

## Capítulo 6

# Propuesta metodológica MIEC para la identificación y estimación del modelo SETAR.

La construcción de un modelo es un proceso estadístico complejo que podemos sintetizar en los siguientes pasos (Tong y Lim, 1980):

1. Análisis gráfico de los datos
2. Contrastación de la linealidad
3. Selección del modelo (Identificación y Estimación)
4. Diagnóstico

Dentro de este marco general, la novedad de nuestra propuesta radica en incorporar en el paso 2 los resultados obtenidos por Tsay y proponer, en el paso 3, un nuevo algoritmo que permita la estimación automática de los parámetros estructurales y los coeficientes. Más concretamente, nuestra propuesta consigue la automatización de los siguientes subprocesos:

- A.1. Estimación del orden de cada uno de los procesos autoregresivos que conforman el modelo SETAR  $(l; k_1, \dots, k_l)$ , asumiendo que  $l \geq 2$ .
- A.2. Selección, en cada uno de los procesos autoregresivos, del conjunto de regresores y estimación de los coeficientes correspondientes a dichos regresores.
- A.3. Estimación automática del valor umbral (si consideramos que el número de regímenes es igual a 2).

Para conseguir estas mejoras nuestra propuesta introduce un nuevo proceso algorítmico, a la vez que recoge algunos de los procedimientos de Tong y Lim, Tsay y Thanoon. Antes de pasar a comentar los principios teóricos en los que se basa nuestro algoritmo vamos a presentar la estructura del proceso de estimación e identificación propuesto para un modelo SETAR.

## 6.1 Estructura del proceso de identificación y estimación.

La metodología que presentamos y que vamos a denominar MIEC (Metodología de Identificación y Estimación de Coeficientes), permite la identificación automática de los órdenes, la selección de los regresores y la estimación de los coeficientes de un modelo SETAR. Cuando consideramos modelos SETAR con sólo dos regímenes, completamos la anterior metodología con un conjunto de procesos que permiten la estimación automática del umbral.

Vamos a presentar en primer lugar las etapas en que se estructura MIEC para un modelo SETAR de más de dos regímenes (asumimos que los umbrales ya han sido estimados).

Dada una serie temporal  $\{X_t\}$  con un comportamiento no lineal de tipo SETAR con  $l > 2$  regímenes, y suponemos que conocemos los  $l - 1$  umbrales (pueden haber sido estimados por métodos no automáticos como los propuestos por Tong y Tsay), entonces la identificación y estimación de dicho modelo supone la realización de las siguientes etapas (Figura 6.1):

*Etapas A:* Estimamos el parámetro de retardo  $d$  y el orden máximo  $p$  a partir de la metodología de Tsay.

*Etapas B:* Decidimos el criterio de especificación automática que queremos utilizar.

*Etapa C:* Estimamos los coeficientes  $a_i^{(j)}$  por el método de estimación de mínimos cuadrados condicionados. El proceso evalúa el  $AIC$  para cada posible conjunto de regresores y selecciona aquel para el cual este criterio es mínimo. Se estiman los coeficientes  $\hat{a}_i^{(j)}$ ,  $j = 1, \dots, l$ ;  $i = 0, \dots, \hat{k}_i$  así como los órdenes  $\hat{k}_1, \hat{k}_2, \dots, \hat{k}_l$  del modelo  $SETAR(l; \hat{k}_1, \hat{k}_2, \dots, \hat{k}_l)$ .

La etapa C soporta, tanto a nivel teórico como computacional, el peso de la propuesta metodológica pues en ella se seleccionan las variables autoregresoras, se identifican los órdenes de los procesos autoregresivos y se realiza la estimación de los coeficientes. Para realizar estos procesos hemos implementado un algoritmo de denominamos AIEC (Algoritmo de Identificación y Estimación de Coeficientes) que permite determinar y estimar todas las posibles autoregresiones.

Cuando conocemos o suponemos que el modelo SETAR tiene sólo dos regímenes podemos completar la estructura de MIEC a fin de conseguir incorporar la estimación del umbral de forma automática. Para ello debemos incorporar nuevos procesos, tal y como se refleja en el siguiente esquema:

Dada una serie temporal  $\{X_t\}$  con un comportamiento no lineal de tipo SETAR con  $l = 2$  regímenes, la identificación y estimación <sup>1</sup>de dicho modelo supone la realización de las siguientes etapas (Figura 6.2):

***Etapa 1:* Estimamos el parámetro de retardo  $d$  y el orden máximo  $p$  a partir de la metodología de Tsay.**

***Etapa 2:* Determinamos el conjunto de posibles valores  $(\tau_1, \dots, \tau_j)$  entre los que puede variar el valor umbral  $r$  (el proceso algorítmico que permite construir el conjunto de variación del umbral se detalla en el siguiente capítulo). En esta etapa **se decide también cuál es el criterio de especificación automática que queremos utilizar.****

***Etapa 3:* Consideramos un valor  $\tau_i$  del conjunto  $(\tau_1, \dots, \tau_k)$  y que el criterio de especificación automática elegido es el  $AIC$ <sup>2</sup>. Para cada uno de los regímenes de-**

---

<sup>1</sup>Queremos destacar que la metodología utilizada es la misma que para  $l > 2$ . Para corroborarlo hemos destacado en negrita los procesos comunes.

<sup>2</sup>El razonamiento será el mismo si elegimos cualquier otro criterio.

