



Facultad de Ciencias
Instituto de Estadística
Ingeniería en Estadística

Modelación de la incertidumbre de la economía chilena basada en procesos autorregresivos de umbrales

Diego Chávez Guerrero
14 de diciembre de 2022

Profesor Guía

Javier Contreras. Ph. D.

Instituto de Estadística, Universidad de Valparaíso

Profesor Co-Guía

Byron Idrovo. M. Sc.

Gerencia de Estudios, Cámara Chilena de la Construcción

Proyecto de titulación para optar al:

grado académico de: *Licenciado en Estadística*

título profesional de: *Ingeniero en Estadística*

minor en: *Estadística Financiera*

Resumen

En el siguiente trabajo, se emplea el uso de diversos modelos (ARMA, TAR y ARMA-TGARCH) para los índices de percepción de la economía chilena (ICE e IPECO), con el fin de encontrar un modelo óptimo para estos índices y demostrar que el uso de un modelo ARMA no es suficientemente eficiente para estos datos, ya que se producen problemas de invertibilidad. El modelado por medio de un TAR, si bien cumple con las condiciones del análisis de diagnóstico del mismo, tiene la complejidad de recurrir a un gran número de parámetros para resumir la información de los índices. Es por ello que trabajar con un modelo ARMA-TGARCH para estos índices de expectativas logra resumir la información contenida en ellos de mejor manera. Finalmente, se pone a prueba este modelo por medio de análisis de diagnóstico residual y de sus predicciones, las cuales posteriormente se comparan con los datos reales a través de un análisis de validación cruzada.

Algunas Palabras

Agradezco a toda mi familia. Una mención a mis padres y mi hermana, gracias por todos los consejos brindados, por creer que a pesar de toda situación, podría lograr lo que me propusiera y aquí estoy, en los últimos momentos de esta etapa que marcará un antes y un después en mi vida.

Agradezco a mi pareja, Lis. Me escuchaste, me aconsejaste, me guiaste, me acompañaste en mis altos y bajos, lo importante de todo, es que siempre puedo confiar en ti.

Agradezco a 2 personas que fueron un gran apoyo en el comienzo de mi carrera, hablo de Doña Rina y Don Galo, que confiaron en que podría lograrlo desde el minuto 1 de conocernos.

Agradezco a 2 compañeros, que más que compañeros, se volvieron amigos. Nicolás y Matías, grandes momentos vividos, fueron un apoyo fundamental.

Agradezco a todos los profesores que a lo largo de estos 5 años me hayan ayudado en este gran camino. Una mención especial a mi profesor guía, el Dr. Javier Contreras, fue de enorme ayuda en todo momento.

Agradezco a mis amigos por todos los buenos momentos vividos, por las risas entregadas, por los consejos. Saben que pueden contar conmigo en todo momento.

Agradezco a mis abuelos. Una dedicatoria especial a mi abuela, Graciela Zepeda, de las mujeres más importantes en mi vida. Sé que siempre estás cuidando de mí y lucharé por cumplir las cosas que te prometí.

Índice general

Resumen	2
Algunas Palabras	3
1. Introducción	6
2. Estado del Arte	8
3. Objetivos	10
3.1. Objetivo general	10
3.2. Objetivos específicos	10
3.3. Hipótesis	10
4. Metodología	11
4.1. Índices de percepción económica	11
4.1.1. Índice de confianza empresarial	11
4.1.2. Índice de percepción del consumidor	13
4.2. Modelos ARIMA	15
4.3. Modelos TAR	17
4.3.1. Identificación del modelo	18
4.3.2. Estimación por mínimos cuadrados	20
4.4. Modelos TGARCH	20
4.4.1. Estimación por máxima verosimilitud	21
4.5. Criterio de selección de modelos	23
4.6. Librerías	23
4.6.1. <code>forecast</code>	23
4.6.2. <code>aTSA</code>	23
4.6.3. <code>NTS</code>	24
4.6.4. <code>rugarch</code>	24
5. Resultados	25
5.1. Modelo ARIMA	25
5.1.1. ICE	25
5.1.2. IPECO	26
5.2. Modelo TAR	26
5.2.1. ICE	27
5.2.2. IPECO	28
5.3. Modelo ARMA-TGARCH	29

5.3.1. ICE	29
5.3.2. IPECO	32
6. Conclusiones	40
A. Criterio de Información de Akaike para modelos TAR	41
B. Código	42
Referencias	48

Capítulo 1

Introducción

La economía ha formado parte de la vida de las personas desde sus inicios (López, 2021). Ésta se puede definir como una ciencia la cual se dedica a la administración de recursos, para así satisfacer las necesidades de una sociedad. Debido a que es necesario obtener más información, en ocasiones se trabaja bajo la elaboración de un modelo el cual busca representar un proceso estocástico proveniente de una serie financiera.

Generalmente, en el área financiera, se trabaja con diversos índices, los cuales pueden estar relacionados, por ejemplo, a valores que toman las acciones de una empresa, o a un grupo de empresas las cuales se incluyan en conjunto en algún índice como lo es el SP500. Los ejemplos anteriores radican netamente en índices relacionados con las empresas, pero también existen otro tipo que pueden estar ligados más a un ámbito más general, como lo puede ser información a nivel de países (macroeconómicos) sobre los mismos o un conjunto de ellos. Es aquí que se encuentran aquellos índices como la tasa de desempleo, inflación, entre otros. Estos tienen cierta influencia en los demás, como lo son aquellos relacionados netamente con las empresas. Además, pueden sufrir cambios a partir de posibles comentarios, anuncios, leyes o decretos que pueden generar una ruptura respecto a la serie en su comportamiento, haciendo que este tipo de series sean más volátiles que otras pertenecientes a otra área (meteorología, oceanografía, astronomía).

Lo que podemos observar en el sector económico, es que existen 2 agentes claros, que sin ellos no podría existir la economía que se observa hoy en día. Uno sería el empresariado o las empresas, y por otra parte se encuentra la otra arista que sería con respecto a los consumidores, ya que la unión de estos genera el comercio y el cual se traduce como al intercambio de bienes por algún servicio o producto. En esta tesis, se analizarán estos 2 sectores, el que respecta al empresariado, bajo el cual se considera al Índice de Confianza Empresarial (ICE), y obtenido por medio del Centro de Estudios en Economía y Negocios de la Universidad del Desarrollo (CEEN-UDD). Este permite conocer la mirada de los empresarios y poder estimar posibles escenarios económicos que permitirían hablar de posibles inversiones, o por el contrario, tener una señal de alerta temprana ante un posible decrecimiento económico. En cuanto al otro sector, se considera al Índice de Percepción del Consumidor (IPECO), el cual al igual que el ICE, fue obtenido por el CEEN-UDD. Éste índice tiene como objetivo conocer la percepción de los consumidores con respecto al estado de la economía, las cuales involucran la situación económica actual, desempleo actual, situación económica futura, desempleo futuro, e ingresos futuros. Considerando ambos, nos daría una información general respecto a la mirada de la economía que tienen los 2 sectores.

Es por ello que se pretende elaborar un modelo que pueda condensar de una forma correcta la información contenida en estos índices. Esto por medio del uso de diversos modelos con el fin de comparar los resultados obtenidos por cada uno de estos. Además, realizar predicciones con los modelos que fueron seleccionados y realizar una validación cruzada con el objetivo de conocer si el funcionamiento de estos es el correcto.

Cabe señalar que la aplicación de esta metodología no se ha hecho para analizar el comportamiento de indicadores de expectativas. Por lo general, se han aplicado a series con información pasada (como el PIB y otros indicadores de actividad cuyos registros se publican con rezago).

Capítulo 2

Estado del Arte

[Moreno y Nieto \(2014\)](#) evalúan el desempeño de un modelo autorregresivo de umbrales (TAR), en donde utiliza datos del mercado accionario brasileiro y norteamericano. Además, realiza una comparación con otro tipo de modelo del tipo lineal, como lo es el modelo GARCH. Por otra parte, [Uribe Bravo \(2015\)](#) se enfoca en los modelos $TAR(2, r, r)$, buscando por medio de ejercicios de simulación demostrar que los resultados obtenidos en el pronóstico de un modelo TAR obtenidos mediante el método de verosimilitud predictiva son mejores en cuanto a desempeño que los realizados bajo un modelo AR. [Hansen \(2011\)](#) lleva a cabo una revisión de la literatura en torno al impacto que ha tenido los modelos TAR, STAR y SETAR.

En cuanto a aplicaciones económicas, [Pérez y Velásquez \(2004\)](#) llevan a cabo el estudio sobre el comportamiento dinámico del crecimiento trimestral del PIB español desde 1970 a 1998, basándose en la metodología propuesta por [Tsay \(1998\)](#), en relación de 4 pasos para lograr la construcción de un modelo TAR. Este trabajo concluyó que las predicciones realizadas con modelos TAR, son más precisas que aquellas por medio de modelos lineales, destacando esto a un horizonte de mediano plazo. [Cao y Tsay \(1992\)](#) llevan a cabo un estudio respecto a la medición de volatilidad de rendimientos de acciones, particularmente analiza los rendimientos bursátiles mensuales ponderados (SP500) durante el periodo de enero de 1928 a diciembre de 1989. Se lleva a cabo una comparación de los resultados obtenidos mediante un modelo TAR, comparándolos con otros modelos, como lo son el ARMA, GARCH y EGARCH. Concluyen que las volatilidades de los rendimientos mensuales de las acciones son no lineales, también se habla de una ganancia en la precisión del pronostico utilizando un modelo TAR versus un ARMA.

[Djeddour y Boularouk \(2013\)](#) estudian las exportaciones de petróleo de EE.UU. durante enero de 1991 hasta diciembre de 2004. Estos buscan realizar un modelo TAR para el conjunto de datos analizados, el cual es comparado por otro modelo del tipo lineal como lo es un ARIMA, destacando que en un inicio ya se presentaban inclinaciones respecto a la utilización de un tipo de modelo no lineal frente a uno lineal. Primero comienzan con lo que sería el test de linealidad, en el cual se ocupa como hipótesis nula el hecho de que el conjunto de datos sigue una distribución $TAR(1)$, lo que es equivalente a decir que se tiene un modelo $AR(p)$, por lo que se basan en [Tsay \(1998\)](#) para la resolución de esto. [Tsay \(1998\)](#) considera la autorregresión ordenada y las predicciones para construir pruebas estadísticas residuales que no involucran los parámetros indefinidos, para luego obtener un estadístico de Fisher (F) para el test de linealidad. Si se rechaza dicha hipótesis, se continúa con el proceso y es aquí donde se ocupa este valor como la variable umbral que maximiza el estadístico F . En cuanto a los resultados, [Djeddour y Boularouk \(2013\)](#) obtienen un modelo con

3 regímenes versus un modelo $ARIMA(0, 1, 1)$, en donde por medio del uso de la raíz del error cuadrático (RMSE, por sus siglas en inglés), el modelo TAR obtenido es mejor que el ARIMA planteado en términos de RMSE, concluyendo así que el modelamiento del tipo no lineal es mejor en este caso frente a uno lineal. Por último, se entrega un pronóstico de los próximos 12 meses utilizando el modelo TAR obtenido.

El modelo TAR viene a ser aquel precursor de los modelos de umbrales autorregresivos, en que se tienen algunos derivados del mismo, como lo es el modelo SETAR, es así como [Hoyos, Ramos, y Viyas \(2010\)](#) proponen una aplicación de este modelo sobre datos recogidos del PIB de Colombia. Los resultados obtenidos en este trabajo no son del todo favorables para lo que esperaban, puesto que al realizar la construcción de un modelo SETAR de dos regímenes sobre la tasa de crecimiento del PIB colombiano durante los periodos de 1982 y 2008 y utilizando la metodología de Tsay y Tong, no encontraron evidencia contundente que demuestre algún tipo de superioridad de este modelo a pesar de que el desempeño obtenido por el mismo es superior con respecto a un modelo del tipo lineal.

En el trabajo realizado por [Gibson y Nur \(2011\)](#), se comparan distintos modelos del tipo de umbrales autorregresivos, como lo son el STAR y SETAR. En el cual se tiene un estudio del tipo simulado y otro respecto al índice NIKKEI-225, que corresponde al índice bursátil más popular del mercado japonés. El periodo utilizado consta desde enero del 2000 a septiembre del 2010, cuyos gráficos apuntaban a ser un proceso no lineal. Se lleva a cabo los procedimientos previos para llegar a lo que sería la identificación y estimación del modelo, en donde se plantea un primer modelo $AR(4)$, pero finalmente se escoge lo que sería un modelo SETAR versus el STAR. Finalmente, se llega a la conclusión de que existe un peligro en llevar a cabo la producción de un modelo ARIMA para esos datos, debido principalmente a la no linealidad, que es algo que no se debe omitir en este tipo de análisis de modelos. Por otra parte, se destaca la construcción de modelos SETAR para este tipo de datos, declarando que si se tiene un conjunto de datos similar, éstos podrían beneficiarse de la aplicación de este modelo.

El precursor de los modelos autorregresivos generalizado que captura las agrupaciones de volatilidad de las series de tiempo a través de la varianza condicional (GARCH, por su sigla en inglés) fue [Bollerslev \(1986\)](#). Sin embargo, [Zakoian \(1994\)](#) extendió el modelo GARCH a los GARCH con umbrales (TGARCH, por su sigla en inglés). El modelo TGARCH involucra una componente de umbral definido por regímenes, y un segundo componente definido por la varianza modelada condicionalmente en el tiempo. Se ha utilizado comúnmente en varias aplicaciones; por ejemplo, [Yang y Chang \(2008\)](#) consideró a los modelos TGARCH para el estudio de las interacciones entre la información de los mercados bursátiles y de divisas para encontrar reacciones asimétricas de los rendimientos bursátiles y su variabilidad asociada. [Wu \(2011\)](#) usó los modelos TGARCH para los índices de volatilidad como una variable de umbral; específicamente, empleó 20 acciones de la Major Market Index de China, y encontró que el modelo de umbral con un activador exógeno se ajusta bien a los datos. Posteriormente, [Korap \(2011\)](#) modeló la inflación para el estudio de la incertidumbre de la economía turca.

Capítulo 3

Objetivos

3.1. Objetivo general

Modelar las series de tiempo de las percepciones económicas (empresariales y del consumidor), utilizando el modelo ARMA-TGARCH.

3.2. Objetivos específicos

1. Implementar los modelos ARMA-TGARCH en el software R.
2. Aplicar los datos de los índices de ICE e IPECO al modelo TGARCH.
3. Determinar el mejor modelo ARMA-TGARCH por medio del uso de criterio de información, con sus respectivas estimaciones y análisis de diagnóstico.
4. Comparar los resultados obtenidos de los distintos modelos trabajados (ARMA, TAR y ARMA-TGARCH).
5. Realizar predicciones con el modelo ARMA-TGARCH.

3.3. Hipótesis

Un modelo ARMA-TGARCH permite resumir de una mejor forma la información sobre la volatilidad de las percepciones económicas (empresariales y del consumidor), logrando de esta manera, obtener un modelo más adecuado que utilizando un modelo TAR o ARIMA.

