TEMA 11

HACIA LA AUTOMATIZACIÓN DE LA MODELIZACIÓN DE LAS SERIES TEMPORALES

TRAMO / SEATS

Valencia, mayo 2005

Bernardí Cabrer Borrás
(Análisis Económico)

1. HACIA LA MODELIZACIÓN AUTOMÁTICA DE LAS SERIES TEMPORALES: TRAMO (Time series Regression with ARIMA, Missing observation and Outliers).

Una de las limitaciones en el estudio de series temporales mediante la modelización ARIMA, y en concreto a través del enfoque Box-Jenkins, es la dificultad de identificar correctamente el modelo a partir de una muestra y en su caso la selección del modelo más adecuado. Este obstáculo se ha intentado superar mediante la identificación, estimación y predicción de forma automática de los modelos tipo ARIMA. El proceso se realiza mediante de una batería de test aplicados de forma consecutiva con el fin último de estudiar las pautas de comportamiento y efectuar predicciones de variable objeto de estudio. Otra de las limitaciones en la modelización Box-Jenkins consiste en ignorar el tratamiento previo de las series temporales o análisis descriptivo. Esto es el estudio de la estacionariedad, estacionalidad, efectos calendarios y out-liers. El ignorar la existencia de out-liers en una serie puede acarrear problemas de cara a la identificación correcta del modelo.

Con el fin de superar las dificultades de la identificación de los modelos ARIMA se han ideado procedimientos y programas computacionales que de forma automática seleccionan el modelo más adecuado. Uno de los programas más utilizados[1] en la especificación y estimación de los modelos ARIMA con procedimiento automatizado es el TRAMO desarrollado por Agustín Maravall y Victor Gomez. (1994). Pero además las series temporales económicas se ven sometidas a un conjunto de acontecimientos que son de fácil detección si se realiza un estudio minucioso. Así por ejemplo si se analiza el Consumo de Energía Eléctrica (CEE) mensual de la Comunidad Valencia, va a tomar valores menores en los meses que se contabilice un número menor de días laborables, mientras que en los meses de un número mayor de días laborables el valor que va a tomar la serie en este mes va a ser mas elevado. Pero además algunas veces hay acontecimientos extraordinarios (huelgas, fallos en el suministro de energía eléctrica, olas de calor, olas de frío, etc.) que provocan que el valor del CEE en el mes que ocurre el acontecimiento extraordinario tome valores anómalos (outliers). Desde el punto de vista informático, cabe destacar que en la actualidad existe una versión del TRAMO desarrollada en entorno Windows. El nombre de dicho programa es TSW[2], y ha sido elaborado por G.Caporello y A.Maravall (2001).. Así mismo el programa Eviews-4 ha incorporado una subrutina que permite tratar las series a través de lo procedimiento propuesto por Maravall y Victor Gomez. (1994).

El esquema general del tratamiento de una serie temporal mediante el procedimiento TRAMO es la siguiente:

[1] Tanto el procedimiento TRAMO (estudio y predicción de series temporales) como el SEATS (obtención de la señal ciclo-tendencia de la serie temporal) se utilizan de forma oficial para el tratamiento de series temporales en las siguientes instituciones: EUROSTAT, European Central Bank, Banco de España, INE, Ministerio de Economía y Hacienda, Agencia Tributaria , Banco de Santander Central Hispano, etc.

[2] La versión original del TRAMO desarrollada por Agustín Maravall et al. se ofrece de forma gratuita en la página WEB del Banco de España (***http://www.bde.es***)

VISUALIZACIÓN DE LA SERIE TEMPORAL
(Periodicidad de los datos MQ=4 datos trimestrales; MQ=12 datos mensuales)

TRATAMIENTO DE LAS OBSERVACIONES ANÓMALAS (MISSING DATA)

TRATAMIENTO DEL NÚMERO DE DIAS LABORABLES "TRADING DAIS"
(ITRAD =0 no se considera el efecto días laborables y festivos
ITRAD =1 se consideran días laborables de lunes a viernes y festivos sábados y domingos
ITRAD =2 idem mas la duración de los meses
ITRAD =3 computa el efecto día como efecto diferencial del número de días de cada tipo

ESTUDIO DEL EFECTO PASCUA (SEMANA SANTA)
(IEAST =0 no existe efecto Pascua, IEAST =1 existe efecto Pascua;
IEAST =-1 el programa determina la significatividad del efecto)
[IDUR= número de días de duración del Efecto Pascua]

TRATAMIENTO DE LOS OUTLIERS
AO (Aditivo) IO (Innovación) LS (Cambio de nivel) TC (Cambio Transitorio)
[IATIP = 1

SIGNIFICATIVIDAD DE LA MEDIA
(IMEAN =0 media cero, IMEAN =1 media distinta de cero)

TRATAMIENTO DE LA SERIE (LAM) [test regresión rango media]
(LAM=1 no transformación, LAM=0 se toman logaritmo,
LAM=-1 se selecciona la transformación más adecuada)

INCLUSIÓN DE VARIABLES EXPLICATIVAS (F. DE TRANSFERENCIA)
(IREG = número de variables explicativas incluidos los outliers)

IDENTIFICACIÓN DEL MODELO INICIAL
[por defecto se utiliza el modelo ARIMA (0,1,1) $ARIMA_s$ (0,1,1)]
ARIMA (P,D,Q) $ARIMA_s$ (BP,BD,BQ)]

ESTIMACIÓN DEL MODELO FINAL
(INCON = 0 estimación máximo verosímil
INCON = 1 estimación MCO incondicional)

11.1 Identificación de una serie temporal mediante TRAMO. Estacionariedad, estacionalidad, efectos calendario, efecto pascua, estudio de los valores atípicos (outliers).

Estacionariedad

Con el fin de contrastar si una serie es o no estacionaria se pueden utilizar distintos test entre otros: la representación gráfica de la serie, el diagrama desviación típica/media, la función de autocorrelación, el test Dikey-Fuller, etc.

Así, el diagrama desviación típica/media de la variable permite comprobar si una serie presenta o no tendencia. En el caso de que una serie no presente tendencia la nube de puntos debe estar dispuesta alrededor de un punto (ver Gráfico 11.1). Si la nube de puntos define una línea paralela al eje de ordenadas ésto es indicativo de que la serie no es estacionaria en varianza (ver Gráfico 11.2); pero si ésta define una línea paralela al eje de abscisas indica que la serie no es estacionaria en media (ver Gráfico 11.3). Por último, cuando la nube define una línea oblicua al eje de abscisas, es indicativo de una serie no estacionaria ni en media ni en varianza (ver Gráfico 11.4).

Gráfico 11.1

Estacionaria

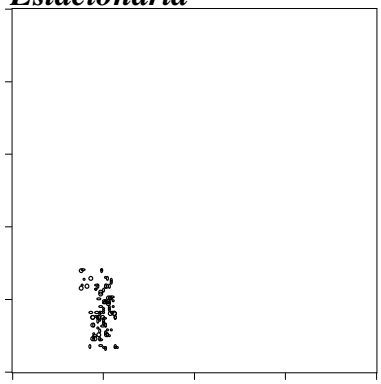


Gráfico 11.2

No estacionaria en varianza

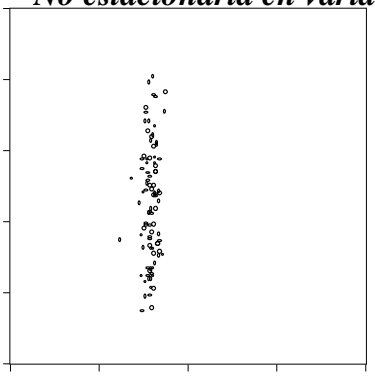


Gráfico 11.3

No estacionaria en media

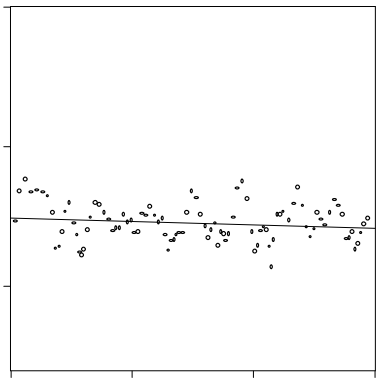
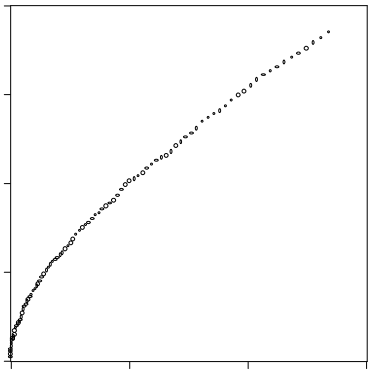


