



**COMILLAS**

UNIVERSIDAD PONTIFICIA

ICAI

# GRADO EN INGENIERÍA EN TECNOLOGÍAS DE TELECOMUNICACIÓN

TRABAJO FIN DE GRADO

## **LA CAUSALIDAD DE GRANGER EN EL ANÁLISIS Y LA PREVISIÓN DE SERIES TEMPORALES CLÁSICAS, DE INTERVALO Y DE HISTOGRAMA. APLICACIÓN EN MERCADOS FINANCIEROS.**

Autor: Cristina Acín Coello de Portugal

Director: Carlos Maté Jiménez

Madrid

Julio 2021



Declaro, bajo mi responsabilidad, que el Proyecto presentado con el título  
La causalidad de Granger en el análisis y la previsión de series temporales clásicas, de  
intervalo y de histograma. Aplicación en los mercados financieros.  
en la ETS de Ingeniería - ICAI de la Universidad Pontificia Comillas en el  
curso académico 2021/22 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.

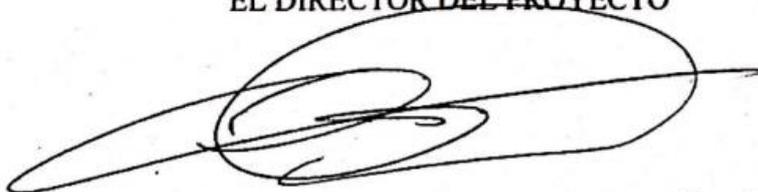


Fdo.: Cristina Acín Coello de Portugal

Fecha: 06 / 07 / 2022

Autorizada la entrega del proyecto

EL DIRECTOR DEL PROYECTO



Fdo.: Carlos Maté Jiménez

Fecha: 06/07/2022





**COMILLAS**

UNIVERSIDAD PONTIFICIA

ICAI

GRADO EN INGENIERÍA EN TECNOLOGÍAS DE  
TELECOMUNICACIÓN

TRABAJO FIN DE GRADO

**LA CAUSALIDAD DE GRANGER EN EL ANÁLISIS Y  
LA PREVISIÓN DE SERIES TEMPORALES  
CLÁSICAS, DE INTERVALO Y DE HISTOGRAMA.  
APLICACIÓN EN MERCADOS FINANCIEROS.**

Autor: Cristina Acín Coello de Portugal

Director: Carlos Maté Jiménez

Madrid

Julio 2021



# Agradecimientos

En primer lugar, me gustaría agradecer a mi familia por su apoyo constante, confianza y cariño. Además, quisiera también dar las gracias a mi tutor, Carlos, por su incansable ayuda durante la realización del proyecto, por todas sus recomendaciones de artículos y trabajos, y por el esfuerzo que ha realizado por mantener nuestras reuniones.

Por último, no quería dejar de mencionar a todos los profesores y compañeros que he tenido a lo largo de estos cinco años de universidad, ya que han contribuido especialmente a mi formación académica y sin ellos nada de esto hubiera sido posible.



# **LA CAUSALIDAD DE GRANGER EN EL ANÁLISIS Y LA PREVISIÓN DE SERIES TEMPORALES CLÁSICAS, DE INTERVALO Y DE HISTOGRAMA. APLICACIÓN EN MERCADOS FINANCIEROS.**

**Autor: Acín Coello de Portugal, Cristina.**

Director: Maté Jiménez, Carlos.

Entidad Colaboradora: ICAI – Universidad Pontificia Comillas.

## **RESUMEN DEL PROYECTO**

La causalidad de Granger es un test estadístico que comprueba si los resultados de una variable sirven para predecir la otra variable, y si tiene resultado unidireccional o bidireccional. Para ello, se tiene que comparar y deducir si el comportamiento actual y pasado de una serie temporal A predice la conducta de otra serie temporal B. Si esto ocurre, se puede afirmar que A causa B, y el comportamiento es unidireccional. Si de la misma manera B ayuda en la predicción de A, la causalidad es bidireccional, afirmando que A causa B y B causa A.