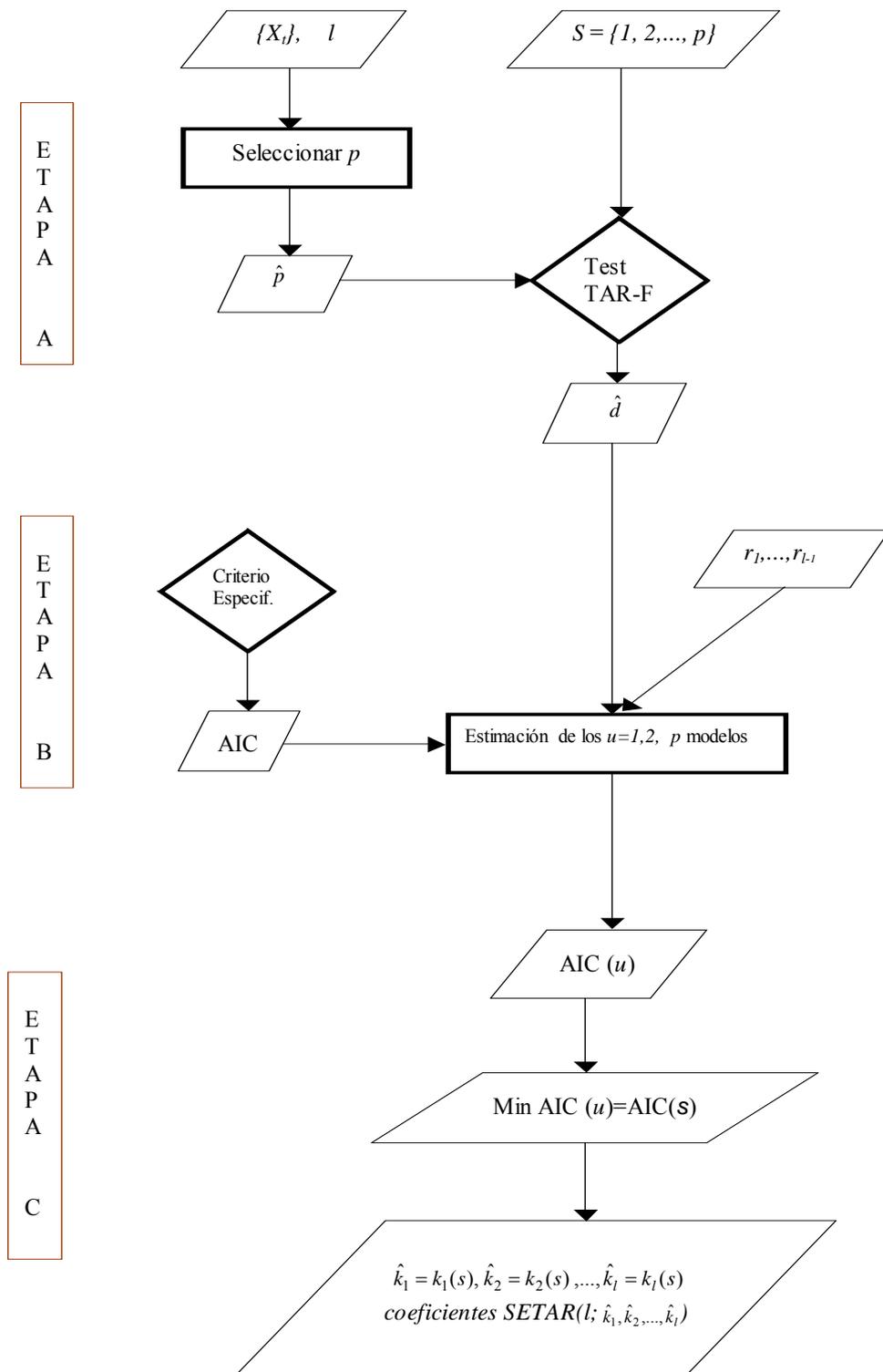


Figura 6-1: Estructura de la metodología MIEC para un  $SETAR(l; k_1, k_2, \dots, k_l)$ .

terminados a partir del valor  $\tau_i$  y de la variable umbral  $X_{t-\hat{d}}$  estimamos, utilizando mínimos cuadrados condicionados, todos los posibles modelos autoregresivos cuyo orden sea menor o igual a  $\hat{p}$ . Calculamos el  $AIC$  de cada uno de los modelos  $SETAR$  estimados y elegimos aquel que tiene  $AIC$  mínimo. El resultado es una  $SETAR(l; k_1(\tau_i), k_2(\tau_i))$  caracterizado por un valor del  $AIC$  que designamos por  $AIC(\tau_i)$ .

*Etapa 4:* Repetimos la etapa 3 para cada valor del conjunto  $(\tau_1, \dots, \tau_k)$  y minimizamos el valor del  $AIC$  de manera que la estimación de  $r$ ,  $\hat{r}$ , es el valor que hace mínimo el  $AIC(\tau_i)$ :

$$\hat{r} = \tau_s \quad \text{tal que} \quad AIC(\tau_s) = \min(AIC(\tau_i))$$

*Etapa 5:* Calculado  $\hat{r}$ , estimamos los coeficientes  $a_i^{(j)}$  por el método de estimación de mínimos cuadrados condicionados. El proceso evalúa el  $AIC$  para cada posible conjunto de regresores y selecciona aquel para el que este criterio es mínimo. Se estiman los coeficientes  $\hat{a}_i^{(j)}$ ,  $j = 1, \dots, l$ ;  $i = 0, \dots, \hat{k}_i$  así como los órdenes  $\hat{k}_1$  y  $\hat{k}_2$  del modelo  $SETAR(l; \hat{k}_1, \hat{k}_2)$ .

Hemos destacado la parte común de ambas estructura para resaltar que la estructura subyacente cuando modelizamos un  $SETAR(2; \hat{k}_1, \hat{k}_2)$  es la misma que la seguida para modelizar un  $SETAR(l; \hat{k}_1, \hat{k}_2, \dots, \hat{k}_l)$ .

## 6.2 Características del proceso algorítmico AIEC.

Tal y como comentábamos al principio de este capítulo, la principal novedad de nuestra propuesta consiste en la elaboración de un nuevo algoritmo que permite implementar de forma automática la selección de los regresores y la estimación de parámetros estructurales y de los coeficientes. Este algoritmo, que hemos denominado AIEC, es el responsable de realizar los procesos que caracterizan la Etapa C (o la Etapa 5, cuando consideramos modelos con dos regímenes y estimamos de forma automática el umbral). Debido a la importancia de estos procesos en el conjunto de nuestra propuesta metodológica, vamos a realizar su justificación

