



GRADO EN INGENIERÍA EN TECNOLOGÍAS DE TELECOMUNICACIÓN

TRABAJO FIN DE GRADO

CAMBIOS DE RÉGIMEN EN MODELOS DE REGRESIÓN CON BASES DE DATOS DE INTERVALOS. APLICACIÓN EN MERCADOS FINANCIEROS.

Autor: Teresa Domingo Górriz

Director: Carlos Maté Jiménez

Madrid

Julio 2021

Declaro, bajo mi responsabilidad, que el Proyecto presentado con el título
**CAMBIOS DE RÉGIMEN EN MODELOS DE REGRESIÓN CON BASES DE DATOS
DE INTERVALOS. APLICACIÓN EN MERCADOS FINANCIEROS.**

en la ETS de Ingeniería - ICAI de la Universidad Pontificia Comillas en el

curso académico 2020/21 es de mi autoría, original e inédito y

no ha sido presentado con anterioridad a otros efectos.

El Proyecto no es plagio de otro, ni total ni parcialmente y la información que ha sido

tomada de otros documentos está debidamente referenciada.



Fdo.: Teresa Domingo Górriz

Fecha: 09/ 07/ 2021

Autorizada la entrega del proyecto

EL DIRECTOR DEL PROYECTO



Fdo.: Carlos Maté Jiménez

Fecha: 09/ 07/ 2021



GRADO EN INGENIERÍA EN TECNOLOGÍAS DE TELECOMUNICACIÓN

TRABAJO FIN DE GRADO

CAMBIOS DE RÉGIMEN EN MODELOS DE REGRESIÓN CON BASES DE DATOS DE INTERVALOS.

APLICACIÓN EN MERCADOS FINANCIEROS.

Autor: Teresa Domingo Górriz

Director: Carlos Maté Jiménez

Madrid

Julio 2021

CAMBIOS DE RÉGIMEN EN MODELOS DE REGRESIÓN CON BASES DE DATOS DE INTERVALOS. APLICACIÓN EN MERCADOS FINANCIEROS.

Autor: Domingo Górriz, Teresa

Director: Maté Jiménez, Carlos

Entidad Colaboradora: ICAI – Universidad Pontificia Comillas

RESUMEN DEL PROYECTO

El modelo de cambio de régimen (Regime Switching, RS), es uno de los modelos de series temporales no lineales más conocidos en la literatura. De acuerdo con este modelo, una tendencia o estructura puede prevalecer durante un periodo aleatorio y ser reemplazada por una tendencia o estructura diferentes cuando tenga lugar un cambio de régimen. Es por esto por lo que este modelo tiene una gran aplicación en mercados financieros, así como en otros campos de estudio distintos.

Los modelos de cambio de régimen con datos clásicos son ampliamente conocidos y han sido estudiados en profundidad en la literatura. Sin embargo, los modelos de cambio de régimen con datos de intervalo (Interval-Valued, IV), formados por los límites superiores e inferiores de un intervalo típico, todavía no han sido explorados de forma muy detallada. Es por esta razón por la que en este proyecto se realizará una investigación de modelos de cambio de régimen (RS) para datos clásicos y para datos IV, con el objetivo de analizar y predecir los precios y la rentabilidad de distintas acciones de cara al día, mes o año siguientes. Por lo tanto, uno de los objetivos perseguidos a lo largo del trabajo es el diseño de un sistema original, básico y visual que ayude a inversores a la hora de analizar el comportamiento de los precios de distintas acciones en el mercado. Se incluirán aplicaciones a diferentes mercados financieros, entre ellos el NASDAQ o el IBEX 35, así como al mercado de divisas FOREX, en diferentes periodos para poder analizarlos y obtener de ellos las conclusiones pertinentes.

Palabras clave: Series Temporales de Intervalos, Cambio de Régimen, Ruptura Estructural, Regresión Lineal con Intervalos, Mercados Financieros, Matlab.

1. Introducción

Un modelo de cambio de régimen es un tipo de modelo de series de tiempo no lineales que se basan en el seguimiento de una tendencia durante un determinado periodo de tiempo, que cambiará a otra tendencia diferente a partir de un momento específico. Es en ese instante cuando se produce un cambio de régimen.

Los datos de intervalo existen desde hace mucho tiempo, sin embargo, no han sido tan explorados como los datos clásicos formados por un único valor. A pesar de ello, los datos clásicos tienen el inconveniente de que cuentan con una mayor probabilidad de poseer cierto error de precisión ya que existen numerosas variables de las que no se puede obtener información completa a partir de un único valor. Es por ello por lo que, en muchas ocasiones, los datos de intervalo pueden proporcionar una información más amplia y precisa que la proporcionada por datos clásicos.

A lo largo de este proyecto se estudiarán los cambios de régimen de un periodo a otro en modelos de regresión, haciendo uso de bases de datos de intervalo. Para poder realizar

este análisis, se evaluarán ciertos hechos ocurridos anteriormente en distintos mercados financieros y de divisas con el objetivo de estudiar los patrones y tendencias que siguen y así, poder detectar cuándo se producen estos cambios de régimen a partir de datos reales.

2. Definición del proyecto

Durante la realización del proyecto, el cual está enfocado a los mercados financieros y de divisas, se investigarán diferentes modelos de datos clásicos, así como modelos de datos de intervalos en los que se producen cambios de régimen, analizando y tratando de estimar los precios, rendimientos y comportamientos de las acciones para unos periodos de tiempo específicos.

Para poder llevar a cabo este estudio, es necesario hacer uso de Matlab, un sistema de programación con lenguaje propio, en el que se podrán diseñar y programar las funciones necesarias. Dichas funciones, basándose en diferentes variables y aplicando distintas formas de operación y análisis a cada una de ellas, serán capaces de detectar cambios de régimen en series de datos valoradas en intervalos.

Todos los métodos desarrollados en Matlab serán aplicados a valores reales de datos pasados a partir de los cuales no solo se podrá comprobar la precisión del modelo, sino que también podrá mejorarse en base a los resultados obtenidos en las diferentes funciones.

A lo largo del proyecto, se definirán un conjunto de capítulos, cada uno de ellos centrado en una temática concreta relacionada con los modelos de cambio de régimen. A lo largo de cada capítulo, se explicarán una serie de conceptos teóricos como los definidos por (Frankie, 2012) que facilitarán la comprensión de los ejemplos prácticos analizados al final de este. Para estudiar los ejemplos prácticos, será necesario utilizar la programación en Matlab y a partir de las funciones creadas, aplicando valores de datos reales, se podrán obtener conclusiones.

De esta forma, se pretende obtener un conjunto de distintas funciones que serán analizadas individual y conjuntamente para saber cuál o cuáles de ellas es capaz de obtener unos resultados más exactos en relación con el tema que se está estudiando. Así, se pretende que las funciones elegidas puedan ser de utilidad para los inversores a la hora de realizar sus movimientos.

3. Descripción del modelo

Como se ha explicado, a lo largo del proyecto se programarán una serie funciones y ficheros en Matlab que ayudarán a estudiar los temas expuestos en los distintos capítulos.

El proyecto se va a centrar en el estudio, por un lado, de métodos de regresión con los que se pueda predecir y, por otro lado, en el análisis de métodos y funciones con los que detectar cambios de régimen, parte en la que se pondrá especial atención. De esta forma, se buscará su posterior combinación en un único sistema para poder realizar predicciones de cambios de régimen de cara al futuro.

Varias funciones, cada una de ellas basada en distintos conceptos, ideas y enfoques, van a ser programadas a lo largo del estudio. Sin embargo, todas ellas siguen el siguiente esquema de elaboración y ejecución.

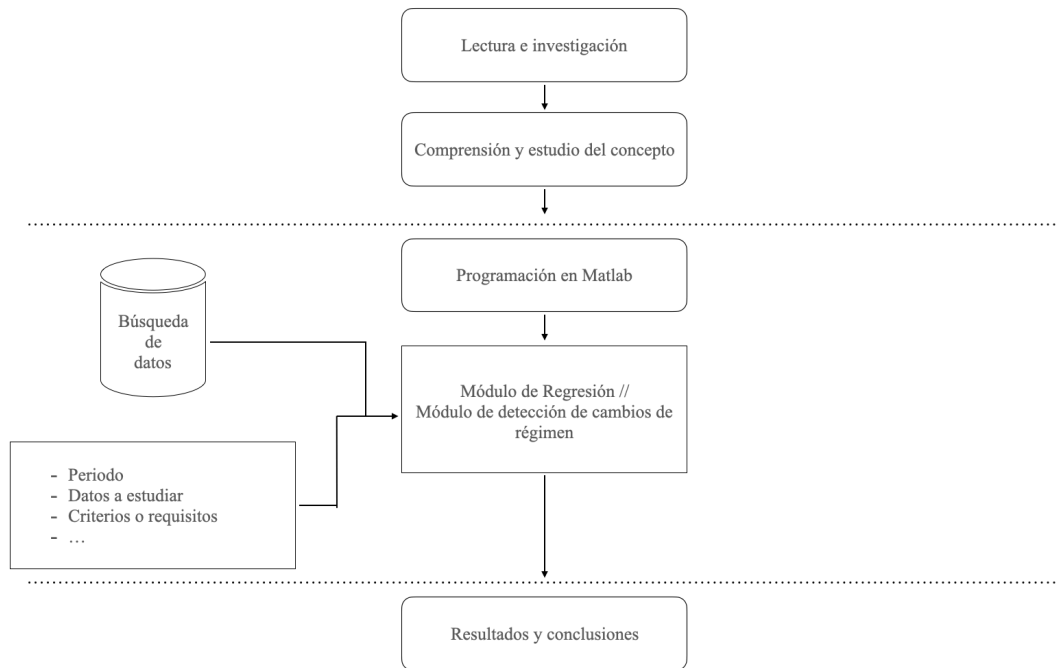


Ilustración 1 Esquema del modelo seguido en el proyecto (Elaboración propia)

Como se puede observar en el esquema, el proceso de estudio de cada capítulo comienza con la investigación sobre el tema tratado en este. Se revisarán revistas, artículos y libros de los que poder extraer la información necesaria para plasmar en el proyecto y con la que resulte sencillo comprender los conceptos en cuestión.

A partir de ahí se programará la función con relación a los conceptos estudiados y se hará una búsqueda exhaustiva de datos con los que poder operar en la función. Además, a esos datos se le aplicarán filtros, límites, reglas o condiciones, entre otras cosas, que se indicarán a la función para concretar el alcance de los datos que el usuario pretenda estudiar.

Finalmente, cada función devolverá una serie de tablas de valores y gráficas, indicando los resultados obtenidos a partir de los datos introducidos, los cuales se estudiarán a fondo para poder obtener todas las conclusiones y razonamientos posibles a partir de los resultados previamente mencionados.

4. Resultados

Tras una profunda evaluación y después de haber programado numerosas funciones y desarrollos en Matlab con el objetivo de obtener resultados destacables, se obtiene lo siguiente.

Como primer resultado del proyecto se ha creado una función basada en el modelo M-ARD de regresión (Capítulo 4.) que calcula una predicción a partir de una serie de tiempo introducida. Un ejemplo de dicha función se muestra a continuación, donde se pueden apreciar los valores para la serie de tiempo real y para la estimada, obtenidos por la función.

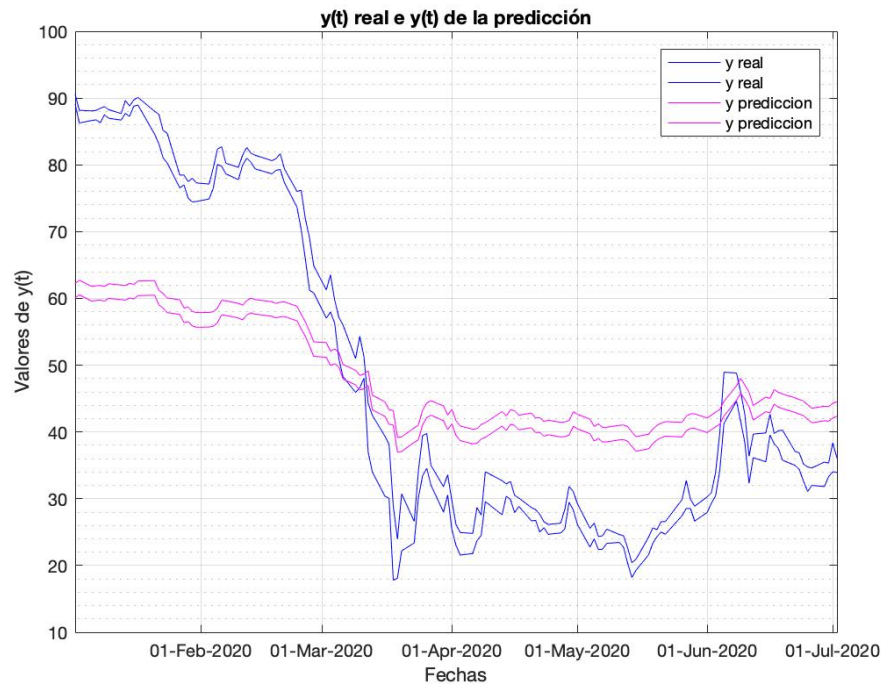


Figura 1.1 Ejemplo predicción con Modelo M-ARD (Elaboración propia)

Posteriormente, se elaboran una serie de funciones según el procedimiento explicado previamente para calcular cambios de régimen.

A continuación, se explican, junto con una serie de ejemplos, las principales funciones que se han creado. Para la Función 1, se utilizarán datos relativos a los precios de las acciones de BBVA y de Santander durante un periodo de tiempo. Por otro lado, para los otros dos casos, se hace uso de los valores de conversión del euro al dólar, EURUSD, durante unos periodos de tiempo establecidos.

Las funciones son las siguientes:

- Función 1 – taus: a partir de una función para calcular la variable umbral según (Järås & Mohammadipour Gishani, 2010) se desarrolla esta función que, tomando una serie de tiempo de datos de intervalo y sus fechas, va a calcular, entre otras cosas, un conjunto de taus que indican los puntos en los que se producen cambios de régimen a lo largo de la serie (Capítulo 5.).

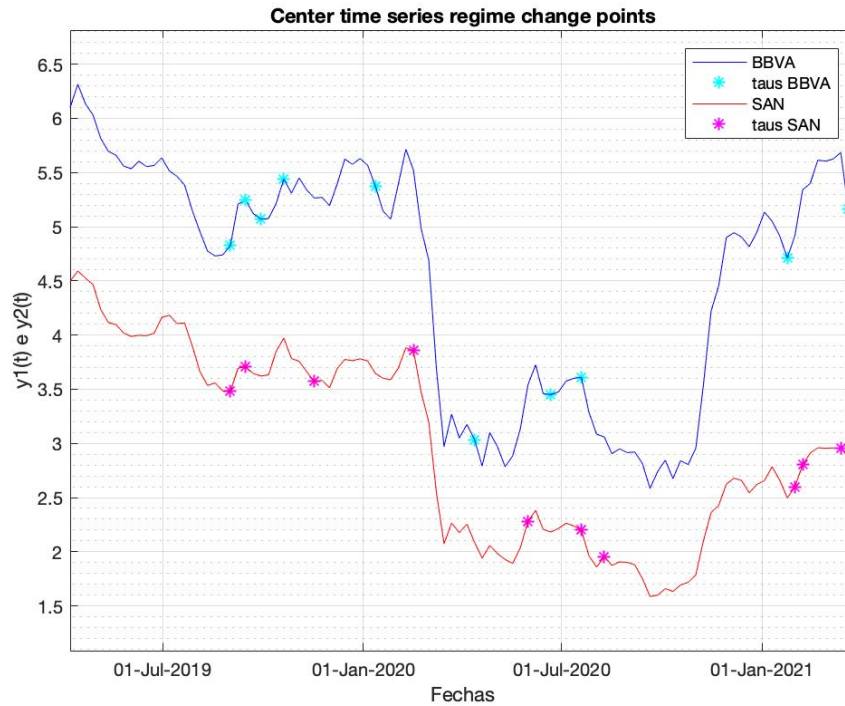


Figura 1.2 Ejemplo Función 1 – taus (Elaboración propia)

- Función 2 – coeficientes: en este caso, la función se va a enfocar hacia el cálculo de dos coeficientes, a partir de los cuales obtendrá los puntos de ruptura estructural que considere oportunos. Estos coeficientes se definen según (Maté, 2021) y son la Tasa de Cubrimiento (Coverage Rate, CR) y la Tasa de Eficiencia (Efficiency Rate, ER). La función deberá detectar todos aquellos instantes en los que ambas tasas tengan un valor inferior a un cierto límite establecido (Capítulo 6.).

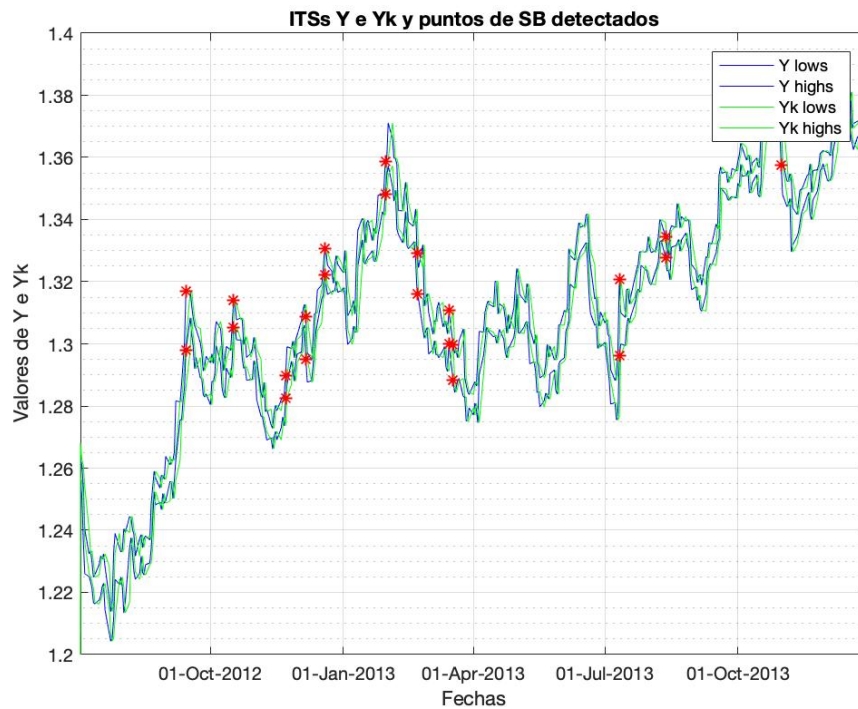


Figura 1.3 Ejemplo Función 2 – coeficientes (Elaboración propia)

- Función 3 – variaciones: en este caso, partiendo de una serie de tiempo valorada en intervalos, la función se va a centrar en el cálculo de la variación del valor de la serie temporal de un periodo a otro para, en función de estas variaciones, calcular aquellos puntos en los que considera que existe un cambio de régimen (Capítulo 6.).

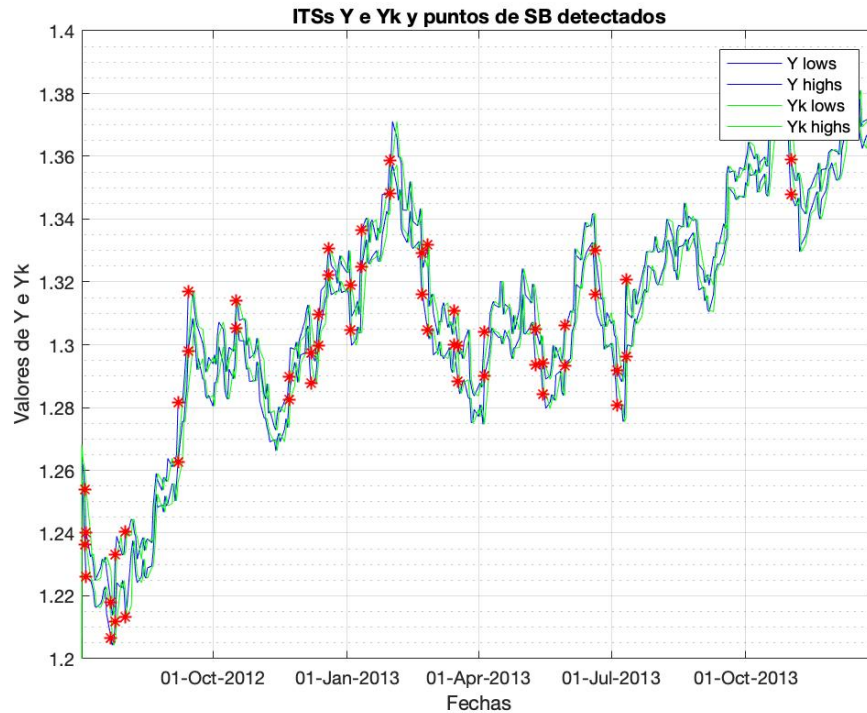


Figura 1.4 Ejemplo Función 3 – variaciones (Elaboración propia)

Cabe mencionar que los diferentes gráficos mostrados reflejan una representación de las series de tiempo introducidas junto con los momentos en los que se detectan cambios de régimen según los criterios de las diferentes funciones presentadas.

5. Conclusiones

Una vez expuestas las funciones anteriores, se estudia un ejemplo (Capítulo 7.) con el que se puede realizar una evaluación de estas. Para ello, se analizan tanto los procedimientos seguidos por las funciones como los resultados obtenidos a partir de ellas y, como prueba clave, se comparan los resultados entre sí para evaluar su efectividad y poder obtener así, las conclusiones del proyecto.

Por un lado, se concluye que los resultados a los que llega el modelo M-ARD a pesar de no ser exactos por tratarse de una predicción, son coherentes ya que en todo momento la serie de tiempo de la predicción sigue las mismas tendencias y fluctuaciones que la serie de tiempo real, aunque lo hace de forma menos pronunciada.

Además, de entre las distintas funciones de detección de cambios de régimen, se obtiene que la más efectiva es la función 2 por ser la que más puntos en común tiene con las

demás, seguida de la función 3 que, de ajustarse correctamente, también proporciona unos resultados con gran acierto.

Por último, la función 1 será la menos conveniente ya que el usuario debe indicarle el número de cambios de régimen que busca y la separación mínima entre ellos por lo que en caso de no introducir unos valores adecuados, la función dejaría de ser efectiva.

Finalmente, como conclusión general se destaca que se ha podido desarrollar un sistema efectivo y visual capaz de ayudar a inversores a realizar sus movimientos en mercados financieros con mayor seguridad y menor riesgo, cumpliendo así con el objetivo principal del proyecto.

6. Referencias

- Frankie, J. (2012). Markov switching time series models. *Handbook of Statistics*, 30, 99-122.
- Järås, J., & Mohammadipour Gishani, A. (2010). *Threshold Detection in Autoregressive Non-linear Models*. Lund University Department of Statistics., Department of Statistics.
- Maté, C. (2021). Combining Interval Time Series Forecasts. A First Step in a Long Way (Research Agenda). *Revista Colombiana de Estadística*, 123-157.

REGIME SWITCHING IN REGRESSION MODELS UNDER INTERVAL-VALUED DATABASES. APPLICATIONS IN FINANCIAL MARKETS.

Author: Domingo Górriz, Teresa

Supervisor: Maté Jiménez, Carlos

Collaborating Entity: ICAI – Universidad Pontificia Comillas

ABSTRACT

The Regime Switching (RS) model is one of the most popular non-linear time series models in the literature. As such, a trend or structure can prevail for a random period, and it will be replaced by a different trend or structure when a regime switching occurs. Therefore, this model has a wide range of application in financial markets, as well as in other different fields of study.

Regime Switching models with crisp data are widely known and have been studied on several occasions in the literature. However, RS models with Interval-Valued (IV) data, which are composed of the upper and lower bounds of a typical interval, have not yet been explored in detail. This is the reason why this project will research RS models for crisp data and for IV data, with the aim of analyzing and predicting the prices and returns of stocks for the day, month, or year ahead. Therefore, one of the objectives of the project is to design an original, basic, and visual system to help investors and traders to analyze the behavior of the prices of different stocks in the market. For this reason, applications to different financial markets will be included, such as NASDAQ and IBEX 35, as well as the FOREX currency market, in several periods to be able to analyze different cases and obtain relevant conclusions from them.

Keywords: Interval-Valued Time Series, Regime Switching, Structural Break, Linear Regression with Intervals, Financial Markets, Matlab.

1. Introduction

A Regime Switching model is a type of non-linear time series model, which is based on following a pattern over a certain period, that will change to a different pattern after a specific point in time. This point represents the moment at which a regime switching takes place.

Interval valued data have been around for a long time; however, it has not been explored as much as simple crisp data. The latter have the disadvantage that they are more likely to contain some accuracy error as there are many variables for which it is not possible to obtain complete information from just a single value. Therefore, on many occasions, IV data can provide broader and more accurate information than that provided by crisp data.

Throughout this project we will study regime switching from one period to another in different regression models, making use of interval-valued databases. To carry out this analysis, certain past events in different financial markets will be evaluated to study the patterns and trends they follow and thus, to be able to detect when these regime shifts occur based on real data.

2. Project definition

During the execution of the project, which is focused on financial and currency markets, different crisp data models will be evaluated, as well as IV data models in which regime switching occur, analyzing and trying to estimate stock prices, returns and behaviors for specific time periods.

To carry out this study, it is necessary to make use of Matlab, a programming system with its own language, in which a series of functions can be designed and programmed. These functions, based on different variables and applying different forms of operation and analysis to each of them, will be capable of detecting regime switching in IV data series.

All the methods developed in Matlab will be applied to real values of past data from which it will not only be possible to check the accuracy of the model, but also to improve it based on the results obtained in the different functions.

Throughout the project, a set of chapters will be defined, each focusing on a specific topic related to regime switching models. Each chapter will define a series of theoretical concepts, such as the ones defined by (Frankie, 2012), that will be explained to facilitate the understanding of the practical examples analyzed at the end of the chapter. To study these practical examples, it will be necessary to use Matlab programming and from the functions created, applying real data, different conclusions will be obtained.

To conclude, the aim of the project is to obtain an original set of different functions that will be analyzed individually and collectively to find out which of them can obtain the most accurate results in relation to the subject being studied. In this way, it is intended that the functions chosen can be useful for investors when making their movements.

3. Model description

As explained above, a series of Matlab functions will be programmed throughout the project to help study the topics covered in the different chapters.

The project will focus on the study, on the one hand, of regression methods with which to predict future values and, on the other hand, on the analysis of methods and functions with which to detect regime switching. The latter will require special attention. In this way, their subsequent combination in a single system will be sought in order to be able to make predictions of regime switching in the future.

Several functions, each based on different concepts, ideas, and approaches, will be scheduled throughout the study. However, all of them follow the next execution scheme.

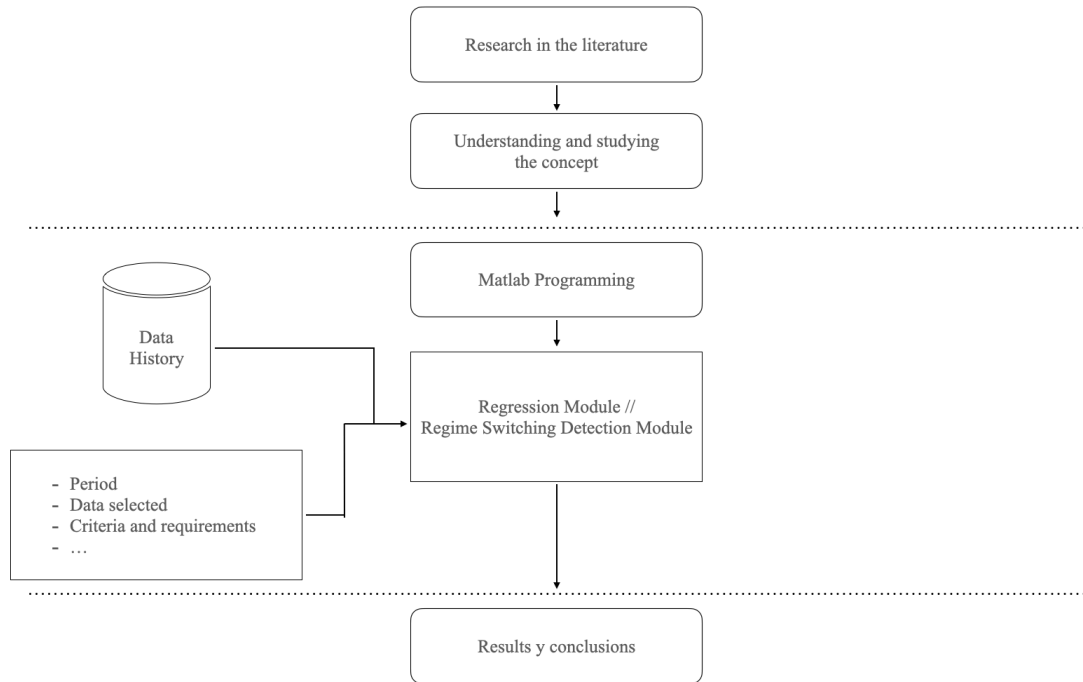


Ilustración 2 Scheme of the model followed during the project

As it can be seen in the outline, the study process for each chapter begins with research on the topic covered in the chapter. Journals, articles, and books will be reviewed to extract the necessary information to be included in the project and with which it will be easy to understand the concepts in question.

From there, the function will be programmed according to the concepts studied and an exhaustive research will be made for data with which to operate and integrate into the function. Furthermore, filters, boundaries, rules, or conditions, among other things, will be applied to these data, which will be indicated to the function to specify the scope of the data that the user intends to study.

Finally, each function will return a series of tables of values and graphs, indicating the results obtained from the data entered, which will be studied to obtain all possible conclusions and reasoning for the results previously mentioned.

4. Results

After a thorough evaluation and after having programmed numerous functions and developments in Matlab with the aim of achieving outstanding results, the following is obtained.

As a first result of the project, a function based on the M-ARD regression model (Capítulo 4.) has been obtained. This function calculates a prediction from an input IV time series. An example of this function is shown below, where the values for the actual and estimated time series obtained by the function are shown.

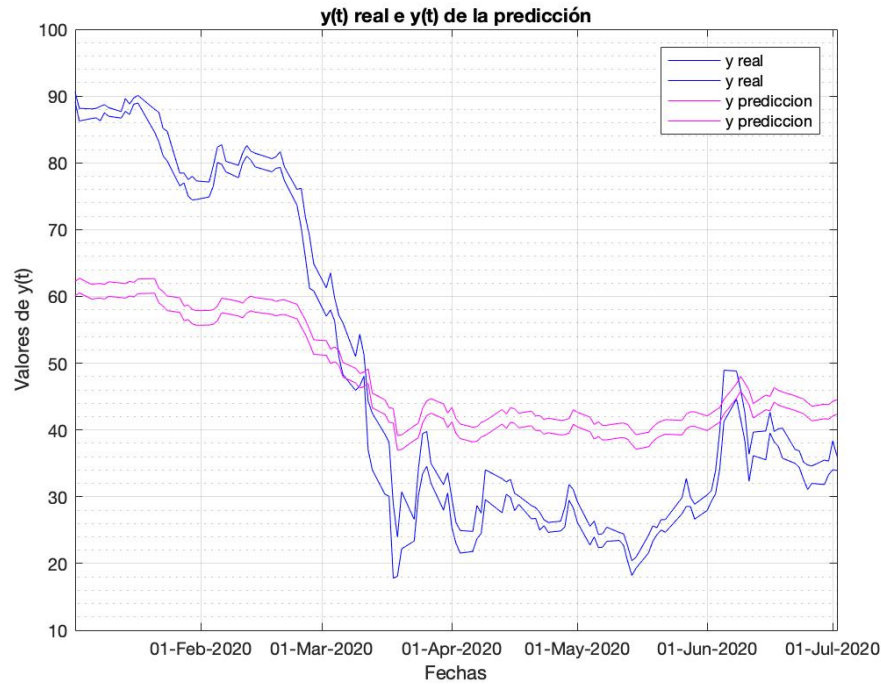


Figura 1.5 Example of M-ARD regression model (Elaboración propia)

In addition, a series of functions are developed according to the procedure explained above to calculate regime switching points.

The main functions that have been created are explained below, together with a series of examples. For Function 1, data relating to the stock prices of BBVA and Santander over a period will be used. For the other two cases, the conversion values of the euro to the dollar, EURUSD, during established time periods are used.

The main functions are explained here:

- Function 1 - taus: based on a function to calculate the threshold variable according to (Järås & Mohammadipour Gishani, 2010), this function is developed which, taking a time series of interval data and its dates, will calculate, among other things, a set of taus indicating the points at which regime switching occur throughout the series (Capítulo 5.).

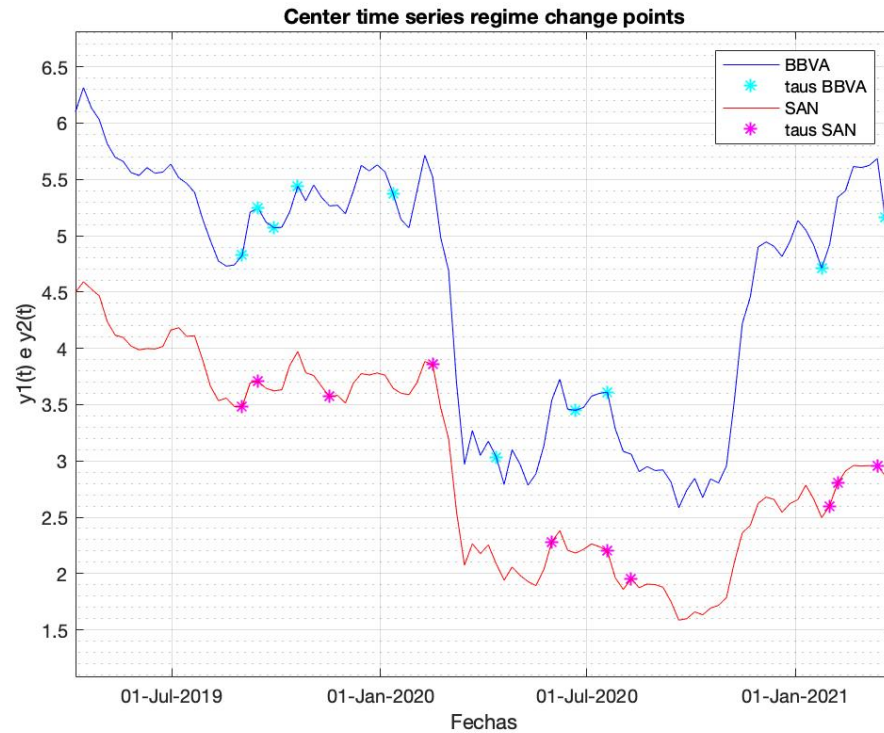


Figura 1.6 Example of Function 1 – taus (Elaboración propia)

- Function 2 - coefficients: in this case, the function will focus on the calculation of two coefficients, from which it will obtain the structural breakpoints it considers appropriate. These coefficients, defined by (Maté, 2021) are the Coverage Rate (CR) and the Efficiency Rate (ER). The function must detect all those moments in which both rates have a value lower than a certain established limit (Capítulo 6.).

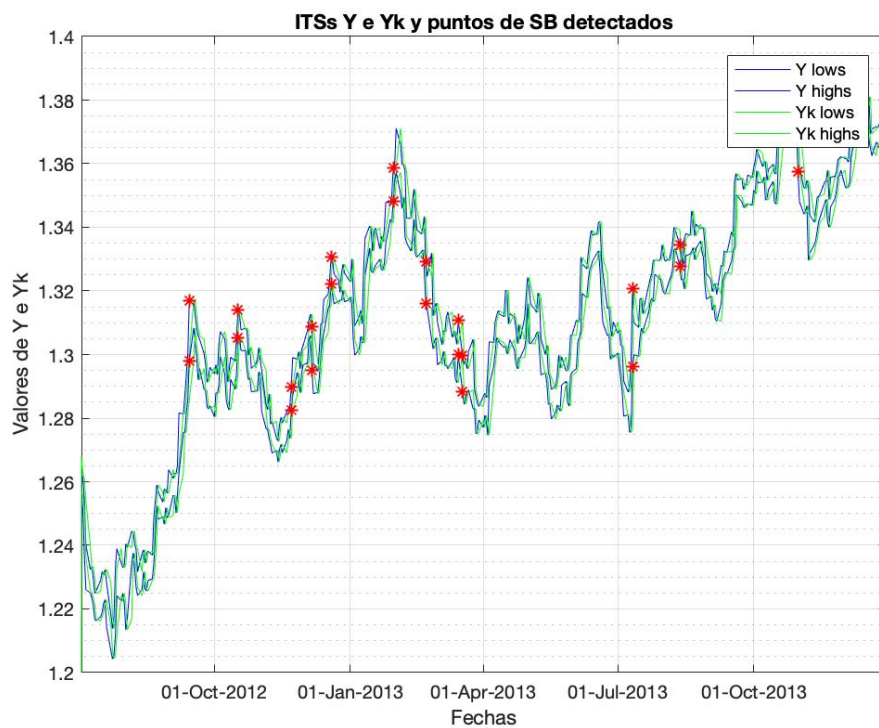


Figura 1.7 Example of Function 2 – coefficients (Elaboración propia)

- Function 3 - variations: in this case, starting from a time series valued in intervals, the function is going to focus on calculating the variation of the value of the time series from one period to another to calculate, based on these variations, those points at which it considers that a regime switching takes place (Capítulo 6.).

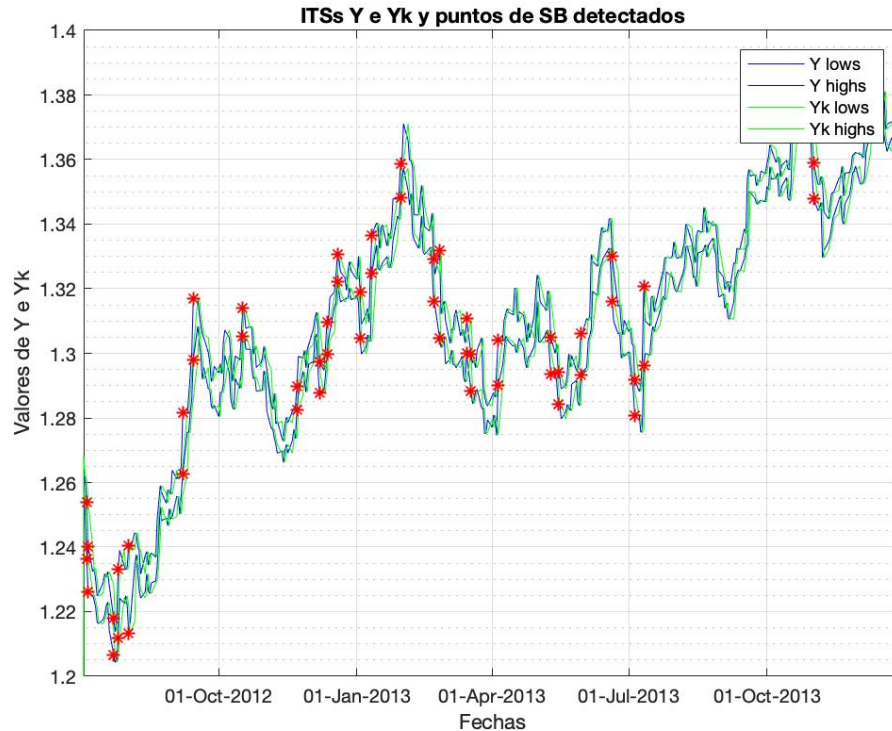


Figura 1.8 Example of Funcion 3 – variations (Elaboración propia)

It should be noted that the graphs shown reflect a representation of the time series introduced to each of the functions together with the moments in which regime switching points are detected according to the criteria used by each of the functions discussed.