La aplicación de la causalidad de Granger a series temporales clásicas ha sido estudiada con profundidad en la literatura, pudiéndose encontrar numerosos estudios de autores prestigiosos sobre ello, algunos de ellos se incorporan en el proyecto. Sin embargo, la causalidad de Granger en series temporales de intervalos y de histogramas todavía no ha sido explorada de forma detallada. Es por esta razón que el principal objetivo a perseguir en este proyecto es la programación en MATLAB de sistemas de análisis causal para la aplicación de la causalidad de Granger a series temporales de intervalos y de histogramas. En este trabajo se incluirán aplicaciones a distintos activos financieros en diferentes periodos para poder analizar su comportamiento y obtener las conclusiones pertinentes.

**Palabras clave:** series temporales de intervalo, series temporales de histograma, regresión, MATLAB, causalidad de Granger, oro, petróleo.

## **1. Introducción**

La causalidad de Granger es un concepto estadístico relevante ya que ayuda a mejorar la predicción de las series temporales. Con el objetivo de determinar si se cumple o no la causalidad de Granger se realiza un contraste de hipótesis, muy común en el campo de la estadística, y se obtiene un p-valor. Un p-valor considerablemente pequeño nos hará rechazar la hipótesis nula (la cual parte de la base que no se cumple la causalidad de Granger) y por tanto podremos afirmar que existe una relación de causalidad.

Después, pasaremos a comprobar si esta relación es en ambos sentidos (bidireccional) o simplemente en uno de ellos (unidireccional). Además, miraremos la fuerza de la causalidad según el grado de confianza al que podemos rechazar la hipótesis nula.

## **2. Definición del Proyecto**

El análisis de intervalos tiene sus orígenes en el mundo clásico, pero no fue hasta la década de 1960 cuando se empezó a investigar sobre ello, dando importancia a los métodos y aplicaciones con datos de intervalo. Desde entonces, los datos de intervalo han adquirido mucha relevancia y hoy en día su estudio es de gran interés, produciéndose de esta manera muchos avances en la década actual. Este tipo de datos son mucho más precisos que los datos clásicos y aportan más información que la que se puede obtener a partir de un único valor.

Los datos de intervalo reflejan la variabilidad de una observación en un determinado instante mediante un rango de valores, sin embargo, no informan de lo que sucede entre los extremos considerados, es decir, no indican cómo se distribuyen las observaciones dentro de dicho rango. Esta carencia es cubierta por los histogramas, que permiten representar de una manera más completa la distribución de datos cuantitativos. El tratamiento con histogramas se vuelve aún más complejo, pero proporciona una información más completa y precisa que la que puede contener un intervalo.

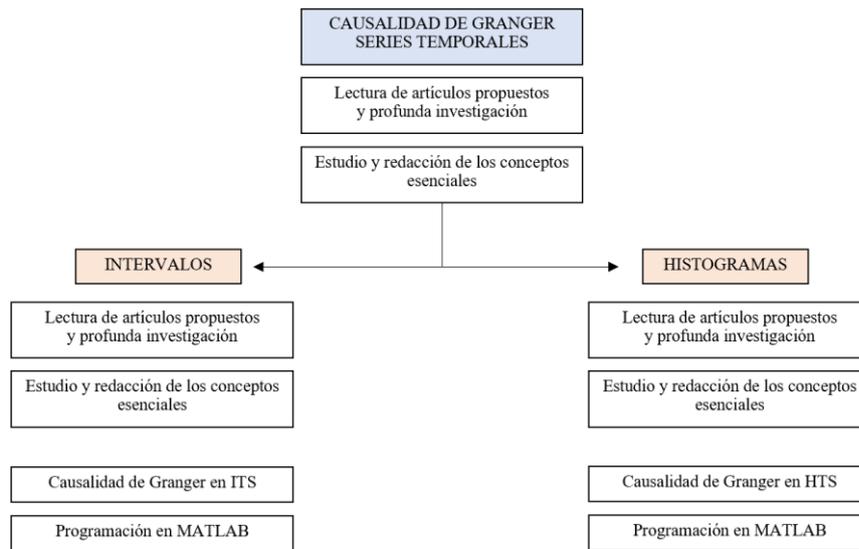
A lo largo de este proyecto, se realizarán los tests de causalidad de Granger aplicándolos a los datos de intervalo y de histograma. Para poder realizar este estudio, se evaluarán ciertos hechos ocurridos en el pasado con datos de diferentes frecuencia y se analizará la causalidad existente entre distintos activos a lo largo del tiempo.