Capítulo 4

Metodología

4.1. Índices de percepción económica

A continuación, se describen los dos índices de percepción de la economía chilena a utilizar en esta tesis.

4.1.1. Índice de confianza empresarial

El Índice de Confianza Empresarial (ICE) mide el estado de la economía, desde el punto de vista empresarial, a través de su valoración del estado de sus empresas, el sector económico y la economía en su conjunto. El índice cubre seis sectores económicos del país: industria, comercio, finanzas, minería, construcción y agricultura. Entre julio de 2003 y enero de 2005, el ICE fue realizado cada tres meses y fue a partir de enero de 2005 es que comenzó a elaborarse mensualmente. Con respecto a la muestra considerada, corresponde a aproximadamente 300 gerentes generales, dueños de empresas o ejecutivos. La información fue recolectada por medio de encuestas telefónicas y correos electrónicos. Se destacar que la muestra trabaja bajo una estructura de panel, lo que quiere decir que los encuestados se repiten con el pasar del tiempo.

El cuestionario está orientado a la empresa del encuestado, de la industria a la que pertenece, a las empresas en general y a la economía nacional. Además, considera la coyuntura económica y las expectativas que tienen los empresarios sobre el futuro. De esta manera, se busca clasificar las respuestas como optimistas, neutrales y pesimistas, para así conocer la proporción de empresarios optimistas y pesimistas sobre estos temas.

Hasta fines del año 2018, las preguntas eran específicas a cada sector. A partir de enero de 2019, se dio un cambio respecto a la visión general de la encuesta, para así realizar preguntas que pudieran ser contestadas por todos los sectores. Además, se eliminaron las preguntas que hacían alusión a responder sobre las percepciones a sólo un año. En la Figura 4.1, se pueden ver los cambios que recibió este instrumento.

Las posibles respuestas de los encuestados son calificadas como optimistas, neutrales y pesimistas, en donde:

- Para respuestas optimistas se tienen : Aumentará / Mejorará/ Mejor que lo esperado.

Tiempo	Pregunta	Sector que no contesta la pregunta	
		Hasta el 2018	2019
Presente	Capacidad instalada	Agro., Const.	
Presente	Situación de oferta y demanda nacional	Agro.	
Presente	Situación de oferta y demanda internacional	Agro., Const.	Const., Finan.
Presente	Nivel de inventarios	Agro.	Agro., Finan.
Presente	Nivel de Producción / Ventas últimos 3 meses	Const.	
Presente	Nivel de precios últimos 3 meses	Com., Ind., Min. y Const.	
Presente	Situación general del negocio últimos 3 meses	Com., Ind., Min. y Finan.	
Presente	Situación general del negocio últimos 3 meses v/s lo esperado (1)	TODOS	
Futuro	Proyección situación económica global en 3 meses más		
Futuro	Y en un año más (2)		
Futuro	Situación general del sector económico en 3 meses más	Agro., Min., Com., Ind. y Finan.	
Futuro	Proyección situación general del negocio en 3 meses más		
Futuro	Proyección producción/ventas en 3 meses más		
Futuro	Proyección precios de venta en 3 meses más		
Futuro	Proyección precios de insumos en 3 meses más	Finan.	
Futuro	Proyección Número de Trabajadores en 3 meses más		
Futuro	Y en un año más (2)		
Futuro	Proyección de salarios en 3 meses más	Agro., Min., Com., Ind. y Finan.	
Futuro	Evolución inversión en maquinaria, equipos e infraestructura en 3 meses más	Const. y Agro.	
Futuro	Y en un año más (2)		
Futuro	Situación financiera en los próximos 3 meses	Com., Ind., Min. y Finan.	
Futuro	Y en un año más (2)		

(1) Pregunta nueva.

(2) Preguntas eliminadas.

Figura 4.1: Cambios en el instrumento. Fuente: CEEN (2020).

- Para respuestas neutrales se tienen: No cambiará / Se mantendrá/ Igual que lo esperado.
- Para respuestas pesimistas se tienen: Disminuirá / Empeorará / Peor que lo esperado.

En lo que se asocia con la elaboración de los índices, el procedimiento para el cálculo de éste se realiza de la siguiente manera. Se genera un subíndice X_i , el cual corresponde a la suma del número de respuestas optimistas, neutrales y pesimistas, para cada pregunta, y según el sector económico y tamaño de la empresa (pequeña, mediana o grande), para obtener:

$$X_i = \frac{OPT_i - PES_i}{TOT_i},$$

donde

- OPT: N° de respuestas optimistas,
- PES: N° de respuestas pesimistas,
- TOT: Total de respuestas.

Posteriormente, estos se combinan mediante un promedio ponderado acorde al tamaño de la empresa, logrando así un nuevo conjunto de subíndices por pregunta y sector económico dado por:

$$Z_i = \sum_{j=1}^3 w_j X_i$$

Tamaño	Ponderador (2005-2018)	Ponderador (Desde 2019)
Grande	0.750	0.540
Mediana	0.125	0.175
Pequeña	0.125	0.285

Cuadro 4.1: Actualización de ponderadores según el tamaño de la empresa. Fuente: [CEEN \(2020\)](#).

Sector Económico	Ponderadores (2005-2018)	Ponderadores (2019-2020)
Agrícola	0.074	0.064
Comercio	0.180	0.201
Construcción	0.131	0.142
Financiero	0.209	0.112
Industria	0.272	0.240
Minería	0.134	0.241

Cuadro 4.2: Actualización de ponderadores según sector económico. Fuente: [CEEN \(2020\)](#).

Es a partir del año 2019 que ocurre un cambio en el cálculo de este ponderador asociado al tamaño de la empresa, ya que ahora es calculado mediante la inclusión del empleo. Estos ponderadores w_i están dado en el Cuadro 4.1, en conjunto con el antiguo ponderador que se tenía.

El índice sectorial (S_i), es obtenido por medio de un promedio simple de los subíndices por pregunta y corresponde a cada sector económico. Además del tamaño de la empresa, también se debe considerar lo referente al sector económico, es por ello que también se realiza un ponderador sobre los índices sectoriales. A partir del año 2005, los sectores considerados son: minería, industria, construcción, comercio, agrícola y financiero. Este ponderador es construido mediante el porcentaje de participación respecto del PIB total que tiene cada sector, para posteriormente obtener lo que sería el ICE:

$$ICE = \sum_{i=1}^6 W_i S_i.$$

El ICE ha sufrido cambios en lo que respecta a su cálculo. Se ha actualizado mediante el recálculo de la participación sobre el PIB que tiene cada sector. En el Cuadro 4.2 se muestran las comparaciones. Con respecto a la interpretación de los valores del ICE, se define la neutralidad con respecto al valor 0. La forma en cómo se define el optimismo y pesimismo es que se crea una escala cualitativa en base a los valores que podría tomar el ICE, lo cual queda representado en el cuadro 4.3.

A continuación se presenta la gráfica 4.2 que resume los datos de este Índice de Confianza Empresarial.

4.1.2. Índice de percepción del consumidor

El Índice de Percepción del Consumidor (IPECO) tiene por objetivo medir la percepción de los consumidores por medio de las evaluaciones sobre la coyuntura económica y de sus expectativas. En cuanto a la muestra, esta consiste en aproximadamente 380 encuestados, localizados en centros

Categoría	Intervalo
Extraordinariamente optimista	≥ 45
Muy optimista	$[35, 45)$
Optimista	$[25, 35)$
Moderadamente optimista	$[15, 25)$
Levemente optimista	$[5, 15)$
Neutral	$[-5, 5)$
Levemente pesimista	$[-25, -15)$
Moderadamente pesimista	$[-35, -25)$
Muy pesimista	$[-45, -35)$
Extraordinariamente pesimista	< -45

Cuadro 4.3: Escala cualitativa. Fuente: [CEEN \(2020\)](#).

comerciales de varias ciudades del país. La forma en la se responde esta encuesta es de forma presencial, es decir, existe un encuestador encargado de ejecutar las preguntas y recoger las respuestas. El cuestionario mide los siguientes 5 conceptos:

- Situación económica actual: En comparación con el año anterior, usted diría que su situación económica actual es: peor / igual / mejor.
- Desempleo actual: En relación al año anterior, hoy el desempleo en el país es: mayor / igual / menor.
- Situación económica futura: Usted diría que en un año más su situación económica será: peor / igual / mejor.
- Desempleo futuro: Usted diría que en un año más el desempleo en el país será: mayor / igual / menor.
- Ingresos futuros: El próximo año cree usted que el ingreso total de su familia será: menor / igual / mayor.

Además, se incluyen en el instrumento preguntas de caracterización socioeconómica. Al igual que en el caso del ICE, el cálculo de este índice se basa en las percepciones de los consumidores (optimistas, neutrales y pesimistas), así que se calcula el porcentaje de ello por cada pregunta del instrumento. El resultado se construye en base a dividir el porcentaje de consumidores optimistas por la suma de los consumidores optimistas y pesimistas.

Finalmente, se promedian los cinco índices de percepciones para obtener el IPECO. Además, el índice coyuntural está relacionado con la situación económica actual, es por eso que su cálculo es a través del promedio de las percepciones de la situación económica y el desempleo actual. Por otra parte, también se considera el futuro en que es relacionado con el índice de expectativas que se calcula como el promedio de las percepciones sobre la economía futura, desempleo futuro e ingresos futuros. Por último, cada índice se divide por su valor inicial y se multiplica por 100.

A continuación se presenta la gráfica 4.3 que resume los datos de este Índice de Confianza Empresarial.

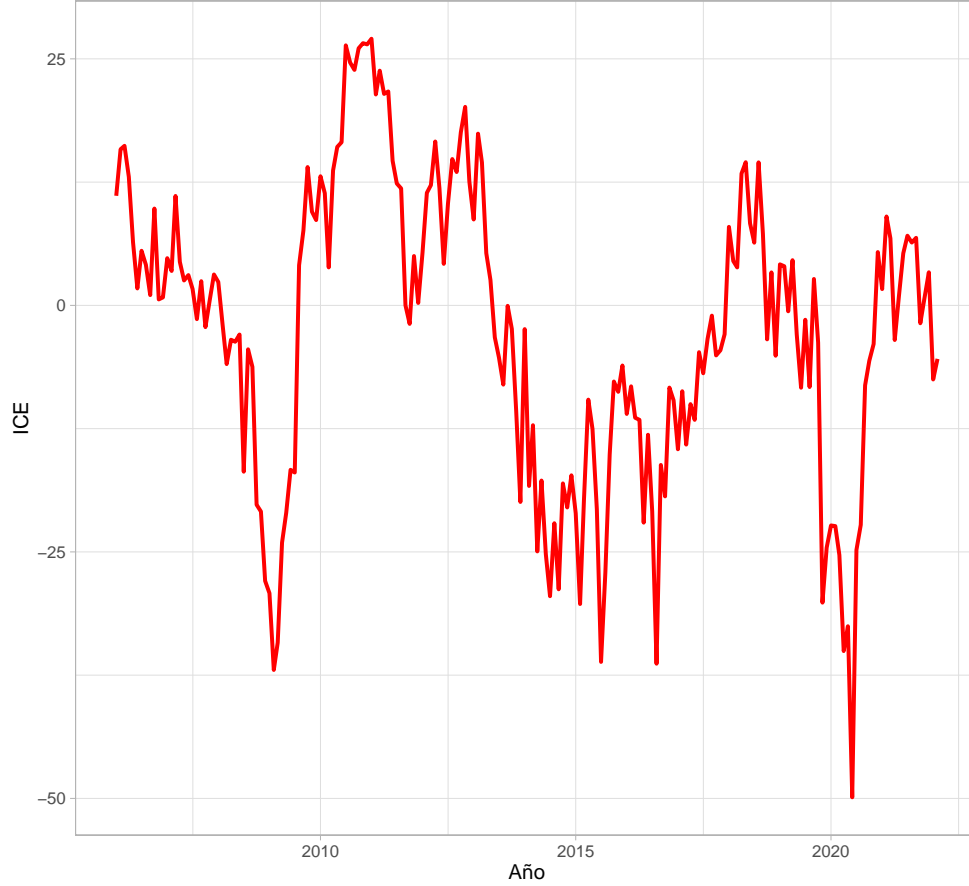


Figura 4.2: Índice de Confianza Empresarial, 2006 - 2022.
Fuente: Elaboración propia (2022).

4.2. Modelos ARIMA

Para empezar a definir lo que se conoce como proceso autorregresivo integrado de media móvil (ARIMA), se debe comenzar hablando un poco de su antecesor, es decir el proceso autorregresivo de medias móviles (ARMA). Para que un proceso y_t siga el orden $ARMA(p, q)$, este debe estar en función de su pasado hasta el retardo p , de la innovación contemporánea y el pasado de la innovación hasta el retardo q , de la siguiente manera:

$$y_t = \phi_1 + \dots + \phi_p y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q}, \quad \epsilon_t \sim RB(0, \sigma^2). \quad (4.1)$$

La expresión anterior, al definirla en términos del operador de retardos se obtiene que

$$(1 - \phi_1 L - \dots - \phi_p L^p) y_t = (1 + \theta_1 L + \dots + \theta_q L^q) \epsilon_t$$

y en forma polinomial como

$$\Phi_p(L) y_t = \Theta_q(L) \epsilon_t, \quad (4.2)$$

donde $\phi_p(L)$ es el polinomio autoregresivo y $\theta_q(L)$ es el polinomio de medias móviles.

Los modelos ARMA se basan en el supuesto de estacionariedad en la varianza/covarianza, esto quiere decir que la media y varianza son constantes y finitas, y las autocovarianzas no dependen

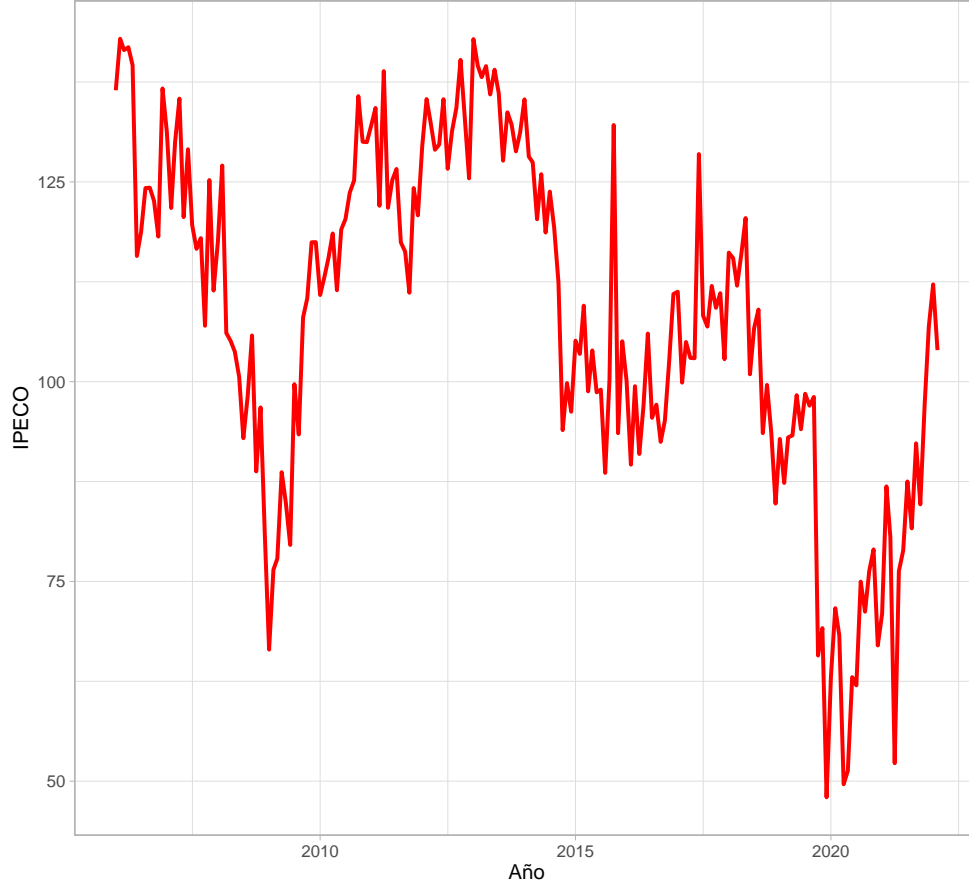


Figura 4.3: Índice de Percepción del Consumidor, 2006 - 2022.
Fuente: Elaboración propia (2022).

del tiempo sino que radican en el número de periodos de separación entre las variables. Particularmente, las series económicas no suelen cumplir con los supuestos de estacionariedad, debido a que no cumplan con las condiciones de invertibilidad y causalidad. Es por ello que surge el modelo Autorregresivo Integrado de Medias Móviles de orden (p, d, q) , denotado por $ARIMA(p, d, q)$.