5. Conclusions

Once the above functions have been explained, an example is studied (Capítulo 7.) with which an evaluation of these functions can be carried out. For this purpose, both the procedures followed by the functions and the results obtained from them are analyzed and, as a key test, the results are compared with each other to evaluate their effectiveness and thus obtain the conclusions of the project.

On the one hand, it is concluded that the results obtained by the M-ARD model, despite not being completely exact because it is a prediction, they are coherent since at every moment the prediction time series follows the same trends and fluctuations as the real time series, being the first one smoothed.

In addition, among the different regime switching detection functions, function 2 is the most effective as it has the most points in common with the others, followed by function 3 which, if set correctly, also provides very accurate results.

Finally, function 1 will be the least convenient since the user must indicate the number of regime switching points he is looking for and the minimum gap between them, so that if appropriate values are not entered, the function will no longer be effective.

Finally, as a general conclusion, it should be noted that it has been possible to develop an effective and visual system capable of helping investors to make their movements in financial markets with greater security and lower risk, thus the main objective of the project has been fulfilled.

6. References

Frankie, J. (2012). Markov switching time series models. *Handbook of Statistics*, 30, 99-122.

Järås, J., & Mohammadipour Gishani, A. (2010). *Threshold Detection in Autoregressive Non-linear Models*. Lund University Department of Statitics., Department of Statitics.

Maté, C. (2021). Combining Interval Time Series Forecasts. A First Step in a Long Way (Research Agenda). *Revista Colombiana de Estadística*, 123-157.

Índice de la memoria

Capítulo 1. Introducción	7
1.1 Motivación del proyecto.....	7
1.2 Objetivos del proyecto.....	8
1.3 Metodología del proyecto.....	10
Capítulo 2. Conceptos generales	11
2.1 Introducción a la predicción y series de tiempo	11
2.1.1 Predicción.....	11
2.1.2 Descomposición en series temporales.....	13
2.1.3 Precisión del pronóstico.....	14
2.1.4 Comparación de métodos de predicción	16
2.2 Regresión.....	17
2.2.1 Regresión simple	17
2.2.2 Regresión múltiple.....	19
2.3 Otros métodos de predicción.....	22
2.3.1 Métodos de promedio	22
2.3.2 Métodos de suavizamiento exponencial	22
2.4 Métodos de predicción no lineales	24
2.4.1 Redes neuronales.....	24
2.4.2 Modelos de heterocedasticidad condicional	26
Capítulo 3. Modelos de cambio de régimen con datos clásicos	31
3.1 Modelos de Markov.....	31
3.2 Modelos Autorregresivos de Umbral, TAR	32
3.3 Modelos Autorregresivos de Media Móvil	34
3.3.1 Conceptos previos	34
3.4 Contraste de Portamanteau y PACF	34
3.5 Estacionariedad y series de datos temporales.....	35
3.6 Modelo de Camino Aleatorio	35
3.7 Introducción a los modelos ARIMA de series temporales de datos.....	36
3.7.1 Modelo Autorregresivo Simple (AR)	36
3.7.2 Modelo Simple de Media Móvil (MA)	37

3.7.3 Modelos Autorregresivos de Media Móvil (ARMA)	37
3.7.4 Modelos Integrados Autorregresivos de Media Móvil (ARIMA)	38
3.8 Contraste de Raíz Unitaria	38
3.9 Identificación del modelo	39
3.9.1 Estimación del parámetro	40
3.9.2 Revisión de la identificación.....	40
3.9.3 Comprobación del diagnóstico.....	41
3.10 Regresión segmentada	41
3.10.1 Regresión segmentada. Explicación.....	41
3.10.2 Procedimiento de pruebas	44
3.10.3 Rango de no-efecto	45
3.11 Ejemplo de regresión segmentada	46
Capítulo 4. Métodos y modelos de regresión lineal con datos de intervalo.....	52
4.1 Conceptos básicos al analizar series de tiempo en intervalos (ITS).....	52
4.2 Conceptos esenciales al analizar varias predicciones para una ITS	54
4.2.1 Error de la distancia media (MDE)	54
4.2.2 Varianza Relativa Media de Intervalo (iARV)	54
4.2.3 Estadístico U de Theil de intervalo	55
4.3 Modelos de Regresión Lineal para Intervalos (ILR).....	55
4.3.1 Método del Centro (CM)	57
4.3.2 Método minMax (mMM).....	58
4.3.3 Método del Centro y Radio (CRM).....	59
4.3.4 Método del Centro y Radio con Restricciones (CCRM).....	60
4.3.5 Regresión Lineal Simple Basada en Aritmética de Intervalos (Modelo M).....	62
4.3.6 Regresión Lineal Simple Basada en Lasso.....	65
4.3.7 Método de Centro y Radio Conjuntos con Restricciones (CCRJM)	67
4.3.8 Enfoque Parametrizado para Regresión Lineal de Datos de Intervalos.....	68
4.4 Ejemplo de Modelo M.....	70
Capítulo 5. Modelos de cambios de régimen con datos de intervalo. Enfoque de datos simbólicos 73	
5.1 Introducción.....	73
5.2 Modelos Autorregresivos de Umbral para series de datos de Intervalos (TARI)	75

5.2.1 Introducción	75
5.2.2 Desarrollo Modelo TARI	77
5.2.3 Estimación Modelo TARI	82
5.2.4 Aplicación empírica del Índice S&P500	85
5.2.5 Conclusiones.....	89
5.3 Ejemplos función TAR (Función 1).....	89
Capítulo 6. Modelos de cambios de régimen con datos de intervalo. Enfoque de aritmética de intervalos.....	101
6.1 Introducción.....	101
6.2 Test de Bondad de Ajuste, CR y ER	102
6.2.1 Tasa de Cubrimiento (CR).....	102
6.2.2 Tasa de Eficiencia (ER).....	102
6.3 Concepto de Variación	104
6.4 Ejemplo función CR y ER (Función 2)	105
6.5 Ejemplo función Variación (Función 3).....	110
6.6 Comparación de los dos ejemplos anteriores	114
Capítulo 7. Aplicaciones a acciones del NASDAQ.....	117
7.1 Ejemplo práctico UAL (United Airlines).....	117
Capítulo 8. Conclusiones.....	129
8.1 Conclusiones	129
8.2 Desarrollos futuros	131
Capítulo 9. Bibliografía.....	132
ANEXO I: ALINEACIÓN DEL PROYECTO CON LOS ODS.....	136

Índice de figuras

Figura 1.1 Ejemplo predicción con Modelo M-ARD (Elaboración propia).....	10
Figura 1.2 Ejemplo Función 1 – taus (Elaboración propia)	11
Figura 1.3 Ejemplo Función 2 – coeficientes (Elaboración propia).....	11
Figura 1.4 Ejemplo Función 3 – variaciones (Elaboración propia).....	12
Figura 1.5 Example of M-ARD regression model (Elaboración propia)	17
Figura 1.6 Example of Function 1 – taus (Elaboración propia)	18
Figura 1.7 Example of Function 2 – coefficients (Elaboración propia)	18
Figura 1.8 Example of Funcion 3 – variations (Elaboración propia)	19
Figura 2.1 Precio del oro (€/onza) en los últimos 5 años (Inversoro.es, 2020).....	12
Figura 2.2 Valor del IBEX 35 en los últimos 5 años (Bolsamanía, 2020)	13
Figura 2.3 Descomposición de series temporales (Vidal, 2016)	14
Figura 2.4 Esquema de una red neuronal (Atria Innovation, 2019)	24
Figura 3.1 Ejemplo de regresión segmentada - Regresión lineal (Wikipedia, 2021).....	42
Figura 3.2 Ejemplo de regresión segmentada - Regresión segmentada (Wikipedia, 2021)	43
Figura 3.3 Regresión segmentada - Ejemplo 1.1 (Elaboración propia).....	47
Figura 3.4 Regresión segmentada - Ejemplo 1.2 (Elaboración propia).....	49
Figura 3.5 Regresión segmentada - Ejemplo 2 (Elaboración propia).....	51
Figura 4.1 Ejemplo Modelo M-ARD (Elaboración propia)	72
Figura 5.1 Resultados de las predicciones de los modelos TARI y CR-SETAR (1) (Sun, Han, Hong, & Wang, 2018)	88
Figura 5.2 Resultados de las predicciones de los modelos TARI y CR-SETAR (2) (Sun, Han, Hong, & Wang, 2018)	88
Figura 5.3 Resultados de las predicciones de los modelos TARI y CR-SETAR (3) (Sun, Han, Hong, & Wang, 2018)	89
Figura 5.4 Valores de los residuos para $y(t)$ (Elaboración propia).....	91
Figura 5.5 Serie de tiempo original y serie con los dos regímenes (Elaboración propia) ...	92
Figura 5.6 iBoxPlot BBVA (Elaboración propia)	94
Figura 5.7 iBoxPlot Santander (Elaboración propia)	95

Figura 5.8 Cambios de régimen en las ITS de los valores máximos (Elaboración propia).	96
Figura 5.9 Cambios de régimen en las ITS de los valores mínimos (Elaboración propia) .	96
Figura 5.10 Cambios de régimen en las ITS de los centros (Elaboración propia)	97
Figura 5.11 Cambios de régimen en las ITS de los radios (Elaboración propia)	97
Figura 5.12 Taus en valores máximos de BBVA, por rectángulos (Elaboración propia) ...	98
Figura 5.13 Fechas con taus en máximos y mínimos de BBVA (Elaboración propia).....	100
Figura 6.1 CR y ER (Maceda, 2018)	103
Figura 6.2 Serie de tiempo real y serie de tiempo con retraso (Elaboración propia)	106
Figura 6.3 Intervalo medio real y estimado y valores de ER y CR (Elaboración propia) .	107
Figura 6.4 CR, ER y warnings (Elaboración propia)	108
Figura 6.5 Representación de los intervalos en las fechas con ruptura estructural (Elaboración propia)	109
Figura 6.6 Puntos de ruptura estructural detectados (Elaboración propia).....	110
Figura 6.7 Representación temporal de $y(t)$ e $y(t-k)$ (Elaboración propia)	111
Figura 6.8 Variaciones de máximos y mínimos (Elaboración propia)	112
Figura 6.9 Variaciones máximos y mínimos y puntos de ruptura (Elaboración propia)...	113
Figura 6.10 Puntos de ruptura estructural detectados (Elaboración propia).....	114
Figura 6.11 Puntos comunes de ruptura estructural detectados (Elaboración propia)	116
Figura 7.1 Predicción con Modelo M-ARD, UAL (Elaboración propia).....	118
Figura 7.2 iBoxPlot, UAL (Elaboración propia)	119
Figura 7.3 Cambios de régimen con Función 1, UAL (Elaboración propia)	120
Figura 7.4 Datos resumen Función 2, UAL (Elaboración propia)	121
Figura 7.5 Cambios de régimen con Función 2, UAL (Elaboración propia)	122
Figura 7.6 Intervalos en puntos de Cambio de Régimen con Función 2, UAL (Elaboración propia).....	122
Figura 7.7 Variaciones en máximos y mínimos con Función 3, UAL (Elaboración propia)	124
Figura 7.8 Cambios de régimen con Función 3, UAL (Elaboración propia)	124

Índice de tablas

Tabla 1 Datos 1, regresión segmentada - Ejemplo 1.1 (Elaboración propia)	46
Tabla 2 Datos 2, regresión sementada - Ejemplo 1.1 (Elaboración propia)	47
Tabla 3 Datos 1, regresión segmentada - Ejemplo 1.2 (Elaboración propia)	48
Tabla 4 Datos 2, regresión sementada - Ejemplo 1.2 (Elaboración propia)	48
Tabla 5 Datos 1, regresión segmentada - Ejemplo 2 (Elaboración propia)	50
Tabla 6 Datos 2, regresión sementada - Ejemplo 2 (Elaboración propia)	50
Tabla 7 Datos de la función (Elaboración propia)	91
Tabla 8 Fechas con taus en máximos y mínimos en BBVA y SAN (Elaboración propia) .	99
Tabla 9 Fechas con puntos de ruptura estructural (Elaboración propia)	109
Tabla 10 Fechas con puntos de ruptura estructural (Elaboración propia)	113
Tabla 11 Fechas con puntos comunes de ruptura estructural (Elaboración propia)	115
Tabla 12 Cambios de régimen en Función 1, UAL (Elaboración propia)	120
Tabla 13 Cambios de régimen en Función 2, UAL (Elaboración propia)	123
Tabla 14 Cambios de régimen en Función 3, UAL (Elaboración propia)	124
Tabla 15 Tabla resumen resultados UAL, datos predicción (Elaboración propia)	125
Tabla 16 Tabla resumen resultados UAL, datos reales (Elaboración propia)	127

Capítulo 1. INTRODUCCIÓN

En la realización de este proyecto, se va a llevar a cabo un estudio de los cambios de un régimen a otro en modelos de regresión haciendo uso de bases de datos de intervalos. Para ello, se va a realizar el análisis de ciertos hechos ocurridos anteriormente en los mercados financieros con el objetivo de estudiar los patrones y tendencias que siguen para poder predecir situaciones futuras.

Un modelo de cambio de régimen es un tipo de modelo de series de tiempo no lineales, los cuales se basan en el seguimiento de una cierta tendencia o estructura durante un determinado periodo de tiempo, que cambiará a una tendencia o estructura diferente a partir de un momento específico. Es ahí cuando se produce un cambio de régimen.

A lo largo del proyecto, el cual está enfocado a los mercados financieros, se investigarán modelos clásicos, así como modelos y métodos de datos en intervalos con cambios de régimen, analizando y tratando de predecir los precios de las acciones y sus rendimientos para un determinado periodo de tiempo. De esta forma, se pretende obtener un sistema original que ayude a los inversores a analizar los mercados y realizar sus movimientos.

1.1 MOTIVACIÓN DEL PROYECTO

Cada vez son más las personas que deciden invertir su dinero en acciones con el objetivo de obtener una cierta rentabilidad al cabo del tiempo. El mercado bursátil es un mercado volátil, es decir, es un mercado que varía con facilidad y que además lo hace de forma impredecible, por lo que se convierte también en un mercado arriesgado en el que es complicado saber lo que va a ocurrir.

Existen muchas formas de estudiar el comportamiento que los precios de las acciones han tenido hasta ahora, así como diferentes indicadores técnicos que ayudan a realizar un

análisis bursátil para tratar de analizar la evolución futura de los precios de los activos financieros. Sin embargo, es necesario hacer uso de otros indicadores o herramientas que permitan realizar una predicción más precisa para poder tomar la decisión más certera posible a la hora de invertir.

El objetivo de la realización de este proyecto, aplicado a mercados financieros, es poder crear un sistema a partir del cual sea posible analizar cómo se comportan y cómo se van a comportar los precios de las acciones, partiendo del estudio de series de datos temporales pasados, detectando cuándo en las series temporales tienen lugar cambios de régimen. De esta forma, se busca crear una herramienta que ayude a los accionistas o inversores a tomar las decisiones referentes a sus inversiones con el objetivo de obtener los mayores rendimientos posibles.

1.2 OBJETIVOS DEL PROYECTO

Con este proyecto se persigue realizar un estudio de los modelos de cambios de régimen en series de tiempo evaluadas en intervalos, centrándose en el estudio de sus aplicaciones en el mercado bursátil, con el objetivo principal de poder evaluar de cara a obtener ideas hacia el futuro de los comportamientos de los precios de las acciones en el mercado.

Para ello, se estudiarán y programarán diferentes métodos de detección de cambios de régimen para poder aplicarlos a casos reales.

Más concretamente, se van a perseguir los siguientes objetivos:

- Capítulo 3. Modelos de cambio de régimen con datos clásicos. Un modelo de cambio de régimen es aquel en el que se produce una diferencia entre la tendencia seguida por los datos de un periodo a otro. Se realizará un estudio de este tipo de modelos, explicando los conceptos más importantes que serán la base para la elaboración del resto del proyecto. Más tarde se hará uso de estos

conceptos teóricos y se programarán funciones en Matlab para el estudio de diferentes casos, (Frankie, 2012), (Hansen, 2011).

- Capítulo 4. Métodos y modelos de regresión lineal con datos de intervalo. La regresión lineal es un modelo matemático utilizado para aproximar la relación de dependencia entre una variable dependiente, las variables independientes y un término aleatorio. En este caso, se estudiará lo anterior haciendo uso de datos por intervalos. Se realizará, inicialmente, una explicación teórica y más tarde se partirá de estos conceptos para la programación de funciones en Matlab (Maté, 2021), (Maceda, 2018).
- Capítulo 5. Modelos de cambios de régimen con datos de intervalo. Enfoque de datos simbólicos. En este capítulo, a diferencia del primero, los modelos estudiados serán aquellos en los que se produzca un cambio de régimen, pero partiendo de datos por intervalo. Se explica el concepto de regresión segmentada, formada a partir de diferentes tramos de regresión lineal (Ryan & Porth, 2007). El capítulo se centrará en el análisis teórico y práctico de los datos simbólicos.
- Capítulo 6. Modelos de cambios de régimen con datos de intervalo. Enfoque de aritmética de intervalos. En relación con el apartado anterior, pero proporcionando un punto de vista diferente, a partir de la aritmética de intervalos y, por tanto, se realizarán distintas funciones con operaciones a partir de los datos de los intervalos.
- Capítulo 7. Aplicaciones a acciones del NASDAQ. Todo lo explicado previamente se va a poder resumir en este capítulo, en el que se va a partir de casos reales de acciones de la bolsa de valores NASDAQ para poder ver en conjunto los conceptos más importantes estudiados.
- Capítulo 8. Conclusiones. Para finalizar, a lo largo de este capítulo se explicarán todas las conclusiones obtenidas a lo largo del proyecto, así como los resultados derivados del capítulo anterior.

1.3 METODOLOGÍA DEL PROYECTO

Una vez definido el proyecto a realizar, es preciso definir cuáles van a ser los objetivos que se persiguen, así como los pasos a seguir para la realización de las distintas tareas y la planificación de estas.

Para llevar a cabo este proyecto, el primer paso necesario es hacer una investigación en profundidad sobre distinta información relacionada con la predicción y las series temporales, así como el aprendizaje de una serie de conceptos estadísticos relacionados que más tarde se aplicarán para el estudio e identificación de los cambios de régimen.

Además, se hará uso de herramientas y programas como Matlab para crear y programar las funciones necesarias que permitan la realización de las tareas definidas. Para ello se partirá de los conceptos investigados previamente de forma que se puedan poner en práctica, proporcionando una más sencilla y rápida comprensión.

Se tomarán también ejemplos reales de datos pasados para poder comprobar los resultados y, a partir de ellos, obtener las conclusiones pertinentes, las cuales se irán exponiendo a lo largo del proyecto y finalmente, se recapitularán en el último capítulo.

A lo largo del desarrollo del proyecto se realizará la memoria correspondiente en la que se expliquen los conceptos más importantes, acompañados de distintos ejemplos que ayuden a su comprensión, así como los pasos seguidos para la realización del sistema. Finalmente, una vez terminado el trabajo se realizará la presentación y defensa de este.

Capítulo 2. CONCEPTOS GENERALES

A lo largo de este apartado se va a introducir al lector en el tema del que va a tratar el resto del proyecto. Para ello, se realiza una explicación de los términos más importantes, así como una descripción de las tecnologías principales que van a tenerse en cuenta de cara al desarrollo del proyecto.

2.1 INTRODUCCIÓN A LA PREDICCIÓN Y SERIES DE TIEMPO

2.1.1 PREDICCIÓN

En el mundo de los negocios y en distintas situaciones administrativas en general, resulta muy útil poder hacer predicciones de las diferentes circunstancias financieras que vayan a ocurrir de forma que se reduzca lo máximo posible la incertidumbre ante ellas.

Los **métodos de predicción** tienen como objetivo ayudar a realizar la toma de decisiones en este ámbito, de forma que, haciendo uso de situaciones pasadas puedan hacer una estimación de lo que va a pasar en el futuro. Existen dos métodos principales de predicción, cualitativos y cuantitativos.

Los **modelos cuantitativos** se ayudan de información pasada que pueden transformar en datos numéricos asumiendo su continuidad. Dentro de los modelos cuantitativos, destacan los modelos explicativos y las series temporales.

Los **modelos explicativos** buscan realizar una descripción de por qué y cómo funciona una cosa. Para ello, hacen uso de otras variables que influyen en ella de forma que pueden ver cómo le afectan. Siempre hay una parte impredecible, representada por el error.

Las **series temporales** son un conjunto de observaciones de una variable ordenadas en el tiempo, con una frecuencia determinada. No buscan saber cuáles son los factores que

afectan al comportamiento de una variable. Analizando los distintos valores pasados y el patrón que siguen, predicen situaciones futuras.

Por otra parte, los **modelos cualitativos** suponen una mayor complejidad. Son útiles para ayudar a realizar el análisis y modelo cuantitativos.

Como ejemplo, es interesante analizar la evolución del precio del oro y del valor del Ibex 35 en los últimos 5 años, donde cabe destacar el contraste entre ambas a partir del primer trimestre de 2020 debido a la pandemia causada por el coronavirus.

En los siguientes gráficos se aprecia como, al comenzar la pandemia, aumenta la demanda del oro (Figura 2.1) y, por tanto, su precio se eleva exponencialmente mientras que el valor del Ibex 35 cae drásticamente desde entonces (Figura 2.2).



Figura 2.1 Precio del oro (€/onza) en los últimos 5 años (Inversoro.es, 2020)



Figura 2.2 Valor del IBEX 35 en los últimos 5 años (Bolsamanía, 2020)

2.1.2 DESCOMPOSICIÓN EN SERIES TEMPORALES

La descomposición de series temporales consiste en deshacer una serie temporal en varias partes, de forma que cada una de ellas represente una categoría subyacente del patrón principal que ha seguido la serie. Las distintas partes son:

- Y_t : valor de la serie de datos a lo largo del periodo t .
- S_t : componente estacional a lo largo del periodo t .
- T_t : componente de tendencia en el periodo t (existe una tendencia cuando hay una dirección persistente de aumento o disminución de los datos).
- E_t : componente irregular (“ruido”) durante el periodo t .

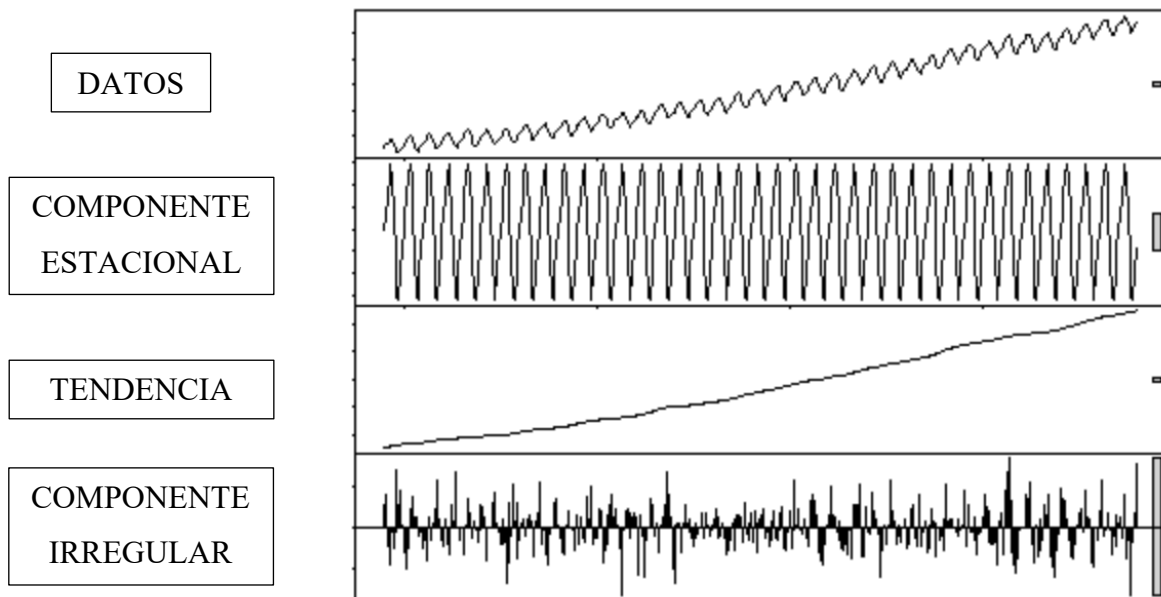


Figura 2.3 Descomposición de series temporales (Vidal, 2016)

Una serie de tiempo con un modelo de descomposición se representa, de forma aditiva, mediante la siguiente ecuación que se muestra.

$$Y_t = S_t + T_t + E_t$$

Ecuación 2.1 Descomposición de series temporales

Podemos realizar un ajuste estacional eliminando S_t de la fórmula anterior, de forma que eliminamos sus efectos de la serie temporal que muestra variaciones debidas a la estacionalidad.

2.1.3 PRECISIÓN DEL PRONÓSTICO

La precisión del pronóstico es el grado de cercanía entre la expresión de una cantidad y el valor real de esa cantidad. Este nivel de exactitud nos puede dar una idea de cómo de adecuado es un método ante un conjunto de datos.

Medidas estadísticas estándar

El **error de predicción** es la diferencia entre la observación real para un periodo t y la predicción para ese mismo periodo.

Como el error medio no es muy preciso porque los valores positivos se cancelan con los negativos, definimos lo siguiente.

El **Error Cuadrático Medio** (MSE, en inglés) de un estimador mide el promedio de los errores al cuadrado, es decir, la diferencia entre el estimador y lo que estima. Cuanto mayor sea el valor, peor será el modelo.

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2$$

Ecuación 2.2 Error Cuadrático Medio (MSE)

Siendo $\hat{Y}_i$ el valor estimado en $t = i$ e Y_i el valor observado en ese mismo momento.

El **Error Absoluto Medio** (MAE, en inglés) calcula el error como un promedio de diferencias absolutas entre los valores objetivo y las predicciones.

$$MAE = \frac{\sum_{i=1}^n |\hat{Y}_i - Y_i|}{n} = \frac{\sum_{i=1}^n |e_i|}{n}$$

Ecuación 2.3 Error Absoluto Medio (MAE)

Siendo, igual que antes, $\hat{Y}_i$ el valor estimado en $t = i$, Y_i el valor observado en ese mismo momento y e_i el error estimado en cada momento.

Con estas dos medidas no es posible realizar una comparación con otras series temporales. Para ello, se definen otros conceptos.

El **Error Medio Porcentual** (MPE, en inglés) es el promedio calculado de los errores porcentuales por los que los pronósticos de un modelo difieren de los valores reales.

El **Error Absoluto Medio Porcentual** (MAPE, en inglés) es el promedio del error absoluto.

2.1.4 COMPARACIÓN DE MÉTODOS DE PREDICCIÓN

Existen métodos sencillos para comparar los distintos métodos de predicción.

- **Naive Forecast 1** (NF1) utiliza la observación más reciente como predicción para el siguiente periodo.
- **Naive Forecast 2** (NF2) elimina la estacionalidad en la serie original.

La comparación entre los métodos más complejos se realiza al calcular MPE y MAPE de los métodos Naive y de los métodos complejos que queremos estudiar, estudiándolos posteriormente.

Estadístico U de Theil

El **coeficiente U de Theil**, que se utiliza como medida de exactitud del pronóstico, tiene en cuenta los posibles errores que puedan surgir y realiza una comparación entre un modelo o método de predicción y un modelo Naive.

$$U = \sqrt{\frac{\sum_{i=1}^{n-1} \left(\frac{\hat{Y}_{i+1} - Y_{i+1}}{Y_i} \right)^2}{\sum_{i=1}^{n-1} \left(\frac{\hat{Y}_{i+1} - Y_i}{Y_i} \right)^2}}$$

Ecuación 2.4 Coeficiente U de Theil

Siendo Y_t el valor actual en t , n el número de observaciones e $\hat{Y}_t$ la predicción.

- Si $U = 1$ el método Naive es igual de adecuado que el que estamos usando.
- Si $U < 1$ el método de predicción es mejor que el Naive y sus pronósticos son fiables.
- Si $U > 1$ el método de predicción no es el más útil y el Naive es mejor.

Función de autocorrelación (ACF) del error de predicción

Otra forma de analizar el error es mediante la **Función de Autocorrelación (ACF)**, que mide la relación entre los valores de una variable en una serie de tiempo y describe lo que tiende a suceder a un valor si ocurre un cambio en el otro.

$$r_k = \frac{\sum_{i=1}^{n-k} (Y_i - \hat{Y})(Y_{i+k} - \hat{Y})}{\sum_{i=1}^n (Y_i - \hat{Y})^2}$$

Ecuación 2.5 Función de Autocorrelación (ACF)

El valor de la ACF debe oscilar entre +1 y -1. Si el valor es cero significa que no existe una relación entre los valores de la variable a lo largo del tiempo. Si el valor sobrepasa los límites, indica que hay información en las series que la predicción no está teniendo en cuenta.

2.2 REGRESIÓN

2.2.1 REGRESIÓN SIMPLE

La **regresión simple** consiste en predecir el comportamiento de una variable endógena Y a partir de la información que tenemos de una variable explicativa X .

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

Ecuación 2.6 Regresión simple

ε representa el error y β_0 y β_1 son los parámetros de la ecuación. El primero expresa el término independiente y es un término de ajuste que no tiene ningún significado; el segundo representa, en este caso, la pendiente, y nos indica cuánto incrementaría Y al incrementar X en una unidad.

Estimación por el método de Mínimos Cuadrados Ordinarios. Coeficiente de correlación (r)

Esta **estimación por Mínimos Cuadrados Ordinarios** consiste en estimar un valor para los β de forma que los residuos tomen el mínimo valor posible, es decir, que la diferencia entre lo que observamos y lo que predecimos sea mínima.

Para ello, buscamos minimizar la suma de los cuadrados de los residuos, es decir, hacer mínimo el valor siguiente:

$$s = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

Ecuación 2.7 Suma de los cuadrados de los residuos

donde Y_i es el valor observado, $\hat{Y}_i$ es el valor estimado por el modelo y n el número de observaciones.

El **Coeficiente de Correlación “r”** es una medida de la dependencia lineal entre dos variables aleatorias cuantitativas.

- Si $r = 1$, existe correlación positiva perfecta, es decir, hay dependencia total entre las dos variables.
- Si $0 < r < 1$, existe correlación positiva.
- Si $r = 0$ no existe correlación lineal. Pueden existir relaciones no lineales.
- Si $-1 < r < 0$, existe correlación negativa.
- Si $r = -1$, existe correlación negativa perfecta, lo cual implica dependencia total inversa entre las dos variables.

El **Coeficiente de Determinación “ R^2 ”** indica la proporción de variabilidad de la Y explicada por la variabilidad de la X . Cuanto mayor es el valor de R^2 mejor es el ajuste que se está realizando.

Contraste de significación conjunta (Contraste F)

Este contraste permite conocer si el modelo que se estudia, en conjunto, es significativo o no, es decir si existe realmente una relación entre las variables.

Cuanto mayor sea F mayor será la relación entre variables.

$$F = \frac{\frac{R^2}{k-1}}{\frac{1-R^2}{n-k}}$$

Ecuación 2.8 Contraste F de significación conjunta

Donde k es el número de parámetros y n es el número de observaciones del modelo.

Contraste de significación individual (Contraste t)

En este caso, se contrasta la hipótesis de si los β asociados a cada variable pueden valer 0 o no, es decir, comprobamos si son individualmente significativos en el modelo.

Cuanto mayor sea el valor de t más significativo será el parámetro.

$$t = \frac{b_i}{S_{b_i}}$$

Ecuación 2.9 Contraste t de significación individual

Donde b_i es el valor estimado del parámetro y S_{b_i} es la desviación típica de la estimación del parámetro.

2.2.2 REGRESIÓN MÚLTIPLE

La **regresión múltiple** consiste en predecir el comportamiento de una variable endógena Y a partir de la información que tenemos de un conjunto de variables explicativas X.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon$$

Ecuación 2.10 Regresión múltiple

El valor de los parámetros lo podemos estimar, igual que en regresión simple, con el método de Mínimos Cuadrados Ordinarios.

Además, podemos hacer uso del coeficiente de determinación “ R^2 ”, para saber la proporción de variabilidad de la Y explicada por la variabilidad de la X .

R^2 será mayor cuantas más variables explicativas tenga el modelo. Cuanto mayor es el valor de R^2 mejor es el ajuste que estamos realizando, sin embargo, este coeficiente solo se puede utilizar para comparar modelos en el caso en el que los modelos estén anidados. Para ello, utilizamos el R^2 corregido.

$$R^2 \text{ corregido} = 1 - (1 - R^2) * \frac{n - 1}{n - k}$$

Ecuación 2.11 R^2 corregido

Donde k es el número de parámetros y n es el número de observaciones del modelo.

De esta forma, se penaliza al modelo con mayor número de variables para que se puedan comparar estando los dos modelos en igualdad de condiciones.

Contraste de significación conjunta (Contraste F)

Se utiliza la misma fórmula que en el caso de regresión simple para saber si el modelo en conjunto que se está utilizando es adecuado o no, pero no se indica la importancia de cada una de las variables.

$$F = \frac{\frac{R^2}{k - 1}}{\frac{1 - R^2}{n - k}}$$

Ecuación 2.12 Contraste F de significación conjunta

Donde k es el número de parámetros y n es el número de observaciones del modelo.

Contraste de significación individual (Contraste t)

Al igual que en regresión simple, permite saber la importancia de cada una de las variables en el modelo. Para ello, se realiza la siguiente operación.

$$t = \frac{b_i}{S_{b_i}}$$

Ecuación 2.13 Contraste t de significación individual

Donde b_i es el valor estimado del parámetro y S_{b_i} es la desviación típica de la estimación del parámetro.

Cabe destacar que la estabilidad de los coeficientes de la regresión depende de la relación entre las variables explicativas, de forma que cuanto más relacionadas estén dos variables, más inestables serán sus coeficientes.

Además, los coeficientes pueden estar interrelacionados por lo que para estudiar uno de ellos será necesario tener en cuenta los demás.

Regresión con series de tiempo

Para detectar la presencia de autocorrelación de los residuos que puede surgir a la hora de predecir series temporales podemos utilizar el **Estadístico de Durbin-Watson**.

$$d = \frac{\sum_{i=2}^T (e_i - e_{i+1})^2}{\sum_{i=1}^T e_i^2}$$

Ecuación 2.14 Estadístico de Durbin-Watson

Siendo e_i el residuo asociado a la observación en t y T el número de observaciones.

- Si $2 < d < 4$ hay autocorrelación positiva.
- Si $d = 2$ existe dependencia de los errores.
- Si $0 < d < 2$ hay autocorrelación negativa.

2.3 OTROS MÉTODOS DE PREDICCIÓN

2.3.1 MÉTODOS DE PROMEDIO

Se basan en el uso de datos pasados para crear un método de predicción para el futuro. Algunos de ellos se explican a continuación.

La **media** consiste en hacer una media de los datos pasados para predecir el valor del siguiente periodo.

La **media móvil** (MA, en inglés) consiste en escoger un conjunto de datos creando series de promedios. Utiliza una serie de datos pasados y especifica cuántas observaciones se van a estudiar, de forma que reduce la influencia de datos antiguos. Cada vez que hay disponible una nueva observación, la más antigua se elimina de la media, consiguiendo tener continuamente un valor actualizado.

Ninguno de ellos es adecuado para series temporales con un patrón de tendencia o estacionalidad.

2.3.2 MÉTODOS DE SUAVIZAMIENTO EXPONENCIAL

Se utilizan, en lugar de los métodos de promedio anteriores, porque dan más importancia a las observaciones obtenidas más recientemente que a las más antiguas.

Suavizamiento Exponencial Simple (SES)

Utiliza la predicción del periodo anterior y la ajusta al actual haciendo uso del error de predicción.

$$\hat{Y}_{T+1|T} = \alpha Y_T + \alpha(1 - \alpha)Y_{T-1} + \alpha(1 - \alpha)^2 Y_{T-2} + \dots$$

Ecuación 2.15 Suavizamiento Exponencial Simple (SES)

α es un parámetro entre 0 y 1 que, cuanto mayor sea su valor, mayor peso estará dando a las observaciones más recientes.

Suavizamiento Exponencial Simple (SES), un enfoque adaptativo

Permite la modificación del parámetro durante el proceso cuando el patrón o tendencia de los datos cambie. De esta forma se reduce la posibilidad de cometer grandes errores si la serie tiene distintas tendencias o estacionalidad.

Método lineal de Holt

Se caracteriza porque puede adaptarse fácilmente a cambios de tendencia. Es un tipo de método de SES.

Por ejemplo, es un método utilizado por su sencillez en muchas compañías para pronosticar la demanda a corto plazo cuando los datos de venta contienen tendencias y patrones estacionales de un modo subyacente.

$$\hat{Y}_{t+h|t} = \ell_t + hb_t$$

Ecuación 2.16 Método lineal de Holt – Ecuación de predicción

$$\ell_t = \alpha Y_t + (1 - \alpha)(\ell_{t-1} + b_{t-1})$$

Ecuación 2.17 Método lineal de Holt - Ecuación de nivel

$$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)b_{t-1}$$

Ecuación 2.18 Método lineal de Holt - Ecuación de tendencia

ℓ_t es la estimación de nivel de la serie en t , b_t es la estimación de pendiente (tendencia) de la serie en t y α y β^* representan los parámetros de nivel y tendencia, respectivamente, y tienen valores entre 0 y 1.

Método de tendencia y estacionalidad de Holt-Winter

Es el único método útil cuando la serie de datos presenta estacionalidad.

Se establecen 3 ecuaciones (para nivel, tendencia y estacionalidad) y 3 constantes. Podemos distinguir dos tipos de estacionalidad: multiplicativa y aditiva.

2.4 MÉTODOS DE PREDICCIÓN NO LINEALES

2.4.1 REDES NEURONALES

Las **Redes Neuronales Artificiales** consisten en un conjunto de unidades (neuronas artificiales) conectadas entre sí para transmitirse señales. Son modelos no paramétricos basados en datos con escasas suposiciones previas de los procesos a partir de los cuales se generan los datos.

