
Un nuevo método para el control de calidad de los datos en series temporales

Este artículo ha sido elaborado por Cristina Luna, de la Oficina de Instituciones Financieras, y Agustín Maravall, del Servicio de Estudios (1).

1. INTRODUCCIÓN

Este trabajo resume la aplicación de una derivación del programa TRAMO (2) denominada TERROR (*TRAMO for errors*), al control de calidad de los datos que el Banco de España recibe regularmente de entidades, y que sirven de base para la construcción de sus series agregadas. Dado que cada mes y cada trimestre se reciben cientos de miles de datos nuevos, es preciso realizar un esfuerzo para detectar posibles errores en los mismos. Las causas de estos errores son múltiples y muy variadas; pueden ir desde un error en la colocación de una coma, hasta la confusión, por poner un ejemplo, por parte de algún banco, de sus activos bancarios frente a Costa de Marfil con sus activos bancarios frente a las Islas Caimán. El problema del control de calidad de los datos ha sido objeto de una atención creciente en los últimos años. Sin embargo, las técnicas disponibles hasta el presente para tratar conjuntos grandes de series presentan limitaciones importantes, que implican que la detección de errores sea fundamentalmente manual y se base en unos criterios simples, que utilizan una cantidad pequeña de la información disponible. Por una parte, la detección se realiza, en esencia, comparando el crecimiento de los dos últimos períodos (ignorando, pues, la historia completa de la serie de que se trate), y, por otro lado, la comparación se realiza a un nivel de agregación relativamente elevado. Una implicación de la dificultad de detección de errores que ha habido en el pasado es que algunos de los grupos de series de base presentan un aspecto errático, que hace sospechar que pueden contener un número considerable de errores. Conviene señalar que métodos semejantes, fundamentalmente manuales, son también empleados en la mayoría de los bancos centrales e instituciones tales como el Banco de Pagos Internacionales.

En la detección de errores es obvio que la información exógena *a priori* es importante (por ejemplo, pueden haberse fusionado dos bancos). Pero sería importante disponer de un método que fuese automático y que permitiese tratar gran cantidad de series en muy poco tiem-

(1) Queremos agradecer la ayuda de Luis Zapatero, Fernando Sánchez, Gianluca Caporello y Víctor Gómez. Estos dos últimos programaron las modificaciones necesarias de TRAMO.

(2) *Time Series Regression with Arima noise, Missing values, and Outliers*, Gómez y Maravall (1997 y 1998a).

po, utilizando la información disponible de forma eficiente. Un método de este tipo podría proporcionar una relación básica de partida sobre cuáles de los nuevos datos son sospechosos de ser erróneos.

Cualquier método automático tiene que construirse sobre una definición precisa de qué significa que un dato es «sospechoso». Dado que se trata, en nuestro caso, de series temporales, parece natural plantear el problema en los siguientes términos: dada la evolución que ha seguido la serie, se obtiene una predicción óptima para el próximo período, así como una medida de la desviación típica del error de predicción. Si, al llegar el nuevo dato, este resulta «excesivamente» alejado de la predicción, se clasifica como sospechoso. El concepto de «excesivo» es fácil de precisar. Por ejemplo, se puede considerar excesivo un error de predicción de un tamaño tal que, dadas las características de la serie, solo esperaría uno encontrarlo una de cada 100.000 veces.

2. TERROR: UNA APLICACIÓN DEL PROGRAMA TRAMO

La idea de detectar errores en la producción de datos por medio de los errores de predicción en series temporales no es nueva, y se han desarrollado procedimientos de interés basados en métodos de predicción más o menos *ad-hoc*, como el llamado EWMA (*Exponentially Weighted Moving Average method*), o que utilizan modelos pequeños y muy simples (en muchos casos, autorregresivos con uno o dos parámetros); véanse, por ejemplo, Alwan y Roberts (1988), Lucas y Saccucci (1990) y Montgomery y Mastrangelo (1991). Aunque la ventaja que supondría obtener predicciones ARIMA se menciona con frecuencia, su aplicación se descarta, dados los altos requisitos de complejidad computacional y analistas especializados, que se prevén difíciles de superar.

