

Capítulo 9

LA IDENTIFICACIÓN DE LOS POSIBLES MODELOS ARIMA

John Tukey (1915-2000)

Estadístico estadounidense. Uno de los creadores del análisis espectral de series temporales, inventor de la transformada rápida de Fourier. Director fundador del departamento de estadística en Princeton, simultaneó la docencia universitaria con su trabajo en los laboratorios Bell de ATT. Pionero en los estudios de robustez y creador de términos científicos (como software o bit), sus contribuciones han influido mucho en la evolución de la estadística en la segunda mitad del siglo XX.

9.1 Introducción

En las secciones anteriores hemos estudiado las propiedades teóricas de los procesos ARIMA. Vamos a analizar a continuación como ajustar estos modelos a series reales. Box y Jenkins (1976) propusieron realizar este ajuste en tres etapas.

La primera etapa consiste en *identificar* el posible modelo ARIMA que sigue la serie, lo que requiere: (1) decidir qué transformaciones aplicar para convertir la serie observada en una serie estacionaria; (2) determinar un modelo ARMA para la serie estacionaria, es decir, los órdenes p y q de su estructura autorregresiva y de media móvil y, si el proceso es estacional, los órdenes P, Q de la estructura ARMA estacional.

Una vez seleccionado provisionalmente un modelo para la serie estacionaria pasaremos 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 etapa es la de *diagnos*, donde se comprueba que los residuos no tienen estructura de dependencia y representan un proceso de ruido blanco. Si aceptamos que los residuos no contienen información el proceso termina, mientras que si existe, el modelo se modifica para incorporarla y se repiten las tres etapas anteriores.

Estas tres etapas supusieron un avance muy importante, ya que la estimación de los parámetros de un modelo ARMA requería mucho tiempo de cálculo, por lo que era conveniente asegurarse que el modelo a estimar podía ser el adecuado. En la actualidad, la estimación de un modelo ARIMA por máxima verosimilitud es inmediata, por lo que es más simple estimar todos los modelos que consideremos posibles para explicar la serie, y después elegir entre ellos con un criterio de selección, como veremos en el capítulo siguiente. Esta es la filosofía de los criterios automáticos de selección de modelos ARIMA, que funcionan bien en la práctica en muchos casos, y que son imprescindibles cuando se desea modelar y obtener predicciones para un gran conjunto de series. Sin embargo, cuando el número de series a modelar sea pequeño, es conveniente realizar la etapa de identificación que describimos a continuación, para comprender mejor y familiarizarse con la estructura dinámica de las series de interés. El objetivo no es ahora seleccionar un modelo para la serie, sino más bien identificar un conjunto de modelos posibles que sean compatibles con el gráfico de la serie y sus funciones de autocorrelación simple y parcial. Estos modelos serán estimados en la etapa siguiente y seleccionaremos el mejor. En la tercera fase de *diagnos* nos aseguraremos que este mejor modelo no tiene deficiencias claras, como veremos en el Capítulo 11.

9.2 Identificación de la estructura no estacionaria

Supondremos que la serie observada es una realización de un proceso que sigue un modelo ARIMA. La identificación de la estructura no estacionaria consiste en detectar que transformaciones hay que aplicar para conseguir un proceso ARMA estacionario. En particular consiste en determinar si es necesario:

1. Transformar la serie para que tenga varianza constante;
2. Diferenciar la serie para que tenga media constante. Esto equivale a seleccionar el parámetro, d , número de diferencias regulares, que es habitualmente cero, uno o dos;
3. Diferenciar la serie estacionalmente para eliminar la estacionalidad. Cuando la serie sea estacional con período s suele ser necesario aplicar una diferencia $(1 - B^s)$ para convertirla en estacionaria.

Vamos a analizar estos tres aspectos.

9.2.1 Determinación de la transformación para estabilizar la varianza

Típicamente muchas series económicas presentan una variabilidad tanto mayor cuanto mayor sea el nivel de la serie. Por ejemplo, la figura 9.1 presenta la serie de matriculación de automóviles en España y se observa que la variabilidad aumenta aproximadamente de forma lineal con el nivel. Esta impresión visual se confirma haciendo un gráfico de la media cada año y la desviación típica de los datos con relación a la media del año. La figura 9.2 presenta este gráfico y puede verse que la relación entre la desviación típica cada año y la media es lineal. Cuando se de esta relación entre la variabilidad y el nivel de la serie se consigue una serie con variabilidad constante tomando logaritmos. La figura 9.3 presenta la serie de matriculación de vehículos expresada ahora en logaritmos y se observa que con este transformación la variabilidad es aproximadamente constante.

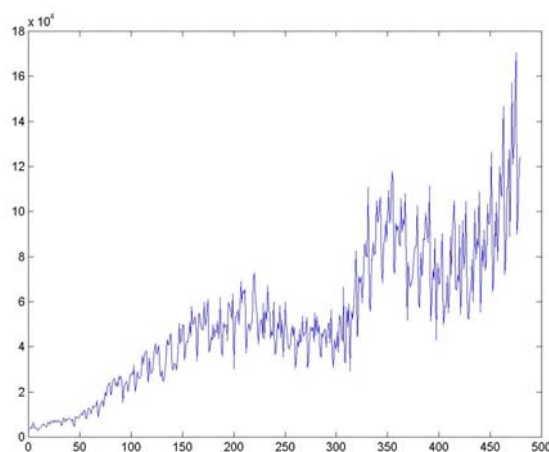


Figura 9.1: Gráfico de la serie de matriculación de automóviles de turismo en España, 1960 a 1999.

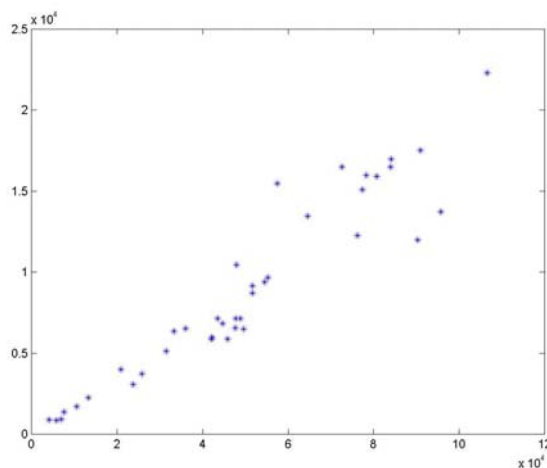


Figura 9.2: Relación entre la desviación típica de las matriculaciones un año dado y la media para ese año.

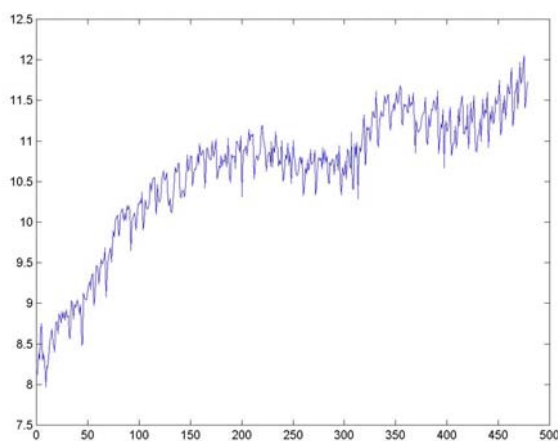


Figura 9.3: La serie de matriculación de vehículos en logaritmos.