Gráfico 11.4

No estacionaria en media ni en varianza



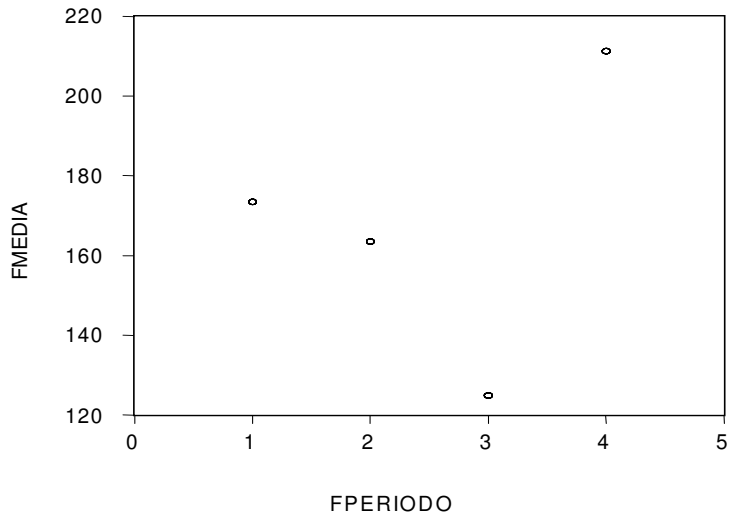
Estacionalidad

La verificación de la existencia de la componente estacional en una serie se puede efectuar, entre otros, a través de los siguientes test: representación gráfica de la serie, diagrama media de los valores estacionales (por ejemplo media del primer trimestre, media del segundo trimestre, …) frente a los periodos estacionales.

El contraste de la existencia de estacionalidad a través del diagrama media de los valores estacionales respecto a los períodos estacionales se utiliza para contrastar la existencia de movimientos de periodicidad inferior al año (de frecuencia estacional). Para ello se debe calcular la media de los valores de la serie en cada período estacional (si se trata de una serie trimestral). Se debe calcular la media de los valores del primer periodo (primer trimestre), después la del segundo período (segundo trimestre) y así sucesivamente hasta completar todos los períodos estacionales (los cuatro trimestres). Una vez calculadas las medias de los períodos estacionales se representa, a través de un diagrama, los períodos estacionales (los cuatro trimestres) frente a la media de los períodos estacionales. Si los valores del diagrama resultante están dispuestos de forma que dibujen una poligonal paralela al eje de abscisas, indicaría que la serie no presenta estacionalidad.

En el Gráfico 11.5 se observa que en el diagrama definido por períodos estacionales, las medias están dispuestos de forma que dibujan una poligonal que no es paralela al eje de abscisas, indicando que la serie presenta estacionalidad

Gráfico 11.5 **DIAGRAMA MEDIA PERÍODO ESTACIONAL**



Análisis de los outliers e intervención

Las series temporales se ven sometidas a una serie de acontecimientos externos, tales como: catástrofes naturales, huelgas, fiestas, vacaciones, cambios de política económica, etc. y en general por acontecimientos extraordinarios y/o de calendario. En el contexto de las series temporales el estudio de los acontecimientos extraordinarios de una serie temporal se les puede plantear bajo dos perspectivas diferentes, cuando se conocen de antemano el hecho extraordinario, conociendo además la fecha exacta del suceso o bien cuando se ignora de antemano si existe o ha existido algún acontecimiento extraordinario a lo largo de la serie objeto del análisis. Desde el punto de vista del análisis de series temporales el estudio de los acontecimientos extraordinarios cuando se conoce de antemano la fecha y duración del suceso se le denomina ***análisis de intervención***. Mientras que el estudio de los acontecimientos extraordinarios cuando se desconoce la fecha y las causas que lo han provocado se le denomina el ***análisis de los outliers***. Por su parte el estudio de las incidencias sobre una serie temporal de los acontecimientos derivados de las festividades, vacaciones, ferias, fiestas patronales, etc. se les denomina ***análisis del efecto calendario***. Además en algunos casos nos podemos encontrar con el hecho de que alguna o algunas observaciones de la serie se hayan omitido o perdido. El estudio de este problema se le denomina ***análisis de observaciones anómalas*** o ***análisis de missing data.***

Las técnicas estadísticas que configuran el ***análisis de intervención*** tienen por objetivo evaluar los efectos de los acontecimientos extraordinarios en una serie temporal. En general la incidencia de los hechos o acontecimientos extraordinarios en la serie temporal se manifiesta, a través de:

- El período de inicio del impacto (inicio de la intervención)
- Duración y perfil del impacto (perfil de la intervención)

Análisis de intervención.

Recordemos que la característica esencial del ***análisis de intervención*** es que se sospecha ***a priori*** que existe algún suceso singular que ha influido en la serie. En este caso, además de la fecha en que ocurrió el acontecimiento, se conoce la duración del suceso.

Existen distintos tipos de variables de intervención las más simples o sencillas son a través de la variable impulso (pulse variable) o bien a través de la variable escalón (step variable). La variable impulso, ver Gráfico 11.6, se especifica mediante una variable dummy que toma el valor uno para el periodo en que acontece el suceso extraordinario y cero para el resto de los periodos:

$$I_t = \begin{cases} 1 & \textit{en el período en que ocurre el suceso extraordinario} \\ 0 & \textit{en el resto de los períodos} \end{cases}$$

La variable escalón, ver Gráfico 11.7, se especifica mediante una variable que toma el valor cero para todos los períodos temporales anteriores al acontecimiento del suceso extraordinario y el valor uno para todos los periodos temporales posteriores al acontecimiento:

$$I_t = \begin{cases} 0 & \textit{para los períodos anteriores al que ocurre el suceso extraordinario} \\ 1 & \textit{para los períodos posteriores al que ocurre el suceso extraordinario} \end{cases}$$

Gráfico 11.6

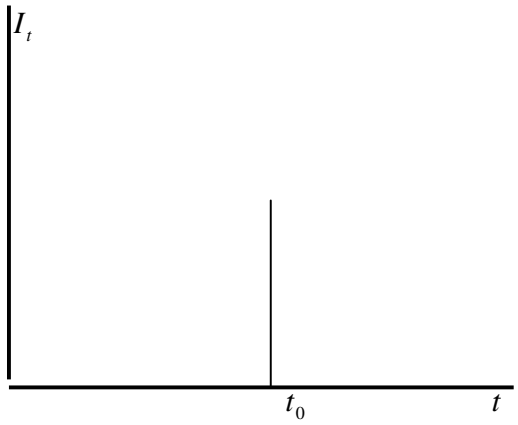
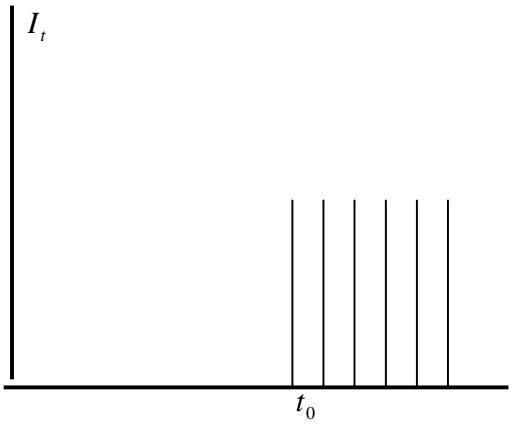


