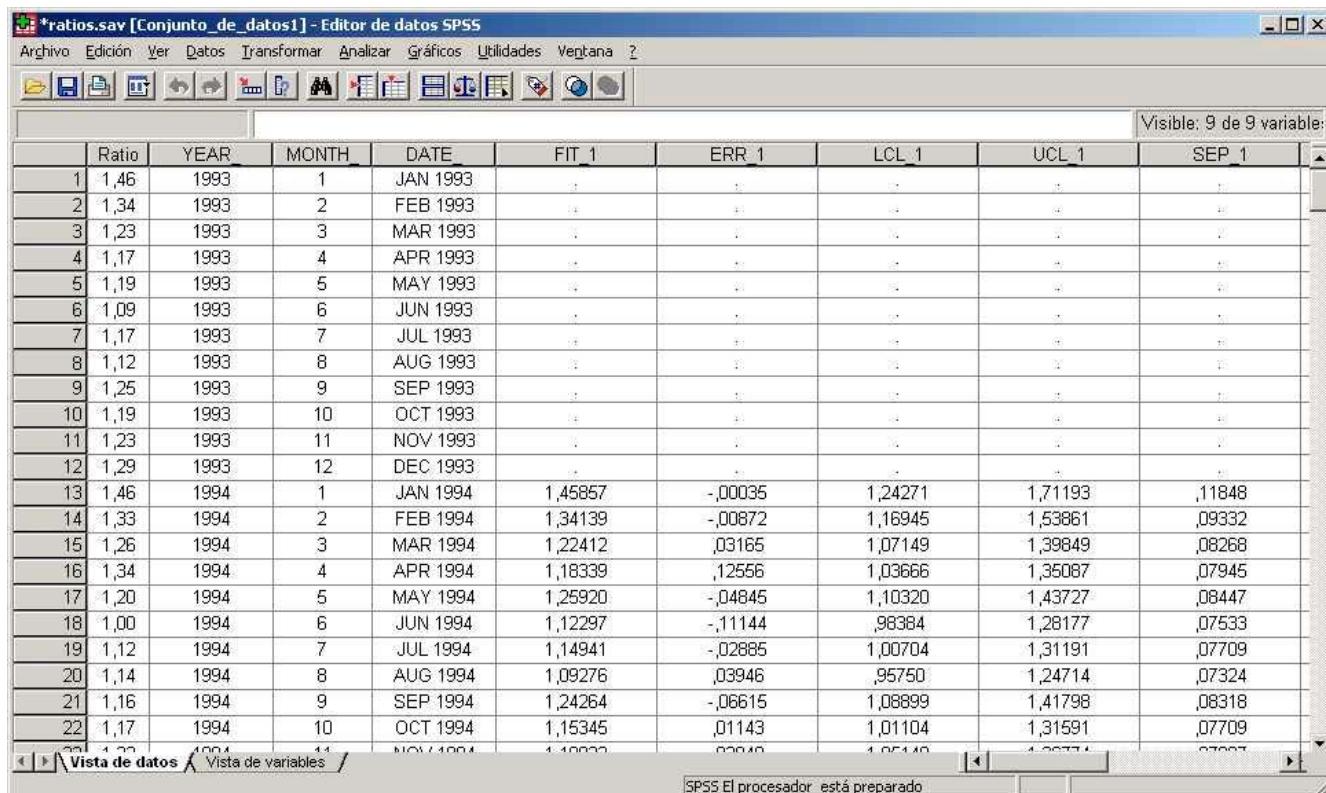
Se ha utilizado el algoritmo de Melard para la estimación.

Matriz de correlaciones

		Retardos no estacionales		Retardos estacionales
		AR1	MA1	Seasonal MA1
Retardos no estacionales	AR1	1,000	,690	,117
	MA1	,690	1,000	-,010
Retardos estacionales	Seasonal MA1	,117	-,010	1,000

Se ha utilizado el algoritmo de Melard para la estimación.

En el *Editor de datos*:



La variable **FIT_1** ha generado las predicciones hasta febrero de 2014, la variable **ERR_1** ha generado las estimaciones del término de error del modelo, las variables **LCL_1**, **UCL_1** han generado límites inferiores y superiores de los intervalos de confianza al 95% de fiabilidad, la variable **SEP_1** contiene los errores estándar para las predicciones.

Para obtener la representación de la serie original y la serie de predicciones **FIT_1** se recurre a **Analizar/Series temporales/Gráfico de secuencia**

*ratios.sav [Conjunto_de_datos1] - Editor de datos SPSS

Archivo Edición Ver Datos Transformar Analizar Gráficos Utilidades Ventana ?

8 :

	Ratio	YEAR	MON
1	1,46	1993	1
2	1,34	1993	2
3	1,23	1993	3
4	1,17	1993	4
5	1,19	1993	5
6	1,09	1993	6
7	1,17	1993	7
8	1,12	1993	8
9	1,25	1993	9
10	1,19	1993	10
11	1,23	1993	11
12	1,29	1993	12
13	1,46	1994	1
14	1,33	1994	2
15	1,26	1994	3
16	1,34	1994	4
17	1,20	1994	5
18	1,00	1994	6
19	1,12	1994	7
20	1,14	1994	8
21	1,16	1994	9
22	1,17	1994	10
		NOV 1994	1,499999

Vista de datos Vista de variables Secuencia

SPSS El procesador está preparado

Analizar Series temporales Gráficos de secuencia

Visible: 9 de 9 variables

ERR_1 LCL_1 UCL_1 SEP_1 var

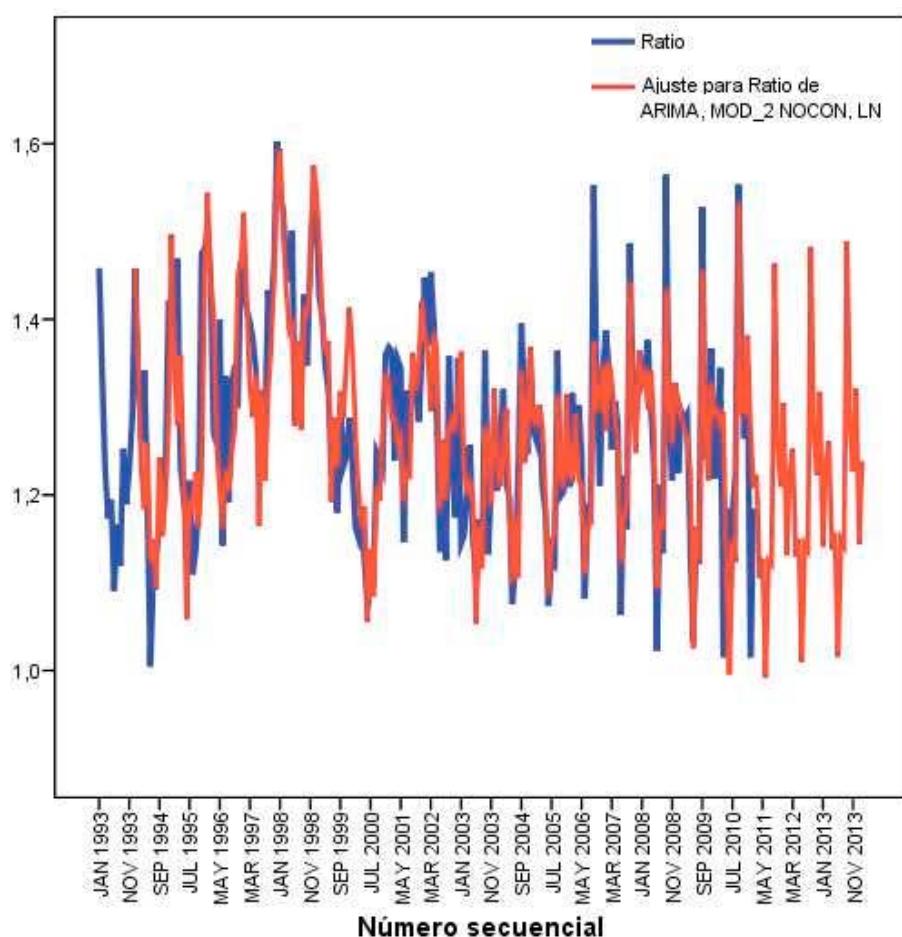
Variables: Ratio Ajuste para Ratio de AF

Aceptar Pegar Restablecer Cancelar Ayuda

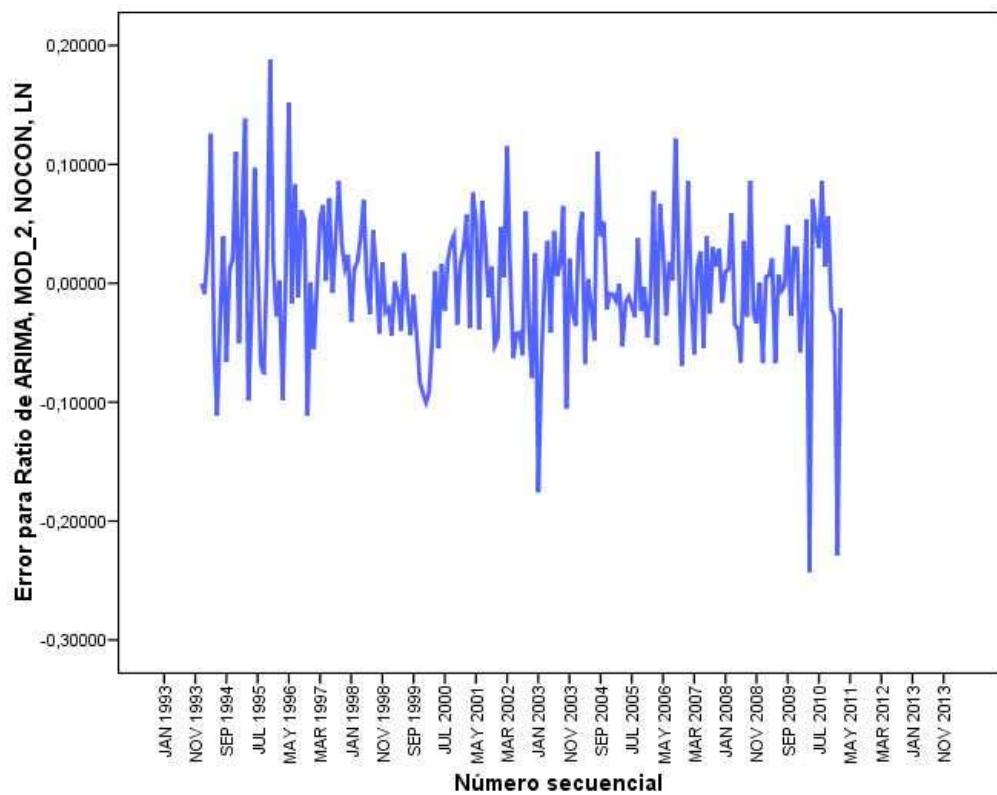
Etiquetas del eje del tiempo:

Transformar Transformación log natural Diferenciar Diferenciar ciclo Periodicidad actual: Ninguna

Un gráfico por variable Líneas temporales... Formato...



Análogamente, con la instrucción **Analizar/Series temporales/Gráfico de secuencia** se obtiene la representación de los errores del modelo estimado, que presenta una estructura aleatoria, hecho favorable como verificación del diagnóstico de la modelización ARIMA realizada.