En general, puede demostrarse que cuando existe una relación entre las medias, μ_i , de un conjunto de variables aleatorias, x_i , y sus desviaciones típicas, σ_i , del tipo

$$\sigma_i = k\mu_i^\alpha \quad (9.1)$$

al transformar las variables x_i en las variables y_i mediante

$$y_i = \frac{x_i^{1-\alpha} - 1}{1-\alpha} \quad (9.2)$$

se obtienen variables con varianza constante. Esta es la familia Box-Cox de transformaciones con parámetro $\lambda = 1 - \alpha$. Esta familia incluye las transformaciones potenciales y, al escribirla de esta forma, incorpora también el logaritmo. En efecto, cuando α es muy próximo a uno, o λ próximo a cero, puede comprobarse tomando límites en la expresión (9.2) que se obtiene el logaritmo de la variable. La demostración de esta propiedad puede verse en Peña (2001, pag 155).

La transformación necesaria para conseguir variabilidad constante cuando en la serie se observa una relación entre el nivel y la variabilidad puede estimarse haciendo grupos de observaciones consecutivas de la serie, (de 5 o 6 datos o si la serie es estacional con periodo s de s datos para hacer comparaciones homogéneas), calculando en cada grupo la media muestral \bar{x}_i y la desviación típica, s_i , y representando estas variables en un gráfico para estimar α . En particular, tomando logaritmos en (9.1), la pendiente de la regresión

$$\log s_i = c + \alpha \log \bar{x}_i$$

nos estimará el valor de α . Por ejemplo, la figura 9.4 presenta la relación entre estas variables para la serie de matriculaciones y ya se observa que la pendiente es próxima a la unidad. Si estimamos la regresión se obtiene $c = -2.17$ y $\alpha = 1.04$, confirmando la necesidad del logaritmo para estabilizar la varianza de esta serie.

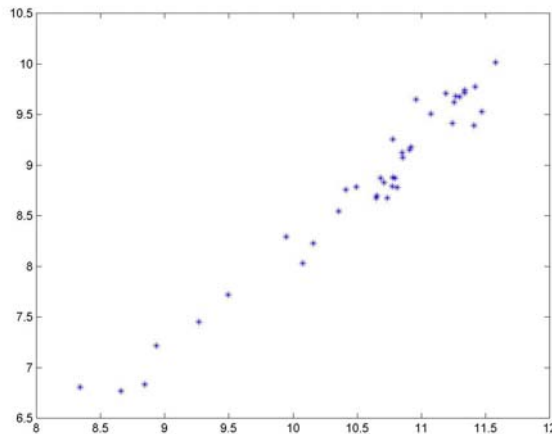


Figura 9.4: Relación entre los logaritmos de las desviaciones típicas y los de las medias para la serie de matriculaciones

El gráfico media-variabilidad es especialmente útil cuando en el de la serie no se capta la variabilidad a simple vista. Por ejemplo, la figura 9.5 representa la serie de paro en España y no es obvio en este gráfico si la variabilidad es constante o no. La figura 9.6 presenta la relación entre la variabilidad y la media, indicando una relación lineal y la necesidad de tomar logaritmos para modelar esta serie.

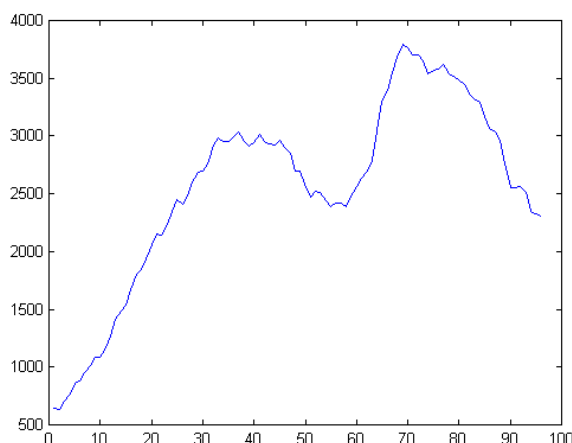


Figura 9.5: Serie de paro en España, desde el primer cuatrimestre de 1977 al cuarto del 2000

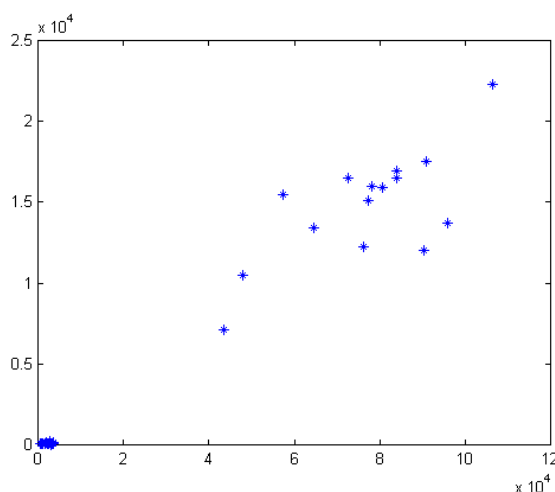


Figura 9.6: Relación enter la desviación típica y la media para la serie de paro en los años 1977 a 2000 con datos cuatrimestrales

La falta de constancia en la variabilidad puede ser el resultado de que la serie se genera como producto (en lugar de la suma) de un componente sistemático o predecible y un término de perturbación. En este caso:

$$z_t = f(z_{t-1}, \dots) u_t \quad (9.3)$$

y aunque u_t tenga una varianza constante producirá una variabilidad en la serie observada z_t tanto mayor cuan mayor sea el valor de la parte sistemática $f(z_{t-1}, \dots)$, que establece el nivel de la serie. Este problema se resuelve tomando logaritmos, ya que entonces, llamando $a_t = \ln u_t$:

$$y_t = \ln z_t = \ln f(z_{t-1}, \dots) + a_t$$

y se obtiene una descomposición aditiva para la variable y_t , que tendrá varianza constante. Si en lugar de (9.3) suponemos un modelo generado con ruidos aditivos y multiplicativo del tipo:

$$z_t = f(z_{t-1}, \dots) u_t + a_t$$

será necesaria una transformación de Box-Cox con λ entre cero y uno para conseguir varianza constante.

En resumen, para obtener una descomposición aditiva entre la parte explicada y no explicada puede ser necesario transformar la serie, y la manifestación de esta necesidad será que la varianza de la serie dependa monotonamente del nivel. Para determinar λ , se forman grupos de entre 6 y 12 observaciones consecutivas, se calcula la desviación típica y la media en estos grupos y se estima una regresión entre los logaritmos de las desviaciones típicas y los de las medias. La pendiente estimada es $1 - \lambda$, donde λ es el parámetro de la transformación Box-Cox necesaria para convertir la serie en estacionaria.

Existen series heterocedásticas donde la varianza cambia sin relación con el nivel. Por ejemplo, aparecen cambios de varianza por tramos, o la varianza en el instante t depende de los valores previos de la serie, lo que se conoce como heterocedasticidad condicional. Estas situaciones no pueden resolverse con estas transformaciones y serán estudiadas posteriormente.