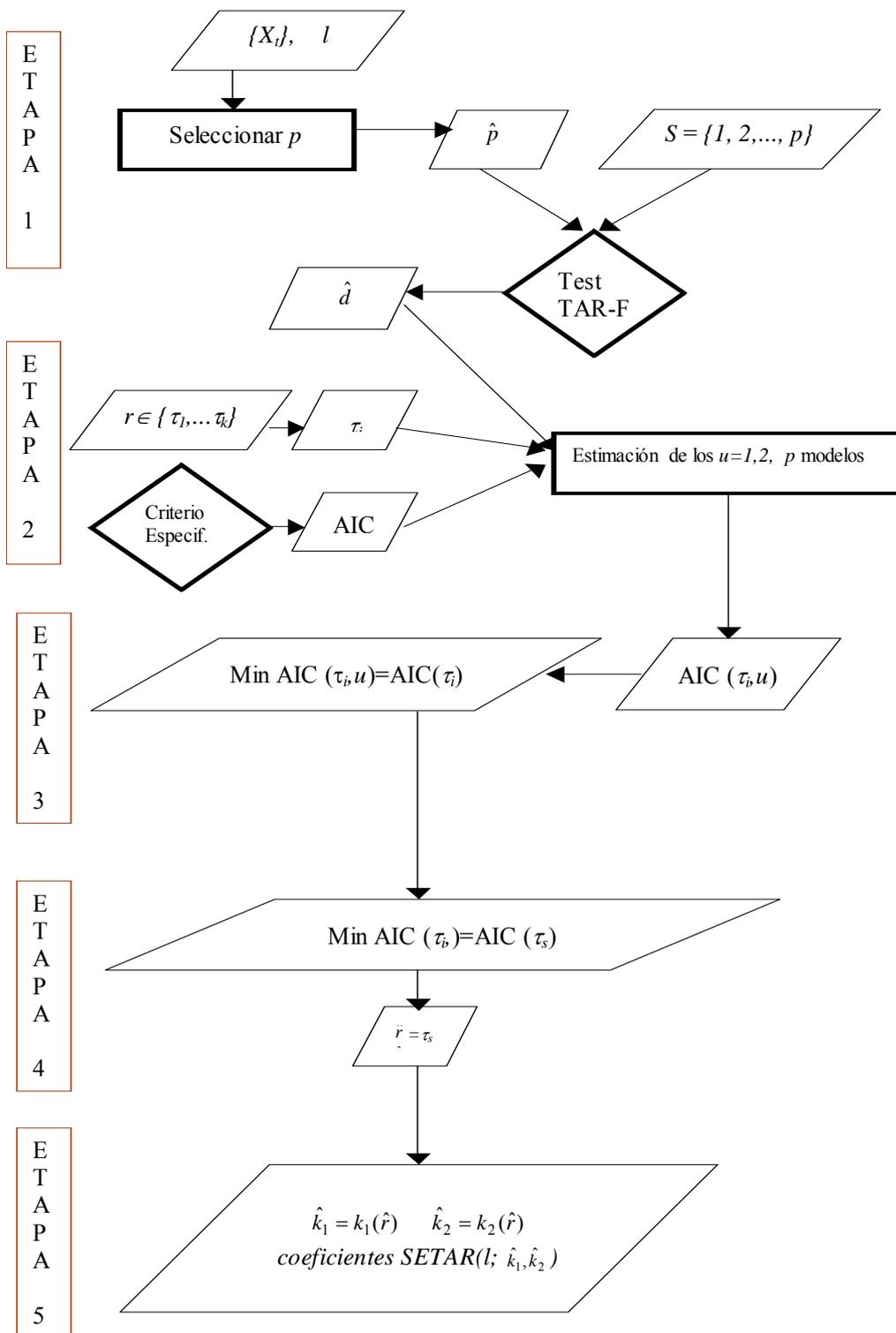


Figura 6-2: Estructura de la metodología MIEC para un  $SETAR(2; k_1, k_2)$ .

a nivel teórico centrándonos en la estimación de los coeficientes y la selección de las variables regresoras<sup>3</sup>.

### 6.2.1 Estimación de los coeficientes.

El modelo SETAR es como ya hemos comentado un modelo localmente lineal, el comportamiento no lineal es introducido por la variable retardo al imponer el cambio de un régimen a otro, pero si nos centramos en la estimación de los coeficientes autoregresivos de un régimen, este proceso es equivalente al de estimación de los coeficientes de un modelo AR lineal.

Los diferentes métodos de estimación para los coeficientes de un modelo AR se diferencian esencialmente en el grado de precisión; basándose en este criterio Priestley (1981) los clasifica como sigue:

1. Función de verosimilitud exacta.
2. Estimaciones por mínimos cuadrados.
3. Estimaciones por mínimos cuadrados aproximados (usando  $\bar{x} = \mu$ ).
4. Estimaciones a partir de las ecuaciones de Yule-Walker.

En esta clasificación el mayor grado de precisión corresponde al primer método y, la menor precisión a las estimaciones de Yule-Walker. Las estimaciones mínimo cuadráticas proporcionan estimadores idénticos a los que obtendríamos al maximizar la función de verosimilitud condicionada y, cuando el número de observaciones es moderadamente grande, se observa que las diferencias entre las estimaciones obtenidas por los diferentes métodos son muy pequeñas (Box y Jenkins, 1970). Brockwell y Davis, (1997) demuestran que para un AR(p) los estimadores obtenidos por el método de Yule-Walker tienen las mismas propiedades asintóticas que los estimadores máximo verosímiles. Otro resultado interesante es el obtenido por Priestley (1981), este autor prueba que, asumiendo la normalidad de los errores, el criterio de los mínimos cuadrados puede ser considerado como una aproximación de la estimación por máxima verosimilitud.

---

<sup>3</sup>En el siguiente capítulo presentaremos la estructura del algoritmo y comentaremos como se ha implementado en un código de programación.

Teniendo en cuenta los resultados anteriores y considerando la sencillez de automatización del proceso de estimación por mínimos cuadrados condicionados, elegimos este método para obtener la estimación de los coeficientes en nuestra propuesta ( también es utilizado por Tong y Tsay en sus respectivas metodologías).

Thanoon en su trabajo no concreta como se realiza la estimación de los coeficientes pero, si analizamos el algoritmo de Haggan-Oyetunji (utilizado en la etapa 2 del algoritmo de Thanoon), deducimos que la estimación de la matriz de covarianzas se realiza a partir de las ecuaciones de Yule-Walker.

### 6.2.2 Análisis de todas las posibles autoregresiones.

Si consideramos un proceso autoregresivo de orden máximo  $p$ , entendemos por todas las posibles autoregresiones las obtenidas al considerar cada uno de los subconjuntos formados a partir de las variables autoregresoras  $X_{t-1}, X_{t-2}, \dots, X_{t-\hat{p}}$  (Thanoon, 1990). De entre éstas se va a elegir (mediante un criterio de especificación automática) aquella que mejor se ajusta a los datos<sup>4</sup>.

Más concretamente, supongamos que queremos ajustar un modelo SETAR a un conjunto de datos  $\{X_t\}$  :

$$\mathbf{X}_t = a_0^{(j)} + \sum_{i=1}^{k_i} a_i^{(j)} X_{t-i} + \varepsilon_t^{(j)} \quad \text{si} \quad X_{t-d} \in R_j, \quad j = 1, \dots, l$$

donde  $R_j = (r_{j-1}, r_j)$  y consideramos  $r_0 = -\infty$  y  $r_l = +\infty$

Hemos estimado previamente el número de regímenes  $\hat{l}$ , los  $\hat{l} - 1$  umbrales  $\hat{r}_1, \dots, \hat{r}_{l-1}$ , el valor del retardo  $\hat{d}$  y el orden máximo de los procesos autoregresivos  $\hat{p}$ . Debemos ahora estimar los coeficientes  $\hat{a}_i^{(j)}$   $j = 1, \dots, \hat{l}; \quad i = 0, \dots, \hat{p}$ .

Si tomamos como punto de partida el proceso autoregresivo definido para un régimen

---

<sup>4</sup>El proceso algorítmico que permite considerar todas las posibles regresiones y elegir el modelo más adecuado se detalla en el siguiente capítulo.

cualquiera  $j$ :

$$\mathbf{X}_t = a_0^{(j)} + a_1^{(j)}X_{t-1} + a_2^{(j)}X_{t-2} + \cdots + a_{\hat{p}}^{(j)}X_{t-\hat{p}} + \varepsilon_t^{(j)} \quad \text{si} \quad X_{t-d} \in R_j \quad (6.1)$$

la estimación de los coeficientes supone la resolución del siguiente sistema de ecuaciones:

$$Y = X\beta + \varepsilon \quad (6.2)$$

donde

$$\beta = ( a_0^{(j)}, a_1^{(j)}, a_2^{(j)}, \dots, a_{\hat{p}}^{(j)} )'$$

$$X = \begin{pmatrix} 1 & x_{t-1} & x_{t-2} & \cdots & x_{t-\hat{p}} \\ 1 & x_{t+1-1} & x_{t+1-2} & \cdots & x_{t+1-\hat{p}} \\ 1 & x_{t+2-1} & x_{t+2-2} & \cdots & x_{t+2-\hat{p}} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n-1} & x_{n-2} & \cdots & x_{n-\hat{p}} \end{pmatrix}$$

$$Y = ( x_t, x_{t+1}, x_{t+2}, \dots, x_n )'$$

$$\varepsilon = ( \varepsilon_t^{(j)}, \varepsilon_{t+1}^{(j)}, \varepsilon_{t+2}^{(j)}, \dots, \varepsilon_n^{(j)} )'$$

$x_t$  son las observaciones del proceso  $\{X_t\}$  tal que  $x_{t-d} \in R_j$ ,  $t = \max(\hat{p}, \hat{d})$  y  $n$  el número de observaciones en el régimen  $j$ .