## **3. Descripción del modelo/sistema/herramienta**

La única herramienta que se ha utilizado a lo largo del proyecto ha sido MATLAB. Este lenguaje de programación se ha empleado con diferentes fines.

En primer lugar, para el análisis y representación de las series temporales de intervalo y de histograma. Además, también se han programado funciones para la predicción de los valores en función de diferentes métodos de regresión. Por último, se han utilizado funciones propias de MATLAB para realizar los tests de causalidad de Granger y de esta manera obtener los resultados.

El proceso de estudio de cada una de las dos partes en las que se podría dividir el TFG (intervalos e histogramas) ha sido muy similar. Se muestra una imagen en la que se ha realizado un esquema del proceso seguido:



*Ilustración 1. Esquema del proceso seguido a lo largo del proyecto*

Tal y como se puede observar en la Ilustración 1, el proyecto parte del estudio del concepto de la causalidad de Granger y las series temporales en general. Para ello se han leído numerosos artículos, revistas, libros y páginas web, referenciadas a lo largo del proyecto.

Una vez se han comprendido dichos conceptos, se ha procedido a estudiar los intervalos y las series temporales de intervalo (ITS). Después se aplica la causalidad de Granger con ejemplos de la vida real. Se ha realizado de una igual manera con los histogramas y las series temporales de histograma (HTS). Todo ello se ha realizado con funciones programadas en MATLAB.

## 4. Resultados

Tras una profunda evaluación y después de haber programado y realizado numerosos desarrollos en MATLAB se han obtenido los siguientes resultados.

### 4.1 Series Temporales de Intervalo (ITS)

En relación con las series temporales de intervalo, se han realizado dos estudios paralelos.

En el primero de ellos se ha estudiado la causalidad de Granger en las series temporales de intervalo de precios del oro y el petróleo desde el año 2010 hasta el 2021. Los resultados obtenidos en este estudio indican que no existe causalidad entre estas dos series temporales de intervalo, o lo que es lo mismo, ninguna de ellas ayuda a predecir la otra. Estos resultados reflejan cierta robustez ya que se han tenido en cuenta cinco métodos de regresión lineal diferentes y todos ellos presentan los mismos resultados de no causalidad.

En el segundo, se ha analizado la causalidad de Granger en las series temporales de intervalo de precios del oro y el petróleo en los momentos de antes, durante y después de la pandemia.

Los resultados obtenidos indican que en el año 2019 existía una causalidad débil entre el oro y el petróleo, al igual que en el año 2022. Sin embargo en el año 2020 la causalidad entre las series temporales de intervalo de precios del oro y el petróleo fue mucho más fuerte. Hay que destacar que los resultados de causalidad obtenidos han sido en todo momento unidireccionales, es decir, únicamente del petróleo hacia el oro, y no en sentido contrario.

Además, con el objetivo de determinar el método de regresión lineal de series temporales de intervalo más preciso se han obtenido algunas de las métricas más características. Se han obtenido los siguientes resultados, recogidos en la Ilustración 2. El estudio incluye cinco métodos de regresión de intervalos (CM, mM, CRM, CCRM y MG), que se explican más en profundidad a lo largo del desarrollo de trabajo. Las métricas iARV y CER miden la precisión del desarrollo de estos métodos. Un valor menor del iARV indica mayor precisión, al igual que un valor más alto en el CER.

MÉTRICAS	CM	mM	CRM	CCRM	MG
<b>iARV</b>	0.0603	0.0603	0.0517	0.0519	<b>0.0497</b>
<b>CER</b>	0.6193	0.6073	0.6507	0.6440	<b>0.6637</b>

*Ilustración 2. Resultados métricas iARV y CER para cada uno de los métodos de regresión lineal de intervalos*

#### **4.2 Series Temporales de Histograma (HTS)**

Un análisis similar se ha realizado para estudiar la causalidad de Granger en las series temporales de histograma. Los resultados en este caso indican que, al realizar un análisis en un periodo de tiempo amplio (más de 10 años), prácticamente no apreciamos causalidad en ninguno de los sentidos.

Sin embargo, si reducimos este periodo de tiempo a los años de la pandemia, aproximadamente los dos años más recientes, encontramos resultados de causalidad de Granger bidireccional entre las dos series temporales del oro y del petróleo.