De todas formas, aunque se pudiese resolver el problema de utilizar modelos ARIMA de forma fiable y rápida, existen algunos problemas previos importantes. En primer lugar, la relación entre error de predicción, observación atípica y error en los datos es bien conocida [(véase, por ejemplo, Barnett y Lewis (1998)] y, en bases de datos que no han sido cuidadosamente controladas (es decir, en la gran mayoría), los errores (no corregidos) del pasado tenderán a presentarse como valores atípicos. Podrán existir también otros fenómenos atípicos que afecten a las observaciones, como, por ejemplo, cambios en las normas de computación del dato o algún tipo de acontecimiento especial. La presencia de estas anomalías en los datos es sabido que puede

distorsionar seriamente la identificación y estimación de modelos ARIMA y, por tanto, también las predicciones. En segundo lugar, las series desagregadas de las bases de datos con frecuencia no están completas, es decir, contienen valores ausentes.

Se trataría, por tanto, de obtener la predicción ARIMA correcta de la serie, una vez interpolados los valores ausentes y corregido el efecto de los valores atípicos. El procedimiento requiere la identificación conjunta, a partir de los datos, del modelo ARIMA y de las observaciones atípicas aun cuando la serie no esté completa.

La idea anterior se puede plasmar, en la práctica, de una forma muy sencilla por medio del programa TRAMO (3). Este programa calcula predicciones óptimas de series utilizando modelos ARIMA combinados con regresión. Permite un funcionamiento enteramente automático, en cuyo caso el procedimiento consiste en lo siguiente. El programa determina la transformación adecuada de la serie, identifica el modelo apropiado para representarla, a la vez que detecta y corrige las observaciones atípicas (o anómalas) que pueden existir, interpola de forma óptima observaciones ausentes y estima, en su caso, efectos especiales de tipo calendario, pascua y días festivos. Los valores atípicos que considera son de tres tipos: aditivos, cambios de nivel y desplazamientos transitorios. El primer tipo representa valores atípicos aislados que afectan solo a un período; el segundo tipo captura un salto en el nivel medio de la serie, que se mantiene en los períodos siguientes; finalmente, el tercero representa saltos en el nivel de la serie que, de forma amortiguada, eventualmente regresan a su nivel anterior. La efectividad del método para «limpiar» series mensuales y trimestrales es grande [(véase, por ejemplo, Fiorentini y Maravall (1996)]. Por último, basándose en el modelo identificado, se obtienen predicciones óptimas (con error cuadrático medio mínimo) para la serie corregida de las distorsiones producidas por los valores atípicos. TRAMO es un programa muy rápido que permite tratar, en un PC ordinario, más de 500.000 series en un día. El programa está fundamentalmente orientado a series mensuales y trimestrales, con un máximo de 600 observaciones y un mínimo de 36 observaciones si la serie es mensual, y 15 si es trimestral.

De todas maneras, la aplicación de TRAMO al problema de la detección de errores requiere

(3) La aplicación de TRAMO para detectar posibles errores nos fue sugerida originalmente por el Banco Nacional de Bélgica. Queremos expresar nuestro agradecimiento a Daniel Desie e Yves de Lombaerde, pioneros en esta aplicación.

prestar un cuidado especial a algunos aspectos del programa. Mencionaremos dos particularmente relevantes. El primero se refiere a la detección de valores atípicos en la última (o penúltima) observación. Supongamos que la serie disponible es $[x(1), \dots, x(T-1)]$ y basándose en ella se obtiene la predicción para el período T. Si el programa detecta un valor atípico para T-1 (el último período de la muestra), es imposible que pueda determinar si ese valor atípico es del tipo aditivo y durará solo un período, o se trata de un cambio de nivel, que permanecerá en la serie. Obviamente, que la observación atípica sea de un tipo u otro afectará drásticamente a la predicción (el valor atípico aditivo prácticamente se ignora, mientras que el cambio de nivel se incorpora plenamente a la predicción). En el período T-1 es imposible identificar el tipo de valor atípico para ese período, y el programa lo trata como aditivo, es decir, no lo considera. Una identificación mínimamente razonable requiere un mínimo de dos o tres períodos adicionales. En consecuencia, parece sensato utilizar una opción que permita detectar, pero no corregir, valores atípicos en la(s) última(s) observaciones. Es una forma de protegerse frente a la incertidumbre de cuál será la corrección final.

