

CIMAT

Centro de Investigación en Matemáticas, A.C.

Pronósticos en modelos autorregresivos con umbral

T E S I S

Que para obtener el grado de:

Doctor en ciencias

Con orientación en:

Probabilidad y Estadística

P r e s e n t a:

María Guadalupe Russell Noriega

Guanajuato, Gto. México. Septiembre de 2006

Pronósticos en modelos autorregresivos con umbral

A mis hijos, Julia y Emilio

A Suito

A mis padres y hermanos

Agradecimientos

A mis asesores, Dra. Graciela González Farías y Dr. Jesús Gonzalo, porque gracias a ellos pude iniciar y finalizar mis estudios de doctorado. Gracias por el apoyo, la disponibilidad y el esfuerzo que me dedicaron.

A mis sinodales, los doctores:

Joaquín Ortega Sánchez, Rogelio Ramos Quiroga, Manuel Mendoza Ramírez y Rolando Cavazos Cadena. Gracias por sus valiosas sugerencias, comentarios y por la disponibilidad mostrada en todo momento.

Al Dr. Víctor Guerrero por ser parte de mi comité de tesis tutorial y por fungir como sinodal lector.

A mis profesores durante el doctorado: Dra. Graciela González Farías, Dr. Gabriel Huerta, Dr. Andrés Christen Gracia, Dr. Daniel Hernández Hernández y Stephanie Dunbar.

Al Dr. Ignacio Barradas por creer en mí y contratarme como profesor de la FAMAT, antes de concluir con mis estudios doctorales.

Al Dr. Adolfo Sánchez Valenzuela por apoyarme y autorizar las actividades relacionadas a la culminación de mis estudios.

Al CONACyT por mi beca No. 86663 y al PROMEP por mi beca-tesis.

A la Universidad de Guanajuato y CIMAT por los apoyos financieros otorgados para asistir a las diferentes actividades académicas a lo largo de mis estudios. Y por usar sus infraestructuras para alcanzar mis objetivos.

A los apoyos recibidos de los proyectos de investigación del CONACyT: 39017E y 45974-F.

Al Ministerio de Ciencia y Tecnología DGYCIT (SEC 01-0890) y Secretaria de Educación de España SEJ2004-04101-ECON por el soporte financiero parcial en mis estancias académicas al departamento de Econometría de la Universidad Carlos III de Madrid.

A mi amigo Alonso Núñez Páez por su ayuda en el trabajo de cómputo, por escucharme y darme ánimo.

A mi cuñado Ramón por su ayuda en el trabajo de cómputo y a su esposa Ely por el apoyo brindado.

A mi querido esposo, por apoyarme, aguantarme y ayudarme incondicionalmente tanto académicamente como personalmente.

A mis queridos hijos Julia y Emilio por el siempre hecho de estar aquí.

A mis padres y suegros por cuidar a mis hijos cuando yo no estuve cerca y por todo el apoyo que me han brindado desde siempre.

Resumen

Los modelos de series de tiempo no lineal constituyen una herramienta útil para describir y pronosticar una gran variedad de fenómenos naturales. Uno de los modelos más utilizados y estudiados en la literatura de modelos no lineales corresponde a los modelos autorregresivos tipo umbral ("threshold"), identificados en la literatura como modelos TAR.

Ante la siguiente problemática, identificada en la literatura de modelos no lineales, ¿por qué, dado un proceso generador de datos de un modelo no lineal, los pronósticos a través de un modelo lineal resultan más satisfactorios en términos de su error cuadrático medio? nos planteamos el objetivo general de explorar nuevas alternativas de obtención de pronósticos en modelo TAR. El cumplimiento de dicho objetivo se logra al proponer métodos de estimación y pronóstico que tomen en cuenta la estructura no lineal del modelo TAR. Esto se verifica en la evaluación de los desempeños de los pronósticos bajo cada alternativa considerada.

Específicamente, una vez identificado el problema proponemos dos alternativas para generar y cuantificar los desempeños de los pronósticos h -pasos adelante en los modelos TAR: verosimilitud predictiva y regresión por cuantiles. A diferencia de los procedimientos tradicionales para generar pronósticos en modelos TAR, verosimilitud predictiva pronostica conjuntamente la serie TAR y la variable umbral que determina el comportamiento en cada uno de los regímenes en el modelo, aprovechando así las características de no linealidad que caracterizan al proceso. Verosimilitud predictiva se basa en el principio de máxima verosimilitud y consiste en calcular la verosimilitud predictiva del modelo TAR, de manera que obtenemos los pronósticos h -pasos adelante vía la maximización de la verosimilitud predictiva perfil de los valores futuros desconocidos. La segunda alternativa considera al modelo TAR como un modelo de regresión por cuantiles manejando en forma adecuada la no linealidad que caracteriza al modelo TAR dentro y fuera de la muestra. Esta alternativa identificada como regresión por cuantiles es una propuesta de distribución libre que puede tomar en cuenta la asimetría de los errores en forma natural. Consiste en minimizar la función de cuantiles resultante, la cual es una función asimétrica como función de los pesos asociados a los errores positivos y negativos del modelo.

Demostremos las propiedades asintóticas de los estimadores de máxima verosimilitud y de regresión por cuantiles bajo los modelos TAR. Implementamos también un amplio ejercicio de simulación para comparar los desempeños de los pronósticos h -pasos adelante en los modelos TAR contra los desempeños de los pronósticos en los modelos AR, para cada una de las alternativas de pronósticos. Valoramos dichos desempeños mediante las funciones de pérdida error cuadrático medio de pronóstico, que corresponde a una función de pérdida simétrica, y por medio de la función de pérdida de cuantiles, la cual considera el caso de pérdidas asimétricas, así como el caso simétrico, dado por desviaciones absolutas del error.

Contenido

Prefacio	1
1 Introducción	5
1.1 Pronósticos en modelos no lineales	5
1.2 Técnicas de evaluación y obtención de pronósticos	10
1.3 Motivación	14
1.4 Identificación del problema y alternativas de solución	16
2 El modelo TAR	19
2.1 Algunas definiciones y resultados	19
2.2 Error cuadrático medio bajo especificaciones incorrectas	24
2.2.1 Clasificación incorrecta de las observaciones en los regímenes bajo el modelo TAR simplificado	24
2.2.2 Especificación incorrecta del verdadero proceso generador de datos .	29
3 Estimación en modelos TAR	51
3.1 Estimación por mínimos cuadrados	51
3.1.1 Propiedades de los estimadores de mínimos cuadrados	52
3.2 Estimación mediante máxima verosimilitud	56
3.2.1 Obtención de los estimadores de máxima verosimilitud	57
3.2.2 Propiedades asintóticas de los estimadores de máxima verosimilitud .	60
4 Pronósticos en modelos TAR aproximando la esperanza condicional	63
4.1 Modelos de regresión para series binarias	65
5 Pronósticos en modelos TAR mediante verosimilitud predictiva	71
5.1 Verosimilitud predictiva en el modelo lineal AR(1)	72
5.2 Verosimilitud predictiva para el modelo TAR(2; 1, 1)	74

6	Regresión por cuantiles en modelos TAR: estimación y obtención de pronósticos	79
6.1	Introducción	79
6.2	Regresión por cuantiles	81
6.3	El modelo de cuantiles para series de tiempo no lineal	84
6.3.1	Regresión por cuantiles para el modelo TAR(2; 1, 1)	85
6.3.2	Regresión por cuantiles para el modelo TAR más simple	86
6.3.3	Propiedades asintóticas	95
6.3.4	Pruebas de Hipótesis	97
7	Estudio de simulación	103
7.1	Descripción del algoritmo de simulación	104
7.2	Presentación y análisis de los resultados	108
7.2.1	Resultados asumiendo que $Z_{t-1} \sim iidU(0, 1)$	109
7.2.2	Resultados asumiendo que Z_{t-1} sigue un proceso AR(1)	127
8	Conclusiones y discusión	147
A	Prueba de la propiedad de consistencia de los EMV	151
B	Condiciones de regularidad, modelo de cuantiles	159
C	Pruebas de los Teoremas 6.1 a 6.5 del Capítulo 6	163

Prefacio

Esta tesis está orientada al estudio y a la obtención de los pronósticos en modelos tipo umbral autorregresivos (Threshold Autoregressive (TAR) models).

Históricamente el análisis de series de tiempo se ha desarrollado alrededor de modelos de series de tiempo lineales; sin embargo en una gran cantidad de aplicaciones para diversas áreas como: economía (ciclos de negocios), oceanografía (niveles de agua en el mar, temperatura del agua en el mar, el fenómeno del niño), hidrología (fluidos), ciencias ambientales (contaminación del agua), física (caos, rotación de la tierra, magnitudes de huracanes), epidemiología (salmonela, sarampión, rubéola), entre otras, se ha observado un esquema de comportamiento común, el cual puede ser descrito por un comportamiento lineal por pedazos en el espacio de la llamada variable *umbral*¹, pero no lineal en el tiempo. El uso de modelos tipo umbral para describir distintos fenómenos de la naturaleza crece día a día; sabemos por ejemplo que en economía y crecimiento de poblaciones animales, existen fases de expansión y contracción, las cuales difícilmente pueden ser descritas por la misma dinámica lineal. Asimismo, suponer que un crecimiento de la temperatura diaria promedio de 0°C a 2°C tiene el mismo impacto que el aumento de la temperatura en los flujos de un río de -12°C a -10°C si hay un glaciar dentro de la zona de influencia del río, es poco realista. La linealidad ignora las fases de transición; mientras que asumir una dinámica simple y distinta (del tipo autorregresivo, por ejemplo) dentro de cada una de las fases, nos ayuda a comprender y describir mejor el fenómeno. La consideración de estos comportamientos distintos se da en general por cambios de un régimen a otro y tomando en cuenta los comportamientos de las distintas fases de transición del fenómeno en estudio. Tong y Lim (1980) y Tong (1983) introducen los llamados modelos de regímenes autorregresivos, capaces de considerar comportamientos distintos en cada uno de los regímenes como función del comportamiento de una variable o covariable, identificada como variable umbral y activado por el cruce de dicha variable en función de un valor crítico llamado valor del umbral. Posteriormente, Tong (1983, 1990) da una introducción detallada de los modelos de regímenes autorregresivos, enfatizando lo amplio de las aplicaciones de dichos modelos, debido a que estos contemplan diversos fenómenos como: ciclos límites, resonancias, distorsiones armónicas, modulación de efectos y

¹traducción de la palabra threshold

caos. Sin embargo, cabe aclarar que en los trabajos de Tong (1983, 1990), así como la gran mayoría de la bibliografía actual, consideran el caso donde la serie misma identifica a qué régimen corresponde la t -ésima observación, es decir, la variable umbral que determina el cambio en el régimen está dada por rezagos o función de rezagos de ella misma. Por claridad de exposición, identificamos a este tipo particular de modelos como SETAR (Self-Exciting Threshold AutoRegressive model). Por otro lado, si la variable umbral es una función exógena de la variable de interés identificamos al modelo de regímenes como TAR (Threshold AutoRegressive model). Las propiedades teóricas de los modelos SETAR, así como los resultados asintóticos de los estimadores para cada uno de los parámetros en el modelo, han sido ampliamente estudiadas y se deben principalmente a los trabajos de Chan, K. S., Petrucci, J. D, Tong, H. y Woolford, S. W. (1985), Chan y Tong (1985), Petrucci y Woolford (1984), Chan (1993), Chen y Tsay (1991), Hansen (1997, 2000), Chan y Tsay (1998), Qian (1998), entre otros. La literatura en torno a la generación y evaluación de pronósticos h -pasos hacia adelante en modelos SETAR es extensa, por ejemplo, Clements y Smith (1997), Clements, M. P., Frances, P. H., Smith, J. y Van Dijk D. (2003), Diebold, F. X., Gunther, T. A. y Tay, A.S. (1998), Diebold y Mariano (1995), Dacco y Sanchell (1999), entre otros. Una de las conclusiones coincidentes en dichos trabajos es que los modelos SETAR pueden describir adecuadamente el comportamiento del fenómeno correspondiente, pero al momento de evaluar el desempeño de los pronósticos fuera de la muestra observada, resulta que dichos pronósticos no se comportan en general mejor que los pronósticos provenientes de un modelo del tipo lineal autorregresivo (AR por sus siglas en inglés). La función de pérdida de mayor uso, para cuantificar los desempeños de los pronósticos h -pasos hacia adelante, es el error cuadrático medio de pronóstico (ECMP). Mientras que los procedimientos de obtención de pronósticos más estudiados se basan en métodos del tipo Monte Carlo los cuales aproximan la esperanza condicional resultante para cada horizonte de pronóstico, mediante procedimientos recursivos.

Motivados por el hecho de que en algunas situaciones reales el considerar que los cambios en los regímenes se dan por rezagos de la serie misma puede no ser muy realista, estudiamos el caso más general de modelos autorregresivos de regímenes. Es decir, el caso en que la variable umbral que determina a qué régimen corresponde la t -ésima observación está dada por una variable exógena, que denotaremos por Z_t . La variable umbral Z_t posee su propia dinámica, la cual es conocida y relativamente más simple que el modelo TAR de interés. Contrario al caso SETAR, a la fecha existe un número pequeño de artículos en torno al estudio de los modelos TAR, como por ejemplo González y Gonzalo (1998, 1999), Hansen (1997, 2000). Suponer que la variable umbral es distinta a rezagos o función de rezagos de la serie misma, conlleva a consideraciones teóricas distintas a las estudiadas en el modelo SETAR. El espacio probabilístico asociado al modelo TAR debe considerar la estructura probabilística de la variable umbral. Las condiciones de regularidad, así como las demostraciones correspondientes, para garantizar estacionariedad y ergodicidad en los

modelos TAR, se establecen en González y Gonzalo (1998).

El objetivo general del presente trabajo de tesis consiste en explorar nuevas alternativas de obtención de pronósticos en modelos TAR. El alcance de dicho objetivo trae como consecuencia los siguientes objetivos particulares: proponer métodos de estimación y obtención de pronósticos en modelos TAR, que tomen en cuenta la estructura no linealidad del modelo y evaluar los desempeños de los pronósticos bajo cada alternativa considerada.

Analizamos tres procedimientos para la obtención de los parámetros estimados en el modelo TAR: mínimos cuadrados, máxima verosimilitud y regresión por cuantiles. Para los primeros dos métodos obtenemos las propiedades asintóticas de los coeficientes autorregresivos para el caso de umbral conocido y desconocido. En regresión por cuantiles estudiamos sólo las propiedades asintóticas de los estimadores para el caso de umbral conocido. Estudiamos los comportamientos de los pronósticos h -pasos adelante para los modelos TAR, bajo cada alternativa de obtención de pronósticos y considerando las funciones de pérdida ECM y función de pérdida de cuantiles. El estudio de los desempeños de los pronósticos en los modelos TAR se contrasta con los respectivos desempeños de los pronósticos en modelos lineales AR. A diferencia de los modelos SETAR, los pronósticos en los modelos TAR implican el conocer el comportamiento futuro de la variable umbral. Proponemos dos nuevas alternativas de obtención de pronósticos en los modelos TAR, verosimilitud predictiva y regresión por cuantiles. La primera de ellas se basa en el principio de máxima verosimilitud y consiste en obtener la verosimilitud predictiva del modelo TAR bajo estudio. Bajo esta alternativa los pronósticos h -pasos hacia adelante para el modelo de interés, serán los valores futuros más plausibles en el modelo TAR dada la información observada hasta el tiempo t , en el sentido de que dichos valores maximizan la ecuación objetivo resultante e identificada en la literatura como verosimilitud predictiva perfil, ver Lauritzen (1974), Hinkley (1979), Butler (1986) y Bjørnstad (1990). La segunda alternativa se basa en considerar al modelo TAR como un modelo de regresión por cuantiles, justificado por el hecho de que ante una especificación errónea del modelo, ocasionada por considerar un modelo del tipo lineal como el verdadero proceso generador de datos y acrecentada por la influencia de probabilidades con valores grandes de mala clasificación en el régimen. Concretamente, identificamos que el proceso de error ante la mala especificación no se comporta como un proceso de ruido blanco. La función de pérdida asociada al modelo de regresión por cuantiles es asimétrica y contempla el caso simétrico dado por las desviaciones absolutas del error de predicción, Koul (1996) y Mukherjee (1999). Realizamos un amplio ejercicio de simulación, bajo el supuesto de que el parámetro de umbral es conocido, a fin de estudiar los comportamientos de los pronósticos h -pasos adelante, bajo cada una de las alternativas de obtención de pronósticos. Cuantificamos los desempeños de los pronósticos mediante los errores cuadráticos medios de pronósticos y calculando también la función de pérdida definida en el modelo de regresión por cuantiles. Análogo al caso SETAR, contrastamos nuestras dos alternativas para la obtención de los pronósticos en el modelo TAR, contra los pronósticos obtenidos del

método de Monte Carlo recursivo. El procedimiento implementado es similar a los métodos de Monte Carlo estudiados por Granger y Teräsvirta (1993), para modelos no lineales. Es decir, aproximamos los valores esperados condicionales para los h horizontes de interés en el modelo TAR, con la diferencia de que ahora incluimos el pronóstico de la variable umbral Z_t o bien el valor de la función indicadora definida en términos de si la variable umbral al tiempo t está por arriba o por abajo del valor del umbral.

Una conclusión general de nuestro ejercicio de simulación es que los pronósticos en el modelo TAR, obtenidos por medio de verosimilitud predictiva presentan mejores desempeños que los pronósticos bajo el modelo AR, ya que los valores de las funciones de pérdidas para los distintos horizontes de pronóstico son más pequeños que sus respectivos valores bajo el modelo lineal.

La organización de esta tesis es la siguiente: el Capítulo 1 contiene una revisión del tema de pronósticos para modelos no lineales. En el Capítulo 2 se describe el modelo TAR, sus propiedades teóricas, como por ejemplo estacionariedad y ergodicidad. Así como una comparación teórica de las expresiones de los errores cuadrático medio de pronóstico bajo un esquema TAR simplificado. El Capítulo 3 aborda el problema de estimación en los modelos TAR, se estudian los métodos de mínimos cuadrados y de máxima verosimilitud; en ambos métodos de estimación se discuten las propiedades asintóticas de los estimadores bajo parámetro de umbral conocido y desconocido. El Capítulo 4 describe el método recursivo de Monte Carlo para la aproximación de los pronósticos en el modelo TAR, describiendo los métodos correspondientes para el pronóstico de la variable umbral. En el Capítulo 5 introducimos el concepto de verosimilitud predictiva y obtenemos bajo el supuesto de normalidad en los errores la correspondiente función de verosimilitud predictiva para el modelo AR(1) y el modelo TAR(2; 1, 1). El Capítulo 7 se refiere a las ideas de regresión por cuantiles, específicamente expresamos el modelo TAR como un modelo de regresión por cuantiles, establecemos las propiedades asintóticas de los parámetros estimados en el modelo y describimos el procedimiento de obtención de los pronósticos bajo esta alternativa. En el Capítulo 8 describimos el ejercicio de simulación implementado para la comparación de las distintas alternativas de obtención de pronósticos, así como la cuantificación de sus desempeños, como función de los valores de las funciones de pérdida. Los resultados de dicho ejercicio se presentan en forma gráfica y tabulada. Finalmente en el Capítulo 9 resaltamos algunas observaciones en torno al problema de estimación y pronósticos en modelos TAR, así como un resumen de las conclusiones de este trabajo de tesis.

Capítulo 1

Introducción

Los modelos de series de tiempo no lineal son una herramienta útil para describir y pronosticar una gran variedad de series de tiempo económicas. Sin embargo, se ha observado en una gran variedad de trabajos empíricos, que estos modelos ajustan bien dentro de la muestra pero difícilmente presentan una ganancia sustancial en sus predicciones fuera de la muestra. Las predicciones se comparan con las predicciones correspondientes en modelos lineales autorregresivos. Algunos autores, Pesaran y Potter (1997) y Clements y Smith (1999), sugieren que los modelos no lineales pueden ser mejores competidores de desempeños de pronósticos puntuales como función de sus respectivos ECMP si se consideran las varianzas condicionales para cada uno de los regímenes. Clements y Smith (1999) calculan los ECMP condicionando en cada uno de los regímenes y calculando también estadísticos relacionados con la dirección del cambio. El estudio de Pesaran y Potter (1997) sugiere que los modelos no lineales pueden ser mejores al pronosticar momentos de orden alto, lo cual se entiende como: el desempeño de los intervalos de pronósticos y de la densidad de pronósticos es mejor que su correspondiente competidor lineal.

1.1 Pronósticos en modelos no lineales

Sea Y_t una serie de tiempo estacionaria dada por la siguiente ecuación,

$$Y_t = g(Y_{t-1}) + \varepsilon_t,$$

con g una función arbitraria y ε_t una secuencia de variables aleatorias independientes e idénticamente distribuidas con distribución D , media cero y varianza σ^2 y que denotaremos como: $\varepsilon_t \sim iid D(0, \sigma^2)$. Dado un conjunto de T observaciones sobre el tiempo, el predictor de Y_{T+h} con error cuadrático medio mínimo está dado por la esperanza condicional, es decir,

$$\hat{Y}_{T+h} = E[g(Y_{T+h})|Y_s; s \leq T]. \quad (1.1)$$

Si $g(\cdot)$ es una función lineal, pensemos por ejemplo que Y_t sigue un proceso autorregresivo de orden uno (AR(1)) estacionario y dado por la expresión: $Y_t = \theta Y_{t-1} + \eta_t$, con proceso de error $\eta_t \sim iidN(0, \sigma_\eta^2)$, tenemos que a partir de la ecuación recursiva:

$$Y_{T+h} = \theta^h Y_T + \sum_{j=0}^{h-1} \theta^j \eta_{T+h-j},$$

el pronóstico optimal h -pasos adelante está dado por:

$$\begin{aligned} \hat{Y}_{T+h} &= E[Y_{T+h} | Y_s; s \leq T] \\ &= \theta^h Y_T. \end{aligned}$$

En la práctica es común que el parámetro θ sea desconocido, y es necesario estimarlo del conjunto de datos observados. La estimación de θ puede realizarse, por ejemplo, mediante los métodos de mínimos cuadrados o por máxima verosimilitud, entre otros métodos. El estimador de θ , bajo el supuesto de normalidad en los errores para ambos procedimientos de estimación es el mismo, y satisface las propiedades de consistencia y normalidad asintótica. De esta forma el pronóstico h -pasos adelante para el modelo AR(1), se obtiene de la siguiente expresión:

$$\hat{Y}_{T+h}^{AR(1)} = \hat{\theta}^h Y_T,$$

donde $\hat{\theta}$ es un estimador consistente de θ . La varianza del error de predicción o error cuadrático medio de pronóstico, se obtiene de la siguiente expresión:

$$\text{Var} \left(Y_{T+h} - \hat{Y}_{T+h}^{AR(1)} | Y_T \right) \approx \hat{\sigma}_\eta^2 \frac{(1 - \hat{\theta}^{2h})}{(1 - \hat{\theta}^2)},$$

con $\hat{\sigma}_\eta^2$ la varianza estimada del proceso de error en el modelo AR(1).

La obtención de los pronósticos en modelos no lineales es más difícil que en el caso lineal, por ejemplo, consideremos el modelo no lineal dado por:

$$Y_t = g(Y_{t-1}) + \varepsilon_t, \text{ con } \varepsilon_t \sim iidD(0, \sigma^2), \quad (1.2)$$

pero ahora la función $g(\cdot)$ es una función no lineal. Sea \mathfrak{F}_T la σ -álgebra generada por el conjunto de información hasta el tiempo T , digamos $\{Y_1, \dots, Y_T\}$. El pronóstico optimal un paso adelante está dado por:

$$\hat{Y}_{T+1} = E[Y_{T+1} | \mathfrak{F}_T] = g(Y_T).$$

Sin embargo, dada la no linealidad de la función g , el pronóstico dos pasos adelante está dado por:

$$\hat{Y}_{T+2} = E[Y_{T+2} | \mathfrak{F}_T] = E[g(Y_{T+1}) | \mathfrak{F}_T] \neq g(E[Y_{T+1} | \mathfrak{F}_T]) = g(\hat{Y}_{T+1}).$$

O equivalentemente,

$$\begin{aligned}\hat{Y}_{T+2} &= E[Y_{T+2}|\mathfrak{F}_T] = E[g(Y_{T+1})|\mathfrak{F}_T] = E[g(g(Y_T) + \varepsilon_{T+1})|\mathfrak{F}_T] \\ &= E\left[g\left(\hat{Y}_{T+1} + \varepsilon_{T+1}\right)|\mathfrak{F}_T\right] \neq g\left(\hat{Y}_{T+1}\right)\end{aligned}$$

Similarmente, podemos expresar el pronóstico h -pasos adelante mediante la siguiente igualdad:

$$\begin{aligned}\hat{Y}_{T+h} &= E[Y_{T+h}|\mathfrak{F}_T] = E[g(Y_{T+h-1})|\mathfrak{F}_T] \\ &= E[g(g(g(\dots g(Y_T) + \varepsilon_{T+1})) + \dots + \varepsilon_{T+h-3}) \\ &\quad + \varepsilon_{T+h-2}) + \varepsilon_{T+h-1})|\mathfrak{F}_T] \\ &= E\left[g\left(g\left(g(\dots g(\hat{Y}_{T+1} + \varepsilon_{T+1})) + \dots + \varepsilon_{T+h-3}\right) \right. \right. \\ &\quad \left. \left. + \varepsilon_{T+h-2}) + \varepsilon_{T+h-1}\right)|\mathfrak{F}_T\right],\end{aligned}$$

donde la función g es evaluada $h - 1$ veces.

Los métodos más utilizados en la literatura para generar pronósticos en modelos no lineales, y en particular en los modelos SETAR, se basan en aproximar la esperanza condicional para diferentes horizontes de pronóstico. Algunas de las alternativas más estudiadas se describen a continuación, Granger y Teräsvirta (1993).

Método Exacto. Supongamos que la función de distribución acumulada D , para los errores ε_t en el modelo es conocida y dada por $\Phi(\cdot)$, entonces el pronóstico dos pasos adelante por el método exacto se obtiene como:

$$\hat{Y}_{E,T+2} = E[Y_{T+2}|\mathfrak{F}_T] = \int_{-\infty}^{\infty} g(\hat{Y}_{T+1} + Z) d\Phi(Z).$$

Equivalentemente, tenemos que:

$$\begin{aligned}\hat{Y}_{E,T+2} &= E[Y_{T+2}|\mathfrak{F}_T] = \int_{-\infty}^{\infty} g(y_{T+1}) \eta(y_{T+1}|\mathfrak{F}_T) dy_{T+1} \\ &= \int_{-\infty}^{\infty} E(Y_{T+2}|Y_{T+1}) \eta(y_{T+1}|\mathfrak{F}_T) dy_{T+1},\end{aligned}$$

donde $\eta(y_{T+1}|\mathfrak{F}_T)$ es la función de densidad de Y_{t+1} condicional a la σ -álgebra \mathfrak{F}_T . Notemos que la función de densidad anterior es la misma función de densidad de ε_{T+1} , sólo que ahora la media de la variable aleatoria es $g(Y_t)$.

Generalizando, el pronóstico h pasos hacia adelante para Y_t está dado por:

$$\hat{Y}_{E,T+h} = E[Y_{T+h}|\mathfrak{F}_T] = \int_{-\infty}^{\infty} g(y_{T+h-1}) \eta(y_{T+h-1}|\mathfrak{F}_T) dy_{T+h-1},$$

de modo que para obtener el pronóstico h -pasos adelante debemos determinar la integral anterior en cada punto por medio de integración numérica.

Una alternativa para calcular la integral y obtener los pronósticos h -pasos adelante $\hat{Y}_{E,T+h}$ para Y_{T+h} , consiste en utilizar la relación de Chapman-Kolmogorov para calcular los pronósticos h -pasos adelante en el método exacto, es decir, sea

$$\eta(y_{T+h}|\mathfrak{F}_T) = \int_{-\infty}^{\infty} \eta(y_{T+h}|Y_{T+h-1}) \eta(y_{T+h-1}|\mathfrak{F}_T) dy_{T+h-1},$$

donde $\eta(y_{T+h}|Y_{T+h-1})$ es la función de densidad de Y_{T+h} dado Y_{T+h-1} , entonces,

$$\begin{aligned} \hat{Y}_{E,T+h} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(y_{T+h-1}) \eta(y_{T+h-1}|Y_{T+h-2}) \times \\ &\quad \eta(y_{T+h-2}|\mathfrak{F}_T) dy_{T+h-1} dy_{T+h-2} \\ &\quad \vdots \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(y_{T+h-1}) \eta(y_{T+h-1}|Y_{T+h-2}) \times \\ &\quad \cdots \eta(y_{T+1}|Y_T) \eta(y_T|\mathfrak{F}_T) dy_{T+h-1} dy_{T+h-2} \cdots dy_T. \end{aligned}$$

Dado que no existe una expresión analítica general para las integrales anteriores, es necesario aproximar dichas integrales mediante métodos numéricos, Tong (1990). Las integrales se complican cada vez más si el número de pronósticos hacia adelante aumenta. En este sentido es que se han propuesto una serie de métodos numéricos encaminados a resolver el problema de pronósticos para modelos SETAR y en general para modelos no lineales.

Método Naive. Supongamos que $\varepsilon_{T+1} = 0$, esto significa que podemos intercambiar el operador esperanza por la función g , dado que se cumplen las siguientes igualdades,

$$\begin{aligned} Y_{T+1} &= g(Y_T) = \hat{Y}_{T+1} \\ E[g(Y_{T+1})|\mathfrak{F}_T] &= E\left[g(\hat{Y}_{T+1})|\mathfrak{F}_T\right] = g(\hat{Y}_{T+1}). \end{aligned}$$

De modo que el pronóstico dos pasos adelante está dado por:

$$\hat{Y}_{N,T+2} = E[Y_{T+2}|\mathfrak{F}_T] = E[g(Y_{T+1})|\mathfrak{F}_T] = g(Y_{T+1}) = g(\hat{Y}_{T+1}).$$

Equivalentemente, si $\varepsilon_{T+1} = \varepsilon_{T+2} = 0$, el pronóstico tres pasos adelante para Y_{T+3} es:

$$\hat{Y}_{N,T+3} = E[Y_{T+3}|\mathfrak{F}_T] = E[g(Y_{T+2})|\mathfrak{F}_T] = g(\hat{Y}_{N,T+2}).$$

Finalmente, si $\varepsilon_{T+1} = \dots = \varepsilon_{T+h-1} = 0$, el pronósticos h -pasos adelante se obtiene de la siguiente igualdad:

$$\hat{Y}_{N,T+h} = E[Y_{T+h} | \mathfrak{F}_T] = g\left(\hat{Y}_{N,T+h-1}\right),$$

donde el subíndice N en el pronóstico h -pasos adelante corresponde a la identificación del método naive.

Método de Monte Carlo. Una alternativa para calcular pronósticos h -pasos hacia adelante consiste en aproximar la esperanza condicional respectiva por medio de un procedimiento computacional recursivo del tipo Monte Carlo, por ejemplo, el pronóstico dos pasos adelante se aproxima mediante la siguiente ecuación:

$$\hat{Y}_{MC,t+2} = \frac{1}{N} \sum_{j=1}^N g\left(\hat{Y}_{T+1} + Z_j\right)$$

donde Z_j , $j = 1, \dots, N$ son extraídos de una función de distribución D conocida.

Bajo este esquema el pronóstico Monte Carlo para Y_{T+h} con $h > 1$ es de la forma:

$$\hat{Y}_{MC,T+h} = \frac{1}{N} \sum_{j=1}^N \underbrace{g(g(\dots g(\hat{Y}_{T+1} + Z_j^1) \dots))}_{h-1 \text{ veces}} + Z_j^{h-1},$$

con $Z_j^1, Z_j^2, \dots, Z_j^{h-1}$ generados de la distribución D , para $j = 1, 2, \dots, N$. Cabe mencionar que este método es el de mayor uso en la práctica, dada la facilidad de implementación y las propiedades de convergencia para N suficientemente grande.

Método de Remuestreo. Similar al método de Monte Carlo, el método de remuestreo consiste en aproximar la esperanza condicional, mediante la expresión:

$$\hat{Y}_{B,T+2} = \frac{1}{n-1} \sum_{j=1}^{n-1} g\left(\hat{Y}_{T+1} + \hat{\varepsilon}_j\right),$$

donde a diferencia del método de Monte Carlo los $\hat{\varepsilon}_j$, $j = 1, \dots, n-1$ son valores extraídos de los residuales observados $\hat{\varepsilon}_t$ sobre el periodo muestral. Equivalentemente tenemos que el pronóstico h -pasos adelante se aproxima por medio de la siguiente igualdad:

$$\hat{Y}_{B,T+h} = \frac{1}{n-1} \sum_{j=1}^{n-1} \underbrace{g(g(\dots g(\hat{Y}_{T+1} + \hat{\varepsilon}_j^1) \dots))}_{h-1 \text{ veces}} + \hat{\varepsilon}_j^{h-1}, \quad h > 1.$$

donde $\hat{\varepsilon}_j^1, \hat{\varepsilon}_j^2, \dots, \hat{\varepsilon}_j^{h-1}$ son $h-1$ valores aleatorios extraídos de los residuales observados, $j = 1, 2, \dots, n-1$.

Para más detalles ver Granger y Teräsvirta (1993); Clements y Smith (1997); Clements *et al* (2003); Diebold *et al* (1998); Diebold y Mariano (1995).

En la práctica la función $g(\cdot)$ no es conocida y deberá ser especificada y estimada por medio de algún procedimiento de estimación. La función estimada $\hat{g}(\cdot)$ se sustituye en cada una de los métodos anteriores para obtener los pronósticos de interés.

Cada uno de los métodos anteriores son alternativas diferentes, con ventajas y desventajas, por ejemplo: los pronósticos mediante el método naive son fáciles de obtener pero usualmente sesgados e inconsistentes ya que $E[g(\varepsilon)] \neq g[E(\varepsilon)]$. Los pronósticos mediante el método exacto son computacionalmente costosos y complicados, mientras que los pronósticos mediante el método de Monte Carlo, pueden ser sesgados si seleccionamos una función de distribución D incorrecta, sin embargo en condiciones generales suele ser la mejor técnica. Los pronósticos por el método de remuestreo son relativamente fáciles y pueden tener buenas propiedades siempre que $\hat{g}(\cdot)$ sea una buena aproximación a $g(\cdot)$, pero en otras situaciones pueden ser sesgados, Granger y Teräsvirta (1993).

1.2 Técnicas de evaluación y obtención de pronósticos

Gran parte del estudio de obtención y evaluación de pronósticos se ha enfocado a pronósticos puntuales, donde el resumen de la incertidumbre de los pronósticos se mide a través de sus errores estándar, Clements y Smith (1997), Clements *et al* (2003), Diebold *et al* (1998), Diebold y Mariano (1995), Dacco y Sanchell (1999), etc.. Chistosffersen (1998) sugiere algunas formas de evaluar los desempeños de los pronósticos mediante la evaluación de intervalos de pronósticos condicionales; mientras que Diebold *et al* (1998) proponen algunos métodos para evaluar la densidad de pronóstico.

Pronósticos puntuales

La obtención de pronósticos puntuales h -pasos adelante en modelos no lineales involucra, por lo general, la solución analítica de integrales múltiples y de aquí el uso de técnicas de integración numérica o métodos de simulación alternativos para la obtención de los pronósticos. Dentro de los métodos más utilizados se encuentran los métodos Monte Carlo del tipo recursivo los cuales aproximan la esperanza condicional correspondiente mediante el promedio de un número predeterminado de réplicas para cada uno de los horizontes de pronóstico. Ver por ejemplo, Clements y Smith (1997, 1998), De Gooijer y De Bruin (1998), Davies *et al* (1988). La comparación de los desempeños de los pronósticos bajo el modelo TAR y bajo el modelo AR se realiza por lo general tomando el cociente de los ECMP para el modelo AR entre los ECMP para el modelo TAR, considerando cada uno de los horizontes de pronóstico. De esta forma, valores mayores que uno indican que los desempeños de los pronósticos en el modelo TAR son mejores que los pronósticos en el modelo AR. Algunas de las conclusiones

de los trabajos de Clements y Smith (1997), Clements *et al* (2003), Diebold *et al* (1998), Diebold y Mariano (1995), Dacco y Sanchell (1999), en torno al estudio de los pronósticos en modelos SETAR, son las siguientes:

► Se sabe que el comportamiento de los pronósticos en modelos no lineales dentro de la muestra es mejor que el comportamiento de los pronósticos en modelos lineales, pero fuera de la muestra no, a menos que la no linealidad caracterice el periodo de pronóstico.

► Los desempeños de los pronósticos a largo plazo (más de tres pasos hacia adelante), son comparables tanto para modelos lineales autorregresivos como para modelos SETAR.

► Los desempeños de los pronósticos en modelos SETAR son mejores que los desempeños de los pronósticos en modelos AR, siempre que los valores de las pendientes en cada uno de los regímenes sean de signo opuesto y cercanos al círculo unitario (valores ± 0.6 y ± 0.8).

► Si la comparación entre los pronósticos de un modelo SETAR y un modelo AR se realiza condicionando en cada uno de los regímenes se tiene que la ganancia de los pronósticos en el modelo SETAR es mayor que en el modelo AR.

Diebold y Mariano (1995) realizan la comparación de desempeños de pronósticos puntuales mediante el contraste de las alternativas de pronósticos (lineal y no lineal), bajo la hipótesis nula de igualdad de precisión de pronóstico, para alguna función de pérdida arbitraria pero definida en términos de los errores de pronóstico $g(e_{i,t|t-h})$, donde $e_{i,t|t-h} = y_t - \hat{Y}_{i,t|t-h}$, $i = 1, 2$ corresponde a los errores de pronósticos h -pasos adelante bajo la alternativa de pronóstico i . La propuesta consiste en definir la pérdida diferenciada como:

$$d_t = g(e_{i,t|t-h}) - g(e_{j,t|t-h}),$$

con $i, j = 1, 2$, $i \neq j$, si $E(d_t) = 0$ entonces tenemos igualdad en la precisión de los pronósticos. Dada una realización muestral estacionaria en covarianzas $\{d_t\}_{T+h}^{T+n}$ la distribución asintótica de la pérdida diferenciada muestral, \bar{d} , con

$$\bar{d} = \frac{1}{n-h+1} \sum_{T+h}^{T+n} d_t,$$

está dada por: $\sqrt{n-h+1}(\bar{d} - \mu) \xrightarrow{d} N(0, V(\bar{d}))$, donde

$$V(\bar{d}) = \frac{1}{n-h+1} \left(\gamma_0 + 2 \sum_{i=0}^{h-1} \gamma_i \right),$$

bajo el supuesto de que los pronósticos h -pasos adelante exhiben dependencias de orden mayor que $h-1$. De esta forma el estadístico de Diebold y Mariano (1995) se basa en el

siguiente estadístico de prueba para probar la hipótesis nula de igualdad de precisión de pronóstico:

$$\frac{\bar{d}}{\sqrt{\tilde{V}(\bar{d})}} \stackrel{aprox.}{\sim} N(0, 1),$$

donde $\tilde{V}(\bar{d})$ es un estimador consiste de $V(\bar{d})$, basado en una suma ponderada de autocovarianzas muestrales, $\hat{\gamma}_i$, con

$$\hat{\gamma}_i = \frac{1}{n-h+1} \sum_{T+h-i}^{T+n} (d_t - \bar{d})(d_{t-i} - \bar{d}).$$

Clements *et al* (2003) presentan un estudio detallado sobre el comportamiento del estadístico de prueba de Diebold y Mariano (1995) bajo distintos escenarios, mostrando la robustez del mismo.

Evaluación mediante la densidad de pronóstico

Diebold *et al* (1998), proponen evaluar el desempeño de los pronósticos un paso adelante mediante el cálculo de la densidad de pronóstico. La evaluación de la densidad de pronóstico se basa en el análisis de las probabilidades de transformación integral (*pti*) de las realizaciones actuales de las variables sobre el periodo de pronóstico con respecto a la densidad de pronóstico de los modelos SETAR y AR, por ejemplo. Las probabilidades *pti* se denotan por $p_t(y_t)$, $t = T+1, T+2, \dots, T+n$, de modo que,

$$z_t = \int_{-\infty}^{y_t} p_t(u) du, \quad t = T+1, \dots, T+n.$$

Si la densidad de pronóstico corresponde a la verdadera densidad predictiva, dada por el verdadero proceso generador de datos, y denotada por: $f_t(y_t)$, es decir, $p_t(y_t) = f_t(y_t)$, tenemos que la secuencia de probabilidades de transformación integral $\{z_t\}_{T+1}^{T+n}$ son tal que las variables aleatorias z_t son independientes e idénticamente distribuidas con distribución uniforme en el intervalo $[0, 1]$, y que denotaremos como $z_t \sim iidU[0, 1]$. Algunas de las opciones para verificar uniformidad son: por inspección visual de la función de distribución empírica para la serie z_t , y mediante la prueba de Kolmogorov-Smirnov, la cual se basa en la desviación máxima entre la función de distribución empírica de los z_t y la función de distribución teórica. Berkowitz (2001) sugiere tomar las transformaciones inversas de la función de distribución acumulada normal de la serie $\{z_t\}_{T+1}^{T+n}$, y denotadas por $\{z_t^*\}_{T+1}^{T+n}$. Note que ahora los z_t^* bajo la hipótesis nula son variables aleatorias normales estándar e independientes.

Clements *et al* (2003) argumentan que los métodos de evaluación de la densidad predictiva, bajo el supuesto de errores normales en el proceso generador de datos, SETAR, puede verse como un método de evaluación de pronóstico puntual ya que la secuencia de probabilidades normales $\{z_t^*\}$ son equivalentes a los errores de pronósticos escalados. Por ejemplo, supongamos el siguiente modelo SETAR(2; 1, 1),

$$y_t = (\phi_1 y_{t-1} + \varepsilon_t) I(y_{t-1} \leq \gamma) + (\phi_2 y_{t-1} + \varepsilon_t) I(y_{t-1} > \gamma), \quad t = 1, 2, \dots, T + n,$$

donde la función $I(\cdot)$ corresponde a la función indicadora, T denota el tamaño de la muestra y n la longitud de el periodo de pronóstico fuera de la muestra. El pronóstico un paso adelante está dado por:

$$\begin{aligned} \hat{y}_{t+1} &= E(y_{t+1} | \mathcal{F}_t) \\ &= \phi_1 y_t I(y_t \leq \gamma) + \phi_2 y_t I(y_t > \gamma), \end{aligned} \tag{1.3}$$

donde \mathcal{F}_t denota como antes el conjunto de la información disponible hasta el tiempo t o equivalentemente la σ -álgebra generada por el conjunto de observaciones disponible hasta el tiempo t . Bajo el supuesto de errores gaussianos la densidad de pronóstico para el modelo AR(1) es $p_t(y_t) = N(\hat{y}_{t+1}^{AR}, (\hat{\sigma}_t^{AR})^2)$, con $\hat{\sigma}_t^{AR}$ la desviación estándar estimada para el proceso de error y \hat{y}_{t+1}^{AR} el pronóstico un paso adelante para y_t , ambos bajo el modelo AR(1). Para el caso del modelo SETAR(2; 1, 1) tenemos que la secuencia de probabilidades $\{z_t\}$ son obtenidas por $z_t = \Phi\left(\frac{y_t - \hat{y}_{t+1}}{\hat{\sigma}_t}\right)$, con \hat{y}_{t+1} dado en la ecuación (1.3) y $\hat{\sigma}_t^2$ la varianza estimada para el proceso de error bajo el modelo SETAR. Observemos que la serie $\{z_t^*\}$ se obtiene aplicando la inversa de la función de distribución normal estándar a la serie $\{z_t\}$ de tal forma que obtenemos los errores estándar originales $(y_t - \hat{y}_{t+1})/\hat{\sigma}_t$.

Una conclusión general en torno al estudio de los desempeños de los pronósticos en modelos SETAR contra los desempeños de los pronósticos en modelos AR, e independientemente del método de evaluación, es que la ganancia de pronóstico es mayor en el modelo SETAR, siempre que el pronóstico y la medida de desempeño utilizada se obtengan condicionando a cada uno de los regímenes. En nuestro caso, y como se discutirá en la siguiente sección, nos interesa estudiar los desempeños de los pronósticos para modelos del tipo TAR, donde la variable umbral corresponde a una variable exógena con dinámica propia y donde, posiblemente, uno de los regímenes posee una raíz unitaria y de aquí que la noción de condicionamiento para la obtención y evaluación de los pronósticos no es aplicable.

1.3 Motivación

Uno de los objetivos principales en el análisis de series de tiempo es la obtención de pronósticos, basados en modelos significativos en el sentido de un buen ajuste y una coherente interpretación del fenómeno bajo estudio. Nuestro interés particular se centra en aplicaciones económicas, en las cuales se ha observado que los modelos autorregresivos con umbral son capaces de capturar características comúnmente observadas en series económicas, como son: irreversibilidad en el tiempo, asimetrías, persistencia, etc.. Asimismo, desde el punto de vista econométrico, es más realista suponer que los cambios de un régimen a otro a través del tiempo se deben a rezagos o función de rezagos de una variable aleatoria exógena más que a rezagos de la serie misma. Más aún, existen situaciones prácticas donde suponer un modelo lineal autorregresivo conlleva a considerar que dicho fenómeno se comporta como un proceso de raíz unitaria. Por ejemplo, consideremos los datos trimestrales de las tasas de interés de USA, en el periodo de agosto de 1959 a junio de 1999, obtenida de Citibase y analizada en González y Gonzalo (1999). El comportamiento gráfico de los datos para las tasas de interés se da en la Figura 1. Al realizar la prueba de raíz unitaria para dichos datos, concluimos que no se rechaza la hipótesis nula de raíz unitaria para distintos niveles de significancia, como se muestra en la Tabla 1. Sin embargo la teoría económica argumenta que no es realista asumir que las series de tasas de interés sigan un modelo de raíz unitaria, ya que dicho supuesto carece de utilidad en cuanto a la obtención y estudio del comportamiento futuro de la serie. González y Gonzalo (1998, 1999) estudian el comportamiento de las tasas de interés por medio de un modelo TAR de dos regímenes, con la característica de que uno de los regímenes posee una raíz unitaria y denotan a dichos modelos como modelos con umbral de raíz unitaria (TUR models). La ventaja de dicha consideración es que el modelo utilizado es un modelo estacionario, que describe adecuadamente a los datos y con la característica de que la interpretación de cada uno de los regímenes resalta la naturaleza misma de la serie, como por ejemplo el que uno de los regímenes se considere como una fase de recesión o bien como una fase de crecimiento económico. González y Gonzalo (1998, 1999) presentan las condiciones teóricas bajo las cuales los modelos TAR son estacionarios y ergódicos; se dan también algunas sugerencias para posibles variables de umbral, bajo algunos escenarios económicos. Un ejemplo de variable umbral en el estudio de las tasas de interés es la serie que describe los cambios en la inflación. Note que los modelos SETAR no pueden contener una raíz unitaria en algunos de los regímenes ya que el cambio se da en función de valores rezagados de la misma serie y ésta no cumpliría el supuesto de estacionariedad y ergodicidad, requerido para el análisis y el cumplimiento de algunos resultados asintóticos para los parámetros estimados en el modelo de interés.

Tabla1. Prueba Aumentada de Dickey-Fuller para los
datos de las tasas de interés en la Figura 1.

ADF test Statistic = -2.415	Valor crítico del 1%	-3.446
	Valor crítico del 5%	-2.868
	Valor crítico del 10%	-2.570

Comportamiento gráfico de las tasas de interés

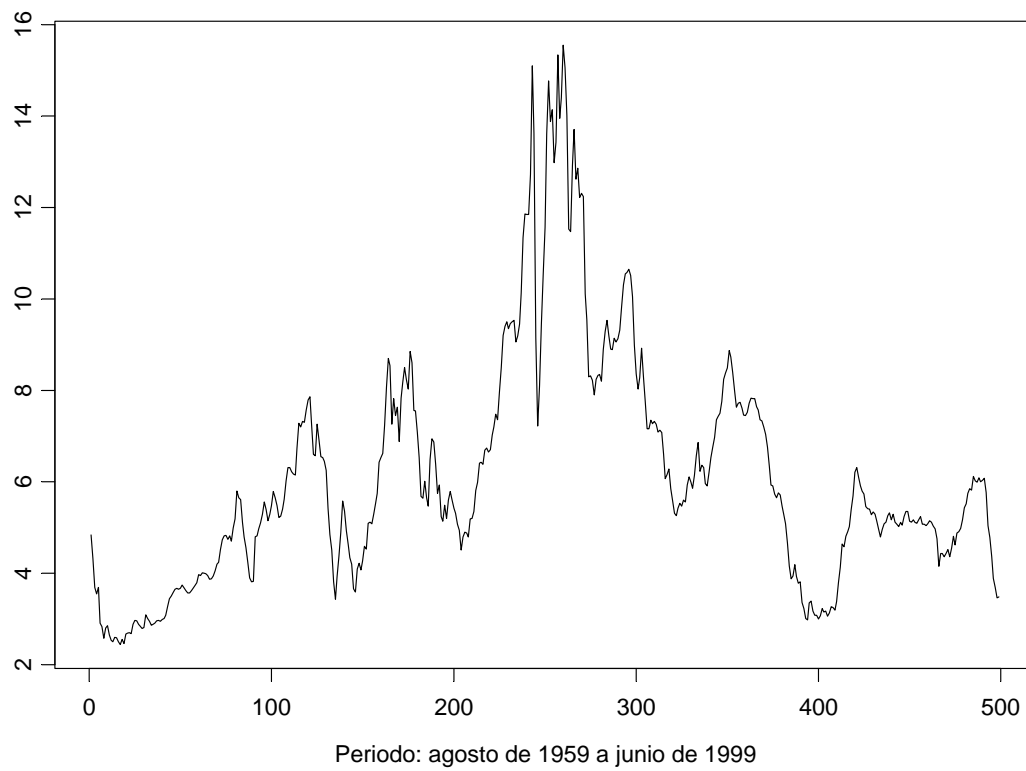


Figura 1. Datos de tasas de interés analizados en González y Gonzalo (1999).

1.4 Identificación del problema y alternativas de solución

La mayor parte de la literatura, en torno a la problemática de obtención y cuantificación sobre los desempeños de pronósticos en modelos TAR contra modelos AR, se ha concentrado en el estudio de los modelos SETAR. Se sabe que los estimadores de los parámetros autorregresivos en el modelo SETAR, bajo parámetro de umbral conocido y desconocido satisfacen las propiedades de consistencia y normalidad asintótica, Petrucci y Woolford (1984), Chan *et al* (1993). En relación a la obtención y cuantificación de los desempeños de los pronósticos en modelos SETAR, se ha encontrado que en general estos no presentan mejores desempeños que los pronósticos generados a partir de un modelo lineal AR. La conclusión general es que los ECM de pronósticos en el modelo lineal AR son menores o iguales que los ECM de pronósticos en el modelo SETAR, aún en el caso en que el verdadero proceso generador de datos (VPGD) este dado por un modelo SETAR. A la fecha, no existe un consenso general en cuanto a, ¿Cuáles son las características inherentes al modelo que originan dicho comportamiento?, o ¿En qué situaciones esperamos que los pronósticos bajo dichos modelos se comporten en forma similar?.

La discusión del mal desempeño de los pronósticos en modelos SETAR, se relaciona con el problema de mala especificación de los regímenes. Dacco y Sanchell (1999) ilustran la problemática anterior en los llamados modelos Regime-Switching (RS) contra los modelos de caminas aleatorias con y sin deriva. Se calculan los ECMP, suponiendo que el verdadero proceso generador de datos es el modelo de RS, pero considerando que el pronóstico un paso adelante se calcula por medio de cada una de las alternativas de modelación. Los ECMP se expresan en función de las probabilidades de mala especificación en los regímenes, observándose que el mal desempeño de los pronósticos se debe principalmente a valores grandes de dichas probabilidades. Se sugiere también que el criterio del ECM puede no ser el criterio apropiado para la evaluación de los desempeños de los pronósticos, y de aquí que necesitamos establecer métodos de evaluación que consideren las características no lineales que caracterizan a los modelos TAR.

En nuestro caso identificamos el problema de mala especificación, debido a clasificaciones erróneas de las observaciones en los regímenes y la mala especificación debida al suponer un modelo distinto del verdadero proceso generador de datos. De esta forma, valores grandes de las probabilidades de mala clasificación implican valores grandes en los ECM. Paralelamente, asumir un modelo erróneo al VPDG implica un aumento en los ECM de pronóstico bajo el modelo incorrecto, más aún, el comportamiento de los residuales no corresponde al comportamiento de un proceso de ruido blanco. El aumento en los valores de los ECM de pronóstico bajo un modelo mal especificado no es suficiente para garantizar que los ECMP en los modelos TAR son de menor magnitud que los ECMP en los modelos AR, esto debido a que la metodología de estimación de los parámetros en el modelo es independiente del

proceso de evaluación de los pronósticos, tal y como se ha estudiado en la literatura. Por ejemplo, una práctica común es obtener las estimaciones de los parámetros vía la metodología de mínimos cuadrados y posteriormente calcular y evaluar los pronósticos por medio de un algoritmo de Monte Carlo.

Nuestras propuestas de solución abordan el problema de estimación, obtención y evaluación de los pronósticos bajo la misma metodología, tomando en consideración las características de no linealidad inherentes en el modelo TAR. La primera de nuestras alternativas es conocida como verosimilitud predictiva, y consiste en obtener los estimadores de los parámetros y valores futuros más plausibles vía la maximización de la función objetivo resultante e identificada como verosimilitud predictiva perfil. La segunda de nuestras propuestas se conoce como regresión por cuantiles, y consiste en minimizar la función objetivo de cuantiles para obtener primeramente los parámetros estimados y posteriormente los pronósticos h -pasos adelante. Esta alternativa considera adecuadamente las características de asimetrías de los modelos TAR, lo cual se refleja en mejores desempeños de los pronósticos bajo el modelo TAR. Los detalles teóricos de ambas alternativas son estudiados en los Capítulos 5 y 6 respectivamente. En el Capítulo 7 presentamos un amplio ejercicio de simulación donde comparamos los desempeños de los pronósticos para cada una de las alternativas antes mencionadas y de donde concluimos que: en términos generales los pronósticos bajo el modelo TAR por medio de verosimilitud predictiva y cuantiles de regresión son mejor comportados que los pronósticos bajo el modelo AR, en virtud de que los valores de las funciones de pérdidas bajo el modelo TAR son menores en magnitud que los valores respectivos bajo el modelo AR.

Capítulo 2

El modelo TAR

Sea $Y = (Y_1, \dots, Y_T)'$ el vector de datos observados, con condiciones iniciales $(Y_0, Y_{-1}, \dots, Y_{-p+1})'$. Se dice 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 forma siguiente:

$$Y_t = \alpha_{0,k} + \sum_{i=1}^{p_k} \alpha_{i,k} Y_{t-i} + \varepsilon_{k,t} \text{ para } \gamma_{k-1} \leq Z_{t-d} < \gamma_k, \quad (2.1)$$

donde $k = 1, 2, \dots, r$ y d un entero positivo, en nuestro caso conocido. Los errores $\{\varepsilon_{k,t}\}$ se consideran una secuencia de variables aleatorias normales independientes con media cero y varianza σ_k^2 , donde $\{\varepsilon_{i,t}\}$ y $\{\varepsilon_{j,t}\}$ son independiente si $i \neq j$. Los números reales γ_j satisfacen $-\infty = \gamma_0 < \gamma_1 < \dots < \gamma_r = \infty$ y forman una partición del espacio de Z_{t-d} . El entero positivo d se conoce como rezago o *retardo*¹ de la variable umbral en el modelo. La partición $\gamma_{k-1} \leq Z_{t-d} < \gamma_k$ se refiere a el k -ésimo régimen del modelo en la ecuación (2.1). Note que el modelo TAR es un modelo lineal por pedazos en el espacio de Z_{t-d} , pero es no lineal por pedazos en el tiempo. Si la variable umbral Z_{t-d} está dada por rezagos o funciones de rezagos de la serie misma Y_t , identificamos al modelo de la ecuación (2.1) como un modelo SETAR.

2.1 Algunas definiciones y resultados

Al igual que en el análisis de series de tiempo lineales, las propiedades de estacionariedad y ergodicidad para el caso de series de tiempo no lineales son fundamentales para garantizar las propiedades asintóticas de los estimadores para cada uno de los parámetros en el modelo. Aún cuando podemos decir que para modelos de series de tiempo lineales es relativamente

¹del inglés delay

sencillo mostrar que la serie en cuestión es estacionaria, en el caso de modelos no lineales dicha propiedad no es fácil de establecer. De esta forma, el que una serie de tiempo no lineal posea o no la propiedad de estacionariedad dependerá del modelo particular de interés. La manera común de verificar los supuestos de estacionariedad y ergodicidad en modelos no lineales es representar a la serie de tiempo como una cadena de Markov y establecer que la cadena es ergódica. La serie de tiempo no lineal es estacionaria dado que una cadena de Markov ergódica es estacionaria. Vale la pena entonces considerar las siguientes definiciones, en torno a los conceptos de interés.

Definición 2.1 (*Cadena de Markov*). Sea $\{\mathbf{Y}_t\}$ un proceso estocástico de valores reales, decimos que $\{\mathbf{Y}_t\}$ es una cadena de Markov si cumple con la propiedad Markoviana, es decir, si la distribución condicional de \mathbf{Y}_{t+1} dado $\{\mathbf{Y}_t, \mathbf{Y}_{t-1}, \dots\}$ depende sólo de $\mathbf{Y}_t, \forall t$. La distribución condicional de $\mathbf{Y}_{t+1}|\mathbf{Y}_t$, digamos $F_t(\cdot|\mathbf{y})$, se conoce como la distribución de transición de la cadena al tiempo t .

Consideremos el siguiente modelo autorregresivo no lineal,

$$Y_t = f(Y_{t-1}, \dots, Y_{t-p}) + \varepsilon_t, \quad (2.2)$$

donde $\{\varepsilon_t\}$ es una secuencia de variables aleatorias independientes e idénticamente distribuidas y ε_t independiente de $\{Y_{t-k}, k \geq 1\}$. De la ecuación (2.2) se sigue que $\{\mathbf{Y}_t\}$ es una cadena de Markov dada por:

$$\mathbf{Y}_t = \mathbf{f}(\mathbf{Y}_{t-1}) + \boldsymbol{\varepsilon}_t, \quad (2.3)$$

con $\mathbf{Y}_t = (Y_{t-1}, \dots, Y_{t-p})'$, $\boldsymbol{\varepsilon}_t = (\varepsilon_t, 0, \dots, 0)'$ y $\mathbf{f}(\mathbf{y}) = (f(y_1), y_1, \dots, y_{p-1})'$ para $\mathbf{y} = (y_1, \dots, y_p)' \in \mathbb{R}^p$.

Definición 2.2 (*Distribución de transición*). Sea $D(\cdot)$ la función de distribución de $\boldsymbol{\varepsilon}_t$, y $F_n(\cdot|\mathbf{y})$ la distribución condicional de \mathbf{Y}_n dado $\mathbf{Y}_0 = \mathbf{y}$, entonces para $n \geq 2$,

$$F_n(\mathbf{x}|\mathbf{y}) = \int D\{\mathbf{x} - \mathbf{f}(\mathbf{u})\} F_{n-1}(d\mathbf{u}|\mathbf{y}),$$

y $F_1(\mathbf{x}|\mathbf{y}) = D\{\mathbf{x} - \mathbf{f}(\mathbf{y})\}$ es conocida como la distribución de transición de la cadena.

Definición 2.3 (*Ergodicidad*). Si existe una distribución F y una constante $\lambda \in (0, 1]$, tal que

$$\lambda^{-n} \|F_n(\cdot|\mathbf{y}) - F(\cdot)\| \longrightarrow 0 \text{ para cualquier } \mathbf{y},$$

decimos que la cadena de Markov \mathbf{Y}_t es ergódica si $\lambda = 1$ y geoméricamente ergódica si $\lambda < 1$. F se conoce como la distribución estacionaria y $\|\cdot\|$ denota la norma de variación total.

Note que ergodicidad geométrica implica ergodicidad y que la ergodicidad de una cadena de Markov depende completamente de su distribución estacionaria.

Resultado 2.4 (*Estacionariedad estricta*). Suponga que la cadena de Markov \mathbf{Y}_t dada en (2.3) es ergódica. Entonces existe una distribución estacionaria (p -dimensional) F tal que la serie de tiempo $\{Y_t, t = 1, 2, \dots\}$ definida por el modelo no lineal dado en (2.2) y con condiciones iniciales $(Y_0, Y_{-1}, \dots, Y_{-p+1})' \sim F$ es estrictamente estacionaria.

Para la demostración del Resultado 2.4 ver Fan y Yao (2003).

Las propiedades de estacionariedad y ergodicidad de los modelos SETAR se deben esencialmente a los trabajos de Petrucci y Woolford (1984), Chan *et al* (1985), Chen y Tsay (1991), entre otros.

El Teorema de ergodicidad para el modelo SETAR($r; 1, 1$) establecido en Chan *et al* (1985) dice lo siguiente:

Teorema 2.5 (*Teorema de ergodicidad, SETAR*). El proceso $\{Y_t\}$ definido en la ecuación (2.1), con $Z_{t-d} = Y_{t-d}$ es ergódico si y sólo si se satisface cualquiera de las siguientes condiciones:

- a) $\alpha_{1,1} < 1, \alpha_{1,r} < 1, \alpha_{1,1}\alpha_{1,r} < 1$;
- b) $\alpha_{1,1} = 1, \alpha_{1,r} < 1, \alpha_{0,1} > 0$;
- c) $\alpha_{1,1} < 1, \alpha_{1,r} = 1, \alpha_{0,r} < 0$;
- d) $\alpha_{1,1} = 1, \alpha_{1,r} = 1, \alpha_{0,r} < 0 < \alpha_{0,1}$;
- e) $\alpha_{1,1}\alpha_{1,r} = 1, \alpha_{1,1} < 0, \alpha_{0,r} + \alpha_{1,r}\alpha_{0,1} > 0$.

En el caso de dos regímenes y constante cero, Petrucci y Woolford (1984), y Chen y Tsay (1991) demuestran que $\{Y_t\}$ es ergódico si y sólo si $\alpha_{1,1} < 1, \alpha_{1,2} < 1$ y $\alpha_{1,1} \alpha_{1,2} < 1$. La demostración del teorema de ergodicidad parte del supuesto de que la serie de interés es una cadena de Markov con espacios de estados $(\mathfrak{R}, \mathcal{B})$, donde \mathcal{B} es la σ -álgebra de Borel en los números reales \mathfrak{R} , y con densidad de transición dada por:

$$p(x, y) = \sum_{k=1}^l I(x \in \gamma_k) f_k(y - \alpha_{0,k} - \alpha_{l,k}x).$$

Nuestro interés particular se refiere al estudio de los modelo TAR de primer orden, sin constante, estudiado a detalle en González y Gonzalo (1998), con variable umbral Z_{t-d} , cuya representación matemática es,

$$\begin{aligned} Y_t &= [\alpha_1 I(Z_{t-d} \leq \gamma_1) + \dots + \alpha_T I(Z_{t-d} > \gamma_{T-1})] Y_{t-1} + \varepsilon_t \\ &= \delta_t Y_{t-1} + \varepsilon_t, \quad t = 1, 2, \dots, \end{aligned} \tag{2.4}$$

donde $\delta_t = \alpha_1 I(Z_{t-d} \leq \gamma_1) + \dots + \alpha_T I(Z_{t-d} > \gamma_{T-1})$, $I(\cdot)$ es la función indicadora, y los procesos ε_t y Z_t satisfacen las siguientes condiciones:

- S1. Los procesos (ε_t, Z_{t-d}) son estrictamente estacionarios y ergódicos.
- S2. $E(\varepsilon_t | \mathfrak{F}_{t-1}) = 0$ y $E(\varepsilon_t^2 | \mathfrak{F}_{t-1}) = \sigma^2$.
- S3. Para algún $\tau > 1$, $E(\varepsilon_t^{2\tau} | \mathfrak{F}_{t-1}) \leq B < \infty$.
- S4. $E(\max(0, \log |\varepsilon_1|)) < \infty$.
- S5. El supremo esencial de $|\varepsilon_1| < \infty$, es decir, $\text{ess sup } |\varepsilon_1| = \inf \{x : \Pr(|\varepsilon_1| > x) = 0\} < \infty$.
- S6. ε_1 admite una función de densidad de probabilidad continua y positiva.

Los supuestos (S1) y (S4) son necesarios para estacionariedad estricta, mientras que el supuesto (S5) es necesario para estacionariedad en covarianza. El supuesto (S6) se requiere para ergodicidad geométrica. Los supuestos (S2) son las consideraciones estándar para especificar que el proceso de error ε_t es una secuencia de martingalas en diferencias condicionalmente homoscedastica. (S2) conjuntamente con (S3) se utiliza para obtener algunos de los resultados asintóticos del proceso TAR.

Las condiciones bajo las cuales el modelo TAR, dado en la ecuación (2.4) es estacionario y ergódico se establecen en González y Gonzalo (1998), así como para algunos casos particulares y de interés práctico, los cuales consideraremos en nuestro ejercicio de simulación.

Teorema 2.6 (*Teorema de estacionariedad TAR*). *Sea Y_t un proceso generado a partir del modelo TAR dado en la ecuación (2.4), donde el proceso de error ε_t satisface los supuestos (S1), (S4) y (S5) enunciados arriba. Sea también Z_{t-d} la variable umbral, la cual cumple el supuesto (S1). Si $E(|\delta_t^2|) < 1$, entonces el proceso Y_t es estrictamente estacionario. Por otra parte, si $\sum_{j=1}^{\infty} \left(E \left| \prod_{T=1}^j \delta_T^2 \right| \right)^{1/2} < \infty$, el proceso es también débilmente estacionario.*

Note que el teorema anterior establece las condiciones bajo las cuales el proceso Y_t es estrictamente estacionario, y éstas se basan en que se cumplan los supuestos (S1) y (S4) para los procesos (ε_t, Z_{t-d}) . Sin embargo, esto no garantiza la existencia de los momentos y de aquí que necesitemos asumir que el proceso $\{\varepsilon_t\}$ cumple con la condición (S5) para garantizar que el proceso Y_t es estacionario en covarianzas. Note también que el Teorema 2.6 no requiere que la variable umbral Z_{t-d} sea una secuencia de observaciones independientes de ε_t . Por ejemplo, el proceso Z_{t-d} podría ser el proceso ε_t y las condiciones del teorema podrían seguir cumpliéndose. El teorema incluye el caso de los modelos SETAR siempre que se cumplan las condiciones $|\alpha_1| < 1$, $|\alpha_T| < 1$, y $|\alpha_1 \alpha_T| < 1$, de otra forma la variable de umbral no sería ergódica y no cumpliría el supuesto (S1).

Corolario 2.7 *Considere el modelo TAR de dos regímenes de primer orden, denotado como $TAR(2; 1, 1)$ con rezago d y dado por la siguiente expresión:*

$$\begin{aligned} Y_t &= [\phi_1 I(Z_{t-d} \leq \gamma) + \phi_2 I(Z_{t-d} > \gamma)] Y_{t-1} + \varepsilon_t \\ &= \delta_t Y_{t-1} + \varepsilon_t, \end{aligned} \quad (2.5)$$

donde la variable umbral Z_{t-d} se supone un proceso de variables aleatorias independientes e idénticamente distribuidas y mutuamente independientes del proceso de errores ε_t , con $p = \Pr(Z_{t-d} \leq \gamma)$. Entonces, si $E(\delta_t^2) < 1$, el proceso Y_t es estacionario en covarianzas.

Corolario 2.8 *Considere el modelo TAR de primer orden, dado en la ecuación (2.5) con Z_{t-d} un proceso de Markov de orden N . Entonces si $E(\delta_t^2) < 1$ y $p_{1|2,\dots,2} \geq p_{1|1,\dots,1}$ el proceso es estacionario en covarianzas. $p_{1|j,\dots,j}$ es la probabilidad de iniciar en el estado uno dado que durante los N periodos anteriores el proceso estuvo en el estado $j = 1, 2$. Cuando $N = 1$, tenemos que para que el proceso sea estacionario en covarianzas debe cumplirse que: $E(\delta_t^2) < 1$ y $p_{1|2} \geq p_{1|1}$.*

Corolario 2.9 *Considere el modelo TAR de primer orden dado en la ecuación (2.5) donde Z_{t-d} es un proceso de Markov de orden N y $\phi_i = 1$ y $|\phi_j| < 1$, $i \neq j$, $i, j = 1, 2$. Entonces el proceso Y_t es estacionario en covarianzas.*

El Corolario 2.9 introduce los modelos TAR con una raíz unitaria en uno de los regímenes, denotados por González y Gonzalo (1998) como modelos TUR, los cuales son una alternativa para la modelación de series económicas con comportamientos de raíces unitarias, como por ejemplo las tasas de interés.

Teorema 2.10 *(Teorema de ergodicidad TAR). Sea Y_t generado a partir del modelo dado en la ecuación (2.5), y que satisface las condiciones del teorema de estacionariedad, más el supuesto S6. Entonces, Y_t es geométricamente ergódico y por ende ergódico.*

Las demostraciones de los Corolarios 2.7, 2.8, 2.9 y los teoremas de estacionariedad y ergodicidad en modelos TAR, Teoremas 2.6 y 2.10, se encuentran en el trabajo de González y Gonzalo (1998).

2.2 Error cuadrático medio bajo especificaciones incorrectas

Recordemos que el problema inicial de este trabajo de tesis es investigar porqué en general los ECM de pronósticos en modelos lineales presentan mejores desempeños que los ECM de pronósticos en modelo TAR. En esta sección obtenemos las expresiones teóricas de los ECM asociados a los pronósticos para el modelo TAR y el modelo AR. Estudiamos el problema de clasificaciones erróneas de los pronósticos en los regímenes para un escenario simplificado, similar al caso estudiado en Dacco y Satchell (1999). Posteriormente analizamos los ECM de pronósticos considerando una mala especificación ocasionada ahora por suponer un modelo distinto del verdadero proceso generador de datos y generar a partir de dicho modelo los pronósticos h -pasos hacia adelante.

2.2.1 Clasificación incorrecta de las observaciones en los regímenes bajo el modelo TAR simplificado

El interés en esta sección consiste en investigar porqué, la ganancia de los pronósticos en los modelos TAR no es generalmente más grande que la ganancia en los modelos AR, en términos de los ECMP, aún en el caso en que el modelo TAR ajuste mejor a los datos. El procedimiento de análisis es como en Dacco y Satchell (1999), quienes presentan un trabajo dirigido al estudio de los modelos Regime-Switching en relación al mal desempeño de los pronósticos.

Supongamos que el verdadero proceso generador de datos está dado por la siguiente versión simplificada de la ecuación (2.4) y que llamaremos modelo TAR simplificado:

$$\begin{aligned} Y_t &= \phi_1 I(Z_{t-1} \leq \gamma) + \phi_2 I(Z_{t-1} > \gamma) + \varepsilon_t \\ &= \delta_t + \varepsilon_t, \end{aligned} \quad (2.6)$$

con $\delta_t = \phi_1 I(Z_{t-1} \leq \gamma) + \phi_2 I(Z_{t-1} > \gamma)$, $\varepsilon_t \sim iidN(0, \sigma^2)$ y Z_{t-1} una serie de tiempo estacionaria y mutuamente independiente del proceso ε_t tal que $\Pr(I(Z_{t-1} \leq \gamma) = 1) = p(\gamma)$ y $\Pr(I(Z_{t-1} \leq \gamma) = 0) = 1 - p(\gamma)$. Para efectos de análisis supongamos que los parámetros en el modelo (2.6) son conocidos.

Los primeros momentos del proceso δ_t , se obtienen como sigue:

$$\begin{aligned} \mu_\delta &= E(\delta_t) \\ &= E[\phi_1 I(Z_{t-1} \leq \gamma) + \phi_2 I(Z_{t-1} > \gamma)] \\ &= \phi_1 E(I(Z_{t-1} \leq \gamma)) + \phi_2 E(I(Z_{t-1} > \gamma)) \\ &= \phi_1 \Pr(Z_{t-1} \leq \gamma) + \phi_2 \Pr(Z_{t-1} > \gamma) \\ &= \phi_1 p(\gamma) + \phi_2 [1 - p(\gamma)], \end{aligned} \quad (2.7)$$

asimismo la condición de estacionariedad en covarianzas o segundo momento para el proceso δ_t es, $E(\delta_t^2) = \phi_1^2 p(\gamma) + \phi_2^2 [1 - p(\gamma)]$, de tal forma que la varianza de δ_t se reduce a:

$$\begin{aligned}\sigma_\delta^2 &= \text{Var}(\delta_t) \\ &= p(\gamma) [1 - p(\gamma)] (\phi_1 - \phi_2)^2.\end{aligned}\tag{2.8}$$

Los momentos del proceso Y_t están dados en función de los momentos de δ_t y ε_t para todo t , y se escriben a continuación:

$$E(Y_t) = E(\delta_t + \varepsilon_t) = E(\delta_t) = \phi_1 p(\gamma) + \phi_2 [1 - p(\gamma)],\tag{2.9}$$

y

$$\begin{aligned}\text{Var}(Y_t) &= \text{Var}(\delta_t + \varepsilon_t) = \text{Var}(\delta_t) + \text{Var}(\varepsilon_t) \\ &= p(\gamma) [1 - p(\gamma)] (\phi_1 - \phi_2)^2 + \sigma^2,\end{aligned}\tag{2.10}$$

ya que la variable umbral Z_{t-1} es independiente del proceso ε_t .

Para analizar el comportamiento de los pronósticos h -pasos hacia adelante, consideremos dos posibles modelos para la variable umbral Z_{t-1} . El primer caso consiste en considerar a Z_{t-1} como un proceso de variables aleatorias independientes e idénticamente distribuidas con distribución uniforme en $(0, 1)$, es decir, $Z_{t-1} \sim iidU(0, 1)$. Para el segundo caso consideramos cierta estructura de dependencia en la variable umbral, descrita por ejemplo por un modelo autorregresivo de orden uno, estacionario con errores gaussianos, es decir, $Z_{t-1} \sim AR(1)$.

Caso 1. $Z_{t-1} \sim iidU(0, 1)$

Note que al tiempo T hemos observado el valor de la variable umbral y de aquí que conocemos $I(Z_T \leq \gamma)$ ó $I(Z_T > \gamma)$, entonces el pronóstico un paso adelante para la serie Y_t está dado por:

$$\begin{aligned}\hat{Y}_{T+1}^{TAR} &= E\{Y_{T+1}|Y_T\} = E\{\phi_1 I(Z_T \leq \gamma) + \phi_2 I(Z_T > \gamma) + \varepsilon_{T+1} | \mathfrak{F}_T\} \\ &= \phi_1 I(Z_T \leq \gamma) + \phi_2 I(Z_T > \gamma),\end{aligned}$$

donde \mathfrak{F}_T contiene toda la información disponible hasta el tiempo T . El ECMP un paso adelante es igual a la varianza del error, es decir,

$$ECMP_1^{TAR} = E\left(Y_{T+1} - \hat{Y}_{T+1}\right)^2 = E(\varepsilon_{T+1})^2 = \sigma^2.$$

De la misma forma, tenemos que el pronóstico h -pasos adelante, se obtiene de la siguiente ecuación:

$$\begin{aligned}
 \hat{Y}_{T+h}^{TAR} &= E(Y_{T+h}|Y_T) = E[\phi_1 I(Z_{T+h-1} \leq \gamma) + \phi_2 I(Z_{T+h-1} > \gamma) + \varepsilon_{T+h} | \mathfrak{F}_T] \\
 &= \phi_1 E[I(Z_{T+h-1} \leq \gamma) | \mathfrak{F}_T] + \phi_2 E[I(Z_{T+h-1} > \gamma) | \mathfrak{F}_T] \\
 &= \phi_1 E[I(Z_{T+h-1} \leq \gamma)] + \phi_2 E[I(Z_{T+h-1} > \gamma)] \\
 &= \phi_1 p(\gamma) + \phi_2 [1 - p(\gamma)] \\
 &= \phi_1 \gamma + \phi_2 (1 - \gamma) \\
 &= E(\delta_t),
 \end{aligned}$$

por la independencia de Z_{t-1} , notando que $p(\gamma) = \gamma$, dado que $Z_{t-1} \sim iidU(0, 1)$. El ECMP h -pasos adelante para el modelo TAR simplificado está dado por la siguiente igualdad,

$$\begin{aligned}
 ECM P_h^{TAR} &= E\left(Y_{t+h} - \hat{Y}_{T+h}\right)^2 \\
 &= \gamma(1 - \gamma)(\phi_1 - \phi_2)^2 + \sigma^2.
 \end{aligned}$$

Para fines de comparación consideremos ahora el modelo lineal simplificado, dado por la siguiente ecuación:

$$Y_t = \phi + \varepsilon_t, \quad (2.11)$$

con $\varepsilon_t \sim iidN(0, \sigma^2)$. El pronóstico h -pasos hacia adelante, bajo el modelo (2.11), es simplemente dado por $\hat{Y}_{T+h}^L = \phi$, con $h \geq 1$. El parámetro ϕ puede estimarse por la media muestral, la cual converge en probabilidad a $\phi_0 = \phi_1 p(\gamma) + \phi_2 [1 - p(\gamma)] = E(\delta_t)$, bajo el hecho de que el modelo dado en la ecuación (2.6) es el VPGD. Finalmente el ECMP para el pronóstico h -pasos adelante, obtenido del modelo (2.11) está dado por:

$$\begin{aligned}
 ECM P_h^L &= E\left[Y_{T+h} - \hat{Y}_{T+h}^L\right]^2 \\
 &= \text{Var}(Y_{T+h}) = \gamma(1 - \gamma)(\phi_1 - \phi_2)^2 + \sigma^2.
 \end{aligned}$$

Entonces, si $Z_{t-1} \sim iidU(0, 1)$ tenemos que $ECM P_h^{TAR} = ECM P_h^L$. Sin embargo, en la práctica tenemos que los parámetros en el modelo son desconocidos, y es necesario estimarlos. La estimación de los parámetros se realiza mediante mínimos cuadrados o máxima verosimilitud y en ambos casos tenemos que los estimadores resultantes son consistentes y asintóticamente normales. Los desempeños de los pronósticos en el modelo TAR simplificado y en el modelo lineal dependen esencialmente de qué tan cercano está el estimador $\hat{\phi}$ del modelo lineal simplificado con el correspondiente valor estimado de $E(\delta_t)$.

Caso 2. Z_{t-1} proceso AR(1)

Consideremos ahora que la variable umbral sigue un proceso AR(1), estacionario de tal forma que los pronósticos para Y_t , en (2.6) dependerán de los pronósticos de la variable umbral. Note que el pronóstico un paso adelante es conocido en virtud de que al tiempo T sabemos si estamos en el régimen uno o régimen dos. Sin embargo, desconocemos los valores futuros de Z_t al tiempo $T + 1, \dots, T + h$ y se hace necesario pronosticarlos. Por ejemplo el pronóstico 2-pasos hacia adelante se obtiene como sigue:

$$\begin{aligned}
 \hat{Y}_{T+2} &= E \{Y_{T+2} | \mathfrak{F}_T\} \\
 &= E [\phi_1 I(Z_{T+1} \leq \gamma) + \phi_2 I(Z_{T+1} > \gamma) + \varepsilon_{T+2} | \mathfrak{F}_T] \\
 &= \phi_1 E [I(Z_{T+1} \leq \gamma) | \mathfrak{F}_T] + \phi_2 E [I(Z_{T+1} > \gamma) | \mathfrak{F}_T] \\
 &= \phi_1 \hat{I}(Z_{T+1} \leq \gamma) + \phi_2 \hat{I}(Z_{T+1} > \gamma),
 \end{aligned}$$

es decir, para pronosticar la serie Y_t dos pasos adelante es necesario pronosticar la función indicadora $I(Z_{T+1} \leq \gamma)$ ó $I(Z_{T+1} > \gamma)$ o de forma equivalentemente pronosticar la variable umbral Z_t un paso adelante. La discusión sobre como simular tales variables se describirá en la Sección 4.1.

Para calcular el ECMP dos pasos adelante definamos la función de distribución conjunta de las variables $\hat{I}(Z_{T+1} \leq \gamma)$ y $I(Z_{T+1} \leq \gamma)$, es decir, sea

$$p_{ij} = \Pr \left(\hat{I}(Z_{T+1} \leq \gamma) = i, I(Z_{T+1} \leq \gamma) = j \right)$$

para $i, j = 0, 1$ y $\sum \sum p_{ij} = 1$. Entonces el ECMP dos pasos adelante en el modelo TAR simplificado dado en la ecuación (2.6) es:

$$\begin{aligned}
 ECM P_2^{TAR} &= E \left[Y_{T+2} - \hat{Y}_{T+2} \right]^2 \\
 &= E \left[Y_{T+2} - \hat{Y}_{T+2} \mid \hat{I}(Z_{T+1} \leq \gamma) = 1, I(Z_{T+1} \leq \gamma) = 1 \right]^2 p_{11} \\
 &\quad + E \left[Y_{T+2} - \hat{Y}_{T+2} \mid \hat{I}(Z_{T+1} \leq \gamma) = 1, I(Z_{T+1} \leq \gamma) = 0 \right]^2 p_{10} \\
 &\quad + E \left[Y_{T+2} - \hat{Y}_{T+2} \mid \hat{I}(Z_{T+1} \leq \gamma) = 0, I(Z_{T+1} \leq \gamma) = 1 \right]^2 p_{01} \\
 &\quad + E \left[Y_{T+2} - \hat{Y}_{T+2} \mid \hat{I}(Z_{T+1} \leq \gamma) = 0, I(Z_{T+1} \leq \gamma) = 0 \right]^2 p_{00} \\
 &= E [\phi_1 + \varepsilon_t - \phi_1]^2 p_{11} + E [\phi_2 + \varepsilon_t - \phi_1]^2 p_{10} \\
 &\quad + E [\phi_1 + \varepsilon_t - \phi_2]^2 p_{01} + E [\phi_2 + \varepsilon_t - \phi_2]^2 p_{00} \\
 &= \sigma^2 p_{11} + (\phi_2 - \phi_1)^2 p_{10} + \sigma^2 p_{10} + (\phi_1 - \phi_2)^2 p_{01} + \sigma^2 p_{01} + \sigma^2 p_{00} \\
 &= \sigma^2 (p_{11} + p_{10} + p_{01} + p_{00}) + (\phi_1 - \phi_2)^2 (p_{10} + p_{01}) \\
 &= \sigma^2 + (\phi_1 - \phi_2)^2 (p_{10} + p_{01}),
 \end{aligned}$$

es decir,

$$ECMP_2^{TAR} = \sigma^2 + (\phi_1 - \phi_2)^2(p_{10} + p_{01}),$$

donde p_{10} y p_{01} se entienden como las probabilidades de mala clasificación en cada uno de los regímenes, en el sentido de que el valor pronosticado de la variable umbral al tiempo $T + 1$ excede o no el valor del parámetro γ conocido.

Comparando los respectivos ECMP en el modelo (2.6) y en el modelo (2.11) tenemos que se cumple la siguiente desigualdad, para $h = 1$, $ECMP_1^{TAR} < ECMP_1^L$. Para garantizar que $ECMP_2^{TAR} \leq ECMP_2^L$ dos pasos hacia adelante tenemos que:

$$ECMP_2^{TAR} \leq ECMP_2^L \text{ siempre que } p_{10} + p_{01} < p(\gamma)[1 - p(\gamma)]. \quad (2.12)$$

A partir de la función de densidad conjunta tenemos que las probabilidades p_{10} y p_{01} corresponden a las probabilidades de los pronósticos mal especificados o clasificados, en el entendido de que el pronóstico futuro de la variable indicadora no corresponde al valor verdadero. Por otro lado la probabilidad $p(\gamma)$ corresponde a la proporción de observaciones en el régimen uno y $[1 - p(\gamma)]$ a la proporción de observaciones en el régimen dos.

