

# Prácticas de la asignatura Series Temporales

## Quinta Entrega

### 1 Identificación y Predicción de modelos ARIMA

En esta entrega veremos como ajustar modelos ARIMA a series de datos reales. Supongamos que tenemos una serie que queremos analizar. El proceso de análisis que se describe a continuación es el propuesto por Box y Jenkins (1976) y que ha sido ampliamente aceptado durante los últimos años:

1. Identificar el modelo ARIMA:
  - (a) Transformar la serie para convertirla en estacionaria.
  - (b) Determinar un modelo ARMA para la serie estacionaria: ordenes  $p$  y  $q$  de la parte regular y ordenes  $P$  y  $Q$  si la serie es estacional.
2. Estimación máximo-verosímil: estimación de los parámetros del modelo seleccionado y de sus errores estándar y los residuos del modelo ajustado.
3. Diagnóstico: se comprueba que los residuos verifican la hipótesis de ruido blanco.

La idea de este proceso es identificar varios modelos posibles en base a la información de la serie (autocorrelaciones, autocorrelaciones parciales,...), estimarlos e seleccionar el mejor. Una vez que hemos estimado dicho modelo, nos aseguraremos que los residuos de dicho modelo no muestran discrepancias con la hipótesis de ruido blanco.

En base al modelo seleccionado y estimado, podremos realizar predicciones para los datos siguientes.

Vemos con series reales como llevar a la práctica este proceso.

### 2 Ejemplo 1: Número de usuarios conectados a Internet en un servidor por minuto

Vemos el ejemplo de los datos del libro de Durbin y Koopman (2001). Es una serie con 100 datos y por lo tanto, creamos un workfile de datos "Undated or irregular" de 110 datos (ya veremos porque tomar un número mayor de datos) y el nombre de la serie, por ejemplo,  $x$ . Una vez importados los datos, vemos el gráfico de la serie mediante View→Linegraph. Lo que podemos ver es una serie que no es estacionaria, que tiene crecimientos y decrecimientos indiscriminados. Vemos los correlogramas para comprobar que efectivamente tenemos no estacionariedad. Tomamos una diferencia por lo que definimos una serie como sigue. En el menú del workfile:

$$Genr \rightarrow dx = d(x, 1)$$

Vemos el gráfico y el correlograma de esta nueva serie, que es aparentemente estacionaria. El gráfico de las autocorrelaciones muestra que la serie tiene memoria bastante larga formando una especie de senoide. Esto descarta que el proceso sea un MA puro. Por otro lado, la función de autocorrelación parcial muestra tres picos significativos. Por lo tanto un posible modelo podría ser un ARIMA(3,1,0), dado que se ajusta a las funciones de autocorrelación y de autocorrelación parcial. De todos modos, como alternativa siempre es conveniente estimar algún otro modelo. En este caso, no es descartable un ARIMA(1,1,1) o algún modelo ARIMA similar. Pasamos a estimarlos.

E-views realiza estimación condicional, es decir, no dispone de estimación máximo verosímil exacta. Para estimar un modelo ARIMA(3,1,0) a la serie x, vamos al menú principal de E-views y marcamos:

*Quick → Estimate Equation*

y escribimos,

*d(x,1) ar(1) ar(2) ar(3)*

Lo que aparece en pantalla es un cuadro donde aparecen las estimaciones y sus errores estándar, así como un estadístico t de dicha estimación para cada uno de los parámetros. En este caso, el modelo estimado resulta ser:

$$(1 - B) x_t = \frac{1.16}{0.09} x_{t-1} - \frac{0.66}{0.13} x_{t-2} + \frac{0.34}{0.09} x_{t-3} + a_t$$

Todos los parámetros son significativos, por lo que no parece haber ningún problema. El valor del AIC=5.14 y del BIC=5.22. Pasamos a la fase de diagnosis. Vamos a:

*Actual, fitted, residual → Actual, fitted, residual graph*

Como se ve, la serie está bien ajustada. Vamos a:

*Residual tests → Correlogram Q – Statistic*

Aquí vemos la función de autocorrelación y de autocorrelación parcial de los residuos. Vemos que no existe ninguno significativo. También aparece el valor del estadístico de Ljung-Box:

$$Q_{LB} = n(n+2) \sum_{j=1}^k \frac{r_j^2}{n-j} \sim \chi_k^2 \text{ (asintóticamente)}$$

Si los p-valores son mayores que 0.05 para todos los retardos, no podremos rechazar la hipótesis de incorrelación para los residuos. Por otro lado:

*Residual tests → Histogram (Normality test)*

Podemos ver el histograma de los residuos. El estadístico de Jarque-Bera no rechaza la hipótesis de normalidad. La desviación típica de dichos residuos es  $\hat{\sigma}_a = 3.07$ .

Si queremos estimar un ARIMA(1,1,1) a los datos, hacemos como sigue:

*Procs → Specify/Estimate*

y marcamos D(x,1) AR(1) MA(1). El modelo estimado es:

$$\left(1 - \frac{0.65B}{0.08}\right) (1 - B) x_t = \left(1 + \frac{0.53B}{0.09}\right) a_t$$

Los parámetros son significativos. El valor del AIC=5.15 y del BIC=5.20, por lo que el AIC del modelo anterior es un poco menor, mientras que el del BIC es algo inferior para este modelo. Pasamos a la fase de diagnosis. Vamos a:

$$Actual, fitted, residual \rightarrow Actual, fitted, residual graph$$

Como se ve, la serie se ajusta de manera muy parecida al caso anterior. Vamos a:

$$Residual tests \rightarrow Correlogram Q - Statistic$$

Aquí vemos la función de autocorrelación y de autocorrelación parcial de los residuos. Vemos que no existe ninguno significativo, y que los valores de los estadísticos de Ljung-Box son mayores que 0.05 para todos los retardos y no podemos rechazar la hipótesis de incorrelación para los residuos. Por otro lado:

$$Residual tests \rightarrow Histogram (Normality test)$$

y comprobamos que el estadístico de Jarque-Bera no rechaza la hipótesis de normalidad y que la desviación típica de dichos residuos es  $\hat{\sigma}_a = 3.12$ , superior a la del modelo anterior.

En base a lo anterior es complicado seleccionar un modelo para estos datos, puesto que parece que ambos modelos ajustan los datos de manera muy similar. Vamos a seleccionar el modelo con menor varianza residual y que tiene p-valores de los estadísticos de Ljung-Box mayores. Este resulta ser el modelo ARIMA(3,1,0). Así que hacemos como sigue:

$$Procs \rightarrow Especify/Estimate$$

y marcamos D(x,1) AR(1) AR(2) AR(3). Seleccionamos en el menú forecast:

$$Forecast \rightarrow \left\{ \begin{array}{ll} \text{Forecast of} & x \\ \text{Forecast name} & xf \\ \text{S.E.} & \text{errores} \\ \text{Sample range to forecast} & 101\ 110 \end{array} \right.$$

Debemos ver un gráfico con los valores predichos para los próximos 10 datos. En el workfile aparecerán dos nuevas series, xf y errores. La serie xf corresponde a la serie x para los datos del 1 al 100 y la predicción para los datos 101 a 110. Podemos ver el gráfico para comprobar que la predicción sigue el recorrido decreciente de los últimos datos. Vemos en la serie errores que son cada vez más grandes.

### 3 Ejemplo 2: Serie de líneas aéreas

Vemos el ejemplo de los datos del libro de Box y Jenkins. Los datos son de 1949:1 a 1960:12. Creamos un workfile de datos mensuales de 1949:1 a 1962:12, para predecir los dos últimos años. Una vez importados los datos, comprobamos como los datos no tienen varianza constante, si no creciente. Tomamos el logaritmo de dichos datos,

$$Genr \rightarrow lx = \log(x)$$

y comprobamos que ahora los datos no aparentan un cambio continuo de varianza. Vemos el correlograma y comprobamos que la serie es no estacionaria, por lo que tomamos una diferencia. Definimos la serie:

$$dlx = d(lx, 1)$$

y vemos como el gráfico muestra muy claramente la estacionalidad. Volvemos a comprobar el correlograma y vemos como claramente necesita una diferencia estacional:

$$ddlx = d(lx, 1, 12)$$

Si vemos ahora el gráfico de la serie, vemos como ésta es aparentemente estacionaria. Del correlograma podemos sacar algunas conclusiones. En la parte regular podemos comprobar que los gráficos de la correlación y de la correlación parcial son muy parecidos, siendo el primer retardo significativo para ambos casos. Esto nos dice que quizás estemos ante un AR(1), un MA(1) o un ARMA(1,1). En la parte estacional vemos que el retardo 12 es significativo y que en sus cercanías hay muy poca correlación. En principio esto da por resultado un proceso con muy poca memoria sugiriendo un proceso MA(1)<sub>12</sub> en la parte estacional. Además, como no existe correlación significativa alrededor del retardo 12, esto nos indica que quizás la parte regular también sea un MA(1). Vayamos por partes. Estimamos un ARIMA(1,1,0)×(0,1,1)<sub>12</sub>, un ARIMA(0,1,1)×(0,1,1)<sub>12</sub> y un ARIMA(1,1,1)×(0,1,1)<sub>12</sub>:

*Quick → Estimate Equation* : d(lx,1,12) ar(1) sma(12)

El primer modelo tiene los dos parámetros significativos, un AIC=-3.75 y un BIC=-3.71, mientras que el correlograma muestra algún estadístico Q significativo (<0.05). La desviación típica de los residuos es 0.0365. Para el segundo modelo,

*Procs → Specify/Estimate* : d(lx,1,12) ma(1) sma(12)

los parámetros son significativos, tiene AIC=-3.75 y un BIC=-3.71, mientras que el correlograma no muestra algún estadístico Q significativo (<0.05). La desviación típica de los residuos es 0.0365. El último modelo:

*Procs → Specify/Estimate* : d(lx,1,12) ar(1) ma(1) sma(12)

tiene parámetros no significativos por lo que lo descartamos. En definitiva los dos primeros modelos dan un ajuste muy parecido de los datos. La única diferencia es que el estadístico de Ljung-Box tiene mayores p-valores para el modelo ARIMA(0,1,1)×(0,1,1)<sub>12</sub>, que es finalmente seleccionado. Obtenemos predicciones para los dos últimos años como vimos:

<i>Forecast</i> →	{	Forecast of	lx
		Forecast name	lxf
		S.E.	errores
		Sample range to forecast	1961:01 1962:12

Debemos ver un gráfico con los valores predichos para los próximos 2 años. En el workfile aparecerán dos nuevas series, lxf y errores. La serie lxf corresponde a la serie lx para los datos del 1941:1 al 1960:12 y la predicción para los datos 1961:01 1962:12. Podemos ver el gráfico para comprobar que la predicción sigue el recorrido anterior de los datos. Vemos en la serie errores que son cada vez más grandes.

## 4 Apéndice: Estimación con el programa TRAMO

Veamos como estimar una serie con el programa TSW (Tramo-Seats de windows). Aparece una ventana con el programa, con un menu de botones y con una serie de ranuras. Para analizar una serie del mismo modo que antes, procedemos como sigue. Lo primero es importar los datos que están en la página web:

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

con el nombre de airpastra.txt. La serie empieza en 1949:1 y acaba en 1960:12 por lo que son 144 datos. 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 lter si lo cambiamos a lter=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)×(0,1,1)<sub>12</sub>, 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.  $BQ=1$ .
6.  $BD=1$ .
7.  $INIT=0$ , este parámetro sirve para inicializar la estimación de los parámetros.
8.  $IMEAN=1$ , este parámetro sirve para corregir la media o no. Esto significa que no corregimos la media.
9.  $LAM=-1$ , este parámetro dice si transformar logarítmicamente 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 así, es para cambiar la escala si los datos son muy pequeños o grandes.

Marcamos Others para nuevos parámetros. De todos estos valores, sólo nos interesará en un futuro, los correspondientes a Outliers. Por ahora marcamos  $IATIP=0$ . Marcamos OK. Marcamos el modelo en la tabla y marcamos Run. En output, vemos la salida.