

Introducción a los Modelos TAR Enfocada a los Modelos SETAR

Johana Gómez G.

May 6, 2019

Consideremos un proceso $\{Z_t\}$

$$Z_t = \sum_{j=0}^{\infty} \psi_j a_{t-j} + \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \psi_{i,j} a_{t-i} a_{t-j} + \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \psi_{i,j,k} a_{t-i} a_{t-j} a_{t-k} + \dots \quad (1)$$

Donde $a_t \sim RB(0, \sigma_a^2)$

Z_t es lineal si solo se tiene el primer término en la ecuación (1).

Z_t es no lineal si tiene más del primer término en la ecuación (1).

Los modelos de series de tiempo no lineales fueron mencionados por primera vez a principios del siglo XX, estos han adquirido gran importancia a través del tiempo. Entre estos, están los modelos no lineales TAR (Threshold AutoRegressive models) los cuales, fueron propuestos por Tong (1978). Los modelos TAR tienen gran popularidad debido a que son relativamente simples de especificar, estimar e interpretar en comparación con muchos otros modelos de series de tiempo no lineales; y además, porque tienen la capacidad de capturar asimetrías, bimodalidad, fenómenos de saltos, irreversibilidad en el tiempo

Especificación Modelos TAR

Dada una partición de la recta real en l intervalos o regímenes, $\mathbb{R} = \bigcup_{j=1}^l R_j$, donde $R_j = (r_{j-1}, r_j]$ y $r_0 := -\infty$ y $r_l := \infty$, entonces, la especificación del modelo TAR esta dada por:

$$X_t = a_0^{(j)} + \sum_{i=1}^{k_j} a_i^{(j)} X_{t-i} + \epsilon_t \quad (2)$$

si $Z_t \in R_j$ para algún $j = 1, \dots, l$.

Los valores r_j con $r_1 < r_2 < \dots < r_{l-1}$ se denominan valores umbrales y definen l regímenes para el proceso Z_t , denominado "proceso de umbrales"; los valores k_1, \dots, k_l son números enteros positivos que representan los órdenes autorregresivos en los l regímenes. Además, los coeficientes autorregresivos $a_i^{(j)}$ $j = 1, \dots, l$; $i = 0, \dots, k_j$, son números reales; $\epsilon_t \sim iid(0, \sigma_j^2)$ y son independientes del proceso de umbrales $\{Z_t\}$.

$TAR(l; k_1, \dots, k_l)$ representará un modelo TAR con l regímenes y los respectivos órdenes autorregresivos para cada régimen k_1, \dots, k_l . Los parámetros del modelo TAR se dividen en dos grupos:

- *Parámetros estructurales*: El número de regímenes l , los $l - 1$ valores umbrales r_1, r_2, \dots, r_{l-1} y los órdenes autorregresivos en los l regímenes k_1, \dots, k_l .
- *Parámetros no estructurales*: Los coeficientes autorregresivos $a_i^{(j)}$ con $j = 1, \dots, l; i = 0, \dots, k_j$, las varianzas $\sigma_1^2, \dots, \sigma_l^2$.

Son definidos como en (2) considerando que el proceso de umbrales es un cierto valor rezagado del propio proceso, esto es, $l = 2$ y $Z_t = X_{t-d}$ donde d es un entero no negativo llamado parámetro de rezago. Luego $TAR(2; k_1, k_2, d)$ será

$$X_t = a_0^{(j)} + \sum_{i=1}^{k_j} a_i^{(j)} X_{t-i} + \epsilon_t; \quad \text{si } X_{t-d} \in R_j \quad j = 1, 2 \quad (3)$$

Con las mismas especificaciones que en (2) para los otros términos.

Ejemplo SETAR

Simulamos una serie de tiempo SETAR de dos regímenes, con desviaciones estándar $\sigma_1 = 1, \sigma_2 = 2$ para los regímenes inferior y superior, respectivamente; y un tamaño de la muestra $n = 100$

$$X_t = \begin{cases} 0.5X_{t-1} + \epsilon_t^{(1)} & X_{t-1} \leq -1 \\ -1.8X_{t-1} + \epsilon_t^{(2)} & X_{t-1} > -1 \end{cases}$$

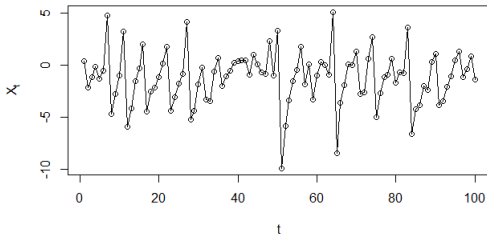


Figure: SETAR(2,1,1)