Fichero: *rations.sav*

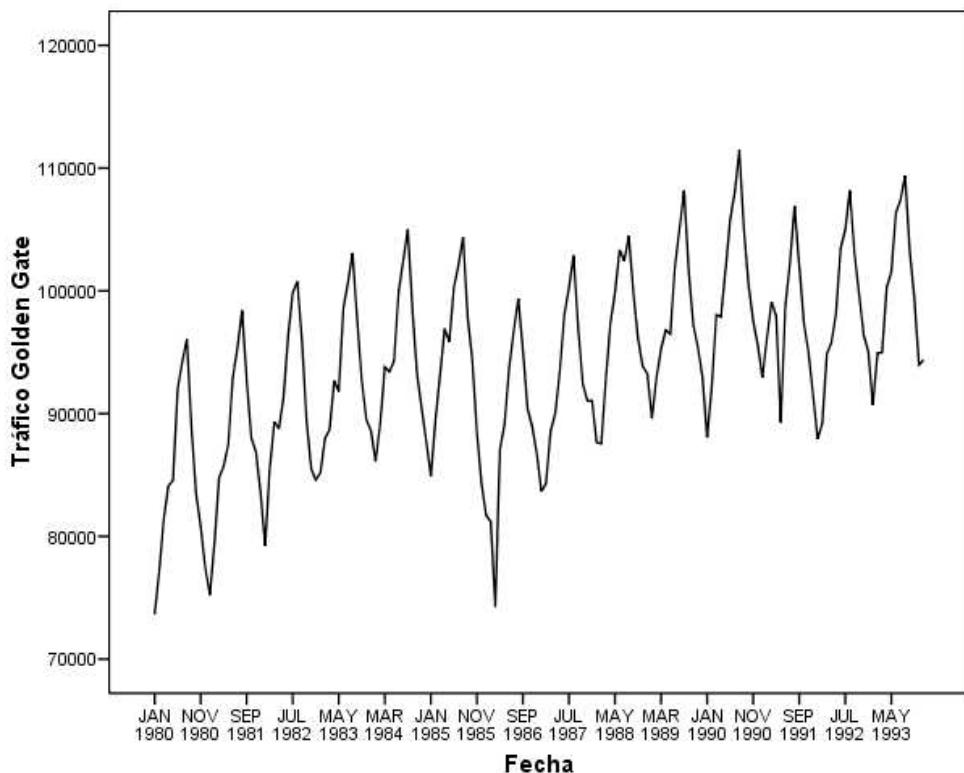
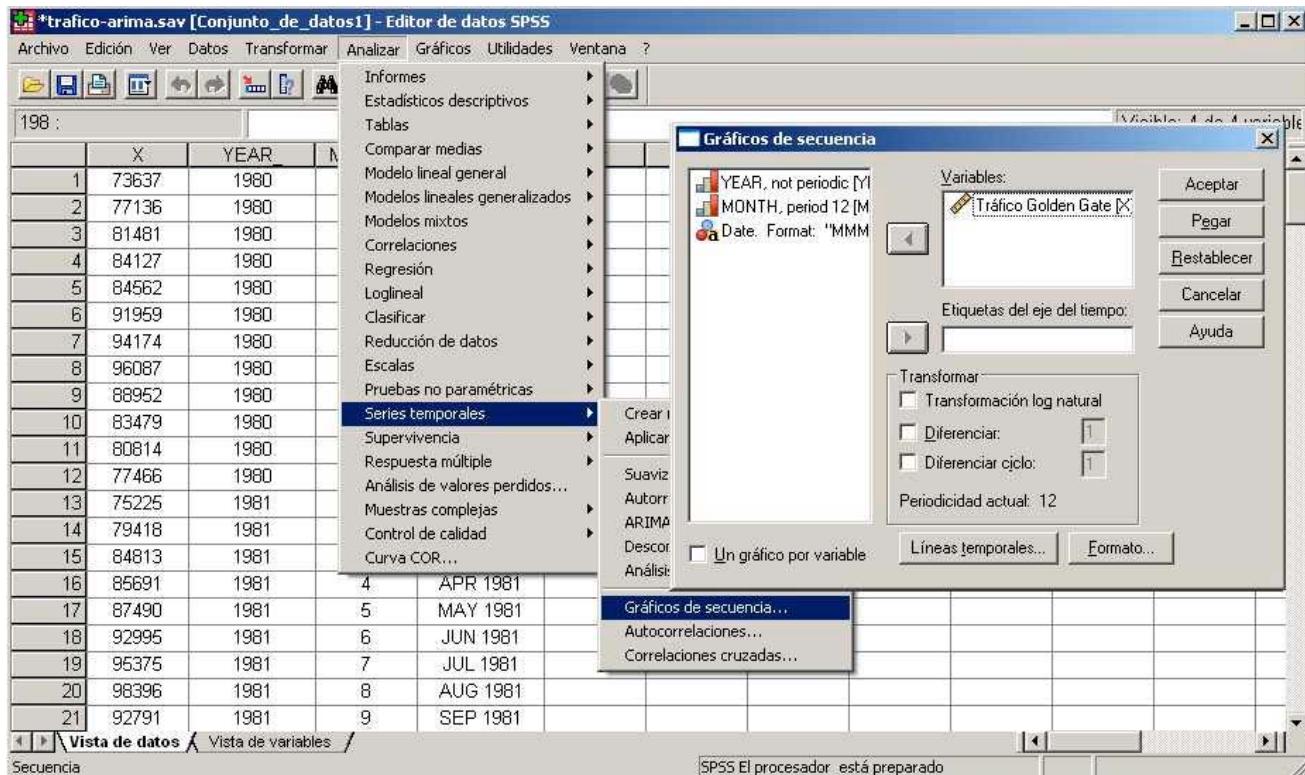
	1993	1994	1995	1996	1997	1998	1999
Enero	1,46	1,46	1,42	1,5	1,44	1,54	1,52
Febrero	1,34	1,33	1,42	1,45	1,41	1,52	1,43
Marzo	1,23	1,26	1,47	1,27	1,4	1,46	1,4
Abril	1,17	1,34	1,23	1,26	1,38	1,44	1,35
Mayo	1,19	1,2	1,2	1,4	1,34	1,5	1,32
Junio	1,09	1	1,17	1,14	1,25	1,28	1,22
Julio	1,17	1,12	1,22	1,34	1,31	1,34	1,27
Agosto	1,12	1,14	1,11	1,19	1,26	1,33	1,18
Septiembre	1,25	1,16	1,14	1,33	1,43	1,43	1,31
Octubre	1,19	1,17	1,19	1,35	1,41	1,35	1,23
Noviembre	1,23	1,22	1,48	1,3	1,47	1,47	1,25
Diciembre	1,29	1,42	1,48	1,47	1,6	1,54	1,29

	2000	2001	2002	2003	2004	2005	2006
Enero	1,23	1,37	1,45	1,14	1,2	1,27	1,21
Febrero	1,16	1,36	1,37	1,15	1,26	1,26	1,32
Marzo	1,15	1,24	1,45	1,2	1,32	1,3	1,23
Abril	1,18	1,35	1,4	1,26	1,21	1,21	1,3
Mayo	1,12	1,34	1,28	1,14	1,2	1,18	1,25
Junio	1,07	1,15	1,13	1,1	1,08	1,07	1,08
Julio	1,11	1,32	1,21	1,17	1,12	1,13	1,18
Agosto	1,11	1,26	1,13	1,13	1,24	1,12	1,18
Septiembre	1,25	1,35	1,36	1,36	1,4	1,36	1,55
Octubre	1,24	1,34	1,24	1,13	1,3	1,2	1,33
Noviembre	1,23	1,28	1,17	1,21	1,25	1,2	1,21
Diciembre	1,36	1,36	1,36	1,29	1,36	1,26	1,34

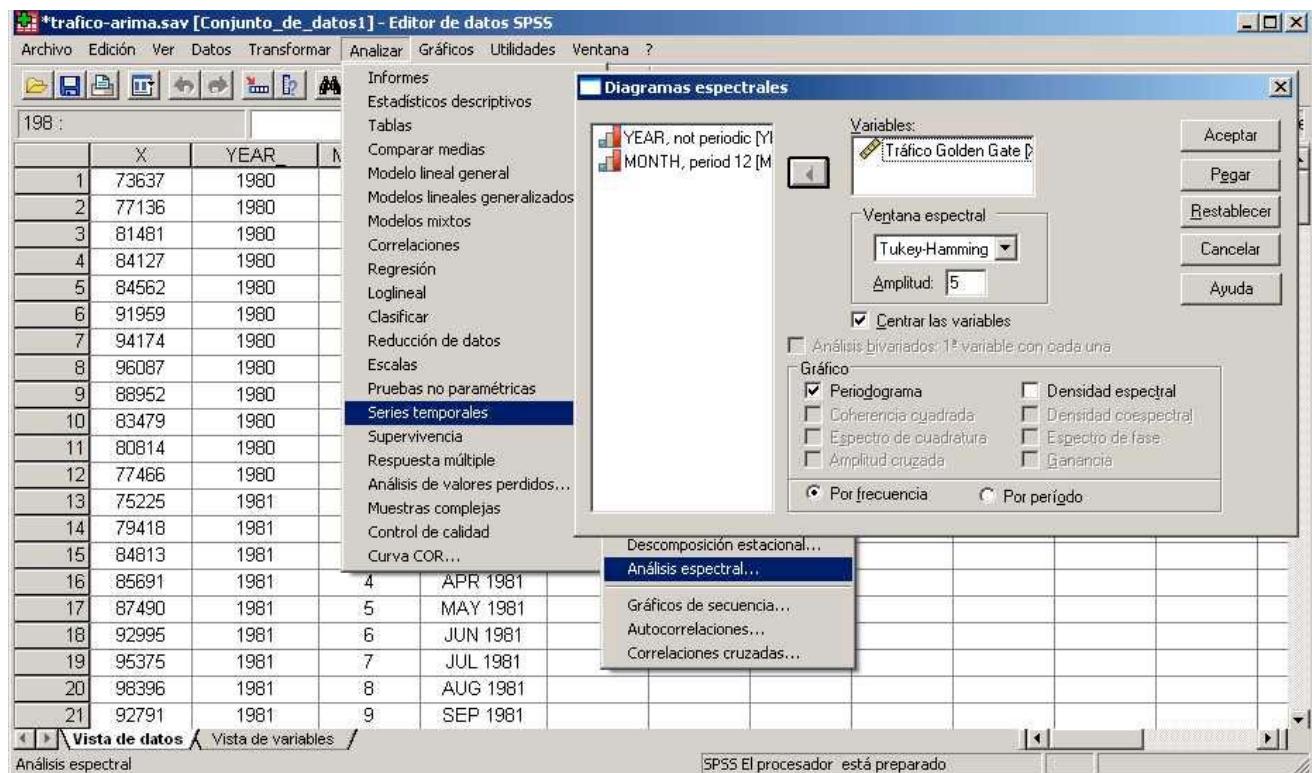
	2007	2008	2009	2010	2011
Enero	1,39	1,34	1,23	1,22	1,02
Febrero	1,34	1,35	1,29	1,28	1,18
Marzo	1,25	1,38	1,28	1,34	
Abril	1,31	1,3	1,29	1,02	
Mayo	1,28	1,23	1,15	1,18	
Junio	1,06	1,02	1,03	1,04	
Julio	1,22	1,21	1,16	1,18	
Agosto	1,16	1,13	1,12	1,22	
Septiembre	1,49	1,56	1,53	1,55	
Octubre	1,31	1,28	1,24	1,37	
Noviembre	1,28	1,22	1,25	1,26	
Diciembre	1,34	1,33	1,37	1,34	

EJERCICIO 2.- El archivo **tráfico-ariama.sav** muestra los datos relativos al tráfico mensual en el puente Golden Gate de enero de 1980 a diciembre de 1993.

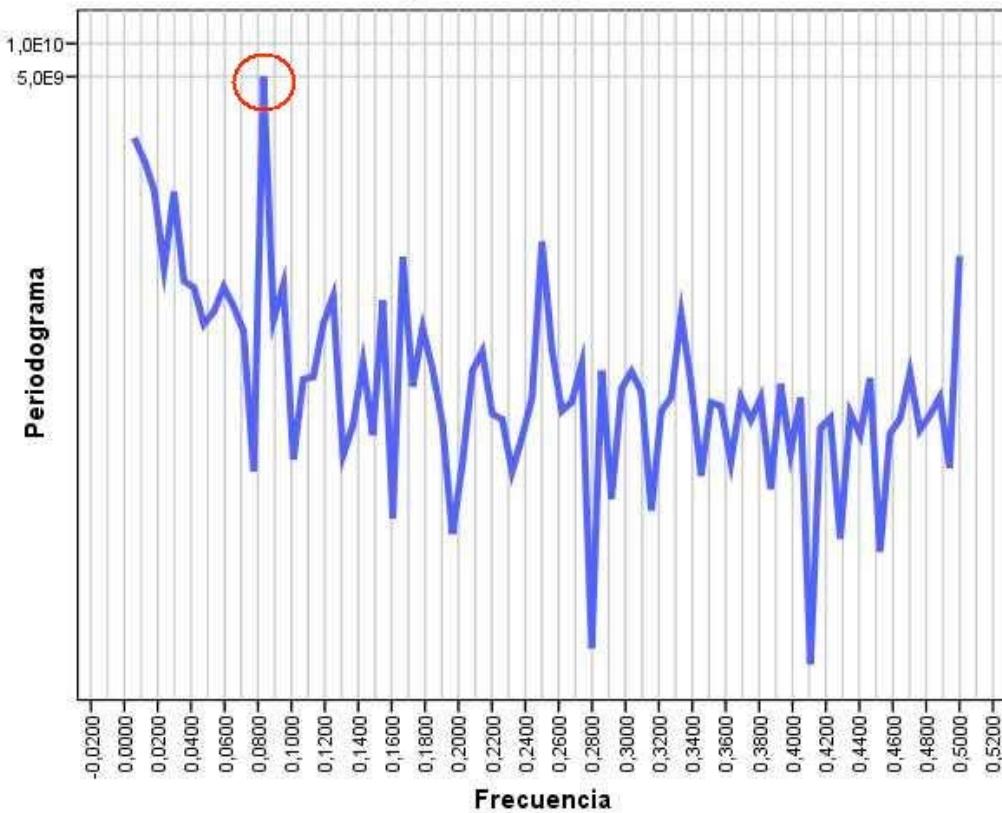
La fase de identificación comienza realizando una representación gráfica de la variable [Analizar/Series temporales/Gráficos de secuencia] con objeto de observar la estacionalidad.



Para observar mejor la estacionalidad, se representa el **PERIODOGRAMA** por frecuencia de la serie mediante [Analizar/Series temporales/Análisis espectral].