El polinomio AR del modelo (4.2) puede ser factorizado en función de sus p raíces, las cuales quedan denotadas por L_1, L_2, \dots, L_p , de la forma

$$\Phi_p(L) = (1 - L_1^{-1}L)(1 - L_2^{-1}L) \dots (1 - L_p^{-1}L).$$

Ahora se supone que las $p - 1$ raíces son estacionarias, es decir, con módulo fuera del círculo unitario, y una de ellas es unitaria, $L_i = 1$, Entonces, los polinomios se pueden reescribir de la forma:

$$\Phi_p(L) = (1 - L_1^{-1}L)(1 - L_2^{-1}L) \dots (1 - L_p^{-1}L) = \varphi_{p-1}(L)(1 - 1^{-1}L)$$

y

$$\Theta_p(L) = \varphi_{p-1}(L)(1 - L),$$

donde el polinomio $\varphi_{p-1}(L)$ resulta del producto de los $p - 1$ polinomios de orden 1 asociados a las raíces L_i con módulo fuera del círculo unitario.

De esta forma, sustituyendo el modelo ARMA (4.2) se tiene:

$$\varphi_{p-1}(L)(1-L)y_t = \Theta_q(L)\epsilon_t \rightarrow \varphi_{p-1}(L)\Delta y_t = \Theta_q(L)\epsilon_t \quad (4.3)$$

donde el polinomio $\varphi_{p-1}(L)$ es estacionario debido a que sus raíces sin excepción tienen módulo fuera del círculo unitario y el polinomio $\Delta = (1-L)$ es el que contiene la raíz unitaria.

Se puede dar que el polinomio AR del modelo (4.2) presente más de una raíz unitaria, por lo que si se generaliza a tener d raíces unitarias, el modelo se puede escribir como:

$$\Theta_p(L) = \varphi_{p-d}(L)(1-L)^d,$$

lo que también se puede observar de la siguiente manera:

$$\varphi_{p-d}(L)\Delta^d y_t = \Theta_q(L)\epsilon_t,$$

donde el polinomio $\varphi_{p-d}(L)$ es estacionario dado que sus $p-d$ raíces tiene módulo fuera del círculo unitario, y el polinomio $\Delta^d = (1-L)^d$ es aquel que contiene las d raíces unitarias no estacionarias. De esta manera y_t corresponde a un proceso integrado de orden d y se denota por $y_t \sim I(d)$.

De esta manera, se puede expresar un modelo autoregresivo integrado de medias móviles de orden (p, d, q) de la siguiente manera:

$$\Phi_p(L)\Delta^d y_t = \delta + \Theta_q(L)\epsilon_t \quad (4.4)$$

donde el polinomio autoregresivo estacionario es $\Phi_p(L)$ y el polinomio de medias móviles $\Theta_q(L)$ no tienen raíces comunes.

Una extensión de los modelos ARIMA contempla una componente estacional tanto en la parte AR como MA. Esta extensión son los llamados modelos SARIMA, que consideran un conjunto de parámetros rezagados relacionados con el proceso y sus ciclos estacionales (Hyndman y Khandakar, 2008). Decimos que $\{y_t\}$ sigue un modelo SARIMA si está compuesto por su parte ARIMA más la inclusión de las componentes estacionales como sigue:

$$\Phi_s(L^s)\Delta_s^D \Phi_p(L)\Delta^d y_t = \delta + \Theta_s(L^s)\Theta_q(L)\epsilon_t, \quad (4.5)$$

donde el polinomio de P parámetros autoregresivos de la parte estacional es $\Phi_s(L^s)$ y el polinomio de Q parámetros de media móvil de la parte estacional es $\Theta_s(L^s)$. Denotamos este modelo como SARIMA(p, d, q) \times (P, D, Q) $_s$.

4.3. Modelos TAR

Tong (1978) propuso este tipo de modelos de umbrales autorregresivos, los cuales se pueden definir de la siguiente manera. Sea y_t un proceso que sigue un modelo TAR($r; p_1, \dots, p_r$), el cual posee una variable umbral Z_{t-d} y r regímenes, definido de la siguiente forma:

$$y_t = \alpha_{0,k} + \sum_{i=1}^{p_k} \alpha_{i,k} y_{t-i} + \epsilon_{k,t}, \quad \gamma_{k-1} \leq Z_{t-d} < \gamma_k, \quad (4.6)$$

en donde se tiene que $k = 1, 2, \dots, r$, d es un entero positivo, y p_1, \dots, p_r corresponden a los órdenes autoregresivos de y_t en cada régimen, los cuales son números enteros no negativos. Por

otra parte, $\epsilon_{k,t}$ son las innovaciones o errores que pertenecen a una sucesión de variables aleatorias independientes y que siguen una distribución normal con media cero y varianza σ_k^2 .

En relación a los umbrales Z_t del proceso, se tienen los valores ordenados $\gamma_0 < \gamma_1 < \dots < \gamma_d$, en donde se definen los d regímenes para el modelo. Debido a esto es que se puede referir al k -ésimo régimen del modelo con la partición de la forma $\gamma_{k-1} \leq Z_{t-d} < \gamma_k$. Si se tiene que la variable umbral Z_{t-d} está dada por rezagos o funciones de la misma serie y_t , se pasaría a tener lo que se conoce como el modelo SETAR (Self Exciting Threshold Autoregressive).

Los procesos TAR cumplen con las propiedades de estacionariedad y ergodicidad (Petrucci y Woolford, 1984), particularmente para el caso de series de tiempo que siguen un comportamiento no lineal. Estas condiciones son fundamentales para las propiedades asintóticas de los estimadores de cada uno de los parámetros del modelo.

4.3.1. Identificación del modelo

Lo que se busca en esta sección, es describir el procedimiento encargado de la identificación del modelo TAR, por lo que lo primero que se debe realizar es conocer si los datos pueden ser resumidos a través de este modelo. Luego, la colección de datos debe cumplir con seguir un comportamiento no lineal, es por ello que se debe realizar un test de no linealidad. El test de Tsay (1998) contrasta las siguientes hipótesis:

$$\begin{aligned} H_0 : & \quad y_t \sim TAR(1), \\ H_1 : & \quad y_t \sim TAR(r), \quad r > 1. \end{aligned}$$

El contraste antes mencionado se basa en un proceso sencillo para la construcción de modelos TAR, lo que permite conocer si los datos se adecuan a un modelo no lineal TAR. La hipótesis alternativa H_1 se explicaría gracias a la presencia de umbrales, mientras que el estadístico es construido por medio de una regresión ordenada, la cual es estimada por el método de mínimos cuadrados, el cual considera los residuos predictivos para el cálculo del estadístico de prueba.

Para comenzar con el proceso de identificación, lo primero es seleccionar la estructura autoregresiva de los datos, es decir, mediante la función de autocorrelación parcial (PACF) se busca el orden p , el cual es ocupado para definir el intervalo de los posibles valores para d , con $0 < d \leq p$. Luego, se debe comprobar la no linealidad, así que se toma el estadístico de prueba propuesto por (Tsay, 1998). Este estadístico es obtenido a través de los siguientes pasos: primero se ordenan de menor a mayor los datos de la variable umbral Z_t por magnitud. Sea $t(i)$ el índice de tiempo correspondiente a la i -ésima observación más pequeña de Z_t , se tiene la siguiente regresión ordenada:

$$y_{t(i)} = X_{t(i)}^\top \beta + \epsilon_{t(i)}, \quad i = 1, \dots, T - p, \quad (4.7)$$

donde $X_{t(i)}^\top = (1, y_{t(i)-1}, \dots, y_{t(i)-k})$ es la matriz de diseño, $t(i)$ es el índice de tiempo de la variable umbral Z_t ordenada, y T es el tamaño de la muestra.

Al usar la regresión ordenada, se está transformando de forma efectiva un modelo de umbrales en un problema de punto de cambio, lo que se traduce en que se tienen los datos agrupados de

forma en que en un grupo, todas las observaciones siguen el mismo modelo lineal AR, ya que esta regresión se presenta de forma ordenada y creciente de la variable umbral de Z_{t-d} .

Posteriormente, la regresión anterior es estimada por medio del método de mínimos cuadrados ordinarios para las m primeras observaciones. Tsay (1998) propone que $m \in (3\sqrt{T}, 5\sqrt{T})$ para obtener el vector de parámetros estimados $\hat{\beta}_m$. Luego, se calcula los residuales predictivos dados por:

$$\hat{e}_{t(m+1)} = y_{t(m+1)} - X_{t(m+1)}^\top \hat{\beta}. \quad (4.8)$$

Por otra parte, los residuos estandarizados se calculan como

$$\hat{\eta}_{t(m+1)} = \hat{e}_{t(m+1)} [1 + X_{t(m+1)}^\top V_m X_{t(m+1)}]^{-1/2}, \quad (4.9)$$

donde

$$V_m = \left[\sum_{i=1}^m X_{t(i)} X_{t(i)}^\top \right]^{-1}, \quad (4.10)$$

$y_{t(m+1)}$ y $X_{t(m+1)}^\top$ son respectivamente la próxima observación de la variable respuesta y de las regresoras, las cuales son incluidas en la regresión ordenada. Luego, se estima la regresión de la forma:

$$\hat{\eta}_{t(l)} = X_{t(l)}^\top \omega + \epsilon_{t(l)}, \quad l = m_0 + 1, \dots, T - p, \quad (4.11)$$

donde m_0 corresponde al punto inicial de la estimación de mínimos cuadrados recurrentes.

A continuación, se considera la prueba de hipótesis con $H_0 : \omega = 0$ y $H_1 : \omega \neq 0$ como hipótesis nula y alternativa, respectivamente. El estadístico de prueba C es:

$$C = (T - 2p - m - 1) \log |S_0| + \log |S_1|, \quad (4.12)$$

donde

$$S_0 = \frac{1}{T - p - m} \sum_{t=m+1}^{T-p} \hat{\eta}_{t(l)}^2, \quad (4.13)$$

y

$$S_1 = \frac{1}{T - n - m} \sum_{t=m+1}^{T-p} \hat{e}_{t(l)}^2. \quad (4.14)$$

Bajo la hipótesis nula de linealidad y algunas condiciones de regularidad, el estadístico construido C es asintóticamente una variable aleatoria χ^2 con $k + 1$ grados de libertad.

Si los resultados apuntan a rechazar la hipótesis nula, es decir, la serie sigue un comportamiento no lineal, es que se pasa a identificar parámetros estructurales tales como el número de regímenes del modelo (r), los ordenes autoregresivos de cada partición (p_1, \dots, p_r) en donde para cada una de estas se define un proceso $AR(p_i)$, además de los valores umbrales $\gamma_1, \dots, \gamma_{r-1}$ los cuales logran determinar cada uno de los regímenes y el parámetro de retardo d de la variable umbral Z_{t-d} . La localización de los umbrales, así como la identificación de los ordenes autoregresivos de las particiones es obtenido mediante el método planteado por (Tsay, 1998).

4.3.2. Estimación por mínimos cuadrados

A continuación, se presenta el método de mínimos cuadrados para realizar el cálculo de los parámetros estructurales de rezago d y los valores de umbrales γ_i . Sin pérdida de generalidad, se expone el funcionamiento de lo propuesto por [Hansen \(2011\)](#) para un modelo TAR(2, p_1, p_2). La estimación, tanto de los coeficientes como del umbral, es realizada por medio de mínimos cuadrados. Para lo primero se usan los mínimos cuadrados ordinarios y para lo segundo sólo los mínimos cuadrados. De esta manera se busca aquel valor que minimice la suma de errores al cuadrado. Lo propuesto por [Hansen \(2011\)](#) también permite probar la existencia de uno o varios umbrales, así como también determinar la distribución asintótica de los coeficientes. Luego, bajo el siguiente modelo:

$$y_t = \begin{cases} X_t^\top \Phi_1 + \sigma_1 \epsilon_t, & \text{si } Z_{t-d} \leq \gamma \\ X_t^\top \Phi_2 + \sigma_2 \epsilon_t, & \text{si } Z_{t-d} > \gamma, \end{cases} \quad (4.15)$$

donde $X_t^\top = (1, y_{t-1}, \dots, y_{t-p})$, $\Phi_1 = (\phi_{10}, \phi_{11}, \dots, \phi_{1p_1})$, $\Phi_2 = (\phi_{20}, \phi_{21}, \dots, \phi_{2p_2})$, y Z_{t-d} se asume estacionaria.

Luego, los estimadores de mínimos cuadrados son:

$$\hat{\Phi}_i(\gamma, d) = \left(\sum_{t=1}^{n_j} X_t X_t^\top \right)^{-1} \left(\sum_{t=1}^{n_j} X_t y_t^\top \right), \quad (4.16)$$

y

$$\hat{\sigma}_i^2(\gamma, d) = \frac{1}{n_i - p_i} \sum_{t=1}^{n_j} (y_t - X_t^\top \hat{\Phi}_i(\gamma, d))(y_t - X_t^\top \hat{\Phi}_i(\gamma, d))^\top. \quad (4.17)$$

La suma de los cuadrados residuales queda expresada por

$$S(\gamma, d) = (n_1 - p_1) \hat{\sigma}_1^2(\gamma, d) + (n_2 - p_2) \hat{\sigma}_2^2(\gamma, d). \quad (4.18)$$

Luego, los estimadores de mínimos cuadrados condicionales de γ y d son obtenidos por

$$(\hat{\gamma}, \hat{d}) = \arg \min_{\gamma, d} S(\gamma, d), \quad (4.19)$$

donde $1 \leq d \leq p$ y $\gamma \in \mathbb{R}_0$.