Supondremos que los ruidos $\epsilon_t \sim N(0, \sigma_j^2)$ $j = 1, 2$, y que los valores r, d están fijos, entonces, los casos de datos se pueden dividir en dos partes de acuerdo a si $X_{t-d} \leq r$ (régimen inferior) ó $X_{t-d} > r$ (régimen superior), así, con los datos en el régimen inferior ($n_1(r, d)$) podemos hacer regresión a X_t con los rezagos desde 1 hasta k_1 y obtener los valores estimados de $\hat{a}_0^{(1)}, \hat{a}_1^{(1)}, \dots, \hat{a}_{k_1}^{(1)}$ y por máxima verosimilitud $\hat{\sigma}_1^2 = \frac{SSE}{n_1}$ De forma similar para datos en el régimen superior ($n_2(r, d)$) se pueden obtener los valores estimados de $\hat{a}_0^{(2)}, \hat{a}_1^{(2)}, \dots, \hat{a}_{k_1}^{(2)}$ y $\hat{\sigma}_2^2 = \frac{SSE}{n_2}$

Estimación del modelo

Sustituyendo estos valores estimados en la función log-verosimilitud se tiene la llamada función de perfil log-verosimilitud de (r,d) :

$$\begin{aligned}l(r, d) &= -\frac{n_1 + n_2}{2}[1 + \log(2\pi)] - \frac{n_1(r, d)}{2}\log((\hat{\sigma}_1(r, d))^2) \\ &\quad - \frac{n_2(r, d)}{2}\log((\hat{\sigma}_2(r, d))^2)\end{aligned}$$

Los estimadores de r y d pueden ser obtenidos por maximizar esta función.

Recordemos que para los modelos ARMA los órdenes AR se pueden estimar minimizando el AIC; y como es evidente, para r y d fijos, el modelo TAR se ajusta esencialmente a dos modelos AR de órdenes k_1 y k_2 , respectivamente, de modo que el AIC se convierte en

$$AIC(k_1, k_2, r, d) = -2l(r, d) + 2(k_1 + k_2 + 2) \quad (4)$$

Donde el número de parámetros excluyendo a r , d , σ_1 y σ_2 es igual a $k_1 + k_2 + 2$

Ejemplo

En la estimación el orden máximo se establece en $p = 4$ y $1 \leq d \leq 4$. Usando el método en el que consideramos el mínimo AIC, en la tabla (1) podemos observar que el valor del AIC es mas pequeño cuando $d=3$, y los órdenes para los regímenes inferior y superior respectivos son $k_1 = 1, k_2 = 4$.

Rezago	AIC	\hat{r}	\hat{k}_1	\hat{k}_2
1	19.04	4.15	2	1
2	12.15	4.048	1	4
3	10.92	4661	1	4
4	18.42	5.096	3	4

Table: AIC Mínimo

Ahora, si ajustamos un modelo $TAR(2,1,4)$ los valores estimados para los parámetros serán:

Parámetro	Valor estimado
\hat{d}	3
\hat{r}	4.661
Régimen inferior	
$\hat{a}_0^{(1)}$	0.262
$\hat{a}_1^{(1)}$	1.02
$\hat{\sigma}_1^2$	0.0548
Régimen Superior	
$\hat{a}_0^{(2)}$	4.20
$\hat{a}_1^{(2)}$	0.708
$\hat{a}_2^{(2)}$	-0.301
$\hat{a}_3^{(2)}$	0.279
$\hat{a}_4^{(2)}$	-0.611
$\hat{\sigma}_2^2$	0.0560

Tabla: Parámetros Estimados

Una gráfica para la serie de tiempo simulada del modelo ajustado $TAR(2,1,4)$ a la serie predator es dada por:

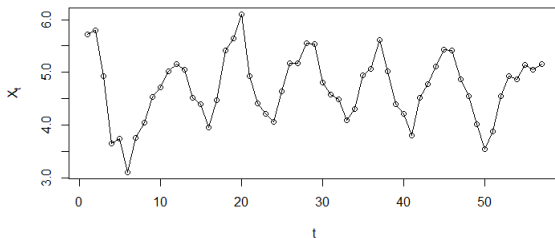


Figure: $TAR(2,1,4)$ simulado

Los residuos brutos son definidos como la resta entre los valores ajustados y los datos, donde el t -ésimo valor ajustado es la media estimada de X_t condicionada a los valores pasados de X , estos residuales estan dados por:

$$\begin{aligned}\hat{\epsilon}_t &= X_t - \{\hat{a}_0^{(1)} + \hat{a}_1^{(1)}X_{t-1} + \cdots + \hat{a}_{k_1}^{(1)}X_{t-k_1}\}I(X_{t-\hat{d}} \leq \hat{r}) \\ &\quad - \{\hat{a}_0^{(2)} + \hat{a}_1^{(2)}X_{t-1} + \cdots + \hat{a}_{k_2}^{(2)}X_{t-k_2}\}I(X_{t-\hat{d}} > \hat{r})\end{aligned}\quad (5)$$

Donde $I()$ es la función indicadora.