En la práctica las probabilidades de mala especificación también dependen de la estimación del parámetro de umbral γ , el cual puede estimarse por ejemplo, mediante regresión secuencial, Hansen (1997), o mediante máxima verosimilitud, específicamente mediante la función de log verosimilitud perfil, como detallaremos en la Sección 3.2.1 correspondiente a la obtención de los estimadores mediante máxima verosimilitud. Chan (1998), y Hansen (1997), demuestran que el estimador del parámetro γ , conocido como estimador por mínimos cuadrados es T -consistente. El mismo resultado se tiene para γ estimado a partir de máxima verosimilitud, como demostramos en la Sección 3.2.2. Sin embargo es importante mencionar que desviaciones pequeñas del estimador γ con respecto al verdadero valor, influyen de manera significativa en la asignación de las observaciones a cada uno de los regímenes, y de aquí que las probabilidades de mala especificación pueden ser grandes o pequeñas, Dacco y Sanchell (1999). Por ejemplo, supongamos ahora que conocemos el valor futuro de la variable umbral al tiempo $T + 1$, pero desconocemos el parámetro γ . Supongamos que el verdadero valor del parámetro γ en el modelo (2.6) es $\gamma = \gamma_0$, pero por medio de algún procedimiento de estimación encontramos que $\gamma = \gamma_1$ con $\gamma_1 > \gamma_0$, entonces podemos calcular las probabilidades de mala especificación como:

$$\begin{aligned} p_{10}^* &= \Pr(I(Z_{T+1} \leq \gamma_1) = 1, I(Z_{T+1} \leq \gamma_0) = 0) \\ &= \Pr(Z_{T+1} \leq \gamma_1, Z_{T+1} > \gamma_0) \\ &= \Pr(\gamma_0 < Z_{T+1} \leq \gamma_1) \\ &= \Pr(Z_{T+1} \leq \gamma_1) - \Pr(Z_{T+1} \leq \gamma_0), \end{aligned}$$

y

$$\begin{aligned}
p_{01}^* &= \Pr(I(Z_{T+1} \leq \gamma_1) = 0, I(Z_{T+1} \leq \gamma_0) = 1) \\
&= \Pr(Z_{T+1} > \gamma_1, Z_{T+1} \leq \gamma_0) \\
&= \Pr(\gamma_1 < Z_{T+1} \leq \gamma_0) \\
&= 0.
\end{aligned}$$

El valor de la probabilidad p_{10}^* depende de la función de densidad asociada al proceso Z_t . Por ejemplo, en el caso en que la distribución asociada a la variable umbral Z_{t-1} es la distribución uniforme (i.e. $Z_{t-1} \sim iidU(0, 1)$) tenemos que la probabilidad de mala especificación estará dada en función de que tan alejado está el valor estimado γ_1 del valor verdadero γ . Por su parte si $Z_{t-1} \sim AR(1)$ estacionario con media cero y errores gaussianos, tenemos que la distribución de Z_{T+1} es $N(0, \sigma_u^2 / (1 - \rho^2))$, con ρ el coeficiente autorregresivo en el modelo.

El problema de la mala especificación aumenta, dado que en la práctica debemos pronosticar el comportamiento de la variable umbral, así como estimar los parámetros en el modelo. Sin embargo, los métodos de estimación descritos en la sección de estimación de parámetros garantizan que dichos estimadores poseen algunas propiedades asintóticas, como son la propiedad de consistencia y normalidad.

2.2.2 Especificación incorrecta del verdadero proceso generador de datos

En la sección anterior estudiamos el comportamiento de los errores cuadráticos medios de pronósticos ante una especificación incorrecta originada esencialmente por valores grandes en las probabilidades de mala clasificación en cada uno de los regímenes. El interés en esta sección es estudiar el problema de especificación incorrecta originado al suponer un modelo lineal a nuestro conjunto de datos provenientes de un modelo $TAR(2; 1, 1)$. Mostramos explícitamente, bajo ciertas condiciones en la variable umbral, que al ajustar erróneamente un proceso lineal AR el supuesto distribucional de ruido blanco para los errores no se satisface. Mostramos que dichos errores siguen un modelo de volatilidad, e identificamos al modelo TAR como un modelo de coeficientes aleatorios. El caso más general, como función de los supuestos en la variable umbral también presenta errores cuyos comportamientos se alejan del supuesto de ruido blanco. Ilustramos dichos comportamientos por medio de un análisis empírico. Tenemos entonces que una especificación incorrecta, del tipo lineal autorregresivo, puede generar errores con comportamientos alejados del supuesto distribucional de ruido blanco lo cual debe tomarse en cuenta al momento de generar pronósticos h -pasos hacia adelante, y al realizar la comparación de los desempeños de dichos pronósticos.

Consideremos ahora la siguiente versión estacionaria del modelo TAR dada en la ecuación

(2.4), y que denotamos por $\text{TAR}(2; 1, 1)$, con $d = 1$,

$$Y_t = \delta_t Y_{t-1} + \varepsilon_t, \quad (2.13)$$

con $\delta_t = \phi_1 I(Z_{t-1} \leq \gamma) + \phi_2 I(Z_{t-1} > \gamma)$, Z_{t-1} mutuamente independiente de $\varepsilon_t \sim iidN(0, \sigma^2)$.

Recordemos de las ecuaciones (2.7) y (2.8) que las expresiones para los primeros momentos de δ_t están dadas por las siguientes ecuaciones:

$$\begin{aligned} \mu_\delta &= \phi_1 p + \phi_2 (1 - p), \\ E(\delta_t^2) &= \phi_1^2 p + \phi_2^2 (1 - p), \\ \sigma_\delta^2 &= p(1 - p)(\phi_1 - \phi_2)^2, \end{aligned} \quad (2.14)$$

con $p = p(\gamma) = \Pr(Z_{t-1} \leq \gamma)$. A partir de los momentos del proceso δ_t y del proceso de errores ε_t podemos calcular los momentos del proceso Y_t . Supongamos que $Y_0 = 0$, y que la variable umbral está dada por una secuencia de variables aleatorias independientes e idénticamente distribuidas, entonces el proceso Y_t está caracterizado por los siguientes momentos:

$$\begin{aligned} E(Y_t) &= E(\delta_t Y_{t-1} + \varepsilon_t) \\ &= E(\delta_t Y_{t-1}) + E(\varepsilon_t) \\ &= E(\delta_t \delta_{t-1} Y_{t-2} + \delta_t \varepsilon_{t-1}) \\ &= E(\delta_t \delta_{t-1} Y_{t-2}) + E(\delta_t \varepsilon_{t-1}) \\ &\vdots \\ &= E(\delta_t \delta_{t-1} \dots \delta_1 Y_0) + E(\delta_t \delta_{t-1} \dots \delta_2 \varepsilon_1) \\ &= E(0) + E(\delta_t \delta_{t-1} \dots \delta_2) E(\varepsilon_1) \\ &= 0, \end{aligned}$$

por la independencia entre los procesos $\{Z_{t-1}\}$ y $\{\varepsilon_t\}$ y dado que $E(\varepsilon_t) = 0, \forall t$. Igualmente la varianza del proceso Y_t es:

$$\begin{aligned} \text{Var}(Y_t) &= \text{Var}(\delta_t Y_{t-1} + \varepsilon_t) \\ &= \text{Var}(\delta_t Y_{t-1}) + \text{Var}(\varepsilon_t) + \text{Cov}(\delta_t Y_{t-1}, \varepsilon_t) \\ &= \text{Var}(\delta_t Y_{t-1}) + \sigma^2, \end{aligned}$$

ya que $\text{Cov}(\delta_t Y_{t-1}, \varepsilon_t) = E[\delta_t Y_{t-1} \varepsilon_t - E(\delta_t Y_{t-1}) E(\varepsilon_t)] = E(\delta_t Y_{t-1} \varepsilon_t) = E(\delta_t Y_{t-1}) E(\varepsilon_t) = 0$ por la independencia entre $\{Z_{t-1}\}$ y $\{\varepsilon_t\}$ y del hecho de que ε_t es independiente de

2.2. ERROR CUADRÁTICO MEDIO BAJO ESPECIFICACIONES INCORRECTAS 31

Y_{t-1}, Y_{t-2}, \dots , retomando la ecuación anterior tenemos que la varianza de Y_t es:

$$\begin{aligned}
 \text{Var}(Y_t) &= \text{Var}(\delta_t Y_{t-1}) + \sigma^2 \\
 &= \text{Var}(\delta_t \delta_{t-1} Y_{t-2} + \delta_t \varepsilon_{t-1}) + \sigma^2 \\
 &= \text{Var}(\delta_t \delta_{t-1} Y_{t-2}) + \text{Var}(\delta_t \varepsilon_{t-1}) + \sigma^2 \\
 &= \text{Var}(\delta_t \delta_{t-1} \delta_{t-2} Y_{t-3}) + \text{Var}(\delta_t \delta_{t-1} \varepsilon_{t-2}) + \text{Var}(\delta_t \varepsilon_{t-1}) + \sigma^2 \\
 &\quad \vdots \\
 &= \text{Var}(\delta_t \delta_{t-1} \delta_{t-2} \cdots \delta_1 Y_0) + \text{Var}(\delta_t \delta_{t-1} \delta_{t-2} \cdots \delta_2 \varepsilon_1) + \cdots + \text{Var}(\delta_t \varepsilon_{t-1}) + \sigma^2 \\
 &= \text{Var}(\delta_t \delta_{t-1} \delta_{t-2} \cdots \delta_2 \varepsilon_1) + \cdots + \text{Var}(\delta_t \varepsilon_{t-1}) + \sigma^2 \\
 &= E(\delta_t \delta_{t-1} \delta_{t-2} \cdots \delta_2 \varepsilon_1)^2 + \cdots + E(\delta_t \delta_{t-1} \varepsilon_{t-2})^2 + E(\delta_t \varepsilon_{t-1})^2 + \sigma^2 \\
 &= E(\delta_t \delta_{t-1} \delta_{t-2} \cdots \delta_2)^2 E(\varepsilon_1)^2 + \cdots \\
 &\quad + E(\delta_t \delta_{t-1})^2 E(\varepsilon_{t-2})^2 + E(\delta_t)^2 E(\varepsilon_{t-1})^2 + \sigma^2 \\
 &= \sigma^2 \left\{ 1 + E(\delta_t^2) + [E(\delta_t^2)]^2 + \cdots + [E(\delta_t^2)]^{t-1} \right\} \\
 &= \sigma^2 \sum_{i=0}^{t-1} [E(\delta_t^2)]^i = \sigma^2 \frac{1 - [E(\delta_t^2)]^t}{1 - E(\delta_t^2)},
 \end{aligned}$$

por ser Z_{t-1} un proceso independiente e idénticamente distribuido, y por la condición de estacionariedad de segundo orden del proceso Y_t , que establece que $E(\delta_t^2) < 1$.

Igualmente tenemos que la covarianza de Y_t y Y_{t-s} , está dada por:

$$\begin{aligned}
 \text{Cov}(Y_t, Y_{t-s}) &= E[(\delta_t Y_{t-1} + \varepsilon_t) Y_{t-s}] \\
 &= E(\delta_t Y_{t-1} Y_{t-s}) + E(\varepsilon_t Y_{t-s}) \\
 &= E(\delta_t Y_{t-1} Y_{t-s}) \\
 &= E(\delta_t \delta_{t-1} \delta_{t-2} \cdots \delta_{t-s+1} Y_{t-s} Y_{t-s}) \\
 &= E(\delta_t \delta_{t-1} \delta_{t-2} \cdots \delta_{t-s+1}) E(Y_{t-s} Y_{t-s}) \\
 &= \mu_\delta^s \text{Var}(Y_{t-s}) \\
 &= \mu_\delta^s \sigma^2 \left(\frac{1 - [E(\delta_t^2)]^{t-s}}{1 - E(\delta_t^2)} \right)
 \end{aligned}$$

Dado que $E(\delta_t^2) < 1$, por la condición de estacionariedad de segundo orden y haciendo tender $t \rightarrow \infty$, resumimos los primeros momentos del proceso Y_t por medio de las siguientes

ecuaciones:

$$\begin{aligned} E(Y_t) &= 0 \\ \text{Var}(Y_t) &= \sigma^2 \frac{1}{1 - E(\delta_t^2)} \\ \text{Cov}(Y_t, Y_{t-s}) &= \sigma^2 \frac{\mu_\delta^s}{1 - E(\delta_t^2)} \end{aligned} \quad (2.15)$$

Recordemos que una de las razones de interés para estudiar los modelos TAR, desde el punto de vista económico por ejemplo, es que dichos modelos pueden describir fenómenos de raíces unitarias, como las tasas de interés o de desempleo, pero con la característica de que el modelo TAR cumple con la condición de ser un proceso globalmente estacionario. Comparemos por ejemplo la varianza del proceso TAR(2; 1, 1), dada en la ecuación (2.15) con la varianza convencional de un modelo AR(1) con coeficiente autorregresivo dado por μ_δ . Denotemos el modelo AR(1) por \tilde{Y}_t , con varianza dada por: $\text{Var}(\tilde{Y}_t) = \frac{\sigma^2}{1 - \mu_\delta^2}$. Podemos descomponer la varianza de Y_t , dada en (2.15), como la suma de la varianza de \tilde{Y}_t y un término adicional que depende exclusivamente de la varianza de δ_t , como sigue:

$$\begin{aligned} \text{Var}(Y_t) &= \frac{\sigma^2}{1 - E(\delta_t^2)} - \frac{\sigma^2}{1 - \mu_\delta^2} + \frac{\sigma^2}{1 - \mu_\delta^2} \\ &= \frac{\sigma^2}{1 - \mu_\delta^2} + \sigma^2 \left[\frac{1}{1 - E(\delta_t^2)} - \frac{1}{1 - \mu_\delta^2} \right] \\ &= \text{Var}(\tilde{Y}_t) + \sigma^2 \left\{ \frac{1 - \mu_\delta^2 - 1 + E(\delta_t^2)}{[1 - E(\delta_t^2)][1 - \mu_\delta^2]} \right\} \\ &= \text{Var}(\tilde{Y}_t) + \frac{\sigma^2}{[1 - E(\delta_t^2)][1 - \mu_\delta^2]} \sigma_\delta^2. \end{aligned}$$

Continuemos con nuestra idea de explorar especificaciones incorrectas ante el hecho de que nuestro VPGD es el modelo TAR(2; 1, 1) pero por alguna razón creemos que los datos provienen de un modelo AR(1). Notemos que el proceso δ_t es igual en distribución al proceso $\mu_\delta + \sigma_\delta V_t$, es decir,

$$\delta_t \stackrel{d}{=} \mu_\delta + \sigma_\delta V_t,$$

con

$$V_t = \left(\frac{\phi_1 - \mu_\delta}{\sigma_\delta} \right) I(Z_{t-1} \leq \gamma) + \left(\frac{\phi_2 - \mu_\delta}{\sigma_\delta} \right) I(Z_{t-1} > \gamma),$$

un proceso estacionario con media 0 y varianza 1, y por consiguiente podemos rescribir el modelo TAR(2; 1, 1) como un modelo linealizado del tipo AR(1) con coeficiente autorregresivo

μ_δ , tal y como se escribe a continuación,

$$Y_t = \mu_\delta Y_{t-1} + \sigma_\delta Y_{t-1} V_t + \varepsilon_t = \mu_\delta Y_{t-1} + r_t \quad (2.16)$$

y $r_t = \sigma_\delta Y_{t-1} V_t + \varepsilon_t$ el nuevo proceso de error. En situaciones prácticas, y dado un conjunto de observaciones provenientes de un modelo TAR(2; 1, 1), comúnmente se asume un modelo del tipo lineal como el dado en la ecuación (2.16), pero considerando que los errores en el modelo lineal cumplen con el supuesto de ser ruido blanco gaussiano. Bajo tal consideración se obtienen las estimaciones de los parámetros involucrados en el modelo lineal, así como la obtención de los pronósticos h -pasos hacia adelante.

A partir de la expresión (2.16) resaltan dos características interesantes para conjeturar cómo será el desempeño de los pronósticos ante una incorrecta especificación en el modelo. Por ejemplo, supongamos que el VPGD es efectivamente el modelo TAR(2; 1, 1) pero por error consideramos que los datos provienen de un modelo lineal autorregresivo, dado por:

$$Y_t^\dagger = \phi Y_{t-1}^\dagger + \varepsilon_t^\dagger,$$

donde la secuencia de errores $\{\varepsilon_t^\dagger\}$ cumple con el supuesto distribucional de ruido blanco gaussiano. Sabemos que $\hat{\phi} \xrightarrow{p} \phi$, con $\hat{\phi}$ estimador de mínimos cuadrados o máxima verosimilitud, entonces bajo el VPGD $\hat{\phi} \xrightarrow{p} \mu_\delta = \phi_1 p(\gamma) + \phi_2 [1 - p(\gamma)]$. Análogamente, retomemos el modelo (2.16), si podemos asumir que el proceso de ruido r_t cumple el supuesto de ruido blanco gaussiano, esperamos que los desempeños de los pronósticos en el modelo lineal, y el modelo TAR estén directamente relacionados con la cercanía entre la estimación del parámetro autorregresivo y el valor μ_δ . Por otra parte, si no se cumple el supuesto de homoscedasticidad en el proceso r_t esperaríamos que el desempeño de los pronósticos en el modelo TAR(2; 1, 1) sea mejor que el modelo lineal AR(1), como función de sus respectivos ECM de pronósticos.

A fin de complementar lo antes mencionado, verifiquemos que el proceso V_t tiene media cero y varianza uno,

$$\begin{aligned}
E(V_t) &= E \left[\left(\frac{\phi_1 - \mu_\delta}{\sigma_\delta} \right) I(Z_{t-1} \leq \gamma) + \left(\frac{\phi_2 - \mu_\delta}{\sigma_\delta} \right) I(Z_{t-1} > \gamma) \right] \\
&= \left(\frac{\phi_1 - \mu_\delta}{\sigma_\delta} \right) E[I(Z_{t-1} \leq \gamma)] + \left(\frac{\phi_2 - \mu_\delta}{\sigma_\delta} \right) E[I(Z_{t-1} > \gamma)] \\
&= \left(\frac{\phi_1 - \mu_\delta}{\sigma_\delta} \right) \Pr(Z_{t-1} \leq \gamma) + \left(\frac{\phi_2 - \mu_\delta}{\sigma_\delta} \right) \Pr(Z_{t-1} > \gamma) \\
&= \left(\frac{\phi_1 - \mu_\delta}{\sigma_\delta} \right) p(\gamma) + \left(\frac{\phi_2 - \mu_\delta}{\sigma_\delta} \right) [1 - p(\gamma)] \\
&= \frac{1}{\sigma_\delta} [\phi_1 p(\gamma) + \phi_2 (1 - p(\gamma)) - \mu_\delta p(\gamma) - \mu_\delta + \mu_\delta p(\gamma)] \\
&= \frac{1}{\sigma_\delta} [\mu_\delta - \mu_\delta] \\
&= 0.
\end{aligned}$$

Asimismo,

$$\begin{aligned}
\text{Var}(V_t) &= E(V_t)^2 \\
&= E \left[\left(\frac{\phi_1 - \mu_\delta}{\sigma_\delta} \right) I(Z_{t-1} \leq \gamma) + \left(\frac{\phi_2 - \mu_\delta}{\sigma_\delta} \right) I(Z_{t-1} > \gamma) \right]^2 \\
&= E \left[\left(\frac{\phi_1 - \mu_\delta}{\sigma_\delta} \right)^2 I(Z_{t-1} \leq \gamma) + \left(\frac{\phi_2 - \mu_\delta}{\sigma_\delta} \right)^2 I(Z_{t-1} > \gamma) \right] \\
&= \left(\frac{\phi_1 - \mu_\delta}{\sigma_\delta} \right)^2 E[I(Z_{t-1} \leq \gamma)] + \left(\frac{\phi_2 - \mu_\delta}{\sigma_\delta} \right)^2 E[I(Z_{t-1} > \gamma)] \\
&= \left(\frac{\phi_1 - \mu_\delta}{\sigma_\delta} \right)^2 p(\gamma) + \left(\frac{\phi_2 - \mu_\delta}{\sigma_\delta} \right)^2 [1 - p(\gamma)] \\
&= \frac{1}{\sigma_\delta^2} [\phi_1^2 p(\gamma) + \phi_2^2 (1 - p(\gamma)) - 2p(\gamma) \phi_1 \mu_\delta + \\
&\quad + p(\gamma) \mu_\delta^2 - 2\phi_2 \mu_\delta (1 - p(\gamma)) + (1 - p(\gamma)) \mu_\delta^2] \\
&= 1.
\end{aligned}$$

El modelo linealizado (2.16) para Y_t corresponde a la expresión matemática de un modelo autorregresivo AR con coeficiente autorregresivo constante μ_δ , e innovaciones “especiales” dadas por el proceso r_t .

Para caracterizar el comportamiento del proceso $r_t = \sigma_\delta v_t Y_{t-1} + \varepsilon_t$ calculemos sus primeros momentos,

$$\begin{aligned}
 E(r_t) &= E[\sigma_\delta V_t Y_{t-1} + \varepsilon_t] \\
 &= E[\sigma_\delta V_t Y_{t-1}] \\
 &= E[E(\sigma_\delta V_t Y_{t-1}) | Y_{t-1}] \\
 &= E[\sigma_\delta Y_{t-1} E(V_t | Y_{t-1})] \\
 &= 0,
 \end{aligned}$$

ya que $E(V_t | Y_{t-1}) = E(V_t) = 0$. El cálculo de la varianza es como sigue:

$$\begin{aligned}
 \text{Var}(r_t) &= \text{Var}(\sigma_\delta V_t Y_{t-1} + \varepsilon_t) = \sigma_\delta^2 \text{Var}(V_t Y_{t-1}) + \sigma^2 \\
 &= \sigma_\delta^2 \text{Var}(V_t) \text{Var}(Y_{t-1}) + \sigma^2 = \sigma_\delta^2 \text{Var}(Y_{t-1}) + \sigma^2 \\
 &= \sigma_\delta^2 \frac{\sigma^2}{1 - E(\delta_t^2)} + \sigma^2 = \sigma^2 \left[\frac{\sigma_\delta^2}{1 - E(\delta_t^2)} + 1 \right] \\
 &= \sigma^2 \left(\frac{1 - [E(\delta_t)]^2}{1 - E(\delta_t^2)} \right),
 \end{aligned}$$

ya que $V_t = \left(\frac{\phi_1 - \mu_\delta}{\sigma_\delta} \right) I(Z_{t-1} \leq \gamma) + \left(\frac{\phi_2 - \mu_\delta}{\sigma_\delta} \right) I(Z_{t-1} > \gamma)$ independiente de Y_{t-1} . Asimismo,

$$\begin{aligned}
 E(r_t r_{t+h}) &= E[(\sigma_\delta V_t Y_{t-1} + \varepsilon_t)(\sigma_\delta V_{t+h} Y_{t+h-1} + \varepsilon_{t+h})] \\
 &= E(\sigma_\delta^2 V_t Y_{t-1} V_{t+h} Y_{t+h-1} + \sigma_\delta V_{t+h} Y_{t+h-1} \varepsilon_t + \sigma_\delta V_t Y_{t-1} \varepsilon_{t+h} + \varepsilon_t \varepsilon_{t+h}) \\
 &= \sigma_\delta^2 E(V_t Y_{t-1} V_{t+h} Y_{t+h-1}) + \sigma_\delta E(V_{t+h} Y_{t+h-1} \varepsilon_t) + \sigma_\delta E(V_t Y_{t-1} \varepsilon_{t+h}) + E(\varepsilon_t \varepsilon_{t+h}) \\
 &= 0,
 \end{aligned}$$

ya que,

$$\begin{aligned}
 E[V_t Y_{t-1} V_{t+h} Y_{t+h-1}] &= E(V_t) E(Y_{t-1} V_{t+h} Y_{t+h-1}) \\
 &= 0, \text{ para } h > 0.
 \end{aligned}$$

Igualmente tenemos que:

$$\begin{aligned}
 E[V_{t+h} Y_{t+h-1} \varepsilon_t] &= E(V_{t+h}) E(Y_{t+h-1} \varepsilon_t) \\
 &= 0, \text{ para } h > 0,
 \end{aligned}$$

$$\begin{aligned}
 E[V_t Y_{t-1} \varepsilon_{t+h}] &= E(V_t) E(Y_{t-1} \varepsilon_{t+h}) \\
 &= 0,
 \end{aligned}$$

y

$$E[\varepsilon_t \varepsilon_{t+h}] = 0.$$

Resumiendo, escribimos los momentos del proceso r_t ,

$$\begin{aligned} E(r_t) &= 0, \\ E(r_t r_{t+h}) &= 0, \quad t > 0, h > 0, \\ \text{Var}(r_t) &= \sigma^2 \left(\frac{1 - \mu_\delta^2}{1 - E(\delta_t^2)} \right) \\ &= \sigma^2 \left(\frac{1 - E(\delta_t^2) + E(\delta_t^2) - \mu_\delta^2}{1 - E(\delta_t^2)} \right) \\ &= \sigma^2 \left(1 + \frac{\sigma_\delta^2}{1 - E(\delta_t^2)} \right). \end{aligned} \tag{2.17}$$

Calculemos ahora los errores cuadráticos medios de pronósticos bajo las siguientes condiciones:

Situación 1: Supongamos que el VPGD es un modelo TAR(2; 1, 1) tal y como se especificó en la ecuación (2.13) y que el pronóstico h -pasos adelante es también generado a partir de dicho modelo, es decir,

$$\begin{aligned} \hat{Y}_{t+h}^{TAR} &= E(Y_{t+h} \mid \mathcal{F}_t) = E(\delta_{t+h} Y_{t+h-1} + \varepsilon_{t+h} \mid \mathcal{F}_t) \\ &= E(\delta_{t+h} Y_{t+h-1} \mid \mathcal{F}_t) + E(\varepsilon_{t+h} \mid \mathcal{F}_t) \\ &= E(\delta_{t+h} Y_{t+h-1} \mid \mathcal{F}_t) \\ &= E(\delta_{t+h} \delta_{t+h-1} Y_{t+h-2} + \delta_{t+h} \varepsilon_{t+h-1} \mid \mathcal{F}_t) \\ &= E(\delta_{t+h} \delta_{t+h-1} Y_{t+h-2} \mid \mathcal{F}_t) \\ &= \dots \\ &= E(\delta_{t+h} \delta_{t+h-1} \dots \delta_{t+1} Y_t \mid \mathcal{F}_t) \\ &= \mu_\delta^h Y_t, \end{aligned}$$

2.2. ERROR CUADRÁTICO MEDIO BAJO ESPECIFICACIONES INCORRECTAS 37

entonces el ECMP, denotado como $ECMP_h^{TAR,TAR}$, se calcula como sigue:

$$\begin{aligned}
 ECM P_h^{TAR,TAR} &= E [Y_{t+h} - E(Y_{t+h} | \mathcal{F}_t)]^2 = E [Y_{t+h} - \hat{Y}_{t+h}^{TAR}]^2 \\
 &= E (\delta_{t+h} Y_{t+h-1} + \varepsilon_{t+h} - \mu_\delta^h Y_t)^2 \\
 &= E (\delta_{t+h}^2 Y_{t+h-1}^2 + \varepsilon_{t+h}^2 + \mu_\delta^{2h} Y_t^2 \\
 &\quad + 2\delta_{t+h} Y_{t+h-1} \varepsilon_{t+h} - 2\mu_\delta^h \delta_{t+h} Y_t Y_{t+h-1} \\
 &\quad - 2\mu_\delta^h \varepsilon_{t+h} Y_t) \\
 &= E (\delta_{t+h}^2 Y_{t+h-1}^2) + E (\varepsilon_{t+h}^2) + E (\mu_\delta^{2h} Y_t^2) \\
 &\quad + 2E (\delta_{t+h} Y_{t+h-1} \varepsilon_{t+h}) - 2\mu_\delta^h E (\delta_{t+h} Y_t Y_{t+h-1}) \\
 &\quad - 2\mu_\delta^h E (\varepsilon_{t+h} Y_t) \\
 &= E (\delta_{t+h}^2) \text{Var} (Y_{t+h-1}) + \sigma^2 + \mu_\delta^{2h} \text{Var} (Y_t) \\
 &\quad - 2\mu_\delta^h E [(\delta_{t+h}) \text{Cov} (Y_t, Y_{t+h-1})] \\
 &= \mu_\delta^2 \frac{\sigma^2}{1 - E (\delta_t^2)} + \sigma^2 + \mu_\delta^{2h} \frac{\sigma^2}{1 - E (\delta_t^2)} \\
 &\quad - 2\mu_\delta^h \mu_\delta \left(\sigma^2 \frac{\mu_\delta^{h-1}}{1 - E (\delta_t^2)} \right) \\
 &= \sigma^2 \left(1 + \frac{\mu_\delta^2}{1 - E (\delta_t^2)} + \frac{\mu_\delta^{2h}}{1 - E (\delta_t^2)} - 2 \frac{\mu_\delta^{2h}}{1 - E (\delta_t^2)} \right) \\
 &= \sigma^2 \left(\frac{1 - E (\delta_t^2) + \mu_\delta^2 - \mu_\delta^{2h}}{1 - E (\delta_t^2)} \right) \\
 &= \sigma^2 \left(\frac{1 - \sigma_\delta^2 - \mu_\delta^{2h}}{1 - E (\delta_t^2)} \right) \\
 &= \sigma^2 \left(\frac{1 - \mu_\delta^{2h}}{1 - E (\delta_t^2)} - \frac{\sigma_\delta^2}{1 - E (\delta_t^2)} \right). \tag{2.18}
 \end{aligned}$$

Situación 2: Consideremos ahora que el VPGD es el modelo TAR(2; 1, 1) linealizado dado en la ecuación (2.16) y que generamos el pronóstico h -pasos adelante mediante el proceso AR(1) con coeficiente autorregresivo μ_δ , entonces el ECMP, denotado por $ECMP_h^{TAR,AR}$, está dado por:

$$\begin{aligned}
ECMP_h^{TAR,AR} &= E [Y_{t+h} - E(Y_{t+h} | \mathcal{F}_t)]^2 \\
&= E [\mu_\delta Y_{t+h-1} + r_{t+h} - \mu_\delta Y_t]^2 \\
&\vdots \\
&= E (\mu_\delta^h Y_t + \mu_\delta^{h-1} r_{t+h-(h-1)} + \cdots + r_{t+h} - \mu_\delta^h Y_t)^2 \\
&= E (\mu_\delta^{h-1} r_{t+h-(h-1)} + \cdots + \mu_\delta r_{t+h-1} + r_{t+h})^2 \\
&= E \left(\sum_{j=1}^h \mu_\delta^{2(h-j)} r_{t+j}^2 \right) + \cdots + 2\mu_\delta^h E(r_{t+j} r_{t+h-1}) + 2\mu_\delta^{h-1} E(r_{t+1} r_{t+h}) \\
&= \sum_{j=1}^h \mu_\delta^{2(h-j)} E(r_{t+j}^2) = \sum_{j=1}^h \mu_\delta^{2(h-j)} \text{Var}(r_{t+j}) \\
&= \sigma^2 \left(\frac{1 - \mu_\delta^2}{1 - E(\delta_t^2)} \right) \sum_{j=1}^h \mu_\delta^{2(h-j)} \\
&= \sigma^2 \left(\frac{1 - \mu_\delta^2}{1 - E(\delta_t^2)} \right) \left(\frac{1 - \mu_\delta^{2h}}{1 - \mu_\delta^2} \right) \\
&= \sigma^2 \left(\frac{1 - \mu_\delta^{2h}}{1 - E(\delta_t^2)} \right). \tag{2.19}
\end{aligned}$$

De las ecuaciones (2.18) y (2.19) podemos observar que $ECMP_h^{TAR,TAR} < ECMP_h^{TAR,AR}$ dado que el cociente $\sigma_\delta^2 / (1 - E(\delta_t^2)) > 0$ por la hipótesis de estacionariedad de segundo orden. Notemos también que los desempeños de los pronósticos h -pasos adelante ante una especificación incorrecta dependerá del comportamiento del proceso de ruido r_t y en particular de que tan alejado estemos del supuesto de estacionariedad de segundo orden. Mostramos que el proceso de error r_t es un proceso de volatilidad, es decir, la varianza del proceso va cambiando en el tiempo. Esta observación es directa para el caso en que $Z_{t-1} \sim iidU(0, 1)$ ya que el modelo TAR(2; 1, 1) dado en la ecuación (2.13) puede pensarse como un modelo autorregresivo de coeficientes aleatorios (RCA, por sus siglas en inglés), Nicholls y Quinn (1982) y Koul y Schick (1996). Sin embargo está característica presentada en los errores también ha sido observada de manera empírica bajo situaciones más generales.

2.2. ERROR CUADRÁTICO MEDIO BAJO ESPECIFICACIONES INCORRECTAS 39

Consideremos el siguiente ejemplo: sea $Z_t \sim iidU(0, 1)$ y Y_t dado por el siguiente modelo TAR(2; 1, 1) :

$$Y_t = \begin{cases} -0.8747Y_{t-1} + \varepsilon_t & \text{si } Z_{t-1} \leq \gamma \\ 1.0747Y_{t-1} + \varepsilon_t & \text{si } Z_{t-1} > \gamma, \end{cases} \quad (2.20)$$

donde el valor del parámetro γ es tal que $\Pr(Z_{t-1} \leq \gamma) = 0.5$, es decir, $\gamma = 0.5$ y

$\varepsilon_t \sim iidN(0, \sigma^2)$, con $\sigma^2 = 1$. Note que el proceso dado en la ecuación (2.20) es estacionario en covarianzas por el Corolario 2.7. del Capítulo 2, ya que los primeros momentos de δ_t son:

$$\begin{aligned} E(\delta_t) &= (\phi_1)(p) + (\phi_2)(1-p) \\ &= (-0.8747)(0.5) + (1.0747)(1-0.5) \\ &= 0.1 \\ E(\delta_t^2) &= (\phi_1)^2(p) + (\phi_2)^2(1-p) \\ &= (-0.8747)^2(0.5) + (1.0747)^2(1-0.5) \\ &= 0.96. \end{aligned}$$

La consideración de este escenario cercano a la condición de no estacionariedad es con la finalidad de acrecentar las características de volatilidad presentadas en el proceso r_t , al usar erróneamente al modelo lineal AR, como el proceso generador de datos. El comportamiento esperado de los residuales bajo dicho modelo será muy distinto al comportamiento de ruido blanco de los residuales bajo el modelo TAR correctamente especificado.

La Figura 2 presenta el comportamiento de una realización del proceso Y_t , dado en la ecuación (2.20), para un tamaño de muestra de 250 observaciones. Note que la serie presenta periodos de volatilidad baja seguidos de periodos de volatilidad alta, es decir, la volatilidad del proceso va cambiando a través del tiempo. Note de la ecuación (2.17) que la variabilidad del proceso r_t depende de los primeros momentos del proceso δ_t . En particular si la condición de estacionariedad de segundo orden ($E(\delta_t^2) \approx 1$) es cercana a uno como, en este experimento, la volatilidad del proceso r_t es notoria.

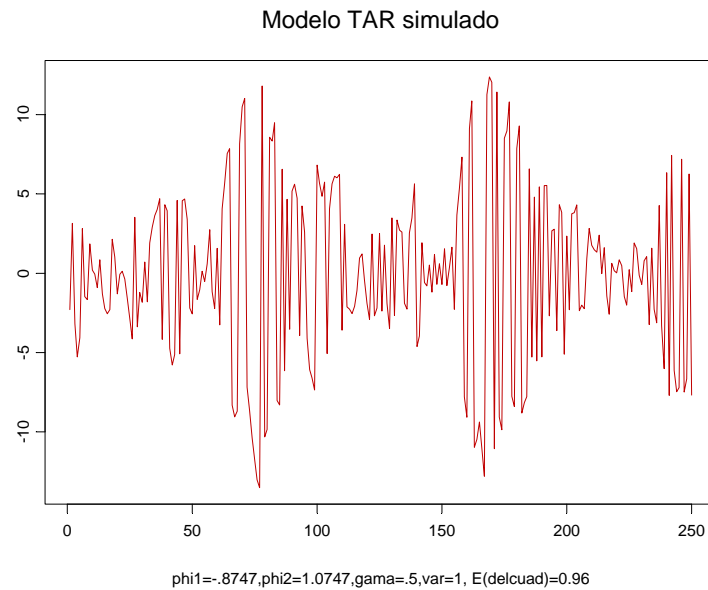


Figura 2. Realización del modelo TAR(2; 1, 1) con $Z_{t-1} \sim iidU(0, 1)$.

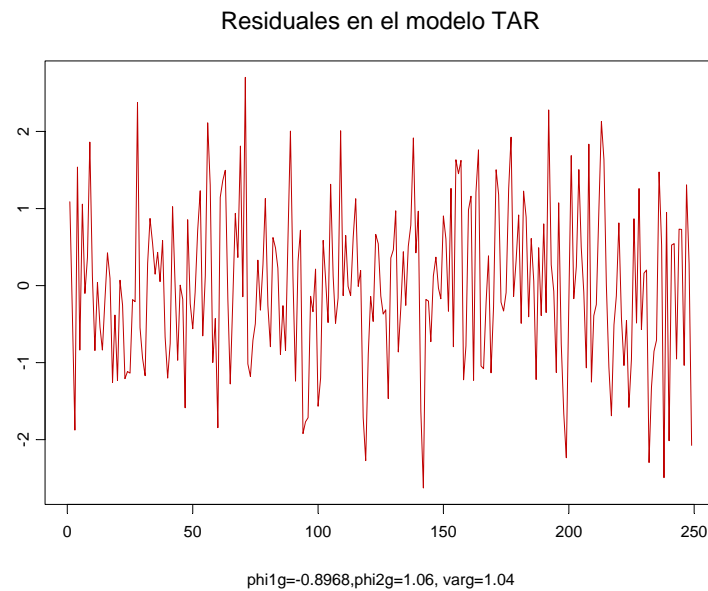


Figura 3. Residuales bajo el modelo TAR(2; 1, 1), con $Z_{t-1} \sim iidU(0, 1)$.

La Figura 3 describe el comportamiento de los residuales obtenidos al ajustar al conjunto de los 250 datos simulados un modelo TAR(2; 1, 1). Los valores estimados de los parámetros en el modelo, se encuentran en la parte inferior de la gráfica y corresponden a las estimaciones

de ϕ_1 , ϕ_2 y σ^2 respectivamente. Observemos que independientemente del comportamiento del proceso de observaciones, los residuales asociados al modelo TAR presentan un comportamiento homogéneo a través del tiempo, el cual es típico de un proceso de ruido blanco gaussiano.

Consideremos el conjunto de datos simulados a partir del modelo (2.20), posteriormente ajustemos a dicho conjunto de datos un modelo lineal AR(1) y AR(2) respectivamente.

Inherente al procedimiento de ajuste de los modelos lineales autorregresivos esta el supuesto de que los errores asociados a dichos modelos son procesos gaussianos. De las gráficas de diagnóstico dadas en las Figuras 4 y 5 observamos que los residuales correspondientes presentan comportamientos de varianza no constante pero dentro de los límites de significancia estadística. Más aún, los valores de las funciones de autocorrelación total y parcial muestran un comportamiento dentro de los límites significativos; al igual que p-valores de la prueba de Ljung-Box, para verificar independencia en los residuales. Podríamos pensar, de los resultados gráficos que los residuales asociados al modelo lineal aparentemente cumplen con el supuesto homoscedasticidad y normalidad en los errores. Sin embargo, al realizar las correspondientes gráficas de cuantiles, observamos que los datos tienden a distribuirse alrededor de la recta de 45 grados, pero con la característica de que el rango de valores de los residuales asociados al modelo lineal está entre ± 10 tal y como se observa en la Figura 6.

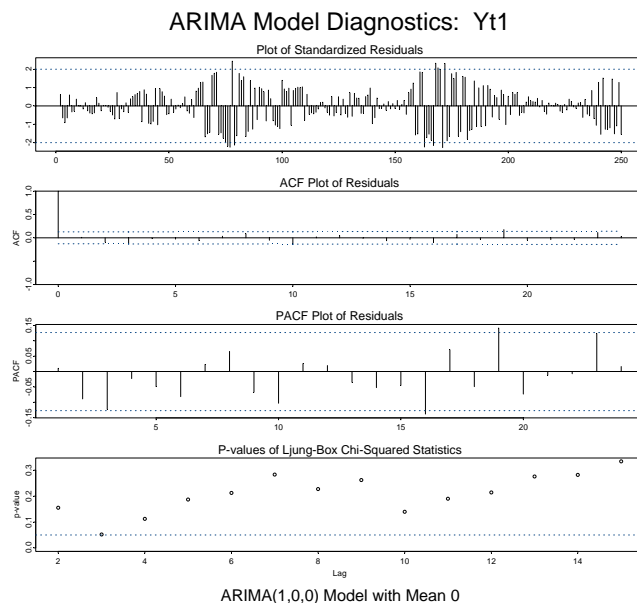


Figura 4. Gráficas de diagnóstico para el modelo AR(1).

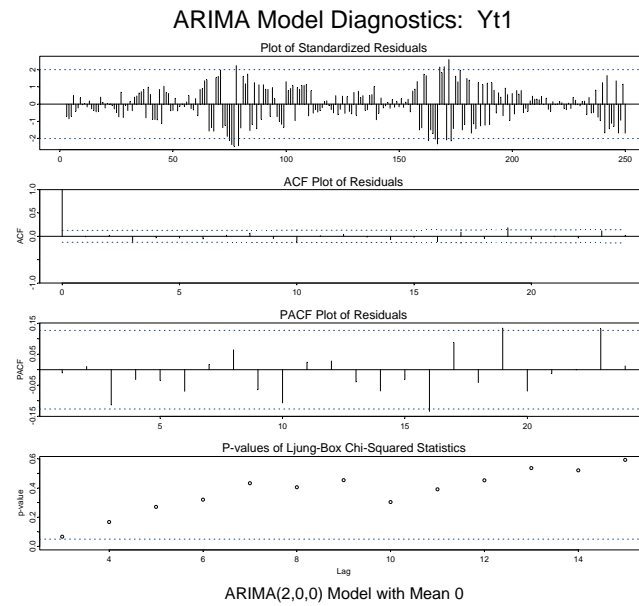


Figura 5. Gráficas de diagnóstico para el modelo AR(2).

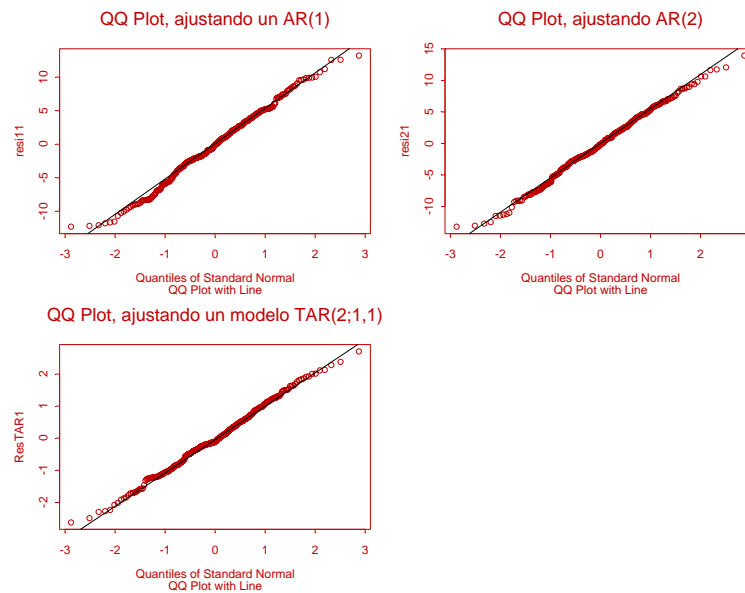


Figura 6. Gráficas de papel normal para los residuales bajo los modelos AR(1), AR(2) y TAR(2; 1, 1) respectivamente.

Además de que los rangos de valores observados en los residuales muestrales bajo los modelos AR(1) y AR(2), son grandes, se observa también que los residuales en los extremos de las gráficas de cuantiles tienden a separarse de la recta de 45 grados, lo cual se presenta

2.2. ERROR CUADRÁTICO MEDIO BAJO ESPECIFICACIONES INCORRECTAS⁴³

en distribuciones de colas más pesadas que la distribución normal. Una recomendación en la literatura de modelos no lineales para identificar estructuras de dependencias de orden mayor es analizar los comportamientos de los residuales al cuadrado. Las Figuras 7, 8 y 9 presentan los comportamientos de los valores para las funciones de autocorrelación de los residuales al cuadrado bajo los modelos $AR(1)$, $AR(2)$ y $TAR(2; 1, 1)$ respectivamente. Observemos que para los modelos $AR(1)$ y $AR(2)$ los valores de las funciones de autocorrelación para los residuales al cuadrado decaen muy lentamente, identificando que dichos residuales presentan una estructura de dependencia. Por su parte las autocorrelaciones para los residuales al cuadrado en el modelo $TAR(2; 1, 1)$ son significativamente iguales a cero ya que éstas caen dentro de los límites de significancia tal y como se observa en la Figura 9.

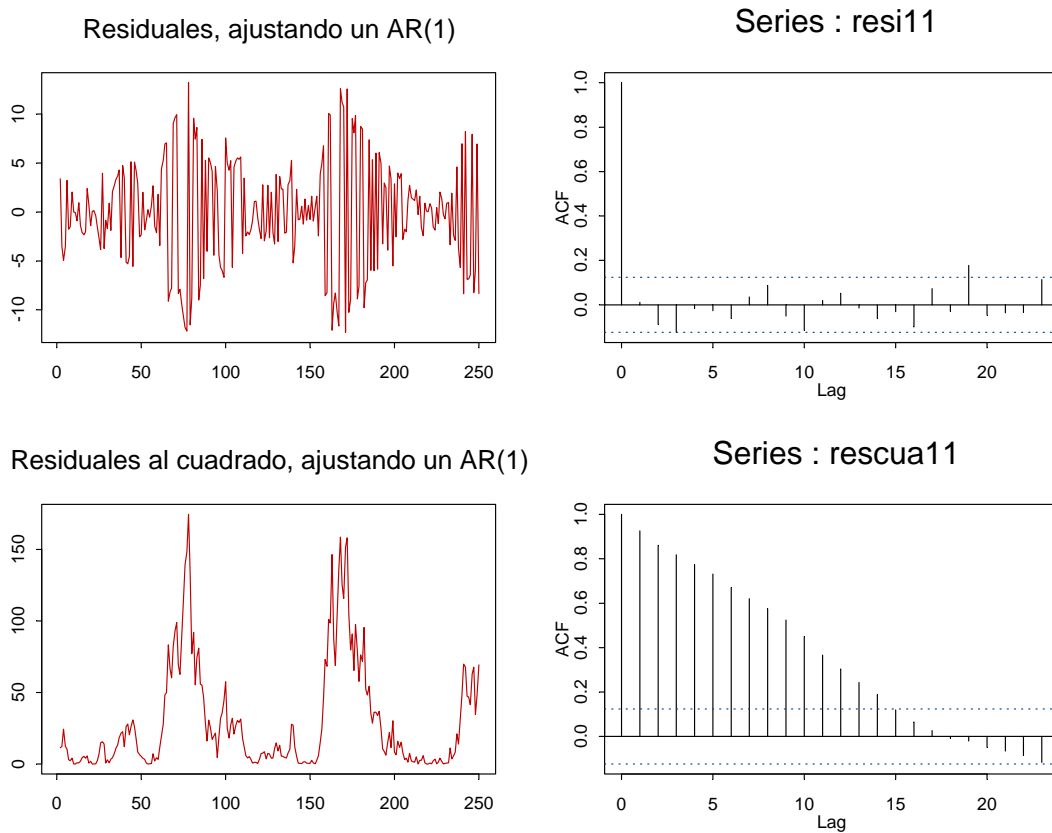


Figura 7. Autocorrelaciones para los residuales y residuales al cuadrado, modelo $AR(1)$.

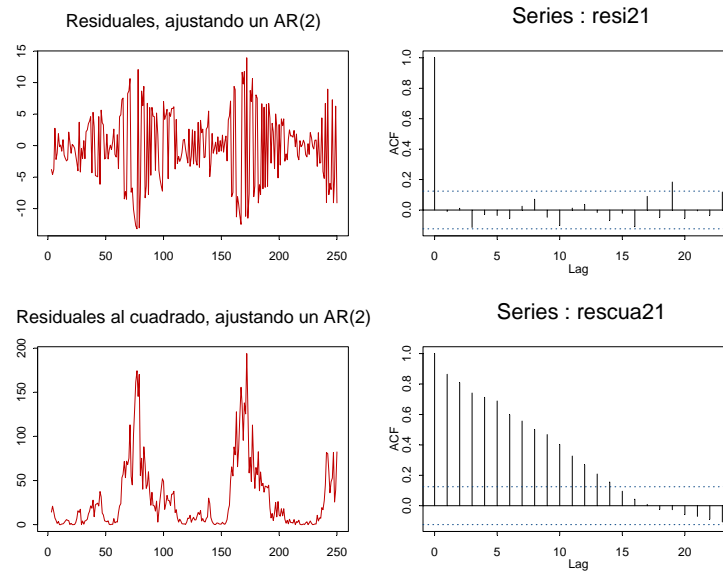


Figura 8. Autocorrelaciones para los residuales y residuales al cuadrado, modelo AR(2).

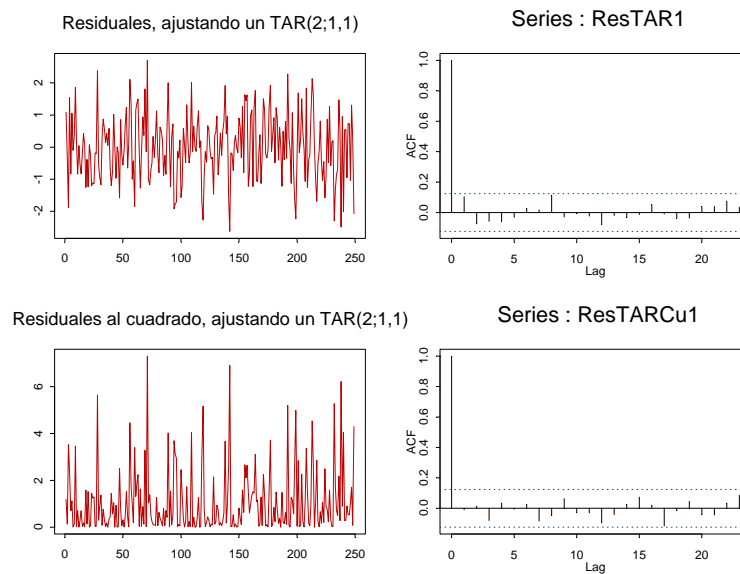


Figura 9. Autocorrelaciones para los residuales y residuales al cuadrado, modelo TAR(2; 1, 1).

2.2. ERROR CUADRÁTICO MEDIO BAJO ESPECIFICACIONES INCORRECTAS 45

Las conclusiones del ejemplo anterior, en relación a que el proceso de ruido asociado al modelo erróneamente especificado es un modelo de volatilidad las mostramos explícitamente; sin embargo cuando la variable umbral Z_{t-1} sigue un proceso AR(1) estacionario con errores gaussianos, también se observa que los residuales presentan comportamientos alejados del supuesto de ruido blanco. Por ejemplo, consideremos el siguiente escenario simulado: sea $Z_t = 0.5Z_{t-1} + u_t$ con $u_t \sim iidN(0, 1)$, y Y_t dado por el siguiente modelo TAR(2; 1, 1) :

$$Y_t = \begin{cases} -0.8Y_{t-1} + \varepsilon_t & \text{si } Z_{t-1} \leq \gamma \\ Y_{t-1} + \varepsilon_t & \text{si } Z_{t-1} > \gamma \end{cases},$$

donde γ es tal que $\Pr(Z_{t-1} \leq \gamma) = 0.5$ y $\varepsilon_t \sim iidN(0, \sigma^2)$, con $\sigma^2 = 1$. El proceso es estacionario por el Corolario 2.9., ya que $\phi_2 = 1$ y $|\phi_1| < 1$, con $E(\delta_t) = 0.1$ y $E(\delta_t^2) = 0.82$.

Simulamos una realización de 250 observaciones a partir del modelo anterior, cuyo comportamiento gráfico se presenta en la Figura 10.

Modelo TAR simulado, phi1=-.8, phi2=1, p=.5, var=1, E(delcuad)=.82

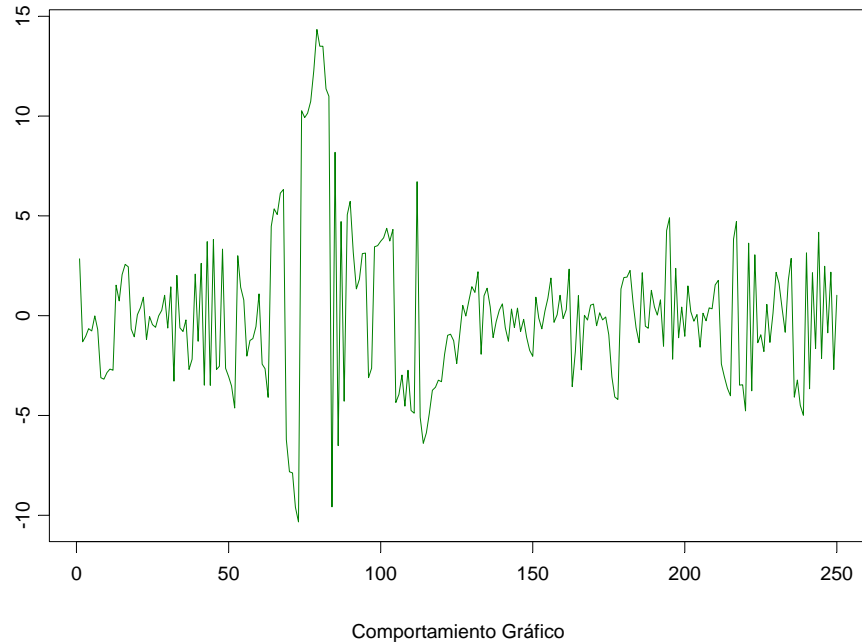


Figura 10. Modelo TAR(2; 1, 1) con $Z_{t-1} \sim AR(1)$ gaussiano.

Como antes consideramos un escenario donde el proceso TAR es cercano a la condición de no estacionariedad para ilustrar los comportamientos de los residuales. Ajustamos un modelo $TAR(2; 1, 1)$, y a partir de los parámetros estimados calculamos los residuales asociados; podemos observar de la Figura 11 que los residuales presentan un comportamiento homogéneo en el tiempo.

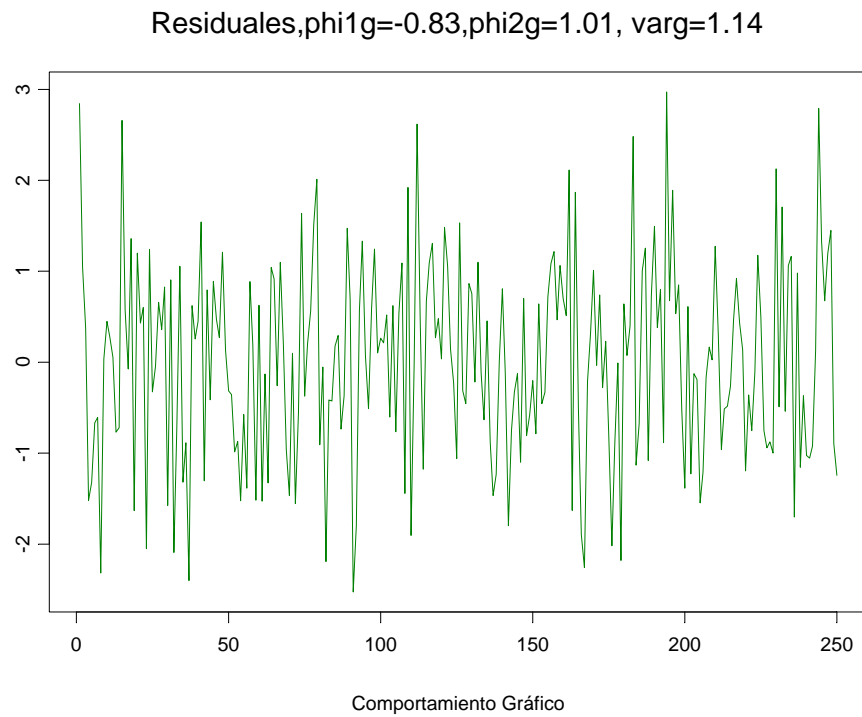


Figura 11. Residuales para el modelo $TAR(2; 1, 1)$ con $Z_{t-1} \sim AR(1)$ gaussiano.

Sin embargo, al ajustar al conjunto de datos, un modelo $AR(1)$ (se ajustó también un $AR(2)$ pero presentamos sólo los resultados para el $AR(1)$ en virtud de que las conclusiones son similares), y a partir de de este calculamos los residuales estimados. La Figura 12 a) presenta los comportamientos de las funciones de autocorrelación para los residuales y residuales al cuadrado bajo el modelo $AR(1)$. Observemos que los valores de las funciones de autocorrelación para los residuales al cuadrado decaen lentamente identificando que existe una estructura de dependencia en los residuales.

2.2. ERROR CUADRÁTICO MEDIO BAJO ESPECIFICACIONES INCORRECTAS47

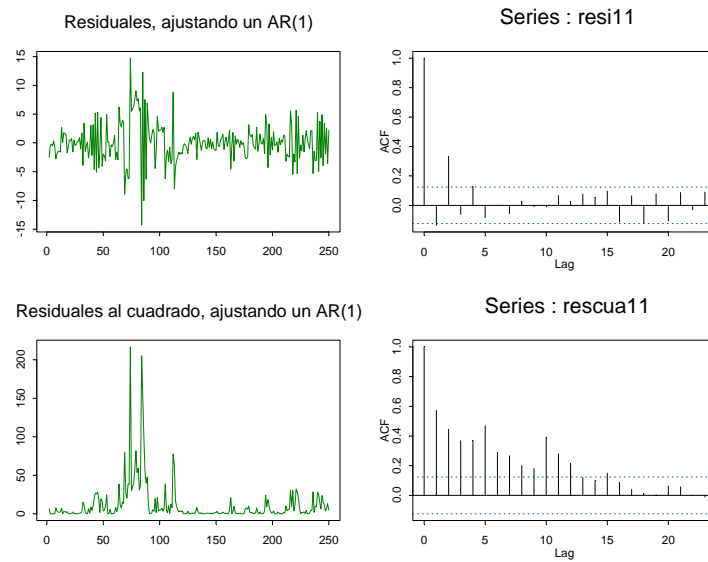


Figura 12 a). Funciones de autocorrelación para los residuales, modelo AR(1).

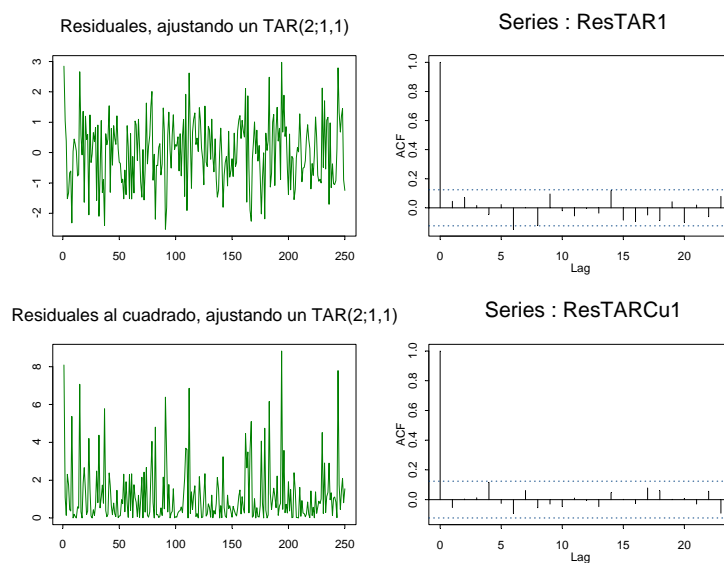


Figura 12 b). Autocorrelaciones para los residuales en el modelo TAR(2;1,1).

Por su parte la Figura 12 b) presenta las funciones de autocorrelación para los residuales y residuales al cuadrado bajo el modelo TAR(2;1,1). Note que las funciones de autocorrelación bajo este modelo pueden considerarse como cero y de aquí que no se identifica estructura de

dependencia en los residuales bajo el modelo TAR.

La Figura 13 presenta las gráficas cuantil cuantil para los residuales bajo los modelos $AR(1)$, $AR(2)$ y $TAR(2; 1, 1)$ respectivamente, notemos que el supuesto de normalidad bajo el esquema lineal no se cumple ya que los comportamientos de los puntos en los extremos de las gráficas se alejan de la línea recta. La gráfica de cuantil cuantil bajo el modelo TAR si reproduce un comportamiento típico de un conjunto de datos provenientes de un proceso gaussiano.

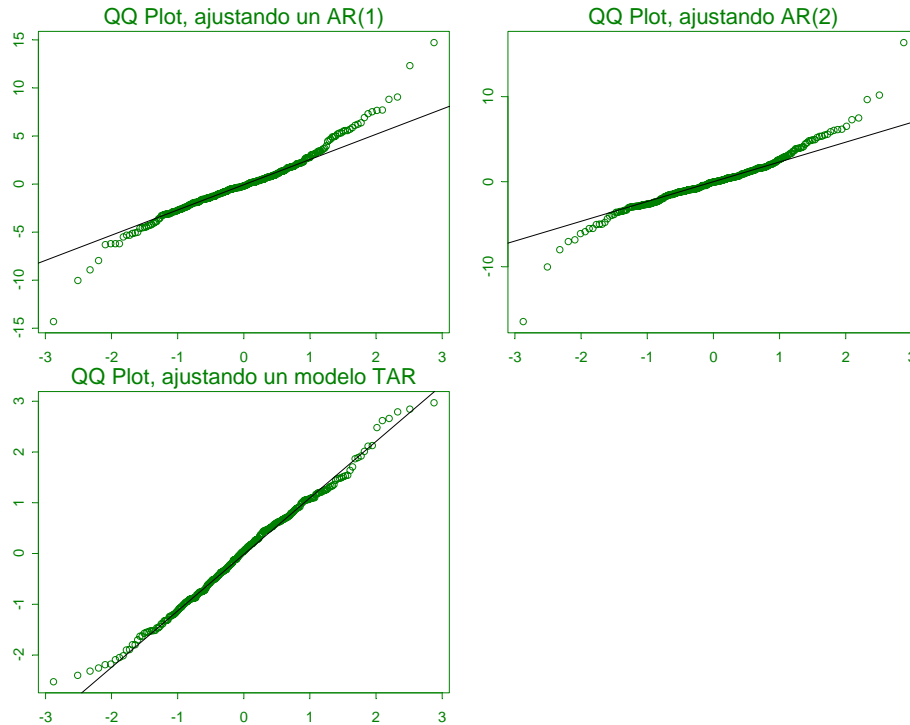


Figura 13. Gráficas de papel normal para los residuales de cada uno de los modelos ajustados.

Demostramos que ante especificaciones incorrectas del verdadero modelo generador de los datos, el proceso de error asociado al modelo incorrecto se comporta como un proceso de volatilidad. Los comportamientos asimétricos observados en una gran variedad de fenómenos reales no pueden describirse por medio de un modelo lineal y de aquí que los residuales ante la mala especificación hereden los comportamientos asimétricos del verdadero proceso generador. Mostramos que los comportamientos de los residuales ante la mala especificación no se comportan como un proceso de ruido blanco. Sin embargo, el incremento de la variabilidad se contrarresta al considerar metodologías de estimación, obtención y evaluación de pronósticos distintas, lo cual favorece al modelo lineal y de aquí que los pronósticos bajo el modelo TAR presenten desempeños pobres comparados con sus competidores lineales.

En el Capítulo 6 retomaremos dicho resultado, y estudiaremos a detalle la alternativa de regresión por cuantiles, como una nueva propuesta para cuantificar los desempeños de los pronósticos en los modelos TAR por medio de la función de pérdida de cuantiles, es decir, cuantificamos los desempeños de los pronósticos en modelos TAR(2; 1, 1) contra los desempeños de los modelos AR(1) por medio de una función de pérdida que maneja en forma adecuada la asimetría natural de los modelos TAR. La función de pérdida de cuantiles corresponde a una función de pérdida asimétrica equivalente a las funciones de pérdida asimétricas estudiadas en Christoffersen y Diebold (1996). Granger (2002), justifica el estudio de medidas de volatilidad, tales como varianzas o retornos absolutos, mediante el estudio de los riesgos relacionados. Una forma útil y simple de estudiar una distribución condicional es considerando sus cuantiles, es decir, estudiando la distribución de los cuantiles en el tiempo.

Capítulo 3

Estimación en modelos TAR

Una de las características prácticas en los modelos de umbral es que comparados con otros modelos no lineales los procedimientos de estimación son relativamente fáciles de implementar computacionalmente. En este capítulo describiremos los procedimientos de máxima verosimilitud y mínimos cuadrados para la estimación de los parámetros en el modelo TAR dado en la ecuación (2.1). Analizamos las propiedades asintóticas de los parámetros estimados para γ fijo, bajo ambos procedimientos de estimación. La estimación del parámetro γ puede realizarse por medio del método de mínimos cuadrados, vía una búsqueda directa, Chan (1993) y Hansen (1997). O bien por medio de la función de log verosimilitud perfil discutida en la Sección 3.2.1.

3.1 Estimación por mínimos cuadrados

Los estimadores de mínimos cuadrados para los modelos SETAR, así como sus propiedades asintóticas, se deben a los trabajos de Petrucci y Woolford (1984), Chan *et al* (1985), Chen y Tsay (1991), Chan (1993), Hansen (1997). Sin embargo, es importante garantizar que dichas propiedades asintóticas siguen cumpliéndose para el caso de los modelos TAR(2; 1, 1), cuya expresión matemática está dada en la ecuación (2.1), para $d = 1$. Suponemos que el proceso es estacionario y ergódico, bajo las condiciones dadas en los respectivos Teoremas de estacionariedad y ergodicidad (Teoremas 2.6. y 2.10 del Capítulo 2). La secuencia de errores $\{\varepsilon_t\}$ y la variable umbral Z_{t-1} son mutuamente independientes y cumplen los supuestos S1 a S6 enunciados en el Capítulo 2, Sección 1.

Sea $\psi = (\phi_1, \phi_2, \sigma)'$ el vector de parámetros en el modelo TAR(2; 1, 1) dado en la ecuación (2.13), los estimadores de mínimos cuadrados condicionales para ψ , considerando que el

parámetro γ es fijo, son aquellos valores que minimizan la siguiente suma de cuadrados,

$$\sum_{t=1}^T [Y_t - E(Y_t | \mathcal{F}_{t-1}; \psi)]^2, \quad (3.1)$$

donde T es el tamaño de la muestra observada, \mathcal{F}_{t-1} es la σ -álgebra generada por Y_{t-1}, Y_{t-2}, \dots , y $E(Y_t | \mathcal{F}_{t-1}; \psi) = [\phi_1 I(Z_{t-1} \leq \gamma) + \phi_2 I(Z_{t-1} > \gamma)] Y_{t-1}$ es la esperanza condicional de Y_t dado \mathcal{F}_{t-1} . Observemos que la esperanza condicional sólo depende de los valores de ϕ_1 y ϕ_2 , entonces dados los valores observados $Y_0, Y_1, Y_2, \dots, Y_T$, tenemos que los estimadores de mínimos cuadrados condicionales $\hat{\phi}_1(\gamma)$ y $\hat{\phi}_2(\gamma)$, para los parámetros ϕ_1 y ϕ_2 respectivamente, se obtienen minimizando la suma de cuadrados de los residuales dados en la ecuación (3.1), es decir, $\hat{\phi}_1(\gamma)$ y $\hat{\phi}_2(\gamma)$ están dados por las siguientes ecuaciones:

$$\hat{\phi}_1(\gamma) = \frac{\sum_{t=1}^T Y_t I(Z_{t-1} \leq \gamma) Y_{t-1}}{\sum_{t=1}^T I(Z_{t-1} \leq \gamma) Y_{t-1}^2} \text{ y } \hat{\phi}_2(\gamma) = \frac{\sum_{t=1}^T Y_t I(Z_{t-1} > \gamma) Y_{t-1}}{\sum_{t=1}^T I(Z_{t-1} > \gamma) Y_{t-1}^2}, \quad (3.2)$$

donde los residuales se obtienen de la igualdad: $\hat{e}_t = Y_t - E(Y_t | \mathcal{F}_{t-1}; \hat{\phi}_1(\gamma), \hat{\phi}_2(\gamma))$, a partir de los cuales obtenemos el estimador natural para σ^2 , dado por:

$$\hat{\sigma}^2(\gamma) = \frac{1}{T} \sum_{t=1}^T \hat{e}_t^2 = \frac{1}{T} \sum_{t=1}^T \left\{ Y_t - [\hat{\phi}_1(\gamma) I(Z_{t-1} \leq \gamma) + \hat{\phi}_2(\gamma) I(Z_{t-1} > \gamma)] Y_{t-1} \right\}^2. \quad (3.3)$$

Observe que como notación nos referimos a los estimadores $\hat{\phi}_1(\gamma)$, $\hat{\phi}_2(\gamma)$ y $\hat{\sigma}^2(\gamma)$ como funciones del parámetro γ , esto es con el fin de tener presente que dichos estimadores dependen del valor de γ considerado.

3.1.1 Propiedades de los estimadores de mínimos cuadrados

Para el modelo SETAR(2; 1, 1), con γ conocido, Petrucci y Woolford (1984) demuestran que los estimadores dados en las ecuaciones (3.2) y (3.3), son estimadores consistentes de ϕ_1 , ϕ_2 y σ^2 respectivamente. Asimismo se prueba que la distribución asintótica de los estimadores autorregresivos es normal. Para el caso del modelo TAR(2; 1, 1) dado en la ecuación (2.13), con variable umbral Z_{t-1} , estacionaria, ergódica y mutuamente independiente de la secuencia de errores $\{\varepsilon_t\}$, también se cumplen las propiedades asintóticas de los estimadores.

Teorema 3.1 (*Consistencia de los estimadores de mínimos cuadrados en el modelo TAR*). Consideremos el modelo TAR(2; 1, 1) estacionario y ergódico, dado en la ecuación (2.13).

Entonces para γ fijo, los estimadores de mínimos cuadrados para ϕ_1 , ϕ_2 y σ^2 dados en las ecuaciones (3.2), y (3.3) respectivamente, son estimadores consistentes.

Demostración. Una forma equivalente de escribir $\hat{\phi}_1(\gamma)$, dado en la ecuación (3.2) es la siguiente:

$$\begin{aligned}\hat{\phi}_1(\gamma) &= \frac{\sum_{t=1}^T \{[\phi_1 I(Z_{t-1} \leq \gamma) + \phi_2 I(Z_{t-1} > \gamma)] Y_{t-1} + \varepsilon_t\} [I(Z_{t-1} \leq \gamma) Y_{t-1}]}{\sum_{t=1}^T I(Z_{t-1} \leq \gamma) Y_{t-1}^2} \\ &= \frac{\sum_{t=1}^T \phi_1 I(Z_{t-1} \leq \gamma) Y_{t-1}^2}{\sum_{t=1}^T I(Z_{t-1} \leq \gamma) Y_{t-1}^2} + \frac{\sum_{t=1}^T \varepsilon_t [I(Z_{t-1} \leq \gamma) Y_{t-1}]}{\sum_{t=1}^T I(Z_{t-1} \leq \gamma) Y_{t-1}^2} \\ &= \phi_1 + \frac{\sum_{t=1}^T \varepsilon_t [I(Z_{t-1} \leq \gamma) Y_{t-1}]}{\sum_{t=1}^T I(Z_{t-1} \leq \gamma) Y_{t-1}^2},\end{aligned}$$

Equivalentemente, podemos escribir $\hat{\phi}_2(\gamma)$ en forma similar y obtener las siguientes expresiones para $\hat{\phi}_1(\gamma)$ y $\hat{\phi}_2(\gamma)$ respectivamente:

$$\begin{aligned}\hat{\phi}_1(\gamma) &= \phi_1 + \frac{\sum_{t=1}^T \varepsilon_t [I(Z_{t-1} \leq \gamma) Y_{t-1}]}{\sum_{t=1}^T I(Z_{t-1} \leq \gamma) Y_{t-1}^2}, \\ \hat{\phi}_2(\gamma) &= \phi_2 + \frac{\sum_{t=1}^T \varepsilon_t [I(Z_{t-1} > \gamma) Y_{t-1}]}{\sum_{t=1}^T I(Z_{t-1} > \gamma) Y_{t-1}^2}.\end{aligned}\tag{3.4}$$

Observemos que por la condición de estacionariedad y ergodicidad de las series involucradas en las expresiones (3.4), también las series en las sumatorias anteriores son ergódicas, y se cumplen los siguientes resultados asintóticos,

$$\begin{aligned}\frac{1}{T} \sum_{t=1}^T I(Z_{t-1} \leq \gamma) Y_{t-1}^2 &\xrightarrow[T \rightarrow \infty]{c.s.} E(I(Z_{t-1} \leq \gamma) Y_{t-1}^2) < \infty, \\ \frac{1}{T} \sum_{t=1}^T I(Z_{t-1} > \gamma) Y_{t-1}^2 &\xrightarrow[T \rightarrow \infty]{c.s.} E(I(Z_{t-1} > \gamma) Y_{t-1}^2) < \infty,\end{aligned}$$

ya que por el supuesto S3 dado en el Sección 2.1, el cual establece que $E(\varepsilon_t^{2\tau} | \mathfrak{F}_{t-1}) \leq B < \infty$ para algún $\tau > 1$ y por la ergodicidad de la series ε_t y Z_t , se obtienen las condiciones de estacionariedad y ergodicidad para Y_t . Asimismo, tenemos que:

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^T \varepsilon_t [I(Z_{t-1} \leq \gamma) Y_{t-1}] &\xrightarrow[T \rightarrow \infty]{c.s.} E(\varepsilon_1) E[I(Z_{t-1} \leq \gamma) Y_{t-1}] = 0, \\ \frac{1}{T} \sum_{t=1}^T \varepsilon_t [I(Z_{t-1} > \gamma) Y_{t-1}] &\xrightarrow[T \rightarrow \infty]{c.s.} E(\varepsilon_1) E[I(Z_{t-1} > \gamma) Y_{t-1}] = 0, \end{aligned}$$

de tal forma que $\hat{\phi}_1(\gamma) \xrightarrow[T \rightarrow \infty]{c.s.} \phi_1$ y $\hat{\phi}_2(\gamma) \xrightarrow[T \rightarrow \infty]{c.s.} \phi_2$.

Rescribamos el estimador de la varianza dado en la ecuación (3.3), como sigue:

$$\begin{aligned} &\hat{\sigma}^2(\gamma) \\ &= \frac{1}{T} \sum_{t=1}^T \left\{ \phi_1 I(Z_{t-1} \leq \gamma) Y_{t-1} + \phi_2 I(Z_{t-1} > \gamma) Y_{t-1} + \varepsilon_t \right. \\ &\quad \left. - \hat{\phi}_1(\gamma) I(Z_{t-1} \leq \gamma) Y_{t-1} - \hat{\phi}_2(\gamma) I(Z_{t-1} > \gamma) Y_{t-1} \right\}^2 \\ &= \frac{1}{T} \sum_{t=1}^T \left\{ -(\hat{\phi}_1(\gamma) - \phi_1) I(Z_{t-1} \leq \gamma) Y_{t-1} - (\hat{\phi}_2(\gamma) - \phi_2) I(Z_{t-1} > \gamma) Y_{t-1} + \varepsilon_t \right\}^2 \\ &= \frac{1}{T} \sum_{t=1}^T \varepsilon_t^2 + \frac{(\phi_1 - \hat{\phi}_1(\gamma))^2}{T} \sum_{t=1}^T I(Z_{t-1} \leq \gamma) Y_{t-1}^2 + \frac{(\phi_2 - \hat{\phi}_2(\gamma))^2}{T} \sum_{t=1}^T I(Z_{t-1} > \gamma) Y_{t-1}^2 \\ &\quad - 2 \frac{(\hat{\phi}_1(\gamma) - \phi_1)}{T} \sum_{t=1}^T I(Z_{t-1} \leq \gamma) Y_{t-1} \varepsilon_t - 2 \frac{(\hat{\phi}_2(\gamma) - \phi_2)}{T} \sum_{t=1}^T I(Z_{t-1} > \gamma) Y_{t-1} \varepsilon_t, \end{aligned}$$

entonces $\hat{\sigma}^2(\gamma) \xrightarrow[T \rightarrow \infty]{c.s.} \sigma^2$, por las propiedades de consistencia de los estimadores $\hat{\phi}_1(\gamma)$ y $\hat{\phi}_2(\gamma)$. De tal forma que los estimadores de mínimos cuadrados para ϕ_1 , ϕ_2 y σ^2 son estimadores consistente, bajo el supuesto de que el parámetro γ es fijo. \square

La distribución asintótica de los estimadores autorregresivos de mínimos cuadrados para ϕ_1 y ϕ_2 se da en el siguiente teorema.

Teorema 3.2 (*Distribución asintótica de los estimadores de mínimos cuadrados*). *Consideremos el modelo $TAR(2; 1, 1)$ estacionario y ergódico, entonces para γ conocido tenemos que $\sqrt{T}(\hat{\phi}_i - \phi_i) \xrightarrow[T \rightarrow \infty]{d} N(0, D_i)$, con $D_i = \frac{\sigma^2}{p_i E(Y_{t-1}^2)}$, $i = 1, 2$, $p_1 = \Pr(Z_{t-1} \leq \gamma)$, y $p_2 = 1 - \Pr(Z_{t-1} > \gamma)$.*

Demostración. A partir de las ecuaciones (3.4) tenemos que:

$$\sqrt{T}(\hat{\phi}_1 - \phi_1) = \sqrt{T} \left[\frac{\sum_{t=1}^T \varepsilon_t [I(Z_{t-1} \leq \gamma) Y_{t-1}]}{\sum_{t=1}^T I(Z_{t-1} \leq \gamma) Y_{t-1}^2} \right],$$

y

$$\sqrt{T}(\hat{\phi}_2 - \phi_2) = \sqrt{T} \left[\frac{\sum_{t=1}^T \varepsilon_t [I(Z_{t-1} > \gamma) Y_{t-1}]}{\sum_{t=1}^T I(Z_{t-1} > \gamma) Y_{t-1}^2} \right],$$

entonces, por el teorema de límite central para series de tiempo estacionarias y ergódicas demostrado por Hannan (1973), concluimos el resultado del teorema. \square

Para el caso de umbral desconocido en los modelos SETAR, Chan (1993) y Hansen (1997) proponen estimar el parámetro de umbral γ por medio de un procedimiento de regresión secuencial, descrito a continuación: consideremos los estimadores de mínimos cuadrados para ϕ_1 , ϕ_2 y σ^2 respectivamente, los cuales dependen del valor del parámetro de umbral γ ,

$$\hat{\phi}_1(\gamma) = \frac{\sum_{t=1}^T Y_t I(Z_{t-1} \leq \gamma) Y_{t-1}}{\sum_{t=1}^T I(Z_{t-1} \leq \gamma) Y_{t-1}^2}, \quad \hat{\phi}_2(\gamma) = \frac{\sum_{t=1}^T Y_t I(Z_{t-1} > \gamma) Y_{t-1}}{\sum_{t=1}^T I(Z_{t-1} > \gamma) Y_{t-1}^2}$$

y

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

Entonces, el estimador de mínimos cuadrados para γ es el valor que minimiza el valor de la varianza, es decir,

$$\hat{\gamma} = \arg \min_{\gamma \in \Gamma} \hat{\sigma}_T^2(\gamma), \quad (3.5)$$

donde $\Gamma = [\underline{\gamma}, \overline{\gamma}]$ es un intervalo acotado.

El problema de minimización de la varianza anterior se resuelve por búsqueda directa y en general la búsqueda se realiza barriendo sobre todos los valores posible entre el cuantil 15% y 85% de los valores de la variable umbral, con la finalidad de garantizar un mínimo de observaciones en el proceso de estimación. Finalmente los estimadores para ϕ_1 , ϕ_2 y σ^2 se obtienen identificando los valores correspondientes en el valor de γ que minimiza la varianza, es decir, $\hat{\phi}_1(\hat{\gamma})$, $\hat{\phi}_2(\hat{\gamma})$ y $\hat{\sigma}_T^2 = \hat{\sigma}_T^2(\hat{\gamma})$ respectivamente.

Para el caso en que el parámetro d es desconocido, el procedimiento de estimación de mínimos cuadrados consiste en estimar d paralelamente con los otros parámetros, de modo que el problema de estimación incluye ahora una búsqueda sobre d y en vez de T regresiones el método de búsqueda requerirá $T\bar{d}$ regresiones, con \bar{d} el valor máximo posible que puede tomar d . Ya que el espacio paramétrico para d es discreto el estimador de mínimos cuadrados \hat{d} será fuertemente consistente. Para propósitos de inferencia en los otros parámetros podemos pensar que d se conoce con certeza, ver Hansen (1997) para más detalles. Chan (1993) demuestra que los estimadores de mínimos cuadrados para los coeficientes autorregresivos con γ desconocido son consistentes y asintóticamente normales. En nuestro caso mostraremos que los estimadores de MV en el modelo TAR(2;1,1) son fuertemente consistencia fuerte, bajo el supuesto de que la distribución de los errores posee cuarto momento finito. En particular la distribución normal cumple con dicho supuesto, lo que implica que también los estimadores de mínimos cuadrados son fuertemente consistentes ya que estos coinciden con los estimadores de máxima verosimilitud bajo el supuesto de normalidad en los errores.

3.2 Estimación mediante máxima verosimilitud

Para fines de ilustración y comprensión, rescribamos nuevamente el modelo autorregresivo de primer orden, con dos regímenes y parámetro $d = 1$ y que denotamos por TAR(2;1,1),

$$Y_t = \begin{cases} \phi_1 Y_{t-1} + \varepsilon_t, & \text{si } Z_{t-1} \leq \gamma, \\ \phi_2 Y_{t-1} + \varepsilon_t & \text{si } Z_{t-1} > \gamma, \end{cases}$$

o equivalentemente,

$$\begin{aligned} Y_t &= [\phi_1 I(Z_{t-1} \leq \gamma) + \phi_2 I(Z_{t-1} > \gamma)] Y_{t-1} + \varepsilon_t \\ &= h(Z_{t-1}, Y_{t-1}, \boldsymbol{\theta}) + \varepsilon_t, \quad t = 1, 2, \dots, \end{aligned} \tag{3.6}$$

donde $h(\cdot) = [\phi_1 I(Z_{t-1} \leq \gamma) + \phi_2 I(Z_{t-1} > \gamma)] Y_{t-1}$, $I(\cdot)$ denota la función indicadora, y en general Z_{t-1} corresponde a una serie de tiempo observada hasta el tiempo t . Identificamos los parámetros en el modelo mediante los vectores $\boldsymbol{\theta}_0$ y $\boldsymbol{\theta}_1$, con $\boldsymbol{\theta}_0 = (\phi_1, \phi_2, \sigma^2, \gamma)$ y $\boldsymbol{\theta}_1$ el vector que contiene los parámetros asociados al proceso Z_{t-1} . González y Gonzalo (1998, 1999) dan las condiciones de estacionariedad y ergodicidad del modelo dado en la ecuación

(3.6) en función de los coeficientes autorregresivos y la variable umbral Z_{t-1} , ver Corolarios 2.7, 2.8, y 2.9, así como los Teoremas 2.6 y 2.10 del Capítulo 2. Los modelos explorados para la variable umbral contemplan el caso en que la variable umbral es independiente e idénticamente distribuida y el caso en que Z_{t-1} sigue una cadena de Markov de orden N . En particular para $N = 1$, asumiremos que Z_{t-1} sigue un proceso autorregresivo gaussiano de orden uno. El caso más general dado en la ecuación (2.4) se trabaja de manera similar.