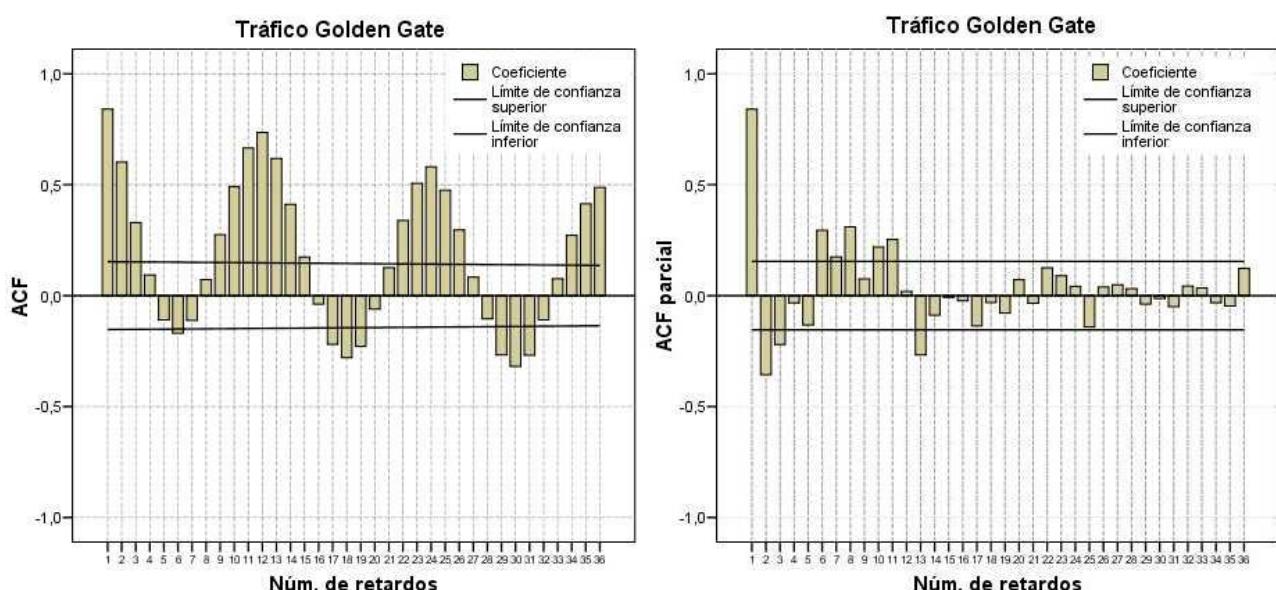
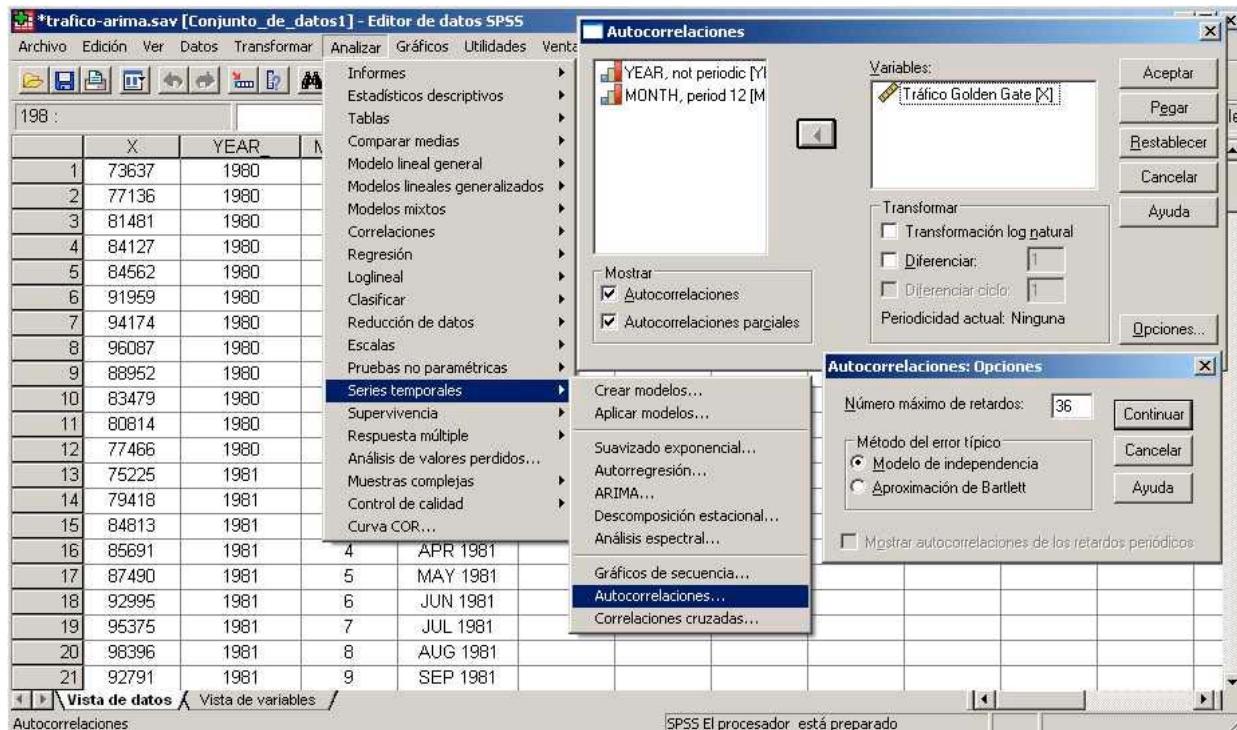


Periodograma de X por frecuencia



El pico señalado corresponde a la frecuencia 0,08, es decir, la estación es $1/0,08 = 12$ meses.

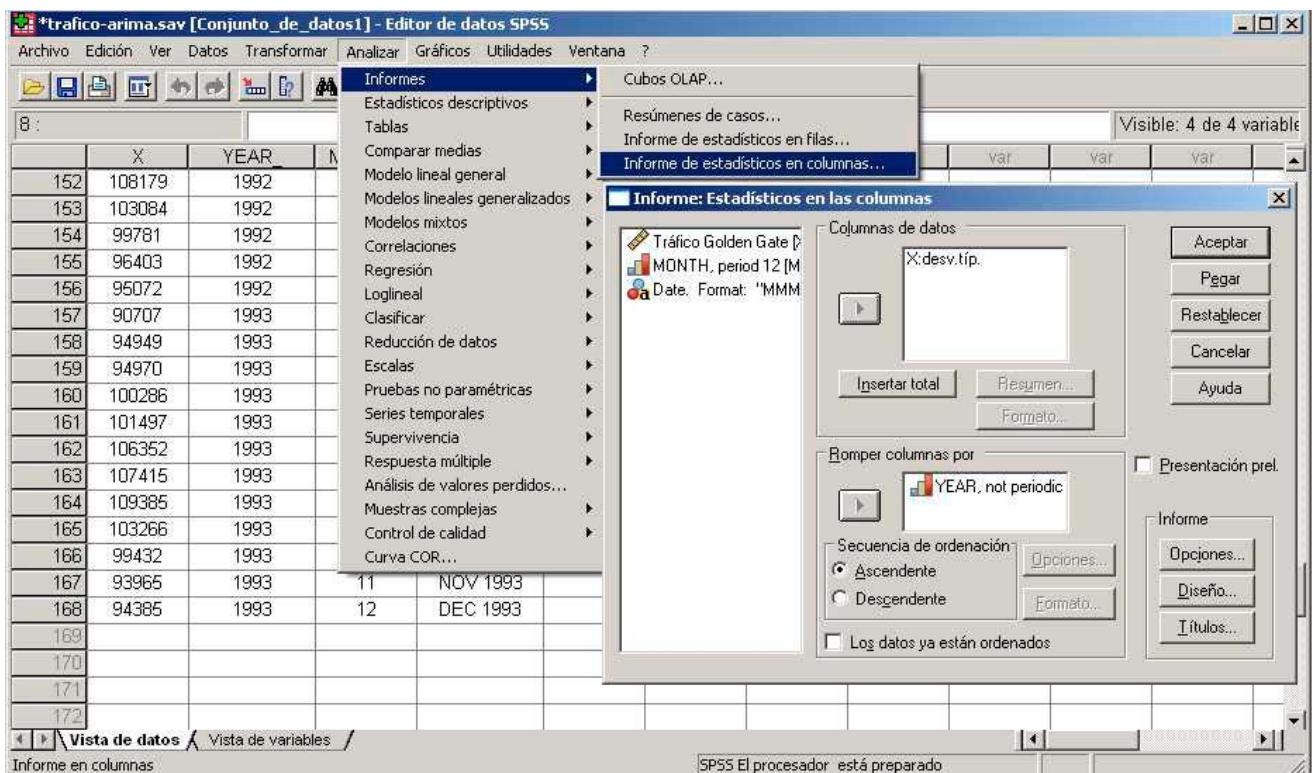
La estacionalidad, así como la estacionariedad, pueden detectarse también mediante las funciones de autocorrelación y autocorrelación parcial estimadas (ACF y ACFP respectivamente). Para ello se elige **Analizar/Series temporales/Autocorrelaciones**. En el botón *Opciones* se escribe 36, número máximo de retardos para representar a la función de autocorrelación ACF con un tramo significativo.



- Se observa que las funciones de autocorrelación (ACF) y autocorrelación parcial (ACFP) estimadas también validan los períodos estacionales porque los coeficientes de la ACF para retardos múltiples del período estacional de la serie son significativamente distintos de cero.
- Además, para una cantidad grande de retardos la ACF 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. Por otro lado, la ACF presenta estructura de coeficientes significativos para retardos periódicos largos. La ACF y ACFP deben considerarse a la vez, puesto que a veces intercambian sus papeles en el comportamiento estacional.

- Los coeficientes de la ACF no decaen rápidamente, indicando falta de estacionariedad en media. Para certificar la falta de estacionariedad en varianza se recurre a **Analizar/Informes/Informes estadísticos en columnas**



Con la finalidad de calcular varianzas por estaciones (años), obteniendo 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.

Con el objetivo de atenuar la falta de estacionariedad en media y en varianza se toman logaritmos y se diferencia la serie original. Para ello, **Analizar/Series temporales/Autocorrelaciones**

*tráfico-arima.sav [Conjunto_de_datos1] - Editor de datos SPSS

Archivo Edición Ver Datos Transformar Analizar Gráficos Utilidades Ventana ?

1 :

	X	YEAR	M
1	73637	1980	
2	77136	1980	
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17	87490	1981	5
18	92995	1981	6
19	95375	1981	7
20	98396	1981	8
21	92791	1981	9
22	88018	1981	10
23	86899	1981	11
24	83636	1981	12
25	79245	1982	1

Vista de datos Vista de variables /

Autocorrelaciones SPSS El procesador está preparado

Autocorrelaciones

Variables: Tráfico Golden Gate [X]

Transformar:

- Transformación log natural
- Diferenciar: 1 **PARTE REGULAR**
- Diferenciar ciclo: 1

Mostrar:

- Autocorrelaciones
- Autocorrelaciones parciales

Periodicidad actual: 12

Autocorrelaciones: Opciones

Número máximo de retardos: 38

Método del error típico:

- Modelo de independencia
- Aproximación de Bartlett

Mostrar autocorrelaciones de los retardos periódicos

Aceptar Pegar Restablecer Cancelar Ayuda

Autocorrelaciones

Visible: 4 díbles

YEAR, not periodic [Y]
MONTH, period 12 [M]

Variables: Tráfico Golden Gate [X]

Transformar:

- Transformación log natural
- Diferenciar: 1
- Diferenciar ciclo: 1

Mostrar:

- Autocorrelaciones
- Autocorrelaciones parciales

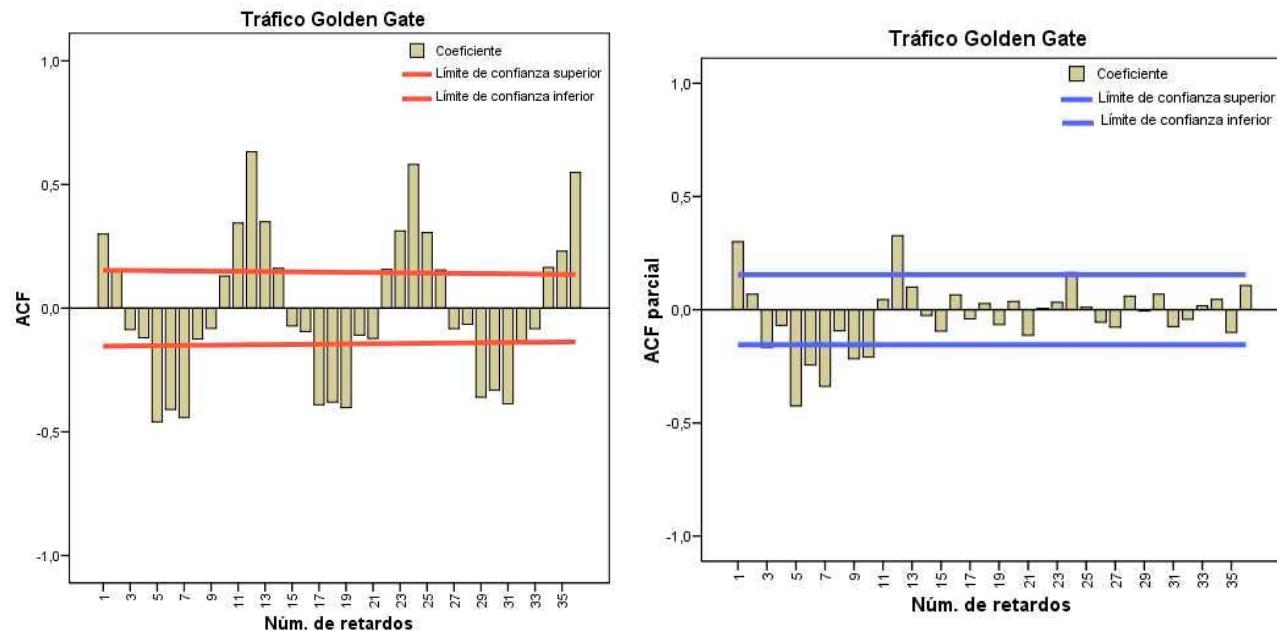
Periodicidad actual: 12

Aceptar Pegar Restablecer Cancelar Ayuda

PARTE ESTACIONAL

Aplicados los logaritmos, como la serie es estacional, el problema es identificar si se diferencia la parte regular de la serie o en la parte estacional. Bajo los supuestos de diferenciación en la parte regular o en la parte estacional (Diferenciar ciclo: 1).

DIFERENCIACIÓN PARTE REGULAR

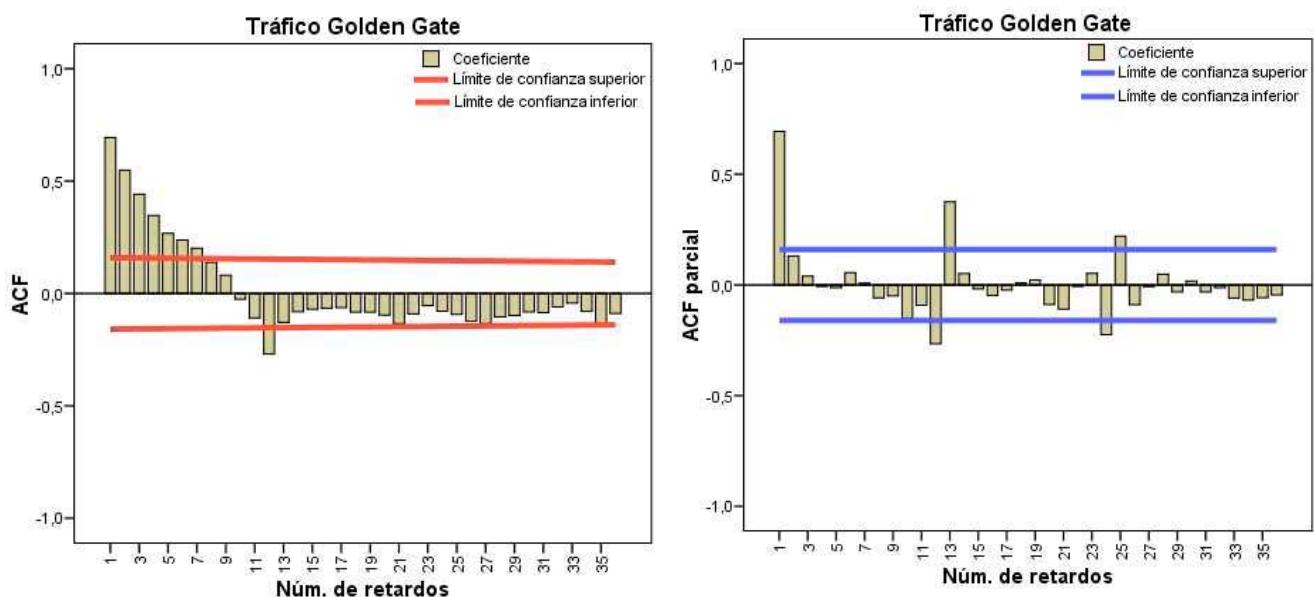


Al diferenciar solo la **parte regular**, las funciones de autocorrelación (ACF) y autocorrelación parcial estimadas (ACFP) **no superan** el problema de la falta de estacionariedad, pues la ACF no decae rápidamente.