9.2.2 Determinar el orden de diferenciación regular

Si la serie tiene tendencia, o muestra cambios de nivel en la media, la diferenciaremos para transformarla en estacionaria. La necesidad de diferenciar es en muchos casos clara a la vista del gráfico de la serie. Por ejemplo, tanto la serie de matriculación de vehículos de la figura 9.1, como la del paro en la figura 9.5, muestran claramente tendencia por tramos, y media no constante. Nótese que como

$$(1 - B) \ln z_t = \ln z_t - \ln z_{t-1} = \ln \left(1 + \frac{z_t - z_{t-1}}{z_{t-1}} \right) \simeq \frac{z_t - z_{t-1}}{z_{t-1}}$$

donde hemos utilizado que $\ln(1+x) \simeq x$ si x es pequeño, estudiar la serie $\nabla \ln z_t$ equivale a estudiar la serie de tasas de incremento relativo a z_t . Las figuras 9.7 y 9.8 presentan la primera diferencia de estas series. La serie de matriculación de vehículos contiene variaciones muy acusadas mes a mes, de hasta 0.8, es decir del 80% de su valor. Esto ocurre por la presencia de una fuerte estacionalidad, que aparece en el gráfico con unas caídas muy fuertes, que corresponden al mes de Agosto de cada año. Aparte de este efecto, la media es aproximadamente constante. La serie de paro tiene unas variaciones trimestrales mucho menores, pero la figura indica que la serie no tiene media constante. Tomaremos por tanto otra diferencia adicional y la figura 9.9 muestra que con esa segunda diferencia la serie parece tener globalmente una media constante, aunque es visible un efecto estacional que se refleja en una repetición de la pauta de variación cada cuatro datos trimestrales.

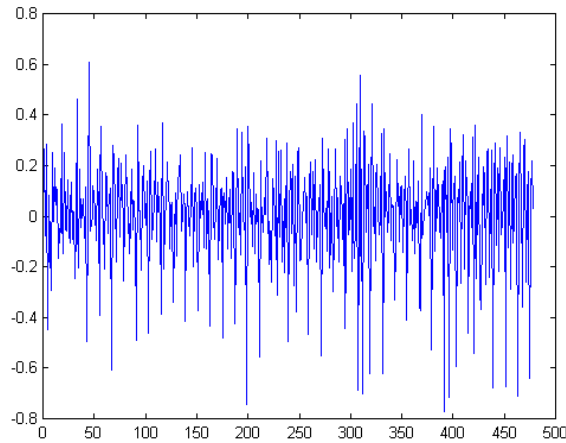


Figura 9.7: Primera diferencia del logaritmo de la serie de matriculaciones

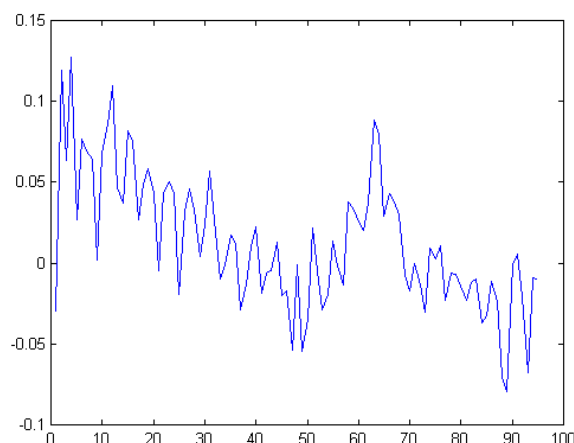


Figura 9.8: Primera diferencia del logaritmo de la serie de paro

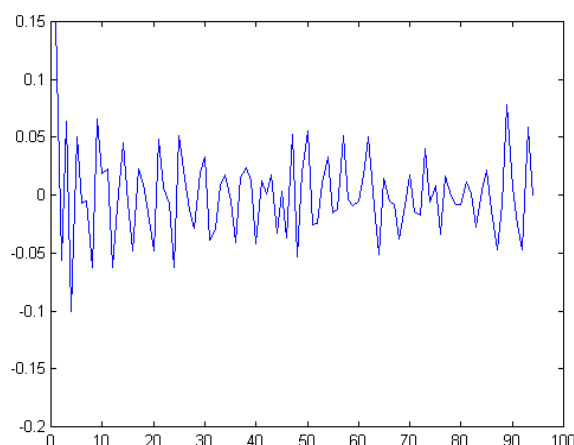


Figura 9.9: Segunda diferencia de la serie del logaritmo del paro

Cuando al analizar el gráfico de la serie la decisión de diferenciar no sea clara, conviene estudiar la función de autocorrelación simple, *fas*, de la serie. Hemos visto que una serie no estacionaria debe mostrar estructura AR positiva con decrecimiento lento y lineal. Por ejemplo, la figura 9.10 presenta la *fas* de la serie de matriculación de vehículos: se observa un decrecimiento lento de los coeficientes de forma lineal, que indica claramente la necesidad de diferenciar.

Es importante señalar que la característica en la *fas* estimada que identifica una serie no estacionaria no es que los coeficientes de autocorrelación sean todos próximos a unidad, sino que decrezcan lentamente y de forma lineal. Puede demostrarse que el valor esperado de los coeficientes de autocorrelación estimados depende del proceso y del tamaño muestral, pudiendo ser mucho menor que la unidad. Por ejemplo, en el proceso IMA(1, 1) si $\theta = 0,8$ y $T = 100$ el valor esperado de los coeficientes de autocorrelación muestral es siempre menor que 0.5. Sin embargo, sigue siendo esperable un decrecimiento muy suave y lineal. En resumen, si la *fas* no se amortigua para retardos altos (mayores de 15 ó 20) es necesario diferenciar para obtener un proceso estacionario.

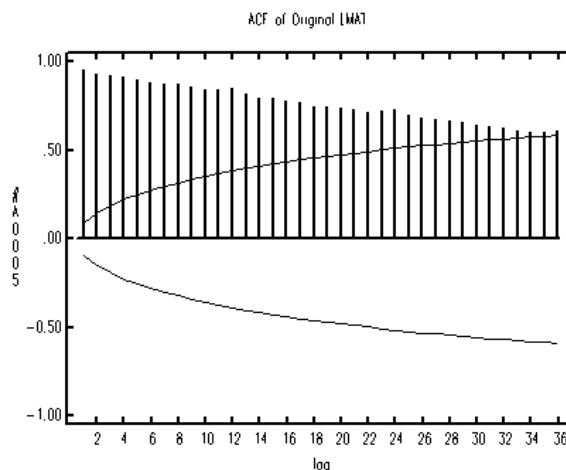


Figura 9.10: Función de autocorrelación para la serie de matriculación de vehículos

Es posible realizar un test para decidir en caso de duda si conviene o no diferenciar la serie. Si el objetivo es la predicción estos contraste no son muy útiles, porque, en caso de duda, conviene siempre diferenciar, ya que las consecuencias negativas de sobrediferenciar son mucho menores que las de infradiferenciar. En efecto, si suponemos que la serie es estacionaria cuando no lo es, los errores de predicción a medio plazo pueden ser enormes, ya que según hemos visto en el capítulo 8 la predicción de una serie estacionaria a medio plazo es su media, mientras que una serie no estacionaria puede alejarse indefinidamente de dicho valor y el error de predicción no está acotado. El modelo no diferenciado será, además, poco robusto y sin capacidad de adaptación ante valores futuros. Por el contrario, si sobrediferenciamos tenemos siempre predictores adaptativos y, aunque perdamos precisión, esta pérdida está acotada. Incluir en el modelo una diferencia de más implica introducir artificialmente un término de media móvil que será no invertible. Como tendremos que estimar un parámetro adicional, perderemos algo de eficiencia en la estimación, pero, como veremos en el capítulo siguiente, los estimadores de todos los parámetros serán consistentes, y el efecto de sobrediferenciar sobre la predicción será pequeño, aunque calcularemos por exceso la incertidumbre de las predicciones. Una razón adicional de tomar diferencias en caso de duda es que puede demostrarse (véase Sánchez y Peña, 2002) que aunque la serie sea estacionaria si el parámetro AR está cerca de la unidad se obtienen mejores predicciones sobrediferenciando la serie que utilizando el modelo verdadero estacionario. Cuando el objetivo no sea prever, sino determinar si una variable es o no estacionaria, situación que aparece con frecuencia en aplicaciones económicas, pueden utilizarse los contraste de raíces unitarias que estudiaremos en la sección siguiente.