El segundo problema que merece la pena mencionar es el de determinar el valor (en términos absolutos) a partir del cual un error de predicción se juzga excesivo. Si, por ejemplo, se pretende aplicar TERROR a 100 series, un tamaño de error que se presenta una de cada mil veces puede ser suficiente. Si se aplica a 100.000 series, la proporción es demasiado alta, ya que (en promedio) se detectarían 100 valores sospechosos que serían completamente espurios; para este caso, una proporción de una en diez millones sería más adecuada. En consecuencia, el criterio para señalar un dato como sospechoso depende, en última instancia, del número de series a analizar. Depende también, por supuesto, del interés del analista, es decir, del grado de finura con que se quieren analizar los posibles errores en el grupo de series de que se trate (lo que el programa llama nivel de sensibilidad).

Con objeto de incorporar estas consideraciones especiales y de utilizar TRAMO de forma eficiente con las especificaciones necesarias, se ha desarrollado una aplicación de TRAMO llamada TERROR, enfocada enteramente al problema de la detección de errores. En la actualidad se encuentra contenida en el programa TRAMO y se ejecuta fijando como parámetro de entrada TERROR = 1 (4). La descripción

(4) El programa TRAMO, con la opción TERROR incluida, se encuentra libremente disponible en la dirección Web del Banco de España (<http://www.bde.es>).

detallada de TERROR puede verse en Luna y Maravall (1999). En esencia, el programa realiza lo siguiente.

El fichero de entrada contiene como última observación (período T) el dato recién recibido, que se pretende clasificar como aceptable o sospechoso de ser erróneo. Para ello, ignorando el último dato, el programa calcula la predicción para T con la serie finalizando en T-1, así como el error estándar de la predicción. Dividiendo el error de predicción (nueva observación menos su predicción) por dicho error estándar, se obtiene un valor t . Si t (en valor absoluto) es menor que un cierto valor t_1 , especificado *a priori*, el dato se clasifica como aceptable. Si $t_1 < |t| < t_2$, donde $t_2 (>t_1)$ también ha sido especificado *a priori*, el dato se clasifica como «posiblemente» erróneo. Finalmente, si $|t| > t_2$, el dato se clasifica como «probablemente» erróneo. El programa permite entrar los valores t_1 y t_2 . Por defecto, son los valores asociados con la probabilidad del 0,01 % y del 0,001 %, respectivamente. El fichero de salida del programa presenta los pares de valores (j, t_j) asociados con las observaciones posiblemente y probablemente erróneas; j indica el número de orden de la serie en el fichero de entrada y t_j representa el valor t asociado con su error de predicción.

Realizamos simulaciones de series generadas por diversos modelos ARIMA y a algunas de las series se les cambió la última observación añadiendo o sustrayendo valores equivalentes a t_1 y t_2 multiplicados por el error estándar residual. De una parte, el objetivo era escoger las opciones de TRAMO más indicadas. Concretamente, se comparó el método de estimación exacta máximo-verosímil con el de mínimos cuadrados incondicionales. Se compararon también diversas combinaciones de los cuatro tipos de anomalías que ofrece TRAMO (los tres descritos anteriormente más el valor atípico denominado «innovacional», que se propaga en función del modelo que sigue la serie). Se compararon diversos valores críticos en la detección de anomalías. Finalmente, se compararon las opciones de detectar y no corregir valores atípicos en la última y en las dos últimas observaciones. Las conclusiones fueron que máxima verosimilitud era preferible a mínimos cuadrados, que la combinación óptima de valores atípicos era la compuesta por los tres tipos descritos al comienzo, que los valores críticos que TRAMO usa por defecto en la detección de anomalías eran apropiados y, por último, se optó por no corregir la última observación de la muestra (es decir, la correspondiente al período anterior al del nuevo dato que se analiza). En este último caso la evidencia no daba una respuesta clara y se trata de un punto que la experiencia ayudará a resolver.

El segundo objetivo del ejercicio de simulación era evaluar el funcionamiento de TERROR aplicado a las series simuladas y, en algunos casos, perturbadas en su última observación. Los resultados obtenidos fueron excelentes. Sin embargo, desde el punto de vista de la aplicación a series reales, la simulación proporcionaba un resultado sesgado porque las series eran generadas precisamente por modelos del tipo de los considerados por TERROR y estos modelos podrían ser poco adecuados para las series reales en muchas aplicaciones, que pueden ser altamente erráticas y muy cortas. Se procedió, por tanto, a realizar un pequeño experimento con un grupo de series reales que presenta de manera acentuada dichos inconvenientes.