Gráfico 11.7



Además de estas dos estructuras sencillas de estudiar la intervención existen otros patrones más complejos teniendo en cuenta la intensidad del inicio del shock o acontecimiento y la duración del efecto. En general y atendiendo la combinación de las características anteriores obtenemos los siguientes cuatro posibles patrones de la intervención:

- Intensidad elevada del shock y persistencia (permanencia) del efecto, que se ilustra a través de los Gráficos 11.8 y 11.9.
- Intensidad elevada del shock y temporalidad del efecto, algunos ejemplos se pueden contemplar a través de los Gráficos 11.10 y 11.11.
- Intensidad gradual del shock y persistencia (permanencia) del efecto, que se ilustra a través de los Gráficos 11.12 y 11.13.
- Intensidad gradual del shock y temporalidad del efecto, cuyo comportamiento a través del tiempo se puede observar a través de los Gráficos 11.14 y 11.15.

Gráfico 11.8 Schok (impulso)

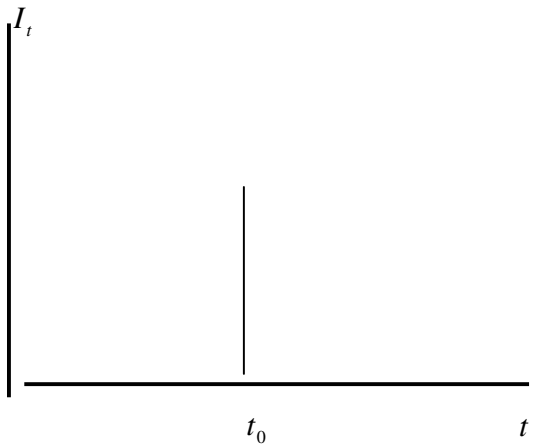


Gráfico 11.9 Respuesta o efecto permanente

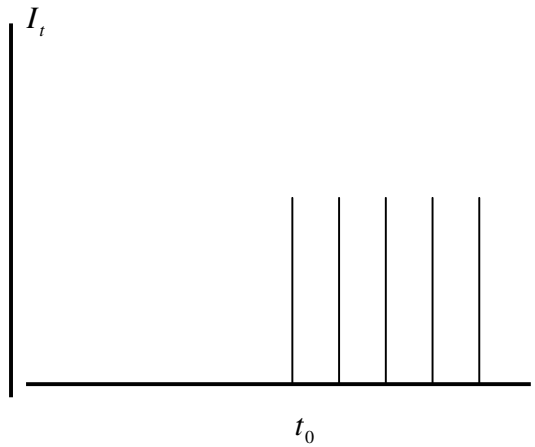


Gráfico 11.10 Schock (impulso)

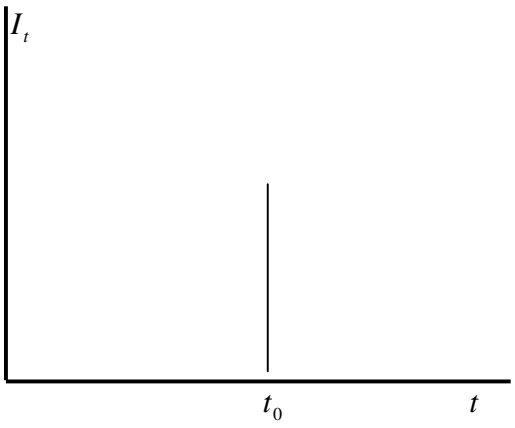


Gráfico 11.11 Respuesta o efecto temporal

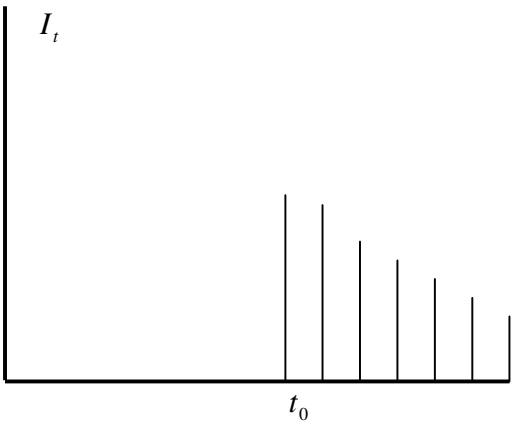


Gráfico 11.12 Shock (impulso)

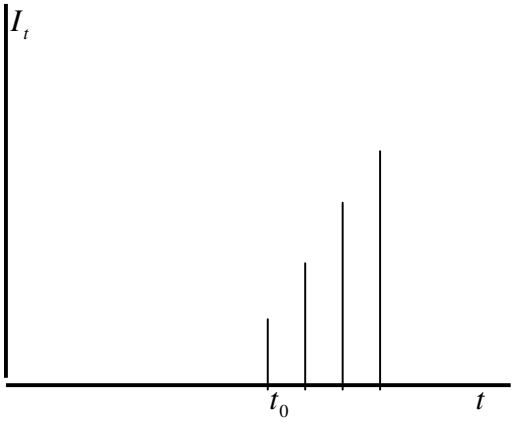


Gráfico 11.13

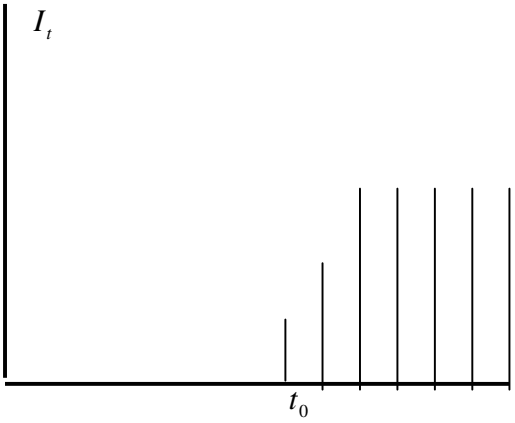


Gráfico 11.14

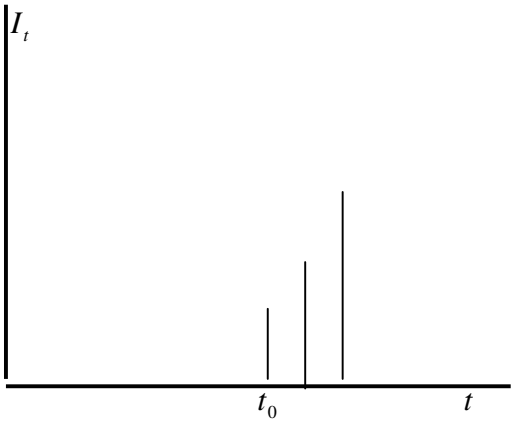
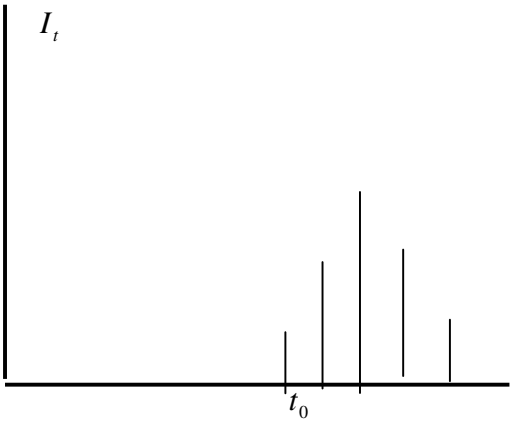