### **5. Conclusiones**

Después de haber obtenido los resultados mostrados anteriormente se llegan a las siguientes conclusiones.

## **5.1 Series Temporales de Intervalo (ITS)**

La conclusión general en el análisis de causalidad de Granger en las ITS es que la causalidad de la serie temporal del petróleo hacia el oro (unidireccional) es mucho más fuerte en aquellos momentos en los que la crisis de la pandemia se ha visto más pronunciada, debilitándose en la recuperación económica.

Se concluye que no se detecta causalidad si analizamos dos series temporales en un extenso periodo de tiempo que abarca varios ciclos de la economía (auge y crisis), sino que la causalidad se hace presente en cuanto nos centramos únicamente en un periodo de crisis, como puede ser la crisis económica mundial provocada por la pandemia del Covid-19.

Por otra parte, se concluye que el método que mejores métricas presenta, como era de esperar al ser el más reciente y complejo, es el modelo MG, presentando el valor CER más alto, y el valor  $iARV$  más bajo. Es por ello que podemos afirmar que el modelo MG es el más preciso y veraz de entre los estudiados. Aun así, los análisis y resultados anteriores se han obtenido con la combinación de todos los métodos, con el objetivo presentar una mayor robustez en los resultados.

## **5.2 Series Temporales de Histograma (HTS)**

Los resultados obtenidos en el estudio de la causalidad de Granger en HTS no son sorprendentes ya que, al igual que habíamos observado en el estudio con ITS, la causalidad es mucho más fuerte en aquellos momentos en los que se haya experimentado una crisis. Es decir, en este análisis no encontramos causalidad cuando estudiamos un periodo extenso en el tiempo, en el que ha habido momentos de auge y momentos de crisis, sin embargo la causalidad es mucho más fuerte en cuanto acortamos el estudio a un periodo de tiempo menor, en el que nos encontrábamos sometidos a una grave crisis económica mundial por causa de la pandemia.

Además, podemos observar que, a diferencia del análisis con series temporales de intervalo, la causalidad en este caso es bidireccional. Con este estudio se puede afirmar que en este periodo de crisis, el petróleo causa al oro tal y como habíamos observado con las ITS, pero además se afirma que el oro también causa al petróleo.

Estos resultados tienen sentido dado que las series temporales de histograma ofrecen mayor información que las de intervalo al reflejar la distribución de los datos. Es por ello que, al dar una información más completa del fenómeno en su conjunto, es capaz de apreciar la causalidad en ambos sentidos.

Por tanto, como reflexión general, las conclusiones obtenidas en el estudio de la causalidad de Granger en HTS reafirman y complementan a las ya obtenidas en el análisis de la causalidad de Granger con las ITS.

## 6. Referencias

Billard, L., & Diday, E. (2000). Regression analysis for interval-valued data. In *Data analysis, classification, and related methods* (pp. 369-374). Springer, Berlin, Heidelberg.

Blanco-Fernández, A., Corral, N., & González-Rodríguez, G. (2011). Estimation of a flexible simple linear model for interval data based on set arithmetic. *Computational Statistics & Data Analysis*, 55(9), 2568-2578.

García-Bárzana, M., Ramos-Guajardo, A. B., Colubi, A., & Kontoghiorghes, E. J. (2020). Multiple linear regression models for random intervals: a set arithmetic approach. *Computational Statistics*, 35(2), 755-773.

Maté, C. G. (2021). Combining Interval Time Series Forecasts. A First Step in a Long Way (Research Agenda). *Revista Colombiana de Estadística*, 44(1), 123-157.

Troster, V. (2018). Testing for Granger-causality in quantiles. *Econometric Reviews*, 37(8), 850-866.

# **GRANGER CAUSALITY WHEN ANALYZING AND FORECASTING CRISP, INTERVAL, AND HISTOGRAM TIME SERIES. APPLICATIONS TO FINANCIAL MARKETS**

**Author: Acín Coello de Portugal, Cristina.**

Supervisor: Maté Jiménez, Carlos.

Collaborating Entity: ICAI – Universidad Pontificia Comillas.

## **ABSTRACT**

Granger causality is a statistical test that checks whether the results of one variable serve to predict the other variable, and whether it has a unidirectional or bidirectional result. To do this, it is necessary to compare and deduce whether the current and past behavior of one time series A predicts the behavior of another time series B. If this is the case, it can be stated that A causes B, and the behavior is unidirectional. If in the same way B assists in the prediction of A, the causality is bidirectional, claiming that A causes B and B causes A.