En general, el grado de diferenciación requerido para convertir en estacionarias series reales es cero, uno o dos.

9.2.3 Determinar la diferenciación estacional

Si la serie tiene componente estacional habrá que aplicar una diferencia estacional, $\nabla_s = 1 - B^s$, para conseguir una serie estacionaria. La estacionalidad se manifiesta:

1. En el gráfico de la serie, que presentará una pauta repetida de periodo s (por ejemplo, con datos mensuales Agosto puede estar siempre por encima de la media del año y Enero por debajo).
2. En la función de autocorrelación simple, que presentará coeficientes positivos que decrecen lentamente en los retardos $s, 2s, 3s, \dots$

Por ejemplo, la figura 9.1 sugiere una pauta estacional porque el valor de la serie es sistemáticamente menor en Agosto. Este efecto es más claramente visible al eliminar la tendencia tomando una diferencia

regular, como ilustra la figura 9.7. Además del gráfico de la serie es conveniente mirar la fas, ya que una serie con estacionalidad mostrará valores de autocorrelación positivos y altos en los retardos estacionales. Por ejemplo, la figura 9.11 muestra el resultado de aplicar una diferencia regular al logartimo de la serie de matriculación de vehículos. Se observa que existen correlaciones altas y persistentes en los retardos estacionales, 12, 24 y 36. Esto sugiere la necesidad de una diferencia estacional para obtener una serie estacionaria.

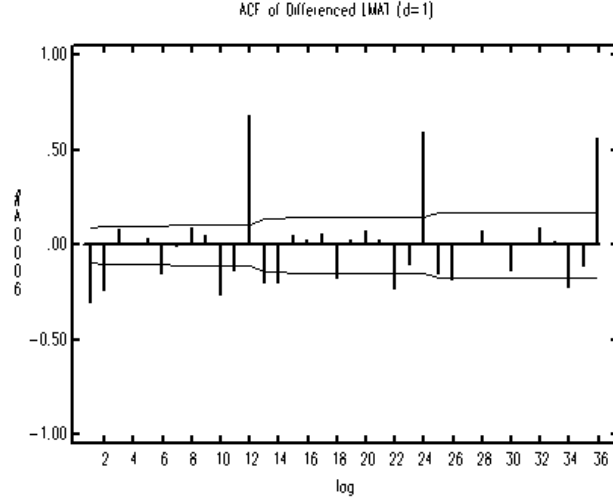


Figura 9.11: Fas de la serie de matriculación en logaritmos despues de una diferencia regular.

Ocurre algo similar en la serie de paro donde el componente estacional sólo comienza a apreciarse con claridad en la serie con dos diferencias (véase la figura 9.9). Los comentarios respecto a las diferencias regulares son también válidos para las estacionales: en caso de duda es mejor diferenciar.

En principio sería teóricamente posible que una serie requiera más de una diferencia estacional, pero en la práctica esto no ocurre, por lo que la decisión se reducirá a aplicar o no una diferencia estacional.

9.3 La identificación de la estructura ARMA

La identificación de los órdenes p y q del modelo ARMA se realiza comparando las funciones de autocorrelación simple y parcial calculadas en la serie observada con el catálogo de funciones teóricas de procesos ARMA estudiadas.

Llamando ω_t , a la serie estacionaria, $\omega_t = \nabla^d \nabla_s^D z_t$, donde en la práctica d toma valores en $(0, 1, 2)$ y D en $(0, 1)$, los coeficientes de autocorrelación muestrales se calculan por:

$$r_k = \frac{\sum_{t=d+sD+1}^{T-k} (\omega_t - \bar{\omega}) (\omega_{t+k} - \bar{\omega})}{\sum_{t=d+sD+1}^T (\omega_t - \bar{\omega})^2}, \quad k = 1, 2, \dots \quad (9.4)$$

donde hemos tenido en cuenta que al diferenciar perderemos $d + sD$ observaciones.

Para juzgar cuándo un coeficiente r_k es distinto de cero necesitamos su error estándar, cuya determinación depende de la estructura del proceso. Una solución simple es tomar como error estándar $1/\sqrt{T}$, que es aproximadamente –según estudiamos en el capítulo 3– el error estándar de un coeficiente de correlación entre variables independientes. Si *todos* los coeficientes teóricos de autocorrelación fuesen nulos, las desviaciones típicas de estimación serían aproximadamente $1/\sqrt{T}$. Por tanto, podemos trazar bandas de confianza a $\pm 2/\sqrt{T}$ y considerar como significativos en primera aproximación a los coeficientes fuera de dichas bandas.

Los coeficientes de autocorrelación parcial se obtienen con las regresiones:

$$\tilde{\omega}_t = \alpha_{k1}\tilde{\omega}_{t-1} + \dots + \alpha_{kk}\tilde{\omega}_{t-k}$$

donde $\tilde{\omega}_t = \omega_t - \bar{\omega}$. La secuencia $\hat{\alpha}_{kk}$ ($k = 1, 2, \dots$) de coeficientes minimocuadráticos es la función de autocorrelación parcial. Utilizaremos siempre en los gráficos de la *fap* los límites asintóticos $\pm 2/\sqrt{T}$, con la salvedad de que se trata de unos límites de referencia aproximados.

Si el proceso es estacional, estudiaremos los coeficientes de la *fas* y *fap* en los retardos $s, 2s, 3s, \dots$, para determinar la estructura ARMA estacional.

La identificación del modelo ARMA es una tarea difícil, pero con la potencia de los ordenadores actuales sólo necesitamos restringir el conjunto de todos los posibles modelos ARMA a un subconjunto que parezca, en principio, adecuado para representar los rasgos principales de la serie. En la práctica la mayoría de las series reales pueden aproximarse bien con modelos ARMA con p y q menores que tres, para series no estacionales, y con P y Q menores que 2, para series estacionales. Los procedimientos automáticos, como el implantado en el programa TRAMO, estiman todos los modelos posibles y seleccionan el mejor con una criterio de selección de modelos, como veremos en el capítulo siguiente. Alternativamente, podemos utilizar la función de autocorrelación simple y parcial para seleccionar un subconjunto pequeño de modelos que sea coherente con la estructura observada. Para ello, conviene:

1. Decidir cuál es el orden máximo de las partes AR y MA a partir de los rasgos obvios de la *fas* y la *fap*: coeficientes claramente significativos, pautas de decrecimiento geométricas o sinusoidales, etc.
2. Evitar la identificación inicial de modelos mixtos ARMA y comenzar con modelos AR o MA preferiblemente de orden bajo.
3. Utilizar las interacciones alrededor de los retardos estacionales para confirmar la concordancia entre la parte regular y la estacional.