Enfatizamos que la variable umbral Z_t es una serie de tiempo estacionaria que rige el comportamiento de la serie Y_t como función del valor del parámetro de umbral, ya que los valores de Y_t dependen de los valores que toma $I(Z_{t-1} \leq \gamma)$. Sin embargo, consideramos que el comportamiento de la serie Z_{t-1} es independiente de los valores de la serie Y_t . En particular supondremos dos escenarios para el proceso Z_{t-1} : Z_{t-1} un proceso estacionario autorregresivo de orden uno ($Z_{t-1} \sim \text{AR}(1)$) y el caso en que Z_{t-1} es un proceso independiente con distribución uniforme en el intervalo $(0,1)$, es decir, $Z_{t-1} \sim \text{iid}U(0,1)$.

3.2.1 Obtención de los estimadores de máxima verosimilitud

Supongamos que los errores $\{\varepsilon_t\}$ en la ecuación (3.6) son variables aleatorias independientes tales que $\varepsilon_t \sim \text{iid}D$ y que ε_t es independiente de $Y_{t-1}, Y_{t-2}, \dots, t \geq 1$, la distribución D tiene media cero y varianza finita. Consideremos también que el proceso Z_{t-1} , estacionario, está dado por: $Z_{t-1} = \rho Z_{t-2} + u_t$ con $u_t \sim \text{iid}N(0, \sigma_u^2)$ y ε_t mutuamente independiente de Z_{t-1} , de tal forma que para tamaños de muestra grande T , tenemos que la función de verosimilitud puede escribirse como:

$$\begin{aligned} & L(\phi_1, \phi_2, \sigma^2, \gamma, \rho, \sigma_u^2; \{Y_t, Z_t\}) \\ &= \prod_{t=2}^T f(Y_t | Y_{t-1}, Z_{t-d}; \phi_1, \phi_2, \sigma^2, \gamma) f_1(Z_{t-d} | Z_{t-d}; \rho, \sigma_u^2) g_{\boldsymbol{\vartheta}}(\mathbf{Y}_1, \mathbf{Z}_{t-d}) \\ &= \prod_{t=2}^T f(Y_t - h(\mathbf{Y}_{t-1}, \mathbf{Z}_{t-d}, \boldsymbol{\vartheta}); \sigma) f_1(Z_{t-1} - h_1(\mathbf{Z}_{t-2}, \rho); \sigma_u) g_{\boldsymbol{\vartheta}}(\mathbf{Y}_1, \mathbf{Z}_{t-d}), \end{aligned}$$

donde $\boldsymbol{\vartheta} = (\phi_1, \phi_2, \gamma)'$. Consideraremos la función de verosimilitud condicional del vector de parámetros $\boldsymbol{\theta} = (\phi_1, \phi_2, \sigma^2, \gamma, \rho, \sigma_u^2)'$, dada por:

$$L(\phi_1, \phi_2, \sigma^2, \gamma, \rho, \sigma_u^2; \{Y_t, Z_t\}) = \prod_{t=2}^T f(Y_t - h(\mathbf{Y}_{t-1}, \mathbf{Z}_{t-1}, \boldsymbol{\vartheta}); \sigma) f_1(Z_{t-1} - h_1(\mathbf{Z}_{t-d-1}, \rho); \sigma_u).$$

Para una serie observada $Y_1, \dots, Y_T, Z_0, \dots, Z_{T-1}$ con D una distribución normal con media cero y varianza σ^2 , obtenemos que la función de verosimilitud condicional de $\boldsymbol{\theta}$, dada la

muestra

$$\begin{aligned}
& L(\phi_1, \phi_2, \sigma^2, \gamma, \rho, \sigma_u^2; Y_t, Z_{t-1}) \\
&= \prod_{t=2}^T f(Y_t | Y_{t-1}, Z_{t-1}; \phi_1, \phi_2, \sigma^2, \gamma) f_1(Z_{t-1} | Z_{t-2}; \rho, \sigma_u^2) \\
&= \frac{1}{\sigma^{T-1}} \exp \left[-\frac{1}{2\sigma^2} \sum_{t=2}^T (Y_t - \delta_t Y_{t-1})^2 \right] \\
&\quad \times \frac{1}{\sigma_u^{T-1}} \exp \left[-\frac{1}{2\sigma_u^2} \sum_{t=2}^T (Z_{t-1} - \rho Z_{t-2})^2 \right]
\end{aligned} \tag{3.7}$$

y de aquí que la función de log verosimilitud se escribe como:

$$\begin{aligned}
\ell(\phi_1, \phi_2, \sigma^2, \gamma, \rho, \sigma_u^2) &= -(T-1) \log \sigma - \frac{1}{2\sigma^2} \sum_{t=2}^T (Y_t - \delta_t Y_{t-1})^2 - (T-1) \log \sigma_u^2 \\
&\quad - \frac{1}{2\sigma_u^2} \sum_{t=2}^T (Z_{t-1} - \rho Z_{t-2})^2.
\end{aligned}$$

De la ecuación anterior resolviendo $\partial \ell(\phi_1, \phi_2, \sigma^2, \gamma, \rho, \sigma_u^2) / \partial \sigma^2 = 0$ obtenemos la siguiente expresión para la varianza estimada:

$$\hat{\sigma}^2(\phi_1, \phi_2, \gamma) = \frac{1}{T-1} \sum_{t=2}^T (Y_t - \delta_t Y_{t-1})^2,$$

donde el término de la derecha satisface la siguiente igualdad,

$$\sum_{t=2}^T (Y_t - \delta_t Y_{t-1})^2 = \sum_{t=2}^T (Y_t^2 - 2\delta_t Y_{t-1} Y_t + \delta_t^2 Y_{t-1}^2),$$

de forma tal que las derivadas parciales de la log verosimilitud con respecto a ϕ_1 y ϕ_2 respectivamente satisfacen la ecuación:

$$\frac{\partial \ell}{\partial \phi_i} = -\frac{1}{2\sigma^2} \sum_{t=2}^T (-2I_t Y_{t-1} Y_t + 2\phi_i I_t Y_{t-1}^2), \quad i = 1, 2.$$

Igualando la ecuación anterior a cero obtenemos las siguientes igualdades para encontrar los

estimadores de máxima verosimilitud (EMV) para ϕ_1 y ϕ_2 como función del parámetro γ ,

$$\begin{aligned}\hat{\phi}_1(\gamma) &= \frac{\sum_{t=2}^T I_t Y_t Y_{t-1}}{\sum_{t=2}^T I_t Y_{t-1}^2}, \\ \hat{\phi}_2(\gamma) &= \frac{\sum_{t=2}^T (1 - I_t) Y_t Y_{t-1}}{\sum_{t=2}^T (1 - I_t) Y_{t-1}^2} = \frac{\sum_{t=2}^T Y_t Y_{t-1} - \sum_{t=2}^T I_t Y_t Y_{t-1}}{\sum_{t=2}^T Y_{t-1}^2 - \sum_{t=2}^T I_t Y_{t-1}^2}\end{aligned}\quad (3.8)$$

con $I_t = I(Z_t \leq \gamma)$, luego el estimador por máxima verosimilitud para σ^2 como función de γ está dado por:

$$\hat{\sigma}^2(\gamma) = \hat{\sigma}^2(\hat{\phi}_1, \hat{\phi}_2, \gamma) = \frac{1}{T-1} \sum_{t=2}^T \left(Y_t - \hat{\delta}_t(\gamma) Y_{t-1} \right)^2, \quad (3.9)$$

con:

$$\hat{\delta}_t(\gamma) = \begin{cases} \hat{\phi}_1(\gamma), & Z_{t-1} \leq \gamma, \\ \hat{\phi}_2(\gamma), & Z_{t-1} > \gamma. \end{cases}$$

Similarmente, los estimadores de MV para ρ y σ_u^2 , respectivamente son:

$$\hat{\rho} = \frac{\sum_{t=2}^T Z_{t-1} Z_{t-2}}{\sum_{t=2}^T Z_{t-2}^2}, \quad \hat{\sigma}_u^2(\hat{\rho}) = \frac{1}{T-1} \sum_{t=2}^T (Z_t - \hat{\rho} Z_{t-1})^2. \quad (3.10)$$

Las expresiones de los estimadores de máxima verosimilitud para los coeficientes ϕ_1 , ϕ_2 y σ^2 están en función de un valor fijo γ . No es posible derivar la función de log verosimilitud con respecto al parámetro γ , debido a que esta no es continua en los puntos Z_0, \dots, Z_{T-1} , en este caso lo recomendable para estimar el parámetro γ es por medio de un método de búsqueda. Una vez calculados los EMV para γ fijo, se evalúa la función de log verosimilitud en estos valores obteniendo la función de log verosimilitud perfil, es decir,

$$\begin{aligned}\ell_p(\gamma) &= \ell \left[\hat{\phi}_1(\gamma), \hat{\phi}_2(\gamma), \hat{\sigma}_\gamma, \hat{\rho}, \hat{\sigma}_u \right] \\ &= -(T-1) \log \hat{\sigma}_\gamma - \frac{1}{2\hat{\sigma}_\gamma^2} \sum_{t=2}^T \left(Y_t - \hat{\delta}_t(\gamma) Y_{t-1} \right)^2 \\ &\quad - (T-1) \log \hat{\sigma}_u^2 - \frac{1}{2\hat{\sigma}_u^2} \sum_{t=2}^T (Z_t - \hat{\rho} Z_{t-1})^2.\end{aligned}$$

Observemos que $\hat{\rho}$ y $\hat{\sigma}_u$ no dependen funcionalmente de γ , por lo que podemos considerar para efectos de maximización respecto al parámetro γ la siguiente función de log verosimilitud perfil:

$$\begin{aligned}\ell_p^*(\gamma) &= -(T-1) \log \hat{\sigma}_\gamma - \frac{1}{2\hat{\sigma}_\gamma^2} \sum_{t=2}^T \left(Y_t - \hat{\delta}_t(\gamma) Y_{t-1} \right)^2 \\ &= -(T-1) \log \hat{\sigma}_\gamma - \frac{T-1}{2\hat{\sigma}_\gamma^2} \hat{\sigma}_\gamma^2 \\ &= -(T-1) \log \hat{\sigma}_\gamma - \frac{T-1}{2} \\ &= -\frac{T-1}{2} \log \left[\frac{1}{T-1} \sum_{t=2}^T \left(Y_t - \hat{\delta}_t(\gamma) Y_{t-1} \right)^2 \right] - \frac{T-1}{2}.\end{aligned}$$

Maximizar $\ell_p^*(\gamma)$ en la ecuación anterior es equivalente a minimizar $-\ell_p^*(\gamma)$, es decir, minimizar la siguiente función objetivo,

$$\frac{T-1}{2} \log \left[\frac{1}{T-1} \sum_{t=2}^T \left(Y_t - \hat{\delta}_t(\gamma) Y_{t-1} \right)^2 \right]$$

con respecto a γ . Este procedimiento nos da una infinidad de valores de γ , la recomendación de Qian (1998) es tomar $\hat{\gamma}$ como el valor de γ que cumpla con la siguiente condición:

$$\inf \arg \min_{\gamma} \frac{T-1}{2} \log \left[\frac{1}{T-1} \sum_{t=2}^T \left(Y_t - \hat{\delta}_t(\gamma) Y_{t-1} \right)^2 \right].$$

Las ecuaciones para el caso en que la variable Z_{t-1} sigue un proceso independiente y con distribución uniforme en $(0, 1)$ pueden escribirse de forma equivalente con la diferencia de que ahora la función de densidad $f_1(Z_{t-1}|Z_{t-2}; \rho, \sigma_u^2) = 1$. Note entonces que las expresiones que contienen a los parámetros ρ y σ_u en las ecuaciones anteriores desaparecen.

3.2.2 Propiedades asintóticas de los estimadores de máxima verosimilitud

En la literatura resaltan tres artículos relacionados con las propiedades asintóticas de los EMV de los estimadores de mínimos cuadrados y máxima verosimilitud para los parámetros de los modelos TAR, bajo el supuesto de γ desconocido y $Z_{t-d} = Y_{t-d}$, es decir, en modelos SETAR. Estos son: Chan (1993), Qian (1998) y Chan y Tsay (1998).

Chan (1993) y Qian (1998) consideran el caso de SETAR discontinuos y Chan y Tsay (1998) el caso SETAR continuo.

Definición 3.3 (*Función autorregresiva discontinua, Chan 1993*). Sean ϕ_1 y ϕ_2 los coeficientes del modelo SETAR(2;1,1), se dice que el modelo tiene una *función autorregresiva discontinua* si existe

$$\mathbf{Z}_* = (1, z_{p-1}, \dots, z_0)', \text{ donde } z_{p-d} = \gamma, \quad (3.11)$$

tal que $(\phi_1 - \phi_2) \mathbf{Z}_* \neq 0$. En este caso, el umbral γ se convierte en el *punto de salto* de la función autorregresiva. Si $(\phi_1 - \phi_2) \mathbf{Z}_* = 0$ para toda \mathbf{Z}_* que satisfaga (3.11) se dice que el modelo tiene una *función autorregresiva continua*.

Observemos que la definición de continuidad en la función autorregresiva está relacionada con un *cambio de régimen suave* cuando el valor de $Y_{t-d} = \gamma$.

El artículo de Chan (1993) considera la estimación por mínimos cuadrados, el cual es equivalente al método de máxima verosimilitud cuando la distribución de la secuencia $\{\varepsilon_t\}$ es gaussiana ($\varepsilon_t \sim iidN(0, \sigma^2)$). Qian (1998) demuestra las propiedades asintóticas de los EMV de los parámetros del modelo SETAR en el caso de que los $\{\varepsilon_t\} \sim iidD$, donde D es una distribución con media cero y varianza finita. Sin embargo, Qian (1998) no considera la estimación de los parámetros de la distribución D . Mientras que Chan (1993) sí estima y estudia las propiedades del estimador de σ .

Conviene resaltar que en el modelo TAR considerado en esta tesis no se tiene el concepto de continuidad.

En el caso del modelo TAR (2.13) no podemos hablar de discontinuidad en la función autorregresiva en el sentido de la Definición 3.3 porque Y_{t-1} puede tomar el valor γ y no involucra cambio de régimen. El cambio de régimen se da cuando $Z_{t-d} = \gamma$ y si quisiéramos que el cambio de régimen sea continuo tendríamos que pedir que $\phi_1 y_{t-1} = \phi_2 y_{t-1}$ cuando $Z_{t-d} = \gamma$ lo cual en el caso continuo tiene probabilidad cero.

En los modelos TAR no podemos auxiliarnos de esta propiedad porque la continuidad o discontinuidad de un TAR no se puede determinar.

Consistencia fuerte de los estimadores de máxima verosimilitud

La demostración se realizará para el caso general donde los errores $\{\varepsilon_t\}$ tienen distribución arbitraria con media cero y cuarto momento finito.

Consideraremos el modelo TAR(2; p, p) dado en la ecuación (2.1) del Capítulo 2.

El modelo (2.1) se puede escribir de la siguiente manera equivalente

$$Y_t = h(Y_{t-1}, Z_{t-1}, \phi) + \varepsilon_t, \quad t \geq 2, \quad (3.12)$$

para algún $\phi = (\phi'_1, \phi'_2, \gamma, d) \in \mathbb{R}^{2p+3} \times \{1, 2, \dots, p\}$, donde $\mathbf{Y}_{t-1} = (Y_0, Y_{-1}, \dots, Y_{-p+1})'$, $\mathbf{Z}_{t-d} = (Z_{t-d}, \dots, Z_{t-d-p+1})'$ $\phi_j = (\phi_{0j}, \phi_{1j}, \dots, \phi_{pj})' \in \mathbb{R}^{p+1}$, $j = 1, 2$ y para $\mathbf{y} \in \mathbb{R}^p$ y $\mathbf{z} \in \mathbb{R}^p$,

$$h(\mathbf{y}, \mathbf{z}, \boldsymbol{\alpha}) = \left(\phi_{01} + \sum_{k=1}^p \phi_{k1} y_k \right) I(z_d \leq \gamma) + \left(\phi_{02} + \sum_{k=1}^p \phi_{k2} y_k \right) I(z_d > \gamma).$$

Los errores $\{\varepsilon_t\}$ en (3.12) son $iidN(0, \sigma^2)$.

Teorema 3.4 (*Consistencia de los estimadores de MV*). *Suponga que $\{Y_t, Z_t\}$ del modelo (3.12) es estacionario y ergódico y que se cumplen las condiciones SC1-SC4 del Apéndice A. Entonces $\hat{\boldsymbol{\theta}}_T$ es un estimador fuertemente consistente de $\boldsymbol{\theta}$.*

Las demostración del Teorema 3.4 se da el Apéndice A.

Capítulo 4

Pronósticos en modelos TAR aproximando la esperanza condicional

En el Capítulo 1 expusimos distintas alternativas para obtener pronósticos multi pasos en modelos no lineales en general, sin embargo nuestro interés se centra en estudiar una clase particular de modelos no lineales conocidos en la literatura como modelos TAR. Recordemos la representación matemática de un modelo TAR de dos regímenes, de orden uno e intercepto igual a cero en cada uno de los regímenes, denotado como TAR(2; 1, 1) y dado por la ecuación:

$$\begin{aligned} Y_t &= \phi_1 Y_{t-1} I(Z_{t-1} \leq \gamma) + \phi_2 Y_{t-1} I(Z_{t-1} > \gamma) + \varepsilon_t \\ &= \delta_t Y_{t-1} + \varepsilon_t, \end{aligned}$$

con, $\delta_t = \phi_1 I(Z_{t-1} \leq \gamma) + \phi_2 I(Z_{t-1} > \gamma)$; Z_t y ε_t procesos estacionarios y ergódicos y Z_t un proceso independiente de ε_t . Consideramos que el proceso Y_t es estacionario y ergódico bajo las condiciones dadas en González y Gonzalo (1998). Al tiempo $T + 1$ escribimos el modelo como sigue:

$$Y_{T+1} = \delta_{T+1} Y_T + \varepsilon_{T+1},$$

de modo que el pronóstico un paso adelante puede calcularse fácilmente mediante la siguiente expresión,

$$\hat{Y}_{T+1} = E(Y_{T+1} | \mathfrak{F}_T) = E(\delta_{T+1} Y_T + \varepsilon_{T+1} | \mathfrak{F}_T) = \delta_{T+1} Y_T,$$

donde \mathfrak{F}_T contiene la información disponible hasta el tiempo T . Al tiempo $T + 2$ la representación del modelo TAR(2; 1, 1), se escribe como:

$$Y_{T+2} = \delta_{T+2} Y_{T+1} + \varepsilon_{T+2},$$

a partir de dicha expresión tenemos que el pronóstico dos pasos adelante es como sigue:

$$\begin{aligned}\hat{Y}_{T+2} &= E(Y_{T+2}|\mathfrak{F}_T) = E(\delta_{T+2}Y_{T+1} + \varepsilon_{T+2}|\mathfrak{F}_T) = E(\delta_{T+2}Y_{T+1}|\mathfrak{F}_T) \\ &= E\{[\phi_1 I(Z_{T+1} \leq \gamma) + \phi_2 I(Z_{T+1} > \gamma)] Y_{T+1}|\mathfrak{F}_T\} \\ &= \phi_1 E[I(Z_{T+1} \leq \gamma) Y_{T+1}|\mathfrak{F}_T] + \phi_2 E[I(Z_{T+1} > \gamma) Y_{T+1}|\mathfrak{F}_T],\end{aligned}$$

observemos que para obtener el pronóstico de Y_{T+2} necesitamos conocer el valor de Z_{T+1} o bien el valor de la variable indicadora $I(Z_{T+1} \leq \gamma)$.

Igualmente al tiempo $T + 3$, describimos el modelo TAR(2; 1, 1) como:

$$Y_{T+3} = \delta_{T+3}Y_{T+2} + \varepsilon_{T+3},$$

de modo que el pronóstico tres pasos adelante,

$$\hat{Y}_{T+3} = E(Y_{T+3}|\mathfrak{F}_T) = E(\delta_{T+3}Y_{T+2} + \varepsilon_{T+3}|\mathfrak{F}_T) = E(\delta_{T+3}Y_{T+2}|\mathfrak{F}_T).$$

Asimismo el pronóstico h -pasos adelante se expresa como sigue:

$$\hat{Y}_{T+h} = E(Y_{T+h}|\mathfrak{F}_T) = E(\delta_{T+h}Y_{T+(h-1)} + \varepsilon_{T+h}|\mathfrak{F}_T) = E(\delta_{T+h}Y_{T+(h-1)}|\mathfrak{F}_T),$$

o equivalentemente,

$$\begin{aligned}\hat{Y}_{T+h} &= E(Y_{T+h}|\mathfrak{F}_T) \\ &= E([\phi_1 I(Z_{T+h-1} \leq \gamma) + \phi_2 I(Z_{T+h-1} > \gamma)] Y_{T+h-1} + \varepsilon_{T+h}|\mathfrak{F}_T) \\ &= \phi_1 E[I(Z_{T+h-1} \leq \gamma) Y_{T+h-1}|\mathfrak{F}_T] + \phi_2 E[I(Z_{T+h-1} > \gamma) Y_{T+h-1}|\mathfrak{F}_T].\end{aligned}\quad (4.1)$$

De la ecuación anterior (4.1) vemos que para obtener el pronóstico h -pasos adelante para Y_t es necesario conocer el valor de la variable umbral Z_t al tiempo $T + h - 1$ o equivalentemente conocer el valor de la variable indicadora $I(Z_{T+h-1} \leq \gamma)$. Note que para el caso en que $Z_{t-1} \sim iidU(0, 1)$ el pronóstico h -pasos para Y_t está dado por:

$$\hat{Y}_{T+h} = \phi_1 I(Z_{T+h-1} \leq \gamma) \hat{Y}_{T+h-1} + \phi_2 I(Z_{T+h-1} > \gamma) \hat{Y}_{T+h-1},$$

y bajo este escenario conocemos los valores que toma $I(Z_t \leq \gamma)$, para $t = T, T + 1, \dots, T + h - 1$.

Por otro lado si Z_{t-1} sigue un AR(1), podemos considerar dicho modelo para pronosticar Z_t utilizando la teoría de pronósticos ampliamente estudiada en el caso lineal, o bien considerar los valores que toma la variable indicadora en el tiempo, y modelar a la variable indicadora $I(Z_{t-1} \leq \gamma)$ como una serie de tiempo binaria. Posteriormente, con base a dicho modelo obtener el comportamiento futuro de la serie misma. De esta forma obtenemos el

pronóstico para Y_t por medio de las siguientes ecuaciones, según utilicemos el pronóstico Z_{T+h-1} o de $I(Z_{T+h-1} \leq \gamma)$ respectivamente,

$$\hat{Y}_{T+h} \simeq \phi_1 I(\hat{Z}_{T+h-1} \leq \gamma) \hat{Y}_{T+h-1} + \phi_2 I(\hat{Z}_{T+h-1} > \gamma) \hat{Y}_{T+h-1},$$

ó

$$\hat{Y}_{T+h} \simeq \phi_1 \hat{I}(Z_{T+h-1} \leq \gamma) \hat{Y}_{T+h-1} + \phi_2 \hat{I}(Z_{T+h-1} > \gamma) \hat{Y}_{T+h-1}.$$

En la práctica los valores de los parámetros en el modelo TAR son desconocidos y se hace necesario estimarlos. Podemos estimar dichos parámetros mediante los procedimientos de mínimos cuadrados o máxima verosimilitud, descritos en Capítulo 2, los cuales son estimadores consistentes y asintóticamente normales. Recordemos también que si los errores en el modelo TAR siguen un proceso gaussiano con media cero y varianza σ^2 , ambos procedimientos de estimación coinciden.

Como parte de nuestro ejercicio de simulación implementamos un procedimiento recursivo del tipo Monte-Carlo para aproximar los pronósticos h -pasos adelante en el modelo TAR, pronosticando Z_t bajo el supuesto de que este sigue un proceso AR(1) estacionario con ruido gaussiano, o bien pronosticando $I(Z_t \leq \gamma)$ mediante la teoría de modelos de regresión para series binarias estudiados en Kedem y Fokianos (2002). Cuando Z_t es un proceso independiente con distribución uniforme $(0, 1)$, simulamos $T + h$ números aleatorios $U(0, 1)$ y tomamos a estos como valores dados para aproximar las esperanzas condicionales.

4.1 Modelos de regresión para series binarias

Consideremos la serie binaria $\{I_t\} = \{I(Z_t \leq \gamma)\}$ que toma los valores de 0 ó 1 dependiendo del valor del parámetro de umbral γ , con información auxiliar o covariables dada por $\{\mathbf{B}_{t-1}\}$, visto como un problema de regresión nos interesa estimar la probabilidad condicional de éxito, denotada por:

$$P_{\alpha}(I_t = 1 \mid \mathfrak{F}_{t-1}), \quad (4.2)$$

donde α es el vector de parámetros p -dimensional y \mathfrak{F}_{t-1} es todo lo que conocemos o hemos observado hasta el tiempo $t - 1$, tanto de la serie de tiempo como de sus posibles covariables. En nuestro caso particular tenemos que la variable binaria esta definida como:

$$I_t = I(Z_t \leq \gamma) = \begin{cases} 1 & \text{si } Z_t \leq \gamma, \\ 0 & \text{si } Z_t > \gamma, \end{cases}$$

y una representación posible para la matriz de covariables es $\mathbf{B}_{t-1} = \{1, Z_{t-1}, I_{t-1}\}$, es decir, la matriz de covariables puede contener rezagos de la variable umbral y rezagos de la variable indicadora. De está forma nuestro interés radica en predecir el comportamiento futuro de los

eventos $\{Z_t \leq \gamma\}$ considerando la información contenida en la matriz de covariables \mathbf{B}_{t-1} . La serie de tiempo binaria I_t puede ser estacionaria o no estacionaria, mientras que la matriz de covariables \mathbf{B}_{t-1} puede representar una o más series, así como funciones de la serie de interés, Kedem y Fokianos (2002), Capítulo 2, páginas 42 y 43. \mathbf{B}_{t-1} no necesita ser estacionaria, sin embargo se debe garantizar que se cumplen los supuestos A, del Capítulo 1, página 16, del libro de Kedem y Fokianos (2002), los cuales garantizan un modelo bien definido y estimadores consistentes bajo condiciones razonables a largo plazo.

Para garantizar que el modelo dado en la ecuación (4.2) estime probabilidades (valores entre cero y uno) debemos elegir una función denominada como función link, digamos F , que mapee la recta real sobre el intervalo $(0, 1)$. Cuando la serie de tiempo binaria se condiciona sólo en función de sus valores pasados, como en nuestro caso, la función de distribución acumulada sugerida o función link, es la función de distribución logística estándar, es decir, se asume el siguiente modelo de regresión logística,

$$\pi_t(\boldsymbol{\alpha}) \equiv \Pr(I_t = 1 \mid \mathfrak{F}_{t-1}) = F_l(\boldsymbol{\alpha}'\mathbf{B}_{t-1}) = \frac{1}{1 + \exp[-\boldsymbol{\alpha}'\mathbf{B}_{t-1}]}, \quad (4.3)$$

donde $\boldsymbol{\alpha}$ es un vector de parámetros de dimensión p igual a las entradas de la matriz de covariables \mathbf{B}_{t-1} . La transformada inversa de la expresión anterior es la función link canónica para datos binarios y se conoce como función logit,

$$\text{Logit}[\pi_t(\boldsymbol{\alpha})] \equiv \log \left\{ \frac{\pi_t(\boldsymbol{\alpha})}{1 - \pi_t(\boldsymbol{\alpha})} \right\} = \boldsymbol{\alpha}'\mathbf{B}_{t-1},$$

o bien modelo logístico.

Observemos que $\pi_t(\boldsymbol{\alpha}) = E(I_t \mid \mathfrak{F}_{t-1})$ y $\text{Var}(I_t \mid \mathfrak{F}_{t-1}) = F_l(\boldsymbol{\alpha}'\mathbf{B}_{t-1})[1 - F_l(\boldsymbol{\alpha}'\mathbf{B}_{t-1})]$, de modo que

$$\Pr(I_t = i_t \mid \mathfrak{F}_{t-1}) = [\pi_t(\boldsymbol{\alpha})]^{i_t} [1 - \pi_t(\boldsymbol{\alpha})]^{1-i_t},$$

y la verosimilitud parcial (PL) de $\boldsymbol{\alpha}$, tal y como la definen en Kedem y Fokianos (2002), dada una muestra observada de tamaño T está dada por el producto de las probabilidades,

$$\begin{aligned} \text{PL}(\boldsymbol{\alpha}) &= \prod_{t=1}^T [\pi_t(\boldsymbol{\alpha})]^{I_t} [1 - \pi_t(\boldsymbol{\alpha})]^{1-I_t} \\ &= \prod_{t=1}^T [F_l(\boldsymbol{\alpha}'\mathbf{B}_{t-1})]^{I_t} [1 - F_l(\boldsymbol{\alpha}'\mathbf{B}_{t-1})]^{1-I_t}. \end{aligned}$$

El estimador de máxima verosimilitud parcial (MVP) para $\boldsymbol{\alpha}$ se encuentra maximizando la verosimilitud anterior o equivalentemente maximizando la log verosimilitud parcial con

respecto al parámetro α , dada por:

$$\begin{aligned}
l(\alpha) &= \sum_{t=1}^T I_t \log(F_t(\alpha' \mathbf{B}_{t-1})) + (1 - I_t) \log(1 - F_t(\alpha' \mathbf{B}_{t-1})) \\
&= \sum_{t=1}^T I_t \log\left(\frac{1}{1 + \exp[-\alpha' \mathbf{B}_{t-1}]}\right) + (1 - I_t) \log\left(1 - \frac{1}{1 + \exp[-\alpha' \mathbf{B}_{t-1}]}\right) \\
&= \sum_{t=1}^T \left\{ -I_t \log(1 + \exp[-\alpha' \mathbf{B}_{t-1}]) + (1 - I_t) \log\left(\frac{\exp[-\alpha' \mathbf{B}_{t-1}]}{1 + \exp[-\alpha' \mathbf{B}_{t-1}]}\right) \right\} \\
&= \sum_{t=1}^T \left\{ -I_t \log(1 + \exp[-\alpha' \mathbf{B}_{t-1}]) + (1 - I_t) (-\alpha' \mathbf{B}_{t-1} - \log(1 + \exp[-\alpha' \mathbf{B}_{t-1}])) \right\} \\
&= \sum_{t=1}^T \left\{ -I_t \log(1 + \exp[-\alpha' \mathbf{B}_{t-1}]) - \alpha' \mathbf{B}_{t-1} - \log(1 + \exp[-\alpha' \mathbf{B}_{t-1}]) \right. \\
&\quad \left. + I_t \alpha' \mathbf{B}_{t-1} + I_t \log(1 + \exp[-\alpha' \mathbf{B}_{t-1}]) \right\} \\
&= \sum_{t=1}^T \left\{ -\alpha' \mathbf{B}_{t-1} - \log(1 + \exp[-\alpha' \mathbf{B}_{t-1}]) + I_t \alpha' \mathbf{B}_{t-1} \right\}.
\end{aligned}$$

A partir de la función de log verosimilitud parcial se tiene que el vector o función de Score para la serie binaria dada por:

$$\begin{aligned}
S_T(\alpha) &= \nabla l(\alpha) \\
&= \frac{\partial \sum_{t=1}^T I_t \log(F(\alpha' \mathbf{B}_{t-1})) + (1 - I_t) \log(1 - F(\alpha' \mathbf{B}_{t-1}))}{\partial \alpha} \\
&= \sum_{t=1}^T I_t \mathbf{B}_{t-1} \frac{f(\alpha' \mathbf{B}_{t-1})}{F(\alpha' \mathbf{B}_{t-1})} + (1 - I_t) \frac{-\mathbf{B}_{t-1} f(\alpha' \mathbf{B}_{t-1})}{1 - F(\alpha' \mathbf{B}_{t-1})} \\
&= \sum_{t=1}^T \frac{I_t \mathbf{B}_{t-1} f(\alpha' \mathbf{B}_{t-1})}{F(\alpha' \mathbf{B}_{t-1})} - \frac{\mathbf{B}_{t-1} f(\alpha' \mathbf{B}_{t-1})}{1 - F(\alpha' \mathbf{B}_{t-1})} + \frac{I_t \mathbf{B}_{t-1} f(\alpha' \mathbf{B}_{t-1})}{1 - F(\alpha' \mathbf{B}_{t-1})} \\
&= \sum_{t=1}^T \frac{I_t \mathbf{B}_{t-1} f(\alpha' \mathbf{B}_{t-1})}{F(\alpha' \mathbf{B}_{t-1}) (1 - F(\alpha' \mathbf{B}_{t-1}))} - \frac{\mathbf{B}_{t-1} f(\alpha' \mathbf{B}_{t-1})}{1 - F(\alpha' \mathbf{B}_{t-1})} \left(\frac{F(\alpha' \mathbf{B}_{t-1})}{F(\alpha' \mathbf{B}_{t-1})} \right) \\
&= \sum_{t=1}^T \mathbf{B}_{t-1} D(\alpha' \mathbf{B}_{t-1}) (I_t - F(\alpha' \mathbf{B}_{t-1})) = \sum_{t=1}^T \mathbf{B}_{t-1} D(\alpha' \mathbf{B}_{t-1}) (I_t - \pi_t(\alpha)),
\end{aligned}$$

con

$$D(\alpha' \mathbf{B}_{t-1}) = \frac{f(\alpha' \mathbf{B}_{t-1})}{F(\alpha' \mathbf{B}_{t-1}) [1 - F(\alpha' \mathbf{B}_{t-1})]}$$

y

$$f(\alpha' \mathbf{B}_{t-1}) = \frac{\partial}{\partial \alpha} F(\alpha' \mathbf{B}_{t-1}).$$

La matriz de información condicional, $G_T(\alpha)$, está dada por

$$\begin{aligned} G_T(\alpha) &= \sum_{t=1}^T \text{Cov}[\mathbf{S}_T(\alpha) \mid \mathfrak{F}_{t-1}] = \sum_{t=1}^T \mathbf{B}_{t-1} D(\alpha' \mathbf{B}_{t-1}) f(\alpha' \mathbf{B}_{t-1}) \mathbf{B}_{t-1}' \\ &= \sum_{t=1}^T \mathbf{B}_{t-1} \frac{f^2(\alpha' \mathbf{B}_{t-1})}{F(\alpha' \mathbf{B}_{t-1}) (1 - F(\alpha' \mathbf{B}_{t-1}))} \mathbf{B}_{t-1}' \\ &= \mathbf{B}_{t-1} \text{Diag} \left(\frac{\exp[\alpha' \mathbf{B}_{t-1}]}{\{1 + \exp[\alpha' \mathbf{B}_{t-1}]\}^2} \right) \mathbf{B}_{t-1}'. \end{aligned}$$

Kedem y Fokianos (2002) demuestran que para T suficientemente grande el estimador por MVP $\hat{\alpha} \xrightarrow{p} \alpha$ (propiedad de consistencia) y $\sqrt{T}(\hat{\alpha} - \alpha) \xrightarrow{dist.} N(0, G^{-1}(\alpha))$ (normalidad asintótica). Notemos que en el caso de regresión logística, y de interés en nuestro caso, se satisfacen las siguientes igualdades,

$$f(x) = \frac{\partial}{\partial x} \left(\frac{e^x}{1 + e^x} \right) = \frac{(1 - e^x)e^x + e^x e^x}{(1 + e^x)^2} = \frac{e^x}{(1 + e^x)^2}$$

y

$$\begin{aligned} F(x)(1 - F(x)) &= \frac{e^x}{1 + e^x} \left(1 - \frac{e^x}{1 + e^x} \right) \\ &= \frac{e^x}{1 + e^x} \left(\frac{1}{1 + e^x} \right) = \frac{e^x}{(1 + e^x)^2} \end{aligned}$$

de tal forma que se cumple las siguientes igualdades:

$$D(\alpha' \mathbf{B}_{t-1}) = \frac{f(\alpha' \mathbf{B}_{t-1})}{F(\alpha' \mathbf{B}_{t-1}) [1 - F(\alpha' \mathbf{B}_{t-1})]} = 1 \quad \text{y} \quad \frac{\partial D(\alpha' \mathbf{B}_{t-1})}{\partial \alpha} = 0.$$

Entonces la función objetivo a maximizar para estimar el parámetro α , se reduce a maximizar la siguiente función:

$$\sum_{t=1}^T \mathbf{B}_{t-1} [I_t - \pi_t(\alpha)], \quad (4.4)$$

con respecto a α . Equivalentemente, podemos también obtener los pronósticos I_t , para $t = T + 1, \dots, T + h$, maximizando la función (4.4) hasta el tiempo $T + h$ con respecto a los valores desconocidos de I_{T+1}, \dots, I_{T+h} , con $\hat{\alpha}$ el estimador por MVP previamente obtenido.

Equivalentemente podemos pronosticar los valores de la variable indicadora por búsqueda directa en todas las posibles combinaciones de unos y ceros. Es decir, pensemos que deseamos pronosticar la serie binaria h -pasos adelante y observemos que podemos listar los 2^h posibles arreglos de unos y ceros. Es decir, si nuestro interés consiste en pronosticar la serie 3-pasos adelante tendremos que el pronóstico obtenido deberá ser uno de los siguientes arreglos:

$$(0, 0, 0), (0, 0, 1), (0, 1, 0), (1, 0, 0), (1, 1, 0), (1, 1, 1),$$

enlistados todos los posibles arreglos podemos sustituir cada una de dichas posibilidades en la verosimilitud parcial de α y maximizar dicha función con respecto a α , de tal forma que nos quedamos con el arreglo de unos y ceros con mayor valor de la función objetivo dada en la ecuación (4.4).

A continuación damos una breve descripción del algoritmo implementado, para la obtención de los pronósticos, en virtud de que este será retomado en el Capítulo 7 correspondiente al ejercicio de simulación.

Primeramente para la variable umbral Z_{t-1} , simulamos 500 realizaciones con 255 observaciones cada una, considerando dos escenarios posibles, $Z_{t-1} \sim iidU(0, 1)$ y Z_{t-1} proceso AR(1) estacionario. A partir de las series simuladas bajo cada escenario, simulamos 500 series considerando el modelo TAR(2; 1, 1) dado en la ecuación (3.6). Las primeras 250 observaciones las utilizamos para la estimación de los parámetros, bajo dos posible modelos: un TAR(2; 1, 1) y un AR(1). Posteriormente, obtenemos los pronósticos de la variable umbral y también de la variable indicadora. Con estos valores y los valores de los parámetros estimados implementamos un procedimiento recursivo para obtener los pronósticos de Y_{T+h} para $h = 1, 2, 3, 4$, y 5, realizamos 500 iteraciones para la obtención de dichos pronósticos. Como parte del algoritmo de Monte Carlo recursivo obtenemos también los correspondientes errores de pronósticos y a partir de estos calculamos las respectivas funciones de pérdida (ECMP y pérdida de cuantiles). Comparamos los desempeños de las pérdidas asociados al modelo TAR(2; 1, 1) contra las pérdidas asociados al modelo AR(1), la idea es obtener los cocientes para $h = 1, 2, 3, 4$, y 5 de los valores para cada una de las funciones de pérdida en el AR(1) entre los valores de cada una de las funciones de pérdidas en el modelo TAR(2; 1, 1), de tal forma que si el valor correspondiente es mayor que uno decimos que los pronósticos en el modelo TAR(2; 1, 1) tienen mejor desempeño que los pronósticos en el modelo AR(1) en el sentido de que los valores de las pérdidas asociadas al modelo TAR(2; 1, 1) son menores para cada horizonte de pronóstico.

Capítulo 5

Pronósticos en modelos TAR mediante verosimilitud predictiva

El problema de predecir valores no observados o valores futuros de variables aleatorias es fundamental en estadística, una alternativa de solución para este problema es vía el concepto de verosimilitud predictiva. Desde el punto de vista Bayesiano, el problema de predicción de valores no observados se resuelve directamente encontrando la densidad predictiva posterior de las variables no observadas dados los datos. El problema de estimación y obtención de pronósticos h -pasos adelante en los modelos SETAR ha sido estudiado desde el punto de vista Bayesiano, por ejemplo: Chen y Lee (1995), Chen (1998) y Sáfadi y Morettin (2000). Sin embargo, como parte de nuestra revisión bibliográfica no contamos con referencias que aborden el problema de predicción en modelos TAR mediante la alternativa de verosimilitud predictiva estudiada en este trabajo. Geisser (2002) presenta un panorama general de inferencia predictiva considerando los enfoques Bayesiano y no Bayesiano.

La propuesta de verosimilitud predictiva para ordenar la plausibilidad de valores futuros, fue desarrollada por Lauritzen (1974), Hinkley (1979), y generalizada por Butler (1986). Aunque, la idea inicial fue introducida por Fisher (1956) para el modelo binomial.

Sea $\mathbf{X} = \mathbf{x}$ el valor observado de la variable aleatoria \mathbf{X} , el problema consiste en predecir valores no observados \mathbf{x}^* de \mathbf{X}^* , la inferencia se realiza sobre los valores \mathbf{x}^* . Suponemos que $(\mathbf{X}, \mathbf{X}^*)'$ tiene una densidad de probabilidad con respecto a la medida de Lebesgue, denotada por $f_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{x}^*)$, o equivalentemente denotada por $f(\mathbf{x}, \mathbf{x}^*; \boldsymbol{\theta})$, donde $\boldsymbol{\theta}$ es el vector de parámetros desconocidos. Igualmente denotamos la función de densidad condicional como $f(\mathbf{x}^*|\mathbf{x}; \boldsymbol{\theta})$. Sea $\hat{\boldsymbol{\theta}}$ el estimador por máxima verosimilitud para $\boldsymbol{\theta}$ basado en los datos \mathbf{x} , y $\hat{\boldsymbol{\theta}}_{\mathbf{w}}$ el estimador de máxima verosimilitud basado en $\mathbf{w} = (\mathbf{x}, \mathbf{x}^*)'$. Supongamos que $\mathbf{X} = (Y_1, \dots, Y_T)'$, contiene la muestra observada de tamaño T y $\mathbf{X}^* = (Y_{T+1}, \dots, Y_{T+h})'$ la muestra no observada sobre la cual deseamos hacer inferencias.

Fundamentalmente el problema de predicción consiste en lo siguiente: consideramos dos

cantidades desconocidas, \mathbf{x}^* y $\boldsymbol{\theta}$, donde el interés primario es obtener información sobre \mathbf{x}^* con $\boldsymbol{\theta}$ jugando el papel de parámetro de estorbo. El principio de verosimilitud para predicción, formulado por Berger y Wolpert (1984), parte de que toda la evidencia sobre $(\mathbf{x}^*, \boldsymbol{\theta})$ está contenida en la función de verosimilitud conjunta,

$$L(\mathbf{x}^*, \boldsymbol{\theta}; \mathbf{x}) = f(\mathbf{x}, \mathbf{x}^*; \boldsymbol{\theta}), \quad (5.1)$$

donde partimos de la verosimilitud base para \mathbf{x} observado, y el objetivo es desarrollar una función de verosimilitud para \mathbf{x}^* , digamos $L(\mathbf{x}^*|\mathbf{x})$ eliminando $\boldsymbol{\theta}$ de la ecuación (5.1). $L(\mathbf{x}^*|\mathbf{x})$ con las características anteriores se conoce como verosimilitud predictiva. Debido a que existen diferentes formas de eliminar el parámetro de estorbo $\boldsymbol{\theta}$, surgen varias propuestas de verosimilitud predictiva. Bjørnstad (1990) contabiliza 14 propuestas de verosimilitud predictiva, sin embargo, algunas de estas son completamente similares, y todas ellas se basan fundamentalmente en una de las siguientes tres operaciones en $L(\mathbf{x}^*, \boldsymbol{\theta}; \mathbf{x})$: integración, maximización o condicionamiento. El enfoque Bayesiano es equivalente a integrar $L(\mathbf{x}^*, \boldsymbol{\theta}; \mathbf{x})$ con respecto a la distribución a priori para $\boldsymbol{\theta}$. Es decir, la densidad predictiva posterior, $f_0(\mathbf{x}^*|\mathbf{x})$ puede pensarse como la verosimilitud integrada marginal tal que, $f_0(\mathbf{x}^*|\mathbf{x}) \propto \int L(\mathbf{x}^*, \boldsymbol{\theta}; \mathbf{x}) d\boldsymbol{\theta}$.

En esta tesis consideramos la función de *verosimilitud predictiva perfil*¹ de las observaciones futuras \mathbf{x}^* , introducida por Mathiasen (1979) y recomendada por Bjørnstad (1990), la cual consiste en eliminar el parámetro de estorbo $\boldsymbol{\theta}$ por medio de maximización de la siguiente función de verosimilitud,

$$L_p(\mathbf{x}^*|\mathbf{x}) = \sup_{\boldsymbol{\theta}} f(\mathbf{x}, \mathbf{x}^*; \boldsymbol{\theta}) = L(\mathbf{x}^*, \hat{\boldsymbol{\theta}}_{\mathbf{w}}; \mathbf{x}). \quad (5.2)$$

Intuitivamente, la motivación de la función de verosimilitud predictiva L_p es como sigue: para \mathbf{x}^* el vector de parámetros de interés y $\boldsymbol{\theta}$ el vector de parámetros de estorbo, se obtienen los valores más probables para $\boldsymbol{\theta}$ dado $(\mathbf{x}, \mathbf{x}^*)'$ dando como resultado la función de verosimilitud dada en la ecuación (5.2). En inferencia paramétrica esta verosimilitud corresponde a la verosimilitud perfil, introducida por Kalbfleisch y Sprott (1970) y de aquí el nombre de verosimilitud predictiva perfil.

5.1 Verosimilitud predictiva en el modelo lineal AR(1)

Dado que nos interesa comparar los desempeños de los pronósticos en los modelo TAR(2; 1, 1) contra los desempeños de los pronósticos en los modelos AR(1), calculamos la función de verosimilitud predictiva para el modelo AR(1) y posteriormente la verosimilitud predictiva para el modelo TAR(2; 1, 1).

¹Del inglés profile predictive likelihood

Consideremos el modelo estacionario AR(1) dado en la siguiente ecuación:

$$Y_t = \rho Y_{t-1} + \eta_t,$$

donde $\eta_t \sim iidN(0, \sigma_\eta^2)$, $|\rho| < 1$ y $Y_1 \sim N(0, \frac{\sigma_\eta^2}{1-\rho^2})$. Escribamos el modelo AR(1) anterior al tiempo $T + h$, en su representación equivalente:

$$Y_{T+h} = \rho^h Y_T + \sum_{j=0}^{h-1} \rho^j \eta_{T+h-j},$$

sea $\boldsymbol{\theta} = (\rho, \sigma_\eta^2)'$ el vector de parámetros desconocidos, $\mathbf{y} = (y_1, y_2, \dots, y_T)'$ el vector de datos observados y $\mathbf{y}^* = (y_{T+1}, \dots, y_{T+h})'$ el vector de observaciones desconocidas futuras sobre las cuales deseamos hacer inferencias.

Tenemos que la función de densidad conjunta para $\mathbf{w} = (\mathbf{y}, \mathbf{y}^*)'$ está dada como:

$$\begin{aligned} f(\mathbf{w}; \boldsymbol{\theta}) &= f(y_1, \dots, y_{T-1}, y_T, y_{T+1}, \dots, y_{T+h}; \boldsymbol{\theta}) \\ &= f(y_1; \boldsymbol{\theta}) \prod_{t=2}^{T+h} f(y_t | y_{t-1}; \boldsymbol{\theta}), \end{aligned}$$

y de aquí que para T suficientemente grande, la función de verosimilitud basada en \mathbf{y} para $\boldsymbol{\theta}$ y \mathbf{y}^* es:

$$\begin{aligned} L(\boldsymbol{\theta}; \mathbf{y}, \mathbf{y}^*) &= \prod_{t=2}^{T+h} f(y_t | y_{t-1}; \boldsymbol{\theta}) \\ &= \prod_{t=2}^{T+h} \frac{1}{\sqrt{2\pi\sigma_\eta^2}} \exp \left[-\frac{1}{2\sigma_\eta^2} (y_t - \rho y_{t-1})^2 \right], \end{aligned}$$

igualmente la log verosimilitud para $\boldsymbol{\theta}$ y \mathbf{y}^* , dada la muestra observada, denotada por $\ell(\boldsymbol{\theta}; \mathbf{y}, \mathbf{y}^*)$ está dada por:

$$\ell(\boldsymbol{\theta}; \mathbf{y}, \mathbf{y}^*) = -(T + h - 1) \log(\sigma_\eta) - \frac{1}{2\sigma_\eta^2} \sum_{t=2}^{T+h} (y_t - \rho y_{t-1})^2. \quad (5.3)$$

derivando la ecuación (5.3) con respecto a σ_η y ρ respectivamente e igualando a cero dichas derivadas, tenemos las siguientes expresiones para los estimadores de σ_η^2 y ρ , respectivamente,

$$\sigma_\eta^2(\mathbf{y}^*; \rho) = \frac{1}{T + h - 1} \sum_{t=2}^{T+h} (y_t - \rho y_{t-1})^2, \quad \text{y} \quad \hat{\rho}(\mathbf{y}^*) = \frac{\sum_{t=2}^{T+h} y_t y_{t-1}}{\sum_{t=2}^{T+h} y_{t-1}^2}.$$

lo que implica que el estimador de la varianza este dado por:

$$\hat{\sigma}_\eta^2(\mathbf{y}^*) = \frac{1}{T+h-1} \sum_{t=2}^{T+h} [y_t - \hat{\rho}(\mathbf{y}^*) y_{t-1}]^2.$$

Entonces la log verosimilitud predictiva perfil para \mathbf{y}^* está dada por:

$$\begin{aligned} \ell_p(\mathbf{y}^*|\mathbf{y}) &= \ell_p(\mathbf{y}^*, \hat{\rho}(\mathbf{y}^*), \hat{\sigma}_\eta^2(\mathbf{y}^*); \mathbf{y}) \\ &= -(T+h-1) \log [\hat{\sigma}_\eta^2(\mathbf{y}^*)] - \frac{1}{2\hat{\sigma}_\eta^2(\mathbf{y}^*)} \sum_{t=2}^{T+h} (y_t - \hat{\rho}(\mathbf{y}^*) y_{t-1})^2 \\ &= -(T+h-1) \log [\hat{\sigma}_\eta^2(\mathbf{y}^*)] - \frac{1}{2}. \end{aligned}$$

Maximizar la ecuación anterior con respecto al vector \mathbf{y}^* es equivalente a minimizar la expresión $-\ell_p$ con respecto a \mathbf{y}^* y esto a su vez es equivalente a minimizar el término:

$$\hat{\sigma}_\eta^2(\mathbf{y}^*) = \frac{1}{T+h-1} \sum_{t=2}^{T+h} [y_t - \hat{\rho}(\mathbf{y}^*) y_{t-1}]^2, \quad (5.4)$$

con respecto a \mathbf{y}^* debido a que la función logaritmo es una función creciente. Finalmente los valores de $\mathbf{y}^* = (y_{T+1}, \dots, y_{T+h})'$ encontrados en el ejercicio de maximización son los pronósticos vía verosimilitud predictiva.

Note que $\ell_p(\mathbf{y}^*|\mathbf{y})$ es máxima en $\hat{\mathbf{y}}^* = \hat{\rho}y_T$, con $\hat{\rho} = (\hat{\rho}, \hat{\rho}^2, \dots, \hat{\rho}^h)$ ya que la ecuación (5.4) es mínima en $\hat{\mathbf{y}}^* = \hat{\rho}y_T$. Es decir, el pronóstico vía la verosimilitud predictiva para el caso AR(1) con errores gaussianos es también el pronóstico vía las ecuaciones de recursión con $\hat{\rho}$ el estimador por mínimos cuadrados o máxima verosimilitud, Clements y Hendry (1998).

5.2 Verosimilitud predictiva para el modelo TAR(2; 1, 1)

Consideremos ahora el modelo TAR(2; 1, 1) dado en la ecuación (3.6), nuestro interés es obtener h pronósticos hacia adelante para Y_t a partir de dicho modelo. Observemos que contamos con T observaciones del proceso y que podemos escribir el modelo de interés para $T+1, \dots, T+h$ mediante la siguiente ecuación:

$$Y_t^* = \delta_t^* Y_{t-1}^* + \varepsilon_t, \quad t = T+1, \dots, T+h,$$

con $\varepsilon_t \sim iidN(0, \sigma^2)$ y Z_t^* un proceso AR(1) estacionario con ruido gaussiano con media cero y varianza σ_u^2 ; y

$$\delta_t^* = \begin{cases} \phi_1, & Z_{t-1}^* \leq \gamma, \\ \phi_2, & Z_{t-1}^* > \gamma. \end{cases}$$

Sea también $\mathbf{y} = (y_1, \dots, y_T)'$, $\mathbf{y}^* = (y_{T+1}^*, \dots, y_{T+h}^*)'$, $\mathbf{z} = (z_0, \dots, z_T)'$, y $\mathbf{z}^* = (z_{T+1}^*, \dots, z_{T+h}^*)'$. Por notación definamos:

$$\mathbf{w} = (\mathbf{y}', \mathbf{y}^{*'})' \text{ y } \mathbf{v} = (\mathbf{z}', \mathbf{z}^{*'})',$$

entonces podemos escribir el modelo TAR(2; 1, 1), para $t = 1, 2, \dots, T, T+1, \dots, T+h$ como:

$$w_t = \delta_t w_{t-1} + \varepsilon_t, \quad (5.5)$$

para el cual la variable δ_t , está dada por:

$$\delta_t = \begin{cases} \phi_1, & v_{t-1} \leq \gamma, \\ \phi_2, & v_{t-1} > \gamma. \end{cases}$$

Podemos escribir la función de verosimilitud para $(\mathbf{y}^*, \mathbf{z}^*, \phi_1, \phi_2, \sigma, \gamma, \rho, \sigma_u)$ en función de (\mathbf{y}, \mathbf{z}) , con T suficientemente grande, y de forma equivalente a la verosimilitud de la ecuación (3.7), como:

$$\begin{aligned} L(\phi_1, \phi_2, \sigma, \gamma, \rho, \sigma_u; \mathbf{y}, \mathbf{y}^*, \mathbf{z}, \mathbf{z}^*) &= \frac{1}{\sigma^{T+h-1}} \exp \left[-\frac{1}{2\sigma^2} \sum_{t=2}^{T+h} (w_t - \delta_t w_{t-1})^2 \right] \\ &\times \frac{1}{\sigma_u^{T+h-1}} \exp \left[-\frac{1}{2\sigma_u^2} \sum_{t=2}^{T+h} (v_t - \rho v_{t-1})^2 \right]. \end{aligned}$$

Luego la función de log verosimilitud para (\mathbf{w}, \mathbf{v}) , está dada por:

$$\begin{aligned} \ell(\phi_1, \phi_2, \sigma, \gamma, \rho, \sigma_u; \mathbf{y}, \mathbf{y}^*, \mathbf{z}, \mathbf{z}^*) &= -(T+h-1) \log \sigma - \frac{1}{2\sigma^2} \sum_{t=2}^{T+h} (w_t - \delta_t w_{t-1})^2 \\ &- (T+h-1) \log \sigma_u - \frac{1}{2\sigma_u^2} \sum_{t=2}^{T+h} (v_t - \rho v_{t-1})^2. \end{aligned}$$

Derivando la función de log verosimilitud con respecto a los parámetros ϕ_1 , ϕ_2 , σ , ρ , y σ_u respectivamente, igualando a cero y resolviendo las respectivas derivadas parciales, obtenemos las siguientes expresiones para los estimadores de ϕ_1 , ϕ_2 , σ , ρ , y σ_u , respectivamente,

$$\hat{\sigma}^2(\mathbf{y}^*; \phi_1, \phi_2, \gamma) = \frac{1}{T+h-1} \sum_{t=2}^{T+h} (w_t - \delta_t w_{t-1})^2,$$

y

$$\hat{\phi}_1(\mathbf{y}^*; \gamma) = \frac{\sum_{t=2}^{T+h} I_t w_t w_{t-1}}{\sum_{t=2}^{T+h} I_t w_{t-1}^2}, \quad \hat{\phi}_2(\mathbf{y}^*; \gamma) = \frac{\sum_{t=2}^{T+h} (1 - I_t) w_t w_{t-1}}{\sum_{t=2}^{T+h} (1 - I_t) w_{t-1}^2},$$

con $I_t = I(v_t \leq \gamma)$, y

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

Lo que implica que sustituyendo $\hat{\phi}_1$ y $\hat{\phi}_2$ en la expresión para $\hat{\sigma}^2$, obtenemos la siguiente igualdad,

$$\begin{aligned} \hat{\sigma}^2(\mathbf{y}^*; \gamma) &= \hat{\sigma}^2(\mathbf{y}^*; \hat{\phi}_1, \hat{\phi}_2, \gamma) \\ &= \frac{1}{T+h-1} \sum_{t=2}^{T+h} \left(w_t - \hat{\delta}_t(\mathbf{y}^*; \gamma) w_{t-1} \right)^2, \end{aligned}$$

con

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

Igualmente, sustituyendo $\hat{\rho}$ en la expresión para $\hat{\sigma}_u^2$, tenemos que:

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

Entonces podemos escribir la log verosimilitud predictiva perfil para $(\mathbf{y}^*, \mathbf{z}^*)$ como:

$$\begin{aligned} \ell_p(\mathbf{y}^*, \mathbf{z}^*; \gamma | \mathbf{y}, \mathbf{z}) &= \ell_p \left[\mathbf{y}^*, \mathbf{z}^*, \hat{\phi}_1(\mathbf{y}^*; \gamma), \hat{\phi}_2(\mathbf{y}^*; \gamma), \hat{\sigma}(\mathbf{y}^*; \gamma), \hat{\rho}(\mathbf{z}^*), \hat{\sigma}^2(\mathbf{z}^*; \hat{\rho}) ; \mathbf{y}, \mathbf{z} \right] \\ &= -\frac{T+h-1}{2} \log \left[\frac{1}{T+h-1} \sum_{t=2}^{T+h} \left(w_t - \hat{\delta}_t(\mathbf{y}^*; \gamma) w_{t-1} \right)^2 \right] \\ &\quad - \frac{T+h-1}{2} \log \left[\frac{1}{T+h-1} \sum_{t=2}^{T+h} (v_t - \hat{\rho}(\mathbf{z}^*) v_{t-1})^2 \right] - (T+h-1). \end{aligned}$$

Como antes, maximizar $\ell_p(\mathbf{y}^*, \mathbf{z}^*; \gamma | \mathbf{y}, \mathbf{z})$ es equivalente a minimizar $-\ell_p(\mathbf{y}^*, \mathbf{z}^*; \gamma | \mathbf{y}, \mathbf{z})$ y esto es equivalente a minimizar, la expresión:

$$\begin{aligned} &\frac{T+h-1}{2} \log \left[\frac{1}{T+h-1} \sum_{t=2}^{T+h} \left(w_t - \hat{\delta}_t(\mathbf{y}^*; \gamma) w_{t-1} \right)^2 \right] \\ &+ \frac{T+h-1}{2} \log \left[\frac{1}{T+h-1} \sum_{t=2}^{T+h} (v_t - \hat{\rho}(\mathbf{y}^*) v_{t-1})^2 \right], \end{aligned} \tag{5.6}$$

con respecto a $y^*, z^*; \gamma$. Bajo el supuesto de que el parámetro de umbral γ es fijo, debemos minimizar la función objetivo (5.6) con respecto a (y^*, z^*) para obtener los pronósticos correspondiente digamos \hat{y}^* , y \hat{z}^* . Los valores \hat{y}^* , \hat{z}^* son los pronósticos por verosimilitud predictiva para y^* y z^* respectivamente, bajo parámetro de umbral γ fijo.

Observe que cuando la variable umbral $Z_{t-1} \sim iidU(0, 1)$ obtenemos los pronósticos por verosimilitud predictiva para y^* bajo el supuesto de que los valores futuros de la variable umbral z^* , son conocidos minimizando la siguiente función objetivo, con respecto a y^* y γ ,

$$-\frac{T+h-1}{2} \log \left[\frac{1}{T+h-1} \sum_{t=2}^{T+h} \left(w_t - \hat{\delta}_t(\mathbf{y}^*; \gamma) w_{t-1} \right)^2 \right] \quad (5.7)$$

o equivalentemente la función objetivo

$$\frac{1}{T+h-1} \sum_{t=2}^{T+h} \left(w_t - \hat{\delta}_t(\mathbf{y}^*; \gamma) w_{t-1} \right)^2, \quad (5.8)$$

en virtud de que la función logaritmo es una función creciente.

El algoritmo implementado para la obtención de los pronósticos h -pasos adelante vía verosimilitud predictiva consiste en lo siguiente: contamos con 500 series simuladas del modelo TAR(2; 1, 1), dado en la ecuación (3.6), con un total de 255 observaciones. Las primeras 250 observaciones las utilizamos para obtener los estimadores de los parámetros bajo el modelo TAR(2; 1, 1) y bajo el modelo AR(1). Obtenemos los pronósticos desde uno hasta 5 pasos adelante, para cada modelo bajo verosimilitud predictiva, lo cual consiste en maximizar las correspondiente funciones objetivo bajo el supuesto de ruido blanco gaussiano. Calculamos a partir de dichos pronósticos y los últimos cinco valores en cada serie, los respectivos errores de pronósticos y a partir de estos calculamos las funciones de pérdida ECMP y pérdida de cuantiles o “check”. Tomamos el cociente de los valores de las funciones de pérdida en el modelo AR(1) para cada horizonte de pronóstico entre los respectivos valores bajo el modelo TAR(2; 1, 1). Si el valor del cociente de las pérdidas es mayor que uno, decimos que el desempeño de los pronósticos bajo el modelo TAR(2; 1, 1) es mejor con respecto al desempeño bajo el modelo AR(1). Este análisis se realiza para una gran cantidad de escenarios, descritos como función de los valores de los parámetros, en el Capítulo 7 dedicado al ejercicio de simulación. Los resultados indican que la alternativa de pronósticos por verosimilitud predictiva es la mejor en el sentido de que los valores de las funciones de pérdidas asociadas a dicha metodología son generalmente menores que las pérdidas asociadas al caso lineal. La ganancia se da también al comparar con las metodologías de Monte Carlo y regresión por cuantiles. La ventaja de verosimilitud predictiva se debe a que se pronostica conjuntamente las observaciones futuras del modelo TAR y de la variable umbral, de modo que la información relevante en torno a las características de no linealidad del modelo TAR son aprovechadas y consideradas por el método.

Capítulo 6

Regresión por cuantiles en modelos TAR: estimación y obtención de pronósticos

6.1 Introducción

Este capítulo aborda el problema del mal desempeño de los pronósticos en los modelos TAR(2; 1, 1) contra los desempeños de los pronósticos en el modelo AR(1) mediante la alternativa de regresión por cuantiles. El análisis del desempeño de los pronósticos se realiza por medio de la función de pérdida de cuantiles, la cual puede ser una función asimétrica. La asimetría de la función de pérdida depende directamente de la distribución considerada para el proceso de errores en el modelo bajo estudio. Específicamente, la función de pérdida de cuantiles está dada por una función lineal en ambos lados del origen, con la característica de que la pendiente de la recta en ambos lados depende del peso asociado a los residuales positivos y a los residuales negativos, respectivamente. Christoffersen y Diebold (1996), estudian el problema de pronósticos y selección de modelos bajo funciones de pérdida asimétricas con una aplicación particular a los modelos autorregresivos condicionalmente heteroscedásticos. Partimos del hecho de que ante una especificación incorrecta del modelo, el comportamiento de los residuales no corresponde a un proceso de ruido blanco, además de que la metodología estándar (Monte Carlo) no considera ni involucra adecuadamente las características de asimetrías presentes en los modelos TAR. Al escribir el modelo TAR(2; 1, 1) como un modelo lineal autorregresivo identificamos a este como un modelo autorregresivo de coeficientes aleatorios o equivalentemente como un modelo de volatilidad, lo cual implica que al ajustar un modelo lineal AR a los datos los residuales correspondientes hereden las características de no linealidad del verdadero proceso generador de datos, lo que se refleja en comportamientos alejados del supuesto de ruido blanco debido. Mostramos que al escribir el

modelo TAR como un modelo linealizado y variable umbral $Z_t \sim iidU(0, 1)$, el nuevo proceso de ruido r_t asociado al modelo linealizado satisface las condiciones de un proceso de volatilidad. Las características de volatilidad en los errores se presentan también en situaciones más generales donde la variable umbral presenta una estructura de dependencia, descrita por ejemplo mediante un proceso autorregresivo estacionario. El grado de volatilidad en el modelo está directamente relacionado con los valores de los parámetros en el modelo y con la dominancia de cada uno de los regímenes. Dicho de otra manera, la volatilidad es grande si $E(\delta_t^2)$ es cercano a uno, es decir, la volatilidad está directamente relacionada con la condición de estacionariedad dada por González y Gonzalo (1998).

Recientemente se han agregado una gran cantidad de trabajos de investigación que consideran modificaciones de los modelos tradicionales de raíces unitarias, incorporando los efectos de diversas causas posibles, ver por ejemplo Tsay (1997), González y Gonzalo (1998), Caner y Hansen (2001). La idea fundamental para el estudio de dichas modificaciones es la consideración de asimetrías sobre las diversas dinámicas económicas, por ejemplo Enders y Granger (1998) estudian diversas series económicas que despliegan comportamientos asimétricos.

En la literatura de modelos de volatilidad, Granger (2002), Engel y Manganelli (2004), proponen una aproximación para la estimación de cuantiles en modelos de valor al riesgo mediante el uso de regresión por cuantiles en modelos de series no lineales. La idea consiste en trabajar la distribución condicional completa vía la modelación de regresión por cuantiles. Granger (2002), justifica el estudio de medidas de volatilidad, tales como varianzas o retornos absolutos, mediante el estudio de los riesgos relacionados. Una forma útil y simple de estudiar una distribución condicional es considerando sus cuantiles, es decir, estudiando la distribución de los cuantiles en el tiempo. Por ejemplo, cuando la varianza varía en el tiempo, los cambios en un punto en el tiempo pueden ocurrir alrededor de una constante, de modo que el nivel medio de la varianza salta a un valor alto o bajo. Engel y Manganelli (2004) postulan distintos modelos para un conjunto de datos de interés e identifican cuáles de los modelos postulados son adecuados para describir al conjunto de datos dentro y fuera de la muestra, en el sentido de que no se rechace la hipótesis correspondiente. Las pruebas de bondad de ajuste consideradas e identificadas como pruebas de cuantil dinámico dentro y fuera de la muestra son pruebas estadísticas asintóticas con distribución asintótica Ji-cuadrada. La implementación de estas pruebas de cuantil dinámico dentro y fuera de la muestra, bajo la hipótesis de que el VPGD es un modelo TAR, identifica que dicho modelo describe adecuadamente al conjunto de observaciones analizadas mientras que el modelo lineal no necesariamente, lo cual nos da indicios de que dicha alternativa presentará un buen comportamiento al momento de abordar el problema de predicción.

El método de regresión por cuantiles no sólo ofrece una alternativa robusta de estimación para la tendencia central de la respuesta, podemos también explorar la distribución condicional de la respuesta de manera completa. En los trabajos clásicos de regresión, la media

condicional $E(Y|X = x)$ se estima generalmente mediante mínimos cuadrados. A menos que hagamos supuestos distribucionales fuertes en la distribución condicional de la respuesta, los métodos de mínimos cuadrados no proporcionan ninguna información más allá de la media condicional. Regresión por cuantiles generaliza la idea de estimar la mediana por regresión u otras funciones de los cuantiles condicionales minimizando la suma de residuales ponderada. Una extensión de los conceptos de regresión por cuantiles en modelos de series de tiempo autoregresivos se debe al trabajo de Koul y Saleh (1995), mientras que Koul (1996) y Mukherjee (1999) generalizan estos resultados para el caso de series de tiempo autorregresivas no lineales.

6.2 Regresión por cuantiles

Koenker y Bassett (1978) desarrollaron la teoría y justificación de regresión por cuantiles como una generalización al problema clásico de estimación por mínimos cuadrados para la media condicional. Se sabe que dada una variable aleatoria de valores reales, Y , ésta puede ser caracterizada por su función de distribución, $F(y) = \Pr(Y \leq y)$. Asimismo para cualquier $0 < \theta < 1$, denotamos a $Q_Y(\theta) = \inf\{y : F(y) \geq \theta\}$ como el θ -cuantil de Y , por ejemplo el cuantil $Q_Y(1/2)$ corresponde al valor de la mediana y al igual que la función de distribución, la función cuantil $Q_Y(\theta)$ caracteriza de forma completa a la variable de interés Y .

Los cuantiles se caracterizan por la solución final de un problema de optimización, donde para cualquier $0 < \theta < 1$ se define la función $\rho_\theta(u) = u[\theta - I(u < 0)] = \theta u^+ + (\theta - 1)u^-$, donde $u^+ = u[I(u > 0)]$ y $u^- = u[I(u < 0)]$ corresponden a las partes positivas y negativas de $u \in \mathbb{R}$, respectivamente. Minimizar $E[\rho_\theta(Y - \xi)]$ con respecto a ξ , arroja soluciones, $\hat{\xi}(\theta)$, donde la solución más pequeña es precisamente el θ -cuantil de Y , $Q_Y(\theta)$.

Nuestro interés versa en torno a la función cuantil condicional, definida como: el θ -ésimo cuantil,

$$Q_{Y|X}(\theta|x) = \inf\{y : F_{Y|X}(y|x) \geq \theta\},$$

bajo el supuesto de que la función de distribución condicional $F_{Y|X}(y|x)$ es estrictamente creciente, con función de densidad $f_{Y|X}(y|x)$, podemos caracterizar a la función cuantil por medio de la siguiente igualdad:

$$Q_{Y|X}(\theta|x) = \arg \min_a E[\rho_\theta(Y - a) | X = x], \quad (6.1)$$

donde la función de pérdida $\rho_\theta(u) = u(\theta - I(u < 0))$ es conocida en la literatura de regresión por cuantiles como la función de pérdida “check” o “tick”. La Figura 14 presenta el comportamiento de la función cuantil, $\rho_\theta(u)$, para tres distintos valores de θ , el caso simétrico corresponde a $\theta = 0.5$ identificado en la gráfica con la línea continua.

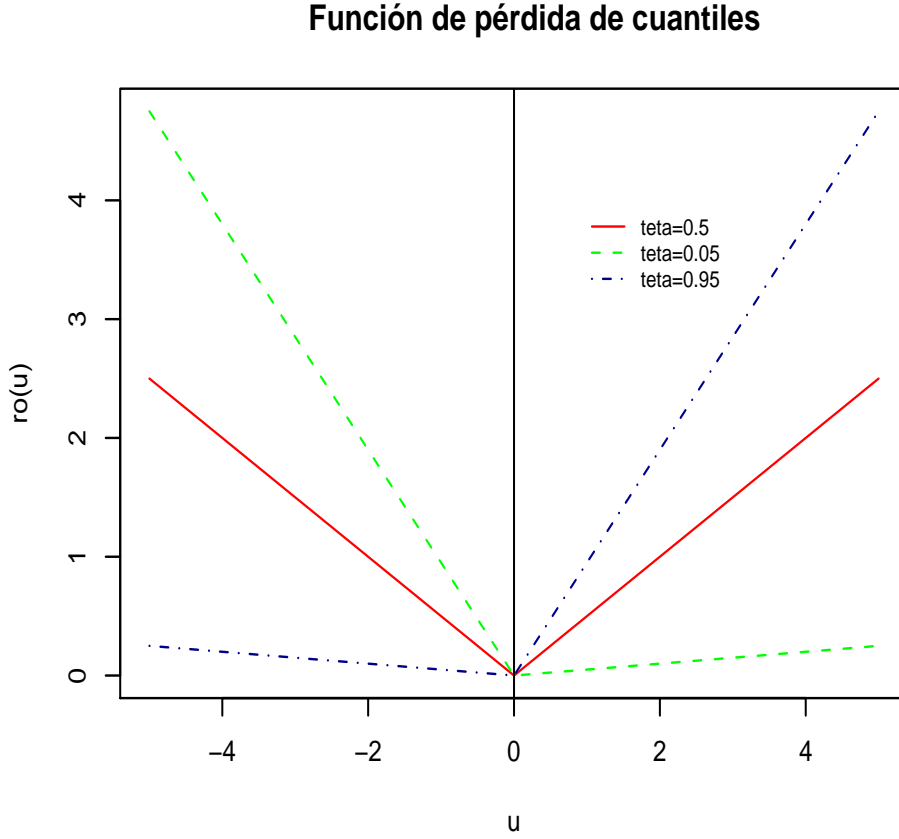


Figura 14. Comportamiento de la función de pérdida de cuantiles para distintos valores de θ .

Para verificar la ecuación (6.1), observemos que:

$$E[\rho_{\theta}(Y - a) | X = x] = (\theta - 1) \int_{-\infty}^a (y - a) f_{Y|X}(y|x) dy + \theta \int_a^{\infty} (y - a) f_{Y|X}(y|x) dy.$$

La condición de primer orden para la igualdad anterior está dada por:

$$-(\theta - 1) \int_{-\infty}^a f_{Y|X}(y|x) dy - \theta \int_a^{\infty} f_{Y|X}(y|x) dy = 0,$$

o equivalentemente por la siguiente igualdad:

$$F(a|x) - \theta = 0,$$

con solución $a^* = F_{Y|X}^{-1}(\theta|x)$, la cual corresponde a la función cuantil. Notemos que, si la función cuantil está dada por $Q_{Y|X}(\theta|x) = x'\beta(\theta)$, la ecuación (6.1) implica que:

$$\beta(\theta) = \arg \min_b E[\rho_\theta(Y - X'b)]. \quad (6.2)$$

Por ejemplo, para $\theta = 1/2$ (mediana) tenemos que la función en (6.2) corresponde a la función valor absoluto, es decir, estamos minimizando una suma simétrica ponderada de errores absolutos con pesos iguales a 0.5 para estimar la función cuantil condicional. Mientras que para valores de $\theta \neq 0.5$ estamos minimizando sumas ponderadas asimétricas de errores absolutos, donde los pesos son funciones de los cuantiles de interés, ver Koenker y Hallock (2001). Ver Figura 14.

Koul y Saleh (1995) extienden los resultados de regresión por cuantiles Koenker y Bassett (1978) al caso de modelos autorregresivos. Posteriormente, Koul (1996) y Mukherjee (1999), generalizan los conceptos de regresión por cuantiles y autorregresión al caso de series de tiempo no lineal. Koul (1996) discute las propiedades asintóticas de algunos estimadores en series de tiempo no lineal, y prueba la normalidad asintótica de ciertas clases de estimadores (M y R estimadores) minimizando la distancia entre los estimadores y los parámetros de interés. Mukherjee (1999), generaliza los conceptos de regresión por cuantiles y autorregresión para modelos de series de tiempo no lineal. La linealización asintótica de los cuantiles no lineales se utiliza, por ejemplo, para obtener la distribución límite de estimadores de mínima desviación absoluta, demostrándose que estos estimadores son más eficientes que los estimadores de mínimos cuadrados condicionales. Los resultados de Mukherjee (1999), son aplicables a modelos de regresión lineales y no lineales; así como también a modelos autorregresivos del tipo lineal y no lineal, incluyendo los modelos TAR con parámetro de umbral conocido.

Reconociendo que el método de regresión por cuantiles es una alternativa para el estudio de dinámicas asimétricas en series económicas, y que los modelos TAR presentan características de asimetrías, consideramos a la metodología de regresión por cuantiles como una metodología viable en el estudio de los modelos TAR. El modelo de cuantiles para el caso TAR, considera adecuadamente la asimetría natural del modelo, así como otras características de asimetrías resultantes al considerar por ejemplo que la función de distribución para los errores en el modelo es asimétrica. En nuestro caso particular partimos de que la distribución asociada a los errores en el modelo TAR es una función simétrica, dado que suponemos que los errores son una secuencia de variables aleatorias gaussianas y de aquí que el valor del peso $0 < \theta < 1$ asociado a la función de pérdida en nuestro caso es $\theta = 0.5$. A partir de la representación de cuantiles obtenemos estimaciones de los parámetros en el modelo TAR(2; 1, 1), así como pronósticos desde uno hasta 5-pasos hacia adelante, bajo la consideración de que el parámetro de umbral γ es conocido. Sin embargo, cabe aclarar que una posibilidad para estimar el parámetro de umbral γ , es por medio de una búsqueda directa en un conjunto compacto, de tal manera que nos quedamos con aquel valor del umbral que minimice la

función cuantil, e identificamos los valores estimados correspondientes para cada uno de los coeficientes autorregresivos en el modelo. Un estimador de la varianza del error está dado por los promedios de los residuales cuadrados asociados a la serie en estudio considerando los estimadores del resto de los parámetros en el modelo, bajo el valor del umbral identificado como parte del proceso de minimización. Esta sugerencia de estimación del parámetro de umbral es equivalente al proceso de estimación de mínimos cuadrados secuencial sugerido en Chan *et al* (1993) y Hansen (1997).

6.3 El modelo de cuantiles para series de tiempo no lineal

Considere el modelo de series de tiempo no lineal de orden p , con valores iniciales Y_{1-p}, \dots, Y_0 , que se rige bajo la siguiente igualdad y estudiado a detalle en Mukherjee (1999),

$$Y_t = f_t(\beta^0, X_{t-1}, Z_t) + \varepsilon_t, \quad 1 \leq t \leq T, \quad (6.3)$$

donde $\beta^{0'} \equiv (\beta_0, \tilde{\beta}')$ $\in \Omega$ es el vector de parámetros verdaderos pero desconocidos y que necesitan ser estimados, Ω es un subconjunto abierto de \mathbb{R}^{1+m} de la forma $\Omega = \mathbb{R}^1 \times \tilde{\Omega}$ y $f_t : \Omega \times \mathbb{R}^p \times \mathbb{R}^q \rightarrow \mathbb{R}^1$ es una función conocida dada por la siguiente ecuación: $f_t(\beta^0, X_{t-1}, Z_t) = \beta_0 + \tilde{f}_t(\tilde{\beta}', X_{t-1}, Z_t)$. Definamos $X_{t-1} = (Y_{t-1}, \dots, Y_{t-p})$, $1 \leq t \leq T$, y $\{Z_t, 1 \leq t \leq T\}$ un arreglo de vectores aleatorios exógenos de dimensión q observables, que representan tendencias o variables de regresión. La secuencia de innovaciones $\{\varepsilon_t, 1 \leq t \leq T\}$ son variables aleatorias independientes e idénticamente distribuidas con distribución F . Entonces el estimador de β asociado al θ -ésimo cuantil del modelo (6.3), con $0 < \theta < 1$, es cualquier vector $\hat{\beta}(\theta)$ que minimice la función $\sum_{t=1}^T \rho_\theta[Y_t - f_t(\beta, X_{t-1}, Z_t)]$, con respecto a $\beta \in \mathbb{R}^{1+m}$ y $\rho_\theta\{u\} \equiv -\theta u [I(u > 0)] - (1 - \theta) u [I(u < 0)] = [\theta - I(u)] u$, es decir, el θ -ésimo cuantil no lineal es aquel que resuelve la siguiente función cuantil:

$$\begin{aligned} & \min_{\beta} \sum_{t=1}^T \{ \theta [Y_t - f_t(\beta, X_{t-1}, Z_t)] I(Y_t > f_t(\beta, X_{t-1}, Z_t)) \\ & \quad - (1 - \theta) [Y_t - f_t(\beta, X_{t-1}, Z_t)] I(Y_t < f_t(\beta, X_{t-1}, Z_t)) \} \\ & = \min_{\beta} \sum_{t=1}^T [\theta - I(Y_t < f_t(\beta, X_{t-1}, Z_t))] [Y_t - f_t(\beta, X_{t-1}, Z_t)]. \end{aligned} \quad (6.4)$$

Observemos que para el caso de modelos autorregresivos no lineales, como por ejemplo en los modelos SETAR, $f_t(\beta, X_{t-1}, Z_t) = f_t(\beta, X_{t-1})$. Mientras que para los modelos TAR la variable Z_t se identifica directamente con la variable umbral que determina en conjunto con el valor del umbral el comportamiento de Y_t en cada uno de los regímenes.

6.3.1 Regresión por cuantiles para el modelo TAR(2; 1, 1)

Una condición suficiente para la existencia de $\hat{\beta}$ estimador consistente en la función objetivo (6.4), es que la función $f_t(\beta, X_{t-1}, Z_t)$ sea una función lineal en β y no lineal en el resto de los argumentos, además de suponer que la serie de tiempo no lineal bajo estudio es estacionaria y ergódica. Tong (1990) muestra que esta condición se satisface para una gran variedad de modelos de series de tiempo no lineal, como son los modelos bilineales, modelos exponenciales autorregresivos (EXPAR) y modelos del tipo TAR. Koul (1996) ejemplifica las propiedades asintóticas de algunos estimadores para el modelo de series de tiempo no lineal, como por ejemplo los estimadores de mínima desviación absoluta. En particular presenta dos aplicaciones de modelos no lineales, el modelo SETAR con parámetro de umbral y rezago conocido y el modelo EXPAR.

Para el modelo TAR(2; 1, 1) dado en la ecuación (3.6), tenemos que el modelo de cuantiles está dado como,

$$\begin{aligned} Y_t &= \delta_t Y_{t-1} + \varepsilon_{t\theta} \\ &= f_t(\beta, Y_{t-1}, Z_{t-1}) + \varepsilon_{t\theta}, \quad t = 1, \dots, T, \end{aligned} \quad (6.5)$$

con $f_t(\beta, Y_{t-1}, Z_{t-1}) = \phi_1 I(Z_{t-1} \leq \gamma) Y_{t-1} + \phi_2 I(Z_{t-1} > \gamma) Y_{t-1}$ y $\beta = (\phi_1, \phi_2)'$ el vector de parámetros en el modelo, considerando el parámetro de umbral γ conocido, con $\varepsilon_{t\theta}$ una secuencia de variables aleatorias independientes e idénticamente distribuidos con distribución normal $N(0, \sigma^2)$.

El vector de parámetros β en el modelo (6.5) se estima considerando dicho modelo en la función cuantil dada en (6.4) y minimizando la función resultante con respecto a β , para $\theta = 0.5$. Por su parte, estimamos la varianza asociada a la secuencia de errores obteniendo el promedio de los residuales al cuadrado en el modelo, calculados a partir de la muestra observada de tamaño T para Y_t , con $t = 1, 2, \dots, T$ y de los coeficientes autorregresivos estimados vía el procedimiento de minimización de la θ -ésima función cuantil. Como comentamos en la introducción de este capítulo, si el parámetro de umbral es desconocido podemos estimar dicho parámetro por medio de una búsqueda directa del umbral en un intervalo compacto de la forma $\Gamma = [-\gamma_0, \gamma_1]$. La búsqueda de γ que minimiza (6.4), en el espacio de valores de la variable umbral Z_{t-1} puede realizarse considerando el valor del cuantil 0.15 y cuantil 0.85 como los límites inferior y superior del intervalo compacto Γ , a fin de garantizar un mínimo de observaciones para la estimación de los coeficiente autorregresivos correspondientes a cada uno de los regímenes.