The application of Granger causality to classical time series has been studied in depth in the literature, and numerous studies by prestigious authors can be found, some of which are incorporated in the project. However, Granger causality in interval and histogram time series has not yet been explored in detail. It is for this reason that the main objective to be pursued in this project is the programming in MATLAB of causal analysis systems for the application of Granger causality to interval and histogram time series. This work will include applications to different financial assets in different periods in order to analyze their behavior and obtain the appropriate conclusions.

**Keywords:** interval time series, histogram time series, regression, MATLAB, Granger causality, gold, brent.

## **1. Introduction**

Granger causality is a relevant statistical concept as it helps to improve the prediction of time series. In order to determine whether Granger causality is satisfied or not, a hypothesis test, very common in the field of statistics, is performed and a p-value is obtained. A considerably small p-value will make us reject the null hypothesis (which assumes that the Granger causality is not satisfied) and therefore we can affirm that there is a causal relationship.

We will then go on to check whether this relationship is in both directions (bidirectional) or just in one of them (unidirectional). In addition, we will look at the strength of causality according to the degree of confidence at which we can reject the null hypothesis.

## **2. Project definition**

Interval analysis has its origins in the classical world, but it was not until the 1960s when research on it began, giving importance to methods and applications with interval data. Since then, interval data have gained much relevance and nowadays their study is of great interest,

thus producing many advances in the current decade. This type of data is much more accurate than classical data and provides more information than can be obtained from a single value.

Interval data reflect the variability of an observation at a given instant through a range of values; however, they do not report what happens between the extremes considered, i.e., they do not indicate how the observations are distributed within that range. This shortcoming is covered by histograms, which allow a more complete representation of the distribution of quantitative data. The treatment with histograms becomes even more complex, but provides more complete and accurate information than a range can contain.

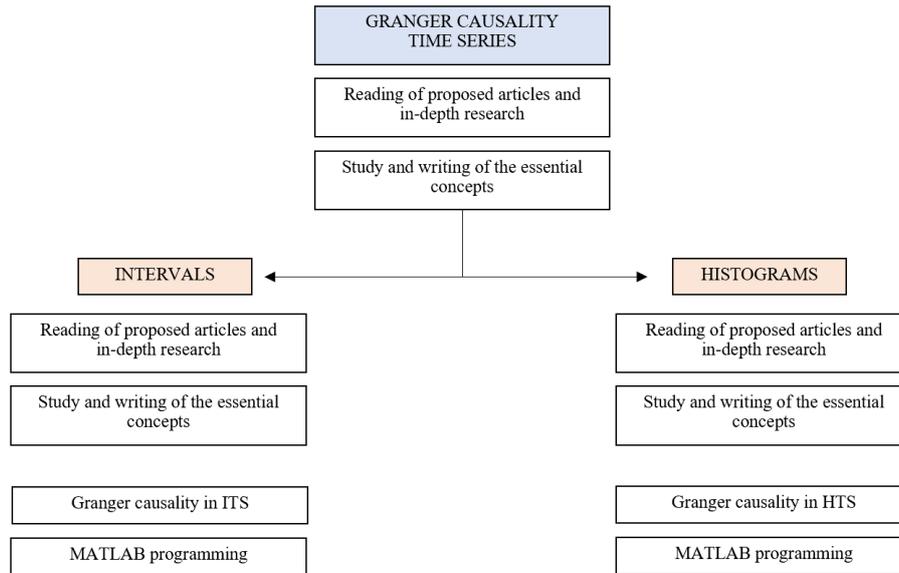
Throughout this project, Granger causality tests will be performed by applying them to interval and histogram data. In order to carry out this study, certain past events will be evaluated with data of different frequencies and the causality between different assets over time will be analyzed.

### **3. Model description**

The only tool used throughout the project was MATLAB. This programming language has been used for different purposes.

Firstly, for the analysis and representation of interval and histogram time series. In addition, functions have also been programmed for the prediction of values based on different regression methods. Finally, MATLAB's own functions have been used to perform the Granger causality tests and thus obtain the results.