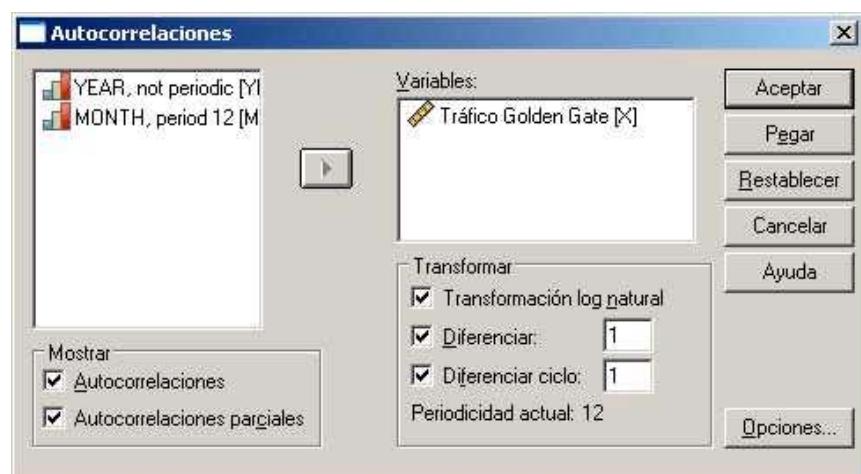
Autocorrelaciones							Autocorrelaciones parciales		
Retardo	Autocorrelación	Error típico ^a	Estadístico de Box-Ljung			Retardo	Autocorrelación parcial	Error típico	
			Valor	gl	Sig. ^b				
1	,300	,077	15,301	1	,000	1	,300	,077	
2	,152	,076	19,258	2	,000	2	,068	,077	
3	-,087	,076	20,560	3	,000	3	-,166	,077	
4	-,119	,076	23,025	4	,000	4	-,070	,077	
5	-,460	,076	59,840	5	,000	5	-,425	,077	
6	-,410	,076	89,240	6	,000	6	-,244	,077	
7	-,442	,075	123,720	7	,000	7	-,339	,077	
8	-,124	,075	126,461	8	,000	8	-,094	,077	
9	-,082	,075	127,665	9	,000	9	-,217	,077	
10	,130	,075	130,703	10	,000	10	-,209	,077	
11	,345	,074	152,190	11	,000	11	,044	,077	
12	,632	,074	224,954	12	,000	12	,327	,077	
13	,350	,074	247,346	13	,000	13	,100	,077	
14	,161	,074	252,149	14	,000	14	-,026	,077	
15	-,072	,073	253,113	15	,000	15	-,094	,077	
16	-,094	,073	254,780	16	,000	16	,066	,077	
17	-,390	,073	283,447	17	,000	17	-,040	,077	
18	-,380	,073	310,797	18	,000	18	,027	,077	
19	-,402	,072	341,616	19	,000	19	-,066	,077	
20	-,109	,072	343,904	20	,000	20	,036	,077	
21	-,122	,072	346,803	21	,000	21	-,113	,077	
22	,157	,072	351,587	22	,000	22	,005	,077	
23	,312	,071	370,675	23	,000	23	,033	,077	
24	,581	,071	437,321	24	,000	24	,164	,077	
25	,305	,071	455,827	25	,000	25	,011	,077	
26	,154	,071	460,592	26	,000	26	-,055	,077	
27	-,083	,070	461,978	27	,000	27	-,078	,077	
28	-,065	,070	462,844	28	,000	28	,060	,077	
29	-,360	,070	489,329	29	,000	29	-,006	,077	
30	-,331	,070	511,883	30	,000	30	,068	,077	
31	-,387	,069	542,940	31	,000	31	-,074	,077	
32	-,133	,069	546,655	32	,000	32	-,043	,077	
33	-,083	,069	548,096	33	,000	33	,017	,077	
34	,165	,069	553,869	34	,000	34	,045	,077	
35	,231	,068	565,237	35	,000	35	-,101	,077	
36	,549	,068	630,143	36	,000	36	,107	,077	

Al diferenciar solo una vez la **parte estacional**, las funciones de autocorrelación (ACF) y autocorrelación parcial estimadas (ACFP) **no superan** el problema de la falta de estacionariedad, pues la ACF no decae rápidamente.

DIFERENCIANDO PARTE ESTACIONAL



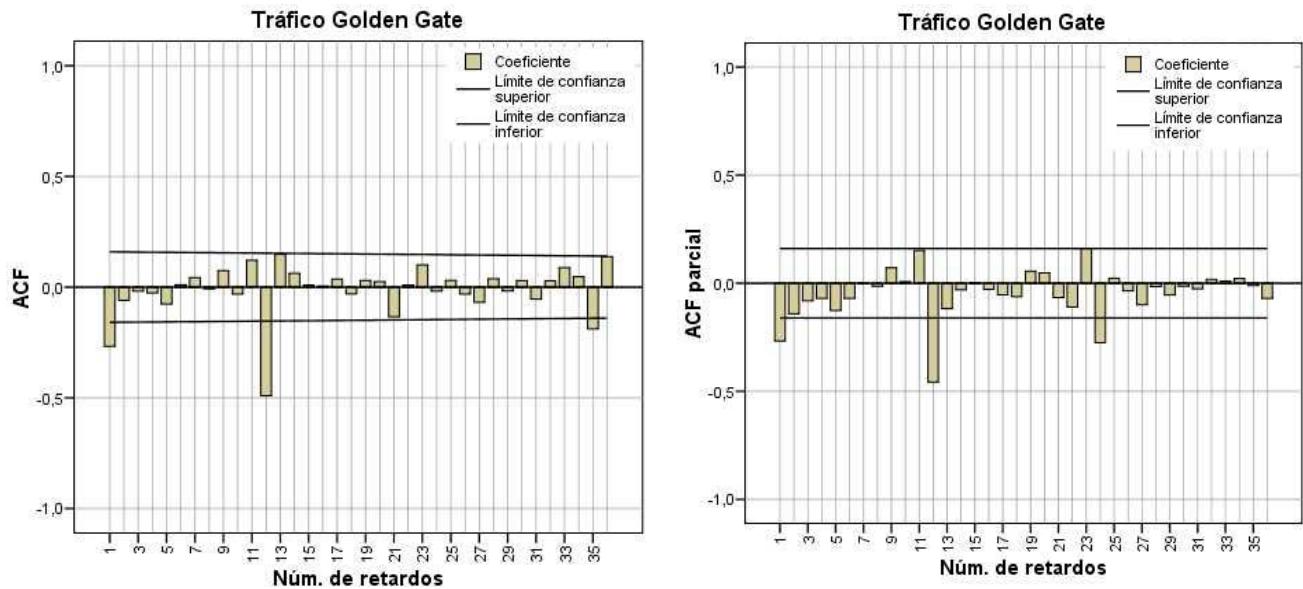
Serie: Tráfico Golden Gate Autocorrelaciones							Autocorrelaciones parciales		
Retardo	Autocorrelación	Error típico ^a	Estadístico de Box-Ljung			Retardo	Autocorrelación parcial	Error típico	
			Valor	gl	Sig. ^b				
1	,893	,079	76,456	1	,000	1	,693	,080	
2	,549	,079	124,621	2	,000	2	,131	,080	
3	,442	,079	156,038	3	,000	3	,040	,080	
4	,347	,079	175,526	4	,000	4	-,007	,080	
5	,268	,078	187,233	5	,000	5	-,012	,080	
6	,238	,078	196,502	6	,000	6	,055	,080	
7	,201	,078	203,189	7	,000	7	,008	,080	
8	,139	,077	206,386	8	,000	8	-,059	,080	
9	,081	,077	207,488	9	,000	9	-,050	,080	
10	-,026	,077	207,598	10	,000	10	-,151	,080	
11	-,109	,077	209,633	11	,000	11	-,091	,080	
12	-,270	,076	222,079	12	,000	12	-,266	,080	
13	-,129	,076	224,947	13	,000	13	,375	,080	
14	-,081	,076	226,090	14	,000	14	,052	,080	
15	-,070	,076	226,956	15	,000	15	-,018	,080	
16	-,065	,075	227,709	16	,000	16	-,048	,080	
17	-,063	,075	228,423	17	,000	17	-,023	,080	
18	-,083	,075	229,651	18	,000	18	,009	,080	
19	-,083	,075	230,878	19	,000	19	,022	,080	
20	-,097	,074	232,569	20	,000	20	-,088	,080	
21	-,134	,074	235,850	21	,000	21	-,109	,080	
22	-,091	,074	237,364	22	,000	22	-,007	,080	
23	-,054	,073	237,908	23	,000	23	,053	,080	
24	-,079	,073	239,061	24	,000	24	-,225	,080	
25	-,093	,073	240,693	25	,000	25	,221	,080	
26	-,124	,073	243,592	26	,000	26	-,089	,080	
27	-,134	,072	247,039	27	,000	27	-,008	,080	
28	-,104	,072	249,136	28	,000	28	,048	,080	
29	-,098	,072	250,992	29	,000	29	-,032	,080	
30	-,082	,071	252,314	30	,000	30	,018	,080	
31	-,085	,071	253,751	31	,000	31	-,032	,080	
32	-,060	,071	254,472	32	,000	32	-,013	,080	
33	-,043	,071	254,845	33	,000	33	-,059	,080	
34	-,080	,070	256,124	34	,000	34	-,068	,080	
35	-,143	,070	260,262	35	,000	35	-,058	,080	
36	-,089	,070	261,889	36	,000	36	-,045	,080	



Dado que no se cumplen las condiciones de estacionariedad en media, se aplican logaritmos diferenciando una vez la parte estacional y la parte regular.

En **Opciones** se indica que el número máximo de retardos es 36.

DIFERENCIANDO PARTE REGULAR Y PARTE ESTACIONAL



Se observa que la ACF no presenta claramente retardos significativos a lo largo de los períodos y la ACFP presenta como mucho un retardo significativo a lo largo de la ACF y ambas funciones tienen estructura sinusoidal, lo que conduce a pensar en una estructura ARIMA(1,1,0) para la parte regular y la misma estructura para la para la parte estacional. La estructura final para la serie será **ARIMA(1, 1, 0)(1, 1, 0)₁₂**

El problema de la estacionalidad y la estacionariedad en media y en varianza se resuelve aplicando logaritmos, diferenciando una vez la parte estacional y la parte regular. Con lo cual, la parte regular de la serie en logaritmos es integrada de orden uno **I(1)** y la parte estacional es integrada de orden uno **I(1)**.

El orden de la parte autorregresiva **AR** y la parte de medias móviles **MA** se realiza observando que los coeficientes de las últimas ACF y ACFP no se anulan bruscamente con periodicidades y que sus estructuras se ajustan claramente a un modelo **ARMA(1,0)(1,0)₁₂**.

La parte **MA(0)** de la parte regular proviene de que la **ACF** no presenta un solo retardo significativo, mientras que la parte **AR(1)** de la **ACF** proviene de las ondas sinusoidales.

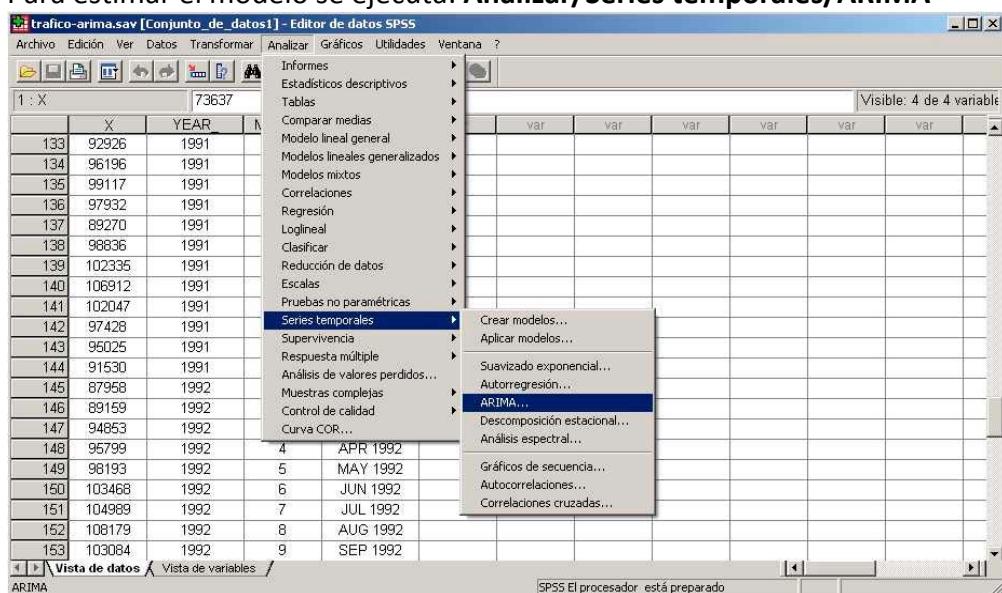
Autocorrelaciones
Serie: Tráfico Golden Gate