Luego de la identificación de los umbrales y los ordenes autorregresivos, se pasa a ajustar el modelo TAR por medio de la estimación de parámetros de cada partición o régimen, gracias al uso de mínimos cuadrados. En cuanto a las propiedades asintóticas de estos estimadores, mediante el Teorema de consistencia de los estimadores realizada por ([González-Farías, Gonzalo, y Russell-Noriega, 2007](#)), se demuestra que para un γ conocido son estimadores consistentes y asintóticamente normales. Por otro lado, si γ es desconocido, bajo ciertas condiciones de regularidad también se logra demostrar que los estimadores son consistentes.

4.4. Modelos TGARCH

El modelo de umbrales heterocedástico TGARCH fue propuesto por [Zakoian \(1994\)](#), el cual menciona que una forma natural para introducir asimetrías en la volatilidad es mediante una función de las partes negativas y positivas del proceso. El autor se basa en un artículo de [Davidian y Carroll \(1987\)](#) sobre la estimación de la función de varianza.

De esta manera el proceso de umbral heterocedástico se define como:

$$y_t = \sigma_t \epsilon_t, \quad (4.20)$$

$$\sigma_t = \alpha_0 + \sum_{i=1}^q \alpha_i^+ y_{t-i}^+ - \alpha_i^- y_{t-i}^- + \sum_{j=1}^p \beta_j \sigma_{t-j}, \quad (4.21)$$

donde (α_i^+) , (α_i^-) y (β_j) son secuencias de escalares reales, $i = 1, \dots, q$, $j = 1, \dots, p$. Además, los errores ϵ_t son i.i.d. con media 0 y varianza 1, y ϵ_t es independiente de y_{t-1} para todo t .

El enfoque de este modelo se encuentra estrechamente relacionado con lo desarrollado por [Tong \(1978\)](#) sobre la modelización de la media condicional. En lo que respecta a las ventajas de este enfoque, se tiene que al modelar la variable de escala σ_t en lugar de la varianza condicional, se da del hecho de que no serían necesarias restricciones en cuanto a la positividad en la definición de las distintas variables: la varianza condicional de σ_t es no negativa por construcción, generando una simplificación respecto a los procedimientos de inferencia numérica. Sin embargo, en lo que refiere a propiedades probabilísticas, el estudio de complica si se considera a σ_t como no positiva, dado esto es [Zakoian \(1994\)](#) propone incluir al modelo planteado ciertas restricciones de positividad:

$$\alpha_0 > 0, \quad \alpha_i^+ \geq 0, \quad \alpha_i^- \geq 0, \quad \beta_i \geq 0, \quad \forall i. \quad (4.22)$$

Consideremos en particular, el modelo TGARCH(1, 1), definido por

$$\sigma_t = \alpha_0 + \alpha_1^+ y_{t-1}^+ - \alpha_1^- y_{t-1}^- + \beta \sigma_{t-1}, \quad \alpha_0 > 0, \quad \alpha_1^+ \geq 0, \quad \alpha_1^- \geq 0. \quad (4.23)$$

El modelo TGARCH más sencillo (pero útil para muchas aplicaciones) establece reescribir [\(4.23\)](#) para tener una ecuación $AR(1)$ para σ_t con un componente aleatoria que depende de Z :

$$\sigma_t = \alpha_0 + B(Z_{t-1}) \sigma_{t-1}, \quad (4.24)$$

con $B(Z_{t-1}) = \alpha_1^+ Z_{t-1}^+ - \alpha_1^- Z_{t-1}^- \beta$. Esta relación se utiliza para investigar problemas relacionados con la estacionariedad estricta y de segundo orden. Para la obtención de la condicionariedad estricta, se supone la existencia de $E[\log B(Z_t)]$. Siguiendo el enfoque entregado por [Nelson \(1990\)](#), se demuestra la existencia de una solución estacionaria estricta para el modelo TGARCH dependiente del signo de esta cantidad.

Existe una reparametrización del modelo [4.23](#) dada en la ecuacion (26) de [Ghalanos \(2022\)](#) :

$$\sigma_t = \alpha_0 + \sum_{j=1}^q \alpha_j \sigma_{t-j} |U_{t-j}| + \sum_{j=1}^p \beta_j \sigma_{t-j}. \quad (4.25)$$

Se considerará esta última versión del modelo TGARCH(p,q) para la estimación y presentación de los resultados en el capítulo 5.

4.4.1. Estimación por máxima verosimilitud

Sea $\omega^\top = \{\alpha_0, \alpha_1^+, \dots, \alpha_q^+, \alpha_1^-, \dots, \alpha_q^-, \beta_1, \dots, \beta_p\}$ y $\theta = [\varphi, \omega^\top] \in \Theta$ un subespacio compacto de \mathbb{R}^{2q+p+1} . Asumiendo innovaciones normales, la función de quasi-log-verosimilitud para una muestra de T observaciones, condicionada sobre la primera q , es (quitando términos constantes)

$$\log l(Y; \theta) = - \sum_{t=q+1}^T \log \sigma_t - \frac{1}{2} \sum_{t=q+1}^T \left(\frac{y_t}{\sigma_t} \right)^2. \quad (4.26)$$

En comparación con el caso del modelo GARCH, es importante notar que la función es continua en θ , diferenciable con respecto a ω pero no siempre con respecto a φ debido a la presencia de umbrales. El estimador máximo verosímil $\hat{\theta}_T$ del parámetro θ es entonces un caso particular del estimador-M basado en una función continua diferenciable por la derecha. Se sabe que bajo ciertas condiciones de regularidad este estimador es consistente y asintóticamente normal:

$$\sqrt{T}(\hat{\theta}_T - \theta_T) \rightarrow^d N(0, J^{-1}), \quad (4.27)$$

con

$$J = E_0 \left[\frac{\partial^+ \log l(Y; \theta_0)}{\partial \theta} \frac{\partial^+ \log l(Y; \theta_0)}{\partial \theta^t} \right] = \left[\frac{\partial}{\partial \theta} E_0 \left[\frac{\partial^+ \log l(Y; \theta_0)}{\partial \theta} \right] \right]_{\theta=\theta_0},$$

donde $\partial^+/\partial\theta$ denota la derivación por la derecha, componente por componente, y $l(Y; \theta)$ la función de densidad de y_t condicionada a y_{t-1} . La diferenciación con respecto a los parámetros relacionados con la varianza da como resultado:

$$\frac{\partial \log l(Y; \theta_T)}{\partial \omega} = \sum_{t=q+1}^T \frac{1}{\sigma_t^3} \frac{\partial \sigma_t}{\partial \omega} [y_t^2 - \sigma_t^2]. \quad (4.28)$$

Otro posible modelo es el llamado ARMA-TGARCH, el cual modela $\{y_t\}$ como un modelo ARMA (4.1), y sus residuos $\{\epsilon_t\}$ son modelados mediante un modelo TGARCH (4.20)-(4.25). En resumen, el modelo queda como:

$$y_t = \sum_{i=1}^p \phi_i Y_{t-i} + \sum_{j=1}^q \theta_j \epsilon_{t-j} + \epsilon_t, \quad (4.29)$$

$$\epsilon_t = \sigma_t \nu_t, \quad \nu_t \sim N(0, 1), \quad (4.30)$$

$$\sigma_t = \alpha_0 + \sum_{j=1}^{q_g} \alpha_j \sigma_{t-j} |U_{t-j}| + \sum_{j=1}^{p_g} \beta_j \sigma_{t-j}, \quad (4.31)$$

el cual se denota por ARMA(p, q)-TGARCH(p_g, q_g).

En esta tesis, consideraremos los modelos ARMA-TGARCH con innovaciones basadas en distribuciones normales. Sin embargo, cuando tenemos observaciones con alta volatilidad, es posible que los modelos heterocedásticos tengan innovaciones con distribución de colas pesadas y asimetría producto de las observaciones atípicas asociadas a la alta volatilidad. Por lo tanto, consideraremos también la distribución Student- t asimétrica (Azzalini y Capitanio, 2003), denotada como $ST(0, 1, \lambda, m)$ y dada por:

$$t_\lambda(x; m) = 2t(x; m)T \left(\lambda x \sqrt{\frac{m+1}{m+x^2}}; m+1 \right),$$

con media 0 y varianza 1, donde

$$t(x; m) = \frac{\Gamma(\frac{m+1}{2})}{\sqrt{m\pi}\Gamma(\frac{m}{2})} \left(1 + \frac{x^2}{m} \right)^{-(m+1)/2},$$

donde m ($m > 2$) son los grados de libertad, y $T(\cdot; m)$ es la función de distribución acumulada de la densidad $t(\cdot; m)$. Valores inferiores de m , digamos $m < 10$, proporciona una densidad de colas

pesadas; mientras que valores altos, tales como $m \geq 10$, proveen colas livianas. Cuando $m \rightarrow \infty$, se obtiene la distribución skew-normal (Azzalini y Capitanio, 2003).

Sin pérdida de generalidad, la función de quasi-log-verosimilitud dada en (4.26) está basada en innovaciones normales, por lo que claramente cambiará dependiendo de la distribución que se escoga para las innovaciones, como es el caso de usar la ST.

4.5. Criterio de selección de modelos

El criterio utilizado para la selección de modelos corresponde al Criterio de Información de Akaike (AIC), este se puede definir de la siguiente forma. Sea $\{M_k\}$, $k = 1, 2, \dots, K$, un conjunto de modelos competidores indexados por k . Entonces el criterio:

$$AIC(M_k) = -2 \log L(\hat{\theta}_k) + 2r, \quad (4.32)$$

donde r es el número de parámetros del modelo M_k y $\hat{\theta}_k$ es su conjunto de parámetros. Se considera al mejor modelo al que tiene menor AIC.

4.6. Librerías

En lo que respecta a recursos utilizados para la obtención de los modelos a utilizar, se basaron en librerías pertenecientes al software R, las cuales son las siguientes.

4.6.1. forecast

La librería `forecast` (Hyndman y Khandakar, 2008) provee funciones de predicción para series temporales y modelos lineales, además de métodos y herramientas para la visualización y el análisis de predicciones de series temporales univariantes, incluido el alisamiento exponencial mediante modelos de espacio de estado y modelización automática (S)ARIMA.

La función a utilizar en esta tesis es `auto.arima`. Esta es utilizada para determinar los modelos ARIMA y SARIMA para ajustar los índices trabajados, ya que por medio de criterios de información como lo pueden ser AIC, AICc o BIC, devuelve el mejor modelo ARIMA o SARIMA.

4.6.2. aTSA

La librería `aTSA` realiza un análisis alternativo de series temporales. Esta contiene algunas herramientas para probar, analizar datos de series temporales y modelos de series temporales populares como ARIMA, AR, entre otros.

La función a utilizar es `adf.test` (Dickey y Fuller, 1979), la cual realiza la prueba de Dickey–Fuller aumentada para la hipótesis nula de una raíz unitaria de una serie temporal univariada.

4.6.3. NTS

La librería `NTS` es un paquete de R para el análisis de series temporales no lineales.

En cuanto a las funciones a utilizar corresponden a 3. La primera es `thr.test`, la cual es la prueba de no-linealidad del umbral de una serie escalar (Tsay, 1998). Como segunda función de este paquete se tiene a `uTAR` el cual estima modelos univariados TAR de dos regímenes incluyendo el umbral. Por último se encuentra `uTAR.est`, que estima modelos TAR con múltiples regímenes con umbrales conocidos.

4.6.4. rugarch

La librería `rugarch` (Ghalanos, 2022) permite trabajar con los modelos GARCH univariados. De este paquete se utilizan dos funciones, la primera es `ugarchspec` este corresponde a un método para crear un objetivo de especificación de un GARCH univariado antes del ajuste. Por otra parte, la otra función a utilizar es la encargada de realizar el ajuste en virtud de lo propuesto en la primera función, es decir, se define como un método para ajustar una variedad de modelos GARCH univariado.

Capítulo 5

Resultados

Como se mencionó antes, este trabajo aborda la búsqueda de un modelo el cual se base en procesos autoregresivos de umbrales para la modelación de los índices de percepción económica de Chile. Estos corresponden al Índice de Confianza Empresarial (ICE) y al Índice de Percepción del Consumidor (IPECO). La naturaleza de los datos obtenidos de ambos índices corresponden al del tipo mensual, los cuales consideran los periodos de enero de 2006 hasta febrero de 2022.

Lo primero que se realizó fue un modelado por medio del tipo de modelos lineal más conocido en el área de la series de tiempo, el modelo ARIMA. Lo obtenido se pueden observar en la siguiente sección.

5.1. Modelo ARIMA

Los resultados obtenidos fueron llevados a cabo por medio de la utilización de la función `auto.arima` el cual pertenece a la librería `forecast` incluida en el software R, el cual trabaja bajo criterios de información.

5.1.1. ICE

Para el caso de los datos de ICE, lo primero que se realizó fue la Prueba Aumentada de Dickey–Fuller, esto con el fin de conocer si la serie en cuestión es estacionaria o no, por medio de testear la existencia de raíces unitarias. El valor p obtenido fue de 0.1156, lo que se traduce en el rechazo de la hipótesis nula con un 95% de nivel de confianza, determinando de esta manera que el Índice de Percepción del Consumidor no es estacionario.

En cuanto al modelo planteado por la función, entregó un $SARIMA(1, 0, 1) \times (0, 0, 2)_{12}$. En los modelos SARIMA se incluyen factores estacionales autoregresivos, diferenciación estacional y media móvil estacional, en este caso el modelo indica el orden de media móvil estacional igual a 2. Los parámetros estimados de este modelo SARIMA se pueden apreciar en la tabla 5.1. Se puede apreciar que el parámetro autoregresivo estimado es cercano a 1, lo que se traduce en que se pueden estar presentando problemas de invertibilidad para el polinomio AR.

Parámetro	Coeficiente	s.e.
ϕ_1	0.9255	0.0296
θ_1	-0.1888	0.0749
θ_1^s	0.0966	0.0736
θ_2^s	-0.2022	0.0770

Cuadro 5.1: Parámetros estimados modelo SARIMA para ICE. Fuente: Elaboración propia (2022).

5.1.2. IPECO

En lo que respecta al Índice de Percepción del Consumidor, la manera de abordar el modelado de este fue de igual manera por medio de la función `auto.arima` perteneciente al software R.

En lo que respecta a la Prueba Aumentada de Dickey–Fuller, el valor p obtenido fue de 0.514, por lo que al igual que en el caso del ICE, se rechaza la hipótesis nula y de esta manera se determinó que IPECO no es estacionaria.

Lo entregado por la función `auto.arima` indica que por medio del criterio de información, el modelo óptimo para estos datos corresponden a un ARIMA(1, 0, 1), es decir, no hay componentes estacionales. En cuanto a los parámetros estimados por este se pueden observar en la tabla 5.2. Cabe señalar de que a pesar de que el test indicó que la serie no es estacionaria, no se incluyó diferenciaciones en la serie para lograr esta estacionariedad. Aún así, al igual que en el caso anterior, el parámetro autoregresivo es inclusive más cercano a 1, por lo que se está nuevamente en presencia de los problemas de invertibilidad en el polinomio AR. De esta manera, este tipo de modelos ARIMA no son lo más apropiado para estos datos en particular.