Además de los órdenes p, q, P, Q del modelo debemos comprobar en esta etapa la serie estacionaria ω_t , tiene media distinta de cero. Para ello se calcula la media muestral

$$\bar{\omega} = \frac{\sum \omega_t}{T_c}$$

donde T_c es el número de sumandos (normalmente $T_c = T - d - sD$), y su desviación típica, que es, aproximadamente:

$$s(\bar{\omega}) \simeq \frac{s_\omega}{\sqrt{T}} (1 + 2r_1 + \dots + 2r_k)^{1/2}$$

siendo s_ω la desviación típica de ω_t , y suponiendo que los primeros k coeficientes autocorrelación son significativos. Si $\bar{\omega} \geq 2s(\bar{\omega})$ admitiremos que la media distinta de cero y la incluiremos como un parámetro a estimar; en caso contrario supondremos que $E[\omega_t] = 0$.

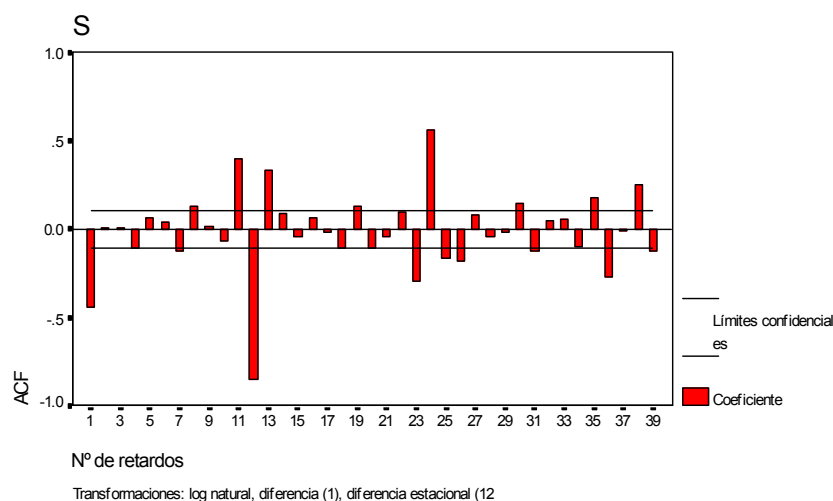


Figura 9.12: Función de autocorrelación simple de la serie de matriculaciones en logaritmos y con una diferencia regular y otra estacional

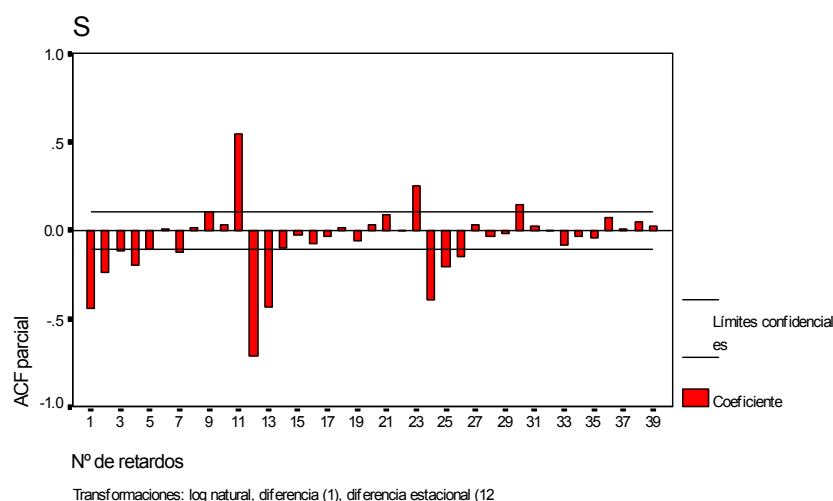


Figura 9.13: Función de autocorrelación parcial para la serie de matriculación de vehículos en logaritmo y con una diferencia regular y otra estacional

Como ilustración del proceso de identificación consideraremos la serie de matriculación de vehículos. Su *fas*, que se presenta en la figura 9.12 muestra como rasgos más destacados: (1) un coeficiente r_1 significativo; (2) coeficientes significativos en los retardos estacionales, r_{12} , r_{24} y r_{36} ; (3) interacción alrededor de los retardos estacionales. Además de estos rasgos destacados, hay otros coeficientes significativos de los que prescindiremos en primera estancia. La parte regular sugiere un modelo MA(1), o como mucho MA(2). La parte estacional es más compleja, ya que la estructura observada es compatible con un AR(1)₁₂ con coeficiente negativo y también con AR más largos o ARMA(1,1). El problema es que sólo tenemos tres coeficientes para obtener la estructura. La *fap* de esta serie, que se presenta en la figura 9.13, confirma la estructura MA(1) para la parte regular y es compatible con un AR(2) o ARMA(1,1) para la parte estacional. Por tanto, pasaremos a estimar modelos con parte MA(1) para la parte regular y AR(2) o ARMA(1,1) para

la parte estacional.

9.4 Contrastes de raíces unitarias

Si se desea decidir si un proceso es o no estacionario podemos realizar un contraste de raíces unitarias, como describimos a continuación. Estos contrastes permiten decidir si debemos o no tomar una diferencia adicional en una serie para convertirla en estacionaria. Un procedimiento alternativo de tomar esta decisión es utilizar los criterios de selección de modelos que presentaremos en el capítulo siguiente.

9.4.1 Contraste de Dickey-Fuller

Presentaremos inicialmente el contraste en el caso más simple. Supongamos que se desea decidir entre el proceso no estacionario

$$\nabla z_t = a_t \quad (9.5)$$

y el estacionario

$$(1 - \rho B) z_t = c + a_t \quad (9.6)$$

El test de Dickey-Fuller se ha desarrollado porque los procedimientos tradicionales para elegir entre ambos modelos no son aplicables en este caso. En efecto, si estimamos ambos modelos y escogemos el de menor varianza es fácil concluir que el modelo adecuado es el (9.6), ya que en (9.5) no estimamos ningún parámetro mientras que en (9.6) el parámetro se estima para que la varianza de los residuos sea mínima. Puede demostrarse que si generamos muchas muestras con (9.5) y estimamos en cada una un modelo como el (9.6), dos tercios de las veces este modelo estimado tendrá una varianza residual menor que la obtenida suponiendo que los datos siguen un paseo aleatorio y tomando una diferencia. Como ilustración, la figura 9.14 presenta la distribución del estimador de mínimos cuadrados del parámetro ρ en muestras de tamaño 100 que siguen un paseo aleatorio. La distribución se ha obtenido generando 20 000 replicaciones del paseo aleatorio y estimando en cada una de ellas el parámetro suponiendo que el proceso es AR(1). Se han considerado dos estimadores. En el caso (a) se ha supuesto que la media de la serie es cero, en cuyo caso en (9.6) la constante c es cero, y el parámetro se estima mediante $\hat{\rho} = \sum z_t z_{t-1} / \sum z_t^2$; en el caso (b), más realista en la práctica, no se supone que la media sea cero y por tanto el parámetro se estima por $\hat{\rho} = \sum (z_t - \bar{z})(z_{t-1} - \bar{z}) / \sum (z_t - \bar{z})^2$. Se observa que la distribución es muy asimétrica y que, especialmente con el segundo estimador, podemos obtener valores del parámetro bastante menores que uno. Por tanto, comparar las varianzas de ambos modelos no es un buen método para elegir entre ellos, y, especialmente, si queremos cubrirnos respecto al error de no diferenciar.