El objetivo del modelo es aprender mientras se modifica automáticamente a si mismo, llegando a poder realizar ciertas tareas complejas que no podrían ser realizadas mediante la clásica programación basada en reglas. De esta forma, se pueden automatizar funciones que en un principio solo podrían ser realizadas por personas. Su funcionamiento se basa en que las neuronas reciben un conjunto de valores, lo que se denomina datos de entrada, los cuales atraviesan las distintas capas, y se irán modificando hasta llegar a la salida, es decir, a la solución del problema planteado (Figura 2.4).

Se conectan las entradas a una o más neuronas, que transmitirán la información a través de otras neuronas en las capas ocultas hasta llegar a la salida.

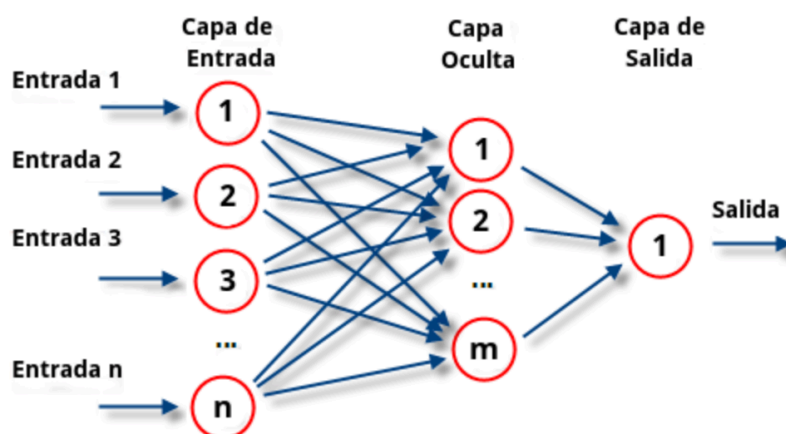


Figura 2.4 Esquema de una red neuronal (Atria Innovation, 2019)

Una red neuronal se caracteriza porque puede recordar experiencias y hacer que estén disponibles para un uso futuro.

Se pueden distinguir dos tipos de señales:

- Señales de función: atraviesan la red desde la entrada hasta la salida.
- Señales de error: atraviesan la red desde la salida hasta la entrada.

Una red neuronal procesa información de una capa a la siguiente mediante la función de activación.

La función del nodo de salida es la siguiente:

$$y = \sum_{i=1}^n w_i * x_i + \theta$$

Ecuación 2.19 Nodo de salida de una red neuronal

donde n es el número de entradas, w es el peso de cada entrada, las x son los nodos de entrada y θ representa el sesgo.

Para utilizar redes neuronales es necesario decidir cuántas capas ocultas y nodos se van a utilizar y probar la red estimando sus sesgo y pesos, normalmente buscando minimizar el valor de los Mínimos Cuadrados Ordinarios de los errores.

El proceso de aprendizaje de una neurona consiste en:

1. Normalizar.
2. Elegir un valor inicial para su sesgo y pesos.
3. Usar el primer ejemplo del conjunto de datos estimado e introducirlo en la red.
4. Evaluar el error final.
5. Modificar sesgo y pesos.
6. Repetir el proceso para el resto de las series de datos estimadas.
7. Evaluar el error total y repetir todo el proceso hasta que se minimice.

Por tanto, para llevar a cabo el proceso los datos se deben dividir en dos muestras, la primera para estimar los parámetros y la segunda para realizar las predicciones.

No es tarea fácil establecer la arquitectura de una red neuronal. Normalmente una capa oculta es suficiente. A continuación, habrá que decidir el número de neuronas de la capa y su relación con el número de nodos de entrada.

Las redes con muchos nodos ocultos no son muy útiles para realizar predicciones.

2.4.2 MODELOS DE HETEROCEDASTICIDAD CONDICIONAL

Los **Modelos Autorregresivos de Heterocedasticidad Condicional** se emplean comúnmente en el modelado de series de tiempo financieras que presentan agrupaciones de volatilidad variables en el tiempo, es decir, periodos de oscilaciones entremezclados con periodos de relativa calma.

Su objetivo en este ámbito es modelar la volatilidad del rendimiento de un activo para poder mejorar la estimación de los parámetros y la exactitud de la predicción del intervalo. La volatilidad es la desviación estándar del rendimiento de un activo subyacente, en un horizonte temporal específico.

Conceptos básicos: estructura del modelo

Media condicional:

$$\mu_t = E(r_t | \Omega_{t-1})$$

Ecuación 2.20 Media condicional

Varianza condicional:

$$\sigma_t^2 = E((r_t - \mu_t)^2 | \Omega_{t-1})$$

Ecuación 2.21 Varianza condicional

donde r_t son los rendimientos de los activos en t y Ω_{t-1} representa la información disponible en $t - 1$ que normalmente consiste en todas las funciones lineales de los rendimientos pasados.

Además, considerando que la ecuación de la media sigue un modelo ARMA (p,q), tenemos la siguiente ecuación:

$$r_t = \mu_t + a_t$$
$$\mu_t = \phi_0 + \sum_{i=1}^p \phi_i \mu_{t-i} - \sum_{i=1}^q \theta_i a_{t-i} + \sum_{i=1}^k \beta_i x_{i,t}$$

Ecuación 2.22 Rendimientos de los activos r_t

combinando esta ecuación junto con las anteriores, se obtiene la ecuación de volatilidad.

$$\sigma_t^2 = \text{Var}(r_t | \Omega_{t-1}) = \text{Var}(a_t | \Omega_{t-1})$$

Ecuación 2.23 Ecuación de la volatilidad de r_t

El modelo para σ_t^2 representa la ecuación de la volatilidad de r_t .

Los modelos de heterocedasticidad condicional tratan la forma en la que σ_t^2 evoluciona a lo largo del tiempo, siendo este factor el que distingue un modelo de volatilidad de otro.

Los modelos pueden clasificarse en dos tipos, los que usan una función generalizada para describir la evolución de σ_t^2 (modelos GARCH) y los que usan ecuaciones estocásticas (modelos de volatilidad estocástica).

Para construir estos modelos se deben seguir los siguientes pasos:

1. Especificar la ecuación de la media buscando la dependencia serial de los datos.
2. Usar los residuos de la ecuación de la media para contrastar la existencia de los efectos de ARCH.
3. Especificar un modelo de volatilidad en el que los efectos de ARCH sean significativos y llevar a cabo una estimación conjunta de ecuaciones para media y volatilidad.

4. Comprobar el modelo ajustado y hacer algún ajuste si es posible.

Para contrastar los efectos de ARCH, tomamos a_t como residuo de la ecuación de la media en t . Los cuadrados de los residuos $(a_t)^2$ se utilizan para contrastar la heterocedasticidad condicional, es decir, si existen efectos ARCH.

Modelo ARCH

La idea básica de los modelos ARCH es que el a_t del rendimiento de un activo está serialmente incorrelado pero es dependiente y la dependencia de a_t puede ser descrita mediante una función cuadrática de sus retardos. El modelo ARCH (m) supone:

$$a_t = \sigma_t \varepsilon_t$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^m \alpha_i a_{t-i}^2$$

$$\text{donde } \varepsilon_t \text{ iid} \sim N(0,1), \alpha_0 > 0, \alpha_i \geq 0, \sum_{i=1}^m \alpha_i < 1$$

Ecuación 2.24 Modelo ARCH

Algunas características del modelo son que la esperanza incondicional de a_t es cero o que shocks largos suelen ir seguidos de otro shock largo. El orden m del modelo determina cuánto va a condicionar el shock las siguientes varianzas.

Algunas desventajas son que se asume que shocks positivos y negativos tienen el mismo impacto en la volatilidad. ARCH es muy restrictivo con sus parámetros, no ofrece información adicional para saber la fuente de las variaciones y suelen “sobre-predecir” la volatilidad.

Modelo GARCH

En ocasiones, el modelo ARCH requiere otros parámetros para describir la volatilidad del proceso del rendimiento de un activo. Para reducir estas estimaciones existe una extensión del modelo ARCH, el modelo ARCH generalizado (GARCH), de forma que a_t sigue un GARCH (m,s) si:

$$a_t = \sigma_t \varepsilon_t$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^m \alpha_i a_{t-i}^2 + \sum_{j=1}^s \beta_j \sigma_{t-j}^2$$

$$\text{donde } \varepsilon_t \text{ iid} \sim N(0,1), \alpha_0 > 0, \alpha_i \geq 0, \beta_j \geq 0, \sum_{i=1}^{\max(m,s)} (\alpha_i + \beta_i) < 1$$

Ecuación 2.25 Modelo GARCH

Esta última restricción implica que la varianza incondicional de a_t es finita. Consideramos α_i parámetros ARCH y β_j parámetros GARCH.

Modelo IGARCH

El modelo GARCH integrado es una versión restrictiva de GARCH, donde los parámetros suman 1 y tiene una raíz unitaria.

Modelo GARCH-M

El modelo GARCH-M añade un término de heterocedasticidad en la ecuación de la media.

Modelo EGARCH

El modelo GARCH exponencial es otra forma de modelo GARCH que surgió debido a que los coeficientes estimados del modelo lineal ARCH tenían muchas veces un resultado negativo.

La ventaja de este modelo es que las restricciones no implican asegurar que la varianza condicional sea positiva.

El modelo EGARCH (p,q) se define por:

$$\log \sigma_t^2 = \omega + \sum_{k=1}^q \beta_k g(Z_{t-k}) + \sum_{k=1}^p \alpha_k \log \sigma_{t-k}^2$$

$$\text{con } g(Z_t) = \theta Z_t + \lambda(|Z_t| - E|Z_t|).$$

Ecuación 2.26 Modelo EGARCH

Donde σ_t^2 es la varianza condicional, ω , β , α , θ y λ son los coeficientes. Z_t puede ser la variable normal estándar o venir de una distribución de error generalizada.

Modelo TGARCH

El modelo Umbral (Threshold, en inglés) GARCH es otro modelo de volatilidad utilizado para manejar los efectos del apalancamiento e implica lo siguiente:

$$\sigma_t = K + \delta \sigma_{t-1} + \alpha_1^+ \epsilon_{t-1}^+ + \alpha_1^- \epsilon_{t-1}^-$$

Ecuación 2.27 Modelo TGARCH

donde $\epsilon_{t-1}^+ = \epsilon_{t-1}$ si $\epsilon_{t-1} > 0$ y $\epsilon_{t-1}^+ = 0$ si $\epsilon_{t-1} \leq 0$. De la misma forma, $\epsilon_{t-1}^- = \epsilon_{t-1}$ si $\epsilon_{t-1} \leq 0$ y $\epsilon_{t-1}^- = 0$ si $\epsilon_{t-1} > 0$.

El modelo utiliza el cero como umbral para separar los impactos de shocks pasados.

Capítulo 3. MODELOS DE CAMBIO DE RÉGIMEN CON DATOS CLÁSICOS

Un cambio de régimen tiene lugar en una serie de tiempo cuando esta parece ser estacionaria a lo largo de diferentes intervalos, pero el mecanismo de generación de sus datos tiene cambios repentinos en ciertos instantes temporales. El cambio de régimen ocurre cuando la apariencia visual de los datos cambia de forma más o menos abrupta para alcanzar de nuevo un aspecto homogéneo.

3.1 *MODELOS DE MARKOV*

La característica principal de los modelos de cambio de régimen, según (Frankie, 2012) es que contienen una serie de estados o regímenes, representados por distintos mecanismos que generan los datos, entre los cuales el sistema cambia continuamente. A diferencia de los procesos estacionarios, en los que las variaciones se producen de una manera gradual, en el caso de los modelos de cambio de régimen, que siguen procesos no estacionarios, las transiciones entre los regímenes se producen de repente o en un periodo de tiempo muy breve.

Existen diferentes formas de modelar una serie de tiempo $\{X_t\}$ que alterna distintos regímenes. Estos cambios de estado se controlan con una variable de conmutación Q_t , que es una serie de tiempo en si misma que asume un número finito de valores. A continuación, se muestra un modelo generalizado en el caso univariante:

$$X_t = F(X_{t-1}, \dots, X_{t-m}, Q_t, \epsilon_t; \theta)$$

Ecuación 3.1 Modelo generalizado de modelado de una serie de tiempo univariante

donde ϵ_t son las innovaciones independientes e idénticamente distribuidas, con distribución conocida y θ es el vector de parámetros del modelo. Se asume que la distribución de X_t dados los datos previos del proceso hasta el momento $t - 1$, depende explícitamente de las m observaciones anteriores.

De acuerdo con este modelo, hay dos enfoques diferentes según la dependencia entre la variable de conmutación y los datos de interés.

El primero de estos enfoques es el de los **modelos de cambio por observación**, que asumen que, dados los datos, la variable de conmutación Q_t no depende de los valores de la variable de conmutación pasados Q_s con $s < t$.

$$\mathbb{P}(Q_t = k, \epsilon_t \in B | Q_{t-1}, \dots, Q_0, X_{t-1}, \dots, X_0) = \mathbb{P}(Q_t = k | X_{t-1}, \dots, X_0) \mathbb{P}(\epsilon_t \in B)$$

Ecuación 3.2 Modelo de cambio por observación

Este tipo de modelos incluyen los modelos TAR, que se explicarán más adelante.

Por otro lado, el segundo enfoque es el de los **modelos de Markov ocultos**. En este caso, las variables de conmutación Q_t forman una cadena de Markov de longitud finita y dependiendo la distribución de Q_t , dada hasta el momento $t - 1$, únicamente de Q_{t-1} y no de los datos $X_{t-1}, \dots, X_0$.

Los modelos de cambios de régimen con series de datos temporales no solo han generado un gran interés en la estadística en los últimos años, sino que también están teniendo una especial importancia en sus aplicaciones en Machine Learning.

3.2 MODELOS AUTORREGRESIVOS DE UMBRAL, TAR

Los **Modelos Autorregresivos de Umbral**, (TAR, del inglés Threshold Autoregression) son modelos de autorregresión que permiten la existencia de dos o más ramas gobernadas por una variable umbral, es decir, permiten utilizar diferentes modelos

autorregresivos separados por una variable umbral. Esto hace posible un comportamiento asimétrico que no puede ser explicado por un solo modelo ARMA.

Los modelos TAR tienen una gran importancia y aplicación en el mundo de la economía. Existe una tendencia entre los economistas a pensar que las aplicaciones deberían usar modelos lineales a no ser que haya una evidencia suficiente para apoyar a un determinado modelo no lineal. Es decir, se tiende a probar la linealidad en vez de la alternativa de los modelos TAR.

Las series de tiempo relacionadas con la economía suelen ser no-estacionarias. Un objetivo de gran importancia en la economía aplicada es poder distinguir las series de tiempo no lineales de las no estacionarias.

Es muy común hacer uso de modelos TAR en la economía empírica. Por ejemplo, para estudiar el aumento de la producción. (Beaundry & Koop, 1993) hacen uso de modelos TAR para modelar o calcular la producción añadida, teniendo en cuenta las tasas de crecimiento del PNB (Producto Nacional Bruto). Otro ejemplo en este mismo ámbito sería el de (Enders, Falk, & Siklos, 2007) que emplean un modelo TAR para estimar el PIB (Producto Interior Bruto) de Estados Unidos y mediante las estimaciones diseñan métodos para construir intervalos de confianza para los parámetros del modelo.

Por otro lado, los modelos TAR son frecuentes a la hora de predecir. Una de sus principales aplicaciones es la predicción de variables económicas.

Muchos autores también han hecho uso de modelos TAR para estudiar los precios del mercado. Normalmente, los precios de los bienes relacionados deberían cambiar simultáneamente, sin embargo, debido a la presencia de los costes de transacción, se puede producir un efecto anormal donde solo las desviaciones producidas por encima de un umbral tendrán un efecto en los cambios del precio.

3.3 MODELOS AUTORREGRESIVOS DE MEDIA MÓVIL

3.3.1 CONCEPTOS PREVIOS

Se dice que una serie de tiempo sigue un **Modelo de Ruido Blanco** si sus valores en dos periodos distintos están incorrelados. Si la serie sigue una distribución normal de media 0 y varianza μ , se dice que sigue un modelo de ruido blanco y Gaussiano. Para una serie de ruido blanco, la ACF debe tener un valor próximo a cero.

Reconocer la estacionalidad de una serie de tiempo es importante al elegir el método de predicción. La **estacionalidad** es un patrón que se repite a lo largo de unos intervalos de tiempo fijos. Se puede encontrar si hay una gran autocorrelación en el retraso estacional.

3.4 CONTRASTE DE PORTAMANTEAU Y PACF

El **Contraste de Portmanteau** estudia si hay alguna autocorrelación distinta de cero en vez de estudiar cada autocorrelación por separado.

Las **Funciones de Autocorrelación Parcial** (PACF, del inglés) se utilizan para medir el grado de asociación entre una serie de tiempo estacional con sus propios valores retardados.

Dada una serie de tiempo z_t , la autocorrelación parcial del retraso k , denotado $\alpha(k)$ es la autocorrelación entre z_t y z_{t+k} , quitando los efectos de otros retrasos de tiempo a través de $k - 1$.

$$\alpha(1) = \text{Cor}(z_t, z_{t+1})$$

$$\alpha(k) = \text{Cor}(z_{t+k} - P_{t,k}(z_{t+k}), z_t - P_{t,k}(z_t)) \text{ para } k \geq 2.$$

Ecuación 3.3 Función de Autocorrelación Parcial (PACF)

Con $P_{t,k}(x)$ como la proyección de x en el espacio abarcado por $z_{t+1} \dots z_{t+k-1}$.

3.5 ESTACIONARIEDAD Y SERIES DE DATOS TEMPORALES

La **estacionariedad** implica que los datos de una serie no varían con respecto al tiempo, es decir, fluctúan alrededor de una media constante y mantienen una varianza y una autocorrelación constantes a lo largo del tiempo.

Las tendencias y patrones no estacionarios provocan grandes autocorrelaciones y es importante eliminarlos antes de construir un modelo de serie temporal.

Una de las formas de eliminar la no estacionariedad es mediante el **método de diferenciación**, que consiste en crear una serie a partir de la diferencia entre observaciones consecutivas de la serie original.

$$y'_t = y_t - y_{t-1}$$

Ecuación 3.4 Método de diferenciación

Este proceso se repite hasta que la serie siga un modelo de ruido blanco ($y_t - y_{t-1} = \varepsilon_t$).

Si los datos tienen un patrón estacional y no son estacionarios, se pueden tomar **diferencias estacionales**, que consisten en el cambio en una observación y la correspondiente observación de la estación anterior.

3.6 MODELO DE CAMINO ALEATORIO

El modelo **Camino Aleatorio**, (Random Walk Model, en inglés) es muy utilizado con datos no estacionarios y consiste en que en cada periodo la variable se compone del valor anterior de la misma y de un término de error definido como ruido blanco ($y_t = y_{t-1} + \varepsilon_t$).

3.7 INTRODUCCIÓN A LOS MODELOS ARIMA DE SERIES TEMPORALES DE DATOS

El **Modelo Autorregresivo Integrado de Media Móvil (ARIMA)** es un modelo que utiliza variaciones y regresiones de datos estadísticos con el fin de encontrar patrones para una predicción hacia el futuro. Las estimaciones futuras se explican por datos del pasado y no por variables independientes.

Un modelo ARIMA(p,d,q) es la combinación de un modelo AR (autorregresivo) y un modelo MA (medias móviles) para datos no estacionarios. Para datos estacionarios, utilizamos el modelo ARMA(p,q).

El modelo ARIMA depende de tres parámetros: p es el orden de la parte autorregresiva, d es el grado de la primera diferenciación en el modelo y q es el orden de la parte de la media móvil.

3.7.1 MODELO AUTORREGRESIVO SIMPLE (AR)

Un **Modelo Autorregresivo Simple (AR)** es una representación de un proceso aleatorio en el que la variable de interés depende de sus observaciones pasadas.

La notación AR(p) representa un modelo autorregresivo de orden p, que se define como:

$$X_t = c + \sum_{i=1}^p \varphi_i X_{t-i} + \varepsilon_t$$

Ecuación 3.5 Modelo Autorregresivo Simple, AR(p)

siendo φ_i los parámetros del modelo, c una constante y ε_t el ruido blanco.

En la práctica, el orden de un modelo ARIMA es desconocido y se debe estimar mirando ACF o PACF.

3.7.2 MODELO SIMPLE DE MEDIA MÓVIL (MA)

La **Media Móvil Simple (MA)** es la media aritmética de los n datos anteriores. Cuando mayor sea n , mayor será la influencia de los datos antiguos para la predicción. Para obtener la predicción nos basamos en términos de error previos y en el actual.

La notación $MA(q)$ se refiere al modelo de media móvil de orden q , que se define como:

$$X_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$$

Ecuación 3.6 Modelo Simple de Media Móvil, $MA(q)$

siendo μ la media de la serie, θ los parámetros del modelo y ε_t los valores de ruido blanco.

3.7.3 MODELOS AUTORREGRESIVOS DE MEDIA MÓVIL (ARMA)

Los modelos ARMA surgen de la combinación de AR y MA. Se refiere a un modelo con p términos autorregresivos y q términos de media móvil y que, para obtener los valores a predecir, se basa en valores previos de la variable que queremos predecir y en los valores de error previos. Son utilizables con series estacionarias en media y varianza. La ecuación por la que se representan es la siguiente:

$$X_t = \varepsilon_t + \sum_{i=1}^p \phi_i X_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i}$$

Ecuación 3.7 Modelo Autorregresivo de Media Móvil, $ARMA(p,q)$

siendo ε_t el error, ϕ_i el parámetro de autorregresión y θ_i el parámetro de la media móvil.

3.7.4 MODELOS INTEGRADOS AUTORREGRESIVOS DE MEDIA MÓVIL (ARIMA)

Cuando los modelos ARMA no pueden ser utilizados porque estar ante una serie de datos no estacionarios, se introduce un nuevo grado diferenciador a estos modelos ARMA, con el que se añade la no-estacionariedad, para crear los modelos ARIMA.

$$Y_t = -(\Delta^d Y_t - Y_t) + \phi_0 + \sum_{i=1}^p \phi_i \Delta^d Y_{t-i} - \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \varepsilon_t$$

Ecuación 3.8 Modelo Integrado Autorregresivo de Media Móvil, ARIMA(p,q,d)

Donde d indica las diferencias necesarias para convertir la serie original en estacionaria, $\phi_1, \dots, \phi_p$ son los parámetros pertenecientes a la parte autorregresiva del modelo, $\theta_1, \dots, \theta_q$ son los parámetros pertenecientes a la parte de medias móviles del modelo, ϕ_0 es una constante y ε_t es el término de error. Además, $\Delta Y_t = Y_t - Y_{t-1}$.

3.8 CONTRASTE DE RAÍZ UNITARIA

En estadística, un **Contraste de Raíz Unitaria** (Unit Root Test) comprueba si una variable de serie temporal no es estacionaria y posee una raíz unitaria. La hipótesis nula describe la presencia de una raíz unitaria mientras que la alternativa implica estacionariedad o una gran raíz.

Contraste de Dickey Fuller

El **Contraste de Dickey Fuller** (Dickey Fuller Test) comprueba si una raíz unitaria se encuentra en un modelo autorregresivo o no, de forma que pueden ver si la serie es estacionaria. Un modelo AR(1) simple es el siguiente:

$$y_t = \rho y_{t-1} + u_t$$

Ecuación 3.9 Modelo AR(1)

si $\rho = 1$ la raíz unitaria está presente y el modelo no sería estacionario. El modelo de regresión también se expresa como:

$$\nabla y_t = (\rho - 1)y_{t-1} + u_t = \delta y_{t-1} + u_t$$

Ecuación 3.10 Modelo de regresión

donde ∇ es el operador de primera diferencia. Este modelo se puede estimar y comprobar para la raíz unitaria, lo que es equivalente a comprobar si $\delta = 0$.

Hay distintas versiones del Contraste. Cada una tiene su valor crítico que depende del tamaño de la muestra y en todas ellas, la hipótesis nula es que existe una raíz unitaria.

3.9 IDENTIFICACIÓN DEL MODELO

Hay una serie de pasos que ayudan a decidir qué modelo es el más apropiado para utilizar. Dichos pasos se definen a continuación:

1. Mostrar los datos en una gráfica y buscar observaciones inusuales. Transformar los datos para buscar estacionariedad en la varianza si es necesario.
2. Representar gráficamente ACF y PACF para comprobar si hay estacionariedad.
3. Si no hay estacionariedad, buscarla mediante diferenciación, repitiendo el proceso hasta encontrar la estacionariedad.
4. Cuando se alcance la estacionariedad, volver a mostrar las autocorrelaciones para ver si algún patrón permanece. Pueden ocurrir varias posibilidades:
 - a. Si las autocorrelaciones en los retrasos temporales son distintas de cero y tienen un valor alto, habrá estacionalidad.
 - b. Si no hay autocorrelaciones significativas después del retraso “q”, utilizar un modelo MA(q).
 - c. Si no hay autocorrelaciones parciales significativas después del retraso “p”, utilizar un modelo AR(p).

- d. Si no hay un modelo claro de AR o MA, utilizar una combinación de ambos (ARMA o ARIMA).

3.9.1 ESTIMACIÓN DEL PARÁMETRO

Una vez determinado el orden del modelo ARIMA es necesario determinar los parámetros de AR y MA de la mejor manera posible.

Un posible método es el de Mínimos Cuadrados Ordinarios, pero los modelos que incluyen MA deberán necesitar un proceso iterativo para obtener las estimaciones del parámetro.

Otra forma de estimarlos es mediante el **Método de Máxima Verosimilitud (L)**, que encuentra los valores de los parámetros que hacen mayor la verosimilitud L. Esta estimación también necesita un proceso iterativo.

$$L(\vec{\beta}, \vec{\sigma}) = \frac{1}{\sqrt{(2\pi\sigma^2)r_0 \dots r_T}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{j=1}^T \frac{(y_j - \hat{y}_j)^2}{r_{j-1}} \right\}$$

Ecuación 3.11 Verosimilitud

3.9.2 REVISIÓN DE LA IDENTIFICACIÓN

Una vez estimado el modelo ARIMA es necesario revisar si el modelo elegido puede ser mejorado. Puede ocurrir lo siguiente:

- Algunos de los parámetros estimados pueden ser insignificantes.
- Se puede considerar extender el modelo a un ARIMA mixto.
- Puede haber más de un modelo aceptable y necesitamos un método para saber cuál es el mejor.

3.9.3 COMPROBACIÓN DEL DIAGNÓSTICO

Es necesario verificar que el modelo es adecuado. Para ello, se analizan los residuos y para comprobar si existe todavía algún patrón más. Para que sea un buen modelo, estos residuos deberían ser únicamente ruido blanco.

3.10 REGRESIÓN SEGMENTADA

3.10.1 REGRESIÓN SEGMENTADA. EXPLICACIÓN

La **regresión segmentada** es una forma de análisis de regresión en la que la variable independiente se divide en intervalos, ajustando en cada uno de ellos una línea o curva de datos. La regresión segmentada permite hacer uso de varios modelos lineales, cada uno de ellos en un rango del eje de la variable independiente x , para ajustar un conjunto de datos.

Este tipo de regresión puede aplicarse también en el caso de la regresión múltiple, es decir, con más de una variable independiente, realizando la segmentación de cada una de ellas.

La **regresión segmentada lineal** es aquella en la cual la relación entre la variable dependiente e independiente de cada uno de sus segmentos se obtiene mediante regresión lineal

La regresión segmentada es de gran utilidad en casos en los que la relación entre la variable dependiente y , y la variable independiente x presenta tendencias diferentes en distintos segmentos del eje de las x .

El punto en el que se produce el cambio de un segmento a otro o, lo que es lo mismo, el cambio en la pendiente de la función lineal, se denomina punto de ruptura/cambio (breakpoint). Normalmente, antes de realizar el análisis del modelo estos puntos son desconocidos y deben ser estimados. A continuación, se presenta un ejemplo (Figura 3.1) gráfico de regresión lineal. En el ejemplo, se está analizando la relación entre la fuerza

compresiva (y) de $n = 18$ lotes de hormigón y la proporción de agua mezclada con cemento en ellos (x).

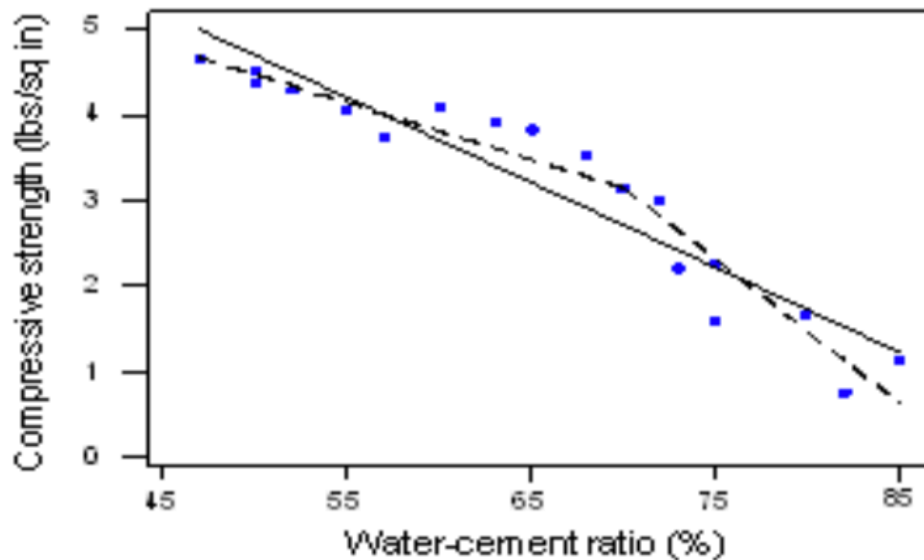


Figura 3.1 Ejemplo de regresión segmentada - Regresión lineal (Wikipedia, 2021)

Por un lado, se observa la forma que tendría la relación entre x e y en caso de utilizar la regresión lineal. A pesar de que en una vista general el modelo se puede considerar aceptable, analizando más detenidamente la gráfica, se observa que el modelo podría ser más claro.

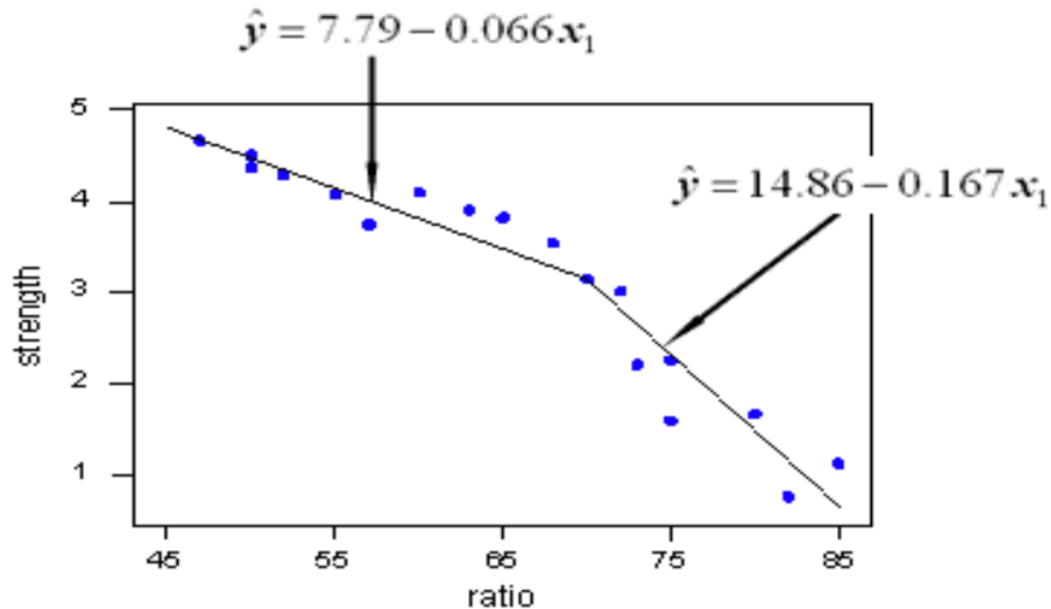


Figura 3.2 Ejemplo de regresión segmentada - Regresión segmentada (Wikipedia, 2021)

Por otro lado, se aprecia el modelo de regresión lineal segmentada, el cual sigue dos tendencias distintas. Cada una de ellas, siendo ambas lineales, es representada por una función lineal diferente y, por tanto, por un segmento. En la imagen se puede observar el punto de ruptura entre ambos segmentos.

Teniendo en cuenta el aspecto que puede tener un modelo de regresión segmentada y, asumiendo que la función sea continua, a pesar de que puede no serlo, un modelo genérico se puede escribir de la siguiente manera.

$$y = a_1 + b_1 * x \quad \text{para } x \leq c$$

$$y = a_2 + b_2 * x \quad \text{para } x > c.$$

Para que la función sea continua, es necesario que en el punto de ruptura (c) ambas ecuaciones sean iguales, por tanto:

$$a_1 + b_1 * c = a_2 + b_2 * c ;$$

$$a_2 = a_1 + c * (b_1 - b_2) ;$$

Reemplazando a_2 en el sistema, se alcanza el sistema siguiente:

$$y = a_1 + b_1 * x \quad \text{para } x \leq c$$

$$y = a_1 + c * (b_1 - b_2) + b_2 * x \quad \text{para } x > c.$$

Ecuación 3.12 Sistema de ecuaciones de regresión segmentada

Donde y es el valor esperado de la variable observada (dependiente) para un cierto valor de x , variable independiente, a_1 y a_2 son las constantes de regresión en sus respectivos segmentos, indicadoras del valor de y cuando $x = 0$ y b_1 y b_2 son los coeficientes de regresión, que indican la inclinación de sus respectivos segmentos en la gráfica.

Podría hacerse uso a continuación del Método de los Mínimos Cuadrados y calcular los coeficientes de correlación R^2 para saber cómo de bueno es el ajuste que se está realizando.

3.10.2 PROCEDIMIENTO DE PRUEBAS

Para determinar el tipo de tendencia existente, se pueden realizar una serie de pruebas estadísticas.

- Significatividad estadística del punto de ruptura: se expresa el punto como una función de los coeficientes de regresión (b_1 y b_2), los promedios Y_1 e Y_2 de los datos y , correspondiendo cada Y a uno de los segmentos, los promedios X_1 y X_2 de los datos x , haciendo uso de las leyes de propagación de errores en adiciones y multiplicaciones para computar el error estándar (SE) del punto de ruptura y una prueba t de Student.
- Significatividad estadística de b_1 y b_2 aplicando la prueba t de Student y el error estándar de ellos.
- Significatividad estadística de la diferencia entre b_1 y b_2 aplicando t de Student y el error estándar de la diferencia.
- Significatividad estadística de la diferencia entre Y_1 e Y_2 mediante t de Student y el error estándar de la diferencia.

Además, se hace también uso del coeficiente de correlación de los datos (R), el coeficiente de determinación (R^2), intervalos de confianza de las funciones de regresión y un análisis de la varianza (ANOVA).

El **coeficiente de determinación** de los datos (R^2), que debe maximizarse bajo las condiciones de las pruebas estadísticas, se escribe como:

$$R^2 = \frac{1 - \sum (y - Y_r)^2}{\sum (y - Y_a)^2}$$

Ecuación 3.13 Coeficiente de determinación, regresión segmentada

donde Y_r es el valor esperado de y e Y_a es el promedio de todos los valores de y .

El valor de R^2 se puede encontrar entre 0, lo que implica que no hay ninguna explicación de la regresión segmentada, y 1, indicando perfecta explicación. En el caso de una regresión lineal sin segmentación, los valores de del coeficiente de determinación (R^2) y el cuadrado del coeficiente de correlación (R) son siempre iguales, sin embargo, en el caso de la regresión segmentada, R debe tener un valor mayor para poder justificar la segmentación.

La optimización del punto de ruptura se alcanza probando una serie de puntos posibles y seleccionando el que presente un valor de R mayor.

3.10.3 RANGO DE NO-EFECTO

En numerosas ocasiones se hace uso de la regresión segmentada para detectar el rango sobre el cual la variable independiente x , no causa un efecto sobre la variable dependiente y , mientras que fuera de ese rango si que existe una relación entre ambas.

Para el análisis del no-efecto se puede hacer uso del método de regresión parcial progresiva sobre el rango afectado, extendiéndolo poco a poco hasta que el coeficiente de regresión (la pendiente de la línea) se vuelve significativamente cero.

3.11 EJEMPLO DE REGRESIÓN SEGMENTADA

A continuación, se va a explicar la regresión segmentada mediante un ejemplo. A partir de una serie de datos de intervalos de los precios de NVAX correspondientes al mes de junio de 2020, se va a analizar si existe un cambio significativo en las pendientes de la primera quincena (A) y segunda quincena (B) del mes.

Ejemplo 1.1

Para empezar, teniendo en cuenta los precios máximos del intervalo, se obtienen a partir de los datos, una serie de resultados.

	A	B
Slope	-0,278441048	2,672744939
n	11	11
SE (regression)	2,77599347	3,689823829
SE (slope)	0,192396444	0,246237105

Tabla 1 Datos 1, regresión segmentada - Ejemplo 1.1 (Elaboración propia)

Donde slope es la pendiente, n es el número de elementos con los que se realizan los cálculos y SE representa al error estándar.

Por otro lado, se obtiene la siguiente tabla de resultados.

Difference	2,951185987
SE (Difference)	0,312488566
t-statistic	9,444140711

df	18
p	0,000000021

Tabla 2 Datos 2, regresión sementada - Ejemplo 1.1 (Elaboración propia)

Donde df es el número de grados de libertad y p es el p-valor del estadístico t calculado previamente.

A partir de los datos obtenidos y teniendo en cuenta el p-valor, siendo que este tiene un valor muy por debajo de 0,05, se obtiene la conclusión de que se va a rechazar la hipótesis nula que dice que ambas pendientes son iguales y, por tanto, va a existir diferencia significativa entre las dos pendientes. A continuación, se puede apreciar su representación gráfica.