3. LA APLICACIÓN EMPÍRICA

En principio, uno esperaría que la mejora que se pudiera producir por el uso de TERROR resultaría del hecho de que permitiría llegar a niveles mucho más desagregados. Al nivel de agregación al que se realiza el control en el presente cabría pensar que un análisis manual cuidadoso debería producir resultados mejores que un método puramente automático. Tiene interés, por tanto, comparar, para ese nivel de agregación, los resultados del control actual y los que produce la aplicación rutinaria y automática de TERROR.

Para esta comparación escogimos un grupo de series trimestrales sobre actividad bancaria internacional. Se trata de unas series que se envían al Banco de Pagos Internacionales y que se utilizan también en Balanza de Pagos y en Contabilidad Nacional. Consideramos las series para cada país, agregando las entidades y el total monedas. Una descripción más completa de la aplicación se encuentra en Luna y Maravall (1999).

Previamente, se eliminaron del grupo las series con menos de 15 observaciones, así como las de poca importancia (aquellas cuya media y últimas observaciones eran inferiores a 1.000 millones de pesetas) y las series cuyos últimos años estaban fuertemente dominadas por valores iguales a cero o no disponibles. Se obtuvo de esta forma un grupo de 436 series que, a pesar de ser muy heterogéneo en lo que se refiere al comportamiento de las series, presentan dos características importantes:

- Las series tienen, en general, una erraticidad grande, con fuertes oscilaciones que no parecen seguir pauta alguna.
- Se trata de series cortas que contienen entre 15 y 47 observaciones.

Se aplicó TERROR a la vez que se iniciaba el procedimiento de control manual, para los 4 trimestres de 1997. Se obtuvieron, por tanto, un total de 1744 comparaciones. Antes de comentar estas es importante resumir los resultados que produjo TERROR al tratar automáticamente las 436 series. Concretamente, interesa analizar el ajuste de los modelos identificados, y el diagnóstico de su estimación.

3.1. Identificación

Un problema que se plantea para la identificación automática del modelo es que, para 103 de las series, el número de observaciones es insuficiente. Para estas series el programa adopta automáticamente el modelo de defecto, dado por el modelo ARIMA multiplicativo $(0,1,1)(0,1,1)_4$, también llamado, desde que Box y Jenkins (1970) lo popularizaran, modelo de las Líneas Aéreas. Se trata del modelo más frecuentemente usado para series trimestrales [véase el estudio a gran escala de Fischer y Planas (1998)] y presenta la virtud de aproximar bien una gran cantidad de modelos [véase Gómez y Maravall (1998b)]. Los resultados que se presentan incluyen estos 103 casos y sesgan a la baja, por tanto, el resultado de la modelización automática, puesto que, cuando se hayan observado algunos trimestres más, los resultados para estas 103 series podrán ser, en muchos casos, mejorados.

De las 436 series, la gran mayoría (92 %) muestra un comportamiento no-estacionario y tan solo un 8 % pueden verse como oscilaciones en torno a un nivel relativamente fijo. Por otra parte, aproximadamente un 50 % contiene estacionariedad y un 50 % no la contiene. Si representa la diferencia entre dos períodos consecutivos y Δ_4 la diferencia interanual, cerca de un 50 % de las series no-estacionarias necesita la transformación Δ_4 para transformarse en estacionarias; del 50 % restante, un 90 % requiere tan solo la transformación Δ , un 6 %, la transformación Δ_4 , y un 4 %, la transformación Δ^2 . Para las 436 series, el programa identificó un total de 38 modelos diferentes. Sin tener en cuenta los 103 casos en que el número de observaciones no llegaba al mínimo para la identificación automática, en cuyo caso el programa impone el modelo de Líneas Aéreas mencionado anteriormente, los modelos que con más frecuencia identifica el programa son: a) Líneas Aéreas, 31 % de los casos; b) ARIMA $(0,1,1)$, 24 % de los casos; c) ARIMA $(1,1,0)$, 17 % de los casos, y d) AR (1) , 8 % de los casos. El 20 % restante se lo reparten los otros 34 modelos. Por último, es de destacar que el número medio de parámetros por modelo (sin incluir la varianza residual) es

CUADRO 1

Observaciones atípicas por serie

	Anomalía aditiva	Despla- zamiento transitorio	Cambio de nivel	Total
Media	0,84	0,63	1,09	2,56
Desviación típica	1,20	0,90	1,39	2,33
Máximo	6	5	8	12
Mínimo	0	0	0	0

1,5, con un máximo de cuatro y un mínimo de cero.