Parámetro	Coeficiente	s.e.
ϕ_1	0.9651	0.0193
θ_1	-0.4003	0.0669

Cuadro 5.2: Parámetros estimados modelo ARIMA para IPECO. Fuente : Elaboración propia (2022).

5.2. Modelo TAR

Dado que los resultados entregados por los modelos ARIMA no fueron los más acertado para estos datos, producto de los problemas presentados en su elaboración tales como la invertibilidad del polinomio autoregresivo, es que se pensó en un primer proceso autoregresivo de umbrales para abarcar esta problemática. De esta manera, en esta sección de los resultados se resume lo obtenido por medio de la generación de modelos del tipo TAR. De igual manera que en el caso de los modelos ARIMA, se verán estas conclusiones primero para los datos de ICE y posteriormente se pasará al IPECO.

Antes de todo, se debe realizar un test de no-linealidad para ver si el modelo se puede adecuar de buena forma a este tipo de modelos. La librería NTS contiene una función llamada `thr.test`, la cual por medio de una prueba de hipótesis permite conocer si los datos pueden ser modelados por

medio de un modelo TAR (ver la sección 4.3.1).

La prueba de hipótesis para los datos de ICE a través de la función `thr.test`, dieron como resultado un valor p igual a 0.000 y el estadístico igual a 7.786899, mientras que para IPECO el valor p fue de un 0.002 y el estadístico igual a 6.22295, por lo que en ambos casos se concluye en el rechazo de la hipótesis nula con un 95 % de nivel de confianza. Por lo tanto, decimos que tanto ICE como IPECO podrían ser representados por un modelo TAR(r), con $r > 1$.

5.2.1. ICE

El primer paso consistió en definir el número de particiones apropiado para este modelo, de esta manera de concluyó un número de 3 particiones (inspección visual de la serie, ver Figura 4.2).

Ya con la cantidad definida de segmentos, se pasó al segundo paso: la elaboración del modelo. Para ello se utilizaron los criterios de información, particularmente el criterio de AIC. Posteriormente se realiza un proceso iterativo en el software R, el cual consiste en que por cada paso se va generando un modelo del tipo TAR con un orden autoregresivo distinto para cada partición, y luego guardar su respectivo AIC. En las pruebas iniciales de esto, se definió un máximo de 5 ordenes autoregresivos para todos los segmentos; sin embargo, los primeros resultados arrojaban que en lo que respecta a las 2 primeras particiones, estas no generaban una ganancia al sobrepasar los 2 ordenes autorregresivos (es decir, en vez de mejorar al modelo, lo empeoraban), esto en base al AIC entregado. Por otra parte, la última partición, si generaba una ganancia es términos de AIC al aumentar el número de parámetros para ésta.

Finalmente, los resultados en términos de AIC indicaron que el modelo TAR más óptimo para los datos de ICE correspondía a un TAR(2, 1, 5). Los demás resultados pueden ser observados en la tabla 5.3.

p ₁	p ₂	p ₃	AIC	p ₁	p ₂	p ₃	AIC
1	1	1	736.6063	2	1	1	735.6555
1	1	2	738.6053	2	1	2	737.6545
1	1	3	736.4068	2	1	3	735.4560
1	1	4	731.8181	2	1	4	730.8673
1	1	5	729.6451	2	1	5	728.6943
1	2	1	737.2708	2	2	1	736.3200
1	2	2	739.2698	2	2	2	738.3190
1	2	3	737.0713	2	2	3	736.1205
1	2	4	732.4827	2	2	4	731.5319
1	2	5	730.3096	2	2	5	729.3588

Cuadro 5.3: Modelos TAR para ICE. Fuente: Elaboración propia (2022).

Análisis de diagnóstico del modelo TAR para ICE

Para llevar a cabo el análisis de diagnóstico del modelo seleccionado, se llevó a cabo una prueba de Ljung-Box (Ljung y Box, 1978), esto con el fin de conocer si los residuos son ruido blanco y que no se esté generando algún tipo de correlación serial entre estos, ver figura 5.1. Para todos los rezagos (lags) y considerando un nivel de significancia de 95 %, se concluye que no existe evidencia

estadísticamente suficiente para no rechazar la hipótesis nula, en otras palabras, los residuos son ruido blanco.

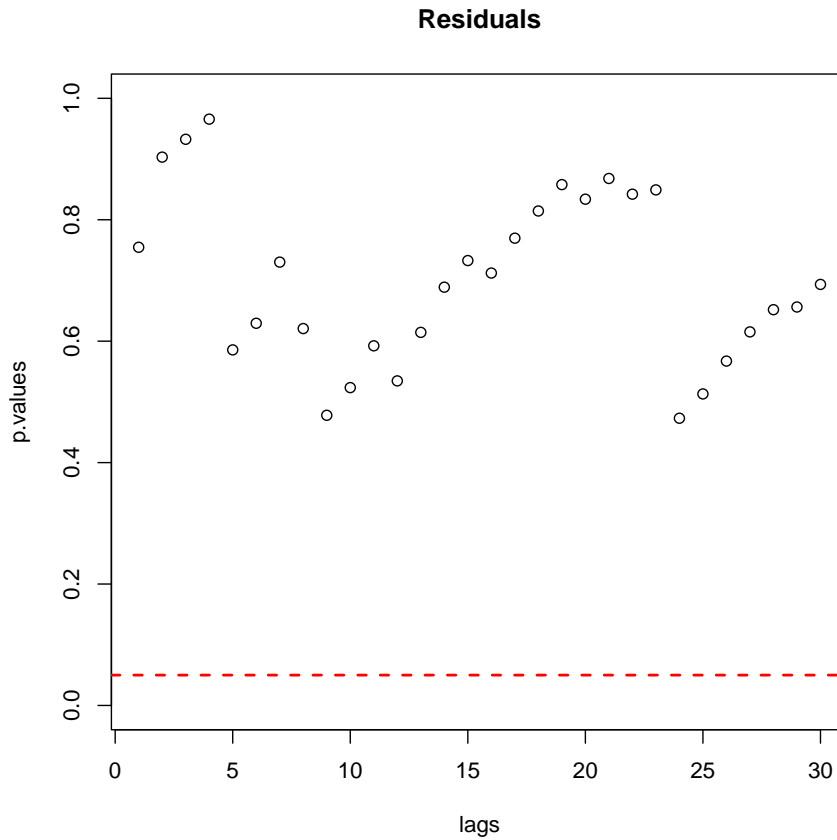


Figura 5.1: Test de Ljung–Box para el modelo TAR para ICE. Fuente : Elaboración propia (2022).

5.2.2. IPECO

La forma de abordar este índice fue similar a la vista en el caso de ICE. Sin embargo, este no consideró 3 particiones sino que se pensó en 2 (inspección visual de la serie, ver Figura 4.3). En lo que respecta a la obtención de resultados esto fue con el uso del mismo algoritmo generado para el caso anterior.

El número de ordenes autoregresivos a probar en el algoritmo fue de 5 para ambas particiones. Los resultados de las distintas combinaciones se logran apreciar en la tabla 5.4. Para el primer segmento, se ve que un aumento en el número de parámetros no ayuda a los resultados, mientras que para el segundo segmento, es decir, aquel que resume la información perteneciente de esta partición, se ve la necesidad de recurrir a más parámetros para la obtención de un mejor AIC. De esta manera el modelo TAR determinado para IPECO corresponde a un TAR(1, 5).

p1	p2	AIC
1	1	849.3376
1	2	842.3831
1	3	831.2180
1	4	828.2918
1	5	826.5194
2	1	845.1154
2	2	844.0723
2	3	832.9071
2	4	829.9810
2	5	828.1809

Cuadro 5.4: Modelos TAR para IPECO. Fuente : Elaboración propia (2022).

Análisis de diagnóstico Modelo TAR para IPECO

Al igual que lo visto anteriormente en el análisis de los residuos del modelo TAR para el otro índice, es que se realizó una prueba de Ljung–Box para conocer si los residuos eran ruido blanco, los resultados se resumen en la figura 5.2.

Se puede observar de la gráfica que para todos los rezagos (lags), se concluye que considerando un nivel de significancia del 95 %, no existe evidencia estadísticamente suficiente para no rechazar la hipótesis nula, en otras palabras, los residuos son ruido blanco.

Para resumir los resultados de este tipo de modelos utilizado a los datos, se pudo apreciar de que a pesar de que se obtuvo un modelo TAR para ambas series, estos modelos presentaban la dificultad de tener un gran número de parámetros (por lo menos 5). Así que se tomó la decisión de explorar nuevas opciones, es por ello que se pensó en un modelo del tipo heterocedástico pero que considera umbrales para su elaboración. De esta manera es que se consideran a los modelos TGARCH en la siguiente sección, para tratar de resumir de mejor forma la información de los datos.

5.3. Modelo ARMA-TGARCH

Al igual que lo visto en los modelos TAR, la elección de los modelos ARMA-TGARCH se lleva a cabo por medio de los respectivos AIC, para ello es que se generaron distintos modelos ARMA-TGARCH en los que iba variando los ordenes de la parte GARCH del modelo.

5.3.1. ICE

En lo que respecta a este índice, los resultados obtenidos se observan en la tabla 5.5. Observando los valores de AIC, se determinó un modelo ARMA(1,1)-TGARCH(1,1) como el más óptimo, ya que este en términos de AIC fue el que obtuvo el menor valor. Destacar que todos los resultados fueron en base a considerar la parte ARMA del modelo como un ARMA(1,1), ya que fue esta combinación la que generaba mejores resultados en términos de bondad de ajuste. Además, la distribución utilizada para los residuos fue la Normal.

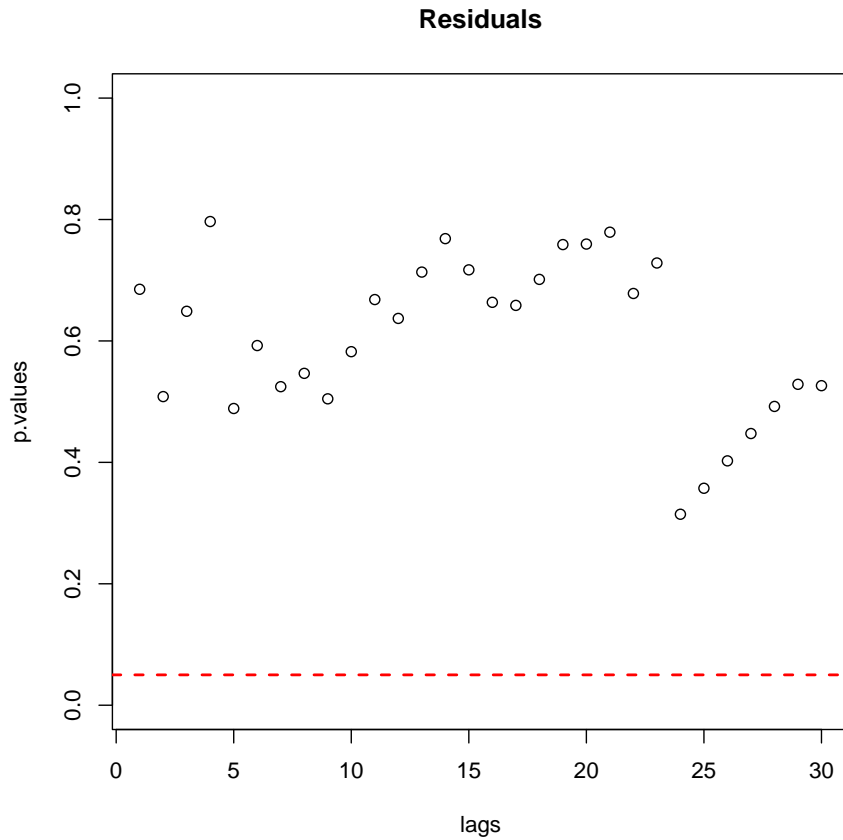


Figura 5.2: Test de Ljung–Box para el modelo TAR para IPECO. Fuente : Elaboración propia (2022).

Estimación de parámetros del modelo ARMA-TGARCH para ICE

La función `ugarchfit` permite conocer los parámetros estimados por el modelo, los cuales se pueden observar en la tabla 5.6. La tabla entrega un resumen de los parámetros estimados por medio de la función `ugarchfit`, en la cual se observa que los parámetros significativos fueron ϕ_1 , α_1 y β_1 . Mientras que θ_1 se encuentra muy cerca del límite de considerarlo como significativo.

Análisis de diagnóstico del modelo ARMA-TGARCH para ICE

Ya con el modelo seleccionado es que se lleva a cabo un análisis de diagnóstico sobre éste, con el fin de conocer si el modelo se está adecuando de una forma correcta a los datos trabajados. En la figura 5.3, se presenta un histograma en el que se ven reflejados los residuos estandarizados entregados por el modelo en cuestión. Se logra observar que los residuos son bien ajustados a una distribución normal. Se incluyen las respectivas ACF tanto de los residuos estandarizados como de los residuos cuadráticos estandarizados, las cuales se pueden ver en la figura 5.4. La gráfica de la ACF de los residuos estandarizados muestra que estos se encuentran completamente dentro de las bandas de Bartlett (en color rojo). Por otro lado, para la ACF de los residuos cuadráticos estandarizados, se ve que existe una leve alza con respecto a estas bandas.

p_g	q_g	AIC	p_g	q_g	AIC
1	1	6.6832	3	5	6.7260
1	2	6.6842	4	1	6.7270
1	3	6.6956	4	2	6.7265
1	4	6.7073	4	3	6.7323
1	5	6.7126	4	4	6.7426
2	1	6.7034	4	5	6.7466
2	2	6.7048	5	1	6.7467
2	3	6.7257	5	2	6.7475
2	4	6.7279	5	3	6.7513
2	5	6.7328	5	4	6.7616
3	1	6.7051	5	5	6.7611
3	2	6.7039			
3	3	6.7095			
3	4	6.7220			

Cuadro 5.5: Modelos ARMA-TGARCH para ICE. Fuente : Elaboración propia (2022).

	Estimación	Std. Error	t valor	p-valor
μ	5.862567	4.642946	1.2627	0.206703
ϕ_1	0.940140	0.033559	28.0142	0.000000
θ_1	-0.162081	0.084247	-1.9239	0.054370
α_0	0.497425	0.331837	1.4990	0.133872
α_1	0.080679	0.033315	2.4217	0.015449
β_1	0.856903	0.066226	12.9390	0.000000

Cuadro 5.6: Parámetros estimados para el modelo ARMA(1,1)-TGARCH(1,1) para ICE. Fuente: Elaboración propia (2022).

Luego, se realizó una prueba de Ljung–Box tanto para los residuos estandarizados como para los residuos cuadráticos estandarizados, los cuales se pueden observar en la tabla 5.7. Lo expuesto en la tabla forma parte del resumen entregado luego del ajuste del modelo a los datos. De esta manera se llega a la conclusión de que considerando un nivel de significancia del 95 %, no existe evidencia estadísticamente significativa para no rechazar la hipótesis nula, lo que dice que los residuos son ruido blanco.