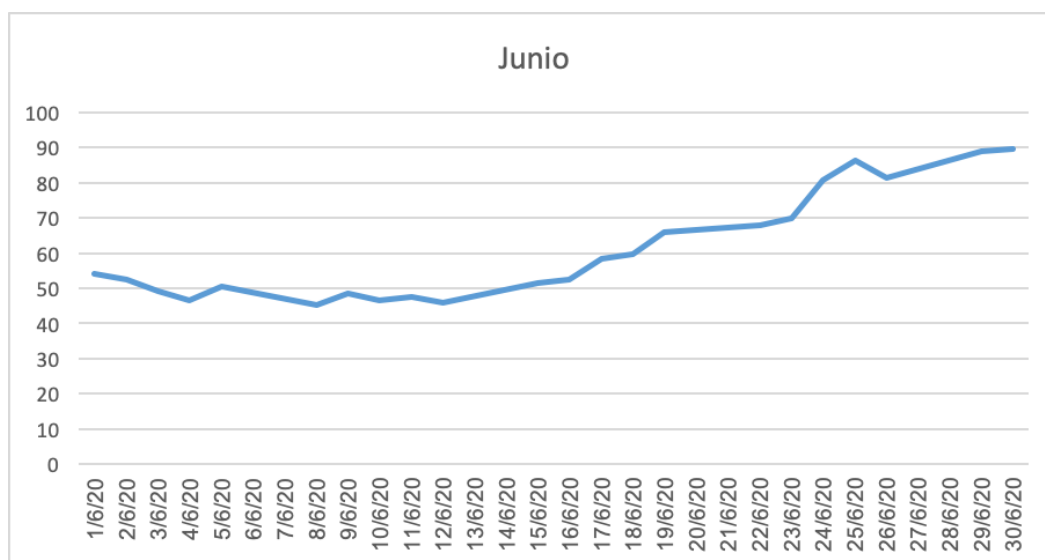


Figura 3.3 Regresión sementada - Ejemplo 1.1 (Elaboración propia)

Visualmente, se puede apreciar cómo la tendencia de la primera quincena del mes es aparentemente horizontal mientras que, a partir de la segunda quincena, la tendencia comienza a ser creciente.

Ejemplo 1.2

Por otro lado, al realizar el mismo ejercicio, pero teniendo en cuenta únicamente los datos mínimos del intervalo para los mismos periodos A y B del mismo mes. Se obtienen los siguientes datos.

	A	B
Slope	-0,059873362	2,269757085
n	11	11
SE (regression)	2,01475307	4,17738942
SE (slope)	0,139636974	0,278774361

Tabla 3 Datos 1, regresión segmentada - Ejemplo 1.2 (Elaboración propia)

Por otro lado, se calcula además la siguiente tabla de resultados.

Difference	2,329630447
SE (Difference)	0,311791001
t-statistic	7,47176935
df	18
p	0,00000064

Tabla 4 Datos 2, regresión sementada - Ejemplo 1.2 (Elaboración propia)

De las tablas anteriores, se puede determinar que, según el valor del p-valor, siendo este muy inferior a 0,05, la hipótesis nula de que ambas pendientes son iguales puede

rechazarse concluyendo, por lo que existe diferencia significativa entre las pendientes de las dos quincenas.

Su gráfica queda representada a continuación.

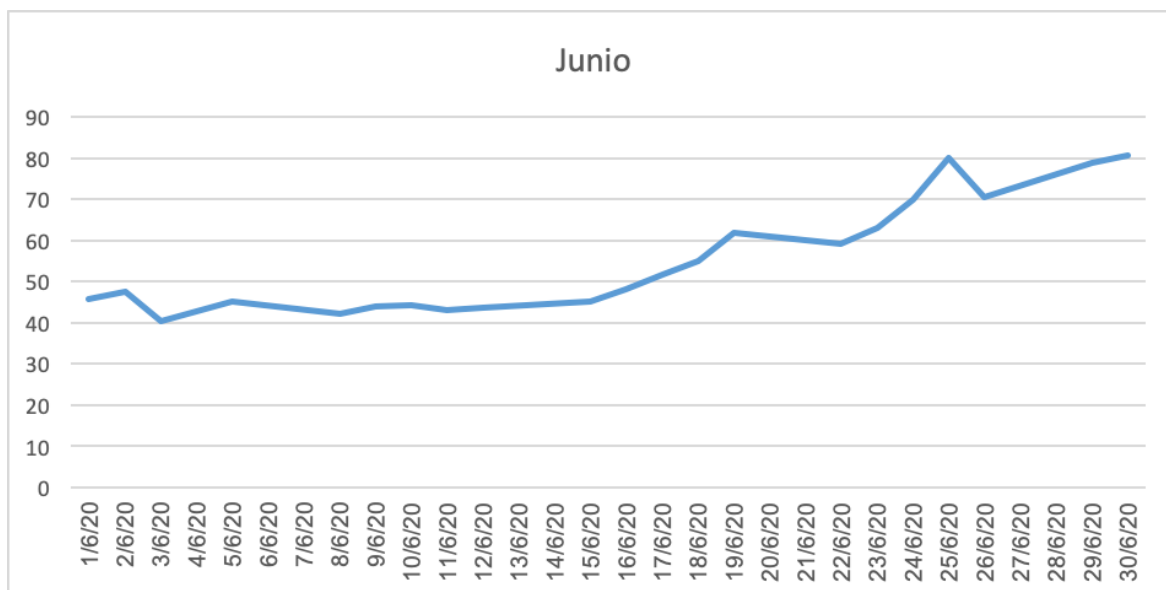


Figura 3.4 Regresión segmentada - Ejemplo 1.2 (Elaboración propia)

Cabe destacar que, aunque a simple vista no se podría obtener un resultado, sí que puede diferenciarse cómo, al igual que en el ejemplo anterior, durante la primera quincena del mes (periodo A) la tendencia de los datos sigue una línea horizontal mientras que durante el periodo B, la tendencia seguida aparenta tener cierta pendiente positiva.

Ejemplo 2

Por último, recapitulando lo anterior, se realiza el análisis de las pendientes de los mínimos del mes de junio y los máximos de ese mismo mes para poder compararlas y observar si existe diferencia significativa entre ellas. En la siguiente tabla se muestran los datos obtenidos.

MODELOS DE CAMBIO DE RÉGIMEN CON DATOS CLÁSICOS

	High	Low
Slope	1,48545889	1,328639479
n	22	22
SE (regression)	7,628699104	6,31859061

Tabla 5 Datos 1, regresión segmentada - Ejemplo 2 (Elaboración propia)

Por otro lado, se obtiene también la siguiente tabla de valores.

Difference	0,156819411
SE (Difference)	0,238089126
t-statistic	0,658658435
df	40
p	0,5139

Tabla 6 Datos 2, regresión sementada - Ejemplo 2 (Elaboración propia)

De acuerdo con el valor del p-valor obtenido en este caso, siendo muy diferente a los anteriores, la conclusión que se puede extraer es que, al ser una cifra por encima de 0,05, no va a haber evidencia suficiente para rechazar la hipótesis nula de que ambas pendientes son iguales y, por tanto, no se va a poder rechazar esta idea.

Los datos quedan representados en la siguiente gráfica.

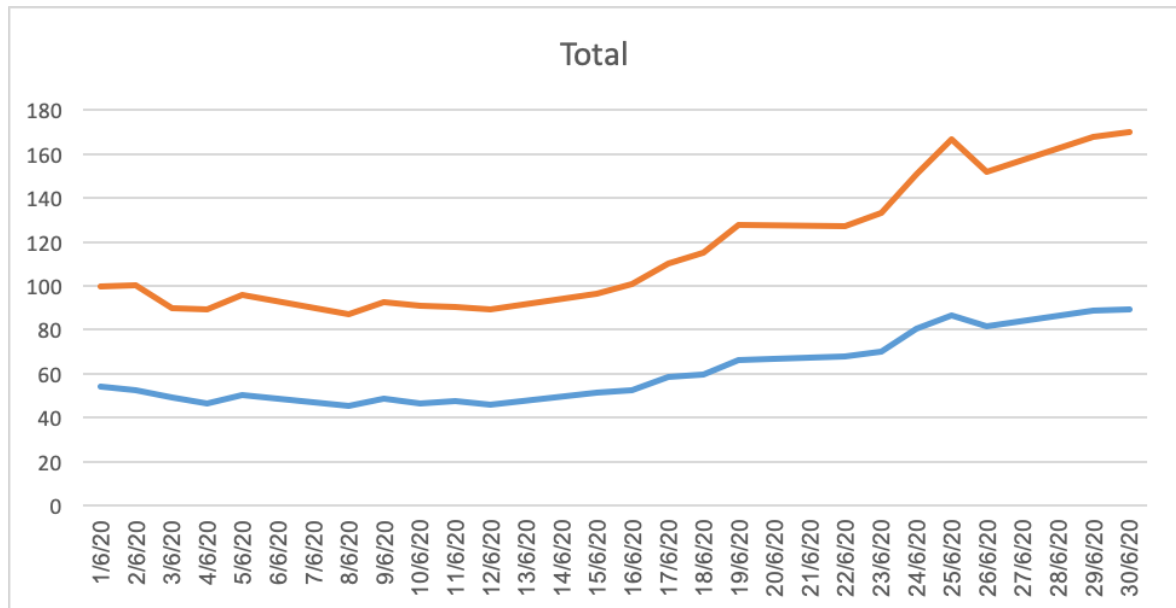


Figura 3.5 Regresión segmentada - Ejemplo 2 (Elaboración propia)

Gráficamente se puede apreciar cómo las tendencias de ambas series temporales son similares a lo largo del mes, por lo que se puede concluir que los resultados obtenidos son coherentes.

Capítulo 4. MÉTODOS Y MODELOS DE REGRESIÓN LINEAL CON DATOS DE INTERVALO

El mundo de hoy en día está, entre otras cosas, caracterizado por resultar cada día más incierto. Es por esto por lo que es de gran ayuda para todos poder realizar pronósticos o predicciones en diversos ámbitos de la vida diaria. No solo resultan de gran utilidad las predicciones puntuales sino también aquellas que se pueden ofrecer con intervalos.

4.1 CONCEPTOS BÁSICOS AL ANALIZAR SERIES DE TIEMPO EN INTERVALOS (ITS)

A continuación, se van a exponer una serie de conceptos relacionados con las series de tiempo de intervalo o ITS, según (Maté, 2021).

Un **intervalo** $[x]$ en una base $(E, \leq)$ es una pareja $[x] = [x^L, x^U]$ donde $x^L, x^U \in E$ son los límites superior e inferior del intervalo, respectivamente, de forma que $x^L \leq x^U$. Un intervalo puede ser representado por su centro o por su radio, siendo estos, respectivamente, $x^C = \frac{1}{2} * (x^L + x^U)$ y $x^R = \frac{1}{2} * (x^U - x^L)$.

Una **serie de tiempo de intervalos** (ITS) es un tipo de proceso estocástico de intervalo, siendo este una colección de variables aleatorias de intervalo que se ordenan a lo largo de un tiempo, y se denota como $\{[x_t]\} = \{[x^L, x^U]\} = \{< x_t^C, x_t^R >\}$ para $t = 1, 2, \dots, T$.

Un **Camino Aleatorio de intervalo** (iRW) se puede explicar como que una ITS $[Y_t]$ es un iRW con una cierta desviación si $[Y_t] = [\mu] + [Y_{t-1}] + [\varepsilon_t]$.

Como consecuencia de ello, se obtiene:

$$[Y_t^L] = [\mu^L] + [Y_{t-1}^L] + [\varepsilon_t^L]$$

$$[Y_t^U] = [\mu^U] + [Y_{t-1}^U] + [\varepsilon_t^U]$$

Ecuación 4.1 iRW

El caso en el que $[\mu] = 0$, implica iRW sin desviación.

El modelo iRW para ITS proporciona la predicción más simple para cada instante en la serie temporal, es decir, un modelo de Naive Forecast.

Un modelo Naive Forecast de intervalo para una ITS $[Y_t]$ para $t = 2, 3, \dots, T$; viene dado por $[\widehat{Y}_t] = [Y_{t-1}]$, de donde se puede obtener:

$$\widehat{Y}_t^L = Y_{t-1}^L$$

$$\widehat{Y}_t^U = Y_{t-1}^U$$

Ecuación 4.2 Naive Forecast de intervalo

Medida de distancia para datos de intervalo

Para medir la precisión del pronóstico se utiliza una medida de la distancia entre la predicción y el valor real. En el caso de las series de tiempo clásicas, la diferencia entre la observación y la predicción es la distancia entre ellas, sin embargo, para el caso de intervalos, existen varios tipos de medidas de la distancia, siendo uno de los principales la **distancia Euclídea**.

$$d_E([x], [y]) = \frac{1}{\sqrt{2}} * \sqrt{(x^L - y^L)^2 + (x^U - y^U)^2}$$

Ecuación 4.3 Distancia Euclídea

4.2 CONCEPTOS ESENCIALES AL ANALIZAR VARIAS PREDICCIONES PARA UNA ITS

Considerando una ITS $\{[x_t]\}$ y su ITS pronosticada $\{[\hat{x}_t]\}$, con $t = 1, \dots, T$.

4.2.1 ERROR DE LA DISTANCIA MEDIA (MDE)

Una forma de evaluar la precisión de un método de predicción de ITS es calculando la distancia en cada momento y la media de todas las distancias. Se obtiene así el **Error de la Distancia Media** (MDE), que se utiliza para calcular la precisión del método de predicción.

$$MDE^2 = \frac{1}{2} * \frac{\sum_{t=1}^T [(x_t^L - \hat{x}_t^L)^2 + (x_t^U - \hat{x}_t^U)^2]}{T}$$

Ecuación 4.4 Error de Distancia Media (MDE)

El MDE es equivalente al RMSE para evaluar la precisión de un método de predicción de ITS.

4.2.2 VARIANZA RELATIVA MEDIA DE INTERVALO (iARV)

El valor de la varianza relativa media de intervalo (iARV) se expresa de la siguiente forma:

$$iARV = \frac{\sum_{t=1}^T (x_t^L - \hat{x}_t^L)^2 + \sum_{t=1}^T (x_t^U - \hat{x}_t^U)^2}{\sum_{t=1}^T (\bar{x}_t^L - \bar{x}_t^L)^2 + \sum_{t=1}^T (\bar{x}_t^U - \bar{x}_t^U)^2}$$

Ecuación 4.5 Varianza Relativa Media de intervalo (iARV)

donde T representa el número de intervalos ajustados, $[\hat{x}_t] = [\hat{x}_t^L, \hat{x}_t^U]$ es el intervalo ajustado número t , $[\bar{x}] = [\bar{x}^L, \bar{x}^U]$ es el intervalo de los promedios de las muestras y $\bar{x}^L$ y $\bar{x}^U$ son los promedios del límite inferior y superior, respectivamente.

El iARV es un estadístico que se encarga de comparar las predicciones del modelo con las predicciones dadas por los valores medios de una variable de intervalo. Cuanto menor

sea el valor de iARV mejor será la predicción, siendo el modelo de predicción ideal el caso en el que $iARV = 0$.

4.2.3 ESTADÍSTICO *U* DE THEIL DE INTERVALO

El **estadístico U de Theil** es utilizado para la comparación de distintos modelos de predicción con un modelo Naive. Se utiliza iU , para intervalos, para comparar el modelo utilizado con el modelo de Camino Aleatorio en intervalo (iRW).

$$iU = \sqrt{\frac{\sum_{t=2}^T (x_t^L - \hat{x}_t^L)^2 + \sum_{t=2}^T (x_t^U - \hat{x}_t^U)^2}{\sum_{t=2}^T (x_t^L - x_{t-1}^L)^2 + \sum_{t=2}^T (x_t^U - x_{t-1}^U)^2}}$$

Ecuación 4.6 Estadístico U de Theil de intervalo (iU)

Según el valor obtenido en iU se puede llegar a las siguientes conclusiones en cuanto al modelo de predicción:

- Si $iU = 1$, el modelo observado se comporta igual que el de iRW .
- Si $iU > 1$, el modelo observado es peor que el de iRW .
- Si $iU < 1$, el modelo observado es mejor que el de iRW .

Los tres estadísticos explicados, MDE, $iARV$ e iU , son medidas del tamaño de la desviación de las predicciones que realizan con respecto al valor real, por tanto, cuanto menores sean estos valores mejor serán las predicciones.

4.3 MODELOS DE REGRESIÓN LINEAL PARA INTERVALOS (*ILR*)

Como explica (Maceda, 2018), hasta ahora el análisis de datos clásicos o puntuales ha sido determinante para el estudio y la predicción de diferentes situaciones de la vida diaria, siendo estos utilizados en estadística, economía, Machine Learning y otras herramientas y softwares elaborados para trabajar con este tipo de datos. Sin embargo, gracias a los datos de intervalos es posible obtener una información más amplia que con los

datos puntuales, especialmente para aquellos elementos o características que no se pueden explicar con un solo número.

Según explica (Maté, 2012) el primer análisis de intervalos tuvo lugar con Ramon Moore, en la década de 1960, con un enfoque inicial hacia los resultados computacionales surgidos de datos de intervalo. Desde entonces el estudio de los intervalos ha adquirido gran interés en el mundo de la investigación, que sigue todavía en desarrollo.

Una de las áreas que más interés ha generado en el estudio de datos de intervalos ha sido el análisis de series temporales de intervalo, especialmente aplicadas al campo de las finanzas, donde se han realizado profundos estudios de los datos significativos, como los valores de la bolsa o la fluctuación de la moneda, así como el análisis de diferentes métodos de predicción.

A continuación, se van a exponer una serie de ideas relacionadas con la regresión lineal de intervalos. Para empezar, cabe mencionar que, al hablar de la regresión lineal de intervalos, se deben destacar 4 paradigmas diferentes:

- Paradigma 1: $\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n = y$, donde tanto las variables independientes x_n como la variable dependiente y representan datos puntuales. Esta es la regresión que se ha explicado hasta ahora en este proyecto.
- Paradigma 2: $\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n = [Y]$, donde las variables independientes x_n representan datos puntuales, mientras que la variable dependiente $[Y]$ indica la presencia de un intervalo.
- Paradigma 3: $\beta_0 + \beta_1 [X_1] + \dots + \beta_n [X_n] = y$, donde las variables independientes $[X_n]$ representan datos de intervalo, mientras que la variable dependiente y es una variable puntual.
- Paradigma 4: $\beta_0 + \beta_1 [X_1] + \dots + \beta_n [X_n] = [Y]$, donde tanto las variables independientes $[X_n]$ como la variable dependiente $[Y]$, representan datos de intervalo.

A partir del paradigma 4, se han diseñado una serie de métodos de regresión lineal para intervalos (ILR), algunos de los cuales se exponen a continuación.

4.3.1 MÉTODO DEL CENTRO (CM)

El **Método del Centro**, en inglés, Centre Method (CM), fue presentado por (Billard & Diday, 2000) y se basa en un modelo de regresión a partir de los centros de los intervalos para, posteriormente, aplicarlo a los límites superiores e inferiores de los intervalos y poder obtener la predicción $[\hat{Y}]$.

El método parte de:

$$y_{li} = \beta_0 + \beta_1 x_{l1} + \dots + \beta_n x_{ln} + \epsilon_{li}$$

$$y_{ui} = \beta_0 + \beta_1 x_{u1} + \dots + \beta_n x_{un} + \epsilon_{ui}$$

Ecuación 4.7 Regresión en el Método del Centro (CM)

Además, la suma de los cuadrados de las desviaciones para el modelo es la siguiente:

$$S_{CM} = \sum_{i=1}^n (\epsilon_{li} + \epsilon_{ui})^2$$

Ecuación 4.8 Suma de los cuadrados de las desviaciones en CM

Teniendo esto en cuenta, es posible obtener los valores de los β para minimizar S_{CM} a partir de una regresión lineal de los centros de $[Y]$ y $[X_1, X_2, \dots, X_n]$. Los valores resultantes de $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_n)$ van a predecir $\hat{Y}_{CM}$ como se muestra a continuación:

$$\hat{y}_{li} = \hat{\beta}_0 + \hat{\beta}_1 x_{l1} + \dots + \hat{\beta}_n x_{ln}$$

$$\hat{y}_{ui} = \hat{\beta}_0 + \hat{\beta}_1 x_{u1} + \dots + \hat{\beta}_n x_{un}$$

Ecuación 4.9 Predicción con CM

4.3.2 MÉTODO MINMAX (MMM)

El **método minMax** (mMM) fue, al igual que el Método del Centro anterior, expuesto por (Billard & Diday, 2002) donde proponían una alternativa mejorada y más desarrollada al Método del Centro.

En este caso, el método propone hacer uso de distintos parámetros para obtener los límites superior e inferior de $[\hat{Y}]$. Esto implica suponer la independencia entre los valores del límite superior y del inferior, tratándolos como variables diferentes a la hora de hacer la predicción.

La relación de regresión lineal es la siguiente:

$$y_{li} = \beta_0^l + \beta_1^l x_{l1} + \dots + \beta_n^l x_{ln} + \epsilon_{li}$$

$$y_{ui} = \beta_0^u + \beta_1^u x_{u1} + \dots + \beta_n^u x_{un} + \epsilon_{ui}$$

Ecuación 4.10 Regresión en Método minMax

Además, al igual que en el método anterior, la suma de los cuadrados de las desviaciones para el modelo es la siguiente:

$$S_{MmnMax} = \sum_{i=1}^n (\epsilon_{li})^2 + \sum_{i=1}^n (\epsilon_{ui})^2$$

Ecuación 4.11 Suma de los cuadrados de las desviaciones en Método minMax

Minimizando la ecuación anterior y teniendo en cuenta los límites superiores $[Y_u]$ y $[X_{u1}, X_{u2}, \dots, X_{un}]$ de forma independiente a los límites inferiores $[Y_l]$ y $[X_{l1}, X_{l2}, \dots, X_{ln}]$ mediante dos regresiones lineales independientes, es posible obtener los valores de los β , $\beta_0^l, \beta_1^l, \dots, \beta_n^l$ y $\beta_0^u, \beta_1^u, \dots, \beta_n^u$ que hacen mínimo el valor de S_{minMax} .

A partir de aquí, se obtienen los valores de $\hat{\beta}^l = (\hat{\beta}_1^l, \hat{\beta}_2^l, \dots, \hat{\beta}_n^l)$ y de $\hat{\beta}^u = (\hat{\beta}_1^u, \hat{\beta}_2^u, \dots, \hat{\beta}_n^u)$ para predecir $\hat{Y}_{minMax}$ como se muestra a continuación:

$$\hat{y}_{li} = \hat{\beta}_0^l + \hat{\beta}_1^l x_{l1} + \dots + \hat{\beta}_n^l x_{ln}$$

$$\hat{y}_{ui} = \hat{\beta}_0^u + \hat{\beta}_1^u x_{u1} + \cdots + \hat{\beta}_n^u x_{un}$$

Ecuación 4.12 Predicción con Método minMax

4.3.3 MÉTODO DEL CENTRO Y RADIO (CRM)

El Método del Centro y Radio o Centre and Range Method (CRM), en inglés, fue presentado por (Lima Neto & De Carvalho, 2008) y fue propuesto como una alternativa mejorada al CM de (Billard & Diday, 2000).

Por un lado, el CRM se basa en el CM y a partir de él, añade el concepto de radios para poder llevar a cabo la predicción de forma mejorada.

Por otro lado, el CRM también hace uso de la idea del método minMax para añadir más información que el CM y poder satisfacer las necesidades que supone la complejidad de los datos de intervalo. En este caso el método, como su propio nombre indica, se basa en el centro y en el radio.

A continuación, se muestra cómo se va a representar el intervalo $[W]$ que, en vez de representarse por sus límites superiores e inferiores, $[W_L, W_U]$, se indicará a partir de su centro y radio, es decir, $[W_c, W_r]$, siendo estos:

$$W_c = (W_L + W_U)/2$$

Ecuación 4.13 Centro del intervalo en CRM

$$W_r = (W_U - W_L)/2$$

Ecuación 4.14 Radio del intervalo en CRM

Para este modelo, la relación de la regresión lineal se muestra de la siguiente forma:

$$y_{ci} = \beta_0^c + \beta_1^c x_{c1} + \cdots + \beta_n^c x_{cn} + \epsilon_{ci}$$

$$y_{ri} = \beta_0^r + \beta_1^r x_{r1} + \cdots + \beta_n^r x_{rn} + \epsilon_{ri}$$

Ecuación 4.15 Regresión en CRM

Además, al igual que en los casos anteriores, es importante saber la suma de los cuadrados de las desviaciones para realizar la predicción posterior. Esta suma se calcula del siguiente modo:

$$S_{CRM} = \sum_{i=1}^n ((\epsilon_{ci})^2 + (\epsilon_{ri})^2)$$

Ecuación 4.16 Suma de los cuadrados de las desviaciones en CRM

Por último, al minimizar la ecuación anterior teniendo en cuenta que los centros $[Y_c]$ y $[X_{c1}, X_{c2}, \dots, X_{cn}]$ como en el CM, de forma independiente a los radios $[Y_r]$ y $[X_{r1}, X_{r2}, \dots, X_{rn}]$ con dos regresiones lineales diferentes, se obtienen los valores de $\beta_0^c, \beta_1^c, \dots, \beta_n^c$ y $\beta_0^r, \beta_1^r, \dots, \beta_n^r$ que minimizan S_{CRM} .

Los valores resultantes de $\hat{\beta}^c = (\hat{\beta}_1^c, \hat{\beta}_2^c, \dots, \hat{\beta}_n^c)$ y de $\hat{\beta}^r = (\hat{\beta}_1^r, \hat{\beta}_2^r, \dots, \hat{\beta}_n^r)$ predecirán $\hat{Y}_{CRM}$ como se indica:

$$\hat{y}_{ci} = \hat{\beta}_0^c + \hat{\beta}_1^c x_{c1} + \dots + \hat{\beta}_n^c x_{cn}$$

$$\hat{y}_{ri} = \hat{\beta}_0^r + \hat{\beta}_1^r x_{r1} + \dots + \hat{\beta}_n^r x_{rn}$$

Ecuación 4.17 Predicción con CRM

Como se aprecia en el desarrollo de este método, se puede destacar que es igual que el anterior con la única diferencia de que en vez de hacer uso de los límites superiores e inferiores para llevar a cabo la predicción, en este caso utiliza el centro y radio, calculados a partir de estos límites.

4.3.4 MÉTODO DEL CENTRO Y RADIO CON RESTRICCIONES (CCRM)

El **Método de Centro y Radio con Restricciones**, en inglés, Constrained Centre and Range Method (CCRM) fue, al igual que el anterior, expuesto por (Lima Neto & De Carvalho, Constrained linear regression models for symbolic interval-valued variables, 2010) proponiendo una versión mejorada de su propio CRM, presentado anteriormente.

El nuevo método se basa en el mismo principio que el CRM haciendo uso del centro y radio de $[X_1, X_2, \dots, X_n]$ para predecir $[\hat{Y}]$. Sin embargo, en este caso incluye una restricción para impedir que el problema llegue a un punto en el que matemáticamente no tenga sentido. Esta restricción consiste en que el radio tenga que ser mayor o igual a cero, lo cual no era obligatorio en CRM y podía dar lugar a problemas.

En este caso, la regresión lineal se muestra a continuación y es igual a la del método anterior.

$$y_{ci} = \beta_0^c + \beta_1^c x_{c1} + \dots + \beta_n^c x_{cn} + \epsilon_{ci}$$

$$y_{ri} = \beta_0^r + \beta_1^r x_{r1} + \dots + \beta_n^r x_{rn} + \epsilon_{ri}$$

Ecuación 4.18 Regresión en CCRM

Sin embargo, la suma de los cuadrados de los residuos es diferente en este caso con respecto al CRM y se define debajo.

$$S_{CCRM} = \sum_{i=1}^n ((\epsilon_{ci})^2 + (\epsilon_{ri})^2)$$

$$s. t. \quad \beta_i^r \geq 0, \quad i = 0, \dots, p$$

Ecuación 4.19 Suma de los cuadrados de las desviaciones en CCRM

Teniendo en cuenta lo siguiente:

$$\epsilon_{ci} = y_{ci} - \beta_0^c - \beta_1^c x_{c1} - \dots - \beta_n^c x_{cn}$$

$$\epsilon_{ri} = y_{ri} - \beta_0^r - \beta_1^r x_{r1} - \dots - \beta_n^r x_{rn}$$

Ecuación 4.20 Valores de ϵ_{ci} y ϵ_{ri}

A partir de Ecuación 4.19 y Ecuación 4.20 se accede al siguiente problema de optimización:

$$\min \| Y_c - X_c \beta^c \|^2 + \| Y_r - X_r \beta^r \|^2$$

$$s. t. \quad \beta^r \geq 0$$

Ecuación 4.21 Problema de optimización resultante de CCRM

Los valores óptimos de β^c se pueden obtener resolviendo un problema de regresión sabiendo que la restricción de la Ecuación 4.21 no depende de β^c . Es por eso por lo que β^c será igual que en CRM.

Por otro lado, la primera norma de la Ecuación 4.21 se puede considerar constante para β^r . La estimación de β^r se obtiene, por tanto, resolviendo un problema de regresión con restricciones, el cual se puede considerar como un caso particular de mínimos cuadrados no-negativos. En conclusión, CCRM se basa en la resolución de dos problemas de regresión independientes.

Los valores resultantes de $\hat{\beta}^c$ y $\hat{\beta}^r$ van a predecir el el valor de $\hat{Y}_{CCRM}$ según el siguiente esquema:

$$\hat{y}_{ci} = \hat{\beta}_0^c + \hat{\beta}_1^c x_{c1} + \dots + \hat{\beta}_n^c x_{cn}$$

$$\hat{y}_{ri} = \hat{\beta}_1^r X_r$$

Ecuación 4.22 Predicción con CCRM

La ventaja de este método, en comparación con los dos anteriores, es que no va a permitir llegar a resultados ilógicos debido a las restricciones impuestas.

4.3.5 REGRESIÓN LINEAL SIMPLE BASADA EN ARITMÉTICA DE INTERVALOS (MODELO M)

La **Regresión Lineal Simple Basada en Aritmética** o Arithmetic-based Simple Linear Regression (Modelo M) fue presentada por (Sinova, Colubi, & Gonzalez-Rodriguez, 2012) y difiere de los métodos expuestos previamente en que el enfoque de la regresión es aritmético.

Para empezar, cabe destacar que un modelo lineal basado en la aritmética de intervalos habitual se centra al mismo tiempo en el centro y en el radio del intervalo de datos, asegurando que el modelo de intervalo basado en aritmética está bien definido. En cambio, se podrían examinar el centro y el radio por separado o tratarlos como variables aleatorias independientes. Sin embargo, de esta forma no se podría garantizar que un modelo lineal para el radio tuviera sentido, debido a las causas que se han explicado en el método anterior.

Para una primera aproximación, asumiendo un modelo lineal, el problema tratará de encontrar un valor real para $\hat{a}$ y unos valores de intervalo para $[\hat{B}]$, de forma que quede $\hat{a}[X] + [\hat{B}]$ lo más cerca posible a la muestra $[Y]$.

Teniendo en cuenta que el problema de mínimos cuadrados para encontrar $\hat{a}$ y $[\hat{B}]$ es el siguiente:

$$\frac{1}{n} \sum_{i=1}^n (d_{\theta}([Y], \hat{a}[X] + [\hat{B}]))$$

Ecuación 4.23 Problema de mínimos cuadrados en el modelo M

A partir de aquí, para calcular $\hat{a}$ es necesario seguir los siguientes pasos:

Paso 1: calcular las estimaciones de la muestra de los momentos siguientes de las variables aleatorias reales centro y radio.

- Las medias de los centros y radios de $[X]$ e $[Y]$:

$$\overline{centx} = \frac{1}{n} \sum_{i=1}^n centx_i, \quad \overline{radx} = \frac{1}{n} \sum_{i=1}^n radx_i$$

Ecuación 4.24 Medias de los centros y radios de $[X]$ en Modelo M

- Las varianzas de los centros y radios de $[X]$ e $[Y]$:

$$s_{centx}^2 = \frac{1}{n} \sum_{i=1}^n (centx_i - \overline{centx})^2, \quad s_{radx}^2 = \frac{1}{n} \sum_{i=1}^n (radx_i - \overline{radx})^2,$$

Ecuación 4.25 Varianzas de los centros y radios de $[X]$ en Modelo M

- Las covarianzas de los centros y radios de $[X]$ e $[Y]$:

$$s_{centx,centy} = \frac{1}{n} \sum_{i=1}^n (centx_i - \overline{centx})(centy_i - \overline{centy})$$

$$s_{radx,rad y} = \frac{1}{n} \sum_{i=1}^n (radx_i - \overline{radx})(rad y_i - \overline{rad y})$$

Ecuación 4.26 Covarianzas de los centros y radios de $[X]$ e $[Y]$ en Modelo M

Paso 2: calcular las estimaciones de la muestra de los momentos siguientes del conjunto aleatorio en intervalos.

- Las medias de $[X]$ e $[Y]$:

$$[\bar{X}] = [\overline{centx} - \overline{radx}, \overline{centx} + \overline{radx}], \quad [\bar{Y}] = [\overline{centy} - \overline{rad y}, \overline{centy} + \overline{rad y}]$$

Ecuación 4.27 Medias de $[X]$ e $[Y]$ en Modelo M

- La varianza θ -Fréchet de $[X]$:

$$\hat{\sigma}_x^2 = s_{centx}^2 + \theta s_{radx}^2$$

Ecuación 4.28 Varianza θ -Fréchet de $[X]$ en Modelo M

- La covarianza θ de $[X]$ e $[Y]$, y de $[-X]$ e $[Y]$:

$$\widehat{\sigma}_{x,y} = s_{centx,centy} + \theta s_{radx,rad y}, \quad \widehat{\sigma}_{-x,y} = -s_{centx,centy} + \theta s_{radx,rad y}$$

Ecuación 4.29 Covarianza θ a de $[X]$ e $[Y]$ y de $[-X]$ e $[Y]$ en Modelo M

Paso 3: calcular las estimaciones de la muestra $\widehat{a}_0$ y $\widehat{a}$ para los datos disponibles:

- Si todas las x_i son valores reales, entonces $\widehat{a}_0 = \infty$.
- Si no,

$$\widehat{a}_0 = \min_{i: radx_i \neq 0} \frac{centx_i}{radx_i}$$

Ecuación 4.30 Estimación de $\widehat{a}_0$ en Modelo M

- Si $\widehat{\sigma}_{x,y} < 0$ y $\widehat{\sigma}_{-x,y} < 0$, entonces $\hat{a} = 0$.
- Si $\widehat{\sigma}_{x,y} \geq 0$ y $\widehat{\sigma}_{x,y} \geq \widehat{\sigma}_{-x,y}$, entonces:

$$\hat{a} = \min \left\{ \widehat{a}_0, \frac{\widehat{\sigma}_{x,y}}{\widehat{\sigma}_x^2} \right\}$$

Ecuación 4.31 Valor de $\hat{a}$ si $\widehat{\sigma}_{x,y} \geq 0$ y $\widehat{\sigma}_{x,y} \geq \widehat{\sigma}_{-x,y}$ en Modelo M

- Si $\widehat{\sigma}_{-x,y} \geq 0$ y $\widehat{\sigma}_{-x,y} \geq \widehat{\sigma}_{x,y}$, entonces:

$$\hat{a} = \min \left\{ \widehat{a}_0, \frac{\widehat{\sigma}_{-x,y}}{\widehat{\sigma}_x^2} \right\}$$

Ecuación 4.32 Valor de $\hat{a}$ si $\widehat{\sigma}_{-x,y} \geq 0$ y $\widehat{\sigma}_{-x,y} \geq \widehat{\sigma}_{x,y}$ en Modelo M

Por último, una vez que $\hat{a}$ ha sido calculado, es el turno de obtener $[\hat{B}]$.

$$[\hat{B}] = [Y] - \hat{a}[X]$$

Ecuación 4.33 Valor de $[\hat{B}]$ en Modelo M

A partir de estos valores es posible predecir $[\hat{Y}_{\text{Modelo M}}]$ como se muestra a continuación:

$$[\hat{Y}] = \hat{a}[X] + [\hat{B}]$$

Ecuación 4.34 Predicción en Modelo M

4.3.6 REGRESIÓN LINEAL SIMPLE BASADA EN LASSO

En estadística, Lasso (Least Absolute Shrinkage and Selection Operator) es un método de análisis de regresión que realiza la selección de variables y regularización para mejorar la exactitud e interpretabilidad del modelo estadístico producido por este.

El método de **Regresión Lineal Basada en la Técnica de Lasso** fue expuesto por (Giordani, 2015) y puede tener un enfoque similar al de CCRM porque el problema consiste en un proceso de optimización que trata de minimizar la función objetiva. Este proceso se lleva a cabo gracias al uso de unos específicos coeficientes de regresión para los radios, de

forma que se resuelvan todas las situaciones en las que la pendiente es diferente a la propagación de la imprecisión.

La regresión lineal para este modelo se expresa a continuación:

$$y_c = b_c X_c + \epsilon_c$$

$$y_r = b_r X_r + \epsilon_r = (b_c + b_a) X_r + \epsilon_r$$

Ecuación 4.35 Regresión en Lasso

donde b_c y b_r son los vectores de los coeficientes de regresión para el centro y el radio, respectivamente, y b_a el vector de los coeficientes añadidos. Es por eso por lo que los coeficientes del radio b_r , van a ser igual a la suma de los del centro b_c más los añadidos, b_a , es decir, $b_r = b_c + b_a$.

A continuación, se presenta la forma de minimizar el problema:

$$\min_{b_c, b_r} \|y_c - X_c b_c\|^2 + \theta \|y_r - X_r(b_c + b_a)\|^2$$

Ecuación 4.36 Problema a minimizar en Lasso

La Ecuación 4.36 anterior requiere una serie de restricciones para poder garantizar que los radios que se estiman no son negativos y que b_a es lo más pequeño posible. Es por eso por lo que se aplica la siguiente restricción:

$$X_r(b_c + b_a) \geq 0$$

Ecuación 4.37 Restricción en Lasso

Como se ha explicado previamente, la técnica de Lasso es un método de estimación mediante regresión que trata de disminuir el valor de algunos coeficientes de regresión y poner otros a cero. Este proceso se lleva a cabo al minimizar la suma de los cuadrados de las desviaciones, teniendo en cuenta la restricción de que la suma de los valores absolutos de los coeficientes de regresión es menor que un límite. Por tanto, teniendo en cuenta las

ecuaciones obtenidas previamente, el problema basado en Lasso se presenta de la siguiente forma:

$$\min_{b_c, b_r} \|y_c - X_c b_c\|^2 + \theta \|y_r - X_r(b_c + b_a)\|^2$$

$$\text{s. t. } X_r(b_c + b_a) \geq 0$$

Ecuación 4.38 Problema basado en Lasso

La predicción de $\hat{Y}_{Lasso}$ es, por tanto, la siguiente:

$$\hat{y}_c = b_c X_c$$

$$\hat{y}_r = (b_c + b_a) X_r$$

Ecuación 4.39 Predicción en el problema basado en Lasso

4.3.7 MÉTODO DE CENTRO Y RADIO CONJUNTOS CON RESTRICCIONES (CCRJM)

El **Método de Centro y Radio Conjuntos con Restricciones**, en inglés, **Constrained Centre and Range Joint Method (CCRJM)** fue introducido por (Hao & Guo, 2017). Este método es similar al anterior ya que busca un coeficiente para el centro y el radio, pero, la diferencia es que, en este caso, lo hace conjuntamente, es decir, utilizando los valores de los radios para el modelo del centro y viceversa.