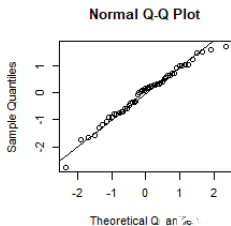
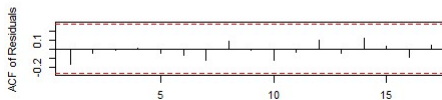
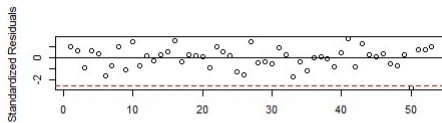
Los residuales estandarizados son obtenidos de normalizar los residuales brutos con su desviación estándar apropiada:

$$\hat{e}_t = \frac{\hat{\epsilon}_t}{\hat{\sigma}_1 I(X_{t-\hat{d}} \leq \hat{r}) + \hat{\sigma} I(X_{t-\hat{d}} > \hat{r})} \quad (5)$$

Al igual que en el caso lineal, si el modelo TAR es el apropiado, entonces los residuales deben comportarse de manera consistente con el modelo, es decir, los residuales estandarizados deberán ser aproximadamente independientes e idénticamente distribuidos

Ejemplo

Observamos las gráficas



Para los modelos ARIMA con errores normales, se tiene que las distribuciones predictivas son normales, lo que simplifica el cálculo de un intervalo de predicción, ya que es suficiente con encontrar la media y la varianza de la distribución predictiva. Sin embargo, para los modelos no lineales, las distribuciones predictivas no son, generalmente, normales, y a menudo son difíciles de tratar. Por lo tanto, un intervalo de predicción para los modelos TAR puede tener que ser calculado por la fuerza bruta mediante simulación.

Ejemplo

Basados en el modelo ajustado $TAR(2,1,4)$ con rezago $d = 3$, y haciendo uso de la función `predict` en R, computamos los intervalos de predicción para la serie de tiempo predator transformada logarítmicamente.

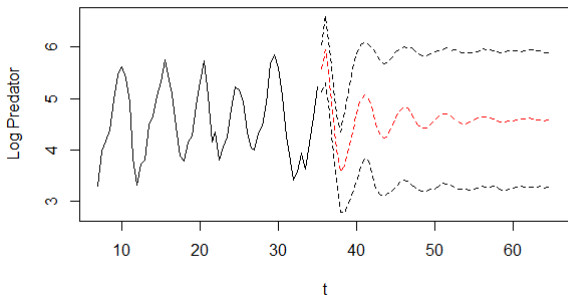









Figure: Pronóstico del modelo $TAR(2,1,4)$

- El proceso de umbrales $\{Z_t\}$ en el modelo TAR es exógeno para $\{X_t\}$, en el sentido de que no hay retroalimentación de $\{X_t\}$ hacia $\{Z_t\}$
- Es posible tener estacionariedad en el modelo TAR aún cuando en un régimen se pueda ver no estacionariedad. (Ejemplo 1).
- Para el modelo SETAR se puede tener que en cada régimen sea un modelo lineal autorregresivo; mientras que el modelo TAR no es necesariamente lineal a trozos, como podría sugerirse de la especificación de este.
- Para los modelos ARIMA con errores normales, se tiene que las distribuciones predictivas son normales, lo que simplifica el cálculo de un intervalo de predicción, Sin embargo, para los modelos no lineales, en particular los modelos TAR, las distribuciones predictivas no son, generalmente, normales, y a menudo son difíciles de tratar.

-  Borja Lafuente Rego, *Implementación integrada de modelos de predicción semiparamétricos de series de tiempo*
-  Edna Carolina Moreno Lopez, *Una aplicación del modelo TAR en series de tiempo financieras*
-  Hanwen Zhang, *Estimación de los modelos TAR cuando el proceso del ruido sigue una distribución t*
-  Hanwen Zhang, Fabio Humberto Nieto, *A new R package for TAR modeling*
-  Holmes fabian Molina alape, *Comparación de Pronósticos con Modelos TAR y TARX en Algunas Series Económicas Colombianas*
-  Jonathan D. Cryer, Kung-Sik Chan, *Time Series Analysis With Applications in R*
-  Robert H.Shumway, David S. Stoffer, *Time Series Analysis and Its Applications With R Examples*