

CIMAT

Centro de Investigación
en Matemáticas, A.C.

ESTIMACIÓN DE PARÁMETROS
Y PRONÓSTICOS EN MODELOS
TAR CON ERRORES t -STUDENT

Tesis

Para obtener el título de:

Maestría en Ciencias con Especialidad
Probabilidad y Estadística

Presenta:

Miguel Ángel Sánchez Ovando

Director de tesis:

Dra. Graciela González Farías

Guanajuato, Gto., Julio 2015.

Agradecimientos

A Dios por estar a mi lado y darme la paz y fuerza que necesitaba para terminar con éxito ésta etapa de mi vida.

A mis padres y a mis sobrinos por estar siempre conmigo, por apoyarme en todo momento, por su cariño y amor. Porque aun estando lejos se convirtieron en un pilar importante para terminar la maestría y este trabajo. ¡Les amo!.

A la Dra. Graciela Gonzalez por su excelente dirección en este trabajo, su paciencia, dedicación, amabilidad y muestras de ánimos. De corazón ¡Muchas gracias!.

Al Dr. Nelson Muriel y al Dr. José Batún por su tiempo dedicado para leer este trabajo y por sus excelentes correcciones que llevaron a mejorar la versión final de la tesis.

A Kourai, por su comprensión, cariño y apoyo en la elaboración de este trabajo. ¡Te quiero!.

A mis compañeros de maestría que compartieron conmigo el cubículo N6, por sus ánimos, apoyos y desvelos. En especial quiero agradecerles a mis compañeros y amigos: Adán Uribe, Nicolas Kuschinski, Javier Rivera, Ernesto Ramos y Miriam Baez.

Por último, quiero agradecer al Centro de Investigación en Matemáticas A.C. (CIMAT) por la excelente preparación académica que me brindó y al Consejo Nacional de Ciencia y Tecnología (CONACYT) por la beca No. 280963 que me otorgó a lo largo de mis estudios de maestría.

Índice general

1. Introducción	4
2. Modelo TAR	8
2.1. Estimación de parámetros del modelo TAR(2; 1, 1)	9
2.1.1. Mínimos Cuadrados	10
2.1.2. Máxima Verosimilitud	12
3. Distribuciones Elípticas	15
3.1. Definición	15
3.2. Algunas propiedades	17
3.3. Ejemplos	18
3.3.1. Distribución Normal	18
3.3.2. Distribución t -student	18
3.4. Estimación de parámetros de la distribución t -student por máxima verosimilitud	20
3.5. Regresión con distribuciones esféricas	25

4. Estimación de parámetros en modelos TAR ($r : p_1, p_2, \dots, p_r$)	30
4.1. Estimación por máxima verosimilitud	32
4.2. Simulaciones	37
5. Identificación del modelo	49
5.1. Selección del número máximo de regímenes	50
5.2. Selección de los umbrales y órdenes autorregresivos	53
5.3. Simulaciones	56
6. Pronósticos mediante verosimilitud predictiva	65
6.1. Verosimilitud predictiva perfil de un TAR(2; 1, 1)	67
6.2. Verosimilitud predictiva perfil de un modelo TAR($r; p_1, \dots, p_r$)	72
6.3. ECMP	74
6.4. Simulaciones	78
6.5. Resumen	112
7. Conclusiones	114
8. Apéndice	116

Capítulo 1

Introducción

Hasta antes de finales de la década de los 70's el estudio de las series de tiempo se había basado asumiendo linealidad, usando esta suposición se propusieron diversos modelos, como por ejemplo, los autorregresivos (AR) o promedios móviles (MA), además, ya se contaba con una teoría desarrollada para estimación de los parámetros, intervalos de confianza para los parámetros, pronósticos, etc. Sin embargo, muchos datos pueden presentar comportamientos tales como no normalidad, ciclos asimétricos, relaciones no lineales con los retrasos, etc., los cuales son difíciles de describir usando modelos lineales. Algunas de las características anteriores fueron detectadas en los datos de lince de Canadá o en el de manchas solares, datos que abrieron el camino hacia el estudio de series de tiempo no lineales.

De acuerdo con Tsay(2002) un modelo puramente estocástico para una serie de tiempo y_t es una sucesión *iid* que consiste de los shocks presentes y pasados,

$$y_t = f(\epsilon_t, \epsilon_{t-1}, \dots),$$

donde cualquier función no lineal $f(\cdot)$ representa un modelo no lineal para

y_t .

A lo largo de las últimas cuatro décadas este tipo de modelos han sido extensamente estudiado debido a su gran aplicabilidad en áreas como ecología, econometría, hidrología, finanzas, sociología, etc.. Muchos modelos han sido formulados entre los cuales se encuentran los modelos bilineales de Granger y Anderson, los modelos de Markov cambiantes de Hamilton, los modelos umbrales propuestos Tong y Lim, los modelos ARCH de Engle, entre otros.

En economía, uno de los modelos no lineales más utilizados son los modelos umbrales y la razón se debe principalmente a las características que presentan los datos económicos. Prueba de la popularidad y aceptación de los modelos umbrales es la gran cantidad de artículos sobre teoría y aplicaciones que se han escrito. Un breve resumen de las publicaciones actuales se encuentra en Hansen(2011).

La idea detrás de los modelos autorregresivos umbrales propuestos por Tong (1978) y Tom y Lim (1980) es la linealización por partes de la serie, lo cual se hace introduciendo una variable llamada umbral y que tiene como propósito hacer que la serie se comporte como un modelo autorregresivo dentro de cada régimen que es delimitado por valores denominados umbrales.

La mayoría de las veces cuando se hacen supuestos distribucionales en los modelos autorregresivos umbrales, se asume que los errores tienen distribución normal con cierta media μ y varianza σ^2 . Como bien sabemos, el uso de la distribución normal en los errores suele ser muy restrictivo y presenta problemas con datos que tienen colas pesadas, un ejemplo de ellos son los datos financieros.

Una idea para solucionar este tipo de problemas podría ser emular lo hecho en análisis de regresión, que para atacar los problemas de no normalidad una alternativa es cambiar la distribución asociada al proceso de errores a una clase de distribuciones más amplia como pueden ser las distribuciones elípticas.

Para resolver la problemática con datos que presentan colas pesadas, Zhang (2012) propone que los errores tengan una distribución t -student y adopta un enfoque Bayesiano para la estimación de los parámetros suponiendo que se conocen los parámetros de la variable umbral y el valor umbral. Como resultado de su estudio de simulación obtiene buenos resultados en la estimación de los parámetros del modelo bajo el supuesto que los errores del modelo siguen una distribución t -student de 5 grados de libertad, sin embargo no realizan comparaciones suponiendo errores normales ni pronósticos.

En el presente trabajo abordaremos el modelo TAR en el cual la distribución de sus errores pertenece a la familia elíptica y como caso especial nos enfocaremos en la distribución t -student. El modificar la distribución de los errores de una distribución normal a una distribución t nos podría ayudar para describir mejor datos que presentan colas pesadas. Sin embargo, la función de verosimilitud se vuelve complicada de tratar analíticamente y por lo que se tendrá que recurrir a métodos de optimización para encontrar los estimadores máximos verosímiles.

A lo largo de este trabajo presentaremos la estimación de los parámetros del modelo TAR con errores t mediante el método de máxima verosimilitud, el ajuste del modelo y los pronósticos a h horizontes, suponiendo adicionalmente que la variable umbral sigue un modelo autorregresivo de orden 1.

De forma más concreta, la estructura de la tesis es la siguiente: en el capítulo 2 daremos la definición del modelo TAR y nos centraremos en la estimación de parámetros para el caso TAR(2; 1, 1) por medio de mínimos cuadrados y máxima verosimilitud. En el capítulo 3 se presentará la definición, propiedades y ejemplos de las distribuciones elípticas. Asimismo mostraremos las técnicas usadas para la estimación de parámetros de la distribución t -student y el modelo de regresión con errores con distribuciones esféricas. Posteriormente, en el capítulo 4 calcularemos la función de verosimilitud para el caso del modelo TAR con r regímenes y errores independientes t y se estimarán los parámetros mediante el método de máxima verosimilitud. Adicionalmen-

te, se harán algunas simulaciones y se presentará una comparación con los resultados de los parámetros obtenidos al suponer que el modelo del cual provienen los datos es TAR con errores normales. En el capítulo 5 abordaremos el problema de la identificación del modelo adecuado para un conjunto de datos, presentaremos una metodología para resolverlo que combina las ideas de selección del valor umbral dadas por Qian(1998) con los criterios de información, como el de Akaike y Bayesiano. Se darán algunos ejemplos simulados para su ajuste y se compararán los modelos seleccionados usando el criterio de Akaike y Bayesiano. Por último, en el capítulo 6 haremos los pronósticos utilizando la técnica de verosimilitud predictiva perfil, daremos expresiones para los errores cuadráticos medios de pronósticos y se hará un estudio de simulación donde contrastaremos los ECMP obtenidos si suponemos errores normales y errores t en el modelo.

Capítulo 2

Modelo TAR

Los modelos autorregresivos umbrales (o simplemente TAR por sus siglas en inglés) fueron publicados por primera vez en el artículo de Tong y Lim(1980), "Threshold Autoregression, Limit Cycles and Cyclical data" y surgen de la idea de que si una serie no es lineal en el tiempo, entonces podemos hacer que tenga comportamiento lineal por pedazos donde la dinámica de la serie estará dada por alguna otra variable llamada umbral y esta a su vez puede ser representada por retrasos de la misma serie o alguna otra serie exógena. El comportamiento anterior se puede ver, por ejemplo, en las poblaciones de animales donde hay ciertas fases de contracción y expansión o en economía cuando la inflación supera cierto valor crítico, lo cual afecta al comportamiento de las tasas de interés.

Estos modelos han tenido gran aceptación debido a su fácil interpretación y además de que son capaces de considerar diferentes comportamientos de la serie en cada régimen como función de otra variable llamada umbral (comúnmente denotada por Z_t), donde cada comportamiento se activa cuando Z_t cruza un valor crítico denominado umbral.

Formalmente, un modelo TAR se define como sigue

Definición 1. Sea $Y = (Y_1, Y_2, \dots, Y_T)'$ el vector de datos observados con condiciones iniciales $(Y_0, Y_{-1}, \dots, Y_{-p+1})'$. Decimos que la serie de tiempo Y_t sigue un modelo TAR($r; p_1, \dots, p_r$) con variable umbral Z_{t-d} y r regímenes si Y_t puede expresarse de la siguiente manera

$$Y_t = \beta_{0,k} + \sum_{i=1}^{p_k} \beta_{i,k} Y_{t-i} + \epsilon_{k,t} \text{ si } \gamma_{k-1} \leq Z_{t-d} < \gamma_k, \quad (2.0.1)$$

donde $k = 1, 2, \dots, r$. Los números reales γ_k satisfacen $-\infty = \gamma_0 < \gamma_1 \leq \dots < \gamma_r = \infty$ y forman una partición del espacio de estado de Z_{t-d} . El intervalo $(\gamma_{k-1}, \gamma_k]$ hace referencia al k -ésimo régimen del modelo. Los errores $\{\epsilon_{k,t}\}$ forman una sucesión de variables aleatorias independientes con cierta distribución D , con media cero y varianza σ^2 , además $\{\epsilon_{i,t}\}$ es independiente de $\{\epsilon_{j,t}\}$ para $i \neq j$. El número d es un entero positivo y se conoce como retardo de la variable umbral Z_t .

A lo largo de este trabajo vamos a suponer que el número de retardos d es conocido.

2.1. Estimación de parámetros del modelo TAR(2; 1, 1)

La estimación de los parámetros en los modelos TAR es uno de los problemas que ha sido estudiado en diversos artículos como en Hansen(1997). A continuación presentaremos las técnicas utilizadas en la estimación de los parámetros del modelo dado en la ecuación (2.0.1) suponiendo que $r = 2$, $d = 1$, es decir, un modelo TAR(2; 1, 1) el cual se puede escribir como

$$Y_t = \begin{cases} \beta_1 Y_{t-1} + \epsilon_t & \text{si } Z_{t-1} \leq \gamma \\ \beta_2 Y_{t-1} + \epsilon_t & \text{si } Z_{t-1} > \gamma \end{cases}, \quad \epsilon_t \sim iidN(0, \sigma^2). \quad (2.1.1)$$

Para la estimación usaremos dos métodos diferentes: mínimos cuadrados y máxima verosimilitud.

2.1.1. Mínimos Cuadrados

Para la estimación de los parámetros del modelo TAR(2;1,1) vamos a suponer que el proceso es estacionario y ergódico, además, asumiremos que los errores $\{\epsilon_t\}$ y la variable umbral Z_{t-d} son independientes. Sea $\Theta = (\beta_1, \beta_2, \sigma^2)'$ el vector de parámetros a estimar y consideremos fijo el parámetro umbral γ dado en la ecuación (2.1.1). Entonces, el estimador de mínimos cuadrados para Θ condicional a γ fijo, está dado por aquellos valores que minimizan la siguiente suma de cuadrados:

$$S = \sum_{t=1}^T (Y_t - E(Y_t | \mathcal{F}_{t-1}, \Theta))^2,$$

donde $\mathcal{F}_{t-1} = \sigma(Y_s | s \leq t-1)$, que se interpreta como toda la información pasada al tiempo T del proceso Y_t . Notemos que

$$\begin{aligned} E(Y_t | \mathcal{F}_{t-1}, \Theta) &= E([\beta_1 I(Z_{t-1} \leq \gamma) + \beta_2 I(Z_{t-1} > \gamma)] Y_{t-1} + \epsilon_t | \mathcal{F}_{t-1}, \Theta), \\ &= [\beta_1 I(Z_{t-1} \leq \gamma) + \beta_2 I(Z_{t-1} > \gamma)] Y_{t-1}. \end{aligned}$$

El estimador para β_1 se encuentra al resolver $\frac{\partial S}{\partial \beta_1} = 0$,

$$\begin{aligned}
\frac{\partial S}{\partial \beta_1} &= \frac{\partial \left(\sum_{t=1}^T (Y_t - E(Y_t | \mathcal{F}_{t-1}, \Theta))^2 \right)}{\partial \beta_1} = 0, \\
0 &= \frac{\partial \left(\sum_{t=1}^T (Y_t^2 - 2Y_t E(Y_t | \mathcal{F}_{t-1}, \Theta) + E(Y_t | \mathcal{F}_{t-1}, \Theta)^2) \right)}{\partial \beta_1}, \\
0 &= \sum_{t=1}^T (-2Y_t I(Z_{t-1} \leq \gamma) Y_{t-1} + 2\beta_1 I(Z_{t-1} \leq \gamma) Y_{t-1}^2),
\end{aligned}$$

por lo que

$$\hat{\beta}_1(\gamma) = \frac{\sum_{t=1}^T Y_t I(Z_{t-1} \leq \gamma) Y_{t-1}}{I(Z_{t-1} \leq \gamma) Y_{t-1}^2}. \quad (2.1.2)$$

Análogamente, se llega a que el estimador para β_2 es

$$\hat{\beta}_2(\gamma) = \frac{\sum_{t=1}^T Y_t I(Z_{t-1} \geq \gamma) Y_{t-1}}{I(Z_{t-1} \geq \gamma) Y_{t-1}^2}. \quad (2.1.3)$$

Mientras que el estimador para σ^2 se obtiene de la siguiente expresión

$$\hat{\sigma}^2(\gamma) = \frac{1}{T} \sum_{i=1}^T \hat{e}^2, \quad (2.1.4)$$

en la cual $\hat{e} = Y_t - E(Y_t | \mathcal{F}_{t-1}; \hat{\beta}_1(\gamma), \hat{\beta}_2(\gamma))$.

Petrucelli y Woolford (1984) demuestran que los estimadores obtenidos en (2.1.2), (2.1.3), (2.1.4) son consistentes, repectivamente, para β_1 , β_2 , σ^2 suponiendo γ fijo y $Z_{t-1} = Y_{t-1}$.

2.1.2. Máxima Verosimilitud

Notemos que el modelo TAR(2; 1, 1) dado en (2.1.1) se puede reescribir como

$$\begin{aligned} Y_t &= [\beta_1 I(Z_{t-1} \leq \gamma) + \beta_2 I(Z_{t-1} > \gamma)] Y_{t-1} + \epsilon_t \\ &= h(Y_{t-1}, Z_{t-1}; \beta_1, \beta_2, \sigma^2, \gamma, \rho, \sigma_u^2) + \epsilon_t, \end{aligned} \quad (2.1.5)$$

donde $h(\cdot) = [\beta_1 I(Z_{t-1} \leq \gamma) + \beta_2 I(Z_{t-1} > \gamma)] Y_{t-1}$. Supongamos que el proceso Z_{t-1} es un AR(1) estacionario, es decir,

$$Z_{t-1} = \rho Z_{t-2} + \psi_t, \quad \psi_t \sim iidN(0, \sigma_u^2), \quad (2.1.6)$$

y asumimos que ϵ_t es independiente de Z_{t-1} .

Dadas las observaciones Y_1, Y_2, \dots, Y_T y Z_0, Z_1, \dots, Z_{T-1} de los procesos (2.1.5) y (2.1.6), se tiene que la función de verosimilitud está dada de la siguiente manera

$$\begin{aligned} L(\beta_1, \beta_2, \sigma^2, \gamma, \rho, \sigma_u^2; Y_t, Z_{t-1}) &= \prod_{i=2}^T f(Y_t | Y_{t-1}, Z_{t-1}; \beta_1, \beta_2, \gamma, \sigma^2) \\ &\quad \times g(Z_{t-1} | Z_{t-2}; \rho, \sigma_u^2), \\ (Z_{t-1} | Z_{t-2}; \rho, \sigma_u^2) &= \frac{1}{\sigma^{T-1}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{t=2}^T (Y_t - \delta_t Y_{t-1})^2 \right\} \\ &\quad \times \frac{1}{\sigma_u^{T-1}} \exp \left\{ -\frac{1}{2\sigma_u^2} \sum_{t=2}^T (Z_{t-1} - \rho Z_{t-2})^2 \right\}. \end{aligned}$$

Con $\delta_t = [\beta_1 I(Z_{t-1} \leq \gamma) + \beta_2 I(Z_{t-1} > \gamma)]$

Mientras que la función de log versomilitud como

$$\begin{aligned} \ell(\beta_1, \beta_2, \sigma^2, \gamma, \rho, \sigma_u^2; Y_t, Z_{t-1}) &= (T-1)\log(\sigma) - \frac{1}{2\sigma^2} \sum_{t=2}^T (Y_t - \delta_t Y_{t-1})^2 \\ &\quad + (T-1)\log(\sigma_u) - \frac{1}{2\sigma_u^2} \sum_{t=2}^T (Z_{t-1} - \rho Z_{t-2})^2. \end{aligned}$$

Derivando ℓ con respecto a σ^2 e igualando a cero se obtiene que,

$$\hat{\sigma}^2(\beta_1, \beta_2, \phi_2, \gamma) = \frac{1}{T-1} \sum_{t=2}^T (Y_t - [\beta_1 I(Z_{t-1} \leq \gamma) + \beta_2 I(Z_{t-1} > \gamma)] Y_{t-1})^2. \quad (2.1.7)$$

Luego, derivando ℓ con respecto a β_1, β_2 e igualando a cero encontramos que sus estimadores máximos verosímiles son

$$\hat{\beta}_1(\gamma) = \frac{\sum_{t=1}^T Y_t I(Z_{t-1} \leq \gamma) Y_{t-1}}{I(Z_{t-1} \leq \gamma) Y_{t-1}^2}. \quad (2.1.8)$$

$$\hat{\beta}_2(\gamma) = \frac{\sum_{t=1}^T Y_t I(Z_{t-1} \geq \gamma) Y_{t-1}}{I(Z_{t-1} \geq \gamma) Y_{t-1}^2}. \quad (2.1.9)$$

De la misma manera, los estimadores máximos verosímiles para σ_u^2 y ρ son

$$\begin{aligned} \hat{\rho} &= \frac{\sum_{t=2}^T Z_{t-1} Z_{t-1}}{\sum_{t=2}^T Z_{t-2}^2}, \\ \hat{\sigma}_u^2(\hat{\rho}) &= \frac{1}{T-1} \sum_{t=2}^T (Z_{t-1} - \hat{\rho} Z_{t-2})^2. \end{aligned}$$

Sustituyendo (2.1.8) y (2.1.9) en (2.1.7) se llega a que el estimador máximo verosímil para σ^2 como función solo del parámetro umbral γ es

$$\hat{\sigma}^2(\gamma) = \frac{1}{T-1} \sum_{t=2}^T \left(Y_t - \left[\hat{\beta}_1 I(Z_{t-1} \leq \gamma) + \hat{\beta}_2 I(Z_{t-1} > \gamma) \right] Y_{t-1} \right)^2.$$

Notemos que no podemos derivar la verosimilitud con respecto a γ debido a que no es continua en los puntos Z_0, \dots, Z_T . Una solución al problema anterior está dada en Qian(1998) donde propone utilizar $\hat{\gamma}$ como el valor que cumpla

$$\inf_{\gamma} \arg \min \frac{T-1}{2} \log \left[\frac{1}{T-1} \sum_{t=2}^T \left(Y_t - \left[\hat{\beta}_1 I(Z_{t-1} \leq \gamma) + \hat{\beta}_2 I(Z_{t-1} > \gamma) \right] Y_{t-1} \right)^2 \right].$$

En Russel(2006) puede verse que los estimadores obtenidos para $\beta_1, \beta_2, \gamma, \sigma, \rho, \sigma_u$ usando el método de máxima verosimilitud son consistentes cuando el valor del umbral es conocido.

Capítulo 3

Distribuciones Elípticas

En este capítulo daremos una breve revisión a las distribuciones elípticas: su definición, algunas propiedades y ejemplos. Además, presentaremos las metodologías utilizadas en la estimación de parámetros para la distribución t -student univariada y para modelos de regresión lineal con distribuciones esféricas. Lo anterior nos dará ideas de posibles alternativas utilizadas ante la no normalidad, colas pesadas de los datos o presencia de datos atípicos, así como también un panorama sobre las técnicas y problemáticas que se ocasionan al cambiar de una distribución normal a otra, como por ejemplo una distribución t -student.

3.1. Definición

Las distribuciones elípticas se presentan como una extensión a la clase de distribuciones normales y las cuales nos permiten atacar problemas como la no normalidad o colas pesadas en los datos, pero sin embargo manteniéndonos dentro de una familia de distribuciones simétricas.

Fang(1990) define a las distribuciones elípticas usando la relación que existe con las distribuciones esféricas. Por lo tanto, comenzaremos dando la defini-

ción de las distribuciones esféricas.

Definición 2. Un vector aleatorio \mathbf{x} de dimensión $n \times 1$ se dice que tiene distribución esférica simétrica o simplemente distribución esférica si para todo $\mathbf{\Gamma} \in \mathcal{O}(n)$ se tiene que

$$\mathbf{\Gamma}\mathbf{x} \stackrel{d}{=} \mathbf{x},$$

donde $\mathcal{O}(n)$ denota el conjunto de matrices ortogonales $n \times n$.

Geométricamente la definición anterior nos dice que una distribuciones esférica es invariante bajo rotaciones.

El siguiente teorema dado en Fang(1990) nos ayuda a saber que distribuciones pertenecen a esta familia fijándonos en su función característica, $\psi(\mathbf{t})$.

Teorema 3. *Un vector n dimensional \mathbf{x} tiene distribución esférica si y sólo si su función característica $\psi(\mathbf{t})$ satisface una de las siguientes condiciones equivalentes*

(i) $\psi(\mathbf{\Gamma}'\mathbf{t}) = \psi(\mathbf{t})$ para cualquier $\mathbf{\Gamma} \in \mathcal{O}(n)$.

(ii) Existe una función $\phi(\cdot)$ de una variable escalar tal que $\psi(\mathbf{t}) = \phi(\mathbf{t}'\mathbf{t})$.

Escribiremos $\mathbf{x} \sim S_n(\phi)$ que significa que \mathbf{x} tiene distribución esférica con función característica de la forma $\phi(\mathbf{t}'\mathbf{t})$ donde $\phi(\cdot)$ es una función de variable escalar llamada el generador característico de la distribución. Con base a la definición anterior se caracteriza a las distribuciones elípticas de la siguiente manera,

Definición 4. Se dice que un vector aleatorio \mathbf{x} de dimensión $n \times 1$ tiene distribución elíptica con parámetros $\boldsymbol{\mu}_{n \times 1}$ y $\boldsymbol{\Sigma}_{n \times n}$ si

$$\mathbf{x} = \boldsymbol{\mu} + \mathbf{A}'\mathbf{y}, \mathbf{y} \sim S_k(\phi),$$

de donde $\mathbf{A}_{k \times n}$, $\mathbf{A}'\mathbf{A} = \boldsymbol{\Sigma}$ con $\text{rang}(\boldsymbol{\Sigma}) = k$. Escribiremos $\mathbf{x} \sim EC_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \phi)$.

3.2. Algunas propiedades

Dos propiedades que poseen las distribuciones elípticas semejantes a las distribuciones normales se enuncian en los teoremas (5) y (6).

Teorema 5. Si $\mathbf{x} \sim EC_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \phi)$ con $\text{rang}(\boldsymbol{\Sigma}) = k$, \mathbf{B} es una matriz de dimensión $n \times m$, y \mathbf{v} es un vector $n \times 1$ entonces

$$\mathbf{v} + \mathbf{B}'\mathbf{x} \sim EC(\mathbf{v} + \mathbf{B}'\boldsymbol{\mu}, \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}'; \phi).$$

Otra propiedad importante es que todas las distribuciones marginales de las distribuciones elípticas también pertenecen a esta familia. Más formalmente tenemos el siguiente teorema.

Teorema 6. Si $\mathbf{x} \sim EC_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \phi)$ y particionado como

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix}, \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} k \times 1 \\ (n-k) \times 1 \end{bmatrix}$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}, \begin{bmatrix} k \times k & k \times (n-k) \\ (n-k) \times k & (n-k) \times (n-k) \end{bmatrix},$$

entonces $\mathbf{x}_1 \sim EC(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}; \phi)$ y $\mathbf{x}_2 \sim EC(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22}; \phi)$.

3.3. Ejemplos

En lo siguiente presentaremos dos ejemplos de distribuciones conocidas que pertenecen a la familia de distribuciones elípticas.

3.3.1. Distribución Normal

La distribución normal multivariada es una de las distribuciones que pertenecen a las distribuciones elípticas. Para comprobarlo primero demostraremos que la distribución normal estándar multivariada es parte de la familia de distribuciones esféricas y posteriormente utilizando resultados de combinaciones lineales de distribuciones normales se obtiene lo deseado.

Consideremos $\mathbf{x} = (x_1, x_2, \dots, x_n) \sim N_n(\mathbf{0}, \mathbf{I}_n)$. Como la función característica de x_1 es $\exp(-\frac{t^2}{2})$, entonces la función característica de \mathbf{x} es

$$\exp(-\frac{1}{2}(t_1^2 + t_2^2 + \dots + t_n^2)) = \exp(-\frac{1}{2}\mathbf{t}'\mathbf{t}),$$

del Teorema 3 se llega a que \mathbf{x} tiene distribución esférica, $S_n(\phi)$, con generador característico $\phi(u) = \exp(-u/2)$.

Recordemos que si \mathbf{x} tiene distribución normal multivariada $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ entonces se puede descomponer de la siguiente manera:

$$\mathbf{x} \stackrel{d}{=} \boldsymbol{\mu} + \mathbf{A}'\mathbf{y}$$

donde $\boldsymbol{\mu} \in \mathbb{R}^n$, $\mathbf{y} \sim N_m(\mathbf{0}, \mathbf{I}_m)$ y $\mathbf{A}_{m \times n}$ con $\boldsymbol{\Sigma} = \mathbf{A}'\mathbf{A}$. Como se tiene que $\mathbf{y} \sim S_m(\phi)$ con $\phi(u) = \exp(-\frac{u}{2})$ entonces $\mathbf{x} \sim EC_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$.

3.3.2. Distribución t -student

Otro caso especial de distribuciones elípticas son las distribuciones t -student, que describiremos en breve en su caso multivariado estándar.

Consideremos $\mathbf{z} \sim N_n(\mathbf{0}, \mathbf{I}_n)$, $\mathbf{s} \sim \chi_\nu^2$ independientes y sea

$$\mathbf{y} = \nu^{\frac{1}{2}} \frac{\mathbf{z}}{\mathbf{s}},$$

entonces decimos que \mathbf{y} tiene distribución t -multivariada estándar con ν grados de libertad y escribimos $Mt_n(\nu, \mathbf{0}, \mathbf{I}_n)$.

Ahora sea

$$\mathbf{x} = \boldsymbol{\mu} + \mathbf{A}'\mathbf{y},$$

donde $\boldsymbol{\mu} \in \mathbb{R}^n$ y \mathbf{A} es de dimensión $n \times n$. Entonces decimos que \mathbf{x} tiene distribución t -multivariada con parámetros $\boldsymbol{\mu}$, $\boldsymbol{\Sigma} = \mathbf{A}'\mathbf{A}$, ν grados de libertad y lo representamos por $\mathbf{x} \sim Mt_n(\nu, \boldsymbol{\mu}, \boldsymbol{\Sigma})$.