Gráfico 11.15



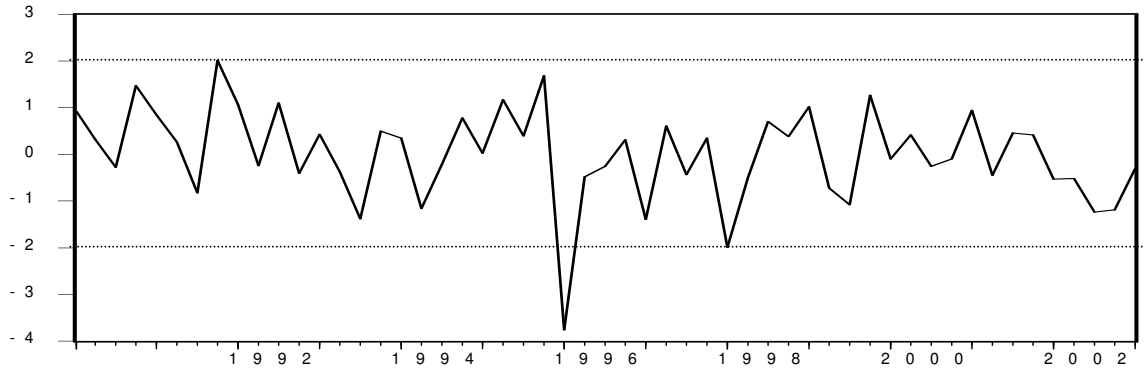
Recordemos que el análisis de intervención tan solo se puede aplicar cuando se conoce a priori la fecha y duración del hecho extraordinario. A menudo este tipo de análisis se utiliza para cuantificar los efectos económicos de los acontecimientos extraordinarios, por ejemplo las repercusiones económicas sobre la economía española de las Olimpiadas de Barcelona en 1992; otro ejemplo es la repercusión sobre el turismo de los acontecimientos del día 11 de septiembre (atentado de las Torres Gemelas).

Análisis de las observaciones atípicas (outliers).

En el contexto del análisis de series temporales se entiende por dato atípico u outlier de una serie que aquellos son la consecuencia de factores externos. En esencia los outliers son el resultado de un suceso o acontecimiento extraordinario que afecta de algún modo la variable analizada. La característica esencial de las observaciones atípicas es que no se conoce con exactitud la fecha de inicio del acontecimiento extraordinario ni tampoco la duración del mismo. Desde el punto de vista estadístico la consecuencia de un outlier es que viola la hipótesis de normalidad de la serie. Esta va a ser la base para detectar los valores anómalos. Así, por ejemplo, a través de la representación gráfica de una serie temporal normalizada (se le resta a la serie original la media y se la divide por la desviación típica) se puede detectar fácilmente los outliers. En efecto, en el Gráfico 11.16 visualizar una serie estandarizada[3], en la que se puede observar que el dato correspondiente al primer semestre del año 1996 supera con creces el doble valor de la desviación típica (casi la cuadriplica). Esta observación desde el punto de vistas estadístico se le puede considerar un outlier u observación atípica.

Gráfico 11.16

Serie estacionaria tipificada



La detección y tratamiento de los outliers es importante ya que permite obtener mejores resultados en la especificación y estimación del modelo. Recordemos que a partir de estos resultados se construyen las predicciones del modelo y en el caso de considerar explícitamente los outliers se obtendrían predicciones más acordes con la realidad.

[3] Una variable estandarizada se caracteriza por tener la media igual a cero y la desviación típica igual a la unidad.

Así pues, la detección y tratamiento de los outliers u observaciones anómalas permite mejorar:

- El conocimiento de la serie, así como el comportamiento o evolución de la misma. En efecto, la detección de los outliers suministra información sobre las incidencias externas, facilitando información sobre el momento en que acontece el suceso extraordinario, al igual que la forma de propagación del mismo. Así pues el análisis de outliers permite detectar las observaciones atípicas de una serie tales como: errores en la reproducción de la información, cambios metodológicos en la elaboración de las series, etc.
- La especificación y estimación del modelo. En este caso los valores anómalos de la serie pueden provocar distorsiones en la identificación del modelo así como en la estimación del mismo. Además las observaciones anómalas pueden alterar la estructura y propiedades de los estimadores de los parámetros del modelo.
- El análisis de intervención, ya que en algunos casos no se tiene la seguridad de la fecha exacta del suceso extraordinario así de su propagación, además en algunos casos no existe sincronía entre el suceso extraordinario y la incidencia en la serie temporal. Así por ejemplo la repercusión de la Guerra del Golfo (Enero 1991) en el precio del petróleo no coincidió ya que meses antes de la declaración de la Guerra se incrementaron los precios ya existían expectativas de que iba a suceder un suceso extraordinario.
- La calidad de los resultados, en concreto los outliers pueden afectar seriamente la calidad de predicciones que se obtienen a partir del modelo especificado. En consecuencia si se tiene en cuenta en el análisis las observaciones anómalas la calidad de las predicciones van verse afectadas favorablemente.

En la literatura económica usualmente se distinguen cuatro tipos de outliers: Outlier Aditivo (AO), Outlier Innovacional (IO), Cambio de nivel (LS) y Cambio Temporal (TC). Las características esenciales de estos son:

El *outlier aditivo* (AO) es un acontecimiento externo a la serie que incide en la serie en un solo momento temporal y que se puede especificar a través instante de una variable dummy. Una de las características del AO es que su efectos es independiente de la evolución de la serie

El *outlier innovacional* (IO) es un acontecimiento que afecta a la serie a partir de un momento temporal y se propaga al resto de valores a partir de que suceda el acontecimiento. Una de las características del IO es que su efectos están ligados a la evolución de la serie.

El *cambio de nivel* (LS) es un acontecimiento que afecta a la serie a partir de un momento temporal y su efecto persiste a lo largo del tiempo (tiene carácter permanente). El LS se caracteriza por provocar un cambio brusco en la serie en un momento determinado y por el hecho de que sus efectos son independientes de la evolución de la serie.

El *cambio temporal* (TC) es un acontecimiento que afecta a la serie a través de un cambio brusco en un momento determinado, pero a partir de este instante temporal su efecto decae (se amortigua) progresivamente de acuerdo con un factor de progresión (tiene carácter transitorio). Se caracteriza por el hecho de sus efectos son independientes de la evolución de la serie.

Es conveniente insistir que la diferencia esencial entre el análisis de las observaciones atípicas (outliers) y el análisis de intervención radica en que en el primer caso se desconoce la existencia de los acontecimientos o sucesos extraordinarios que afectan a la serie (se desconoce la fecha y la duración del suceso), mientras en el análisis de intervención se conoce a priori la fecha y permanencia del acontecimiento o suceso.

Contraste observaciones anómalas (intervención).

Desde el punto de vista del análisis de series temporales el estudio de los acontecimientos extraordinarios, cuando se conoce a ciencia cierta la fecha y duración de los efectos del suceso se puede contrastar estos hechos median un modelo en el que se incluye un regresor que recoja dichos acontecimientos.

Así, en el caso de que el acontecimiento a contrastar es un hecho puntual (outlier aditivo) que no tiene efectos duraderos el regresor que se debe construir es una variable dummy AO_t en que todos sus valores son cero a excepción del valor uno para el periodo en que acontece el hecho extraordinario. Un posible ejemplo se establece a partir de la serie Y_t , número de pernoctaciones mensuales causadas por los extranjeros de origen europeo en la Comunidad Valenciana, en la que se pretende contrastar el efecto del *once de septiembre* sobre dichas pernoctaciones. En este caso la variable AO_t es una variable que toma el valor cero a excepción del mes de septiembre de 2001 que toma el valor uno. El modelo que deberíamos especificar es el siguiente:

$$Y_t = f(AO_t) + u_t = \beta_0 + \beta_1 AO_t + u_t$$

Una vez estimado el modelo especificado se debe contrastar la significatividad del coeficiente de regresión que afecta a la variable, es decir β_1 . En el caso de que sea significativa probabilisticamente se concluye que el efecto del *once de septiembre* influye de forma transitoria (puntual) sobre la variable pernoctaciones.