Así los valores estimados para  $\beta$ , por el método de mínimos cuadrados condicionados, se

obtienen como sigue

$$\widehat{\beta} = (X'X)^{-1}(X'Y) \quad (6.3)$$

donde

$$\widehat{\beta} = ( \widehat{a}_0^{(j)}, \widehat{a}_1^{(j)}, \widehat{a}_2^{(j)}, \dots, \widehat{a}_{\widehat{p}}^{(j)} )'$$

A partir de la ecuación 6.2 podemos obtener todas las posibles autoregresiones construidas a partir de los subconjuntos de  $\{X_{t-1}, X_{t-2}, \dots, X_{t-\widehat{p}}\}$ . El número de posibles subconjuntos es  $2^p - 1$  y por tanto podemos plantearnos  $2^p - 1$  autoregresiones.

De esta manera para cada régimen debemos resolver  $2^p - 1$  sistemas de ecuaciones de la forma

$$Y_u = X_u \beta_u + \varepsilon \quad (6.4)$$

donde el subíndice  $u = 1, \dots, 2^p - 1$  nos informa del sistema de ecuaciones en que nos encontramos, es decir, nos permite identificar el subconjunto de variables regresoras que estamos considerando. Para un valor concreto  $u$  las variables regresoras son el subconjunto  $X_{t-p_1}, X_{t-p_2}, \dots, X_{t-p_k}$ , donde  $p_1, p_2, \dots, p_k$  están incluidos en el conjunto de los enteros  $\{1, 2, \dots, \widehat{p}\}$  con  $1 \leq p_1 < p_2 < \dots < p_k \leq \widehat{p}$ . La matriz  $X_u$ , que es una submatriz obtenida a partir de la matriz  $X$ , se construye como sigue:

$$X_u = \begin{pmatrix} 1 & x_{t-p_1} & x_{t-p_2} & \cdots & x_{t-p_k} \\ 1 & x_{t+1-p_1} & x_{t+1-p_2} & \cdots & x_{t+1-p_k} \\ 1 & x_{t+2-p_1} & x_{t+2-p_2} & \cdots & x_{t+2-p_k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n-p_1} & x_{n-p_2} & \cdots & x_{n-p_k} \end{pmatrix}$$

$Y_u$  es un subvector del vector  $Y$

$$Y_u = ( x_{t-p_1+1}, \quad x_{t+1-p_1+1}, \quad x_{t+2-p_1+1}, \quad \dots, \quad x_{n-p_1+1} )'$$

y  $\beta_u$  el vector de coeficientes se puede expresar como

$$\beta_u = ( a_{p_0}^{(j)}, \quad a_{p_1}^{(j)}, \quad a_{p_2}^{(j)}, \quad \dots, \quad a_{p_k}^{(j)} )'$$

La solución del sistema de ecuaciones tiene ahora la forma:

$$\beta_u = (X_u' X_u)^{-1} (X_u' Y_u)$$

La resolución de cada uno de estos sistemas además de proporcionarnos los coeficientes, nos proporcionará también el AIC, AICc, BIC, los residuos estimados y la varianza residual que denotaremos respectivamente como  $AIC(u, j)$ ,  $AICc(u, j)$ ,  $BIC(u, j)$ ,  $\hat{r}_{u,j}$  y  $\hat{\sigma}^2(u, j)$ . También nos va a permitir determinar el orden máximo de cada sistema  $h$ .

Con anterioridad (etapa 2) hemos elegido un criterio de especificación automática, supongamos que el criterio seleccionado es el AIC, entonces para el régimen  $j$  se calcula el AIC de los modelos de este régimen y se selecciona aquel que tiene un AIC mínimo.

Más explícitamente el modelo estimado para el régimen  $j$  es aquel modelo  $s$  tal que

$$AIC(s, j) = \min \{ AIC(u, j), \quad u = 1, \dots, 2^p - 1 \}$$

así, una vez seleccionado el modelo  $s$ , obtenemos los coeficientes autoregresivos estimados

$$\hat{\beta}_{s_j} = ( \hat{a}_{s_0}^{(j)}, \quad \hat{a}_{s_1}^{(j)}, \quad \hat{a}_{s_2}^{(j)}, \quad \dots, \quad \hat{a}_{s_k}^{(j)} )'$$

donde  $s_1, s_2, \dots, s_k$  es un subconjunto del conjunto de los enteros  $\{1, 2, \dots, \hat{p}\}$  con  $1 \leq s_1 < s_2 < \dots < s_k \leq \hat{p}$  y  $\hat{a}_{s_0}^{(j)}$  es el coeficiente del término independiente.

También el proceso nos proporciona el orden del proceso autoregresivo  $h_{s_j}$  y el subconjunto de las variables regresoras  $\{X_{t-s_1}, X_{t-s_2}, \dots, X_{t-s_k}\}$  que intervienen en el proceso. Finalmente para el modelo estimado se hallan los residuos estimados  $\hat{r}_{s_j,j}$  y la varianza residual  $\hat{\sigma}^2(s_j, j)$ .

El proceso descrito se repite para cada régimen, obteniéndose la estimación de un modelo  $SETAR(l; h_{s1}, \dots, h_{sl})$  en el que de forma automática se ha conseguido determinar los órdenes de cada uno de los procesos autoregresivos  $h_{sj}$ , los coeficientes autoregresivos de los procesos, la varianza residual global  $\hat{\sigma}^2$  y también el  $AIC$  global,  $AICc$  global y  $BIC$  global.

Al considerar todos los esquemas autoregresivos posibles, el anterior proceso supone un gran número de cálculos, pensemos que si el orden máximo  $p$  es igual a 10 deberemos resolver para cada régimen  $2^{10} - 1 = 1023$  sistemas de ecuaciones; como el número de regímenes es mayor o igual que 2, como mínimo se van a resolver 2046 sistemas de ecuaciones si consideramos un único umbral.

No olvidemos, sin embargo, que la resolución de cada sistema tiene como objetivo determinar el valor del criterio de información para proceder después a su minimización, este proceso es según Sin y White (1996) consistente, pero tiene como desventaja el elevado número de cálculos.

Un estudio detallado de la expresión del AIC nos va a permitir la simplificación del proceso acotando el número de sistemas de ecuaciones a resolver. Los resultados de dicho estudio se sintetizan en los teoremas 24 y 25 que a continuación enunciamos y demostramos.

El teorema 24 asegura que el valor mínimo del AIC en un modelo SETAR se obtiene cuando los regresores tienen retardos consecutivos.