La función de densidad de una variable $\mathbf{x} \sim Mt_n(\nu, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ es la siguiente,

$$f_{\mathbf{x}}(\mathbf{x}) = \frac{\Gamma\left(\frac{\nu+n}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right) \nu^{n/2} \pi^{n/2} |\boldsymbol{\Sigma}|^{n/2}} \left(1 + \frac{1}{\nu} (\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right)^{-\left(\frac{\nu+n}{2}\right)}, \quad \mathbf{x} = (x_1, x_2, \dots, x_n).$$

Notemos que al tomar $n = 1$ obtenemos el caso de la función de distribución de una variable t -student univariada. Además, particionando \mathbf{x} como en el Teorema 6, se puede ver en Kotz(2004) que $\mathbf{x}_1 \sim Mt_k(\nu, \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$ y $\mathbf{x}_2 \sim Mt_{n-k}(\nu, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$. Similar al caso normal tenemos que si consideramos transformaciones lineales como en el caso del teorema 5, se tiene que $\mathbf{v} + \mathbf{B}\mathbf{x} \sim Mt_n(\nu, \mathbf{v} + \mathbf{B}'\boldsymbol{\mu}, \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}')$. Sin embargo, tenemos que el producto de dos o más distribuciones t -student univariadas independientes e idénticamente distribuidas generalmente no da como resultado una distribución t multivariada y para ello basta con ver que el producto de las funciones de densidad de las variables no se factoriza en una función de densidad t multivariada, véase Ahsanullah(2014) para una expresión de la función de distribución de dos variables t independientes. Como a lo largo de este trabajo estaremos trabajando con errores t independientes e idénticamente distribuidos y como su producto no nos genera el caso multivariado, entonces tiene

sentido enfocarnos de aquí en adelante solo al caso univariado.

Para darnos una idea de los problemas que se pueden presentar en la estimación de los parámetros del modelo TAR suponiendo errores t , se revisaron las técnicas usadas en la estimación de los parámetros vía máxima verosimilitud de los parámetros de la distribución t dadas en Liu(1995) y en la estimación de los parámetros de un modelo regresión lineal con errores on distribución simétrica dadas en Cysneiros(2005). Liu propone utilizar un algoritmo EM mientras que Cysneiros sugiere emplear el método Scoring de Fisher.

3.4. Estimación de parámetros de la distribución t -student por máxima verosimilitud

Esta sección presentaremos la estimación de los parámetros de una distribución t -student por medio del método máxima verosimilitud.

Recordemos que la función de densidad de la distribución t -student univariada con parámetro de localización μ , parámetro de escala λ , y ν grados de libertad está dada por

$$f(x|\mu, \lambda, \nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \left(\frac{\lambda}{\pi\nu}\right)^{\frac{1}{2}} \left(1 + \frac{\lambda(x-\mu)^2}{\nu}\right)^{-\left(\frac{\nu+1}{2}\right)}. \quad (3.4.1)$$

Denotemos por $\Theta = (\mu, \lambda, \nu)$ al vector de parámetros. La función de verosimilitud dadas T observaciones iid de una distribución t es

$$L(\Theta; \mathbf{X} = (x_1, x_2, \dots, x_T)) = p(\Theta|\mathbf{X}) = \prod_{i=1}^T f(x_i|\Theta). \quad (3.4.2)$$

Encontrar el estimador máximo verosímil para Θ suele resultar muy complicado por la forma que presenta la verosimilitud. Debido a lo anterior se han escrito algunos artículos, como por ejemplo Aeschliman(2010) o Liu(1995), donde se presentan diferentes métodos para la estimación de los parámetros para Θ . A continuación presentamos la idea dada en Scheffer(2008) donde propone utilizar el hecho de que la distribución t se puede ver como una mezcla de distribuciones para posteriormente aplicar el algoritmo EM.

Notemos que (3.4.1) se puede escribir como:

$$f(x|\mu, \lambda, \nu) = \int_0^\infty g(x|\mu, (\lambda\eta)^{-1}) h\left(\eta|\frac{\nu}{2}, \frac{\nu}{2}\right) d\eta,$$

$$g(x|\mu, (\lambda\eta)^{-1}) = \frac{1}{\sqrt{2\pi(\lambda\eta)^{-1}}} \exp\left(-\frac{1}{2(\lambda\eta)^{-1}}(x-\mu)^2\right),$$

$$h\left(\eta|\frac{\nu}{2}, \frac{\nu}{2}\right) = \frac{1}{\Gamma(\frac{\nu}{2})} \left(\frac{\nu}{2}\right)^{\frac{\nu}{2}} \eta^{\frac{\nu}{2}-1} e^{-\frac{\nu\eta}{2}}.$$

Es decir, la distribución t se puede ver como mezcla infinita de distribuciones *Normal* $(\mu, (\lambda\eta)^{-1})$ con distribución mezclante o latente *Gamma* $(\frac{\nu}{2}, \frac{\nu}{2})$. De lo anterior, tenemos que la verosimilitud de los datos completos está dada por la siguiente expresión

$$L(\Theta; \mathbf{X}, \mathbf{Z}) = p(\mathbf{X}, \mathbf{Z} | \Theta) = \prod_{i=1}^T g(x_i|\mu, (\lambda\eta)^{-1}) h\left(\eta_i|\frac{\nu}{2}, \frac{\nu}{2}\right), \quad (3.4.3)$$

donde $\mathbf{Z} = (\eta_1, \dots, \eta_T)$, las η_i son variables latentes y $p(\Theta|\mathbf{X}) = \int_{\mathbf{Z}} p(\mathbf{X}, \mathbf{Z} | \Theta)$.

Recordemos que el algoritmo EM busca encontrar el EMV para Θ aplicando iterativamente los siguientes dos pasos hasta que se cumpla un criterio de convergencia:

- **Paso E.** Calcular el valor esperado de la función de log verosimilitud con respecto a la distribución condicional de \mathbf{Z} dado \mathbf{X} bajo el estimador de parámetros $\Theta^{(t)}$

$$Q\left(\Theta \mid \Theta^{(t)}\right) = E_{\mathbf{Z} \mid \mathbf{X}, \Theta^{(t)}}(\log L(\Theta; \mathbf{X}, \mathbf{Z})).$$

- **Paso M.** Encontrar los parámetros que maximizan $Q\left(\Theta \mid \Theta^{(t)}\right)$, estos es

$$\Theta^{(t+1)} = \underset{\Theta}{\operatorname{argmáx}} Q\left(\Theta \mid \Theta^{(t)}\right).$$

Paso E.

Recordemos que la verosimilitud de los datos completos está dada por

$$L(\Theta; \mathbf{X}, \mathbf{Z}) = p(\mathbf{X}, \mathbf{Z} \mid \Theta) = \prod_{i=1}^T g(x_i \mid \mu, (\lambda\eta)^{-1}) h\left(\eta_i \mid \frac{\nu}{2}, \frac{\nu}{2}\right).$$

Y la log verosimilitud queda de la siguiente manera

$$\begin{aligned} \log p(\mathbf{X}, \mathbf{Z} \mid \Theta) &= \sum_{i=1}^T \left[\log(g(x_i \mid \mu, (\lambda\eta)^{-1})) + \log\left(h\left(\eta_i \mid \frac{\nu}{2}, \frac{\nu}{2}\right)\right) \right], \quad (3.4.4) \\ &= \sum_{i=1}^T \left[-\frac{1}{2} \log(2\pi) + \frac{1}{2} \log(\lambda) + \frac{1}{2} \log(\eta_i) - \frac{\lambda\eta_i}{2} (x_i - \mu)^2 \right] + \\ &\quad \sum_{i=1}^T \left[+\frac{\nu}{2} \log\left(\frac{\nu}{2}\right) + \left(\frac{\nu}{2} - 1\right) \log(\eta_i) - \frac{\nu\eta_i}{2} - \log\left(\Gamma\left(\frac{\nu}{2}\right)\right) \right]. \end{aligned}$$

Luego

$$\begin{aligned} p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\Theta}) &\propto p(\mathbf{Z}, \mathbf{X}|\boldsymbol{\Theta}) \\ &= \prod_{i=1}^T g(x_i|\mu, (\lambda\eta)^{-1}) h\left(\eta_i|\frac{\nu}{2}, \frac{\nu}{2}\right). \end{aligned}$$

Como la distribución Gamma es una distribución apriori conjugada de una distribución Normal con parámetro de precisión desconocido, entonces

$$p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\Theta}) \propto \text{Gamma}\left(\eta_i|\frac{\nu+1}{2}, \left[\frac{\nu}{2} + \frac{\lambda}{2}(x_i - \mu)^2\right]\right).$$

Ahora calcularemos $E_{\mathbf{Z}|\mathbf{X}, \boldsymbol{\Theta}^{(t)}}(\log L(\boldsymbol{\Theta}; \mathbf{X}, \mathbf{Z}))$. De (3.4.4) tenemos que solo es necesario calcular $E_{\mathbf{Z}|\mathbf{X}, \boldsymbol{\Theta}}(\eta_i)$ y $E_{\mathbf{Z}|\mathbf{X}, \boldsymbol{\Theta}}(\log(\eta_i))$.

Para simplificar la notación usaremos

$$E_{\mathbf{Z}}(\eta_i) := E_{\mathbf{Z}|\mathbf{X}, \boldsymbol{\Theta}}(\eta_i), \quad E_{\mathbf{Z}}(\log(\eta_i)) := E_{\mathbf{Z}|\mathbf{X}, \boldsymbol{\Theta}}(\log(\eta_i)).$$

Como $\eta_i|[\frac{\nu+1}{2}, [\frac{\nu}{2} + \frac{\lambda}{2}(x_i - \mu)^2]] \sim \text{Gamma}(\frac{\nu+1}{2}, [\frac{\nu}{2} + \frac{\lambda}{2}(x_i - \mu)^2])$ entonces

$$E_{\mathbf{Z}}(\eta_i) = \frac{\frac{\nu+1}{2}}{[\frac{\nu}{2} + \frac{\lambda}{2}(x_i - \mu)^2]}, \quad (3.4.5)$$

y

$$E_{\mathbf{Z}}(\log(\eta_i)) = \psi\left(\frac{\nu+1}{2}\right) - \log\left(\frac{\nu}{2} + \frac{\lambda}{2}(x_i - \mu)^2\right), \quad (3.4.6)$$

donde

$$\psi\left(\frac{\nu+1}{2}\right) = \frac{d \ln \Gamma(x)}{dx} = \frac{\Gamma'(x)}{\Gamma(x)}.$$

Así

$$\begin{aligned}
E_{\mathbf{Z}}(\log p(\mathbf{X}, \mathbf{Z} | \boldsymbol{\Theta})) &= \sum_{i=1}^T \left[-\frac{1}{2} \log(2\pi) + \frac{1}{2} \log(\lambda) + \frac{1}{2} E_{\mathbf{Z}}(\log(\eta_i)) - \frac{\lambda E_{\mathbf{Z}}(\eta_i)}{2} (x_i - \mu)^2 \right] \\
&+ \sum_{i=1}^T \left[+\frac{\nu}{2} \log\left(\frac{\nu}{2}\right) + \left(\frac{\nu}{2} - 1\right) E_{\mathbf{Z}}(\log(\eta_i)) \right] \\
&+ \sum_{i=1}^T \left[-\frac{\nu E_{\mathbf{Z}}(\eta_i)}{2} - \log\left(\Gamma\left(\frac{\nu}{2}\right)\right) \right].
\end{aligned}$$

Las expresiones para $E_{\mathbf{Z}}(\log(\eta_i))$ y $E_{\mathbf{Z}}(\eta_i)$ están dadas por (3.4.5) y (3.4.6) respectivamente.

Paso M.

Caculando $\frac{\partial Q}{\partial \mu}$, $\frac{\partial Q}{\partial \lambda}$, $\frac{\partial Q}{\partial \nu}$.

$$\begin{aligned}
\frac{\partial Q}{\partial \mu} &= \lambda E_{\mathbf{Z}}(\eta_i) \sum_{i=1}^T (x_i - \mu) = 0, \\
\Rightarrow \hat{\mu} &= \frac{\sum_{i=1}^T x_i E_{\mathbf{Z}}(\eta_i)}{\sum_{i=1}^T E_{\mathbf{Z}}(\eta_i)}.
\end{aligned}$$

$$\begin{aligned}
\frac{\partial Q}{\partial \lambda} &= \frac{N}{2\lambda} - \sum_{i=1}^T \frac{E_{\mathbf{Z}}(\eta_i)}{2} (x_i - \mu)^2 = 0, \\
\Rightarrow \hat{\lambda} &= \frac{N}{\sum_{i=1}^T \frac{E_{\mathbf{Z}}(\eta_i)}{2} (x_i - \hat{\mu})^2},
\end{aligned}$$

Luego,

$$\frac{\partial Q}{\partial \nu} = \frac{1}{2} \sum_{i=1}^T (E_{\mathbf{Z}}(\log(\eta_i)) - E_{\mathbf{Z}}(\eta_i)) - \frac{N}{2} \psi\left(\Gamma\left(\frac{\nu}{2}\right)\right) + \frac{N}{2} + \frac{N}{2} \log\left(\frac{\nu}{2}\right),$$

Igualando a cero obtenemos

$$\begin{aligned} \frac{N}{2}\psi\left(\Gamma\left(\frac{\nu}{2}\right)\right) - \frac{N}{2} - \frac{N}{2}\log\left(\frac{\nu}{2}\right) &= \frac{1}{2}\sum_{i=1}^T (E_{\mathbf{Z}}(\log(\eta_i)) - E_{\mathbf{Z}}(\eta_i)), \\ \psi\left(\Gamma\left(\frac{\nu}{2}\right)\right) - \log\left(\frac{\nu}{2}\right) - 1 &= \frac{1}{N}\sum_{i=1}^T (E_{\mathbf{Z}}(\log(\eta_i)) - E_{\mathbf{Z}}(\eta_i)) \end{aligned} \quad (3.4.7)$$

Debido a la forma que presenta la ecuación (3.4.7) es necesario encontrar el estimador de ν numéricamente. Como complemento, Scheffer(2008) simula datos que contienen observaciones atípicas y realiza una comparación entre la distribución t -student y la distribución normal obtenidas de estimar los parámetros mediante máxima verosimilitud. Se observa que el estimador de la varianza en la distribución normal aumenta conforme se incrementa el número de datos atípicos, a diferencia, el estimador de la varianza de la distribución t se mantiene casi constante conforme se aumenta el número de este tipo de datos, en este sentido el modelo t es más robusto que el normal.

3.5. Regresión con distribuciones esféricas

Es común que en los modelos de regresión lineal se asuma que la distribución de los errores, ϵ_i , siga una distribución normal. Sin embargo, puede ocurrir que el supuesto de normalidad, que nos da propiedades deseables y cálculos menos complejos al estimar los parámetros, no se cumpla. Por lo que cambiar la distribución de los errores puede ser una solución al problema.

En esta sección abordaremos el caso en el que la distribución asociada a los errores en un modelo de regresión pertenece a la familia de distribuciones simétricas. Como bien podemos sospechar, cambiar de distribución a una que pertenezca a la familia de distribuciones simétricas puede provocar que la función de verosimilitud cambie y en consecuencia que los cálculos de estimación de los parámetros se compliquen. Por lo tanto, esto hace necesario explorar diferentes métodos para su estimación.

Las ideas presentadas en esta sección son de interés puesto que algo análogo podríamos proponer en los modelos TAR como alternativa de solución a problemas de datos que presenten colas pesadas o problemas de no normalidad, es decir cambiar la distribución normal a una familia de distribuciones más amplia como las distribuciones elípticas estudiadas en este capítulo.

Consideremos el siguiente modelo de regresión

$$y_i = \mathbf{x}_i' \boldsymbol{\beta} + \epsilon_i, \quad i = 1, \dots, n \quad (3.5.1)$$

donde y_i es nuestra variable de interés, $\mathbf{x}_i = (x_1, x_2, \dots, x_p)'$ es el vector de las p variables explicativas y $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)'$ el vector de parámetros. Además, suponemos que ϵ_i tiene distribución que pertenece a la familia de distribuciones elípticas simétricas $\epsilon_i \sim S(0; g, \phi)$, por lo que su función de densidad tiene la siguiente forma

$$f_{\epsilon_i}(x) = \frac{g\left(\frac{x^2}{\phi}\right)}{\sqrt{\phi}} \quad .$$

Denotemos por $\boldsymbol{\Theta} = (\boldsymbol{\beta}', \phi)'$ al vector de parámetros. Usando la notación anterior y del hecho de que $\epsilon_i \sim S(0; g, \phi)$ se llega a que la función de verosimilitud, basa en n observaciones del modelo (3.5.1) se escribe como sigue

$$\begin{aligned}
L(\Theta) &= \prod_{i=1}^n f_{\epsilon_i} \left((y_i - \mathbf{x}'_i \boldsymbol{\beta})^2 \right) \\
&= \prod_{i=1}^n \frac{1}{\sqrt{\phi}} g \left(\frac{(y_i - \mathbf{x}'_i \boldsymbol{\beta})^2}{\phi} \right) \\
&= \phi^{-\frac{n}{2}} \prod_{i=1}^n g \left(\frac{(y_i - \mathbf{x}'_i \boldsymbol{\beta})^2}{\phi} \right). \tag{3.5.2}
\end{aligned}$$

Tomando el logaritmo de (3.5.2) obtenemos la función de log-verosimilitud,

$$\ell(\Theta) = \left(-\frac{n}{2} \right) \log(\phi) + \sum_{i=1}^n \log(g(u_i)), \quad u_i = \frac{(y_i - \mathbf{x}'_i \boldsymbol{\beta})^2}{\phi}.$$

Luego, las primeras derivadas parciales respecto a $\boldsymbol{\beta}$ y a ϕ denotadas por $\mathbf{U}_{\boldsymbol{\beta}}(\Theta)$, $\mathbf{U}_{\phi}(\Theta)$ son

$$\begin{aligned}
\mathbf{U}_{\boldsymbol{\beta}}(\Theta) &= \frac{1}{\phi} \mathbf{X}' \mathbf{V} (\mathbf{y} - \boldsymbol{\mu}) \\
\mathbf{U}_{\phi}(\Theta) &= (2\phi)^{-1} \{ (Q_v(\boldsymbol{\beta}, \phi) / \phi) - n \}
\end{aligned}$$

$$\begin{aligned}
\mathbf{X} &= (x_1, x_2, \dots, x_n)', \quad \mathbf{y} = (y_1, y_2, \dots, y_n)', \quad \boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_n)' \\
\mathbf{V} &= \text{diag} \{ \nu_1, \nu_2, \dots, \nu_n \}, \quad \nu_i = -2W_g(u_i), \quad W_g(u) = \frac{\frac{dg(u)}{du}}{g(u)}
\end{aligned}$$

$$Q_v(\boldsymbol{\beta}, \phi) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{V} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

La matriz que contiene las segundas derivadas de $\ell(\Theta)$, que denotaremos por I , se expresa como

$$I = \begin{pmatrix} \frac{\partial \ell}{\partial \beta \partial \beta} & \frac{\partial \ell}{\partial \beta \partial \phi} \\ \frac{\partial \ell}{\partial \phi \partial \beta} & \frac{\partial \ell}{\partial \phi \partial \phi} \end{pmatrix}$$

$$\frac{\partial \ell}{\partial \beta \partial \beta} = -\frac{1}{\phi} \mathbf{X}' \mathbf{D}_1 \mathbf{X}$$

$$\frac{\partial \ell}{\partial \beta \partial \phi} = 2\phi^{-1} \mathbf{X}' \mathbf{b}$$

$$\frac{\partial \ell}{\partial \phi \partial \phi} = \phi^{-2} \left(\frac{n}{2} + \mathbf{u}' \mathbf{D}_2 \mathbf{u} - \phi^{-1} Q_v(\beta) \right)$$

con

$$\mathbf{D}_1 = \text{diag}(a_1, a_2, \dots, a_n), \quad a_i = -2 \left(W_g(u_i) + 2u_i \frac{\partial}{\partial u_i} W_g(u_i) \right)$$

$$\mathbf{b} = (b_1, b_2, \dots, b_n)', \quad b_i = \left(W_g(u_i) + u_i \frac{\partial}{\partial u_i} W_g(u_i) \right) \epsilon_i$$

$$\mathbf{D}_2 = \text{diag}(c_1, c_2, \dots, c_n), \quad c_i = \frac{\partial}{\partial u_i} W_g(u_i)$$

$$\mathbf{u} = (u_1, u_2, \dots, u_n)'$$

Paula y Cysneiros(2005) desarrollan los cálculos para calcular la matriz de información de Fisher,

$$\begin{aligned} I(\boldsymbol{\Theta}) &= -E(I|\boldsymbol{\Theta}) \\ &= \begin{pmatrix} \mathbf{K}_{\beta\beta} & \mathbf{0} \\ \mathbf{0} & \mathbf{K}_{\phi\phi} \end{pmatrix} \end{aligned}$$

y obtienen que

$$\begin{aligned} \mathbf{K}_{\beta\beta} &= \frac{4d_g}{\phi} \mathbf{X}' \mathbf{X} \\ \mathbf{K}_{\phi\phi} &= \frac{n}{4\phi^2} (4f_g - 1), \end{aligned}$$

donde $d_g = E(W_g^2(Z^2)Z^2)$, $f_g = E(W_g^2(Z^2)Z^4)$ con $Z \sim S(0, 1)$. Además, calculan las expresiones para d_g y f_g cuando ϵ tiene una distribución t y se presentan en (3.5.3) y (3.5.4) respectivamente.

$$d_g = \frac{\nu + 1}{4(\nu + 3)} \quad (3.5.3)$$

$$f_g = \frac{3(\nu + 1)}{4(\nu + 3)} \quad (3.5.4)$$

Una vez obtenida la matriz de información esperada de Fisher se procede a la estimación de los parámetros mediante un método iterativo como Newton-Rapshon o Scoring de Fisher. El proceso iterativo para obtener $\hat{\Theta}$ se detalla a continuación y proponen que los estimadores en que cada paso estén dados como

$$\begin{aligned} \beta^{(m+1)} &= [\mathbf{X}'(\mathbf{V}^{(m)}) \mathbf{X}]^{-1} \mathbf{X}'(\mathbf{V}^{(m)}) \mathbf{y} \\ \phi^{(m+1)} &= \frac{1}{n} Q_v(\beta^{(m+1)}, \phi^{(m)}). \end{aligned}$$

Como resultado, Cysneiros (2005) obtiene que los modelos de regresión que tienen errores con distribución de colas pesadas como la t -student pueden reducir la influencia de datos extremos y por lo tanto son menos sensibles a datos atípicos que los modelos de regresión con errores cuya distribución es normal.

Capítulo 4

Estimación de parámetros en modelos TAR ($r : p_1, p_2, \dots, p_r$)

Recordemos que el modelo TAR($r; p_1, p_2, \dots, p_r$) con variable umbral Z_{t-1} se expresa de la siguiente manera:

$$Y_t = \begin{cases} \beta_{11}Y_{t-1} + \beta_{12}Y_{t-2} + \dots + \beta_{1p_1}Y_{t-p_1} + \epsilon_t & Z_{t-1} \leq \gamma_1 \\ \vdots & \vdots \\ \beta_{r1}Y_{t-1} + \beta_{r2}Y_{t-2} + \dots + \beta_{rp_r}Y_{t-p_r} + \epsilon_t & \gamma_{r-1} < Z_{t-1} \leq \gamma_r \end{cases}, \quad \epsilon_t \sim iidN(0, \sigma^2)$$

Notemos que la ecuación anterior se puede reescribir como sigue,

$$Y_t = I(Z_{t-1} \leq \gamma_1) \left(\sum_{i=1}^{p_1} \beta_{1i} Y_{t-i} \right) + \dots + I(\gamma_{r-1} < Z_{t-1} \leq \gamma_r) \left(\sum_{i=1}^{p_r} \beta_{ri} Y_{t-i} \right) + \epsilon_t. \quad (4.0.1)$$

Como hemos mencionado a lo largo de este trabajo, el asumir una distribución normal en (4.0.1) podría resultar restrictivo en situaciones donde los datos presenten comportamientos atípicos, con colas pesadas o no normalidad. Razón por la cual podemos usar las ideas presentadas en el capítulo

anterior, es decir, cambiar la distribución del proceso Y_t por una que pertenezca a la familia de distribuciones elípticas y como caso particular nos centraremos en la distribución t -student. Así, en el modelo (4.0.1) asumiremos que $\{\epsilon_t\} \sim iid t_\nu$.

Una de las partes esenciales al hacer un modelo estadístico es la estimación de parámetros. En el caso de los modelos SETAR con errores normales, Petrucci y Woolford(1984) hacen el cálculo vía mínimos cuadrados y además demuestran sus propiedades asintóticas. En modelos TAR especialmente en el caso en que se asumen errores con distribución t -student hay poco material desarrollado, uno de los artículos publicados fue elaborado por Zhang(2012) y aborda el problema de estimación de los parámetros del modelo TAR con r regímenes, $TAR(r; p_1, p_2, \dots, p_r)$, desde el punto de vista Bayesiano y considerando que los valores de los umbrales $\gamma_1, \gamma_2, \dots, \gamma_{r-1}$ son conocidos (basándose en el conocimiento previo que se tiene del fenómeno) y proponiendo distribuciones a priori para los parámetros de los órdenes autorregresivos y los grados de libertad de la distribución t . Como resultado de su estudio de simulación para un caso particular de un $TAR(2; 1, 1)$, se obtiene que la metodología desarrollada da buenos resultados al estimar los parámetros, sin embargo el tiempo computacional es demasiado. Como vimos en la sección 2, existen otros dos métodos utilizados para la estimación de parámetros en los modelos TAR: mínimos cuadrados y máxima verosimilitud.

En este capítulo nos enfocaremos en la estimación de los parámetros del modelo (4.0.1) vía máxima verosimilitud. Posteriormente, se presentarán algunos ejemplos con datos simulados donde se estimarán los parámetros utilizando las ideas presentadas en este capítulo y suponiendo que los datos provienen de un modelo TAR donde los errores ϵ_t tienen distribución t -student con ν grados de libertad. Como complemento al estudio de simulación, haremos el mismo procedimiento de estimación de los parámetros a excepción de que en el modelo TAR asumiremos errores normales. Lo anterior se realizará con la finalidad de saber si dada una serie de datos que presente colas pesadas existe diferencia entre los valores estimados de los parámetros de (4.0.1) si

suponemos un modelo TAR donde los errores tienen distribución t -student o normal.

4.1. Estimación por máxima verosimilitud

A continuación calcularemos la expresión de la función de verosimilitud dada una muestra de (4.0.1), y que nos será de gran utilidad para la estimación de los parámetros del modelo, así como también nos ayudará para la selección del modelo adecuado de una muestra de datos observados basándonos en algún criterio de información como el de Akaike o Bayesiano, esto último lo desarrollaremos en el capítulo 5.

Empezaremos por plantear las ideas en un modelo TAR(2; 1, 1) para luego, extenderlo al caso de un modelo TAR en general con r regímenes. Consideremos que Z_{t-1} sigue un proceso AR(1), es decir,

$$Z_{t-1} = \rho Z_{t-2} + \psi_t, \quad \psi_t \sim iidN(0, \sigma_u^2), \quad |\rho| < 1. \quad (4.1.1)$$

Sea $\mathbf{Y} = (Y_1, \dots, Y_T)'$ una muestra proveniente de un TAR(2; 1, 1) con $\mathbf{Z} = (Z_0, \dots, Z_T)'$ los valores de la variable umbral dada en (4.1.1). Usando la representación en (4.0.1) se tiene la siguiente expresión para el modelo TAR(2; 1, 1),

$$Y_t = I(Z_{t-1} \leq \gamma) \beta_{11} Y_{t-1} + I(Z_{t-1} > \gamma) \beta_{21} Y_{t-1} + \epsilon_t,$$

donde la distribución para ϵ_t es t -student estándar con ν grados de libertad cuya función de densidad se obtiene de tomar $\lambda = 1$, $\mu = 1$ en (3.4.1), es decir

$$f_{\epsilon}(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \left(\frac{1}{\pi\nu}\right)^{\frac{1}{2}} \left(1 + \frac{x^2}{\nu}\right)^{-\left(\frac{\nu+1}{2}\right)}.$$

Asumamos también que el comportamiento de Y_t no afecta al de Z_t y que ψ_i y ϵ_i son mutuamente independientes. Además, consideremos que las funciones de densidades de Y_1 y Z_0 son conocidas.