En caso de que el acontecimiento a contrastar es un hecho permanente, es decir que tiene efecto duradero, la variable explicativa que se debe construir es una variable LS_t que toma valores cero desde el origen hasta el periodo que acontece el hecho extraordinario, a partir de este periodo el valor de la serie es uno. Un posible ejemplo es contrastar el efecto del parque de ocio *Terra Mítica* en la serie mensual Y_t , formada por el número de pernoctaciones causadas en Benidorm. En este caso la variable LS_t es una variable que toma el valor cero desde el origen de la serie hasta el mes de julio de 2001 y a partir de dicho mes la variable toma el valor uno. El modelo que deberíamos especificar es el siguiente:

$$Y_t = f(LS_t) + u_t = \beta_0 + \beta_1 LS_t + u_t$$

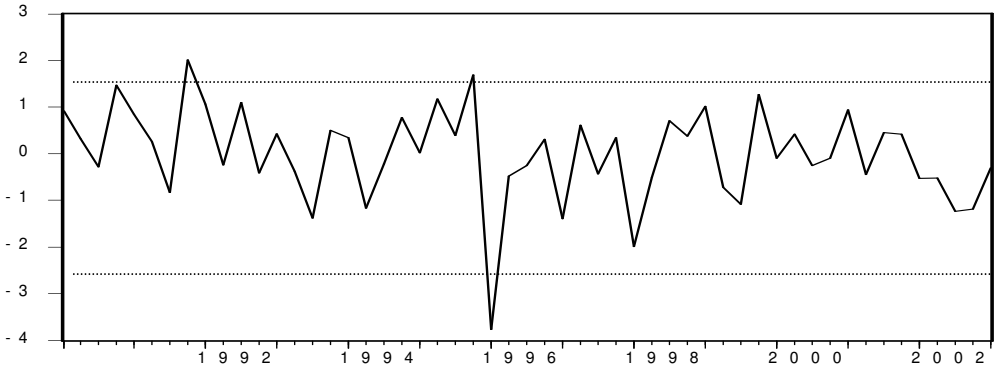
una vez estimado el modelo se debe contrastar la significatividad del coeficiente de regresión que afecta a la variable LS_t , es decir β_1 . En el caso de que sea significativa probabilisticamente se concluye que el efecto *Terra Mítica* influye de forma permanente sobre la variable pernoctaciones.

Contraste observaciones atípicas (outliers)

El estudio de los acontecimientos extraordinarios cuando se desconoce la fecha y las causas que lo han provocado se le denomina el ***análisis de los outliers***. La característica esencial de las observaciones atípicas es que no se conoce con exactitud la fecha de inicio del acontecimiento extraordinario ni tampoco la duración del mismo. Desde el punto de vista estadístico la consecuencia de un outlier, entre otras, es que viola la hipótesis de normalidad de la serie. Esta será la base para detectar los valores anómalos. Así, por ejemplo, a través de la representación gráfica de una serie temporal normalizada (se le resta a la serie original la media y se la divide por la desviación típica) se puede detectar fácilmente los outliers. En efecto, en el Gráfico 11.40 visualizar una serie estandarizada[4], en la que se puede observar que el dato correspondiente al primer semestre del año 1996 supera con creces el doble valor de la desviación típica (casi la cuadriplica). Esta observación desde el punto de vistas estadístico se le puede considerar un outlier o valor atípico.

Gráfico 11.17

Residuos tipificados del modelo



Desde el punto de vista estadístico el contraste de un valor de una serie atípico aislado u outlier Y_0 ***se puede efectuar mediante el siguiente estadístico:***

$$\lambda = \frac{Y_0}{\sqrt{\text{var}(Y_t)}}$$

además, se sabe que el estadístico λ se distribuye asintóticamente según una normal tipificada.

Así para efectuar el contraste de existencia de un outlier se puede establecer el siguiente contraste probabílistico:

H_O: El valor Y_0 de la serie no es un outlier **H_A**: El valor Y_0 de la serie es un outlier

[4] Una variable estandarizada se caracteriza por tener la media igual a cero y la desviación típica igual a la unidad.

Que se puede docimar a través de la siguiente desigualdad probabilistica:

$$\Pr ob(-N_{\alpha/2} < \lambda < N_{\alpha/2}) = 1-\alpha$$

en el caso de cumplirse la desigualdad probabilistica se concluye aceptando la hipótesis nula con una probabilidad del $1-\alpha$.

Análisis de los efectos calendario.

Se entiende por efecto calendario a todas los acontecimientos derivados de las festividades, vacaciones, ferias, fiestas patronales, etc. Dentro del efecto calendario existen varios tipos de acontecimientos que tienen una relevancia considerable. Entre ellos destacan el ***Efecto Pascua***, el ***Efecto Calendario Laboral***.

Efecto Pascua tiene dos derivaciones, la primera nace del hecho de que la festividad de Pascua de Resurrección (Semana Santa) es una fiesta móvil que puede acontecer en el mes de marzo o bien en el mes de abril y dicha festividad tiene una incidencia considerable en la actividad económica, por ejemplo sobre la producción de las industrias o bien sobre la actividad turística. Además, no en todas los países y regiones el periodo vacacional coincide en el tiempo y duración del mismo. Así por ejemplo en Andalucía el Efecto Pascua (vacaciones de Semana Santa) se inician en el domingo de Ramos y termina el Domingo de Pascua, mientras que en Valencia el inicio del período festivo es el Viernes Santo y se prolonga a lo largo de la Semana de San Vicente.

La forma de especificar ***Efecto Calendario Pascua*** ECP_t es a través de una variable que toma el valor cuatro en el mes que coincide con dicha fiesta y cero en el resto de los meses del año. El modelo que deberíamos especificar es el siguiente:

$$Y_t = f(ECP_t) + u_t$$

una vez estimado el modelo se debe contrastar la significatividad del coeficiente de regresión que afecta a la variable ECP_t. En el caso de que sea significativa probabilisticamente se concluye que el ***Efecto Pascua*** influye en la variable objeto del análisis.

En el caso concreto de la Comunidad Valenciana se puede elaborar, a partir del calendario para los años 2000-2003, ver Cuadro 11.1, la variable ***Efecto Pascua***, ver Cuadro 11.2.

Cuadro 11.2 ***Efecto Pascua para la Comunidad Valenciana***

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Sept	Oct	Nov	Dic
2000	0	0	0	4	0	0	0	0	0	0	0	0
2001	0	0	3	1	0	0	0	0	0	0	0	0
2002	0	0	0	4	0	0	0	0	0	0	0	0
2003	0	0	0	4	0	0	0	0	0	0	0	0

Fuente: Elaboración propia a partir del Cuadro 11.1

El ***Efecto Calendario Laboral o Efecto Laboral*** nace del hecho que el número de días laborables de los distintos meses es distinto, la diferencia proviene de que los meses tienen distinto número de días laborables (unos meses tienen treinta idas y otros treinta y uno, febrero veintiocho). Pero además en algunas ocasiones unos meses tienen cuatro domingos mientras que otros tienen cinco domingos. Estos hechos provocan que el calendario laboral (número de idas laborables en un mes) vayan cambiando a lo largo de la historia.

Otras incidencias que se tienen que tener en cuenta dentro del efecto calendario son las fiestas patronales y las ferias, así por ejemplo en Pamplona se celebran los San Fermín, en Sevilla la Feria de Abril, en Valencia las Fallas, etc. y que estas fiestas en algunas ocasiones son móviles y tienen su incidencia sobre la actividad económica en general (producción, actividad turística, suministro de energía, etc.)

