

PRÁCTICA 11. VALORES ANÓMALOS “OUTLIERS” Y EFECTO CALENDARIO

Los modelos ARIMA constituyen una forma práctica de representar la dinámica de una serie. Ya hemos introducido en las prácticas anteriores el modo de detectar la existencia de tendencia y estacionalidad y cómo tratar estos componentes en el marco de los modelos ARIMA. Sin embargo, antes de considerar un modelo ARIMA válido para realizar predicciones puede ser conveniente introducir algunos ajustes previos. En particular, en esta práctica vamos a hablar de cómo incorporar el efecto calendario y cómo tratar a los valores anómalos u *outliers*.

EFECTO CALENDARIO

Por efecto calendario se entiende el efecto de las fechas del calendario sobre la serie temporal. Ejemplos:

Efecto Pascua. Puede ser relevante para explicar el comportamiento de ciertas series temporales, como el tráfico por carretera o la entrada de turistas en una zona. A diferencia de otras vacaciones, las fechas de Pascua cambian año tras año, afectando a los meses de marzo y abril. Su efecto se puede recoger incorporando una variable explicativa que contendrá observaciones sobre los días de vacaciones mensuales debidas a la Pascua (por lo tanto, las observaciones de los distintos meses serán cero, excepto las de marzo y abril que irán variando año tras año)

Efecto calendario laboral. Las series relacionadas con la producción, tanto a nivel microeconómico como macroeconómico dependen del número de días laborales de cada mes. Estos días irán variando en función de los fines de semana que contiene el mes, de si se trata de un mes con 30 o 31 días, de los días festivos, de los años bisiestos etc. Para refinar el modelo que caracteriza a algunas series puede incluirse una variable explicativa que capte el número de días laborales de cada mes del periodo considerado (incluyendo el periodo de predicción)

OUTLIERS

Los outliers representan cambios abruptos que no pueden explicarse bajo el supuesto de normalidad subyacente en los modelos ARIMA.

- Importancia de la detección de outliers

Si una serie temporal tiene observaciones anormalmente elevadas (o reducidas), y no se tratan de modo especial, podrían dominar los resultados obtenidos de la estimación. A modo de intuición recuérdese que MCO minimiza la suma de cuadrados de los residuos, y el cuadrado del residuo de estos valores será anormalmente alto. Por lo tanto, el estimador tenderá a sobreponderar estos valores anómalos, afectando a los parámetros estimados y reduciendo el ajuste total del modelo.

En los modelos ARIMA, los outliers pueden distorsionar la identificación del modelo, sesgar la estimación de los parámetros, de los componentes tendencia y estacionalidad y afectar a la predicción.

- Modo de detección de outliers

¿Cómo detectar desviaciones de una pauta de comportamiento normal?

Con carácter general se pueden estandarizar los valores de una serie y compararlos con un nivel mínimo preasignado C . Lo que es equivalente a comparar las observaciones de la serie con unas bandas de confianza construidas a partir de la media y un múltiplo preasignado (C) de la desviación típica. Como regla de aproximación: 3 veces la desviación típica ($C=3$), lo que implica que la probabilidad de que la observación haya sido extraída de una distribución normal es menor de 1/1000. El Gráfico 1 representa una serie en la que se ha recogido un intervalo de confianza con $C=2$. El Gráfico 2 representa una serie con unos intervalos de confianza de 2, 3 y 4 veces la desviación típica. ¿Detecta algún outlier en el gráfico 1? ¿Y en el gráfico 2? ¿Tienen la misma media y varianza la serie 1 y la serie 2?

En modelos ARIMA se hace uso de la propiedad de normalidad del ruido blanco Gaussiano para utilizar los residuos como método de detección.

- Causas de los outliers
- a) La aparición de un outlier puede tener un motivo identificable. Ejemplos: el consumo de energía de los hogares puede dispararse durante un invierno especialmente frío; la inversión productiva privada puede verse afectada por acontecimientos políticos internacionales relevantes; la entrada de turistas en una región puede sufrir cambios importantes después de la apertura de una nueva vía de comunicación.
 - b) En algunas ocasiones no es sencillo encontrar la causa que provoca el outlier. Se trata de casos en los que es probable que el outlier se deba a algún error en el proceso de tratamiento de la información

- Tratamiento de los outliers
- a) Sustituir el outlier por una media móvil ponderada *ad-hoc* de los valores cercanos. Esta solución no es eficiente, pues descarta información relevante.
 - b) Incorporar el outlier en el modelo por medio de variables ficticias, para controlar por su efecto.