Una vez encontrado β que minimiza la función (6.4) para el modelo TAR(2; 1, 1) con parámetro de umbral conocido sustituimos dicho valor como conocido en la función cuantil a minimizar y encontramos los h -valores hacia adelante más plausibles para el modelo TAR(2; 1, 1) en el sentido de que dichos valores minimizan la función (6.4) con respecto a los valores desconocidos de Y_{T+1}, \dots, Y_{T+h} , dados los valores conocidos para $Z_{T+1}, \dots, Z_{T+h-1}$

o equivalentemente dado que conocemos los valores que toma la variable indicadora. En general, conocemos solamente los valores que toma la variable umbral hasta el tiempo T e igual que el caso de la obtención de pronósticos por el método iterativo, pronosticamos Z_{t-1} considerando su modelo asociado o bien $I(Z_{t-1} \leq \gamma)$ mediante la teoría de pronósticos de series de tiempo binarias. Finalmente, tomamos los valores de Y_{T+1}, \dots, Y_{T+h} que minimizan (6.4) como los pronósticos de Y_{T+1}, \dots, Y_{T+h} , respectivamente.

Los resultados del estudio de simulación presentados en el Capítulo 7 muestran que los desempeños de los pronósticos bajo esta alternativa presentan en general mejor comportamiento que los pronósticos de su contraparte lineal. La ganancia es mayor cuando consideramos los pronósticos de la variable indicadora como valores conocidos y a partir de estos minimizamos la función cuantil para obtener los pronósticos de Y_{T+1}, \dots, Y_{T+h} . La cuantificación de los desempeños de dichos pronósticos se realiza mediante la función de pérdida de cuantiles, así como por medio de los ECMP para cada uno de los horizontes de pronóstico.

Con la finalidad de una mejor comprensión de la alternativa de regresión por cuantiles en los modelos TAR presentamos un análisis del método de regresión por cuantiles para el caso del modelo TAR más simple dado en la ecuación (2.6) del Capítulo 2. Sin embargo los resultados para el modelo TAR(2; 1, 1) dado en la ecuación (6.5) son equivalentes al caso presentado, con la diferencia de que bajo este escenario la característica de simetría en los residuales bajo el modelo lineal AR se acentúa pero con un comportamiento de colas más pesadas que la normal. Esperamos que al aplicar la metodología de regresión por cuantiles a los modelos TAR, independientemente del porcentaje de observaciones en cada uno de los regímenes, presente comportamientos simétricos en los residuales en virtud de que suponemos errores gaussianos. A fin de corroborar que la asimetría natural del modelo TAR es manejada adecuadamente por el método de regresión por cuantiles, el escenario simulado considera que los coeficientes autorregresivos son cercanos al círculo unitario pero de signos contrarios. Posteriormente para cada conjunto de datos simulados ajustamos, mediante el método de regresión por cuantiles con $\theta = 0.5$, un modelo TAR con parámetro de umbral conocido, así como un modelo AR. La robustez del método garantiza que los valores estimados de los parámetros en el modelo TAR serán cercanos a los valores verdaderos, tal y como se muestra en la Tabla S1. Obtenemos los valores de los residuales bajo cada modelo estimado para contrastar los comportamientos de los mismos. Para el modelo AR(1) debemos observar que residuales presentan comportamientos asimétricos ya que la asimetría del VPGD se hereda al proceso de error.

6.3.2 Regresión por cuantiles para el modelo TAR más simple

Consideremos el modelo TAR más simple dado en la ecuación (2.6) y reescrito a continuación

$$Y_t = \phi_1 I(Z_{t-1} \leq \gamma) + \phi_2 I(Z_{t-1} > \gamma) + \varepsilon_t,$$

con $\varepsilon_t \sim iidN(0, \sigma^2)$ y variable umbral Z_{t-1} descrita por un proceso AR(1) estacionario.

Simulamos 500 observaciones a partir del modelo anterior con $\phi_1 = 0.9$, $\phi_2 = -0.9$ y $\sigma^2 = 0.25$, para $p = \Pr(I(Z_{t-1} \leq \gamma))$ igual a 0.1, 0.25, y 0.5 respectivamente, es decir, el porcentaje de observaciones en el régimen uno corresponde al 10%, 25%, y 50% respectivamente. Un número mayor de posibilidades en cuanto a los porcentajes de observaciones en los regímenes fueron exploradas, pero las conclusiones son equivalentes. La Figura S1 presenta las realizaciones del proceso bajo distintos porcentajes de observaciones en el régimen uno, resaltando la asimetría del escenario considerado. La línea continua identifica a las observaciones en el régimen uno mientras que la línea punteada a las observaciones en el régimen dos.

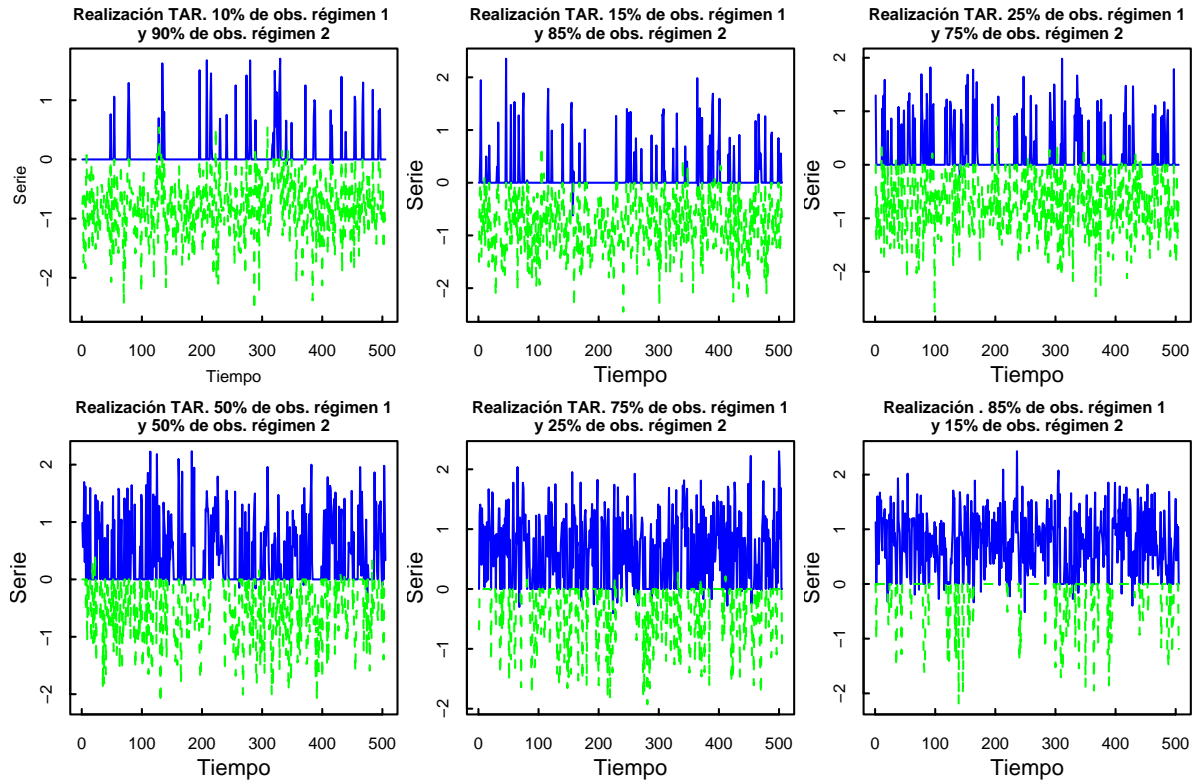


Figura S1. Realizaciones del modelo TAR más simple, $\phi_1 = 0.9$ y $\phi_2 = -0.9$.

Para cada conjunto de datos ajustamos mediante el método de regresión por cuantiles, un modelo lineal dado por $Y_t = \phi + \varepsilon_t$, así como el modelo TAR con $\theta = 0.5$. Complementariamente, en el caso lineal estimamos también el parámetro ϕ por medio de MV. Los resultados de dichas estimaciones se presentan en la Tabla S1, donde: $\hat{\phi}_1, \hat{\phi}_2$ y $\hat{\sigma}_{TAR}^2$ se refieren a los valores estimados de los parámetros bajo el modelo TAR por medio del método de regresión por cuantiles. $\hat{\phi}_{linealCR}$ y $\hat{\sigma}_{linealCR}^2$ corresponden a los valores estimados de los parámetros

bajo el modelo lineal por medio del método de regresión por cuantiles. Finalmente $\hat{\phi}_{linealMV}$ y $\hat{\sigma}_{linealMV}^2$ son las estimaciones de los parámetros bajo el modelo lineal por medio del método de máxima verosimilitud.

Tabla S1. Estimaciones bajo el modelo TAR simplificado, y modelo lineal simplificado respectivamente.

p	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\sigma}_{TAR}^2$	$\hat{\phi}_{linealCR}$	$\hat{\sigma}_{linealCR}^2$	$\hat{\phi}_{linealMV}$	$\hat{\sigma}_{linealMV}^2$
0.10	0.853	-0.870	0.24	-0.821	0.55	-0.708	0.74
0.15	0.899	-0.91	0.26	-0.799	0.71	-0.641	0.85
0.25	0.979	-0.915	0.27	-0.704	0.83	-0.513	0.91
0.50	0.905	-0.850	0.25	-0.164	1.06	-0.025	1.03
0.75	0.880	-0.903	0.26	0.555	0.93	0.323	0.96
0.85	0.902	-0.916	0.24	0.765	0.70	0.580	0.84
0.90	0.887	-0.893	0.23	0.788	0.56	0.680	0.75

La Figura S2 muestra el comportamiento de los residuales bajo el modelo TAR, considerando distintos porcentajes de observaciones en cada uno de los regímenes, los cuales se especifican en cada gráfica. El primer histograma en cada renglón de la Figura S2 corresponde al total de residuales bajo el modelo TAR, mientras que los últimos dos histogramas de cada línea corresponden a los residuales bajo el régimen uno y dos respectivamente. Aún cuando los comportamientos de los residuales para cada uno de los regímenes presentan ligeras asimetrías, y que dicha asimetría es mayor bajo el régimen minoritario, notamos que al considerar dichos residuales en forma conjunta, el histograma obtenido es simétrico. La simetría se presenta independientemente del porcentaje de observaciones en cada uno de los regímenes y de la asimetría natural del proceso simulado.

Contrario a los comportamientos simétricos de los residuales bajo el modelo TAR, tenemos que al ajustar un modelo lineal a cada conjunto de datos simulados e independientemente del procedimiento de estimación de los parámetros los comportamientos de los residuales son asimétricos. Para corroborar dicha observación presentamos dichos residuales en la Figuras S3. Los comportamientos asimétricos de dichos residuales dependen de los porcentajes de observaciones en cada uno de los regímenes y lo pesado de las colas a la derecha o a la izquierda depende de si el régimen dominante es el régimen uno o dos respectivamente. Esto es de esperarse ya que de la Tabla S1, notamos que la estimación de los parámetros correspondientes bajo el modelo lineal se debe principalmente a las observaciones bajo el régimen dominante. De esta forma tenemos que la asimetría observada en los residuales asociados al modelo lineal se debe a la asimetría natural del modelo de regímenes bajo el cual simulamos los conjuntos de datos.

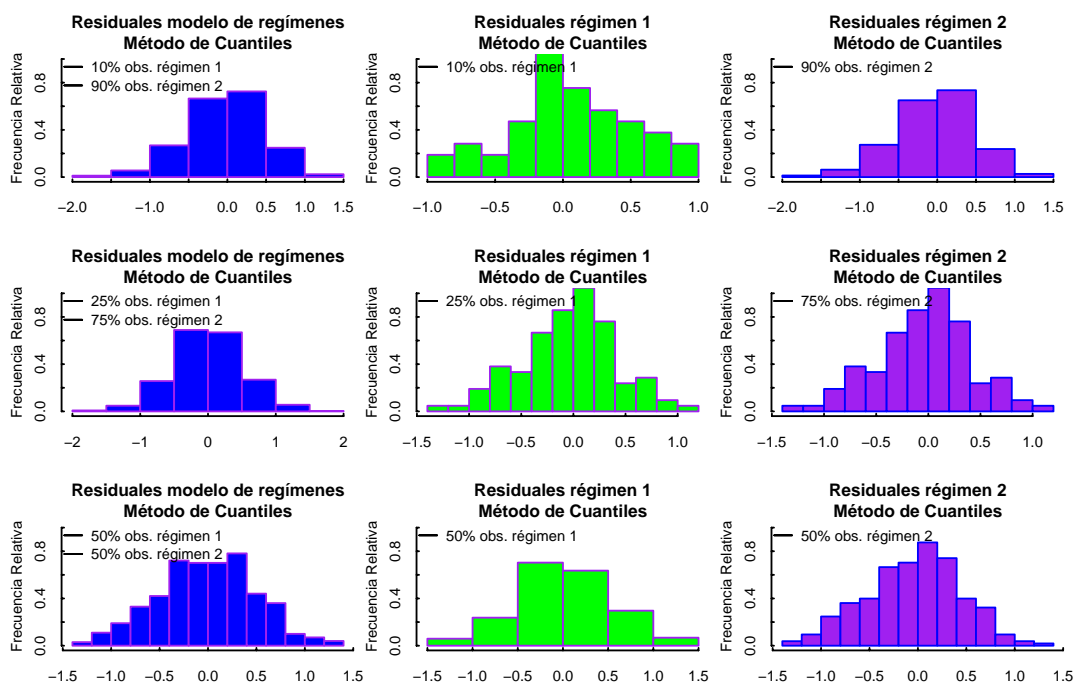


Figura S2. Residuales bajo el modelo TAR simplificado.

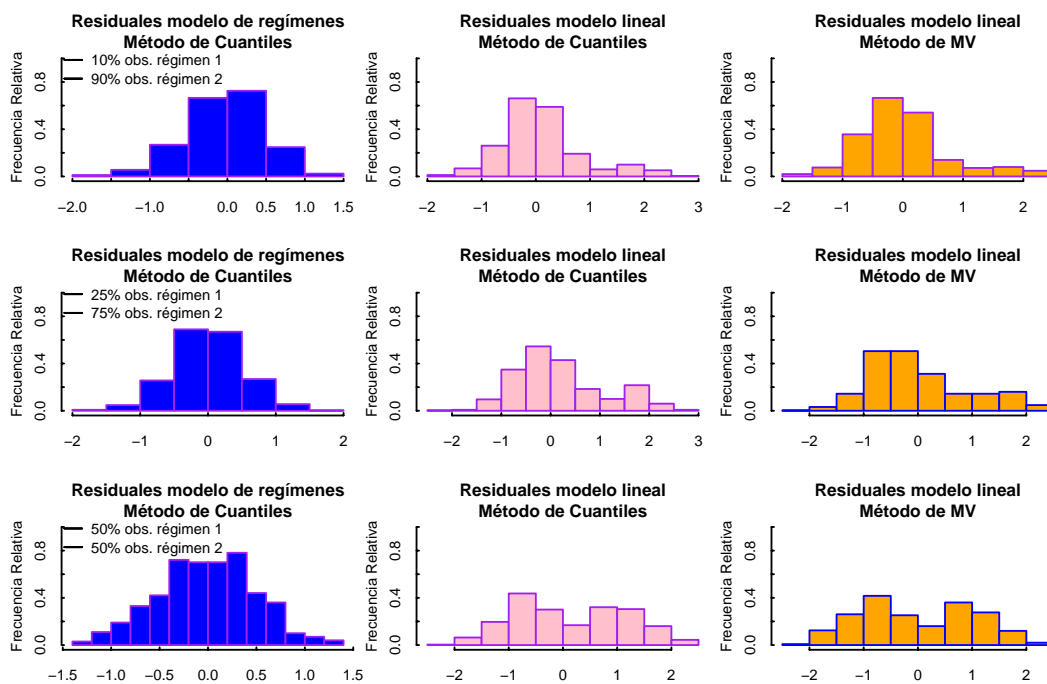


Figura S3. Residuales bajo los modelos lineal y TAR simplificados.

La Figura S4 complementa las observaciones hechas a partir de la Figura S3, ya que presenta las correspondientes gráficas de cuantil cuantil para cada uno de los histogramas dados en la Figura S3. De donde corroboramos que los residuales bajo el modelo lineal presentan comportamientos asimétricos y de colas pesadas, lo cual los aleja del supuesto de normalidad en los errores, contrario a los residuales bajo el modelo correcto.

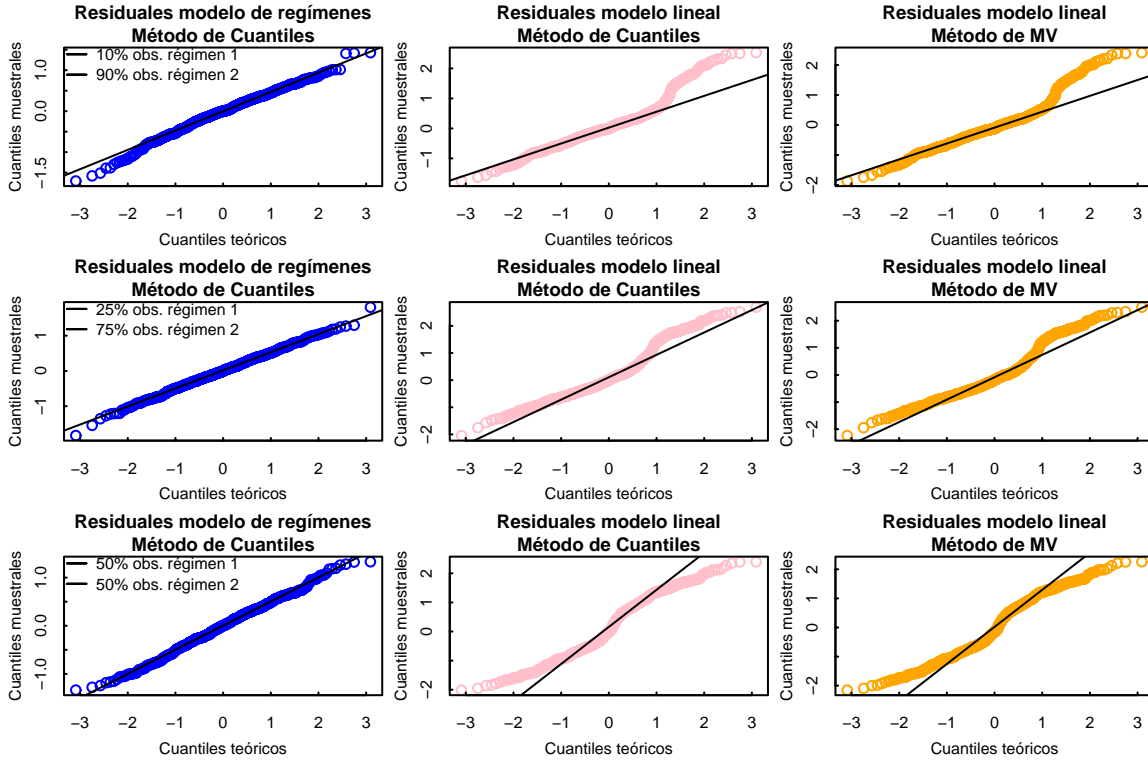


Figura S4. Gráficas de probabilidad para los residuales bajo cada modelo

Realizamos un análisis similar para el caso donde el proceso generador de datos es el modelo $TAR(2; 1, 1)$ de la ecuación (6.5). A diferencia del caso simplificado, observamos que la asimetría de los residuales suponiendo un modelo lineal $AR(1)$ se acentúa, sin embargo las colas de los histogramas se vuelven más pesadas. En las Secciones 6.3.3. y 6.3.4. describiremos las propiedades asintóticas de los estimadores mediante regresión por cuantiles, y se establecerán pruebas de hipótesis sobre el VPGD. Las pruebas de bondad de ajuste establecidas se definen en términos de la función hit dada en la ecuación (6.10) la cual contabiliza el número de aciertos bajo el modelo correcto. Bajo el VPGD los aciertos son no correlacionados y no se comete errores de medición en las pruebas de hipótesis implementadas. La implementación de dichas pruebas dentro y fuera de la muestra favorece al modelo TAR como el VPGD, como se verá en la Sección 6.3.4.

Una ventaja adicional del método de regresión por cuantiles es que no requiere del supuesto de distribuciones simétricas con colas no pesadas, como la distribución normal.

Sin embargo el cálculo de la función de pérdida de cuantiles para pronósticos h -pasos adelante en forma teórica no es fácil aún en el caso del modelo TAR simplificado y lineal simplificado tal y como se ilustra a continuación.

Función de pérdida de cuantiles para el modelo TAR más simple

El objetivo de esta sección es obtener expresiones teóricas para la función de pérdida de cuantiles esperada para los errores de pronósticos dos pasos adelante, similar al caso explorado para la función de pérdida error cuadrático medio. Recordemos que la función de pérdida de cuantiles está dada por la siguiente ecuación:

$$\rho_{\theta}(u) = u[\theta - I(u < 0)] = \theta u^+ + (\theta - 1)u^-, \quad (6.6)$$

donde $u^+ = u[I(u > 0)]$ y $u^- = u[I(u < 0)]$, y consideremos el modelo TAR simplificado dado en la (2.6)

$$Y_t = \phi_1 I(Z_{t-1} \leq \gamma) + \phi_2 I(Z_{t-1} > \gamma) + \varepsilon_t,$$

con $\varepsilon_t \sim iidN(0, \sigma^2)$ y Z_{t-1} , una serie de tiempo estacionaria AR(1). Recordemos también que el pronóstico 2-pasos hacia adelante, si ϕ_1 y ϕ_2 son conocidos, está dado por: $\hat{Y}_{T+2} = \phi_1 \hat{I}(Z_{T+1} \leq \gamma) + \phi_2 \hat{I}(Z_{T+1} > \gamma)$. Y que definida la función de distribución conjunta de las variables $\hat{I}(Z_{T+1} \leq \gamma)$ y $I(Z_{T+1} \leq \gamma)$, tal y como se realizó en el Capítulo 1,

$$p_{ij} = \Pr\left(\hat{I}(Z_{T+1} \leq \gamma) = i, I(Z_{T+1} \leq \gamma) = j\right)$$

para $i, j = 0, 1$ y $\sum \sum p_{ij} = 1$, podemos obtener expresiones para la función de pérdida de los pronósticos dos pasos adelante.

Denotemos la esperanza de la función de pérdida de cuantiles para el modelo TAR más simple dos pasos hacia adelante, por RC_2^{TAR} de tal forma que para $u = Y_{T+2} - \hat{Y}_{T+2}$, tenemos que:

$$RC_2^{TAR} = \theta E[uI(u > 0)] + (\theta - 1)E[uI(u < 0)],$$

donde,

$$\begin{aligned}
& E[uI(u > 0)] \\
&= E\left[\left(Y_{T+2} - \hat{Y}_{T+2}\right) I\left(Y_{T+2} - \hat{Y}_{T+2} > 0\right)\right] \\
&= E\left[\left(Y_{T+2} - \hat{Y}_{T+2}\right) I\left(Y_{T+2} - \hat{Y}_{T+2} > 0\right) \mid \hat{I}(Z_{T+1} \leq \gamma) = 1, I(Z_{T+1} \leq \gamma) = 1\right] p_{11} \\
&\quad + E\left[\left(Y_{T+2} - \hat{Y}_{T+2}\right) I\left(Y_{T+2} - \hat{Y}_{T+2} > 0\right) \mid \hat{I}(Z_{T+1} \leq \gamma) = 1, I(Z_{T+1} \leq \gamma) = 0\right] p_{10} \\
&\quad + E\left[\left(Y_{T+2} - \hat{Y}_{T+2}\right) I\left(Y_{T+2} - \hat{Y}_{T+2} > 0\right) \mid \hat{I}(Z_{T+1} \leq \gamma) = 0, I(Z_{T+1} \leq \gamma) = 1\right] p_{01} \\
&\quad + E\left[\left(Y_{T+2} - \hat{Y}_{T+2}\right) I\left(Y_{T+2} - \hat{Y}_{T+2} > 0\right) \mid \hat{I}(Z_{T+1} \leq \gamma) = 0, I(Z_{T+1} \leq \gamma) = 0\right] p_{00} \\
&= E[(\phi_1 + \varepsilon_{T+2} - \phi_1) I(\phi_1 + \varepsilon_{T+2} - \phi_1 > 0)] p_{11} \\
&\quad + E[(\phi_2 + \varepsilon_{T+2} - \phi_1) I(\phi_2 + \varepsilon_{T+2} - \phi_1 > 0)] p_{10} \\
&\quad + E[(\phi_1 + \varepsilon_{T+2} - \phi_2) I(\phi_1 + \varepsilon_{T+2} - \phi_2 > 0)] p_{01} \\
&\quad + E[(\phi_2 + \varepsilon_{T+2} - \phi_2) I(\phi_2 + \varepsilon_{T+2} - \phi_2 > 0)] p_{00} \\
&= E[\varepsilon_{T+2} I(\varepsilon_{T+2} > 0)] (p_{11} + p_{00}) + E[(\varepsilon_{T+2} - \mu) I(\varepsilon_{T+2} - \mu > 0)] p_{10} \\
&\quad + E[(\varepsilon_{T+2} + \mu) I(\varepsilon_{T+2} + \mu > 0)] p_{01} \\
&= E[u_1 I(u_1 > 0)] (p_{11} + p_{00}) + E[u_2 I(u_2 > 0)] p_{10} + E[u_3 I(u_3 > 0)] p_{01},
\end{aligned}$$

donde $u_1 = \varepsilon_{T+2} \sim iidN(0, \sigma^2)$; $u_2 = \varepsilon_{T+2} - \mu \sim iidN(-\mu, \sigma^2)$ y $u_3 = \varepsilon_{T+2} + \mu \sim iidN(\mu, \sigma^2)$. Paralelamente, notemos que si $x \sim iidN(\mu, \sigma^2)$, la esperanza truncada de x en (a, b) está dada por:

$$\begin{aligned}
E[xI(a < x < b)] &= \int_a^b x \phi(x; \mu, \sigma^2) dx \\
&= -\sigma^2 [\phi(b; \mu, \sigma^2) - \phi(a; \mu, \sigma^2)] + \mu [\Phi(b; \mu, \sigma^2) - \Phi(a; \mu, \sigma^2)].
\end{aligned}$$

Para mostrar la igualdad anterior notemos que a partir de la función de densidad para x denotada por $\phi(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$ y derivando con respecto a x ,

$$\begin{aligned}
\frac{d}{dx} \phi(x; \mu, \sigma^2) &= -\frac{(x - \mu)}{\sigma^2} \phi(x; \mu, \sigma^2) \\
&= \frac{\mu}{\sigma^2} \phi(x; \mu, \sigma^2) - \frac{x}{\sigma^2} \phi(x; \mu, \sigma^2),
\end{aligned}$$

entonces, integrando de a a b en ambos lados de la igualdad anterior, obtenemos lo siguiente:

$$\begin{aligned}
\phi(b; \mu, \sigma^2) - \phi(a; \mu, \sigma^2) &= \int_a^b \left[\frac{\mu}{\sigma^2} \phi(x; \mu, \sigma^2) - \frac{x}{\sigma^2} \phi(x; \mu, \sigma^2) \right] dx \\
&= \frac{\mu}{\sigma^2} \int_a^b \phi(x; \mu, \sigma^2) dx - \frac{1}{\sigma^2} \int_a^b x \phi(x; \mu, \sigma^2) dx.
\end{aligned}$$

Luego despejando,

$$\begin{aligned}\int_a^b x\phi(x; \mu, \sigma^2) dx &= -\sigma^2 [\phi(b; \mu, \sigma^2) - \phi(a; \mu, \sigma^2)] + \mu \int_a^b \phi(x; \mu, \sigma^2) dx \\ &= -\sigma^2 [\phi(b; \mu, \sigma^2) - \phi(a; \mu, \sigma^2)] + \mu [\Phi(b; \mu, \sigma^2) - \Phi(a; \mu, \sigma^2)].\end{aligned}$$

Asimismo tenemos las siguientes expresiones:

$$\begin{aligned}\int_a^\infty x\phi(x; \mu, \sigma^2) dx &= \sigma^2\phi(a; \mu, \sigma^2) + \mu [1 - \Phi(a; \mu, \sigma^2)] \\ &= \mu + \sigma^2\phi(a; \mu, \sigma^2) - \mu\Phi(a; \mu, \sigma^2),\end{aligned}$$

$$\int_{-\infty}^b x\phi(x; \mu, \sigma^2) dx = -\sigma^2\phi(b; \mu, \sigma^2) + \mu\Phi(b; \mu, \sigma^2).$$

Si $a = 0$,

$$\begin{aligned}E[xI(x > 0)] &= \int_0^\infty x\phi(x; \mu, \sigma^2) dx \\ &= \sigma^2\phi(0; \mu, \sigma^2) + \mu [1 - \Phi(0; \mu, \sigma^2)] \\ &= \frac{\sigma}{\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}\mu^2} + \mu \left[1 - \Phi\left(\frac{-\mu}{\sigma^2}; 0, 1\right) \right]\end{aligned}\tag{6.7}$$

Si $b = 0$

$$\begin{aligned}E[xI(x < 0)] &= \int_{-\infty}^0 x\phi(x; \mu, \sigma^2) dx \\ &= -\sigma^2\phi(0; \mu, \sigma^2) + \mu\Phi(0; \mu, \sigma^2) \\ &= -\frac{\sigma}{\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}\mu^2} + \mu\Phi\left(\frac{-\mu}{\sigma}; 0, 1\right),\end{aligned}\tag{6.8}$$

Ya que $\Phi(0; \mu, \sigma^2) = \int_{-\infty}^0 \phi(x; \mu, \sigma^2) dx = \int_{-\infty}^0 \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx = \int_{-\infty}^{-\mu/\sigma} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy = \Phi\left(-\frac{\mu}{\sigma}; 0, 1\right)$.

Podemos ahora expresar los valores esperados para cada una de las expresiones involucradas en el cálculo de $E[uI(u > 0)]$, a partir de la expresión dada en la ec. (6.7), es

decir,

$$\begin{aligned}
E[u_1 I(u_1 > 0)] &= \frac{\sigma}{\sqrt{2\pi}}, \text{ ya que aquí } \mu = 0. \\
E[u_2 I(u_2 > 0)] &= \frac{\sigma}{\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}\mu^2} - \mu \left[1 - \Phi\left(\frac{\mu}{\sigma}; 0, 1\right) \right] \\
E[u_3 I(u_3 > 0)] &= \frac{\sigma}{\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}\mu^2} + \mu \left[1 - \Phi\left(\frac{-\mu}{\sigma}; 0, 1\right) \right] \\
&= \frac{\sigma}{\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}\mu^2} + \mu \Phi\left(\frac{\mu}{\sigma}; 0, 1\right),
\end{aligned}$$

así,

$$\begin{aligned}
&E[u I(u > 0)] \\
&= \frac{\sigma}{\sqrt{2\pi}}(p_{11} + p_{00}) + \left\{ \frac{\sigma}{\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}\mu^2} - \mu \left[1 - \Phi\left(\frac{\mu}{\sigma}; 0, 1\right) \right] \right\} p_{10} \\
&\quad + \left\{ \frac{\sigma}{\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}\mu^2} + \mu \Phi\left(\frac{\mu}{\sigma}; 0, 1\right) \right\} p_{01}.
\end{aligned}$$

Igualmente de la ecuación (6.8) tenemos que $-\frac{\sigma}{\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}\mu^2} + \mu \Phi\left(\frac{-\mu}{\sigma}; 0, 1\right)$

$$\begin{aligned}
E[u_1 I(u_1 < 0)] &= -\frac{\sigma}{\sqrt{2\pi}}, \text{ ya que aquí } \mu = 0. \\
E[u_2 I(u_2 < 0)] &= -\frac{\sigma}{\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}\mu^2} - \mu \left[1 - \Phi\left(\frac{\mu}{\sigma}; 0, 1\right) \right] \\
E[u_3 I(u_3 < 0)] &= -\frac{\sigma}{\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}\mu^2} + \mu \left[1 - \Phi\left(\frac{-\mu}{\sigma}; 0, 1\right) \right] \\
&= -\frac{\sigma}{\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}\mu^2} + \mu \Phi\left(\frac{\mu}{\sigma}; 0, 1\right),
\end{aligned}$$

luego

$$\begin{aligned}
E[u I(u < 0)] &= -\frac{\sigma}{\sqrt{2\pi}}(p_{11} + p_{00}) + \left\{ -\frac{\sigma}{\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}\mu^2} - \mu \left[1 - \Phi\left(\frac{\mu}{\sigma}; 0, 1\right) \right] \right\} p_{10} \\
&\quad + \left\{ -\frac{\sigma}{\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}\mu^2} + \mu \Phi\left(\frac{\mu}{\sigma}; 0, 1\right) \right\} p_{01}.
\end{aligned}$$

Finalmente, sustituyendo en la expresión correspondiente, obtenemos la pérdida esperada dos pasos adelante:

$$\begin{aligned}
\text{RC}_2^{TAR} &= \theta E[u I(u > 0)] + (\theta - 1) E[u I(u < 0)] \\
&= \frac{\sigma}{\sqrt{2\pi}}(p_{11} + p_{00}) + (p_{10} + p_{01}) \left\{ \frac{\sigma}{\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}\mu^2} - \mu \Phi\left(\frac{\mu}{\sigma}; 0, 1\right) + 2\mu \theta \Phi\left(\frac{\mu}{\sigma}; 0, 1\right) \right\} \\
&\quad - 2\mu p_{10} \theta.
\end{aligned}$$

En Capítulo 7, como parte de nuestro ejercicio de simulación obtenemos los valores de las funciones de pérdida de cuantiles para cada uno de los escenarios simulados. La comparación de los desempeños de los pronósticos se realiza mediante los cocientes de los valores de las funciones de pérdida asociadas al modelo lineal AR contra los valores de las funciones de pérdida asociadas al modelo TAR. La observación en torno a esto es que el modelo TAR presenta valores menores en sus funciones de pérdida lo cual le da ventajas en comparación al modelo lineal.

6.3.3 Propiedades asintóticas

A continuación presentamos las propiedades asintóticas de los estimadores de los parámetros en el modelo de cuantiles obtenidos por medio de la minimización de la función cuantil dada en (6.4). Los Teoremas 6.1 y 6.2 muestran que la estimación de β^0 mediante el método de regresión por cuantiles no lineal es consistente y asintóticamente normal. El Teorema 6.3 nos da un estimador consistente de la matriz de varianzas y covarianzas. Las pruebas de los teoremas se dan en el Apéndice C y las condiciones de regularidad para la función f como parte de la función objetivo a minimizar en (6.4) se dan en el Apéndice B. Como notación sea $h_t(0|\mathfrak{F}_t)$ la densidad condicional de $\varepsilon_{t\theta}$ evaluada en 0 y \mathfrak{F}_t la σ -álgebra generada por el conjunto de información hasta el tiempo t . Para $X_{t-1} = Y_{t-1}$, y Z_{t-1} la variable exógena o variable umbral, denote el gradiente de $f_t(\beta, Y_{t-1}, Z_{t-1})$ de orden $1 \times p$ por $\nabla f_t(\beta, Y_{t-1}, Z_{t-1})$ y defina $\nabla f(\beta, Y_{t-1}, Z_{t-1})$ como la matriz de $T \times p$ con renglones $\nabla f_t(\beta, Y_{t-1}, Z_{t-1})$.

Teorema 6.1 (*Consistencia*). *En el modelo (6.3) y bajo los supuestos C0 a C7 enunciados en el apéndice A, se cumple que $\hat{\beta}_T \xrightarrow{p} \beta^0$, donde $\hat{\beta}_T$ es la solución a:*

$$\min_{\beta} T^{-1} \sum_{t=1}^T [\theta - I(Y_t < f_t(\beta, Y_{t-1}, Z_{t-1}))][Y_t - f_t(\beta, Y_{t-1}, Z_{t-1})].$$

Demostración. Ver la demostración en el Apéndice C.

Teorema 6.2 (*Normalidad Asintótica*). *En el modelo (6.3) y bajo los supuestos AN1 a AN4, dados en el Apéndice B y las condiciones del Teorema 6.1, se tiene que:*

$$\sqrt{T} \mathbf{A}_T^{-1/2} \mathbf{D}_T (\hat{\beta} - \beta^0) \xrightarrow{d} N(\mathbf{0}, \mathbf{I}),$$

donde

$$\begin{aligned}\mathbf{A}_T &\equiv E \left[T^{-1} \theta (1 - \theta) \sum_{t=1}^T \nabla' f_t(\boldsymbol{\beta}^0, Y_{t-1}, Z_{t-1}) \nabla f_t(\boldsymbol{\beta}^0, Y_{t-1}, Z_{t-1}) \right], \\ \mathbf{D}_T &\equiv E \left[T^{-1} \sum_{t=1}^T h_t(0|\Omega_t) \nabla' f_t(\boldsymbol{\beta}^0, Y_{t-1}, Z_{t-1}) \nabla f_t(\boldsymbol{\beta}^0, Y_{t-1}, Z_{t-1}) \right],\end{aligned}$$

y $\hat{\boldsymbol{\beta}}$ se calcula como en el Teorema 6.1.

Demostración. Ver la demostración en el Apéndice C.

Teorema 6.3 (*Estimación de la matriz de varianzas y covarianzas*). Bajo los supuestos VC1 a VC3 dados en el Apéndice B y las condiciones de los Teoremas 6.1 y 6.2, $\hat{\mathbf{A}}_T - \mathbf{A}_T \xrightarrow{p} \mathbf{0}$ y $\hat{\mathbf{D}}_T - \mathbf{D}_T \xrightarrow{p} \mathbf{0}$, donde

$$\begin{aligned}\hat{\mathbf{A}}_T &= T^{-1} \theta (1 - \theta) \nabla' f_t(\hat{\boldsymbol{\beta}}, Y_{t-1}, Z_{t-1}) \nabla f_t(\hat{\boldsymbol{\beta}}, Y_{t-1}, Z_{t-1}), \\ \hat{\mathbf{D}}_T &= (2T\hat{c}_T)^{-1} \sum_{t=1}^T I \left[\left| Y_t - f_t(\hat{\boldsymbol{\beta}}, Y_{t-1}, Z_{t-1}) \right| < \hat{c}_T \right] \\ &\quad \times \nabla' f_t(\hat{\boldsymbol{\beta}}, Y_{t-1}, Z_{t-1}) \nabla f_t(\hat{\boldsymbol{\beta}}, Y_{t-1}, Z_{t-1}),\end{aligned}$$

\mathbf{A}_T y \mathbf{D}_T se definieron en el Teorema 6.2 y \hat{c}_T se define en el supuesto VC1 en el Apéndice B.

Demostración. Ver la demostración en el Apéndice C

En la prueba del Teorema 6.1, se establecen los resultados de consistencia de modelos no lineales de regresión por cuantiles en un contexto dinámico aplicando el Corolario 5.12 del libro de White (1996), enunciado en el Apéndice C. El supuesto C1, requiere continuidad en el vector de parámetros $\boldsymbol{\beta}$ de las especificaciones de los cuantiles. Los supuestos C3 y C7 son condiciones de identificación que son comunes en la literatura de regresión por cuantiles. Los supuestos C4 y C5 son condiciones de dominio para dejar fuera del análisis comportamientos explosivos.

La derivación de la distribución asintótica se construye mediante la aproximación de los gradientes discontinuos de la función objetivo con funciones suaves diferenciables, además de que se pueden realizar expansiones de Taylor. Los supuestos AN1 y AN2 imponen condiciones de suficiencia de la función $f_t(\boldsymbol{\beta}, Y_{t-1}, Z_{t-1})$ y de la función de densidad condicional de los términos de error para asegurar que está aproximación suavizada sea bien comportada.

En relación a la matriz de varianzas y covarianzas, note que $\hat{\mathbf{A}}_T$ es simplemente el producto externo del gradiente. La estimación de la matriz \mathbf{D}_T es menos directa, ya que involucra el término $h(0|\mathfrak{F}_t)$.

6.3.4 Pruebas de Hipótesis

Si el modelo (6.3) es el verdadero proceso generador de datos (VPGD), entonces $\Pr [Y_t < f_t(\beta^0, Y_{t-1}, Z_{t-1})] = \theta$ para todo t , esto es equivalente a requerir que la secuencia de funciones indicadoras $\{I[Y_t < f_t(\beta^0, Y_{t-1}, Z_{t-1})]\}_{t=1}^T$ sea independiente e idénticamente distribuida (*iid*). La idea aquí es proveer un filtro para transformar series de tiempo con posible correlación serial y heteroscedasticidad sobre secuencias serialmente independientes de funciones indicadoras. Una forma natural para probar la validez del modelo de pronóstico es verificar si la secuencia $\{I[Y_t < f_t(\beta^0, Y_{t-1}, Z_{t-1})]\}_{t=1}^T \equiv \{I_t\}_{t=1}^T$ es *iid*, como fue realizado, por ejemplo, en Christoffersen (1998). A través de estas pruebas podemos detectar la presencia de correlación serial de las funciones indicadoras $\{I_t\}_{t=1}^T$, esta condición es solo necesaria pero no suficiente para juzgar o valorar el comportamiento de un modelo de cuantiles. Sin embargo, no es difícil generar una secuencia de variables indicadoras $\{I_t\}_{t=1}^T$ independientes a partir de una secuencia dada $\{Y_t\}_{t=1}^T$. Para lograr lo anterior es suficiente con definir una secuencia de variables aleatorias independientes $\{w_t\}_{t=1}^T$, tal que

$$w_t = \begin{cases} 1 & \text{con probabilidad } \theta \\ -1 & \text{con probabilidad } (1 - \theta) \end{cases}, \quad (6.9)$$

de modo que podemos tomar $f_t(\beta^0, Y_{t-1}, Z_{t-1}) = Kw_t$, para K grande. Note sin embargo, que una vez que w_t es observado, la probabilidad de exceder el cuantil es conocida y será casi cero o uno. Así las probabilidades no condicionales son correctas y serialmente no correlacionadas, pero las probabilidades condicionales dado los cuantiles no lo son. Este ejemplo es un caso extremo de errores en la medición de los cuantiles, cualquier ruido introducido en la estimación de los cuantiles cambiará la probabilidad condicional de un acierto dado la estimación del mismo.

Por lo tanto, ninguna de estas pruebas tiene potencia contra esta forma de mala especificación y ninguna puede ser extendida a examinar otras variables explicatorias. La propuesta de Engle y Manganelli (2004) consiste en definir la siguiente función, la cual contabiliza en cierta forma los aciertos o las clasificaciones correctas ante una mala especificación,

$$Hit_t(\beta^0) \equiv I(Y_t < f_t(\beta^0, Y_{t-1}, Z_{t-1})) - \theta. \quad (6.10)$$

La función $Hit_t(\beta^0)$ asume valores de $(1 - \theta)$ cada vez que $Y_t < f_t(\beta^0, Y_{t-1}, Z_{t-1})$ y $-\theta$ en otros casos. Note que $E[Hit_t(\beta^0)] = 0$ y de la definición de función cuantil, la esperanza condicional de $Hit_t(\beta^0)$ dada cualquier información conocida hasta el tiempo $t - 1$ debe de ser también igual a cero. Observemos además que para todo $\omega_{t-1} \in \mathfrak{F}_{t-1}$, tenemos que

$$\begin{aligned} E[Hit_t(\beta^0) \omega_{t-1}] &= E\{E[Hit_t(\beta^0) \omega_{t-1} | \mathfrak{F}_{t-1}]\} \\ &= \omega_{t-1} E\{E[Hit_t(\beta^0) | \mathfrak{F}_{t-1}]\} = 0, \end{aligned}$$

es decir, la función $Hit_t(\beta^0)$ es no correlacionada con cualquier cosa que pertenezca al conjunto de información, en particular es no correlacionada con rezagos de ella misma y con $f_t(\beta^0, Y_{t-1}, Z_{t-1})$. Si la función $Hit_t(\beta^0)$ satisface estas condiciones de momentos, entonces no habrá autocorrelaciones con los aciertos, ni errores de medición como en la ecuación (6.9). Una forma natural para el conjunto de pruebas discutidas es verificar cuando el estadístico de prueba $T^{-1/2}\mathbf{X}'(\hat{\beta})\mathbf{Hit}(\hat{\beta})$ es significativamente diferente de cero, donde $\mathbf{X}_t(\hat{\beta})$, $t = 1, \dots, T$, es un renglón típico de $\mathbf{X}(\hat{\beta})$ (posiblemente dependiente en $\hat{\beta}$), y es un q -vector medible en \mathfrak{F}_t y $\mathbf{Hit}(\hat{\beta}) \equiv [Hit_1(\hat{\beta}), \dots, Hit_T(\hat{\beta})]'$. Podemos pensar en cierto modo que realizamos una regresión ficticia para verificar si el estadístico de prueba $T^{-1/2}\mathbf{X}'(\hat{\beta})\mathbf{Hit}(\hat{\beta})$ es significativo, con $\mathbf{X}(\hat{\beta})$ la matriz de diseño y $\mathbf{Hit}(\hat{\beta})$ la variable de respuesta.

Sea $\mathbf{M}_T \equiv (\mathbf{X}'(\beta^0) - E[T^{-1}\mathbf{X}'(\beta^0)\mathbf{H}\nabla f(\beta^0, Y_{t-1}, Z_{t-1})]\mathbf{D}_T^{-1}\nabla' f(\beta^0, Y_{t-1}, Z_{t-1})),$ donde \mathbf{H} es una matriz diagonal con entradas típicas dadas por $h_t(0|\mathfrak{F}_t)$.

Teorema 6.4 (*Prueba de cuantil dinámico dentro de la muestra*). *Bajo los supuestos del Teorema 6.1, Teorema 6.2 y los supuestos DQ1 a DQ6, enunciados en el Apéndice B, se sigue el siguiente resultado:*

$$[\theta(1-\theta)E(T^{-1}\mathbf{M}_T\mathbf{M}_T')^{-\frac{1}{2}}T^{-\frac{1}{2}}\mathbf{X}'(\hat{\beta})\mathbf{Hit}(\hat{\beta})] \stackrel{d}{\sim} N(\mathbf{0}, \mathbf{I}).$$

Si se cumplen además las condiciones del Teorema 6.3 y el supuesto DQ7, se cumple también el siguiente resultado asintótico:

$$CD_{DM} \equiv \frac{\mathbf{Hit}'(\hat{\beta})\mathbf{X}(\hat{\beta})(\hat{\mathbf{M}}_T\hat{\mathbf{M}}_T')^{-1}\mathbf{X}'(\hat{\beta})\mathbf{Hit}(\hat{\beta})}{\theta(1-\theta)} \stackrel{d}{\sim} \chi_q^2,$$

cuando $T \rightarrow \infty$, y donde,

$$\begin{aligned} \hat{\mathbf{M}}_T &\equiv - \left\{ (2T\hat{c}_T)^{-1} \sum_{t=1}^T I \left[\left| Y_t - f_t(\hat{\beta}, Y_{t-1}, Z_{t-1}) \right| < \hat{c}_T \right] \mathbf{X}'_t(\hat{\beta}) \nabla f_t(\hat{\beta}, Y_{t-1}, Z_{t-1}) \right\} \\ &\quad \times \hat{\mathbf{D}}_T^{-1} \nabla' f(\hat{\beta}, Y_{t-1}, Z_{t-1}). \end{aligned}$$

Demostración. La demostración se encuentra en el Apéndice C.

Si $\mathbf{X}(\hat{\beta})$ contiene $m < q$ rezagos $Hit_{t-1}(\hat{\beta})$ ($i = 1, \dots, m$), entonces $\mathbf{X}(\hat{\beta})$, $\mathbf{Hit}(\hat{\beta})$ y $\nabla f_t(\hat{\beta}, Y_{t-1}, Z_{t-1})$ no son conformables, ya que $\mathbf{X}(\hat{\beta})$ contiene sólo $(T - m)$ elementos. Suponemos entonces, sin pérdida de generalidad, que las matrices son conformables eliminando los primeros m renglones de $\nabla f_t(\hat{\beta}, Y_{t-1}, Z_{t-1})$ y $\mathbf{Hit}(\hat{\beta})$.

Para derivar la prueba de cuantil dinámico fuera de la muestra, sea T_R el número de observaciones dentro de la muestra y sea N_R el número de observaciones fuera de la muestra, donde la dependencia entre T_R y N_R en R se da por el siguiente supuesto: $\lim_{R \rightarrow \infty} T_R = \infty$, $\lim_{R \rightarrow \infty} N_R = \infty$ y $\lim_{R \rightarrow \infty} N_R/T_R = 0$. Defina el q -vector Ω_t medible $\mathbf{X}_n(\hat{\beta}_{T_R})$, $n = T_R + 1, \dots, T_R + N_R$, como un renglón típico de $\mathbf{X}(\hat{\beta}_{T_R})$, posiblemente dependiente en $\hat{\beta}_{T_R}$, y $\mathbf{Hit}(\hat{\beta}_{T_R}) \equiv [\text{Hit}_{T_R+1}(\hat{\beta}_{T_R}), \dots, \text{Hit}_{T_R+N_R}(\hat{\beta}_{T_R})]'$.

Teorema 6.5 (*Prueba de cuantil dinámico fuera de la muestra*). *Bajo los supuestos de los Teoremas 6.1 y 6.2 y los supuestos DQ1 a DQ3, DQ8, y DQ9, en el Apéndice B, se tiene el siguiente resultado asintótico:*

$$CD_{FM} \equiv \frac{\mathbf{Hit}'(\hat{\beta}_{T_R})\mathbf{X}(\hat{\beta}_{T_R}) \left[\mathbf{X}'(\hat{\beta}_{T_R})\mathbf{X}(\hat{\beta}_{T_R}) \right]^{-1} \mathbf{X}'(\hat{\beta}_{T_R})\mathbf{Hit}(\hat{\beta}_{T_R})}{\theta(1-\theta)} \stackrel{d}{\sim} \chi_q^2,$$

cuando $R \rightarrow \infty$.

Demostración. Ver la demostración en el Apéndice C.

Nota: La propuesta original en el artículo de Engle y Manganelli (2004) dice que CD_{FM} está dado por:

$$CD_{FM} \equiv N_R^{-1} \frac{\mathbf{Hit}'(\hat{\beta}_{T_R})\mathbf{X}(\hat{\beta}_{T_R}) \left[\mathbf{X}'(\hat{\beta}_{T_R})\mathbf{X}(\hat{\beta}_{T_R}) \right]^{-1} \mathbf{X}'(\hat{\beta}_{T_R})\mathbf{Hit}(\hat{\beta}_{T_R})}{\theta(1-\theta)} \stackrel{d}{\sim} \chi_q^2,$$

cuando $R \rightarrow \infty$. Sin embargo, al momento de escribir la demostración del teorema en el Apéndice C, el término N_R^{-1} no aparece.

A manera de ilustración, conviene mencionar que aún cuando el objetivo de esta tesis está encaminado al estudio de los desempeños y obtención de pronósticos h -pasos adelante para los modelos TAR hemos explorado empíricamente los desempeños de las pruebas de cuantil dinámico dentro y fuera de la muestra. Es decir, dado un conjunto de datos simulados a partir de un modelo TAR, dividimos el conjunto de datos en dos de manera tal que el primer grupo de observaciones constituye el número de observaciones en la muestra, las cuales se utilizan para el proceso de estimación; mientras que el segundo grupo forma el conjunto de observaciones fuera de la muestra. Específicamente simulamos muestras de tamaños 1500 a partir de la representación del modelo TAR dado en la ec. (6.5). Para cada conjunto de datos y asumiendo que el parámetro de umbral es conocido. Las primeras 1000 observaciones se utilizan para la estimación del vector de parámetros $\beta = (\phi_1, \phi_2)'$, mediante la minimización de la función objetivo dada en la ecuación (6.4), dichas observaciones se identifican como

las observaciones dentro de la muestra. Las últimas 500 observaciones son las observaciones fuera de la muestra y las utilizamos para calcular el estadístico de prueba de cuantil dinámico fuera de la muestra y tomar la decisión, siempre que sea posible, sobre cual modelo es más adecuado para el conjunto de datos analizado. En nuestro caso los modelos candidatos son el modelo TAR(2; 1, 1) y el modelo AR(1). Suponemos que la variable umbral Z_{t-1} es conocida, así como el valor del umbral γ . Presentamos tres escenarios distintos, los cuales son estudiados en el Capítulo 7 como parte del ejercicio de simulación implementado y que denotamos por caso 1-1, caso 3-1 y caso 8-1 respectivamente. Consideremos que los regímenes son igualmente dominantes y un nivel de significancia asociado a las pruebas de hipótesis de 0.05, mientras que el número de columnas en las matrices $\mathbf{X}(\hat{\beta})$ y $\mathbf{X}(\hat{\beta}_{T_R})$ es 4, las cuales corresponden a los primeros rezagos de la matriz $\mathbf{Hit}(\hat{\beta})$. De lo anterior el valor de tablas para realizar la prueba tanto dentro y fuera de la muestra tiene $q = 4$ grados de libertad asociados, es decir, tomamos la decisión de no rechazar la hipótesis de que el modelo es adecuado siempre que el estadístico de prueba CD_{DM} y CD_{FM} sean menores que 9.487729. Para visualizar los resultados rápidamente calculamos los p -valores en cada una de las pruebas, de modo que si el p -valor es mayor que 0.05 no se rechaza la hipótesis de que el modelo en cuestión es adecuado para describir el conjunto de datos considerados.

La Tablas 2 y 3 presentan los resultados para el caso 1-1 ($\phi_1 = 0.1$, $\phi_2 = 1$), caso 3-1 ($\phi_1 = -0.8$, $\phi_2 = -0.6$) y caso 8-1 ($\phi_1 = -0.95$, $\phi_2 = 0.95$). $\gamma = 0.5$, y $\theta = 0.5$ para cada uno de los casos, considerando que el modelo está dado por un modelo TAR(2; 1, 1) o un modelo AR(1), respectivamente.

Tabla 2. Resultados de las pruebas de cuantil dinámico dentro y fuera de la muestra.

Bajo un TAR(2; 1, 1)	$\hat{\phi}_1$	$\hat{\phi}_2$	CD_{DM}	$p\text{-valor}_{DM}$	DQ_{FM}	$p\text{-valor}_{FM}$
caso 1-1	-0.12	1.04	3.31	0.51	4.08	0.40
caso 3-1	-0.84	-0.57	1.39	0.85	6.47	0.17
caso 8-1	-0.95	0.94	3.09	0.54	5.56	0.23

Tabla 3. Resultados de las pruebas de cuantil dinámico dentro y fuera de la muestra.

Bajo un AR(1)	$\hat{\phi}$	CD_{DM}	$p\text{-valor}_{DM}$	DQ_{FM}	$p\text{-valor}_{FM}$
caso 1-1	0.32	188	0.000	10.4	0.03
caso 3-1	-0.77	4.73	0.316	5.43	0.25
caso 8-1	0.04	16.17	0.002	26.09	0.00003

Los resultados de las Tablas 2 y 3 nos muestran que para los casos 1-1 y 8-1, podemos decir que el modelo TAR(2; 1, 1) es adecuado para describir el comportamiento de los datos dentro y fuera de muestra, mientras que el modelo AR(1) no es adecuado ya que sus p -valores son mucho menores que el nivel de significancia estipulado para las pruebas (0.05).

Por su parte el caso 3-1 identifica que tanto el modelo $TAR(2; 1, 1)$ como el modelo $AR(1)$ son modelos adecuados para describir los conjunto de datos simulados e identificados como dentro y fuera de la muestra. En el Capítulo 7 mostramos también que los desempeños de los pronósticos para el caso 3-1 bajo TAR y AR son iguales como función de las funciones de pérdida consideradas.

A partir de los resultados anteriores esperamos que la alternativa de regresión por cuantiles para generar y evaluar los pronósticos en modelos TAR presente mejores desempeños que los pronósticos obtenidos bajo el modelo AR.

Capítulo 7

Estudio de simulación

El propósito principal de este capítulo es analizar los comportamientos de los pronósticos h -pasos adelante en los modelos TAR, considerando cada una de las alternativas de pronóstico descritas en los capítulos previos. Cuantificamos los desempeños de los pronósticos por medio de las funciones de pérdida ECMP y PCR; donde PCR denota la función de pérdida asociada a la función de regresión por cuantiles. Cabe aclarar que la función de pérdida PCR es conocida en la literatura como función de pérdida “check” o “tick”, dicha función de pérdida es simétrica cuando $\theta = 0.5$, es decir, el caso en que la función de pérdida es el valor absoluto de los errores de pronóstico. Si $0 < \theta < 1$ y $\theta \neq 0.5$, la función de pérdida PCR, describe situaciones de asimetrías como función de los valores de θ tal y como se mostró en la Figura 14. Ligado al proceso de comparación y obtención de los pronósticos esta el proceso de estimación de los parámetros en cada uno de los modelos, por tal razón incluimos como parte de nuestro estudio de simulación, el análisis correspondiente para los estimadores bajo cada alternativa. La comparación de las funciones de pérdida (ECMP y PCR) para los modelos TAR(2; 1, 1) y AR(1) se realiza considerando los pronósticos para cada modelo y bajo el mismo método de obtención de los pronósticos. Los métodos o procedimientos estudiados son: procedimiento recursivo, procedimiento de verosimilitud predictiva y procedimiento de regresión por cuantiles, los cuales denotaremos como PR, PVP, y PCR respectivamente. El párrafo anterior significa, por ejemplo, que si los pronósticos en el modelo TAR(2; 1, 1) son obtenidos mediante verosimilitud predictiva, entonces también los pronósticos en el modelo AR(1) se obtienen por verosimilitud predictiva, considerando así la misma metodología para el proceso de estimación de los parámetros involucrados y de obtención y evaluación de los pronósticos. Los resultados del amplio ejercicio de simulación nos darán herramientas para identificar cuál de las alternativas estudiadas es la más adecuada y bajo que condiciones o características se observa dicha mejora.

7.1 Descripción del algoritmo de simulación

Nuestro ejercicio de simulación consiste en generar 500 series de tiempo de tamaño 255 considerando el modelo que describe a la variable umbral y a partir de éstas simular las 500 series correspondientes bajo el modelo $TAR(2; 1, 1)$, tal y como se especificó en el Capítulo 1. Los escenarios simulados, se dan en función de los valores de los parámetros en el modelo y de las probabilidades de dominancia en cada uno de los regímenes, equivalentemente podemos resumir esta información considerando los valores de los primeros dos momentos de la variable $\delta_t = \phi_1 I(Z_{t-1} \leq \gamma) + \phi_2 I(Z_{t-1} > \gamma)$. Las situaciones analizadas abarcan una gran variedad de situaciones posibles como son: una raíz unitaria en uno de los regímenes, coeficientes autorregresivos con signos iguales y signos diferentes, si los signos son diferentes, consideramos las posibilidades de que estos sean iguales o diferentes en magnitud, así como, igual dominancia en cada uno de los regímenes (es decir, el mismo porcentaje de observaciones en el régimen 1 y 2), un régimen dominante, varianzas del proceso de error grandes y pequeñas, etc..

Para cada una de las 500 series simuladas, y considerando el parámetro de umbral γ conocido, usamos las primeras 250 observaciones para estimar los coeficientes autorregresivos, así como la varianza del proceso de error bajo los modelos candidatos, un $TAR(2; 1, 1)$ y un $AR(1)$. Contamos con tres procedimientos de estimación para dichos parámetros, bajo cada modelo, mínimos cuadrados, máxima verosimilitud y regresión por cuantiles. Máxima verosimilitud y mínimos cuadrados coinciden bajo normalidad en los errores y de aquí que presentamos solamente los resultados para los estimadores por máxima verosimilitud. Contemplamos dos posibilidades para la variable umbral Z_{t-1} , el caso de $Z_{t-1} \sim iidU(0, 1)$ y el caso de Z_{t-1} un modelo $AR(1)$ estacionario con ruido blanco gaussiano. Para cada uno de los modelos de interés, obtenemos pronósticos para $h = 1, 2, 3, 4$, y 5 pasos hacia adelante por medio de los métodos identificados como método recursivo, método de verosimilitud predictiva y método de regresión por cuantiles, bajo el supuesto de que el parámetro de umbral γ es conocido. El procedimiento recursivo PR, consiste en aproximar la esperanza condicional dada en la ecuación (4.1) por medio de un procedimiento recursivo del tipo Monte Carlo, equivalente al procedimiento general descrito en el Capítulo 1, Granger y Teräsvirta, (1993). Comparamos los desempeños de cada uno de los pronósticos, calculando los ECMP y la función de pérdida “check” para los últimos cinco valores simulados en cada una de las 500 series. Promediamos los 500 valores obtenidos de las respectivas funciones de pérdida para cada uno de los horizontes de pronósticos. Posteriormente tomamos los cocientes de los valores de las pérdidas asociadas al modelo lineal $AR(1)$ entre los valores de las pérdidas asociadas al modelo $TAR(2; 1, 1)$, para los horizontes de pronóstico $h = 1, 2, 3, 4$, y 5. Valores de los cocientes mayores que uno, implican que la pérdida asociada al modelo $TAR(2; 1, 1)$ para el horizonte de pronóstico en cuestión es menor en magnitud que la pérdida asociada al modelo $AR(1)$, concluyendo que los desempeños de los pronósticos en el modelo $TAR(2; 1, 1)$ son mejores

que los pronósticos en el modelo AR(1), bajo la alternativa de pronóstico considerada.

Método recursivo. La descripción del método recursivo, utilizado para la aproximación de las esperanzas condicionales para $h = 1, 2, 3, 4$, y 5 en la ecuación (4.1), consiste en el siguiente proceso iterativo: bajo el supuesto de que la variable umbral $Z_{t-1} \sim iidU(0, 1)$ tenemos que conocemos con certeza el comportamiento de la serie en los distintos horizontes de pronósticos, sin embargo para el caso en que $Z_{t-1} \sim AR(1)$ estacionario con errores gaussianos debemos pronosticar como primer paso los valores de $Z_{T+1}, \dots, Z_{T+h-1}$ partiendo del modelo asociado para Z_{t-1} o bien pronosticar los valores de $I(Z_{t-1} \leq \gamma)$ con γ fijo, modelando la variable indicadora como una serie binaria, Kedem y Fokianos (2002). En el Capítulo 4 definimos las herramientas necesarias para el cálculo de dichos pronósticos. Rescribamos el modelo TAR(2; 1, 1),

$$Y_t = [\phi_1 I(Z_{t-1} \leq \gamma) + \phi_2 I(Z_{t-1} > \gamma)]Y_{t-1} + \varepsilon_t,$$

donde el pronóstico optimal h -pasos adelante, bajo pérdida cuadrática, está dado por la esperanza condicional:

$$\hat{Y}_{t+h} = E(Y_{t+h} | \mathfrak{F}_t) = \phi_1 E[I(Z_{t+h-1} \leq \gamma) Y_{t+h-1} | \mathfrak{F}_t] + \phi_2 E[I(Z_{t+h-1} > \gamma) Y_{t+h-1} | \mathfrak{F}_t].$$

Dado un conjunto de T observaciones tenemos que el pronóstico $T + 1$ condicional a la información hasta el tiempo T , para el modelo TAR(2; 1, 1) se obtiene fácilmente ya que el régimen es conocido con certeza y el pronóstico un paso adelante por PR es simplemente:

$$\hat{Y}_{T+1}^{PR} = [\hat{\phi}_1 I(Z_T \leq \gamma) + \hat{\phi}_2 I(Z_T > \gamma)]Y_T$$

con $\hat{\phi}_1$ y $\hat{\phi}_2$ los estimadores de los coeficientes autorregresivos en el modelo TAR(2; 1, 1), obtenidos por medio de las ecuaciones de verosimilitud bajo el supuesto de normalidad en los errores. Al tiempo $T + 2$ el pronóstico del régimen está determinado por el valor de Z_{T+1} , de modo que sustituimos el correspondiente pronóstico, generado a partir del modelo que rige la dinámica de la variable umbral. Lo mismo resulta si sustituimos el valor pronosticado de la indicadora $I(Z_{T+1} \leq \gamma)$. Entonces se define el pronóstico para $Y_{T+2}, Y_{T+3}, \dots, Y_{T+h}$ para una realización del proceso de error, $\zeta_{i,j} \sim N(0, \hat{\sigma}^2)$, donde $\hat{\sigma}^2$ es el estimador de la varianza del error bajo la consideración del modelo TAR(2; 1, 1), como:

$$\hat{Y}_{T+2}^{PRZt_j} = [\hat{\phi}_1 I(\tilde{Z}_{T+1} \leq \gamma) + \hat{\phi}_2 I(\tilde{Z}_{T+1} > \gamma)]\hat{Y}_{T+1}^{PRZt_j} + \zeta_{2,j},$$

ó

$$\hat{Y}_{T+2}^{PRIZt_j} = [\hat{\phi}_1 \tilde{I}(Z_{T+1} \leq \gamma) + \hat{\phi}_2 \tilde{I}(Z_{T+1} > \gamma)]\hat{Y}_{T+1}^{PRIZt_j} + \zeta_{2,j},$$

dependiendo de si usamos el pronóstico de la variable umbral \tilde{Z}_{T+1} o el valor del pronóstico de la variable indicadora $\tilde{I}(Z_{T+1} \leq \gamma)$, al tiempo $T + 1$. Igualmente al tiempo $T + h$, el pronóstico h -pasos adelante para Y_t se aproxima por medio de las siguientes ecuaciones:

$$\hat{Y}_{T+h}^{PRZt_j} = [\hat{\phi}_1 I(\tilde{Z}_{T+h-1} \leq \gamma) + \hat{\phi}_2 I(\tilde{Z}_{T+h-1} > \gamma)]\hat{Y}_{T+h-1}^{PRZt_j} + \zeta_{h,j} \quad (7.1)$$

6

$$\hat{Y}_{T+h}^{PRIZt_j} = [\hat{\phi}_1 \tilde{I}(Z_{T+h-1} \leq \gamma) + \hat{\phi}_2 \tilde{I}(Z_{T+h-1} > \gamma)] \hat{Y}_{T+h-1}^{PRIZt_j} + \zeta_{h,j}, \quad (7.2)$$

donde $PRZt_j$ se refiere al método recursivo usando el pronóstico de la variable Z_t , mientras que $PRIZt_j$ indica que el método recursivo utiliza el pronóstico de la variable indicadora $I(Z_{T+h-1} \leq \gamma)$. Cada uno de los h pronósticos hacia adelante se realiza j veces en forma iterativa, de modo que el pronóstico final se obtiene promediando dichos valores para cada uno de los h periodos sobre las $j = 1, \dots, N$ iteraciones. Denotamos a \hat{Y}_{T+h}^{PRZt} ó \hat{Y}_{T+h}^{PRIZt} como el pronóstico h -pasos adelante para Y_t vía el método recursivo dependiendo de si utilizamos el pronóstico de la variable umbral o de la variable indicadora y dado por el promedio muestral:

$$\hat{Y}_{T+h}^{PRZt} = \frac{1}{N} \sum_{j=1}^N \hat{Y}_{T+h}^{PRZt_j}, \quad \hat{Y}_{T+h}^{PRIZt} = \frac{1}{N} \sum_{j=1}^N \hat{Y}_{T+h}^{PRIZt_j}, \quad h > 1.$$

Esta alternativa para generar pronósticos, así como variantes de la misma, han sido estudiadas ampliamente para los modelos SETAR, De Gooijer y De Bruin (1998), Clements y Smith (1997,1999). Inherente al proceso de aproximación de los pronósticos, obtenemos en cada iteración los correspondientes errores de pronósticos para cada una de las 500 series simuladas y posteriormente obtenemos los valores de las funciones de pérdida ECMP y PCR. Para comparar los pronósticos en el modelo TAR(2; 1, 1), bajo este método, generamos los pronósticos h -pasos adelante bajo el modelo AR(1) por medio de la ecuación recursiva correspondiente, obtenemos al igual que el modelo TAR los valores de las funciones de pérdida ECMP y PCR.

Verosimilitud predictiva. Los pronósticos vía verosimilitud predictiva para el modelo TAR(2; 1, 1) los obtenemos por medio del siguiente procedimiento: recordemos que por el principio de verosimilitud, los pronósticos para los h valores no observados serán aquellos valores que minimicen el negativo de la función de log verosimilitud perfil, dada en la ecuación (5.6) o en la ecuación (5.7) dependiendo de si suponemos que la variable umbral sigue un proceso estacionario AR(1) o bien si este es un proceso independiente e idénticamente distribuido $U(0, 1)$ respectivamente. Al igual que el método recursivo, si $Z_{t-1} \sim iidU(0, 1)$ consideramos los últimos 5 valores simulados en cada una de las 500 series para Z_{t-1} como dados y optimizamos la función correspondiente a la verosimilitud predictiva como función de los valores Y_{T+1}, \dots, Y_{T+h} , de tal manera que los valores encontrados corresponden a aquellos valores con mayor plausibilidad en la verosimilitud predictiva. Realizamos el proceso de optimización para cada una de las series simuladas, contamos entonces con 500 realizaciones de pronósticos h -pasos adelante por VP para el modelo TAR(2; 1, 1). Por su parte, bajo el supuesto de que la variable umbral posee una estructura de dependencia dada por: $Z_{t-1} \sim AR(1)$ con errores gaussianos, obtenemos los pronósticos de Y_t y Z_{t-1} por VP de forma conjunta, es decir, optimizamos la VP en función de los valores de Y_{T+1}, \dots, Y_{T+h} y $Z_T, Z_{T+1}, \dots, Z_{T+h-1}$ para la función objetivo correspondiente. Posteriormente, calculamos los errores de pronóstico considerando los últimos cinco valores simulados para Y_t . A

continuación calculamos las funciones de pérdida ECMP y PCR promediando los 500 valores de las funciones de pérdida para cada horizonte de pronóstico. Para la comparación con el modelo alternativo lineal AR(1), hacemos lo propio, es decir, obtenemos los pronósticos vía el método de VP, calculamos los errores de pronóstico considerando las últimas cinco observaciones para cada una de las 500 series TAR(2; 1, 1) simuladas y posteriormente obtenemos los valores de las pérdidas ECMP y PCR.

Regresión por cuantiles. Finalmente, el criterio para generar pronósticos h -pasos adelante para el modelo TAR(2;1,1), considerando las herramientas de regresión por cuantiles consistirá en la optimización de la función cuantil como función de los valores desconocidos Y_{T+1}, \dots, Y_{T+h} , asumiendo que los valores futuros de la variable umbral o de la variable indicadora son conocidos. Como en el caso del método recursivo, si $Z_t \sim iidU(0, 1)$ los valores de Z_t al tiempo $T, T+1, \dots, T+h-1$ son conocidos, mientras que si Z_{t-1} sigue un proceso AR(1) estacionario dichos valores serán pronosticados y sustituidos en la función cuantil así como los valores estimados de los parámetros en el modelo bajo la misma metodología de estimación. Paralelamente obtenemos los errores de pronóstico para calcular las pérdidas ECMP y PCR para $h = 1, 2, 3, 4$, y 5 respectivamente. La función cuantil se caracteriza por el valor de θ , ya que este determina la asimetría de la función de cuantil, sin embargo dado que en nuestro ejercicio de simulación consideramos errores $\varepsilon_t \sim iidN(0, \sigma^2)$, tomamos el valor de $\theta = \Pr(\varepsilon_t \leq 0) = 0.5$, es decir, la función cuantil es la función de pérdida simétrica valor absoluto. Los pronósticos por medio del procedimiento CR para el modelo TAR(2; 1, 1) son comparados con los pronósticos para el modelo AR(1) también obtenidos por CR, la comparación se realiza al igual que en las otras alternativas, es decir, en función de los ECMP y de la función de pérdida asociada al caso de regresión por cuantiles.

7.2 Presentación y análisis de los resultados

La Tabla 4 contiene cada uno de los escenarios considerados en el ejercicio de simulación, nuestro interés es analizar los comportamientos de los pronósticos en el modelo TAR(2;1,1) contra los pronósticos en el modelo AR(1) bajo cada una de las alternativas de pronóstico, con parámetro de umbral γ conocido, para cada uno de los renglones de valores en la Tabla 4.

Tabla 4. Valores de los parámetros utilizados en el ejercicio de simulación.

Identificación	ϕ_1	ϕ_2	$\Pr(Z_{t-1} \leq \gamma)$	$E(\delta_t)$	$E(\delta_t^2)$
caso 1-1	-0.1	1	0.50	0.45	0.505
caso 1-2	-0.1	1	0.75	0.175	0.25
caso 1-3	1	-0.1	0.75	0.725	0.753
caso 2-1	1	0.1	0.50	0.55	0.505
caso 2-2	1	0.1	0.75	0.775	0.753
caso 2-3	0.1	1	0.75	0.325	0.258
caso 3-1	-0.8	-0.6	0.50	-0.7	0.500
caso 4-1	-0.8	0.6	0.50	-0.1	0.500
caso 4-2	-0.8	0.6	0.75	-0.45	0.570
caso 4-3	0.6	-0.8	0.75	0.25	0.430
caso 5-1	-0.4	-0.6	0.50	-0.5	0.260
caso 6-1	-0.4	0.6	0.50	0.1	0.260
caso 6-2	-0.4	0.6	0.75	-0.15	0.210
caso 6-3	0.6	-0.4	0.75	0.35	0.310
caso 7-1	-0.4	0.4	0.50	0	0.160
caso 7-2	-0.4	0.4	0.75	-0.2	0.160
caso 7-3	0.4	-0.4	0.75	0.2	0.160
caso 8-1	-0.95	0.95	0.50	0	0.903
caso 8-2	-0.95	0.95	0.75	-0.475	0.903
caso 8-3	0.95	-0.95	0.75	0.475	0.903
caso 9-1	-0.9	1	0.50	0.05	0.905
caso 9-2	-0.9	1	0.75	-0.425	0.858
caso 9-3	1	-0.9	0.75	0.525	0.952
caso 10-1	0.9	1	0.50	0.95	0.905
caso 11-1	0.9	0.7	0.50	0.8	0.650

Consideramos dos valores para la varianza asociada al proceso de error, $\sigma^2 = 1$ y $\sigma^2 = 4$. Analizamos cada uno de los casos contenidos en la Tabla 4 para los dos valores de la varianza del error y bajo el supuesto de que $Z_{t-1} \sim iidU(0,1)$ y $Z_{t-1} \sim AR(1)$ con ruido estándar

gaussiano y coeficiente autorregresivo igual a 0.5. Observe que en la Tabla 4, en la columna de identificación no se consideran por ejemplo los casos identificados como: caso 3-2 y caso 3-3, la justificación para no considerar dichos casos es que, por la cercanía en la magnitud de los valores de ϕ_1 y ϕ_2 , además de que estos son del mismo signo, tenemos que $E(\delta_t)$ y $E(\delta_t^2)$ son prácticamente iguales a los valores correspondiente en el caso 3-1, como se observa a continuación:

Identificación	ϕ_1	ϕ_2	$\Pr(Z_{t-1} \leq \gamma)$	$E(\delta_t)$	$E(\delta_t^2)$
caso 3-1	-0.8	-0.6	0.5	-0.7	0.5
caso 3-2	-0.8	-0.6	0.75	-0.75	0.57
caso 3-3	-0.6	-0.8	0.75	-0.65	0.43

Por tal razón es suficiente con analizar los resultados para el caso 3-1, ya que esperamos que el comportamiento de las medidas de desempeño sea esencialmente el mismo. La misma observación es válida para el resto de los casos faltantes en la Tabla correspondiente.

Primeramente presentamos los comportamientos de los estimadores para cada uno de los parámetros, bajo los métodos de máxima verosimilitud y regresión por cuantiles. Posteriormente, analizamos los comportamientos de los pronósticos bajo cada alternativa considerada. Por tal razón la primer parte del análisis y presentación de los resultados consiste en analizar los comportamientos de los parámetros estimados bajo el supuesto de $Z_{t-1} \sim iidU(0, 1)$ y $Z_{t-1} \sim AR(1)$ estacionario, respectivamente.

7.2.1 Resultados asumiendo que $Z_{t-1} \sim iidU(0, 1)$

Si $Z_{t-1} \sim iidU(0, 1)$ tenemos que $\Pr(Z_{t-1} \leq \gamma) = \gamma$, es decir la probabilidad de dominancia en el régimen uno es el valor del umbral γ . Consideramos también que los últimos cinco valores simulados de Z_{t-1} para cada una de las 500 series simuladas se consideran como valores conocidos y a partir de estos obtenemos los pronósticos h -pasos hacia adelante para Y_t , el modelo TAR(2; 1, 1). El análisis bajo el supuesto de que la secuencia $Z_{t-1} \sim iidU(0, 1)$ tiene fines didácticos, ya que bajo este escenario conocemos los valores futuros de la variable umbral y el valor del parámetro de umbral γ . Las probabilidades de mala clasificación son cero, es decir, generamos los pronósticos h -pasos para el modelo TAR(2; 1, 1) bajo el verdadero régimen. Evaluamos los desempeños de los pronósticos como función de los valores de los parámetros en el modelo o equivalentemente en cuanto a los primeros dos momentos de la variable δ_t , los ECMP del modelo AR(1) son mayores o menores que los ECMP en el modelo TAR(2; 1, 1). Recordemos que bajo el supuesto de que el coeficiente autorregresivo del modelo AR(1) es igual al $E(\delta_t)$, obtuvimos expresiones de los ECMP, las cuales están dadas como función de los momentos de δ_t .

La Tabla 5 presenta los promedios de los 500 valores estimados en cada una de las series simuladas, por máxima verosimilitud, mientras que la Tabla 6 se refiere a los estimadores

obtenidos mediante regresión por cuantiles para $\theta = 0.5$. Recordemos que independiente del VPGD, ajustamos a cada una de las 500 series simuladas dos modelos candidatos, el modelo TAR(2; 1, 1) y el modelo AR(1), contamos entonces con 500 estimaciones para cada uno de los parámetros bajo cada modelo. Los valores $\hat{\phi}$ y $\hat{\sigma}_{AR}^2$ corresponden entonces a los estimadores de los parámetros en el modelo AR(1). Mientras que $\hat{\phi}_1$, $\hat{\phi}_2$ y $\hat{\sigma}_{TAR}^2$ corresponden a los estimadores de los parámetros bajo el modelo TAR(2; 1, 1), donde $E(\hat{\delta}_t)$ y $E(\hat{\delta}_t^2)$ se calculan sustituyendo los valores estimados de los parámetros en las fórmulas respectivas.

Tabla 5. Promedios de los estimadores por MV, bajo los modelos TAR(2; 1, 1) y AR(1).

Identificación	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\sigma}_{TAR}^2$	$E(\hat{\delta}_t)$	$E(\hat{\delta}_t^2)$	$\hat{\phi}$	$\hat{\sigma}_{AR}^2$
caso 1-1	-0.101	0.989	0.993	0.443	0.498	0.431	1.587
caso 1-2	-0.097	0.994	0.995	0.176	0.260	0.168	1.294
caso 1-3	0.995	-0.100	0.983	0.721	0.747	0.702	1.882
caso 2-1	0.991	0.097	0.995	0.544	0.500	0.536	1.399
caso 2-2	0.995	0.094	0.989	0.770	0.748	0.760	1.612
caso 2-3	0.103	0.986	0.986	0.324	0.257	0.323	1.186
caso 3-1	-0.797	-0.600	0.989	-0.699	0.502	-0.698	1.013
caso 4-1	-0.796	0.600	0.995	-0.098	0.500	-0.098	1.987
caso 4-2	-0.796	0.595	0.997	-0.448	0.568	-0.448	1.854
caso 4-3	0.594	-0.789	0.986	0.248	0.425	0.249	1.614
caso 5-1	-0.396	-0.591	0.993	-0.493	0.258	-0.493	1.009
caso 6-1	-0.401	0.601	0.987	0.091	0.267	0.098	1.333
caso 6-2	-0.396	0.601	0.991	-0.146	0.214	-0.144	1.234
caso 6-3	0.593	-0.402	0.993	0.344	0.309	0.345	1.264
caso 7-1	-0.401	0.394	0.988	-0.003	0.165	-0.001	1.181
caso 7-2	-0.398	0.404	0.988	-0.198	0.167	-0.197	1.137
caso 7-3	0.392	-0.398	0.985	0.194	0.162	0.194	1.128
caso 8-1	-0.940	0.943	0.985	0.001	0.887	0.001	9.623
caso 8-2	-0.943	0.939	0.995	-0.472	0.888	-0.473	7.618
caso 8-3	0.942	-0.939	0.997	0.472	0.888	0.471	7.606
caso 9-1	-0.894	0.991	0.988	0.049	0.891	0.045	10.11
caso 9-2	-0.895	0.996	0.988	-0.422	0.850	-0.431	5.668
caso 9-3	0.992	-0.891	0.989	0.050	0.889	0.047	9.867
caso 10-1	0.892	0.990	0.982	0.941	0.889	0.940	1.009
caso 11-1	0.889	0.697	0.988	0.793	0.641	0.793	1.018

Una primera observación en la Tabla 5 es que en todos los casos analizados el estimador promedio del coeficiente autorregresivo $\hat{\phi}$ es aproximadamente igual al valor $E(\hat{\delta}_t)$. Además tenemos que al alejarnos de la condición de estacionariedad en el modelo TAR(2; 1, 1), como

función de $E(\delta_t^2)$, la varianza estimada para el error en el modelo AR(1) es grande en comparación con el valor verdadero, σ^2 . Esperamos entonces que los ECMP bajo el modelo AR(1) tiendan a ser más grandes que los ECMP bajo el modelo TAR(2; 1, 1). Si las varianzas estimadas bajo el modelo AR(1) son cercanas al valor de σ^2 verdadero, esperamos que los desempeños de los pronósticos bajo ambos modelos sea similar.

Tabla 6. Promedio de las estimaciones por CR bajo los modelos TAR(2; 1, 1) y AR(1).

Identificación	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\sigma}_{TAR}^2$	$E(\hat{\delta}_t)$	$E(\hat{\delta}_t^2)$	$\hat{\phi}$	$\hat{\sigma}_{AR}^2$
caso 1-1	-0.100	0.987	0.998	0.443	0.499	0.438	1.597
caso 1-2	-0.095	1.001	0.999	0.179	0.266	0.108	1.304
caso 1-3	0.994	-0.099	0.987	0.721	0.747	0.836	1.964
caso 2-1	0.990	0.095	0.999	0.545	0.505	0.540	1.408
caso 2-2	0.994	0.090	0.994	0.768	0.747	0.854	1.656
caso 2-3	0.101	0.988	0.990	0.322	0.261	0.283	1.193
caso 3-1	-0.796	-0.603	0.994	-0.699	0.505	-0.698	1.015
caso 4-1	-0.797	0.597	0.999	-0.100	0.502	-0.105	2.003
caso 4-2	-0.796	0.597	1.001	-0.448	0.570	-0.598	1.916
caso 4-3	0.595	-0.792	0.990	0.248	0.429	0.372	1.647
caso 5-1	-0.393	-0.592	0.997	-0.492	0.262	-0.492	1.012
caso 6-1	-0.400	0.603	0.992	0.101	0.271	0.096	1.337
caso 6-2	-0.399	0.595	0.996	-0.150	0.218	-0.190	1.240
caso 6-3	0.590	-0.399	0.998	0.343	0.309	0.391	1.271
caso 7-1	-0.396	0.397	0.993	0.0009	0.168	0.0008	1.185
caso 7-2	-0.398	0.409	0.992	-0.196	0.172	-0.219	1.141
caso 7-3	0.393	-0.403	0.989	0.194	0.167	0.218	1.132
caso 8-1	-0.943	0.943	0.989	-0.0001	0.890	-0.0019	10.87
caso 8-2	-0.942	0.941	0.999	-0.471	0.888	-0.825	8.973
caso 8-3	0.942	-0.942	1.001	0.471	0.887	0.825	8.966
caso 9-1	-0.895	0.991	0.993	0.048	0.893	0.043	11.51
caso 9-2	-0.895	0.995	0.993	-0.422	0.849	-0.762	6.521
caso 9-3	0.991	-0.893	0.994	0.048	0.891	0.058	11.176
caso 10-1	0.892	0.989	0.986	0.941	0.889	0.939	1.012
caso 11-1	0.889	0.696	0.993	0.793	0.643	0.795	1.020

Como mencionamos arriba, la Tabla 6 contiene los promedios de las estimaciones para cada uno de los parámetros en los modelos TAR(2; 1, 1) y AR(1) respectivamente, bajo el método de regresión por cuantiles con $\theta = 0.5$. Los valores promedios de las estimaciones para cada uno de los parámetros en ambos modelos bajo MV y CR son numéricamente muy cercanos, sin embargo hemos observado que cuando el valor de $E(\delta_t^2)$ es cercano a

uno los valores de los coeficientes autorregresivos, estimados por CR presentan un comportamiento alejado del supuesto de normalidad en los errores pero en promedio el estimador $\hat{\phi}$ es aproximadamente iguala a $E(\hat{\delta}_t)$.