**Teorema 24:** *Sea  $\{X_t\}$   $t = 1, \dots, N$  una serie temporal que sigue un modelo autoregresivo SETAR con  $l$  regímenes cada uno de ellos de orden máximo  $p$ . El MAICE (mínimo valor del AIC estimado) se obtiene para un modelo  $SETAR(l; k_1, k_2, \dots, k_l)$  donde los órdenes  $k_i$  ( $k_i \leq p$ ,  $i = 1, 2, \dots, l$ ) de cada uno de los procesos autoregresivos coinciden con el número de variables autoregresoras de cada proceso.*

Este resultado supone que si consideramos cualquier submodelo SSETAR con un número de variables autoregresoras inferior al orden del proceso, entonces el AIC de este modelo no será el mínimo.

Para demostrar el teorema anterior vamos a considerar, en primer lugar, que el criterio de especificación automática elegido es el AIC y que se trata de un modelo autoregresivo AR. Bajo estas hipótesis enunciamos el teorema 25 a continuación:

**Teorema 25:** Sea  $\{X_t\} t = 1, \dots, N$  una serie temporal que sigue un modelo autoregresivo AR de orden máximo  $p$ . De entre todos los subconjuntos de  $k$  variables regresoras posibles ( $k \leq p$ ), el subconjunto  $X_{t-p_1}, X_{t-p_2}, \dots, X_{t-p_k}$ , que hace mínimo el valor del AIC es aquel cuyos retardos cumplen la siguiente condición  $1 = p_1 < p_2 < \dots < p_k = k \leq p$ , es decir, son los primeros  $k$  retardos consecutivos.

Antes de pasar a la demostración de este teorema vamos a enmarcar el problema:

Si  $\{X_t\}$  es una serie temporal que sigue un modelo autoregresivo AR de orden máximo  $p$ , podemos considerar como punto de partida el que vamos a denominar "modelo completo" o modelo con  $p$  regresores:

$$X_t = a_0^{(j)} + a_1^{(j)} X_{t-1} + a_2^{(j)} X_{t-2} + \dots + a_p^{(j)} X_{t-p} + \varepsilon_t^{(j)}$$

a partir de este modelo podemos considerar  $(2^p - 1)$  modelos posibles, tantos como conjuntos de posibles regresores.

Supongamos fijado el número de regresores  $k$  ( $k \leq p$ ), sin tener en cuenta su retardo (por ejemplo, si  $p = 7$  y  $k = 3$ , podríamos considerar los conjuntos  $\{X_{t-1}, X_{t-2}, X_{t-3}\}$  o  $\{X_{t-4}, X_{t-5}, X_{t-6}\}$  o  $\{X_{t-1}, X_{t-4}, X_{t-5}\}$ , entre otros), el subconjunto de regresores para el que el AIC del modelo alcanza su valor mínimo debemos demostrar que es  $X_{t-1}, X_{t-2}, \dots, X_{t-k}$ .

Recordemos la expresión del AIC de un proceso autoregresivo  $AR(u)$ :

$$AIC(u) = n \ln(\sigma^2) + 2(k + 1) \tag{6.5}$$

donde  $n$  es el número de efectivos y  $k + 1$  el número de parámetros estimados.

Notemos que el orden del proceso  $u$  es independiente del número de regresores  $k$ , por ejemplo si los regresores son  $\{X_{t-1}, X_{t-2}, X_{t-3}\}$  el orden  $u$  es igual a 3, si elegimos el conjunto  $\{X_{t-4}, X_{t-5}, X_{t-6}\}$  el orden es 6, y si el subconjunto es el  $\{X_{t-1}, X_{t-4}, X_{t-5}\}$  el orden  $u$  es igual a 5, pero en cambio en todos ellos el número de parámetros estimados sería  $k + 1$ , los  $k$  coeficientes que acompañan a los regresores y el término constante  $a_0$ . Por tanto para cada uno de los modelos se mantendrá fijo el valor de  $k$  en la expresión (6.5).

Podríamos pensar que al variar el orden  $u$  para cada conjunto de  $k$  regresores el número de efectivos  $n = N - u$ , va variando, pero dado que a priori conocemos el orden máximo  $p$  vamos a considerar como efectivo  $n = N - p$ , de esta manera todos los posibles modelos autoregresivos se calculan a partir del mismo conjunto de datos<sup>5</sup>. Por tanto en la expresión (6.5) el único valor que puede cambiar al considerar diferentes subconjuntos de  $k$  regresores es el valor de  $\sigma$ , es decir fijado  $k$  y  $p$ , el valor del criterio de información de Akaike depende de la varianza residual  $\sigma^2$ , que varía entre 0 y 1, en esta situación

$$-\infty < \ln(\sigma^2) < 0$$

Por tanto, entre todos los modelos con  $k$  regresores tendrá un *AIC* mínimo aquel para el que la varianza residual sea menor. Más concretamente queremos demostrar que entre todos los modelos con  $k$  regresores, el modelo

$$\mathbf{X}_t = a_0^{(j)} + a_1^{(j)} X_{t-1} + a_2^{(j)} X_{t-2} + \dots + a_k^{(j)} X_{t-k} + \varepsilon_t^{(j)} \quad (6.6)$$

es el que tiene varianza residual mínima o lo que es lo mismo, la varianza explicada por el modelo es máxima.

Sin ninguna duda la varianza del modelo se descompone en

$$\text{Varianza total} = \text{Varianza explicada} + \text{Varianza residual}$$

si el número de regresores es mayor, aumenta la varianza explicada por el modelo y en consecuencia la varianza residual disminuye. En nuestro caso el número de regresores es siempre el mismo  $k$ , pero no así el orden de cada proceso, por lo que los retardos de los regresores implicados pueden ser diferentes.

Supongamos un conjunto cualquiera de  $k$  regresores  $X_{t-p_1}, X_{t-p_2}, \dots, X_{t-p_k}$ , diferentes

---

<sup>5</sup>El considerar  $n = N - p$  no penaliza el proceso de minimización. Si para un cierto  $u$  el valor del  $AIC(u) = (N - p) \ln(\sigma^2) + 2(k + 1)$  es mínimo aún será más pequeño si sustituimos  $(N - p)$  por  $(N - u)$  pues  $\ln(\sigma^2) < 0$ .

a  $X_{t-1}, X_{t-2}, \dots, X_{t-k}$ , siempre se cumplirá que el orden del proceso correspondiente a este conjunto será mayor que  $k$ . Dado el carácter autoregresivo del modelo propuesto cada una de las variables autoregresoras puede ser explicada por su pasado, es decir,  $X_{t-p_1}$  puede obtenerse a partir de  $X_{(t-p_1)-p_1}, X_{(t-p_1)-p_2}, \dots, X_{(t-p_1)-p_k}$ , y así sucesivamente para cada una de ellas hasta llegar a un momento en que la variable ya no puede ser más explicada por que ya no tiene pasado. Vamos a ilustrar esta situación:

Sean  $x_0, x_1, \dots, x_n$  observaciones  $X_t$ , si esta variable se explica por  $X_{t-1}, X_{t-2}, \dots, X_{t-k}$ , podemos expresar cada observación en función de las anteriores de manera que la observación del período  $k$ ,  $x_k$ , es la primera observación que podemos explicar a partir del pasado ( $k-1, k-2, \dots, k-k=0$ ), lo que supone una pérdida de  $k$  observaciones.