La regresión lineal considerada es la siguiente:

$$Y_c = X\beta^c + \epsilon_c$$

$$Y_r = X\beta^r + \epsilon_r$$

Ecuación 4.40 Regresión lineal en CCRJM

En este caso, X representa lo siguiente, $X = [1 \parallel X^c \parallel X^r]$, es decir, X representa conjuntamente los centros y los radios.

$$X^c = \begin{bmatrix} x_{11}^c & \cdots & x_{1p}^c \\ \vdots & \ddots & \vdots \\ x_{n1}^c & \cdots & x_{np}^c \end{bmatrix} \quad X^r = \begin{bmatrix} x_{11}^r & \cdots & x_{1p}^r \\ \vdots & \ddots & \vdots \\ x_{n1}^r & \cdots & x_{np}^r \end{bmatrix}$$

Ecuación 4.41 X^c y X^r en CCRJM

Ya que los vectores de β^c y β^r necesitarán un coeficiente para el centro y el radio de cada variable X, tendrán dimensión $2p+1$.

El problema de minimización viene dado por:

$$\min \sum_{i=1}^n \epsilon_{ci}^2 + \epsilon_{ri}^2$$

Ecuación 4.42 Problema de minimización en CCRJM

Además, es necesario añadir la restricción de no-negatividad del radio para evitar problemas.

$$s. t. \quad X\beta^r \geq 0$$

Ecuación 4.43 Restricción de no negatividad en CCRJM

Por último, se obtiene la predicción de $\hat{Y}_{CCRJM}$ como:

$$\hat{Y}_c = X\beta^c$$

$$\hat{Y}_r = X\beta^r$$

Ecuación 4.44 Predicción en CCRJM

4.3.8 ENFOQUE PARAMETRIZADO PARA REGRESIÓN LINEAL DE DATOS DE INTERVALOS.

El **Enfoque Parametrizado para la Regresión Lineal de Datos de Intervalos** fue introducido por (Souza, Souza, Amaral , & Silva Filho, 2017). Al igual que en el método anterior (CCRJM), este método supone un avance del método minMax ya que busca un

coeficiente para los límites superior e inferior, como el minMax pero lo hace de forma conjunta, es decir, utiliza los valores del límite inferior para el modelo del límite superior y los valores del límite superior para el modelo del límite inferior. De esta forma, tiene en cuenta que los datos de intervalos tienen múltiples componentes, al igual que ocurría en los métodos anteriores, pero también confirma la relación entre estos componentes.

Considerando la relación de la regresión lineal como:

$$Y_u = X\beta^u + \epsilon_u$$

$$Y_l = X\beta^l + \epsilon_l$$

Ecuación 4.45 Relación de regresión lineal en el Enfoque Parametrizado de Regresión Lineal con Datos de Intervalos

donde:

$$X = \begin{bmatrix} 1 & x_{11}^l & x_{11}^u & x_{21}^l & \cdots & x_{p1}^l & x_{p1}^u \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & x_{1n}^l & x_{1n}^u & x_{2n}^l & \cdots & x_{pn}^l & x_{pn}^u \end{bmatrix}$$

$$\beta^l = [\beta_0^l \quad \alpha_1^l \quad \omega_1^l \quad \alpha_2^l \quad \omega_2^l \quad \dots \quad \alpha_p^l \quad \omega_p^l]^T$$

$$\beta^u = [\beta_0^u \quad \alpha_1^u \quad \omega_1^u \quad \alpha_2^u \quad \omega_2^u \quad \dots \quad \alpha_p^u \quad \omega_p^u]^T$$

Al igual que en el método anterior, los vectores β^l y β^u tendrán dimensión $2p+1$ ya que necesitarán un coeficiente para los límites inferiores y otro para los límites superiores de cada variable X .

Como este método tiene en cuenta todos los parámetros del modelo, el modelo de minimización es independiente para obtener β^u y β^l , y es el siguiente:

$$S^l = \sum_{i=1}^n (\epsilon_{li})^2$$

$$S^u = \sum_{i=1}^n (\epsilon_{ui})^2$$

Ecuación 4.46 Minimización en el Enfoque Parametrizado de Regresión Lineal con Datos de Intervalos

La solución para β^l será la siguiente:

$$\beta^l = ((X)^T X)^{-1} (X)^T y^l$$

Ecuación 4.47 β^l

Y la solución para β^u será:

$$\beta^u = ((X)^T X)^{-1} (X)^T y^u$$

Ecuación 4.48 β^u

Por último, cabe destacar que la predicción para $\hat{Y}_{param}$ será:

$$\hat{Y}_l = X \beta^l$$

$$\hat{Y}_u = X \beta^u$$

Ecuación 4.49 Predicción en Enfoque Parametrizado para la Regresión Lineal de Datos de Intervalos

4.4 EJEMPLO DE MODELO M

A continuación, se muestra un ejemplo de predicción práctico que tiene que ver con el uso del modelo M que, como se ha explicado antes, consiste en un modelo de regresión lineal simple, que para predecir hace uso de aritmética de intervalos.

Para su estudio se ha partido de una función basada en la definición del modelo M propuesta por (Sinova, Colubi, & Gonzalez-Rodriguez, 2012) en la que a partir de una serie de operaciones aritméticas que tienen en cuenta el centro y el radio de un intervalo, es posible detectar los coeficientes $\hat{a}$ y $\hat{B}$. Para ello, es necesaria la realización de los 3 pasos explicados en el apartado que define este modelo.

La función, por tanto, tomará como entradas el fichero Excel que contenga todos los datos y las columnas en las que se encuentren las ITS, $[Y]$ y $[X]$ que se quieran evaluar. Siendo $[Y]$ la ITS observada y $[X]$, para el ejemplo que se va a mostrar, la misma serie retrasada un periodo. Además, la función también recibirá como entradas las fechas de inicio y fin a las que se va a aplicar el modelo M para más tarde, obtener los valores de la predicción $[\hat{Y}]$ en las fechas posteriores.

Por tanto, si $[Y]$ es $[Y(t)]$ y $[X]$ es $[Y(t - 1)]$, se obtiene el modelo M-AR (autorregresivo). Además, la función va a ir variando el conjunto de estimación, donde se encuentran las estimaciones de $[\hat{B}]$ y $\hat{a}$ para obtener el modelo M-AR dinámico (M-ARD).

Teniendo esto en cuenta, la función devolverá el vector de coeficientes $\hat{a}$ y el vector de coeficientes en intervalos $[\hat{B}]$, calculados según los procedimientos ya explicados y, a partir de ellos, podrá calcular y devolver al usuario el la ITS de la predicción, $[\hat{Y}]$.

En el ejemplo estudiado, se ha evaluado una base de datos que contiene los valores de conversión del euro al dólar, EURUSD, entre 2010 y 2019. Para analizarla, se ha escogido un periodo que comienza el 1 de enero de 2012 y finaliza el 30 de junio de 2013.

Una vez indicados estos datos a la función, ésta realiza el cálculo de los coeficientes correspondientes a esas fechas para posteriormente, aplicando el modelo M-AR dinámico, obtener los valores de la predicción para las fechas posteriores.

El resultado obtenido es una gráfica en la que se aprecian, en color azul, la ITS real y en color rosa, la ITS que se obtiene con la predicción. Se puede observar como la ITS de la predicción sigue unos patrones similares a los de la ITS real, pero de forma más suave que ella, llegando incluso a predecir con total exactitud algunos valores a lo largo de la serie, los cuales se pueden encontrar en las zonas en las que ambas series se cruzan.

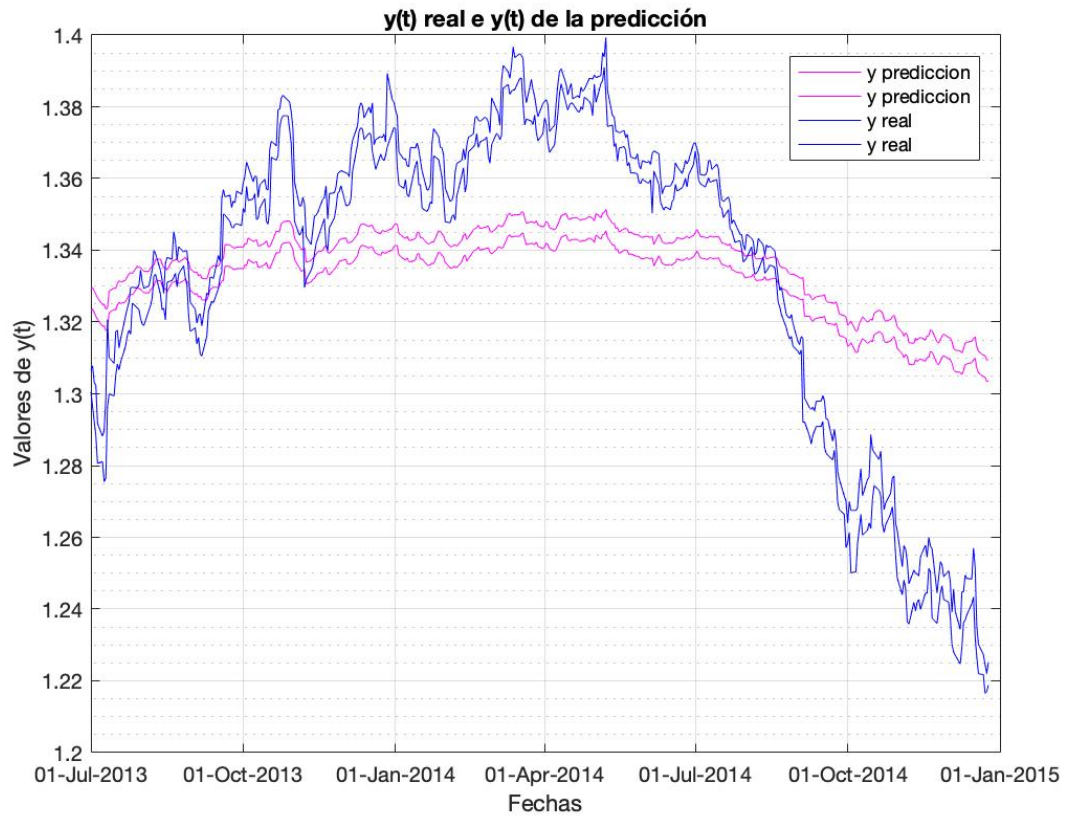


Figura 4.1 Ejemplo Modelo M-ARD (Elaboración propia)

Capítulo 5. MODELOS DE CAMBIOS DE RÉGIMEN

CON DATOS DE INTERVALO. ENFOQUE DE

DATOS SIMBÓLICOS

5.1 INTRODUCCIÓN

El modelado y la predicción de datos de ITS ha recibido una importante atención en econometría y estadística. Los intervalos, así como otras variables multi-valor que contienen información de variabilidad e incertidumbre, son de un gran interés en el dominio del “Análisis de Datos Simbólicos” (Symbolic Data Analysis, SDA) en estadística. Los datos de intervalos surgen en situaciones muy diversas como en los precios máximos y mínimos diarios de las acciones (Yang & Han, 2015) los precios de oferta y demanda de los activos en mercados financieros o datos imprecisos que vienen de medidas repetidas o estimaciones de intervalos de confianza (Brito, 2007). Otros grupos o clases de este tipo de datos surgen cuando valores reales son muestras generalizadas de intervalos o medidas imprecisas de ciertas variables puntuales (Teles & Brito, 2015).

La ventaja de una ITS sobre una serie de tiempo con datos clásicos es que la ITS contiene tanto la información del patrón o tendencia seguidos como la volatilidad de la información, que ayuda a saber la frecuencia o intensidad de los cambios de precio de un activo, midiendo la variabilidad de sus trayectorias, mientras que en las series de datos clásicos se produce una cierta pérdida de información. Por ejemplo, los precios de cierre de las acciones que se recogen al final del día en cada mercado, es decir, cada cierto periodo de tiempo, pero no recogen la información de los valores obtenidos a lo largo del día (Arroyo, Espinola, & Maté, 2011). Además, una observación de intervalos puede evitar los ruidos que se producen en los puntos de las frecuencias más altas de los intervalos observados. Por

ejemplo, al analizar la información de un intervalo recogiendo los precios máximos y mínimos en un día, se evitan ruidos indeseables que se producen durante ese día y, además, se adquiere más información que en observaciones puntuales.

Existen un gran número de métodos estadísticos para modelar y estimar procesos de intervalos desde que (Billard & Diday, 2000) propusieron un modelo de regresión lineal para los puntos centrales de los intervalos que fue más tarde desarrollado por (De Carvalho, Neto, & Tenorio, 2004) al modelar dos procesos puntuales univariantes para el centro y el radio. Más tarde, (Brito, 2007) propuso dos regresiones diferentes para los límites izquierdo y derecho de un intervalo, (Lima Neto & De Carvalho, 2008) consideraron un modelo puntual bivariante para los dos límites o para el centro y rango de un intervalo. Además, para garantizar que el límite izquierdo ajustado es menor que el límite derecho ajustado, (Lima Neto & De Carvalho, 2008) impusieron unas restricciones de no negatividad para los coeficientes de regresión del radio. Por último, (Gonzalez-Rivera & Lin, 2013) propusieron un modelo de regresión más general, restrictivo, bivariante y puntual para intervalos.

Sin embargo, existe un problema con los enfoques mencionados anteriormente, que se basan en modelado de series de datos puntuales, obtenidos a partir de una ITS. El problema es que puede ocurrir que no se haga un uso eficiente de la información que contiene el intervalo.

La información existente sobre modelado de intervalos ha considerado únicamente el modelado lineal de la media condicional de un proceso de intervalo, ignorando posibles no linealidades de la ITS. Las no linealidades, se reconocen ampliamente en un rango de series de datos puntuales, y se han propuesto diversos modelos de series de tiempo no lineales entre los que se incluyen los modelos Autorregresivos de Umbral (TAR), los modelos de Cadenas de Markov de Cambios de Régimen (MCRS) o los modelos Autorregresivos de Transición Suave (STAR). Sin embargo, existen estudios limitados que consideren el modelado de ITS con características no lineales. Una excepción es la de (Rodrigues & Salish, 2015) que propusieron el modelo CR-SETAR para capturar no

linealidades de ITS centrándose en su centro y radio, aunque este método de estimación puede que no utilice la información del intervalo de la manera más eficiente.

5.2 *MODELOS AUTORREGRESIVOS DE UMBRAL PARA SERIES DE DATOS DE INTERVALOS (TARI)*

5.2.1 *INTRODUCCIÓN*

El modelado y la predicción de datos simbólicos, especialmente con series de datos de intervalos (ITS), ha recibido una atención considerable en el campo de la estadística y en otros campos relacionados. Los métodos disponibles para el análisis de ITS se basan en diferentes aplicaciones del modelado lineal convencional. A continuación, según (Sun, Han, Hong, & Wang, 2018) se propone una nueva clase de modelos de autorregresión de umbral por intervalos (TARI), para modelar datos de ITS. De esta forma, al comparar modelos de intervalos con las observaciones de los intervalos, se desarrolla un método de estimación basado en la mínima distancia para los modelos TARI, y se establece la teoría asintótica para los estimadores propuestos.

La teoría asintótica, en estadística, es un marco para evaluar las propiedades de los estimadores y las muestras estadísticas. En este marco, se suele asumir que el tamaño de la muestra, n , puede crecer indefinidamente, por lo que las propiedades de los estimadores y las pruebas se evaluarán bajo el límite de $n \rightarrow \infty$. En la práctica, se considera que una evaluación del límite es aproximadamente válida también para muestras finitas, pero de gran tamaño.

Algunos análisis que se han realizado muestran que el modelo TARI propuesto a continuación ofrece una mayor precisión de pronóstico que el modelo existente de Centro-Radio Autorregresivo de Umbral y Auto-Excitante (CR-SETAR) para series de datos de intervalos.

Para modelar y capturar las características no lineales de un proceso de ITS, (Sun, Han, Hong, & Wang, 2018) propusieron los modelos TARI.

Para empezar, cabe recordar que un modelo TAR es un modelo de regresión lineal por partes, es decir, un modelo de cambios de régimen que consisten en una serie de partes que cambian de un proceso a otro según el valor de una variable observada, la variable umbral.

A partir de aquí, se explican los modelos TARI como una generalización para intervalos de los modelos TAR clásicos, utilizados para series de datos puntuales. Los modelos TARI tratan el intervalo como un conjunto y comparan el modelo de intervalo con los datos del intervalo observado. Se desarrolla entonces un método de estimación basado en la mínima distancia para estos modelos, y se establece la teoría asintótica para los estimadores propuestos.

Comparado con otros modelos de intervalos existentes, los modelos TARI tienen una serie de características llamativas. Para empezar, en comparación con el modelo CR-SETAR, un modelo TARI es parsimonioso, es decir, es lo más sencillo posible ya que considera la observación del intervalo como un conjunto en vez de modelar puntualmente el centro y el radio de este por separado, o los límites izquierdo y derecho. Gracias a esto, se permite proponer un estimador para el modelo TARI minimizando la distancia entre el modelo de intervalo y los datos del intervalo. De hecho, con un uso más eficiente de los datos del intervalo se obtienen asintóticamente unos procedimientos de inferencia más eficientes. Por otro lado, al introducir el concepto de intervalos extendidos, se relaja la restricción de que el límite izquierdo de un intervalo nunca sea mayor que el derecho, lo cual cubre una serie de aplicaciones tanto en economía como en finanzas. Además, un modelo TARI es capaz de detectar no linealidades como respuestas asimétricas ante algunas novedades o un efecto en la tendencia asimétrica.

5.2.2 DESARROLLO MODELO TARI

Existen numerosos trabajos de modelado y estimación de ITS. (Han, Hong, & Wang, 2017) propusieron un tipo de modelo de Intervalo Condicional lineal Autorregresivo (ACI) para el cual desarrollaron un método de estimación por mínima distancia. Suponiendo que $\{Y_t\}$ es una serie de tiempo de intervalo extendido y estacionaria, generalizada para modelar vectores de intervalo extendidos. Un intervalo extendido es una serie de números ordenados, donde el límite izquierdo debe ser menor que el límite derecho del intervalo. Un modelo ACI de orden (p,q) para el que $[Y_t]$ se especifica como:

$$[Y_t] = \alpha_0 + \beta_0[I_0] + \sum_{j=1}^p \beta_j[Y_{t-j}] + \sum_{j=1}^q \gamma_j[u_{t-j}] + [u_t], \quad t = 1, 2, \dots$$

Ecuación 5.1 Modelo ACI de orden (p,q)

donde α_0, β_j ($j = 0, \dots, p$) y γ_j ($j = 1, \dots, q$) son parámetros escalares desconocidos; $[I_0] = [-\frac{1}{2}, \frac{1}{2}]$ es un intervalo; $[u_t]$ es una innovación de intervalo con $E([u_t]|I_{t-1}) = [0,0]$, donde I_{t-1} es la información disponible en $t - 1$. Este modelo es una generalización para intervalos del popular modelo ARMA (p,q) para series de tiempo puntuales y puede ser utilizado para predecir procesos de intervalos.

Cabe destacar que en análisis de series de tiempo o en predicción, la innovación es la diferencia entre el valor observado de una variable en un tiempo t y la predicción óptima de este valor, basada en la información disponible anterior al tiempo t .

Sin embargo, como se ha explicado anteriormente, hay pocos estudios en la literatura que hayan considerado el estudio de posibles no linealidades de procesos de ITS, a diferencia de en series de tiempo puntuales, donde sí que se han desarrollado diversos modelos nombrados previamente.

Fueron (Rodrigues & Salish, 2015) quienes propusieron un modelo para capturar no linealidades en datos de ITS, el modelo CR-SETAR. A continuación, se muestra un modelo general CR-SETAR que identifica dos regímenes diferentes de una ITS:

$$Y_t^c = \begin{cases} \mu_1 + \sum_{i=1}^{p^c} \alpha_{1i} Y_{t-i}^c + \sum_{j=1}^{p^r} \beta_{1j} Y_{t-j}^r + u_t^c, & \text{si } Z_{t-1}^c \leq \tau^c \\ \mu_2 + \sum_{i=1}^{p^c} \alpha_{2i} Y_{t-i}^c + \sum_{j=1}^{p^r} \beta_{2j} Y_{t-j}^r + u_t^c, & \text{si } Z_{t-1}^c > \tau^c \end{cases}$$

Ecuación 5.2 Modelo CR-SETAR para el centro

$$Y_t^r = \begin{cases} \lambda_1 + \sum_{i=1}^{q^c} \phi_{1i} Y_{t-i}^c + \sum_{j=1}^{q^r} \phi_{1j} Y_{t-j}^r + u_t^r, & \text{si } Z_{t-1}^r \leq \tau^r \\ \lambda_2 + \sum_{i=1}^{q^c} \phi_{2i} Y_{t-i}^c + \sum_{j=1}^{q^r} \phi_{2j} Y_{t-j}^r + u_t^r, & \text{si } Z_{t-1}^r > \tau^r \end{cases}$$

Ecuación 5.3 Modelo CR-SETAR para el radio

donde Z_{t-1}^c y Z_{t-1}^r son las funciones del historial pasado del intervalo $[Y_t]$; Y_t^c es el centro de $[Y_t]$; Y_t^r es el radio de $[Y_t]$; u_t^c y u_t^r son las innovaciones puntuales para Y_t^c y Y_t^r , respectivamente. Los parámetros p^c , p^r , q^c y q^r son los órdenes de los retardos y τ^c y τ^r son los parámetros de umbral (o threshold), que determinan el momento en el que se produce el cambio de régimen.

El modelo CR-SETAR se basa en un sistema bivalente medido puntualmente para el centro y radio, donde cada una de las ecuaciones se divide en dos regímenes gracias a distintos valores de umbral que los delimitan.

A continuación, se muestra un modelo TARI de orden p que se utiliza para capturar posibles no linealidades en una ITS $[Y_t]$.

$$[Y_t] = \begin{cases} \alpha_{01} + \alpha_{11}[I_0] + \beta_{11}[Y_{t-1}] + \dots + \beta_{p1}[Y_{t-p}] + \iota_1[\varepsilon_t], & \text{si } q_t \leq \tau \\ \alpha_{02} + \alpha_{12}[I_0] + \beta_{12}[Y_{t-1}] + \dots + \beta_{p2}[Y_{t-p}] + \iota_2[\varepsilon_t], & \text{si } q_t > \tau \end{cases}$$

Ecuación 5.4 Modelo TARI de orden p

donde $[Y_t] = [Y_t^L, Y_t^R]$ es un proceso estocástico valorado por intervalos; β_{ji} , $j = 0, \dots, p$, $i = 1, 2$, son parámetros escalares desconocidos; ι_i , $i = 1, 2$ son otros parámetros desconocidos de escala; $[I_0] = [-\frac{1}{2}, \frac{1}{2}]$ es un intervalo; $\alpha_{0i} + \alpha_{1i}[I_0] = [\alpha_{0i} - \frac{1}{2}\alpha_{1i}, \alpha_{0i} + \frac{1}{2}\alpha_{1i}]$, $i = 1, 2$, es un intervalo constante; τ es un parámetro escalar desconocido que representa el valor de umbral en el que se cambia de régimen; q_t es una variable endógena o exógena; $[\varepsilon_t] = [\varepsilon_t^L, \varepsilon_t^R]$ es un intervalo de innovación.

Una diferencia clave entre el modelo TARI y el modelo bivalente puntual, CR-SETAR, es que el modelo TARI trata el intervalo como un conjunto, mientras que el modelo CR-SETAR trata a cada valor puntual como un conjunto diferente de parámetros, lo que puede hacer que pierda algunas de las características esenciales del intervalo.

Asumiendo $\iota_1 = \iota_2 = \sigma_u$, a partir de la ecuación anterior se puede escribir:

$$[Y_t] = \begin{cases} Z_t' \beta_1 + [u_t], & \text{si } q_t \leq \tau \\ Z_t' \beta_2 + [u_t], & \text{si } q_t > \tau \end{cases}$$

Ecuación 5.5 Modelo TARI asumiendo $\iota_1 = \iota_2 = \sigma_u$

donde $Z_t = \left([1, 1], \left[-\frac{1}{2}, \frac{1}{2} \right], Y_{t-1}, \dots, Y_{t-p} \right)'$, $\beta_i = (\alpha_{0i}, \alpha_{1i}, \beta_{1i}, \dots, \beta_{pi})' \in \mathcal{R}^{p+2}$, $i = 1, 2$, y $u_t \equiv \sigma_u \varepsilon_t$.

Considerando el caso anterior y asumiendo la misma intersección del intervalo para los dos regímenes, se puede escribir la misma ecuación de la siguiente forma:

$$[Y_t] = \begin{cases} \alpha_0 + \beta_0[I_0] + Z_t' \beta_1 + [u_t], & \text{si } q_t \leq \tau \\ \alpha_0 + \beta_0[I_0] + Z_t' \beta_2 + [u_t], & \text{si } q_t > \tau \end{cases}$$

Ecuación 5.6 Modelo TARI asumiendo $\iota_1 = \iota_2 = \sigma_u$, desarrollado

siendo $Z_t = (Y_{t-1}, \dots, Y_{t-p})'$, $\beta_i = (\beta_{1i}, \beta_{2i}, \dots, \beta_{pi})' \in \mathcal{R}^p, i = 1, 2$.

El modelo TARI(p) es una generalización para intervalos del modelo TAR(p), utilizado para series de datos clásicos. En el caso de TARI, las observaciones de intervalos se dividen en dos regímenes, dependiendo de si la variable umbral q_t es menor o mayor que el umbral γ . Los regímenes se asocian con los coeficientes de autorregresión β_1 y β_2 , respectivamente. En economía y finanzas, los modelos TARI son utilizados para detectar momentos de no linealidades en una ITS. Un modelo TARI(p) ofrece una variedad de flexibilidad a partir de diferentes selecciones de q_t .

Un modelo TARI(p) puede extenderse a un modelo TARIX (p,s) al incorporar variables exógenas explicativas de intervalo:

$$[Y_t] = \begin{cases} \alpha_{01} + \beta_{01}[I_0] + \sum_{j=1}^p \beta_{j1}[Y_{t-j}] + \sum_{l=0}^s \delta'_{l1}[X_{t-j}] + [u_t], & \text{si } q_t \leq \tau \\ \alpha_{02} + \beta_{02}[I_0] + \sum_{j=1}^p \beta_{j2}[Y_{t-j}] + \sum_{l=0}^s \delta'_{l2}[X_{t-j}] + [u_t], & \text{si } q_t > \tau \end{cases}$$

Ecuación 5.7 Modelo TARIX de orden (p,s)

donde $X_t = (X_{1t}, \dots, X_{qt})'$ es un proceso vectorial exógeno de intervalo estrictamente estacionario y $\delta_{ji} = (\delta_{l1i}, \dots, \delta_{lqi})'$, con $l = 0, \dots, s$ e $i = 1, 2$ es un vector de parámetros puntuales desconocidos.

Intuitivamente, un modelo TARI parsimonioso proporciona un marco unificado que puede ser útil para deducir la dinámica de una población de intervalos. Esto puede servir para obtener algunos modelos puntuales como casos especiales. Por ejemplo, al coger la diferencia entre los límites izquierdo y derecho de un modelo TARI(p), se obtiene un modelo TAR(p) para el radio de $[Y_t]$:

$$Y_t^r = \begin{cases} \alpha_{11} + \sum_{j=1}^p \beta_{j1} Y_{t-j}^r + u_t^r, & \text{si } q_t \leq \tau \\ \alpha_{12} + \sum_{j=1}^p \beta_{j2} Y_{t-j}^r + u_t^r, & \text{si } q_t > \tau \end{cases}$$

Ecuación 5.8 Modelo TAR(p) para el radio

donde u_t^r es la secuencia de diferencias martíngala (MDS) tal que $E(u_t^r | I_{t-1}) = 0$ casi seguro, dado $E([u_t] | I_{t-1}) = [0,0]$ casi seguro.

Cabe destacar que una serie estocástica C es una MDS si su esperanza con respecto al pasado es 0.

De manera similar, es posible obtener un modelo TAR(p) para el centro de $[Y_t]$:

$$Y_t^c = \begin{cases} \alpha_{01} + \sum_{j=1}^p \beta_{j1} Y_{t-j}^c + u_t^c, & \text{si } q_t \leq \tau \\ \alpha_{02} + \sum_{j=1}^p \beta_{j2} Y_{t-j}^c + u_t^c, & \text{si } q_t > \tau \end{cases}$$

Ecuación 5.9 Modelo TAR(p) para el centro

donde u_t^c es una MDS tal que $E(u_t^c | I_{t-1}) = 0$ casi seguro, dado $E([u_t] | I_{t-1}) = [0,0]$, casi seguro.

Finalmente, se puede obtener un modelo TAR bivalente de datos puntuales con la misma variable umbral en los dos regímenes de $[Y_t]$:

$$Y_t^c = \begin{cases} \alpha_{01} + \sum_{j=1}^p \beta_{j1} Y_{t-j}^c + u_t^c, & \text{si } q_t \leq \tau \\ \alpha_{02} + \sum_{j=1}^p \beta_{j2} Y_{t-j}^c + u_t^c, & \text{si } q_t > \tau \end{cases}$$

$$Y_t^r = \begin{cases} \alpha_{11} + \sum_{j=1}^p \beta_{j1} Y_{t-j}^r + u_t^r, & \text{si } q_t \leq \tau \\ \alpha_{12} + \sum_{j=1}^p \beta_{j2} Y_{t-j}^r + u_t^r, & \text{si } q_t > \tau \end{cases}$$

Ecuación 5.10 Modelo TAR bivalente

donde $E(u_t^c | u_t^r) = [0,0]$, casi seguro, dado $E([u_t] | I_{t-1}) = [0,0]$, casi seguro.

En comparación con el modelo CR-SETAR, el modelo TARI para Y_t^c e Y_t^r , es más parsimonioso. Tiene la misma variable umbral para Y_t^c e Y_t^r , y un menor número de parámetros desconocidos, lo cual se explica por el hecho de que trata el intervalo $[Y_t]$ como un conjunto. Sin embargo, no hay restricciones para los coeficientes de las regresiones del centro y radio en el modelo CR-SETAR, que difícilmente asegura que los valores previstos para los límites inferiores serán menores que los valores previstos para los límites superiores de los intervalos.

Por último, gracias al tratamiento de los intervalos como un conjunto, se propone un método de estimación para el modelo TARI que consiste en minimizar la distancia entre el modelo del intervalo y el intervalo de datos real.

5.2.3 ESTIMACIÓN MODELO TARI

De acuerdo con lo explicado previamente, se va a evaluar un método de estimación por mínima distancia para el modelo TARI. Además, se van a investigar las propiedades asintóticas de los estimadores propuestos.

Para comenzar, se va a considerar la forma en la que se puede medir la distancia entre dos intervalos. Se propone la distancia D_K como una medida para la distancia entre dos conjuntos de dimensión arbitraria, lo cual incluye los intervalos como un caso especial. A continuación, se muestra un ejemplo de medida de la distancia D_K entre el intervalo observado $[Y_t]$ y el intervalo ajustado $[\hat{Y}_t]$:

$$D_K^2([Y_t], [\hat{Y}_t]) = \int_{(u,v) \in S^0} [s_{[Y_t]}(u) - s_{[\hat{Y}_t]}(u)][s_{[Y_t]}(v) - s_{[\hat{Y}_t]}(v)] dK(u, v) =$$

$$\langle s_{[Y_t] - [\hat{Y}_t]}, s_{[Y_t] - [\hat{Y}_t]} \rangle_K = \| [Y_t] - [\hat{Y}_t] \|^2 = \| [u_t] \|^2$$

Ecuación 5.11 Distancia D_K entre el intervalo observado y el intervalo ajustado

donde $S^0 = \{u \in \mathbb{R}^1, |u| = 1\} = \{1, -1\}$, $K(u, v)$ es una función de ponderación simétrica positiva definida en S^0 para asegurar que $D_K([Y_t], [\widehat{Y}_t])$ es una medida para intervalos extendidos.

(Sun, Han, Hong, & Wang, 2018) proponen un modelo de estimación basado en la mínima distancia D_K para el parámetro umbral γ , el parámetro de retraso d y los parámetros de la pendiente β es un modelo TARI con una variable umbral endógena, es decir, que su valor depende de la evolución del modelo. También se pueden obtener resultados similares en un modelo TARI con una variable umbral exógena, cuyo valor no depende de la evolución del modelo.

Estimación de tau basada en la mínima distancia D_K

Considerando el caso de $\iota_1 = \iota_2 = \sigma_u$ y $q_t = Y_{t-d}^c$. La Ecuación 5.5 se puede escribir como:

$$[Y_t] = \begin{cases} Z_t' \beta_1 + [u_t], & \text{si } Y_{t-d}^c \leq \tau \\ Z_t' \beta_2 + [u_t], & \text{si } Y_{t-d}^c > \tau \end{cases}$$

Ecuación 5.12 Modelo TARI considerando $\iota_1 = \iota_2 = \sigma_u$ y $q_t = Y_{t-d}^c$

donde $Z_t = \left([1, 1], \left[-\frac{1}{2}, \frac{1}{2} \right], Y_{t-1}, \dots, Y_{t-p} \right)'$, $\beta_i = (\alpha_{0i}, \alpha_{1i}, \beta_{1i}, \dots, \beta_{pi})' \in \mathcal{R}^{p+2}$, $i = 1, 2$, e $Y_{t-d}^c = \frac{1}{2}(Y_{t-d}^L + Y_{t-d}^R)$. Se define ahora el estimador de mínima distancia D_K :

$$\hat{\theta} = \arg \min_{\theta \in \Theta} Q_T(\theta)$$

Ecuación 5.13 Estimador de la mínima distancia D_K

siendo $(T - p)$ $Q_T(\theta)$ la suma de los cuadrados de los residuos del modelo TARI(p) con respecto a la distancia D_K , sabiendo que:

$$Q_T(\theta) = \frac{1}{T-p} \sum_{t=p+1}^T \phi_t(\theta)$$

Ecuación 5.14 Ecuación de $Q_T(\theta)$

donde $\phi_t(\theta) = \| [u_t] \|_K^2$.

A partir de las ecuaciones anteriores, se obtiene que el minimizador $\hat{\theta}$ de $Q_T(\theta)$ es el estimador de mínima distancia D_K de θ_0 , esto es:

$$\hat{\theta} = \arg \min_{\theta} Q_T(\theta) = \arg \min_{\theta} \left[\frac{1}{T-p} \sum_{Y_{t-d}^c \leq \gamma} \| [Y_t] - Z'_t \beta_1 \|_K^2 + \frac{1}{T-p} \sum_{Y_{t-d}^c > \gamma} \| [Y_t] - Z'_t \beta_2 \|_K^2 \right]$$

Ecuación 5.15 Minimizador $\hat{\theta}$

Por tanto, $\hat{\theta}$ se obtiene al minimizar la media de los promedios al cuadrado de los residuos de los intervalos. Analizando la ecuación anterior, se observa que si la variable umbral Y_{t-d}^c es menor o igual que γ , la función que se obtiene es la distancia D_K entre el intervalo real $[Y_t]$ y $Z'_t \beta_1$. En caso contrario, es decir, si la variable umbral Y_{t-d}^c es mayor que γ , se mide la distancia D_K entre el intervalo real $[Y_t]$ y $Z'_t \beta_2$. Como la función de regresión es no lineal y discontinua, se emplea un método de estimación condicional secuencial para obtener $\hat{\theta}$. Inicialmente, para unos valores datos de τ y d , se minimiza Q_T y se obtiene el estimador $\hat{\theta}(\tau, d)$, como se muestra a continuación:

$$\hat{\theta}(\tau, d) = \left(\frac{1}{T-p} \sum_{t=p+1}^T \langle S_{Z_t^\tau}, S'_{Z_t^\tau} \rangle_K \right)^{-1} \left(\frac{1}{T-p} \sum_{t=p+1}^T \langle S_{Z_t^\tau}, S_{Y_t} \rangle_K \right)$$

Ecuación 5.16 Estimador $\hat{\theta}(\tau, d)$,

donde $Z_t^\tau = ([1, 1], [I_0], \dots, Y_{t-p})'$.

Además, los estimadores de γ y d son los valores que minimizan la Ecuación 5.14:

$$(\hat{\tau}, \hat{d}) = \arg \min_{\theta} Q_T(\theta) |_{\beta=\hat{\theta}(\tau, d)}$$

Ecuación 5.17 Ecuación de $(\hat{\tau}, \hat{d})$

El estimador de θ_0 se obtiene, por tanto, como $\hat{\theta} = \hat{\theta}(\hat{\tau}, \hat{d}) = (\hat{\beta}', \hat{\tau}, \hat{d})'$, donde $\hat{\beta} = \hat{\beta}(\hat{\tau}, \hat{d})$. De acuerdo con este procedimiento, se puede demostrar que $Q_T(\theta)$ alcanza su mínimo global cuando $\theta = \hat{\theta}$. Finalmente, una vez que se haya obtenido el valor de $\hat{\theta}$, se estimará el valor de σ_u^2 .

$$\hat{\sigma}_u^2 = \frac{1}{T_1} \sum_{t=p+1}^T \| [Y_t] - Z'_t \beta_1 \|_K^2 + \frac{1}{T-p} \sum_{Y_{t-d}^c > \gamma} \| [Y_t] - Z'_t \beta_2 \|_K^2$$

$$\hat{\sigma}_u^2 = \begin{cases} \frac{1}{T_1} \sum_{t=p+1}^T \| [Y_t] - Z'_t \hat{\beta}_1 \|_K^2, & \text{si } Y_{t-d}^c \leq \hat{\tau} \\ \frac{1}{T_2} \sum_{t=p+1}^T \| [Y_t] - Z'_t \hat{\beta}_2 \|_K^2, & \text{si } Y_{t-d}^c > \hat{\tau} \end{cases}$$

Ecuación 5.18 Ecuación de $\hat{\sigma}_u^2$

donde $T_1 = \sum_{t=p+1}^T$ si $Y_{t-d}^c \leq \hat{\tau}$ y $T_2 = \sum_{t=p+1}^T$ si $Y_{t-d}^c > \hat{\tau}$. La estimación de σ_u^2 se puede aplicar a $\hat{l}_1$ y $\hat{l}_2$.

5.2.4 APLICACIÓN EMPÍRICA DEL ÍNDICE S&P500

En su trabajo, (Sun, Han, Hong, & Wang, 2018) realizan una aplicación del modelo TARI al índice de precio S&P500. El índice S&P 500 es uno de los índices bursátiles más importantes de Estados Unidos y es considerado el índice más representativo de la situación real del mercado. Se basa en la capitalización bursátil de 500 grandes empresas que poseen acciones que cotizan en las bolsas NYSE o NASDAQ, y captura aproximadamente el 80% de toda la capitalización del mercado de Estados Unidos.