Al comparar las Tablas 5 y 6 con los verdaderos valores de los parámetros bajo cada escenario simulado de la Tabla 4, observamos que dichos valores son prácticamente lo mismo. Un análisis gráfico de los estimadores para cada uno de los escenarios analizados complementaría el análisis de los comportamientos de los estimadores en los modelos de interés. Por ejemplo, podemos graficar los 500 valores estimados en cada uno de los escenarios simulados, y para cada uno de los parámetros estimados para analizar el comportamiento distribucional de dichos estimadores; recordemos que teóricamente la distribución asintótica de los coeficientes autorregresivos es gaussiana. Sin embargo, el número de escenarios simulados es grande e incluir los resultados para cada uno de ellos implicaría una gran cantidad de información. Mostraremos entonces los resultados completos para algunos de los casos identificados en la Tabla 4, los cuales contemplan escenarios muy distintos entre sí, como función de los valores de los parámetros, así como características interesantes en relación al modelo supuesto. Cabe mencionar que para el resto de los escenarios simulados podemos extrapolar los resultados encontrados.

Resultados para el caso 1-1

Recordemos que los valores de los parámetros identificados como caso 1-1 son: $\phi_1 = -0.1$, $\phi_2 = 1$, $\sigma^2 = 1$, $\Pr(Z_{t-1} \leq \gamma) = \gamma = 0.5$, $E(\delta_t) = 0.45$ y $E(\delta_t^2) = 0.505$, con valores estimados bajo el modelo TAR(2; 1, 1) y AR(1) vía MV y RC, dados en la tabla siguiente:

Método	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\sigma}_{TAR}^2$	$E(\hat{\delta}_t)$	$E(\hat{\delta}_t^2)$	$\hat{\phi}$	$\hat{\sigma}_{AR}^2$
MV	-0.101	0.989	0.993	0.443	0.498	0.431	1.587
RC	-0.100	0.987	0.998	0.443	0.499	0.438	1.597

Este caso corresponde al llamado modelo TUR introducido en González y Gonzalo (1998). La Figura 15 presenta el comportamiento gráfico para nueve posibles realizaciones del modelo TAR(2; 1, 1) bajo este escenario.

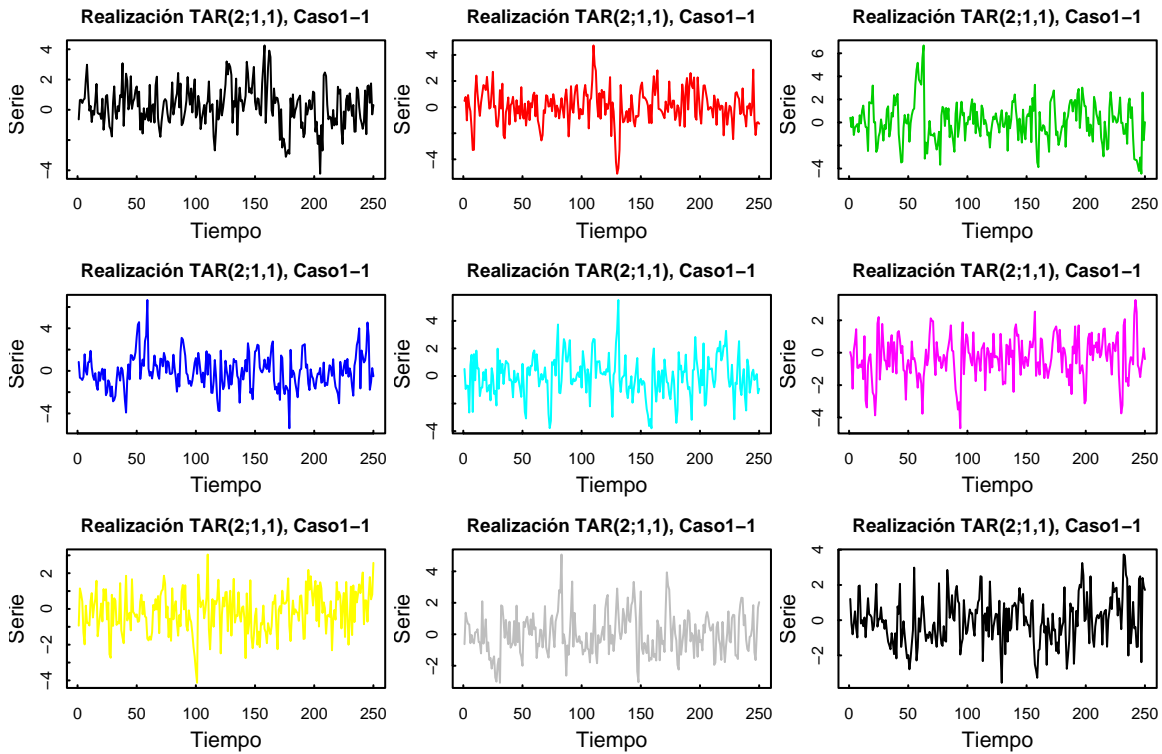


Figura 15. Realizaciones del modelo TAR(2; 1, 1) para el caso 1-1 con $Z_{t-1} \sim iidU(0, 1)$.

La Figura 16 presenta los histogramas de las 500 estimaciones de los parámetros para los modelos TAR(2; 1, 1) y AR(1), el primer renglón de histogramas corresponden a las estimaciones vía MV y el segundo renglón de gráficas a los histogramas de las estimaciones vía RC.

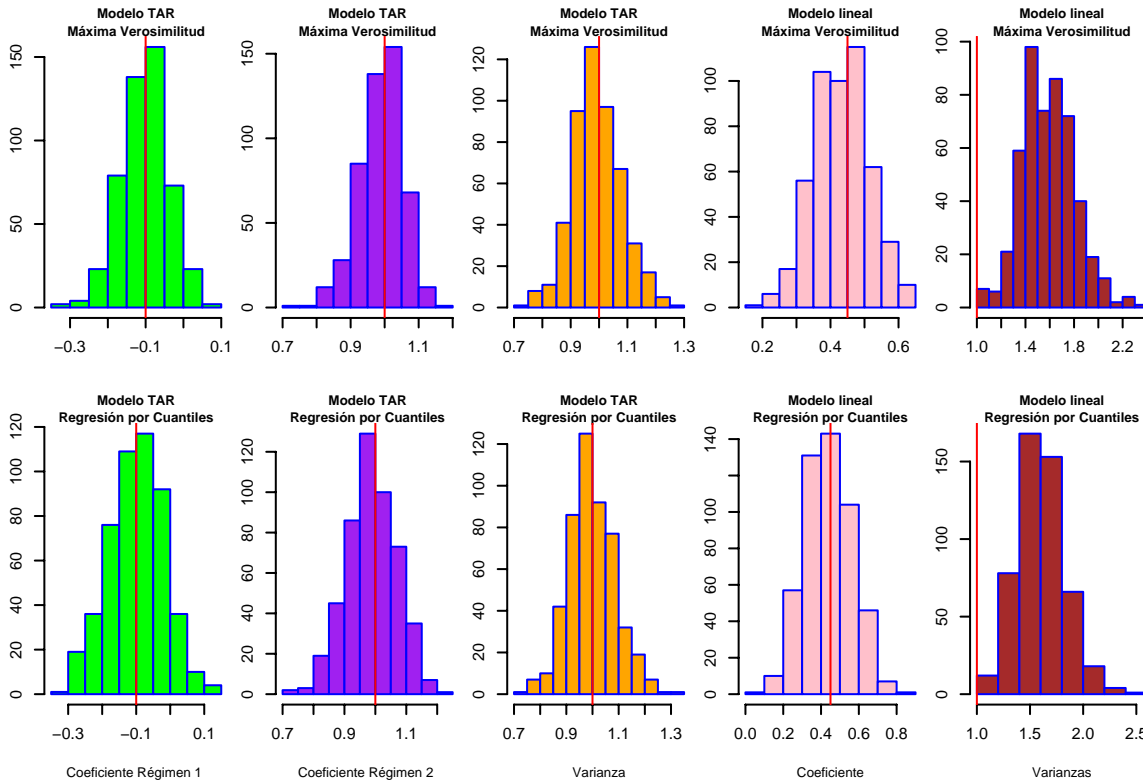


Figura 16. Histogramas de los parámetros estimados, bajo los modelos TAR(2; 1, 1) y AR(1), caso 1-1 con $Z_{t-1} \sim iidU(0, 1)$.

De la Figura 16 observamos que las varianzas estimadas para el modelo AR(1) considerando cada método de estimación son mayores que las varianzas estimadas para el modelo TAR(2, 1, 1) y de aquí que los ECMP en el modelo TAR(2; 1, 1), para cada uno de los métodos de pronóstico son menores que los ECMP en el modelo AR(1). Las líneas verticales en cada uno de los histogramas asociados al modelo TAR corresponden a los verdaderos valores de los parámetros, mientras que para el histograma correspondiente a los valores estimados del coeficiente AR la línea vertical indica el valor verdadero de $E(\delta_t)$.

Para enriquecer la observación anterior graficamos los errores de pronóstico, bajo cada alternativa de obtención en cada modelo candidato y para cada horizonte de pronóstico. La Figura 17 contiene las gráficas de cajas y bigotes para los errores de pronóstico, de donde observamos que la variabilidad de los errores de pronóstico para el caso lineal son más grandes que en el caso TAR(2; 1, 1), note además que los errores de pronósticos bajo cada modelo e independientemente del mecanismo generador de los pronósticos presentan prácticamente las mismas variabilidades para cada horizonte de pronóstico. El título en cada gráfica identifica el modelo y el método empleado para la obtención del pronóstico: TAR de modelo TAR(2; 1, 1); AR de modelo AR(1); y R, RC y VP, de método recursivo, método de regresión por cuantiles y verosimilitud predictiva respectivamente. Es decir, TAR_R

significa que los errores de pronóstico en la gráfica están asociados al modelo TAR y que los pronósticos h -pasos adelante los obtuvimos por medio del método recursivo.

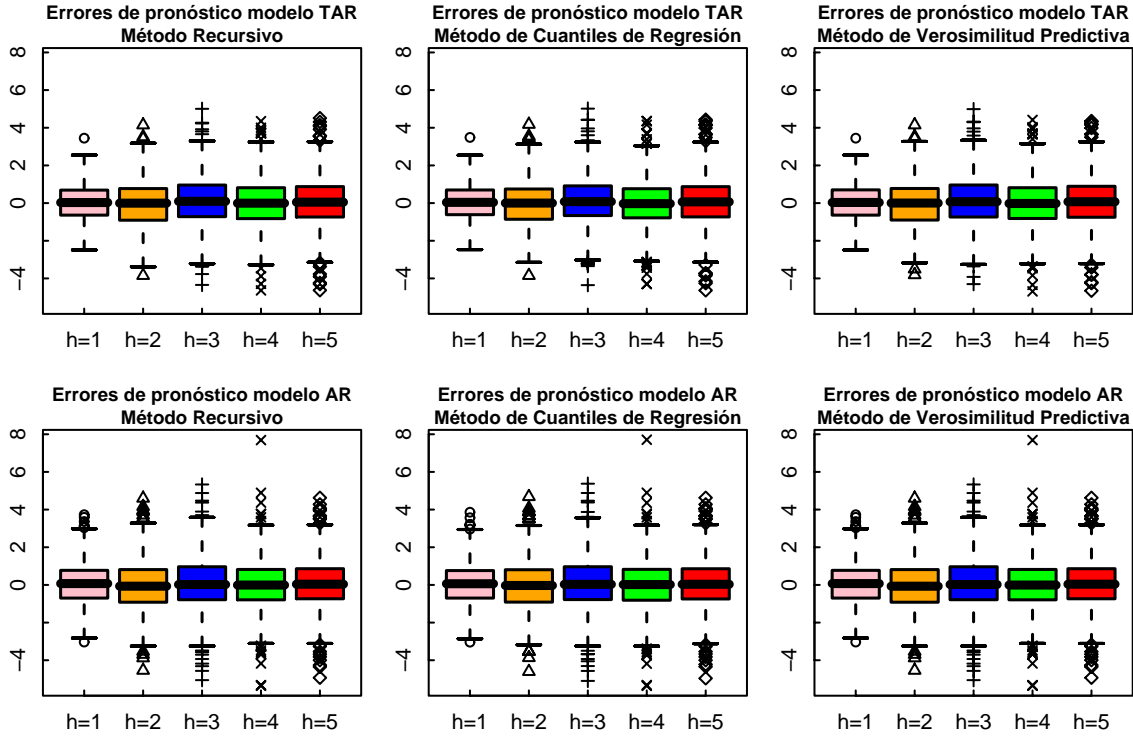


Figura 17. Errores de pronóstico bajo cada alternativa de pronóstico para el caso 1-1 con $Z_{t-1} \sim iidU(0, 1)$.

La Figura 18 contiene los comportamientos gráficos de los promedios de las funciones de pérdida ECMP y función cuantil o check, a manera de complementar las observaciones discutidas arriba. De nuevo los títulos en las gráficas identifican el modelo bajo estudio y el método utilizado para generar los pronósticos h -pasos hacia adelante. Igualmente en el eje horizontal se encuentra la identificación de la función de pérdida ECMP o pérdida de regresión por cuantiles. La tercer gráfica dentro de la Figura 18 corresponde al comportamiento promedio de los pronósticos para cada alternativa de pronóstico y bajo cada modelo, así como los promedios de los últimos cinco valores simulados para cada una de las 500 series los cuales no se utilizaron en la fase de estimación y que identificamos como valores “reales”.

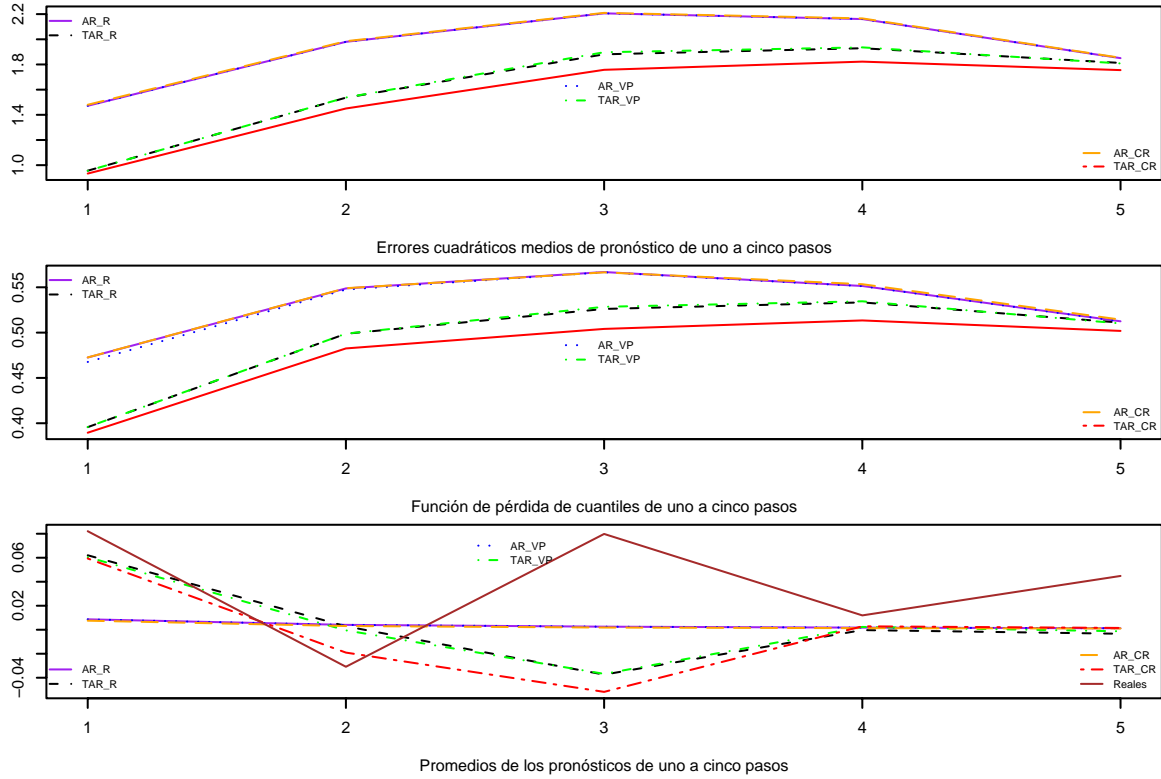


Figura 18. Funciones de pérdida para cada horizonte de pronóstico, caso 1-1 con $Z_{t-1} \sim iidU(0, 1)$.

Observemos de la Figura 18 que los valores de las funciones de pérdida ECMP y pérdida de cuantiles en el modelo $TAR(2; 1, 1)$ son menores que los valores de las funciones de pérdida ECMP y pérdida de cuantiles en el modelo $AR(1)$, independientemente del proceso de obtención de los pronósticos. En el caso lineal $AR(1)$, los valores de los ECMP bajo las distintas alternativas de pronósticos son prácticamente los mismos e igualmente para la función de pérdida de cuantiles. Para el modelo $TAR(2; 1, 1)$ tenemos que los ECMP son prácticamente los mismos valores para los pronósticos obtenidos vía VP y el método R, mientras que para la función de pérdida de cuantiles tenemos que los comportamientos para los pronósticos por VP son iguales que los pronósticos bajo el método recursivo.

La Tabla 7 contiene los cocientes de los valores para las funciones de pérdida ECMP y función cuantil en el modelo $AR(1)$, entre los valores respectivos en el modelo $TAR(1)$, para el caso 1-1 con $\sigma^2 = 1$. Mét. Rec. equivale a método recursivo; Mét. VP. a método de verosimilitud predictiva y Mét. CR. a método de regresión por cuantiles. Los números entre corchetes son los cocientes de los ECMP para cada uno de los métodos de pronósticos y distintos valores de h ; mientras que los números entre paréntesis corresponden a los cocientes bajo la función de pérdida de regresión por cuantiles.

Tabla 7. Cocientes de las funciones de pérdida, caso 1-1.

h	Mét. Rec.	Mét. VP.	Mét. CR.
1	[1.54] (1.19)	[1.54] (1.18)	[1.59] (1.21)
2	[1.29] (1.10)	[1.29] (1.10)	[1.37] (1.13)
3	[1.17] (1.07)	[1.16] (1.07)	[1.26] (1.12)
4	[1.12] (1.03)	[1.12] (1.03)	[1.19] (1.07)
5	[1.02] (1.00)	[1.02] (1.00)	[1.06] (1.02)

Los resultados de la Tabla 7, muestran que independientemente del método para generar los pronósticos h -pasos adelante, los pronósticos provenientes del modelo TAR(2; 1, 1) presentan mayor ganancia en cuanto a las funciones de pérdida consideradas.

Para completar el análisis del escenario uno en la Tabla 4, presentamos las Tablas 8 y 9 que contienen los cocientes de las funciones de pérdida para los casos 1-2 y 1-3, lo que cambia en relación al caso 1-1 es que la proporción de observaciones en el régimen uno es del 75% y 25% respectivamente. Como arriba los valores entre corchetes corresponden a los cocientes de los ECMP y los valores entre paréntesis a la función de pérdida de cuantiles.

Tabla 8. Cocientes de las funciones de pérdida, caso 1-2
75% de observaciones en el régimen 1.

h	Mét. Rec.	Mét. VP	Mét. CR
1	[1.29] (1.12)	[1.30] (1.12)	[1.30] (1.11)
2	[1.02] (1.01)	[1.03] (1.01)	[1.06] (1.03)
3	[0.98] (1.00)	[0.98] (1.00)	[1.03] (1.01)
4	[0.98] (1.00)	[0.98] (1.00)	[1.01] (1.01)
5	[1.00] (1.00)	[1.00] (1.00)	[1.03] (1.01)

Tabla 9. Cocientes de las funciones de pérdida, caso 1-3
25% de observaciones en el régimen 1.

h	Mét. Rec.	Mét. VP	Mét. CR
1	[1.61] (1.24)	[1.62] (1.21)	[1.66] (1.23)
2	[1.52] (1.21)	[1.52] (1.22)	[1.83] (1.34)
3	[1.32] (1.14)	[1.32] (1.15)	[1.62] (1.30)
4	[1.32] (1.14)	[1.30] (1.13)	[1.59] (1.28)
5	[1.23] (1.08)	[1.23] (1.09)	[1.45] (1.21)

Como podemos observar de las Tablas 8 y 9, cuando el porcentaje de observaciones en el régimen de raíz unitaria es del 75% (Tabla 9) la ganancia de los pronósticos h -pasos adelante e independientemente del método de pronóstico y de la función de pérdida es mucho mayor que para el caso en que el porcentaje de observaciones en el régimen de raíz unitaria es del 25% (Tabla 8).

Resultados para el caso 3-1

A continuación presentamos los resultados para el caso 3-1; bajo este escenario de simulación resulta que los coeficientes autorregresivos en cada régimen son muy cercanos en magnitud y del mismo signo lo que equivale a pensar que el modelo $TAR(2; 1, 1)$ simulado, sin considerar el significado de γ , tiende a comportarse como un modelo AR. Los valores de los parámetros son: $\phi_1 = -0.8$, $\phi_2 = -0.6$, $\sigma^2 = 1$, $\Pr(Z_{t-1} \leq \gamma) = \gamma = 0.5$, $E(\delta_t) = -0.7$ y $E(\delta_t^2) = 0.5$, mientras que los valores estimados promedios bajo MV y RC son:

Método	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\sigma}_{TAR}^2$	$E(\hat{\delta}_t)$	$E(\hat{\delta}_t^2)$	$\hat{\phi}$	$\hat{\sigma}_{AR}^2$
MV	-0.797	-0.600	0.989	-0.699	0.502	-0.698	1.013
RC	-0.796	-0.603	0.994	-0.699	0.505	-0.698	1.015

Notemos que las estimaciones de las varianzas bajo el modelo $AR(1)$, $\hat{\sigma}_{AR}^2$, son muy cercanas al valor utilizado para el ejercicio de simulación, además de que $\hat{\phi} \approx E(\hat{\delta}_t)$, estamos también ante un modelo $TAR(2; 1, 1)$ bien comportado en el sentido de que nos alejamos de no estacionariedad.

La Figura 19 presenta el comportamiento gráfico de algunas realizaciones del modelo $TAR(2; 1, 1)$ bajo los valores verdaderos de los parámetros.

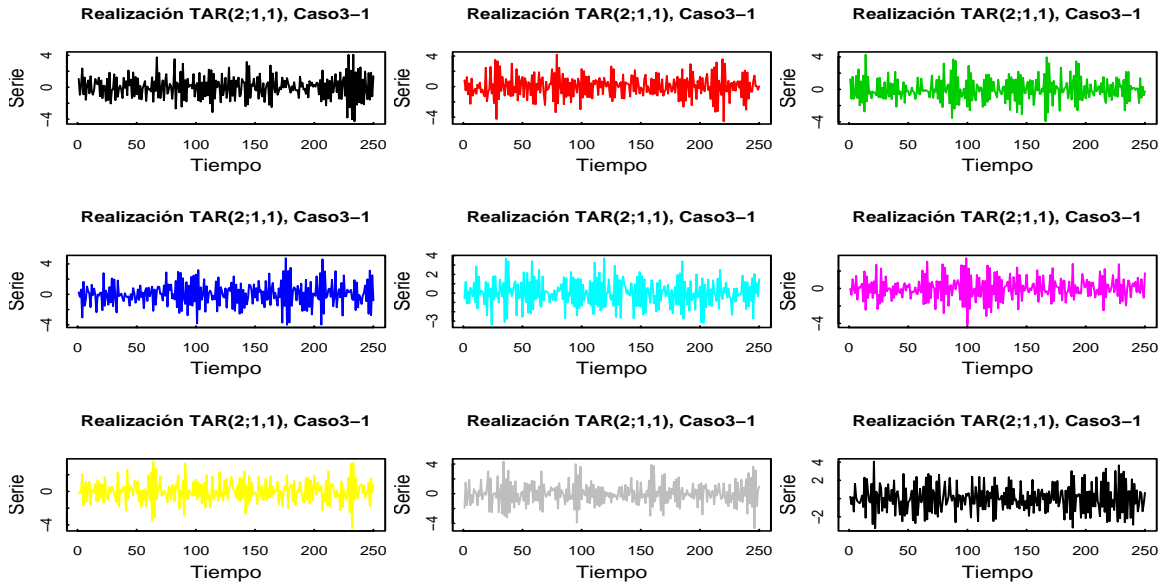


Figura 19. Realizaciones del modelo $TAR(2; 1, 1)$ para el caso 3-1 con $Z_{t-1} \sim iidU(0, 1)$.

La Figura 20 contiene los histogramas de los parámetros estimados bajo MV y RC para los modelos TAR(2;1,1) y AR(1), note que a diferencia del caso 1 las varianzas estimadas por ambos métodos para los errores bajo el modelo AR(1) son muy parecidas a las varianzas estimadas bajo el modelo TAR(2; 1, 1) y que éstas están alrededor del valor verdadero $\sigma^2 = 1$. Como antes los títulos en las gráficas identifican el método de estimación y el parámetro estimado bajo el modelo correspondiente.

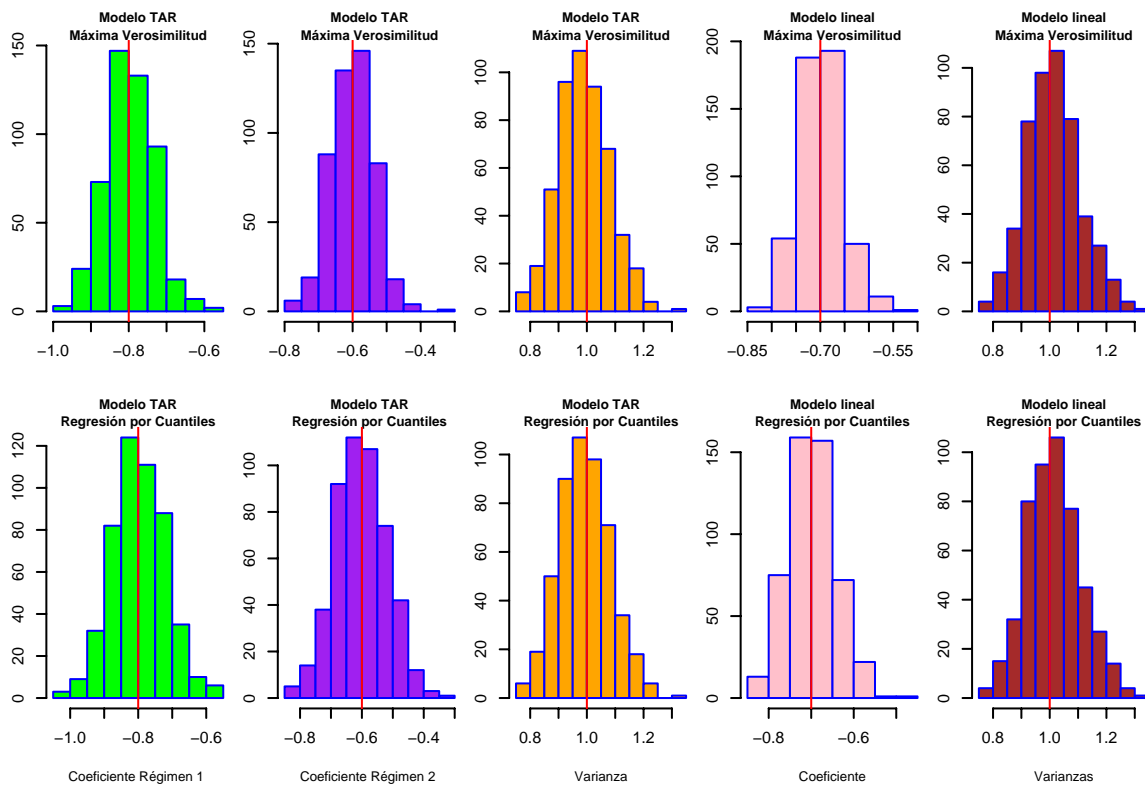


Figura 20. Histogramas de los parámetros estimados, bajo los modelos TAR(2; 1, 1) y AR(1), caso 3-1 con $Z_{t-1} \sim iidU(0, 1)$

El comportamiento esperado de los errores de pronóstico bajo este escenario es similar bajo los dos modelos analizados, tal y como se muestra en la Figura 21, que contiene las gráficas de cajas y bigotes para los pronósticos h -pasos adelante para cada alternativa de generación y bajo cada modelo analizado.

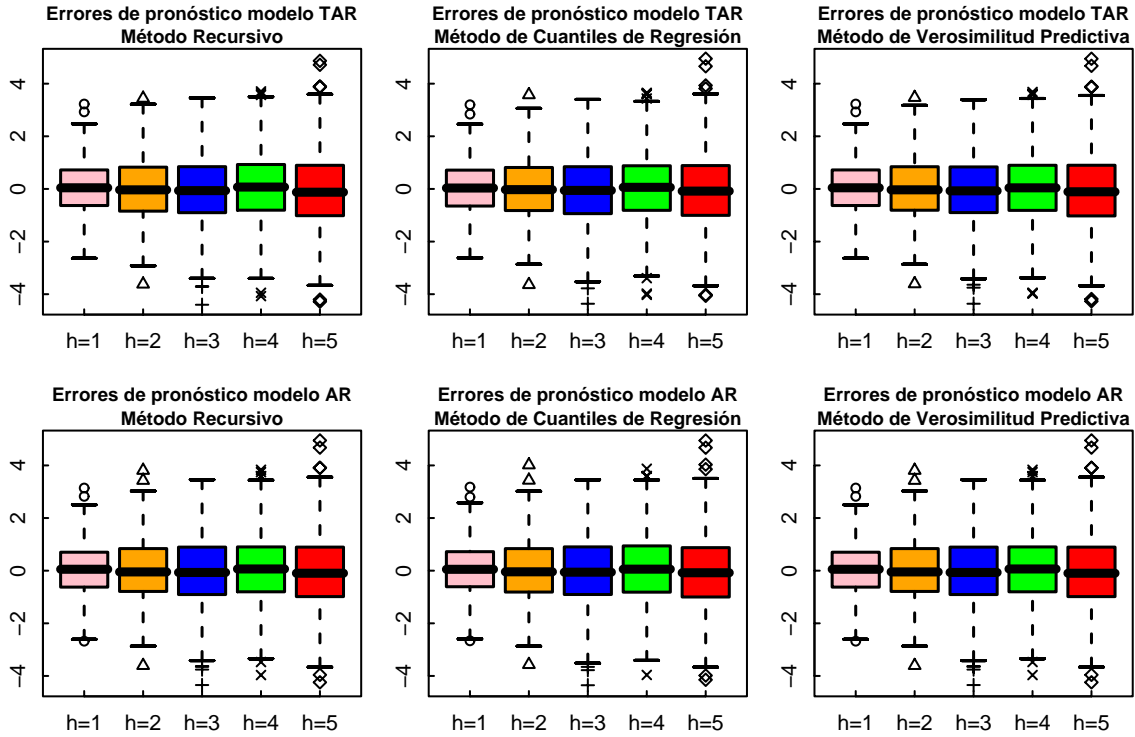


Figura 21. Errores de pronóstico bajo cada alternativa de pronóstico para el caso 3-1 con $Z_{t-1} \sim iidU(0, 1)$.

Los valores de las funciones de pérdida promedio se encuentran en la Figura 22 y como podemos observar bajo pérdida ECMP los modelos $AR(1)$ y $TAR(2; 1, 1)$ independientemente del proceso de obtención de los pronósticos tienen el mismo desempeño. A diferencia, en el caso de la función de pérdida de regresión por cuantiles se observa que la pérdida bajo el modelo $AR(1)$ obteniendo los pronósticos con VP son más pequeños en magnitud que el resto de las alternativas y que la pérdida asociada para los pronósticos por VP para el modelo $TAR(2; 1, 1)$.

La Tabla 10 contiene los cocientes de los valores para las funciones de pérdida ECMP y función cuantil en el modelo $AR(1)$, entre los valores respectivos en el modelo $TAR(1)$, para el caso 3-1 con $\sigma^2 = 1$. Observemos que los valores de los ECMP para el modelo $TAR(2; 1, 1)$ en general son menores que para el modelo $AR(1)$; sin embargo esta ganancia es mínima, y podemos pensar que los desempeños de los pronósticos bajo ambos modelos son equivalentes a reserva de que la interpretación práctica del parámetro de umbral γ en el modelo $TAR(2; 1, 1)$.

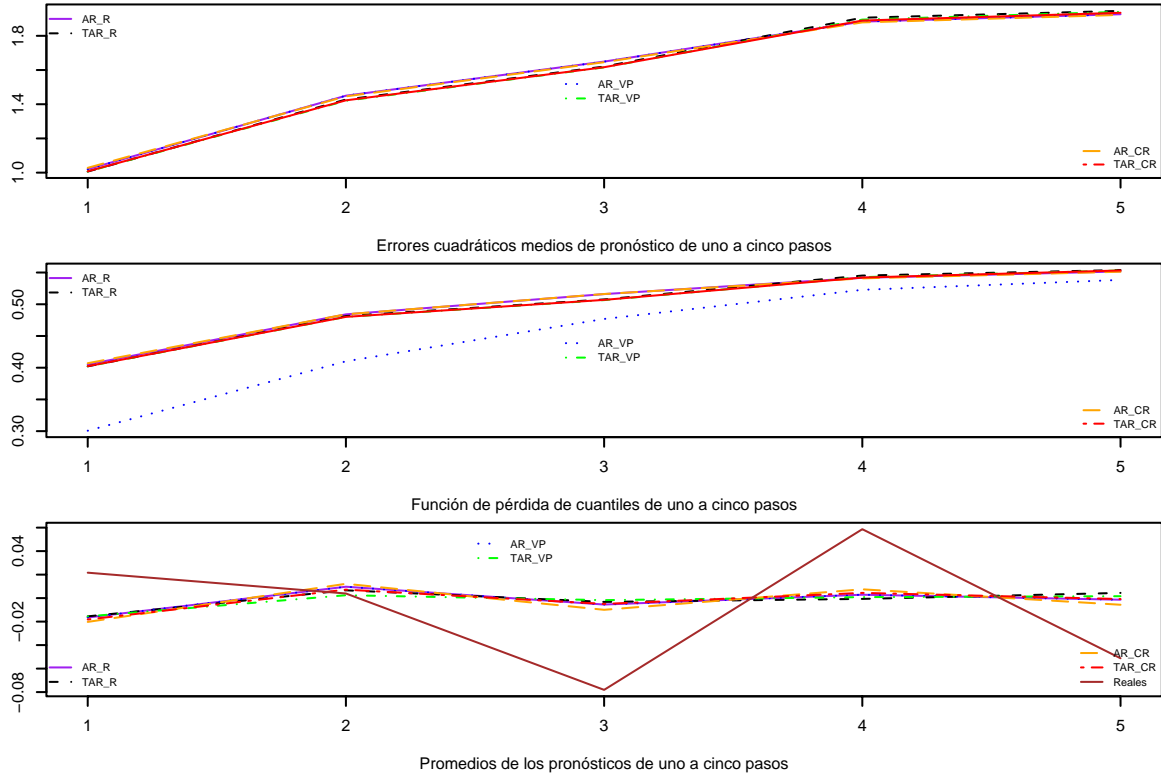


Figura 22. Funciones de pérdida para cada horizonte de pronóstico, caso 3-1 con $Z_{t-1} \sim iidU(0, 1)$.

Tabla 10. Cocientes de las funciones de pérdida, caso 3-1.

h	Mét. Rec.	Mét. VP	Mét. CR
1	[1.01] (1.01)	[1.01] (0.75)	[1.02] (1.01)
2	[1.01] (1.02)	[1.02] (0.86)	[1.02] (1.01)
3	[1.02] (1.02)	[1.02] (0.94)	[1.02] (1.01)
4	[0.99] (1.00)	[1.00] (0.96)	[1.00] (1.00)
5	[0.99] (1.00)	[1.00] (0.97)	[1.00] (1.00)

A diferencia del caso 1-1 recordemos que bajo este escenario los momentos de la variable δ_t son esencialmente iguales independientemente del porcentaje de observaciones en cada uno de los regímenes y de aquí que sólo mostramos los resultados para el caso de igual dominancia en el régimen.

Resultados para el caso 8-1

Bajo este escenario tenemos que los valores de los parámetros son: $\phi_1 = -0.95$, $\phi_2 = 0.95$, $\sigma^2 = 1$, $\Pr(Z_{t-1} \leq \gamma) = \gamma = 0.5$, $E(\delta_t) = 0$ y $E(\delta_t^2) = 0.903$, con valores estimados bajo el modelo TAR(2; 1, 1) y AR(1) vía MV y RC respectivamente, dados en la tabla siguiente:

Método	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\sigma}_{TAR}^2$	$E(\hat{\delta}_t)$	$E(\hat{\delta}_t^2)$	$\hat{\phi}$	$\hat{\sigma}_{AR}^2$
MV	-0.940	0.943	0.985	0.001	0.887	0.001	9.623
RC	-0.943	0.943	0.989	-0.0001	0.890	-0.0019	10.87

Las características relevantes de este caso son que el coeficiente autorregresivo en cada uno de los regímenes en el modelo TAR(2; 1, 1) son cercanos al caso de raíz de unitaria, sin embargo globalmente el modelo es estacionario, aunque cabe mencionar que el valor de la condición de estacionariedad de segundo orden, $E(\delta_t^2) = 0.903$ es cercano a uno y por ende la variabilidad asociada al proceso presenta comportamientos de crecimiento y decrecimiento tal y como se observa en las gráficas en la Figura 23 que contiene algunas realizaciones del proceso en cuestión.

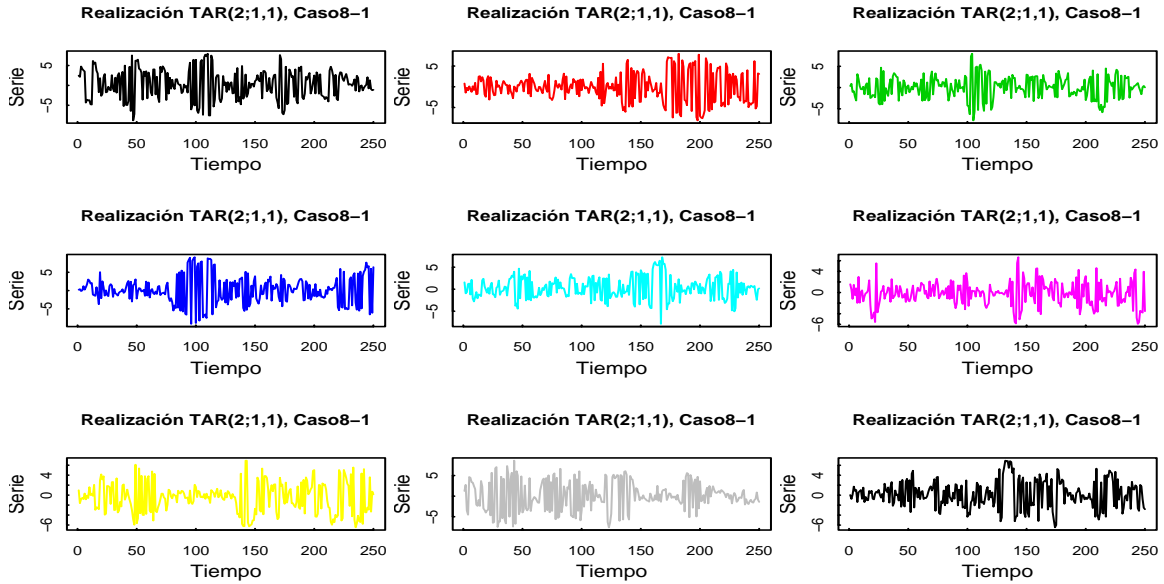


Figura 23. Realizaciones del modelo TAR(2; 1, 1) para el caso 8-1 con $Z_{t-1} \sim iidU(0, 1)$.

Observemos también que el valor promedio de los coeficientes autorregresivos en el modelo AR(1) son cercanos a $E(\delta_t) = 0$, pero la estimación promedio de la varianza asociada es muy grande comparada con el valor verdadero, $\sigma^2 = 1$. El valor grande de la varianza estimada en el modelo AR(1) se debe a que bajo este escenario la mala especificación al asumir un modelo lineal, paga el precio de aumento en la varianza estimada y de que no se cumpla el supuesto de ruido blanco para los residuales.

La Figura 24 muestra los histogramas para los 500 valores estimados para cada uno de los parámetros de interés. Como comentamos arriba, observe que los valores de las varianzas estimadas para el modelo AR(1) son valores grandes.

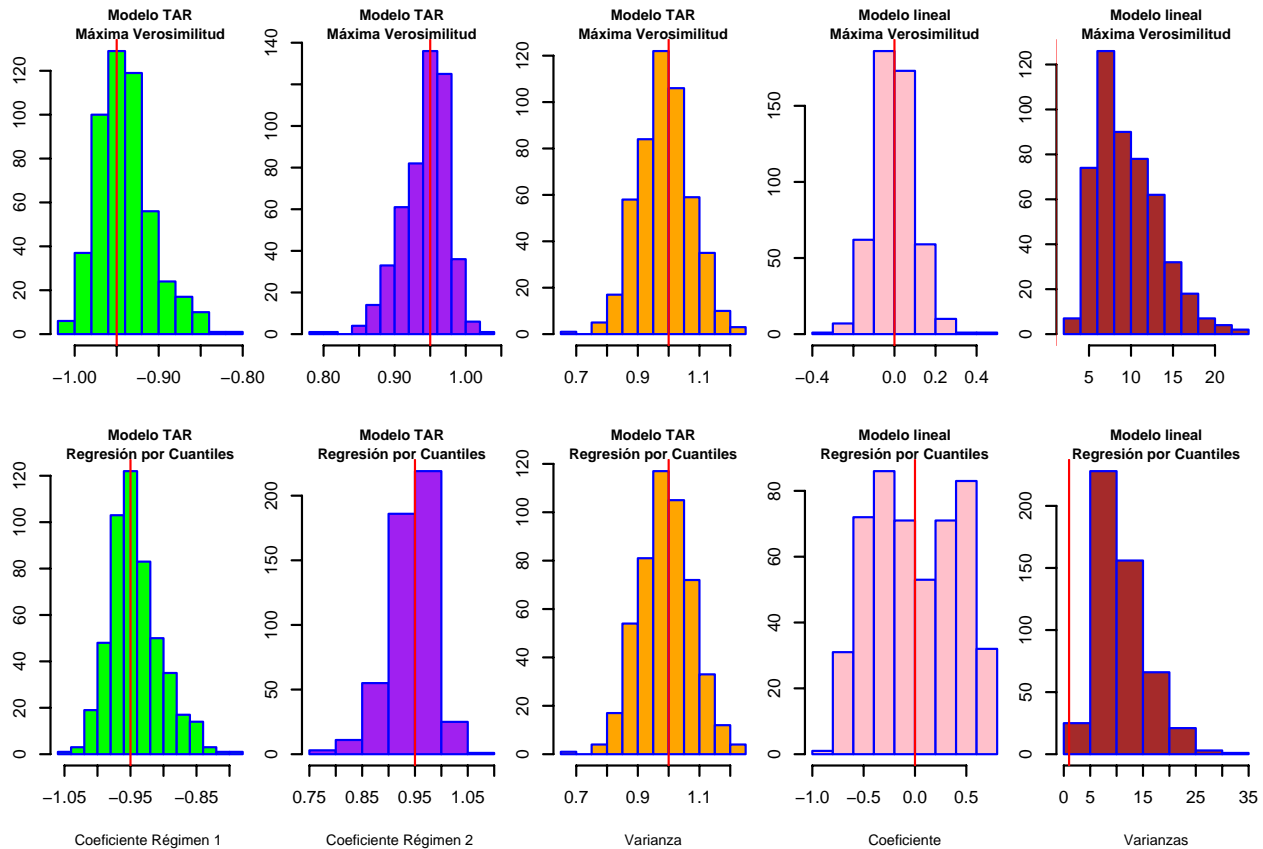


Figura 24. Histogramas de los parámetros estimados, bajo los modelos TAR(2; 1, 1) y AR(1), caso 8-1 con $Z_{t-1} \sim iidU(0, 1)$

Los errores de pronósticos bajo los modelos TAR(2; 1, 1) y AR(1), para cada uno de los métodos estudiados se presentan en la Figura 25. Observemos que la variabilidad asociada a los errores de pronósticos h -pasos adelante para los pronósticos bajo el modelo TAR(2; 1, 1) son mucho menores que bajo el modelo AR(1), independientemente del método de pronóstico utilizado y del horizonte de pronóstico.

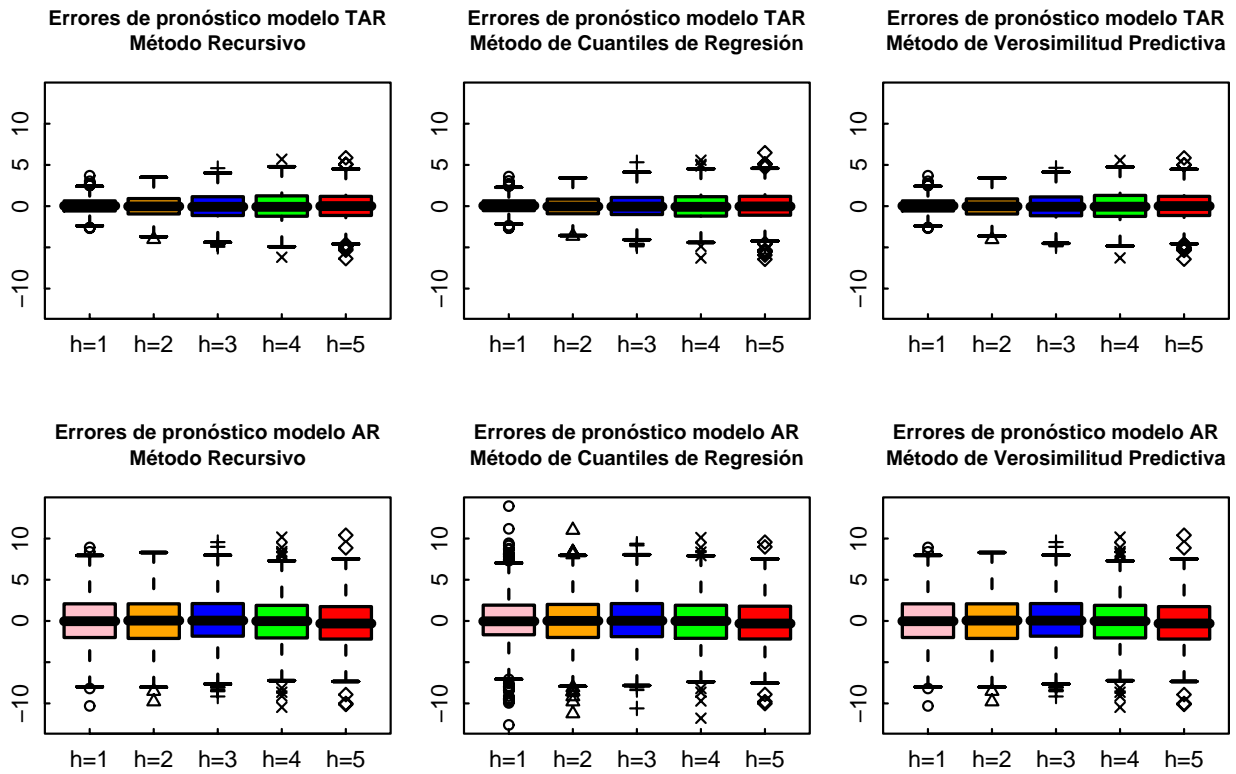


Figura 25. Errores de pronóstico bajo cada alternativa de pronóstico para el caso 8-1, con $Z_{t-1} \sim iidU(0, 1)$.

A partir de las gráficas de cajas y bigotes de la Figura 25 queda claro que los valores de las funciones de pérdida para el modelo TAR(2; 1, 1) serán menores que los respectivos valores de las funciones de pérdida bajo el modelo AR(1). La Figura 26 contiene los comportamientos de la funciones de pérdida para cada uno de los horizontes de pronóstico, observe también que el comportamiento promedio de los pronósticos bajo cada modelo y cada método de pronóstico es cercano al comportamiento del promedio de los llamados valores “reales”.

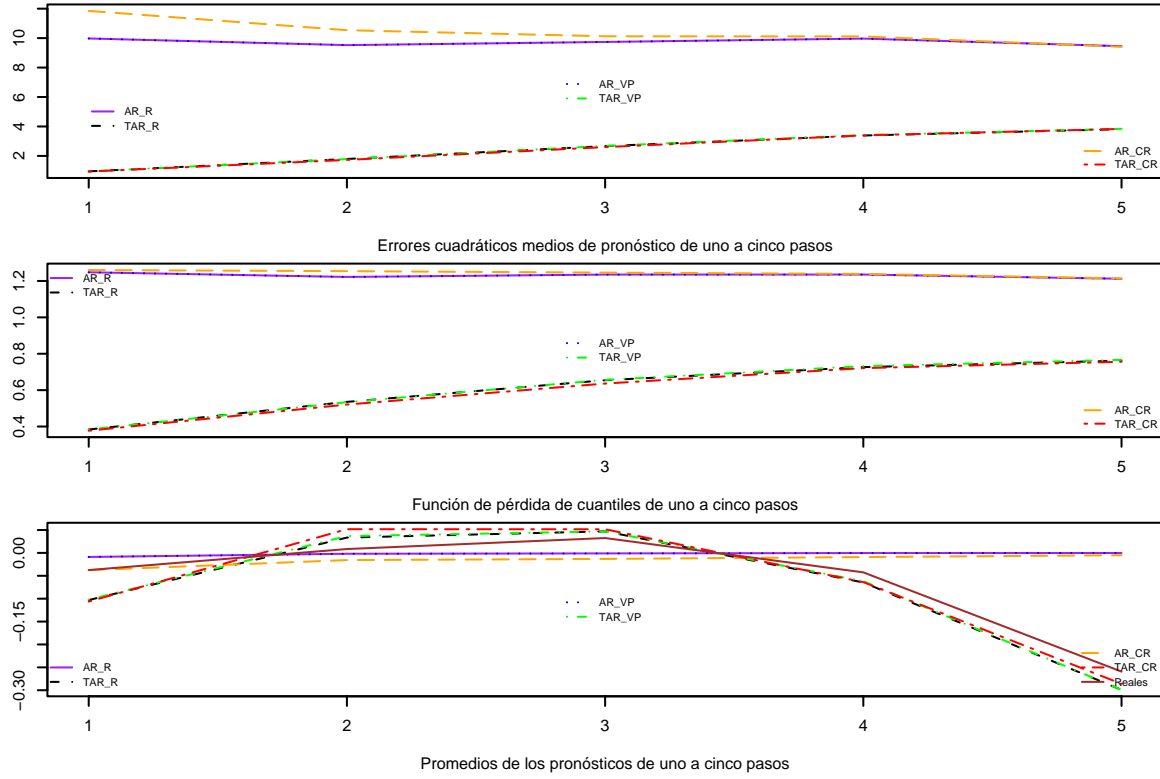


Figura 26. Funciones de pérdida para cada horizonte de pronóstico, caso 8-1 con $Z_{t-1} \sim iidU(0, 1)$.

Bajo este escenario tenemos una separación grande en magnitud entre los valores de las funciones de pérdida bajo el modelo TAR(2; 1, 1) en relación a las funciones de pérdida bajo el modelo AR(1), como se mostró en la figura 26 y como podemos observar en la Tabla 11, donde la ganancia de las funciones de pérdida ECMP y pérdida de cuantiles en el modelo TAR(2; 1, 1) es grande. El desempeño de los pronósticos bajo el modelo TAR(2; 1, 1) es mucho mejor que el desempeño de los pronósticos bajo el modelo lineal AR(1) para cada uno de los métodos de obtención de pronósticos.

Tabla 11. Cocientes de las funciones de pérdida, caso 8-1.

h	Mét. Rec.	Mét. VP	Mét. CR
1	[10.48] (3.25)	[10.47] (3.25)	[12.62] (3.34)
2	[5.31] (2.28)	[5.33] (2.29)	[6.09] (2.40)
3	[3.65] (1.88)	[3.64] (1.88)	[3.89] (1.95)
4	[2.94] (1.70)	[2.92] (1.69)	[2.98] (1.71)
5	[2.47] (1.60)	[2.46] (1.58)	[2.45] (1.60)

Para complementar el análisis del caso 8, presentamos los cocientes de los valores de las funciones de pérdida cuando el porcentaje de observaciones en el régimen con coeficiente negativo es del 75%, identificado en la Tabla 4 como caso 8-2, e igualmente para el caso en que el porcentaje de observaciones en el régimen con coeficiente positivo es del 75% identificado en la Tabla 4 como caso 8-3. Los resultados se encuentran en las Tablas 12 y 13 respectivamente y la conclusión es al igual que en el caso 8-1, los pronósticos bajo el modelo TAR(2; 1, 1) presentan mejor desempeño que los pronósticos bajo el modelo AR(1) debido que los primeros presentan valores menores de las respectivas funciones de pérdida.

Tabla 12. Cocientes de las funciones de pérdida, caso 8-2
75% de observaciones en el régimen 1 ($\phi_1 = -0.95$).

h	Mét. Rec.	Mét. VP	Mét. CR
1	[9.11] (2.72)	[9.12] (2.67)	[11.40] (2.50)
2	[5.26] (2.27)	[5.26] (2.26)	[6.98] (2.30)
3	[3.615] (1.90)	[3.62] (1.90)	[4.72] (2.02)
4	[2.71] (1.64)	[2.73] (1.64)	[3.40] (1.76)
5	[2.17] (1.50)	[2.20] (1.50)	[2.68] (1.61)

Tabla 13. Cocientes de las funciones de pérdida, caso 8-3
75% de observaciones en el régimen 1 ($\phi_1 = 0.95$)

h	Mét. Rec.	Mét. VP	Mét. CR
1	[6.41] (2.35)	[6.38] (2.34)	[6.83] (2.02)
2	[4.16] (2.03)	[4.14] (2.01)	[4.98] (1.96)
3	[2.93] (1.70)	[2.92] (1.70)	[3.48] (1.71)
4	[2.57] (1.62)	[2.57] (1.62)	[2.87] (1.65)
5	[2.24] (1.51)	[2.24] (1.50)	[2.44] (1.51)

7.2.2 Resultados asumiendo que Z_{t-1} sigue un proceso AR(1)

Recordemos que las alternativas de pronósticos identificadas como método recursivo y método de regresión por cuantiles, consideran que conocemos los valores futuros de la variable de umbral Z_{t-1} , bajo el supuesto de que la estructura de dependencia de Z_{t-1} está dada por un proceso AR(1) estacionario con ruido blanco gaussiano. Pronosticamos el comportamiento futuro de Z_{t-1} a partir de las ecuaciones recursivas o por medio de verosimilitud predictiva, ya que ambos procedimientos son equivalentes bajo el supuesto de normalidad. La opción de pronosticar la función indicadora consiste en considerar a dicha variable como una serie de tiempo binaria, para γ fijo y suponemos el siguiente modelo para $I(Z_{t-1} \leq \gamma)$, $t = 1, 2, \dots, T$, con T el número total de observaciones en la muestra, en nuestro caso $T = 250$,

$$\pi_t(\boldsymbol{\alpha}) \equiv \Pr(I(Z_{t-1} \leq \gamma) = 1 \mid \mathfrak{F}_{t-1}) = F_l(\boldsymbol{\alpha}'\mathbf{B}_{t-1}) = \frac{1}{1 + \exp[-\boldsymbol{\alpha}'\mathbf{B}_{t-1}]},$$

consideramos que la matriz de covariables $\mathbf{B}_{t-1} = (\mathbf{1}, I(Z_{t-2} \leq \gamma), I(Z_{t-3} \leq \gamma))'$ está formada por un vector de unos, más los vectores resultantes de considerar los valores de la serie binaria para uno y dos rezagos. A partir de dicha matriz de covariables, estimamos el vector de parámetros $\boldsymbol{\alpha}$, por medio de las correspondientes ecuaciones de verosimilitud parcial, propuestas por Kedem y Fokianos (2002), posteriormente obtenemos los valores más plausibles en la verosimilitud parcial, para la variable indicadora bajo $\boldsymbol{\alpha}$ conocido.

El método de verosimilitud predictiva para el modelo TAR(2; 1, 1) descrito en la sección correspondiente, involucra conjuntamente al modelo de TAR(2; 1, 1) para Y_t y el modelo para la variable umbral y de aquí que al momento de obtener los pronósticos bajo esta alternativa, obtenemos los pronósticos conjuntos para Y_{T+h} y Z_{T+h-1} con $h = 1, 2, 3, 4, 5$, bajo el criterio de que los valores encontrados son los valores más plausibles con respecto a la verosimilitud predictiva para el modelo. La ventaja de verosimilitud predictiva sobre los métodos recursivo y de regresión por cuantiles, radica precisamente en el hecho de pronosticar de manera conjunta la variable umbral y la serie TAR.

Recordemos que si $Z_{t-1} \sim iidU(0, 1)$, el porcentaje de observaciones en el régimen uno es igual al valor del parámetro γ . Para el caso en que $Z_{t-1} \sim AR(1)$ sigue un proceso autorregresivo estacionario con errores ruido blanco gaussiano, es decir, $Z_{t-1} = \rho Z_{t-2} + u_t$ con $|\rho| < 1$ y $u_t \sim iidN(0, \sigma_u^2)$ tenemos que los valores del parámetro dependerán de los porcentajes de observaciones postulados para cada uno de los regímenes, por ejemplo para garantizar una proporción de p observaciones en el régimen uno y $(1 - p)$ en el régimen dos, tenemos que $\Pr(Z_{t-1} \leq \gamma) = p$, entonces $\gamma = \Phi^{-1}(p)$, donde $Z_{t-1} \sim N(0, \sigma_u^2/(1 - \rho^2))$.

La primera parte del análisis consiste en obtener los estimadores de los parámetros para los dos modelos considerados, por medio de los métodos de máxima verosimilitud y regresión por cuantiles. La Tabla 14 muestra los promedios de los 500 estimadores calculados por máxima verosimilitud para cada uno de los parámetros involucrados bajo el modelo correspondiente y bajo cada escenario descrito en la Tabla 4. El mismo comentario se tiene para

los valores en la Tabla 15, sólo que ahora los valores estimados se obtienen por medio del método de regresión por cuantiles. Similar al caso en que $Z_{t-1} \sim iidU(0, 1)$ podemos concluir de manera general que ambos procedimientos de estimación presentan comportamientos adecuados como función de las propiedades teóricas asintóticas.

Tabla 14. Promedios de las estimaciones por MV bajo los modelos TAR(2; 1, 1) y AR(1).

Identificación	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\sigma}_{TAR}^2$	$E(\hat{\delta}_t)$	$E(\hat{\delta}_t^2)$	$\hat{\phi}_{AR}$	$\hat{\sigma}_{AR}^2$
caso 1-1	-0.096	0.99	0.994	0.447	0.496	0.54	1.71
caso 1-2	-0.099	0.989	0.984	0.173	0.252	0.239	1.366
caso 1-3	0.989	-0.096	0.995	0.717	0.736	0.712	1.888
caso 2-1	0.988	0.101	0.988	0.545	0.493	0.631	1.467
caso 2-2	0.992	0.104	0.991	0.770	0.740	0.816	1.632
caso 2-3	0.096	0.983	0.990	0.318	0.248	0.380	1.253
caso 3-1	-0.786	-0.594	0.996	-0.690	0.486	-0.697	1.019
caso 4-1	-0.791	0.597	0.994	-0.097	0.491	-0.138	1.973
caso 4-2	-0.79	0.596	0.991	-0.446	0.561	-0.471	1.792
caso 4-3	0.595	-0.792	0.989	0.248	0.422	0.227	1.658
caso 5-1	-0.400	-0.591	0.990	-0.495	0.255	-0.499	1.006
caso 6-1	-0.399	0.591	0.995	0.095	0.254	0.114	1.333
caso 6-2	-0.396	0.581	0.993	-0.151	0.202	-0.133	1.236
caso 6-3	0.589	-0.400	0.993	0.342	0.300	0.352	1.255
caso 7-1	-0.403	0.396	0.986	-0.003	0.160	-0.002	1.181
caso 7-2	-0.398	0.392	0.988	-0.200	0.157	-0.202	1.130
caso 7-3	0.394	-0.381	0.988	0.200	0.153	0.201	1.125
caso 8-1	-0.941	0.943	0.990	0.0008	0.888	-0.0006	9.664
caso 8-2	-0.941	0.941	0.989	-0.470	0.885	-0.468	7.528
caso 8-3	0.941	-0.942	0.988	-0.0005	0.887	0.006	9.762
caso 9-1	-0.893	0.992	0.990	0.0495	0.891	0.096	10.576
caso 9-2	-0.894	0.988	0.994	-0.423	0.843	-0.3871	5.981
caso 9-3	0.993	-0.894	0.992	0.521	0.939	0.541	14.43
caso 10-1	0.893	0.993	0.997	0.943	0.892	0.945	1.027
caso 11-1	0.895	0.692	0.989	0.794	0.640	0.802	1.023

Tabla 15. Promedios de las estimaciones por CR bajo los modelos TAR(2; 1, 1) y AR(1).

Identificación	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\sigma}_{TAR}^2$	$E(\hat{\delta}_t)$	$E(\hat{\delta}_t^2)$	$\hat{\phi}_{AR}$	$\hat{\sigma}_{AR}^2$
caso 1-1	-0.094	0.992	0.998	0.448	0.496	0.59	1.74
caso 1-2	-0.09	0.989	0.988	0.173	0.252	0.163	1.382
caso 1-3	0.988	-0.094	0.999	0.717	0.735	0.825	1.954
caso 2-1	0.988	0.102	0.993	0.545	0.494	0.663	1.479
caso 2-2	0.990	0.103	0.995	0.768	0.738	0.896	1.672
caso 2-3	0.095	0.991	0.995	0.319	0.252	0.334	1.263
caso 3-1	-0.788	-0.591	1.001	-0.690	0.486	-0.696	1.022
caso 4-1	-0.794	0.594	0.998	-0.100	0.491	-0.156	1.994
caso 4-2	-0.793	0.597	0.995	-0.445	0.561	-0.602	1.840
caso 4-3	0.592	-0.789	0.993	0.247	0.418	0.357	1.696
caso 5-1	-0.399	-0.590	0.994	-0.495	0.254	-0.501	1.008
caso 6-1	-0.400	0.591	0.999	0.095	0.255	0.113	1.338
caso 6-2	-0.394	0.582	0.997	-0.150	0.201	-0.178	1.243
caso 6-3	0.587	-0.399	0.997	0.340	0.298	0.401	1.262
caso 7-1	-0.402	0.401	0.991	-0.0007	0.161	0.003	1.185
caso 7-2	-0.399	0.390	0.992	-0.201	0.157	-0.225	1.135
caso 7-3	0.390	-0.382	0.992	0.197	0.150	0.222	1.129
caso 8-1	-0.942	0.944	0.995	0.0009	0.889	-0.0071	11.175
caso 8-2	-0.942	0.9392	0.993	-0.472	0.885	-0.819	8.9163
caso 8-3	0.941	-0.942	0.992	-0.0005	0.887	0.031	11.296
caso 9-1	-0.892	0.992	0.995	0.049	0.890	0.184	12.490
caso 9-2	-0.895	0.989	0.999	-0.424	0.846	-0.742	7.046
caso 9-3	0.993	-0.894	0.997	0.521	0.940	0.908	17.455
caso 10-1	0.892	0.993	1.001	0.943	0.891	0.945	1.030
caso 11-1	0.893	0.691	0.994	0.792	0.637	0.801	1.026

Al igual que en el caso de $Z_{t-1} \sim iidU(0, 1)$ mostramos algunos resultados gráficos, así como los cocientes de las funciones de pérdida para el modelo AR(1) entre las funciones de pérdida para el modelo TAR(2; 1, 1), para los casos 1-1, 3-1 y 8-1 respectivamente.

Resultados para el caso 1-1

Recordemos que los valores de los parámetros identificados como caso 1-1 son: $\phi_1 = -0.1$, $\phi_2 = 1$, $\Pr(Z_{t-1} \leq \gamma) = 0.5$, $E(\delta_t) = 0.45$ y $E(\delta_t^2) = 0.505$, con valores estimados bajo el modelo TAR(2; 1, 1) y AR(1) vía MV y RC, dados en la Tabla siguiente:

Método	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\sigma}_{TAR}^2$	$E(\hat{\delta}_t)$	$E(\hat{\delta}_t^2)$	$\hat{\phi}$	$\hat{\sigma}_{AR}^2$
MV	-0.096	0.99	0.994	0.447	0.496	0.54	1.71
RC	-0.094	0.992	0.998	0.448	0.496	0.59	1.74

Observemos de la tabla anterior que los valores de $\hat{\phi}$ son mayores que los valores de $E(\hat{\delta}_t)$. La Figura 27 presenta distintas realizaciones del proceso bajo los valores verdaderos de los parámetros.

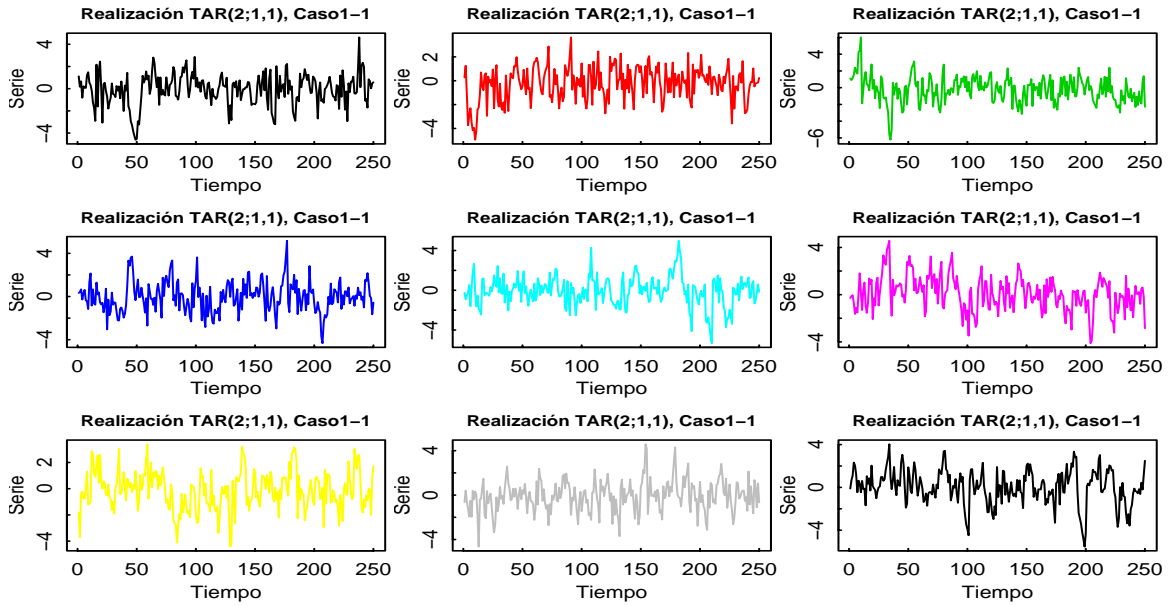


Figura 27. Realizaciones del modelo TAR(2; 1, 1) para el caso 1-1 con $Z_{t-1} \sim \text{AR}(1)$.

El comportamiento gráfico de los estimadores correspondientes al caso 1-1 se muestra en la Figura 28. Las líneas verticales, en el caso de los histogramas asociados al modelo TAR(2; 1, 1) que corresponden a los primeros tres histogramas en cada renglón de la figura, indican los verdaderos valores de los parámetros bajo este escenario. Para los histogramas

correspondientes a los valores estimados bajo el modelo lineal AR(1), la línea vertical indica el valor de $E(\delta_t)$, mientras que dicha línea en los histogramas para las varianzas estimadas bajo el modelo lineal corresponde a la varianza del error considerada en el ejercicio de simulación, $\sigma^2 = 1$.

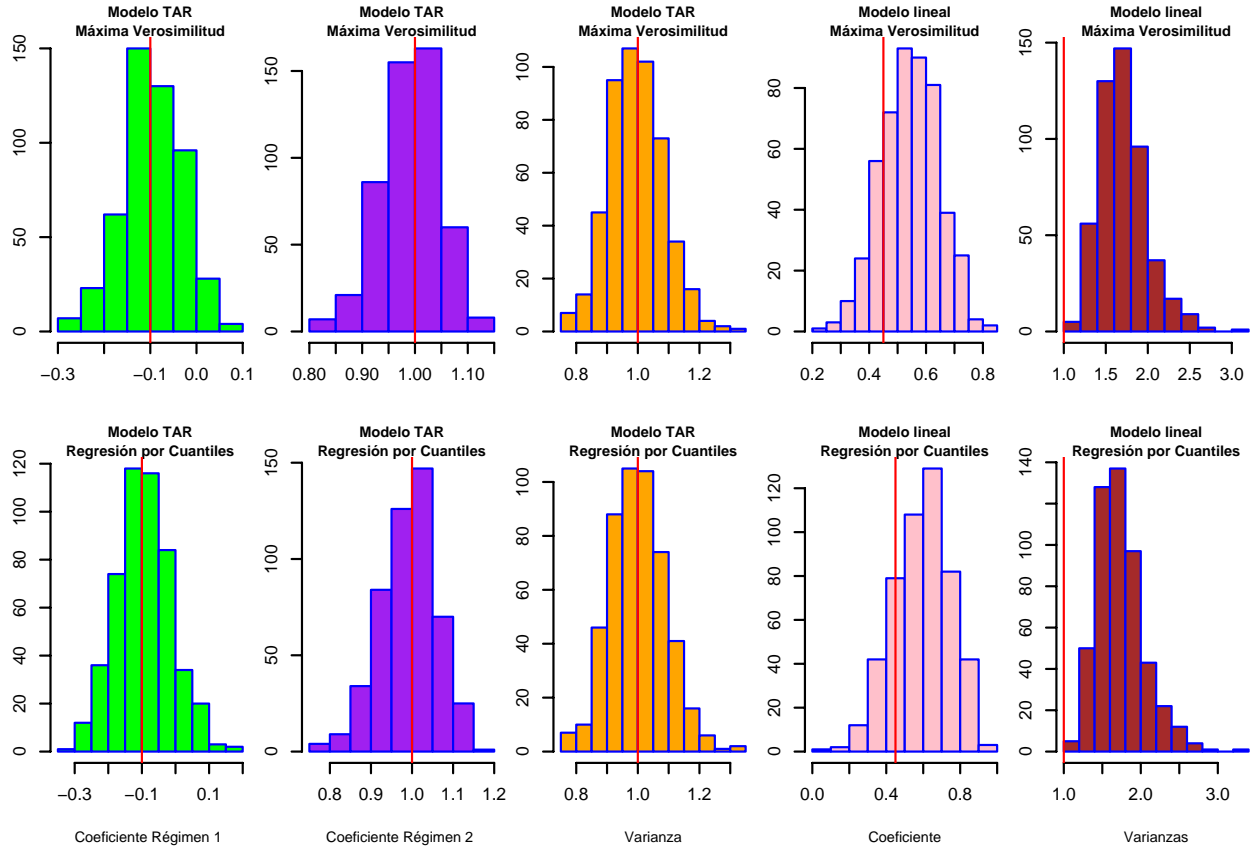


Figura 28. Histogramas de los parámetros estimados, bajo los modelos TAR(2; 1, 1) y AR(1), caso 1-1 con $Z_{t-1} \sim \text{AR}(1)$

Los comportamientos de las estimaciones para los parámetros en el caso TAR(2; 1, 1), para ambos procedimientos de estimación, se distribuyen alrededor del verdadero valor. A diferencia del caso $Z_{t-1} \sim \text{iid}U(0, 1)$ observamos que la mayor parte de los valores de las estimaciones del coeficiente autorregresivo están por arriba del $E(\hat{\delta}_t)$ mientras que los valores de las estimaciones de las varianzas son mucho mayores que el valor verdadero.

La Figura 29 presenta la variabilidad de los errores de pronósticos bajo cada alternativa de pronóstico, para los modelos TAR(2; 1, 1) y AR(1) para los pronósticos $h = 1, 2, 3, 4, 5$ -pasos hacia adelante. A partir de esta gráfica podemos inferir que bajo verosimilitud predictiva los ECMP presentarán un buen comportamiento en el sentido de ECMP más pequeño que los respectivos ECMP bajo las otras alternativas.

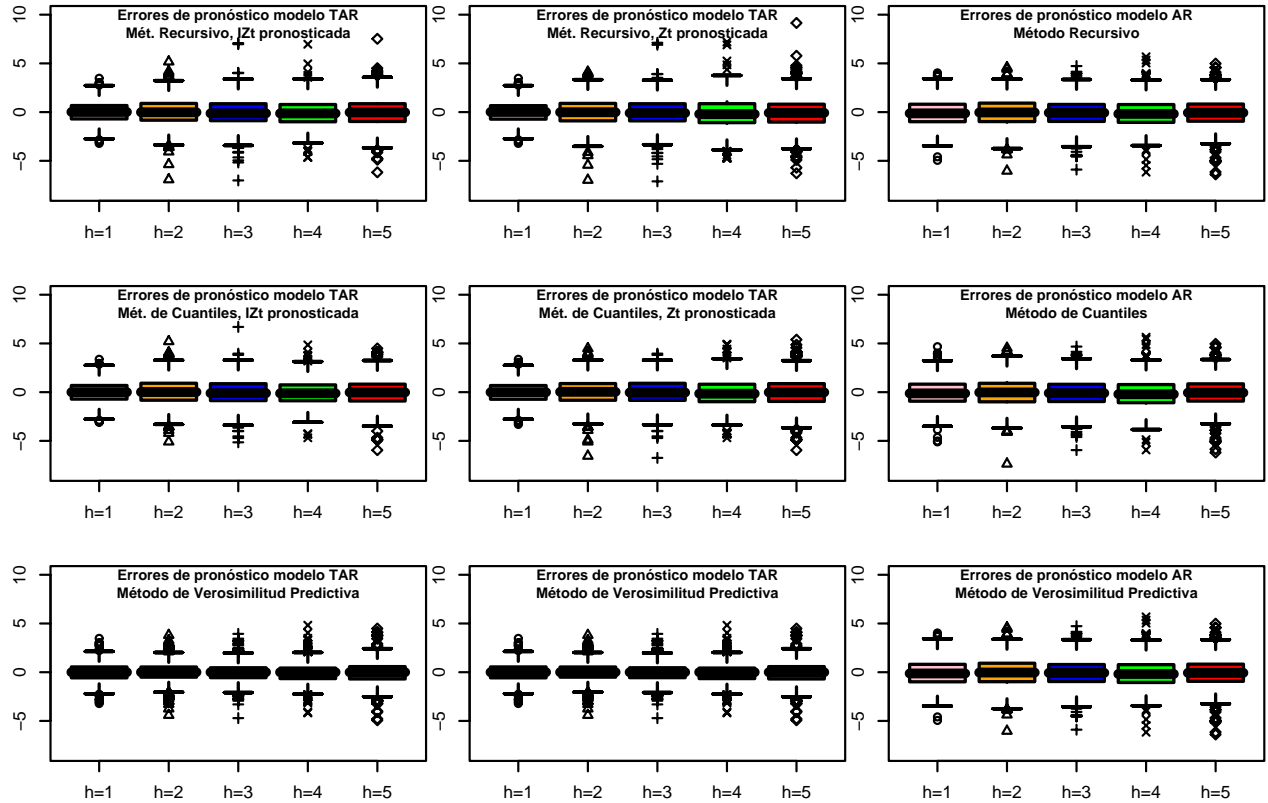


Figura 29. Errores de pronóstico bajo cada alternativa de pronóstico para el caso 1-1, con $Z_{t-1} \sim \text{AR}(1)$.

En la Figura 30 presentamos los promedios de los ECMP, de la función de pérdida de cuantiles y de los pronósticos h -pasos adelante para cada una de las alternativas de pronóstico bajo los modelos TAR(2; 1, 1) y AR(1). La línea verde punteada identificada con la etiqueta TAR_VP está asociada al modelo TAR(2; 1, 1) con pronósticos generados por verosimilitud predictiva. Observamos que para ambas funciones de pérdida, la ganancia del método de verosimilitud predictiva es clara, siguiéndole el método de cuantiles bajo los pronósticos de la variable indicadora identificado por la línea continua verde fuerte y identificada con la etiqueta TAR_CRIZt. La tercera gráfica en la Figura 30 corresponde a los pronósticos

promedios para cada alternativa y cada método de pronóstico, la línea rosa identifica a los promedios de los últimos cinco valores bajo las series simuladas, es decir, promediamos los 500 valores simulados al tiempo $T + 1, T + 2, \dots, T + 5$.

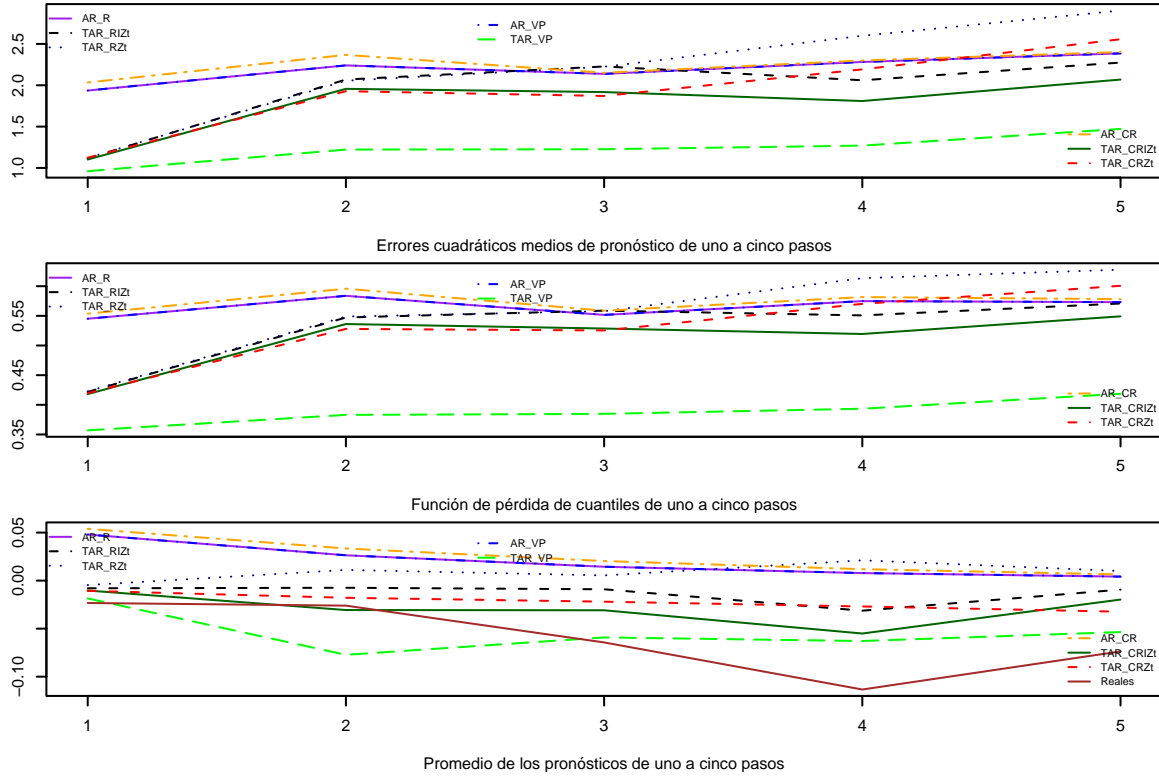


Figura 30. Funciones de pérdida para cada horizonte de pronóstico, caso 1-1 con $Z_{t-1} \sim \text{AR}(1)$.

La Tabla 16 se lee como sigue: Mét. Rec., Método Vero. Pred y Mét. CR. significa que el método usado para la obtención de los pronósticos h -pasos adelante corresponden al método recursivo, método de verosimilitud predictiva y método de cuantiles. $\tilde{I}(Z_{t-1} \leq \gamma)$ y $I(\tilde{Z}_{t-1} \leq \gamma)$ significan que dentro de los métodos recursivo y de regresión por cuantiles se utilizan los pronósticos de $I(Z_{t-1} \leq \gamma)$ o de Z_{t-1} respectivamente. Los números entre corchetes corresponden a los cocientes de los ECMP para el modelo $\text{AR}(1)$ entre los ECMP para el modelo $\text{TAR}(2; 1, 1)$, mientras que los números entre paréntesis corresponden al cociente de la función de pérdida “check” en el modelo $\text{AR}(1)$ entre la función de pérdida “check” en el modelo $\text{TAR}(2; 1, 1)$. Los porcentajes para cada horizonte de pronósticos corresponden a la proporción de veces en que el régimen se identificó correctamente, es decir, dado que conocemos los valores de la variable umbral al tiempo $T, T + 1, \dots, T + h - 1$ podemos obtener los valores de la variable indicadora para los distintos horizontes de pronóstico, bajo γ conocido, luego comparamos estos valores con los correspondientes valores al

considerar los pronósticos $I(\tilde{Z}_{t-1} \leq \gamma)$ y $\tilde{I}(Z_{t-1} \leq \gamma)$. Observe que en los métodos recursivo y de regresión por cuantiles los porcentajes de correcta identificación del régimen, columnas 2 y 5 de la Tabla 16 toman los mismos valores, igualmente para las columnas 3 y 6, lo cual se debe a que los pronósticos para la variable umbral o para la variable indicadora en ambos métodos son los mismos.

Tabla 16. Cocientes de las funciones de pérdida, y porcentajes de veces en que se identifica correctamente al régimen bajo cada esquema de obtención de los pronósticos, caso 1-1.