La forma de tener en cuenta o introducir en el modelo de series temporales los efectos calendario es la inclusión de una o mas variables explicativas o inputs que recojan dichos efectos.

La forma de tener en cuenta o introducir en el modelo de series temporales los ***Efectos Calendario Laboral*** ECL_t es a través de una variable que toma en cada mes un valor igual al número de días laborables. El modelo que deberíamos especificar es el siguiente:

$$Y_t = f(ECL_t) + u_t$$

una vez estimado el modelo se debe contrastar la significatividad del coeficiente de regresión que afecta a la variable ECL_t. En el caso de que sea significativa probabilisticamente se concluye que el ***Efecto Calendario Laboral*** influye en la variable objeto del análisis.

En el caso concreto de la Comunidad Valenciana se puede elaborar, a partir del calendario para los años 2000-2003, ver Cuadro 11.1, la variable ***Efecto Calendario Laboral***, ver Cuadro 11.3.

Cuadro 11.3 ***Efecto Calendario Laboral para la Comunidad Valenciana***

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Sept	Oct	Nov	Dic
2000	20	21	23	18	22	22	21	22	21	20	21	18
2001	22	20	21	19	22	21	22	22	20	21	21	19
2002	22	20	19	21	22	20	23	21	21	22	20	20
2003	20	20	20	20	21	21	23	20	22	21	20	21

Fuente: Elaboración propia a partir del Cuadro 11.1

En otros casos, por ejemplo, en las series relacionadas con el turismo lo que interesa es estudiar la incidencia de los días festivos sobre la actividad turística. A tal fin se debería construir una variable ***efecto calendario*** en que en cada mes se contabilice el número de días festivos.

11.2. Estimación y validación del modelo (análisis de los residuos, contraste de la estabilidad / invertibilidad, significatividad,)

Una vez depurada la serie temporal de la componente(s) irregular(es), que podrían distorsionar los resultados, se procede a la identificación de la serie temporal transformada y estacionaria.

El procedimiento seguido por TRAMO es el de estimar diferentes especificaciones del modelo y se elige el más adecuado. Siempre y cuando supere las hipótesis establecidas en el modelo. Así pues, se trata de un proceso iterativo que culmina con la estimación de un modelo que supera todos y cada una de los test a que se ve sometido el modelo estimado.

El proceso iterativo de estimación se inicia con la especificación del modelo de *Líneas Aéreas*, es decir, un modelo ARIMA(0,1,1)SARISMA(0,1,1). Se inicia con la especificación de este modelo ya que la mayoría de series económicas son susceptibles de ser modelizadas a través de dicha ecuación.

1. Significatividad aislada de cada uno de los coeficientes del modelo incluidos los outliers, efecto calendario, efecto Pascua (t-Student).
2. Significatividad conjunta de los coeficientes del modelo (F-Snedecor).
3. Estudio de la estacionariedad e invertibilidad del modelo (Calcular los módulos de las raíces de los polinomios característicos)
4. Contraste la hipótesis de la aleatoriedad (test de rachas)
5. Contraste la hipótesis de normalidad (test de Bera-Jarque).
6. Contraste la hipótesis de no autocorrelación a partir de la función de autocorrelación de los residuos (test Q de Ljung-Box)
7. Contraste la hipótesis de homoscedasticidad a partir de la función de autocorrelación de los residuos elevados al cuadrado (test Q de Ljung-Box)

Si el modelo estimado supera todos y cada uno de los test descritos éste seria el modelo elegido. De lo contrario se debería especificar un modelo alternativo. El proceso concluye cuando se alcanza un modelo que supere cada una de las hipótesis y test establecidos.

En el presente caso se va a estructurar la exposición agrupando los comentarios según la presentación de los resultados (la salida) del programa[5] TRAMO para una serie que presenta estacionalidad que consiste en las siguientes etapas:

[5] La especificación del modelo ARIMA(p,d,q) en el programa TRAMO es la siguiente:
$(1+\phi_1 L+\phi_2 L^2+...+\phi_p L^p)(1-L)^d Y_t = (1+\theta_1 L+\theta_2 L^2+...+\theta_q L^p)\varepsilon_t$

1. Presentación de la serie original.

```
  TIME SERIES REGRESSION MODELS WITH ARIMA ERRORS, MISSING VALUES AND OUTLIERS.
                              BETA VERSION (*)

                                      BY

                          VICTOR GOMEZ & AGUSTIN MARAVALL

                    with the programming assistance of G. CAPORELLO

(*) Copyright : V. GOMEZ, A. MARAVALL (1994,1996)

 SERIES TITLE=evtramo

  SINCE LONGER FORECAST FUNCTION IS REQUIRED
  BY SEATS, NPRED CHANGED TO (24)

 ORIGINAL SERIES

 NUMBER OF OBSERVATIONS: 135
```

YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
1991										113.3	107.4	96.8
1992	99.6	101.7	105.2	98.5	100.3	102.9	106.0	54.3	100.1	101.7	97.1	86.9
1993	85.5	91.2	99.3	90.2	94.0	97.1	98.3	53.1	97.5	96.8	99.7	89.9
1994	89.5	94.8	103.7	96.7	102.6	104.6	102.0	59.8	104.9	103.2	108.5	97.4
1995	102.6	102.2	116.7	97.3	111.9	112.3	108.1	64.1	105.2	107.6	109.5	92.5
1996	102.6	103.9	107.5	98.3	108.6	105.0	112.6	62.2	106.4	115.1	108.5	97.0
1997	104.8	102.4	106.5	116.9	112.1	114.8	121.2	66.9	116.4	125.5	116.4	105.2
1998	108.9	114.9	122.0	113.5	118.7	123.9	128.9	72.3	121.9	125.2	122.4	109.1
1999	111.8	116.3	126.7	116.2	124.1	129.0	131.3	76.1	126.8	124.7	129.9	115.1
2000	116.5	126.8	139.5	115.6	134.6	135.1	129.7	80.3	125.7	127.5	134.0	108.9
2001	125.2	122.0	134.5	116.1	134.3	131.4	131.7	84.0	123.6	134.1	128.9	104.3
2002	124.3	122.7	118.9	127.8	130.4	123.8	135.2	79.8	125.5	137.8	127.1	107.2

2. Parámetros control para la especificación del modelo a través de los cuales se definen las distintas alternativas del modelo.

```
SEATS CANNOT BE RUN WITH AIO=0              AIO CHANGED TO 2

 MODEL PARAMETERS
 ----------------
  MQ= 12          IMEAN=  1       LAM= -1       D=  1        BD=  1
  P=  0           BP=  0          Q=  1         BQ=  1       IREG=  2
  ITRAD=  2       IEAST=  0       IDUR=  0      M= 36        QM= 24
  INCON=  0       NBACK=  0       NPRED= 24     INTERP=  2   INIT=  0
  IFILT=  2       IDENSC=  1      IROOT=  2     INIC=  3     ICONCE=  1
  ICDET=  1       IATIP=  1       IMVX=  0      IDIF=  3     PG=  0
  AIO=  2         INT1=  1        INT2= 135     RSA=  0      SEATS=  2
  VA= 3.50        TOL=  0.100E-03               PC=  0.143E+00
  NOADMISS= 1     BIAS= 1         SMTR= 0
  THTR= -0.400    RMOD=  0.500    MAXBIAS=  0.500

TH  =          -0.10
BTH =          -0.10
```

3. Estudiar la estacionariedad de la serie y la transformación más adecuada.
3.1. Contraste para analizar la significatividad de la media o termino constante.

```
MEAN IS NOT SIGNIFICANT:
 IMEAN CHANGED TO 0
```

3.2. Estudio de la conveniencia de transformar la serie mediante logaritmos neperianos y diferencias. El test utilizado es el de rango-media.