En cuanto a los valores atípicos, el cuadro 1 presenta, para cada uno de los tres tipos de anomalías, y para las 436 series, el número medio detectado por serie, la desviación típica de esta media, así como los números más altos y más bajos detectados en cada caso.

El número medio de valores atípicos detectados por serie es 2,56, con una frecuencia relativa ligeramente más alta para los cambios de nivel (aproximadamente, uno por serie), seguido de las anomalías aditivas. El gráfico 1 presenta la distribución del número de observaciones atípicas.

En resumen, los modelos que TRAMO, en la modalidad TERROR, identifica para las series son considerablemente parsimoniosos. Dada la erraticidad de las series de partida, uno o dos parámetros y dos o tres valores atípicos por serie resultan razonablemente bajos.

3.2. Estimación

El cuadro 2 presenta los resultados de la estimación. La primera columna contiene el error estándar medio (y su desviación típica) de los residuos de la estimación (dividido por el valor medio de los últimos ocho trimestres cuando el modelo se ha estimado para los niveles y no para los logaritmos). En promedio, los modelos explican una proporción importante de las desviaciones que experimenta la serie. El gráfico 2 presenta unos ejemplos de series originales y de los valores que ajusta el modelo a la serie; los ejemplos corresponden a modelos con un número de parámetros y de valores atípicos cercano al medio. Los ajustes obtenidos son claramente satisfactorios.

3.3. Diagnóstico

Pasando a la fase de diagnóstico, la segunda columna del cuadro 2 presenta el valor me-

CUADRO 2

Resultados de la modelización automática

	Error estándar de los residuos	Estadístico Ljung-Box	Test de normalidad	Tiempo de ejecución (en segundos)
Media	0,24	12,5	2,3	0,11
Desviación típica	0,25	4,9	7,2	0,05
Valores críticos aprox. (al 95 %)	—	18	6	—

dio (y su desviación típica) del estadístico Q de Ljung-Box, utilizado para contrastar la hipótesis de que los residuos no contienen autocorrelación significativa. Dado que el valor crítico, al 95 % de significatividad, es aproximadamente 18, los residuos de los modelos parecen perfectamente compatibles con la hipótesis de ausencia de autocorrelación. El gráfico 3 presenta el histograma de los 436 valores de Q, y lo compara con la función Chi-cuadrado que sería la distribución asintótica de Q para el modelo «promedio». La proximidad del histograma con la distribución asintótica es evidente.

La tercera columna del cuadro 2 presenta el estadístico de Behra-Jarque para contrastar la hipótesis de que los residuos se distribuyen normalmente. La distribución asintótica de este estadístico es una variable Chi-cuadrado con dos grados de libertad, de forma que el valor crítico (también al 95 %) es aproximadamente igual a seis. De nuevo en una gran mayoría de las series los residuos se pueden aceptar como normales, aunque en este caso se producen algunos valores aislados excesivamente altos. El gráfico 4 muestra cómo, también para este contraste, el histograma del estadístico y la distribución asintótica están considerablemente cercanos.