Denotemos por $\mathbf{X} = (\mathbf{Y}, \mathbf{Z})$ al vector de observaciones de los procesos Y_t y Z_t hasta el tiempo T , por $\Theta_{\mathbf{Y}} = (\beta_1, \beta_2, \nu)$ y $\Theta_{\mathbf{Z}} = (\rho, \sigma_u)$. Podemos escribir la función de verosimilitud como

$$\begin{aligned} L(\Theta_{\mathbf{Y}}, \Theta_{\mathbf{Z}}, \gamma; \mathbf{X}) &= f(\mathbf{X}; \Theta_{\mathbf{Y}}, \Theta_{\mathbf{Z}}, \gamma) \\ &= f(\mathbf{Y}|\mathbf{Z}; \Theta_{\mathbf{Y}}, \Theta_{\mathbf{Z}}, \gamma) g(\mathbf{Z}; \Theta_{\mathbf{Z}}), \end{aligned}$$

donde

$$\begin{aligned} g(\mathbf{Z}; \Theta_{\mathbf{Z}}) &= g(Z_{T-1}|Z_{T-2}, \dots, Z_1; \Theta_{\mathbf{Z}}) g(Z_{T-2}|Z_{T-3}, \dots, Z_1; \Theta_{\mathbf{Z}}) \dots g(Z_0; \Theta_{\mathbf{Z}}) \\ &= g(Z_{T-1} - \rho Z_{T-2}; \sigma_u) g(Z_{T-2} - \rho Z_{T-3}; \sigma_u) \dots g(Z_0; \Theta_{\mathbf{Z}}) \end{aligned}$$

Mientras que

$$\begin{aligned} f(\mathbf{Y}|\mathbf{Z}; \Theta_{\mathbf{Y}}, \Theta_{\mathbf{Z}}, \gamma) &= f(Y_T|Y_{T-1}, \dots, Y_1, Z_{T-1} \dots Z_0; \Theta_{\mathbf{Y}}, \Theta_{\mathbf{Z}}, \gamma) \dots \\ &\times f(Y_2|Y_1, Z_1; \Theta_{\mathbf{Y}}, \Theta_{\mathbf{Z}}; \gamma) f(Y_1|Y_0, Z_0; \Theta_{\mathbf{Y}}, \Theta_{\mathbf{Z}}, \gamma) \end{aligned}$$

Recordando que $\psi_t \sim N(0, \sigma_u^2)$ y $\epsilon_t \sim t_{\nu}$ se obtiene

$$g(\mathbf{Z}; \Theta_{\mathbf{Z}}, \gamma) = \prod_{t=2}^T \frac{1}{\sqrt{2\pi}\sigma_u} \exp\left(-\frac{1}{2\sigma_u^2} (Z_i - \rho Z_{i-1})^2\right),$$

y

$$f(\mathbf{Y}|\mathbf{Z}; \boldsymbol{\Theta}_Y, \boldsymbol{\Theta}_Z, \gamma) = \prod_{t=2}^T \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\nu\pi}} \left(1 + \frac{(Y_t - (Z_{t-1} \leq \gamma)Y_{t-1} + I(Z_{t-1} > \gamma)Y_{t-1})^2}{\nu} \right)^{-\frac{\nu+1}{2}}$$

Por lo tanto la función de verosimilitud para el conjunto de datos \mathbf{X} del modelo TAR(2; 1, 1) es:

$$\begin{aligned} L(\boldsymbol{\Theta}_Y, \boldsymbol{\Theta}_Z, \gamma; \mathbf{X}) &= \prod_{t=2}^T \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\nu\pi}} \left(1 + \frac{(Y_t - (Z_{t-1} \leq \gamma)y_{t-1} + I(Z_{t-1} > \gamma)Y_{t-1})^2}{\nu} \right)^{-\frac{\nu+1}{2}} \\ &\quad \times \frac{1}{\sigma_u^{T-1}} \exp \left(-\frac{1}{2\sigma^2} \sum_{t=2}^T (Z_t - \rho Z_{t-1})^2 \right). \end{aligned}$$

Usando el procedimiento anterior se puede extender el cálculo de la verosimilitud al caso general TAR($r; p_1, p_2, \dots, p_r$). Para este caso consideraremos fijos el número de regímenes r , los órdenes de los procesos autorregresivos $\mathbf{p} = (p_1, p_2, \dots, p_r)$ y los valores umbrales $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_{r-1})$. Supongamos conocidas las funciones de densidad de cada una de las observaciones

$$Y_1, Y_2, \dots, Y_{t.max}, Z_0, Z_1, \dots, Z_{t.max-1},$$

$t.max = \max(p_1, p_2, \dots, p_r)$ y denotemos por $\boldsymbol{\Theta}_Y$ al conjunto de parámetros de Y_t ,

$$\boldsymbol{\Theta}_Y = (\beta_{11}, \dots, \beta_{1p_1}, \dots, \beta_{r1}, \dots, \beta_{rp_r}, \nu),$$

por $\boldsymbol{\Theta}_Z$ al conjunto de parámetros de Z_{t-1} .

La función de verosimilitud para la muestra observada $\mathbf{W} = (\mathbf{Y}, \mathbf{Z})$ al tiempo T es

$$\begin{aligned} L(\boldsymbol{\Theta}_Y, \boldsymbol{\Theta}_Z, \boldsymbol{\gamma}; \mathbf{W}) &= f(\mathbf{W}; \boldsymbol{\Theta}_Y, \boldsymbol{\Theta}_Z, \boldsymbol{\gamma}) \\ &= f(\mathbf{Y}|\mathbf{Z}; \boldsymbol{\Theta}_Y, \boldsymbol{\Theta}_Z, \boldsymbol{\gamma}) g(\mathbf{Z}; \boldsymbol{\Theta}_Z), \end{aligned}$$

donde, de nueva cuenta, como en el caso del TAR (2; 1, 1) tenemos

$$\begin{aligned} g(\mathbf{Z}; \boldsymbol{\Theta}_Z) &= g(Z_{T-1}|Z_{T-2}, \dots, Z_1; \boldsymbol{\Theta}_Z) g(Z_{T-2}|Z_{T-3}, \dots, Z_1; \boldsymbol{\Theta}_Z) \dots g(Z_{t.max-1}; \boldsymbol{\Theta}_Z) \\ &= g(Z_{T-1} - \rho Z_{T-2}; \sigma_u) g(Z_{T-2} - \rho Z_{T-3}; \sigma_u) \dots g(Z_{t.max-1}; \boldsymbol{\Theta}_Z) \end{aligned}$$

y

$$\begin{aligned} f(\mathbf{Y}|\mathbf{Z}; \boldsymbol{\Theta}_Y, \boldsymbol{\Theta}_Z, \gamma) &= f(Y_T|Y_{T-1}, \dots, Y_1, Z_{T-1} \dots Z_0; \boldsymbol{\Theta}_Y, \boldsymbol{\Theta}_Z, \gamma) \dots \\ &\times f(Y_{t.max+1}|Y_1, Z_1; \boldsymbol{\Theta}_Y, \boldsymbol{\Theta}_Z, \gamma) f(Y_{t.max}; \boldsymbol{\Theta}_Y, \boldsymbol{\Theta}_Z, \gamma) \end{aligned}$$

con f es la función de densidad de una distribución t -student con ν grados de libertad y g es la función de densidad de una distribución normal con media cero y varianza σ_u^2 .

Recordando que $\psi_t \sim N(0, \sigma_u^2)$ y $\epsilon_t \sim t_\nu$ se llega a que la función de verosimilitud es

$$L(\boldsymbol{\Theta}_Y, \boldsymbol{\Theta}_Z; \mathbf{W}) = \prod_{t=t.max+1}^T \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2}) \sqrt{\nu\pi}} \left(1 + \frac{X_t^2}{\nu}\right)^{-\frac{\nu+1}{2}} \quad (4.1.2)$$

$$\times \frac{1}{\sigma^{T-1}} \exp\left(-\frac{1}{\sigma^2} \sum_{t=t.max+1}^T (Z_{t-1} - \rho Z_{t-2})^2\right). \quad (4.1.3)$$

con

$$X_t = Y_t - \left[I(Z_{t-d} \leq \gamma_1) \left(\sum_{i=1}^{p_1} \beta_{1i} Y_{t-i} \right) + \dots + I(\gamma_{r-1} \leq Z_{t-1}) \left(\sum_{i=1}^{p_r} \beta_{ri} Y_{t-i} \right) \right],$$

Notemos que (4.1.2) depende en demasía de los valores r y \mathbf{p} . Por ende, un cambio en esos parámetros modificará las estimaciones de los parámetros $\boldsymbol{\Theta}_Y$ y $\boldsymbol{\Theta}_Z$, de ahí la importancia de tener el modelo especificado correctamente, y que trataremos en el capítulo 5.

Ahora, el método de máxima verosimilitud establece que los estimadores para Θ_Y , Θ_Z , bajo el supuesto de que γ , r , \mathbf{p} están fijos, son aquellos que maximizan la función de verosimilitud dada en (4.1.2). Regularmente es mucho más fácil maximizar el logaritmo de la verosimilitud dada en (4.1.4):

$$\ell(\Theta_Y, \Theta_Z; Y_t, Z_{t-1}) = \sum_{t=t.max+1}^T \log(f(X_t)) \quad (4.1.4)$$

$$+ \sum_{t=t.max+1}^T \log(g(Z_{t-1} - \rho Z_{t-2})) \quad (4.1.5)$$

Para el cálculo de los parámetros de Θ_Z no se presenta ninguna complicación ya que se puede derivar la log verosimilitud respecto de cada parámetro de Θ_Z y, posteriormente encontrarse los estimadores máximos verosímiles de forma analítica al igual a cero dichas derivadas. De hecho, los estimadores de Θ_Z son mismos que se calcularon en el capítulo 2,

$$\hat{\rho} = \frac{\sum_{t=2}^T Z_{t-1} Z_{t-1}}{\sum_{t=2}^T Z_{t-2}^2},$$

y

$$\hat{\sigma}^2(\hat{\rho}) = \frac{1}{T-1} \sum_{t=2}^T (Z_{t-1} - \hat{\rho} Z_{t-2})^2.$$

Sin embargo, la expresión dada en (4.1.2) resulta ser intratable y difícil de derivar con respecto a cada elemento de Θ_Y por lo que se recurrirá a métodos iterativos como Newton-Raphson o recocido simulado para encontrar estimadores para los parámetros Θ_Y del modelo. Como la verosimilitud no es derivable en los puntos γ entonces estimar el valor del o los umbrales es una tarea complicada. Una alternativa para la estimación de γ se encuentra

en Qian(1998), donde se propone que un estimador para es

$$\inf_{\gamma} \argmin \ell \left(\hat{\Theta}_{\mathbf{Y}}, \hat{\Theta}_{\mathbf{Z}} \right),$$

como ya mencionamos en el capítulo 2.

4.2. Simulaciones

En esta sección presentaremos 4 ejemplos simulados, en los primeros dos estudiaremos datos provenientes de un proceso TAR(2; 1, 1) , y en los dos restantes, datos de un proceso TAR(2; 2, 2). Se considerarán los casos en que la distribución de los errores es t -student con 4 y 5 grados de libertad. Se estimarán los parámetros del proceso y además, para cada uno de ellos se calcularán los promedios y desviación estándar muestrales, así como también los cuantiles $q_{0.05}$ y $q_{0.95}$. Primero se hará la estimación de los parámetros suponiendo que el VPGD¹ es un TAR que tiene errores con distribución t -student y después el caso en el cual ϵ_t es tratado como si fuese normal aunque el VPGD tenga errores t -student independientes.

TAR(2; 1,1) con $\nu = 5$

Consideremos el siguiente proceso TAR(2; 1, 1)

$$Y_t = \begin{cases} 0.5Y_{t-1} + \epsilon_t & Z_{t-1} \leq 1 \\ -0.7Y_{t-1} + \epsilon_t & Z_{t-1} > 1 \end{cases} \quad \epsilon_t \sim t_5. \quad (4.2.1)$$

Con variable umbral

$$Z_{t-1} = 0.5Z_{t-2} + \psi_t, \quad \psi_t \sim iidN(0, 1). \quad (4.2.2)$$

¹Verdadero proceso generador de los datos

Se simularon² 300 datos del proceso y las gráficas de Z_t y de Y_t se encuentran en la figura (4.2.1) y (4.2.2) respectivamente.

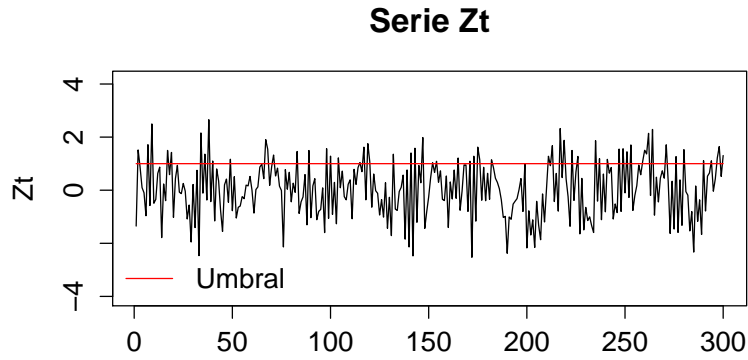


Figura 4.2.1: Realización de $Z_{t-1} = 0.5Z_{t-2} + \psi_t$

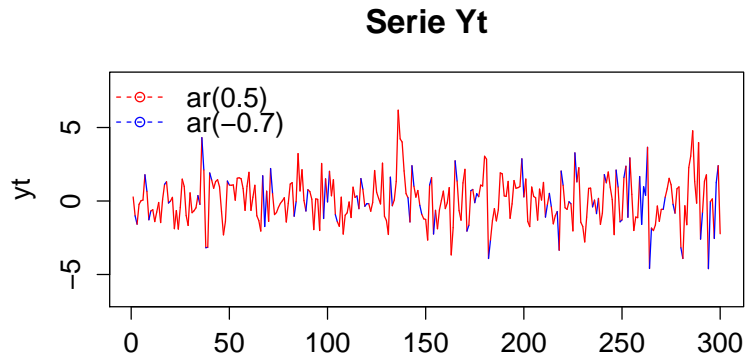


Figura 4.2.2: En rojo el proceso $Y_t = 0.5Y_{t-1} + \epsilon_t$ y en azul $Y_t = -0.7Y_{t-1} + \epsilon_t$.

Luego, la función de verosimilitud del proceso dado en (4.2.1) con 300 datos

²En los datos simulados de todos los ejemplos de este trabajo se eliminaron las primeras 2,000 observaciones con la finalidad de contrarestar los efectos de los valores iniciales de la serie.

y denotando por

$$\Theta_Y = (\beta_1, \beta_2, \nu), \quad \Theta_Z = (\rho, \sigma),$$

es

$$\begin{aligned} L(\Theta_Y, \Theta_Z; Y_T, Z_{T-1}, \gamma) &= \prod_{t=2}^{300} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right) \sqrt{\nu\pi}} \left(1 + \frac{X_t^2}{\nu}\right)^{-\frac{\nu+1}{2}} \\ &\times \frac{1}{\sigma^{299}} \exp\left(-\frac{1}{\sigma^2} \sum_{t=2}^{299} (Z_{t-1} - \rho Z_{t-2})^2\right). \end{aligned}$$

donde,

$$X_t = Y_t - [I(Z_{t-1} \leq \gamma) (\beta_{11}Y_{t-1} + \beta_{12}Y_{t-2}) + I(Z_{t-1} > \gamma) (\beta_{21}Y_{t-1} + \beta_{22}Y_{t-2})].$$

El cálculo de los parámetros Θ_Y , Θ_Z y γ se realizó de la siguiente manera:

1. Se encontró un intervalo en el cual se buscará el estimador de γ , dicho intervalo está dado por los valores mínimos y máximos de Z_t , que en este caso fueron $[-2.527, 2.655]$.
2. Luego, se realizó una partición del 1000 puntos del intervalo anterior en los cuales iremos variando los valores de γ .
3. Para cada valor de la partición se calcularon los estimadores para Θ_Z y Θ_Y , así como la evaluación de la logverosimilitud en los estimadores encontrados.
4. Por último, se seleccionó como estimador de γ al valor de la partición que minimiza la logverosimilitud. Y los estimadores de Θ_Z y Θ_Y son aquellos que se estiman suponiendo ese valor de γ .

Utilizando la función `optim` de R encontramos que el valor de γ que minimiza

la logverosimilitud es 1.002 y en la figura (4.2.3) se encuentra los valores del negativo de la logverosimilitud evaluada en diferentes valores de γ .

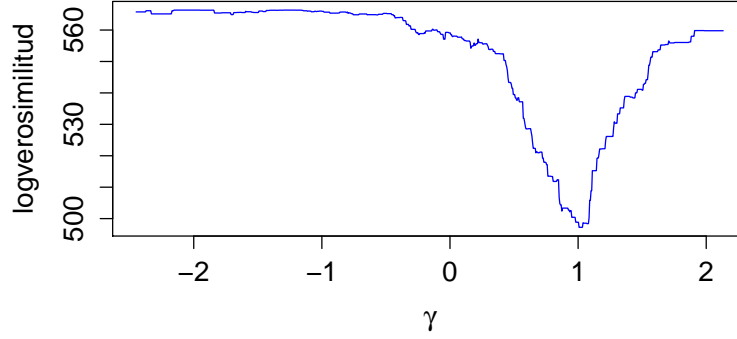


Figura 4.2.3: Valor de $-\ell(\hat{\Theta}_Y, \hat{\Theta}_Z)$ variando γ .

Los parámetros estimados para β_{11} , β_{21} , ν suponiendo que el valor verdadero del umbral γ es 1.002 fueron

$$\hat{\beta}_1 = 0.4728, \hat{\beta}_2 = -0.834, \hat{\nu} = 4.842$$

Después, se realizaron 500 simulaciones del proceso y se tomaron los promedios, desviaciones estándar y los cuantiles $q_{0.5}$ y $q_{0.95}$ para cada uno de los parámetros. Los resultados obtenidos fueron los siguientes:

	Valor real	Promedio	$q_{.05}$	$q_{.95}$	D.E. ³
β_1	0.5	0.499	0.406	0.588	0.063
β_2	-0.7	-0.667	-0.434	-0.864	0.145
ν	5	5.428	3.572	8.495	1.762
γ	1	0.958	0.764	1.078	0.148

Tabla 4.1: Resultados obtenidos para cada parámetro del modelo.

Para hacer una comparativa se calcularon los estimadores suponiendo que en (4.2.1) los errores ϵ_t siguen una distribución normal con media 0 y varianza σ^2 . Para lo anterior se realizó el mismo procedimiento hecho con el caso t -student y los resultados se presentan en la siguiente tabla.

	Valor real	Promedio	$q_{.05}$	$q_{.95}$	D. E.
β_1	0.5	0.497	0.394	0.583	0.063
β_2	-0.7	-0.661	-0.885	-0.361	0.164
σ	1.291	1.289	1.133	1.48	0.114
γ	1	0.959	0.762	1.094	0.151

Tabla 4.2: Resultados obtenidos para cada parámetro del modelo.

Podemos observar de los cuadros (4.1) y (4.2) que la diferencia entre los estimadores obtenidos es casi nula.

TAR(2; 1,1) con $\nu = 4$

En las simulaciones anteriores se encuentran diferencias mínimas en las estimaciones de los parámetros suponiendo errores normales o t -student, razón por la cual en el siguiente ejemplo examinaremos el caso en donde la curtosis de la distribución t de los errores es infinita, tal es el caso en que sus grados de libertad son 4. Así, consideremos el siguiente proceso TAR(2; 1, 1)

$$Y_t = \begin{cases} 0.5Y_{t-1} + \epsilon_t & Z_{t-1} \leq 1 \\ -0.7Y_{t-1} + \epsilon_t & Z_{t-1} \geq 1 \end{cases} \quad \epsilon_t \sim t_4. \quad (4.2.3)$$

Se simularon 300 datos del proceso y las gráficas de Z_t y de Y_t se encuentran en la figura (4.2.1) y (4.2.5) respectivamente.

³En todas las tablas usaremos D. E. para la desviación estándar.

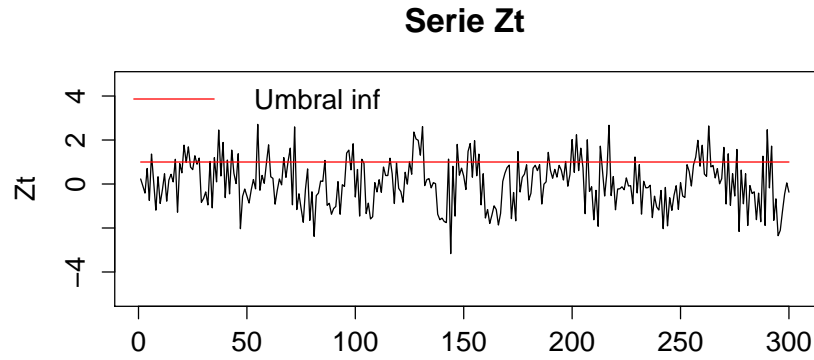


Figura 4.2.4: Realización de $Z_{t-1} = 0.5Z_{t-2} + \psi_t$

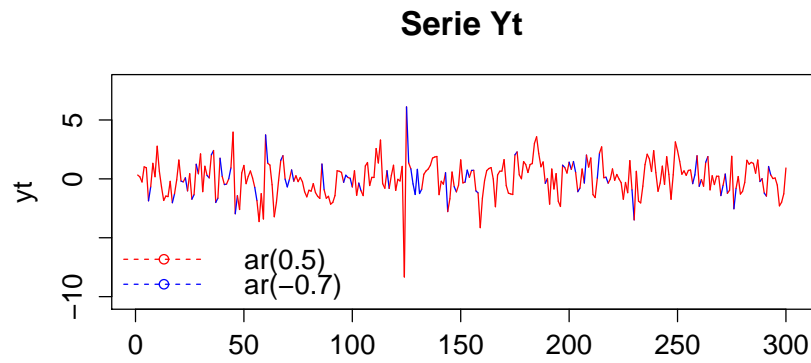


Figura 4.2.5: En rojo el proceso $Y_t = 0.5Y_{t-1} + \epsilon_t$ y en azul $Y_t = -0.7Y_{t-1} + \epsilon_t$

Se realizó el mismo procedimiento para el cálculo de los parámetros hecho en el proceso TAR(2; 1, 1) con $\nu = 5$. Los resultados se presentan en la siguiente tabla

	Valor real	Promedio	$q_{.05}$	$q_{.95}$	D. E.
β_1	0.5	0.499	0.408	0.585	0.045
β_2	-0.7	-0.699	-0.874	-0.517	0.109
ν	4	4.248	3.109	6.146	0.997
γ	1	0.991	0.919	1.078	0.045

y suponiendo errores normales los resultados fueron

	Valor real	Promedio	$q_{.05}$	$q_{.95}$	D. E.
β_1	0.5	0.498	0.400	0.587	0.054
β_2	-0.7	-0.702	-0.892	-0.499	0.122
σ	1.414	1.400	1.205	1.671	0.141
γ	1	0.992	0.915	1.084	0.045

Como podemos notar, en este caso las estimaciones de los parámetros β_1 , β_2 , γ son más precisas, y lo cual se debe a la forma de la distribución que consideramos. Sin embargo, existe un aumento en la desviación estándar muestral del parámetro σ del modelo con distribución normal con respecto a su análogo con $\nu = 5$.

TAR(2; 2, 2) con $\nu = 5$

Ahora estudiaremos el proceso TAR(2;2,2) dado por

$$Y_t = \begin{cases} 1.2Y_{t-1} - 0.4Y_{t-2} + \epsilon_t & Z_{t-1} \leq 1.5 \\ 1.5Y_{t-1} - 0.7Y_{t-2} + \epsilon_t & Z_{t-1} \geq 1.5 \end{cases} \quad \epsilon_t \sim t_5. \quad (4.2.4)$$

Su función de verosimilitud considerando

$$\Theta_Y = (\beta_{11}, \beta_{12}, \beta_{21}, \beta_{22}, \nu), \quad \Theta_Z = (\rho, \sigma),$$

es

$$L(\Theta_Y, \Theta_Z; Y_T, Z_{T-1}, \gamma) = \prod_{t=3}^{300} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2}) \sqrt{\nu\pi}} \left(1 + \frac{X_t^2}{\nu}\right)^{-\frac{\nu+1}{2}} \\ \times \frac{1}{\sigma^{299}} \exp\left(-\frac{1}{\sigma^2} \sum_{t=2}^{299} (Z_{t-1} - \rho Z_{t-2})^2\right),$$

donde

$$X_t = Y_t - [I(Z_{t-1} \leq \gamma) (\beta_{11}Y_{t-1} + \beta_{12}Y_{t-2}) + I(Z_{t-1} > \gamma) (\beta_{21}Y_{t-1} + \beta_{22}Y_{t-2})].$$

Se simularon 500 series con 300 datos de proceso (4.2.4). Las gráficas de una realización de Y_t y Z_t se presentan en las figuras (4.2.6) y (4.2.7) respectivamente.

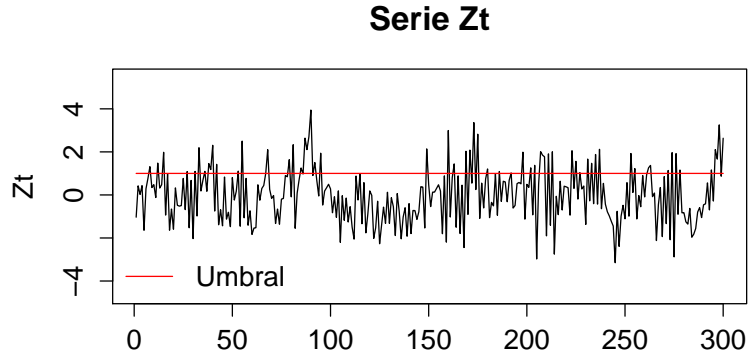


Figura 4.2.6: $Z_{t-1} = 0.5Z_{t-2} + \psi_t$

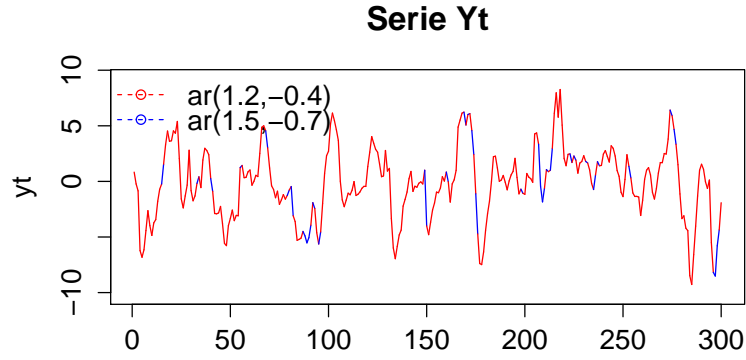


Figura 4.2.7: En rojo $1.2Y_{t-1} - 0.4Y_{t-2} + \epsilon_t$ y en azul $1.5Y_{t-1} - 0.7 + \epsilon_t$.

Los resultados de estimadores obtenidos suponiendo un modelo TAR con errores t -student fueron

	Valor real	Promedio	$q_{.025}$	$q_{.925}$	D. E.
β_{11}	1.2	1.189	1.102	1.282	0.054
β_{12}	-0.4	-0.392	-0.483	-0.300	0.054
β_{21}	1.5	1.514	1.101	1.871	0.236
β_{22}	-0.7	-0.720	-1.067	-0.317	0.240
ν	5	5.935	3.649	9.293	6.182
γ	1.5	1.502	1.350	1.68	0.109

Y bajo el supuesto de normalidad en los errores se obtuvo lo siguiente

	Valor real	Promedio	$q_{.05}$	$q_{.95}$	D. E.
β_{11}	1.2	1.187	1.083	1.286	0.054
β_{12}	-0.4	-0.391	-0.490	-0.289	0.054
β_{21}	1.5	1.513	1.064	1.896	0.248
β_{22}	-0.7	-0.717	-1.107	-0.261	0.262
σ	1.291	1.270	1.125	1.447	0.054
γ	1.5	1.502	1.350	1.682	0.109

TAR(2; 2,2) con $\nu = 4$

Ahora consideremos el mismo proceso TAR(2;2,2) dado en (4.2.4)

$$Y_t = \begin{cases} 1.2Y_{t-1} - 0.4Y_{t-2} + \epsilon_t & Z_{t-1} \leq 0 \\ 1.5Y_{t-1} - 0.7Y_{t-2} + \epsilon_t & Z_{t-1} \geq 0 \end{cases} \quad \epsilon_t \sim t_4. \quad (4.2.5)$$

En las gráficas (4.2.8) y (4.2.9) se encuentra una realización del proceso TAR anterior.

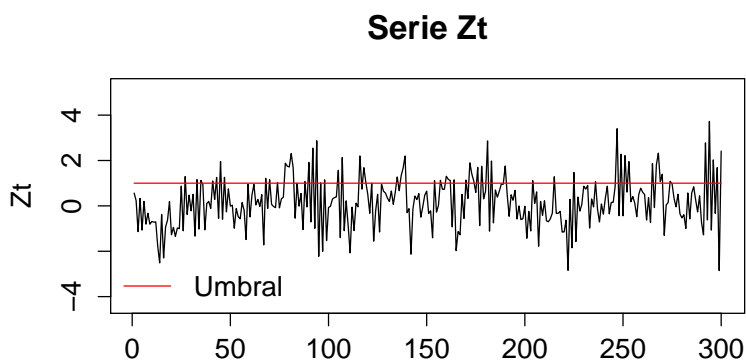


Figura 4.2.8: $Z_{t-1} = 0.5Z_{t-2} + \psi_t$

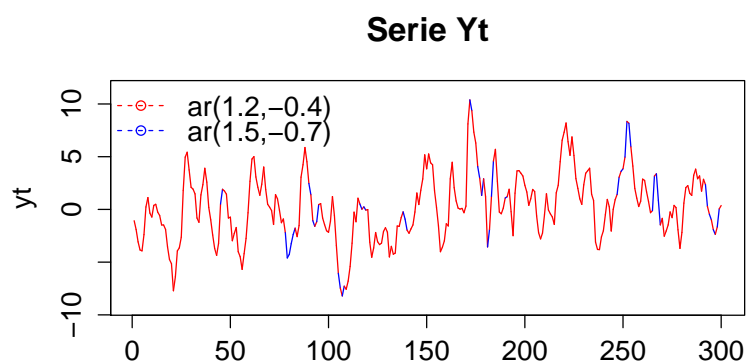


Figura 4.2.9: En rojo $1.2Y_{t-1} - 0.4Y_{t-2} + \epsilon_t$ y en azul $1.5Y_{t-1} - 0.7 + \epsilon_t$

Los resultados de los estimaciones de lo parámetros en las 500 simulaciones se resumen en la siguiente tabla

	Valor real	Promedio	$q_{.025}$	$q_{.975}$	D. E.
β_{12}	1.2	1.194	1.112	1.27	0.048
β_{12}	-0.4	-0.396	-0.477	-0.311	0.05
β_{21}	1.5	1.540	1.128	1.901	0.228
β_{22}	-0.7	-0.742	-1.118	-0.325	0.230
ν	4	4.200	3.045	5.889	1.013
γ	1.5	1.503	1.35	1.682	0.109

Y suponiendo el modelo TAR errores normales los resultados fueron

	Valor real	Promedio	$q_{.05}$	$q_{.95}$	D. E.
β_{12}	1.2	1.195	1.093	1.293	0.054
β_{12}	-0.4	-0.398	-0.493	-0.295	0.054
β_{21}	1.5	1.541	1.050	1.971	0.275
β_{22}	-0.7	-0.743	-1.173	-0.238	0.272
σ	1.414	1.403	1.224	1.639	0.141
γ	1.5	1.513	1.350	1.686	0.109

Como pudimos ver en los ejemplos anteriores, la metodología propuesta para estimar los parámetros del modelo, Θ_Y , Θ_Z y γ , da buenas aproximaciones de los valores verdaderos. Sin embargo, la ganancia que se obtiene de utilizar el modelo con errores t versus el modelo con errores normales es mínima a diferencia de otros modelos mencionados en la literatura, como los modelos de regresión, en los cuales existe una clara ventaja de utilizar errores t para la estimación de parámetros y pronósticos en datos que presentan colas pesadas.