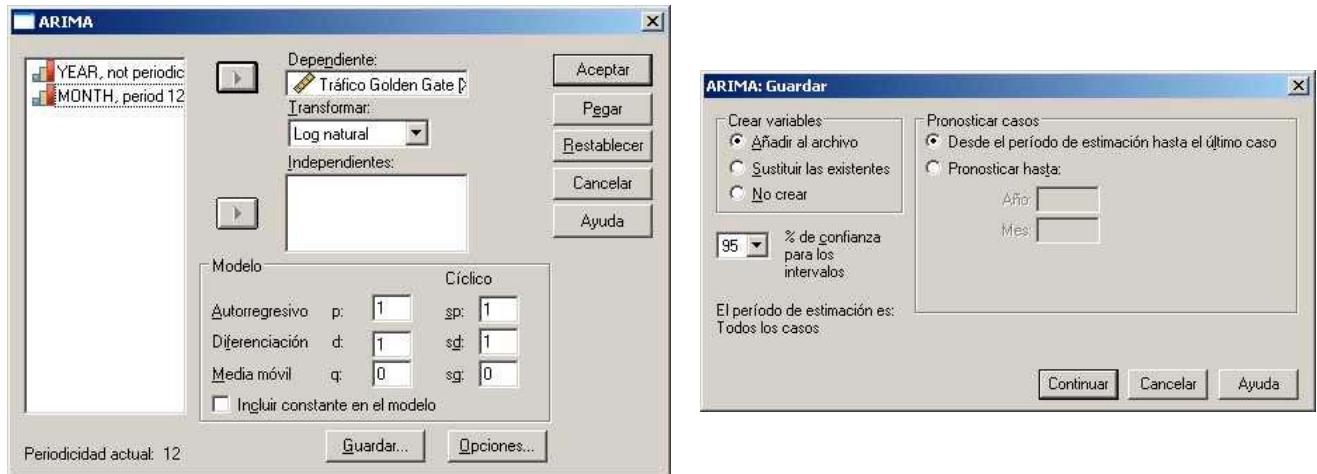
Retardo	Autocorrelación	Error típico ^a	Estadístico de Box-Ljung		
			Valor	gl	Sig. ^b
1	-,267	,080	11,271	1	,001
2	-,060	,079	11,843	2	,003
3	-,018	,079	11,895	3	,008
4	-,026	,079	12,004	4	,017
5	-,077	,079	12,965	5	,024
6	,011	,078	12,983	6	,043
7	,043	,078	13,283	7	,066
8	-,008	,078	13,292	8	,102
9	,075	,077	14,224	9	,115
10	-,031	,077	14,387	10	,156
11	,121	,077	16,862	11	,112
12	-,490	,077	57,732	12	,000
13	,148	,076	61,480	13	,000
14	,063	,076	62,163	14	,000
15	,008	,076	62,175	15	,000
16	,005	,076	62,179	16	,000
17	,037	,075	62,419	17	,000
18	-,031	,075	62,586	18	,000
19	,030	,075	62,750	19	,000
20	,024	,074	62,857	20	,000
21	-,134	,074	66,102	21	,000
22	,009	,074	66,117	22	,000
23	,101	,074	67,984	23	,000
24	-,018	,073	68,042	24	,000
25	,030	,073	68,211	25	,000
26	-,031	,073	68,397	26	,000
27	-,069	,073	69,289	27	,000
28	,038	,072	69,572	28	,000
29	-,017	,072	69,628	29	,000
30	,030	,072	69,799	30	,000
31	-,054	,071	70,378	31	,000
32	,028	,071	70,537	32	,000
33	,088	,071	72,073	33	,000
34	,046	,071	72,507	34	,000
35	-,189	,070	79,767	35	,000
36	,138	,070	83,647	36	,000

Autocorrelaciones parciales
Serie: Tráfico Golden Gate

Retardo	Autocorrelación parcial	Error típico
1	-,267	,080
2	-,141	,080
3	-,081	,080
4	-,070	,080
5	-,126	,080
6	-,071	,080
7	-,003	,080
8	-,014	,080
9	,071	,080
10	,008	,080
11	,151	,080
12	-,457	,080
13	-,117	,080
14	-,031	,080
15	-,002	,080
16	-,029	,080
17	-,054	,080
18	-,062	,080
19	,055	,080
20	,048	,080
21	-,066	,080
22	-,110	,080
23	,160	,080
24	-,275	,080
25	,024	,080
26	-,036	,080
27	-,099	,080
28	-,015	,080
29	-,055	,080
30	-,014	,080
31	-,026	,080
32	,018	,080
33	,010	,080
34	,022	,080
35	-,011	,080
36	-,071	,080

Para estimar el modelo se ejecuta: **Analizar/Series temporales/ARIMA**





Se elige **Guardar** para crear nuevas variables que contengan valores pronosticados (**FIT_1**), residuos (**ERR_1**), intervalos de confianza (**LCL_1**, **UCL_1**), errores estándar para las predicciones (**SEP_1**). Todas las variables se añaden al Editor de datos como nuevas columnas:

*trafico-arima.sav [Conjunto_de_datos1] - Editor de datos SPSS

	MONTH	DATE	FIT_1	ERR_1	LCL_1	UCL_1	SEP_1
1	1	JAN 1980					
2	2	FEB 1980					
3	3	MAR 1980					
4	4	APR 1980					
5	5	MAY 1980					
6	6	JUN 1980					
7	7	JUL 1980					
8	8	AUG 1980					
9	9	SEP 1980					
10	10	OCT 1980					
11	11	NOV 1980					
12	12	DEC 1980					
13	1	JAN 1981					
14	2	FEB 1981	78799,45679	,00782	73761,50301	84181,50577	2635,26951
15	3	MAR 1981	83723,87927	,01292	78543,63453	89245,78042	2706,75376
16	4	APR 1981	87322,78911	-,01886	81919,88913	93082,05174	2823,10483
17	5	MAY 1981	86612,74005	,01008	81253,75291	92325,17231	2800,14928
18	6	JUN 1981	94763,64803	-,01884	88900,34005	101013,66298	3063,66431
19	7	JUL 1981	95793,06878	-,00437	89866,06748	102110,97785	3096,94500
20	8	AUG 1981	97275,72083	,01145	91256,98350	103691,41627	3144,87842
21	9	SEP 1981	90832,16620	,02134	85212,11135	96822,88451	2936,56097
22	10	OCT 1981	86669,13047	,01544	81306,66444	92385,27036	2801,96688
23	11	NOV 1981	84980,25609	,02233	79722,42755	90584,84729	2747,28984

Se obtiene el ajuste a un modelo de **Box-Jenkins ARIMA(1,1,0)**:

Diagnóstico residual

Criterios de finalización de las iteraciones

Cambio máximo en los parámetros menor que	,001
Máxima constante de Marquardt mayor que	1E+009
Cambio en el porcentaje de la suma de cuadrados menor que	,001%
Número de iteraciones igual a	10

Número de residuos	155
Número de parámetros	2
GL residuales	153
Suma de cuadrados residual corregida	,126
Suma de cuadrados residual	,126
Varianza residual	,001
Error típico del modelo	,028
Log-verosimilitud	331,429
Criterio de información de Akaike (AIC)	-658,859
Criterio bayesiano de Schwarz (BIC)	-652,772

Estimaciones de los parámetros

		Estimaciones	Error típico	t	Sig. aprox.
Retardos no estacionales	AR1	-,256	,078	-3,298	,001
Retardos estacionales	Seasonal AR1	-,477	,070	-6,775	,000

Se ha utilizado el algoritmo de Melard para la estimación.

Matriz de covarianza

		Retardos no estacionales	Retardos estacionales
		AR1	Seasonal AR1
Retardos no estacionales	AR1	,006	,000
Retardos estacionales	Seasonal AR1	,000	,005

Se ha utilizado el algoritmo de Melard para la estimación.

Matriz de correlaciones

		Retardos no estacionales	Retardos estacionales
		AR1	Seasonal AR1
Retardos no estacionales	AR1	1,000	,002
Retardos estacionales	Seasonal AR1	,002	1,000

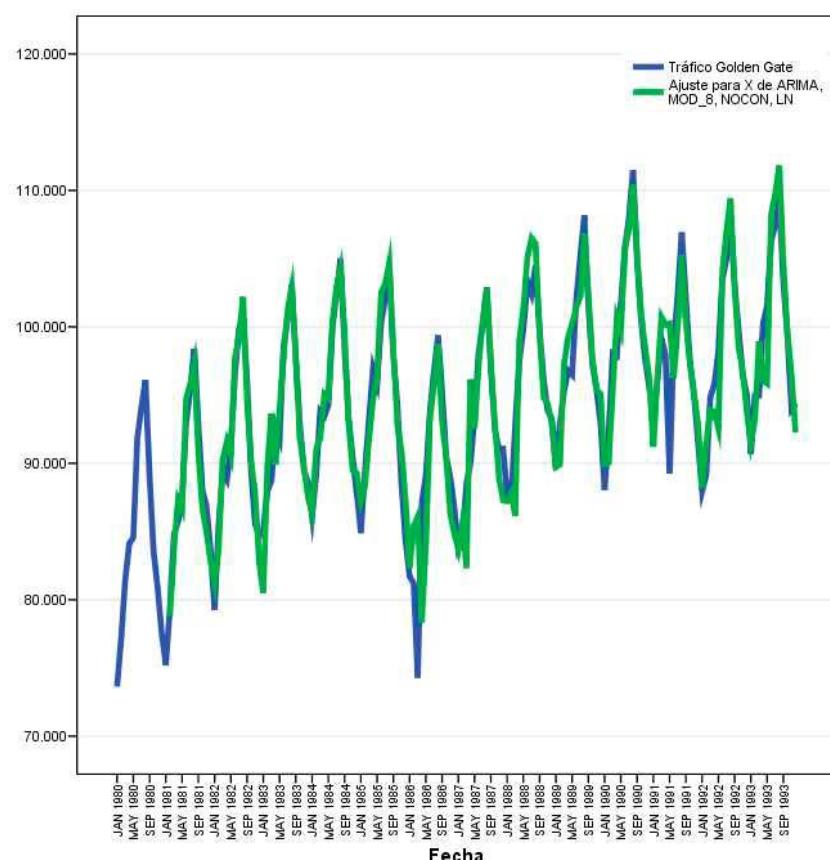
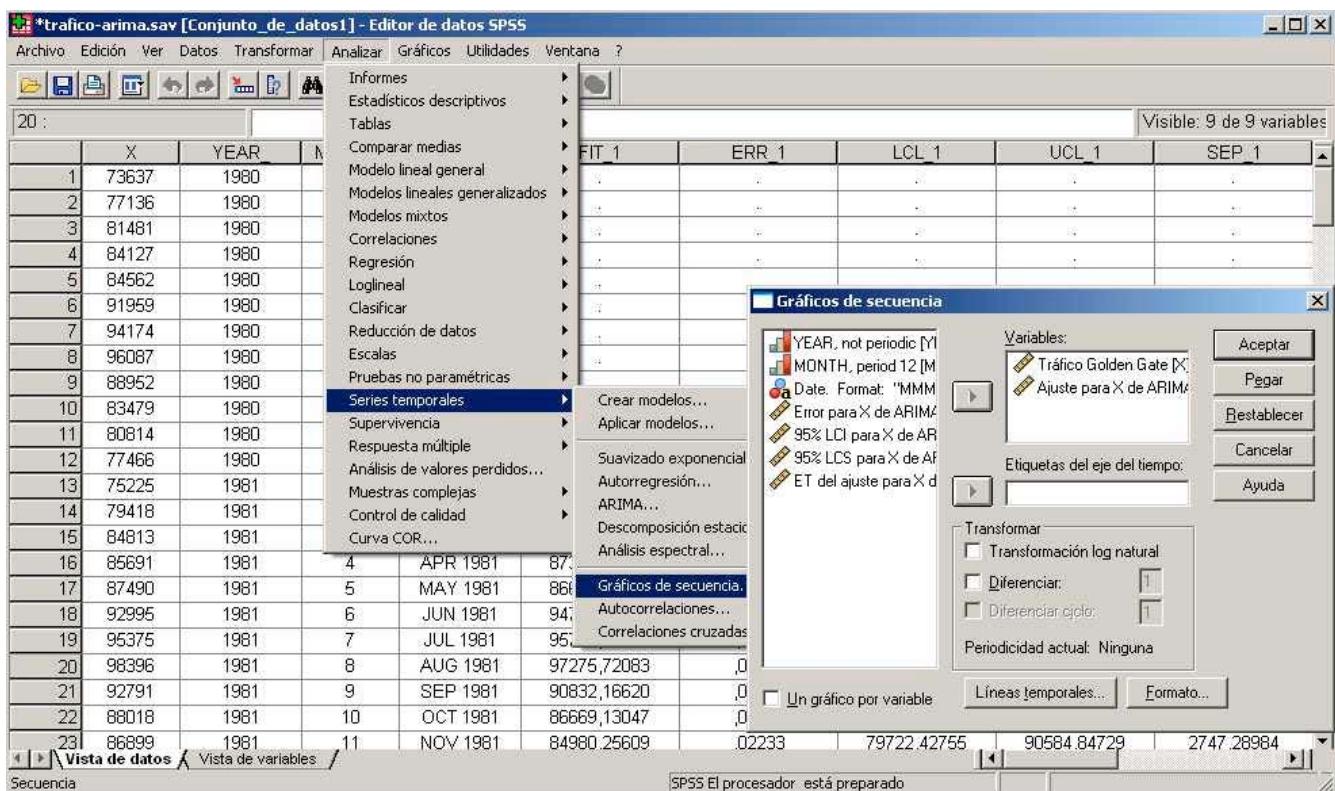
Se ha utilizado el algoritmo de Melard para la estimación.

$$\text{Matriz covarianzas: } \begin{pmatrix} & \text{AR1} & \text{SAR1} \\ \text{AR1} & 0,06 & 0 \\ \text{SAR1} & 0 & 0,05 \end{pmatrix}$$

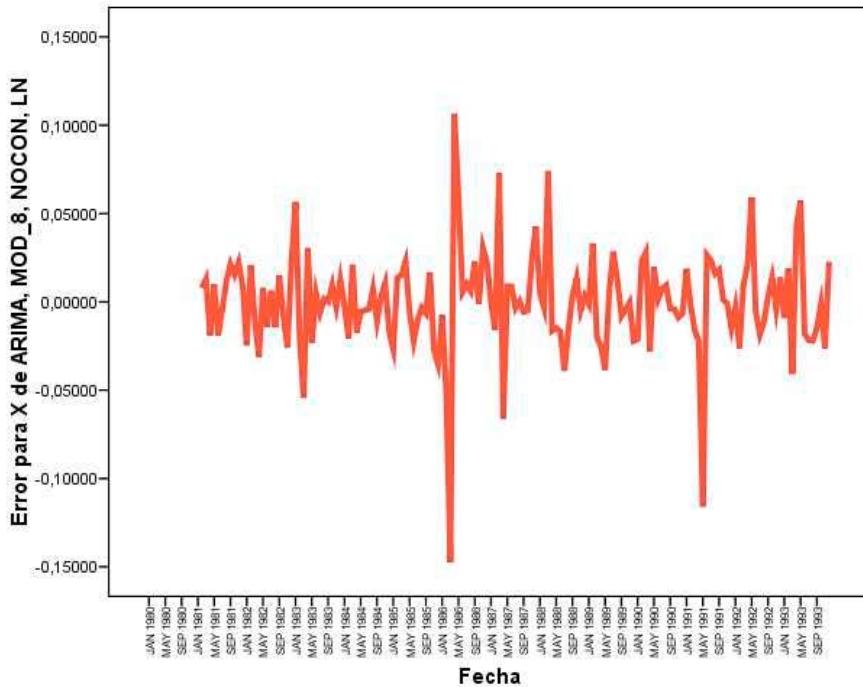
$$\text{Matriz correlaciones: } \begin{pmatrix} & \text{AR1} & \text{SAR1} \\ \text{AR1} & 1 & 0,02 \\ \text{SAR1} & 0,02 & 1 \end{pmatrix}$$

El ajuste ha resultado muy bueno con una significatividad del parámetro MA altísima (p-valor nulo), el diagnóstico del modelo es correcto.