Horiz. pronós.	Mét. Rec. $\tilde{I}(Z_{t-1} \leq \gamma)$	Mét. Rec. $I(\tilde{Z}_{t-1} \leq \gamma)$	Método Vero. Pred.	Mét. CR $\tilde{I}(Z_{t-1} \leq \gamma)$	Mét. CR $I(\tilde{Z}_{t-1} \leq \gamma)$
$h = 1$ Ide.rég.	[1.74] (1.30) 100%	[1.73] (1.29) 100%	[2.02] (1.53) 100%	[1.81] (1.32) 100%	[1.85] (1.32) 100%
$h = 2$ Ide.rég.	[1.08] (1.06) 64.2%	[1.09] (1.06) 66.6%	[1.83] (1.52) 76.4%	[1.23] (1.11) 64.2%	[1.21] (1.13) 66.6%
$h = 3$ Ide.rég.	[0.96] (0.98) 62.0%	[0.96] (0.99) 57.8%	[1.74] (1.43) 69.8%	[1.15] (1.05) 62.0%	[1.12] (1.06) 57.8%
$h = 4$ Ide.rég.	[1.11] (1.04) 99.6%	[0.88] (0.94) 56.6%	[1.80] (1.46) 64.2%	[1.05] (1.12) 99.6%	[1.27] (1.02) 56.6%
$h = 5$ Ide.rég.	[1.05] (1.01) 99.4%	[0.82] (0.92) 54.2%	[1.62] (1.36) 60.0%	[0.94] (1.05) 99.4%	[1.16] (0.97) 54.2%

Los resultados de la Tabla 16 indican que los desempeños de los pronósticos para el modelo TAR(2; 1, 1) considerando las dos funciones de pérdida y en comparación con los desempeños de los pronósticos bajo el modelo AR(1) son mejores. Notemos también que los desempeños de los pronósticos bajo el método de regresión por cuantiles favorece al modelo TAR(2; 1, 1). Recordemos que para los métodos de pronósticos recursivo y de regresión por cuantiles, los pronósticos de la variable umbral o su indicadora se obtienen como un paso previo, sin embargo notemos que los pronósticos para Y_t bajo TAR(2; 1, 1) en el método de regresión por cuantiles presentan mejores desempeños que bajo el método recursivo.

Para completar los resultados bajo este escenario presentamos el equivalente a la Tabla 16 para los casos caso 1-2 y caso 1-3 que difieren del caso 1-1 en los porcentajes de observaciones en los regímenes.

La Tabla 17 contiene los resultados de los cocientes de las funciones de pérdida y los porcentajes de veces en que el régimen se identificó correctamente, para los valores de los parámetros dados en el caso 1-2, para $\sigma^2 = 1$. La diferencia de este caso con el caso 1-1 es que el porcentaje de observaciones en el régimen uno ($\phi_1 = -0.1$) es del 75%.

Tabla 17. Cocientes de las funciones de pérdida, y porcentajes de veces en que se identifica correctamente al régimen, bajo cada esquema de obtención de los pronósticos, caso 1-2.

Horiz. pronós.	Mét. Rec. $\tilde{I}(Z_t \leq \gamma)$	Mét. Rec. $I(\tilde{Z}_{t-1} \leq \gamma)$	Método Vero. Pred.	Mét. CR $\tilde{I}(Z_t \leq \gamma)$	Mét. CR $I(\tilde{Z}_{t-1} \leq \gamma)$
$h = 1$ Ide.rég.	[1.32] (1.15) 100%	[1.32] (1.15) 100%	[1.33] (1.15) 100%	[1.32] (1.14) 100%	[1.32] (1.14) 100%
$h = 2$ Ide.rég.	[1.07] (1.03) 81.6%	[0.98] (1.00) 76.8%	[1.16] (1.06) 81.6%	[1.09] (1.03) 81.6%	[1.01] (1.00) 76.8%
$h = 3$ Ide.rég.	[1.04] (1.00) 75.0%	[0.98] (0.99) 72.8%	[1.20] (1.07) 79.4%	[1.09] (1.02) 75.0%	[0.98] (1.00) 72.8%
$h = 4$ Ide.rég.	[1.03] (1.00) 100%	[1.00] (1.00) 72.8%	[1.20] (1.06) 78.4%	[1.09] (1.02) 100%	[1.00] (1.00) 72.8%
$h = 5$ Ide.rég.	[0.98] (1.00) 99.6%	[0.99] (1.00) 76.4%	[1.09] (1.04) 80.8%	[1.03] (1.01) 99.6%	[1.00] (1.00) 76.4%

La Tabla 18 muestra los resultados de los cocientes de las funciones de pérdida, así como los porcentajes de veces en que el régimen se identificó correctamente, para los valores de los parámetros dados en el caso 1-3, para $\sigma^2 = 1$. La diferencia de este caso con el caso 1-1 es que el porcentaje de observaciones en el régimen de raíz unitaria ($\phi_1 = 1$) es del 75%.

Tabla 18. Cocientes de las funciones de pérdida, y porcentajes de veces en que se identifica correctamente al régimen, bajo cada esquema de obtención de los pronósticos, caso 1-3.

Horiz. pronós.	Mét. Rec. $\tilde{I}(Z_t \leq \gamma)$	Mét. Rec. $I(\tilde{Z}_{t-1} \leq \gamma)$	Método Vero. Pred.	Mét. CR $\tilde{I}(Z_t \leq \gamma)$	Mét. CR $I(\tilde{Z}_{t-1} \leq \gamma)$
$h = 1$ Ide.rég.	[2.02] (1.33) 100%	[2.02] (1.33) 100%	[2.02] (1.33) 100%	[2.17] (1.35) 100%	[2.06] (1.31) 100%
$h = 2$ Ide.rég.	[1.19] (1.13) 72.4%	[1.25] (1.15) 69.8%	[1.67] (1.28) 76.2%	[1.36] (1.20) 72.4%	[1.41] (1.21) 69.8%
$h = 3$ Ide.rég.	[1.09] (1.06) 71.4%	[1.09] (1.06) 66.8%	[1.44] (1.20) 72.0%	[1.29] (1.16) 71.4%	[1.30] (1.15) 66.8%
$h = 4$ Ide.rég.	[1.21] (1.10) 100%	[0.96] (0.96) 64.6%	[1.34] (1.13) 71.6%	[1.46] (1.25) 100%	[1.18] (1.11) 64.6%
$h = 5$ Ide.rég.	[1.15] (1.08) 100%	[0.83] (0.90) 65.8%	[1.22] (1.09) 73.0%	[1.37] (1.20) 100%	[1.04] (1.03) 65.8%

Al considerar los resultados de las Tablas 17 y 18 concluimos que modelo TAR(2; 1, 1) bajo el método de verosimilitud predictiva obtiene mejores resultados en cuanto al desempeño de sus pronósticos h -pasos adelante, bajo las funciones de pérdida ECMP y pérdida de cuantiles. Notamos también el buen desempeño de los pronósticos en el modelo TAR(2; 1, 1) bajo el método de regresión por cuantiles para las dos funciones de pérdida; es importante mencionar que los desempeños son mejores al considerar los pronósticos de la variable indicadora. Para el método recursivo no podemos extender las conclusiones anteriores ya que el modelo TAR(2; 1, 1) se desempeña mejor que el AR(1) para algunos horizontes de pronósticos en el caso en que el porcentaje de observaciones en el régimen estacionario sea mayor que el de raíz unitaria (Tabla 17) mientras que para el caso en que el régimen no estacionario (Tabla 18) posee el 75% de las observaciones es el modelo TAR(2; 1, 1) el que presenta mejor desempeño.

Resultados para el caso 3-1

A continuación presentamos los resultados para el caso 3-1, donde los valores de los parámetros considerados para la simulación de las 500 series son: $\phi_1 = -0.8$, $\phi_2 = -0.6$, $\sigma^2 = 1$, $\Pr(Z_{t-1} \leq \gamma) = 0.5$, $E(\delta_t) = -0.7$ y $E(\delta_t^2) = 0.5$. Los valores estimados promedios bajo los métodos de MV y RC respectivamente, se muestran en la siguiente tabla:

Método	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\sigma}_{TAR}^2$	$E(\hat{\delta}_t)$	$E(\hat{\delta}_t^2)$	$\hat{\phi}$	$\hat{\sigma}_{AR}^2$
MV	-0.786	-0.594	0.996	-0.690	0.486	-0.697	1.019
RC	-0.788	-0.591	1.001	-0.690	0.486	-0.696	1.022

Notemos que el promedio de las estimaciones de los coeficientes autoregresivos bajo el modelo TAR son muy cercanas a los verdaderos. Las estimaciones de las varianzas bajo el modelo AR(1), $\hat{\sigma}_{AR}^2$, son muy cercanas al valor utilizado para el ejercicio de simulación, y a diferencia del caso 1-1 tenemos que el valor de $\hat{\phi}$ es aproximadamente igual al valor de $E(\hat{\delta}_t)$, estamos también ante un modelo TAR(2; 1, 1) bien comportado en el sentido de que nos alejamos de no estacionariedad.

La Figura 31 muestra algunas realizaciones del proceso bajo el escenario descrito arriba, observemos el comportamiento homogéneo de las observaciones a través del tiempo.

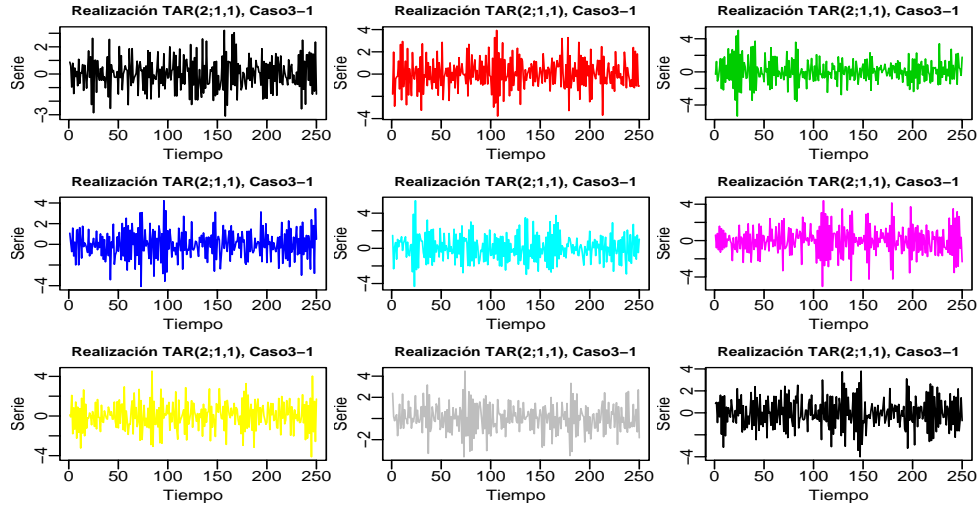


Figura 31. Realizaciones del modelo TAR(2; 1, 1) para el caso 3-1 con $Z_{t-1} \sim \text{AR}(1)$.

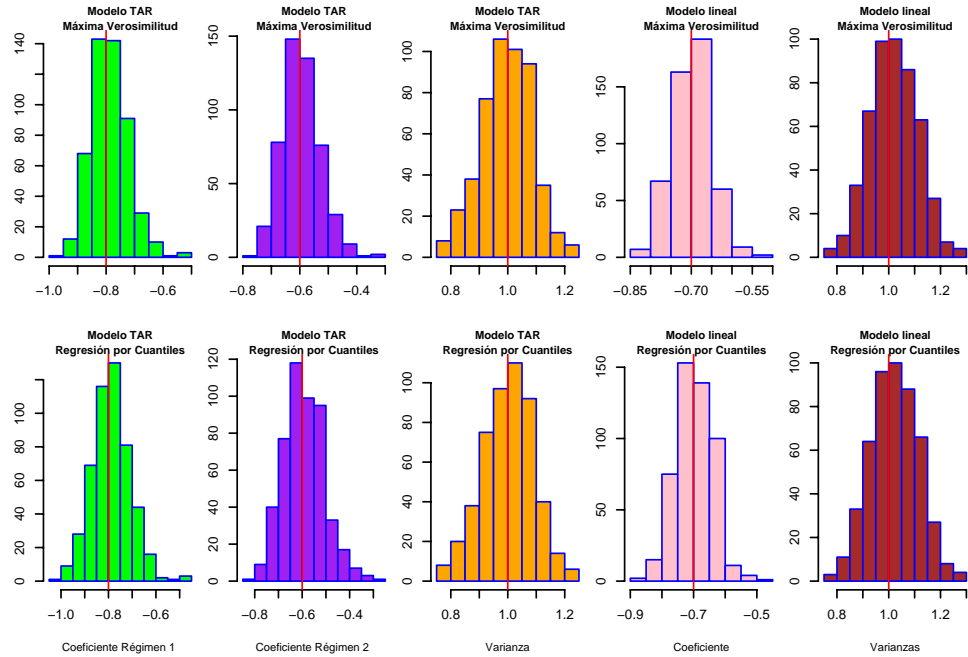


Figura 32. Histogramas de los parámetros estimados, bajo los modelos TAR(2; 1, 1) y AR(1), caso 3-1 con $Z_{t-1} \sim \text{AR}(1)$.

En la Figura 32 presentamos los comportamientos de las estimaciones para cada uno de los parámetros en los modelos TAR(2;1,1) y AR(1), bajo los métodos de MV y RC respectivamente. Observemos que bajo este escenario, el comportamiento de los estimadores para la varianza bajo el modelo TAR(2;1,1) y AR(1) son muy parecidos además de que $\hat{\phi}_1 \approx E(\delta_t)$, y de aquí que esperamos que los desempeños de los pronósticos h -pasos adelante bajo ambos modelos sean aproximadamente iguales.

Los errores de pronósticos para cada horizonte de pronóstico se encuentran en la Figura 33, de la cual observamos que los rangos de variación de los pronósticos bajo los distintos esquemas de pronósticos y bajo ambos modelos son esencialmente lo mismo.

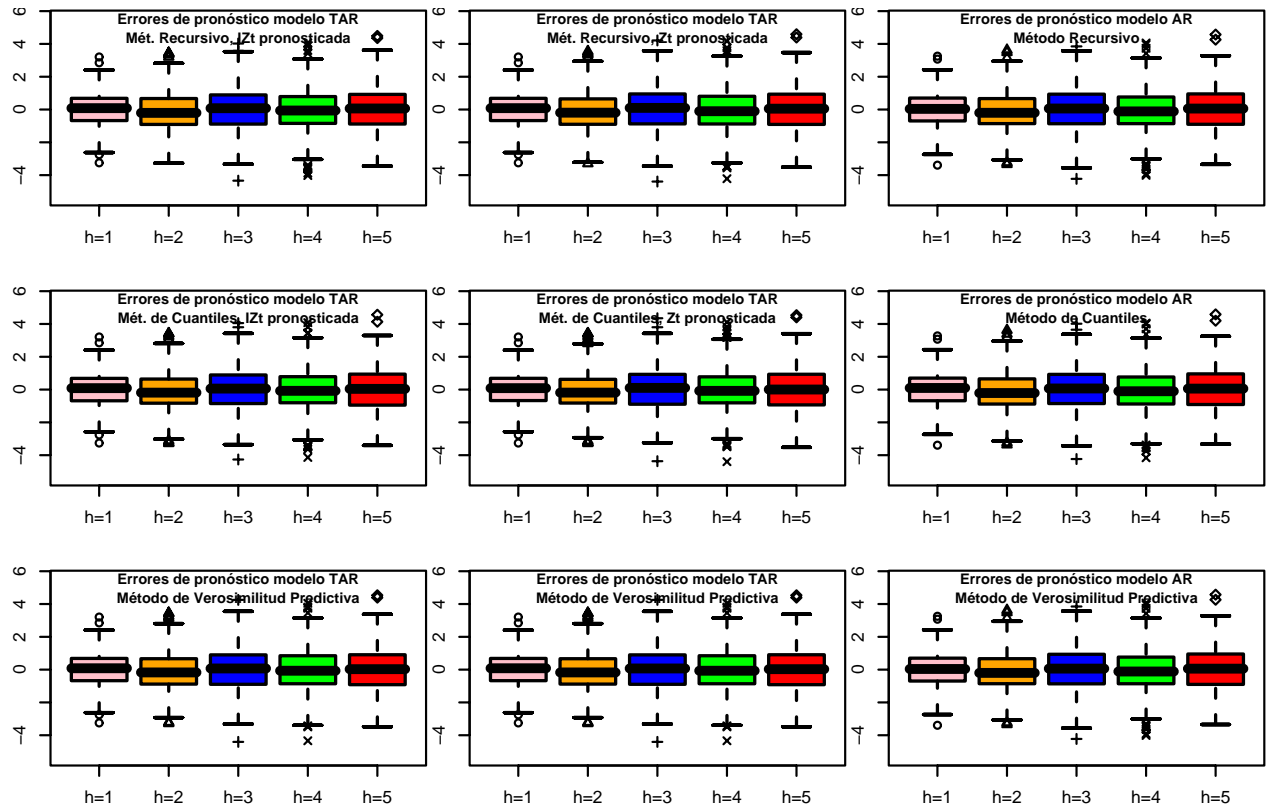


Figura 33. Errores de pronóstico bajo cada alternativa de pronóstico para el caso 3-1, con $Z_{t-1} \sim \text{AR}(1)$.

Los comportamientos de las funciones de pérdida para cada horizonte de pronóstico se presentan en las dos primeras gráficas de la Figura 34. Salvo el caso del modelo TAR(2; 1, 1) bajo el método recursivo y considerando los pronósticos de Z_{t-1} , identificado por la línea punteada azul fuerte, observamos que las funciones de pérdida toman prácticamente los mismos valores.

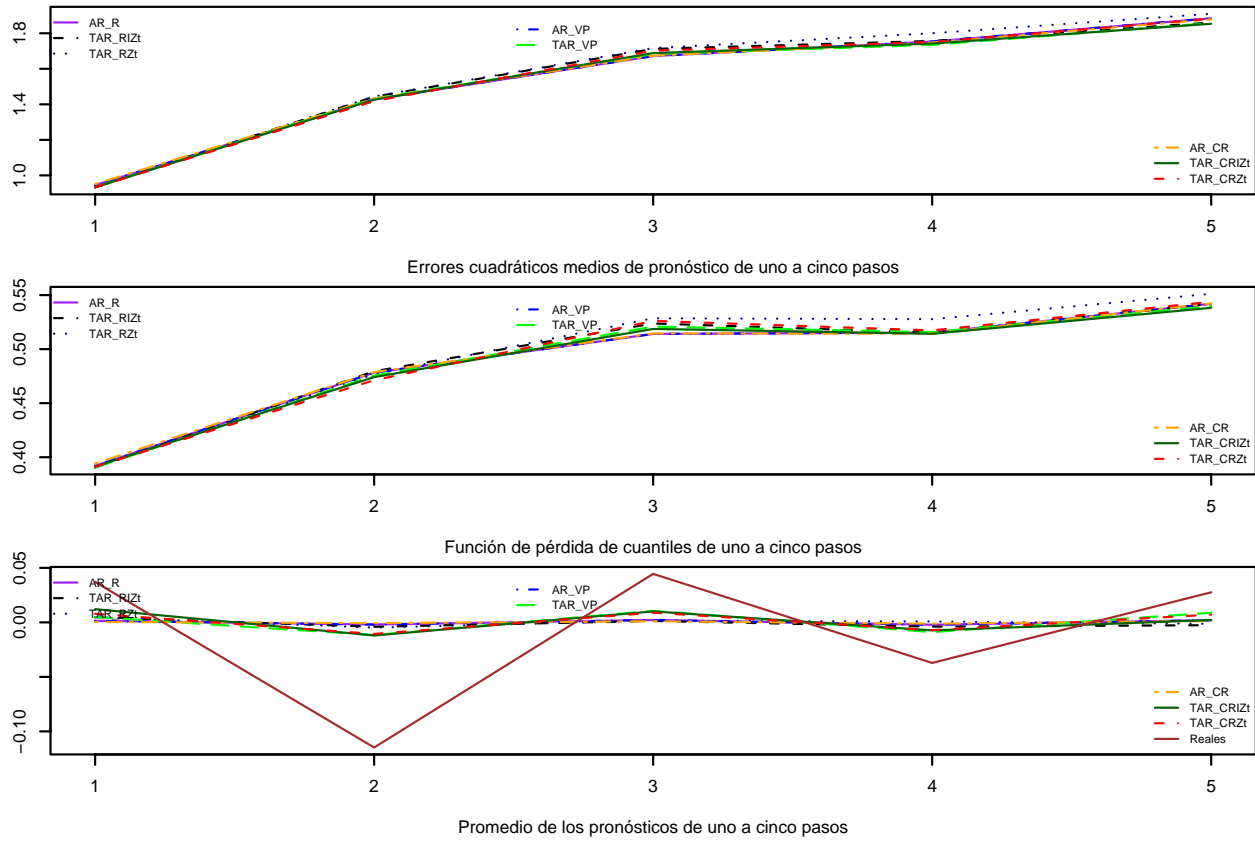


Figura 34. Funciones de pérdida para cada horizonte de pronóstico, caso 3-1 con $Z_{t-1} \sim \text{AR}(1)$.

El resumen del análisis gráfico se encuentra en la Tabla 19, que contiene los cocientes de las funciones de pérdida en el modelo AR(1) entre las funciones de pérdida en el modelo TAR(2; 1, 1), así como los porcentajes en que el régimen es correctamente identificado.

Tabla 19. Cocientes de las funciones de pérdida, y porcentajes de veces en que se identifica correctamente al régimen, bajo cada esquema de obtención de los pronósticos, caso 3-1.

Horiz. pronós.	Mét. Rec. $\tilde{I}(Z_t \leq \gamma)$	Mét. Rec. $I(\tilde{Z}_{t-1} \leq \gamma)$	Método Vero. Pred.	Mét. CR $\tilde{I}(Z_t \leq \gamma)$	Mét. CR $I(\tilde{Z}_{t-1} \leq \gamma)$
1 Ide.rég.	[1.01] (1.00) 100%	[1.01] (1.00) 100%	[1.01] (1.01) 100%	[1.02] (1.01) 100%	[1.02] (1.01) 100%
2 Ide.rég.	[0.99] (1.00) 59.4%	[0.99] (1.00) 64.4%	[1.00] (1.01) 64.0%	[1.01] (1.01) 59.4%	[1.01] (1.02) 64.4%
3 Ide.rég.	[0.98] (0.99) 59.0%	[0.97] (0.97) 59.8%	[1.00] (0.99) 59.8%	[1.00] (1.00) 59.0%	[0.98] (0.98) 59.8%
4 Ide.rég.	[1.00] (1.00) 98.6%	[0.97] (0.98) 55.6%	[1.01] (1.00) 55.6%	[1.01] (1.00) 98.6%	[1.00] (1.00) 55.6%
5 Ide.rég.	[1.01] (1.00) 98.8%	[0.99] (0.98) 54.4%	[1.01] (1.01) 54.2%	[1.01] (1.01) 98.8%	[1.00] (1.00) 54.4%

De los resultados de la Tabla 19 podemos concluir que al utilizar los método de verosimilitud predictiva y de regresión por cuantiles para generar los pronósticos h -pasos adelante el modelo TAR(2; 1, 1) presenta ligeras ganancias en los desempeños de los pronósticos, sin embargo estrictamente hablando tenemos que los desempeños de los pronósticos bajo el modelo AR(1) son equivalente a los desempeños bajo el TAR(2; 1, 1). Si el método de pronóstico utilizado corresponde al método recursivo y en particular si usamos los pronósticos para Z_{t-1} observamos que el desempeño del AR(1) es mejor que el desempeño bajo el TAR(2; 1, 1).

Resultados para el caso 8-1

El caso 8-1 corresponde a los siguientes valores de los parámetros son: $\phi_1 = -0.95$, $\phi_2 = 0.95$, $\sigma^2 = 1$, $\Pr(Z_{t-1} \leq \gamma) = 0.5$, $E(\delta_t) = 0$ y $E(\delta_t^2) = 0.903$, con valores estimados bajo el modelo TAR(2; 1, 1) y AR(1) vía MV y RC respectivamente, dados en la Tabla siguiente:

Método	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\sigma}_{TAR}^2$	$E(\hat{\delta}_t)$	$E(\hat{\delta}_t^2)$	$\hat{\phi}$	$\hat{\sigma}_{AR}^2$
MV	-0.941	0.943	0.990	0.0008	0.888	-0.0006	9.664
RC	-0.942	0.944	0.995	0.0009	0.889	-0.0071	11.175

A partir de los valores de los coeficientes autorregresivos para cada uno de los regímenes notamos que en ambos regímenes estamos cercanos al caso de raíz de unitaria, recuerde sin embargo que globalmente el modelo es estacionario. Podemos observar el comportamiento del modelo TAR(2; 1, 1) bajo este escenario de simulación en las realizaciones del proceso dadas en la Figura 35.

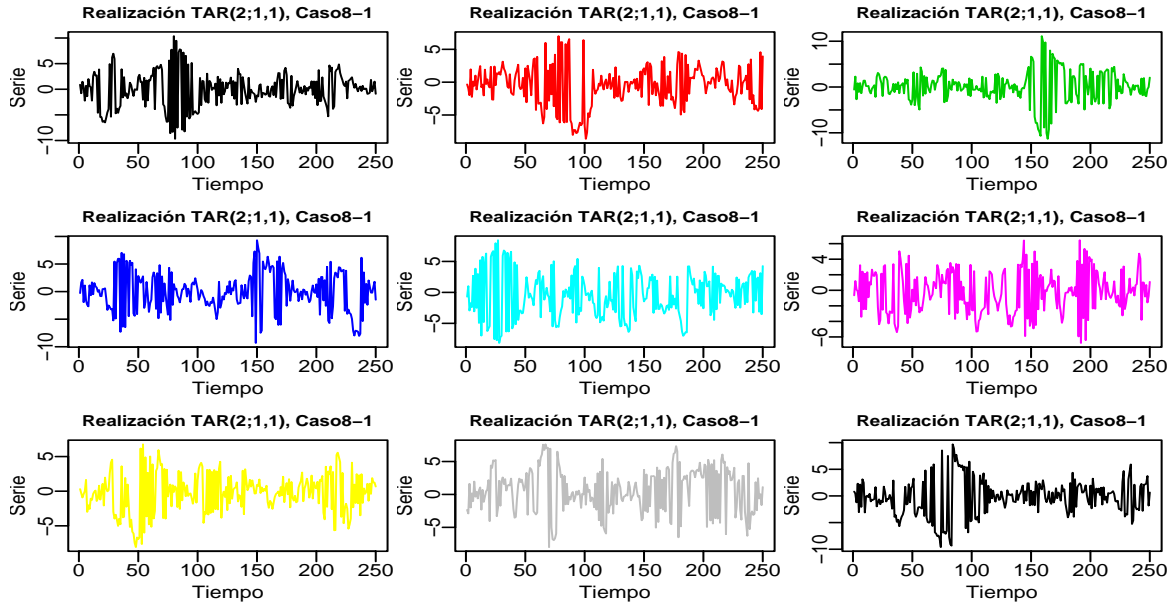


Figura 35. Realizaciones del modelo TAR(2; 1, 1) para el caso 8-1 con $Z_{t-1} \sim \text{AR}(1)$.

Los histogramas correspondientes a las estimaciones de los parámetros bajo cada modelo y considerando los métodos de estimación de MV y CR se presentan en la Figura 36. Las líneas verticales se interpretan como en los otros casos analizados. Notemos que la distribución de los coeficientes autorregresivos en el modelo TAR(2; 1, 1) presentan ligeros sesgos, a la derecha para el caso en que el signo del coeficiente autorregresivo es negativo y a la izquierda cuando el signo del coeficiente autorregresivo es positivo, originado por la cercanía del a la condición de no estacionariedad.

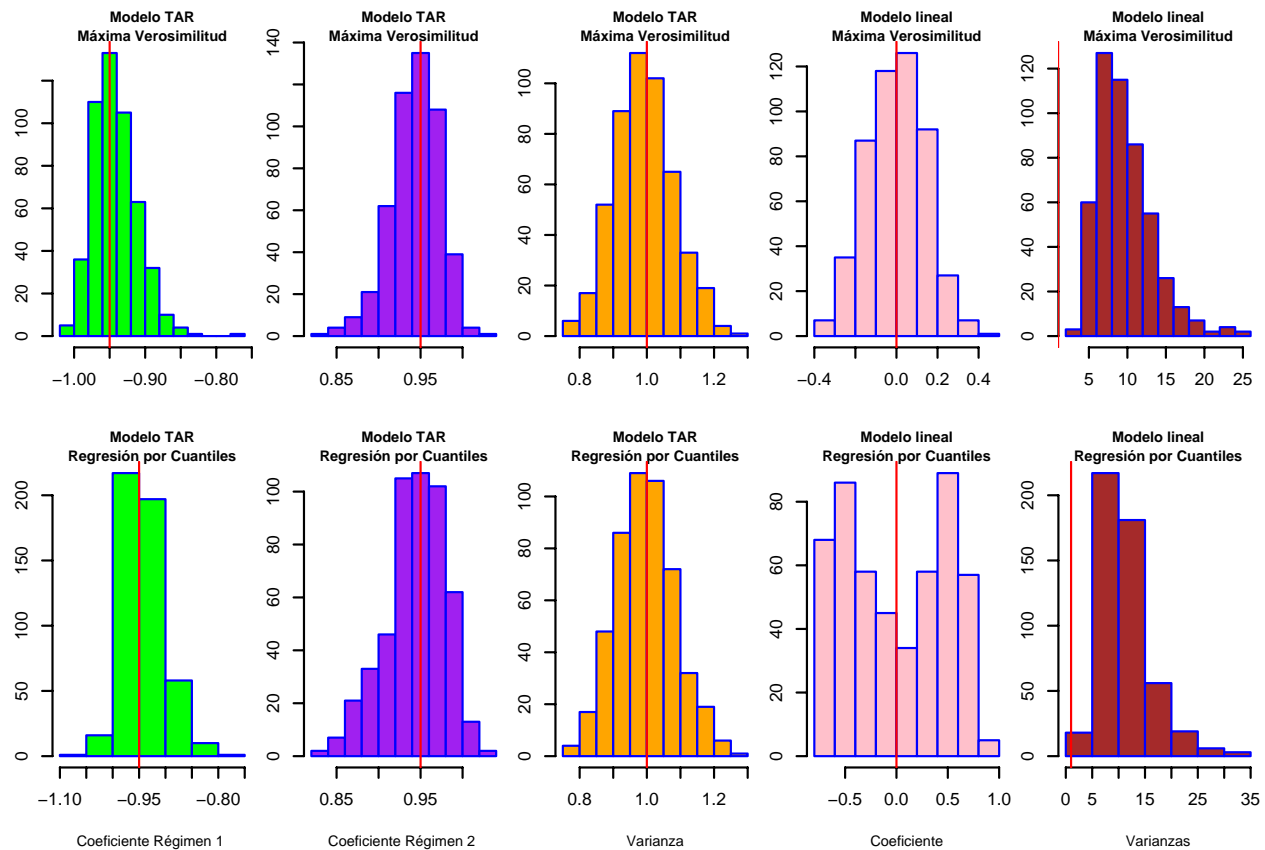


Figura 36. Histogramas de los parámetros estimados, bajo los modelos TAR(2; 1, 1) y AR(1), caso 8-1 con $Z_{t-1} \sim \text{AR}(1)$.

Al calcular los errores de pronósticos graficados en la Figura 37 observamos que los errores de pronóstico bajo verosimilitud predictiva presenta poca variabilidad alrededor de cero, mientras que bajo los métodos de regresión por cuantiles y recursivo la variabilidad de los errores de pronóstico se ve afectada por los valores que exceden el rango intercuantil, veremos la influencia de dichos valores al momento de calcular los valores de las respectivas funciones de pérdida para cuantificar los desempeños de los pronósticos.

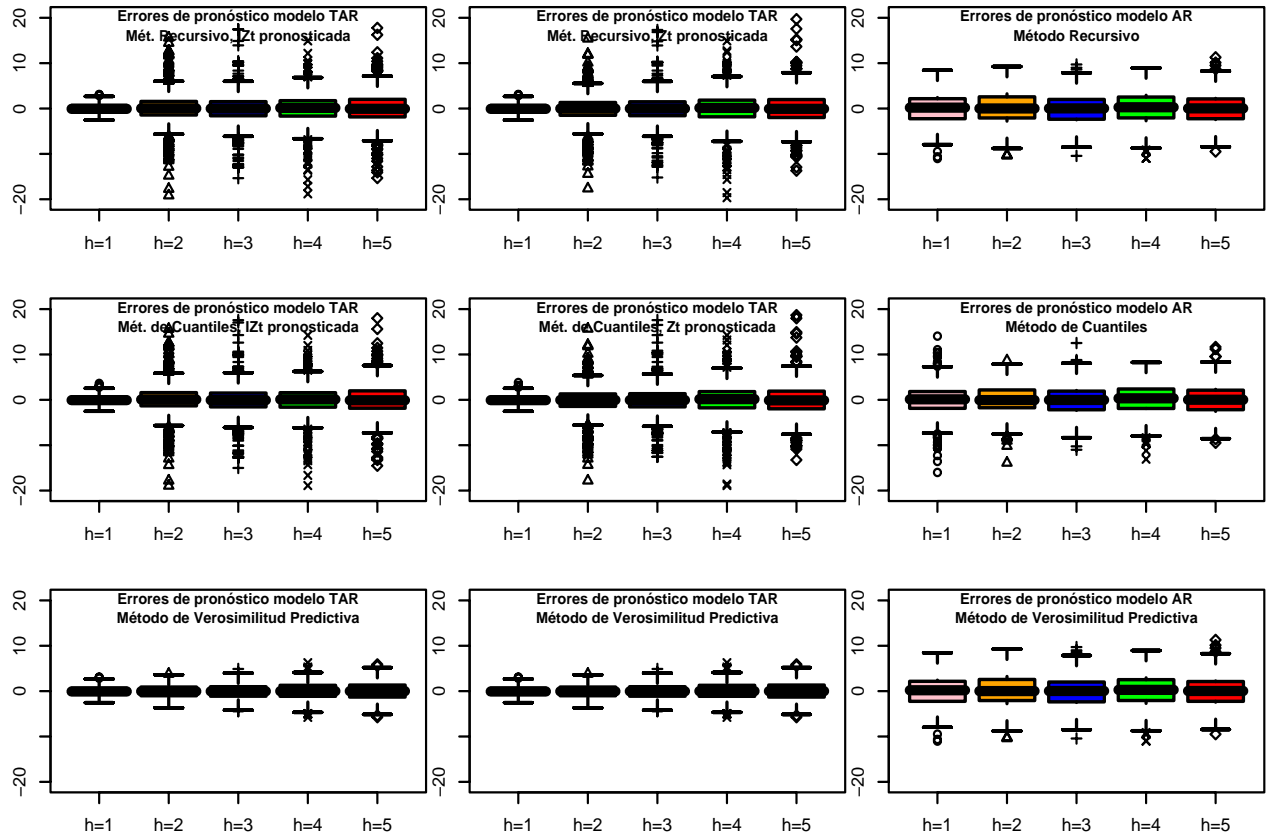


Figura 37. Errores de pronóstico bajo cada alternativa de pronóstico para el caso 8-1, con $Z_{t-1} \sim \text{AR}(1)$.

La Figura 38 presenta los promedios de los valores de las pérdidas ECMP y de cuantiles bajo cada horizonte de pronóstico y para cada alternativa de obtención de los pronósticos. Resalta el comportamiento de las pérdidas bajo verosimilitud predictiva, indicando que la ganancia al utilizar este método para genera los pronósticos h -pasos adelante identifica que el desempeño de los pronósticos en el TAR(2; 1, 1) es superior al caso AR(1). Una observación por demás interesante en relación a la función de pérdida de cuantiles es que, bajo la función

de pérdida ECMP los desempeños de los pronósticos en el modelo TAR(2;1,1) bajo los métodos recursivo y de cuantiles es pobre comparado con el caso similar, pero bajo el modelo AR(1), mientras que bajo la función de pérdida de cuantiles la ganancia en los pronósticos se da para el modelo TAR(2;1,1).

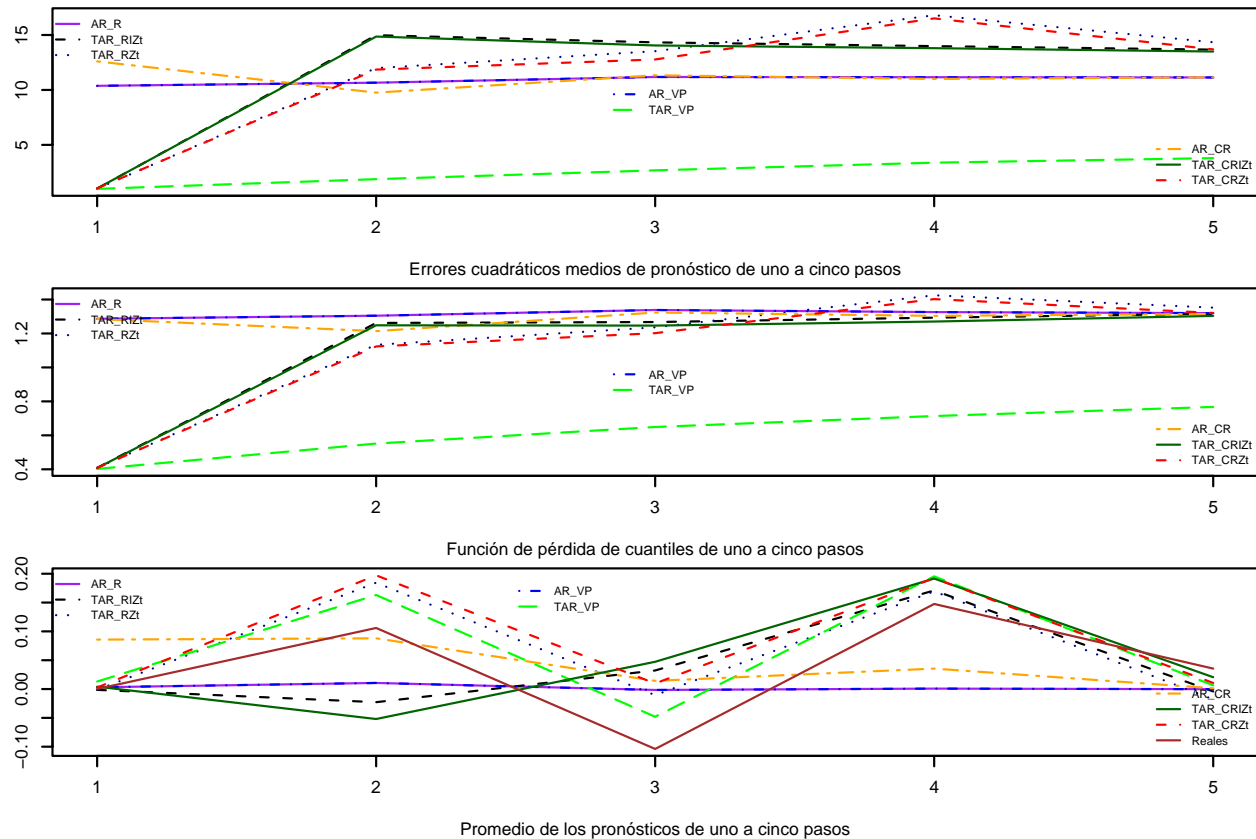


Figura 38. Funciones de pérdida para cada horizonte de pronóstico, caso 8-1 con $Z_{t-1} \sim \text{AR}(1)$.

La Tabla 20 presenta los cocientes de las funciones de pérdida bajo el modelo AR(1) entre las funciones de pérdida bajo el modelo TAR(2; 1, 1), podemos ver la ganancia del método de verosimilitud como función del ECMP y de la pérdida de cuantiles. Notemos además que los desempeños de los pronósticos bajo los métodos recursivos y de cuantiles bajo la pérdida ECMP favorece al modelo AR(1), mientras que bajo la pérdida de regresión por cuantiles es el modelo TAR(2; 1, 1) el que presenta el mejor desempeño.

Tabla 20. Cocientes de las funciones de pérdida, y porcentajes de veces en que se identifica correctamente al régimen, bajo cada esquema de obtención de los pronósticos, caso 8-1.

Horiz. pronós.	Mét. Rec. $\tilde{I}(Z_t \leq \gamma)$	Mét. Rec. $I(\tilde{Z}_{t-1} \leq \gamma)$	Método Vero. Pred.	Mét. CR $\tilde{I}(Z_t \leq \gamma)$	Mét. CR $I(\tilde{Z}_{t-1} \leq \gamma)$
$h = 1$ Ide.rég.	[10.18] (3.15) 100%	[10.18] (3.15) 100%	[10.41] (3.20) 100%	[12.19] (3.15) 100%	[12.05] (3.15) 100%
$h = 2$ Ide.rég.	[0.71] (1.03) 60.6%	[0.89] (1.15) 66.0%	[5.67] (2.37) 93.4%	[0.66] (0.97) 60.6%	[0.82] (1.08) 66.0%
$h = 3$ Ide.rég.	[0.78] (1.06) 60.6%	[0.83] (1.08) 59.4%	[4.17] (2.06) 94.2%	[0.81] (1.06) 60.6%	[0.89] (1.10) 59.4%
$h = 4$ Ide.rég.	[0.80] (1.03) 99.8%	[0.66] (0.93) 57.0%	[3.30] (1.86) 94.2%	[0.80] (1.03) 99.8%	[0.67] (0.93) 57.0%
$h = 5$ Ide.rég.	[0.81] (1.00) 99.6%	[0.78] (0.98) 51.2%	[2.93] (1.72) 92.2%	[0.83] (1.01) 99.6%	[0.81] (1.00) 51.2%

Los resultados para el caso de $\sigma^2 = 4$ son similares y no se reportan aquí. Las conclusiones generales del ejercicio de simulación muestran que los pronósticos vía el método de verosimilitud predictiva son los de mejor desempeño en virtud de que estos presentan valores más pequeños, de las respectivas funciones de pérdida ECMP y pérdida de regresión por cuantiles que sus competidores lineales AR bajo la misma alternativa de pronóstico. La situación explorada en el caso 3-1, muestra la robustez de la alternativa de regresión por cuantiles, ya que identifica que el modelo TAR presenta mejor desempeño de los pronósticos a pesar de la cercanía aparente en los modelo TAR y AR como función de los valores de los coeficientes autorregresivos en el modelo TAR.

Capítulo 8

Conclusiones y discusión

En este trabajo de tesis hemos identificado algunas de las características de mayor peso al momento de cuantificar y estudiar los desempeños de los pronósticos en modelos TAR. Ante el problema identificado en la literatura de pronósticos no lineales, de que aún cuando el VPGD es efectivamente un modelo no lineal no hay garantía de que los pronósticos asociados a dichos modelos presenten mejores desempeños, e comparación con los pronósticos asociados a modelos lineales, y cuantificados por medio de los ECMP, Clements y Smith (1997), Dacco y Satchell (1999), Clements *et al* (2003). Proponemos dos alternativas basadas en la teoría estadística para obtener los estimadores de los parámetros y los pronósticos en modelos TAR: verosimilitud predictiva y regresión por cuantiles. La consideración de dichas alternativas resuelve el problema identificado en la literatura en virtud de que bajo éstas, los desempeños de los pronósticos en modelos TAR son mejor comportados que los pronósticos correspondientes bajo el modelo lineal, considerando las funciones de pérdida error cuadrático medio y la función cuantil.

Una característica de los modelos TAR es que bajo ciertos escenarios (como función de las magnitudes y signos de los parámetros autoregresivos) el modelo lineal provee una buena aproximación al modelo no lineal y de aquí que ante la simplicidad del modelo lineal para obtener los pronósticos, erróneamente se tome a este como el mejor modelo. Sin embargo, hemos mostrado que bajo el modelo especificado incorrectamente el proceso de errores asociados se rige por un modelo de volatilidad. La variabilidad observada de dicho proceso a través del tiempo se identifica directamente como función de los primeros dos momentos de la variable δ_t , en virtud de que dichos momentos determinan la condición de estacionariedad del modelo TAR, es decir, entre más alejado estemos de la condición de no estacionariedad ($E(\delta_t^2) = 1$) mejor será la aproximación del modelo lineal. Aún en estos casos, los métodos de verosimilitud predictiva y regresión por cuantiles presentan un buen desempeño de sus pronósticos, el cual mejora si la cuantificación de los desempeños se realiza por medio de la función de pérdida de cuantiles como reflejo de la robustez de dicha alternativa al considerar

adecuadamente la asimetría natural presentada en los modelo TAR.

Identificamos que ante especificaciones erróneas, al suponer un modelo lineal, se paga el precio de que los residuales asociados al modelo lineal no se comportan como ruido blanco. La distribución identificada en los errores corresponde a distribuciones asimétricas de colas pesadas, y de aquí que el supuesto de errores con distribución normal no se presenta, la gravedad del problema esta en función de la parametrización específica del modelo TAR, como procesos generador de los datos.

El incremento de la variabilidad identificado bajo el modelo mal especificado, así como el no cumplimiento distribucional de los residuales respectivos se ve rebasado al considerar metodologías distintas en los procedimientos de estimación y obtención de pronósticos tal y como se había venido realizando en gran parte de la literatura. Nuestras alternativas, verosimilitud predictiva y regresión por cuantiles, abordan el problema de estimación de parámetros y obtención de pronósticos bajo la misma metodología estadística y de aquí que al momento de evaluar los desempeños de los pronósticos bajo el VPGD estos presenten mejores desempeños que bajo un modelo incorrecto.

Paralelamente, debemos considerar que existen situaciones donde el proceso de estimación se da hasta después de la identificación del modelo, de manera que hay que tomar en cuenta otras características del fenómeno de interés al momento de elegir un modelo lineal en lugar de un modelo no lineal. Otro punto importante es que en muchos fenómenos naturales la interpretación de los pronósticos lineales carece de sentido práctico.

Una ventaja de la alternativa de regresión por cuantiles es que podemos considerar distribuciones asimétricas para los errores, y que esta asimetría está asociada directamente con los valores del parámetro θ en la función cuantil.

Un estudio exhaustivo sobre el efecto de distribuciones no normales en los errores debe realizarse el cual nos lo planteamos como un trabajo futuro, aunque es claro que no es un factor para el método de regresión por cuantiles, solo para el de verosimilitud predictiva.

Para resolver el problema que dió origen a este trabajo de tesis, fue necesario realizar un estudio detallado del proceso estimación de los parámetros de los modelos TAR, demostramos las propiedades de consistencia y distribución asintótica de los estimadores, considerando γ conocido para cada uno de los procedimientos de estimación, mínimos cuadrados, máxima verosimilitud y regresión por cuantiles. Para γ desconocido mostramos sólo la consistencia de los estimadores para el método de máxima verosimilitud.

Las pruebas para las distribuciones asintóticas en mínimos cuadrados y máxima verosimilitud descansan fuertemente en el concepto de continuidad y discontinuidad para SETAR.

Este concepto no tiene una clara extensión al caso TAR. Las demostraciones para las distribuciones asintóticas de los EMV en el caso de γ desconocido no fueron estudiadas, y se deja como trabajo futuro.

Hay que conocer la estructura probabilística particular del modelo TAR para aplicar los resultados del caso SETAR al caso TAR. Tomando esto en cuenta las demostraciones en esta tesis son aportaciones a la literatura de modelos no lineales.

Bajo el supuesto de normalidad, las estimaciones vía mínimos cuadrados y máxima verosimilitud son iguales, sin embargo los métodos de pronósticos en modelos TAR, asociados a dichos métodos de estimación no son iguales, mientras que en el caso lineal sí. Verosimilitud predictiva considera toda la información de manera conjunta, utilizando así las propiedades probabilísticas de la variable umbral, lo cual hace más eficiente el proceso de optimización. Regresión por cuantiles y el método recursivo presuponen conocida la variable umbral, ya que el proceso de obtención de los pronósticos de dicha variable se realiza como un paso previo a la obtención de los pronósticos en el modelo TAR. Como parte de los resultados del ejercicio de simulación notamos que los pronósticos TAR mediante regresión por cuantiles y el método recursivo son más eficientes si utilizamos los pronósticos de la variable indicadora en vez de los pronósticos de la variable umbral. Esto debido a que las probabilidades de mala clasificación en los regímenes bajo el pronóstico de la variable indicadora son muy pequeños comparados con la mala clasificación de los regímenes bajo el pronóstico de la variable umbral.

Una aplicación inmediata de las alternativas de pronósticos aquí propuestas, es aplicar las técnicas de verosimilitud predictiva y regresión por cuantiles a los modelos SETAR, de los cuales no se tienen precedentes en la literatura actual.

Apéndice A

Prueba de la propiedad de consistencia de los EMV

Notación

La función f es la función de densidad de ε_2 y F la función de distribución correspondiente a f .

La esperanza bajo $\boldsymbol{\vartheta}$ se denota por E .

Sea \mathbb{R} la recta real $(-\infty, \infty)$ y $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$, entonces $\bar{\mathbb{R}}$ es compacto bajo la métrica $d(\cdot, \cdot)$ definida por $d(x, y) = |\arctan x - \arctan y|$.

Una función φ es $\text{Lip}(1)$ si $\forall x, y \in \mathbb{R}$, existe $L \geq 0$, tal que $|\varphi(x) - \varphi(y)| \leq L|x - y|$.

El complemento de A se denota por A^c .

El producto interno de \mathbf{x} y \mathbf{y} se denota por $\mathbf{x}'\mathbf{y}$.

Se omitirá el subíndice $t = 2, \dots, T$ a menos que se especifique lo contrario.

Preliminares al método de estimación de máxima verosimilitud

Supondremos que ε_2 tiene distribución conocida y que σ se involucra en el modelo modificando la escala, i.e., El modelo (2.1) se puede escribir de la siguiente manera

$$Y_t = h(\mathbf{Y}_{t-1}, \mathbf{Z}_{t-d}, \boldsymbol{\phi}, \gamma) + \sigma \varepsilon_t, \quad t \geq 2, \quad (\text{A.1})$$

para algún $\boldsymbol{\vartheta} = (\boldsymbol{\phi}'_1, \boldsymbol{\phi}'_2, \gamma) \in \mathbb{R}^{2p+3}$, $d \in \{1, 2, \dots, p\}$, donde $\mathbf{Y}_{t-1} = (Y_{t-1}, \dots, Y_{t-p})'$, $\mathbf{Z}_{t-d} = (Z_{t-d}, \dots, Z_{t-d-p+1})'$, $\boldsymbol{\phi}_j = (\alpha_{0j}, \alpha_{1j}, \dots, \alpha_{pj})' \in \mathbb{R}^{p+1}$, $j = 1, 2$ y para $\mathbf{y} \in \mathbb{R}^p$ y $\mathbf{z} \in \mathbb{R}^p$,

$$h(\mathbf{y}, \mathbf{z}, \boldsymbol{\vartheta}) = \left(\phi_{01} + \sum_{k=1}^p \phi_{k1} y_k \right) I(z_d \leq \gamma) + \left(\phi_{02} + \sum_{k=1}^p \phi_{k2} y_k \right) I(z_d > \gamma).$$

Los errores $\{\varepsilon_t\}$ en (A.1) son $iidN(0, 1)$ y

$$Z_t = \rho Z_{t-1} + u_t,$$

donde $\{u_t\}$ son $iidN(0, \sigma_u^2)$.

La demostración de las propiedades asintóticas de los EMV de $\phi, \gamma, d, \sigma, \rho, \sigma_u$ se hará suponiendo $\sigma = 1$, por conveniencia en las expresiones involucradas en las demostraciones del Lema A.1 y el Teorema 3.4. En el Lema A.2 se probará para el caso $\sigma > 0$ general.

Debido a que Z_t tiene su propia dinámica se tiene que los estimadores de ρ y σ_u cumplen las propiedades de los modelos AR(1): fuertemente consistentes y asintóticamente normales ver Teorema 8.2.1. de Fuller (1996).

Supondremos que $\boldsymbol{\vartheta} = (\phi', \gamma, d)'$ es un punto interior del espacio paramétrico $\mathbb{R}^{2p+2} \times \bar{\mathbb{R}} \times \{1, 2, \dots, p\}$.

Dado que $\boldsymbol{\vartheta}$ es un punto interior se sigue que existe un conjunto compacto $K \subset \mathbb{R}^{2p+2}$ tal que $\boldsymbol{\vartheta}$ es un punto interior de $\Lambda = K \times \bar{\mathbb{R}} \times \{1, 2, \dots, p\}$, Λ es un conjunto compacto.

Sea $\mathbf{v} = (\beta', r, q)'$ un punto interior de Λ . Sea $(\mathbf{Y}'_t, \mathbf{Z}'_t)' = (Y_t, \dots, Y_{t-p+1}, Z_t, \dots, Z_{t-p+1})'$, $\{\mathbf{Z}_t\}$ y $\{\mathbf{Y}_t, \mathbf{Z}_t\}$ son cadenas de Markov. Sea $g_{\boldsymbol{\vartheta}}(\mathbf{Y}_1, \mathbf{Z}_{2-d})$ la densidad inicial de \mathbf{Y}_1 y \mathbf{Z}_{2-d} bajo $\boldsymbol{\vartheta}$, la función de transición a un paso, iniciando en $(\mathbf{Y}'_1, \mathbf{Z}'_{2-d})'$ es

$$f(Y_t - h(\mathbf{Y}_{t-1}, \mathbf{Z}_{t-d}, \boldsymbol{\vartheta})) f_1(Z_{t-d} - h_1(\mathbf{Z}_{t-d-1}, \rho)), t \geq 2.$$

La función de transición a un paso, iniciando en \mathbf{Z}'_{2-d} es $f_1(Z_t - h_1(\mathbf{Z}_{t-1}, \rho)) w(Z_t | Z_{t-1})$, $t \geq 2$.

Sea f_1 la función de densidad de u_t , $g_1(\mathbf{Z}_{2-d})$ es la densidad inicial de \mathbf{Z}_{2-d} bajo $(\rho, \sigma_u^2)'$,

Si observamos $(\mathbf{Y}'_1, Y_2, \dots, Y_T, \mathbf{Z}_{2-d}, Z_{2-d+1}, \dots, Z_{T-d+1})'$ se sigue que la función de verosimilitud bajo $\boldsymbol{\vartheta}$ es

$$\prod_{t=2}^T f(Y_t - h(\mathbf{Y}_{t-1}, \mathbf{Z}_{t-d}, \phi, \gamma)) f_1(Z_{t-d} - h_1(\mathbf{Z}_{t-d-1}, \rho)) g_{\boldsymbol{\vartheta}}(\mathbf{Y}_1, \mathbf{Z}_{2-d}).$$

La estimación de $\boldsymbol{\vartheta}$ no se ve afectada por la estimación de ρ , por lo que consideraremos la función de verosimilitud condicional de $\boldsymbol{\vartheta}$ dada por

$$L_n(\boldsymbol{\vartheta}) = \prod_{t=d}^T f(Y_t - h(\mathbf{Y}_{t-1}, \mathbf{Z}_{t-d}, \phi, \gamma)). \quad (\text{A.2})$$

Sea $\hat{\boldsymbol{\vartheta}}_T = (\hat{\boldsymbol{\alpha}}', \hat{\gamma}, \hat{d})'$ cualquier función medible de $(\mathbf{Y}_1, Y_2, \dots, Y_T, \mathbf{Z}_{2-d}, Z_{2-d+1}, \dots, Z_{T-d+1})'$ de \mathbb{R}^{T+2p} a Λ tal que $\hat{\boldsymbol{\vartheta}}_T$ maximice, sobre Λ , la función de verosimilitud condicional dada en (A.2).

Supuestos

SC1. $f(y) > 0$, para toda $y \in \Re$ y es absolutamente continua. f' existe casi en todas partes, $\varphi = f'/f$ y $I(f) = \int_{-\infty}^{\infty} \varphi^2(y) f(y) dy < \infty$.

SC2. φ es Lip(1).

SC3. φ es derivable y la derivada φ' es Lip(1).

SC4. $E|\varepsilon_2|^4 < \infty$ y $E|u_2|^4 < \infty$.

Utilizaremos el hecho de que $E|\varepsilon_2|^k < \infty$, para $k = 1, 2, 3, 4$, implica que $E|\mathbf{Y}_{d-1}|^k < \infty$.

Probaremos la consistencia fuerte del EMV de $\hat{\boldsymbol{\vartheta}}_T$. Sea ℓ_T la razón de log-verosimilitud condicional:

$$\ell_T(\mathbf{v}) = \ell_T(\boldsymbol{\beta}, r, q) = \frac{1}{T} \sum \ln \frac{f(Y_t - h(\mathbf{Y}_{t-1}, \mathbf{Z}_{t-q}, \boldsymbol{\beta}, r))}{f(Y_t - h(\mathbf{Y}_{t-1}, \mathbf{Z}_{t-d}, \boldsymbol{\phi}, \gamma))}, \quad \mathbf{v} \in \Lambda \quad (\text{A.3})$$

La ecuación invariante

$$g_{\boldsymbol{\vartheta}}(\mathbf{y}, \mathbf{z}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(y - h(u, v, \boldsymbol{\vartheta})) f_1(z - h_1(u, \rho)) g_{\boldsymbol{\vartheta}}(u, v) dudv$$

y el supuesto SC1 implican que $g_{\boldsymbol{\vartheta}}$ es acotada fuera de $[0, \infty]^2$ sobre conjuntos compactos. $I(f) < \infty$ implica que f es acotada.

$$\psi(Y_{t-1}, Z_{t-q}, \varepsilon_t, \boldsymbol{\beta}, r) = \ln \frac{f(\varepsilon_t + h(\mathbf{Y}_{t-1}, \mathbf{Z}_{t-d}, \boldsymbol{\phi}, \gamma) - h(\mathbf{Y}_{t-1}, \mathbf{Z}_{t-q}, \boldsymbol{\beta}, r))}{f(\varepsilon_t)}, \quad 1 \leq t \leq n.$$

Obsérvese que

$$\ell_T(\mathbf{v}) = \frac{1}{T} \sum \psi(\mathbf{Y}_{t-1}, \mathbf{Z}_{t-q}, \varepsilon_t, \boldsymbol{\beta}, r), \quad \mathbf{v} \in \Lambda.$$

Sea $\mathbf{W} = (1, \mathbf{y}')'$ y

$$\begin{aligned} \nabla_{\boldsymbol{\phi}} h(\mathbf{y}, \mathbf{z}_q; \boldsymbol{\beta}, s) &= \frac{\partial}{\partial \boldsymbol{\beta}} h(\mathbf{y}, \mathbf{z}_q; \boldsymbol{\beta}, r) \\ &= (\mathbf{W}' I(z_d \leq r), \mathbf{W}' I(z_d > r))'. \end{aligned}$$

Por lo tanto

$$|\nabla_{\boldsymbol{\phi}} h(\mathbf{y}, \mathbf{z}_q; \boldsymbol{\beta}, r)| = \sqrt{1 + |\mathbf{y}|^2}. \quad (\text{A.4})$$

Tenemos que

$$|h(\mathbf{y}, \mathbf{z}_q; \boldsymbol{\beta}, r)| \leq |\boldsymbol{\beta}| \sqrt{1 + |\mathbf{y}|^2}, \quad (\text{A.5})$$

puesto que

$$h(\mathbf{y}, \mathbf{z}_q; \boldsymbol{\beta}, s) = \boldsymbol{\beta}' \nabla_{\boldsymbol{\beta}} h(\mathbf{y}, \mathbf{z}_q; \boldsymbol{\beta}, s). \quad (\text{A.6})$$

Adicionalmente, para cualquier $s \in \bar{\mathfrak{R}}$, $r \in \bar{\mathfrak{R}}$, $q \in \{1, \dots, p\}$

$$\begin{aligned} |\nabla_{\phi} h(\mathbf{y}, \mathbf{z}_q; \boldsymbol{\beta}, r) - \nabla_{\phi} h(\mathbf{y}, \mathbf{z}_q; \boldsymbol{\beta}, s)| &\leq \sqrt{2(1 + |\mathbf{y}|^2)} I(\min(s, r) < z_q < \max(s, r)) \\ &\leq \sqrt{2(1 + |\mathbf{y}|^2)} I(|z_q - s| \leq |r - s|). \end{aligned} \quad (\text{A.7})$$

Lema A.1. *Bajo las suposiciones del Teorema 3.4, para cualquier $\mathbf{v} = (\boldsymbol{\beta}', r, q)' \in \Lambda$ con $\Lambda = K \times \bar{\mathfrak{R}} \times \{1, 2, \dots, p\}$ y una vecindad abierta $U_{\mathbf{v}}$,*

$$E \left[\sup_{\mathbf{v}^* \in U_{\mathbf{v}}} |\psi(\mathbf{Y}_1, \mathbf{Z}_{2-q}, \varepsilon_2, \boldsymbol{\beta}^*, r^*) - \psi(\mathbf{Y}_1, \mathbf{Z}_{2-q}, \varepsilon_2, \boldsymbol{\beta}, r)| \right] \longrightarrow 0, \quad (\text{A.8})$$

cuando $U_{\mathbf{v}}$ se contrae a \mathbf{v} .

Demostración. Sea

$$U_{\mathbf{v}}(\eta) = \{ \mathbf{v}^* = (\boldsymbol{\beta}^*, r^*, q)' \in \Lambda : |\boldsymbol{\beta}^* - \boldsymbol{\beta}| < \eta, d(r^*, r) < \eta \}, \quad \eta > 0.$$

Sin pérdida de generalidad probaremos el lema para

$$E \sup_{\mathbf{v}^* \in U_{\mathbf{v}}(\eta)} |\psi(\mathbf{Y}_1, \mathbf{Z}_{2-q}, \varepsilon_2, \mathbf{v}^*) - \psi(\mathbf{Y}_1, \mathbf{Z}_{2-q}, \varepsilon_2, \mathbf{v})| \longrightarrow 0 \text{ cuando } \eta \rightarrow 0.$$

Sea $\varepsilon_2(\mathbf{v}) = Y_2 - h(\mathbf{Y}_1, \mathbf{Z}_{2-q}, \boldsymbol{\beta}, r)$ y $\delta(\mathbf{y}, \mathbf{z}_q, \boldsymbol{\beta}^*, r^*) = h(\mathbf{y}, \mathbf{z}_q, \boldsymbol{\beta}, r) - h(\mathbf{y}, \mathbf{z}_q, \boldsymbol{\beta}^*, r^*)$. Para cualquier $\mathbf{v} = (\boldsymbol{\beta}', r, q)'$ y para cualquier $\mathbf{y} \in \mathbb{R}^p$, por (A.4) y (A.6) concluimos que

$$|h(\mathbf{y}, \mathbf{z}_q, \boldsymbol{\beta}, r) - h(\mathbf{y}, \mathbf{z}_q, \boldsymbol{\beta}^*, r^*)| \leq |\boldsymbol{\beta} - \boldsymbol{\beta}^*| \sqrt{1 + |\mathbf{y}|^2}, \quad (\text{A.9})$$

y por ecuaciones (A.5) y (A.7)

$$\begin{aligned} |h(\mathbf{y}, \mathbf{z}_q, \boldsymbol{\beta}^*, r) - h(\mathbf{y}, \mathbf{z}_q, \boldsymbol{\beta}^*, r^*)| &\leq |\boldsymbol{\beta}^*| \sqrt{2(1 + |\mathbf{y}|^2)} I(\min(r, r^*) < z_q < \max(r, r^*)) \\ &\leq |\boldsymbol{\beta}^*| \sqrt{2(1 + |\mathbf{y}|^2)} I(|z_q - r| \leq |r^* - r|). \end{aligned}$$

Entonces para $\mathbf{v}^* \in U_{\mathbf{v}}(\eta)$ y para $r \in \bar{\mathfrak{R}}$, $\mathbf{y} \in \mathbb{R}^p$

$$\begin{aligned} |\delta(\mathbf{y}, \mathbf{z}_q, \mathbf{v}^*)| &= |h(\mathbf{y}, \mathbf{z}_q, \boldsymbol{\beta}, r) - h(\mathbf{y}, \mathbf{z}_q, \boldsymbol{\beta}^*, r^*)| \\ &= |h(\mathbf{y}, \mathbf{z}_q, \boldsymbol{\beta}, r) - h(\mathbf{y}, \mathbf{z}_q; \boldsymbol{\beta}, r^*) + h(\mathbf{y}, \mathbf{z}_q; \boldsymbol{\beta}, r^*) - h(\mathbf{y}, \mathbf{z}_q, \boldsymbol{\beta}^*, r^*)| \\ &\leq |h(\mathbf{y}, \mathbf{z}_q; \boldsymbol{\beta}, r) - h(\mathbf{y}, \mathbf{z}_q; \boldsymbol{\beta}, r^*)| + |h(\mathbf{y}, \mathbf{z}_q; \boldsymbol{\beta}, r^*) - h(\mathbf{y}, \mathbf{z}_q, \boldsymbol{\beta}^*, r^*)| \\ &\leq \left[\sqrt{2} |\boldsymbol{\beta}| I(|z_q - r| \leq |r^* - r|) + |\boldsymbol{\beta} - \boldsymbol{\beta}^*| \right] \sqrt{1 + |\mathbf{y}|^2} \\ &\leq \left[\sqrt{2} |\boldsymbol{\beta}| I(|z_q - r| \leq |r_0(\eta) - r|) + \eta \right] \sqrt{1 + |\mathbf{y}|^2}, \end{aligned} \quad (\text{A.10})$$

donde $r_0(\eta)$ es tal que $d(r_0(\eta), r) = \eta$.

Definamos

$$\Delta(\mathbf{y}, \mathbf{z}_q, \eta) \equiv \left[\sqrt{2} |\boldsymbol{\beta}| I(|z_q - r| \leq |r_0(\eta) - r|) + \eta \right] \sqrt{1 + |\mathbf{y}|^2}. \quad (\text{A.11})$$

Los supuestos SC1 y SC2, $E|\mathbf{Y}_{d-1}| < \infty$ y la ecuación (A.5) implican que existe una constante L tal que para cualquier $\mathbf{v} \in \Lambda$,

$$\begin{aligned} E\varphi^2(\varepsilon_2(\mathbf{v})) &= E\{[\varphi(\varepsilon_2 + h(\mathbf{Y}_1, \mathbf{Z}_{2-d}, \boldsymbol{\phi}, \gamma) - h(\mathbf{Y}_1, \mathbf{Z}_{2-q}, \boldsymbol{\beta}, r)) - \varphi(\varepsilon_2)] + \varphi(\varepsilon_2)\}^2 \\ &= E\{[\varphi(\varepsilon_2 + h(\mathbf{Y}_1, \mathbf{Z}_{2-d}, \boldsymbol{\phi}, \gamma) - h(\mathbf{Y}_1, \mathbf{Z}_{2-q}, \boldsymbol{\beta}, r)) - \varphi(\varepsilon_2)]^2 + \varphi^2(\varepsilon_2) \\ &\quad + 2[\varphi(\varepsilon_2 + h(\mathbf{Y}_1, \mathbf{Z}_{2-d}, \boldsymbol{\phi}, \gamma) - h(\mathbf{Y}_1, \mathbf{Z}_{2-q}, \boldsymbol{\beta}, r)) - \varphi(\varepsilon_2)]\varphi(\varepsilon_2)\} \\ &\leq 2E[\varphi^2(\varepsilon_2) + L|h(\mathbf{Y}_1, \mathbf{Z}_{2-d}, \boldsymbol{\phi}, \gamma) - h(\mathbf{Y}_1, \mathbf{Z}_{2-q}, \boldsymbol{\beta}, r)|^2] \\ &\leq 2I(f) + 4L(|\boldsymbol{\phi}|^2 + |\boldsymbol{\beta}|^2)E(1 + |\mathbf{Y}_1|^2) < \infty \end{aligned} \quad (\text{A.12})$$

Del supuesto SC1 y la ecuación (A.10) se sigue que $\ln f$ es absolutamente continua, de aquí

$$|\psi(\mathbf{Y}_1, \mathbf{Z}_{2-q}, \varepsilon_2, \mathbf{v}^*) - \psi(\mathbf{Y}_1, \mathbf{Z}_{2-q}, \varepsilon_2, \mathbf{v})| \leq \int_{-\Delta(\mathbf{y}, \mathbf{z}_q, \eta)}^{\Delta(\mathbf{y}, \mathbf{z}_q, \eta)} |\varphi(\varepsilon_2(\mathbf{v}) + v)| dv, \quad (\text{A.13})$$

donde $\Delta(\mathbf{y}, \mathbf{z}_q, \eta)$ está definido en (A.11).

Por lo tanto, de las ecuaciones (A.12) y (A.13) y de la desigualdad de Cauchy-Bunyakovski-Schwarz se sigue que

$$\begin{aligned} &E \left[\sup_{\mathbf{v}^* \in U_{\mathbf{v}}} |\psi(\mathbf{Y}_1, \mathbf{Z}_{2-q}, \varepsilon_2, \mathbf{v}^*) - \psi(\mathbf{Y}_1, \mathbf{Z}_{2-q}, \varepsilon_2, \mathbf{v})| \right] \\ &\leq E\{[2\varphi(\varepsilon_2(\mathbf{v}))] + L\Delta(\mathbf{y}, \mathbf{z}_q, \eta)\} \Delta(\mathbf{y}, \mathbf{z}_q, \eta) \\ &\leq 2(E\varphi^2(\varepsilon_2(\mathbf{v})))^{1/2} (E\Delta^2(\mathbf{y}, \mathbf{z}_q, \eta))^{1/2} + LE\Delta^2(\mathbf{y}, \mathbf{z}_q, \eta). \end{aligned}$$

Debido a que

$$\begin{aligned} E\Delta^2(\mathbf{y}, \mathbf{z}_q, \eta) &= E\left\{(1 + |\mathbf{y}|^2) \left[\sqrt{2} |\boldsymbol{\beta}| I(|z_q - r| \leq |r_0(\eta) - r|) + \eta \right]\right\} \\ &\xrightarrow{\eta \rightarrow 0} 0. \end{aligned} \quad (\text{A.14})$$

De (A.12) y (A.14) se sigue (A.8).

Si $r = \infty$, tenemos un resultado similar a (A.10)

$$\begin{aligned} |h(\mathbf{y}, \mathbf{z}_q, \boldsymbol{\beta}^*, r) - h(\mathbf{y}, \mathbf{z}_q, \boldsymbol{\beta}^*, r^*)| &\leq |\boldsymbol{\beta}^*| \sqrt{2(1 + |\mathbf{y}|^2)} I(\min(\infty, r^*) < z_q < \max(\infty, r^*)) \\ &\leq |\boldsymbol{\beta}^*| \sqrt{2(1 + |\mathbf{y}|^2)} I(z_q > r^*). \end{aligned}$$

$$\begin{aligned}
|\delta(\mathbf{y}, \mathbf{z}_q, \mathbf{v}^*)| &= |h(\mathbf{y}, \mathbf{z}_q, \boldsymbol{\beta}, r) - h(\mathbf{y}, \mathbf{z}_q, \boldsymbol{\beta}^*, r^*)| \\
&\leq |h(\mathbf{y}, \mathbf{z}_q; \boldsymbol{\beta}, r) - h(\mathbf{y}, \mathbf{z}_q; \boldsymbol{\beta}, r^*)| + |h(\mathbf{y}, \mathbf{z}_q; \boldsymbol{\beta}, r^*) - h(\mathbf{y}, \mathbf{z}_q; \boldsymbol{\beta}^*, r^*)| \\
&\leq \left[\sqrt{2} |\boldsymbol{\beta}| I(z_q > r^*) + |\boldsymbol{\beta} - \boldsymbol{\beta}^*| \right] \sqrt{1 + |\mathbf{y}|^2} \\
&\leq \left[\sqrt{2} |\boldsymbol{\beta}| I(z_q > r_0(\eta)) + \eta \right] \sqrt{1 + |\mathbf{y}|^2},
\end{aligned}$$

donde $d(r_0(\eta), \infty) = \eta$. De nuevo si $\Delta_1(\mathbf{y}, \mathbf{z}_q, \eta) = \left[\sqrt{2} |\boldsymbol{\beta}| I(z_q > r_0(\eta)) + \eta \right] \sqrt{1 + |\mathbf{y}|^2}$ se puede probar que

$$E\Delta_1^2(\mathbf{y}, \mathbf{z}_q, \eta) \longrightarrow 0 \text{ cuando } \eta \longrightarrow 0.$$

Si $r = -\infty$

$$\begin{aligned}
|h(\mathbf{y}, \mathbf{z}_q, \boldsymbol{\beta}^*, r) - h(\mathbf{y}, \mathbf{z}_q, \boldsymbol{\beta}^*, r^*)| &\leq |\boldsymbol{\beta}^*| \sqrt{2(1 + |\mathbf{y}|^2)} I(\min(-\infty, r^*) < z_q < \max(-\infty, r^*)) \\
&\leq |\boldsymbol{\beta}^*| \sqrt{2(1 + |\mathbf{y}|^2)} I(z_q < r^*).
\end{aligned}$$

$$\begin{aligned}
|\delta(\mathbf{y}, \mathbf{z}_q, \mathbf{v}^*)| &= |h(\mathbf{y}, \mathbf{z}_q, \boldsymbol{\beta}, r) - h(\mathbf{y}, \mathbf{z}_q, \boldsymbol{\beta}^*, r^*)| \\
&\leq \left[\sqrt{2} |\boldsymbol{\beta}| I(z_q < r^*) + |\boldsymbol{\beta} - \boldsymbol{\beta}^*| \right] \sqrt{1 + |\mathbf{y}|^2} \\
&\leq \left[\sqrt{2} |\boldsymbol{\beta}| I(z_q < r_0(\eta)) + \eta \right] \sqrt{1 + |\mathbf{y}|^2},
\end{aligned}$$

Por último si $\Delta_2(\mathbf{y}, \mathbf{z}_q, \eta) = \left[\sqrt{2} |\boldsymbol{\beta}| I(z_q < r_0(\eta)) + \eta \right] \sqrt{1 + |\mathbf{y}|^2}$ se puede probar que

$$E\Delta_2^2(\mathbf{y}, \mathbf{z}_q, \eta) \longrightarrow 0 \text{ cuando } \eta \longrightarrow 0.$$

□

Demostración del Teorema 3.4 con $\sigma = 1$.

Sea $\alpha(\mathbf{v}) = E\psi(\mathbf{Y}_1, \mathbf{Z}_{2-q}, \varepsilon_2, \mathbf{v})$ para $\mathbf{v} \in \mathbf{\Lambda}$. Los supuestos SC1 y SC2, el teorema del valor medio, la independencia de ε_2 y Y_1 y la desigualdad de Cauchy-Bunyakovski-Schwarz implican que $E|(\mathbf{Y}_1, \mathbf{Z}_{2-q}, \varepsilon_2, \mathbf{v})| < \infty$. así, α está bien definida. Se cumple que $\alpha(\boldsymbol{\vartheta}) = 0$ y que $\ln y \leq y - 1$, $y \neq 1$. Para cualquier vecindad abierta V de $\boldsymbol{\vartheta}$ en $\mathbf{\Lambda}$ y cualquier $\mathbf{v} \in V^c$,

un argumento condicional da

$$\begin{aligned}
\alpha(\mathbf{v}) &= E\psi(\mathbf{Y}_1, \mathbf{Z}_{2-q}, \varepsilon_2, \mathbf{v}) \\
&= E \ln \frac{f(\varepsilon_t + h(\mathbf{Y}_1, \mathbf{Z}_{2-q}, \boldsymbol{\vartheta}) - h(\mathbf{Y}_1, \mathbf{Z}_{2-q}, \mathbf{v}))}{f(\varepsilon_t)} \\
&= E \left\{ E \left[\ln \frac{f(\varepsilon_t + h(\mathbf{Y}_1, \mathbf{Z}_{2-q}, \boldsymbol{\vartheta}) - h(\mathbf{Y}_1, \mathbf{Z}_{2-q}, \mathbf{v}))}{f(\varepsilon_t)} \middle| \mathbf{Y}_1, \mathbf{Z}_{2-q} \right] \right\} \\
&< E \left\{ \int [f(y + h(\mathbf{Y}_1, \mathbf{Z}_{2-q}, \boldsymbol{\vartheta}) - h(\mathbf{Y}_1, \mathbf{Z}_{2-q}, \mathbf{v})) - f(y)] dy \right\} \\
&= 0.
\end{aligned}$$

Por el Lema A.1, la función α es continua y de aquí, por la compacidad de V^c , existe $\mathbf{v}_0 \in V^c$, tal que

$$\sup_{\mathbf{v} \in V^c} \alpha(\mathbf{v}) = \alpha(\mathbf{v}_0) < 0.$$

Sea $\delta_0 = -\alpha(\mathbf{v}_0)/3$. Para cualquier $\mathbf{v} \in V^c$, por el Lema A.1 tenemos que existe $\eta_0 > 0$ tal que

$$\begin{aligned}
E \sup_{\mathbf{v}^* \in U_{\mathbf{v}}(\eta_0)} \psi(\mathbf{Y}_1, \mathbf{Z}_{2-q}, \varepsilon_2, \mathbf{v}^*) &\leq E\psi(\mathbf{Y}_1, \mathbf{Z}_{2-q}, \varepsilon_2, \mathbf{v}) + \delta_0 \\
&\leq \alpha(\mathbf{v}_0) + \delta_0 \\
&= -2\delta_0.
\end{aligned} \tag{A.15}$$

La compacidad de V^c implica que existe un número finito M de vecindades $U_{\mathbf{v}_j}(\eta_0)$, $\mathbf{v}_j \in V^c$, $j = 1, 2, \dots, M$ tal que

$$\bigcup U_{\mathbf{v}_j}(\eta_0) = V^c.$$

Luego, por el Teorema de ergodicidad y por (A.15), existe un T_0 tal que para cualquier $T \geq T_0$, $1 \leq j \leq M$,

$$\begin{aligned}
\sup_{\mathbf{v}^* \in U_{\mathbf{v}_j}(\eta_0)} \ell_T(\mathbf{v}^*) &\leq \frac{1}{T} \sum \sup_{\mathbf{v}^* \in U_{\mathbf{v}_j}(\eta_0)} \psi(Y_{t-1}, Z_{t-d}, \varepsilon_t, \boldsymbol{\beta}^*, r^*) \\
&\leq E \sup_{\mathbf{v}^* \in U_{\mathbf{v}_j}(\eta_0)} \psi(\mathbf{Y}_1, \mathbf{Z}_{2-q}, \varepsilon_2, \mathbf{v}^*) + \delta_0 \\
&\leq -\delta_0, \quad \text{c.s.}
\end{aligned}$$

Pero

$$\sup_{\mathbf{v} \in V} \ell_T(\mathbf{v}) \geq \ell_T(\boldsymbol{\vartheta}) = 0.$$

Entonces para cualquier vecindad V de $\boldsymbol{\vartheta}$ en $\boldsymbol{\Lambda}$, existe T_0 tal que para toda $T \geq T_0$,

$$\begin{aligned} \sup_{\boldsymbol{v}^* \in V^c} \ell_T(\boldsymbol{v}^*) &\leq \max_{1 \leq j \leq M} \sup_{\boldsymbol{v}^* \in U_{\boldsymbol{v}_j}(\eta_0)} \ell_T(\boldsymbol{v}^*) \\ &\leq -\delta_0 \\ &< 0 \\ &\leq \sup_{\boldsymbol{v} \in V} \ell_T(\boldsymbol{v}). \end{aligned}$$

Lo cual implica que

$$\hat{\boldsymbol{\vartheta}}_T \in V, \quad \text{c.s. para toda } V \quad \text{y para toda } T \geq T_0.$$

Debido a que V es arbitrario se sigue que $\hat{\boldsymbol{\vartheta}}_T \rightarrow \boldsymbol{\vartheta}$ c.s. □

Demostración del Teorema 3.4 con $\sigma > 0$.

Sean $\boldsymbol{\vartheta} = (\boldsymbol{\phi}', \gamma, d, \sigma)$ y $\boldsymbol{v} = (\boldsymbol{\beta}', r, q, \tau)$.

Se puede obtener la demostración del Lema A.1. con $\sigma > 0$ de manera similar las demostración mostrada. También se puede generalizar la demostración del caso $\sigma = 1$ al caso $\sigma > 0$ con los siguientes modificaciones.

Defina

$$\ell_T(\boldsymbol{v}) = \ell_T(\boldsymbol{\beta}, r, q, \tau) = \frac{1}{T} \sum \ln \frac{\frac{1}{\tau} f\left(\frac{Y_t - h(Y_{t-1}, Z_{t-q}, \boldsymbol{\beta}, r)}{\tau}\right)}{\frac{1}{\sigma} f\left(\frac{Y_t - h(Y_{t-1}, Z_{t-d}, \boldsymbol{\phi}, \gamma)}{\sigma}\right)}, \quad \boldsymbol{v} \in \boldsymbol{\Lambda} \times \Re$$

y

$$\psi(Y_{t-1}, Z_{t-q}, \varepsilon_t, \boldsymbol{\beta}, r) = \ln \frac{\frac{1}{\tau} f(\sigma \varepsilon_t + h(Y_{t-1}, Z_{t-d}, \boldsymbol{\phi}, \gamma) - h(Y_{t-1}, Z_{t-q}, \boldsymbol{\beta}, r))}{\frac{1}{\sigma} f(\sigma \varepsilon_t)}, \quad 1 \leq t \leq n.$$

Dado que si $\varepsilon_t^* = \sigma \varepsilon_t$ y f es la función de densidad de ε_t entonces la función de densidad de ε_t^* está dada por $f^* = \frac{1}{\sigma} f(y/\sigma)$., f^* definida así cumple los supuestos SC1-SC4. □

Apéndice B

Condiciones de regularidad, modelo de cuantiles

Supuestos para consistencia

Sean $f_t(\boldsymbol{\beta}) = f_t(\boldsymbol{\beta}, Y_{t-1}, Z_{t-1})$ y Ω_t el conjunto de información disponible hasta el tiempo t .

C0. (Ω, F, P) es un espacio de probabilidad completo y $\{\varepsilon_{\theta t}, \mathbf{x}_t\}$, $t = 1, 2, \dots$, son vectores aleatorios en este espacio.

C1. La función $f_t(\boldsymbol{\beta}) : \Re^{k_t} \times B \longrightarrow \Re$ es tal que para cada $\boldsymbol{\beta} \in B$, con B un subconjunto compacto de \Re^p , $f_t(\boldsymbol{\beta})$ es medible con respecto al conjunto de información Ω_t y $f_t(\cdot)$ es continua en B , $t = 1, 2, \dots$, para una elección dada de variables explicatorias $\{y_{t-1}, \mathbf{x}_{t-1}, \dots, y_1, \mathbf{x}_1\}$.

C2. Condicional en el conjunto de información pasada Ω_t , el término de error $\varepsilon_{\theta t}$ forma un proceso estacionario, con densidad condicional continua $h_t(\varepsilon|\Omega_t)$.

C3. Existe $h > 0$ tal que para todo t , $h_t(0|\Omega_t) \geq h$.

C4. $|f_t(\boldsymbol{\beta})| < K(\Omega_t)$ para cada $\boldsymbol{\beta} \in B$, y para todo t , donde $K(\Omega_t)$ es alguna función posiblemente estocástica de variables que pertenecen al conjunto de información, tal que $E|K(\Omega_t)| \leq K_0 < \infty$ para alguna constante K_0 .

C5. $E[|\varepsilon_{\theta t}|] < \infty$ para todo t .

C6. $\{[\theta - I(y_t < f_t(\boldsymbol{\beta}))][y_t - f_t(\boldsymbol{\beta})]\}$ obedece la Ley Uniforme de los Grandes Números.

C7. Para cada $\xi > 0$, existe un $\tau > 0$ tal que si $\|\boldsymbol{\beta} - \boldsymbol{\beta}^0\| \geq \xi$, entonces el

$$\liminf_{T \rightarrow \infty} T^{-1} \sum [P|f_t(\boldsymbol{\beta}) - f_t(\boldsymbol{\beta}^0) > \tau|] > 0.$$

Supuestos para normalidad asintótica

AN1. $f_t(\beta)$ es diferenciable en B y para todo β y γ en una vecindad v_0 de β^0 , tal que $\|\beta - \gamma\| \leq d$ para d suficientemente pequeño y para todo t .

(a) $\|\nabla f_t(\beta)\| \leq F(\Omega_t)$, donde $F(\Omega_t)$ es alguna función estocástica de variables que pertenecen a el conjunto de información y $E[F(\Omega_t)^3] \leq F_0 < \infty$, para alguna constante F_0 .