Si la variable  $X_t$ , se explica por un conjunto cualquiera de  $k$  regresores  $X_{t-p_1}, \dots, X_{t-p_k}$ , diferentes a los anteriormente considerados, la primera observación que podemos explicar a partir de su pasado es  $x_{p_k}$ , lo que supone una pérdida de  $p_k$  observaciones. Como  $p_k > k$ , el número de datos disponibles es menor, y por tanto la varianza explicada por este modelo será siempre menor que la varianza explicada cuando se consideran como regresores  $X_{t-1}, X_{t-2}, \dots, X_{t-k}$ .

Vamos a proponer un ejemplo para explicitar la relación que se establece entre las varianzas explicadas: Suponemos  $k = 3$  regresores y concretamente se toma como conjuntos de regresores  $\{X_{t-1}, X_{t-2}, X_{t-3}\}$  y  $\{X_{t-1}, X_{t-3}, X_{t-4}\}$ ; la Tabla 6.1 muestra en cada caso la dependencia del pasado de cada observación  $x_t$  en función del conjunto de regresores considerados:

Tabla 6.1:

Modelo A				Modelo B			
$X_t$	$X_{t-1}$	$X_{t-2}$	$X_{t-3}$	$X_t$	$X_{t-1}$	$X_{t-3}$	$X_{t-4}$
$x_3$	$x_2$	$x_1$	$x_0$	$x_4$	$x_3$	$x_1$	$x_0$
$x_4$	$x_3$	$x_2$	$x_1$	$x_5$	$x_4$	$x_2$	$x_1$
$x_5$	$x_4$	$x_3$	$x_2$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$x_{t-4}$	$x_{t-5}$	$x_{t-7}$	$x_{t-8}$
$x_{t-4}$	$x_{t-5}$	$x_{t-6}$	$x_{t-7}$	$x_{t-3}$	$x_{t-4}$	$x_{t-6}$	$x_{t-7}$
$x_{t-3}$	$x_{t-4}$	$x_{t-5}$	$x_{t-6}$	$x_{t-2}$	$x_{t-3}$	$x_{t-5}$	$x_{t-6}$
$x_{t-2}$	$x_{t-3}$	$x_{t-4}$	$x_{t-5}$	$x_{t-1}$	$x_{t-2}$	$x_{t-4}$	$x_{t-5}$
$x_{t-1}$	$x_{t-2}$	$x_{t-3}$	$x_{t-4}$	$x_t$	$x_{t-1}$	$x_{t-3}$	$x_{t-4}$
$x_t$	$x_{t-1}$	$x_{t-2}$	$x_{t-3}$				

Podemos observar como todos los datos que intervienen en la explicación del segundo conjunto de regresores están contenidos en los que explican el modelo A. Por tanto, la varianza explicada por este primer modelo será siempre mayor.

Aunque ya hemos expuesto las razones que justifican el resultado del teorema 25 vamos a formalizarlas en la siguiente demostración.

**Demostración del teorema 25:**

Sea  $\{X_t\}$  una serie temporal que sigue un modelo autoregresivo AR de orden máximo  $p$ . Supongamos dos posibles modelos autoregresivos de  $k$  regresores ( $k \leq p$ ), que denominamos *modelo A*:

$$\mathbf{X}_t = a_0^{(j)} + a_1^{(j)} X_{t-1} + a_2^{(j)} X_{t-2} + \dots + a_k^{(j)} X_{t-k} + \varepsilon_t^{(j)}$$

y *modelo B*:

$$\mathbf{X}_t = b_0^{(j)} + b_1^{(j)} X_{t-p_1} + b_2^{(j)} X_{t-p_2} + \dots + b_k^{(j)} X_{t-p_k} + \varepsilon_t^{(j)}$$

donde  $k < p_k \leq p$ .

Si  $n$  es el número de observaciones, el *modelo A* recoge para su explicación  $n - k$  y el *modelo B* un número inferior  $n - p_k$ . Dado el carácter autoregresivo del proceso las  $n - p_k$  observaciones del *modelo B* están contenidas en las  $n - k$  del *modelo A* y no aportan por tanto información adicional. Podemos afirmar que la varianza explicada por  $X_{t-1}, X_{t-2}, \dots, X_{t-k}$ , que vamos a designar a partir de este momento  $\sigma^2(X_{t-1}, X_{t-2}, \dots, X_{t-k})$ , es mayor que la varianza explicada por  $X_{t-p_1}, X_{t-p_2}, \dots, X_{t-p_k}$ , que simbolizamos por  $\sigma^2(X_{t-p_1}, X_{t-p_2}, \dots, X_{t-p_k})$ . Si consideramos  $\sigma^2(X_t)$  la varianza total del proceso entonces:

$$\sigma^2(X_t) = \sigma^2(X_{t-1}, X_{t-2}, \dots, X_{t-k}) + \sigma_\varepsilon^2$$

y también

$$\sigma^2(X_t) = \sigma^2(X_{t-p_1}, X_{t-p_2}, \dots, X_{t-p_k}) + \sigma_{\varepsilon'}^2$$

donde  $\sigma_\varepsilon^2$  y  $\sigma_{\varepsilon'}^2$  son las varianzas residuales respectivas de cada modelo. Como

$$\sigma^2(X_{t-1}, X_{t-2}, \dots, X_{t-k}) > \sigma^2(X_{t-p_1}, X_{t-p_2}, \dots, X_{t-p_k}) \Rightarrow \sigma_\varepsilon^2 < \sigma_{\varepsilon'}^2$$

por tanto la varianza residual mínima es  $\sigma_\varepsilon^2$ , es decir, la obtenida al considerar como regresores  $X_{t-1}, X_{t-2}, \dots, X_{t-k}$ .

Si,

$$\sigma_\varepsilon^2 < \sigma_{\varepsilon'}^2 \Rightarrow \ln(\sigma_\varepsilon^2) < \ln(\sigma_{\varepsilon'}^2) \quad (6.7)$$

y recordando la expresión (6.5) del AIC para cada modelo obtenemos:

$$AIC(k) = (n - p) \ln(\sigma_\varepsilon^2) + 2(k + 1) \quad (6.8)$$

$$AIC(p_k) = (n - p) \ln(\sigma_{\varepsilon'}^2) + 2(k + 1) \quad (6.9)$$

De las anteriores expresiones (6.7-6.8-6.9) se deduce que

$$AIC(k) < AIC(p_k) \quad \forall p_k, \quad p_k > k$$

y queda así demostrado el teorema.

■

El resultado del teorema 25 es aplicable a modelos autoregresivos (AR); como los procesos SETAR se formulan a partir de procesos AR, vamos a utilizar el teorema 25 en la demostración del resultado del teorema 24.