Predicciones usando el modelo ARMA-TGARCH para ICE

Con el respectivo análisis ya realizado, se procedió a probar la capacidad predictiva del modelo ARMA-TGARCH. Por lo que, por medio de una validación cruzada, se comparó los resultados de las predicciones entregadas por este modelo en contraste con los datos reales del ICE. Esto se observa en la figura 5.5. La figura muestra que el modelo ARMA(1,1)-TGARCH(1,1) para el ICE entrega predicciones cercanos a las observaciones originales, ya que logra capturar de gran manera la heterocedasticidad que se estaba presentando en los datos. En cuanto al error cuadrático medio este es fue de 13.09217.

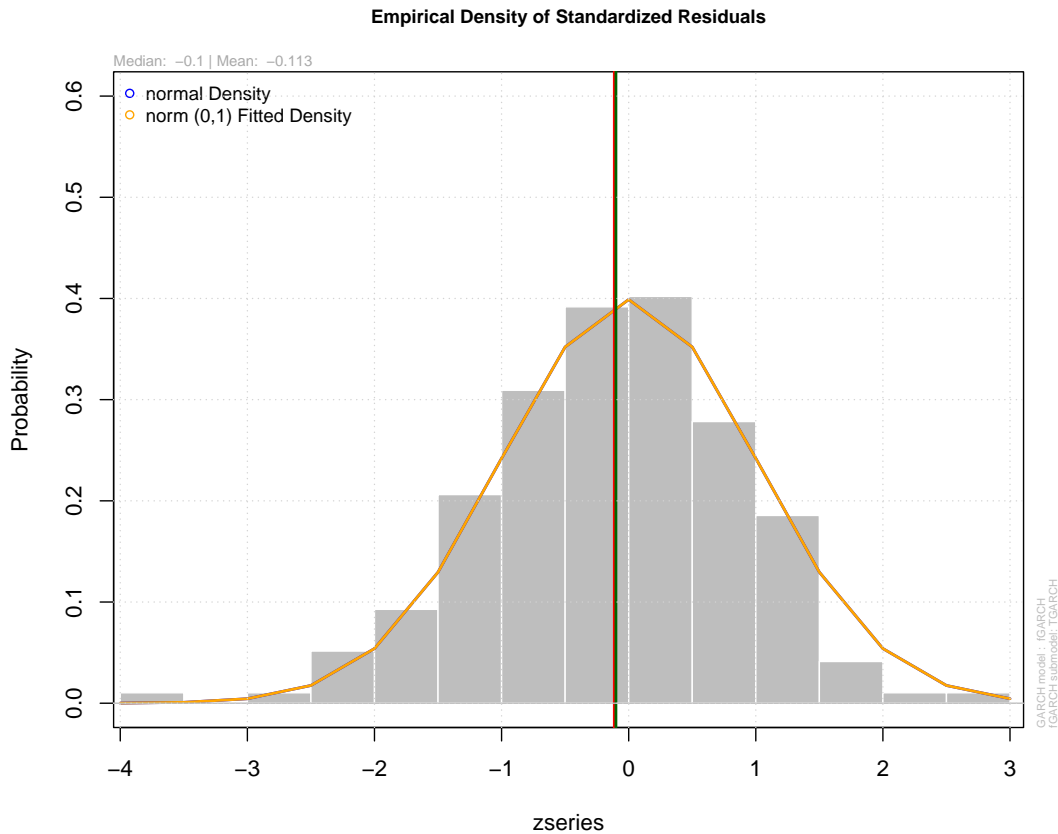


Figura 5.3: Histograma de residuos estandarizados modelo ARMA(1,1)-TGARCH(1,1) para ICE. Fuente : Elaboración propia (2022).

5.3.2. IPECO

En el caso de IPECO, se trabajó con la misma función del caso de ICE, de esta manera el resumen de esto se ve en la tabla 5.8. El modelo seleccionado para IPECO fue un ARMA(1,1)-TGARCH(2,2) y, al igual que para los otros datos, también se consideró la parte ARMA del modelo como un ARMA(1,1) ya que también presentaba los mejores resultados en términos de bondad de ajuste. Sin embargo, la función utilizada entrega la opción de trabajar con distintos tipos de distribuciones para los residuos, es por ello que se consideró incluir esta variación al modelado; en particular, la distribución skew- t .

En el caso del modelo trabajado para los datos de ICE, este no presentaba una gran ganancia en lo que respecta al criterio de selección del modelo, por ende se consideró la distribución por defecto con la que trabaja estos residuos, la distribución normal. Mientras que para IPECO si existió una ganancia en base al criterio utilizado, lo cual se puede observar en la tabla 5.9. La tabla sugiere utilizar una distribución skew- t , pues ésta logra capturar de una mejor forma estos residuos.

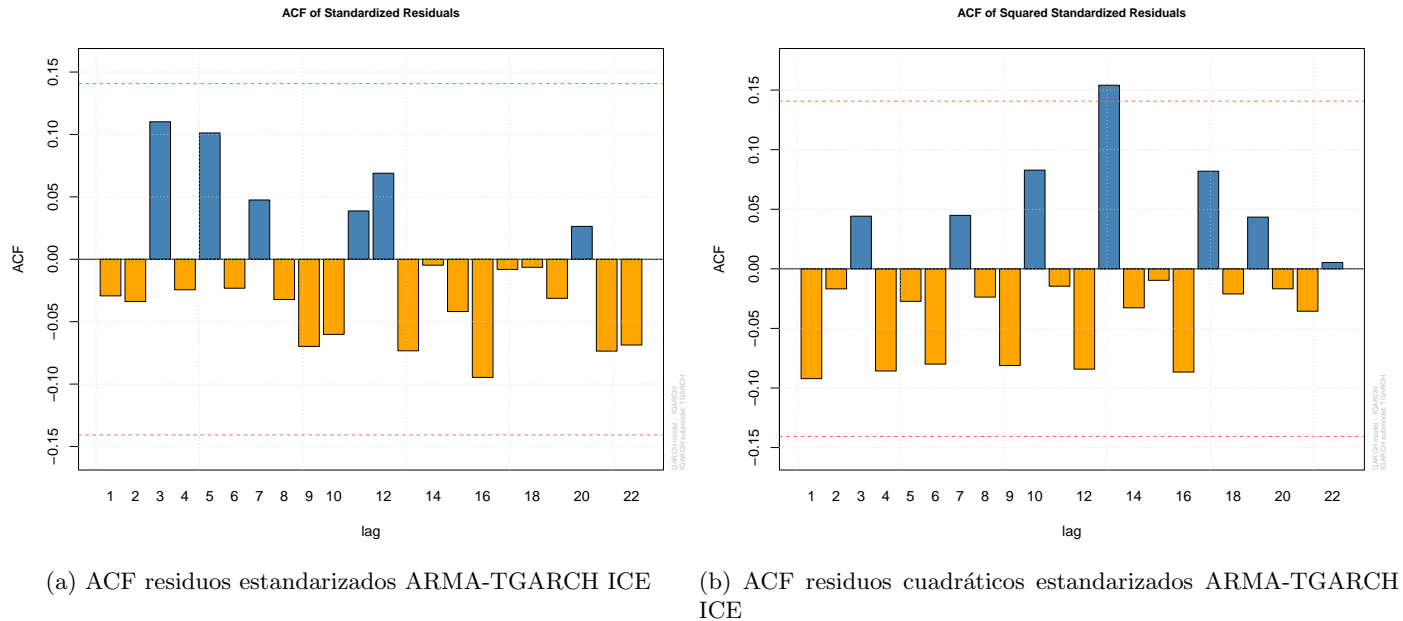


Figura 5.4: ACF de residuos estandarizados y residuos estandarizados cuadráticos ARMA(1,1)-TGARCH(1,1) para ICE. Fuente : Elaboración propia (2022).

Residuos estandarizados			Residuos estandarizados cuadráticos		
Lag	Estadístico	p-valor	Lag	Estadístico	p-valor
Lag 1	0.1696	0.6805	Lag 1	1.674	0.1957
Lag 5	2.2603	0.8870	Lag 5	2.571	0.4910
Lag 9	3.8340	0.7293	Lag 9	3.978	0.5926

Cuadro 5.7: Test de Ljung-Box para el modelo ARMA(1,1)-TGARCH(1,1) para ICE, aplicado a residuos estandarizados y residuos cuadráticos estandarizados. Fuente : Elaboración propia (2022).

Estimación de parámetros Modelo ARMA-TGARCH para IPECO

La función `ugarchfit` permite conocer los parámetros estimados por el modelo, los cuales se pueden observar en la tabla 5.10. La tabla entrega un resumen de los parámetros estimados por medio de la función `ugarchfit`, en la cual se observa que los parámetros que se presentan como significativos son μ , ϕ_1 , θ_1 , β_1 , λ , y m . Esto implica que α_0 , α_1 y α_2 no fueron significativos, por lo que el modelo sugerido es uno más simple: un ARMA(1,1)-TGARCH(2,2), es decir, la componente de umbrales no aporta al modelo. Por otro lado, λ y m fueron cercanos a 1 y 5, lo que implica que hay presencia de asimetría y colas pesadas en los residuos.

Análisis de diagnóstico del modelo ARMA-TGARCH para IPECO

Pasando a lo que es el respectivo análisis de diagnóstico del modelo ARMA(1,1)-TGARCH(2,2) propuesto para el Índice de Percepción del Consumidor, tenemos el histograma de la figura 5.6, en el que se aprecian los residuos estandarizados que entrega el modelo ajustado. El histograma muestra que los residuos estandarizados se ajustan tanto a la distribución normal como a la skew- t ,

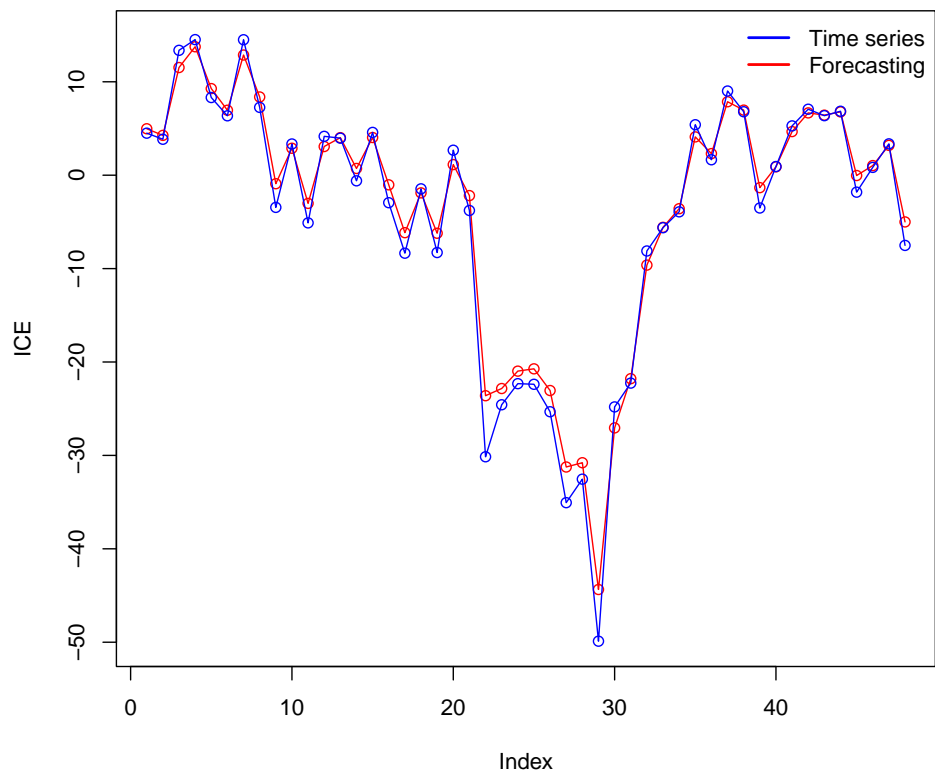


Figura 5.5: Predicciones versus datos originales del ICE. Fuente : Elaboración propia (2022).

donde esta última logra ajustar de mejor forma que la distribución normal (tiene mayores problemas para ajustar los residuos cercanos a 0).

De igual manera que en el caso del ICE, se incluyen las ACF de los residuos estandarizados y los residuos cuadráticos estandarizados, como se aprecia en la figura 5.7. La figura indica que en lo que respecta a las ACF tanto de los residuos estandarizados como los cuadráticos estandarizados, se presentan dentro de las bandas de Bartlett (en color rojo).

Además, los resultados de la prueba de Ljung-Box aplicada tanto para los residuos estandarizados como para los residuos cuadráticos estandarizados, se observan en la tabla 5.11. Lo expuesto en la tabla representa lo entregado a través de la función `ugarchfit`, de esta manera se llega a la conclusión de que considerando un nivel de significancia del 95 %, no existe evidencia estadísticamente significativa para no rechazar la hipótesis nula, lo que dice que los residuos son ruido blanco.

Predicciones con el modelo ARMA-GARCH para IPECO

De igual manera que lo visto en el caso de ICE, es que se puso a prueba el modelo ARMA(1,1)-TGARCH(2,2) ajustado al IPECO. Así que a través de una validación cruzada se compararon los datos originales de IPECO y se compararon por medio de una gráfica a los entregados por la

p_g	q_g	AIC	p_g	q_g	AIC
1	1	7.3196	3	5	7.3439
1	2	7.3305	4	1	7.3415
1	3	7.3405	4	2	7.3180
1	4	7.3501	4	3	7.3621
1	5	7.3163	4	4	7.3396
2	1	7.3018	4	5	7.3646
2	2	7.2869	5	1	7.3607
2	3	7.3249	5	2	7.3389
2	4	7.2992	5	3	7.3743
2	5	7.3103	5	4	7.3609
3	1	7.3258	5	5	7.3698
3	2	7.3046			
3	3	7.3345			
3	4	7.3190			

Cuadro 5.8: Modelos ARMA-TGARCH para IPECO y sus respectivos AIC. Fuente : Elaboración propia (2022).

Distribución	AIC
<i>Normal</i>	7.2869
<i>Skew-t</i>	7.2523

Cuadro 5.9: AIC de los modelos ARMA(1, 1)-TGARCH(2, 2) para IPECO, considerando el tipo de distribución en los residuos. Fuente : Elaboración propia (2022).

predicción realizada con el modelo propuesto, como se aprecia en la figura 5.8. La figura muestra que, a diferencia de lo visto en el caso de ICE, este tiene algunas diferencias entre el valor observado y el predicho por el modelo planteado, aunque de igual manera logra controlar de cierta forma la heterocedasticidad que se presenta. En lo que respecta al error cuadrático medio se obtiene 31.22406.

Resumiendo los resultados, se apreció que con el modelo ARMA-TGARCH planteado se logró resolver el problema visto con los modelos TAR, ya que se disminuyó la cantidad de parámetros a estimar en ambos modelos. Los modelos tuvieron un análisis de diagnóstico en los que se presentó independencia de los residuos, no existían algún tipo de correlación serial en ellos. Por último, las predicciones realizadas dejaron ver que los modelos lograron resumir la información presente en los datos de una buena forma.