TRANSFORMED SERIES (LOGARITHMS OF THE DATA)

YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
1991										4.730	4.677	4.573
1992	4.601	4.622	4.656	4.590	4.608	4.634	4.663	3.995	4.606	4.622	4.576	4.465
1993	4.449	4.513	4.598	4.502	4.543	4.576	4.588	3.972	4.580	4.573	4.602	4.499
1994	4.494	4.552	4.642	4.572	4.631	4.650	4.625	4.091	4.653	4.637	4.687	4.579
1995	4.631	4.627	4.760	4.578	4.718	4.721	4.683	4.160	4.656	4.678	4.696	4.527
1996	4.631	4.643	4.677	4.588	4.688	4.654	4.724	4.130	4.667	4.746	4.687	4.575
1997	4.652	4.629	4.668	4.761	4.719	4.743	4.797	4.203	4.757	4.832	4.757	4.656
1998	4.690	4.744	4.804	4.732	4.777	4.819	4.859	4.281	4.803	4.830	4.807	4.692
1999	4.717	4.756	4.842	4.755	4.821	4.860	4.877	4.332	4.843	4.826	4.867	4.746
2000	4.758	4.843	4.938	4.750	4.902	4.906	4.865	4.386	4.834	4.848	4.898	4.690
2001	4.830	4.804	4.902	4.754	4.900	4.878	4.881	4.431	4.817	4.899	4.859	4.647
2002	4.823	4.810	4.778	4.850	4.871	4.819	4.907	4.380	4.832	4.926	4.845	4.675

4. Identificación automática de la serie.
4.1. Determinación de los ordenes de la parte autoregresiva, diferenciación y de medias móviles del modelo regular ARIMA(p,d,q) y los ordenes de la parte autoregresiva, diferenciación y de medias móviles del modelo estacional SARISMA(P,D,Q). En el presente caso se ha seleccionado como modelo más adecuado ARIMA(0,1,2)*SARISMA(0,1,1).

AUTOMATIC MODEL IDENTIFICATION BEGINS

MODEL FINALLY CHOSEN: (0,1,2)(0,1,1)

4.2. Inclusión o no del término independiente o constante en el modelo. Teniendo en cuenta el resultado obtenido d el punto 3.1.

WITHOUT MEAN

4.3. Inclusión o no de efectos calendario, observaciones atípicas y del posible análisis de intervención realizado. Informando además de las características de las observaciones atípicas. En el presente caso, la serie analizada no presenta observaciones atípicas.

```
WITH TRADING DAY CORRECTION

 WITHOUT EASTER CORRECTION

 OUTLIERS

  67 AO     ( 4 1997)
 127 AO     ( 4 2002)
 126 TC     ( 3 2002)
```

5. Estimaciones de los parámetros del modelo ARIMA así como el valor del ratio "t" bajo la hipótesis nula de no significatividad.

```
METHOD OF ESTIMATION: EXACT MAXIMUM LIKELIHOOD

 PARAMETER      ESTIMATE          STD ERROR      T RATIO   LAG

 MA1  1         -.38801          0.93842E-01      -4.13      1
 MA1  2         0.098765         0.95204E-01       1.04      2
 MA2  1         -.33403          0.93371E-01      -3.58     12
```

6. Validación del modelo estimado.

6.1. Análisis de la estacionariedad e invertibilidad del modelo estimado a través del estudio de las raíces del polinomio característico . En el presente caso dado que las raíces en modulo son inferiores a la unidad el modelo estimado es estacionario.

```
REGULAR MA INVERSE ROOTS ARE
 NO.     REAL P.     IMAG.P.     MODULUS    ARGUMENT      PERIOD
  1    0.1940054  -0.2472385   0.3142689  -51.8791070   -6.9392097
  2    0.1940054   0.2472385   0.3142689   51.8791070    6.9392097

 SEASONAL MA INVERSE ROOTS ARE
 NO.     REAL P.     IMAG.P.     MODULUS    ARGUMENT      PERIOD
  1     0.33403      0.0000     0.33403      0.0000           -
```

6.2. Análisis de la correlación entre los parámetros estimados del modelo estimado.

CORRELATIONS OF THE ESTIMATES

```
  1.0000   -0.3443    0.0145
 -0.3443    1.0000   -0.0262
  0.0145   -0.0262    1.0000
```

6.3. Criterios para analizar la significatividad conjunta del modelo estimado. Los criterios utilizados son: AIC (AIC de Akaike), BIC (Bayesian Information Criterium) y la función objetivo.

```
AIC
   -565.3547

 BIC
    -7.3010

 FINAL VALUE OF OBJECTIVE FUNCTION:
              0.60867E-01

 ITERATIONS:            2

  NUMBER OF FUNCTION EVALUATIONS:            9
```

7. Estimaciones de los parámetros de los efectos calendario y outliers así como el valor del ratio "t" bajo la hipótesis nula de no significatividad.

```
ESTIMATES OF REGRESSION PARAMETERS
 CONCENTRATED OUT OF THE LIKELIHOOD

 PARAMETER         VALUE        ST. ERROR      T VALUE
 TRAD   1        0.007305      (   0.00047)   15.50
 TRAD   2        0.025996      (   0.00954)    2.73
 OUT 1 ( 67)     0.12577       (   0.01540)    8.17   AO     ( 4 1997)
 OUT 2 (127)     0.12517       (   0.01974)    6.34   AO     ( 4 2002)
 OUT 3 (126)     -.096275      (   0.02181)   -4.41   TC     ( 3 2002)

 COVARIANCE MATRIX OF ESTIMATORS

 0.222E-06  -0.772E-07  -0.486E-06  -0.149E-05   0.111E-05
-0.772E-07   0.909E-04  -0.663E-05  -0.119E-05  -0.105E-04
-0.486E-06  -0.663E-05   0.237E-03   0.534E-05  -0.166E-05
-0.149E-05  -0.119E-05   0.534E-05   0.390E-03  -0.119E-03
 0.111E-05  -0.105E-04  -0.166E-05  -0.119E-03   0.476E-03
```

8. Validación de las hipótesis del modelo a partir del análisis de los residuos.
8.1. Presentación de los residuos del modelo estimado

```
NUMBER OF WHITE NOISE RESIDUALS          117
  WHITE NOISE RESIDUALS

-0.0035   0.0194   0.0144   0.0104   0.0064   0.0202   0.0112   0.0235
 0.0305   0.0299   0.0070   0.0053   0.0501  -0.0051   0.0065  -0.0077
 0.0302  -0.0026  -0.0121   0.0232   0.0377   0.0265  -0.0308   0.0302
-0.0564  -0.0014   0.0183  -0.0093   0.0172  -0.0394  -0.0299  -0.0065
-0.0266  -0.0024  -0.0179   0.0068  -0.0252   0.0063   0.0212   0.0119
 0.0128   0.0050   0.0279  -0.0168  -0.0107  -0.0018   0.0062  -0.0358
 0.0228   0.0364   0.0257   0.0167   0.0312  -0.0213   0.0190   0.0087
-0.0372  -0.0088   0.0545   0.0053   0.0037  -0.0152  -0.0317   0.0025
 0.0292  -0.0310  -0.0288  -0.0048  -0.0013   0.0169  -0.0142  -0.0116
 0.0095   0.0204  -0.0056   0.0029  -0.0062  -0.0040  -0.0242   0.0066
 0.0232   0.0007  -0.0116   0.0519  -0.0268  -0.0221   0.0039  -0.0278
 0.0031  -0.0098  -0.0303   0.0278  -0.0179   0.0177  -0.0080  -0.0075
-0.0265   0.0063   0.0073  -0.0143   0.0528  -0.0201  -0.0035  -0.0512
-0.0345   0.0368   0.0251   0.0082   0.0245  -0.0035  -0.0244   0.0110
-0.0098  -0.0069   0.0342  -0.0228  -0.0267
```