(b) $\|\nabla f_t(\beta) - \nabla f_t(\gamma)\| \leq M(\Omega_t, \beta, \gamma) = O(\|\beta - \gamma\|)$, donde $M(\Omega_t, \beta, \gamma)$ es alguna función tal que $E[M(\Omega_t, \beta, \gamma)]^2 \leq M_0 \|\beta - \gamma\| < \infty$ y

$$E[M(\Omega_t, \beta, \gamma) F(\Omega_t)] \leq M_1 \|\beta - \gamma\| < \infty \text{ para alguna constante } M_0 \text{ y } M_1.$$

AN2.

(a) $h_t(\varepsilon|\Omega_t) \leq N < \infty \quad \forall t$, para alguna constante N .

(b) $h_t(\varepsilon|\Omega_t)$ satisface la condición de Lipschitz $|h_t(\lambda_1|\Omega_t) - h_t(\lambda_2|\Omega_t)|$

$$\leq L(\lambda_1 - \lambda_2) \text{ para alguna constante } L < \infty \quad \forall t.$$

AN3. Las matrices $\mathbf{A}_T \equiv E \left[T^{-1} \theta (1 - \theta) \sum_{t=1}^T \nabla' f_t(\beta^0) \nabla f_t(\beta^0) \right]$ y

$\mathbf{D}_T \equiv E \left[T^{-1} \sum_{t=1}^T h_t(0|\Omega_t) \nabla' f_t(\beta^0) \nabla f_t(\beta^0) \right]$ tienen los eigenvalores más pequeños acotados por abajo por una constante positiva para T suficientemente grande.

AN4. La secuencia $\left\{ T^{-1/2} \sum_{t=1}^T [\theta - I(y_t < f_t(\beta^0))] \nabla' f_t(\beta^0) \right\}$ obedece el Teorema Central del Límite.

Supuestos para la estimación de la matriz de varianzas y covarianzas.

VC1. $\hat{c}_T/c_T \xrightarrow{p} 1$, donde la secuencia positiva no estocástica c_T satisface que $c_T = o(1)$ y $c_T^{-1} = o(T^{1/2})$.

VC2. $E[|F(\Omega_t)|^4] \leq F_1 < \infty$ para todo t y para alguna constante F_1 , donde $F(\Omega_t)$ se definió en el supuesto AN1 (a).

VC3. $T^{-1} \theta (1 - \theta) \sum_{t=1}^T \nabla' f_t(\beta^0) \nabla f_t(\beta^0) - \mathbf{A}_T \xrightarrow{p} \mathbf{0}$ y

$$T^{-1} \sum_{t=1}^T h_t(0|\Omega_t) \nabla' f_t(\beta^0) \nabla f_t(\beta^0) - \mathbf{D}_T \xrightarrow{p} \mathbf{0}.$$

Supuestos para la prueba de cuantil dinámico dentro de la muestra (CD_{DM}).

DQ1. $\mathbf{X}_t(\boldsymbol{\beta})$ posee elementos diferentes de $\nabla f_t(\boldsymbol{\beta})$, es Ω_t medible y $\|\mathbf{X}_t(\boldsymbol{\beta})\| \leq W(\Omega_t)$, donde $W(\Omega_t)$ es alguna función (posiblemente) estocástica de variables que pertenecen al conjunto de información, tal que:

$E[W(\Omega_t) \times M(\Omega_t, \boldsymbol{\beta}, \boldsymbol{\gamma})] \leq W_0 \|\boldsymbol{\beta} - \boldsymbol{\gamma}\| < \infty$ y $E\{[W(\Omega_t) \times F(\Omega_t)]^2\} < W_1 < \infty$ para algunas constantes finitas W_0 y W_1 , y $F(\Omega_t)$ y $M(\Omega_t, \boldsymbol{\beta}, \boldsymbol{\gamma})$ se definen en AN1.

DQ2. $\|\mathbf{X}_t(\boldsymbol{\beta}) - \mathbf{X}_t(\boldsymbol{\gamma})\| \leq S(\Omega_t, \boldsymbol{\beta}, \boldsymbol{\gamma})$, donde $E[S(\Omega_t, \boldsymbol{\beta}, \boldsymbol{\gamma})] \leq S_0 \|\boldsymbol{\beta} - \boldsymbol{\gamma}\| < \infty$, $E[W(\Omega_t) S(\Omega_t, \boldsymbol{\beta}, \boldsymbol{\gamma})] \leq S_1 \|\boldsymbol{\beta} - \boldsymbol{\gamma}\|$ para algunas constantes S_0 y S_1 .

DQ3. Sea $\{\varepsilon_t^1, \dots, \varepsilon_t^{j_i}\}$ el conjunto de valores para el cual $\mathbf{X}_t(\boldsymbol{\beta})$ es no diferenciable, entonces $\Pr(\varepsilon_{\theta t} = \varepsilon_t^j) = 0$ para $j = 1, \dots, j_i$. Siempre que la derivada exista, $\|\nabla \mathbf{X}_t(\boldsymbol{\beta})\| \leq Z(\Omega_t)$, donde $Z(\Omega_t)$ es alguna función (posiblemente) estocástica de variables que pertenecen al conjunto de información, tal que $E[Z(\Omega_t)^r] < Z_0 < \infty$, $r = 1, 2$, y Z_0 alguna constante.

DQ4. $T^{-1} \mathbf{X}'(\boldsymbol{\beta}^0) \mathbf{H} \nabla f(\boldsymbol{\beta}^0) - E[T^{-1} \mathbf{X}'(\boldsymbol{\beta}^0) \mathbf{H} \nabla f(\boldsymbol{\beta}^0)] \xrightarrow{p} \mathbf{0}$.

DQ5. $T^{-1} \mathbf{M}_T \mathbf{M}'_T - E(T^{-1} \mathbf{M}_T \mathbf{M}'_T) \xrightarrow{p} \mathbf{0}$ donde

$$\mathbf{M}_T \equiv \mathbf{X}'(\boldsymbol{\beta}^0) - E[T^{-1} \mathbf{X}'(\boldsymbol{\beta}^0) \mathbf{H} \nabla f(\boldsymbol{\beta}^0)] \mathbf{D}_T^{-1} \nabla' f(\boldsymbol{\beta}^0).$$

DQ6. La secuencia $\{T^{-1} \mathbf{M}_T \mathbf{H} \mathbf{I} \mathbf{t}(\boldsymbol{\beta}^0)\}$ obedece el teorema de límite central.

DQ7. $T^{-1} E(\mathbf{M}_T \mathbf{M}'_T)$ es una matriz no singular.

Supuestos para la prueba de cuantil dinámico fuera de la muestra (CD_{FM}).

DQ8. $\lim_{R \rightarrow \infty} T_R = \infty$, $\lim_{R \rightarrow \infty} N_R = \infty$ y $\lim_{R \rightarrow \infty} \frac{N_R}{T_R} = 0$.

DQ9. La secuencia $[N_R^{-1/2} \mathbf{X}'(\boldsymbol{\beta}^0) \mathbf{H} \mathbf{I} \mathbf{t}(\boldsymbol{\beta}^0)]$ obedece el teorema de límite central.

Apéndice C

Pruebas de los Teoremas 6.1 a 6.5 del Capítulo 6

Teorema 6.1 (Consistencia). *En el modelo (6.3) y bajo los supuestos C0 a C7 enunciados en el Apéndice B, se cumple que $\hat{\beta} \xrightarrow{p} \beta^0$, donde $\hat{\beta}$ es la solución de la siguiente función:*

$$\min_{\beta} T^{-1} \sum_{t=1}^T [\theta - I(y_t < f_t(\beta))] [y_t - f_t(\beta)].$$

Demostración. Este resultado es una consecuencia del Corolario 5.12 de White (1996, página 75), el cual dice lo siguiente: Sea $Q_T(\beta) \equiv T^{-1} \sum_{t=1}^T q_t(\beta)$, donde $q_t(\beta) \equiv [\theta - I(y_t < f_t(\beta))] [y_t - f_t(\beta)]$. Bajo el supuesto C0 (los datos observados son una realización de un proceso estocástico en el espacio de probabilidad completo (Ω, F, P)), y sea $f_t(\beta)$ que cumple C1, C4, C6 y C7, y $\hat{\beta}$ solución de (6.4). Si se satisfacen los supuestos, 3.1 y 3.2 del Corolario de White, los cuales dicen que:

Supuesto 3.1. Para cada β en B $E[q_t(\beta)]$ existe y es finito para $t = 1, 2, \dots$;

(a) $E[q_t(\beta)]$ es continuo en B , $t = 1, 2, \dots$; y

(b) $\{q_t(\beta)\}$ obedece la Ley Uniforme de los Grandes Números (supuesto C6).

Supuesto 3.2. $E[Q_T(\beta) - Q_T(\beta^0)]$ es únicamente minimizado en β^0 para T suficientemente grande. Entonces $\hat{\beta} \xrightarrow{p} \beta^0$ cuando $T \rightarrow \infty$.

Primero necesitamos mostrar que $E[q_t(\beta)]$ existe y es finito cada β . Esto es como sigue:

$$\begin{aligned} E[q_t(\beta)] &< E[y_t - f_t(\beta)] \\ &= E[y_t - f_t(\beta^0) + f_t(\beta^0) - f_t(\beta)] \\ &\leq E|\varepsilon_{\theta t}| + E|f_t(\beta)| + E|f_t(\beta^0)| < \infty, \end{aligned}$$

por los supuestos C4 y C5. Se tiene además que, $E[q_t(\beta)]$ es continuo en β , ya que por el supuesto C1 f_t es continua en β y $q_t(\beta)$ es continua en $f_t(\beta)$.

Resta mostrar que $E[V_T(\beta)] = E[Q_T(\beta) - Q_T(\beta^0)]$ se minimiza de forma única en β^0 para T suficientemente grande. Sea

$$v_T(\beta) \equiv q_T(\beta) - q_T(\beta^0),$$

observemos que $q_T(\beta)$ puede reexpresarse como:

$$\begin{aligned} q_t(\beta) &\equiv \{\theta - I[y_t < f_t(\beta)]\} [y_t - f_t(\beta)] \\ &= \{\theta - I[y_t - f_t(\beta^0) < f_t(\beta) - f_t(\beta^0)]\} \\ &\quad \times \{y_t - f_t(\beta^0) - [f_t(\beta) - f_t(\beta^0)]\} \\ &= \{\theta - I[\varepsilon_{\theta t} < \tau_t(\beta)]\} \{\varepsilon_{\theta t} - \tau_t(\beta)\}, \end{aligned}$$

con $\tau_t(\beta) = f_t(\beta) - f_t(\beta^0)$. Entonces,

$$\begin{aligned} v_T(\beta) &= \{\theta - I[\varepsilon_{\theta t} < \tau_t(\beta)]\} \{\varepsilon_{\theta t} - \tau_t(\beta)\} - [\theta - I(\varepsilon_{\theta t} < 0)] \varepsilon_{\theta t} \\ v_T(\beta) &= \begin{cases} (\theta - 1)[\varepsilon_{\theta t} - \tau_t(\beta)] - (\theta - 1)\varepsilon_{\theta t} & \text{si } \varepsilon_{\theta t} < \tau_t(\beta) \text{ y } \varepsilon_{\theta t} < 0 \\ (\theta - 1)[\varepsilon_{\theta t} - \tau_t(\beta)] - \theta\varepsilon_{\theta t} & \text{si } \varepsilon_{\theta t} < \tau_t(\beta) \text{ y } \varepsilon_{\theta t} > 0 \\ \theta[\varepsilon_{\theta t} - \tau_t(\beta)] - (\theta - 1)\varepsilon_{\theta t} & \text{si } \varepsilon_{\theta t} > \tau_t(\beta) \text{ y } \varepsilon_{\theta t} < 0 \\ \theta[\varepsilon_{\theta t} - \tau_t(\beta)] - \theta\varepsilon_{\theta t} & \text{si } \varepsilon_{\theta t} > \tau_t(\beta) \text{ y } \varepsilon_{\theta t} > 0 \end{cases} \\ v_T(\beta) &= \begin{cases} (1 - \theta)\tau_t(\beta) & \text{si } \varepsilon_{\theta t} < \tau_t(\beta) \text{ y } \varepsilon_{\theta t} < 0 \\ (1 - \theta)\tau_t(\beta) - \varepsilon_{\theta t} & \text{si } \varepsilon_{\theta t} < \tau_t(\beta) \text{ y } \varepsilon_{\theta t} > 0 \\ \varepsilon_{\theta t} - \theta\tau_t(\beta) & \text{si } \varepsilon_{\theta t} > \tau_t(\beta) \text{ y } \varepsilon_{\theta t} < 0 \\ -\theta\tau_t(\beta) & \text{si } \varepsilon_{\theta t} > \tau_t(\beta) \text{ y } \varepsilon_{\theta t} > 0 \end{cases}. \end{aligned}$$

Luego el valor esperado de $v_T(\beta)$ condicional a la información se obtiene como sigue:

$$E[v_T(\beta) | \Omega_t] = E \left[\begin{array}{cc|c} (1 - \theta)\tau_t(\beta) & \text{si } \varepsilon_{\theta t} < \tau_t(\beta) \text{ y } \varepsilon_{\theta t} < 0 & \\ (1 - \theta)\tau_t(\beta) - \varepsilon_{\theta t} & \text{si } \varepsilon_{\theta t} < \tau_t(\beta) \text{ y } \varepsilon_{\theta t} > 0 & \\ \varepsilon_{\theta t} - \theta\tau_t(\beta) & \text{si } \varepsilon_{\theta t} > \tau_t(\beta) \text{ y } \varepsilon_{\theta t} < 0 & \\ -\theta\tau_t(\beta) & \text{si } \varepsilon_{\theta t} > \tau_t(\beta) \text{ y } \varepsilon_{\theta t} > 0 & \end{array} \middle| \Omega_t \right].$$

Supongamos que $\tau_t(\beta) < 0$, entonces se sigue lo siguiente:

$$\begin{aligned} \varepsilon_{\theta t} &< \tau_t(\beta) < 0 \Rightarrow -\infty < \varepsilon_{\theta t} < \tau_t(\beta) \\ \varepsilon_{\theta t} &< \tau_t(\beta) < 0 \text{ y } \varepsilon_{\theta t} > 0 &\text{ es el vacío} \\ \tau_t(\beta) &< \varepsilon_{\theta t} < 0 \\ \varepsilon_{\theta t} &> 0 > \tau_t(\beta) \Rightarrow 0 < \varepsilon_{\theta t} < \infty \end{aligned}$$

y de aquí que:

$$\begin{aligned}
& E[v_T(\beta) | \Omega_t] \\
&= \int_{-\infty}^{\tau_t(\beta)} (1 - \theta) \tau_t(\beta) h_t(\lambda | \Omega_t) d\lambda + \int_{\tau_t(\beta)}^0 [\lambda - \theta \tau_t(\beta)] h_t(\lambda | \Omega_t) d\lambda \\
&\quad - \int_0^{\infty} \theta \tau_t(\beta) h_t(\lambda | \Omega_t) d\lambda \\
&= \tau_t(\beta) \int_{-\infty}^{\tau_t(\beta)} h_t(\lambda | \Omega_t) d\lambda - \theta \tau_t(\beta) \int_{-\infty}^{\tau_t(\beta)} h_t(\lambda | \Omega_t) d\lambda \\
&\quad + \int_{\tau_t(\beta)}^0 \lambda h_t(\lambda | \Omega_t) d\lambda - \theta \tau_t(\beta) \int_{\tau_t(\beta)}^0 h_t(\lambda | \Omega_t) d\lambda \\
&\quad - \theta \tau_t(\beta) \int_0^{\infty} h_t(\lambda | \Omega_t) d\lambda \\
&= \tau_t(\beta) \int_{-\infty}^{\tau_t(\beta)} h_t(\lambda | \Omega_t) d\lambda - \theta \tau_t(\beta) \int_{-\infty}^0 h_t(\lambda | \Omega_t) d\lambda \\
&\quad + \int_{\tau_t(\beta)}^0 \lambda h_t(\lambda | \Omega_t) d\lambda + \theta \tau_t(\beta) \int_{-\infty}^0 h_t(\lambda | \Omega_t) d\lambda \\
&= \tau_t(\beta) \int_{-\infty}^{\tau_t(\beta)} h_t(\lambda | \Omega_t) d\lambda + \int_{\tau_t(\beta)}^0 \lambda h_t(\lambda | \Omega_t) d\lambda \\
&= \tau_t(\beta) \int_{-\infty}^0 h_t(\lambda | \Omega_t) d\lambda - \tau_t(\beta) \int_{\tau_t(\beta)}^0 h_t(\lambda | \Omega_t) d\lambda \\
&\quad + \int_{\tau_t(\beta)}^0 \lambda h_t(\lambda | \Omega_t) d\lambda.
\end{aligned}$$

Supongamos ahora que $\tau_t(\beta) > 0$, de forma tal que:

$$\begin{aligned}
\varepsilon_{\theta t} &< \tau_t(\beta) > 0 \text{ y } \varepsilon_{\theta t} < 0 \Rightarrow -\infty < \varepsilon_{\theta t} < 0 \\
0 &< \varepsilon_{\theta t} < \tau_t(\beta) \\
0 &< \tau_t(\beta) < \varepsilon_{\theta t} < 0 \text{ vacío} \\
\varepsilon_{\theta t} &> \tau_t(\beta) > 0 \Rightarrow \tau_t(\beta) < \varepsilon_{\theta t} < \infty
\end{aligned}$$

y

$$\begin{aligned}
& E[v_T(\beta) | \Omega_t] \\
= & \int_{-\infty}^0 (1 - \theta) \tau_t(\beta) h_t(\lambda | \Omega_t) d\lambda + \int_0^{\tau_t(\beta)} [(1 - \theta) \tau_t(\beta) - \lambda] h_t(\lambda | \Omega_t) d\lambda \\
& - \int_{\tau_t(\beta)}^{\infty} \theta \tau_t(\beta) h_t(\lambda | \Omega_t) d\lambda \\
= & \tau_t(\beta) \int_{-\infty}^0 h_t(\lambda | \Omega_t) d\lambda - \theta \tau_t(\beta) \int_{-\infty}^0 h_t(\lambda | \Omega_t) d\lambda \\
& + \tau_t(\beta) \int_0^{\tau_t(\beta)} h_t(\lambda | \Omega_t) d\lambda - \theta \tau_t(\beta) \int_0^{\tau_t(\beta)} h_t(\lambda | \Omega_t) d\lambda \\
& - \int_0^{\tau_t(\beta)} \lambda h_t(\lambda | \Omega_t) d\lambda - \theta \tau_t(\beta) \int_{\tau_t(\beta)}^{\infty} h_t(\lambda | \Omega_t) d\lambda \\
= & \tau_t(\beta) \int_{-\infty}^0 h_t(\lambda | \Omega_t) d\lambda - \theta \tau_t(\beta) \int_{-\infty}^0 h_t(\lambda | \Omega_t) d\lambda \\
& + \tau_t(\beta) \int_0^{\tau_t(\beta)} h_t(\lambda | \Omega_t) d\lambda - \theta \tau_t(\beta) \int_0^{\infty} h_t(\lambda | \Omega_t) d\lambda \\
& - \int_0^{\tau_t(\beta)} \lambda h_t(\lambda | \Omega_t) d\lambda \\
= & \tau_t(\beta) \int_{-\infty}^0 h_t(\lambda | \Omega_t) d\lambda - \theta \tau_t(\beta) \int_{-\infty}^0 h_t(\lambda | \Omega_t) d\lambda \\
& + \tau_t(\beta) \int_0^{\tau_t(\beta)} h_t(\lambda | \Omega_t) d\lambda + \theta \tau_t(\beta) \int_{-\infty}^0 h_t(\lambda | \Omega_t) d\lambda \\
& - \int_0^{\tau_t(\beta)} \lambda h_t(\lambda | \Omega_t) d\lambda \\
= & \tau_t(\beta) \int_{-\infty}^0 h_t(\lambda | \Omega_t) d\lambda + \tau_t(\beta) \int_0^{\tau_t(\beta)} h_t(\lambda | \Omega_t) d\lambda \\
& - \int_0^{\tau_t(\beta)} \lambda h_t(\lambda | \Omega_t) d\lambda
\end{aligned}$$

finalmente podemos expresar el valor esperado de $v_T(\beta) | \Omega_t$ como sigue:

$$\begin{aligned}
& E[v_T(\beta) | \Omega_t] \\
&= -|\tau_t(\beta)| \int_{-\infty}^0 h_t(\lambda | \Omega_t) d\lambda + I[\tau_t(\beta) < 0] \int_{-|\tau_t(\beta)|}^0 [\lambda + |\tau_t(\beta)|] h_t(\lambda | \Omega_t) d\lambda \\
&\quad + |\tau_t(\beta)| \int_{-\infty}^0 h_t(\lambda | \Omega_t) d\lambda + I[\tau_t(\beta) > 0] \int_0^{|\tau_t(\beta)|} [|\tau_t(\beta)| - \lambda] h_t(\lambda | \Omega_t) d\lambda \\
&= I[\tau_t(\beta) < 0] \int_{-|\tau_t(\beta)|}^0 [\lambda + |\tau_t(\beta)|] h_t(\lambda | \Omega_t) d\lambda + I[\tau_t(\beta) > 0] \int_0^{|\tau_t(\beta)|} [|\tau_t(\beta)| - \lambda] \\
&\quad \times h_t(\lambda | \Omega_t) d\lambda.
\end{aligned}$$

Dado que $h_t(\cdot | \Omega_t)$ es una función continua y por el supuesto C3, el cual asegura que existe una $h > 0$ tal que para todo t , $h_t(0 | \Omega_t) \geq h$, implican que existe un $h_1 > 0$ tal que $h_t(\lambda | \Omega_t) > h_1$ siempre que $|\lambda| < h_1$. De lo anterior para cualquier $0 < \tau < h_1$, se cumple la siguiente desigualdad

$$\begin{aligned}
E[v_T(\beta) | \Omega_t] &\geq I[\tau_t(\beta) < -\tau] \int_{-\tau}^0 [\lambda + \tau] h_1 d\lambda + I[\tau_t(\beta) > \tau] \int_0^{\tau} [\tau - \lambda] h_1 d\lambda \\
&= I[\tau_t(\beta) < -\tau] h_1 \left\{ \int_{-\tau}^0 \lambda d\lambda + \tau \int_{-\tau}^0 d\lambda \right\} + I[\tau_t(\beta) > \tau] \\
&\quad \times h_1 \left\{ \int_0^{\tau} \tau d\lambda - \int_0^{\tau} \lambda d\lambda \right\} \\
&= I[\tau_t(\beta) < -\tau] h_1 \left\{ \frac{\lambda^2}{2} \Big|_{-\tau}^0 + \tau^2 \right\} + I[\tau_t(\beta) > \tau] h_1 \left\{ \tau^2 - \frac{\lambda^2}{2} \Big|_0^{\tau} \right\} \\
&= I[\tau_t(\beta) < -\tau] h_1 \left\{ -\frac{\tau^2}{2} + \tau^2 \right\} + I[\tau_t(\beta) > \tau] h_1 \left\{ \tau^2 - \frac{\tau^2}{2} \right\} \\
&= I[\tau_t(\beta) < -\tau] h_1 \frac{\tau^2}{2} + I[\tau_t(\beta) > \tau] h_1 \frac{\tau^2}{2} \\
&= \frac{\tau^2}{2} h_1 I[|\tau_t(\beta)| > \tau].
\end{aligned}$$

Tomando la esperanza no condicional, se sigue que:

$$\begin{aligned}
 E \{ E [v_T (\boldsymbol{\beta}) | \Omega_t] \} &= E [v_T (\boldsymbol{\beta})] \\
 &\geq E \left\{ \frac{\tau^2}{2} h_1 I [|\tau_t (\boldsymbol{\beta})| > \tau] \right\} \\
 &= \frac{\tau^2}{2} h_1 \Pr [|\tau_t (\boldsymbol{\beta})| > \tau] \\
 &= \frac{\tau^2}{2} h_1 \Pr [|f_t (\boldsymbol{\beta}) - f_t (\boldsymbol{\beta}^0)| > \tau].
 \end{aligned}$$

luego,

$$\begin{aligned}
 E [V_T (\boldsymbol{\beta})] &\equiv E \left[T^{-1} \sum_{t=1}^T v_t (\boldsymbol{\beta}) \right] \\
 &\geq \left\{ \frac{\tau^2}{2} h_1 T^{-1} \sum_{t=1}^T \Pr [|f_t (\boldsymbol{\beta}) - f_t (\boldsymbol{\beta}^0)| > \tau] \right\} \\
 &> 0,
 \end{aligned}$$

ya que por el supuesto C7 si $\|\boldsymbol{\beta} - \boldsymbol{\beta}^0\| \geq \xi$ entonces

$$\liminf_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T \Pr [|f_t (\boldsymbol{\beta}) - f_t (\boldsymbol{\beta}^0)| > \tau] > 0,$$

de donde concluimos que $E [V_T (\boldsymbol{\beta})] = E [Q_T (\boldsymbol{\beta}) - Q_T (\boldsymbol{\beta}^0)]$ se minimiza de forma única en $\boldsymbol{\beta}^0$ para T suficientemente grande. \square

Teorema 6.2 (Normalidad asintótica). *En el modelo (6.3) y bajo los supuestos AN1 a AN4 en el Apéndice B y las condiciones del Teorema 6.1., tenemos el siguiente resultado asintótico:*

$$\sqrt{T} \mathbf{A}_T^{-1/2} \mathbf{D}_T (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0) \xrightarrow{d} N(\mathbf{0}, \mathbf{I}),$$

donde

$$\begin{aligned}
 \mathbf{A}_T &\equiv E \left[T^{-1} \theta (1 - \theta) \sum_{t=1}^T \nabla' f_t (\boldsymbol{\beta}^0) \nabla f_t (\boldsymbol{\beta}^0) \right], \\
 \mathbf{D}_T &\equiv E \left[T^{-1} \sum_{t=1}^T h_t (0 | \Omega_t) \nabla' f_t (\boldsymbol{\beta}^0) \nabla f_t (\boldsymbol{\beta}^0) \right],
 \end{aligned}$$

y $\hat{\boldsymbol{\beta}}$ estimado como en el Teorema 6.1..

Demostración. Definamos $Hit_t (\boldsymbol{\beta}) \equiv I [y_t < f_t (\boldsymbol{\beta})] - \theta$ y $g_t (\boldsymbol{\beta}) \equiv Hit_t (\boldsymbol{\beta}) \nabla' f_t (\boldsymbol{\beta})$,

$\lambda_T (\boldsymbol{\beta}) \equiv T^{-1} \sum_{t=1}^T E [g_t (\boldsymbol{\beta})]$ y $\mu_t (\boldsymbol{\beta}, d) \equiv \sup_{\|\boldsymbol{\tau} - \boldsymbol{\beta}\| \leq d} \|g_t (\boldsymbol{\tau}) - g_t (\boldsymbol{\beta})\|$. Siguiendo la estrategia de Ruppert y Carroll (1980), Lemas A1 y A2, sea $\{\mathbf{e}_j\}_{j=1}^p$ una base estándar de

\mathbb{R}^p y defina $Q_j(a) \equiv -T^{-1/2} \sum_{t=1}^T q_t(\hat{\beta} + a\mathbf{e}_j)$, donde a es un escalar. Sea $G_j(a)$ la derivada (finita) de $Q_j(a)$, esto es,

$$G_j(a) = \frac{\partial \left[-T^{-1/2} \sum_{t=1}^T q_t(\hat{\beta} + a\mathbf{e}_j) \right]}{\partial \hat{\beta}},$$

recordemos que por definición

$$q_t(\hat{\beta} + a\mathbf{e}_j) \equiv \{\theta - I[y_t < f_t(\hat{\beta} + a\mathbf{e}_j)]\}[y_t - f_t(\hat{\beta} + a\mathbf{e}_j)]$$

y de aquí que

$$G_j(a) = -T^{-1/2} \sum_{t=1}^T \nabla_j f_t(\hat{\beta} + a\mathbf{e}_j) \text{Hit}_t(\hat{\beta} + a\mathbf{e}_j).$$

Debido a que $Q_j(a)$ es continua en a y alcanza su máximo en 0, se cumple que en una vecindad del 0, para algún $\xi > 0$,

$$\begin{aligned} |G_j(0)| &\leq G_j(\xi) - G_j(-\xi) \\ &= T^{-1/2} \sum_{t=1}^T \left[-\nabla_j f_t(\hat{\beta} + \xi\mathbf{e}_j) \text{Hit}_t(\hat{\beta} + \xi\mathbf{e}_j) \right. \\ &\quad \left. + \nabla_j f_t(\hat{\beta} - \xi\mathbf{e}_j) \text{Hit}_t(\hat{\beta} - \xi\mathbf{e}_j) \right]. \end{aligned}$$

Tomando el límite cuando $\xi \rightarrow 0$, tenemos lo siguiente:

$$\begin{aligned} |G_j(0)| &\leq \lim_{\xi \rightarrow 0} T^{-1/2} \sum_{t=1}^T \left[-\nabla_j f_t(\hat{\beta} + \xi\mathbf{e}_j) \text{Hit}_t(\hat{\beta} + \xi\mathbf{e}_j) \right. \\ &\quad \left. + \nabla_j f_t(\hat{\beta} - \xi\mathbf{e}_j) \text{Hit}_t(\hat{\beta} - \xi\mathbf{e}_j) \right] \\ &= T^{-1/2} \sum_{t=1}^T \left| \nabla_j f_t(\hat{\beta}) \right| I[y_t = f_t(\hat{\beta})] \\ &\leq T^{-1/2} \left[\max_{1 \leq t \leq T} F(\Omega_t) \right] \sum_{t=1}^T I[y_t = f_t(\hat{\beta})], \end{aligned}$$

donde $T^{-1/2} [\max_{1 \leq t \leq T} F(\Omega_t)] = o_p(1)$ por el supuesto AN1 (a) y $\sum_{t=1}^T I[y_t = f_t(\hat{\beta})] = O_{a.s.}(1)$ por el supuesto C2. De lo anterior se sigue que $G_j(0) \xrightarrow{p} 0$ para cada j , lo cual implica que $T^{-1/2} \sum_{t=1}^T g_t(\hat{\beta}) = o_p(1)$.

Por otro lado $\lambda_T(\beta^0)$ está bien definida ya que β^0 es un punto interior de B por el supuesto AN1, entonces por la ley de las esperanzas iteradas,

$$E[g_t(\beta^0)] = E[E(\text{Hit}_t(\beta^0) | \Omega_t) \nabla' f_t(\beta^0)] = 0,$$

ya que asumimos que el modelo (6.3) está correctamente especificado.

Usando el teorema del valor medio para expandir $\lambda_T(\beta)$ alrededor de β^0 y aplicando la regla de Leibnitz para diferenciar bajo el signo integral, obtenemos la siguiente ecuación:

$$\lambda_T(\beta) - \lambda_T(\beta^0) = T^{-1} \frac{\partial \sum_{t=1}^T E[g_t(\beta)]}{\partial \beta} \Big|_{\beta=\beta^*} (\beta - \beta^0),$$

con $\lambda_T(\beta^0) = 0$ ya que $E[g_t(\beta^0)] = 0$, de modo que podemos escribir la ecuación anterior como:

$$\begin{aligned} \lambda_T(\beta) &= T^{-1} \frac{\partial \sum_{t=1}^T E[g_t(\beta)]}{\partial \beta} \Big|_{\beta=\beta^*} (\beta - \beta^0) \\ \lambda_T(\beta) &= T^{-1} \frac{\partial \sum_{t=1}^T E[Hit_t(\beta) \nabla' f_t(\beta)]}{\partial \beta} \Big|_{\beta=\beta^*} (\beta - \beta^0) \\ \lambda_T(\beta) &= T^{-1} \frac{\partial \sum_{t=1}^T E\{I[y_t < f_t(\beta)] - \theta\} \nabla' f_t(\beta)}{\partial \beta} \Big|_{\beta=\beta^*} (\beta - \beta^0) \\ &= T^{-1} \frac{\partial \sum_{t=1}^T E\{I[\varepsilon_{\theta t} < \tau_t(\beta)] - \theta\} \nabla' f_t(\beta)}{\partial \beta} \Big|_{\beta=\beta^*} \\ &= T^{-1} \frac{\partial \sum_{t=1}^T E\{E\{I[\varepsilon_{\theta t} < \tau_t(\beta)] - \theta\} \nabla' f_t(\beta) | \Omega_t\}}{\partial \beta} \Big|_{\beta=\beta^*} \\ &= T^{-1} \frac{\partial \sum_{t=1}^T E\{\nabla' f_t(\beta) [\int_{-\infty}^{\tau_t(\beta)} h_t(\lambda | \Omega_t) d\lambda - \int_{-\infty}^0 h_t(\lambda | \Omega_t) d\lambda]\}}{\partial \beta} \Big|_{\beta=\beta^*}, \\ &= T^{-1} \sum_{t=1}^T \frac{\partial}{\partial \beta} \{E\{\nabla' f_t(\beta) [I(\tau_t(\beta) > 0) \int_0^{\tau_t(\beta)} h_t(\lambda | \Omega_t) d\lambda \\ &\quad - I(\tau_t(\beta) < 0) \int_{\tau_t(\beta)}^0 h_t(\lambda | \Omega_t) d\lambda]\}\} \Big|_{\beta=\beta^*}, \end{aligned}$$

luego por la Regla de Leibnitz, tenemos la siguiente igualdad:

$$\begin{aligned} \lambda_T(\beta) &= \left\{ T^{-1} \sum_{t=1}^T E\{\nabla' f_t(\beta^*) \nabla f_t(\beta^*) \right. \\ &\quad \left[I(\tau_t(\beta^*) > 0) \int_0^{\tau_t(\beta^*)} h_t(\lambda | \Omega_t) d\lambda \right. \\ &\quad \left. \left. - I(\tau_t(\beta^*) < 0) \int_{\tau_t(\beta^*)}^0 h_t(\lambda | \Omega_t) d\lambda \right] \right\} \\ &\quad + T^{-1} \sum_{t=1}^T E\{\nabla' f_t(\beta^*) \nabla f_t(\beta^*) h_t(\tau_t(\beta^*) | \Omega_t)\} (\beta - \beta^0) \\ &= \Lambda_T(\beta^*) (\beta - \beta^0), \end{aligned}$$

donde β^* está entre β y β^0 . Ahora mostraremos que $\Lambda_T(\beta^*) = \mathbf{D}_T + O(\|\beta - \beta^0\|)$, donde

\mathbf{D}_T es como se define en el supuesto AN3.

$$\begin{aligned} \Lambda_T(\beta^*) - \mathbf{D}_T &= \left\| T^{-1} \sum_{t=1}^T E\{\nabla' f_t(\beta^*) \nabla f_t(\beta^*) \right. \\ &\quad \times [I(\tau_t(\beta^*) > 0) \int_0^{\tau_t(\beta^*)} h_t(\lambda|\Omega_t) d\lambda \\ &\quad \left. - I(\tau_t(\beta^*) < 0) \int_{\tau_t(\beta^*)}^0 h_t(\lambda|\Omega_t) d\lambda] \right\} \\ &\quad + T^{-1} \sum_{t=1}^T E\{\nabla' f_t(\beta^*) \nabla f_t(\beta^*) h_t(\tau_t(\beta^*)|\Omega_t) \\ &\quad - \nabla' f_t(\beta^0) \nabla f_t(\beta^0) h_t(0|\Omega_t)\|. \end{aligned}$$

Observemos que el primer término es $O(\|\beta^* - \beta^0\|)$, para lo cual tomamos la expansión en serie de Taylor de $\int_0^{\tau_t(\beta^*)} h_t(\lambda|\Omega_t) d\lambda$ alrededor de β^0 , es decir,

$$\begin{aligned} \int_0^{\tau_t(\beta^*)} h_t(\lambda|\Omega_t) d\lambda &= \int_0^{\tau_t(\beta^0)} h_t(\lambda|\Omega_t) d\lambda \text{ (esta parte es cero)} \\ &\quad + \frac{\partial \int_0^{\tau_t(\beta^*)} h_t(\lambda|\Omega_t) d\lambda}{\partial \beta^*} \Big|_{\beta^*=\beta^{**}} (\beta^* - \beta^0) \\ &= [h_t(\tau_t(\beta^{**})|\Omega_t) \nabla f_t(\beta^{**}) - h_t(0|\Omega_t) \times 0] (\beta^* - \beta^0) \\ &= h_t(\tau_t(\beta^{**})|\Omega_t) \nabla f_t(\beta^{**}) (\beta^* - \beta^0), \end{aligned}$$

e igualmente

$$\begin{aligned} &\int_{\tau_t(\beta^*)}^0 h_t(\lambda|\Omega_t) d\lambda \\ &= \int_{\tau_t(\beta^0)}^0 h_t(\lambda|\Omega_t) d\lambda + \frac{\partial \int_{\tau_t(\beta^*)}^0 h_t(\lambda|\Omega_t) d\lambda}{\partial \beta^*} \Big|_{\beta^*=\beta^{**}} (\beta^* - \beta^0) \\ &= [h_t(0|\Omega_t) \times 0 - h_t(\tau_t(\beta^{**})|\Omega_t) \nabla f_t(\beta^{**})] (\beta^* - \beta^0) \\ &= -h_t(\tau_t(\beta^{**})|\Omega_t) \nabla f_t(\beta^{**}) (\beta^* - \beta^0), \end{aligned}$$

con β^{**} entre β^* y β^0 . De modo que,

$$\begin{aligned}
& \left\| T^{-1} \sum_{t=1}^T E \left\{ \nabla' f_t(\beta^*) \nabla f_t(\beta^*) \left[I(\tau_t(\beta^*) > 0) \int_0^{\tau_t(\beta^*)} h_t(\lambda | \Omega_t) d\lambda \right. \right. \right. \\
& \quad \left. \left. \left. - I(\tau_t(\beta^*) < 0) \int_{\tau_t(\beta^*)}^0 h_t(\lambda | \Omega_t) d\lambda \right] \right\} \right\| \\
& \leq T^{-1} \sum_{t=1}^T E \{ \| 2 \nabla' f_t(\beta^*) \nabla f_t(\beta^*) h_t(\tau_t(\beta^{**}) | \Omega_t) \nabla f_t(\beta^{**}) (\beta^* - \beta^0) \| \} \\
& \leq T^{-1} \sum_{t=1}^T E \{ \| 2 F(\Omega_t)^3 N (\beta^* - \beta^0) \| \} \\
& \leq T^{-1} \sum_{t=1}^T N F_0 \| \beta^* - \beta^0 \| = O(\| \beta^* - \beta^0 \|).
\end{aligned}$$

luego dado que β^* está entre β y β^0 , entonces también el primer término es de orden $O(\| \beta - \beta^0 \|)$, mientras que el segundo término podemos rescribirlo como:

$$\begin{aligned}
& \left\| T^{-1} \sum_{t=1}^T E \{ \nabla' f_t(\beta^*) \nabla f_t(\beta^*) h_t(\tau_t(\beta^*) | \Omega_t) \right. \\
& \quad - \nabla' f_t(\beta^0) \nabla f_t(\beta^*) h_t(\tau_t(\beta^*) | \Omega_t) + \nabla' f_t(\beta^0) \nabla f_t(\beta^*) h_t(\tau_t(\beta^*) | \Omega_t) \\
& \quad - \nabla' f_t(\beta^0) \nabla f_t(\beta^0) h_t(\tau_t(\beta^*) | \Omega_t) + \nabla' f_t(\beta^0) \nabla f_t(\beta^0) h_t(\tau_t(\beta^*) | \Omega_t) \\
& \quad \left. - \nabla' f_t(\beta^0) \nabla f_t(\beta^0) h_t(0 | \Omega_t) \} \right\| \\
& \leq T^{-1} \sum_{t=1}^T E \{ \| \nabla' f_t(\beta^*) - \nabla' f_t(\beta^0) \| \| \nabla f_t(\beta^*) \| h_t(\tau_t(\beta^*) | \Omega_t) \\
& \quad + \| \nabla' f_t(\beta^0) \| \| \nabla f_t(\beta^*) - \nabla f_t(\beta^0) \| h_t(\tau_t(\beta^*) | \Omega_t) \\
& \quad + \| \nabla' f_t(\beta^0) \nabla f_t(\beta^0) \| [h_t(\tau_t(\beta^*) | \Omega_t) - h_t(0 | \Omega_t)] \} \\
& \leq T^{-1} \sum_{t=1}^T E \{ M(\Omega_t, \beta^*, \beta^0) F(\Omega_t) N \\
& \quad + F(\Omega_t) M(\Omega_t, \beta^*, \beta^0) N + F(\Omega_t)^2 L \|\tau_t(\beta^*)\| \\
& \quad = T^{-1} \sum_{t=1}^T E \{ 2M(\Omega_t, \beta^*, \beta^0) F(\Omega_t) N + F(\Omega_t)^2 L \|f(\beta^*) - f(\beta^0)\| \}
\end{aligned}$$

Por el Teorema del valor medio y siguiendo con la demostración,

$$\begin{aligned}
& \| f(\beta^*) - f(\beta^0) \| \\
& \leq \| \nabla f_t(\beta^*) \| \| \beta^* - \beta^0 \| \leq \| F(\Omega_t) \| \| \beta^* - \beta^0 \| \\
& \leq T^{-1} \sum_{t=1}^T E \{ 2M(\Omega_t, \beta^*, \beta^0) F(\Omega_t) N \} + F(\Omega_t)^3 L \| \beta^* - \beta^0 \| \\
& \leq T^{-1} \sum_{t=1}^T [2NM_1 \| \beta^* - \beta^0 \| + F_0 L \| \beta^* - \beta^0 \|] \\
& \leq T^{-1} \sum_{t=1}^T [2NM_1 + F_0 L] \| \beta^* - \beta^0 \| \\
& = O(\| \beta^* - \beta^0 \|),
\end{aligned}$$

e igualmente tenemos que dicho término es también $O(\| \beta - \beta^0 \|)$. Por consiguiente, se cumple lo siguiente,

$$\lambda_T(\beta) = \mathbf{D}_T(\beta - \beta^0) + O(\| \beta - \beta^0 \|^2). \quad (\text{C.1})$$

Pero debido a que la matriz \mathbf{D}_T es positiva definida para T suficientemente grande se sigue el resultado (parte N3 (i) del Teorema de Weiss (1991)).

Para N3 (ii) en el Teorema de Weiss (1991), note que

$$\begin{aligned}
 |Hit_t(\boldsymbol{\tau}) - Hit_t(\boldsymbol{\beta})| &= |I[y_t < f_t(\boldsymbol{\tau})] - \theta - I[y_t < f_t(\boldsymbol{\beta})] + \theta| \\
 &= |I[y_t < f_t(\boldsymbol{\tau})] - I[y_t < f_t(\boldsymbol{\beta})]| \\
 &\leq |I[y_t - f_t(\boldsymbol{\beta}) < f_t(\boldsymbol{\tau}) - f_t(\boldsymbol{\beta})]| + |I[y_t < f_t(\boldsymbol{\beta})]| \\
 &= I[y_t - f_t(\boldsymbol{\beta}) < f_t(\boldsymbol{\tau}) - f_t(\boldsymbol{\beta})] + I[y_t - f_t(\boldsymbol{\beta}) < 0] \\
 &= I[|y_t - f_t(\boldsymbol{\beta})| < |f_t(\boldsymbol{\tau}) - f_t(\boldsymbol{\beta})|],
 \end{aligned}$$

a partir de lo cual, tenemos que:

$$\begin{aligned}
 \mu_t(\boldsymbol{\beta}, d) &= \sup_{\|\boldsymbol{\tau} - \boldsymbol{\beta}\| \leq d} \|g_t(\boldsymbol{\tau}) - g_t(\boldsymbol{\beta})\| \\
 &= \sup_{\|\boldsymbol{\tau} - \boldsymbol{\beta}\| \leq d} \|Hit_t(\boldsymbol{\tau}) \nabla' f_t(\boldsymbol{\tau}) - Hit_t(\boldsymbol{\beta}) \nabla' f_t(\boldsymbol{\beta})\| \\
 &= \sup_{\|\boldsymbol{\tau} - \boldsymbol{\beta}\| \leq d} \|Hit_t(\boldsymbol{\tau}) \nabla' f_t(\boldsymbol{\tau}) - Hit_t(\boldsymbol{\beta}) \nabla' f_t(\boldsymbol{\beta}) \\
 &\quad + \nabla' f_t(\boldsymbol{\beta}) Hit_t(\boldsymbol{\tau}) - \nabla' f_t(\boldsymbol{\beta}) Hit_t(\boldsymbol{\tau})\| \\
 &= \sup_{\|\boldsymbol{\tau} - \boldsymbol{\beta}\| \leq d} \|Hit_t(\boldsymbol{\tau}) [\nabla' f_t(\boldsymbol{\tau}) - \nabla' f_t(\boldsymbol{\beta})] \\
 &\quad + \nabla' f_t(\boldsymbol{\beta}) [Hit_t(\boldsymbol{\tau}) - Hit_t(\boldsymbol{\beta})]\| \\
 &= \sup_{\|\boldsymbol{\tau} - \boldsymbol{\beta}\| \leq d} \|Hit_t(\boldsymbol{\tau}) [\nabla' f_t(\boldsymbol{\tau}) - \nabla' f_t(\boldsymbol{\beta})]\| \\
 &\quad + \sup_{\|\boldsymbol{\tau} - \boldsymbol{\beta}\| \leq d} \|\nabla' f_t(\boldsymbol{\beta}) [Hit_t(\boldsymbol{\tau}) - Hit_t(\boldsymbol{\beta})]\| \\
 &\leq \sup_{\|\boldsymbol{\tau} - \boldsymbol{\beta}\| \leq d} |Hit_t(\boldsymbol{\tau})| \|\nabla' f_t(\boldsymbol{\tau}) - \nabla' f_t(\boldsymbol{\beta})\| \\
 &\quad + \sup_{\|\boldsymbol{\tau} - \boldsymbol{\beta}\| \leq d} \|\nabla' f_t(\boldsymbol{\beta})\| |Hit_t(\boldsymbol{\tau}) - Hit_t(\boldsymbol{\beta})| \\
 &\leq \sup_{\|\boldsymbol{\tau} - \boldsymbol{\beta}\| \leq d} \|\nabla' f_t(\boldsymbol{\tau}) - \nabla' f_t(\boldsymbol{\beta})\| \\
 &\quad + \sup_{\|\boldsymbol{\tau} - \boldsymbol{\beta}\| \leq d} \|\nabla' f_t(\boldsymbol{\beta})\| I[|y_t - f_t(\boldsymbol{\beta})| < |f_t(\boldsymbol{\tau}) - f_t(\boldsymbol{\beta})|].
 \end{aligned}$$

Así, que se cumple la siguiente desigualdad,

$$\begin{aligned}
 \mu_t(\boldsymbol{\beta}, d) &\leq \sup_{\|\boldsymbol{\tau} - \boldsymbol{\beta}\| \leq d} O \|\boldsymbol{\beta} - \boldsymbol{\tau}\| \\
 &\quad + \sup_{\|\boldsymbol{\tau} - \boldsymbol{\beta}\| \leq d} F(\Omega_t) I[|y_t - f_t(\boldsymbol{\beta})| < |f_t(\boldsymbol{\tau}) - f_t(\boldsymbol{\beta})|] \\
 &\leq O(d) + F(\Omega_t) I[|y_t - f_t(\boldsymbol{\beta})| < |f_t(\boldsymbol{\tau}) - f_t(\boldsymbol{\beta})|]
 \end{aligned}$$

luego, tomando valores esperados a ambos lados de la desigualdad,

$$E[\mu_t(\beta, d)] \leq E[dM(\Omega_t, \beta, \tau)] + E\{F(\Omega_t) I[|\varepsilon_t| < |f_t(\tau) - f_t(\beta)|]\},$$

donde el segundo término es igual a:

$$\begin{aligned} & E\{F(\Omega_t) I[|\varepsilon_t| < |f_t(\tau) - f_t(\beta)|]\} \\ &= E\{E\{F(\Omega_t) I[|\varepsilon_t| < |f_t(\tau) - f_t(\beta)|] | \Omega_t\}\} \\ &= E\{F(\Omega_t) E\{I[|\varepsilon_t| < |f_t(\tau) - f_t(\beta)|] | \Omega_t\}\} \\ &= E\left\{F(\Omega_t) \int_{-\infty}^{\infty} I[|\lambda| < |f_t(\tau) - f_t(\beta)|] h_t(\lambda | \Omega_t) d\lambda\right\} \\ &= E\left\{F(\Omega_t) \int_{-|f_t(\tau) - f_t(\beta)|}^{|f_t(\tau) - f_t(\beta)|} h_t(\lambda | \Omega_t) d\lambda\right\} \\ &\leq E\left\{F(\Omega_t) N \int_{-|f_t(\tau) - f_t(\beta)|}^{|f_t(\tau) - f_t(\beta)|} d\lambda\right\} \\ &= E\{F(\Omega_t) N 2 |f_t(\tau) - f_t(\beta)|\}, \end{aligned}$$

luego por el teorema del valor medio tenemos que:

$$|f_t(\tau) - f_t(\beta)| \leq \|\nabla f_t(\tau^*)\| \|\tau - \beta\|,$$

con τ^* entre τ y β , de esta forma retomando el valor esperado, tenemos que:

$$\begin{aligned} E[\mu_t(\beta, d)] &\leq O(d) + E\{F(\Omega_t) \|\nabla f_t(\tau^*)\| \|\tau - \beta\| N\} \\ &\leq O(d) + E\{F(\Omega_t)^2 dN\} \\ &\leq O(d) + dNE\{F(\Omega_t)^2\} \\ &\leq O(d) + dNH = O(d) \end{aligned}$$

donde d está definido en el supuesto AN1.

Por último para el segundo momento de $\mu_t(\beta, d)$ tenemos lo siguiente,

$$\begin{aligned} E[\mu_t(\beta, d)]^2 &\leq E\{O(d)^2 + F(\Omega_t)^2 \|\nabla f_t(\tau^*)\| \|\tau - \beta\| N^2 \\ &\quad + 2O(d) F(\Omega_t) \|\nabla f_t(\tau^*)\| \|\tau - \beta\| N\} \\ &\leq O(d)^2 + dN^2 E[F(\Omega_t)^2 F(\Omega_t)] \\ &\quad + 2O(d) E(F(\Omega_t) \|\nabla f_t(\tau^*)\| \|\tau - \beta\| N) \\ &\leq O(d) + dN^2 F_0 + 2O(d) \\ &= O(d). \end{aligned}$$

De lo anterior se sigue que (recordemos que $\lambda_T(\beta) \equiv T^{-1} \sum_{t=1}^T E[g_t(\beta)]$)

$$T^{1/2}\lambda_T(\hat{\beta}) = -T^{-1/2} \sum_{t=1}^T g_t(\beta^0) + o_p(1).$$

Por consistencia y el Teorema de Slutsky aplicado a (C.1), tenemos que

$$T^{1/2}\lambda_T(\hat{\beta}) = \mathbf{D}_T T^{1/2}(\hat{\beta} - \beta^0) + o_p(1).$$

De los últimos dos resultados tenemos que

$$\begin{aligned} \mathbf{D}_T T^{1/2}(\hat{\beta} - \beta^0) &= -T^{-1/2} \sum_{t=1}^T g_t(\beta^0) \\ &= -T^{-1/2} \sum_{t=1}^T \text{Hit}_t(\beta^0) \nabla' f_t(\beta^0) \\ &= T^{-1/2} \sum_{t=1}^T \{\theta - I[y_t < f_t(\beta)]\} \nabla' f_t(\beta^0), \end{aligned}$$

luego por el supuesto AN4, $T^{-1/2} \sum_{t=1}^T \{\theta - I[y_t < f_t(\beta)]\} \nabla' f_t(\beta^0)$ obedece el TLC, con valor esperado igual a cero ya que $E[g_t(\beta^0)] = 0$, y varianza

$$E\{-T^{-1/2} \sum_{t=1}^T g_t(\beta^0)\}^2 = T^{-1} \sum_{t=1}^T E[g_t(\beta^0)]^2,$$

ya que los valores de los productos cruzados es cero, luego,

$$\begin{aligned} E[g_t(\beta^0)]^2 &= E[\text{Hit}_t(\beta^0) \nabla' f_t(\beta^0) \nabla f_t(\beta^0) \text{Hit}_t'(\beta^0)] \\ &= E\{E[\text{Hit}_t(\beta^0) \nabla' f_t(\beta^0) \nabla f_t(\beta^0) \text{Hit}_t'(\beta^0)] | \Omega_t\} \\ &= E\{\nabla' f_t(\beta^0) E[\text{Hit}_t(\beta^0) \text{Hit}_t'(\beta^0)] | \Omega_t \nabla f_t(\beta^0)\} \\ &= E\{\theta(1 - \theta) \nabla' f_t(\beta^0) \nabla f_t(\beta^0)\}, \end{aligned}$$

es decir,

$$E\{-T^{-1/2} \sum_{t=1}^T g_t(\beta^0)\}^2 = E\{T^{-1} \theta(1 - \theta) \sum_{t=1}^T \nabla' f_t(\beta^0) \nabla f_t(\beta^0)\} \equiv A_T$$

y de aquí que $T^{1/2} A_T^{-1/2} \mathbf{D}_T(\hat{\beta} - \beta^0) \xrightarrow{d} N(\mathbf{0}, \mathbf{I})$. □

Teorema 6.3 (Estimación de la matriz de varianzas y covarianzas). *Bajo los supuestos VC1 a VC3, en el Apéndice B y las condiciones de los Teoremas 6.1 y 6.2, $\hat{\mathbf{A}}_T - \mathbf{A}_T \xrightarrow{p} \mathbf{0}$ y $\hat{\mathbf{D}}_T - \mathbf{D}_T \xrightarrow{p} \mathbf{0}$, donde*

$$\begin{aligned} \hat{\mathbf{A}}_T &= T^{-1} \theta(1 - \theta) \nabla' f_t(\hat{\beta}) \nabla f_t(\hat{\beta}), \\ \hat{\mathbf{D}}_T &= (2T \hat{c}_T)^{-1} \sum_{t=1}^T I\left[\left|y_t - f_t(\hat{\beta})\right| < \hat{c}_T\right] \nabla' f_t(\hat{\beta}) \nabla f_t(\hat{\beta}), \end{aligned}$$

\mathbf{A}_T y \mathbf{D}_T se definieron en el Teorema 2 y \hat{c}_T se define en el supuesto VC1 enunciado en el Apéndice B.

Demostración. Por el supuesto VC3 tenemos que:

$$T^{-1}\theta(1-\theta)\sum_{t=1}^T\nabla'f_t(\beta^0)\nabla f_t(\beta^0) - E\left[T^{-1}\theta(1-\theta)\sum_{t=1}^T\nabla'f_t(\beta^0)\nabla f_t(\beta^0)\right] \xrightarrow{p} \mathbf{0},$$

además dado que $\hat{\beta} \xrightarrow{p} \beta^0$ y por los supuestos C1 y AN1, tenemos que $\nabla f_t(\hat{\beta}) \xrightarrow{p} \nabla f_t(\beta^0)$ y de aquí que $\hat{\mathbf{A}}_T \xrightarrow{p} \mathbf{A}_T$.

Para probar que $\hat{\mathbf{D}}_T \xrightarrow{p} \mathbf{D}_T$ definamos,

$$\tilde{\mathbf{D}}_T \equiv (2Tc_T)^{-1} \sum_{t=1}^T I(|\varepsilon_{t\theta}| < c_T) \nabla'f_t(\beta^0) \nabla f_t(\beta^0),$$

y mostremos primeramente que $\hat{\mathbf{D}}_T - \tilde{\mathbf{D}}_T = o_p(1)$ y posteriormente que $\tilde{\mathbf{D}}_T - \mathbf{D}_T = o_p(1)$, para concluir así que $\hat{\mathbf{D}}_T - \mathbf{D}_T \xrightarrow{p} \mathbf{0}$.

Defina $\hat{\varepsilon}_t \equiv y_t - f_t(\hat{\beta})$, entonces

$$\begin{aligned} & \left\| \hat{\mathbf{D}}_T - \tilde{\mathbf{D}}_T \right\| \\ = & \left\| (2T\hat{c}_T)^{-1} \sum_{t=1}^T I(|\hat{\varepsilon}_t| < \hat{c}_T) \nabla'f_t(\hat{\beta}) \nabla f_t(\hat{\beta}) \right. \\ & \left. - (2Tc_T)^{-1} \sum_{t=1}^T I(|\varepsilon_{t\theta}| < c_T) \nabla'f_t(\beta^0) \nabla f_t(\beta^0) \right\| \\ = & \left\| (2T\hat{c}_T)^{-1} \sum_{t=1}^T I(|\hat{\varepsilon}_t| < \hat{c}_T) \nabla'f_t(\hat{\beta}) \nabla f_t(\hat{\beta}) \right. \\ & - (2Tc_T)^{-1} \sum_{t=1}^T I(|\varepsilon_{t\theta}| < c_T) \nabla'f_t(\beta^0) \nabla f_t(\beta^0) \\ & + (2T\hat{c}_T)^{-1} \sum_{t=1}^T I(|\varepsilon_{t\theta}| < c_T) \{ \nabla'f_t(\hat{\beta}) \nabla f_t(\hat{\beta}) \\ & - \nabla'f_t(\hat{\beta}) \nabla f_t(\hat{\beta}) - \nabla'f_t(\beta^0) \nabla f_t(\hat{\beta}) \\ & + \nabla'f_t(\beta^0) \nabla f_t(\hat{\beta}) - \nabla'f_t(\beta^0) \nabla f_t(\beta^0) + \nabla'f_t(\beta^0) \nabla f_t(\beta^0) \} \left. \right\| \\ = & \frac{c_T}{\hat{c}_T} \left\| (2T\hat{c}_T)^{-1} \sum_{t=1}^T \{ [I(|\hat{\varepsilon}_t| < \hat{c}_T) - I(|\varepsilon_{t\theta}| < c_T)] \right. \\ & \times \nabla'f_t(\hat{\beta}) \nabla f_t(\hat{\beta}) + I(|\varepsilon_{t\theta}| < c_T) [\nabla'f_t(\hat{\beta}) - \nabla'f_t(\beta^0)] \nabla f_t(\hat{\beta}) \\ & + I(|\varepsilon_{t\theta}| < c_T) \nabla'f_t(\beta^0) [\nabla f_t(\hat{\beta}) - \nabla f_t(\beta^0)] \\ & \left. + \frac{c_T - \hat{c}_T}{c_T} I(|\varepsilon_{t\theta}| < c_T) \nabla'f_t(\beta^0) \nabla f_t(\beta^0) \} \right\|. \end{aligned}$$

La función indicadora en la primer línea cumple la siguiente desigualdad:

$$\begin{aligned}
 |I(|\hat{\varepsilon}_t| < \hat{c}_T) - I(|\varepsilon_{t\theta}| < c_T)| &\leq I(|\hat{\varepsilon}_t| < \hat{c}_T) - I(|\varepsilon_{t\theta}| < c_T) \\
 &\leq I\left(|\varepsilon_{t\theta} - c_T| < \left|\tau_t(\hat{\beta})\right| + |\hat{c}_T - c_T|\right) \\
 &\quad + I\left(|\varepsilon_{t\theta} + c_T| < \left|\tau_t(\hat{\beta})\right| + |\hat{c}_T + c_T|\right).
 \end{aligned}$$

Así, que:

$$\begin{aligned}
 \left\|\hat{\mathbf{D}}_T - \tilde{\mathbf{D}}_T\right\| &\leq \frac{c_T}{\hat{c}_T} (2Tc_T)^{-1} \sum_{t=1}^T \left\{ I\left(|\varepsilon_t - c_T| < \left|\tau_t(\hat{\beta})\right| + |\hat{c}_T - c_T|\right) \right. \\
 &\quad \left. + I\left(|\varepsilon_t + c_T| < \left|\tau_t(\hat{\beta})\right| + |\hat{c}_T + c_T|\right) \right\} F(\Omega_t)^2 \\
 &\quad + I(|\varepsilon_t| < c_T) M(\Omega_t, \hat{\beta}, \beta^0) F(\Omega_t) + I(|\varepsilon_t| < c_T) F(\Omega_t) M(\Omega_t, \hat{\beta}, \beta^0) \\
 &\quad + \frac{c_T - \hat{c}_T}{c_T} I(|\varepsilon_{t\theta}| < c_T) F(\Omega_t)^2 \\
 &\equiv \frac{c_T}{\hat{c}_T} (A_1 + 2A_2 + A_3).
 \end{aligned}$$

Ahora suponga que para $d > 0$ dado, el cual puede elegirse arbitrariamente pequeño, y T suficientemente grande, tal que se cumple que $\left|\frac{c_T - \hat{c}_T}{c_T}\right| < d$ y que $c_T^{-1} \|\hat{\beta} - \beta^0\| < d$. Es posible mostrar que $E(A_i) = O(d)$, para $i = 1, 2, 3$. Es decir, podemos encontrar funciones acotadas para $\left\|\hat{\mathbf{D}}_T - \tilde{\mathbf{D}}_T\right\|$ las cuales pueden hacerse arbitrariamente pequeñas en probabilidad, eligiendo d suficientemente pequeño. Calculemos el valor esperado para A_1 , para lo cual notemos que por el teorema del valor medio se sigue la siguiente desigualdad:

$$\left|\tau_t(\hat{\beta})\right| = \left|f_t(\hat{\beta}) - f_t(\beta^0)\right| \leq \|\nabla f_t(\beta^*)\| \|\hat{\beta} - \beta^0\|,$$

y de aquí que:

$$\begin{aligned}
 E(A_1) &= E\left\{(2Tc_T)^{-1} \sum_{t=1}^T \left[I\left(|\varepsilon_t - c_T| < \left|\tau_t(\hat{\beta})\right| + |\hat{c}_T - c_T|\right) \right. \right. \\
 &\quad \left. \left. + I\left(|\varepsilon_t + c_T| < \left|\tau_t(\hat{\beta})\right| + |\hat{c}_T + c_T|\right) \right] F(\Omega_t)^2 \right\} \\
 &\leq E\left\{(2Tc_T)^{-1} \sum_{t=1}^T \left[I\left(|\varepsilon_t - c_T| < \|\nabla f_t(\beta^*)\| \|\hat{\beta} - \beta^0\| + |\hat{c}_T - c_T|\right) \right. \right. \\
 &\quad \left. \left. + I\left(|\varepsilon_t + c_T| < \|\nabla f_t(\beta^*)\| \|\hat{\beta} - \beta^0\| + |\hat{c}_T + c_T|\right) \right] F(\Omega_t)^2 \right\} \\
 &\leq E\left\{(2Tc_T)^{-1} \sum_{t=1}^T [I(|\varepsilon_t - c_T| < F(\Omega_t)d + |\hat{c}_T - c_T|) \right. \\
 &\quad \left. + I(|\varepsilon_t + c_T| < F(\Omega_t)d + |\hat{c}_T + c_T|)] F(\Omega_t)^2 \right\} \\
 &\leq E\left\{(2Tc_T)^{-1} \sum_{t=1}^T [I(|\varepsilon_t - c_T| < dc_T[F(\Omega_t) + 1]) \right. \\
 &\quad \left. + I(|\varepsilon_t + c_T| < dc_T[F(\Omega_t) + 1])] F(\Omega_t)^2 \right\},
 \end{aligned}$$

por otro lado observemos que

$$\begin{aligned}
& E [I (|\varepsilon_t - c_T| < dc_T [F(\Omega_t) + 1])] \\
&= E \{E \{I [|\varepsilon_t - c_T| < dc_T (F(\Omega_t) + 1)] | \Omega_t\}\} \\
&= E \{E \{I [c_T - dc_T (F(\Omega_t) + 1) < \varepsilon_t < \\
&\quad dc_T (F(\Omega_t) + 1) + c_T] | \Omega_t\}\} \\
&= E \left\{ \left\{ \int_{c_T - dc_T (F(\Omega_t) + 1)}^{c_T + dc_T (F(\Omega_t) + 1)} h_t(\lambda | \Omega_t) d\lambda \right\} \right\} \\
&= E \{h_t(c_T + dc_T (F(\Omega_t) + 1) | \Omega_t) \\
&\quad - h_t(c_T - dc_T (F(\Omega_t) + 1) | \Omega_t)\} \\
&= E \{c_T + dc_T [F(\Omega_t) + 1] - c_T + dc_T [F(\Omega_t) + 1]\} \\
&= E \{2dc_T [F(\Omega_t) + 1]\},
\end{aligned}$$

e igualmente,

$$E \{I (|\varepsilon_t + c_T| < dc_T [F(\Omega_t) + 1])\} = E \{2dc_T [F(\Omega_t) + 1]\},$$

de está forma retomando la desigualdad para $E(A_1)$, tenemos que,

$$\begin{aligned}
E(A_1) &\leq E \left\{ (2Tc_T)^{-1} \sum_{t=1}^T 4dc_T [F(\Omega_t) + 1] F(\Omega_t)^2 \right\} \\
&\leq E \left\{ 2d(T)^{-1} \sum_{t=1}^T [F(\Omega_t)^3 + F(\Omega_t)^2] \right\} \\
&\leq 2dT^{-1} \sum_{t=1}^T \{E[F(\Omega_t)]^3 + E[F(\Omega_t)]^2\} \\
&\leq 2dT^{-1} \sum_{t=1}^T (F_0 + H) = 2d(F_0 + H) = O(d), \\
H &\neq 0 \text{ constante.}
\end{aligned}$$

Veamos ahora que $E(A_2) = O(d)$,

$$\begin{aligned}
E(A_2) &= E \left\{ (2Tc_T)^{-1} \sum_{t=1}^T I(|\varepsilon_{t\theta}| < c_T) M(\Omega_t, \hat{\beta}, \beta^0) F(\Omega_t) \right\} \\
&\leq E \left\{ (2Tc_T)^{-1} \sum_{t=1}^T M(\Omega_t, \hat{\beta}, \beta^0) F(\Omega_t) \right\} \\
&\leq (2Tc_T)^{-1} \sum_{t=1}^T E \left[M(\Omega_t, \hat{\beta}, \beta^0) F(\Omega_t) \right] \\
&\leq (2Tc_T)^{-1} \sum_{t=1}^T M_1 \|\hat{\beta} - \beta^0\| \\
&\leq (2)^{-1} M_1 c_T^{-1} \|\hat{\beta} - \beta^0\| \leq \frac{M_1 d}{2} = O(d).
\end{aligned}$$

Igualmente $E(A_3) = O(d)$, ya que:

$$\begin{aligned}
E(A_3) &= E \left\{ \frac{c_T - \hat{c}_T}{c_T} (2Tc_T)^{-1} \sum_{t=1}^T I(|\varepsilon_{t\theta}| < c_T) F(\Omega_t)^2 \right\} \\
&\leq E \left\{ d(2Tc_T)^{-1} \sum_{t=1}^T F(\Omega_t)^2 \right\} \\
&\leq d(2Tc_T)^{-1} \sum_{t=1}^T E[F(\Omega_t)]^2 \\
&\leq 2H(c_T)^{-1} d = O(d).
\end{aligned}$$

Mostremos ahora que $\tilde{\mathbf{D}}_T - \mathbf{D}_T = o_p(1)$,

$$\begin{aligned}
&\tilde{\mathbf{D}}_T - \mathbf{D}_T \\
&= (2Tc_T)^{-1} \sum_{t=1}^T I(|\varepsilon_{t\theta}| < c_T) \nabla' f_t(\beta^0) \nabla f_t(\beta^0) \\
&\quad - E \left[T^{-1} \sum_{t=1}^T h_t(0|\Omega_t) \nabla' f_t(\beta^0) \nabla f_t(\beta^0) \right] \\
&= (2Tc_T)^{-1} \sum_{t=1}^T \left\{ I(|\varepsilon_{t\theta}| < c_T) \nabla' f_t(\beta^0) \nabla f_t(\beta^0) \right. \\
&\quad \left. - E[I(|\varepsilon_{t\theta}| < c_T) |\Omega_t] \nabla' f_t(\beta^0) \nabla f_t(\beta^0) \right\} \\
&\quad + T^{-1} \sum_{t=1}^T \left\{ (2c_T)^{-1} E[I(|\varepsilon_{t\theta}| < c_T) |\Omega_t] \nabla' f_t(\beta^0) \nabla f_t(\beta^0) \right. \\
&\quad \left. - E[h_t(0|\Omega_t) \nabla' f_t(\beta^0) \nabla f_t(\beta^0)] \right\}.
\end{aligned}$$

Calculemos la esperanza y la varianza del primer término,

$$\begin{aligned}
&E \left\{ (2Tc_T)^{-1} \sum_{t=1}^T [I(|\varepsilon_{t\theta}| < c_T) \nabla' f_t(\beta^0) \nabla f_t(\beta^0) \right. \\
&\quad \left. - E[I(|\varepsilon_{t\theta}| < c_T) |\Omega_t] \nabla' f_t(\beta^0) \nabla f_t(\beta^0)] \right\} \\
&= (2Tc_T)^{-1} \sum_{t=1}^T \left\{ E[I(|\varepsilon_{t\theta}| < c_T) \nabla' f_t(\beta^0) \nabla f_t(\beta^0)] \right. \\
&\quad \left. - E[E[I(|\varepsilon_{t\theta}| < c_T) |\Omega_t] \nabla' f_t(\beta^0) \nabla f_t(\beta^0)] \right\} \\
&= 0,
\end{aligned}$$

ya que,

$$\begin{aligned}
E[I(|\varepsilon_{t\theta}| < c_T) \nabla' f_t(\beta^0) \nabla f_t(\beta^0)] &= E\{E[I(|\varepsilon_{t\theta}| < c_T) \nabla' f_t(\beta^0) \nabla f_t(\beta^0)] |\Omega_t\} \\
&= E\{E[I(|\varepsilon_{t\theta}| < c_T) |\Omega_t] \nabla' f_t(\beta^0) \nabla f_t(\beta^0)\}.
\end{aligned}$$

Mientras que la varianza está dada por:

$$\begin{aligned}
& \text{Var} \left\{ (2Tc_T)^{-1} \sum_{t=1}^T \{I(|\varepsilon_{t\theta}| < c_T) - E[I(|\varepsilon_{t\theta}| < c_T) | \Omega_t]\} \right. \\
& \quad \times \nabla' f_t(\beta^0) \nabla f_t(\beta^0) \left. \right\} \\
&= E \left\{ (2Tc_T)^{-1} \sum_{t=1}^T \{I(|\varepsilon_{t\theta}| < c_T) - E[I(|\varepsilon_{t\theta}| < c_T) | \Omega_t]\} \right. \\
& \quad \times \nabla' f_t(\beta^0) \nabla f_t(\beta^0) \left. \right\}^2 \\
&= (2Tc_T)^{-2} E \left\{ \sum_{t=1}^T \{[I(|\varepsilon_{t\theta}| < c_T) - E[I(|\varepsilon_{t\theta}| < c_T) | \Omega_t]]\}^2 \right. \\
& \quad \times [\nabla' f_t(\beta^0) \nabla f_t(\beta^0)]^2 \left. \right\},
\end{aligned}$$

esto porque las esperanzas de los dobles productos son cero por la ley de las esperanzas iteradas, entonces continuando tenemos que:

$$\begin{aligned}
& (2Tc_T)^{-2} E \left\{ \sum_{t=1}^T \{[I(|\varepsilon_{t\theta}| < c_T) - E[I(|\varepsilon_{t\theta}| < c_T) | \Omega_t]]\}^2 \right. \\
& \quad \times [\nabla' f_t(\beta^0) \nabla f_t(\beta^0)]^2 \left. \right\} \\
&\leq (2Tc_T)^{-2} \sum_{t=1}^T E [F(\Omega_t)^2]^2 \leq (2Tc_T)^{-2} \sum_{t=1}^T F_1 \\
&= (4Tc_T^2)^{-1} F_1 = o(1),
\end{aligned}$$

de modo que el primer término converge en media cuadrática a cero y de aquí que también converge en probabilidad a cero. Para el segundo término tenemos que por el supuesto VC3 $T^{-1} \sum_{t=1}^T h_t(0|\Omega_t) \nabla' f_t(\beta^0) \nabla f_t(\beta^0) - \mathbf{D}_T \xrightarrow{p} \mathbf{0}$ o equivalentemente que

$$\begin{aligned}
& T^{-1} \sum_{t=1}^T h_t(0|\Omega_t) \nabla' f_t(\beta^0) \nabla f_t(\beta^0) \\
& - E \left[T^{-1} \sum_{t=1}^T h_t(0|\Omega_t) \nabla' f_t(\beta^0) \nabla f_t(\beta^0) \right] \xrightarrow{p} \mathbf{0},
\end{aligned}$$

de tal forma que debemos probar lo siguiente:

$$\begin{aligned}
& T^{-1} \sum_{t=1}^T \{ (2c_T)^{-1} E[I(|\varepsilon_{t\theta}| < c_T) | \Omega_t] \nabla' f_t(\beta^0) \nabla f_t(\beta^0) \\
& - E[h_t(0|\Omega_t) \nabla' f_t(\beta^0) \nabla f_t(\beta^0)] \} \text{ es } o_p(1)
\end{aligned}$$

o equivalentemente que,

$$\begin{aligned}
& T^{-1} \sum_{t=1}^T \{ (2c_T)^{-1} E[I(|\varepsilon_{t\theta}| < c_T) | \Omega_t] \nabla' f_t(\beta^0) \nabla f_t(\beta^0) \\
& - h_t(0|\Omega_t) \nabla' f_t(\beta^0) \nabla f_t(\beta^0) \} \\
&= T^{-1} \sum_{t=1}^T \{ (2c_T)^{-1} E[I(|\varepsilon_{t\theta}| < c_T) | \Omega_t] - h_t(0|\Omega_t) \} \nabla' f_t(\beta^0) \nabla f_t(\beta^0) \\
&= o_p(1).
\end{aligned}$$

Notemos que,

$$\begin{aligned}
& |(2c_T)^{-1} E [I(|\varepsilon_{t\theta}| < c_T) |\Omega_t] - h_t(0|\Omega_t)| \\
&= \left| (2c_T)^{-1} \int_{-c_T}^{c_T} h_t(\lambda|\Omega_t) d\lambda - h_t(0|\Omega_t) \right| \\
&\leq \left| (2c_T)^{-1} 2c_T \max_{\lambda \in [-c_T, c_T]} h_t(\lambda|\Omega_t) - h_t(0|\Omega_t) \right| \\
&= \left| \max_{\lambda \in [-c_T, c_T]} h_t(\lambda|\Omega_t) - h_t(0|\Omega_t) \right| \\
&\leq L |c_T| = o_p(1),
\end{aligned}$$

por el supuesto AN2 (b) que dice: $h_t(\varepsilon|\Omega_t)$ satisface la condición de Lipschitz $|h_t(\lambda_1|\Omega_t) - h_t(\lambda_2|\Omega_t)| \leq L(\lambda_1 - \lambda_2)$ para alguna constante $L < \infty \forall t$. De lo anterior concluimos que también el segundo término converge en probabilidad a cero y de aquí que $\tilde{\mathbf{D}}_T - \mathbf{D}_T \xrightarrow[T \rightarrow \infty]{p} 0$, finalmente dado que en la primer parte de la demostración mostramos que $\hat{\mathbf{D}}_T - \tilde{\mathbf{D}}_T \xrightarrow[T \rightarrow \infty]{p} 0$ concluimos que $\hat{\mathbf{D}}_T - \mathbf{D}_T \xrightarrow[T \rightarrow \infty]{p} 0$. \square

Teorema 6.4 (*Prueba de cuantil dinámico dentro de la muestra, CD_{DM}*). Bajo los supuestos del Teorema 6.1, Teorema 6.2 y los supuestos DQ1 a DQ6, enunciados a en el Apéndice B. Se sigue el siguiente resultado:

$$[\theta(1-\theta) E(T^{-1} \mathbf{M}_T \mathbf{M}_T')]^{-\frac{1}{2}} T^{-\frac{1}{2}} \mathbf{X}'(\hat{\beta}) \mathbf{Hit}(\hat{\beta}) \stackrel{d}{\sim} N(\mathbf{0}, \mathbf{I}).$$

Además, si se cumple el supuesto DQ7, y las condiciones del Teorema 6.3, tenemos el siguiente resultado asintótico:

$$CD_{DM} \equiv \frac{\mathbf{Hit}'(\hat{\beta}) \mathbf{X}(\hat{\beta}) (\hat{\mathbf{M}}_T \hat{\mathbf{M}}_T')^{-1} \mathbf{X}'(\hat{\beta}) \mathbf{Hit}(\hat{\beta})}{\theta(1-\theta)} \stackrel{d}{\sim} \chi_q^2,$$

donde

$$\begin{aligned}
\hat{\mathbf{M}}_T &\equiv \mathbf{X}'(\hat{\beta}) - \left\{ (2T\hat{c}_T)^{-1} \sum_{t=1}^T I \left[|Y_t - f_t(\hat{\beta})| < \hat{c}_T \right] \mathbf{X}'_t(\hat{\beta}) \nabla f_t(\hat{\beta}) \right\} \\
&\quad \times \hat{\mathbf{D}}_T^{-1} \nabla' f(\hat{\beta}).
\end{aligned}$$

Demostración. El procedimiento de la prueba es el siguiente: Aproximar la función discontinua $Hit_t(\hat{\beta}) \equiv I[Y_t < f_t(\hat{\beta})] - \theta$ mediante una función continua diferenciable, luego aplicar

el teorema el valor medio alrededor de β^0 y mostrar que el estadístico de prueba aproximado converge en distribución a la distribución normal postulada. finalmente, mostrar que está aproximación converge en probabilidad al estadístico de prueba definido en el teorema.

Definamos,

$$Hit_t^\oplus(\hat{\beta}) = \{1 + \exp[c_T^{-1}\hat{\varepsilon}_t]\}^{-1} - \theta = I^*(\hat{\varepsilon}_t) - \theta,$$

donde $\hat{\varepsilon}_t = y_t - f_t(\hat{\beta})$ y c_T es una secuencia no estocástica tal que $\lim_{T \rightarrow \infty} c_T = 0$, de tal forma que:

$$\begin{aligned} \nabla_{\hat{\beta}} Hit_t^\oplus(\hat{\beta}) &= c_T^{-1} \exp[c_T^{-1}\hat{\varepsilon}_t] \{1 + \exp[c_T^{-1}\hat{\varepsilon}_t]\}^{-2} \nabla f_t(\hat{\beta}) \\ &\equiv k_{c_T}(\hat{\varepsilon}_t) \nabla f_t(\hat{\beta}). \end{aligned}$$

Note que $k_{c_T}(\hat{\varepsilon}_t)$ es la función de densidad de probabilidad de una logística con media cero y parámetro c_T . En forma matricial, podemos escribir $\nabla_{\hat{\beta}} \mathbf{Hit}^\oplus(\hat{\beta}) = \mathbf{K}(\hat{\varepsilon}_t) \nabla f(\hat{\beta})$, donde $\mathbf{K}(\hat{\varepsilon}_t)$ es una matriz diagonal con entradas $k_{c_T}(\hat{\varepsilon}_t)$.

Debido a que $\mathbf{X}_t(\hat{\beta})$ es acotada en probabilidad y $Hit_t^\oplus(\hat{\beta})$ es acotada entre $-\theta$ y $1 - \theta$, note que:

$$\begin{aligned} T^{-\frac{1}{2}} \mathbf{X}'(\hat{\beta}) \mathbf{Hit}^\oplus(\hat{\beta}) &= T^{-\frac{1}{2}} \sum_{t=1}^T \{ \mathbf{X}'_t(\hat{\beta}) Hit_t^\oplus(\hat{\beta}) \\ &\times [1 - \sum_{j=1}^{J_i} I(\varepsilon_{\theta t} = \varepsilon_t^j)] \} + o_p(1), \end{aligned}$$

ya que los puntos sobre los cuales $\mathbf{X}_t(\hat{\beta})$ es no diferenciable forman un conjunto de medida cero, por el supuesto DQ3. En lo que sigue por simplicidad de notación, supondremos que $\mathbf{X}_t(\hat{\beta})$ es diferenciable en todas partes. El caso de no diferenciabilidad puede completarse trabajando con

$$T^{-\frac{1}{2}} \sum_{t=1}^T \left\{ \mathbf{X}'_t(\hat{\beta}) Hit_t^\oplus(\hat{\beta}) \left[1 - \sum_{j=1}^{J_i} I(\varepsilon_{\theta t} = \varepsilon_t^j) \right] \right\}$$

más que con $\mathbf{X}'(\hat{\beta}) \mathbf{Hit}^\oplus(\hat{\beta})$.

Aplicando el teorema de valor medio a $T^{-\frac{1}{2}} \mathbf{X}'(\hat{\beta}) \mathbf{Hit}^\oplus(\hat{\beta})$ alrededor de β^0 tenemos lo siguiente:

$$\begin{aligned} &T^{-\frac{1}{2}} \mathbf{X}'(\hat{\beta}) \mathbf{Hit}^\oplus(\hat{\beta}) - T^{-\frac{1}{2}} \mathbf{X}'(\beta^0) \mathbf{Hit}^\oplus(\beta^0) \\ &= T^{-\frac{1}{2}} [\nabla \mathbf{X}(\beta^*) \mathbf{Hit}^\oplus(\beta^*) + \mathbf{X}(\beta^*) \nabla \mathbf{Hit}^\oplus(\beta^*)] (\hat{\beta} - \beta^0) \\ &= T^{-\frac{1}{2}} [\nabla \mathbf{X}(\beta^*) \mathbf{Hit}^\oplus(\beta^*) + \mathbf{X}(\beta^*) \mathbf{K}(\varepsilon_t^*) \nabla f(\beta^*)] (\hat{\beta} - \beta^0), \end{aligned}$$

es decir,

$$\begin{aligned} & T^{-\frac{1}{2}} \mathbf{X}'(\hat{\beta}) \mathbf{Hit}^{\oplus}(\hat{\beta}) \\ = & T^{-\frac{1}{2}} \mathbf{X}'(\beta^0) \mathbf{Hit}^{\oplus}(\beta^0) + T^{-\frac{1}{2}} [\nabla \mathbf{X}(\beta^*) \mathbf{Hit}^{\oplus}(\beta^*) + \mathbf{X}(\beta^*) \mathbf{K}(\varepsilon_t^*) \nabla f(\beta^*)] \times (\hat{\beta} - \beta^0), \end{aligned} \quad (\text{C.2})$$

donde β^* está entre $\hat{\beta}$ y β^0 , y las variables involucradas se definen como:

$$\begin{aligned} \varepsilon_t^* &= y_t - f_t(\beta^*), \\ \mathbf{X}'(\beta) &\equiv [\mathbf{X}'_1(\beta), \dots, \mathbf{X}'_T(\beta)], \\ \nabla \mathbf{X}(\beta) &\equiv [\nabla \mathbf{X}_1(\beta), \dots, \nabla \mathbf{X}_T(\beta)], \\ \mathbf{Hit}^{\oplus}(\beta) &\equiv [\text{Hit}_1^{\oplus}(\beta), \dots, \text{Hit}_T^{\oplus}(\beta)]', \\ \mathbf{K}(\varepsilon_t^*) &= \text{diag} \{k_{c_T}(\varepsilon_1^*), \dots, k_{c_T}(\varepsilon_T^*)\} \\ \nabla f(\beta) &\equiv [\nabla' f_1(\beta), \dots, \nabla' f_T(\beta)]'. \end{aligned} \quad (\text{C.3})$$

Sumando y restando términos apropiados en la ecuación (C.2), podemos describir dicha expresión como sigue:

$$\begin{aligned} & T^{-\frac{1}{2}} \mathbf{X}'(\hat{\beta}) \mathbf{Hit}^{\oplus}(\hat{\beta}) \\ = & T^{-\frac{1}{2}} \mathbf{X}'(\beta^0) \mathbf{Hit}(\beta^0) - E [T^{-1} \mathbf{X}(\beta^0) \mathbf{H} \nabla f(\beta^0)] \mathbf{D}_T^{-1} T^{-\frac{1}{2}} \nabla' f(\beta^0) \mathbf{Hit}(\beta^0) \\ & + E [T^{-1} \mathbf{X}(\beta^0) \mathbf{H} \nabla f(\beta^0)] \mathbf{D}_T^{-1} T^{-\frac{1}{2}} \nabla' f(\beta^0) \mathbf{Hit}(\beta^0) \\ & - T^{-1} [\mathbf{X}(\beta^0) \mathbf{H} \nabla f(\beta^0)] \mathbf{D}_T^{-1} T^{-\frac{1}{2}} \nabla' f(\beta^0) \mathbf{Hit}(\beta^0) \\ & + T^{-\frac{1}{2}} \mathbf{X}'(\beta^0) \mathbf{Hit}^{\oplus}(\beta^0) - T^{-\frac{1}{2}} \mathbf{X}'(\beta^0) \mathbf{Hit}(\beta^0) \\ & + T^{-\frac{1}{2}} [\nabla \mathbf{X}(\beta^*) \mathbf{Hit}^{\oplus}(\beta^*) + \mathbf{X}(\beta^*) \mathbf{K}(\varepsilon_t^*) \nabla f(\beta^*)] (\hat{\beta} - \beta^0) \\ & + T^{-1} [\mathbf{X}(\beta^0) \mathbf{H} \nabla f(\beta^0)] \mathbf{D}_T^{-1} T^{-\frac{1}{2}} \nabla' f(\beta^0) \mathbf{Hit}(\beta^0) \end{aligned} \quad (\text{C.4})$$

Por el supuesto DQ4, tenemos que,

$$T^{-1} \mathbf{X}'(\beta^0) \mathbf{H} \nabla f(\beta^0) - E [T^{-1} \mathbf{X}'(\beta^0) \mathbf{H} \nabla f(\beta^0)] \xrightarrow[T \rightarrow \infty]{p} \mathbf{0},$$

de tal forma que los términos

$$E [T^{-1} \mathbf{X}(\beta^0) \mathbf{H} \nabla f(\beta^0)] \mathbf{D}_T^{-1} T^{-\frac{1}{2}} \nabla' f(\beta^0) \mathbf{Hit}(\beta^0)$$

y

$$-T^{-1} [\mathbf{X}(\beta^0) \mathbf{H} \nabla f(\beta^0)] \mathbf{D}_T^{-1} T^{-\frac{1}{2}} \nabla' f(\beta^0) \mathbf{Hit}(\beta^0)$$

son $o_p(1)$.