	Estimación	Std. Error	t valor	p-valor
μ	139.168309	4.160668	33.448551	0.000000
ϕ_1	0.994294	0.009299	106.930029	0.000000
θ_1	-0.411233	0.059429	-6.919763	0.000000
α_0	0.004617	0.061341	0.075265	0.940004
α_1	0.000000	0.000003	0.007033	0.994388
α_2	0.030876	0.021492	1.436621	0.150826
β_1	0.931203	0.000234	3972.813955	0.000000
β_2	0.000004	0.012486	0.000343	0.999726
λ	0.888078	0.076263	11.644982	0.000000
m	4.950907	1.504686	3.290326	0.001001

Cuadro 5.10: Parámetros estimados para el modelo ARMA(1, 1)-TGARCH(2, 2) para IPECO. Fuente: Elaboración propia (2022).

Residuos estandarizados			Residuos estandarizados cuadráticos		
Lag	Estadístico	p-valor	Lag	Estadístico	p-valor
Lag 1	0.3514	0.5533	Lag 1	0.8165	0.3662
Lag 5	1.4675	0.9984	Lag 11	6.0852	0.4244
Lag 9	2.5919	0.9427	Lag 19	9.2987	0.5278

Cuadro 5.11: Test de Ljung-Box para el modelo ARMA(1, 1)-GARCH(2, 2) para IPECO, aplicado a residuos estandarizados y residuos cuadráticos estandarizados. Fuente : Elaboración propia (2022).

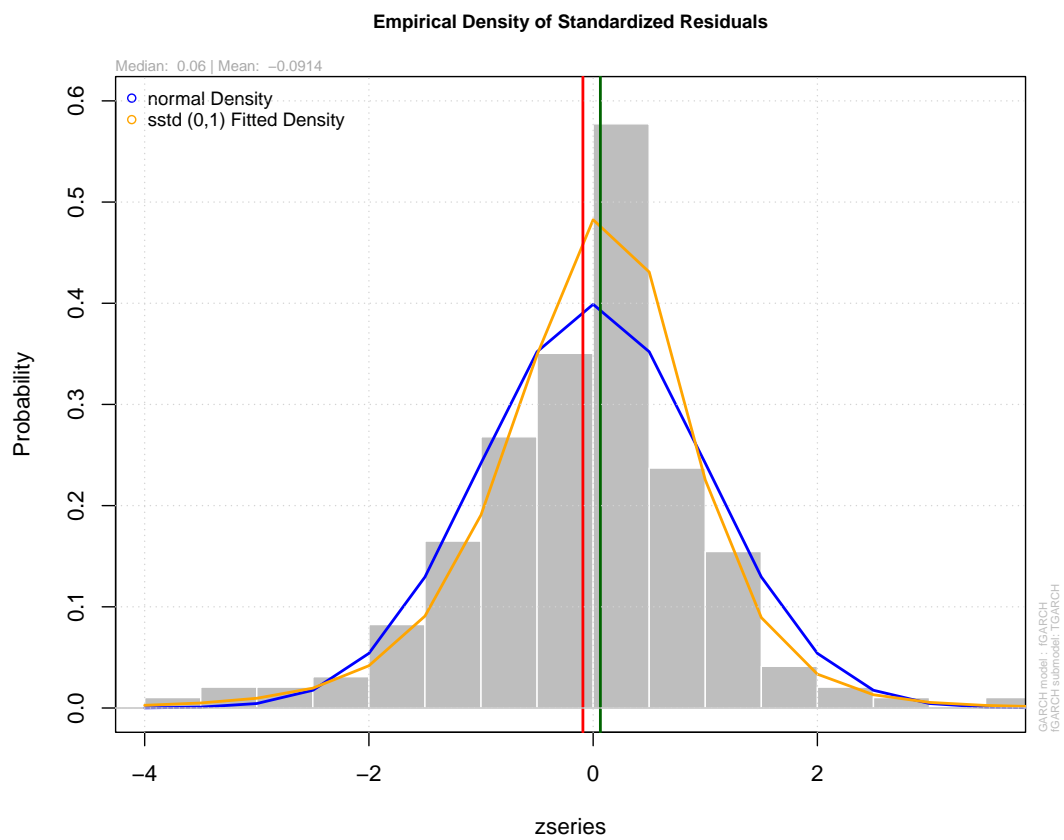
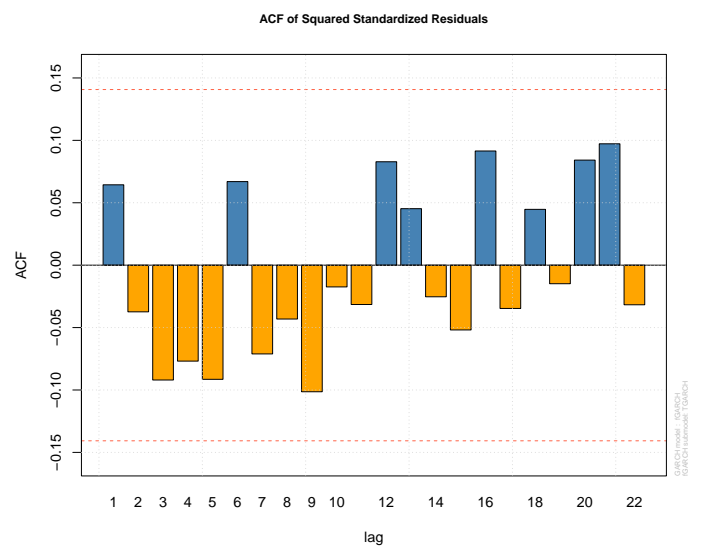
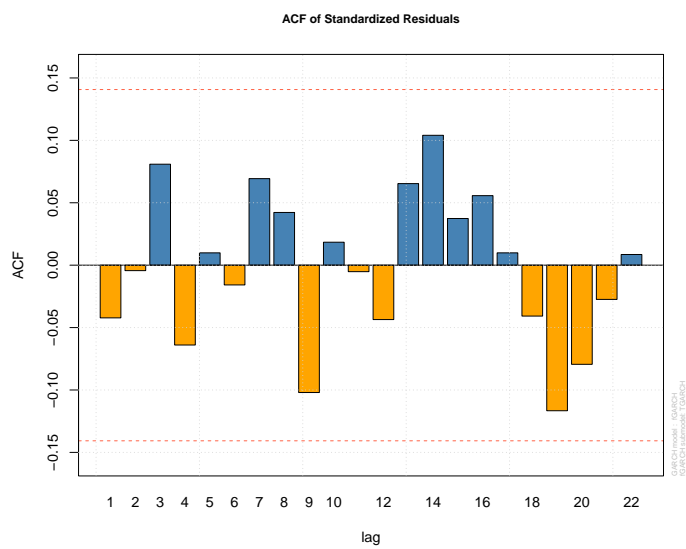


Figura 5.6: Histograma de los residuos estandarizados del modelo ARMA(1, 1)-GARCH(2, 2) para IPECO. Fuente : Elaboración propia (2022).



(a) ACF residuos estandarizados ARMA-TGARCH IPECO

(b) ACF residuos cuadráticos estandarizados ARMA-TGARCH IPECO

Figura 5.7: ACF residuos estandarizados y residuos estandarizados cuadráticos ARMA(1,1)-TGARCH(2,2) para IPECO.

Fuente : Elaboración propia (2022).

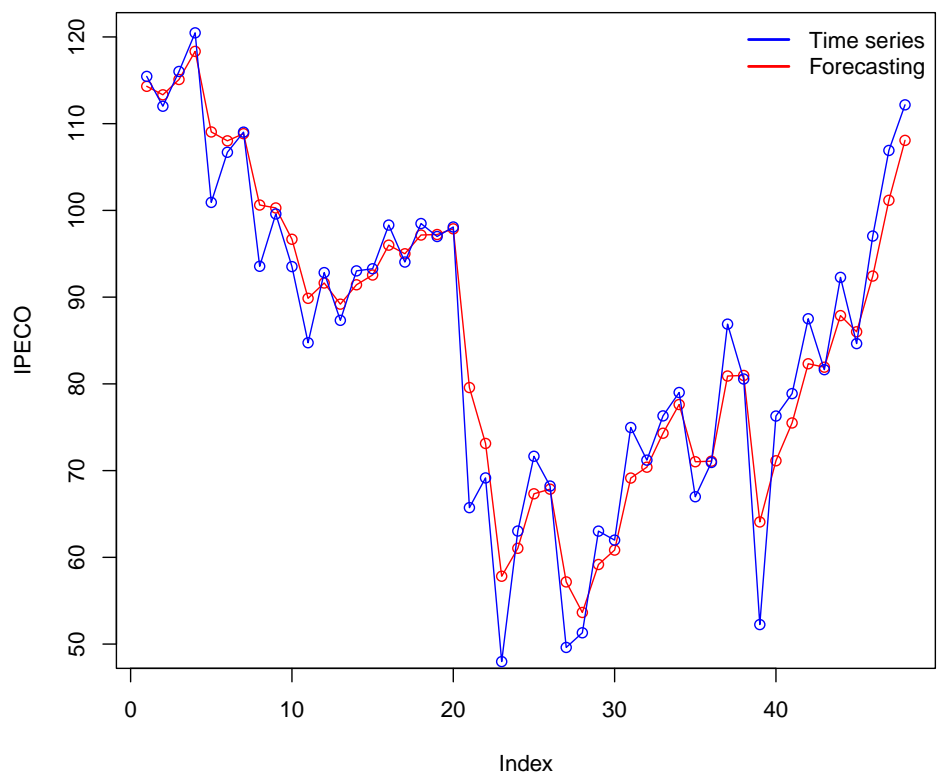


Figura 5.8: Predicciones versus datos originales IPECO. Fuente : Elaboración propia (2022).

Capítulo 6

Conclusiones

En esta tesis, se presentaron 3 tipos de modelos para series de tiempo para las percepciones económicas estudiadas. El primero fue haciendo uso del tipo de modelos más comunes en líneas generales en las series temporales, este corresponde a modelos del tipo ARIMA, el cual presentó complicaciones con respecto a la estacionariedad de la misma, debido a los problemas de invertibilidad relacionados a los parámetros autorregresivos entregados por la función `auto.arima`, tanto en el modelado de ICE como IPECO (fueron cercanos a 1).

Posteriormente, se propuso otro tipo de modelos, el cual trabaja con la idea de que existen valores umbrales presentes en los datos que serían los encargados de generar particiones a lo largo de la serie con el fin de trabajar estos de una manera más minuciosa, ya que no se analizan todos los datos con un solo modelo sino que se generan distintos modelos en virtud de la cantidad de valores umbrales que se presenten. En lo que se refiere a los resultados, los modelos determinados a pesar de cumplir con un buen análisis de diagnóstico de estos, presentaba la dificultad de recurrir a un gran número de parámetros a lo largo de las particiones para resumir de mejor manera la información de los datos.

Finalmente se trabajó con el modelo ARMA-TGARCH, el cual resulta de modelar los datos por medio de un modelo ARMA y los residuos con un modelo TGARCH. En términos del análisis de diagnóstico respectivo para ambos índices, se determinó que los residuos estandarizados eran independientes y estos no presentaban correlación serial, y el número de parámetros se vio disminuido en comparación con lo visto en los TAR. Por último, sobre las predicciones se hizo un análisis de validación cruzada para ver el rendimiento de la predicción, este mostró que en los periodos de pesimismo producidos por la crisis social y el covid, incrementa el sesgo entre el valor observado y las predicciones. Sin embargo, el RMSE propuesto es relativamente pequeño porque los modelos heterocedástico propuesto pueden lidiar con esas observaciones que producen volatilidad en los índices estudiados.

Otros trabajos futuros podrían ser el uso de modelos propuestos para otros índices, como el índice diario S&P500, el índice Hang Seng, los tipos de cambio de divisas ([Ramírez Parietti, Contreras-Reyes, y Idrovo-Aguirre, 2021](#)) u otra serie temporal. Además, la incertidumbre económica podría modelarse mediante un enfoque multivariante ([Contreras-Reyes, 2022](#)). Animamos a los investigadores a considerar los modelos propuestos para el estudio de la incertidumbre económica en otros países, especialmente afectados por crisis similares, como lo pueden ser en Latinoamérica, adaptando otras variantes pertinentes (como los factores externos).

Apéndice A

Criterio de Información de Akaike para modelos TAR

En lo que se refiere a la selección del modelo, se consideran algunos criterios como el de Información de Akaike (AIC) y el de Información Bayesiana (BIC). El primero está definido por:

$$AIC(p_1, \dots, p_r) = \sum_{j=1}^r [n_j \log(\hat{\sigma}_j^2) + 2(p_j + 1)], \quad (\text{A.1})$$

donde

$$\hat{\sigma}_j^2 = \frac{1}{n_j} \sum_{t=1}^{n_j} \hat{e}_{t(j)}^2. \quad (\text{A.2})$$

Se asumen conocidos Z_t y r , con $0 \leq p_i \leq p$ y $1 \leq d \leq p$. Además, $\hat{e}_{t(j)}$ son los residuos (o error de mínimos cuadrados) y n_j son el número de observaciones en el j -ésimo régimen.

Sin embargo, [Hurvich y Tsai \(1989\)](#) proponen una modificación al AIC, buscando así una corrección del sesgo, el cual se denomina Criterio de Información de Akaike corregido (AICc). Es así que [Wong y Li \(1998\)](#) definen el AICc de la forma:

$$AICc(p_1, \dots, p_r) = \sum_{j=1}^r \left[n_j \log(\hat{\sigma}_j^2) + \frac{n_j(n_j + p_j + 1)}{(n_j - p_j - 3)} \right]. \quad (\text{A.3})$$

Los criterios AIC y AICc tienen la desventaja de que la selección de los ordenes autorregresivos no es consistente. Para solucionar este problema de consistencia, se considera el Criterio de Información Bayesiano (BIC), dado por:

$$BIC(p_1, \dots, p_r) = \sum_{j=1}^r [n_j \log(\hat{\sigma}_j^2) + (p_j + 1) \log(n_j)]. \quad (\text{A.4})$$