The process of studying each of the two parts into which the project as a whole could be divided (intervals and histograms) has been very similar. An image is shown in which an outline of the process followed has been made:



*Illustration 3. Scheme of the model followed during the Project*

As can be seen in Illustration 3, the project starts from the study of the concept of Granger causality and time series in general. For this purpose, numerous articles, journals, books and web pages have been read and referenced throughout the project.

Once these concepts were understood, we proceeded to study intervals and interval time series (ITS). Then Granger causality is applied with real life examples. The same was done with histograms and histogram time series (HTS). All this has been done with functions programmed in MATLAB.

## **4. Results**

After a thorough evaluation and after having programmed numerous developments in MATLAB, the following results have been obtained.

### **4.1 Interval Time Series (ITS)**

In relation to interval time series, two parallel studies have been carried out.

In the first one, Granger causality has been studied in the interval time series of gold and Brent prices from 2010 to 2021. The results obtained in this study indicate that there is no causality between these two interval time series, or in other words, neither of them helps to predict the other. These results reflect a certain robustness since five different linear regression methods have been taken into account and all of them show the same results of non-causality.

In the second, Granger causality has been analyzed in the interval time series of gold and Brent prices at the times before, during and after the pandemic. The results obtained indicate that in the year 2019 there was a weak causality between gold and Brent, as well as in the year 2022. However in the year 2020 the causality between gold and Brent price interval time series was much stronger. It should be noted that the causality results obtained were at all times unidirectional, i.e. only from Brent to gold, and not in the opposite direction.

In addition, in order to determine the most accurate interval time series linear regression method, some of the most characteristic metrics have been obtained. The following results have been obtained, shown in Illustration 4. The study includes five interval regression methods (CM, mM, CRM, CCRM and MG), which are explained in more depth throughout the development work. The iARV and CER metrics measure the development accuracy of these methods. A lower value for iARV indicates higher accuracy, as does a higher value for CER.

MÉTRICAS	CM	mM	CRM	CCRM	MG
<b>iARV</b>	0.0603	0.0603	0.0517	0.0519	<b>0.0497</b>
<b>CER</b>	0.6193	0.6073	0.6507	0.6440	<b>0.6637</b>

*Illustration 4. Results iARV and CER metrics for each of the interval linear regression methods*

## 4.2 Histogram Time Series (HTS)

A similar analysis has been performed to study Granger causality in histogram time series. The results in this case indicate that, when performing an analysis over a large time period (more than 10 years), we see virtually no causality in either direction.

However, if we reduce this time period to the pandemic years, roughly the most recent two years, we find bidirectional Granger causality results between the two time series of gold and Brent.

## 5. Conclusions

After obtaining the results shown above, the following conclusions are reached.

### 5.1 Interval Time Series (ITS)

The general conclusion in the Granger causality analysis in the ITS is that the causality of the time series from Brent to gold (unidirectional) is much stronger at those times when the pandemic crisis has been more pronounced, weakening in the economic recovery.

It is concluded that causality is not detected if we analyze two time series over a long period of time spanning several cycles of the economy (boom and bust), but that causality is present when we focus only on a period of crisis, such as the global economic crisis caused by the Covid-19 pandemic.

On the other hand, it is concluded that the method with the best metrics, as expected since it is the most recent and complex, is the MG model, with the highest CER value and the lowest iARV value. This is why we can affirm that the MG model is the most accurate and truthful of those studied. Even so, the previous analyses and results have been obtained with the combination of all the methods, with the objective of presenting a greater robustness in the results.

## **5.2 Histogram Time Series (HTS)**

The results obtained in the study of Granger causality in HTS are not surprising since, as we had observed in the study with ITS, causality is much stronger in those moments in which a crisis has been experienced. That is, in this analysis we do not find causality when we study an extended period in time, in which there have been boom times and crisis times, however the causality is much stronger when we shorten the study to a shorter period of time, in which we were undergoing a severe global economic crisis due to the pandemic.

Moreover, we can observe that, unlike the analysis with interval time series, the causality in this case is bidirectional. With this study it can be stated that in this crisis period, Brent causes gold as we had observed with STIs, but it is further stated that gold also causes Brent.

These results make sense given that histogram time series provide more information than interval time series by reflecting the distribution of the data. Thus, by giving more complete information on the phenomenon as a whole, it is able to appreciate causality in both directions.