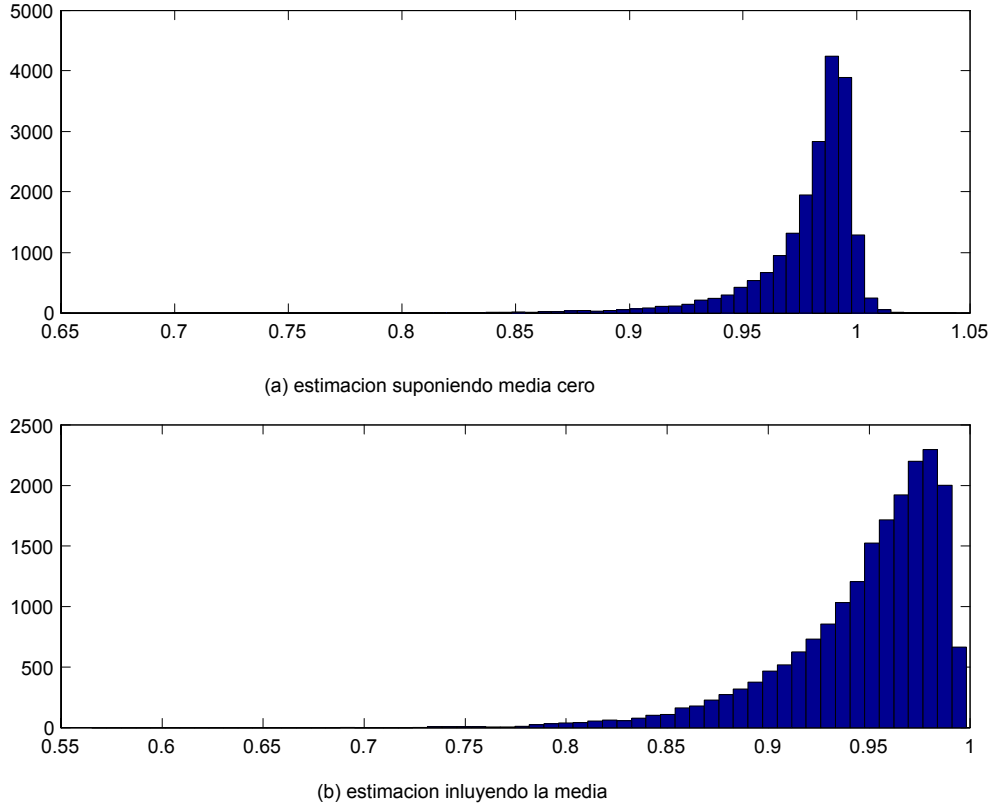


Figura 9.14: Distribución de la estimación del parámetro de un AR(1) cuando los datos se generan por un paseo aleatorio. En el caso se supone que la media es cero y en el (b) se estima con la media muestral.

Podría pensarse que la forma de decidir entre los dos modelos es estimar el modelo (9.6) por mínimos cuadrados y contrastar si el coeficiente ρ es uno, comparando el estimador con su desviación típica estimada. El contraste sería $H_0 : \rho = 1$ frente a la alternativa $H_1 : \rho < 1$, y el estadístico para el contraste:

$$t_\mu = \frac{\hat{\rho} - 1}{\hat{s}_\rho} \quad (9.7)$$

donde \hat{s}_ρ es la desviación típica estimada del estimador de mínimos cuadrados. Este método no es correcto. El problema es que si el proceso es no estacionario tanto la estimación minimicuadrática como la de MV exacta proporcionan un valor erróneo de la varianza del estimador $\hat{\rho}$ y el estadístico (9.7) no sigue una distribución t de Student. En efecto, en este caso si la hipótesis nula es cierta el valor del parámetro se encuentra en el extremo del intervalo (0,1) y las condiciones de regularidad necesarias para las propiedades asintóticas del estimador MV no se aplican. La figura 9.15 presenta el resultado de calcular el estadístico (9.7) en 20 000 muestras de tamaño 100 generadas como paseos aleatorios. Se observa que la distribución difiere mucho de la t de Student. Por tanto, si utilizamos los valores críticos de la t de Student y rechazamos la hipótesis $\rho = 1$ cuando el valor de (9.7) es mayor de 2 o 2.5 tendremos una probabilidad alta de rechazar que el proceso es no estacionario cuando realmente lo es.

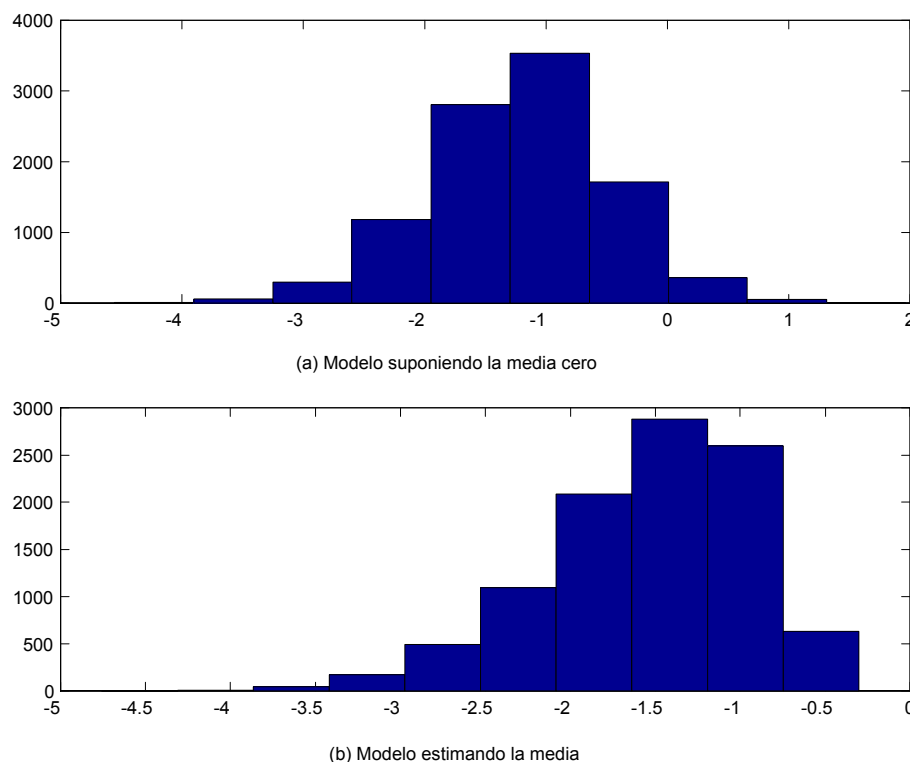


Figura 9.15: Distribución del estadístico t cuando se contrasta que el parámetro es uno en muestras que siguen paseos aleatorios

El test de Dickey-Fuller se basa en los resultados anteriores. Toma como hipótesis nula, H_0 , que es necesario diferenciar, $\rho = 1$, y contrasta si esta hipótesis puede rechazarse a la vista de los datos utilizando la verdadera distribución del contraste indicada en la figura 9.15. Una forma simple de obtener el estadístico (9.7) para el contraste es escribir el modelo a estimar como:

$$\nabla z_t = c + \alpha z_{t-1} + a \quad (9.8)$$

donde $\alpha = \rho - 1$. A este modelo se le conoce como *el modelo de corrección de error*. Si $\alpha = 0$ tenemos un paseo aleatorio y si $\alpha \neq 0$ un proceso estacionario. El contraste consiste en estimar el parámetro α en (9.8) por mínimos cuadrados y rechazar que el proceso es estacionario si el valor t_μ es significativamente pequeño. El contraste (9.7) se escribe ahora:

$$t_\mu = \frac{\hat{\alpha}}{\hat{s}_\alpha} \quad (9.9)$$

donde $\hat{\alpha}$ es la estimación minimocuadrática de α en (9.8) y \hat{s}_α su desviación típica estimada calculada de la forma habitual. Bajo la hipótesis de que $\alpha = 0$ (que implica $\rho = 1$) la distribución de t_μ ha sido tabulada por Dickey y Fuller. Un extracto de sus tablas se presentan en el tabla 9.1. Observemos que la estimación de $\hat{\alpha}$ será negativa ya que $\rho \leq 1$. Según indica la figura 9.15 la distribución del estadístico es distinta cuando incluimos o no una constante en el modelo y, para mayor seguridad, conviene siempre incluirla.