**Demostración del teorema 24:**

Sea una serie  $\{X_t\} t = 1, \dots, N$  y consideremos dos posibles modelos:

- El modelo  $SETAR(l; p_1, p_2, \dots, p_l)$  donde  $p_1, p_2, \dots, p_l$  son los órdenes de cada uno de los procesos autoregresivos ( $p_i \leq p \quad \forall i$ ), cada uno de los procesos autoregresivos de orden  $p_i$  esta explicado por  $k_i$  variables autoregresoras tal que  $\exists i \mid k_i < p_i$ .
- El modelo  $SETAR(l; k_1, k_2, \dots, k_l)$  donde  $k_1, k_2, \dots, k_l$  son los órdenes de cada uno de los procesos autoregresivos y también el número de variables autoregresoras de cada proceso.

El valor del  $AIC$  para cada modelo se obtiene a partir del  $AIC$  de cada uno de los procesos autoregresivos:

$$\begin{aligned} AIC(SETAR(l; p_1, p_2, \dots, p_l)) &= \sum_{i=1}^l n_i \ln(\hat{\sigma}_i^2) + 2 \sum_{i=1}^l (k_i + 1) \\ &= AIC(p_1) + AIC(p_2) + \dots + AIC(p_l) \end{aligned}$$

donde  $AIC(p_i)$  es el  $AIC$  del proceso autoregresivo de orden  $p_i$

$$\begin{aligned}
AIC(SETAR(l; k_1, k_2, \dots, k_l)) &= \sum_{i=1}^l n_i \ln(\hat{\sigma}_i^2) + 2 \sum_{i=1}^l (k_i + 1) \\
&= AIC(k_1) + AIC(k_2) + \dots + AIC(k_l)
\end{aligned}$$

y  $AIC(k_i)$  es el AIC del proceso autoregresivo de orden  $k_i$

Debemos demostrar que

$$AIC(SETAR(l; p_1, p_2, \dots, p_l)) > AIC(SETAR(l; k_1, k_2, \dots, k_l)) \quad (6.10)$$

o lo que es equivalente, que

$$\exists i \mid AIC(p_i) > AIC(k_i)$$

Por hipótesis el  $AIC(p_i)$  y el  $AIC(k_i)$  corresponden a dos procesos autoregresivos definidos para el mismo conjunto de datos (los del régimen  $i$ ) que tienen distinto orden pero el mismo número de variables regresoras  $k_i$  y además suponíamos que en el modelo  $SETAR(l; p_1, p_2, \dots, p_l)$

$\exists i \mid p_i > k_i$ .

Aplicando el teorema 25 obtenemos que

$$\exists i \mid AIC(k_i) < AIC(p_i)$$

por tanto la expresión 6.10 es cierta y en consecuencia sólo se podrá alcanzar el AIC mínimo cuando en el modelo  $SETAR(l; p_1, p_2, \dots, p_l)$   $p_i = k_i \quad \forall i$ . Podemos concluir que

$$MAICE(SETAR(l; p_1, p_2, \dots, p_l)) = AIC(SETAR(l; k_1, k_2, \dots, k_l))$$

expresión que asegura que el mínimo valor del AIC estimado se obtiene para un modelo  $SETAR$  donde los órdenes  $k_i$  de cada uno de los procesos autoregresivos coincide con el número de variables autoregresoras de cada proceso, es decir, un  $SETAR(l; k_1, k_2, \dots, k_l)$ .■

### Comentarios:

- Este teorema asegura que fijado el número de regresores  $k$  el AIC mínimo se obtiene siempre para el modelo

$$\mathbf{X}_t = a_0^{(j)} + a_1^{(j)} X_{t-1} + a_2^{(j)} X_{t-2} + \cdots + a_k^{(j)} X_{t-k} + \varepsilon_t^{(j)}$$

esto contradice los resultados obtenidos por Thanoon (1990) que considera posible obtener valores  $(l; p_1^1, \dots, p_{k_1}^1; p_1^2, \dots, p_{k_2}^2; \dots; p_1^l, \dots, p_{k_l}^l)$  donde  $p_1^i, \dots, p_{k_i}^i$  no tienen por qué ser valores consecutivos

- Si el criterio de especificación elegido es el BIC o el AICc los resultados de los anteriores teoremas siguen siendo válidos.

### Ejemplos:

Para ilustrar la tesis expuesta en el teorema 24 hemos calculado<sup>6</sup> para cada conjunto de posibles autoregresores el valor del AIC, BIC y de la varianza residual. El conjunto de datos objeto de estudio es la serie Linces Canadienses, y para que el número de autoregresiones consideradas no sea muy elevado vamos a suponer que el orden máximo<sup>7</sup> es  $p = 5$ , por tanto el número de subconjuntos de variables autoregresoras es  $2^p - 1 = 31$ . Hemos designado con  $u$  ( $u = 1, 2, \dots, 2^p - 1$ ) al valor que permite identificar el conjunto de variables autoregresoras. Los valores obtenidos se recogen en la Tabla 6.2 y permiten comprobar los resultados del teorema 24 y 25, a partir de ellos se observa que:

- El valor de la varianza residual es mínimo cuando consideramos el número máximo de regresores (Para  $u = 31$  el valor de la varianza residual es en el primer régimen 0,0273 y en el segundo 0,0482).
- El mínimo del *AIC*, *BIC* y la varianza residual no tiene por qué alcanzarse para el mismo conjunto de regresores. Podemos observar que en el *régimen 1* el valor del *MAICE*

---

<sup>6</sup>Estos cálculos han sido realizados mediante un código escrito en Fortran77.

<sup>7</sup>Queremos remarcar que con este ejemplo solo se pretende comprobar el resultado obtenido en el teorema 24. Si nuestro interés fuera estimar un modelo SETAR sería necesario considerar orden máximo  $p=10$ .

u	k	regresores*	Régimen 1			Régimen 2		
			AIC	BIC	Varianza	AIC	BIC	Varianza
1	1	(1)	-188,54	-184,28	0,0448	-103,26	-99,56	0,1021
2	1	(2)	-158,59	-154,33	0,0726	-33,98	-30,28	0,4457
4	1	(3)	-43,69	-39,43	0,4634	-3,93	-0,23	0,8447
8	1	(4)	-24,26	-20,01	0,6339	-8,32	-4,62	0,7694
16	1	(5)	-20,50	-16,25	0,6736	-5,06	-1,36	0,8247
3	2	(1,2)	-202,39	-196,01	0,0347	-134,41	-128,86	0,0504
5	2	(2,3)	-191,87	-185,49	0,0411	-112,01	-106,46	0,0812
6	2	(1,3)	-133,36	-126,98	0,1056	-74,42	-68,87	0,1807
9	2	(1,4)	-175,85	-169,47	0,0532	-97,08	-91,53	0,1116
10	2	(2,4)	-109,51	-103,13	0,1552	-44,16	-38,61	0,3440
12	2	(3,4)	-66,55	-60,17	0,3103	-28,44	-22,89	0,4806
17	2	(1,5)	-179,96	-173,58	0,0498	-93,41	-87,86	0,1206
18	2	(2,5)	-82,74	-76,36	0,2390	-21,86	-16,31	0,5528
20	2	(3,5)	-36,95	-30,57	0,5002	-14,11	-8,56	0,6519
24	2	(4,5)	-29,94	-23,56	0,5601	-9,46	-3,91	0,7197
7	3	(1,2,3)	-207,20	-198,69	0,0311	-132,53	-125,13	0,0503
11	3	(1,2,4)	-201,35	-192,84	0,0342	-132,45	-125,05	0,0504
13	3	(1,3,4)	-196,67	-188,16	0,0368	-107,49	-100,09	0,0857
14	3	(2,3,4)	-132,85	-124,34	0,1031	-71,66	-64,26	0,1836
19	3	(1,2,5)	-200,53	-192,02	0,0346	-131,22	-123,82	0,0517
21	3	(1,3,5)	-189,99	-181,48	0,0410	-108,49	-101,09	0,0839
22	3	(2,3,5)	-123,52	-115,01	0,1199	-64,89	-57,49	0,2120
25	3	(1,4,5)	-164,52	-156,01	0,0619	-86,84	-79,44	0,1329
26	3	(2,4,5)	-107,85	-99,35	0,1543	-42,45	-35,05	0,3418
28	3	(3,4,5)	-64,92	-56,41	0,3085	-27,26	-19,86	0,4723
15	4	(1,2,3,4)	-209,55	-198,91	0,0290	-130,58	-121,32	0,0502
23	4	(1,2,3,5)	-206,29	-195,65	0,0305	-128,81	-119,56	0,0522
27	4	(1,2,4,5)	-189,69	-179,06	0,0399	-128,13	-118,88	0,0529
29	4	(1,3,4,5)	-186,98	-176,34	0,0417	-102,09	-92,84	0,0921
30	4	(2,3,4,5)	-131,24	-120,60	0,1025	-69,49	-60,24	0,1843
31	5	(1,2,3,4,5)	-211,30	-198,53	0,0273	-130,50	-119,40	0,0482