Analizar/Series temporales/Gráfico de secuencia para obtener la representación de la serie original X_t y la serie de las predicciones **FIT_1**



Análogamente, con la instrucción **Analizar/Series temporales/Gráfico de secuencia** se obtiene la representación de los errores del modelo estimado, que presenta una estructura aleatoria, hecho favorable como verificación del diagnóstico de la modelización ARIMA realizada.



DATOS (en miles) *trafico-ariama.sav*

	1980	1981	1982	1983	1984	1985	1986
Enero	73637	75225	79245	85169	86141	84883	81714
Febrero	77136	79418	85536	87973	89156	89476	81222
Marzo	81481	84813	89313	88696	93795	93191	74295
Abril	84127	85691	88785	92686	93422	96919	87121
Mayo	84562	87490	91307	91807	94277	95869	89099
Junio	91959	92995	96394	98593	99927	100272	93743
Julio	94174	95375	99864	100677	102451	102103	96741
Agosto	96087	98396	100744	103084	105003	104381	99366
Septiembre	88952	92791	96009	97644	98477	97779	95184
Octubre	83479	88018	89428	92648	93219	94369	90407
Noviembre	80814	86899	85518	89486	90413	88432	88968
Diciembre	77466	83636	84603	88593	87763	84314	86765

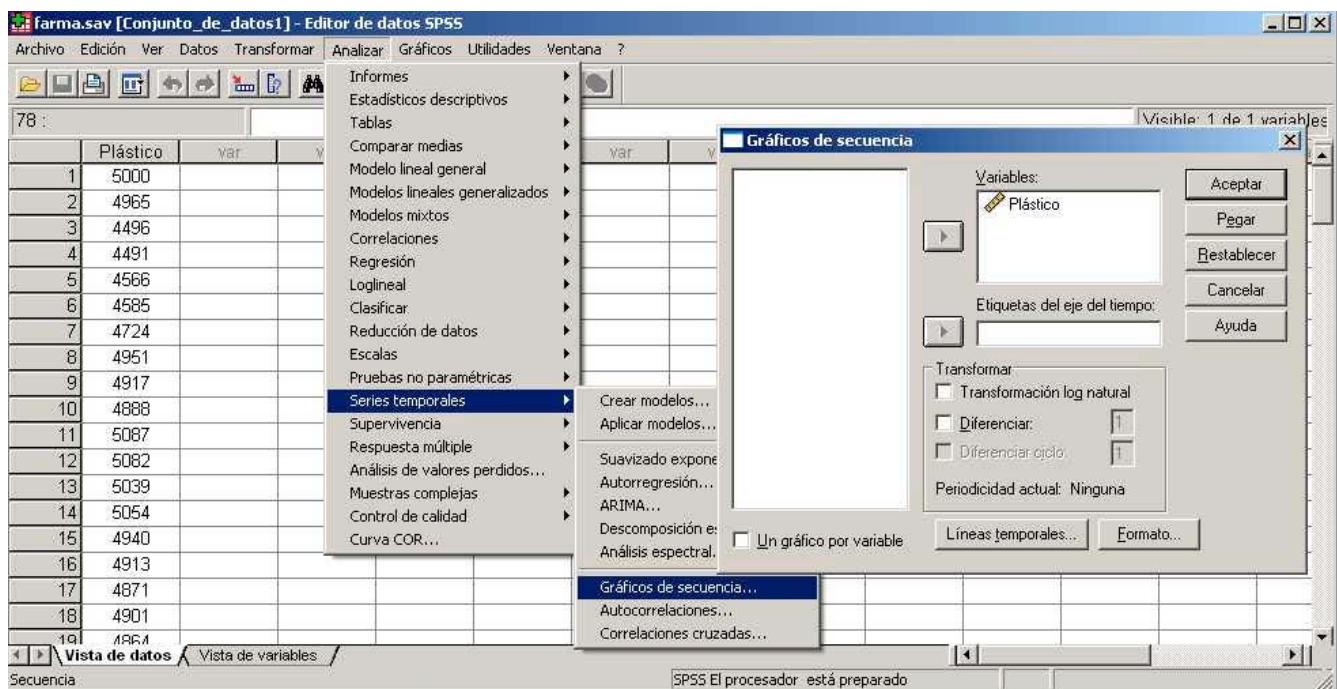
	1987	1988	1989	1990	1991	1992	1993
Enero	83674	87656	89609	88047	92926	87958	90707
Febrero	84265	87534	92934	92172	96196	89159	94949
Marzo	88551	92763	95272	98050	99117	94853	94970
Abril	89998	97427	96791	97874	97932	95799	100286
Mayo	93421	99845	96478	101700	89270	98193	101497
Junio	98109	103349	101938	105794	98836	103468	106352
Julio	100235	102438	105094	108016	102335	104989	107415
Agosto	102866	104467	108158	111475	106912	108179	109385
Septiembre	96903	99744	101625	104945	102047	103084	103266
Octubre	92426	96101	97255	100496	97428	99781	99432
Noviembre	91024	93858	95378	97555	95025	96403	93965
Diciembre	91072	93265	92997	95556	91530	95072	94385

Ejercicio 3.- El archivo **farma.sav** recoge cien datos relativos a la demanda mensual de contenedores de plástico que utilizan las compañías farmacéuticas desde enero de 2002.

El objetivo es predecir el número de contenedores que serán demandados en los próximos diez primeros meses con vistas a la producción.

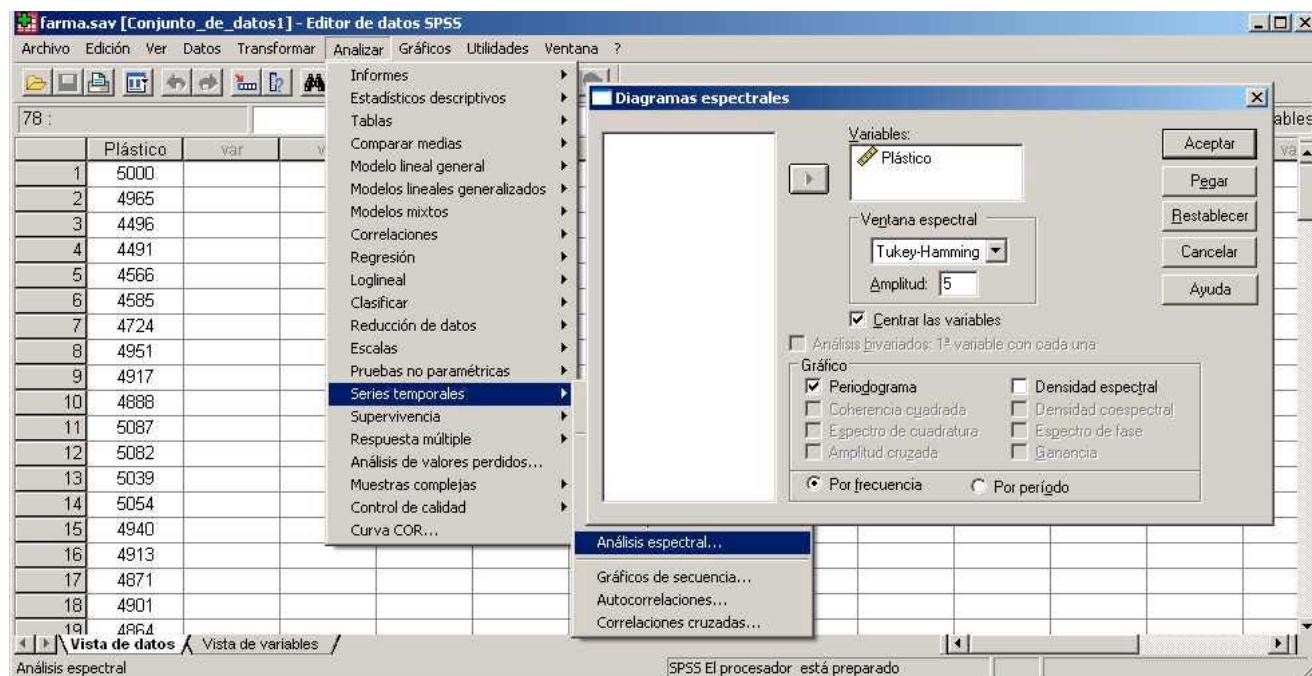
- Utilizando la **metodología de Box-Jenkins**

La fase de identificación comienza realizando una representación gráfica de la variable [Analizar/Series temporales/Gráficos de secuencia] con objeto de observar la estacionalidad.

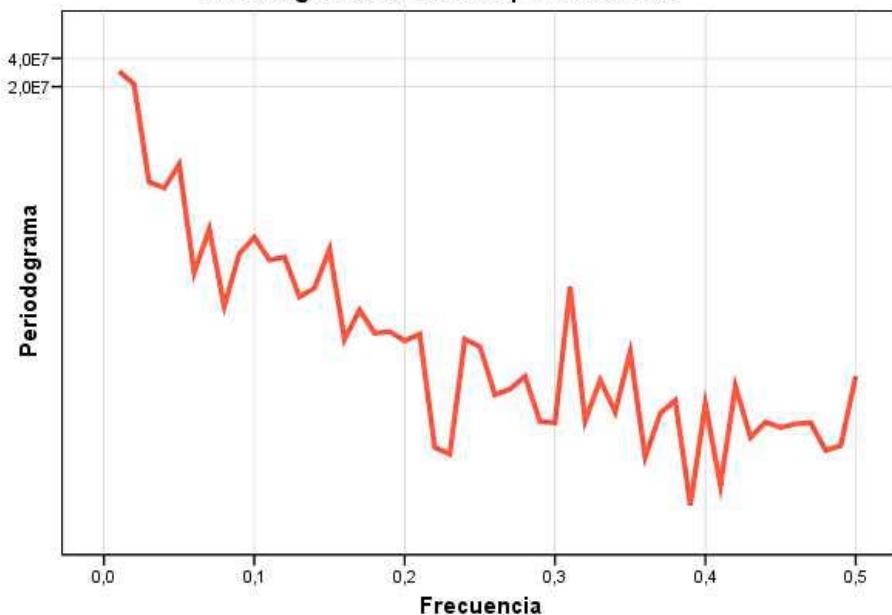


La estructura de la serie temporal es **no estacional**.

Para observar mejor la estacionalidad, se representa el PERIODOGRAMA por frecuencia de la serie mediante [Analizar/Series temporales/Análisis espectral].