GRÁFICO 1

Distribución de anomalías

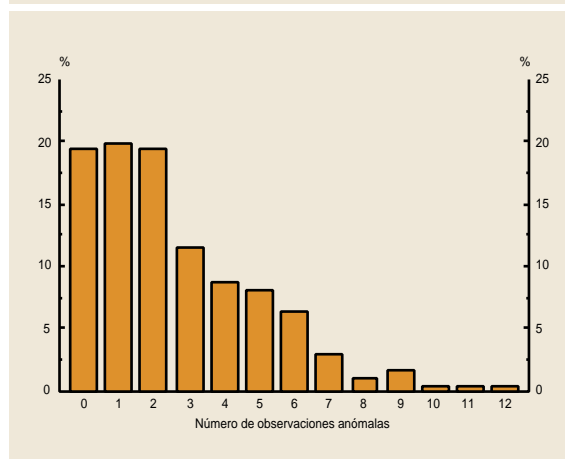
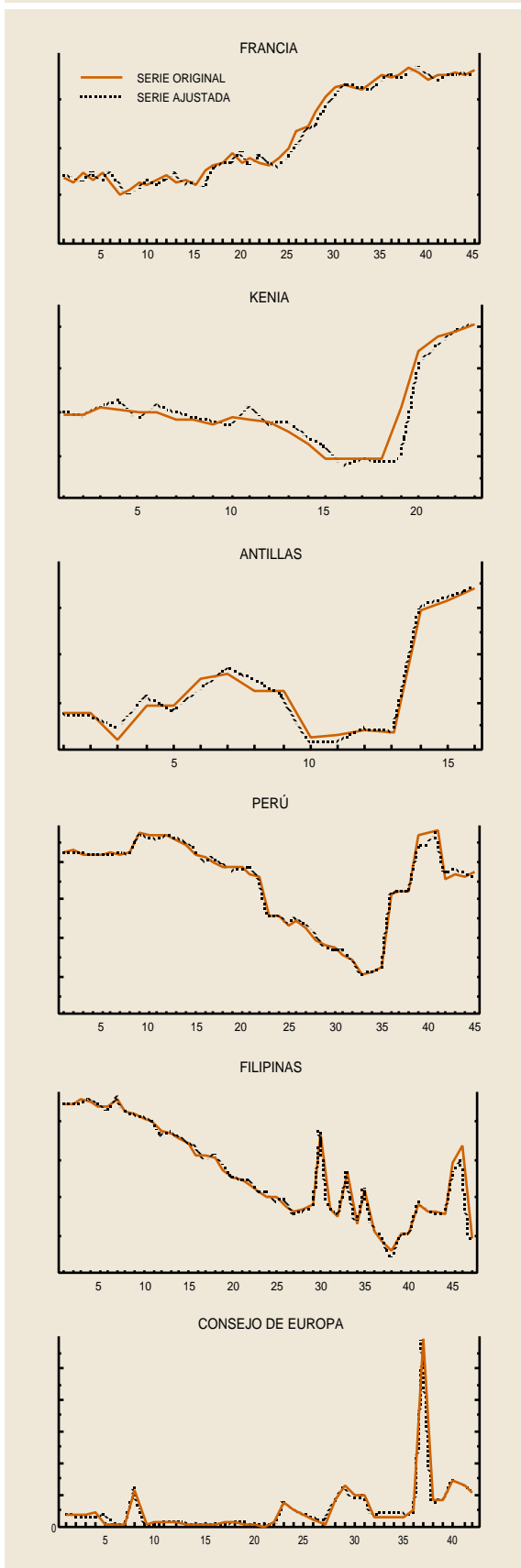


GRÁFICO 2

Series originales y ajustadas



Merece la pena señalar que los diagnósticos, así como los histogramas, se han obtenido incluyendo también los 103 casos en que la longitud de la serie no permite identificación automática y se usa por tanto el modelo de defecto. Los resultados muestran el buen comportamiento que presenta este modelo, y que ya mencionamos anteriormente.

Como resumen del análisis, de los resultados que produce TERROR sobre este grupo de series, es de destacar que:

- Los modelos obtenidos son parsimoniosos, con pocos parámetros y relativamente pocos valores atípicos corregidos.
- Los modelos capturan gran parte de la variación de las series y producen buenos ajustes.
- Los diagnósticos obtenidos son excelentes.
- Merece mención el buen funcionamiento del modelo de defecto en series muy cortas.

4. COMPARACIÓN DE LOS RESULTADOS DE TERROR CON LOS DEL MÉTODO MANUAL

Los nuevos datos correspondientes a los cuatro trimestres de 1997 se chequearon para las 436 series, lo cual significa un total de 1.744 chequeos. El procedimiento de chequeo actual de este grupo de series detectó, en total, 132 datos sospechosos. TERROR detectó 114, y solamente coincidieron los dos procedimientos en 24 ocasiones, un porcentaje sorprendentemente bajo.