(Sun, Han, Hong, & Wang, 2018) estudian esta aplicación con el objetivo de capturar fenómenos de no linealidades y de comparar las actuaciones a la hora de predecir entre un modelo TARI y otros modelos alternativos de intervalos, para el índice de precio semanal del S&P 500.

Predicción del índice de precio semanal del S&P 500: TARI vs. CR-SETAR

Se analiza la precisión de la predicción fuera de la muestra de un modelo TARI en comparación con otros modelos de intervalos. En concreto, se compara con el modelo CR-SETAR.

Basado en los patrones del centro y radio en el intervalo de muestra, se incluyen cuatro retardos en la variable de intervalo del modelo TARI. De hecho, las estimaciones de los parámetros para los mismos valores en los modelos del centro y radio podrían tener signos diferentes si no se aplicaran restricciones. Concretamente, para los datos de la muestra utilizados en (Rodrigues & Salish, 2015), basándose en un método de Mínimos Cuadrados Condicionales (CLS), los centros de la variable que causa la regresión, Y_{t-1} , y de la variable dependiente de la regresión, es decir $[Y_t]$, están correlados negativamente, mientras que los radios están correlados positivamente. Es por eso por lo que se considera la variable de intervalo “inversa” $[\widetilde{Y}_t] = [-Y_t^R, -Y_t^L]$, como una forma para reemplazar $[Y_t]$.

Por tanto, se obtiene un modelo TARI transformado de la siguiente forma:

$$\Delta Y_t = \begin{cases} \alpha_0 + \beta_0[I_0] + Z_t\beta_1 + [u_t], & \text{si } \Delta Y_{t-1}^c \leq \tau \\ \alpha_0 + \beta_0[I_0] + Z_t\beta_2 + [u_t], & \text{si } \Delta Y_{t-1}^c > \tau \end{cases}$$

Ecuación 5.19 Modelo TARI transformado

donde $\Delta Y_t = \{Y_t - Y_{t-1} : [Y_t] = [Y_t^L, Y_t^R]\}$, $Z_t = [\Delta \widetilde{Y}_{t-1}, \Delta \widetilde{Y}_{t-2}, \Delta \widetilde{Y}_{t-3}, \Delta \widetilde{Y}_{t-4}]$, $\Delta \widetilde{Y}_t = [Y_{t-1}^R - Y_t^R, Y_{t-1}^L - Y_t^L]$ y ΔY_{t-1}^c es el punto medio de ΔY_{t-1} . Los coeficientes de $\Delta \widetilde{Y}_{t-1}$ en los modelos de centros y radios ahora tienen los mismos signos.

El modelo CR-SETAR realiza la predicción del centro y del rango de una ITS. En comparación con el modelo CR-SETAR, el modelo TARI es parsimonioso, es decir, es

mucho más sencillo, ya que el primero emplea coeficientes diferentes para cada ecuación como se muestra a continuación:

$$\Delta Y_t^c = \begin{cases} \mu_1 + \sum_{i=1}^4 \alpha_{1i} \Delta Y_{t-i}^c + \sum_{j=1}^4 \beta_{1j} Y_{t-j}^r + [\varepsilon_t], & \text{si } \Delta Y_{t-1}^c \leq \tau^c \\ \mu_2 + \sum_{i=1}^4 \alpha_{2i} \Delta Y_{t-i}^c + \sum_{j=1}^4 \beta_{2j} Y_{t-j}^r + [\varepsilon_t], & \text{si } \Delta Y_{t-1}^c > \tau^c \end{cases}$$

Ecuación 5.20 CR-SETAR para los centros

$$Y_t^r = \begin{cases} \lambda_1 + \sum_{i=1}^4 \phi_{1i} Y_{t-i}^c + \sum_{j=1}^4 \phi_{1j} Y_{t-j}^r + [\xi_t], & \text{si } \Delta Y_{t-1}^r \leq \tau^r \\ \lambda_2 + \sum_{i=1}^4 \phi_{2i} Y_{t-i}^c + \sum_{j=1}^4 \phi_{2j} Y_{t-j}^r + [\xi_t], & \text{si } \Delta Y_{t-1}^r > \tau^r \end{cases}$$

Ecuación 5.21 CR-SETAR para los radios

donde ΔY_t^c es el centro de $\Delta Y_t = [Y_t^L - Y_{t-1}^L, Y_t^R - Y_{t-1}^R]$ e $Y_t^r = \frac{1}{2}(Y_t^R - Y_t^L)$ es el radio del intervalo.

Para estudiar la actuación de la predicción, se dividen inicialmente los índices de precio semanales, por intervalos (de máximos y mínimos) del S&P 500, observando los valores desde el 3 de enero de 2003 hasta el 12 de noviembre de 2010. Para ello, se dividen en dos conjuntos diferentes: el primero de ellos, comenzando el 3 de enero de 2003 hasta el 28 de diciembre de 2007, utilizado para estimar los parámetros del modelo; y el segundo de ellos, comenzando el 4 de enero de 2008 y terminando el 12 de noviembre de 2010, utilizado para evaluar la predicción realizada. Es interesante destacar que el corte entre un periodo y el otro se realiza en el momento en el que comienza la crisis económica mundial de 2008.

Para empezar, se realiza la comparación de ambos modelos, y se obtienen las siguientes gráficas:

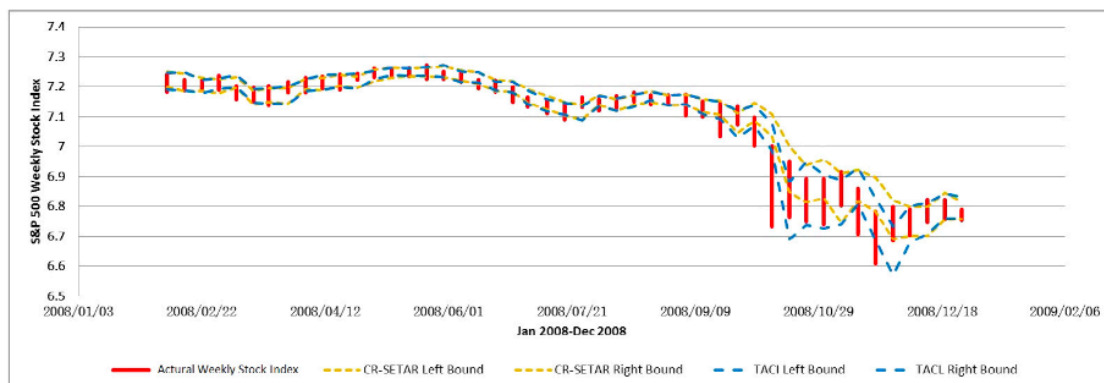


Figura 5.1 Resultados de las predicciones de los modelos TARI y CR-SETAR (1) (Sun, Han, Hong, & Wang, 2018)

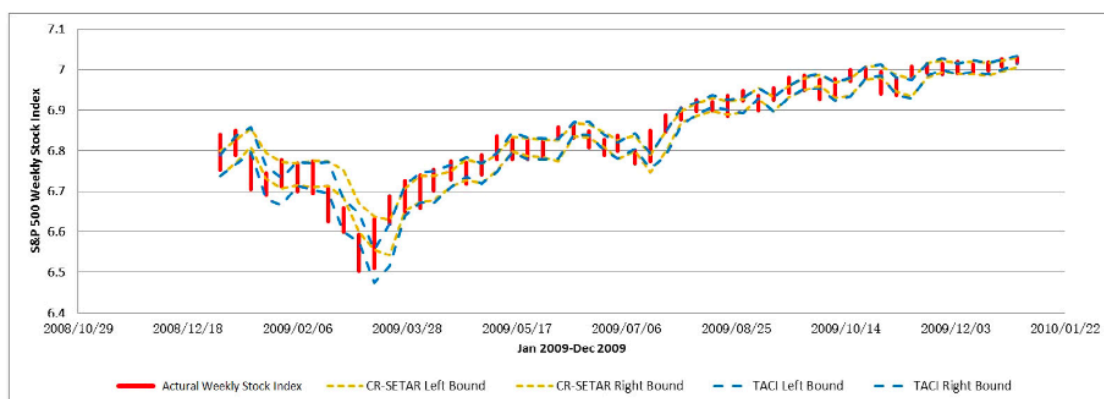


Figura 5.2 Resultados de las predicciones de los modelos TARI y CR-SETAR (2) (Sun, Han, Hong, & Wang, 2018)

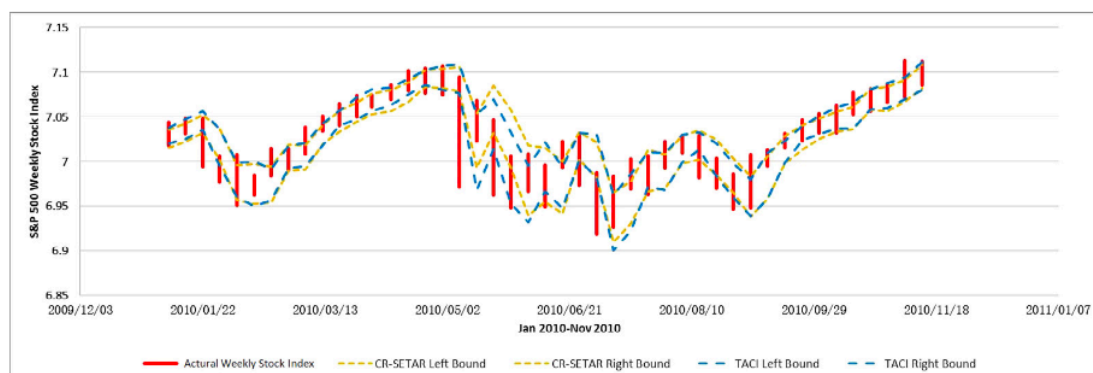


Figura 5.3 Resultados de las predicciones de los modelos TARI y CR-SETAR (3) (Sun, Han, Hong, & Wang, 2018)

De las representaciones anteriores se pueden obtener una serie de conclusiones. Para empezar, se puede observar que ambos modelos tienen actuaciones similares al predecir puntos de inflexión. En algunas semanas en particular, como el 17 de octubre de 2008 o el 13 de marzo de 2009, se observa que el modelo TARI es mejor que el CR-SETAR en términos de precisión de pronóstico, particularmente durante periodos bajistas del mercado de valores. Sin embargo, en los periodos alcistas, los resultados de los dos modelos muestran que las predicciones del índice de precio de las acciones van por detrás de los datos reales, esto ocurre el 10 de octubre de 2008 o el 21 de noviembre de ese mismo año, entre otros.

5.2.5 CONCLUSIONES

En este capítulo se ha estudiado un modelo TAR aplicado a series de tiempo de intervalos, desarrollando para el mismo un método de estimación por mínima distancia basado en comparar el modelo de intervalo con los datos de intervalo observados. Se muestra que el estimador del parámetro umbral es asintóticamente independiente de otros estimadores de parámetros del modelo. Se confirma además la teoría.

Se demuestra también que el modelo TARI mejora las predicciones fuera de muestra de varios modelos competidores para el rendimiento de acciones. Por último, se concluye que el modelo TARI propuesto supera el modelo existente no lineal de CR-SETAR para datos de ITS.

5.3 EJEMPLOS FUNCIÓN TAR (FUNCIÓN 1)

Para comprender de una forma más clara toda la explicación anterior, se ha desarrollado una función en Matlab que permite ilustrar estos conceptos de forma más visual.

Para empezar, se parte de una función para calcular la variable umbral según (Järås & Mohammadipour Gishani, 2010). Dicha función consiste en introducir una serie de tiempo

de datos clásicos y las fechas correspondientes a los mismos. A partir de ahí, la función estudia la serie de tiempo para encontrar en qué punto se produce el cambio de régimen.

Ejemplo 1 – taus para serie de tiempo de datos clásicos

Para ver un ejemplo, la función hará un análisis de los residuos obtenidos después de haber hecho uso del método de regresión. La función devuelve como resultado el punto de cambio de régimen (tau), los coeficientes (b) del modelo de autorregresión estimados en cada uno de los dos regímenes que se calculan, el residuo (minr) correspondiente al umbral estimado y las dos series de tiempo que se obtienen una vez calculada la variable umbral, y1 para el régimen menor e y2 para el régimen mayor.

Además, la función devuelve gráficamente dos figuras.

- Figura A: muestra los valores de los residuos correspondientes a los valores de la serie de tiempo $y(t)$ que se ha introducido, habiendo ordenado previamente los valores de $y(t)$ de menor a mayor para poder obtener una representación más clara de los resultados.
- Figura B: va a mostrar dos gráficas diferentes. La primera representa la serie de tiempo introducida $y(t)$. La gráfica siguiente muestra cada uno de los dos regímenes, y1 e y2, en los que se ha dividido la serie original.

A continuación, se muestra un ejemplo gráfico de lo que se acaba de explicar. Partiendo de los datos diarios del índice los precios de las acciones de la compañía farmacéutica BCRX y, en concreto, los valores desde el 3 de marzo de 2020 hasta el 12 de marzo de 2021. Se introduce en la función la serie de tiempo de los datos máximos de estos periodos y devuelve la siguiente información.

tau	8,98
b	$\begin{bmatrix} 0,0028 & 1,0055 \\ 1,5749 & 0,8632 \end{bmatrix}$

minr	37,177
-------------	--------

Tabla 7 Datos de la función (Elaboración propia)

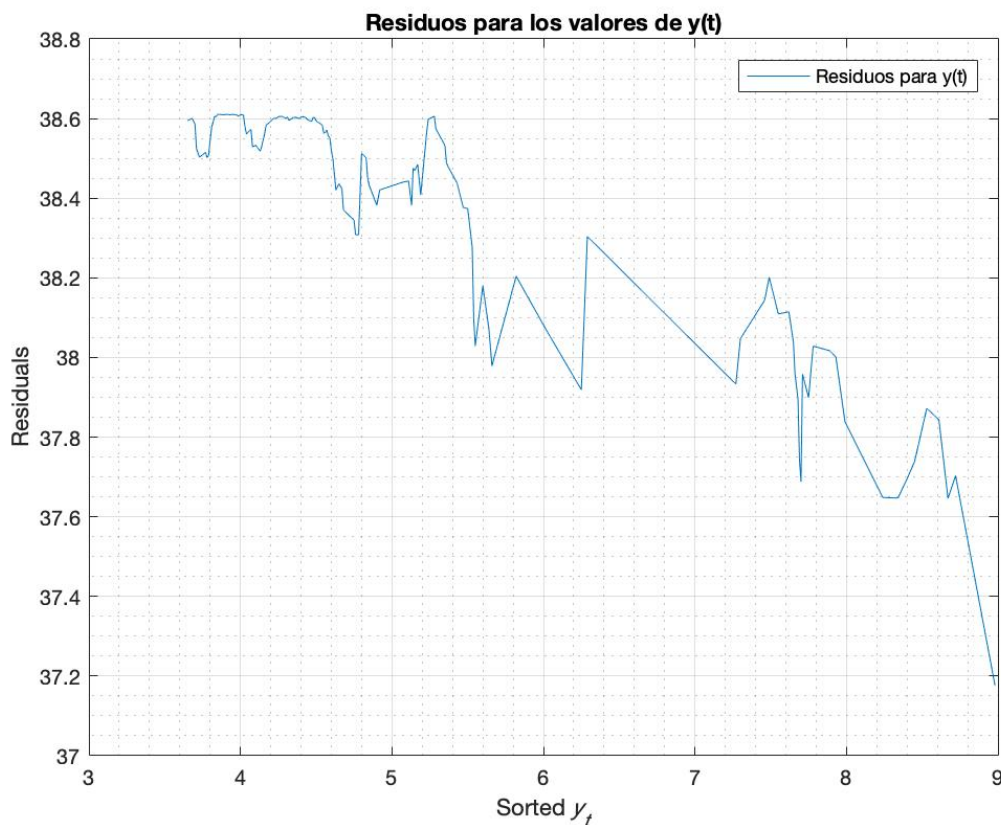


Figura 5.4 Valores de los residuos para $y(t)$ (Elaboración propia)

De esta figura, se puede derivar la obtención del tau ya que aprecia el residuo mínimo, obtenido también en la tabla ($\text{minr} = 37,177$) y, además, muestra el valor de $y(t)$ con el que se corresponde, es decir, el valor de tau, que como se ha obtenido en la tabla es igual a 8,98.

Por otro lado, se obtiene la segunda figura a continuación.

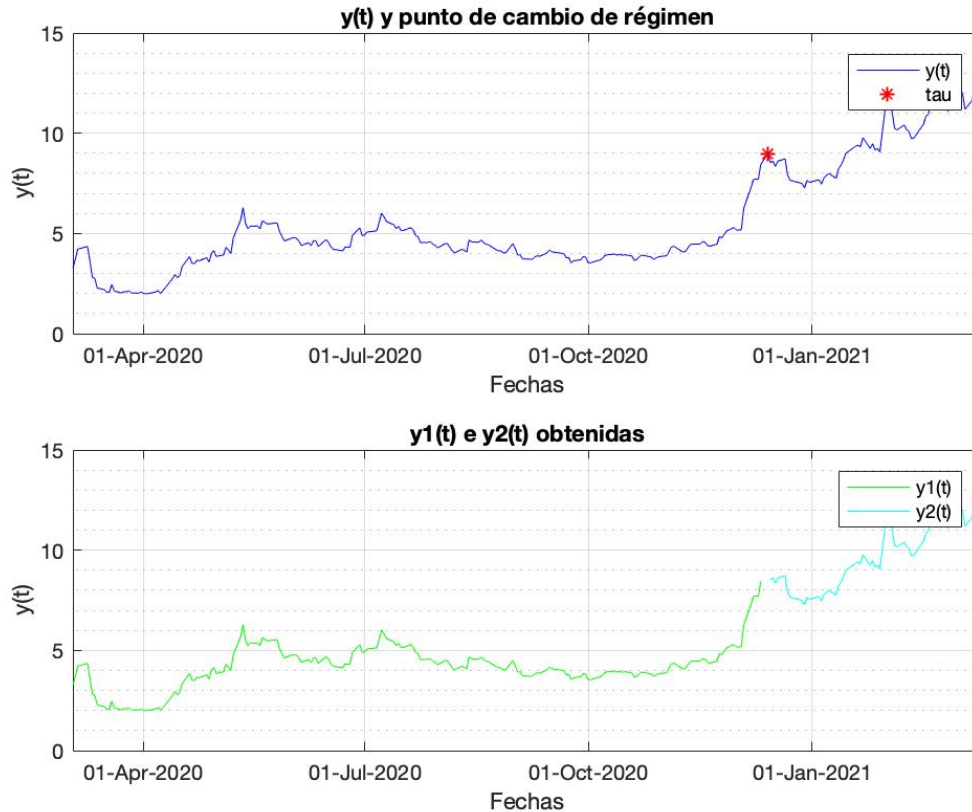


Figura 5.5 Serie de tiempo original y serie con los dos regímenes (Elaboración propia)

Como se ha explicado previamente, en la representación superior se encuentra la serie de tiempo y se indica el tau en el que se produce un cambio de régimen y en la inferior, se observan con dos colores diferentes, las series de tiempo y_1 e y_2 que se han obtenido una vez calculado el tau.

Ejemplo 2 – taus para serie de tiempo de datos de intervalo

Todo lo anterior resulta de gran utilidad para poder seguir progresando con el trabajo. Analizar con la función anterior las series de tiempo con datos puntuales permite estudiar y entender más profundamente algunos de los comportamientos de estas, sin embargo, dentro del alcance de este proyecto se encuentran las series de datos de intervalos. Es por eso por lo que es de especial interés extender la aplicación de la función anterior a los intervalos. Para ello, una vez desarrollada la función, se estudia cómo poder aplicarla con las ITS.

Para poder trabajar con la nueva función, se deberán introducir en ella los valores máximos y mínimos del intervalo en cuestión, durante el periodo de tiempo que sea de interés. Una vez seleccionados estos datos, la función calculará sus centros y sus radios y a partir de ellos podrá obtener conclusiones de mayor profundidad.

Además, a diferencia de la anterior, esta función será capaz de calcular más de un tau, es decir, más de un cambio de régimen, ya que en cualquier serie de tiempo que se estudie, es interesante calcular todos los puntos en los que se produce el cambio de régimen, para poder adquirir una mejor comprensión y nuevas ideas sobre las tendencias y sus cambios, para poder aplicarlas en el futuro.

En este caso, se deberán introducir en la función, además de la serie de tiempo a analizar, el número 't' de taus que el usuario quiere que calcule su función, además de un número de días 's' que indica la separación mínima, que el usuario quiere que exista entre dos cambios de régimen diferentes. Este valor lo elegirá el usuario en función de la estabilidad del mercado que esté estudiando. Cuanto más estable sea el mercado, mayor podrá ser el valor de s ya que no habrá cambios de régimen muy seguidos. Sin embargo, hay ciertos sectores en los que la estabilidad es muy variable y, por tanto, poner un valor de s muy elevado supondría la no detección de un gran número de cambios de régimen, por lo que los resultados obtenidos no serían de gran valor.

Se deberán analizar los taus resultantes para poder ver si en las distintas series de tiempo se obtienen taus en las mismas fechas, lo cual aumentaría la probabilidad de que, en el conjunto del intervalo, se produzcan cambios de régimen en un momento determinado.

Además, la función es capaz de calcular y comparar al mismo tiempo dos series de datos diferentes por lo que, para este ejemplo, tomamos los valores de las acciones de BBVA y del Banco Santander a lo largo de un mismo periodo.

Para empezar, la función va a devolver dos gráficos iBoxPlot, (Maté, 2021) correspondientes a cada uno de los bancos, en los que se indican:

- Intervalo mínimo: desde el valor mínimo de los mínimos hasta el valor mínimo de los máximos.
- Intervalo máximo: incluye desde el valor máximo de los mínimos hasta el valor máximo de los máximos.
- Intervalo de la mediana: desde la mediana de los límites inferiores hasta la mediana de los límites superiores.
- Intervalo del primer cuartil (Q1): desde el primer cuartil de los límites inferiores hasta el primer cuartil de los límites superiores.
- Intervalo del tercer cuartil (Q3): desde el tercer cuartil de los límites inferiores hasta el tercer cuartil de los límites superiores.

Gracias a esta representación se puede ver la posible presencia de valores atípicos cuando haya cambios abruptos al principio o al final de la pendiente que une los valores máximos y/o la pendiente que une los valores mínimos de cada uno de los intervalos.

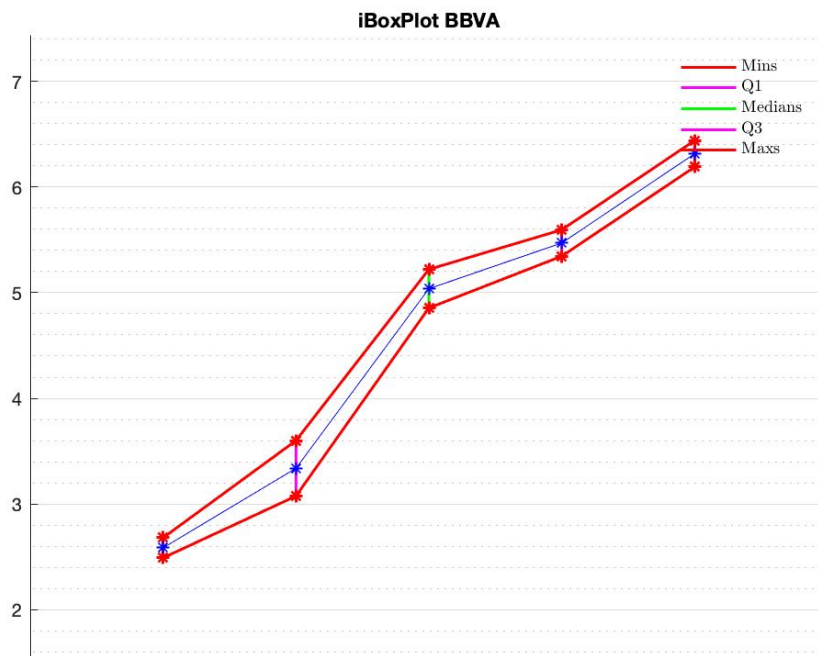


Figura 5.6 iBoxPlot BBVA (Elaboración propia)

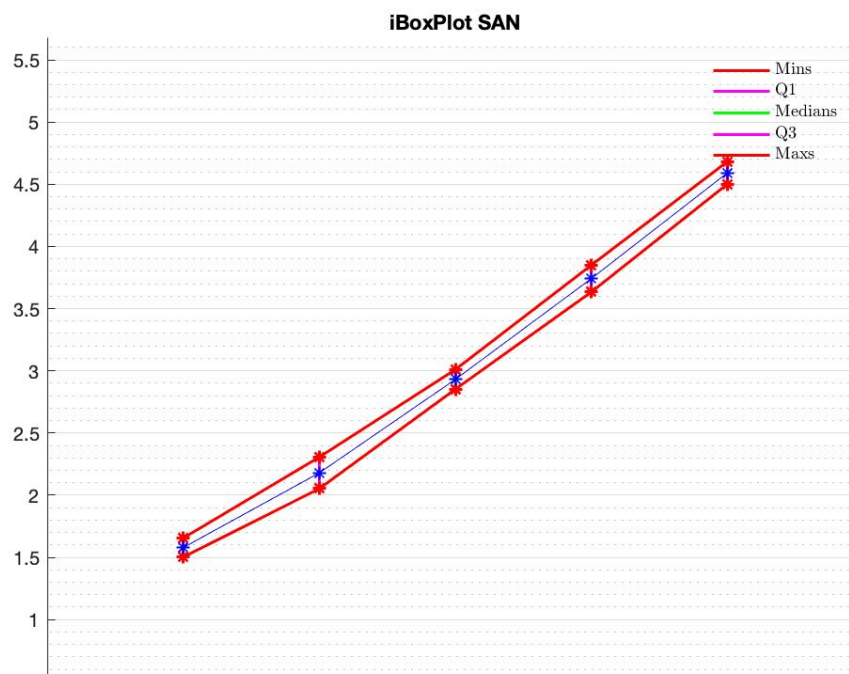


Figura 5.7 iBoxPlot Santander (Elaboración propia)

De las figuras anteriores, se observa que mientras que en la del Santander las pendientes superiores e inferiores parecen ser bastante continuas, en el caso de BBVA llaman la atención ciertos cambios abruptos que se producen en ellas, destacando, por tanto, que es muy probable la presencia de datos atípicos en los precios de las acciones de BBVA durante el periodo estudiado.

A continuación, cabe destacar que para ver el ejemplo se ha escogido una ‘t’ igual a 10 taus y, además, la separación mínima entre estos taus, ‘s’, se ha elegido igual a 5 días.

Sabiendo esto, se obtienen las siguientes gráficas que van a mostrar las ITS de los máximos, mínimos, centros y radios de ambos bancos junto con los taus encontrados en cada una de ellas.

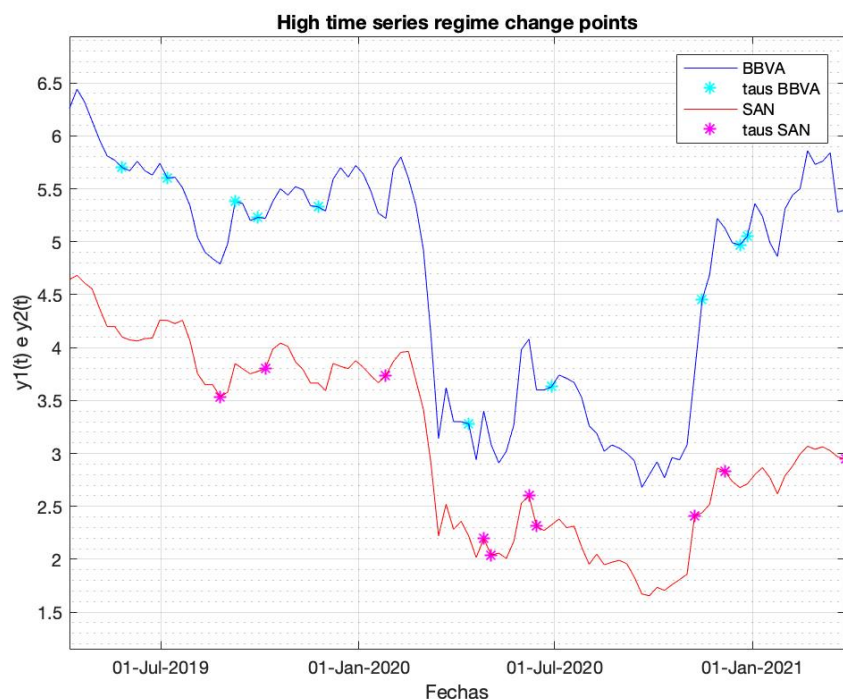


Figura 5.8 Cambios de régimen en las ITS de los valores máximos (Elaboración propia)

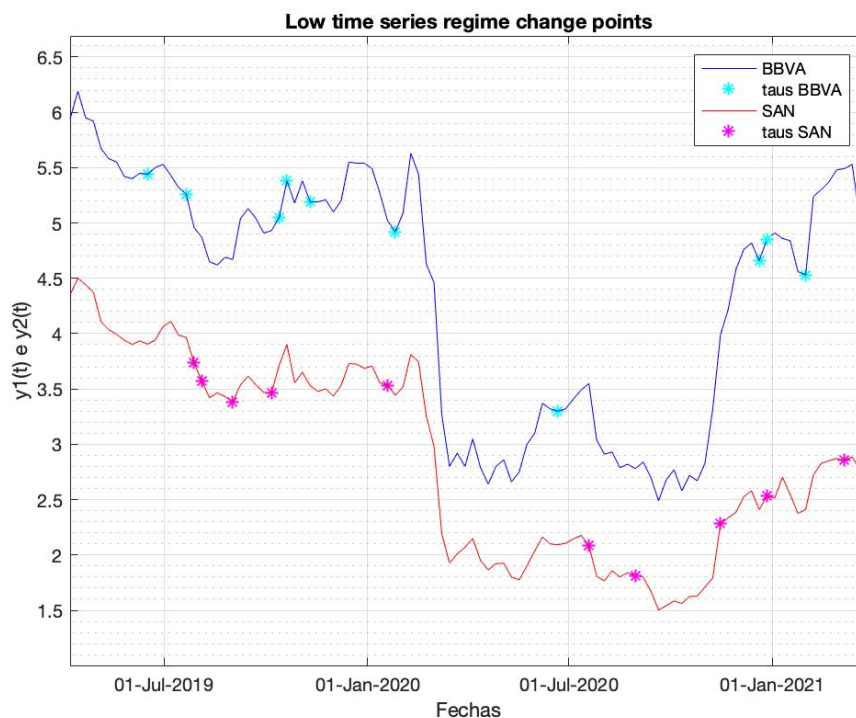


Figura 5.9 Cambios de régimen en las ITS de los valores mínimos (Elaboración propia)

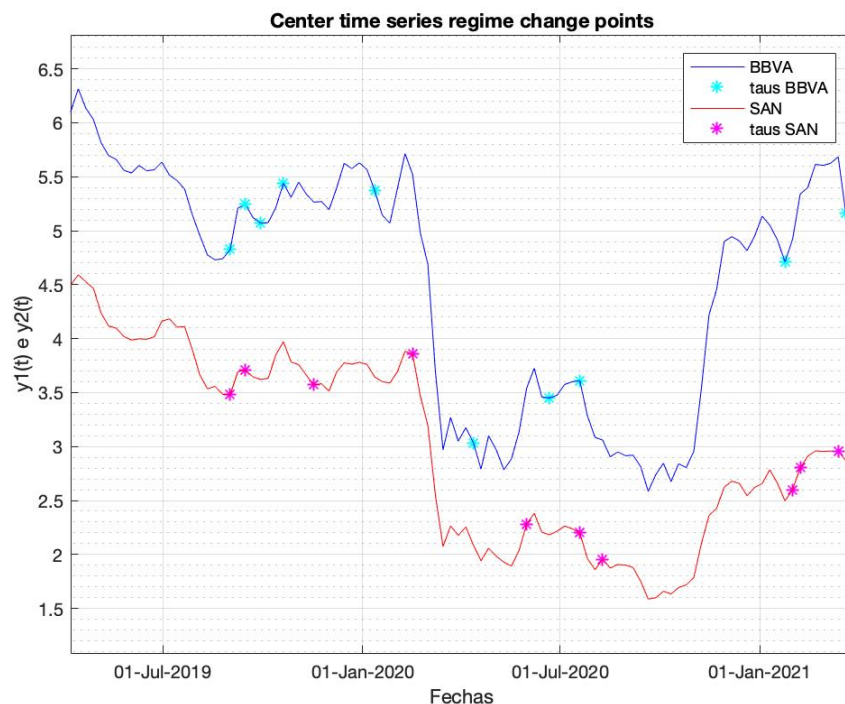


Figura 5.10 Cambios de régimen en las ITS de los centros (Elaboración propia)

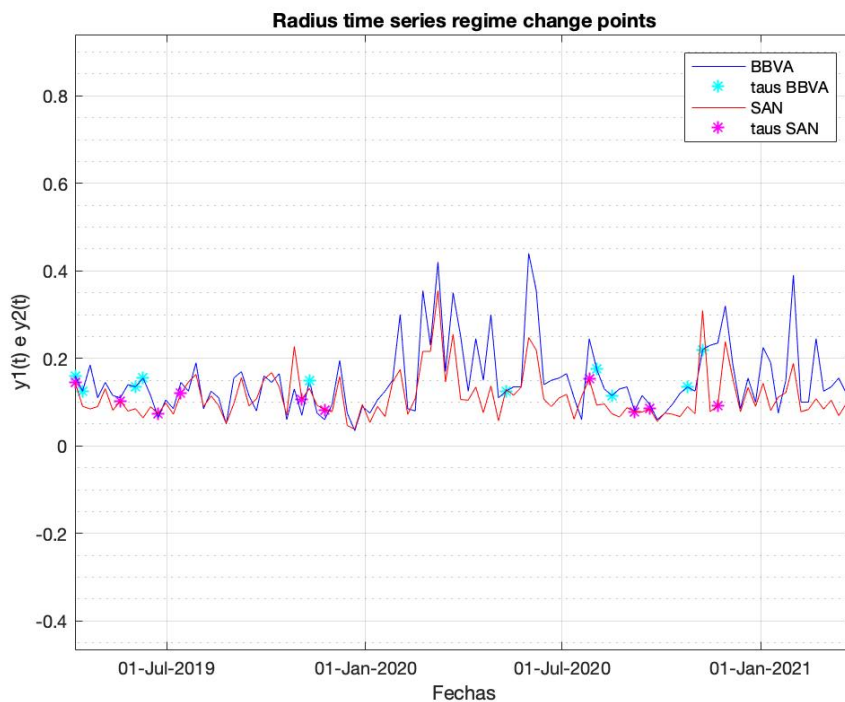


Figura 5.11 Cambios de régimen en las ITS de los radios (Elaboración propia)

Además, se puede ver más claramente las diferentes tendencias detectadas gracias a la función de representación por rectángulos elaborada. En ella se puede apreciar como las pendientes más prolongadas y pronunciadas quedan separadas en los distintos rectángulos.

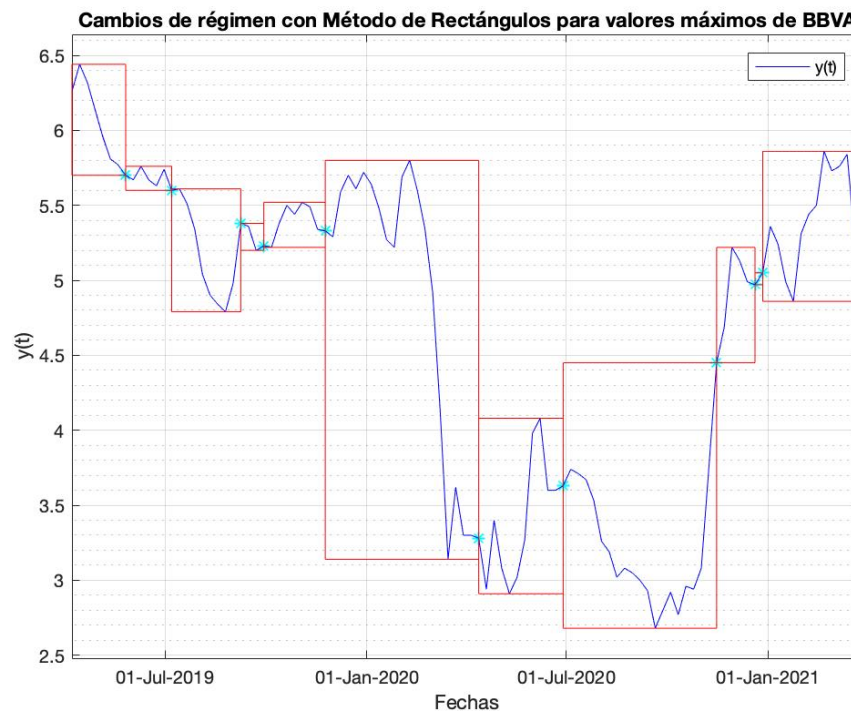


Figura 5.12 Taus en valores máximos de BBVA, por rectángulos (Elaboración propia)

Es preciso, sabiendo los taus y las fechas en las que ocurren, comparar estas fechas en las que se producen los taus de un mismo banco en los máximos, mínimos, centros y radios para encontrar los puntos en los que es más probable que estos cambios de régimen estén ocurriendo realmente. Este proceso lo realiza la propia función, devolviendo aquellas fechas en las que coinciden los taus en los mínimos, máximos, centros o radios.

Para empezar, busca los puntos en los que los taus de máximos y mínimos son iguales y los muestra en una tabla. Los resultados se muestran a continuación:

MODELOS DE CAMBIOS DE RÉGIMEN CON DATOS DE INTERVALO. ENFOQUE DE DATOS SIMBÓLICOS

BBVA	SAN
'20-Dec-2020'	'06-Oct-2019'
'27-Dec-2020'	

Tabla 8 Fechas con taus en máximos y mínimos en BBVA y SAN (Elaboración propia)

Es interesante estudiar estas coincidencias con más profundidad para ver en qué casos estas fechas también coinciden con las obtenidas para los centros, logrando así la denominada **ruptura estructural de nivel**. Sin embargo, la función ha determinado que las tablas son nulas para ambos casos por lo que, con estos datos, no habría ninguna ruptura de nivel para BBVA ni para Santander.

Por último, se pueden obtener las representaciones de los intervalos (del valor mínimo al máximo) de BBVA y de Santander, resaltando aquellos puntos en los que coinciden las fechas de cambio de régimen para los valores mínimos y para los valores máximos.