- Modelización

$$z_t^* = \omega v(L)I_t(T) + z_t \quad (1)$$

siendo z^* la serie en presencia de outlier, z la serie libre de outlier, I_t una dummy que indica el momento del tiempo en que se produce el outlier, $v(L)$ el proceso que caracteriza al outlier (su efecto a lo largo del tiempo), w la magnitud del outlier.

- Tipos de outliers

- a) Outlier aditivo (AO): $v(L)=1$
- b) Cambio temporal (TC): $v(L)=1/(1-aL)$ $a<1$
- c) Cambio de nivel (LS): $v(L)=1/(1-L)$

(recordar representación gráfica)

- Estimación en presencia de outliers

Un problema en la detección y tratamiento de outliers es que el modelo estimado puede condicionar los outliers detectados que a su vez influyen en la identificación y estimación del modelo. Por ello, el programa TRAMO utiliza un método iterativo (ver Chen y Liu, 1993) que, estima conjuntamente los parámetros del modelo y el efecto de los outliers y que opera en tres etapas:

ETAPA I

Todos los potenciales outliers son identificados basados en estimaciones preliminares de los parámetros del modelo.

ETAPA II

Se obtiene la estimación conjunta de los parámetros del modelo y de los efectos de los outliers usando la información acumulada sobre los outliers de la etapa I.

ETAPA II

Se vuelven a identificar los outliers (ETAPA I) y a estimar sus efectos (ETAPA II) basándose en estimaciones del valor de los parámetros menos contaminadas que las anteriores.

La detección de outliers por el método anterior depende del valor crítico utilizado (que hemos llamado C) ¿Cuál es un valor sensato de C ? Solución de compromiso entre un valor reducido que implicaría considerar outliers a observaciones que no lo son, y un valor elevado que dejaría fuera de consideración observaciones que sí son outliers.

Como una regla de aproximación, el número de outliers suele ser menor que el 3% - 4% de las observaciones totales.

Analizar la sensibilidad de los resultados a cambios en C . Valores razonables: (2.5-3.5)

Referencia:

Chen, C. Y Liu, L.M. (1993): "Joint estimation of model parameter and outlier effects in time series". *Journal of the American Statistical Association*, 88, 284-297

Gráfico 1

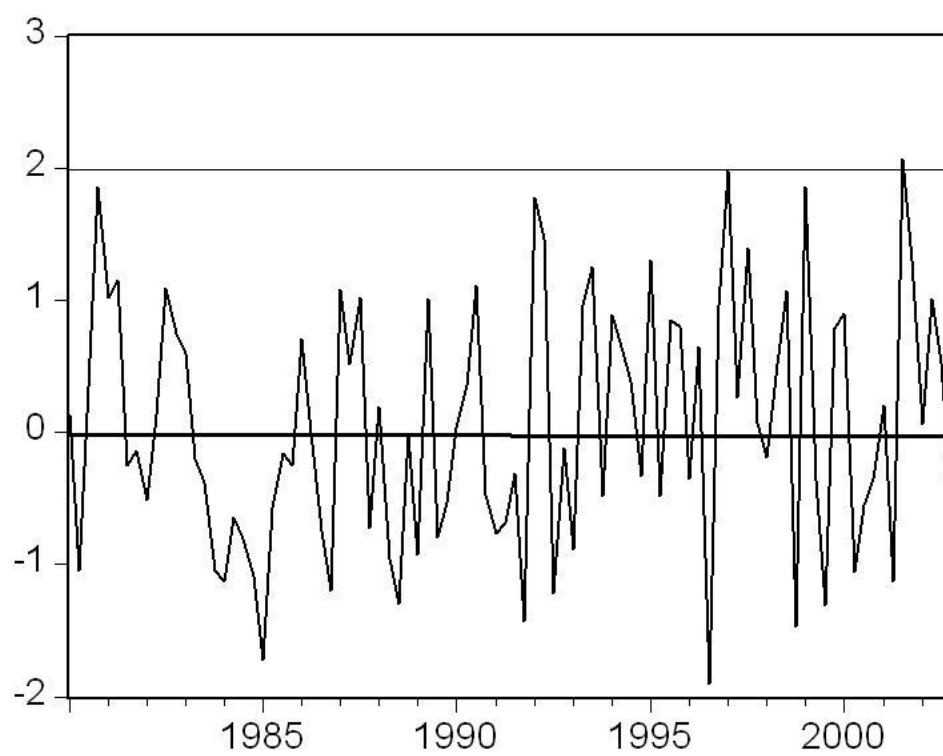


Gráfico 2

