

UNIVERSIDAD DE ANTIOQUIA
FACULTAD DE CIENCIAS EXACTAS Y NATURALES

3/Mayo/2019

Series de Tiempo: Modelos TAR

PRESENTADO POR:

JOHANA GÓMEZ GÓMEZ

CC:1036947051

Introducción a los Modelos TAR Con Base en los Modelos SETAR

1. RESUMEN

Los modelos que ya hemos trabajado anteriormente como ARMA y ARIMA son ejemplos de modelos lineales de series temporales; en los últimos años, ha sido de total interés el análisis de series de tiempo no lineales, ya que, algunos fenómenos naturales y sociales como: series meteorológicas, económicas, financieras, entre otras, pueden ser explicados por medio de estos. En este trabajo nos centraremos en el estudio y conocimiento de los modelos de series de tiempo no lineales paramétricos de umbral autorregresivos (TAR), considerando la especificación, estimación, predicción y validez, para un caso particular de estos modelos, los modelos SETAR, y se observarán posteriormente los resultados estudiados con un ejemplo.

2. PALABRAS CLAVES

Modelos TAR, Modelos SETAR, Rezago, Parámetros estructurales y no estructurales, Valores y Procesos umbrales.

3. OBJETIVO

Conocer, estudiar y aprender a cerca de los modelos de series de tiempo no lineales paramétricos de umbral autorregresivos (TAR), teniendo en cuenta resultados para el caso particular SETAR.

4. INTRODUCCIÓN

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.

En el estudio y trabajo con los modelos de series de tiempo no lineales paramétricos de umbral autorregresivos se han obtenido varios resultados, Tong (2015) ha dado una amplia revisión de la literatura que se ha desarrollado respecto a los modelos TAR desde su aparición; para los cuales, Chen y Lee (1995) han presentado una metodología Bayesiana para estimar sus parámetros en conjunto, Nieto (2005) ha implementado una metodología Bayesiana y técnicas MCMC (Monte Carlo Markov Chain) para estimar sus parámetros y datos faltantes en forma simultánea, Nieto (2008) presenta la fase de pronósticos en estos modelos, Hansen (2011) da una revisión extensiva de estos modelos en economía y Chen et al. (2011) dan una revisión similar de los modelos en finanzas, Hoyos (2007) presenta un estudio entre la tasas de interés y el índice de producción industrial para Colombia con dichos modelos, Vivas (2011) haciendo uso de este modelo, presenta un análisis del desempleo trimestral y el índice de confianza industrial trimestral en Colombia, Salamanca y Monroy (2008) analizan la relación existente entre la deuda externa pública y la inversión privada en Colombia a partir de estos modelos, siguiendo la metodología de Nieto (2008) pero sin desarrollar la fase de pronósticos.

Como ya se ha mencionado, el principal objetivo de este trabajo es conocer y estudiar algunos resultados sobre los modelos de series de tiempo no lineales paramétricos de umbral autorregresivos (TAR), en particular para los modelos SETAR, y así, visualizar lo aprendido por medio de algunos ejemplos. Dicho esto, procedamos.

5. MARCO TEÓRICO

5.1. Especificación del modelo

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:

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

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 , el cual es 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 $\{Z_t\}$.

Usamos el símbolo $TAR(l; k_1, \dots, k_l)$ para representar un modelo TAR con l regímenes y los respectivos órdenes autorregresivos para cada régimen k_1, \dots, k_l .

Los parámetros del anterior modelo TAR se pueden dividir 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 ordenes 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 $\sigma_1^2, \dots, \sigma_l^2$.

El modelo TAR describe una relación no lineal entre las variables X_t y Z_t , la cual no es necesariamente lineal a trozos, como podría sugerirse de la especificación del modelo o como sucede en el modelo SETAR (definidos más adelante).

6. METODOLOGÍA

Con el fin de estudiar un poco más a fondo y ver algunas aplicaciones de los modelos de umbrales autorregresivos, en este trabajo se enfocará la atención en los modelos de umbral autorregresivo auto-excitante SETAR (Self-Exciting Threshold Autoregressive) con dos regímenes, los cuales son definidos como en (1) 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

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

Con las mismas especificaciones que en (1) para los otros términos.

El proceso es un modelo lineal autorregresivo en cada régimen y es no lineal cuando hay al menos dos regímenes con diferentes modelos lineales.

Este modelo se convierte en un modelo AR lineal no homogéneo cuando solo las variaciones del ruido difieren para diferentes regímenes; y se convierte en un modelo de cambio de nivel aleatorio cuando solo los términos constantes a_0 difieren para diferentes regímenes.

Con base en estos modelos, procederemos a conocer resultados y técnicas para hacer estimación, diagnóstico y predicción sobre ellos.

