Prácticas de la asignatura Series Temporales

Sexta Entrega

1 Análisis de intervención en series temporales

Hasta ahora hemos estudiado modelos para series temporales que únicamente dependían del pasado de la propia serie. En esta entrega vamos a iniciar el análsis de una serie temporal teniendo en cuenta algún tipo de información externa. El análisis de modelos ARIMA nos permite estudiar las correlaciones y correlaciones parciales de una serie en función de su pasado. Basandonos en estas correlaciones obteniamos un modelo y una vez estimado, predicciones. Veamos un ejemplo donde claramente este procedimiento va a fallar claramente sin el uso de información exterior. Supongamos que medimos el número de viajeros diarios en tren. Si un día $t = h$ se produce una huelga de maquinistas, el número de viajeros va a sufrir un bajón fortuito el día $t = h$. Por lo tanto, en el tiempo $t = h$, el número de viajeros varió de alguna forma debido a una intervención externa a la dinámica de la serie. Este tipo de efectos es lo que estudia el análisis de intervención. Por ejemplo, en el caso anterior, podemos medir el efecto producido por la huelga, mediante una variable que tome el valor 1 en el tiempo $t = h$. Esta variable es conocida como variable impulso. Sea y_t una serie que sigue un modelo ARIMA, $\phi(B)(1-B)^d(y_t-\mu) = \theta(B)a_t$, que podemos representar por $(y_t-\mu) = \phi(B)^{-1}(1-B)^{-d}\theta(B)a_t = \psi(B)a_t$. Entonces, supongamos que la serie y_t está afectada por un efecto impulso. Entonces la serie que observamos es:

$$z_t = y_t + w_0 I_t^{(h)}$$

donde $I_t^{(h)}$ vale 1 si $t = h$ y vale 0 si $t \neq h$. Por lo tanto, en tiempo $t = h$, observamos $z_h = y_h + w_0$. Este efecto puede ser más complicado e ir desapareciendo con el tiempo. Por ejemplo, supongamos que la huelga dura varios días, de tal manera que el número de maquinistas en huelga va decreciendo con el tiempo. Esto produce que el número de viajeros se vaya incrementando a lo largo de los días. Entonces podemos escribir:

$$z_t = y_t + w(B)I_t^{(h)} = y_t + (w_0 + w_1 B + \ldots + w_m B^m)I_t^{(h)}$$

de tal manera que observamos,

$$\begin{aligned} z_h &= y_h + w_0 \\ z_{h+1} &= y_{h+1} + w_1 \\ z_{h+2} &= y_{h+2} + w_2 \\ &\vdots \end{aligned}$$

de tal manera que $w_j \to 0$, y el efecto desaparece de la serie. Como esto incluye varios parámetros a estimar, se suele utilizar el siguiente modelo que se da en llamar cambio transitorio:

$$z_t = y_t + \frac{w_0}{1-\delta B}I_t^{(h)} = y_t + w_0 I_t^{(h)} + \delta w_0 I_{t-1}^{(h)} + \delta^2 w_0 I_{t-2}^{(h)} + \cdots$$

donde $0 < \delta < 1$. De esta manera únicamente debemos estimar w_0.

Otro tipo de efecto es el siguiente. Supongamos la serie que mide las ventas mensuales de un producto. Supongamos que es conocido que en un momento $t = h$ se realiza una promoción de dicho producto de tal manera que su precio bajó y que además se llevó a cabo una campaña en prensa, radio y televisión. Esto produjo un aumento de las ventas de dicho producto. Por lo tanto, en el tiempo $t = h$, el nivel de las ventas varió de alguna forma debido a una intervención externa. Este tipo de efectos es lo que estudia el análisis de intervención. Por ejemplo, en el caso anterior, podemos medir el efecto producido por la promoción del producto, mediante una variable que tome el valor 1 en los tiempos $t = h,\ h + 1,\ \ldots, n$. Esta variable es conocida como variable escalón:

$$z_t = y_t + w_0 S_t^{(h)}$$

donde $S_t^{(h)}$ vale 1 si $t \geq h$ y vale 0 si $t < h$. Por lo tanto, para los tiempos $t \geq h$, observamos $z_t = y_t + w_0$. Es decir, el nivel de la serie aumenta w_0 en todo punto a partir de $t = h$. Este efecto puede ser más complicado de tal manera que el nivel de la serie varíe hasta llegar al definitivo. Por ejemplo, supongamos que la promoción aumenta con los días, de tal manera que las ventas del producto van aumentando hasta alcanzar el nivel definitivo. Entonces podemos escribir:

$$z_t = y_t + w(B) S_t^{(h)} = y_t + (w_0 + w_1 B + \ldots + w_m B^m) S_t^{(h)}$$

de tal manera que observamos,

$$\begin{aligned}
z_h &= y_h + w_0 \\
z_{h+1} &= y_{h+1} + (w_0 + w_1) \\
&\vdots \\
z_{h+m} &= y_{h+m} + (w_0 + \cdots + w_m) \\
z_{h+m+1} &= y_{h+m+1} + (w_0 + \cdots + w_m) \\
&\vdots
\end{aligned}$$

de tal manera que pasados m datos alcanzamos el nivel definitivo de la serie. También podemos utilizar el modelo:

$$z_t = y_t + w(B) S_t^{(h)} = y_t + w_0 S_t^{(h)} + w_1 S_{t-1}^{(h)} + \ldots + w_m S_{t-m}^{(h)}$$

Cómo antes también podemos utilizar el modelo:

$$z_t = y_t + \frac{w_0}{1 - \delta B} S_t^{(h)} = y_t + w_0 I_t^{(h)} + (1 + \delta) w_0 I_{t-1}^{(h)} + \left(1 + \delta + \delta^2\right) w_0 I_{t-2}^{(h)} + \cdots$$

donde $0 < \delta < 1$. De esta manera únicamente debemos estimar w_0.

2 Estimación de modelos ARIMA con análisis de intervención con el programa TRAMO

Veamos como estimar una serie con análisis de intervención con el programa TSW (Tramo-Seats de windows). Abrimos el programa y vemos como aparece una ventana con el programa, con un menu de botones y con una serie de ranuras. Lo primero es importar los datos que están en la página web:

http://halweb.uc3m.es/esp/Docencia/serieslic.html

con el nombre de cresttramo.txt. La serie son datos semanales que empiezan en la primera semana de Enero de 1958 y acaban en la última semana de marzo de 1963 por lo que son 275 datos. Esta serie se puede modelizar mediante un ARIMA(0,1,1). En la última semana del mes de Julio de 1960, se produjo un aumento de ventas debido a que se anuncio que la inclusión de fluor en el dentífrico era beneficioso para los dientes lo que hizo que las ventas del dentífrico de crest aumentaran, mientras que disminuian las de colgate, que no incluía fluor.

Vemos como el programa TSW busca este tipo de efectos de manera directa sin necesidad de que nosotros los determinemos. Marcamos Series y a continuación seleccionamos la serie. En Series list, vemos que tenemos la serie incluida en el programa. Si marcamos la serie aparece el nombre de la serie y ciertos atributos que podemos cambiar. Por ejemplo, el parámetro Iter si lo cambiamos a Iter=1, a la misma serie le podemos ajustar diferentes modelos. Lo mismo con tramo para utilizar solo el programa tramo. También aparece el gráfico de la serie, lo que nos permite llevar a cabo una rápida inspección. Puesto que ya hemos determinado el modelo ARIMA(0,1,1), marcamos +Model. El parámetro RSA lo dejamos en 0. Pasamos a ARIMA model y marcamos:

1. P=0.
2. Q=1.
3. D=1.
4. BP=0.
5. BQ=0.
6. BD=0.
7. INIT=0, este parámetro sirve para inicializar la estimación de los parámetros.
8. IMEAN=0, este parámetro sirve para corregir la media o no. Esto significa que no corregimos la media.
9. LAM=1, este parámetro dice si tranformar logaritmicamente los datos. Vale 0 si se quiere transformar, 1 si no, y -1 si el programa realiza un contraste sobre la necesidad de transformar.
10. FCT=1, relacionado con el contraste, se deja así.
11. TYPE=0, para estimar MLE exacta. Vale 1 para LSE.
12. UNITS=0, se deja asi, es para cambiar la escala si los datos son muy pequeños o grandes.