Therefore, as a general reflection, the conclusions obtained in the study of Granger causality in HTS reaffirm and complement those already obtained in the analysis of Granger causality with ITS.

## **6. References**

Billard, L., & Diday, E. (2000). Regression analysis for interval-valued data. In *Data analysis, classification, and related methods* (pp. 369-374). Springer, Berlin, Heidelberg.

Blanco-Fernández, A., Corral, N., & González-Rodríguez, G. (2011). Estimation of a flexible simple linear model for interval data based on set arithmetic. *Computational Statistics & Data Analysis*, 55(9), 2568-2578.

García-Bárzana, M., Ramos-Guajardo, A. B., Colubi, A., & Kontoghiorghes, E. J. (2020). Multiple linear regression models for random intervals: a set arithmetic approach. *Computational Statistics*, 35(2), 755-773.

Maté, C. G. (2021). Combining Interval Time Series Forecasts. A First Step in a Long Way (Research Agenda). *Revista Colombiana de Estadística*, 44(1), 123-157.

Troster, V. (2018). Testing for Granger-causality in quantiles. *Econometric Reviews*, 37(8), 850-866.

## Índice de la memoria

<b>Capítulo 1. Introducción .....</b>	<b>7</b>
Motivación del proyecto .....	8
Objetivos del proyecto .....	9
1.1.1 Teórico.....	9
1.1.2 Práctico .....	9
Metodología del proyecto .....	10
<b>Capítulo 2. Causalidad de Granger .....</b>	<b>11</b>
Conceptos generales.....	11
2.1.1 Series temporales univariantes y multivariantes.....	11
2.1.2 Modelos VAR.....	12
2.1.3 Concepto de la causalidad de Granger.....	14
Análisis estadístico (contrastes de hipótesis).....	14
2.1.4 Leave-one-out Test .....	16
2.1.5 Exclude-all Test.....	17
2.1.6 Block Exogeneity Test.....	18
Ejemplos en los mercados financieros .....	20
<b>Capítulo 3. Análisis de datos valorados por intervalos .....</b>	<b>26</b>
Características y notación de los datos de intervalo .....	28
Series temporales de intervalo (ITS).....	29
3.1.1 Conceptos estadísticos de las ITS.....	30
Visualización .....	32
3.1.2 Visualización de intervalos transversales .....	32
3.1.3 Visualización de series temporales de intervalos.....	33
<b>Capítulo 4. Regresión lineal simple y múltiple para datos de intervalo .....</b>	<b>35</b>
Paradigmas.....	37
4.1.1 Paradigma 1 .....	37
4.1.2 Paradigma 2 .....	37
4.1.3 Paradigma 3 .....	38
4.1.4 Paradigma 4.....	38

Métodos de regresión lineal para intervalos .....	38
4.1.5 Método de centro (CM-Centre Method).....	38
4.1.6 Método de los mínimos y máximos (mM-MinMax Method).....	40
4.1.7 Método del centro y rango (CRM-Centre and Range Method).....	42
4.1.8 Método del centro y rango restringido (CCRM-Constrained Centre and Range Method) .....	44
4.1.9 Modelo MG.....	46
<b>Capítulo 5. Causalidad de Granger en ITS .....</b>	<b>53</b>
Estudio de la causalidad de Granger en las series temporales de intervalo del oro y el petróleo (2010-2021) en función de los diferentes métodos de regresión .....	53
5.1.1 Análisis.....	54
5.1.2 Resultados .....	56
La causalidad de Granger bajo el enfoque de series temporales de intervalo. El caso del oro y el petróleo antes, durante y después de la crisis de la pandemia del Covid-19.....	59
5.1.3 Análisis .....	59
5.1.4 Resultados .....	60
Estudio de las métricas de precisión iARV Y CER para cada uno de los métodos de regresión de intervalos.....	70
5.1.5 Varianza relativa media del intervalo (Interval Average Relative variance - iARV).....	70
5.1.6 Tasa de cubrimiento (Coverage Rate - CR) .....	71
5.1.7 Tasa de eficiencia (Efficiency Rate - ER).....	71
5.1.8 Ratio de cubrimiento y eficiencia (Coverage and Efficiency Rate – CER) .....	72
5.1.9 Resultados .....	73
<b>Capítulo 6. Análisis de datos valorados por histogramas .....</b>	<b>75</b>
Ventajas de los histogramas.....	77
6.1.1 Representación cercana a los datos originales.....	77
6.1.2 Versatilidad .....	77
6.1.3 Precisión.....	77
6.1.4 Facilidad en el tratamiento computacional.....	77
Tipos de histogramas .....	78
6.1.5 Histogramas equiespaciados.....	78
6.1.6 Histogramas equifrecuenciales .....	78
6.1.7 Histogramas contruidos sobre una partición del dominio de la variable .....	78
6.1.8 Histogramas definidos a partir de una secuencia de cuartiles .....	78