Demostremos ahora que $T^{-\frac{1}{2}}\mathbf{X}'(\beta^0)\mathbf{Hit}^\oplus(\beta^0) - T^{-\frac{1}{2}}\mathbf{X}'(\beta^0)\mathbf{Hit}(\beta^0) \xrightarrow[T \rightarrow \infty]{p} 0$, recordemos que por definición,

$$I^*(|\varepsilon_t|) = \{1 + \exp[c_T^{-1}|\varepsilon_t|]\}^{-1}$$

o equivalentemente (para simplificar notación),

$$I^*(|a|) = \{1 + \exp[|b|]\}^{-1} = \frac{1}{1 + \exp[|b|]},$$

Probemos que:

$$\begin{aligned} I^*(|a|) &= \frac{1}{1 + \exp[|b|]} \\ &= 1 - \frac{1}{1 + \exp[-|b|]} \\ &= 1 - I^*(-|a|). \end{aligned}$$

si $b > 0$, tenemos que:

$$\begin{aligned} 1 - I^*(-|a|) &= 1 - \frac{1}{1 + \exp[-|b|]} \\ &= 1 - \frac{1}{1 + \exp[-b]} \\ &= \frac{1 + \exp[-b] - 1}{1 + \exp[-b]} \\ &= \frac{\exp[-b]}{1 + \exp[-b]} \\ &= \frac{1}{\exp[b] + 1} \\ &= I^*(|a|). \end{aligned}$$

Si $b < 0$, entonces

$$\begin{aligned} 1 - I^*(-|a|) &= 1 - \frac{1}{1 + \exp[b]} \\ &= \frac{\exp[b]}{1 + \exp[b]} \\ &= \frac{1}{\exp[-b] + 1} \\ &= I^*(|a|). \end{aligned}$$

De aquí que $I^*(|\varepsilon_t|) = 1 - I^*(-|\varepsilon_t|)$. Para cada t , tenemos que;

$$\begin{aligned}
 |\text{Hit}_t^\oplus(\beta^0) - \text{Hit}_t(\beta^0)| &= |I^*(\varepsilon_{\theta t}) - \theta - I(\varepsilon_{\theta t}) + \theta| \\
 &= |I^*(\varepsilon_{\theta t}) - I(\varepsilon_{\theta t})| \\
 &\leq I^*(|\varepsilon_{\theta t}|) [I(|\varepsilon_{\theta t}| \geq T^{-d}) + I(|\varepsilon_{\theta t}| < T^{-d})] \\
 &\equiv C_t + D_t,
 \end{aligned}$$

donde d es un entero positivo más grande que $1/2$, tal que $\lim_{T \rightarrow \infty} T^d = 0$, y $C_t = I^*(|\varepsilon_{\theta t}|) \times I(|\varepsilon_{\theta t}| \geq T^{-d})$ y $D_t = I^*(|\varepsilon_{\theta t}|) \times I(|\varepsilon_{\theta t}| < T^{-d})$; por lo tanto,

$$\begin{aligned}
 &T^{-\frac{1}{2}} \|\mathbf{X}'(\beta^0) [\mathbf{Hit}^\oplus(\beta^0) - \mathbf{Hit}(\beta^0)]\| \\
 &\leq T^{-\frac{1}{2}} \sum_{t=1}^T \|\mathbf{X}_t(\beta^0) [\text{Hit}_t^\oplus(\beta^0) - \text{Hit}_t(\beta^0)]\| \\
 &\leq T^{-\frac{1}{2}} \sum_{t=1}^T \|\mathbf{X}_t(\beta^0)\| |[\text{Hit}_t^\oplus(\beta^0) - \text{Hit}_t(\beta^0)]| \\
 &\leq T^{-\frac{1}{2}} \sum_{t=1}^T W(\Omega_t) |C_t + D_t|.
 \end{aligned}$$

Notemos que por definición $I^*(|\varepsilon_{\theta t}|)$ es decreciente en $|\varepsilon_{\theta t}|$, tenemos que $C_t \leq I^*(T^{-d})$, por consiguiente,

$$\begin{aligned}
 T^{-\frac{1}{2}} \sum_{t=1}^T E[W(\Omega_t) C_t] &\leq T^{-\frac{1}{2}} \sum_{t=1}^T E[W(\Omega_t) \times I^*(T^{-d})] \\
 &= T^{-\frac{1}{2}} \sum_{t=1}^T E[W(\Omega_t)] [1 + \exp(c_T^{-1} T^{-d})]^{-1} \\
 &\leq T^{-\frac{1}{2}} \sum_{t=1}^T W_0 [1 + \exp(c_T^{-1} T^{-d})]^{-1} \\
 &= T^{\frac{1}{2}} W_0 [1 + \exp(c_T^{-1} T^{-d})]^{-1} \xrightarrow{T \rightarrow \infty} 0.
 \end{aligned}$$

Para D_t , note que $D_t \leq I(|\varepsilon_{\theta t}| < T^{-d})$, ya que $I^*(|\varepsilon_{\theta t}|)$ es una función acotada entre 0 y 1,

de está forma, para cualquier $\xi > 0$, y por la desigualdad de Markov, con $p = 1$ tenemos que

$$\begin{aligned}
& T^{-\frac{1}{2}} \sum_{t=1}^T \Pr [W(\Omega_t) D_t > \xi] \\
& \leq T^{-\frac{1}{2}} \sum_{t=1}^T \Pr [W(\Omega_t) I(|\varepsilon_{\theta t}| < T^{-d}) > \xi] \\
& \leq T^{-\frac{1}{2}} \sum_{t=1}^T \frac{E[W(\Omega_t) I(|\varepsilon_{\theta t}| < T^{-d})]}{\xi} \\
& \leq T^{-\frac{1}{2}} \xi^{-1} \sum_{t=1}^T E[W(\Omega_t)] \Pr[I(|\varepsilon_{\theta t}| < T^{-d})] \\
& = T^{-\frac{1}{2}} \xi^{-1} \sum_{t=1}^T \left\{ E[W(\Omega_t)] \int_{-T^{-d}}^{T^{-d}} h_t(\lambda|\Omega_t) d\lambda \right\} \\
& \leq T^{-\frac{1}{2}} \xi^{-1} \sum_{t=1}^T \left\{ W_0 \int_{-T^{-d}}^{T^{-d}} h_t(\lambda|\Omega_t) d\lambda \right\} \\
& \leq T^{-\frac{1}{2}} \xi^{-1} \sum_{t=1}^T \left\{ W_0 \int_{-T^{-d}}^{T^{-d}} N d\lambda \right\} \text{ por supuesto AN2} \\
& \leq T^{-\frac{1}{2}} \xi^{-1} T^{-1} \{W_0 2T^{-d} N\} \\
& = 2W_0 \xi^{-1} N T^{-d+\frac{1}{2}} \xrightarrow{T \rightarrow \infty} 0, \text{ ya que } d > \frac{1}{2},
\end{aligned}$$

de está forma mostramos que $T^{-\frac{1}{2}} \mathbf{X}'(\beta^0) \mathbf{Hit}^\oplus(\beta^0) - T^{-\frac{1}{2}} \mathbf{X}'(\beta^0) \mathbf{Hit}(\beta^0) \xrightarrow[T \rightarrow \infty]{p} 0$.

Reescribiendo el término

$$T^{-\frac{1}{2}} [\nabla \mathbf{X}(\beta^*) \mathbf{Hit}^\oplus(\beta^*) + \mathbf{X}(\beta^*) \mathbf{K}(\varepsilon_t^*) \nabla f(\beta^*)] (\hat{\beta} - \beta^0),$$

en la ecuación (C.4), tenemos que:

$$\begin{aligned}
& T^{-\frac{1}{2}} [\nabla \mathbf{X}(\beta^*) \mathbf{Hit}^\oplus(\beta^*) + \mathbf{X}(\beta^*) \mathbf{K}(\varepsilon_t^*) \nabla f(\beta^*)] (\hat{\beta} - \beta^0) \\
& = T^{-\frac{1}{2}} [\nabla \mathbf{X}(\beta^*) \mathbf{Hit}^\oplus(\beta^*) + \mathbf{X}(\beta^*) \mathbf{K}(\varepsilon_t^*) \nabla f(\beta^*)] T^{-\frac{1}{2}} \mathbf{D}_T^{-1} \mathbf{D}_T T^{\frac{1}{2}} (\hat{\beta} - \beta^0).
\end{aligned}$$

Entonces las dos últimas líneas en la ecuación (C.4) se cancelan asintóticamente siempre que se cumpla lo siguiente:

$$\mathbf{D}_T T^{\frac{1}{2}} (\hat{\beta} - \beta^0) = -T^{-\frac{1}{2}} \nabla' f(\beta^0) \mathbf{Hit}(\beta^0) + o_p(1),$$

y

$$\begin{aligned} & T^{-1} [\nabla \mathbf{X}(\beta^*) \mathbf{Hit}^\oplus(\beta^*) + \mathbf{X}(\beta^*) \mathbf{K}(\varepsilon_t^*) \nabla f(\beta^*)] \mathbf{D}_T^{-1} \\ = & T^{-1} [\mathbf{X}(\beta^0) \mathbf{H} \nabla f(\beta^0)] \mathbf{D}_T^{-1}, \end{aligned}$$

la primer igualdad se cumple ya que en la demostración del Teorema 6.2 se mostró que:

$$\mathbf{D}_T T^{\frac{1}{2}} (\hat{\beta} - \beta^0) = -T^{-\frac{1}{2}} \sum_{t=1}^T \text{Hit}_t(\beta^0) \nabla' f_t(\beta^0) + o_p(1).$$

Para la segunda igualdad, tenemos que por ser $\hat{\beta}$ un estimador consistente de β^0 se cumple el siguiente resultado:

$$\begin{aligned} & T^{-1} [\nabla \mathbf{X}(\beta^*) \mathbf{Hit}^\oplus(\beta^*) + \mathbf{X}(\beta^*) \mathbf{K}(\varepsilon_t^*) \nabla f(\beta^*)] \\ & - T^{-1} [\nabla \mathbf{X}(\beta^0) \mathbf{Hit}^\oplus(\beta^0) + \mathbf{X}(\beta^0) \mathbf{K}(\varepsilon_t) \nabla f(\beta^0)] \xrightarrow[T \rightarrow \infty]{p} \mathbf{0}. \end{aligned}$$

Equivalente a la demostración de que $T^{-\frac{1}{2}} \mathbf{X}'(\beta^0) \mathbf{Hit}^\oplus(\beta^0) - T^{-\frac{1}{2}} \mathbf{X}'(\beta^0) \mathbf{Hit}(\beta^0) \xrightarrow[T \rightarrow \infty]{p} 0$, podemos mostrar que $T^{-1} \nabla \mathbf{X}(\beta^0) \mathbf{Hit}^\oplus(\beta^0) - T^{-1} \nabla \mathbf{X}(\beta^0) \mathbf{Hit}(\beta^0) \xrightarrow[T \rightarrow \infty]{p} 0$. Calculemos los primeros dos momentos, y mostremos que $E [T^{-1} \nabla \mathbf{X}(\beta^0) \mathbf{Hit}(\beta^0)] = 0$ y $\text{Var} [T^{-1} \nabla \mathbf{X}(\beta^0) \mathbf{Hit}(\beta^0)] \xrightarrow[T \rightarrow \infty]{} 0$, tal y como se muestra a continuación.

$$\begin{aligned} E \left[T^{-1} \sum_{t=1}^T \nabla \mathbf{X}_t(\beta^0) \text{Hit}_t(\beta^0) \right] &= E \left\{ T^{-1} \sum_{t=1}^T \nabla \mathbf{X}_t(\beta^0) E [\text{Hit}_t(\beta^0) | \Omega_t] \right\} \\ &= 0, \end{aligned}$$

y

$$\begin{aligned}
& \text{Var} \left[T^{-1} \sum_{t=1}^T \nabla \mathbf{X}_t(\beta^0) \text{Hit}_t(\beta^0) \right] \\
&= T^{-2} \sum_{t=1}^T E \left\{ [\nabla \mathbf{X}_t(\beta^0) \text{Hit}_t(\beta^0)] [\nabla \mathbf{X}_t(\beta^0) \text{Hit}_t(\beta^0)]' \right\} \\
&= T^{-2} \sum_{t=1}^T E \left\{ E [\nabla \mathbf{X}_t(\beta^0) \text{Hit}_t^2(\beta^0) \nabla' \mathbf{X}_t(\beta^0) | \Omega_t] \right\} \\
&= T^{-2} \sum_{t=1}^T E [\nabla \mathbf{X}_t(\beta^0) \nabla' \mathbf{X}_t(\beta^0)] \theta(1 - \theta) \\
&\leq T^{-2} \sum_{t=1}^T E [\nabla \mathbf{X}_t(\beta^0) \nabla' \mathbf{X}_t(\beta^0)] \quad (\text{ya que } \theta(1 - \theta) < 1) \\
&\leq T^{-2} \sum_{t=1}^T E [Z(\Omega_t) Z(\Omega_t)], \text{ por DQ3} \\
&\leq T^{-2} \sum_{t=1}^T Z_0 = T^{-1} Z_0 \xrightarrow{T \rightarrow \infty} 0,
\end{aligned}$$

de lo anterior concluimos que $T^{-1} \nabla \mathbf{X}(\beta^0) \mathbf{H} \text{it}(\beta^0) \xrightarrow[T \rightarrow \infty]{p} 0$.

Resta sólo probar que $T^{-1} [\mathbf{X}'(\beta^0) \mathbf{K}(\varepsilon_{\theta t}) \nabla f(\beta^0)] - T^{-1} [\mathbf{X}'(\beta^0) \mathbf{H} \nabla f(\beta^0)] = o_p(1)$, veamos,

$$\begin{aligned}
& T^{-1} [\mathbf{X}'(\beta^0) \mathbf{K}(\varepsilon_{\theta t}) \nabla f(\beta^0)] - T^{-1} [\mathbf{X}'(\beta^0) \mathbf{H} \nabla f(\beta^0)] \\
&= T^{-1} \left\{ \sum_{t=1}^T k_{c_T}(\varepsilon_{\theta t}) \mathbf{X}'_t(\beta^0) \nabla f_t(\beta^0) - \sum_{t=1}^T E[k_{c_T}(\varepsilon_{\theta t}) | \Omega_t] \mathbf{X}'_t(\beta^0) \nabla f_t(\beta^0) \right. \\
&\quad \left. + \sum_{t=1}^T E[k_{c_T}(\varepsilon_{\theta t}) | \Omega_t] \mathbf{X}'_t(\beta^0) \nabla f_t(\beta^0) - \sum_{t=1}^T h_t(0 | \Omega_t) \mathbf{X}'_t(\beta^0) \nabla f_t(\beta^0) \right\} \\
&= T^{-1} \left\{ \sum_{t=1}^T \{k_{c_T}(\varepsilon_{\theta t}) - E[k_{c_T}(\varepsilon_{\theta t}) | \Omega_t]\} \mathbf{X}'_t(\beta^0) \nabla f_t(\beta^0) \right. \\
&\quad \left. \sum_{t=1}^T \{E[k_{c_T}(\varepsilon_{\theta t}) | \Omega_t] - h_t(0 | \Omega_t)\} \mathbf{X}'_t(\beta^0) \nabla f_t(\beta^0) \right\}.
\end{aligned}$$

Mostraremos primeramente que $E[k_{c_T}(\varepsilon_{\theta t}) | \Omega_t] \xrightarrow{T \rightarrow \infty} h_t(0 | \Omega_t)$. Sea $k(u) = e^u (1 + e^u)^{-2}$

(recordemos que $k_{c_T}(\varepsilon_{\theta t}) = c_T^{-1} \exp[c_T^{-1} \varepsilon_{\theta t}] \{1 + \exp[c_T^{-1} \varepsilon_{\theta t}]\}$), entonces

$$E[k_{c_T}(\varepsilon_{\theta t}) | \Omega_t] = \int_{-\infty}^{\infty} k(u) h_t(u c_T | \Omega_t) du,$$

tomando la expansión de Taylor de $h_t(u c_T | \Omega_t)$ al rededor del cero tenemos que:

$$h_t(u c_T | \Omega_t) = h_t(0 | \Omega_t) + h'_t(0 | \Omega_t) u c_T + o(c_T),$$

luego,

$$\begin{aligned} E[k_{c_T}(\varepsilon_{\theta t}) | \Omega_t] &= \int_{-\infty}^{\infty} k(u) [h_t(0 | \Omega_t) + h'_t(0 | \Omega_t) u c_T + o(c_T)] du \\ &= h_t(0 | \Omega_t) \int_{-\infty}^{\infty} k(u) du + c_T h'_t(0 | \Omega_t) \int_{-\infty}^{\infty} k(u) u du \\ &\quad + o(c_T) \int_{-\infty}^{\infty} k(u) du \\ &= h_t(0 | \Omega_t) + o(c_T), \end{aligned}$$

dado que la función k corresponde a la densidad de una variable aleatoria logística con parámetro c_T y valor esperado igual a cero. Resumiendo hemos mostrado que

$E[k_{c_T}(\varepsilon_{\theta t}) | \Omega_t] \xrightarrow{T \rightarrow \infty} h_t(0 | \Omega_t)$. Demostremos ahora que

$$\begin{aligned} &T^{-1} \sum_{t=1}^T \{k_{c_T}(\varepsilon_{\theta t}) - E[k_{c_T}(\varepsilon_{\theta t}) | \Omega_t]\} \mathbf{X}'_t(\beta^0) \nabla f_t(\beta^0) \\ &= o_p(1), \end{aligned}$$

para lo cual calcularemos sus primeros dos momentos,

$$\begin{aligned}
& E \left\{ T^{-1} \sum_{t=1}^T \{k_{c_T}(\varepsilon_{\theta t}) - E[k_{c_T}(\varepsilon_{\theta t}) | \Omega_t]\} \mathbf{X}'_t(\beta^0) \nabla f_t(\beta^0) \right\} \\
&= T^{-1} \sum_{t=1}^T E \{ \{k_{c_T}(\varepsilon_{\theta t}) - E[k_{c_T}(\varepsilon_{\theta t}) | \Omega_t]\} \mathbf{X}'_t(\beta^0) \nabla f_t(\beta^0) \} \\
&= T^{-1} \sum_{t=1}^T E \{ \{E\{k_{c_T}(\varepsilon_{\theta t}) - E[k_{c_T}(\varepsilon_{\theta t}) | \Omega_t]\} \mathbf{X}'_t(\beta^0) \nabla f_t(\beta^0)\} | \Omega_t \} \\
&= T^{-1} \sum_{t=1}^T E \{ E\{k_{c_T}(\varepsilon_{\theta t}) - E[k_{c_T}(\varepsilon_{\theta t}) | \Omega_t] | \Omega_t\} \mathbf{X}'_t(\beta^0) \nabla f_t(\beta^0) \} \\
&= T^{-1} \sum_{t=1}^T E \{ E[k_{c_T}(\varepsilon_{\theta t}) | \Omega_t] - E[k_{c_T}(\varepsilon_{\theta t}) | \Omega_t] \mathbf{X}'_t(\beta^0) \nabla f_t(\beta^0) \} \\
&= 0.
\end{aligned}$$

y

$$\begin{aligned}
& \left\| \text{Var} \left\{ T^{-1} \sum_{t=1}^T \{k_{c_T}(\varepsilon_{\theta t}) - E[k_{c_T}(\varepsilon_{\theta t}) | \Omega_t]\} \times \mathbf{X}'_t(\beta^0) \nabla f_t(\beta^0) \right\} \right\| \quad (\text{C.5}) \\
&= \left\| E \left\{ T^{-1} \sum_{t=1}^T \{k_{c_T}(\varepsilon_{\theta t}) - E[k_{c_T}(\varepsilon_{\theta t}) | \Omega_t]\} \mathbf{X}'_t(\beta^0) \nabla f_t(\beta^0) \right\}^2 \right\| \\
&= \left\| T^{-2} \sum_{t=1}^T E \left\{ \{k_{c_T}(\varepsilon_{\theta t}) - E[k_{c_T}(\varepsilon_{\theta t}) | \Omega_t]\}^2 [\mathbf{X}'_t(\beta^0) \nabla f_t(\beta^0)]^2 \right\} \right\| \\
&= \left\| T^{-2} \sum_{t=1}^T E \left\{ E \left\{ \{k_{c_T}(\varepsilon_{\theta t}) - E[k_{c_T}(\varepsilon_{\theta t}) | \Omega_t]\}^2 [\mathbf{X}'_t(\beta^0) \nabla f_t(\beta^0)]^2 \right\} | \Omega_t \right\} \right\| \\
&= \left\| T^{-2} \sum_{t=1}^T E \left\{ E \left\{ \{k_{c_T}(\varepsilon_{\theta t}) - E[k_{c_T}(\varepsilon_{\theta t}) | \Omega_t]\}^2 | \Omega_t \right\} [\mathbf{X}'_t(\beta^0) \nabla f_t(\beta^0)]^2 \right\} \right\|,
\end{aligned}$$

ya que los valores esperados de los productos cruzados son cero ($E\{E[k_{c_T}(\varepsilon_{\theta t}) | \Omega_t]\} = E(k_{c_T}(\varepsilon_{\theta t})) = 0$) y por la ley de las esperanzas iteradas. Veamos ahora que

$$\begin{aligned}
E \{ \{ k_{c_T}(\varepsilon_{\theta t}) - E[k_{c_T}(\varepsilon_{\theta t}) | \Omega_t] \}^2 | \Omega_t \} &= O(c_T^{-1}), \text{ es decir,} \\
&E \{ \{ k_{c_T}(\varepsilon_{\theta t}) - E[k_{c_T}(\varepsilon_{\theta t}) | \Omega_t] \}^2 | \Omega_t \} \\
&= E[k_{c_T}(\varepsilon_{\theta t})^2 | \Omega_t] - \{ E[k_{c_T}(\varepsilon_{\theta t}) | \Omega_t] \}^2 \\
&= E[k_{c_T}(\varepsilon_{\theta t})^2 | \Omega_t] - [h(0 | \Omega_t) + o(c_T)]^2 \\
&= \int_{-\infty}^{\infty} k_{c_T}(\lambda)^2 h(\lambda | \Omega_t) d\lambda - [h(0 | \Omega_t)]^2 + o(c_T) \\
&= c_T^{-1} \int_{-\infty}^{\infty} k(u)^2 h(uc_T | \Omega_t) du - [h(0 | \Omega_t)]^2 + o(c_T),
\end{aligned}$$

tomando la expansión en series de Taylor de $h(uc_T | \Omega_t)$ alrededor del cero y notando que $k(u) \leq 1/4$ para $u \leq 0$, tenemos la siguiente desigualdad:

$$\begin{aligned}
&E \{ \{ k_{c_T}(\varepsilon_{\theta t}) - E[k_{c_T}(\varepsilon_{\theta t}) | \Omega_t] \}^2 | \Omega_t \} \\
&\leq \frac{1}{4} c_T^{-1} \int_{-\infty}^{\infty} k(u) \left[\begin{array}{c} h_t(0 | \Omega_t) \\ + h'_t(0 | \Omega_t) uc_T + o(c_T) \end{array} \right] du \\
&\quad - [h(0 | \Omega_t)]^2 + o(c_T) \\
&\leq \frac{1}{4} c_T^{-1} [h_t(0 | \Omega_t) + o(c_T)] - [h(0 | \Omega_t)]^2 + o(c_T) \\
&= O(c_T^{-1}).
\end{aligned}$$

De está forma retomando al ecuación para la varianza dada en (C.5), tenemos que

$$\begin{aligned}
&\left\| Var \left\{ T^{-1} \sum_{t=1}^T \{ k_{c_T}(\varepsilon_{\theta t}) - E[k_{c_T}(\varepsilon_{\theta t}) | \Omega_t] \} \mathbf{X}'_t(\beta^0) \nabla f_t(\beta^0) \right\} \right\| \\
&\leq \left\| T^{-2} \sum_{t=1}^T E \left\{ O(c_T^{-1}) [\mathbf{X}'_t(\beta^0) \nabla f_t(\beta^0)]^2 \right\} \right\| \\
&\leq T^{-2} \sum_{t=1}^T O(c_T^{-1}) E \left\| [\mathbf{X}'_t(\beta^0) \nabla f_t(\beta^0)] \right\|^2 \\
&\leq T^{-2} \sum_{t=1}^T O(c_T^{-1}) E \{ W(\Omega_t) F(\Omega_t) \}^2 \\
&\leq T^{-2} \sum_{t=1}^T W_1 O(c_T^{-1}) = T^{-1} W_1 O(c_T^{-1}) \xrightarrow{T \rightarrow \infty} 0.
\end{aligned}$$

Luego por la desigualdad de Chebyshev, se sigue que

$$T^{-1} \sum_{t=1}^T \{ k_{c_T}(\varepsilon_{\theta t}) - E[k_{c_T}(\varepsilon_{\theta t}) | \Omega_t] \} \mathbf{X}'_t(\beta^0) \nabla f_t(\beta^0) = o_p(1).$$

De los resultados de arriba podemos describir la ecuación (C.4) como sigue:

$$\begin{aligned}
& T^{-\frac{1}{2}} \mathbf{X}'(\hat{\beta}) \mathbf{Hit}^{\oplus}(\hat{\beta}) \\
&= T^{-\frac{1}{2}} [\mathbf{X}'(\beta^0) \mathbf{Hit}(\beta^0) - E [T^{-1} \mathbf{X}(\beta^0) \mathbf{H} \nabla f(\beta^0)] \mathbf{D}_T^{-1} \nabla f(\beta^0) \mathbf{Hit}(\beta^0)] + o_p(1) \\
&\equiv T^{-\frac{1}{2}} \mathbf{M}_T \mathbf{Hit}(\beta^0) + o_p(1),
\end{aligned}$$

con $\mathbf{M}_T \equiv \mathbf{X}'(\beta^0) \mathbf{Hit}(\beta^0) - E [T^{-1} \mathbf{X}(\beta^0) \mathbf{H} \nabla f(\beta^0)] \mathbf{D}_T^{-1} \nabla f(\beta^0)$.

Finalmente, análogo a como mostramos que

$$T^{-\frac{1}{2}} \mathbf{X}(\beta^0) \mathbf{Hit}^{\oplus}(\beta^0) - T^{-\frac{1}{2}} \mathbf{X}(\beta^0) \mathbf{Hit}(\beta^0) = o_p(1),$$

podemos mostrar que $T^{-\frac{1}{2}} \mathbf{X}'(\hat{\beta}) \mathbf{Hit}^{\oplus}(\hat{\beta}) - T^{-\frac{1}{2}} \mathbf{X}'(\hat{\beta}) \mathbf{Hit}(\hat{\beta}) = o_p(1)$, de tal forma que llegamos a la siguiente ecuación:

$$T^{-\frac{1}{2}} \mathbf{X}'(\hat{\beta}) \mathbf{Hit}(\hat{\beta}) = T^{-\frac{1}{2}} \mathbf{M}_T \mathbf{Hit}(\beta^0) + o_p(1), \quad (\text{C.6})$$

recordemos que por el supuesto DQ6, $T^{-\frac{1}{2}} \mathbf{M}_T \mathbf{Hit}(\beta^0)$ obedece el TLC, con primeros dos momentos como sigue:

$$\begin{aligned}
E [T^{-\frac{1}{2}} \mathbf{M}_T \mathbf{Hit}(\beta^0)] &= E \left\{ E [T^{-\frac{1}{2}} \mathbf{M}_T \mathbf{Hit}(\beta^0) | \Omega_t] \right\} \\
&= E \left\{ T^{-\frac{1}{2}} \mathbf{M}_T E [\mathbf{Hit}(\beta^0) | \Omega_t] \right\} \\
&= \mathbf{0},
\end{aligned}$$

ya que

$$\begin{aligned}
E [\mathbf{Hit}_t(\beta^0) | \Omega_t] &= E \{ I [y_t < f_t(\beta^0)] - \theta | \Omega_t \} \\
&= \Pr [y_t < f_t(\beta^0)] - \theta \\
&= \theta - \theta \\
&= 0,
\end{aligned}$$

para $t = 1, \dots, T$. Mientras que la varianza está dada como:

$$\begin{aligned}
\text{Var} [T^{-\frac{1}{2}} \mathbf{M}_T \mathbf{Hit}(\beta^0)] &= E \left\{ [T^{-\frac{1}{2}} \mathbf{M}_T \mathbf{Hit}(\beta^0)] [T^{-\frac{1}{2}} \mathbf{M}_T \mathbf{Hit}(\beta^0)]' \right\} \\
&= E [T^{-1} \mathbf{M}_T \mathbf{Hit}(\beta^0) \mathbf{Hit}(\beta^0)' \mathbf{M}_T'] \\
&= E \{ E [T^{-1} \mathbf{M}_T \mathbf{Hit}(\beta^0) \mathbf{Hit}(\beta^0)' \mathbf{M}_T' | \Omega_t] \} \\
&= E \{ T^{-1} \mathbf{M}_T E [\mathbf{Hit}(\beta^0) \mathbf{Hit}(\beta^0)' | \Omega_t] \mathbf{M}_T' \},
\end{aligned}$$

con

$$\begin{aligned} & \mathbf{Hit}(\beta^0)\mathbf{Hit}(\beta^0)' \\ = & \begin{pmatrix} Hit_1^2(\beta^0) & Hit_1(\beta^0)Hit_2(\beta^0) & \cdots & Hit_1(\beta^0)Hit_T(\beta^0) \\ Hit_2(\beta^0)Hit_1(\beta^0) & Hit_2^2(\beta^0) & \cdots & Hit_2(\beta^0)Hit_T(\beta^0) \\ \vdots & \vdots & \ddots & \vdots \\ Hit_T(\beta^0)Hit_1(\beta^0) & Hit_T(\beta^0)Hit_2(\beta^0) & \cdots & Hit_T^2(\beta^0) \end{pmatrix}, \end{aligned}$$

tomando valores esperados condicionales en Ω_t , tenemos que

$$\begin{aligned} E[\mathbf{Hit}(\beta^0)\mathbf{Hit}(\beta^0)'|\Omega_t] &= \begin{pmatrix} \theta(1-\theta) & 0 & \cdots & 0 \\ 0 & \theta(1-\theta) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \theta(1-\theta) \end{pmatrix} \\ &= \theta(1-\theta)\mathbf{I}_{T \times T}, \end{aligned}$$

y de aquí que la varianza se reduce a lo siguiente:

$$\begin{aligned} \text{Var}\left[T^{-\frac{1}{2}}\mathbf{M}_T\mathbf{Hit}(\beta^0)\right] &= E\left\{T^{-1}\mathbf{M}_T\theta(1-\theta)\mathbf{I}_{T \times T}\mathbf{M}_T'\right\} \\ &= \theta(1-\theta)E\left[T^{-1}\mathbf{M}_T\mathbf{M}_T'\right]. \end{aligned}$$

Resumiendo, llegamos a que $T^{-\frac{1}{2}}\mathbf{M}_T\mathbf{Hit}(\beta^0) \sim N(\mathbf{0}, \theta(1-\theta)E[T^{-1}\mathbf{M}_T\mathbf{M}_T'])$, pero por la igualdad dada en la ecuación (C.6) podemos concluir que

$$\frac{T^{-\frac{1}{2}}\mathbf{X}'(\hat{\beta})\mathbf{Hit}(\hat{\beta})}{\{\theta(1-\theta)E[T^{-1}\mathbf{M}_T\mathbf{M}_T']\}^{\frac{1}{2}}} \xrightarrow{d} N(\mathbf{0}, \mathbf{I}_{T \times T}).$$

Para finalizar la prueba debemos mostrar que $T^{-1}\hat{\mathbf{M}}_T\hat{\mathbf{M}}_T' - E[T^{-1}\mathbf{M}_T\mathbf{M}_T'] \xrightarrow{p} \mathbf{0}$, es decir, por la definición de convergencia en probabilidad debemos mostrar que: para cualquier $\xi > 0$

$$\lim_{T \rightarrow \infty} \Pr\left(\left|T^{-1}\hat{\mathbf{M}}_T\hat{\mathbf{M}}_T' - E[T^{-1}\mathbf{M}_T\mathbf{M}_T']\right| \leq \xi\right) = 1$$

donde

$$\hat{\mathbf{M}}_T \equiv \mathbf{X}'(\hat{\beta}) - \hat{\mathbf{G}}_T\hat{\mathbf{D}}_T^{-1}\nabla'f(\hat{\beta})$$

y

$$\begin{aligned} \hat{\mathbf{G}}_T &= (2T\hat{c}_T)^{-1} \sum_{t=1}^T I\left[\left|y_t - f_t(\hat{\beta})\right| < \hat{c}_T\right] \\ &\quad \times \mathbf{X}_t'(\hat{\beta})\nabla f_t(\hat{\beta}). \end{aligned}$$

Realicemos el producto de las matrices para cada uno de los términos involucrados,

$$\begin{aligned}
& T^{-1}\hat{\mathbf{M}}_T\hat{\mathbf{M}}'_T \\
&= T^{-1}[\mathbf{X}'(\hat{\beta}) - \hat{\mathbf{G}}_T\hat{\mathbf{D}}_T^{-1}\nabla'f(\hat{\beta})][\mathbf{X}'(\hat{\beta}) - \hat{\mathbf{G}}_T\hat{\mathbf{D}}_T^{-1}\nabla'f(\hat{\beta})]' \\
&= T^{-1}[\mathbf{X}'(\hat{\beta}) - \hat{\mathbf{G}}_T\hat{\mathbf{D}}_T^{-1}\nabla'f(\hat{\beta})][\mathbf{X}(\hat{\beta}) - \nabla f(\hat{\beta})\hat{\mathbf{D}}_T^{-1}\hat{\mathbf{G}}'_T] \\
&= T^{-1}[\mathbf{X}'(\hat{\beta})\mathbf{X}(\hat{\beta}) - \hat{\mathbf{G}}_T\hat{\mathbf{D}}_T^{-1}\nabla'f(\hat{\beta})\mathbf{X}(\hat{\beta}) \\
&\quad - \mathbf{X}'(\hat{\beta})\nabla f(\hat{\beta})\hat{\mathbf{D}}_T^{-1}\hat{\mathbf{G}}'_T + \hat{\mathbf{G}}_T\hat{\mathbf{D}}_T^{-1}\nabla'f(\hat{\beta})\nabla f(\hat{\beta})\hat{\mathbf{D}}_T^{-1}\hat{\mathbf{G}}'_T] \\
&= T^{-1}[\mathbf{X}'(\hat{\beta})\mathbf{X}(\hat{\beta}) - 2\mathbf{X}'(\hat{\beta})\nabla f(\hat{\beta})\hat{\mathbf{D}}_T^{-1}\hat{\mathbf{G}}'_T + \hat{\mathbf{G}}_T\hat{\mathbf{D}}_T^{-1}\nabla'f(\hat{\beta})\nabla f(\hat{\beta})\hat{\mathbf{D}}_T^{-1}\hat{\mathbf{G}}'_T],
\end{aligned}$$

y

$$\begin{aligned}
& T^{-1}\mathbf{M}_T\mathbf{M}'_T \\
&= T^{-1}\{\mathbf{X}'(\beta^0) - E[T^{-1}\mathbf{X}'(\beta^0)\mathbf{H}\nabla f(\beta^0)]\mathbf{D}_T^{-1}\nabla'f(\beta^0)\} \\
&\quad \{\mathbf{X}'(\beta^0) - E[T^{-1}\mathbf{X}'(\beta^0)\mathbf{H}\nabla f(\beta^0)]\mathbf{D}_T^{-1}\nabla'f(\beta^0)\}' \\
&= T^{-1}\{\mathbf{X}'(\beta^0) - E[T^{-1}\mathbf{X}'(\beta^0)\mathbf{H}\nabla f(\beta^0)]\mathbf{D}_T^{-1}\nabla'f(\beta^0)\} \\
&\quad \{\mathbf{X}(\beta^0) - \nabla f(\beta^0)\mathbf{D}_T^{-1}E[T^{-1}\mathbf{X}'(\beta^0)\mathbf{H}\nabla f(\beta^0)]'\} \\
&= T^{-1}\{\mathbf{X}'(\beta^0)\mathbf{X}(\beta^0) - 2\mathbf{X}'(\beta^0)\nabla f(\beta^0)\mathbf{D}_T^{-1}E[T^{-1}\mathbf{X}'(\beta^0)\mathbf{H}\nabla f(\beta^0)]' \\
&\quad + E[T^{-1}\mathbf{X}'(\beta^0)\mathbf{H}\nabla f(\beta^0)]\mathbf{D}_T^{-1}\nabla'f(\beta^0)\nabla f(\beta^0)\mathbf{D}_T^{-1}E[T^{-1}\mathbf{X}'(\beta^0)\mathbf{H}\nabla f(\beta^0)]'\}.
\end{aligned}$$

Por el supuesto DQ5 sabemos que $T^{-1}\mathbf{M}_T\mathbf{M}'_T - E(T^{-1}\mathbf{M}_T\mathbf{M}'_T) \xrightarrow[T \rightarrow \infty]{p} 0$, y dado nuestro interés de mostrar que $T^{-1}\hat{\mathbf{M}}_T\hat{\mathbf{M}}'_T - E(T^{-1}\mathbf{M}_T\mathbf{M}'_T) \xrightarrow[T \rightarrow \infty]{p} 0$, el resultado se cumple si probamos que: $T^{-1}\hat{\mathbf{M}}_T\hat{\mathbf{M}}'_T - T^{-1}\mathbf{M}_T\mathbf{M}'_T \xrightarrow[T \rightarrow \infty]{p} 0$, para lo cual debemos probar cada una de las siguientes convergencias en probabilidad:

- 1.- $T^{-1}\mathbf{X}'(\hat{\beta})\mathbf{X}(\hat{\beta}) - T^{-1}\mathbf{X}'(\beta^0)\mathbf{X}(\beta^0) \xrightarrow[T \rightarrow \infty]{p} 0$.
- 2.- $T^{-1}\mathbf{X}'(\hat{\beta})\nabla f(\hat{\beta})\hat{\mathbf{D}}_T^{-1}\hat{\mathbf{G}}'_T - T^{-1}\mathbf{X}'(\beta^0)\nabla f(\beta^0)\mathbf{D}_T^{-1}E[T^{-1}\mathbf{X}'(\beta^0)\mathbf{H}\nabla f(\beta^0)]' \xrightarrow[T \rightarrow \infty]{p} 0$
- 3.- $T^{-1}[\hat{\mathbf{G}}_T\hat{\mathbf{D}}_T^{-1}\nabla'f(\hat{\beta})\nabla f(\hat{\beta})\hat{\mathbf{D}}_T^{-1}\hat{\mathbf{G}}'_T] - E[T^{-1}\mathbf{X}'(\beta^0)\mathbf{H}\nabla f(\beta^0)] [T^{-1}\mathbf{D}_T^{-1}\nabla'f(\beta^0)\nabla f(\beta^0)]\{\mathbf{D}_T^{-1}E[T^{-1}\mathbf{X}'(\beta^0)\mathbf{H}\nabla f(\beta^0)]'\} \xrightarrow[T \rightarrow \infty]{p} 0$.

En el primer caso, tenemos que $T^{-1}\mathbf{X}'(\hat{\beta})\mathbf{X}(\hat{\beta}) - T^{-1}\mathbf{X}'(\beta^0)\mathbf{X}(\beta^0) \xrightarrow[T \rightarrow \infty]{p} 0$, ya que $\hat{\beta} - \beta^0 \xrightarrow[T \rightarrow \infty]{p} 0$ implica que $\mathbf{X}'(\hat{\beta}) - \mathbf{X}'(\beta^0) \xrightarrow[T \rightarrow \infty]{p} 0$ o equivalentemente $\mathbf{X}'(\hat{\beta}) \xrightarrow[T \rightarrow \infty]{p} \mathbf{X}'(\beta^0)$, por el supuesto DQ1 y además dado que $\mathbf{X}(\hat{\beta})$ es una matriz de dimensión $T \times q$, con cada renglón Ω_t medible (por ejemplo, las columnas de la matriz $\mathbf{X}(\hat{\beta})$ puede ser rezagos del vector $\mathbf{Hit}(\hat{\beta})$,

e igualmente la función $f(\hat{\beta})$, de manera que se garantiza la convergencia). De esta forma tenemos que $\mathbf{X}'(\hat{\beta})\mathbf{X}(\hat{\beta}) \xrightarrow[T \rightarrow \infty]{p} \mathbf{X}'(\beta^0)\mathbf{X}(\beta^0)$ y de aquí $\mathbf{X}'(\hat{\beta})\mathbf{X}(\hat{\beta}) - \mathbf{X}'(\beta^0)\mathbf{X}(\beta^0) \xrightarrow[T \rightarrow \infty]{p} \mathbf{0}$. Finalmente $T^{-1}[\mathbf{X}'(\hat{\beta})\mathbf{X}(\hat{\beta}) - \mathbf{X}'(\beta^0)\mathbf{X}(\beta^0)] \xrightarrow[T \rightarrow \infty]{p} \mathbf{0}$ ya que $T^{-1} = O_p(1)$ (**Teorema:** Sea una secuencia $\{Y_n\}_{n=1}^{\infty}$ acotada en probabilidad ($O_p(1)$), si $X_n \xrightarrow{p} 0$ entonces $X_n Y_n \xrightarrow{p} 0$, cuando $n \rightarrow \infty$).

Para la convergencia en el punto 2, observemos que por el supuesto DQ4. se tiene que: $T^{-1}\mathbf{X}'(\beta^0)\mathbf{H}\nabla f(\beta^0) - E[T^{-1}\mathbf{X}'(\beta^0)\mathbf{H}\nabla f(\beta^0)] \xrightarrow{p} \mathbf{0}$, entonces probar 2 es equivalente a probar que:

$$T^{-1}\mathbf{X}'(\hat{\beta})\nabla f(\hat{\beta})\hat{\mathbf{D}}_T^{-1}\hat{\mathbf{G}}'_T - T^{-1}\mathbf{X}'(\beta^0)\nabla f(\beta^0)\mathbf{D}_T^{-1}T^{-1}\nabla f(\beta^0)'\mathbf{H}\mathbf{X}(\beta^0) \xrightarrow[T \rightarrow \infty]{p} \mathbf{0}.$$

Dicha convergencia se alcanza siempre que $\hat{\mathbf{G}}_T - T^{-1}\mathbf{X}(\beta^0)\mathbf{H}\nabla f(\beta^0) \xrightarrow{p} \mathbf{0}$, es decir,

$$(2T\hat{c}_T)^{-1} \sum_{t=1}^T I[|\hat{\varepsilon}_t| < \hat{c}_T] \mathbf{X}'_t(\hat{\beta})\nabla f_t(\hat{\beta}) - T^{-1}\mathbf{X}(\beta^0)\mathbf{H}\nabla f(\beta^0) \xrightarrow[T \rightarrow \infty]{p} \mathbf{0},$$

con $\hat{\varepsilon}_t \equiv y_t - f_t(\hat{\beta})$. Procedamos como en la demostración del Teorema 6.3, de tal forma que definamos

$$\tilde{\mathbf{G}}_T = (2Tc_T)^{-1} \sum_{t=1}^T I[|\varepsilon_{t\theta}| < c_T] \mathbf{X}'_t(\beta^0)\nabla f_t(\beta^0),$$

y mostremos que $\hat{\mathbf{G}}_T - \tilde{\mathbf{G}}_T = o_p(1)$ y $\tilde{\mathbf{G}}_T - T^{-1}\mathbf{X}(\beta^0)\mathbf{H}\nabla f(\beta^0) = o_p(1)$ y por ende $\hat{\mathbf{G}}_T - T^{-1}\mathbf{X}(\beta^0)\mathbf{H}\nabla f(\beta^0) \xrightarrow{p} \mathbf{0}$.

Mostremos ahora que $\hat{\mathbf{G}}_T - \tilde{\mathbf{G}}_T = o_p(1)$,

$$\begin{aligned}
& \left\| \hat{\mathbf{G}}_T - \tilde{\mathbf{G}}_T \right\| \\
= & \left\| (2T\hat{c}_T)^{-1} \sum_{t=1}^T \{I[|\hat{\varepsilon}_t| < \hat{c}_T] \mathbf{X}'_t(\hat{\beta}) \nabla f_t(\hat{\beta}) \right. \\
& \left. - (2Tc_T)^{-1} \sum_{t=1}^T I[|\varepsilon_{t\theta}| < c_T] \mathbf{X}'_t(\beta^0) \nabla f_t(\beta^0) \right\| \\
= & \left\| (2T\hat{c}_T)^{-1} \sum_{t=1}^T I[|\hat{\varepsilon}_t| < \hat{c}_T] \mathbf{X}'_t(\hat{\beta}) \nabla f_t(\hat{\beta}) \right. \\
& \left. - (2Tc_T)^{-1} \sum_{t=1}^T \{I[|\varepsilon_{t\theta}| < c_T] \mathbf{X}'_t(\beta^0) \nabla f_t(\beta^0) \right. \\
& + (2T\hat{c}_T)^{-1} \sum_{t=1}^T I[|\varepsilon_{t\theta}| < c_T] \{ \mathbf{X}'_t(\hat{\beta}) \nabla f_t(\hat{\beta}) - \mathbf{X}'_t(\beta^0) \nabla f_t(\hat{\beta}) \\
& - \mathbf{X}'_t(\beta^0) \nabla f_t(\hat{\beta}) + \mathbf{X}'_t(\beta^0) \nabla f_t(\hat{\beta}) - \mathbf{X}'_t(\beta^0) \nabla f_t(\beta^0) \\
& \left. + \mathbf{X}'_t(\beta^0) \nabla f_t(\beta^0) \} \right\| \\
= & \frac{c_T}{\hat{c}_T} \left\| (2Tc_T)^{-1} \sum_{t=1}^T \{ [I(|\hat{\varepsilon}_t| < \hat{c}_T) - I(|\varepsilon_{t\theta}| < c_T)] \mathbf{X}'_t(\hat{\beta}) \nabla f_t(\hat{\beta}) \right. \\
& + I(|\varepsilon_{t\theta}| < c_T) [\mathbf{X}'_t(\hat{\beta}) - \mathbf{X}'_t(\beta^0)] \nabla f_t(\hat{\beta}) \\
& + I(|\varepsilon_{t\theta}| < c_T) \mathbf{X}'_t(\beta^0) [\nabla f_t(\hat{\beta}) - \nabla f_t(\beta^0)] \\
& \left. + \frac{c_T - \hat{c}_T}{c_T} I(|\varepsilon_{t\theta}| < c_T) \mathbf{X}'_t(\beta^0) \nabla f_t(\beta^0) \right\|.
\end{aligned}$$

Cómo antes la función indicadora en la primer línea cumple la siguiente desigualdad:

$$\begin{aligned}
& |I(|\hat{\varepsilon}_t| < \hat{c}_T) - I(|\varepsilon_{t\theta}| < c_T)| \\
\leq & I(|\varepsilon_{t\theta} - c_T| < |\delta_t(\hat{\beta})| + |\hat{c}_T - c_T|) + \\
& + I(|\varepsilon_{t\theta} + c_T| < |\delta_t(\hat{\beta})| + |\hat{c}_T + c_T|).
\end{aligned}$$

Así

$$\begin{aligned}
\left\| \hat{\mathbf{G}}_T - \tilde{\mathbf{G}}_T \right\| &\leq \frac{c_T}{\hat{c}_T} (2T c_T)^{-1} \\
&\times \sum_{t=1}^T \left\{ I \left(|\varepsilon_t - c_T| < \left| \delta_t(\hat{\boldsymbol{\beta}}) \right| + |\hat{c}_T - c_T| \right) \right. \\
&+ I \left(|\varepsilon_t + c_T| < \left| \delta_t(\hat{\boldsymbol{\beta}}) \right| + |\hat{c}_T + c_T| \right) \left. \right\} W(\Omega_t) F(\Omega_t) \\
&+ I(|\varepsilon_t| < c_T) W(\Omega_t) M(\Omega_t, \hat{\boldsymbol{\beta}}, \boldsymbol{\beta}^0) \\
&+ \frac{c_T - \hat{c}_T}{c_T} I(|\varepsilon_{t\theta}| < c_T) W(\Omega_t) F(\Omega_t) \\
&+ I(|\varepsilon_t| < c_T) S(\Omega_t, \hat{\boldsymbol{\beta}}, \boldsymbol{\beta}^0) F(\Omega_t) \\
&\equiv \frac{c_T}{\hat{c}_T} (B_1 + B_2 + B_3 + B_4),
\end{aligned}$$

donde B_i $i = 1, 2, 3$, y 4 corresponde a cada uno de los sumando en el lado derecho de la desigualdad anterior. Ahora suponga que para $d > 0$ dado, el cual puede elegirse arbitrariamente pequeño, y T suficientemente grande se cumple que $\left| \frac{c_T - \hat{c}_T}{c_T} \right| < d$ y que $c_T^{-1} \left\| \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0 \right\| < d$. Es posible mostrar que $E(B_i) = O(d)$, para $i = 1, 2, 3, 4$. Es decir podemos encontrar funciones acotadas para $\left\| \hat{\mathbf{G}}_T - \tilde{\mathbf{G}}_T \right\|$ las cuales pueden hacerse arbitrariamente pequeñas en probabilidad, eligiendo d suficientemente pequeño. Por ejemplo, calculemos el valor esperado para B_1 , para lo cual notemos que por el teorema del valor medio se sigue la siguiente desigualdad:

$$\left| \delta_t(\hat{\boldsymbol{\beta}}) \right| = \left| f_t(\hat{\boldsymbol{\beta}}) - f_t(\boldsymbol{\beta}^0) \right| \leq \left\| \nabla f_t(\boldsymbol{\beta}^*) \right\| \left\| \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0 \right\|,$$

y de aquí que:

$$\begin{aligned}
E(B_1) &= E \left\{ (2Tc_T)^{-1} \sum_{t=1}^T [I(|\varepsilon_t - c_T| < |\delta_t(\hat{\beta})| + |\hat{c}_T - c_T|) \right. \\
&\quad \left. + I(|\varepsilon_t + c_T| < |\delta_t(\hat{\beta})| + |\hat{c}_T + c_T|)] F(\Omega_t) W(\Omega_t) \right\} \\
&\leq E \left\{ (2Tc_T)^{-1} \sum_{t=1}^T \left[I(|\varepsilon_t - c_T| < \|\nabla f_t(\beta^*)\| \|\hat{\beta} - \beta^0\| \right. \right. \\
&\quad \left. \left. + |\hat{c}_T - c_T|) + I(|\varepsilon_t + c_T| < \|\nabla f_t(\beta^*)\| \|\hat{\beta} - \beta^0\| + |\hat{c}_T + c_T|) \right] \right\} \\
&\quad \times F(\Omega_t) W(\Omega_t) \} \\
&\leq E \left\{ (2Tc_T)^{-1} \sum_{t=1}^T [I(|\varepsilon_t - c_T| < dc_T[F(\Omega_t) + 1]) \right. \\
&\quad \left. + I(|\varepsilon_t + c_T| < dc_T[F(\Omega_t) + 1])] F(\Omega_t) W(\Omega_t) \right\} \\
&\leq E \left\{ (2Tc_T)^{-1} \sum_{t=1}^T 4Hdc_T[F(\Omega_t) + 1] F(\Omega_t) W(\Omega_t) \right\} \\
&\leq E \left\{ (2Tc_T)^{-1} \sum_{t=1}^T 4Hdc_T[F(\Omega_t)]^2 W(\Omega_t) \right\} \\
&\leq E \left\{ T^{-1} \sum_{t=1}^T 2Hd[F(\Omega_t)]^2 W(\Omega_t) \right\} \\
&= 2HdT^{-1} \sum_{t=1}^T E \{ [F(\Omega_t)]^2 W(\Omega_t) \} \\
&\leq 2HdT^{-1} \sum_{t=1}^T W_1 = 2HdW_1 = O(d). \quad H \neq 0 \text{ constante.}
\end{aligned}$$

Igualmente $E(B_2) = O(d)$, ya que:

$$\begin{aligned}
E(B_2) &= E \left\{ (2Tc_T)^{-1} \sum_{t=1}^T I(|\varepsilon_t| < c_T) W(\Omega_t) M(\Omega_t, \hat{\beta}, \beta^0) \right\} \\
&\leq E \left\{ (2Tc_T)^{-1} \sum_{t=1}^T W(\Omega_t) M(\Omega_t, \hat{\beta}, \beta^0) \right\} \\
&\leq (2Tc_T)^{-1} \sum_{t=1}^T E \{ W(\Omega_t) M(\Omega_t, \hat{\beta}, \beta^0) \} \\
&\leq (2Tc_T)^{-1} TW_0 \|\hat{\beta} - \beta^0\| \\
&\leq 2dW_0 = O(d).
\end{aligned}$$

Similarmente podemos mostrar que $E(B_3) = O(d)$ y $E(B_4) = O(d)$.

Mostremos ahora que $\tilde{\mathbf{G}}_T - \mathbf{G}_T = o_p(1)$,

$$\begin{aligned}
& \tilde{\mathbf{G}}_T - \mathbf{G}_T \\
&= (2Tc_T)^{-1} \sum_{t=1}^T I(|\varepsilon_{t\theta}| < c_T) \mathbf{X}'_t(\boldsymbol{\beta}^0) \nabla f_t(\boldsymbol{\beta}^0) \\
&\quad - E[T^{-1} \sum_{t=1}^T h_t(0|\Omega_t) \mathbf{X}'_t(\boldsymbol{\beta}^0) \nabla f_t(\boldsymbol{\beta}^0)] \\
&= (2Tc_T)^{-1} \sum_{t=1}^T \{I(|\varepsilon_{t\theta}| < c_T) \mathbf{X}'_t(\boldsymbol{\beta}^0) \nabla f_t(\boldsymbol{\beta}^0) \\
&\quad - E[I(|\varepsilon_{t\theta}| < c_T) |\Omega_t] \mathbf{X}'_t(\boldsymbol{\beta}^0) \nabla f_t(\boldsymbol{\beta}^0)\} \\
&\quad + T^{-1} \sum_{t=1}^T \{(2c_T)^{-1} E[I(|\varepsilon_{t\theta}| < c_T) |\Omega_t] \mathbf{X}'_t(\boldsymbol{\beta}^0) \nabla f_t(\boldsymbol{\beta}^0) \\
&\quad - E[h_t(0|\Omega_t) \mathbf{X}'_t(\boldsymbol{\beta}^0) \nabla f_t(\boldsymbol{\beta}^0)]\}.
\end{aligned}$$

Calculemos la esperanza y la varianza del primer término,

$$\begin{aligned}
& E \left\{ (2Tc_T)^{-1} \sum_{t=1}^T \begin{bmatrix} I(|\varepsilon_{t\theta}| < c_T) \mathbf{X}'_t(\boldsymbol{\beta}^0) \nabla f_t(\boldsymbol{\beta}^0) \\ -E[I(|\varepsilon_{t\theta}| < c_T) |\Omega_t] \mathbf{X}'_t(\boldsymbol{\beta}^0) \nabla f_t(\boldsymbol{\beta}^0) \end{bmatrix} \right\} \\
&= (2Tc_T)^{-1} \sum_{t=1}^T \left\{ \begin{bmatrix} E[I(|\varepsilon_{t\theta}| < c_T) \mathbf{X}'_t(\boldsymbol{\beta}^0) \nabla f_t(\boldsymbol{\beta}^0)] \\ -E[E[I(|\varepsilon_{t\theta}| < c_T) |\Omega_t] \mathbf{X}'_t(\boldsymbol{\beta}^0) \nabla f_t(\boldsymbol{\beta}^0)] \end{bmatrix} \right\} \\
&= 0,
\end{aligned}$$

ya que

$$\begin{aligned}
E[I(|\varepsilon_{t\theta}| < c_T) \mathbf{X}'_t(\boldsymbol{\beta}^0) \nabla f_t(\boldsymbol{\beta}^0)] &= E\{E[I(|\varepsilon_{t\theta}| < c_T) \mathbf{X}'_t(\boldsymbol{\beta}^0) \nabla f_t(\boldsymbol{\beta}^0)] |\Omega_t\} \\
&= E\{E[I(|\varepsilon_{t\theta}| < c_T) |\Omega_t] \mathbf{X}'_t(\boldsymbol{\beta}^0) \nabla f_t(\boldsymbol{\beta}^0)\}.
\end{aligned}$$

Mientras que la varianza está dada por:

$$\begin{aligned}
& \text{Var}\{(2Tc_T)^{-1} \sum_{t=1}^T \{[I(|\varepsilon_{t\theta}| < c_T) - E[I(|\varepsilon_{t\theta}| < c_T) | \Omega_t]] \\
& \quad \times \mathbf{X}'_t(\beta^0) \nabla f_t(\beta^0)\}\}\} \\
&= E\{(2Tc_T)^{-1} \sum_{t=1}^T \{[I(|\varepsilon_{t\theta}| < c_T) - E[I(|\varepsilon_{t\theta}| < c_T) | \Omega_t]] \\
& \quad \times \mathbf{X}'_t(\beta^0) \nabla f_t(\beta^0)\}^2\} \\
&= (2Tc_T)^{-2} E\left\{\sum_{t=1}^T \{[I(|\varepsilon_{t\theta}| < c_T) - E[I(|\varepsilon_{t\theta}| < c_T) | \Omega_t]]\}^2\right. \\
& \quad \left. \times [\mathbf{X}'_t(\beta^0) \nabla f_t(\beta^0)]^2\right\},
\end{aligned}$$

esto porque las esperanzas de los dobles son cero por la ley de las esperanzas iteradas, entonces continuando tenemos que:

$$\begin{aligned}
& (2Tc_T)^{-2} E\left\{\sum_{t=1}^T \{[I(|\varepsilon_{t\theta}| < c_T) - E[I(|\varepsilon_{t\theta}| < c_T) | \Omega_t]]\}^2\right. \\
& \quad \left. \times [\mathbf{X}'_t(\beta^0) \nabla f_t(\beta^0)]^2\right\} \\
&\leq (2Tc_T)^{-2} \sum_{t=1}^T E[F(\Omega_t) W(\Omega_t)]^2 \leq (2Tc_T)^{-2} \sum_{t=1}^T W_1 \\
&= (4Tc_T^2)^{-1} W_1 = o(1),
\end{aligned}$$

de modo que el primer término converge en media cuadrática a cero y de aquí que también converge en probabilidad a cero. Para el segundo término tenemos que por el supuesto DQ4

$$\begin{aligned}
& T^{-1} \sum_{t=1}^T h_t(0|\Omega_t) \mathbf{X}'_t(\beta^0) \nabla f_t(\beta^0) \\
& - E\left[T^{-1} \sum_{t=1}^T h_t(0|\Omega_t) \mathbf{X}'_t(\beta^0) \nabla f_t(\beta^0)\right] \xrightarrow{p} \mathbf{0},
\end{aligned}$$

de tal forma que debemos probar que

$$\begin{aligned}
& T^{-1} \sum_{t=1}^T \{ (2c_T)^{-1} E [I(|\varepsilon_{t\theta}| < c_T) |\Omega_t] \mathbf{X}'_t(\beta^0) \nabla f_t(\beta^0) \\
& - E [h_t(0|\Omega_t) \mathbf{X}'_t(\beta^0) \nabla f_t(\beta^0)] \} \\
& = T^{-1} \sum_{t=1}^T \{ (2c_T)^{-1} E [I(|\varepsilon_{t\theta}| < c_T) |\Omega_t] - h_t(0|\Omega_t) \} \\
& \quad \times \mathbf{X}'_t(\beta^0) \nabla f_t(\beta^0) \\
& = o_p(1).
\end{aligned}$$

Como antes (Teorema 6.3), notemos que

$$\begin{aligned}
& |(2c_T)^{-1} E [I(|\varepsilon_{t\theta}| < c_T) |\Omega_t] - h_t(0|\Omega_t)| \\
& = \left| (2c_T)^{-1} \int_{-c_T}^{c_T} h_t(\lambda|\Omega_t) d\lambda - h_t(0|\Omega_t) \right| \\
& \leq \left| (2c_T)^{-1} 2c_T \max_{\lambda \in [-c_T, c_T]} h_t(\lambda|\Omega_t) - h_t(0|\Omega_t) \right| \\
& = \left| \max_{\lambda \in [-c_T, c_T]} h_t(\lambda|\Omega_t) - h_t(0|\Omega_t) \right| \\
& \leq L |c_T| = o_p(1),
\end{aligned}$$

por el supuesto AN2 (b) que dice: $h_t(\varepsilon|\Omega_t)$ satisface la condición de Lipschitz $|h_t(\lambda_1|\Omega_t) - h_t(\lambda_2|\Omega_t)| \leq L(\lambda_1 - \lambda_2)$ para alguna constante $L < \infty \forall t$. De lo anterior concluimos que también el segundo término converge en probabilidad a cero y de aquí que $\hat{\mathbf{G}}_T - \mathbf{G}_T \xrightarrow{p} 0$, finalmente dado que en la primer parte de la demostración mostramos que $\hat{\mathbf{G}}_T - \tilde{\mathbf{G}}_T \xrightarrow{p} 0$ concluimos que $\hat{\mathbf{G}}_T - \mathbf{G}_T \xrightarrow{p} 0$ y la convergencia en el punto dos se cumple. Finalmente reexpresando la diferencia en el punto tres, obtenemos que,

$$\begin{aligned}
& T^{-1} [\hat{\mathbf{G}}_T \hat{\mathbf{D}}_T^{-1} \nabla' f(\hat{\beta}) \nabla f(\hat{\beta}) \hat{\mathbf{D}}_T^{-1} \hat{\mathbf{G}}_T'] \\
& - T^{-1} [\mathbf{G}_T \mathbf{D}_T^{-1} \nabla' f(\beta^0) \nabla f(\beta^0) \mathbf{D}_T^{-1} \mathbf{G}_T] \xrightarrow{p} 0,
\end{aligned}$$

porque cada una de las componentes etiquetadas con gorro converge en probabilidad a su respectiva contraparte teórica. Luego, $T^{-1} \hat{\mathbf{M}}_T \hat{\mathbf{M}}_T' - E [T^{-1} \mathbf{M}_T \mathbf{M}_T'] \xrightarrow{p} 0$ y el estadístico

$$CD_{DM} = \frac{\mathbf{Hit}'(\hat{\beta}) \mathbf{X}(\hat{\beta}) (\hat{\mathbf{M}}_T \hat{\mathbf{M}}_T')^{-1} \mathbf{X}'(\hat{\beta}) \mathbf{Hit}(\hat{\beta})}{\theta(1-\theta)} \underset{T \rightarrow \infty}{\overset{d}{\sim}} \chi_q^2.$$

□

Teorema 5 (*Prueba de cuantil dinámico fuera de la muestra, CD_{FM}*). Bajo los supuestos de los Teoremas 6.1 y 6.2, y los supuestos de DQ1 a DQ3 y DQ8 y DQ9, enunciados en el Apéndice B, se tiene que:

$$CD_{FM} \equiv \frac{\mathbf{Hit}'(\hat{\beta}_{T_R})\mathbf{X}(\hat{\beta}_{T_R}) \left[\mathbf{X}'(\hat{\beta}_{T_R})\mathbf{X}(\hat{\beta}_{T_R}) \right]^{-1} \mathbf{X}'(\hat{\beta}_{T_R})\mathbf{Hit}(\hat{\beta}_{T_R})}{\theta(1-\theta)} \underset{R \rightarrow \infty}{\overset{d}{\rightsquigarrow}} \chi_q^2.$$

Demostración. Al igual que en el Teorema 6.4, lo primero es aproximar la función $\mathbf{Hit}(\hat{\beta}_{T_R})$ mediante una función continua diferenciable y posteriormente aplicar el teorema del valor medio a dicha aproximación,

$$\begin{aligned} & N_R^{-1/2} \mathbf{X}'(\hat{\beta}_{T_R}) \mathbf{Hit}^\oplus(\hat{\beta}_{T_R}) \\ &= N_R^{-1/2} \{ \mathbf{X}'(\beta^0) \mathbf{Hit}^\oplus(\beta^0) \\ & \quad + [\nabla \mathbf{X}(\beta^*) \mathbf{Hit}^\oplus(\beta^*) + \mathbf{X}(\beta^*) \mathbf{K}(\varepsilon_t^*) \nabla f(\beta^*)] (\hat{\beta}_{T_R} - \beta^0) \} \end{aligned}$$

donde β^* está entre $\hat{\beta}_{T_R}$ y β^0 y el resto de las variables son definidas en las ecuaciones (C.3). Luego

$$\begin{aligned} & \lim_{R \rightarrow \infty} N_R^{-1/2} \mathbf{X}'(\hat{\beta}_{T_R}) \mathbf{Hit}^\oplus(\hat{\beta}_{T_R}) \\ &= \lim_{R \rightarrow \infty} \{ N_R^{-1/2} \{ \mathbf{X}'(\beta^0) \mathbf{Hit}^\oplus(\beta^0) \\ & \quad + [\nabla \mathbf{X}(\beta^*) \mathbf{Hit}^\oplus(\beta^*) + \mathbf{X}(\beta^*) \mathbf{K}(\varepsilon_t^*) \nabla f(\beta^*)] (\hat{\beta}_{T_R} - \beta^0) \} \\ &= \lim_{R \rightarrow \infty} N_R^{-1/2} \mathbf{X}'(\beta^0) \mathbf{Hit}^\oplus(\beta^0) + \lim_{R \rightarrow \infty} \left(\frac{N_R}{T_R} \right)^{1/2} \\ & \quad \times \frac{1}{N_R} [\nabla \mathbf{X}(\beta^*) \mathbf{Hit}^\oplus(\beta^*) + \mathbf{X}(\beta^*) \mathbf{K}(\varepsilon_t^*) \nabla f(\beta^*)] T_R^{1/2} (\hat{\beta}_{T_R} - \beta^0). \end{aligned}$$

Por el supuesto DQ8, así como el hecho de que $\hat{\beta}_{T_R}$ es un estimador consistente de β^0 y el Teorema de Slutsky se sigue que:

$$\lim_{R \rightarrow \infty} \left(\frac{N_R}{T_R} \right)^{1/2} \frac{1}{N_R} [\nabla \mathbf{X}(\beta^*) \mathbf{Hit}^\oplus(\beta^*) + \mathbf{X}(\beta^*) \mathbf{K}(\varepsilon_t^*) \nabla f(\beta^*)] T_R^{1/2} (\hat{\beta}_{T_R} - \beta^0) = \mathbf{0},$$

y de aquí que

$$\lim_{R \rightarrow \infty} N_R^{-1/2} \mathbf{X}'(\hat{\beta}_{T_R}) \mathbf{Hit}^\oplus(\hat{\beta}_{T_R}) = \lim_{R \rightarrow \infty} N_R^{-1/2} \mathbf{X}'(\beta^0) \mathbf{Hit}^\oplus(\beta^0).$$

En el Teorema 6.4 demostramos que

$$T_R^{-1/2} \mathbf{X}'(\beta^0) \mathbf{Hit}^\oplus(\beta^0) - T_R^{-1/2} \mathbf{X}'(\beta^0) \mathbf{Hit}(\beta^0) = o_p(1),$$

de forma equivante podemos mostrar que

$$N_R^{-1/2} \mathbf{X}'(\beta^0) \mathbf{Hit}^\oplus(\beta^0) - N_R^{-1/2} \mathbf{X}'(\beta^0) \mathbf{Hit}(\beta^0) = o_p(1)$$

y similarmente

$$N_R^{-1/2} \mathbf{X}'(\hat{\beta}_{T_R}) \mathbf{Hit}^\oplus(\hat{\beta}_{T_R}) - N_R^{-1/2} \mathbf{X}'(\hat{\beta}_{T_R}) \mathbf{Hit}(\hat{\beta}_{T_R}) = o_p(1),$$

lo que implica que

$$\lim_{R \rightarrow \infty} N_R^{-1/2} \mathbf{X}'(\hat{\beta}_{T_R}) \mathbf{Hit}(\hat{\beta}_{T_R}) = \lim_{R \rightarrow \infty} N_R^{-1/2} \mathbf{X}'(\beta^0) \mathbf{Hit}(\beta^0).$$

Sabemos además que $N_R^{-1/2} \mathbf{X}'(\beta^0) \mathbf{Hit}(\beta^0)$ obedece el teorema del límite central y que sus primeros dos momentos son como sigue:

$$\begin{aligned} E \left[N_R^{-1/2} \mathbf{X}'(\beta^0) \mathbf{Hit}(\beta^0) \right] &= E \left\{ E \left[N_R^{-1/2} \mathbf{X}'(\beta^0) \mathbf{Hit}(\beta^0) \right] | \Omega_T \right\} \\ &= E \left\{ N_R^{-1/2} \mathbf{X}'(\beta^0) E \left[\mathbf{Hit}(\beta^0) \right] | \Omega_T \right\} \\ &= \mathbf{0}, \end{aligned}$$

y

$$\begin{aligned} &\text{Var} \left[N_R^{-1/2} \mathbf{X}'(\beta^0) \mathbf{Hit}(\beta^0) \right] \\ &= E \left[N_R^{-1/2} \mathbf{X}'(\beta^0) \mathbf{Hit}(\beta^0) \right] \left[N_R^{-1/2} \mathbf{X}'(\beta^0) \mathbf{Hit}(\beta^0) \right]' \\ &= E \left[N_R^{-1} \mathbf{X}'(\beta^0) \mathbf{Hit}(\beta^0) \mathbf{Hit}'(\beta^0) \mathbf{X}(\beta^0) \right] \\ &= E \{ E \left[N_R^{-1} \mathbf{X}'(\beta^0) \mathbf{Hit}(\beta^0) \mathbf{Hit}'(\beta^0) \mathbf{X}(\beta^0) | \Omega_T \right] \} \\ &= E \{ N_R^{-1} \mathbf{X}'(\beta^0) E \left[\mathbf{Hit}(\beta^0) \mathbf{Hit}'(\beta^0) | \Omega_T \right] \mathbf{X}(\beta^0) \} \\ &= E \left[N_R^{-1} \mathbf{X}'(\beta^0) \theta(1-\theta) \mathbf{I}_{N_R \times N_R} \mathbf{X}(\beta^0) \right] \\ &= \theta(1-\theta) N_R^{-1} \mathbf{X}'(\beta^0) \mathbf{X}(\beta^0). \end{aligned}$$

Notemos que similar a la demostración del Teorema 6.4, se tiene que

$$N_R^{-1} \mathbf{X}'(\hat{\beta}_{T_R}) \mathbf{X}(\hat{\beta}_{T_R}) - N_R^{-1} \mathbf{X}'(\beta^0) \mathbf{X}(\beta^0) \xrightarrow{p} 0,$$

de modo que,

$$\frac{N_R^{-1/2} \mathbf{X}'(\hat{\beta}_{T_R}) \mathbf{Hit}(\hat{\beta}_{T_R})}{\left\{ \theta(1-\theta) N_R^{-1} \mathbf{X}'(\hat{\beta}_{T_R}) \mathbf{X}(\hat{\beta}_{T_R}) \right\}^{1/2}} \stackrel{a}{\sim} N(\mathbf{0}, \mathbf{I}_{N_R \times N_R}),$$

y por ende,

$$\begin{aligned} & N_R^{-1/2} \mathbf{X}'(\hat{\beta}_{T_R}) \mathbf{Hit}(\hat{\beta}_{T_R}) \left\{ \theta(1-\theta) N_R^{-1} \mathbf{X}'(\hat{\beta}_{T_R}) \mathbf{X}(\hat{\beta}_{T_R}) \right\}^{-1} \\ & \times N_R^{-1/2} \mathbf{Hit}'(\hat{\beta}_{T_R}) \mathbf{X}(\hat{\beta}_{T_R}) \\ = & \frac{\mathbf{X}'(\hat{\beta}_{T_R}) \mathbf{Hit}(\hat{\beta}_{T_R}) \left\{ \mathbf{X}'(\hat{\beta}_{T_R}) \mathbf{X}(\hat{\beta}_{T_R}) \right\}^{-1} \mathbf{Hit}'(\hat{\beta}_{T_R}) \mathbf{X}(\hat{\beta}_{T_R})}{\theta(1-\theta)} \stackrel{a}{\underset{R \rightarrow \infty}{\sim}} \chi_q^2. \end{aligned}$$

Lo cual concluye la demostración del teorema. □

Bibliografía

1. Berger, J. O. and Wolpert, R. L. (1984). *The Likelihood Principle*. IMS, Hayward, Calif..
2. Berkowitz, J. (2001). Testing Density Forecasts with Applications to Risk Management. *Journal of Business & Economic Statistics*. **19**, 465-474.
3. Butler, R. W. (1986). Predictive Likelihood Inference with Applications. *Journal of the Royal Statistical Society. Series B (Methodological)*. **48**, No.1, 1-38.
4. Bjørnstad, J. F. (1990). Predictive Likelihood: A Review. *Statistical Science*. **5**, No. 1, 242-265.
5. Caner, M. and Hansen, B. H. (2001). Threshold Autoregression with a Unit Root. *Econometrica*. **69**, No. 6, 1555-1596.
6. Clements, M. P., Frances, P. H., Smith, J. and Van Dijk D. (2003). On SETAR Non-linearity and Forecasting. *Journal of Forecasting*. **22**, 359-375.
7. Clements, M. P. and Hendry, D. F. (1998). *Forecasting Economic Time Series*. Cambridge University Press.
8. Clements, M. P. and Smith, J. (1997). The performance of alternative forecasting methods for SETAR models. *International Journal of Forecasting*. **13**, 463-475.
9. Clements, M. P. and Smith, J. (1999). A Monte Carlo Study of the Forecasting Performance of Empirical SETAR Models. *Journal of Applied Econometrics*. **14**, 123-141.
10. Chan, K. S. (1993). Consistency and limiting distribution of the least square estimation of a threshold autoregressive model. *The Annals of Statistics*. **21**, No. 1, 520-533.
11. Chan, K. S., Petrucci, J. D, Tong, H. and Woolford, S. W. (1985). A Multiple Threshold AR(1) Model. *Journal Applied Probability*. **22**, 267-279.

12. Chan, K. S. and Tong, H. (1985). On the use of the deterministic lyapunov functions for the ergodicity of stochastic difference equations. *Advances in Applied Probability*. **17**, 666-678.
13. Chan, K. S. and Tsay, R. S. (1998). Limiting Properties of the Least Square Estimator of a Continuous Threshold Autoregressive Model. *Biometrika*. **85**, No. 2, 413-426.
14. Chen, C. W. S. and Lee J. C. (1995). Bayesian Inference of Threshold Autoregressive Models. *Journal of Time Series Analysis*. **16**, No. 5, 483-492.
15. Chen, C. W. S. (1998). A Bayesian analysis of generalized threshold autoregressive models. *Statistics & Probability Letters*. **40**, 15-22.
16. Chen, R. and Tsay, R. S. (1991). On the Ergodicity of Tar(1) Processes. *The Annals of Applied Probability*. **1**, No. 4, 613-634.
17. Christoffersen, P. F. and Diebold, F. X. (1996). Further Results on Forecasting and Models Selection under Asymmetric Loss. *Journal of Applied Econometrics*. **11**, 561-571.
18. Dacco, R. and Satchell, S. (1999). Why do Regime-Switching Models Forecast so Badly. *Journal of Forecasting*. **18**, 1-16.
19. Davies, N., Pemberton, D. and Petrucci, J. D. (1988). An automatic procedure for identification, estimation and forecasting univariate self exiting threshold autoregressive models. *The Statistician*. **37**, 199-204.
20. De Gooijer, J. G. and De Bruin, P. T. (1998). On Forecasting SETAR processes. *Statistics & Probability Letters*. **37**, 7-14.
21. Diebold, F. X. and Mariano, R. S. (1995). Comparing Predictive Accuracy. *Journal of Business & Economic Statistics*. **13**, Num. 3.
22. Diebold, F. X., Gunther, T. A. and Tay, A.S. (1998). Evaluating Density Forecast with Applications to financial Risk Management. *International Economic Review*. **39**, 863-883.
23. Enders, W. and Granger, C. (1998). Unit Root tests and asymmetric adjustment with an example sing the term structure of interest rates. *Journal of Business & Economic Statistics*. **16**, No. 3, 304-311.
24. Engel, R. F. and Manganelli, S. (2004). CAVIAR: Conditional Autoregressive Value at Risk by Regression Quantiles. *Journal of Business & Economic Statistics*. **22**, No. 4, 367-381.

25. Fan, J. and Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer Series in Statistics.
26. Fisher, R. A. (1956). *Statistical Methods and Scientific Inference*. Oliver and Boyd, London.
27. Fuller, W. A. (1996). *Introduction to Statistical Time Series. Second Edition*. Wiley Series in Probability and Statistics.
28. Geisser, S. (2002). *Predictive Inference: An Introduction*. Monographs on Statistics and Applied Probability 55, Chapman & Hall.
29. González, M. and Gonzalo, J. (1998). Threshold Unit Root Models. *Working paper, Universidad Carlos III*.
30. González, M. and Gonzalo, J. (1999). Threshold Unit Root (TUR) Models for Interest Rates. *Working paper, Universidad Carlos III*.
31. Granger, C. W. J. and Teräsvirta, T. (1993). *Modelling Nonlinear Economic Relationships*. Advanced Texts in Econometrics.
32. Granger, C. W. J. (2002). Long Memory, Volatility, Risk and Distribution. *Mimeo*.
33. Hannan, E. J. (1973). Central limit theorems for time series regression. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*. **26**, 157-170. Título posterior: Probability Theory and Related Fields.
34. Hansen, B. E. (1997). Inference in TAR Models. *Studies in Nonlinear Dynamics and Econometrics*. **2**(1), 1-14.
35. Hansen, B. E. (2000). Sample Splitting and Threshold Estimation. *Econometrica*. **68**, No. 3., 575-603.
36. Hinkley, D. V. (1979). Predictive Likelihood. *Annals of Statistics*. **7**, 718-728. Corrigendum **8** 694.
37. Kalbfleisch, J. D. and Sprott, D. A. (1970). Applications of likelihood methods to models involving large numbers of parameters (with discussion). *Journal Royal Statistics Society Serie B*. **32**, 175-208.
38. Kadem, B. and Fokianos, K. (2002). *Regression Models for Time Series Analysis*. Wiley Series in Probability and Statistics.
39. Koenker, R. and Bassett, G. (1978). Regression Quantiles. *Econometrica*, **46**, 33-50.

40. Koenker, R. and Hallock, K. (2001). Quantile Regression: An Introduction.
www.econ.uiuc.edu/~roger/research/intro.html.
41. Koul H. L. (1996). Asymptotics of Some Estimators and Sequential Residual Empiricals in Nonlinear Time Series. *The Annals of Statistics*. **24**, No.1, 380-404.
42. Koul H. L., and Saleh A. K. Md. E. (1995). Autoregression Quantiles and Related Rank-Scores Processes. *The Annals of Statistics*. **23**, No.2, 670-689.
43. Koul, H. L. and Schick, A. (1996). Adaptative Estimation in Random Cefficient Autoregressive Model. *The Annals of Statistics*. **24**, No. 3, 1025-1052.
44. Lauritzen, S. L. (1974). Sufficiency, prediction and extreme models. *Scand. Journal Statistics*. **1**, 128-134.
45. Mathiasen, P. E. (1979). Prediction functions. *Scand. Journal Statistics*. **6**, 1-21.
46. Mukherjee K. (1999). Asymptotics of Quantiles and Rank Scores in Nonlinear Time Series. *Journal of Time Series Analysis*. **20**, 173-192.
47. Nicholls, D. F. and Quinn, B. G. (1982). *Random coefficient autoregressive models: an introduction*. Springer Verlag, Lectures Notes in Statistics. **11**.
48. Pesaran, M. H. and Potter, S. M. (1997). A floor and ceiling model of US output. *Journal of Economic Dynamics and Control*. **21**, 661-695.
49. Petrucci J. D. and Woolford S. W. (1984). A Threshold AR(1) Models. *Journal of Applied Probability*. **21**, No. 2, 270-286.
50. Qian L. (1998). On maximum likelihood estimators for a threshold autoregression. *Journal of Statistical Planning and Inference*. **75**, 21-46.
51. Ruppert, D. and Carroll, R. J. (1980). Trimmed Least Square Estimation in the Linear Model. *Journal of the American Statistical Association*. **75**, 828-838.
52. Sáfadi, T. and Morettin, P. A. (2000). Bayesian Analysis of Threshold Autoregressive Moving Average Models. *Sankhya: The Indian Journal of Statistics*. **62**, Series B, Pt. 3, 353-371.
53. Tong, H. and Lim, K. S. (1980). Threshold autoregressive, limit cycles and cyclical data. *Journal Royal Statistic Society Serie B*. **42**, 245-292.
54. Tong, H. (1983). *Threshold Models in Non-Linear Time-Series Analysis*. Lecture Notes in Statistics. **21**, Springer Verlag, New York.

55. Tong, H. (1990). *Non-linear Time Series, A Dynamical System Approach*. Oxford Science Publications.
56. Tsay, R. (1997). Unit root tests with Threshold Innovations. Preprint, University of Chicago.
57. Weiss, A. (1991). Estimating Nonlinear Dynamic Models Using Least Absolute Error Estimation. *Econometric Theory*. **7**, 46-68.
58. White, H. (1996). *Estimation, Inference and Specification Analysis*. Cambridge, U.K.. Cambridge University Press. Econometric monographs; 22.