Capítulo 5

Identificación del modelo

Una parte importante cuando hacemos uso de los modelos TAR es la identificación del modelo adecuado para un conjunto de datos del cual, mediante pruebas como los residuales predictivos propuesto por Tsay(1998) y en base a la experiencia del experto, sabemos que el fenómeno que los genera puede ser apropiado para este tipo de modelos. Dicha especificación se puede hacer, a grandes rasgos, de la siguiente manera:

1. Identificar un número máximo de regímenes, que denotaremos por r_{max} . Para cada i fijo, $i = 2, 3, \dots, r_{max}$ hacer los pasos 2, 3, 4, 5.
2. Buscar posibles intervalos para los valores de cada uno de los $i - 1$ umbrales, que indicaremos por I_i , y establecer los órdenes máximos probables, $p_{1,max}, p_{2,max}, \dots, p_{i,max}$, de los i procesos autorregresivos.
3. Luego, para cada $(i - 1)$ -tupla, $\gamma = (\gamma_1, \dots, \gamma_{i-1})$, construida tomando valores $\gamma_i \in I_i$ y cumpliendo la condición $\gamma_1 < \gamma_2 < \dots, \gamma_{i-2} < \gamma_{i-1}$, hacer el siguiente paso.
4. Para cada modelo

$$\text{TAR}(i; p_1, p_2, \dots, p_i), p_i = 1, 2, \dots, p_{i,max},$$

estimar sus parámetros para después calcular el valor de la verosimilitud evaluada en los parámetros encontrados.

5. Elegimos los parámetros umbrales, $\gamma = (\gamma_1, \dots, \gamma_{i-1})$ como aquellos que minimicen el valor negativo de la logverosimilitud. Asimismo los valores de los parámetros asociados a los procesos autorregresivos en cada régimen serán los que se obtengan de suponer el valor de γ encontrado anteriormente. Después, calcular el valor del criterio de información, como por ejemplo el de Akaike (AIC) o Bayesiano (BIC), asociado al modelo.
6. De todos los posibles modelos

$$\text{TAR}(i; p_1, p_2, \dots, p_i), \quad i = 1, 2, \dots, r_{max}$$

seleccionamos aquel que haga mínimo el criterio de información.

En las siguientes secciones desarrollaremos las ideas para identificación del modelo, así como también presentaremos ejemplos con datos simulados en los que ajustaremos un modelo TAR de acuerdo a los pasos descritos anteriormente. Adicionalmente, haremos una comparación entre los modelos seleccionados para datos simulados de un $\text{TAR}(2; p_1, p_2)$ usando el criterio de información de Akaike y el Bayesiano.

5.1. Selección del número máximo de regímenes

Para la selección del número máximo de regímenes, una solución se propone en Nieto(2005) y la cual consiste en ajustar una regresión no paramétrica teniendo como variable de respuesta la serie Y_t y como variable regresora a Z_t . Después, se grafica la función de regresión y en base a los diferentes

comportamientos(o quiebres) que se observen entonces se procederá a proponer un número máximo de regímenes, digamos r_{max} . De la misma manera, usando la gráfica de la regresión también se puede dar una aproximación para posibles valores de los umbrales y los cuales consistirán en intervalos cercanos a los valores de Z_t donde se observan comportamientos diferentes en Y_t . Para ejemplificar lo anterior, consideremos el siguiente modelo TAR (3; 1, 1, 1)

$$Y_t = \begin{cases} 0.1 + \epsilon_t & Z_{t-1} < -0.6 \\ -0.7 + \epsilon_t & -0.6 \leq Z_{t-1} < 0.6, \epsilon_t \sim t_5 \\ 0.4 + \epsilon_t & Z_{t-1} \geq 0.6 \end{cases} \quad (5.1.1)$$

Se realizó una simulación de proceso y las gráficas obtenidas para Y_t y Z_t se presentan en las figuras (5.1.1) y (5.1.2)

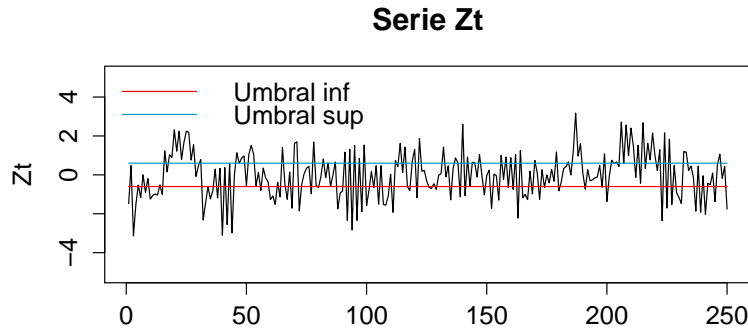


Figura 5.1.1: Proceso Z_t

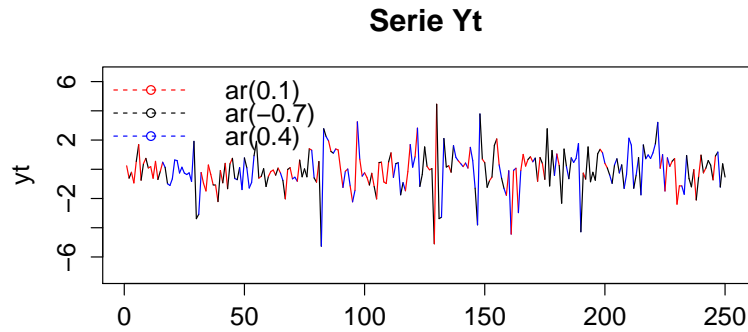


Figura 5.1.2: Proceso Y_t .

Al hacer la regresión no paramétrica de Y_t teniendo como covariable Z_t obtenemos la siguiente gráfica de la función de regresión

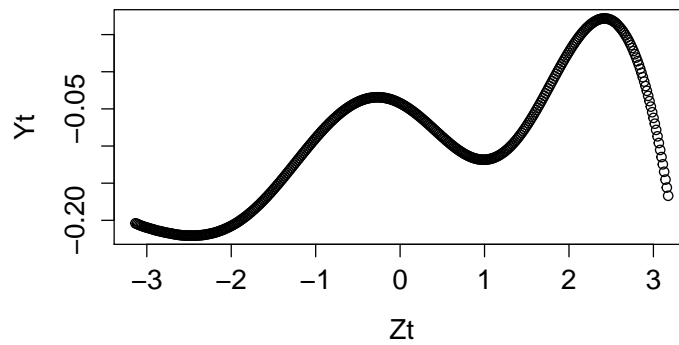


Figura 5.1.3: Regresión no paramétrica entre Y_t y Z_t

de donde observamos 4 quiebres y que de acuerdo con lo mencionado en Nieto(2005) se propone que un número máximo de umbrales serían 4 y por consiguiente un modelo TAR con máximo 5 regímenes. Sin embargo, al considerar modelos TAR con 5 regímenes notamos que 2 de ellos quedan con

pocos datos, misma situación sucede al considerar modelos TAR con 4 regímenes donde uno de ellos contiene pocos datos, razón por la cual se omitieron los modelos TAR con 3 y 4 umbrales. Así, se consideró un número máximo de umbrales igual a 2 y por lo tanto se exploraron modelos TAR con 2 y 3 regímenes. Ahora, los posibles intervalos para los valores de los dos umbrales del modelo TAR $(3; p_1, p_2, p_3)$, denotados por I_1 e I_2 , son aquellos que están conformados por valores de Z_t que se encuentran alrededor de los puntos de quiebres de la gráfica (5.1.3), por ejemplo,

$$\begin{aligned} I_1 &= [-0.8, 0] \\ I_2 &= [0.5, 1.3]. \end{aligned}$$

Mientras que para el valor umbral del modelo TAR $(2; p_1, p_2)$ se utilizó un intervalo de búsqueda comprendido entre los cuantiles $q_{0.15}$ y $q_{0.85}$ de los valores observados de Z_t .

5.2. Selección de los umbrales y órdenes autorregresivos

Como mencionamos anteriormente, la gráfica de la regresión no paramétrica nos ayuda para dar intervalos de valores en los cuales podemos buscar los estimadores de $\gamma_1, \dots, \gamma_{i-1}$. Otra posible solución, aunque computacionalmente menos eficiente, es homologar lo propuesto por Qian(1998) y que consiste en hacer una búsqueda ordena de los umbrales en el intervalo $[\min(Z_t), \max(Z_t)]$.

Una vez obtenidos los intervalos de búsqueda para los valores umbrales, fijamos los ordenes de los procesos autorregresivos con valores entre $1 \leq p_i \leq p_{i,max}$, $i = 1, \dots, r$.

Luego, hacemos una partición del intervalo correspondiente a γ_i , $i = 1, \dots, r$ —

1. Para todas las posibles $\gamma = (\gamma_1, \dots, \gamma_{i-1})$ construidas con valores de la partición en los intervalos correspondientes y considerando que $\gamma_1 < \gamma_2 < \dots, \gamma_{i-2} < \gamma_{i-1}$, calculamos la función de verosimilitud, que de acuerdo con lo hecho en el capítulo anterior tiene la siguiente expresión,

$$L(\Theta_Y, \Theta_Z; Y_t, Z_{t-1}) = \prod_{t=t.max}^T \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\nu\pi}} \left(1 + \frac{X_t^2}{\nu}\right)^{-\frac{\nu+1}{2}} \quad (5.2.1)$$

$$\times \frac{1}{\sigma^{T-1}} \exp\left(-\frac{1}{\sigma^2} \sum_{t=t.max}^T (Z_{t-1} - \rho Z_{t-2})^2\right). \quad (5.2.2)$$

donde,

$$X_t = Y_t - \left[I(Z_{t-1} \leq \gamma_1) \left(\sum_{j=1}^{p_1} \beta_{1j} Y_{t-j} \right) + \dots + I(\gamma_{i-2} \leq Z_{t-1} \leq \gamma_{i-1}) \left(\sum_{k=1}^{p_r} \beta_{rk} Y_{t-k} \right) \right],$$

$$t.max = \text{máximo}(p_1, p_2, \dots, p_r) + 1.$$

Una vez obtenida la función de verosimilitud para el modelo TAR $(i; p_1, p_2, \dots, p_i)$, se minimiza el negativo del logaritmo de dicha función para encontrar los estimadores máximos verosímiles de Θ_Z y Θ_Y . Posteriormente, se obtiene el valor del negativo del logaritmo de $L(\Theta_Y, \Theta_Z; Y_t, Z_{t-1})$ evaluada en los parámetros encontrados. Después, repetimos el procedimiento anterior variando γ dentro de los valores de la partición y seleccionamos como su estimador a aquel que minimice el negativo de la función de log versimilitud.

Ya que se hallan elegidos los estimadores de γ y de los parámetros de los procesos autorregresivos para el modelo TAR $(i; p_1, p_2, \dots, p_r)$, se calcula el valor del criterio de información. En este trabajo consideraremos los criterios de Akaike(AIC) y el Bayesiano(BIC), cuya expresión matemática están dadas por

$$AIC = \sum_{i=1}^r \left(2k_i - 2\log(\hat{L}_i) \right) \quad (5.2.3)$$

$$BIC = \sum_{i=1}^r \left(-2\log(\widehat{L}_i) + k_i \log(T_i) \right) \quad (5.2.4)$$

donde \widehat{L}_i representa la verosimilitud del modelo autorregresivo referente al i -ésimo régimen evaluada en los parámetros estimados, k_i y T_i son, respectivamente, el número de parámetros y el número de datos, en el i -ésimo régimen.

Luego, repetimos todo el proceso anterior variando los valores de los órdenes de los procesos autorregresivos.

Por último la estimación de los órdenes, p_1, \dots, p_r de los procesos autorregresivos de cada régimen y el número de umbral, i , se puede hacer mediante la minimización del criterio de información.

Para nuestro ejemplo dado en (5.1.1) como ya habíamos mencionado se consideró un número máximo de regímenes $r_{max} = 3$ y los órdenes máximos para cada uno de los procesos autorregresivos fue 5. Los resultados obtenidos según el criterio Bayesiano y el de Akaike se resumen en las siguientes tablas

Modelo	BIC	Modelo	BIC
$TAR(2; 1, 1)$	823.8598	$TAR(3; 1, 1, 1)$	793.2021
$TAR(2; 1, 3)$	823.8653	$TAR(3; 1, 1, 3)$	793.3414
$TAR(2; 1, 2)$	825.7739	$TAR(3; 1, 1, 2)$	795.1696

Tabla 5.1: Resultados usando criterio Bayesiano

Modelo	AIC	Modelo	AIC
$TAR(2; 1, 5)$	805.1820	$TAR(3; 1, 1, 5)$	771.6654
$TAR(2; 1, 4)$	805.2453	$TAR(3; 1, 1, 3)$	772.2126
$TAR(2; 1, 3)$	806.2580	$TAR(3; 1, 2, 4)$	772.9403

Tabla 5.2: Resultados usando criterio de Akaike

donde podemos observar que el modelo que hace mínimo el criterio Bayesiano

es un TAR (3; 1, 1, 1) con los siguiente parámetros

	β_1	β_2	β_3	ν	γ_1	γ_2
Real	0.1	-0.7	0.4	5	-0.6	0.6
Estimado	0.2041	-0.5589	0.5276	5.4764	-0.608	0.608

Tabla 5.3: Parámetros del modelo TAR (3; 1, 1, 1)

Mientras que si usamos el criterio de Akaike el modelo seleccionado es un TAR (3; 1, 1, 5) y sus parámetros se encuentran en la tabla (5.4)

	β_{11}	β_{21}	β_{31}	β_{33}	β_{33}	β_{33}	β_{35}	ν	γ_1	γ_2
Real	0.1	-0.7	0.4					5	-0.6	0.6
Estimados	0.2058	-0.558	0.4492	-0.0626	0.2398	-0.007	-0.0220	5.4500	-0.608	0.608

Tabla 5.4: Parámetros del modelo TAR (3; 1, 1, 5)

De este ejemplo podemos ver que el criterio Bayesiano estima de manera correcta el modelo y da una buena aproximación a los parámetros reales, mientras que el criterio de Akaike sobreestima el modelo verdadero del cual provienen los datos. Lo anterior es similar a lo obtenido por Liew & Chong(2003) para modelos TAR y STAR.¹

5.3. Simulaciones

En esta sección presentaremos algunos ejemplos en donde ajustaremos modelos TAR (2; p_1, p_2) y TAR (3; p_1, p_2, p_3) para un conjunto de datos simulados. Los modelos reales del cual provienen los datos se detallan en los siguientes cuadros y en todos ellos consideraremos que la variable Z_t es un proceso AR(1) dado por:

¹Smooth Transition Autoregressive

$$Z_t = 0.5Z_{t-1} + \psi_t, \psi_t \sim N(0, 1)$$

Etiqueta	γ	β_1	β_2	ν
Ejemplo 1	1	0.8	-0.1	5
Ejemplo 2	0.897	0.5	-0.7	5

Tabla 5.5: Modelos TAR(2;1,1) para ajustar.

Etiqueta	γ	β_{11}	β_{12}	β_{21}	β_{22}	ν
Ejemplo 3	1	0.4	0.2	-0.3	0.4	5
Ejemplo 4	1	1.2	-0.4	0.1	-0.7	5

Tabla 5.6: Modelos TAR(2;2,2) para ajustar.

Etiqueta	γ	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}	ν
Ejemplo 5	0	1.2	-0.4	-0.3	-0.7	0.5	5

Tabla 5.7: Modelo TAR(2;3,2) para ajustar.

Recordemos que por ejemplo para el TAR(2; p_1, p_2) la función logverosimilitud es

$$\ell(\Theta_Y, \Theta_Z; Y_t, Z_{t-1}, \gamma) = \sum_{t=\max(p_1, p_2)}^T \log(f_1(X_t)) \quad (5.3.1)$$

$$+ \sum_{t=\max(p_1, p_2)}^T \log(f_2(Z_{t-1} - \rho Z_{t-2})). \quad (5.3.2)$$

con

$$X_t = Y_t - \left[I(Z_{t-1} \leq \gamma) \left(\sum_{i=1}^{p_1} \beta_{1i} Y_{t-i} \right) + I(Z_{t-1} > \gamma) \left(\sum_{i=1}^{p_2} \beta_{2i} Y_{t-i} \right) \right]$$

y f_1, f_2 representando las funciones de densidad t -student con media cero y ν grados de libertad y normal con media cero y varianza σ_u^2 , respectivamente.

Para cada uno de los modelos que se presentan en los cuadros (5.5), (5.6) y (5.7) se simuló una serie con 300 observaciones, y se realizó lo siguiente para la selección mejor modelo $TAR(2; p_1, p_2, p_3)$ que los ajusta:

- Se encontró el intervalo en el cual buscaremos el valor del umbral γ , es decir $[\min(Z_t), \max(Z_t)]$.
- Se realizó una partición de 500 puntos de $[\min(Z_t), \max(Z_t)]$. Para cada punto de la partición se alternaron los órdenes de los procesos autorregresivos, $p_1 = 1, 2, \dots, 5$, $p_2 = 1, 2, \dots, 5$, y se calcularon los parámetros del modelo así como el valor del negativo de la log verosimilitud.
- Se escogió como valor umbral γ a aquel valor de la partición que minimizó el negativo de la log verosimilitud dada en (5.3.1).
- Y se eligieron p_1 y p_2 como aquellos valores que minimizan el criterio de información de Akaike o Bayesiano.

Mientras que para la selección del mejor modelo $TAR(3; p_1, p_2, p_3)$ se hizo lo siguiente

- Se encontraron los intervalos en el cual buscaremos los valores de los umbrales γ_1 y γ_2 , es decir $I_1 = [\min(Z_t), \max(Z_t)]$ e $I_2 = [\min(Z_t), \max(Z_t)]$.
- Se realizó una partición de 500 puntos para I_1 e I_2 . Para cada $\gamma = (\gamma_1, \gamma_2)$, construida tomando valores γ_1 y γ_2 en la partición del intervalo correspondiente y cumpliendo la condición $\gamma_1 < \gamma_2$ se alternaron los órdenes de los procesos autorregresivos, $p_1 = 1, 2, \dots, 5$, $p_2 = 1, 2, \dots, 5$, $p_3 = 1, 2, \dots, 5$ y se calcularon los parámetros del modelo así como el valor del negativo de la log verosimilitud.

- Se escogió como valor umbral $\gamma = (\gamma_1, \gamma_2)$ a aquel valor de la partición que minimizó el negativo de la log verosimilitud.
- Y se eligieron p_1 , p_2 y p_3 como aquellos valores que minimizan el criterio de información de Akaike o Bayesiano.

Ejemplo 1

Para este caso notamos que considerar modelos TAR con 3 regímenes nos deja al menos uno de ellos con pocos datos por lo que solo se considerará un modelo TAR con dos regímenes y por lo tanto nos enfocaremos en la estimación de los órdenes del modelo, en los parámetros de los modelos autorregresivos, y el valor umbral.

Se probaron modelos TAR $(2; p_1, p_2)$ alternando $p_1 = 1, 2, \dots, 5, p_2 = 1, 2, \dots, 5$ y los 3 modelos con menor valor del criterio BIC se encuentran la tabla siguiente

Modelo	BIC	$\hat{\gamma}$
$TAR(2; 1, 1)$	949.821	1.009
$TAR(2; 1, 2)$	951.152	1.009
$TAR(2; 2, 1)$	951.727	1.009

Tabla 5.8: Resultados para el Ejemplo 1 usando el BIC

El mejor modelo para el conjunto de datos de acuerdo con el criterio BIC es un TAR $(2; 1, 1)$ cuyos parámetros son

	β_{12}	β_{21}	γ	ν
Real	0.8	-0.1	1	5
Estimado	0.755	-0.15	1.009	5.453

Para el criterio de Akaike se empleo la misma metodología y se alternaron diferentes modelos TAR $(2; p_1, p_2)$ con $p_1 = 1, 2, \dots, 5, p_2 = 1, 2, \dots, 5$. Los

3 modelos con menor valor del criterio AIC fueron

Modelo	AIC	$\hat{\gamma}$
$TAR(2; 5, 1)$	932.367	1.009
$TAR(2; 1, 5)$	933.520	1.009
$TAR(2; 5, 2)$	933.695	1.009

Tabla 5.9: Resultados para el Ejemplo 1 usando el AIC

Y por lo tanto, el mejor modelo usando el criterio AIC es un $TAR(2; 5, 1)$ con los siguientes parámetros

	β_{11}	β_{12}	β_{13}	β_{14}	β_{15}	β_{21}	γ	ν
Real	0.8					-0.1	1	5
Estimado	0.717	0.072	0.019	-0.004	-0.059	-0.145	1.009	5.575

Ejemplo 2

En la tabla (5.10) se presentan los 3 modelos con menor valor en el criterio BIC y su respectivo umbral estimado para los datos simulados del Ejemplo 2.

Modelo	BIC	$\hat{\gamma}$	Modelo	BIC	$\hat{\gamma}$	
$TAR(2; 1, 1)$	952.566	0.895	$TAR(3; 5, 1, 1)$	956.617	-1.767	0.897
$TAR(2; 1, 2)$	953.579	0.895	$TAR(3; 4, 1, 1)$	957.399	-1.800	0.897
$TAR(2; 2, 1)$	955.877	0.895	$TAR(3; 3, 1, 1)$	958.329	-1.800	0.897

Tabla 5.10: Resultados para el Ejemplo2 usando el BIC

De acuerdo con lo anterior, el modelo ajustado por el BIC es un $TAR(2; 1, 1)$ con los siguientes parámetros

	β_{11}	β_{21}	γ	ν
Valor real	0.5	-0.7	0.897	5
Estimado	0.535	-0.825	0.895	6.259

Mientras que en la tabla (5.11) se presenta la misma información que en (5.10) solo que utilizando el criterio de Akaike.

Modelo	AIC	γ	Modelo	AIC	γ	
$TAR(2; 1, 5)$	936.596	0.895	$TAR(3; 5, 1, 2)$	941.378	-1.767	0.897
$TAR(2; 1, 4)$	937.306	0.895	$TAR(3; 5, 1, 1)$	941.683	-1.767	0.897
$TAR(2; 2, 5)$	938.596	0.895	$TAR(3; 1, 1, 5)$	942.223	-1.767	0.897

Tabla 5.11: Resultados para el Ejemplo 2 usando el AIC

Y el modelo ajustado por el AIC es un $TAR(2; 1, 5)$ con los siguientes parámetros

	β_{11}	β_{21}	β_{22}	β_{23}	β_{24}	β_{25}	γ	ν
Valor real	0.5	-0.7					0.897	5
Estimado	0.534	-0.917	0.164	-0.013	-0.002	-0.052	0.895	6.243

Ejemplo 3

Para el Ejemplo 3 obtuvimos que los tres modelos con menor valor del criterio Bayesiano para los procesos $TAR(2; p_1, p_2)$ y $TAR(3; p_1, p_2, p_3)$ fueron

Modelo	BIC	$\hat{\gamma}$	Modelo	BIC	$\hat{\gamma}_1$	$\hat{\gamma}_2$
$TAR(2; 2, 2)$	1028.251	1.011	$TAR(3; 2, 2, 2)$	1037.085	-0.352	0.99
$TAR(2; 2, 3)$	1029.941	1.011	$TAR(3; 4, 2, 2)$	1037.733	-0.400	0.99
$TAR(2; 3, 2)$	1030.759	1.011	$TAR(3; 3, 2, 2)$	1037.890	-0.400	0.99

Tabla 5.12: Resultados para el Ejemplo 3 usando el BIC

y en consecuencia, el mejor modelo para los datos según el criterio BIC fue un $TAR(2; 2, 2)$ y los parámetros se presentan en el siguiente cuadro

	β_{11}	β_{12}	β_{21}	β_{22}	ν	γ
Real	0.4	0.2	-0.3	0.4	5	1
Estimado	0.391	0.2209	-0.440	0.458	4.100	1.011

En la tabla (5.13) se encuentran los 3 modelos con menor valor del criterio Akaike para los procesos TAR (2; p_1, p_2) y TAR (3; p_1, p_2, p_3)

Modelo	AIC	$\hat{\gamma}$	Modelo	AIC	$\hat{\gamma}_1$	$\hat{\gamma}_2$
$TAR(2; 5, 2)$	1009.775	1.011	$TAR(3; 5, 2, 2)$	1009.385	-0.4	0.99
$TAR(2; 4, 2)$	1010.780	1.011	$TAR(3; 4, 2, 2)$	1009.660	-0.4	0.99
$TAR(2; 3, 2)$	1011.279	1.011	$TAR(3; 5, 2, 3)$	1011.000	-0.4	1.06

Tabla 5.13: Resultados para el Ejemplo 3 usando el AIC

y de acuerdo con lo anterior, el modelo seleccionado según el criterio AIC es un TAR (3; 5, 2, 2) con los parámetros dados en la próxima tabla

	β_{11}	β_{12}	β_{13}	β_{14}	β_{15}	β_{21}	β_{22}	β_{31}	β_{32}	ν	γ_1	γ_2
Real	0.4	0.2				-0.3	0.4			5	1	
Estimado	0.331	0.255	-0.060	-0.120	-0.020	0.459	0.286	-0.420	0.398	4.084	-0.4,	0.99

Ejemplo 4

Para el Ejemplo 4, los tres modelos con menor valor en cuanto al criterio BIC se encuentran en la tabla (5.14)

Modelo	BIC	$\hat{\gamma}$	Modelo	BIC	$\hat{\gamma}$	
$TAR(2; 2, 3)$	997.447	0.972	$TAR(3; 2, 2, 3)$	1008.237	-0.1	0.976
$TAR(2; 2, 4)$	998.812	0.972	$TAR(3; 2, 3, 2)$	1009.197	-0.1	0.976
$TAR(2; 3, 2)$	999.327	0.972	$TAR(3; 3, 2, 2)$	1015.317	-0.1	0.976

Tabla 5.14: Resultados para el Ejemplo 4 usando el BIC

El modelo estimado según el criterio Bayesiano fue un TAR (2; 2, 3) con los parámetros de la tabla siguiente

	β_{11}	β_{12}	β_{21}	β_{22}	β_{23}	ν	γ
Real	1.2	-0.4	0.1	-0.7		5	1
Estimado	1.179	-0.380	0.074	-0.565	0.052	5.172	0.972

Para el caso en el que se utilizó el criterio de Akaike se obtuvo que los 3 modelos que lo hacen mínimo se presentan a continuación

Modelo	AIC	$\hat{\gamma}$	Modelo	AIC	$\hat{\gamma}_1$	$\hat{\gamma}_2$
$TAR(2; 2, 5)$	978.063	0.972	$TAR(3; 2, 5, 3)$	981.834	-0.07	1
$TAR(2; 2, 4)$	978.118	0.972	$TAR(3; 2, 3, 3)$	982.2610	-0.07	1
$TAR(2; 5, 2)$	978.393	0.972	$TAR(3; 2, 2, 5)$	982.925	-0.10	0.976

Tabla 5.15: Resultados para el Ejemplo 4 usando el AIC

El modelo estimado según el criterio AIC es un $TAR(2; 2, 5)$ con los siguientes parámetros

	β_{11}	β_{12}	β_{21}	β_{22}	β_{23}	β_{24}	β_{25}	ν	γ
Real	1.2	-0.4	0.1	-0.7				5	1
Estimado	1.183	-0.383	0.070	-0.554	0.016	0.042	-0.008	5.144	0.972

Ejemplo 5

Para el Ejemplo 5 se obtuvo que los 3 modelos con menor valor en el criterio Bayesiano fueron

Modelo	BIC	γ	Modelo	BIC	γ	
$TAR(2; 3, 2)$	1000.510	0.0004	$TAR(3; 4, 3, 2)$	1001.598	-2.067	0.006
$TAR(2; 4, 2)$	1001.187	0.0004	$TAR(3; 5, 3, 2)$	1001.68	-2.067	0.006
$TAR(2; 5, 2)$	1003.754	0.0004	$TAR(3; 2, 3, 2)$	1004.185	-2.067	0.006

Tabla 5.16: Resultados para el Ejemplo 5 usando el BIC

y por lo cual se seleccionó un modelo $TAR(2; 3, 2)$ como aquel que ajusta mejor a los datos con los siguientes parámetros

	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}	ν	γ
Real	1.2	-0.4	-0.3	0.1	-0.7	5	0
Estimado	1.188	-0.437	-0.254	0.059	-0.701	4.191	0.0004

Mientras que los 3 modelos con menor valor del criterio AIC se presentan en la siguiente tabla

Modelo	AIC	$\hat{\gamma}$	Modelo	AIC	$\hat{\gamma}$	
$TAR(2; 5, 2)$	976.903	0.0004	$TAR(3; 5, 3, 2)$	979.186	-2.067	0.006
$TAR(2; 4, 2)$	977.270	0.0004	$TAR(3; 4, 3, 2)$	979.401	-2.067	0.006
$TAR(2; 3, 5)$	978.471	0.0004	$TAR(3; 2, 5, 2)$	980.266	-2.067	0.006

Tabla 5.17: Resultados para el Ejemplo 5 usando el AIC

y de acuerdo al criterio AIC se seleccionó un modelo $TAR(2; 5, 2)$ como aquel que ajusta mejor a los datos con los parámetros dados en la próxima tabla

	β_{11}	β_{12}	β_{13}	β_{14}	β_{15}	β_{21}	β_{22}	ν	γ
Real	1.2	-0.4	-0.3			0.1	-0.7	5	0
Estimado	1.195	-0.457	0.226	-0.023	0.011	0.060	-0.702	4.197	0.004

Como pudimos notar en los ejemplos de simulación anteriores, la metodología propuesta en este capítulo donde se combinan las ideas de Qian(1998) con el criterio de información bayesiana da buenos resultados para la selección del verdadero modelo generador de los datos del proceso TAR con errores t -student, mientras que el criterio de Akaike tiende a seleccionar modelos cuyos ordenes de los procesos autorregresivos son mayores a los del VPGD.