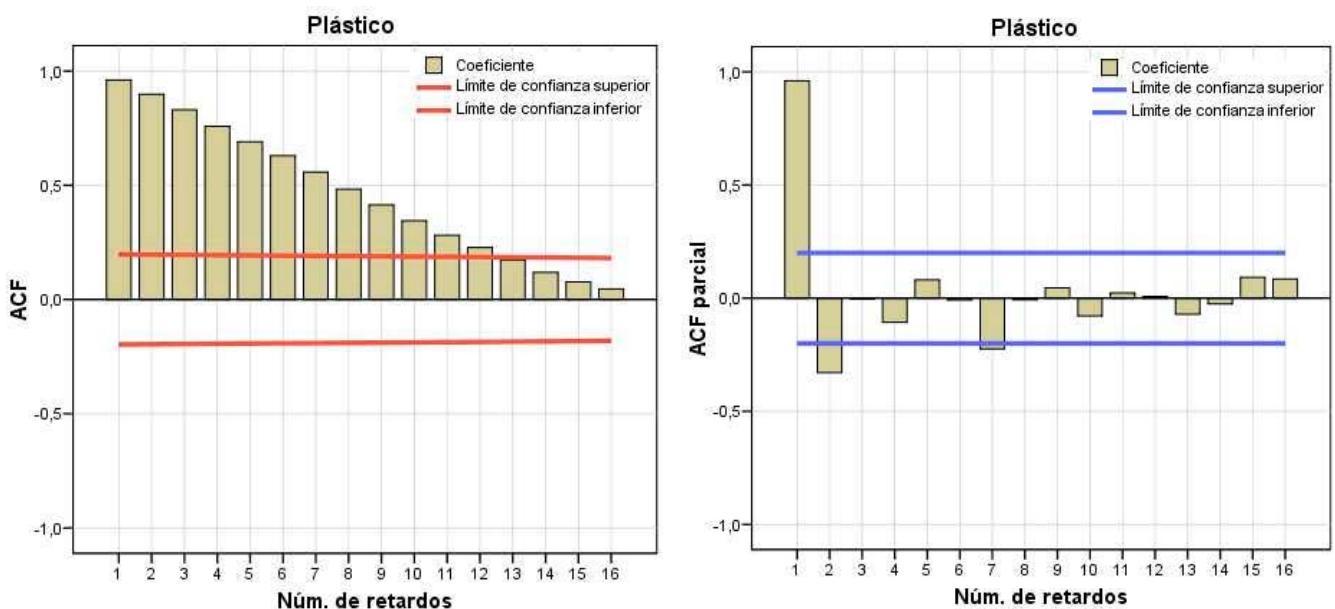
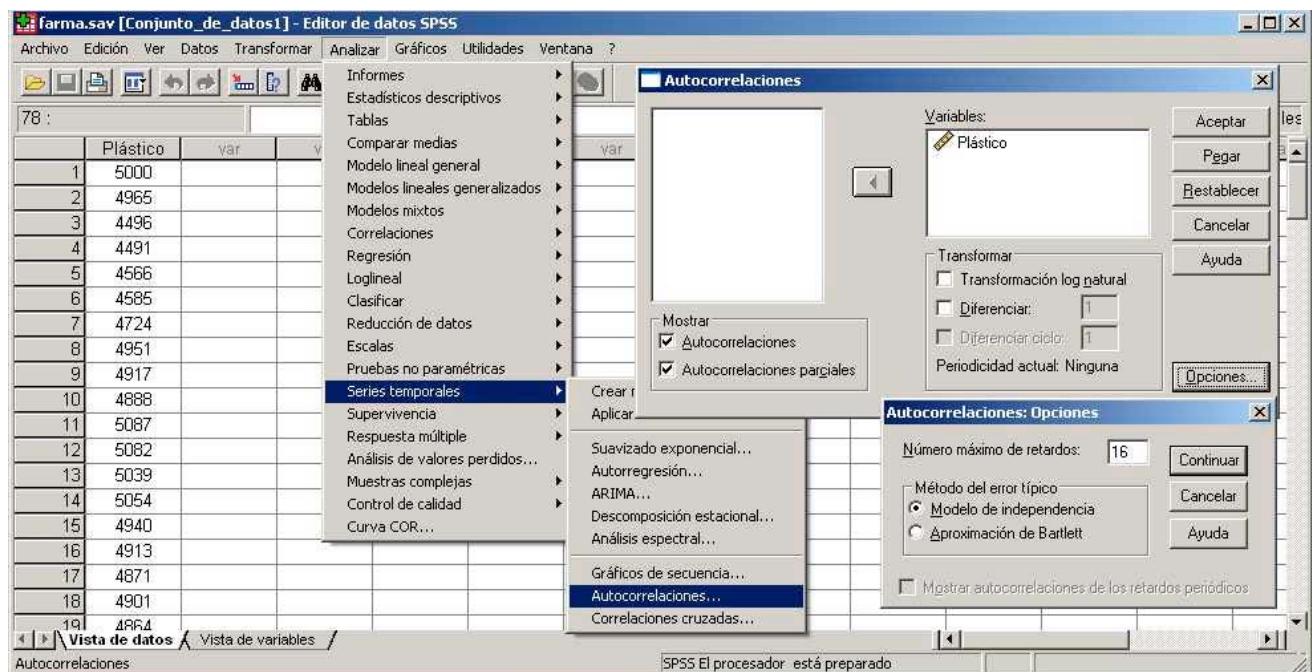
Periodograma de Plástico por frecuencia



Tiene puntos máximos para valores de la frecuencia muy pequeños, cuyos inversos producen unos posibles períodos estacionales más elevados incluso que la longitud de la serie. Circunstancia que indica que no hay estacionalidad, hecho que se extraía de la representación gráfica anterior.

La **estacionalidad**, así como la **estacionariedad**, pueden detectarse también mediante las funciones de **autocorrelación** y **autocorrelación parcial estimadas (ACF y ACFP respectivamente)**.

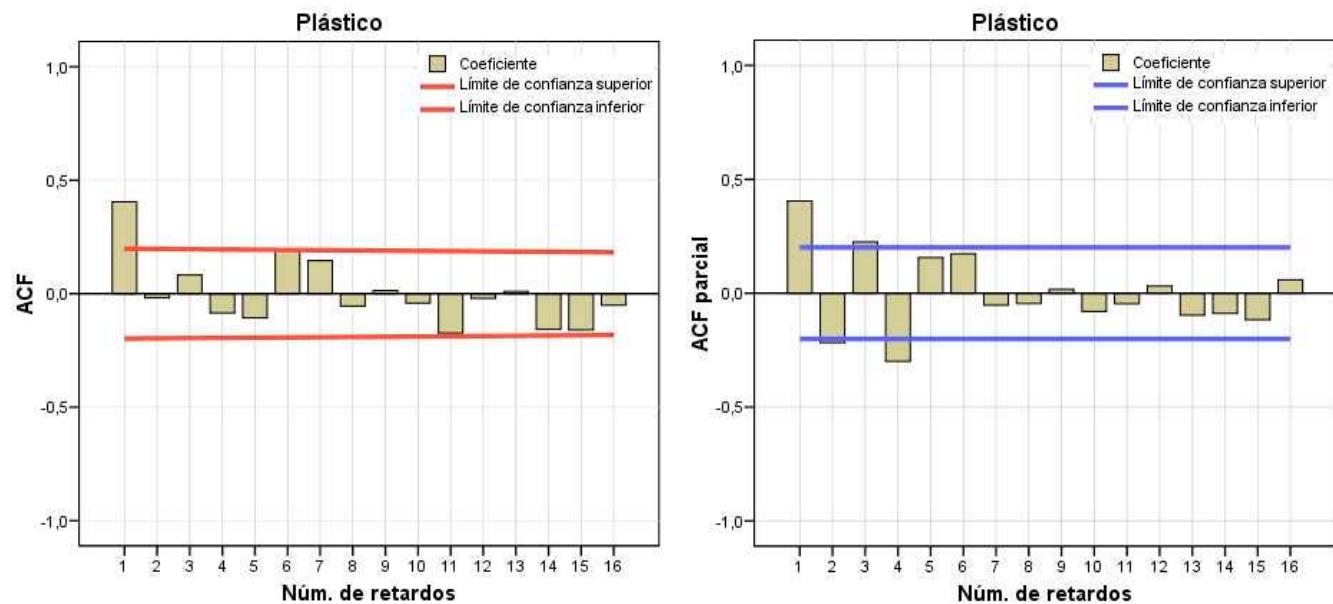
Para ello se elige **Analizar/Series temporales/Autocorrelaciones**. En el botón **Opciones** se deja **16**, valor por defecto para representar la función de autocorrelación **ACF** con un tramo significativo.



Se observa que los coeficientes de la función de autocorrelación ACF no decaen rápidamente, indicando **falta de estacionariedad en media**. En consecuencia, se diferencia la serie original.



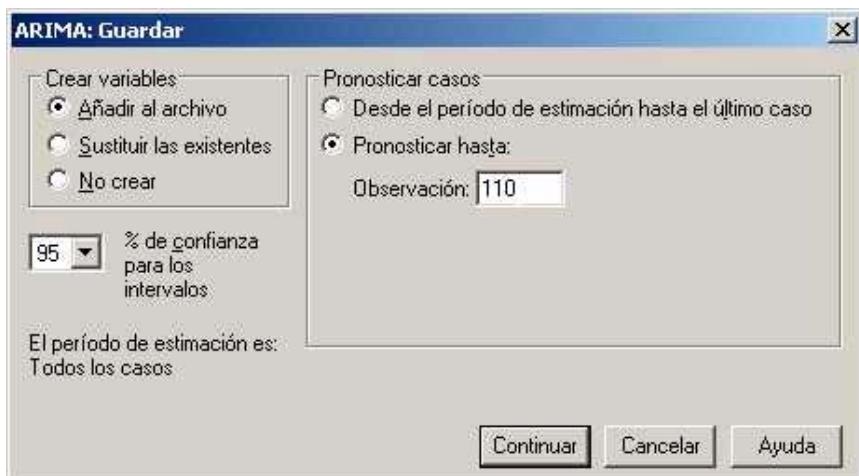
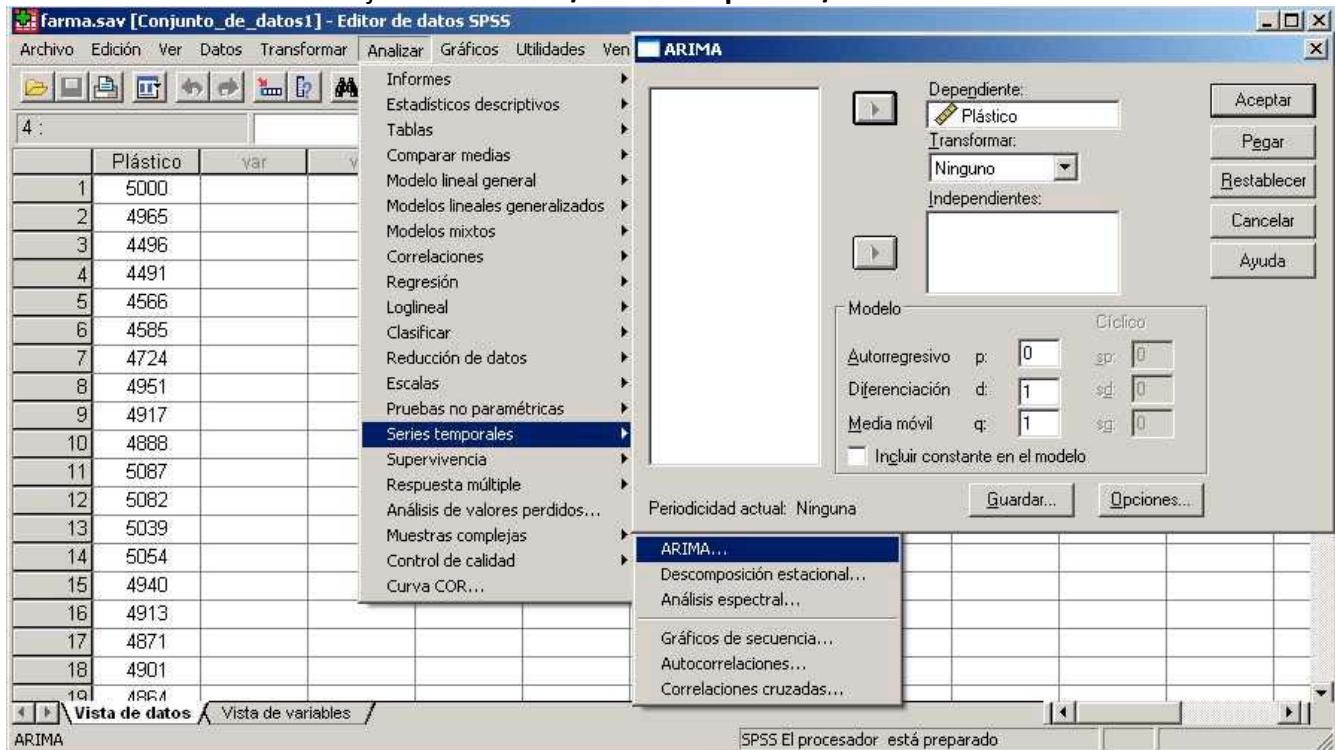
Autocorrelaciones						Autocorrelaciones parciales		
Retardo	Autocorrelación	Error típico ^a	Estadístico de Box-Ljung			Retardo	Autocorrelación parcial	Error típico
			Valor	gl	Sig. ^b			
1	,405	,099	16,709	1	,000	1	,405	,101
2	-,018	,098	16,743	2	,000	2	-,217	,101
3	,082	,098	17,447	3	,001	3	,225	,101
4	-,084	,097	18,198	4	,001	4	-,299	,101
5	-,107	,097	19,410	5	,002	5	,156	,101
6	,195	,096	23,495	6	,001	6	,173	,101
7	,145	,096	25,785	7	,001	7	-,052	,101
8	-,055	,095	26,114	8	,001	8	-,045	,101
9	,014	,095	26,136	9	,002	9	,017	,101
10	-,042	,094	26,330	10	,003	10	-,080	,101
11	-,173	,094	29,730	11	,002	11	-,046	,101
12	-,021	,093	29,779	12	,003	12	,031	,101
13	,010	,093	29,790	13	,005	13	-,097	,101
14	-,157	,092	32,703	14	,003	14	-,089	,101
15	-,160	,092	35,740	15	,002	15	-,116	,101
16	-,050	,091	36,047	16	,003	16	,059	,101



- Los retardos de la función de autocorrelación ACF decaen tan rápidamente que sólo el primero es significativo, con lo que **no existen problemas de estacionariedad en la serie diferenciada**. En concreto, la *serie diferenciada* es $I(0)$ y la serie original es $I(1)$.
- Respecto a la identificación de la parte de la media móvil de la serie, solo el primer retardo de la ACF es significativo y el decrecimiento de los retardos de la ACFP es muy rápido. En consecuencia, la parte de *media móvil* se modeliza como un proceso MA(1).
- Para la identificación de la parte autorregresiva se observa que aunque hay tres retardos de la ACF estimada ninguno de ellos es claramente significativo, decreciendo rápido los coeficientes significativos de la ACF. La parte *autorregresiva* se modeliza como un proceso AR(0).
- Considerando las dos funciones de autocorrelación en conjunto, se observa que sus retardos no se anulan demasiado bruscamente. Por tanto, es una estructura ARMA(0, 1) para la serie diferenciada, concluyendo que la serie original se ajusta a un modelo ARIMA(0, 1, 1).

Para estimar el modelo ARIMA(0,1, 1) se ejecuta el procedimiento ARIMA (*Modelo Autorregresivo Integrado de Media Móvil*) mediante la instrucción **Analizar/Series temporales/ARIMA**

Para estimar el modelo se ejecuta: **Analizar/Series temporales/ARIMA**



En el campo *Pronosticar hasta* se incluyen 110 observaciones como se deseaba.

Se elige **Guardar** para crear nuevas variables que contengan valores pronosticados (**FIT_1**), residuos (**ERR_1**), intervalos de confianza (**LCL_1 , UCL_1**), errores estándar para las predicciones (**SEP_1**). Todas las variables se añaden al Editor de datos como nuevas columnas.