6.1. Estimación del Modelo

Para este caso asumiremos que los ruidos $\epsilon_t \sim N(0, \sigma_j^2)$ $j = 1, 2$, esto implica que la respuesta es condicionalmente normal; asumiremos además, que el valor umbral del modelo r , y el parámetro de rezago d , son conocidos, 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,

digamos que hay n_1 , podemos retroceder X_t con los rezagos desde 1 hasta k_1 y encontrar los valores estimados de $\hat{a}_0^{(1)}, \hat{a}_1^{(1)}, \dots, \hat{a}_{k_1}^{(1)}$ y por máxima verosimilitud un estimador para la varianza del ruido $\hat{\sigma}_1^2 = \frac{SSE}{n_1}$; n_1 y los parámetros estimados, generalmente dependen de r y d , luego por claridad es mejor escribir $n_1(r, d)$. De forma similar para los n_2 datos en el régimen superior se pueden obtener los valores estimados de $\hat{a}_0^{(2)}, \hat{a}_1^{(2)}, \dots, \hat{a}_{k_1}^{(2)}$ y $\hat{\sigma}_2^2 = \frac{SSE}{n_2}$. Sustituyendo estos valores estimados en la función log-verosimilitud se tiene la llamada función log-verosimilitud de perfil de (r, d) :

$$l(r, d) = -\frac{n_1 + n_2}{2} [1 + \log(2\pi)] - \frac{n_1(r, d)}{2} \log((\hat{\sigma}_1(r, d))^2) - \frac{n_2(r, d)}{2} \log((\hat{\sigma}_2(r, d))^2) \quad (3)$$

Los estimadores de r y d pueden ser obtenidos por maximizar la función log-verosimilitud de perfil anterior.

Por otro lado, como se especifica en el modelo y como se verá en la práctica, los órdenes del AR en los dos regímenes no necesariamente son iguales o conocidos, por lo tanto, será esencial encontrar un procedimiento de estimación eficiente que los estime. 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 $k_1 + k_2 + 2$

6.2. Diagnósticos del Modelo

Como sabemos de los modelos ya conocidos, la bondad de ajuste de un modelo estadístico a un conjunto de datos, se mide comparando los valores observados con los valores ajustados obtenidos del modelo ajustado, por lo tanto, se visualizarán algunos enfoques estadísticos formales para modelar diagnósticos mediante análisis residual.

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 están dados por:

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

Donde $I(\cdot)$, es la función indicadora.

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

$$\hat{e}_t = \frac{\hat{\epsilon}_t}{\hat{\sigma}_1 I(X_{t-\hat{d}} \leq \hat{r}) + \hat{\sigma}_2 I(X_{t-\hat{d}} > \hat{r})} \quad (5)$$

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. La suposición de independencia de los errores estandarizados puede verificarse examinando la ACF de los residuales estandarizados, la heterocedasticidad puede ser chequeada examinando la ACF de los residuales estandarizados al cuadrado.

6.3. Predicción

Ahora se considerará el problema de predecir valores futuros de un proceso TAR. En la práctica, la predicción es basada sobre un modelo ajustado TAR; procediendo como si el modelo ajustado fuera el verdadero modelo.

La incertidumbre de un valor futuro, digamos X_{t+l} se caracteriza completamente por su distribución de probabilidad condicionada por las observaciones pasadas, X_t, X_{t-1}, \dots , referida como la distribución predictiva l pasos adelante.

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. (Veremos mas adelante en un modelo aplicado)

7. APLICACIÓN

En esta sección vamos a ver 2 ejemplos.

1. En el primero, simulamos una serie de tiempo SETAR de dos regímenes, $l = 1$, con parámetro de rezago $d = 1$, valor umbral $r = -1$ el proceso es un AR(1) con coeficientes $a_0^{(1)} = 0, a_1^{(1)} = 0,5, a_0^{(2)} = 0, a_1^{(2)} = -1,8$ y desviación estándar $\sigma_1 = 1, \sigma_2 = 2$ para los regímenes inferior y superior, respectivamente; el tamaño de la muestra es de $n = 100$.(ver figura1)

Note que el modelo TAR simulado en el ejemplo, es condicionalmente heteroscedástico, ya que la desviación estándar para el régimen superior es el doble que para la del régimen inferior.

El coeficiente autorregresivo del submodelo en el régimen superior es igual a -1.8, pero los datos simulados parecen ser estacionarios (ver figura1), lo que puede ser inesperado desde una perspectiva lineal.