Capítulo 6

Pronósticos mediante verosimilitud predictiva

La idea central de este capítulo es estimar los pronósticos para h valores futuros dada una muestra de datos de un proceso TAR ($r; p_1, p_2, \dots, p_r$). Para realizar lo anterior haremos uso de la técnica de verosimilitud predictiva perfil, que nos permite obtener estimaciones de manera conjunta para valores futuros de la serie Y_t y de la variable umbral Z_t . Esta técnica se encuentra desarrollada para modelos TAR(2; 1, 1) en Russel(2006) y la cual extendaremos al caso general con r regímenes. Después, presentaremos las expresiones del error cuadrático medio de pronóstico (ECMP) de un TAR(2; 1, 1) que nos será de utilidad para medir las diferencias entre estimaciones de pronósticos al suponer una distribución t o normal en los errores de un modelo TAR cuyo verdadero proceso de errores tiene distribución t -student. Por último, presentaremos algunos ejemplos simulados de modelos TAR (2; 1, 1) y TAR (2; 2, 2) donde abordaremos diferentes escenarios como por ejemplo la presencia de datos atípicos, los cuales que podrían emular los comportamientos bruscos que presentan algunas series económicas. Para cada caso calcularemos los pronósticos para $h = 1, 2, 3$ pasos adelante y su respectivo ECMP obtenido utilizando lo desarrollado en el capítulo. A diferencia de

como se ha hecho a lo largo de este trabajo, primero se realizarán los pronósticos asumiendo un modelo TAR con errores normales y luego suponiendo que tienen errores con distribución t -student, con ello pretendemos primero ejemplificar la metodología para luego ver que ajustes son necesarios para el caso de errores independientes t -student.

Verosimilitud predictiva perfil

Comenzaremos dando las ideas de verosimilitud predictiva perfil presentadas en Bjørnstad(1990) para después aplicarla a los modelos TAR bajo el supuesto de errores normales y t .

Supongamos que contamos con una muestra \mathbf{x} de una variable aleatoria \mathbf{X} , y nuestro problema se centra en querer pronosticar valores \mathbf{x}^* de \mathbf{X}^* . Asumamos también que $\mathbf{W} = (\mathbf{X}, \mathbf{X}^*)'$ tiene una densidad de probabilidad la cual denotaremos por $f(\mathbf{x}, \mathbf{x}^*; \Theta)$ con Θ el vector de parámetros desconocidos, y $f(\mathbf{x}^*|\mathbf{x}; \Theta)$ representará la función de densidad condicional a los datos observados \mathbf{x} . Sea $\hat{\Theta}$ el estimador de máxima verosimilitud para Θ considerando los datos conocidos \mathbf{x} , y $\hat{\Theta}_w$ el estimador considerando en conjunto los datos conocidos y los datos a pronosticar, y que denotaremos por $\mathbf{w} = (\mathbf{x}, \mathbf{x}^*)$. En este caso podemos suponer que $\mathbf{X} = (Y_1, Y_2, \dots, Y_T)$ representa una muestra observada de tamaño T y $\mathbf{X}^* = (Y_{T+1}, Y_{T+2}, \dots, Y_{T+h})$ representa los datos no observados, los cuales deseamos predecir.

Berger y Wolpert(1984) formularon el principio de verosimilitud para predicción asumiendo que toda evidencia acerca de w está contenida en la función de verosimilitud conjunta

$$L(\mathbf{x}^*, \Theta; \mathbf{x}) = f_{\Theta}(\mathbf{x}, \mathbf{x}^*) \quad (6.0.1)$$

El objetivo es desarrollar una verosimilitud para \mathbf{x}^* , $L(\mathbf{x}^*|\mathbf{x})$, eliminando Θ

de (6.0.1). A $L(\mathbf{x}^*|\mathbf{x})$ se le conoce como verosimilitud predictiva. Hay muchas maneras de resolver el problema anterior, un ejemplo es el desarrollado por Mathiasen (1979) conocido como verosimilitud predictiva perfil y el cual consiste en eliminar Θ mediante la maximización de la siguiente función de verosimilitud.

$$L_p(\mathbf{x}^*|\mathbf{x}) = \sup_{\Theta} f(\mathbf{x}, \mathbf{x}^*; \Theta) = L(\mathbf{x}^*, \hat{\Theta}_w; \mathbf{x}). \quad (6.0.2)$$

Russel(2006) utiliza (6.0.2) para desarrollar los pronóstico en el modelo TAR(2; 1, 1) con errores normales y que detallaremos a continuación.

6.1. Verosimilitud predictiva perfil de un TAR(2; 1, 1)

Supongamos que se tienen T observaciones del proceso TAR (2; 1, 1) dado por

$$Y_t = \beta_1 Y_{t-1} I(Z_{t-1} \leq \gamma) + \beta_2 Y_{t-1} I(Z_{t-1} > \gamma) + \epsilon_t, \quad \epsilon_t \sim iidN(0, \sigma^2),$$

con variable umbral

$$Z_t = \rho Z_{t-1} + \psi_t, \quad \psi_t \sim iidN(0, \sigma_u^2).$$

Nuestro interés se enfoca en estimar las observaciones $T + 1, \dots, T + h$. El proceso de los h datos posteriores al tiempo T , lo podemos escribir como

$$Y_t^* = \beta_1 Y_{t-1}^* I(Z_{t-1}^* \leq \gamma) + \beta_2 Y_{t-1}^* I(Z_{t-1}^* > \gamma) + \epsilon_t, \quad \epsilon \sim iidN(0, \sigma^2),$$

para $t = T + 1, \dots, T + h$.

Primero, denotaremos a la muestra observada de Y_t y Z_t , respectivamente, por

$$\mathbf{y} = (y_1, y_2, \dots, y_T)', \quad \mathbf{z} = (z_0, z_1, \dots, z_{T-1})'.$$

A sus pronósticos a estimar por

$$\mathbf{y}^* = (y_{T+1}^*, y_{T+2}^*, \dots, y_{T+h}^*)', \quad \mathbf{z}^* = (z_{T+1}^*, \dots, z_{T+h}^*)'.$$

Mientras que $\Theta_Y = (\beta_1, \beta_2, \nu)$ representará a los parámetros de Y_t , y $\Theta_z = (\rho, \sigma_u^2)$ a los parámetros de Z_t .

Sea $\mathbf{w} = (\mathbf{y}, \mathbf{y}^*)$, $\mathbf{v} = (\mathbf{z}, \mathbf{z}^*)$ y definamos

$$\delta_t = \begin{cases} \beta_1, & v_{t-1} \leq \gamma \\ \beta_2, & v_{t-1} > \gamma \end{cases}.$$

Con la notación anterior podemos escribir el proceso TAR(2; 1, 1) como,

$$w_t = \delta_t w_{t-1} + \epsilon_t, \quad t = 1, 2, T + h.$$

$$v, \nu$$

Luego, la función de verosimilitud se expresa como

$$\begin{aligned} L(\Theta_Y, \Theta_Z, \gamma; \mathbf{w}, \mathbf{v}) &= \prod_{t=2}^{T+h} f(w_t | w_{t-1}; \Theta_Y, \Theta_Z, \gamma) g(v_{t-1} | v_{t-2}; \Theta_Z) \\ L(\Theta_Y, \Theta_Z, \gamma; \mathbf{w}, \mathbf{v}) &= \frac{1}{\sigma^{T+h-1}} \exp \left(-\frac{1}{2\sigma^2} \sum_{t=2}^{T+h} (w_t - \delta_t w_{t-1})^2 \right) \\ &\times \frac{1}{\sigma_u^{T+h-1}} \exp \left(-\frac{1}{2\sigma_u^2} \sum_{t=2}^{T+h} (v_t - \rho v_{t-1})^2 \right). \end{aligned}$$

Tomando logaritmo en la ecuación anterior obtenemos la función de log-verosimilitud:

$$\begin{aligned}\ell(\Theta_Y, \Theta_Z, \gamma; \mathbf{w}, \mathbf{v}) &= -(T+h-1)\log(\sigma) - \frac{1}{2\sigma^2} \sum_{t=2}^{T+h} (w_t - \delta_t w_{t-1})^2 \\ &\quad - (T+h-1)\log(\sigma_u) - \frac{1}{2\sigma_u^2} \sum_{t=2}^{T+h} (v_t - \rho v_{t-1})^2.\end{aligned}$$

Derivando de ℓ con respecto de σ obtenemos que

$$\begin{aligned}\hat{\sigma}^2(\mathbf{y}^*; \beta_1, \beta_2, \gamma) &= \frac{1}{T+h-1} \sum_{t=2}^{T+h} (w_t - [\beta_1 I(v_{t-1} \leq \gamma) + \beta_2 I(v_{t-1} > \gamma)] w_{t-1})^2 \\ &= \frac{1}{T+h-1} \sum_{t=2}^{T+h} (w_t - \delta_t w_{t-1})^2.\end{aligned}\tag{6.1.1}$$

Luego, derivando ℓ con respecto a β_1, β_2 e igualando a cero encontramos que sus estimadores máximos verosímiles son

$$\hat{\beta}_1(\mathbf{y}^*; \gamma) = \frac{\sum_{t=2}^{T+h} w_t I(v_{t-1} \leq \gamma) w_{t-1}}{\sum_{t=2}^{T+h} I(v_{t-1} \leq \gamma) w_{t-1}^2}.\tag{6.1.2}$$

$$\hat{\beta}_2(\mathbf{y}^*; \gamma) = \frac{\sum_{t=2}^{T+h} w_t I(v_{t-1} > \gamma) w_{t-1}}{\sum_{t=2}^{T+h} I(v_{t-1} > \gamma) w_{t-1}^2}.\tag{6.1.3}$$

De la misma manera obtenemos que

$$\hat{\rho}(\mathbf{z}^*) = \frac{\sum_{t=2}^{T+h} v_{t-1} v_{t-1}}{\sum_{t=2}^{T+h} v_{t-2}^2}$$

$$\hat{\sigma}_u^2(\mathbf{z}^*; \hat{\rho}) = \frac{1}{T+h-1} \sum_{t=2}^{T+h} (v_{t-1} - \hat{\rho} v_{t-2})^2$$

Sustituyendo (6.1.2) y (6.1.3) en (6.1.1) se llega a lo siguiente

$$\hat{\sigma}^2(\mathbf{y}^*; \gamma) = \frac{1}{T+h-1} \sum_{t=2}^{T+h} \left(w_t - \left[\hat{\beta}_1 I(v_{t-1} \leq \gamma) + \hat{\beta}_2 I(v_{t-1} > \gamma) \right] w_{t-1} \right)^2.$$

Posteriormente, la logverosimilitud predictiva perfil para $(\mathbf{y}^*, \mathbf{z}^*)$ sustituyendo los estimadores anteriores se reduce a

$$\begin{aligned} \ell_p(\mathbf{y}^*, \mathbf{z}^*, \gamma | \mathbf{y}, \mathbf{z}) &= -(T+h-1) \log \left(\frac{1}{T+h-1} \sum_{t=2}^{T+h} (w_t - \delta_t w_{t-1})^2 \right) \\ &\quad - (T+h-1) \log \left(\frac{1}{T+h-1} \sum_{t=2}^{T+h} (v_t - \rho v_{t-1})^2 \right) \\ &\quad - (T+h-1). \end{aligned}$$

donde

$$\hat{\delta}_t(\mathbf{y}^*; \gamma) = \begin{cases} \hat{\beta}_1(\mathbf{y}^*; \gamma) & v_{t-1} \leq \gamma \\ \hat{\beta}_2(\mathbf{y}^*; \gamma) & v_{t-1} > \gamma \end{cases}$$

Notemos que el último término $T+h-1$ permanece constante por lo cual maximizar ℓ_p es equivalente a maximizar la siguiente expresión

$$\begin{aligned}
& - (T + h - 1) \log \left(\frac{1}{T + h - 1} \sum_{t=2}^{T+h} \left(w_t - \hat{\delta}_t(\mathbf{y}^*; \gamma) w_{t-1} \right)^2 \right) \\
& - (T + h - 1) \log \left(\frac{1}{T + h - 1} \sum_{t=2}^{T+h} (v_t - \hat{\rho}(\mathbf{z}^*) v_{t-1})^2 \right).
\end{aligned}$$

Que a su vez es equivalente a minimizar la siguiente función respecto a $(\mathbf{y}^*, \mathbf{z}^*)$

$$\begin{aligned}
& (T + h - 1) \log \left(\frac{1}{T + h - 1} \sum_{t=2}^{T+h} \left(w_t - \hat{\delta}_t(\mathbf{y}^*; \gamma) w_{t-1} \right)^2 \right) \\
& + (T + h - 1) \log \left(\frac{1}{T + h - 1} \sum_{t=2}^{T+h} (v_t - \hat{\rho}(\mathbf{z}^*) v_{t-1})^2 \right).
\end{aligned}$$

Ahora para el caso t - student, realizando pasos análogos a los de la capítulo 4, se llega que la función de verosimilitud predictiva tiene la siguiente forma

$$\begin{aligned}
L(\boldsymbol{\Theta}_Y, \boldsymbol{\Theta}_Z, \gamma; \mathbf{y}, \mathbf{y}^*, \mathbf{z}, \mathbf{z}^*) &= \prod_{t=2}^{T+h} f(w_t | w_{t-1}, Z_{t-1}; \boldsymbol{\Theta}_Y, \gamma) g(v_{t-1} | v_{t-2}, \boldsymbol{\Theta}_Z) \\
&= \prod_{t=2}^{T+h} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2}) \sqrt{\nu\pi}} \left(1 + \frac{(w_t - \delta_t w_{t-1})^2}{\nu} \right)^{-\frac{\nu+1}{2}} \\
&\times \frac{1}{\sigma_u^{T-1}} \exp \left(-\frac{1}{\sigma_u^2} \sum_{t=2}^{T+h} (v_{t-1} - \rho v_{t-2})^2 \right)
\end{aligned}$$

Mientras que la función de log-verosimilitud es

$$\begin{aligned}
\ell(\Theta_Y, \Theta_Z, \gamma; \mathbf{y}, \mathbf{y}^*, \mathbf{z}, \mathbf{z}^*) &= \sum_{t=2}^{T+h} \log \left(\frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2}) \sqrt{\nu\pi}} \right) \\
&+ \sum_{t=2}^{T+h} \left(-\frac{\nu+1}{2} \right) \log \left(1 + \frac{(w_t - \delta_t w_{t-1})^2}{\nu} \right) \\
&- (T-1) \log(\sigma_u) - \frac{\sum_{t=2}^{T+h} (v_{t-1} - \rho v_{t-2})^2}{\sigma_u^2}
\end{aligned} \tag{6.1.4}$$

En este caso no podemos realizar el mismo procedimiento analítico que se hace cuando suponemos errores normales por lo que se maximizará la log verosimilitud predictiva (6.1.4) de manera numérica para encontrar los estimadores $\hat{\beta}_1(\mathbf{y}^*; \gamma)$, $\hat{\beta}_2(\mathbf{y}^*; \gamma)$, $\hat{\nu}(\mathbf{y}^*; \beta_1, \beta_2, \gamma)$, $\hat{\rho}(\mathbf{z}^*)$ y $\hat{\sigma}_u^2(\mathbf{z}^*; \hat{\rho})$. Después se sustituirán en (6.1.4) para obtener la log verosimilitud predictiva perfil y de nueva cuenta se maximizará numéricamente con respecto a \mathbf{y}^* , \mathbf{z}^* para así obtener los estimadores de los pronósticos a h pasos.

6.2. Verosimilitud predictiva perfil de un modelo TAR($r; p_1, \dots, p_r$)

Ahora calcularemos la verosimilitud predictiva perfil para el caso en el cual tenemos un modelo TAR con r regímenes. Recordemos que la expresión para el proceso TAR($r; p_1, p_2, \dots, p_r$) es la siguiente

$$Y_t = I(Z_{t-1} \leq \gamma_1) \left(\sum_{i=1}^{p_1} \beta_{1i} Y_{t-i} \right) + \dots + I(\gamma_{r-1} < Z_{t-1} \leq \gamma_r) \left(\sum_{i=1}^{p_r} \beta_{ri} Y_{t-i} \right) + \epsilon_t,$$

con variable umbral

$$Z_t = \rho Z_{t-1} + \psi_t, \quad \psi_t \sim iidN(0, \sigma_u^2).$$

Usaremos la misma notación que en la sección anterior, es decir, denotaremos a la muestra observada de Y_t y Z_t , respectivamente, por

$$\mathbf{y} = (y_1, y_2, \dots, y_T)', \quad \mathbf{z} = (z_0, z_1, \dots, z_{T-1})'.$$

A sus pronósticos a estimar por

$$\mathbf{y}^* = (y_{T+1}^*, y_{T+2}^*, \dots, y_{T+h}^*)', \quad \mathbf{z}^* = (z_T^*, \dots, z_{T+h-1}^*)'.$$

Por último, por $\boldsymbol{\Theta}_Y = (\beta_{11}, \dots, \beta_{rp_r}, \nu)$ a los parámetros de Y_t , a los parámetros de Z_t por $\boldsymbol{\Theta}_Z = (\rho, \sigma_u^2)$ y a los valores umbrales por $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \dots, \gamma_{r-1})$.

Sea $\mathbf{w} = (\mathbf{y}, \mathbf{y}^*)$, $\mathbf{v} = (\mathbf{z}, \mathbf{z}^*)$, con la notación anterior, y usando las ideas del capítulo 4 podemos escribir la verosimilitud predictiva del proceso TAR $(r; p_1, p_2, \dots, p_r)$ como

$$\begin{aligned} L(\boldsymbol{\Theta}_Y, \boldsymbol{\Theta}_Z, \boldsymbol{\gamma}; \mathbf{y}, \mathbf{y}^*, \mathbf{z}, \mathbf{z}^*) &= \prod_{t=2}^{T+h} f(w_t | w_{t-1}, Z_{t-1}, \boldsymbol{\Theta}_Z) g(v_{t-1} | v_{t-2}, \boldsymbol{\Theta}_Z) \\ &= \prod_{t=2}^{T+h} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2}) \sqrt{\nu\pi}} \left(1 + \frac{(X_t)^2}{\nu}\right)^{-\frac{\nu+1}{2}} \\ &\times \frac{1}{\sigma^{T-1}} \exp\left(-\frac{1}{\sigma^2} \sum_{t=2}^{T+h} (v_{t-1} - \rho v_{t-2})^2\right) \end{aligned} \quad (6.2.1)$$

donde

$$X_t = w_t - \left[I(v_{t-1} \leq \gamma_1) \left(\sum_{i=1}^{p_1} \beta_{1i} w_{t-i} \right) + \dots + I(\gamma_{r-1} \leq v_{t-1}) \left(\sum_{i=1}^{p_r} \beta_{ri} Y_{t-i} \right) \right].$$

Tomando logaritmo de (6.2.1) obtenemos la función de logverosimilitud siguiente

$$\begin{aligned}
\ell(\boldsymbol{\Theta}_Y, \boldsymbol{\Theta}_Z, \gamma; \mathbf{y}, \mathbf{y}^*, \mathbf{z}, \mathbf{z}^*) &= \sum_{t=2}^{T+h} \log \left(\frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2}) \sqrt{\nu\pi}} \right) \\
&+ \sum_{t=2}^{T+h} \left(-\frac{\nu+1}{2} \right) \log \left(1 + \frac{(X_t)^2}{\nu} \right) \\
&- (T-1) \log(\sigma_u) - \frac{1}{\sigma_u^2} \sum_{t=2}^{T+h} (v_{t-1} - \rho v_{t-2})^2 \quad (6.2.2)
\end{aligned}$$

Como en el caso del TAR(2;1,1) maximizaremos de manera numérica (6.2.2) para encontrar los estimadores de los parámetros. Posteriormente los sustituiremos los parámetros obtenidos en (6.2.2) para formar la función de verosimilitud predictiva perfil $\ell_p(\mathbf{y}^*, \mathbf{z}^*, \gamma|\mathbf{y}, \mathbf{z})$ y de nueva cuenta se maximizará numéricamente dicha función con respecto a $\mathbf{y}^*, \mathbf{z}^*$ para obtener los pronósticos para h pasos adelante, el cual era nuestro objetivo.

6.3. ECMP

El error cuadrático medio de pronóstico se define como

$$\begin{aligned}
ECMP_h &= E \left(Y_{t+h} - \hat{Y}_{t+h} \right)^2 \\
&= E \left((Y_{t+h} - E(Y_{t+h}|\mathcal{F}_t)) \right)^2 \quad (6.3.1)
\end{aligned}$$

Para el cálculo del ECMP del modelo TAR(2;1,1) necesitamos calcular la varianza, la función de covarianzas del proceso Y_t y el pronóstico a h pasos que está dado por $E(Y_{t+h}|\mathcal{F}_t)$.

Comenzaremos con el cálculo de la varianza de Y_t asumiendo que $Y_0 = 0$.

Sabemos que

$$Y_t = [\beta_1 I(Z_{t-1} \leq \gamma) + \beta_2 I(Z_{t1} > \gamma)] Y_{t-1} + \epsilon_t$$

y definimos

$$\delta_t = \beta_1 I(Z_{t-1} \leq \gamma) + \beta_2 I(Z_{t1} > \gamma)$$

por lo que

$$Y_t = \delta_t Y_{t-1} + \epsilon_t$$

Ahora,

$$\begin{aligned} \text{Var}(Y_t) &= \text{Var}(\delta_t Y_{t-1}) + \text{Var}(\epsilon_t) \\ &= \text{Var}(\delta_t Y_{t-1}) + \sigma^2 \\ &= \text{Var}(\delta_t (\delta_{t-1} Y_{t-2} + \epsilon_{t-1})) + \sigma^2 \\ &= \text{Var}(\delta_t \delta_{t-1} Y_{t-2} + \delta_t \epsilon_{t-1}) + \sigma^2 \\ &= \text{Var}(\delta_t \delta_{t-1} Y_{t-2}) + \text{Var}(\delta_t \epsilon_{t-1}) + \sigma^2 \\ &= \text{Var}(\delta_t \delta_{t-1} (\delta_{t-2} Y_{t-3} + \epsilon_{t-2})) + \text{Var}(\delta_t \epsilon_{t-1}) + \sigma^2 \\ &= \text{Var}(\delta_t \delta_{t-1} \delta_{t-2} Y_{t-3}) + \text{Var}(\delta_t \delta_{t-1} \epsilon_{t-2}) + \text{Var}(\delta_t \epsilon_{t-1}) + \sigma^2 \\ &= \text{Var}(\delta_t \delta_{t-1} \dots \delta_1 Y_0) + \text{Var}(\delta_t \delta_{t-1} \dots \delta_2 \epsilon_1) + \dots + \text{Var}(\delta_t \epsilon_{t-1}) + \sigma^2 \\ &= \text{Var}(\delta_t \delta_{t-1} \dots \delta_2 \epsilon_1) + \dots + \text{Var}(\delta_t \epsilon_{t-1}) + \sigma^2 \\ &= E((\delta_t \delta_{t-1} \dots \delta_2 \epsilon_1)^2) + \dots + E((\delta_t \epsilon_{t-1})^2) + \sigma^2 \\ &= E(\delta_t^2 \delta_{t-1}^2 \dots \delta_2^2) E(\epsilon_1^2) + \dots + E(\delta_t^2) E(\epsilon_{t-1}^2) + \sigma^2 \\ &= \sigma^2 [E(\delta_t^2 \delta_{t-1}^2 \dots \delta_2^2) + \dots + E(\delta_t^2) + 1] \end{aligned}$$

Calculando la función de covarianzas

$$\begin{aligned}
Cov(Y_t, Y_{t-s}) &= E(Y_t Y_{t-s}) - E(Y_t) E(Y_{t-s}) \\
&= E(Y_t Y_{t-s}) \\
&= E((\delta_t Y_{t-1} + \epsilon_t) Y_{t-s}) \\
&= E(\delta_t Y_{t-1} Y_{t-s}) + E(\epsilon_t Y_{t-s}) \\
&\vdots \\
&= E(\delta_t \dots \delta_{t-s+1} Y_{t-s} Y_{t-s}) \\
&= E(\delta_t \dots \delta_{t-s+1} Y_{t-s}^2)
\end{aligned}$$

$$\begin{aligned}
Y_{t-s}^2 &= (\delta_{t-s} Y_{t-s-1} + \epsilon_{t-s})^2 \\
&= (\delta_{t-s} (\delta_{t-s-1} Y_{t-s-2} + \epsilon_{t-s-1}) + \epsilon_{t-s})^2 \\
&= (\delta_{t-s} \delta_{t-s-1} Y_{t-s-2} + \delta_{t-s} \epsilon_{t-s-1} + \epsilon_{t-s})^2 \\
&= (\delta_{t-s} \delta_{t-s-1} (\delta_{t-s-2} Y_{t-s-3} + \epsilon_{t-s-2}) + \delta_{t-s} \epsilon_{t-s-1} + \epsilon_{t-s})^2 \\
&= (\delta_{t-s} \delta_{t-s-1} \delta_{t-s-2} Y_{t-s-3} + \delta_{t-s} \delta_{t-s-1} \epsilon_{t-s-2} + \delta_{t-s} \epsilon_{t-s-1} + \epsilon_{t-s})^2 \\
&\vdots \\
&= (\delta_{t-s} \delta_{t-s-1} \delta_{t-s-2} \dots \delta_1 Y_0 + \dots \delta_{t-s} \delta_{t-s-1} \epsilon_{t-s-2} + \delta_{t-s} \epsilon_{t-s-1} + \epsilon_{t-s})^2
\end{aligned}$$

El pronóstico de Y_t para h pasos adelante se obtiene por

$$\begin{aligned}
E(Y_{t+h}|\mathcal{F}_t) &= E(\delta_{t+h}Y_{t+h-1} + \epsilon_{t+h}|\mathcal{F}_t) \\
&= E(\delta_{t+h}Y_{t+h-1}|\mathcal{F}_t) + E(\epsilon_{t+h}|\mathcal{F}_t) \\
&= E(\delta_{t+h}Y_{t+h-1}|\mathcal{F}_t) \\
&= E(\delta_{t+h}(\delta_{t+h-1}Y_{t+h-2} + \epsilon_{t+h-1})|\mathcal{F}_t) \\
&= E(\delta_{t+h}\delta_{t+h-1}Y_{t+h-2}|\mathcal{F}_t) \\
&\vdots \\
&= E(\delta_{t+h}\delta_{t+h-1}\dots\delta_{t+1}Y_t|\mathcal{F}_t) \\
&= Y_tE(\delta_{t+h}\delta_{t+h-1}\dots\delta_{t+1})
\end{aligned}$$

Ahora el ECMP es

$$\begin{aligned}
ECMP_h &= E((Y_{t+h} - Y_tE(\delta_{t+h}\delta_{t+h-1}\dots\delta_{t+1})))^2 \\
&= E((\delta_{t+h}Y_{t+h-1} + \epsilon_{t+h} - Y_tE(\delta_{t+h}\delta_{t+h-1}\dots\delta_{t+1}))^2) \\
&= E(\delta_{t+h}^2Y_{t+h-1}^2) + E(\epsilon_{t+h}^2) + E(Y_t^2(E(\delta_{t+h}\delta_{t+h-1}\dots\delta_{t+1}))^2) \\
&+ 2E(\delta_{t+h}Y_{t+h-1}\epsilon_{t+h}) - 2E(\epsilon_{t+h}Y_tE(\delta_{t+h}\delta_{t+h-1}\dots\delta_{t+1})) \\
&- 2E(\delta_{t+h}Y_{t+h-1}Y_tE(\delta_{t+h}\delta_{t+h-1}\dots\delta_{t+1})) \\
&= Var(Y_{t+h}) + (E(\delta_{t+h}\delta_{t+h-1}\dots\delta_{t+1}))^2 Var(Y_t) \\
&- 2E(\epsilon_{t+h}Y_t)E((\delta_{t+h}\delta_{t+h-1}\dots\delta_{t+1})) - 2E(\delta_{t+h}Y_{t+h-1}Y_t)E(\delta_{t+h}\delta_{t+h-1}\dots\delta_{t+1}) \\
&= Var(Y_{t+h}) + (E(\delta_{t+h}\delta_{t+h-1}\dots\delta_{t+1}))^2 Var(Y_t) \\
&- 2E(\delta_{t+h}Y_{t+h-1}Y_t)E(\delta_{t+h}\delta_{t+h-1}\dots\delta_{t+1}) \\
&= Var(Y_{t+h}) + (E(\delta_{t+h}\delta_{t+h-1}\dots\delta_{t+1}))^2 Var(Y_t) \\
&- 2Cov(Y_{t+h}, Y_t)E(\delta_{t+h}\delta_{t+h-1}\dots\delta_{t+1})
\end{aligned}$$

donde las expresiones para la varianza y covarianza fueron calculadas anteriormente.

6.4. Simulaciones

En esta sección presentaremos 8 ejemplos con datos simulados de procesos TAR(2; p_1, p_2), donde para cada uno simulamos 500 series de 253 datos. En este estudio de simulación nos centraremos en la estimación de los pronósticos para $h = 1, 2, 3$ pasos adelante, mediante la técnica de verosimilitud predictiva perfil presentada en las secciones anteriores. Adicionalmente, para cada serie se calculará su respectivo error cuadrático medio de pronóstico para cada uno de los horizontes $h = 1, 2, 3$ y por último se promediarán los 500 ECMP obtenidos. En cada ejemplo suponemos que el valor umbral y los órdenes de los procesos autorregresivos son conocidos.