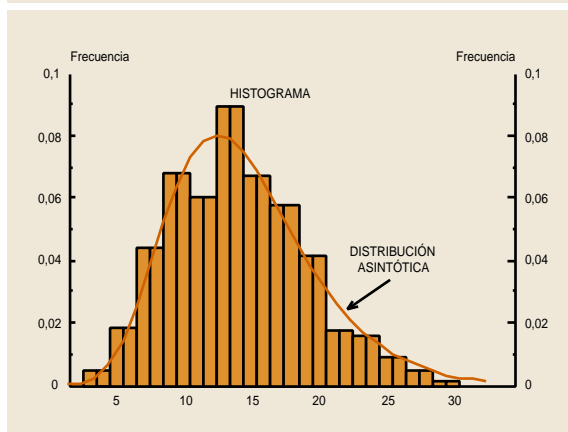
Se realizaron consultas para 85 de los 132 casos detectados manualmente. Una de las consultas no obtuvo respuesta y, de las 84 restantes, seis datos erróneos fueron encontrados y corregidos. La proporción es, por tanto, de un error detectado por cada 14 consultas.

De los 114 datos sospechosos detectados por TERROR, tan solo se realizaron 42 consultas, de las cuales 6 no obtuvieron respuesta. De las 36 contestaciones se obtuvieron los seis errores anteriores (es decir, todos los detectados por el procedimiento usual), más tres errores adicionales que la detección manual no había capturado. Es decir, en el caso de TERROR, se detectó un error por cada cuatro consultas.

Aunque la muestra es pequeña, los resultados de la comparación son elocuentes:

GRÁFICO 3

Distribución del test de autocorrelación



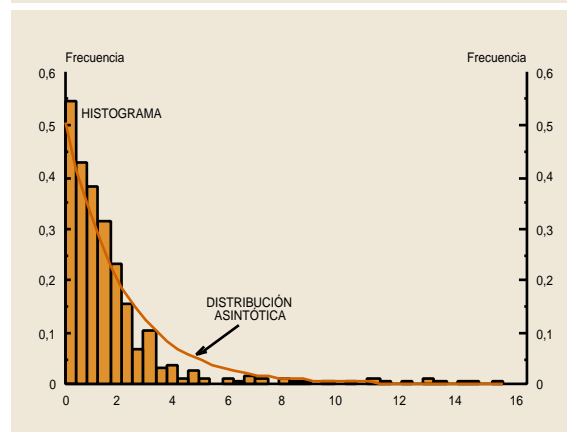
- TERROR reduce el número de consultas innecesarias.
- TERROR captura los errores detectados con el procedimiento actual y un 50 % de errores adicionales.

Dado que el tiempo medio empleado por TERROR es de 0,1 segundos por serie en un Pentium II, el nuevo método para detectar errores parece considerablemente más preciso y más eficiente. En términos prácticos, el método nuevo implicaría sustituir el programa, que hoy se usa para obtener el primer listado con las tasas de crecimiento, por TERROR; la parte manual desaparecería totalmente.

En el grupo de series analizado, la frecuencia de error que detecta la aplicación de TERROR es de nueve errores en 42 consultas realizadas sobre 114 observaciones sospechosas de ser erróneas, de un total de 1.744 observaciones. Es difícil traducir el resultado a una simple proporción del total de errores existentes en las observaciones originales, pero el resultado final parece estar razonablemente de acuerdo con la frecuencia de error, en bases de datos controladas, fuera del campo de la economía [a *grosso modo*, entre un 1 % y un 5 % de los datos; véase Barnett y Lewis (1998)]. En la reducida aplicación que hemos analizado, la frecuencia de error parece quizá ligeramente alta, pero esto se puede explicar fácilmente por la combinación de tres factores: la inestabilidad de las series, el buen ajuste proporcionado por TERROR, que implica errores estándar de predicción relativamente bajos, y los valores relativamente altos usados para t_1 y t_2 . Cuanto más elevados sean los valores de t_1 y t_2 , menos observaciones sospechosas serán detectadas, pero la probabilidad de que una observación sospechosa sea un error aumenta. (Una venta-

GRÁFICO 4

Distribución del test de normalidad



ja adicional de TERROR es que su aplicación sistemática, al cabo de cierto tiempo, podría mejorar la estructura de las series, que se harían más estables. Las series serían, por tanto, más modelizables, sus predicciones más precisas y los errores grandes, como consecuencia, más detectables.)