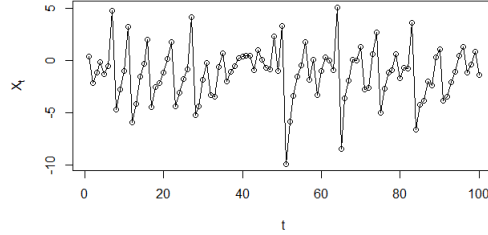


Figura 1: SETAR(2,1,1)

2. En el segundo ejemplo, simularemos un modelo SETAR, considerando la segunda serie (predators; en realidad se trabajará con la transformación logarítmica de esta serie) del Data frame veillex en R, el cual consiste de una serie de tiempo bivariada de un experimento que estudió la fluctuación de la población de un sistema depredador-presa; y observamos:

- a) **Estimación:** 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	4.661	1	4
4	18.42	5.096	3	4

Cuadro 1: AIC Mínimo

Ahora, si ajustamos el modelo $TAR(2, 1, 4)$ considerando lo anterior, tendremos que los respectivos valores estimados para los parámetros serán: (ver tabla2)

Donde los coeficientes autorregresivos son ajustados por mínimos cuadrados y la desviación estándar por máxima verosimilitud, en cada régimen.

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

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

Cuadro 2: Parámetros Estimados

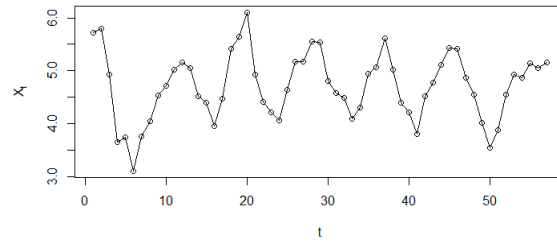


Figura 2: TAR(2,1,4) simulado

- b) **Diagnóstico** Haciendo uso de los residuales y los residuales estandarizados definidos, procedemos a hacer diagnóstico sobre el modelo ajustado $TAR(2,1,4)$. En la figura (3) observamos las gráficas de los valores ajustados de la serie vs los residuales estandarizados y la ACF de los residuales estandarizados, respectivamente. De estas gráficas, se puede observar uniformidad de los residuales alrededor del cero y como las autocorrelaciones caen en el interior de las bandas $\pm 1,96/\sqrt{n}$, donde n es el tamaño de la muestra.

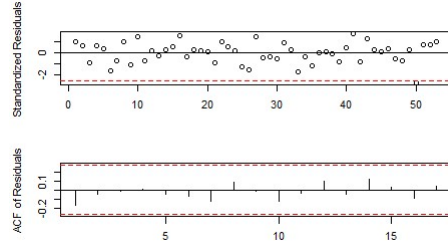


Figura 3: Diagnóstico del modelo $TAR(2,1,4)$

La figura(4) muestra el gráfico "Normal Q-Q plot" de los residuales estandarizados, en esta se puede ver que los puntos están, aproximadamente, sobre una línea recta, por lo tanto, se comprueba el supuesto de normalidad de los residuales.

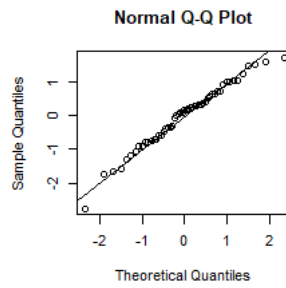


Figura 4: Diagnóstico del modelo $TAR(2,1,4)$

Con estos análisis, se tiene entonces que los residuales se comportan de manera consistente con el modelo, luego, se puede afirmar que el modelo es apropiado.

- c) **Predicción:** 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 logáritmicamente. (ver figura5)

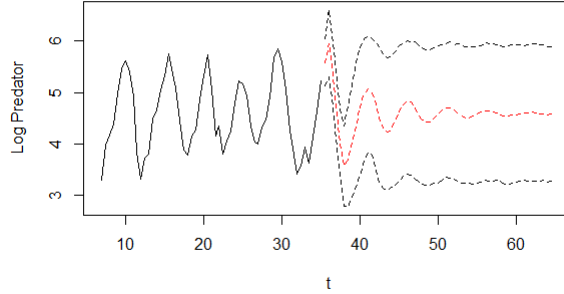


Figura 5: Pronóstico del modelo TAR(2,1,4)

Donde la línea discontinua roja es la mediana de la distribución predictiva y las otras líneas discontinuas nos dan un intervalo de confianza para el pronóstico.

8. CONCLUSIONES

- 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.

Referencias

- [1] Borja Lafuente Rego, *Implementación integrada de modelos de predicción semiparamétricos de series de tiempo*
- [2] Edna Carolina Moreno Lopez, *Una aplicación del modelo TAR en series de tiempo financieras*

- [3] Hanwen Zhang, *Estimación de los modelos TAR cuando el proceso del ruido sigue una distribución t*
- [4] Hanwen Zhang, Fabio Humberto Nieto, *A new R package for TAR modeling*
- [5] Holmes fabian Molina alape, *Comparación de Pronósticos con Modelos TAR y TARX en Algunas Series Económicas Colombianas*
- [6] Jonathan D. Cryer, Kung-Sik Chan, *Time Series Analysis With Applications in R*
- [7] Robert H.Shumway, David S. Stoffer, *Time Series Analysis and Its Applications With R Examples*