El algoritmo de simulación utilizado para los pronósticos es el siguiente

- Para cada serie Y_t se tomaron las primeras 250 observaciones para estimar los parámetros del modelo y hacer los pronósticos mediante verosimilitud predictiva. Los últimos 3 datos se usaron para el cálculo del error cuadrático medio de pronóstico.
- Se calculó la función de verosimilitud dada en (6.2.2), considerando fijos γ, p_1, p_2 , y se maximizó respecto a los parámetros $\Theta_Y = (\beta_{11}, \dots, \beta_{2p_2}, \nu)$, $\Theta_Z = (\rho, \sigma_u^2)$.
- Los resultados obtenidos en el paso previo se sustituyeron en (6.2.2) y se maximizó la función de verosimilitud predictiva perfil con respecto a

$$\mathbf{y}^* = (y_{T+1}, y_{T+2}, y_{T+3}), \text{ y } \mathbf{z}^* = (z_{T+1}, z_{T+1}, z_{T+3}).$$

- Una vez obtenidos los pronósticos de $\mathbf{y}^*, \mathbf{z}^*$ se calcula el ECMP que tiene por expresión $ECMP = (y_{observado} - y_{pronosticado})^2$.

Como complemento, para cada una de las series simuladas se realizó el procedimiento anterior con la variante de suponer los errores del verdadero proceso generador de los datos, Y_t , tienen distribución normal.

Los ejemplos considerados en el ejercicio de simulación se encuentran en las tablas (6.1), (6.2) y (6.3), en todos ellos asumiremos que la variable umbral sigue un proceso autorregresivo de orden 1 con la siguiente expresión,

$$Z_t = 0.5Z_{t-1} + \psi_t, \psi_t \sim iidN(0, 1).$$

Etiqueta	β_1	β_2	γ	ν	Etiqueta	β_1	β_2	γ	ν
TAR 1	0.5	-0.7	0	5	TAR 3	0.95	-0.95	0	5
	0.5	-0.7	0	4		0.95	-0.95	0	4
	0.5	-0.7	0.897	5		0.95	-0.95	0.897	5
	0.5	-0.7	0.897	4		0.95	-0.95	0.897	4
TAR 2	-0.8	-0.6	0	5	TAR 4	-0.5	0.5	0	5
	-0.8	-0.6	0	4		-0.5	0.5	0	4
	-0.8	-0.6	0.897	5		-0.5	0.5	0.897	5
	-0.8	-0.6	0.897	4		-0.5	0.5	0.897	4

Tabla 6.1: Ejemplos considerados para un modelo $TAR(2, 1, 1)$

Etiqueta	β_1	β_2	γ	ν	Etiqueta	β_1	β_2	γ	ν
TAR 5	0.8	-0.2	0	5	TAR 6	0.7	-0.7	0	5
	0.8	-0.2	0.897	4		0.7	-0.7	0	4
	0.8	-0.2	0	5		0.7	-0.7	0.897	5
	0.8	-0.2	0.897	4		0.7	-0.7	0.897	4

Tabla 6.2: Ejemplos considerados para un modelo $TAR(2; 1, 1)$ con datos atípicos

Et.	β_{11}	β_{12}	β_{21}	β_{22}	γ	ν	Et.	β_{11}	β_{12}	β_{21}	β_{22}	γ	ν
TAR 7	1.2	-0.4	1.6	-0.8	0	5	TAR 8	0.3	-0.5	0.6	-0.7	0	5
	1.2	-0.4	1.6	-0.8	0	4		0.3	-0.5	0.6	-0.7	0	4
	1.2	-0.4	1.6	-0.8	0.897	5		0.3	-0.5	0.6	-0.7	0.897	5
	1.2	-0.4	1.6	-0.8	0.897	4		0.3	-0.5	0.6	-0.7	0.897	4

Tabla 6.3: Ejemplos considerados para un modelo $TAR(2; 2, 2)$

Como podemos notar, para cada ejemplo manejamos 4 diversos escenarios: el porcentaje de observaciones en régimen inferior igual a 50 y 75, y grados de libertad iguales a 4 y 5. Además, en los ejemplos TAR5 y TAR6 se presentan datos simulados con datos atípicos para emular movimientos bruscos de las series. Los resultados obtenidos para cada ejemplo se presentan a continuación.

TAR 1

Se simularon 500 series del modelo TAR 1 y algunas gráficas de las realizaciones del proceso Y_t bajo los 4 escenarios abordados se encuentran en la gráfica siguiente

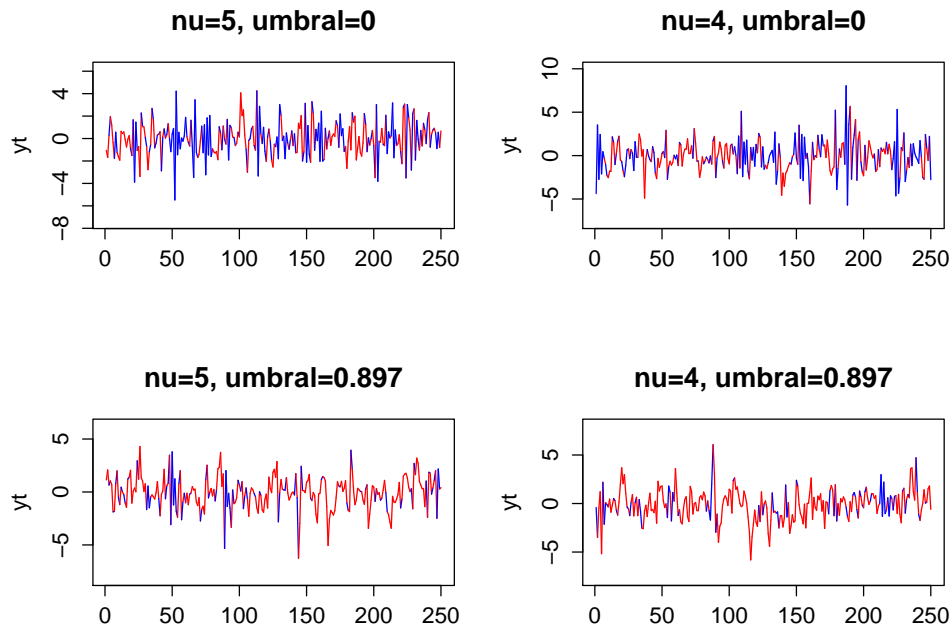


Figura 6.4.1: Realizaciones de Y_t .

Los resultados obtenidos para la estimación de los parámetros de cada uno de los modelos se resumen en las tablas (6.4), (6.5), (6.6), (6.7).

Caso $\gamma = 0$		Modelo t -student		Modelo Normal	
Parámetro	Real	Promedio	D. E.	Promedio	D. E.
β_1	0.5	0.495	0.065	0.493	0.072
β_2	-0.7	-0.695	0.062	-0.694	0.069
ν	5	5.518	1.739	-	-
σ	1.291	1.291	0.100	1.277	0.104

Tabla 6.4: Resultados para los parámetros del modelo con $\nu = 5$.

Caso $\gamma = 0.897$		Modelo t -student		Modelo Normal	
Parámetro	Real	Promedio	D. E.	Promedio	D. E.
β_1	0.5	0.499	0.053	0.498	0.060
β_2	-0.7	-0.693	0.116	-0.692	0.121
ν	5	5.496	1.737	-	-
σ	1.291	1.291	0.099	1.276	0.095

Tabla 6.5: Resultados para los parámetros del modelo con $\nu = 5$.

Caso $\gamma = 0$		Modelo t -student		Modelo Normal	
Parámetro	Real	Promedio	D. E.	Promedio	D. E.
β_1	0.5	0.497	0.060	0.494	0.071
β_2	-0.7	-0.694	0.066	-0.694	0.072
ν	4	4.328	1.033	-	-
σ	1.414	1.408	0.135	1.387	0.156

Tabla 6.6: Resultados para los parámetros del modelo con $\nu = 4$.

Caso $\gamma = 0.897$		Modelo t -student		Modelo Normal	
Parámetro	Real	Promedio	D. E.	Promedio	D. E.
β_1	0.5	0.494	0.052	0.492	0.061
β_2	-0.7	-0.700	0.102	-0.695	0.119
	4	4.332	0.953	-	-
σ	1.414	1.409	0.143	1.386	0.156

Tabla 6.7: Resultados para los parámetros del modelo con $\nu = 4$.

Podemos observar que en todos los escenarios considerados las estimaciones para cada uno de los parámetros son prácticamente los mismos.

Ahora, en la figura (6.4.2) se presentan algunos de los pronósticos simulados, mientras que en la gráfica (6.4.3) se encuentran los promedios de los pronósticos para horizonte considerado.

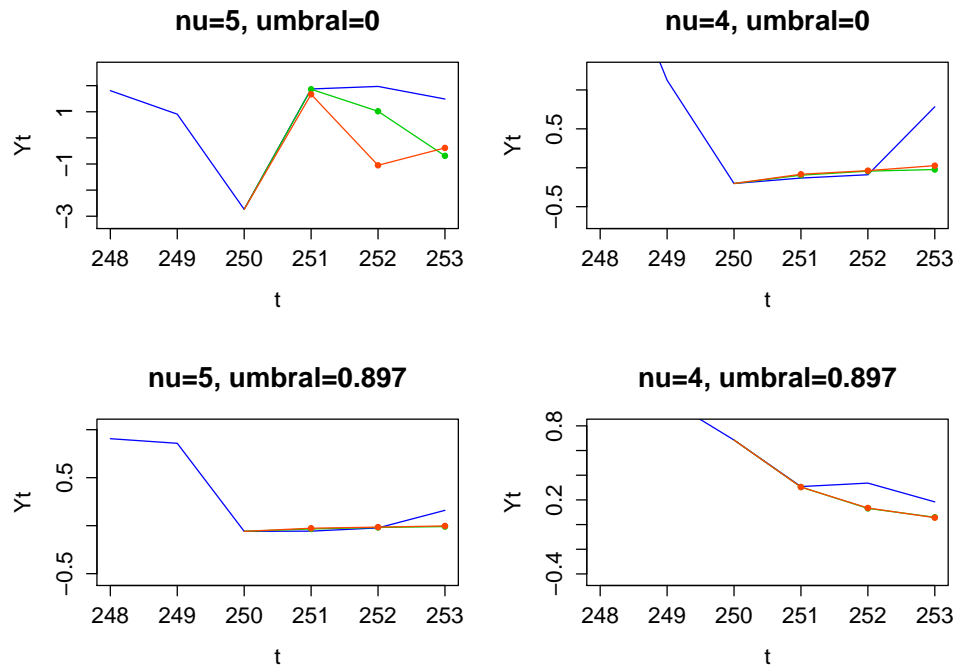


Figura 6.4.2: En color verde los pronósticos bajo el modelo t y en rojo los del modelo normal

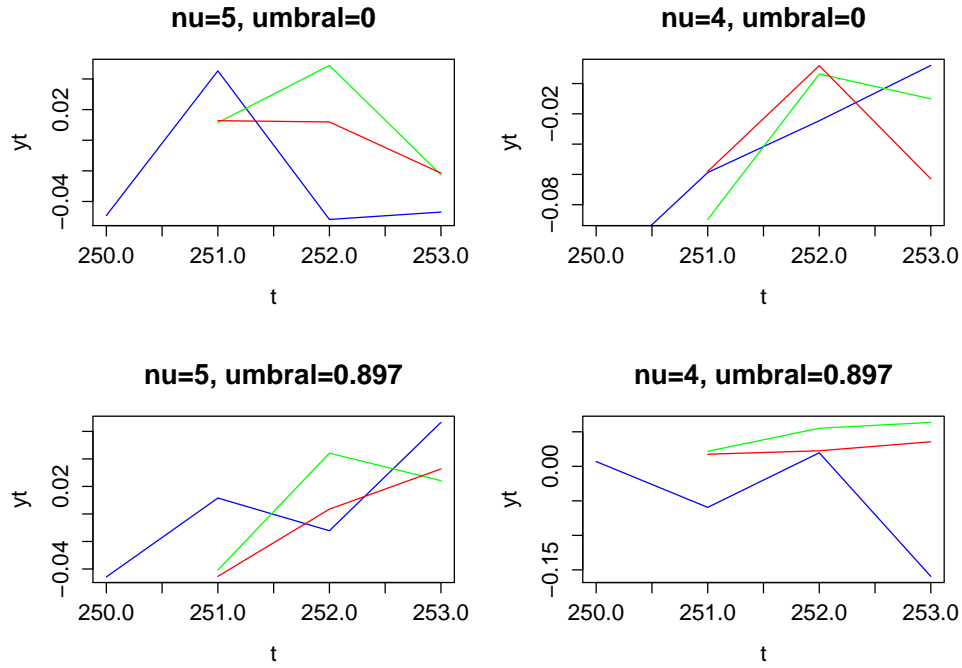


Figura 6.4.3: Promedio de los pronósticos

Por último, se calculó el error cuadrático medio de pronósticos (ECMP) para los pronósticos del modelo t y normal. Los cocientes entre ellos se encuentran en las próximas dos tablas.

Caso $\gamma = 0$	$h = 1$	$h = 2$	$h = 3$	Caso $\gamma = 0.897$	$h = 1$	$h = 2$	$h = 3$
ecmpt/ecmpn	1.643	0.919	1.063	ecmpt/ecmpn	0.998	1.010	0.988

Tabla 6.8: Cocientes de los ECMP para el caso $\nu = 5$

Caso $\gamma = 0$	$h = 1$	$h = 2$	$h = 3$	Caso $\gamma = 0.897$	$h = 1$	$h = 2$	$h = 3$
ecmpt/ecmpn	1.007	1.096	0.958	ecmpt/ecmpn	1.079	1.036	1.445

Tabla 6.9: Cocientes de los ECMP para el caso $\nu = 4$

Podemos notar que en la mayoría de los casos el modelo normal presenta un mejor rendimiento sobre el modelo t , aunque la diferencia es mínima.

TAR 2

En el ejemplo TAR 2 consideramos un modelo en el cual los procesos autorregresivos de cada régimen presentan el mismo signo. Una simulación para cada uno de los 4 escenarios contemplados se encuentran en la figura (6.4.4).

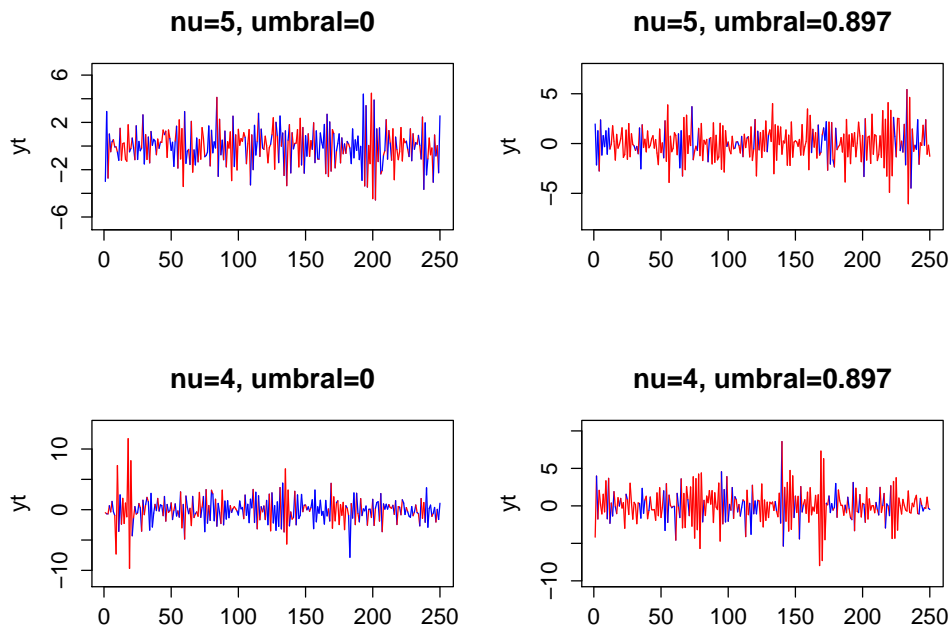


Figura 6.4.4: Realizaciones del procesos Y_t

Los resultados de los promedios y varianzas muestrales para cada uno de los parámetros se encuentra en las tablas (6.10), (6.11), (6.12), (6.13).

Caso $\gamma = 0$		Modelo t -student		Modelo Normal	
Parámetro	Real	Promedio	D. E.	Promedio	Varianza
β_1	-0.8	-0.791	0.059	-0.791	0.066
β_2	-0.6	-0.592	0.059	-0.592	0.066
ν	5	5.756	2.995	-	-
σ	1.291	1.288	0.112	1.274	0.111

Tabla 6.10: Resultados para el modelo con $\nu = 5$.

Caso $\gamma = 0.897$		Modelo t -student		Modelo Normal	
Parámetro	Real	Promedio	D. E.	Promedio	D. E.
β_1	-0.8	-0.792	0.042	-0.792	0.048
β_2	-0.6	-0.585	0.091	-0.586	0.095
ν	5	5.567	1.740	-	-
σ	1.291	1.287	0.100	1.274	0.099

Tabla 6.11: Resultados para el modelo $\nu = 5$

Caso $\gamma = 0$		Modelo t -student		Modelo Normal	
Parámetro	Real	Promedio	D. E.	Promedio	D. E.
β_1	-0.8	-0.794	0.053	-0.793	0.065
β_2	-0.6	-0.598	0.051	-0.597	0.058
ν	4	4.294	1.032	-	-
σ	1.414	1.418	0.150	1.387	0.146

Tabla 6.12: Resultados para el modelo $\nu = 4$

$\gamma = 0.897$		Modelo t -student		Modelo Normal	
Parámetro	Real	Promedio	D. E.	Promedio	D. E.
β_1	-0.8	-0.794	0.041	-0.793	0.046
β_2	-0.6	-0.593	0.076	-0.592	0.090
ν	4	4.301	1.076	-	-
σ	1.414	1.423	0.157	1.392	0.148

Tabla 6.13: Resultados para el modelo $\nu = 4$

De la misma manera que en el ejemplo anterior, notamos que las diferencias en las medias de la estimaciones de los parámetros es casi nula.

Alguno de los pronósticos realizados se encuentran a continuación

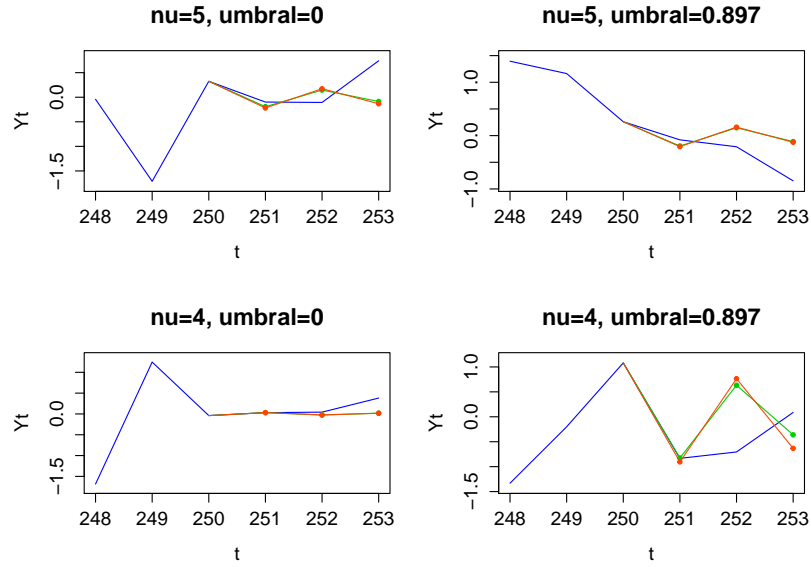


Figura 6.4.5: En verde pronósticos bajo el modelo t y en rojo bajo el modelo normal

Mientras que en la figura (6.4.6) están los promedios para cada uno de los horizontes de lo pronósticos.

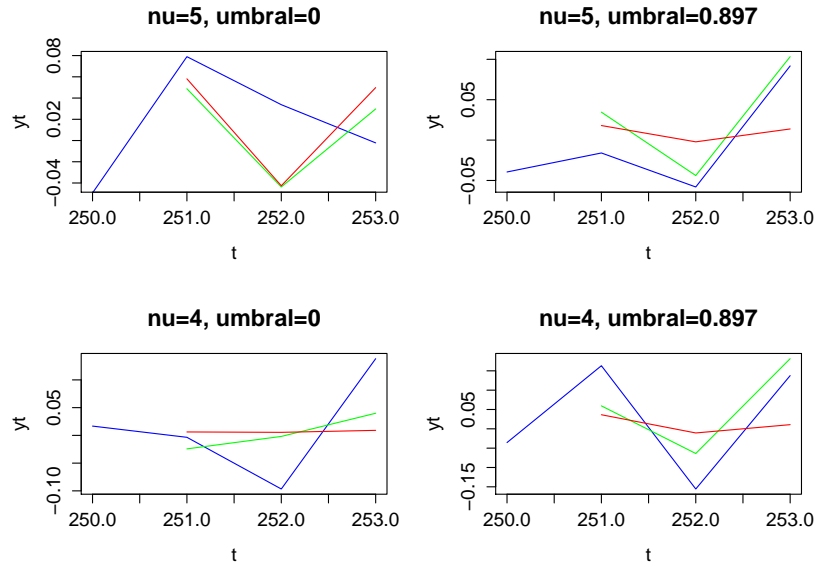


Figura 6.4.6: En verde promedio de los pronósticos del modelo t y en rojo del modelo normal

Por último, en las tablas (6.14), (6.15) se presentan los cocientes de los ECMP del modelo t y normal.

Caso $\gamma = 0$	$h = 1$	$h = 2$	$h = 3$	Caso $\gamma = 0.897$	$h = 1$	$h = 2$	$h = 3$
ecmpt/ecmpn	0.976	1.008	0.958	ecmpt/ecmpn	1.030	1.023	1.824

Tabla 6.14: Cocientes de los ECMP para el caso $\nu = 5$

Caso $\gamma = 0$	$h = 1$	$h = 2$	$h = 3$	Caso $\gamma = 0.897$	$h = 1$	$h = 2$	$h = 3$
ecmpt/ecmpn	1.110	1.039	1.431	ecmpt/ecmpn	1.079	1.036	1.445

Tabla 6.15: Cocientes de los ECMP para el caso $\nu = 4$

Como podemos notar ver, en la mayoría de los casos el modelo normal presenta un mejor desempeño, sin embargo la diferencia entre ambos modelos es mínima para dos pasos adelante.

TAR 3

Para el ejemplo TAR3 consideremos un proceso donde los parámetros de los procesos autorregresivos están cercanos a 1. Una gráfica de las realizaciones de Y_t para cada escenario considero, se muestra en la próxima gráfica.

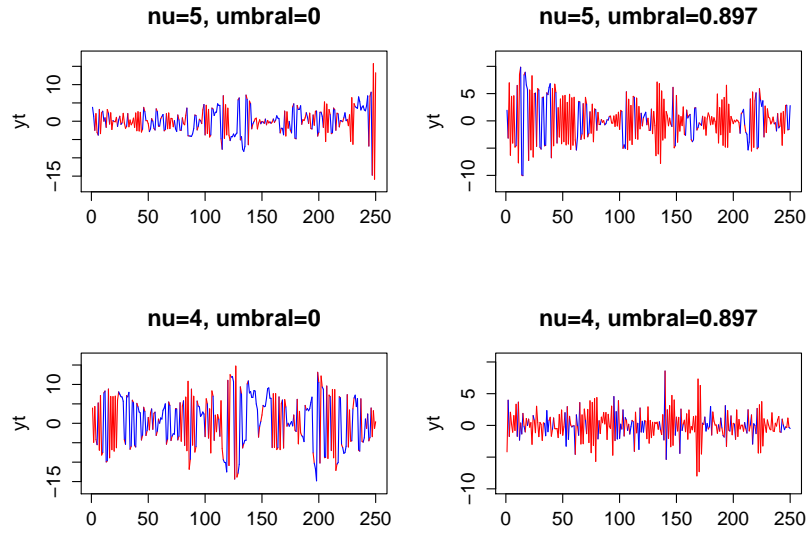


Figura 6.4.7: Realizaciones de los procesos Y_t

Los promedios y varianzas muestrales de las 500 estimaciones para cada uno de los parámetros del modelo TAR3 se pueden observar en las tablas siguientes

Caso $\gamma = 0$		Modelo t -student		Modelo Normal	
Parámetros	Real	Promedio	D. E.	Promedio	D. E.
β_1	-0.95	-0.943	0.029	-0.942	0.031
β_2	0.95	0.943	0.027	0.941	0.031
ν	5	5.530	1.883	-	-
σ	1.291	1.293	0.108	1.274	0.099

Tabla 6.16: Resultados de los parámetros para el modelo $\nu = 5$

Caso $\gamma = 0.897$		Modelo t -student		Modelo Normal	
Parámetros	Real	Promedio	D. E.	Promedio	D. E.
β_1	-0.95	-0.942	0.022	-0.941	0.025
β_2	0.95	0.941	0.044	0.940	0.048
ν	5	5.427	1.713	-	-
σ	1.291	1.298	0.106	1.271	0.093

Tabla 6.17: Resultados de los parámetros del modelo con $\nu = 5$

Caso $\gamma = 0$		Modelo t -student		Modelo Normal	
Parámetros	Real	Promedio	D. E.	Promedio	D. E.
β_1	-0.95	-0.943	0.026	-0.942	0.031
β_2	0.95	0.944	0.026	0.941	0.032
ν	4	4.290	1.163	-	-
σ	1.414	1.419	0.140	1.383	0.122

Tabla 6.18: Resultados de los parámetros del modelo con $\nu = 4$

		Modelo t -student		Modelo Normal	
Parámetro	Real	Promedio	D. E.	Promedio	D. E.
β_1	-0.95	-0.946	0.019	-0.945	0.023
β_2	0.95	0.945	0.038	0.943	0.046
ν	4	4.141	1.076	-	-
σ	1.414	1.455	0.180	1.407	0.162

Tabla 6.19: Resultados de los parámetros con $\nu = 4$

Al igual que en los ejemplos anteriores, las diferencias entre las estimaciones obtenidas en cada modelo son mínimas.

Algunas realizaciones de los pronósticos se pueden ver a continuación

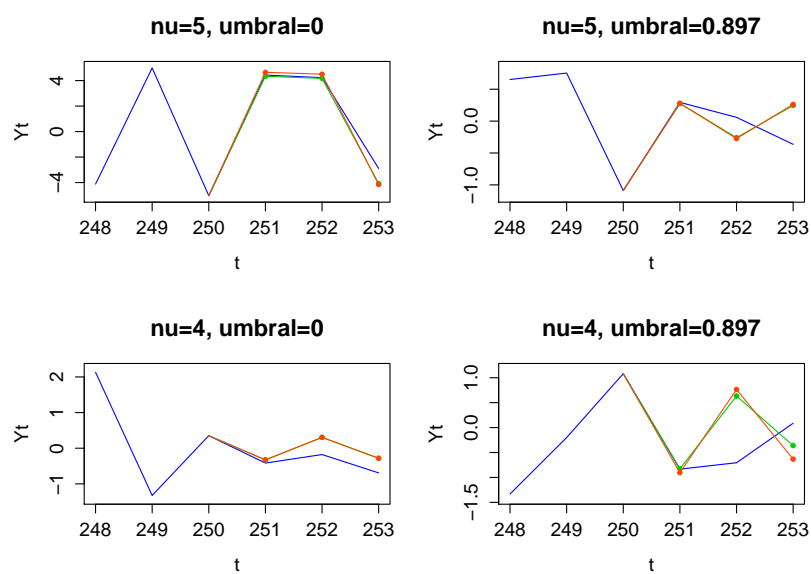


Figura 6.4.8: En verde pronósticos bajo el modelo t y en rojo bajo el modelo normal

Y una gráfica de los promedios de los pronósticos se muestra en (6.4.9)

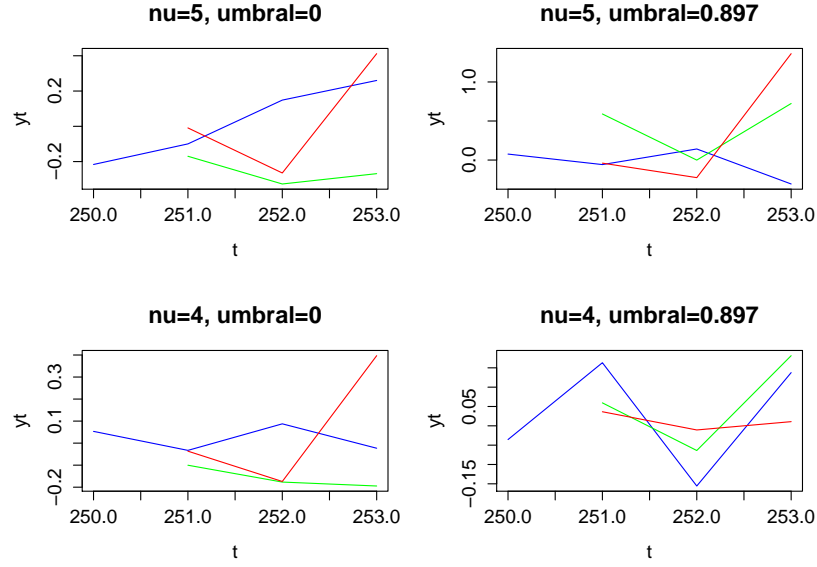


Figura 6.4.9: En verde promedio de los pronósticos bajo el modelo t y en rojo pronósticos bajo el normal

En las siguiente tablas se presenta los cocientes de los ECMP del modelo t con el normal.

Caso $\gamma = 0$	$h = 1$	$h = 2$	$h = 3$	Caso $\gamma = 0.897$	$h = 1$	$h = 2$	$h = 3$
ecmpt/ecmpn	4.903	0.955	1.014	ecmpt/ecmpn	8.335	0.999	1.127

Tabla 6.20: Cocientes de los ECMP para el caso $\nu = 5$

Caso $\gamma = 0$	$h = 1$	$h = 2$	$h = 3$	Caso $\gamma = 0.897$	$h = 1$	$h = 2$	$h = 3$
ecmpt/ecmpn	6.509	1.036	1.059	ecmpt/ecmpn	13.140	1.137	1.115

Tabla 6.21: Cocientes de los ECMP para el caso $\nu = 4$

Notamos que en la mayoría de los casos el modelo normal presenta un mejor desempeño, especialmente un paso adelante en el pronóstico.