Al pulsar **Aceptar** se obtiene el ajuste a un modelo de **Box-Jenkins ARIMA(0,1,1)**:

Diagnóstico residual

Criterios de finalización de las iteraciones

Cambio máximo en los parámetros menor que	,001
Máxima constante de Marquardt mayor que	1E+009
Cambio en el porcentaje de la suma de cuadrados menor que	,001%
Número de iteraciones igual a	10

Número de residuos	99
Número de parámetros	1
GL residuales	97
Suma de cuadrados residual corregida	2464863
Suma de cuadrados residual	2630144
Varianza residual	25218,883
Error típico del modelo	158,805
Log-verosimilitud	-641,544
Criterio de información de Akaike (AIC)	1287,088
Criterio bayesiano de Schwarz (BIC)	1292,278

Estimaciones de los parámetros

	Estimaciones	Error típico	t	Sig. aprox.
Retardos no estacionales MA1	-,727	,074	-9,874	,000
Constante	7,501	27,443	,273	,785

Se ha utilizado el algoritmo de Melard para la estimación.

El parámetro MA tiene una significación muy alta (p-valor asociado aproximadamente de cero), en consecuencia la diagnosis del modelo es correcto.

La ecuación del modelo ARIMA(0, 1, 1) estimada será:

$$(1 - B) Y_t = (1 + 0,727) a_t \Rightarrow Y_t - Y_{t-1} = a_t + 0,727 B a_{t-1}$$

Matriz de covarianza

		Retardos no estacionales	Constante
		MA1	
Retardos no estacionales	MA1	,005	0 ^a
Constante		0 ^a	753,142

Se ha utilizado el algoritmo de Melard para la estimación.

- a. La estimación del parámetro ARMA y la estimación de los parámetros de la regresión no están correlacionados asintóticamente.

Matriz de correlaciones

		Retardos no estacionales	Constante
		MA1	
Retardos no estacionales	MA1	1,000	0 ^a
Constante		0 ^a	1,000

Se ha utilizado el algoritmo de Melard para la estimación.

- a. La estimación del parámetro ARMA y la estimación de los parámetros de la regresión no están correlacionados asintóticamente.

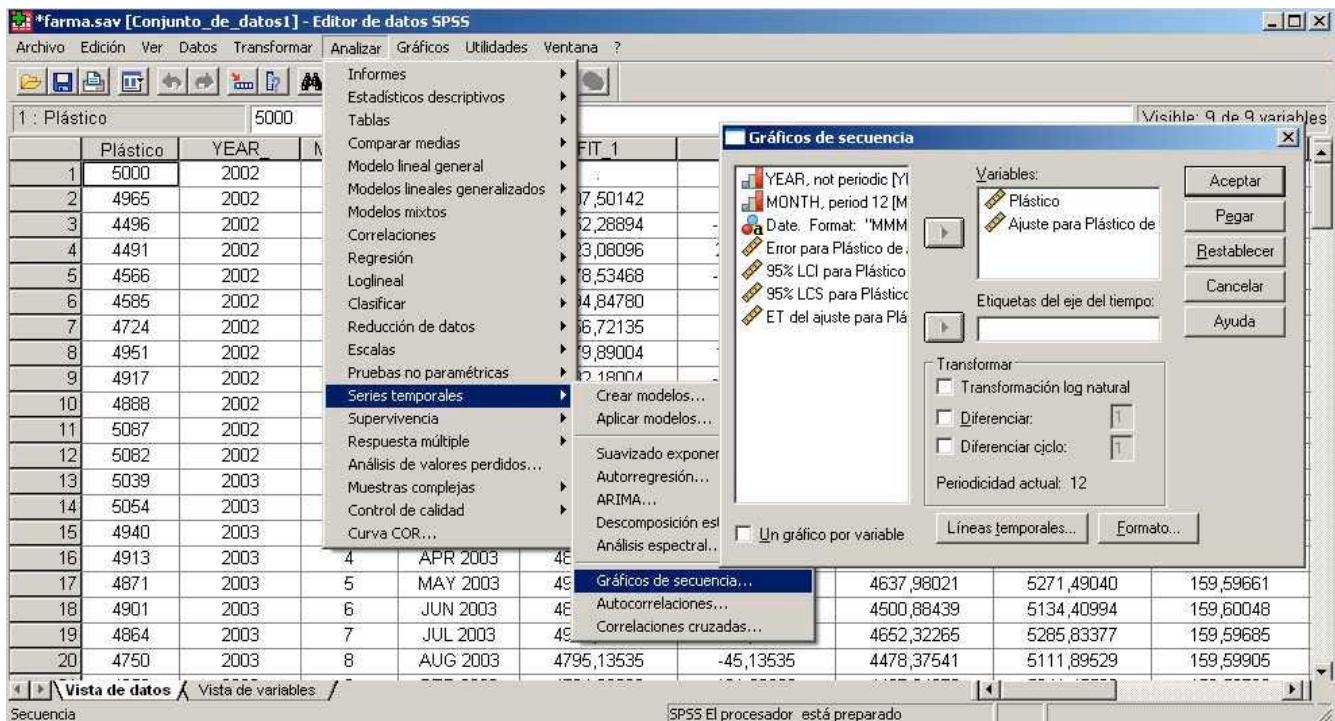
En el *Editor de datos*:

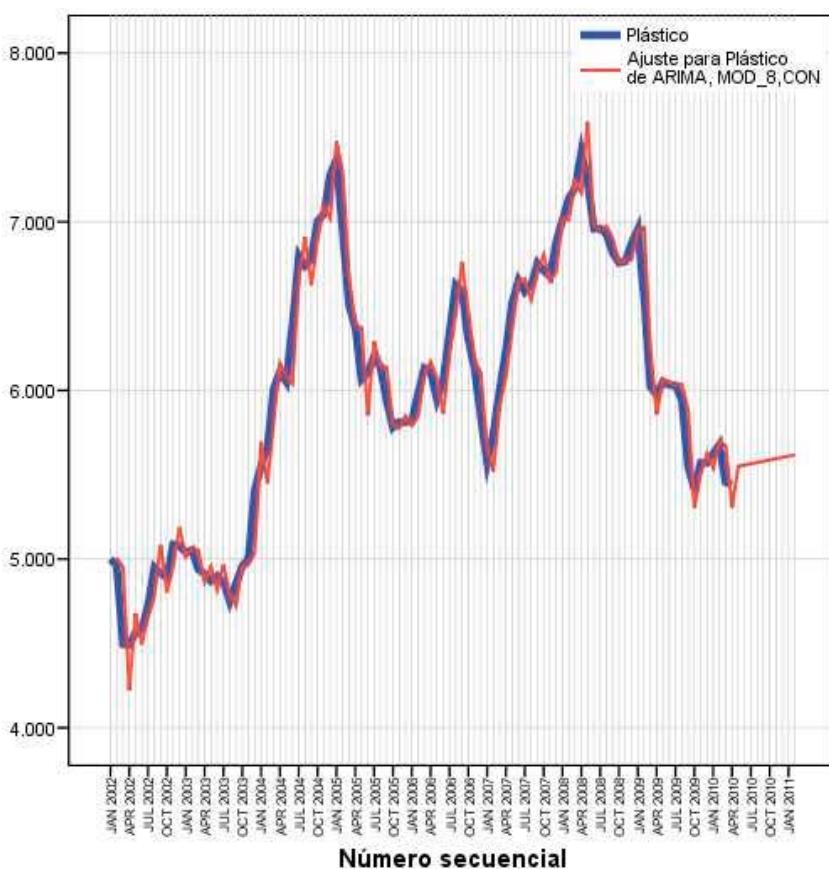
	Plástico	YEAR	MONTH	DATE	FIT_1	ERR_1	LCL_1	UCL_1	SEP_1
92	5944	2009	8	AUG 2009	6038,84799	-94,84799	5722,09044	6355,80553	159,59784
93	5543	2009	9	SEP 2009	5882,56988	-339,56988	5565,81234	6199,32742	159,59784
94	5416	2009	10	OCT 2009	5303,71627	112,28373	4986,95873	5620,47382	159,59784
95	5571	2009	11	NOV 2009	5505,10452	65,89548	5188,34698	5821,86206	159,59784
96	5571	2009	12	DEC 2009	5626,39149	-55,39149	5309,63395	5943,14903	159,59784
97	5627	2010	1	JAN 2010	5538,24521	88,75479	5221,48767	5885,00275	159,59784
98	5679	2010	2	FEB 2010	5699,00468	-20,00468	5382,24714	6015,76222	159,59784
99	5455	2010	3	MAR 2010	5671,96286	-216,96286	5355,20532	5988,72040	159,59784
100	5443	2010	4	APR 2010	5304,82192	138,17808	4988,06437	5621,57946	159,59784
101		2010	5	MAY 2010	5550,92345		5234,16591	5887,68099	159,59784
102		2010	6	JUN 2010	5558,42487		4923,64875	6193,20100	319,83106
103		2010	7	JUL 2010	5565,92629		4722,43381	6409,41877	424,99250
104		2010	8	AUG 2010	5573,42771		4560,53491	6586,32051	510,34462
105		2010	9	SEP 2010	5580,92912		4420,60265	6741,25560	584,62887
106		2010	10	OCT 2010	5588,43054		4295,10279	6881,75830	651,64138
107		2010	11	NOV 2010	5595,93196		4179,96259	7011,90133	713,43419
108		2010	12	DEC 2010	5603,43338		4072,69000	7134,17676	771,26291
109		2011	1	JAN 2011	5610,93480		3971,63167	7250,23793	825,96059
110		2011	2	FEB 2011	5618,43621		3875,62594	7361,24649	878,11252
111									

La variable **FIT_1** ha generado las predicciones hasta febrero de 2011.

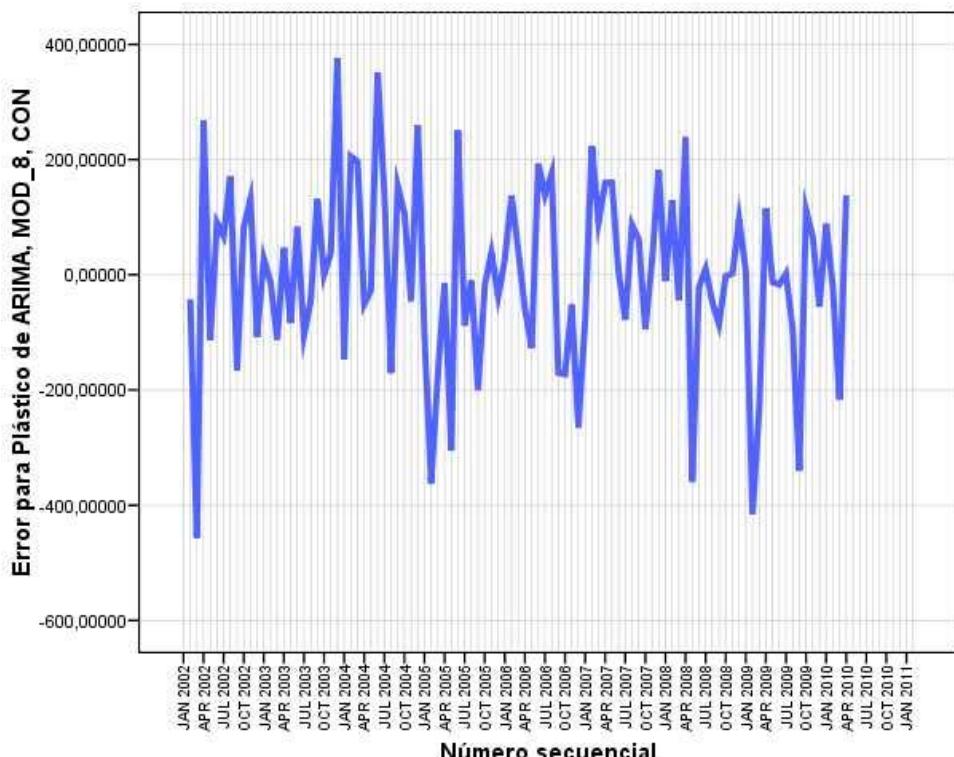
La variable **ERR_1** ha generado las estimaciones del término de error del modelo, las variables **LCL_1**, **UCL_1** han generado límites inferiores y superiores de los intervalos de confianza al 95% de fiabilidad, la variable **SEP_1** contiene los errores estándar para las predicciones.

Para obtener la representación de la serie original y la serie de predicciones **FIT_1** se recurre a **Analizar/Series temporales/Gráfico de secuencia**





Análogamente, con la instrucción **Analizar/Series temporales/Gráfico de secuencia** se obtiene la representación de los errores del modelo estimado, que presenta una estructura aleatoria, hecho favorable como verificación del diagnóstico de la modelización ARIMA realizada.



Fichero: *farma.sav*

	2002	2003	2004	2005	2006	2007
Enero	5000	5039	5550	7367	5818	5572
Febrero	4965	5054	5657	6934	5982	5744
Marzo	4496	4940	6010	6506	6132	6005
Abril	4491	4913	6109	6374	6111	6239
Mayo	4566	4871	6052	6066	5948	6523
Junio	4585	4901	6391	6102	6056	6652
Julio	4724	4864	6798	6204	6342	6585
Agosto	4951	4750	6740	6138	6626	6622
Septiembre	4917	4856	6778	5938	6591	6754
Octubre	4888	4959	7005	5781	6302	6712
Noviembre	5087	5004	7045	5813	6132	6675
Diciembre	5082	5415	7279	5811	5837	6882

	2008	2009	2010
Enero	7011	6954	5627
Febrero	7140	6551	5679
Marzo	7197	6022	5455
Abril	7411	5974	5443
Mayo	7233	6052	
Junio	6958	6033	
Julio	6960	6030	
Agosto	6927	5944	
Septiembre	6814	5543	
Octubre	6757	5416	
Noviembre	6765	5571	
Diciembre	6870	5571	



Asignatura Grupo
Apellidos Nombre
Ejercicio del día



Universidad Autónoma
de Madrid

Asignatura Grupo
Apellidos Nombre
Ejercicio del día

(IMPRESO EN PAPEL RECICLADO)

UNICAMENTE PARA USO ESCOLAR



Instrumentos Estadísticos Avanzados
Facultad Ciencias Económicas y Empresariales
Departamento de Economía Aplicada
Profesor: Santiago de la Fuente Fernández