Recomendamos realizar el contraste con nivel de significación pequeño, .01, de manera que sólo supongamos que el proceso es estacionario cuando realmente lo sea. Como hemos comentado, incluso aunque el verdadero proceso sea AR(1) estacionario podemos obtener mejores predicciones con el paseo aleatorio si el parámetro está próximo a la unidad.

T	sin constante				con constante			
	.01	.025	.05	.1	.01	.025	.05	.1
25	-2.66	-2.26	-1.95	-1.60	-3.75	-3.33	-3.00	-2.63
50	-2.62	-2.25	-1.95	-1.61	-3.58	-3.22	-2.93	-2.60
100	-2.60	-2.24	-1.95	-1.61	-3.51	-3.17	-2.89	-2.58
250	-2.58	-2.23	-1.95	-1.62	-3.46	-3.14	-2.88	-2.57
500	-2.58	-2.23	-1.95	-1.62	-3.44	-3.13	-2.87	-2.57
∞	-2.58	-2.23	-1.95	-1.62	-3.43	-3.12	-2.86	-2.57

Tabla 9.1: Valores críticos para el contraste de Dickey Fuller de raíz unitaria

9.4.2 EL Contraste de Dickey Fuller aumentado

Este contraste se generaliza para el caso en que hemos diferenciado d veces, con $d \geq 1$, una serie z_t que sigue un proceso AR y dudamos si es necesario diferenciar o no una vez más. Sea $z_t = \nabla^d z_t$, ($d \geq 0$). Supondremos que la elección se plantea entre los modelos:

$$H_0 : \phi_p(B) \nabla (z_t - \mu) = a_t \quad (9.10)$$

$$H_1 : \phi_{p+1}(B) (z_t - \mu) = a_t \quad (9.11)$$

es decir, tenemos dudas sobre la raíz mayor de un $AR(p+1)$ que puede ser igual a uno, como en el modelo (9.10) con lo que el proceso será no estacionario, o menor que uno, como en (9.11) y el proceso es estacionario.

Podríamos estimar directamente el modelo $AR(p+1)$, calcular después las raíces del polinomio AR y ver si la mayor raíz es suficientemente próxima a uno. Sin embargo, para realizar un contraste nos encontramos con el problema que de esta manera obtenemos los errores estándar de los parámetros estimados, pero no el error de estimación de la raíz mayor del operador autorregresivo. Podríamos intentar obtener este error estándar como función de los errores estándar de los coeficientes estimados, pero el proceso sería complicado. Un enfoque más simple es intentar contrastar directamente la condición de una raíz unitaria en el operador $\phi_{p+1}(B)$.

Para implementar esta idea, se descompone el operador $\phi_{p+1}(B)$ como

$$\phi_{p+1}(B) = (1 - \alpha_0 B) - (\alpha_1 B + \dots + \alpha_p B^p) \nabla \quad (9.12)$$

Esta descomposición siempre puede hacerse, porque en ambos lados de la igualdad tenemos un polinomio en B de orden $p+1$ con $p+1$ coeficientes, con lo que identificando potencias de B en ambos miembros podemos obtener los $p+1$ coeficientes $\alpha_0, \dots, \alpha_p$ dados $\phi_1, \dots, \phi_{p+1}$. Vamos a ver que esta descomposición tiene la ventaja de trasladar la condición de una raíz unitaria en el primer miembro a una condición sobre un coeficiente que podemos estimaren el segundo miembro. En efecto, si $\phi_{p+1}(B)$ tiene una raíz unidad, entonces $\phi_{p+1}(1) = 0$, y si hacemos $B = 1$ en (9.12) el término $(\alpha_2 B + \dots + \alpha_{p+1} B^p)(1 - B)$ se anulará y tendrá que verificarse que $(1 - \alpha_0) = 0$, es decir $\alpha_0 = 1$. Por otro lado, si $\alpha_0 = 1$, podemos escribir el polinomio de la izquierda como $(1 - \alpha_1 B - \dots - \alpha_p B^p) \nabla$, y el polinomio tiene una raíz unitaria. Por tanto, hemos comprobado que la condición equivalente a que el polinomio $\phi_{p+1}(B)$ tenga una raíz unidad es que el coeficiente α_0 en la descomposición (9.12) sea uno.

Otra manera de comprobar esta equivalencia es obtener los coeficientes α_i como función de los ϕ igualando potencias en ambos polinomios. Los coeficientes de la derecha son

$$\phi_{p+1}(B) = 1 - \phi_1 B - \dots - \phi_{p+1} B^{p+1} \quad (9.13)$$

y los de la izquierda (9.12) :

$$\phi_{p+1}(B) = (1 - (\alpha_0 + \alpha_1)B - (\alpha_2 - \alpha_1)B^2 - \dots - (\alpha_p - \alpha_{p-1})B^p - (-\alpha_p)B^{p+1}) \quad (9.14)$$

e igualando coeficientes entre (9.13) y (9.14) se obtiene $\alpha_p = -\phi_{p+1}$, $\alpha_{p-1} = -\phi_{p+1} - \phi_p$, y en general:

$$\alpha_i = - \sum_{j=i+1}^{p+1} \phi_j ; \quad i \geq 1$$

y

$$\alpha_0 = \phi_1 + \dots + \phi_{p+1} \quad (9.15)$$

que confirma el resultado que habíamos obtenido antes: la condición $(1 - \phi_1 - \dots - \phi_{p+1}) = 0$, implica la condición $\alpha_0 = 1$.

El modelo (9.11) puede escribirse utilizando (9.12)

$$\phi_{p+1}(B)(z_t - \mu) = (1 - \alpha_0 B)(z_t - \mu) - (\alpha_1 B + \dots + \alpha_p B^p) \nabla z_t = a_t \quad (9.16)$$

donde hemos utilizado que $\nabla z_t = \nabla(z_t - \mu)$. Entonces:

$$(z_t - \mu) = \alpha_0(z_{t-1} - \mu) + \sum_{i=1}^p \alpha_i \nabla z_{t-i}$$

Para obtener directamente el estadístico del contraste podemos, como en el caso anterior, restar $(z_{t-1} - \mu)$ de ambos miembros y escribir el modelo en la *forma de corrección de error*, como:

$$\nabla z_t = c + \alpha z_{t-1} + \alpha_1 \nabla z_{t-1} + \dots + \alpha_p \nabla z_{t-p} + a_t \quad (9.17)$$

donde $\alpha = (\alpha_0 - 1) = -(1 - \phi_1 - \dots - \phi_p)$ y $c = (1 - \alpha_0)\mu$. Esta ecuación (9.17) puede estimarse por mínimos cuadrados y el contraste de que la serie tiene una raíz unitaria, $\alpha_0 = 1$ es equivalente al contraste $\alpha = 0$.