5. CONCLUSIÓN

La sustitución del procedimiento (que hemos denominado) manual por TERROR implica una diferencia metodológica importante. Mientras que la detección manual considera dos observaciones de la serie, TERROR utiliza de forma eficiente la información completa de la misma. Es fácil imaginar situaciones en las que la comparación de las dos últimas tasas puede dar indicaciones erróneas. Un ejemplo sería una serie con una estacionalidad tal que la tasa de crecimiento para el período de que se trata implicase un bajón brusco. Otro ejemplo sería una serie sistemáticamente errática, en la que las tasas aisladas pueden ser aparentemente disparatadas. La modelización que realiza TERROR tiene en cuenta todo este pasado «sistemático» de la serie.

Por otra parte, la velocidad de proceso de información de TERROR es muy alta (permite tratar más de 30.000 series por hora), de forma que es posible analizar series a niveles altos de desagregación, a la vez que permite reducir drásticamente las necesidades de «analistas manuales».

Hemos visto que, incluso al nivel de desagregación al que se realiza la detección manual básica, la aplicación de TERROR implica la realización de menos consultas a las instituciones a la vez que la detección de más errores. Utili-

zando niveles mayores de desagregación, la ganancia sería, sin duda, más grande.

Merece la pena mencionar una posible aplicación del programa TERROR que puede ser de interés. La detección de observaciones sospechosas puede indicar un error en los datos, pero puede también indicar la ocurrencia de algún suceso anómalo de importancia. Por citar un ejemplo, una bajada relativamente generalizada de los activos bancarios frente a un país en concreto podría ser un indicador adelantado de la inminencia de una crisis financiera (o una próxima devaluación) en dicho país. La salida del programa podría ser, pues, de interés complementario en el análisis de coyuntura.

19.5.1999.

BIBLIOGRAFÍA

- ALWAN, L. C. y ROBERTS, H. V. (1988). «Time-Series Modeling for Statistical Process Control», *Journal of Business and Economic Statistics*, 6, pp. 87-95.
- BARNETT, V. y LEWIS, T. (1998). *Outliers in Statistical Data*, Nueva York, J. Wiley and Sons.
- BOX, G. E. P. y JENKINS, G. M. (1970). *Time Series Analysis: Forecasting and Control*, San Francisco, Holden Day.
- FIORENTINI, G. y MARAVALL, A. (1996). «Unobserved Components in ARCH Models: An Application to Seasonal Adjustment», *Journal of Forecasting*, 15, pp. 175-201.
- FISCHER, B. y PLANAS, C. (1998). *Large Scale Fitting of ARIMA Models and Stylized Facts of Economic Time Series*, Eurostat Working Paper 9/1998/A/8.

GÓMEZ, V. y MARAVALL, A. (1997). «Programs TRAMO and SEATS: Instructions for the User», Documento Interno, Servicio de Estudios, Banco de España.

— (1998a). «TRAMO and SEATS: Programs Update, June 1998 version», Documento Interno, Servicio de Estudios, Banco de España

— (1998b). *Automatic Modeling Methods for Univariate Series*, Documento de Trabajo nº 9808, Servicio de Estudios, Banco de España.

LUCAS, J. M. y SACCUCCI, M. S. (1990). «Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements (with discussion)», *Technometrics*, 32, pp. 1-29.

LUNA, C. y MARAVALL, A. (1999). «Control de calidad de los datos en series temporales: TERROR, una aplicación del programa TRAMO», Documento Interno, Servicio de Estudios, Banco de España (de futura publicación como Documento de Trabajo).

MONTGOMERY, D. C. y MASTRANGELO, C. M. (1991). «Some Statistical Process Control Methods for Autocorrelated Data», *Journal of Quality Technology*, 23, pp. 179-204.

Rectificación

En relación con los programas TRAMO y SEATS, el último párrafo de la sección 4.e) del artículo «El Banco de España en la red Internet» (*Boletín económico* de octubre 1996) se deslizó una frase que podría inducir a confusión. Los autores de dichos programas son Agustín Maravall, del Servicio de Estudios del Banco de España, y Víctor Gómez, de la Dirección General de Análisis y Programación Presupuestaria, del Ministerio de Economía y Hacienda.