Figura 6-3: Comprobación empírica de los resultados de los teoremas 24 y 25. (\*) Los números entre parentesis corresponden a los retardos de los regresores.

(mínimo AIC estimado) se obtiene para  $u = 31$  (al igual que la varianza), pero en cambio el *MBICE* (mínimo BIC estimado) corresponde a  $u = 15$ . En el *segundo régimen AIC* y *BIC* alcanzan el mínimo para  $u = 3$  mientras que el valor mínimo de la varianza residual se obtiene para  $u = 31$ .

- Fijado el número de regresores  $k$ , el AIC y el BIC mínimo corresponden al valor de  $u$  menor de los considerados. En este ejemplo  $k$  varía desde 1 a 5, para cada valor de  $k$  el valor de  $u$  en el que hace mínimo *AIC* y *BIC* és:

$$\begin{array}{lll}
 k = 1 & u = 1 & X_{t-1} \\
 k = 2 & u = 3 & X_{t-1}, X_{t-2} \\
 k = 3 & u = 7 & X_{t-1}, X_{t-2}, X_{t-3} \\
 k = 4 & u = 15 & X_{t-1}, X_{t-2}, X_{t-3}, X_{t-4} \\
 k = 5 & u = 31 & X_{t-1}, X_{t-2}, X_{t-3}, X_{t-4}, X_{t-5}
 \end{array}$$

Este resultado coincide con el del teorema 24.

- El modelo *SETAR* con *AIC* mínimo es un *SETAR*(2; 5, 3) donde se estiman los coeficientes de las variables  $X_{t-1}, X_{t-2}, X_{t-3}, X_{t-4}, X_{t-5}$  en el primer régimen y los coeficientes de las variables  $X_{t-1}, X_{t-2}, X_{t-3}$  para el segundo régimen. Si hacemos mínimo el BIC el modelo identificado es un *SETAR*(2; 4, 3) donde las variables autoregresoras que intervienen son  $X_{t-1}, X_{t-2}, X_{t-3}, X_{t-4}$  en el primer régimen y  $X_{t-1}, X_{t-2}, X_{t-3}$  en el segundo.

### 6.3 Mejoras obtenidas

Mejorar la automatización del proceso de estimación e identificación era el objetivo principal de nuestra propuesta. Estas mejoras se han concretado en :

1. Identificación automática de los órdenes de los procesos autoregresivos que conforman el modelo *SETAR*.
2. Selección automática de las variables autoregresoras que explican cada uno de los procesos *AR* del modelo y estimación automática de los coeficientes correspondientes.

### 3. Para modelos de 2 regímenes estimación automática del umbral.

Respecto al primer aspecto enumerado queremos mencionar que, tanto en la metodología de Tong, como en la de Tsay, es preciso un refinamiento final (no automático) para determinar los órdenes de los procesos autoregresivos. Así, Tong estima ya en la primera etapa de su algoritmo los órdenes de los procesos autoregresivos utilizando la minimización del AIC. Esta estimación tan temprana del orden, anterior a la estimación de  $d$  y  $r$ , no proporciona un valor definitivo; es necesario, una vez completado el proceso algorítmico, refinar la estimación de estos parámetros mediante el análisis de las características de los modelos obtenidos al hacer variar los valores de  $L$  (orden máximo de los procesos AR en el algoritmo de Tong).

El algoritmo de Tsay mantiene durante todo el proceso el orden máximo de los procesos autoregresivos  $\hat{p}$  fijado al inicio del mismo y es en la última etapa (una vez estimados  $d$  y  $r$ ) cuando se refina el orden de cada modelo AR haciendo variar ligeramente este parámetro. Este refinamiento se realiza sobre un número menor de modelos al haberse estimado de forma exacta el valor de  $d$ .

Una vez que se ha determinado el orden de cada proceso AR, no tiene por que resultar automáticamente determinado el número de variables regresoras que explican dicho modelo. Tsay considera como regresoras todas las variables retardadas desde 1 hasta  $\hat{p}$ , por tanto, se obtiene la estimación de todos los coeficientes aunque algunos de ellos no sean significativos. El modelo obtenido tiene orden  $\hat{p}$ , y por tanto se habrán estimado  $\hat{p} + 1$  coeficientes, el analista debe decidir mediante la utilización de un test qué coeficientes desprecia y cuales mantiene en el modelo. Como ya se ha comentado en el capítulo 4, la aplicación de un test supone la elección del nivel de significación, lo que introduce un componente subjetivo en la determinación del modelo.

Thanoon ha propuesto una metodología para seleccionar las variables autoregresoras; según este autor si el orden del proceso es  $k$ , es posible que el mejor modelo tenga como variables explicativas un subconjunto de  $k'$  variables, donde  $k' < k$  ( recordemos los resultados obtenidos para la serie Linces Canadienses, Capítulo 5, Sección 3) . El teorema 24 contradice los resultados empíricos obtenidos por Thanoon.

La estimación automática del umbral ha sido abordada por Tong y también existen trabajos que estiman este parámetro a partir de técnicas bayesianas. El interés de nuestro proceso de

estimación del umbral reside en la utilización del mismo criterio de información que hemos utilizado para la estimación del resto de parámetros.

En resumen, nuestra propuesta se presenta esencialmente como una mejora del algoritmo de Tsay, pues parte de éste para su diseño, pero su aportación va más allá. La metodología que hemos elaborado realiza la estimación e identificación del modelo de forma conjunta a partir de un único proceso algorítmico gobernado en todas sus etapas por una misma ley, la minimización del criterio de información elegido. Esta unicidad de criterio en todos los procesos es una mejora conceptual del diseño algorítmico.