Llamando

$$t_\mu = \frac{\hat{\alpha}}{\hat{s}_\alpha} \quad (9.18)$$

donde $\hat{\alpha}$ es la estimación minimocuadrática de α en (9.17) y \hat{s}_α su desviación típica estimada calculada de la forma habitual. La distribución de t_μ cuando la hipótesis de una raíz unitaria, $\rho = 1$ es cierta ha sido tabulada por Dickey y Fuller. De nuevo la distribución depende de que tengamos o no constante en el modelo, pero no es necesario corregir por la inclusión de los retardos adicionales, por lo que podemos utilizar la tabla 1.

Este contraste puede también aplicarse para modelos ARMA. Supongamos que el modelo es $\pi(B)z_t = a_t$, donde $\pi(B) = \phi(B)\theta^{-1}(B)$ y queremos contrastar si una de las raíces de $\pi(B)$ es igual a la unidad, en cuyo caso podemos escribir $\pi(B) = \pi^*(B)\nabla$. Aproximando el operador $\pi(B)$, que puede ser infinito si existe parte MA, por un $AR(k)$ largo, escribiremos $\pi(B) \simeq \phi_k(B)$. La descomposición anterior donde factorizamos un operador extrayendo la raíz unidad es un caso particular de descomposición de cualquier polinomio $a(B) = (1 - \alpha_1 B - \dots)$ como:

$$a(B) = a(1)B + \nabla a^*(B) \quad (9.19)$$

donde si $a(B)$ es infinito $a^*(B) = (1 - \alpha_1^* B - \dots)$ también lo es, y si $a(B)$ es de orden k , $a^*(B)$ es de orden $k - 1$. Los coeficientes de $a^*(B)$ se obtienen a partir de los de $a(B)$ igualando potencias en ambos polinomios y puede comprobarse que:

$$a_j^* = - \sum_{i=j+1}^p a_i ; \quad j \geq 1$$

y utilizando (9.19) podemos escribir el modelo $\pi(B)z_t = a_t$ como

$$\pi(1)z_{t-1} + \pi^*(B)\nabla z_t = a_t$$

o también, llamando $\alpha = -\pi(1)$

$$\nabla z_t = \alpha z_{t-1} + \pi_1^* \nabla z_{t-1} + \dots + \pi_p^* \nabla z_{t-p} + a_t$$

que es análogo al modelo anterior.

En series estacionales hay que tener cuidado en introducir todos los retardos necesarios para que el ruido de la regresión sea ruido blanco al efectuar el contraste.

Como ilustración, vamos a realizar un contraste de raíz unitaria una vez tomada la diferencia estacional en la serie del logartimo de matriculación de vehículos. El gráfico de la serie 9.16 muestra que la serie no tiene nivel constante, pero ya no muestra ninguna tendencia clara. La fas de esta serie, dada en la figura 9.17 muestra también muchos coeficientes positivos y decreciendo lentamente, sugiriendo la necesidad de una diferencia.

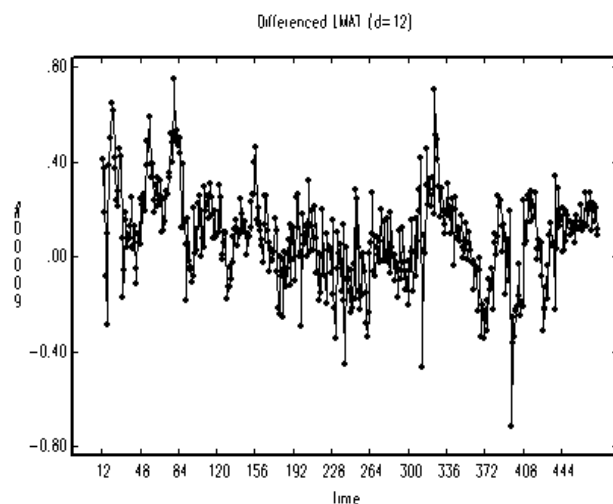


Figura 9.16: Grafico de una diferencia estacional de orden 12 en el logartimo de la serie de matriculación de vehículos

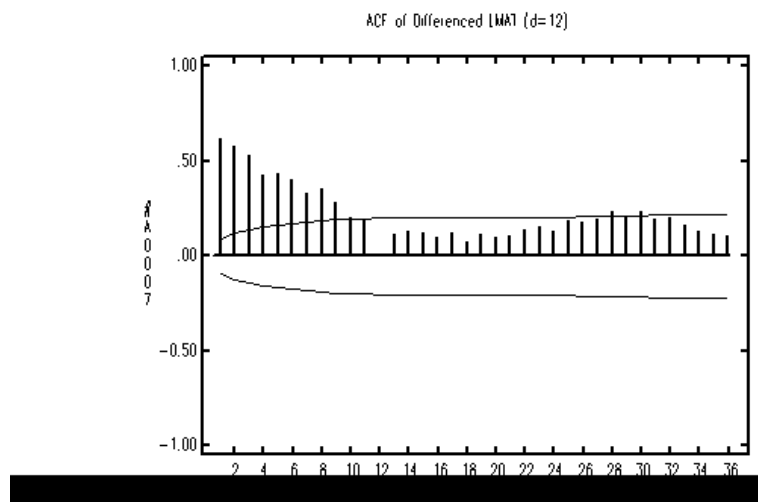


Figura 9.17: Fas de la diferencia de orden 12 de la serie de matriculación de vehiculos en logartimos

Como ilustración vamo a realizar un contraste de raíz unitaria mediante el programa SCA. Supongamos que nos quedamos cortos en el número de retardos y a la vista de la figura 9.17, decidimos introducir sólo los nueve retardos donde observamos coeficientes significativos al principio del gráfico. Entonces el resultado del contraste es

AUGMENTED DICKEY-FULLER TEST WITH CONSTANT

TEST	0.010	0.050	0.100	0.050		
ORDER STATISTIC LEVEL	LEVEL	LEVEL	LEVEL	LEVEL	UROOT	
10	-4.97	-3.44	-2.87	-2.57	-2.87	NO

concluiríamos que no hay una raíz unitaria. Sin embargo, si miramos a la fap vemos que existen coeficientes significativos hasta el retardo 36 e introduciendo hasta ese retardo el resultado del contraste es:

TEST	0.010	0.050	0.100	0.050		
ORDER STATISTIC LEVEL	LEVEL	LEVEL	LEVEL	LEVEL	UROOT	
36	-2.37	-3.44	-2.87	2.57	-2.87	YES

con lo que concluimos que claramente hace falta tomar otra diferencia para convertir la serie en estacionaria.

1. Ejercicios

9.1 Un criterio antiguo para determinar el número de diferencias es el criterio de Titner, que consiste en diferenciar la serie hasta que su varianza aumente al diferenciarla. Demostrar que si partimos de una serie estacionaria pero con autocorrelación de primer orden mayor que .5 la varianza de la serie disminuye al diferenciarla. (Sugerencia, utilice que llamando x_t a la serie original y $n_t = \nabla x_t$, entonces $Var(n_t) = 2\sigma_x^2(1 - \rho_1)$.)

9.2 Identificar un modelo para la serie airline.dat. Esta es la serie que da nombre al modelo de pasajeros de avión.

9.3 Identificar un modelo para la serie accidentes en jornada laboral del fichero accidentes2.dat

9.4 Justificar que en una muestra de 100 observaciones datos generados por el modelo $(1 - .2B)z_t = a_t$ pueden igualmente identificarse como generados por un AR(1), MA(1) o ARMA (1,1) (Sugerencia, exprese el modelo como MA(1) y tenga en cuenta que los límites de los coeficientes de la fas y fap son $T^{-1/2}$).