Marcamos Others para nuevos parámetros. De todos estos valores, sólo nos interesan los correspondientes a Outliers. Por ahora marcamos IATIP=1, lo que significa que además de estimar los parámetros, busque atípicos y estime sus efectos. Marcamos OK. Marcamos el modelo en la tabla y marcamos Run. En output, vemos la salida. Comprobamos como han aparecido los siguientes valores:

$$
\begin{array}{c}
\text{135 } LS \\
\text{166 } AO \\
\text{195 } TC
\end{array}
$$

El dato en 135 corresponde a un escalon (cambio de nivel), al dato en 166, le asigna un impulso y el dato en 195 corresponde a un cambio transitorio, o impulsos sucesivos que van decreciendo con el tiempo.

Vemos como afecto el cambio con la serie colgatetramo.txt, utilizando los mismos parámetros y el mismo modelo. Marcamos el OK y comprobamos que no se detecta ningún tipo de cambio aunque es evidente que el nivel de la serie cambia.

3 Estimación de modelos ARIMA con análisis de intervención con el programa E-views

Veamos como estimar una serie con análisis de intervención con el programa E-views. Abrimos el programa y creamos un workfile. La serie son datos semanales que empiezan en 1:1:1958 y acaban en 3:27:1963 por lo que son 275 datos. Importamos los datos que están en la página web:

$$http://halweb.uc3m.es/esp/Docencia/serieslic.html$$

con el nombre de crestcolgate.txt. El fichero contiene cuatro series, las importamos con los nombre x1, x2, x3 y x4. Primero vemos la serie x1 que es la correspondiente a las ventas de crest. La observación del cambio de nivel corresponde a la observación 7:27:60. Generamos una serie tal que:

$$\begin{aligned} Genr &\rightarrow s1 = 0 \\ Genr &\rightarrow s1 = 1, \text{ sample: } 7:27:1960, 3:27:1963 \end{aligned}$$

y pasamos a estimar el modelo ARIMA(0,1,1) con un cambio de nivel. Estimamos el siguiente modelo ARIMA(0,1,1) con un cambio de nivel en el dato 7:27:60. Para ello y teniendo en cuenta que tenemos que tomar una raiz unitaría, estimamos el modelo como sigue:

$$Quick \rightarrow Estimate\ Equation \rightarrow d(x1,1)\ d(s1,1)\ ma(1)$$

Obtenemos el siguiente modelo:

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(S1,1)	0.13	0.030	4.31	0
MA(1)	-0.76	0.04	-18.66	0

Además podemos ver un efecto que puede ser un impulso en el dato 1/10/1962, es decir, en la segunda semana del mes de enero. Para ello, definimos una serie como sigue:

$$\begin{aligned} Genr &\rightarrow s2 = 0 \\ Genr &\rightarrow s2 = 1, \text{ sample: } 1:10:1962, 1:10:1962 \end{aligned}$$

y podemos estimar el modelo como sigue:

$$\mathrm{Pr}\,ocs \rightarrow Especify/Estimate \rightarrow d(x1,1)\ d(s1,1)\ d(s2,1)\ ma(1)$$

y obtenemos el modelo siguiente:

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(S1,1)	0.13	0.029	4.55	0
D(S2,1)	0.11	0.04	2.87	0.0043
MA(1)	-0.76	0.04	-18.66	0

Si vemos el correlograma y el histograma de los residuos:

Residuals test → Correlogram Statistics
Residuals test → Histogram

que muestran que no existen problemas con los residuos.

¿Que ocurre con la serie de ventas de colgate? Estimamos un modelo con el cambio de nivel en la serie:

$$\mathsf{Pr}\,ocs \rightarrow Especify/Estimate \rightarrow d(x\mathsf{2},\mathsf{1})\ d(s\mathsf{1},\mathsf{1})\ ma(\mathsf{1})$$

y obtenemos el modelo siguiente:

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(S1,1)	-0.08	0.027	-3.14	0
MA(1)	-0.81	0.03	-22.67	0

Si vemos el correlograma y el histograma de los residuos:

Residuals test $\rightarrow$ Correlogram Statistics
Residuals test $\rightarrow$ Histogram

que muestran que no existen problemas con los residuos. Por lo tanto tenemos un efecto escalon que TSW es incapaz de detectar, pero que es significativo con E-views.