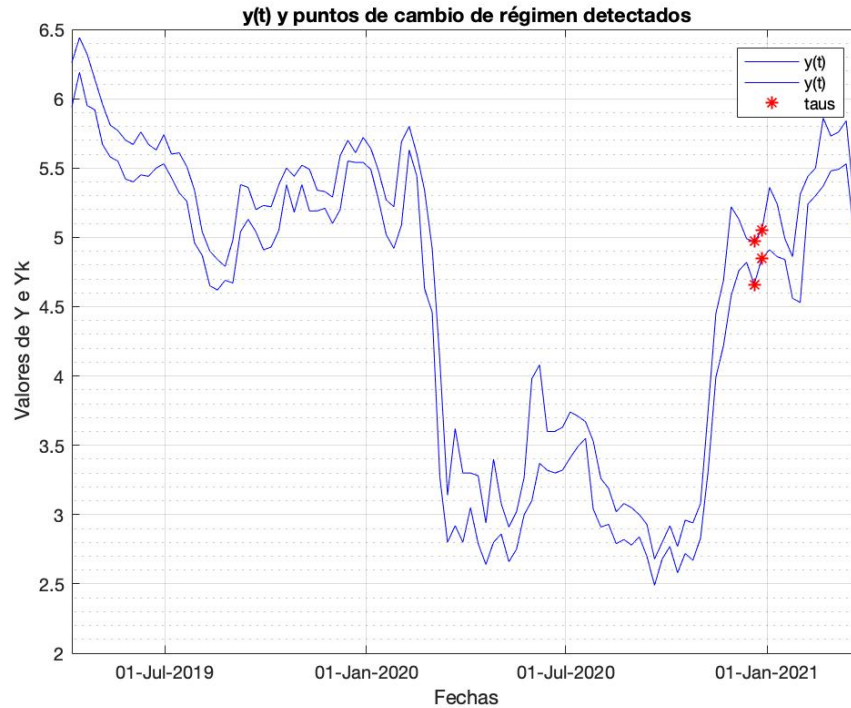


Figura 5.13 Fechas con taus en máximos y mínimos de BBVA (Elaboración propia)

Finalmente, es importante tener en cuenta que los resultados obtenidos dependen de los valores de la ‘t’ (número de taus que se buscan) y de la ‘s’ (separación mínima, en días, entre los taus) seleccionados. Si se hubieran elegido unos valores diferentes para ambas variables, los resultados hubieran sido otros.

Además, hay que destacar que, en el último paso de la función, al calcular los puntos comunes en máximos y mínimos, la función no tiene en cuenta un posible margen de uno o dos días entre puntos, por lo que esto puede restringir el número de puntos finales obtenidos. Este factor será mejorado en aplicaciones posteriores de la función.

Capítulo 6. MODELOS DE CAMBIOS DE RÉGIMEN

CON DATOS DE INTERVALO. ENFOQUE DE

ARITMÉTICA DE INTERVALOS

6.1 INTRODUCCIÓN

Para continuar con los modelos de cambio de régimen con datos de intervalo como los que se han tratado hasta ahora, es interesante realizar un análisis desde el punto de vista de la aritmética de intervalos.

El análisis de intervalos (AI) es una técnica en la que la investigación se ha centrado recientemente, que se utiliza en áreas diferentes áreas, principalmente de la ingeniería como la lógica, la optimización o la robótica. El AI se utiliza como técnica de aproximación numérica tanto para modelos como para problemas matemáticos en los que los datos vienen dados por intervalos y en los cuales se espera que cualquier resultado calculado posea cierto error. El estudio del AI ha provocado un gran desarrollo dentro de los métodos y aplicaciones en el mundo de los datos de intervalo.

La aritmética de intervalos ofrece la teoría básica para poder realizar operaciones conocidas y habituales dentro de cualquier espacio vectorial como, por ejemplo, suma, resta, multiplicación y división.

A lo largo de este capítulo se van a repasar diferentes conceptos y técnicas que tienen que ver con la aritmética de intervalos a partir de los cuales, se llevará a cabo la realización de dos nuevas funciones para detectar cambios de régimen, que se sumarán a la ya estudiada hasta ahora.

6.2 TEST DE BONDAD DE AJUSTE, CR Y ER

A la hora de realizar la comprobación del funcionamiento de un modelo estadístico, se hace uso de la bondad de ajuste para describir lo bien que se ajusta ese modelo a un conjunto de observaciones.

Para realizar esta comprobación, es posible utilizar diferentes parámetros, algunos de ellos orientados a datos clásicos, pudiendo extenderse a datos de intervalos y otros, orientados directamente a datos de intervalos.

A continuación, se van a definir dos de ellos, los cuales van a ser utilizados conjuntamente para poder detectar cambios de régimen.

6.2.1 TASA DE CUBRIMIENTO (CR)

La **Tasa de Cubrimiento** o Coverage Rate (CR) es un coeficiente de intervalo. Como explica (Maté , 2015), la tasa de cubrimiento es la media de los ratios de la intersección entre el valor real de los intervalos con el valor estimado de los mismos, dividido entre la longitud de los intervalos observados o reales.

$$CR = \frac{1}{n} \sum_{i=1}^n \frac{w([\hat{y}_i^l] \cap [y_i^l])}{w([y_i^l])}$$

Ecuación 6.1 Tasa de cubrimiento (CR)

Donde $w([\hat{y}_i^l] \cap [y_i^l])$ es la distancia de la intersección entre el intervalo estimado y el observado y $w([y_i^l])$ es la distancia del valor observado.

6.2.2 TASA DE EFICIENCIA (ER)

La **Tasa de Eficiencia** o Efficiency Rate (ER) representa un coeficiente de intervalo. Según (Maté , 2015), la tasa de eficiencia es la media de todos los ratios de la intersección entre el valor real de los intervalos con los intervalos estimados, dividido entre la longitud de los intervalos estimados.

$$ER = \frac{1}{n} \sum_{i=1}^n \frac{w([\hat{y}_i^l] \cap [y_i^l])}{w([\hat{y}_i^l])}$$

Ecuación 6.2 Tasa de Eficiencia (ER)

Donde $w([\hat{y}_i^l] \cap [y_i^l])$ es la distancia de la intersección entre el intervalo estimado y el observado y $w([\hat{y}_i^l])$ es la distancia del valor estimado.

A continuación, según (Maceda, 2018) se puede apreciar un ejemplo visual, para una mejor comprensión de lo anterior.

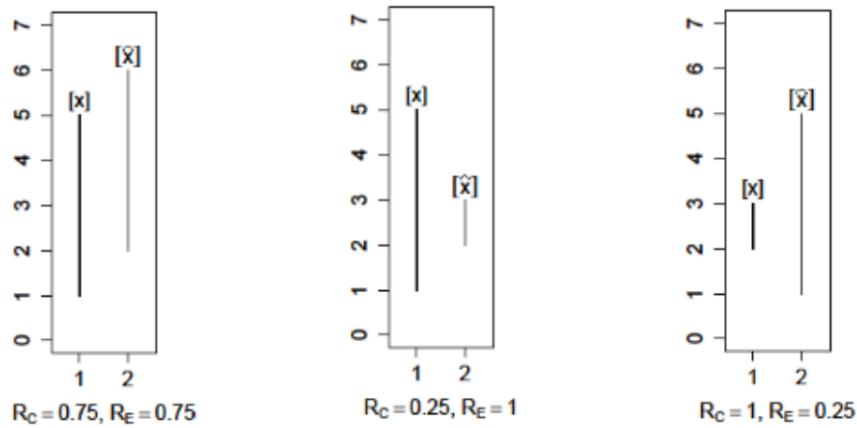


Figura 6.1 CR y ER (Maceda, 2018)

Partiendo de esta información, se ha realizado con Matlab una función capaz de detectar puntos de ruptura estructural a partir del cálculo de los valores de CR y ER en cada periodo de una ITS. Para ello, compara una serie de tiempo valorada en intervalos con esa misma serie de tiempo retardada k posiciones como la predicción para evaluar el CR y ER en cada fecha. Posteriormente, buscará aquellas en las que coincidan ambas por debajo de un cierto valor. Un ejemplo real aplicando dicha función se podrá ver posteriormente en este capítulo.

6.3 CONCEPTO DE VARIACIÓN

A raíz de la experimentación con los coeficientes anteriores, surgió la idea de profundizar en el estudio de detección de cambios de régimen partiendo del análisis de la variación de los valores de una ITS entre un periodo y el anterior.

Una variación representa un cambio o alteración que hace que algo o alguien sea diferente, en algún aspecto, de lo que antes era. Aplicando este concepto a series temporales de datos valoradas en intervalos, puede ser de utilidad calcular el porcentaje de variación de una función en un periodo con respecto al periodo anterior.

De esta forma, es interesante analizar aquellos periodos en los que tanto las cifras de las variaciones correspondientes a los límites superiores del intervalo como las correspondientes a los límites inferiores, superan un valor concreto, que variará según el tipo de datos que se esté evaluando. Por un lado, en el caso de mercados de divisas en los que los grandes cambios se producen a lo largo de un periodo extenso de tiempo, será necesario determinar un límite no muy elevado para poder detectar cuando la variación es significativa de un día a otro. Por otro lado, ocurre lo contrario para el caso de los mercados financieros en los que los valores de los precios de las acciones pueden variar de forma drástica incluso a lo largo de un solo día, por lo que será necesario establecer unos límites más elevados de forma que se detecten únicamente aquellos cambios que realmente son significativos.

Cabe destacar que el porcentaje de variación entre dos periodos consecutivos viene dado por la siguiente ecuación:

$$v(t) = \frac{y(t) - y(t-1)}{y(t-1)} \times 100$$

Ecuación 6.3 Porcentaje de variación entre dos periodos consecutivos

Siguiendo con lo anterior, en el momento en el que las variaciones en máximos y mínimos superen este valor establecido, es posible que la serie se encuentre ante un cambio

de régimen al haber tenido lugar un cambio significativo en los datos que se están analizando entre dos días consecutivos. Es ahí cuando se propondrá un punto de cambio de régimen.

Siguiendo esta idea, se ha desarrollado una función en Matlab que se encarga de recibir una serie de datos valorada en intervalos y detectar todos aquellos puntos que, siguiendo con la explicación previa, sean posibles candidatos de cambio de régimen. Para ello, la función calculará el porcentaje de variación en cada periodo con respecto al anterior, tanto en los valores máximos como en los mínimos y, aquellos puntos en los que coincidan que ambos superen el límite establecido, serán propuestos como puntos de cambio de régimen.

Un ejemplo práctico de lo que se ha explicado será mostrado más adelante para poder facilitar su comprensión a través de datos reales.

6.4 EJEMPLO FUNCIÓN CR Y ER (FUNCIÓN 2)

Como se ha indicado previamente, para analizar este ejemplo se parte de una función realizada en Matlab, capaz de detectar puntos de ruptura estructural. Inicialmente, la función detectaba únicamente los puntos donde tanto ER como CR eran igual a cero. Sin embargo, se observó que de esta manera muchos puntos con posibles rupturas estructurales quedaban fuera del alcance de la función por lo que se realizaron las modificaciones pertinentes para que los puntos detectados fueran aquellos en los que los valores de CR y ER se encontraban por debajo de un límite cercano al cero, pero no necesariamente cero, pudiendo así alcanzar un mayor rango de posibilidades.

Para ello, la función compara una serie de tiempo valorada en intervalos con esa misma serie de tiempo retardada k posiciones como la predicción para evaluar el CR y ER en cada fecha. Posteriormente, busca aquellas en las que ambas tasas estén por debajo del límite seleccionado.

Para realizar este análisis, se van a evaluar los datos de la conversión EURUSD entre el euro y el dólar americano, en el periodo que abarca desde el 2 de julio de 2012 y el 30 de diciembre de 2013.

Para comenzar, la función muestra una gráfica que representa los valores de la serie de tiempo introducida ($y(t)$) junto con la serie retardada k posiciones ($y(t-k)$), que en este caso es igual a 1, para ayudar al usuario a hacerse una idea de la forma que tiene la ITS en cuestión a lo largo del tiempo.

Aquellas fechas en la que no exista una intersección entre ambos intervalos, según las definiciones de CR y ER que se han explicado previamente, o las fechas en las que la intersección sea muy leve, son las que aparecerán más tarde como posibles puntos de ruptura estructural.

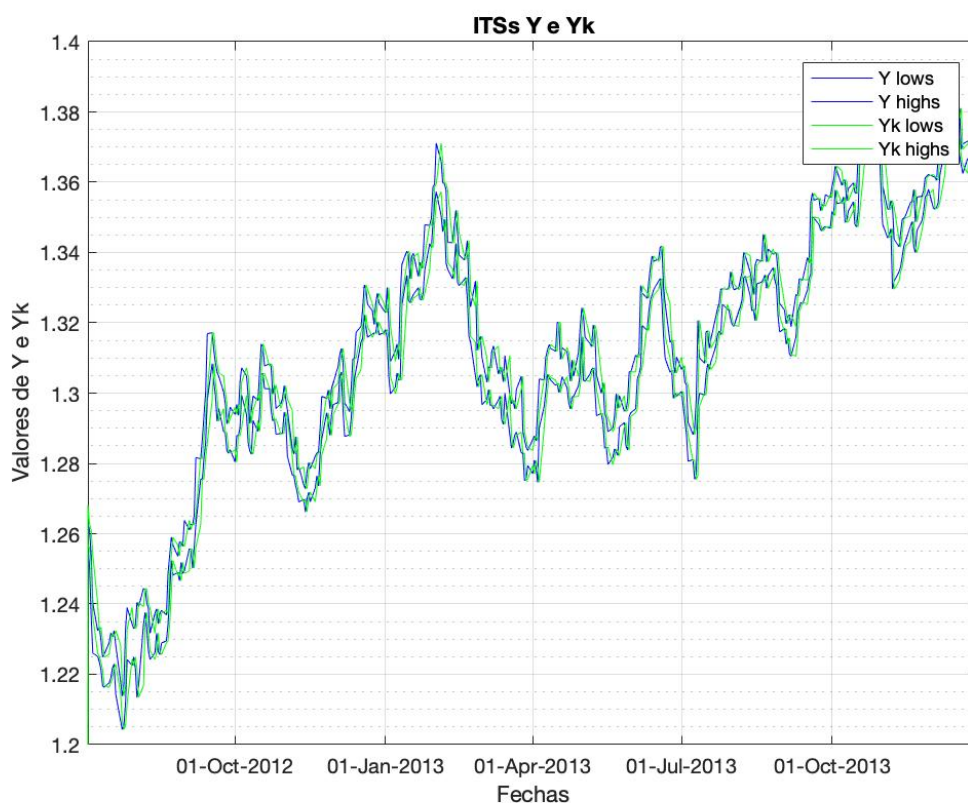


Figura 6.2 Serie de tiempo real y serie de tiempo con retraso (Elaboración propia)

A continuación, se muestra en una gráfica el valor medio de cada uno de los intervalos junto con las tasas medias calculadas de la serie temporal en conjunto para obtener una idea general de la ITS introducida.

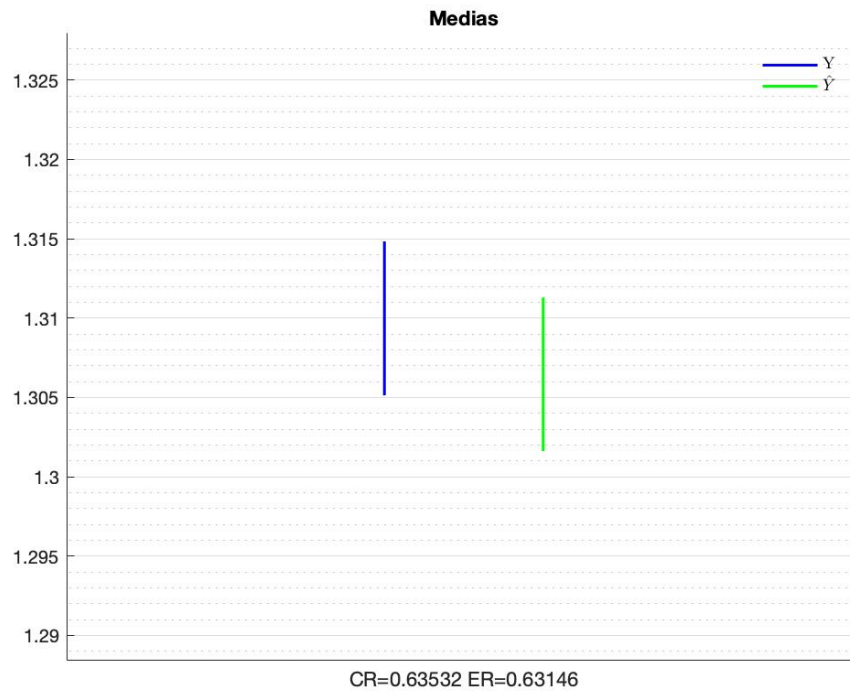


Figura 6.3 Intervalo medio real y estimado y valores de ER y CR (Elaboración propia)

El siguiente paso es obtener con mayor precisión los valores de CR y ER en cada una de las fechas para las cuales existen datos, de forma que posteriormente se puedan analizar los resultados en profundidad y obtener conclusiones. Para ello, la función devuelve la siguiente gráfica con los valores de ambos coeficientes indicando aquellos puntos en los que los valores empiezan a estar por debajo del límite. En este caso el límite se ha establecido en 0,2.

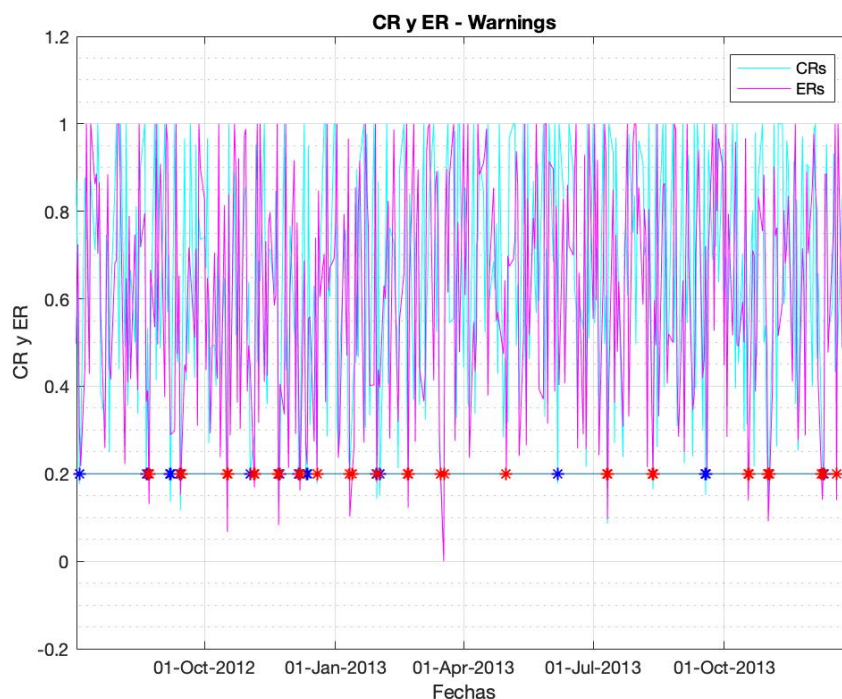


Figura 6.4 CR, ER y warnings (Elaboración propia)

A partir de las gráficas anteriores, la función devuelve una tabla en la que se muestran aquellas fechas en las que se considera que existe una ruptura estructural.

EURUSD	
'14-Sept-2012'	'15-Mar-2013'
'17-Oct-2012'	'18-Mar-2013'
'22-Nov-2012'	'11-Jul-2013'
'06-Dec-2012'	'12-Aug-2013'
'19-Dec-2012'	'31-Oct-2013'
'30-Ene-2013'	'09-Dec-2013'
'21-Feb-2013'	

Tabla 9 Fechas con puntos de ruptura estructural (Elaboración propia)

Los intervalos existentes para estas fechas se muestran en la gráfica siguiente. En todos ellos, es posible comprobar que existe ninguna o muy poca intersección entre el intervalo real y el estimado para fechas indicadas. Es por ello por lo que tanto su valor de CR como su valor de ER son nulos o casi nulos.

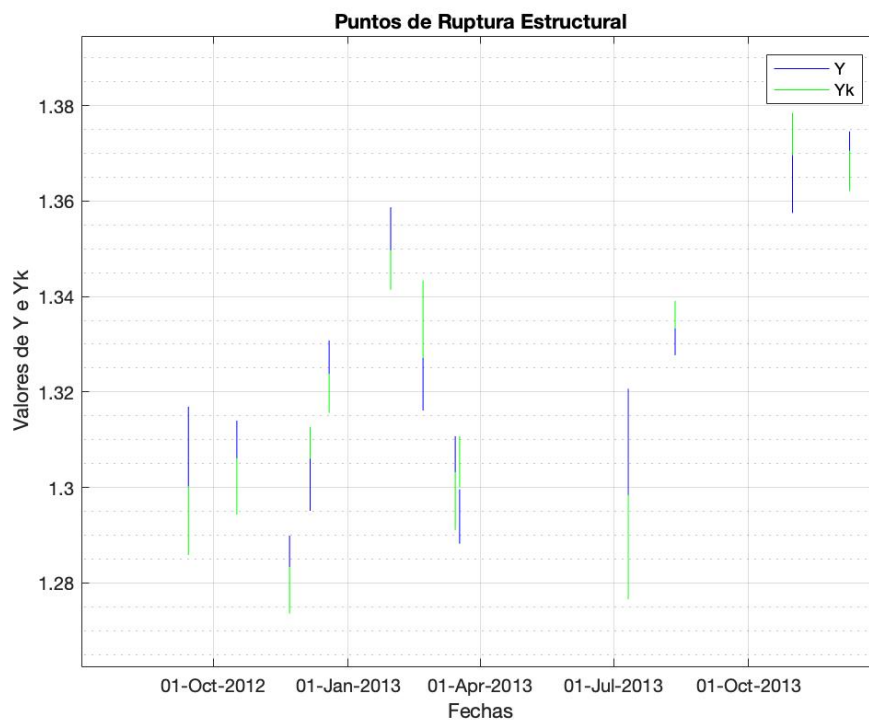


Figura 6.5 Representación de los intervalos en las fechas con ruptura estructural (Elaboración propia)

Por último, la función devuelve una gráfica en la que los puntos obtenidos quedan representados en la serie de tiempo original.

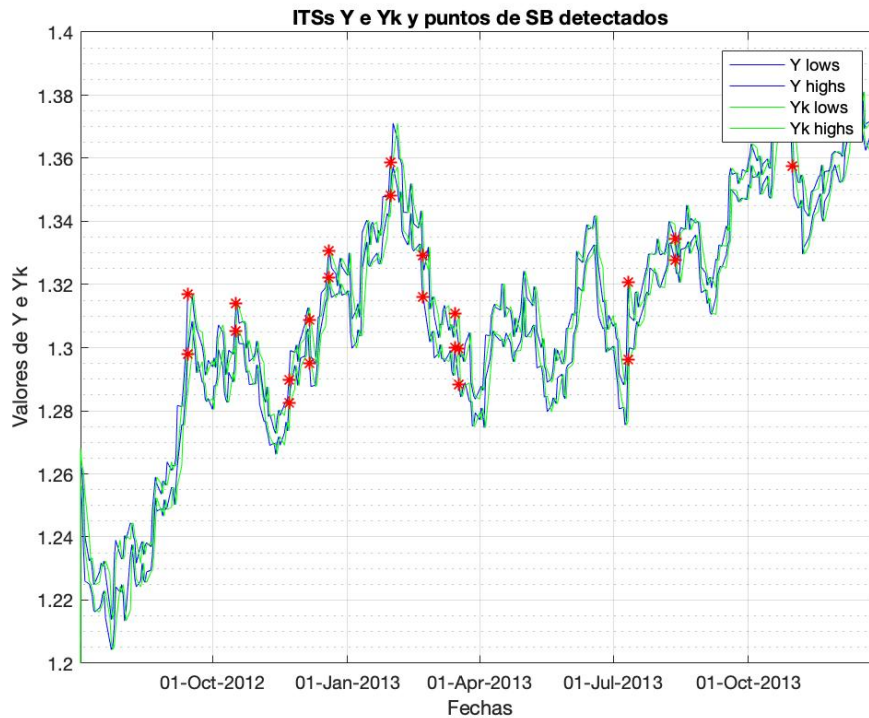


Figura 6.6 Puntos de ruptura estructural detectados (Elaboración propia)

Se observa, por tanto, en esta figura, como la función a simple vista parece fiable ya que muchos de los puntos detectados se encuentran en picos en los que la tendencia de la serie cambia de ascendente a descendente o viceversa.

6.5 EJEMPLO FUNCIÓN VARIACIÓN (FUNCIÓN 3)

Por último, se va a mostrar la última función elaborada para poder detectar cambios de régimen, la cual tiene que ver con el cálculo de los porcentajes de variación de las ITS entre periodos consecutivos, explicada previamente.

Para comenzar, la función requiere conocer el intervalo a estudiar y las fechas en las que quiere estudiarse. Además, debe indicarse un valor k que representa el retraso con respecto al cual se va a comparar la función en el periodo actual. En este caso se ha tomado $k=1$.

Para este ejemplo, se van a tomar los mismos datos que en el ejemplo anterior de CR-ER, es decir, se toman los datos de conversión EURUSD entre el euro y el dólar americano, en el periodo que ocupa desde el 2 de julio de 2012 hasta el 30 de diciembre de 2013.

Inicialmente, la función muestra la representación en el tiempo de la serie temporal real, $y(t)$, y serie con k periodos de retraso, $y(t-k)$, de forma que el usuario pueda hacerse una idea del aspecto de la función en el tiempo.

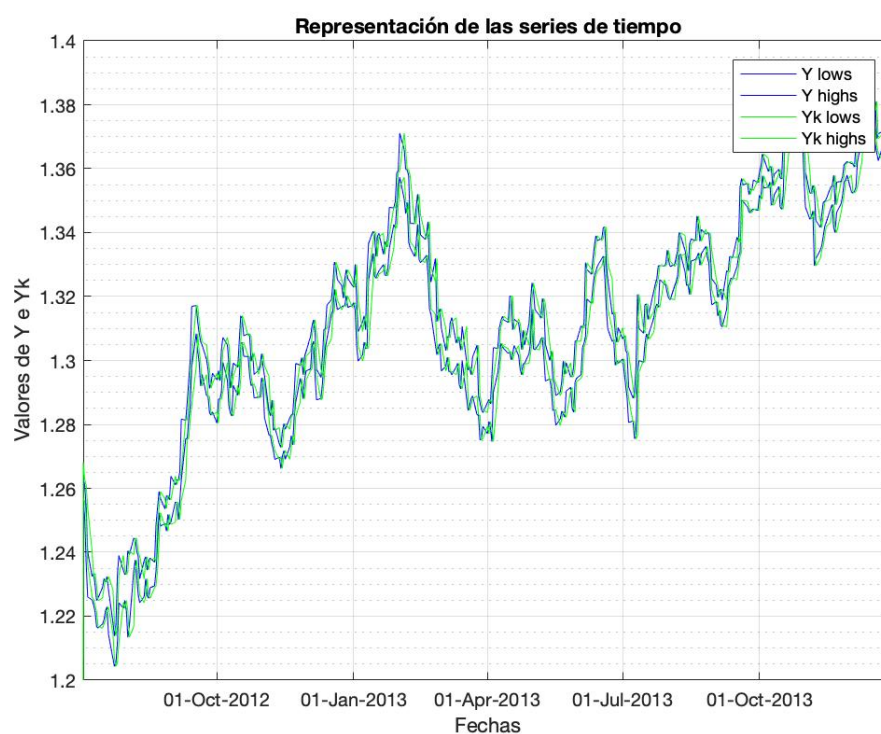


Figura 6.7 Representación temporal de $y(t)$ e $y(t-k)$ (Elaboración propia)

A continuación, la función calcula las variaciones, correspondientes a cada periodo y las muestra gráficamente indicando en color verde aquellas que son crecientes y en color rojo las que son decrecientes.

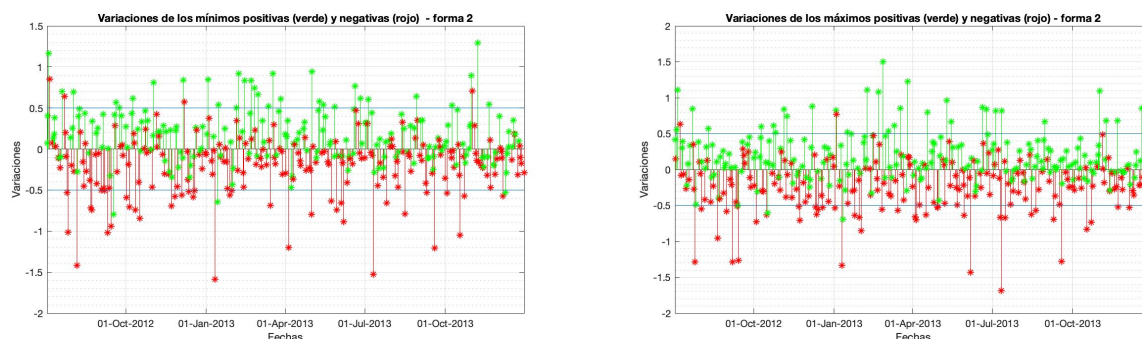


Figura 6.8 Variaciones de máximos y mínimos (Elaboración propia)

Es también interesante analizar la siguiente figura en la que se observan las variaciones conjuntamente. En este caso es el 50% ya que los datos corresponden a cambios de divisas en los que las variaciones grandes se producen a lo largo de periodos extensos. Aquellos puntos en las que las variaciones de máximos y mínimos superen la línea establecida, serán los propuestos para cambios de régimen y están marcados en la gráfica en rojo.



Figura 6.9 Variaciones máximos y mínimos y puntos de ruptura (Elaboración propia)

A continuación, a raíz de la representación de la Figura 6.9, la función devuelve la tabla con las fechas en las que ha detectado puntos de ruptura estructural, los cuales se muestran a continuación.

EURUSD	
'02-Jul-2012'	'11-Jan-2013'
'05-Jul-2012'	'30-Jan-2013'
'06-Jul-2012'	'21-Feb-2013'
'23-Jul-2012'	'25-Feb-2013'
'26-Jul-2012'	'15-Mar-2013'
'02-Aug-2012'	'18-Mar-2013'
'07-Sept-2012'	'05-Apr-2013'
'14-Sept-2012'	'10-May-2013'
'17-Oct-2012'	'15-May-2013'
'22-Nov-2012'	'30-May-2013'
'07-Dec-2012'	'20-Jun-2013'
'12-Dec-2012'	'05-Jul-2013'
'19-Dec-2012'	'11-Jul-2013'
'03-Jan-2013'	'01-Nov-2013'

Tabla 10 Fechas con puntos de ruptura estructural (Elaboración propia)

Finalmente, la función devuelve una gráfica en la que muestra estos puntos detectados sobre las ITS a lo largo del tiempo, de forma que se pueda apreciar visualmente el resultado obtenido.

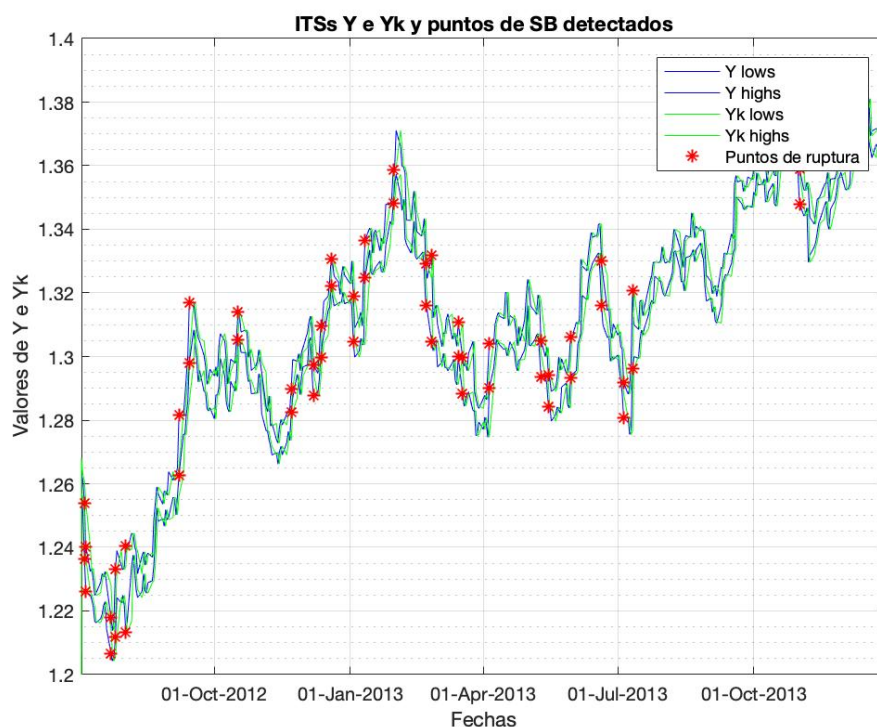


Figura 6.10 Puntos de ruptura estructural detectados (Elaboración propia)

Para concluir, se observa como muchos de los puntos detectados coinciden con picos en la serie temporal, por lo que los resultados serían coherentes ya que ahí se produce un cambio de tendencia.

6.6 COMPARACIÓN DE LOS DOS EJEMPLOS ANTERIORES

Como último paso, resulta de gran interés analizar aquellos puntos en los que coinciden las fechas de cambios de régimen para ambos ejemplos. Se analizarán en conjunto gracias a una tercera función que tiene en cuenta los resultados obtenidos en ambas funciones y almacena y devuelve aquellos que son iguales, dejando un margen de 1 día de diferencia.

Los valores comunes obtenidos se muestran a continuación.

EURUSD

'14-Sept-2012'	'21-Feb-2013'
'17-Oct-2012'	'15-Mar-2013'
'22-Nov-2012'	'18-Mar-2013'
'06-Dec-2012'	'11-Jul-2013'
'19-Dec-2012'	'31-Oct-2013'
'30-Jan-2013'	

Tabla 11 Fechas con puntos comunes de ruptura estructural (Elaboración propia)

Por último, la función muestra la representación siguiente.

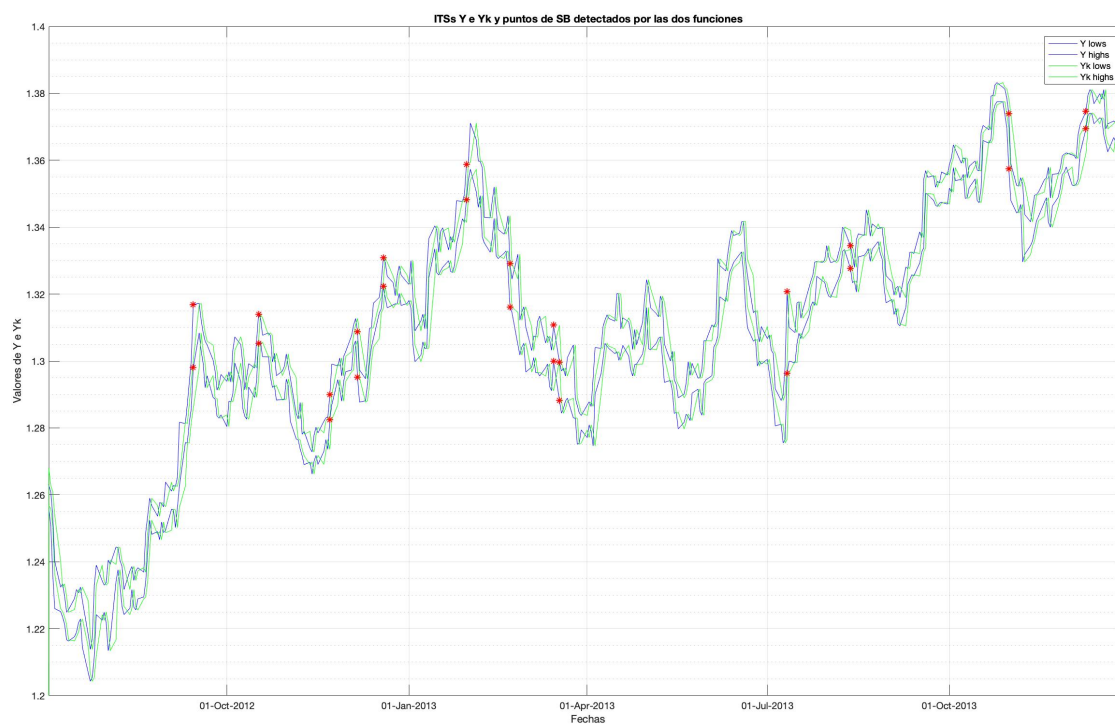


Figura 6.11 Puntos comunes de ruptura estructural detectados (Elaboración propia)

Como conclusión, cabe decir que la gran mayoría de los puntos obtenidos en la función de los coeficientes CR y ER existen también en la función de las variaciones. Siendo 11 de los 13 puntos comunes a la función de las variaciones. De esto último se puede derivar que existe una alta probabilidad de que los puntos obtenidos en la función de los coeficientes como puntos de ruptura estructural sean verdaderos puntos de ruptura al ser también detectados por la función de variabilidad que, además de estos, proporciona en este caso un conjunto de puntos adicionales.

Capítulo 7. APLICACIONES A ACCIONES DEL NASDAQ

7.1 EJEMPLO PRÁCTICO UAL (UNITED AIRLINES)

Para finalizar con el proyecto, se va a evaluar a continuación un ejemplo práctico que pone en relación todas las funciones anteriores y a partir del cual, se obtendrán las conclusiones del trabajo realizado.

Para ello, el primer paso será aplicar la función de regresión del Modelo M, explicada en el Capítulo 4. , a una serie de tiempo en intervalos elegida previamente. Esta función devolverá una estimación para los valores de la ITS en periodos posteriores, los cuales serán enviados a las 3 funciones creadas de detección de cambios de régimen, que son las siguientes:

- Función 1 – Taus, a partir del modelo TAR aplicado a intervalos, definida en el Capítulo 5.
- Función 2 – Coeficientes (CR-ER), que opera a partir del cálculo de los valores de las Tasas de Cubrimiento (CR) y Eficiencia (ER), definida en el Capítulo 6.
- Función 3 – Variaciones, proporciona los puntos de cambios de régimen a partir de las variaciones calculadas de un periodo a otro, expuesta en el Capítulo 6.

Posteriormente, los resultados obtenidos para cada una de estas funciones serán comparados entre sí para detectar similitudes y diferencias.

Además, se realizará el mismo proceso, pero con los valores reales de la misma ITS para el periodo estudiado, es decir, el periodo para el cual se han obtenido antes los valores de la estimación. Una vez aplicadas las 3 funciones previas, se estudiarán en profundidad los resultados alcanzados para poder valorar la efectividad de las funciones programadas.

De acuerdo con lo anterior, se toman los datos de una ITS correspondiente a los precios de acciones de la aerolínea estadounidense United Airlines (UAL), que opera en el mercado de valores NASDAQ. Es interesante analizar estos datos ya que, a partir del mes de marzo de 2020, como consecuencia de la pandemia causada por el Covid-19, los valores de las acciones de las aerolíneas disminuyeron de forma drástica por la imposibilidad de realizar viajes en avión durante un largo periodo de tiempo y en todo el mundo. A continuación, se explican los pasos seguidos para el estudio.