TAR 4

Para el modelo TAR4 consideramos un proceso TAR (2; 1, 1) en el cual los parámetros de los órdenes autorregresivos son de igual magnitud pero signos contrarios. Una gráfica de este proceso lo vemos en la imagen (6.4.10).

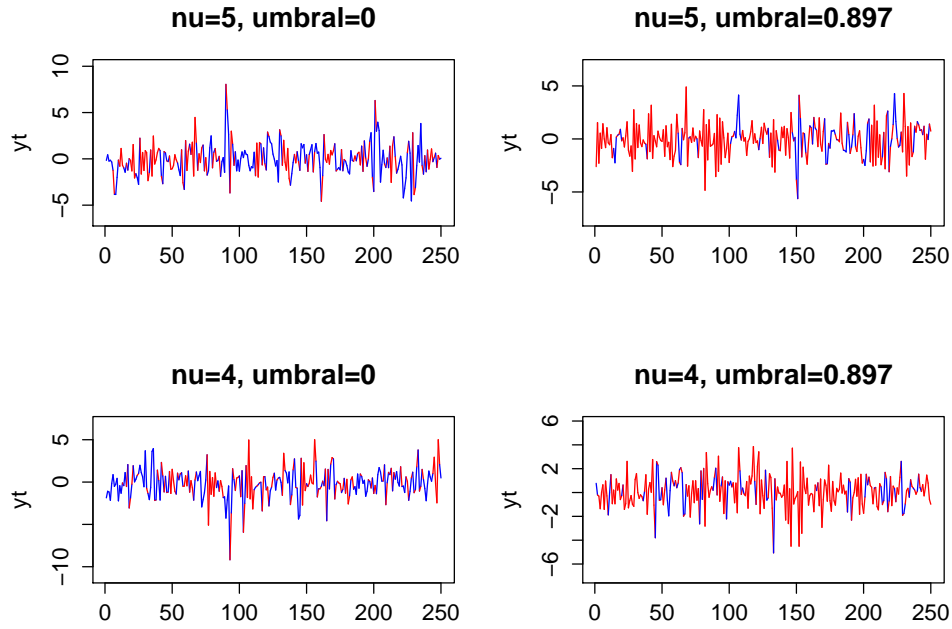


Figura 6.4.10: Realizaciones del proceso Y_t

Los promedios y varianzas muestrales de las 500 estimaciones para cada uno de los parámetros se resumen en las siguientes tablas

Caso $\gamma = 0$		Modelo t -student		Modelo Normal	
Parámetros	Real	Promedio	D. E.	Promedio	D. E.
β_1	-0.5	-0.492	0.069	-0.491	0.077
β_2	0.5	0.502	0.073	0.501	0.080
ν	5	5.568	2.122	-	-
σ	1.291	1.286	0.096	1.273	0.092

Tabla 6.22: Modelo con $\nu = 5$

Caso $\gamma = 0.897$		Modelo t -student		Modelo Normal	
Parámetros	Real	Promedio	D. E.	Promedio	D. E.
β_1	-0.5	-0.498	0.053	-0.499	0.057
β_2	0.5	0.498	0.113	0.493	0.124
ν	5	5.5249	2.294	-	-
σ	1.291	1.296	0.114	1.281	0.114

Tabla 6.23: Modelo con $\nu = 5$

Caso $\gamma = 0$		Modelo t -student		Modelo Normal	
Parámetros	Real	Promedio	D. E.	Promedio	D. E.
β_1	-0.5	-0.497	0.067	-0.499	0.076
β_2	0.5	0.497	0.068	0.496	0.078
ν	4	4.355	1.081	-	-
σ	1.414	1.414	0.158	1.386	0.167

Tabla 6.24: Modelo con $\nu = 4$

Caso $\gamma = 0$		Modelo t -student		Modelo Normal	
Parámetros	Real	Promedio	D. E.	Promedio	D. E.
β_1	-0.5	-0.499	0.052	-0.500	0.060
β_2	0.5	0.502	0.109	0.500	0.127
ν	4	4.308	1.035	-	-
σ	1.414	1.421	-0.159	1.397	0.165

Tabla 6.25: Modelo con $\nu = 4$

Las gráficas de algunos pronósticos se presentan en la figura (6.4.11).

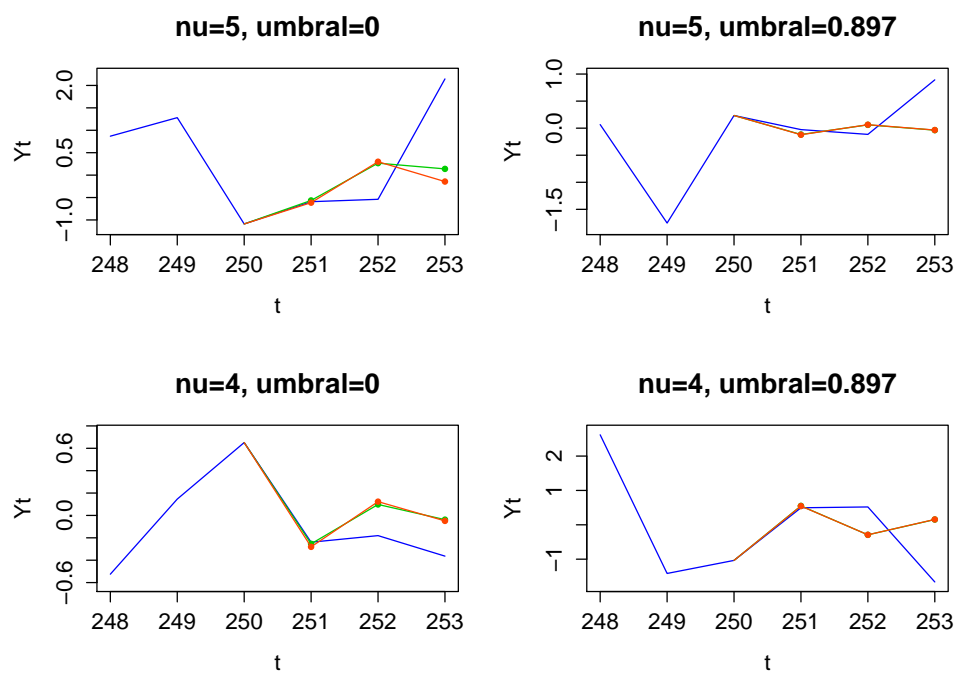


Figura 6.4.11: En verde los pronósticos bajo el modelo t y en rojo bajo el modelo normal

Mientras que los promedios de las estimaciones de los pronósticos se encuentran en (6.4.12)

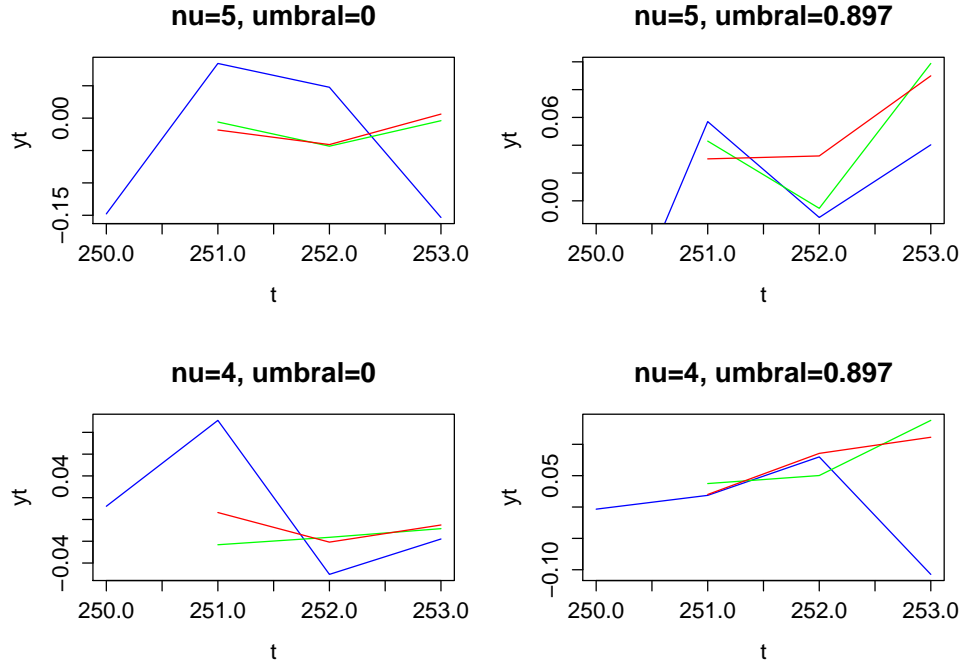


Figura 6.4.12: Promedio de los pronósticos

En las tablas siguientes se encuentran los cocientes ECMP del modelo t y normal para cada uno de los horizontes $h = 1, 2, 3$ y bajo cada uno de los escenarios considerados

Caso $\gamma = 0$	$h = 1$	$h = 2$	$h = 3$	Caso $\gamma = 0.897$	$h = 1$	$h = 2$	$h = 3$
ecmpt/ecmpn	1.006	1.003	0.996	ecmpt/ecmpn	1.216	1.014	1.036

Tabla 6.26: Cocientes de los ECMP para el caso $\nu = 5$

Caso $\gamma = 0$	$h = 1$	$h = 2$	$h = 3$	Caso $\gamma = 0.897$	$h = 1$	$h = 2$	$h = 3$
ecmpt/ecmpn	1.088	1.019	0.972	ecmpt/ecmpn	1.220	0.969	1.096

Tabla 6.27: Cocientes de los ECMP para el caso $\nu = 4$

Como podemos notar, el modelo normal en la mayoría de los casos presenta un mejor desempeño que el modelo t , aunque las diferencias son mínimas.

TAR 5

En el ejemplo TAR5 consideramos un proceso TAR $(2; 1, 1)$ donde los parámetros de los órdenes autorregresivos tienen diferente magnitud y signo. Una vez simulado el proceso, se eligió a elegir al azar 11 datos entre los primeros 250 datos y uno de los tres datos a pronosticar, a ese conjunto de 12 puntos se les adicionó un número aleatorio de una distribución $2t_5$. Algunas realizaciones de los procesos considerados se ubican en la siguiente gráfica

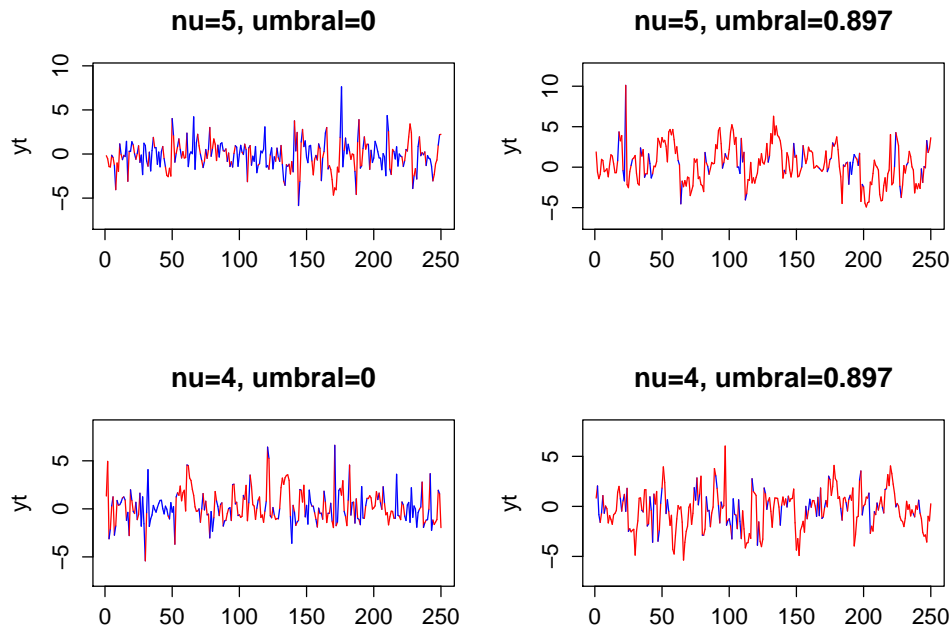


Figura 6.4.13: Gráfica de los procesos Y_t

Mientras que los resultados obtenidos de las simulaciones para los parámetros se resumen en las próximas tablas con sus respectivas medias y varianzas muestrales.

Caso $\gamma = 0$		Modelo t -student		Modelo Normal	
Parámetros	Real	Promedio	D. E.	Promedio	D. E.
β_1	0.8	0.737	0.088	0.656	0.133
β_2	-0.2	-0.167	0.069	-0.162	0.085
ν	5	3.461	0.692	-	-
σ	1.291	1.623	0.243	1.542	0.189

Tabla 6.28: Modelo $\nu = 5$

Caso $\gamma = 0.897$		Modelo t -student		Modelo Normal	
Parámetros	Real	Promedio	D. E.	Promedio	D. E.
β_1	0.8	0.751	0.055	0.681	0.085
β_2	-0.2	-0.171	0.091	-0.170	0.114
ν	5	3.353	0.685	-	-
σ	1.291	1.680	0.318	1.556	0.177

Tabla 6.29: Modelo $\nu = 5$

Caso $\gamma = 0$		Modelo t -student		Modelo Normal	
Parámetros	Real	Promedio	D. E.	Promedio	D. E.
β_1	0.8	0.744	0.080	0.670	0.121
β_2	-0.2	-0.166	0.067	-0.169	0.078
ν	4	3.036	0.524	-	-
σ	1.414	1.850	0.403	1.650	0.250

Tabla 6.30: Modelo $\nu = 4$

Caso $\gamma = 0$		Modelo t -student		Modelo Normal	
Parámetros	Real	Promedio	D. E.	Promedio	D. E.
β_1	0.8	0.762	0.051	0.691	0.086
β_2	-0.2	-0.175	0.086	-0.172	0.103
ν	4	2.929	0.477	-	-
σ	1.414	1.953	0.646	1.673	0.205

Tabla 6.31: Modelo $\nu = 4$

Notamos que con el modelo de datos atípicos(TAR 5) el modelo t da mejores aproximaciones a los parámetros reales que el modelo normal.

A continuación en (6.4.14)encontraremos los promedios para cada uno de los pronósticos

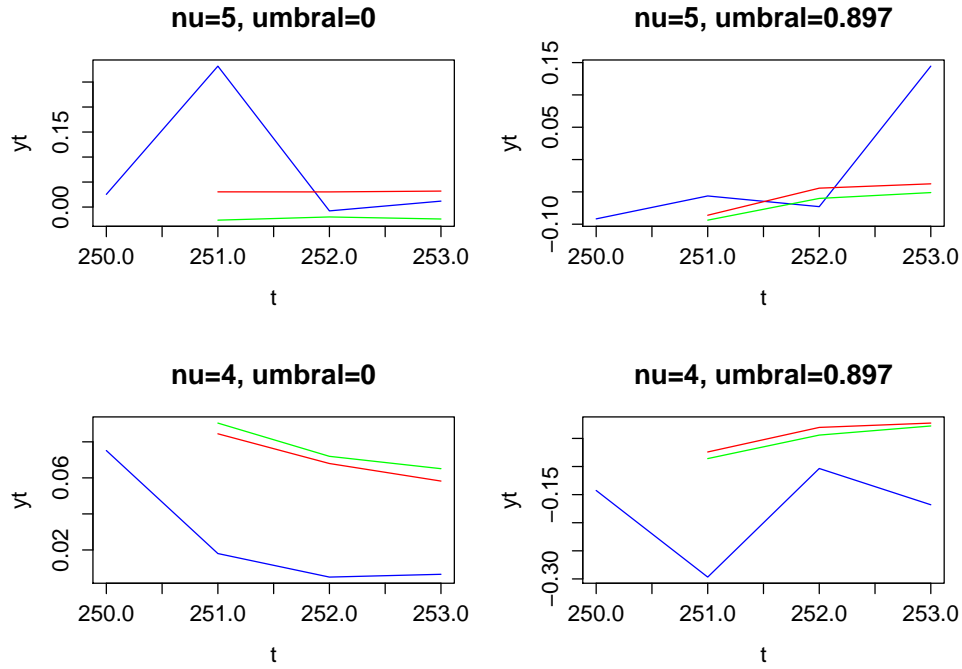


Figura 6.4.14: En verde promedio de pronósticos bajo el modelo t y en rojo bajo el modelo Normal

Los cocientes de los ECMP para cada uno de los horizontes $h = 1, 2, 3$ se encuentran en las siguientes tablas

Caso $\gamma = 0$	$h = 1$	$h = 2$	$h = 3$	Caso $\gamma = 0.897$	$h = 1$	$h = 2$	$h = 3$
ecmpt/ecmpn	0.979	1.002	0.964	ecmpt/ecmpn	0.986	1.000	0.999

Tabla 6.32: Cocientes de los ECMP para el caso $\nu = 5$

Caso $\gamma = 0$	$h = 1$	$h = 2$	$h = 3$	Caso $\gamma = 0.897$	$h = 1$	$h = 2$	$h = 3$
ecmpt/ecmpn	1.013	1.035	1.021	ecmpt/ecmpn	0.995	1.002	1.016

Tabla 6.33: Cocientes de los ECMP para el caso $\nu = 4$

En las tablas anteriores podemos ver que los desempeños de ambos modelos son casi iguales.

TAR 6

En el ejemplo TAR 6 consideramos un proceso en el cual los órdenes de los procesos autorregresivos tiene la misma magnitud pero diferente signo. Una vez simulado el proceso se eligió a elegir al azar 11 datos entre los primeros 250 datos y uno de los tres datos a pronosticar, a ese conjunto de 12 puntos se les adicionó un número aleatorio de una distribución $2t_5$. Un realización para cada escenario considerado aparecen en la siguiente gráfica

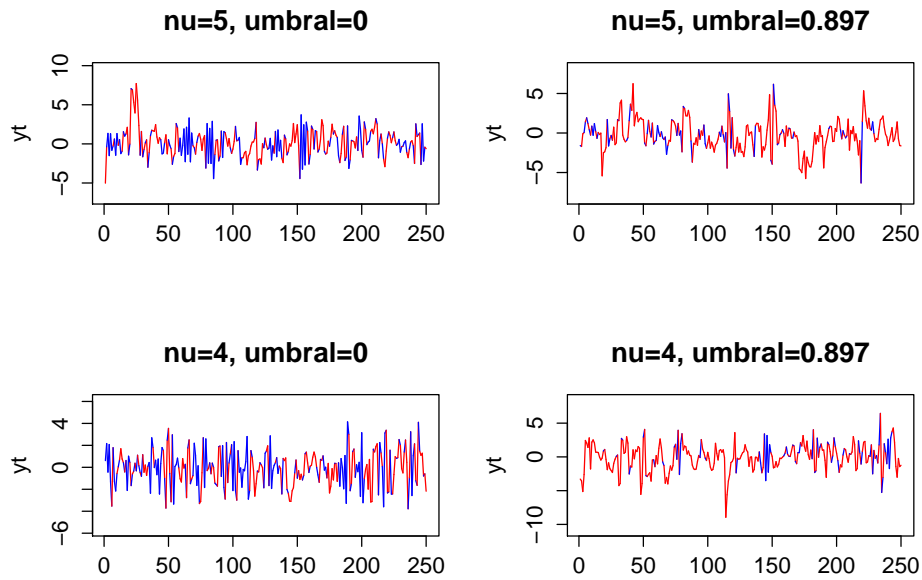


Figura 6.4.15: Realización del proceso

Los promedios y varianzas muestrales de las 500 estimaciones para cada uno de los parámetros se encuentran en las siguientes tablas.

Caso $\gamma = 0$		Modelo t -student		Modelo Normal	
Parámetros	Real	Promedio	D. E.	Promedio	D. E.
β_1	0.7	0.667	0.065	0.638	0.088
β_2	-0.7	-0.664	0.067	-0.639	0.084
ν	5	4.004	0.867	-	-
σ	1.291	1.468	0.168	1.417	0.123

Tabla 6.34: Modelo $\nu = 5$

Caso $\gamma = 0.897$		Modelo t -student		Modelo Normal	
Parámetros	Real	Promedio	D. E.	Promedio	D. E.
β_1	0.7	0.662	0.058	0.638	0.067
β_2	-0.7	-0.660	0.100	-0.639	0.124
ν	5	4.093	0.992	-	-
σ	1.291	1.458	0.166	1.412	0.128

Tabla 6.35: Modelo con $\nu = 5$

Caso $\gamma = 0$		Modelo t -student		Modelo Normal	
Parámetros	Real	Promedio	D. E.	Promedio	D. E.
β_1	0.7	0.672	0.063	0.650	0.081
β_2	-0.7	-0.675	0.057	-0.654	0.074
ν	4	3.459	0.695	-	-
σ	1.414	1.618	0.241	1.517	0.158

Tabla 6.36: Modelo $\nu = 4$

Caso $\gamma = 0.897$		Modelo t -student		Modelo Normal	
Parámetros	Real	Promedio	D. E.	Promedio	D. E.
β_1	0.7	0.665	0.048	0.637	0.074
β_2	-0.7	-0.666	0.103	-0.652	0.129
ν	4	3.405	0.690	-	-
σ	1.414	1.648	0.273	1.542	0.203

Tabla 6.37: Modelo $\nu = 4$

Algunos pronósticos se presentan en la gráfica (6.4.16)

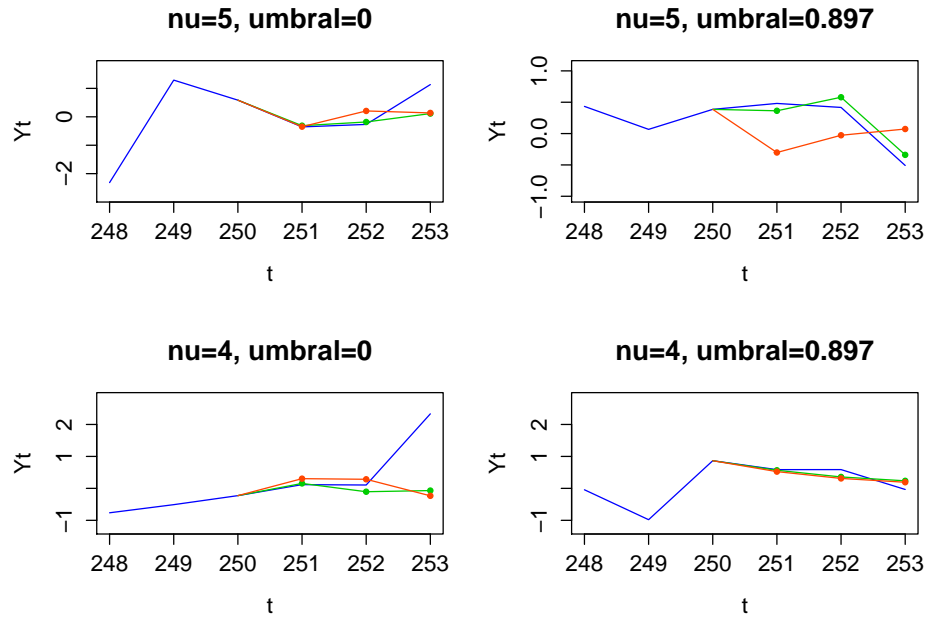


Figura 6.4.16: En verde pronósticos usando el modelo t y en rojo usando el modelo normal

Mientras que el promedio para cada horizonte $h = 1, 2, 3$ se ubica en (6.4.17)

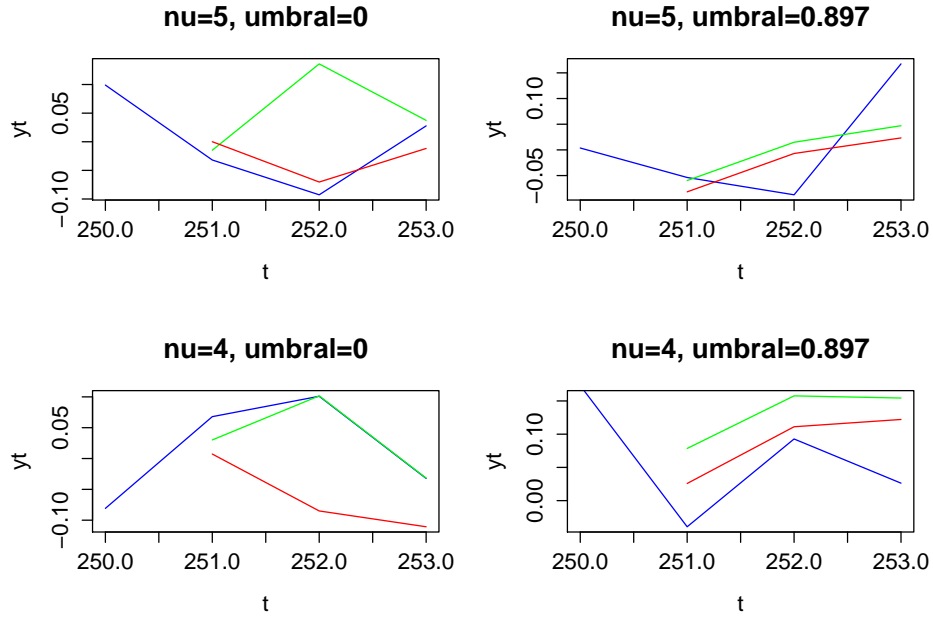


Figura 6.4.17: Promedio de los pronósticos.

Ahora, en las tablas (6.38) y (6.39) se presenta el cociente de los ECMP del modelo con errores t y normal para cada uno de los horizontes considerados.

Caso $\gamma = 0$	$h = 1$	$h = 2$	$h = 3$	Caso $\gamma = 0.897$	$h = 1$	$h = 2$	$h = 3$
ecmpt/ecmpn	1.060	0.990	1.017	ecmpt/ecmpn	1.043	1.010	1.035

Tabla 6.38: Cocientes de los ECMP para el caso $\nu = 5$

Caso $\gamma = 0$	$h = 1$	$h = 2$	$h = 3$	Caso $\gamma = 0.897$	$h = 1$	$h = 2$	$h = 3$
ecmpt/ecmpn	1.207	1.049	0.953	ecmpt/ecmpn	1.046	0.987	1.015

Tabla 6.39: Cocientes de los ECMP para el caso $\nu = 4$

Al igual que en el ejemplo TAR5, notamos que los desempeños de ambos modelos son casi iguales.

TAR 7

Para el ejemplo TAR 7 consideramos un proceso TAR $(2; 2, 2)$ y algunas realizaciones del proceso están en (6.4.18)

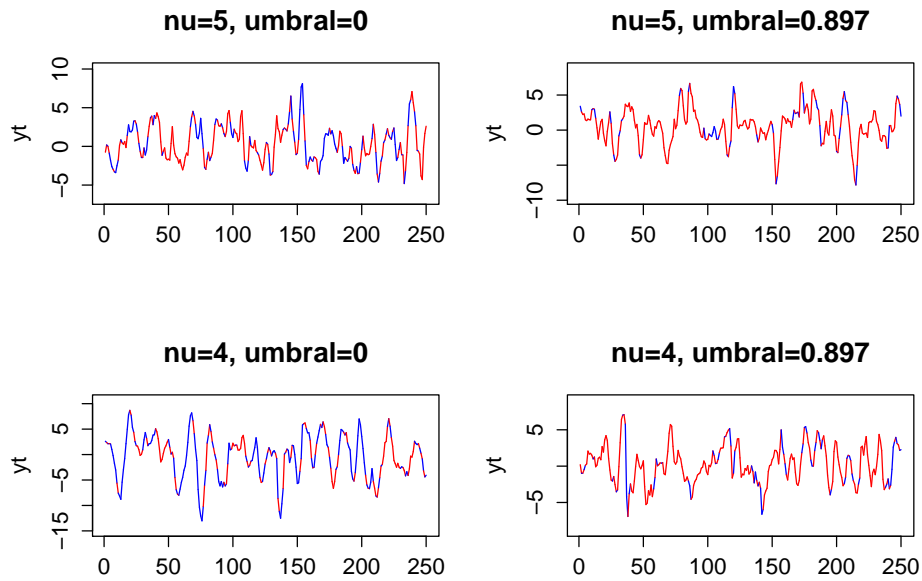


Figura 6.4.18: Proceso Y_t

Los resultados obtenidos para cada uno de los parámetros se resumen en la siguientes tablas

Caso $\gamma = 0$		Modelo t -student		Modelo Normal	
Parámetros	Real	Promedio	D. E.	Promedio	D. E.
β_{11}	1.2	1.195	0.063	1.195	0.069
β_{12}	-0.4	-0.396	0.065	-0.396	0.071
β_{21}	1.6	1.594	0.065	1.591	0.072
β_{22}	-0.8	-0.796	0.065	-0.794	0.005
ν	5	5.681	2.035	-	0.072
σ	1.291	1.286	0.104	1.271	0.102

Tabla 6.40: Modelo con $\nu = 5$

Caso $\gamma = 0.897$		Modelo t -student		Modelo Normal	
Parámetros	Real	Promedio	D. E.	Promedio	D. E.
β_{11}	1.2	1.199	0.056	1.201	0.062
β_{12}	-0.4	-0.402	0.057	-0.404	0.063
β_{21}	1.6	1.592	0.115	1.590	0.126
β_{22}	-0.8	-0.793	0.113	-0.792	0.123
ν	5	5.656	1.994	-	-
σ	1.291	1.286	0.110	1.269	0.105

Tabla 6.41: Modelo con $\nu = 5$

Caso $\gamma = 0$		Modelo t -student		Modelo Normal	
Parámetros	Real	Promedio	D. E.	Promedio	D. E.
β_{11}	1.2	1.194	0.059	1.194	0.067
β_{12}	-0.4	-0.395	0.060	-0.397	0.068
β_{21}	1.6	1.589	0.059	1.588	0.072
β_{22}	-0.8	-0.792	0.061	-0.792	0.072
ν	4	4.354	1.022	-	-
σ	1.414	1.409	0.144	1.381	0.136

Tabla 6.42: Modelo con $\nu = 4$

Caso $\gamma = 0.897$		Modelo t -student		Modelo Normal	
Parámetros	Real	Promedio	D. E.	Promedio	D. E.
β_{11}	1.2	1.193	0.054	1.190	0.062
β_{12}	-0.4	-0.396	0.053	-0.394	0.060
β_{21}	1.6	1.590	0.110	1.583	0.127
β_{22}	-0.8	-0.793	0.114	-0.790	0.131
ν	4	4.351	1.103	-	-
σ	1.414	1.413	0.147	1.389	0.156

Tabla 6.43: Modelo con $\nu = 4$

Algunos pronósticos realizados para cada uno de los escenarios considerados se ubican en la gráfica (6.4.19)

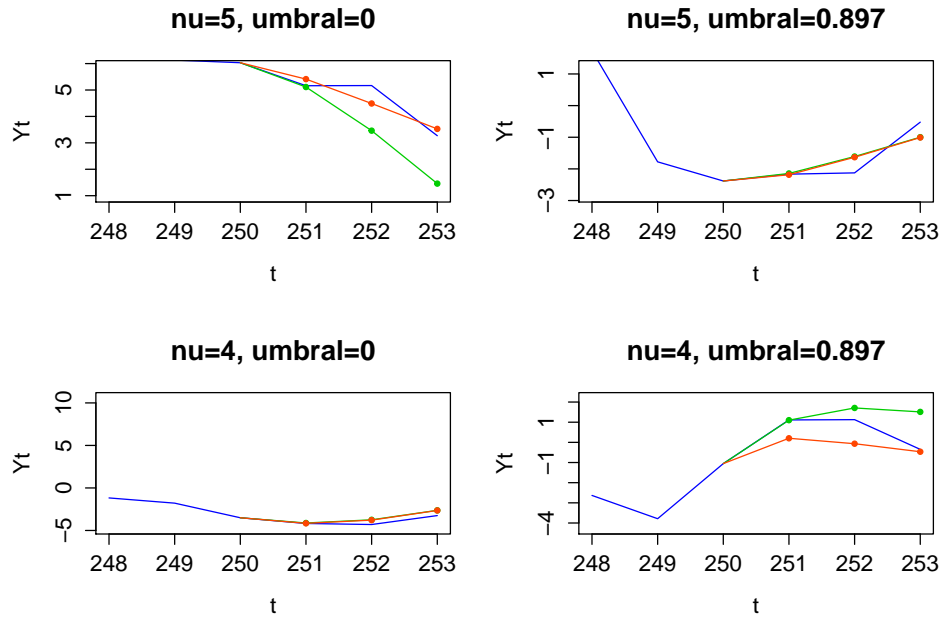


Figura 6.4.19: En verde pronóstico bajo modelo t -student y en rojo bajo modelo normal

Los promedios de los pronósticos para cada horizonte de los modelos considerados se encuentran en la siguiente gráfica.