Apéndice B

Código

```
1
2 #Librerías utilizadas
3 library(ggplot2)
4 library(CADFTest)
5 library(forecast)
6 library(NTS)
7 library(bfast)
8 library(rugarch)
9
10 #definir el directorio
11 setwd("C:/Directorio de trabajo")
12
13 ice <- read.csv("CEEN UDD ICE.csv")
14
15 ipeco <- read.csv("CEEN UDD IPECO.csv")
16
17 #creando un dataframe solo con las columnas de interes, las cuales son fecha e ICE
18 ice_fe_ind=data.frame(ice$ ..Fecha,ice$ICE)
19
20 #Filtrando los datos a partir del año 2006.
21 ice_2006=ice_fe_ind[20:213,]
22
23
24 #Definiendo el formato de fechas.
25 ice_2006$ice. ..Fecha=as.Date(ice_2006$ice. ..Fecha,format="%d-%m-%Y")
26
27 #Formato serie de tiempo.
28 ice_ts <- ts(ice_2006$ice.ICE, frequency=12, start=c(2006,1))
29
30 #Apartando las columnas referentes a la fecha y al índice IPECO.
31 ipeco_fe_ind=data.frame(ipeco$ ..Fecha,ipeco$IpeCo)
32
33
34 #Filtrando los datos a partir del año 2006.
35 ipeco_2006=ipeco_fe_ind[23:216,]
36
37
38 #Definiendo el formato de fechas.
39 ipeco_2006$ipeco. ..Fecha=as.Date(ipeco_2006$ipeco. ..Fecha,format="%d-%m-%Y")
40
41
```

```

42 #Formato serie de tiempo.
43 ipeco_ts <- ts(ipeco_2006$ipeco.IPeCo, frequency=12, start=c(2006,1))
44
45
46 #Grafico de ICE.
47 graf_ice=ggplot(data = ice_2006, aes(x = ice.  ..Fecha, y = ice.ICE)) + geom_line(
  color = "red", size = 1)
48 graf_ice + labs(x = "A o",y = "ICE") + theme_light()
49
50
51 #Grafico de IPECO.
52 graf_ipeco=ggplot(data = ipeco_2006, aes(x = ipeco.  ..Fecha, y = ipeco.IPeCo)) +
  geom_line(color = "red", size = 1)
53 graf_ipeco + labs(x = "A o",y = "IPECO") + theme_light()
54
55
56 ###Modelos ARIMA###
57
58 #ICE
59
60 #Se realiza un test de dickey-fuller para verificar si existe raiz unitaria.
61 adf.test(ice_ts)
62
63 #Modelado ARIMA.
64 modelo_ice_arima=auto.arima(ice_ts, stationary = T)
65 summary(modelo_ice_arima)
66
67 #IPECO
68
69 #Se realiza un test de dickey-fuller para verificar si existe raiz unitaria.
70 adf.test(ipeco_ts)
71
72 #Modelado ARIMA.
73 modelo_ipeco_arima=auto.arima(ipeco_ts, stationary = T)
74 summary(modelo_ipeco_arima)
75
76
77
78 #datos ICE(formato num).
79 datos_ice_num=ice_2006$ice.ICE
80
81 #datos IPECO(formato num).
82 datos_ipeco_num=ipeco_2006$ipeco.IPeCo
83
84 #Test de no linealidad.
85 thr.test(datos_ice_num,p=1,d=2,ini=40,include.mean=TRUE)
86 thr.test(datos_ipeco_num,p=1,d=2,ini=40,include.mean=TRUE)
87
88
89
90 #Vector con valores umbrales identificados en la serie ICE.
91 val_ice=c(datos_ice_num[97],datos_ice_num[165])
92
93
94
95 #Comparacion de modelos TAR mediante AIC.
96
97 #ICE
98

```

```

99 A <- NULL
100
101 #Funcion que genera modelos TAR en los que varia el orden autorregresivo de las
    particiones.
102 for(i in 1:2) for(j in 1:2) for(k in 1:5){
103
104     est=uTAR.est(y=datos_ice_num, arorder=c(i,j,k), thr=val_ice, d=2)
105     A<-rbind(A,c(i,j,k,est$AIC))
106
107 }
108
109
110 #IPECO
111
112 B <- NULL
113
114 #Funcion que genera modelos TAR en los que varia el orden autorregresivo de las
    particiones.
115 for(i in 1:2) for(j in 1:5) {
116
117     thr.est=uTAR(y=datos_ipeco_num, p1=i, p2=j, d=2, thrQ=c(0,1), Trim=c(0.1,0.9),
        method="RLS")
118     est=uTAR.est(y=ipeco_ts, arorder=c(i,j), thr=thr.est$thr, d=2)
119     B<-rbind(B,c(i,j,thr.est$thr,est$AIC))
120
121 }
122
123
124
125 #Definicion de los modelos a utilizar post comparacion.
126
127
128 #ICE
129 ice_modelo=uTAR.est(y=ice_ts, arorder=c(2,1,5), thr=prueba_val_ice, d=2)
130
131
132 #IPECO
133 thr.est=uTAR(y=ice_ts, p1=1, p2=5, d=2, thrQ=c(0,1), Trim=c(0.1,0.9), method="RLS"
    )
134 ipeco_modelo=uTAR.est(y=ice_ts, arorder=c(1,5), thr=thr.est$thr, d=2)
135
136
137 #Prueba de Ljung-Box modelos TAR de ambas series.
138
139 #ICE
140 ice_res <- ice_modelo$residuals
141 acf(ice_res)
142 R <- rep(NA,30)
143 for(i in 1:30) R[i]=Box.test(ice_res, lag = i, type = "Ljung")$p.value
144 plot(R, ylim=c(0,1), ylab="p.values", xlab="lags", main="Residuals")
145 abline(h=0.05, col="red", lty=2, lwd=2)
146
147 #IPECO
148 ipeco_res <- ipeco_modelo$residuals
149 acf(ipeco_res)
150 P <- rep(NA,30)
151 for(i in 1:30) P[i]=Box.test(ipeco_res, lag = i, type = "Ljung")$p.value
152 plot(P, ylim=c(0,1), ylab="p.values", xlab="lags", main="Residuals")
153 abline(h=0.05, col="red", lty=2, lwd=2)

```

```

154
155
156 #Modelo ARMA(1,1)-TGARCH(1,1), en donde se especifica el submodelo a utilizar.
157 spec_modelo_tgarch_ice= ugarchspec(variance.model = list(model="fGARCH",garchOrder
    = c(1,1),submodel="TGARCH")
158                                     ,mean.model = list(armaOrder = c(1, 1),
    external.regressors = NULL))
159
160 modelo_tgarch_ice=ugarchfit(data= ice_ts,spec=spec_modelo_tgarch_ice)
161
162 #Modelo ARMA(1,1)-TGARCH(2,2), en donde se especifica el submodelo a utilizar y la
    distribucion de los residuos.
163 spec_modelo_tgarch_ipeco= ugarchspec(variance.model = list(model="fGARCH",
    garchOrder= c(2,2),submodel="TGARCH")
164                                     ,mean.model = list(armaOrder = c(1, 1),
    external.regressors = NULL),distribution.model = "sstd")
165
166 modelo_tgarch_ipeco=ugarchfit(data= ipeco_ts,spec=spec_modelo_tgarch_ipeco)
167
168 ## Funcion encargada de las predicciones a un paso por medio del modelo ARMA-
    TGARCH para ICE
169 cval_ice <- function(spec,data, p){
170   n = length(data)
171   #q: considerando el valor p para calcular la posicion desde donde se comienza
    con las predicciones
172   q = round(n*p,0)
173   #data1: acotando los datos hasta donde se define el q.
174   data1 = data[1:round(n*p,0)]
175   #m: utilizado para definir el largo del vector que contendra la cantidad de
    predicciones a realizar
176   m = n - q + 1
177   #modelo:ajuste del modelo a los datos acotados
178   model = ugarchfit(data= data1,spec=spec)
179   #pred: vector lleno de "NA" el cual sera utilizado para llenar con las
    predicciones.
180   pred = rep(NA,m)
181
182   for (i in 1:(m-1)){
183     #Llenando el vector pred con las predicciones realizadas a un paso
    considerando los datos acotados
184     pred[i]=ugarchforecast(model,data = data1,n.ahead=1)@forecast$seriesFor
185     #Actualizando los datos acotados con el dato real siguiente
186     data1 <- c(data1,data[q+i])
187     #ajustando nuevamente el modelo considerando la actualizacion de la data
188     model=ugarchfit(data= data1,spec=spec)
189   }
190
191   rmse <- function(x,y) sqrt(sum((x-y)^2,na.rm = T))
192
193   R = rmse(pred[1:(m-1)],data[(q):(n-1)])
194
195   #Grafico que incluye las predicciones realizadas en comparativa con los datos
    reales de ICE
196   plot(pred[1:(m-1)], col = "red", type = "o", ylab = "ICE",ylim = c(-50,15))
197   lines(data[(q):(n-1)], col = "blue",type = "o")
198   legend(x = "topright",
199         legend = c("Time series", "Forecasting"),
200         bty = "n",
201         lty = c(1, 1),

```

```

202     col = c("blue", "red"),
203     lwd = 2)
204
205     return(list(pred=pred,R=R, n=n,q=q,m=m))
206
207 }
208
209 #Argumentos de la funci n: (especificacion del modelo, serie de tiempo, fraccion
    de datos a utilizar desde donde comienzan las predicciones)
210 cval_ice(spec_modelo_tgarch_ice,ice_ts,0.75)
211
212
213 ## Funcion encargada de las predicciones a un paso por medio del modelo ARMA-
    TGARCH para IPECO
214 cval_ice <- function(spec,data, p){
215     n = length(data)
216     #q: considerando el valor p para calcular la posicion desde donde se comienza
        con las predicciones
217     q = round(n*p,0)
218     #data1: acotando los datos hasta donde se define el q.
219     data1 = data[1:round(n*p,0)]
220     #m: utilizado para definir el largo del vector que contendra la cantidad de
        predicciones a realizar
221     m = n - q + 1
222     #modelo:ajuste del modelo a los datos acotados
223     modelo = ugarchfit(data= data1,spec=spec)
224     #pred: vector lleno de "NA" el cual sera utilizado para llenar con las
        predicciones.
225     pred = rep(NA,m)
226
227     for (i in 1:(m-1)){
228         #Llenando el vector pred con las predicciones realizadas a un paso
            considerando los datos acotados
229         pred[i]=ugarchforecast(modelo,data = data1,n.ahead=1)@forecast$seriesFor
230         #Actualizando los datos acotados con el dato real siguiente
231         data1 <- c(data1,data[q+i])
232         #ajustando nuevamente el modelo considerando la actualizacion de la data
233         modelo=ugarchfit(data= data1,spec=spec)
234     }
235
236     rmse <- function(x,y) sqrt(sum((x-y)^2,na.rm = T))
237
238     R = rmse(pred[1:(m-1)],data[(q):(n-1)])
239
240     #Grafico que incluye las predicciones realizadas en comparativa con los datos
        reales de IPECO
241     plot(pred[1:(m-1)], col = "red", type = "o", ylab = "IPECO",ylim = c(50,120))
242     lines(data[(q):(n-1)], col = "blue",type = "o")
243     legend(x = "topright",
244           legend = c("Time series", "Forecasting"),
245           bty = "n",
246           lty = c(1, 1),
247           col = c("blue", "red"),
248           lwd = 2)
249
250     return(list(pred=pred,R=R, n=n,q=q,m=m))
251
252 }
253

```

```
254 #Argumentos de la funcion: (especificacion del modelo, serie de tiempo, fraccion
    de datos a utilizar desde donde comienzan las predicciones)
255 cval_ipeco(spec_modelo_tgarch_ipeco,ipeco_ts,0.75)
```

Referencias

- Azzalini, A., y Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 65(2), 367–389. Descargado 2022-11-22, de <http://www.jstor.org/stable/3647510>
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3), 307–327. doi: [https://doi.org/10.1016/0304-4076\(86\)90063-1](https://doi.org/10.1016/0304-4076(86)90063-1)
- Cao, C. Q., y Tsay, R. S. (1992). Nonlinear time-series analysis of stock volatilities. *Journal of Applied Econometrics*, 7(S1), S165–S185.
- CEEN, U. (2020). Índice de confianza empresarial ice nota metodológica.
- Contreras-Reyes, J. (2022). Rényi entropy and divergence for varfima processes based on characteristic and impulse response functions. *Chaos, Solitons & Fractals*, 160, 112268.
- Davidian, M., y Carroll, R. (1987). Variance function estimation. *Journal of the American Statistical Association*, 82, 1079–1091.
- Dickey, D., y Fuller, W. (1979). Distribution of the estimates for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74, 427–431.
- Djeddour, K., y Boularouk, Y. (2013). Application of threshold autoregressive model: modeling and forecasting using us export crude oil data. *Am. J. Oil Chem. Technol.*, 9(1), 1–11.
- Ghalanos, A. (2022). Introduction to the rugarch package . (version 1 . 3-1) alexios ghalanos..
- Gibson, D., y Nur, D. (2011). Threshold autoregressive models in finance: a comparative approach. Descargado de <https://ro.uow.edu.au/asearc/26/>
- González-Farías, G., Gonzalo, J., y Russell-Noriega, M. (2007). Forecasts by predictive likelihood in threshold autoregressive model. Descargado de <http://cimat.repositorioinstitucional.mx/jspui/handle/1008/639>
- Hansen, B. E. (2011). Threshold autoregression in economics. *Statistics and its Interface*, 4(2), 123–127.
- Hoyos, M., Ramos, J., y Vivas, L. (2010). Un modelo SETAR para el PIB colombiano. *Cuadernos de Economía*, 29(52), 63–78.
- Hurvich, C. M., y Tsai, C.-L. (1989). Regression and time series model selection in small samples. *Biometrika*, 76(2), 297–307.
- Hyndman, R. J., y Khandakar, Y. (2008). Automatic time series forecasting: the forecast package for r. *Journal of statistical software*, 27, 1–22.
- Korap, L. (2011). Threshold garch modeling of the inflation & inflation uncertainty relationship: historical evidence from the turkish economy. *Istanbul Üniversitesi İktisat Fakültesi Mecmuası*, 60(2), 157–172.
- Ljung, G., y Box, G. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65, 297–303.
- López, D. (2021). ¿Cuál es la importancia de la economía en la vida diaria? Descargado de <https://economipedia.com/actual/cual-es-la-importancia-de-la-economia-en>

[-la-vida-diaria.html](#)

- Moreno, E., y Nieto, F. H. (2014). Modelos tar en series de tiempo financieras. *Comunicaciones en Estadística*, 7(2), 223–243.
- Nelson, D. (1990). Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory*, 6, 318–334.
- Pérez, F., y Velásquez, H. (2004). Análisis de cambio de régimen en series de tiempo no lineales utilizando modelos TAR. *Lecturas de Economía*, 61, 101–119.
- Petrucci, J., y Woolford, S. (1984). A threshold AR(1) model. *Journal of Applied Probability*, 21, 270–286.
- Ramírez Parietti, I., Contreras-Reyes, J., y Idrovo-Aguirre, B. (2021). Cross-sample entropy estimation for time series analysis: a nonparametric approach. *Nonlinear Dynamics*, 105, 2485–2508.
- Tong, H. (1978). On a threshold model in pattern recognition and signal processing, ed. *CH Chen, Amsterdam: Sijhoff & Noordhoff*.
- Tsay, R. S. (1998). Testing and modeling multivariate threshold models. *Journal of the American Statistical Association*, 93(443), 1188–1202.
- Uribe Bravo, A. (2015). *Pronósticos en modelos con umbrales*. Publicaciones CIMAT. Descargado de <https://repositorioslatinoamericanos.uchile.cl/handle/2250/2257038>
- Wong, C., y Li, W. K. (1998). A note on the corrected akaike information criterion for threshold autoregressive models. *Journal of Time Series Analysis*, 19(1), 113–124.
- Wu, J. (2011). Threshold garch model: Theory and application. *Ontario: The University of Western*.
- Yang, Y.-L., y Chang, C.-L. (2008). A double-threshold GARCH model of stock market and currency shocks on stock returns. *Mathematics and Computers in Simulation (MATCOM)*, 79(3), 458–474. Descargado de <https://ideas.repec.org/a/eee/matcom/v79y2008i3p458-474.html> doi: 10.1016/j.matcom.2008.01.
- Zakoian, J.-M. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and Control*, 18(5), 931–955. doi: [https://doi.org/10.1016/0165-1889\(94\)90039-6](https://doi.org/10.1016/0165-1889(94)90039-6)