Modelo M.

El primer objetivo que se persigue es que la función de regresión del modelo M-ARD, devuelva la predicción para los precios de las acciones de UAL en el primer semestre de 2020, tomando como referencia datos reales anteriores. Según la explicación que se ha realizado en el Capítulo 4. sobre el funcionamiento de la programación de este modelo en Matlab, se calculan los coeficientes que permiten obtener la predicción a partir de la Ecuación 4.34 y se obtiene la siguiente representación.

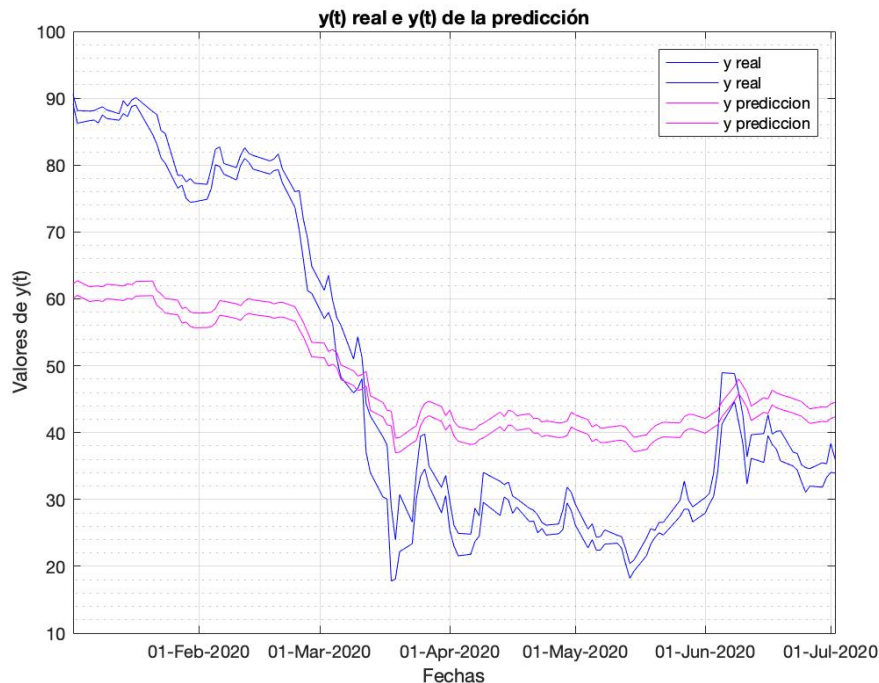


Figura 7.1 Predicción con Modelo M-ARD, UAL (Elaboración propia)

Se puede observar como la predicción sigue en todo momento tendencias similares a la ITS real, aunque la primera lo hace de forma más suavizada.

Una vez obtenido este resultado, serán los valores de la $y(t)$ estimada los que se envíen al resto de funciones para que calculen los puntos de cambios de régimen y poder comparar los resultados.

Función 1 - Taus

Para comenzar, se analiza la primera de las funciones estudiadas en el proyecto. Esta función, a partir de los datos de la predicción devuelve lo siguiente.

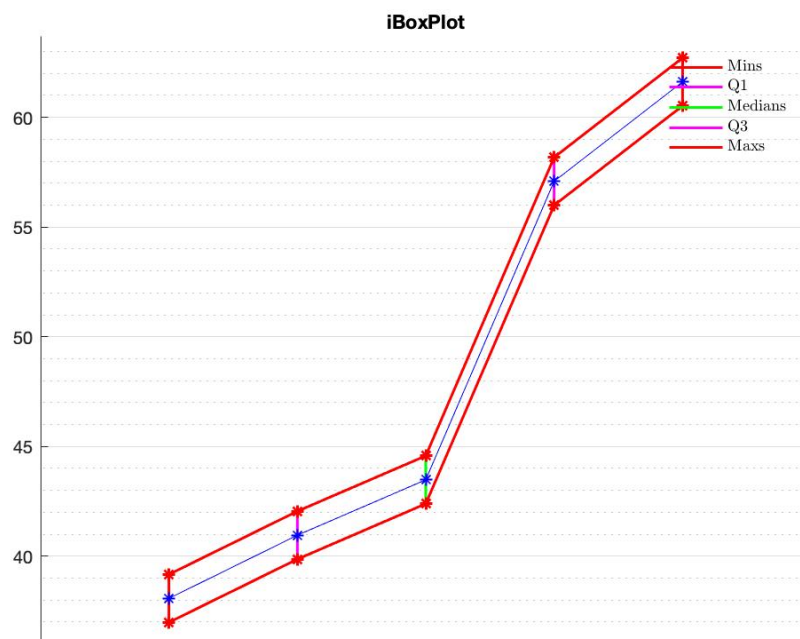


Figura 7.2 iBoxPlot, UAL (Elaboración propia)

De esta representación es posible deducir que seguramente existirán puntos con datos atípicos en los precios de las acciones de UAL en el periodo estudiado, lo cual viene indicado por los cambios abruptos que se producen en las pendientes del iBoxplot.

Además, se obtiene la siguiente gráfica de la ITS estudiada con los cambios de régimen encontrados tanto en los máximos como en los mínimos, los cuales se consideran cambios de régimen de la serie de tiempo.

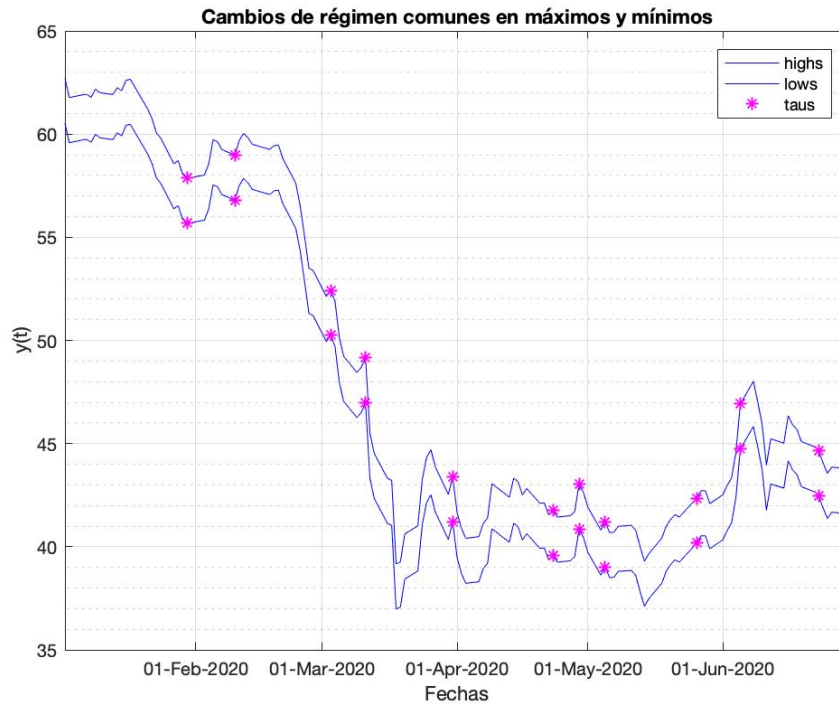


Figura 7.3 Cambios de régimen con Función 1, UAL (Elaboración propia)

UAL – Función 1	
'30-Jan-2020'	'29-Apr-2020'
'10-Feb-2020'	'05-May-2020'
'03-Mar-2020'	'26-May-2020'
'11-Mar-2020'	'05-Jun-2020'
'31-Mar-2020'	'23-Jun-2020'
'23-Apr-2020'	

Tabla 12 Cambios de régimen en Función 1, UAL (Elaboración propia)

Función 2 - Coeficientes

El siguiente paso es calcular los cambios de régimen detectados por la segunda función programada, que hace uso de las Tasas de Eficiencia y Cubrimiento, ER y CR, para obtener los puntos de ruptura estructural según el procedimiento explicado en el Capítulo 6.

De la función se obtienen, a partir de los valores estimados por la función del modelo M-ARD, los siguientes resultados.

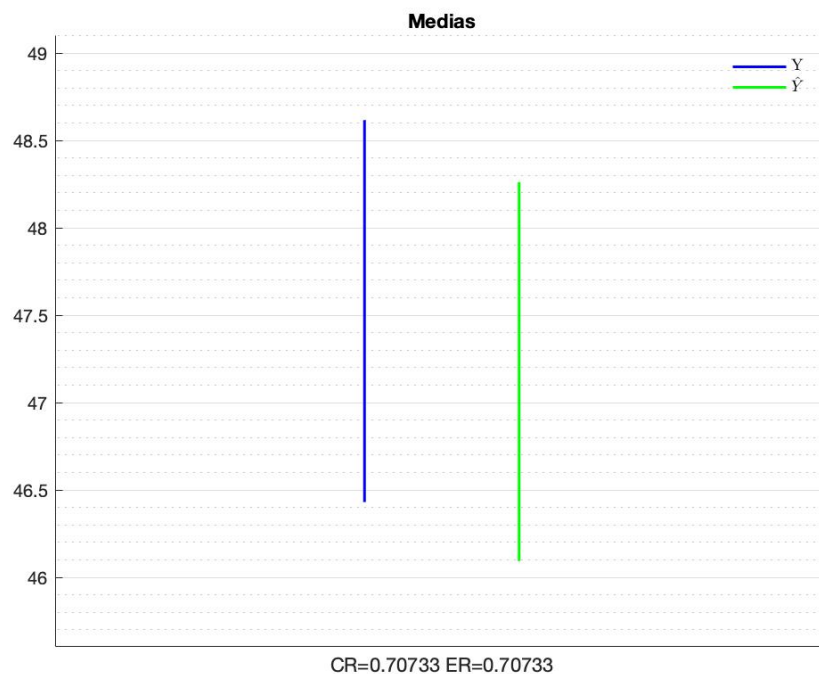


Figura 7.4 Datos resumen Función 2, UAL (Elaboración propia)

Para empezar, se obtiene la gráfica anterior que indica los valores medios de los dos intervalos estudiados y los valores medios de las dos tasas en las que se basa el estudio, donde cabe destacar que tienen resultados exactamente iguales, por lo que la predicción y el valor real serán iguales o muy similares.

A continuación, se representa la serie de tiempo estudiada junto con los puntos detectados de posible ruptura estructural.

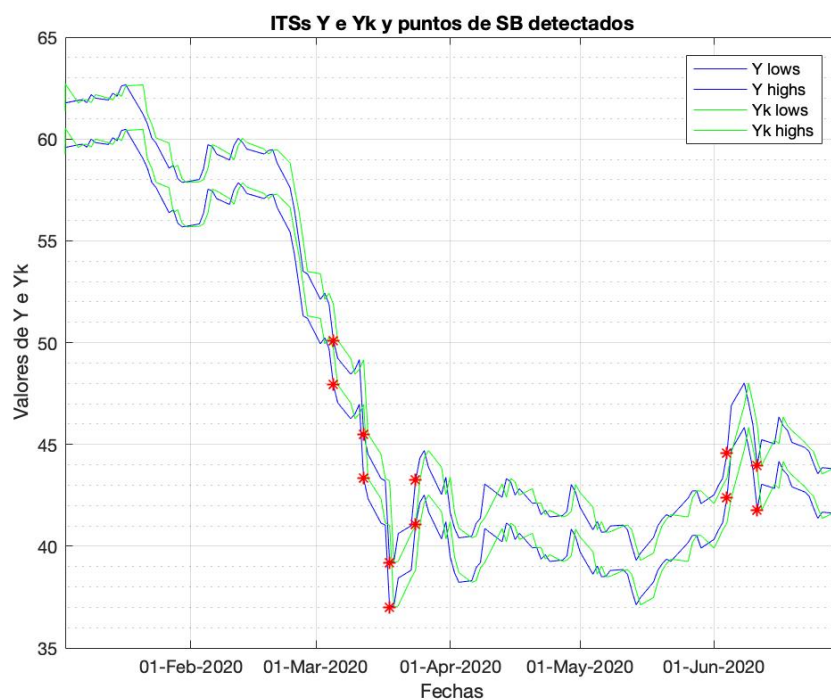


Figura 7.5 Cambios de régimen con Función 2, UAL (Elaboración propia)

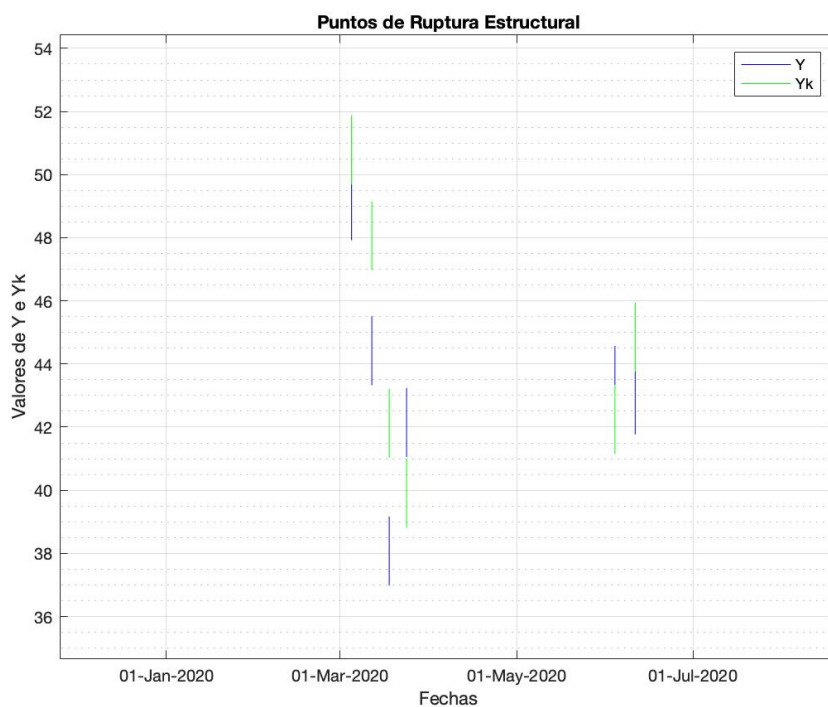


Figura 7.6 Intervalos en puntos de Cambio de Régimen con Función 2, UAL (Elaboración propia)

Es interesante observar como en la Figura 7.6 , los intervalos de las dos series de tiempo prácticamente no interseccionan entre sí o, en caso de hacerlo, lo hacen levemente. Es por ello por lo que los valores de CR y ER serán muy pequeños y, por tanto, se detectan puntos de ruptura estructural en esas fechas, las cuales se muestran a continuación.

UAL – Función 2

'05-Mar-2020'	'24-Mar-2020'
'11-Mar-2020'	'04-Jun-2020'
'18-Mar-2020'	'11-Jun-2020'

Tabla 13 Cambios de régimen en Función 2, UAL (Elaboración propia)

Función 3 - Variaciones

Por último, se utiliza la tercera función de detección de cambios de régimen estudiada, que se desarrolla en el Capítulo 6. y tiene que ver con el cálculo de las variaciones. De esta función se obtienen los resultados que se muestran debajo.

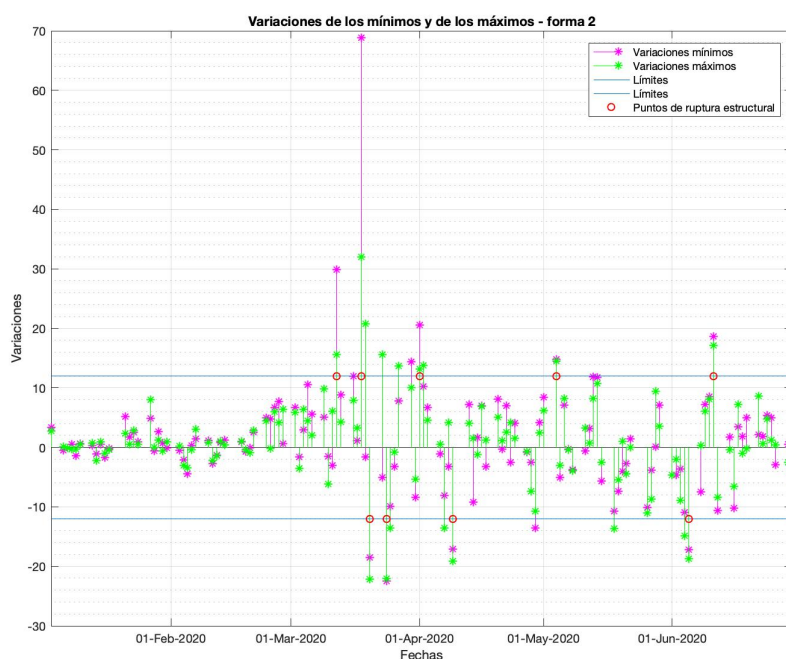


Figura 7.7 Variaciones en máximos y mínimos con Función 3, UAL (Elaboración propia)

Como las variaciones entre periodos consecutivos son muy grandes, se establece un porcentaje muy alto a partir del cual se detectarán los cambios de régimen cuando este valor sea superado por el valor de la variación en los máximos y en los mínimos en un mismo punto. Estos momentos se muestran en la siguiente representación.

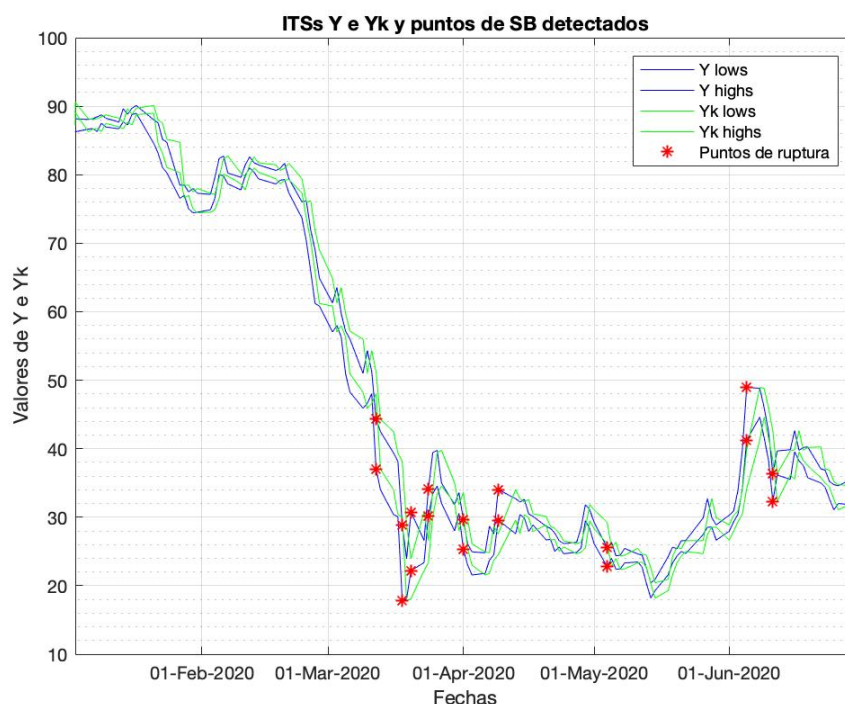


Figura 7.8 Cambios de régimen con Función 3, UAL (Elaboración propia)

UAL – Función 3

'02-Jan-2020'	'01-Apr-2020'
'12-Mar-2020'	'09-Apr-2020'
'18-Mar-2020'	'04-May-2020'
'20-Mar-2020'	'05-Jun-2020'
'24-Mar-2020'	'11-Jun-2020'

Tabla 14 Cambios de régimen en Función 3, UAL (Elaboración propia)

Resultados predicción

Una vez calculados los resultados para las tres funciones, se observa que muchas de las fechas obtenidas en ellas son similares o iguales en al menos dos de ellas.

UAL – Función 1	
'30-Jan-2020'	'29-Apr-2020'
'10-Feb-2020'	'05-May-2020'
'03-Mar-2020'	'26-May-2020'
'11-Mar-2020'	'05-Jun-2020'
'31-Mar-2020'	'23-Jun-2020'
'23-Apr-2020'	
UAL – Función 2	
'05-Mar-2020'	'24-Mar-2020'
'11-Mar-2020'	'04-Jun-2020'
'18-Mar-2020'	'11-Jun-2020'
UAL – Función 3	
'02-Jan-2020'	'01-Apr-2020'
'12-Mar-2020'	'09-Apr-2020'
'18-Mar-2020'	'04-May-2020'
'20-Mar-2020'	'05-Jun-2020'
'24-Mar-2020'	'11-Jun-2020'

Tabla 15 Tabla resumen resultados UAL, datos predicción (Elaboración propia)

Destacan especialmente las fechas destacadas en rojo, correspondientes las al 12 de marzo de 2020, momento en el que se decretó oficialmente un confinamiento prácticamente mundial, donde se visualiza la caída más abrupta en la representación gráfica (Figura 7.3) y el 5 de junio de 2020, momento en el que se detecta un pico de recuperación destacable en las representaciones, cuando los primeros vuelos después del confinamiento comenzaron a tener lugar. Estas fechas coinciden como puntos de cambios de régimen en las tres funciones y, por lo tanto, se puede concluir que son ciertos. Además, hay otros puntos que se marcan en la tabla en los que es probable que un cambio de régimen haya tenido lugar, aunque no sea tan evidente como los anteriores ya que se han detectado en dos de las tres funciones programadas.

Resultados datos reales

A continuación, se ha realizado el mismo proceso que el explicado hasta ahora en este capítulo, pero teniendo en cuenta datos reales para el periodo en cuestión, el cual abarca el primer semestre de 2020. En este caso, a diferencia de en el anterior, no se utilizarán los datos de la predicción calculados por el modelo M-ARD para obtener los cambios de régimen, sino que se hace uso de los datos reales para realizar una comparativa posteriormente.

A partir de los resultados alcanzados, será posible comprobar la efectividad de las funciones programadas en Matlab.

Teniendo esto en cuenta, se obtiene la siguiente tabla.

UAL – Función 1	
'30-Jan-2020'	'27-Mar-2020'
'10-Feb-2020'	'20-Apr-2020'
'02-Mar-2020'	'03-Jun-2020'
'11-Mar-2020'	'19-Jun-2020'

UAL – Función 2

'16-Jan-2020'	'20-Apr-2020'
'21-Jan-2020'	'29-Apr-2020'
'27-Jan-2020'	'04-May-2020'
'05-Feb-2020'	'12-May-2020'
'11-Feb-2020'	'14-May-2020'
'20-Feb-2020'	'18-May-2020'
'24-Feb-2020'	'26-May-2020'
'12-Mar-2020'	'03-Jun-2020'
'18-Mar-2020'	'05-Jun-2020'
'24-Mar-2020'	'10-Jun-2020'
'01-Apr-2020'	'12-Jun-2020'
'09-Apr-2020'	'15-Jun-2020'

UAL – Función 3

'02-Jan-2020'	'01-Apr-2020'
'12-Mar-2020'	'09-Apr-2020'
'18-Mar-2020'	'04-May-2020'
'20-Mar-2020'	'05-Jun-2020'
'24-Mar-2020'	'11-Jun-2020'

Tabla 16 Tabla resumen resultados UAL, datos reales (Elaboración propia)

Para concluir, destaca que los valores obtenidos de forma común a las tres funciones coinciden, siendo estos el 12 de marzo y el 5 de junio de 2020, que además son los que más claramente se pueden detectar a simple vista en la gráfica, por lo que es posible deducir que existe una alta probabilidad de que las funciones calculen correctamente los cambios y también, que en esos días hayan ocurrido cambios de régimen.

Por otro lado, llama la atención que muchos de los puntos que antes aparecían comunes a dos funciones vuelven a serlo ahora (18 de marzo, 5 de mayo, 11 de junio...), por lo que aumenta la probabilidad de que el modelo M programado funcione correctamente y, además, de que las funciones de detección de cambio de régimen devuelvan resultados coherentes.

Capítulo 8. CONCLUSIONES

8.1 CONCLUSIONES

Una vez concluido el proyecto, es momento de extraer y explicar las conclusiones más importantes que se han derivado de él.

Para comenzar, por la parte de regresión, el modelo M fue elegido para poner en práctica en Matlab entre los 8 modelos expuestos en la sección 4.3. Al llevarlo a la práctica en el capítulo anterior, se han obtenido los resultados que se aprecian gráficamente en la Figura 7.1. Según esta representación, se puede observar como la tendencia marcada por la serie de tiempo estimada sigue en todo momento las fluctuaciones de la serie de tiempo real, con la diferencia de que la serie de tiempo estimada tiene unas fluctuaciones en el tiempo más suaves que los valores reales. Sin embargo, bien es cierto que cuando la serie real tiene una pendiente decreciente, la serie estimada también la tiene o que cuando la serie observada tiende a estabilizarse a lo largo del tiempo, la serie estimada hace lo mismo. Por lo tanto, se puede concluir que, a pesar de no ser exactos al 100% los resultados obtenidos de la regresión, lo cual es de esperar porque hasta ahora no existe ningún modelo que pueda predecir por completo valores futuros, el modelo M-ARD desarrollado permite obtener unos resultados de predicción coherentes y con gran similitud a los resultados reales, por lo que será de gran provecho a la hora de estimar valores futuros.

Por otro lado, el otro gran peso del proyecto está puesto en la detección de los cambios de régimen. Tras haber desarrollado numerosos métodos de detección de cambios de régimen en Matlab, se han seleccionado tres métodos principales por considerarse los más interesantes y diferentes entre sí, de forma que se han podido abarcar distintos conceptos teóricos al desarrollarlos. De las tres funciones programadas en Matlab para estos métodos, analizadas en el Capítulo 7. se derivan una serie de conclusiones que se muestran a continuación.

Para empezar, se ha logrado que los cambios de régimen más relevantes, que además suelen ser fácilmente detectables a simple vista, sean señalados por las tres funciones. Este resultado se puede deducir en el capítulo anterior a partir de la detección como puntos de cambio, por parte de las todas las funciones, las fechas de inicio y final del confinamiento debido a la pandemia del año 2020. Todas las funciones coinciden en que los días 11 de marzo y 4 de junio de ese año tuvieron lugar dos cambios de régimen, lo cual es coherente al estar analizando las acciones de una compañía aérea que durante el periodo de confinamiento tuvo que cancelar todos sus vuelos. De este hecho, se deduce que las tres funciones son apropiadas para detectar los cambios de régimen más notables

Además, atendiendo a los resultados visualizados en la Tabla 15 y en la Tabla 16, se advierte que todas las fechas que se pueden apreciar simultáneamente en más de una función aparecen siempre en la función 2, relativa al cálculo de las tasas CR y ER, por lo que se puede decir que los resultados que obtiene esta función tienen alta probabilidad de ser verdaderos puntos de cambio de régimen, siendo ésta, por tanto, la función más efectiva.

Por otro lado, aunque todas las funciones deben ser ajustadas a los datos que se vayan a estudiar, se destaca que la función 1, relativa al cálculo de taus según el modelo TAR, es la que obtendrá unos resultados más dispares dependiendo de los valores indicados por el usuario en los parámetros de entrada. Estos parámetros tienen que ver con el número de taus que busca y la separación, en días, mínima entre ellos. Esto último, sirviendo de gran provecho en muchas ocasiones, en otras hace que los resultados varíen considerablemente de acuerdo con los valores indicados para estos campos y, por tanto, de no estar bien indicados estos parámetros, la función dejará de ser útil ya que podría detectar cambios leves por delante de cambios de mayor importancia.

Por último, es preciso mencionar la función 3, la cual tiene que ver con el cálculo de variaciones entre periodos normalmente consecutivos. Ésta, en caso de ser bien ajustada según la variación que tiendan a sufrir los datos evaluados, recaba unos resultados que presentan gran coincidencia con los de la función 2, por lo que se considera que será una buena candidata para el uso a la hora de detectar cambios de régimen, siendo preferible en

todo caso la segunda función. Es preciso para hacer un uso efectivo de esta función, tener en cuenta la naturaleza de los datos que se están analizando de forma que se pueda establecer el límite de variación correctamente. Por ejemplo, en caso de analizar los datos de cambio diarios de un mercado de divisas, el límite establecido de variación no debe ser muy grande. Si, por otro lado, se están evaluando los precios diarios de unas acciones en un mercado bursátil, el límite de variación que permite detectar los cambios de régimen deberá ser más elevado ya que en este caso, si que puede haber cambios bruscos en los precios de las acciones en días consecutivos. Por lo tanto, de esta función se deduce que será un buen detector de cambios de régimen si el límite de variación para detectar estos cambios se indica coherentemente.

8.2 DESARROLLOS FUTUROS

Finalmente, una vez desarrollado el proyecto completo, los resultados han mostrado que se han obtenido desarrollos y conclusiones de gran interés y con gran cabida para seguir analizando más detalladamente de cara al futuro.

Por un lado, las funciones elaboradas han resultado tener un alto porcentaje de acierto y podrían ser desarrolladas con detenimiento, tiempo y mayor exactitud con el objetivo de poder aumentar su efectividad de aquí en adelante. Para ello, podrían incluirse en el estudio indicadores técnicos como el RSI que aportarían nueva información sobre los datos estudiados. Cabe mencionar que fuera del alcance de este proyecto, se está estudiando el RSI y otros conceptos de cara a la elaboración de un artículo con la incansable ayuda de Carlos Maté, director de este proyecto.

Además, sería interesante en un futuro poder aplicar nuevos datos bursátiles con los que obtener una mayor variedad de resultados y así derivar nuevas conclusiones que ayuden a perfeccionar el trabajo realizado en los últimos meses y que ha resultado en este proyecto que con estas líneas llega a su fin.

Capítulo 9. BIBLIOGRAFÍA

- Arroyo, J., Espinola, R., & Maté, C. (2011). Different approaches to forecast interval time series: a comparison in finance. *Computational Economics*, 37(2), 169-191.
- Atria Innovation. (22 de Octubre de 2019). Recuperado el 2020 de Noviembre, de <https://www.atriainnovation.com/que-son-las-redes-neuronales-y-sus-funciones/>
- Beaundry, P., & Koop, G. (1993). Do recessions permanently change output? *Journal of Monetary Economics*, 31, 149-163.
- Billard, L., & Diday, E. (2000). Regression analysis of interval-valued data. *Data Analysis, Classification and Related Methods: Proceedings of the Seventh Conference of the International Federation of Classifications Societies*, (págs. 369-374).
- Billard, L., & Diday, E. (2002). Symbolic regression analysis. *Data Analysis, Classification, Clustering and Data Analysis, Proceedings of the Eighteenth Conference of the International Federation of Classifications Societies*, (págs. 281-288).
- Bolsamanía. (2020). Recuperado el Octubre de 2020, de <https://www.bolsamania.com/indice/IBEX-35/historico-precios>
- Brito, P. (2007). Modelling and analyzing interval data. *Advances in Data Analysis. Springer Berlin Heidelberg*, 197-208.
- De Carvalho, F., Neto, E., & Tenorio, C. (2004). A new method to fit a linear regression model for interval-valued data. *Advances in Artificial Intelligence*, 295-306.
- Enders, W., Falk, B., & Siklos, P. (2007). A threshold model of real US GDP and the problem of constructing confidence intervals in TAR models. *Studies in Nonlinear Dynamis and Econometrics*.

- Frankie, J. (2012). Markov switching time series models. *Handbook of Statistics*, 30, 99-122.
- Giordani, P. (2015). *Linear regression analysis for interval-valued data based on the Lasso technique*. Sapienza University of Rome.
- Gonzalez-Rivera, G., & Lin, W. (2013). Constrained regression for interval-valued data. *Journal of Business Economic Statistics*, 31(4), 473-490.
- Han, A., Hong, Y., & Wang, S. (2017). *Autoregressive conditional models for interval-valued time series data*. Cornell University, Department of Economics.
- Hansen, B. (2011). Threshold autoregression in economics. *Statistics and its Interface*, 123-127.
- Hao, P., & Guo, J. (2017). Constrained Center and Range Joint Model for Interval-valued symbolic data regression. *Computational Statistics and Data Analysis*, 116-138.
- Inversoro.es*. (2020). Recuperado el Octubre de 2020, de <https://www.inversoro.es/precio-del-oro/grafico-precio-oro-ultimos-5-anos/>
- Järås, J., & Mohammadipour Gishani, A. (2010). *Threshold Detection in Autoregressive Non-linear Models*. Lund University Department of Statistics., Department of Statistics.
- Lima Neto, E., & De Carvalho, F. (2008). Centre and Range method for fitting a linear regression model to symbolic interval data. *Computational Statistics and Data Analysis*, 1500-1515.
- Lima Neto, E., & De Carvalho, F. (2010). Constrained linear regression models for symbolic interval-valued variables. *Computational Statistics and Data Analysis*, 54, 333-347.
- Maceda, P. (2018). Basic analytic system for interval valued data. *Trabajo final de Máster*. Madrid, España.

- Maté , C. (Julio de 2015). Combining interval time series forecasts. An agenda for future research. *1st International Symposium on Interval Data Modelling: Theory and Applications*.
- Maté, C. (Junio de 2012). El análisis de intervalos. Aplicaciones en ingeniería. *Anales de Mecánica y Electricidad, LXXXIX(III)*, 20-27.
- Maté, C. (2021). Combining Interval Time Series Forecasts. A First Step in a Long Way (Research Agenda). *Revista Colombiana de Estadística*, 123-157.
- Maté, C. (2021). The iBoxplot: a visual tool for interval-valued data based on order statistics of lows and highs. Revised version submitted to *Communications in Statistics-Theory and Methods*.
- Rodrigues, P., & Salish, N. (2015). Modeling and forecasting interval time series with threshold models. *Advanced Data Anal. Classification*, 9(1), 41-57.
- Ryan, S., & Porth, L. (2007). *A Tutorial on the Piecewise Regression Approach Applied to Bedload Transport Data*. General Technical Report RMRS-GTR- 189., Department of Agriculture, Forest Service, Rocky Mountain Research Station. , Fort Collins, CO.
- Sinova, B., Colubi, A., & Gonzalez-Rodriguez, G. (2012). Interval arithmetic-based simple linear regression between interval data: Discussion and sensitivity analysis on the choice of the metric. *Information Sciences*, 199, 109-124.
- Souza, L., Souza, R., Amaral , G., & Silva Filho, T. (2017). A Parametrized Approach for Linear Regression of the Interval Data. *Knowledge Based Systems*.
- Sun, Y., Han, A., Hong, Y., & Wang, S. (2018). Threshold autoregressive models for interval-valued time series data. *Journal of Econometrics*, 206(2), 414-446.
- Teles, P., & Brito, P. (2015). Modelling interval time series with space–time processes. *Comm. Statist. Theory Methods*, 44(17), 3599-3627.

Vidal, J. (25 de Agosto de 2016). *DataPrix*. Recuperado el Octubre de 2020, de <https://www.dataprix.com/es/blog-it/juanvidal/descomposicion-series-temporales-sas>

Wikipedia. (Febrero de 2021). Recuperado el Marzo de 2021, de https://es.wikipedia.org/wiki/Regresi3n_segmentada

Yang, W., & Han, A. (2015). A new approach for forecasting the price range with financial interval-valued time series data.

ANEXO I: ALINEACIÓN DEL PROYECTO CON LOS ODS

Los Objetivos de Desarrollo Sostenible (ODS) fueron elaborados en el año 2015 por parte de jefes de Gobierno y de Estado procedentes de diferentes países pertenecientes a las Naciones Unidas, como parte de la Agenda 2030 en la que se proponen una serie de objetivos a cumplir entre todos con ese año como límite.

Estos objetivos se proponen para ampliar los previamente existentes Objetivos de Desarrollo del Milenio (ODM) y, además, buscan alcanzar los que no se habían conseguido hasta entonces.

La idea que promueven es que independientemente de su nivel de riqueza y desarrollo, cada país sea capaz de aportar su granito de arena persiguiendo el cumplimiento de estos objetivos para poder cumplirlos entre todos, de forma que se promueva la prosperidad y se proteja el medioambiente.

De acuerdo con la idea de consecución de los objetivos que se persiguen, todos los ciudadanos deberían colaborar en la medida en la que fuera posible para el desarrollo de estos objetivos y es por esa razón por la cual este trabajo está comprometido con algunos de ellos.

Para comenzar, el objetivo 4, **Educación de Calidad**, lleva a reflexionar sobre que la educación permite una movilidad socioeconómica ascendente y que es clave para el desarrollo económico y para poder salir de situaciones de pobreza. Con el afán de conseguir un ascenso económico y alcanzar altas cifras de dinero en sus cuentas, existe un gran número de personas que, sin tener conocimientos con respecto a este tema, se embarcan a invertir todos sus ahorros en bolsa y terminan arruinados. Con este proyecto se puede colaborar con la educación financiera de las personas haciéndolos conocedores de diversos conceptos estadísticos útiles a la hora de invertir y, además, reduciendo su forma de invertir arriesgando y facilitando que realicen previos estudios sobre lo que puede ocurrir de cara al futuro.

ANEXO I: ALINEACIÓN DEL PROYECTO CON LOS ODS

Por otro lado, el objetivo 8, **Trabajo Decente y Crecimiento Económico**, está directamente alineado con el fundamento de este trabajo. Está demostrado que más del 90% de las personas que deciden invertir su dinero en bolsa acaban perdiéndolo. En este trabajo, se ha desarrollado un sistema original y sencillo capaz de ayudar a los inversores a la hora de realizar sus inversiones, de forma que puedan evitar futuras pérdidas pudiendo observar cuál va a ser el comportamiento de las acciones de cara al futuro. De esta forma, se fomenta un uso cauteloso del dinero, lo que puede incitar a personas con recursos medios a participar en el mercado bursátil y obtener rendimientos, pudiendo impulsar así el crecimiento económico.

Por otro lado, el objetivo 10, **Reducción de las Desigualdades**, hace referencia a las desigualdades tanto dentro de un país como fuera de él, es decir, entre los distintos países. Debido a la pandemia ocasionada por el Covid-19, el umbral de la pobreza ha aumentado en el último año y, por lo tanto, se han intensificado las desigualdades existentes. Además, ha aumentado significativamente el desempleo mundial y los ingresos de muchas familias han disminuido. Como aportación a este objetivo, el trabajo desarrollado en estas páginas puede ayudar a algunas personas que han reducido sus ingresos a actuar con más seguridad en el mundo de la bolsa, minimizando el riesgo de perder su dinero y ayudando a mantener un crecimiento de los ingresos que permita reducir las desigualdades que, en el último año, más que nunca, habían aumentado.

Finalmente, cabe mencionar que cualquier utilización de los métodos desarrollados en este proyecto de cara a minimizar riesgos y ayudar tanto a pequeños como grandes inversores a invertir su dinero con más fiabilidad, podrá ser de gran provecho para reducir grandes desigualdades, fomentar el crecimiento económico y colaborar en el ámbito del aprendizaje y educación de las personas dispuestas a estudiar los conceptos expuestos y continuar desarrollándolos hacia el futuro.