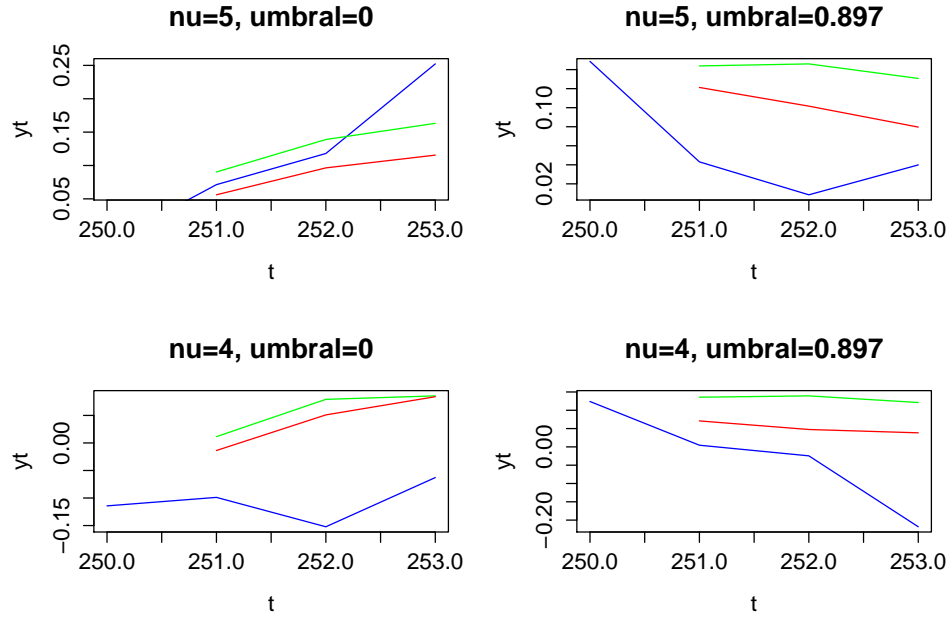


Figura 6.4.20: En verde pronósticos bajo el modelo t y en rojo bajo modelo normal

Ahora, los cocientes de los ECMP del modelo t con el normal se encuentran en (6.44), (6.45).

Caso $\gamma = 0$	$h = 1$	$h = 2$	$h = 3$	Caso $\gamma = 0.897$	$h = 1$	$h = 2$	$h = 3$
ecmpt/ecmpn	1.028	0.998	0.960	ecmpt/ecmpn	0.991	0.984	1.002

Tabla 6.44: Cocientes de los ECMP para el caso $\nu = 5$

Caso $\gamma = 0$	$h = 1$	$h = 2$	$h = 3$	Caso $\gamma = 0.897$	$h = 1$	$h = 2$	$h = 3$
ecmpt/ecmpn	1.057	1.033	0.969	ecmpt/ecmpn	1.302	1.136	1.074

Tabla 6.45: Cocientes de los ECMP para el caso $\nu = 4$

Como podemos ver los cocientes de los ECMP para cada horizonte y bajo los 4 escenarios considerados son muy cercanos a 1 por lo cual el desempeño de los modelos es muy similar.

TAR 8

Una representación gráfica de cada uno de los procesos considerados en el ejemplo TAR 8 la podemos ver en la gráfica (6.4.21)

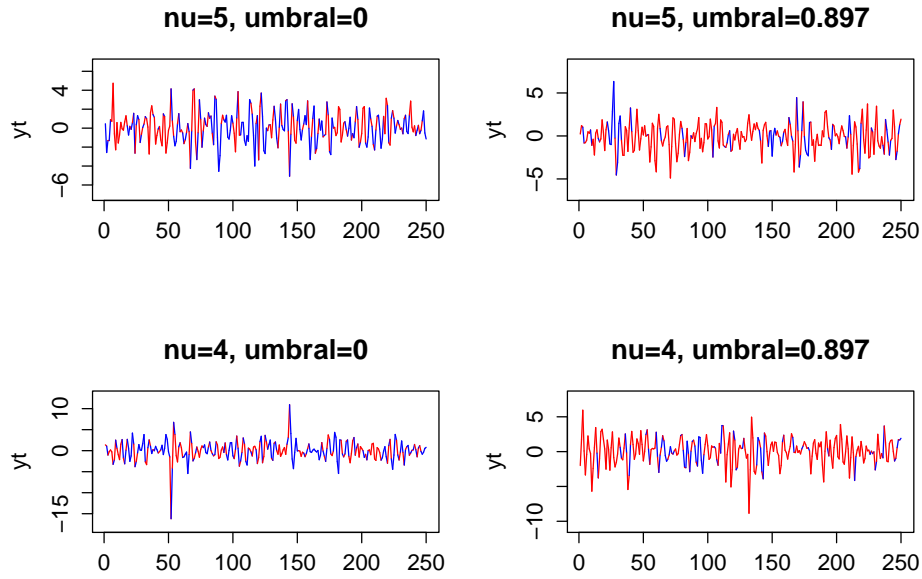


Figura 6.4.21: Procesos Y_t

Y los resultados para cada uno de los parámetros se presentan en las siguientes tablas.

Caso $\gamma = 0$		Modelo t -student		Modelo Normal	
Parámetros	Real	Promedio	D. E.	Promedio	D. E.
β_{11}	0.3	0.300	0.065	0.300	0.070
β_{12}	-0.5	-0.496	0.067	-0.494	0.074
β_{21}	0.6	0.596	0.067	0.595	0.074
β_{22}	-0.7	-0.691	0.064	-0.691	0.070
ν	5	5.659	1.683	-	-
σ	1.291	1.281	0.103	1.266	0.100

Tabla 6.46: Modelo con $\nu = 5$

Caso $\gamma = 0.897$		Modelo t -student		Modelo Normal	
Parámetros	Real	Promedio	D. E.	Promedio	D. E.
β_{11}	0.3	0.297	0.049	0.297	0.057
β_{12}	-0.5	-0.498	0.052	-0.496	0.058
β_{21}	0.6	0.596	0.103	0.597	0.112
β_{22}	-0.7	-0.691	0.104	-0.692	0.114
ν	5	5.663	1.916	-	-
σ	1.291	1.281	0.099	1.270	0.103

Tabla 6.47: Modelo con $\nu = 5$

Caso $\gamma = 0$		Modelo t -student		Modelo Normal	
Parámetros	Real	Promedio	D. E.	Promedio	D. E.
β_{11}	0.3	0.302	0.063	0.302	0.069
β_{12}	-0.5	-0.500	0.064	-0.502	0.074
β_{21}	0.6	0.600	0.062	0.601	0.071
β_{22}	-0.7	-0.695	0.059	-0.695	0.067
ν	4	4.358	1.103	-	-
σ	1.414	1.410	0.143	1.382	0.142

Tabla 6.48: Modelo con $\nu = 4$

Caso $\gamma = 0.897$		Modelo t -student		Modelo Normal	
Parámetros	Real	Promedio	D. E.	Promedio	D. E.
β_{11}	0.3	0.297	0.053	0.299	0.059
β_{12}	-0.5	-0.501	0.051	-0.500	0.058
β_{21}	0.6	0.601	0.101	0.599	0.113
β_{22}	-0.7	-0.703	0.095	-0.701	0.107
ν	4	4.306	1.343	-	-
σ	1.414	1.427	0.157	1.394	0.165

Tabla 6.49: Modelo con $\nu = 4$

En la siguiente gráfica encontraremos algunos de los pronósticos realizados para cada escenario considerado.

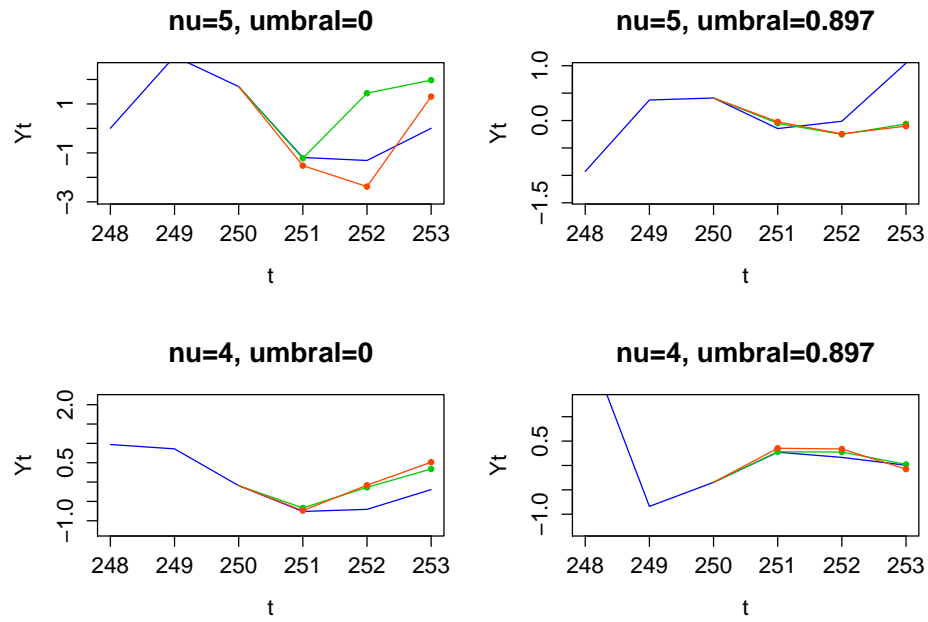


Figura 6.4.22: En verde pronósticos bajo modelo t y en rojo bajo el modelo normal

Los promedios para cada uno de los pronósticos se encuentra en la figura (6.4.23)

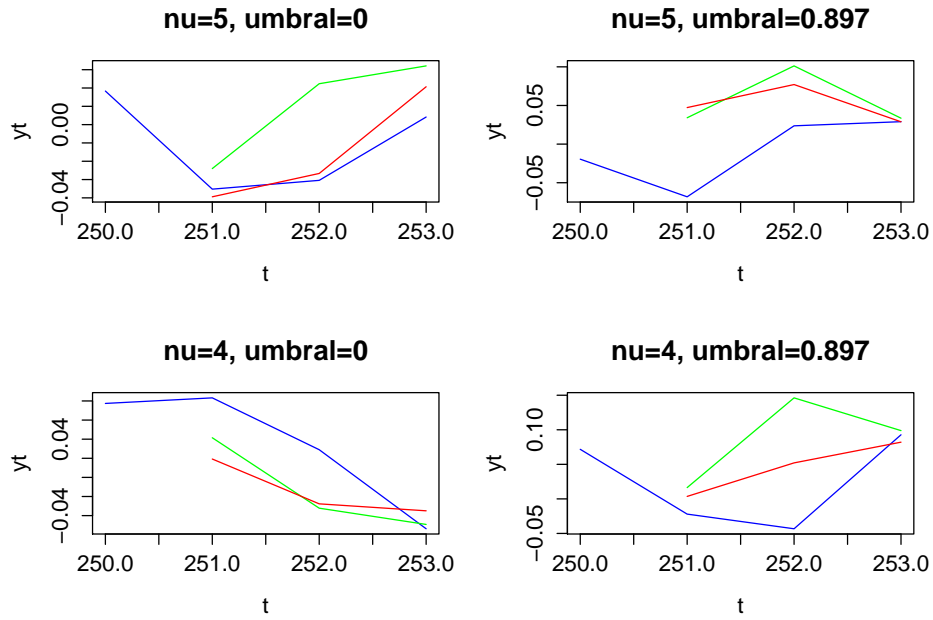


Figura 6.4.23: Promedio de los pronósticos

En las siguientes tablas encontraremos los cocientes de los ECMP del modelo t con el modelo normal.

Caso $\gamma = 0$	$h = 1$	$h = 2$	$h = 3$	Caso $\gamma = 0.897$	$h = 1$	$h = 2$	$h = 3$
ecmpt/ecmpn	0.965	1.143	1.093	ecmpt/ecmpn	0.919	1.084	1.006

Tabla 6.50: Cocientes de los ECMP para el caso $\nu = 5$

Caso $\gamma = 0$	$h = 1$	$h = 2$	$h = 3$	Caso $\gamma = 0.897$	$h = 1$	$h = 2$	$h = 3$
ecmpt/ecmpn	1.139	1.068	1.054	ecmpt/ecmpn	0.965	1.513	1.153

Tabla 6.51: Cocientes de los ECMP para el caso $\nu = 4$

Para este ejemplo el modelo normal presentó un mejor desempeño en la mayoría de los escenarios.

6.5. Resumen

En cada uno de los 8 ejemplos considerados anteriormente analizamos 4 escenarios diferentes, por lo que en total tenemos 32 casos en los cuales calculamos los ECMP de cada uno de los 3 horizontes, $h = 1, 2, 3$, para los modelos t y normal, donde recordemos que la distribución de los errores del VPGD son t -student independientes. En la segunda fila de la tabla (6.52) presentamos los porcentajes en los cuales el modelo TAR con errores normales, basándonos en los ECMP, fue más eficiente que el modelo t , mientras que en la tercera fila se encuentra en porcentaje de las veces que el modelo TAR con errores t fue mejor que el de errores normales. Podemos observar que para cada uno de los horizontes el modelo normal, en más del 70 % de los casos, presenta un mejor desempeño que el modelo t .

Modelo	$h = 1$	$h = 2$	$h = 3$
Normal	71.875 %	75 %	71.875 %
t -student	28.125 %	25 %	28.125 %

Tabla 6.52: Resultados por cada horizonte

En la siguiente tabla se presenta la misma información que en (6.52) con la diferencia de que los porcentajes se calculan considerando los grados de libertad de la distribución t de los errores del VPGD.

Caso $\nu = 5$	$h = 1$	$h = 2$	$h = 3$	Caso $\nu = 4$	$h = 1$	$h = 2$	$h = 3$
Normal	56.25 %	62.5 %	68.75 %	Normal	87.5 %	87.5 %	75 %
t	43.75 %	37.5 %	31.25 %	t	12.5 %	12.5 %	25 %

Tabla 6.53: Resultados por valor de ν

De nueva cuenta el modelo normal presenta un mejor desempeño comparado con el modelo t . Además, cuando $\nu = 4$ se observa que en al menos el 75 % de los casos el modelo normal resultó ser más eficiente.

Por último, consideramos la misma información que en (6.52) con la diferencia que cada uno de los porcentajes de la tabla es calculado considerando el valor de γ . Los resultados se resumen en la tabla (6.54)

$\gamma = 0$	$h = 1$	$h = 2$	$h = 3$	$\gamma = 0.897$	$h = 1$	$h = 2$	$h = 3$
Normal	81.25 %	75 %	50 %	Normal	62.5 %	75 %	93.75 %
t	18.75 %	25 %	50 %	t	37.5 %	25 %	6.25 %

Tabla 6.54: Resultados por porcentaje de observaciones

Al igual que en las tablas anteriores, podemos observar que en nuestros ejemplos simulados el modelo normal, basados en los ECMP, presenta un mejor desempeño sobre el modelo t sin importar que la cantidad teórica de observaciones del régimen inferior sea del 50 % (caso $\gamma = 0$) o del 75 % (caso $\gamma = 0.897$) .

Capítulo 7

Conclusiones

En este trabajo hemos presentado las herramientas para el cálculo de los estimadores, ajuste del modelo adecuado y pronósticos de valores futuros para un conjunto de datos del que suponemos proviene de un modelo TAR($r; p_1, p_2, \dots, p_r$) cuya distribución de los errores es t -student.

En el capítulo 5 presentamos una metodología para el ajuste del modelo adecuado combinando las ideas de Qian(1998) y los criterios de información de Akaike y Bayesiano. En los ejemplos de simulación considerados se mostró que el método propuesto da buenos resultados en la selección del verdadero modelo cuando hacemos uso del criterio Bayesiano, mientras que el criterio de Akaike tiende a sobreestimar el verdadero proceso generador de los datos.

Además, hemos investigado mediante estudios de simulación, las diferencias que hay en las estimaciones de los parámetros suponiendo dos tipos de errores asociados, normales y t , en modelos TAR(2;1,1) y TAR(2;2,2) cuyo verdadero proceso de errores tienen distribución t -student independientes. Como hemos podido notar, las diferencias entre ambos modelos son casi nulas en condiciones en el que los grados de libertad del proceso que genera los datos son 4 o 5. Sin embargo, bajo la presencia de datos atípicos pudimos comprobar, en los ejercicios TAR5 y TAR6 del capítulo pasado, que existe una

ligera diferencia en la estimación de parámetros.

En lo referente a los pronósticos, extendimos la verosimilitud predictiva perfil al caso general e hicimos algunos ejercicios de simulación en donde obtuvimos, utilizando los errores cuadráticos medio de pronósticos, que el modelo normal presenta en la mayoría de los escenarios considerados un mejor desempeño que el modelo t , aunque hay que resaltar que estas diferencias son mínimas.

Es importante remarcar que a pesar de que en otros modelos mencionados en la literatura, como los modelos de regresión, en donde existe una clara ventaja al utilizar errores t versus normales en datos que presentan colas pesadas o atípicos, parece indicar que en los modelos TAR, al menos en las simulaciones presentadas, existe un comportamiento consistente en no mostrar diferencias más allá de cosas mínimas en la estimación y la predicción de valores futuros aún ante la presencia de valores atípicos en las series. Por lo cual, para trabajos futuros se podría investigar si esto es debido a la estructura de los modelos TAR o bien si existe otras razones para ello. Así también, se recomendaría hacer un ejercicio de simulación exhaustivo para la estimación de parámetros y pronósticos de modelos TAR con dos o más regímenes y explorar nuevas alternativas para la estimación de pronósticos como los realizados en Russel(2006) donde se presentan, por ejemplo, regresión por cuantiles.

Capítulo 8

Apéndice

En este apartado se presentará una descripción de las funciones programadas en R y que fueron empleadas para la simulación de datos, estimación de parámetros, cálculo de los criterios de información AIC, BIC y cálculo de los pronósticos.

Función **tar2.sim**

La función **tar2.sim** simula datos de un proceso TAR($2; p_1, p_2$) donde los errores asociados a Y_t tienen distribución t -student y la variable umbral, Z_t , es un proceso autorregresivo de orden 1 con errores normales. Recibe los siguientes argumentos (num, p1, p2, gamma, par.inf, par.sup, nu, rho, sigma), los cuales se detallan a continuación:

num	Número de datos a simular.
p1	Orden del proceso autorregresivo inferior.
p2	Orden del proceso autorregresivo superior.
gamma	Valor del umbral γ .

par.inf	Parámetros del proceso autorregresivo inferior.
par.sup	Parámetros del proceso autorregresivo superior.
nu	Grados de libertad asociados a la distribución t —student de los errores.
rho	Parámetro del proceso Z_t .
sigma	Desviación estándar de la distribución de los errores de Z_t .

La función nos devuelve una matriz X de dimensión $(num \times 3)$ que contiene en la primera columna los datos simulados de Y_t , en la segunda columna están los datos del proceso Z_t y en la tercera columna se encuentra una etiqueta que nos indica si el dato Y_t proviene del proceso autorregresivo inferior o superior.

Función `estima.tar`

estima.tar es una función que, dado un conjunto de datos (Y_t, Z_t) y un valor fijo de γ , estima los parámetros suponiendo que provienen de un proceso TAR(2; p_1, p_2). Recibe los siguientes argumentos (yt, zt, p1, p2, gamma),

yt	Datos que provienen de un modelo TAR(2; p_1, p_2).
zt	Datos de la variable umbral Z_t .
p1	Orden del proceso autorregresivo inferior.
p2	Orden del proceso autorregresivo superior.
gamma	Valor del umbral γ .

Como resultado, la función nos devuelve una matriz X de dimensión $2 \times (p_1 + p_2 + 2)$ que contiene en la primera fila el valor de la logverosimilitud evaluada en los parámetros estimados, los valores estimados para los parámetros de los procesos autorregresivos y el estimador de los grados de libertad, ν . En la segunda fila contiene los mismos valores anteriores, excepto que se calculan suponiendo que los errores del proceso $\text{TAR}(2; p_1, p_2)$ son normales y por lo tanto, en vez de estimar ν se estima la desviación estándar de los errores, σ .

Función `pronosticos.tar`

La función `pronosticos.tar` estima los parámetros y los pronósticos a h pasos de un proceso $\text{TAR}(2; p_1, p_2)$ mediante la técnica de verosimilitud predictiva perfil. La función recibe los siguientes argumentos (`yt`, `zt`, `gamma`, `p1`, `p2`, `h`),

<code>yt</code>	Conjunto de datos que provienen de un proceso $\text{TAR}(2; p_1, p_2)$.
<code>zt</code>	Conjunto de datos de la variable umbral Z_t .
<code>gamma</code>	Valor del umbral γ .
<code>p1</code>	Orden del proceso autorregresivo inferior.
<code>p2</code>	Orden del proceso autorregresivo superior.
<code>h</code>	Número de pronósticos a estimar.

Y nos devuelve una matriz X de dimensión $2 \times (p_1 + p_2 + 2h + 3)$. En la primera fila se encuentran los estimadores de los parámetros del proceso autorregresivo inferior y superior, los grados de libertad, ν , el parámetro ρ del proceso Z_t , la desviación estándar σ de los errores de Z_t y los h estimadores

de los pronósticos Y_{T+1}, \dots, Y_{T+h} y $Z_T, Z_{T+1}, \dots, Z_{T+h-1}$. La segunda fila contiene los valores anteriores, a excepción que se calculan suponiendo que los errores del proceso $\text{TAR}(2; p_1, p_2)$ tienen distribución normal con media cero y varianza σ^2 .

Función `ajuste.tar`

ajuste.tar es una función que, dado un conjunto de datos (Y_t, Z_t) y un valor fijo de γ , nos calcula el valor del criterio de Akaike y el criterio Bayesiano suponiendo que los datos provienen de un modelo $\text{TAR}(2; p_1, p_2)$. La función recibe los siguientes argumentos (`yt`, `zt`, `p1`, `p2`, `gamma`)

<code>yt</code>	Conjuntos de datos que provienen de un proceso $\text{TAR}(2; p_1, p_2)$.
<code>zt</code>	Conjunto de datos de la variable umbral Z_t .
<code>p1</code>	Orden del proceso autorregresivo inferior.
<code>p2</code>	Orden del proceso autorregresivo superior.
<code>gamma</code>	Valor del umbral γ .

Como resultado, **ajuste.tar** nos devuelve una matriz X de dimensión 2×2 donde en la primera fila contiene los valores del criterio de Akaike y el criterio Bayesiano suponiendo que la distribución de los errores del proceso es t -student mientras que la segunda fila contiene los valores del criterio de Akaike y Bayesiano suponiendo que los errores del proceso $\text{TAR}(2; p_1, p_2)$ tienen distribución normal.

Bibliografía

- [1] Aeschliman, C., Park, J. and Kak, A. (2010). A Novel Parameter Estimation Algorithm for the Multivariate t-Distribution and Its Application to Computer Vision. Lecture Notes in Computer Science Volume 6312, 594-607.
- [2] Ahsanullah, M., Kibria, G. and Shakil, M. (2014). Normal and Student's t Distributions and Their Applications. Atlantis Press.
- [3] Berger, J. O. and Wolpert, R. L. (1984). The Likelihood Principle. IMS, Hayward, Calif..
- [4] Bjørnstad, J. F. (1990). Predictive Likelihood: A Review. Statistical Science. 5, No. 1, 242-265.
- [5] Chan, K. S. (2009). Exploration of a nonlinear world. World Scientific Publishing.
- [6] Chan, K. S. (1993). Consistency and limiting distribution of the least square estimation of a threshold autoregressive model. The Annals of Statistics.
- [7] Cysneiros, F.(2005). Regression Models with Symmetrical Errors. International Encyclopedia of Statistical Science .
- [8] Cysneiros, F. and Paula, G. (2005) Restricted methods in symmetrical linear regression models. Computational Statistics & Data Analysis. 49, 689-708.

- [9] Fang, K. T., Kotz, S., and Ng, K.W. (1990). Symmetric Multivariate and Related Distributions. Chapman & Hall.
- [10] Kotz, S. and Nadarajah, S. (2004). Multivariate t distributions and their applications. Cambridge University Press.
- [11] Gonzalo, J. and Pitarakis, J. (2012) Estimation and Infrence in Threshhold Type Regime Switching Models.
- [12] Granger, C. W. J. and Teräsvirta, T. (1993). Modelling Nonlinear Economic Relationships. Advanced Texts in Econometrics.
- [13] Hansen, B. E. (1997). Inference in TAR models. Studies in Nonlinear Dynamics and Econometrics, Vol. 2, Num. 1.
- [14] Hansen, B. E. (2011). Threshold autoregression in economics. Statistics and Its Interface Volume 4, 123–127.
- [15] Liew, V. K. and Chong, T. T. (2003). Effects of STAR and TAR types nonlinearities on order selection criteria. Working paper.
- [16] Liu, C. and Rubin, D. B. (1995). ML estimation of the t distribution using EM and its extensions, ECM and ECME. Statistica Sinica Vol. 5, 19-39.
- [17] Mathiasen, P. E. (1979). Prediction functions. Scand. Journal Statistics. 6, 1-21.
- [18] Moreno, E. C. (2010). Una aplicación del modelo TAR en series de tiempo financieras, Tesis de maestría.UNAL.
- [19] Nieto, F. (2005). Modeling bivariate threshold autoregressive processes in the presence of missing data. Communications in Statistics: Theory and Methods 34, 905–930.
- [20] Petrucci J. D. and Woolford S. W. (1984). A Threshold AR(1)Models.Journal of Applied Probability. 21, No. 2, 270-286.

- [21] Qian L. (1998). On maximum likelihood estimators for a threshold autoregression. *Journal of Statistical Planning and Inference*.
- [22] Russell, M. G. (2006) Pronósticos en modelos autorregresivos con Umbral, Tesis de Doctorado. CIMAT.
- [23] Scheffler, C. (2008). A derivation of the em updates for finding the maximum likelihood parameter estimates of the student's t distribution. Working paper.
- [24] Tong, H. (1978). On a Threshold Model. *Pattern Recognition and Signal Processing*, 575-586.
- [25] Tong, H. and Lim, K. S. (1980). Threshold autoregressive, limit cycles and cyclical data. *Journal Royal Statistic Society Serie B*. 42, 245-292.
- [26] Tong, H. (1983). *Threshold Models in Non-linear Time Series Analysis*. Springer Verlag.
- [27] Tong, H. (1990). *Non-linear Time Series: A Dynamical System Approach*. Oxford University Press.
- [28] Tsay, R. (1998). Testing and modeling multivariate threshold models. *Journal of the American Statistical Association* 93, 1188–1202.
- [29] Tsay, R. (2002). *Analysis of financial time series*. Wiley.
- [30] Zhang, H.(2012) Estimación de los modelos TAR cuando el proceso de ruidos sigue una distribución t . *Comunicaciones en estadística*, Vol. 4, No. 2.