8.2. Test de la normalidad Jarque-Bera a partir de los residuos del modelo estimado

```
 TEST-STATISTICS ON RESIDUALS
          MEAN=   0.0014461
       ST.DEV.=   0.0020908
       OF MEAN
       T-VALUE=  0.6916
 NORMALITY TEST=  0.5652        ( CHI-SQUARED(2) )

      SKEWNESS=  0.0075          ( SE =  0.2265 )
      KURTOSIS=  2.6598          ( SE =  0.4529 )
 SUM OF SQUARES=   0.6008786E-01
```

8.3. Test de Durbin-Watson para analizar la autocorrelación de primer orden.

DURBIN-WATSON= 2.0236

8.4. Suma de los cuadrados de los residuos, desviación típica y varianza del modelo estimado.

```
SUM OF SQUARES=   0.6008786E-01

STANDARD ERROR=   0.2295836E-01
 OF RESID.
  MSE OF RESID.=   0.5270865E-03
```

8.5. Función de autocorrelación (AC) y de autocorrelación parcial (PAC) de los residuos, para analizar la existencia de autocorrelación de orden superior. Se incluye el test de Ljung-Box.

```
AUTOCORRELATIONS
   -0.017 -0.012  0.087 -0.083  0.042  0.0949  0.057  0.021  0.142 -0.071  0.012  0.0423
SE  0.092  0.092  0.092  0.092  0.092  0.0925  0.092  0.092  0.092  0.092  0.092  0.0925
Q   0.04   0.06   0.98   1.84   2.07   3.19    3.62   3.68   6.30   6.96   6.98   7.21
PV -1.00  -1.00  -1.00   0.18   0.36   0.36    0.46   0.60   0.39   0.43   0.54   0.62

   -0.120  0.145 -0.062 -0.104 -0.021 -0.0884  0.065 -0.089 -0.094 -0.027  0.007 -0.1499
SE  0.0925 0.092  0.092  0.092  0.092  0.0925  0.092  0.092  0.092  0.092  0.092  0.0925
Q   9.14  11.99  12.53  14.04  14.10  15.20   15.81  16.97  18.25  18.37  18.38  21.74
PV  0.52   0.36   0.40   0.37   0.44   0.44    0.47   0.46   0.44   0.50   0.56   0.41

    0.088 -0.029 -0.125 -0.027  0.030 -0.033   0.144  0.0566 0.0670 0.046 -0.042 -0.0489
SE  0.0925 0.092  0.0925 0.092  0.092  0.092   0.092  0.0925 0.092  0.092  0.092  0.0925
Q  22.91  23.05  25.50  25.62  25.76  25.95   29.32  29.84  30.59  30.95  31.26  31.67
PV  0.41   0.46   0.38   0.43   0.48   0.52    0.40   0.42   0.44   0.47   0.50   0.53

  LJUNG-BOX Q VALUE OF ORDER 24 IS    21.74 AND IF RESIDUALS ARE RANDOM
IT SHOULD BE DISTRIBUTED AS CHI-SQUARED(21)

 PARTIAL AUTOCORRELATIONS
 -0.017 -0.0130 0.086 -0.081  0.043  0.087  0.076  0.012  0.140 -0.066  0.016  0.0066
SE  0.092  0.0925 0.092  0.092  0.092  0.092  0.0925  0.092  0.092  0.092  0.092  0.0925

   -0.1033 0.116 -0.094 -0.095 -0.065 -0.084  0.0943 -0.123 -0.094  0.006  0.006 -0.1073
SE  0.0925 0.0925 0.092  0.092  0.0925 0.092  0.0925  0.092  0.092  0.092  0.092  0.0925

    0.123 -0.051 -0.021 -0.108  0.114  0.015  0.1443  0.071  0.113  0.012  0.032 -0.0796
SE  0.092  0.0925 0.092  0.092  0.092  0.092  0.0925  0.092  0.092  0.092  0.092  0.0925
```

8.6. Análisis de la aleatoriedad a través del test de rachas aplicados sobre los residuos y sobre la función de autocorrelación.

```
APPROXIMATE TEST OF RUNS ON RESIDUALS
 -------------------------------------
  NUM.DATA= 117
   NUM.(+)=  59
   NUM.(-)=  58
  NUM.RUNS=  57
   T-VALUE= -0.3714
```

```
 APPROXIMATE TEST OF RUNS ON AUTOCORRELATION FUNCTION
 ----------------------------------------------------
  NUM.DATA=  36
   NUM.(+)=  18
   NUM.(-)=  18
  NUM.RUNS=  19
    T-VALUE=  0.0000
```

8.7. Función de autocorrelación (AC) de los residuos elevados al cuadrado, con el fin estudiar la homoscedasticidad (varianzas condicionales constantes). Se incluye el test de Ljung-Box.

```
 SQUARED RESIDUALS:
 ------------------

AUTOCORRELATIONS
  -----------------
    -0.034  0.005  0.017  0.124 -0.035 -0.063 -0.044  0.020 -0.058 -0.051 -0.074 -0.0038
SE   0.092  0.092  0.092  0.092  0.092  0.092  0.092  0.092  0.092  0.092  0.092  0.0925
Q    0.15   0.15   0.19   2.08   2.24   2.75   3.00   3.05   3.50   3.84   4.57   4.57
PV -1.00  -1.00  -1.00   0.15   0.33   0.43   0.56   0.69   0.74   0.80   0.80   0.87

    -0.034 -0.050 -0.083 -0.036  0.056  0.008 -0.125  0.081 -0.090 -0.010 -0.066 -0.1046
SE   0.092  0.092  0.092  0.092  0.092  0.092  0.092  0.092  0.092  0.092  0.092  0.0925
Q    4.73   5.08   6.02   6.20   6.65   6.66   8.89   9.84  11.02  11.03  11.68  13.32
PV   0.91   0.93   0.91   0.94   0.95   0.97   0.92   0.91   0.89   0.92   0.93   0.90

     0.073 -0.010  0.028 -0.048 -0.010 -0.058 -0.017  0.001 -0.044  0.176  0.061  0.0452
SE   0.092  0.092  0.092  0.092  0.092  0.092  0.092  0.092  0.092  0.092  0.092  0.0925
Q   14.14  14.16  14.28  14.64  14.66  15.20  15.25  15.25  15.58  20.78  21.43  21.78
PV   0.90   0.92   0.94   0.95   0.96   0.97   0.98   0.98   0.99   0.92   0.92   0.93

  LJUNG-BOX Q VALUE OF ORDER 24 IS    13.32 AND IF RESIDUALS ARE RANDOM
IT SHOULD BE DISTRIBUTED AS CHI-SQUARED(21)
```

9. Predicción. Se presenta la predicción puntual y su desviación típica, tanto de la serie transformada como en la serie original, con el fin de obtener la predicción por intervalos.

```
 ORECASTS:

 ORIGIN:           135 NUMBER:            12

 OBS     FORECAST       STD ERROR       ACTUAL      RESIDUAL       FORECAST    STD ERROR
               (TR. SERIES)                                          ( ORIGINAL SERIES)
 136     4.81973       0.229619E-01                                  123.931      2.84608
 137     4.80330       0.269369E-01                                  121.912      3.28452
 138     4.87015       0.350285E-01                                  130.341      4.56705
 139     4.78686       0.383436E-01                                  119.924      4.60000
 140     4.88322       0.399241E-01                                  132.055      5.27426
 141     4.87427       0.427012E-01                                  130.878      5.59121
 142     4.92077       0.455513E-01                                  137.108      6.24867
 143     4.36565       0.483959E-01                                   78.700      3.81102
 144     4.86574       0.509573E-01                                  129.767      6.61687
 145     4.91814       0.534736E-01                                  136.749      7.31768
 146     4.82111       0.559671E-01                                  124.102      6.95109
 147     4.70568       0.582369E-01                                  110.574      6.44493
 148     4.79554       0.662848E-01                                  120.970      8.02728
```