Series temporales de histograma (HTS).....	79
Métodos de regresión lineal de las HTS .....	81
Medidas de precisión de las HTS.....	82
<b>Capítulo 7. Causalidad de Granger en HTS.....</b>	<b>85</b>
La causalidad de Granger bajo el enfoque de series temporales de histograma. El caso del oro y el petróleo antes, durante y después de la crisis de la pandemia del Covid-19.....	85
7.1.1 Análisis .....	89
7.1.2 Resultados .....	92
<b>Capítulo 8. Conclusiones y Trabajos Futuros.....</b>	<b>94</b>
Conclusiones .....	94
8.1.1 Conclusiones respecto a los objetivos.....	94
8.1.2 Conclusiones respecto a los resultados.....	94
Desarrollos futuros.....	97
<b>Capítulo 9. Bibliografía.....</b>	<b>98</b>
<b>ANEXO I: ALINEACIÓN DEL PROYECTO CON LOS ODS .....</b>	<b>101</b>
Objetivo 4: Garantizar una educación inclusiva, equitativa y de calidad y promover oportunidades de aprendizaje durante toda la vida para todos:.....	101
Objetivo 8: Promover el crecimiento económico inclusivo y sostenible, el empleo y el trabajo decente para todos:.....	102
Objetivo 10: Reducir la desigualdad en y entre todos los países: .....	102

## Índice de figuras

Figura 1. Resultados ejemplo Leave-One-Out Test .....	17
Figura 2. Resultados ejemplo Exclude-All Test.....	18
Figura 3. Resultados ejemplo Block Exogeneity Test.....	19
Figura 4. Alta correlación positiva entre las series temporales de precios del oro y del petróleo (2000-2007), (Zhang & Wei, 2010) .....	21
Figura 5. Evolución de las series temporales de precios del oro y del petróleo (2009-2015), (Kanjilal & Ghosh, 2017) .....	23
Figura 6. Representación gráfica de los intervalos de una variable (Maceda, 2018).....	33
Figura 7. Representación gráfica de los intervalos de dos variables (Maceda, 2018).....	33
Figura 8. Representación gráfica de dos ITS.....	34
Figura 9. Ejemplo de Método del Centro (Maceda, 2018).....	40
Figura 10. Ejemplo del Método MinMax (Maceda, 2018) .....	42
Figura 11. Ejemplo del Método del Centro y Rango (Maceda, 2018) .....	44
Figura 12. Ejemplo del Método del Centro y Rango Restringido (Maceda, 2018).....	46
Figura 13. Representación del Modelo MG AR (de los Santos, 2021).....	49
Figura 14. Evolución del error del Modelo MG (de los Santos, 2021).....	50
Figura 15. Evolución del $R^2$ del Modelo MG (de los Santos, 2021).....	50
Figura 16. Evolución de los precios de las ITS del oro y del petróleo (2010-2021).....	53
Figura 17. Media y mediana de las ITS del oro y del petróleo (2010-2021) .....	54
Figura 18. Gráficos de causalidad ITS del oro y del petróleo (2010-2021). Métodos CM, mM y CRM .....	57
Figura 19. Gráficos de causalidad ITS del oro y del petróleo (2010-2021). Métodos CCRM y MG.....	58
Figura 20. Evolución del PIB de España (1996-2021) (Instituto Nacional de Estadística, 2022).....	68
Figura 21. Ejemplos de Ratios de Cobertura y Eficiencia (Maceda, 2018) .....	72

Figura 22. Resultados CER para cada uno de los métodos .....	73
Figura 23. Ejemplos de HTS (Arroyo & Maté, 2009).....	76
Figura 24. HTS de las precipitaciones mensuales en China (1953-1954) (Arroyo & Maté, 2009).....	80
Figura 25. HTS de la distribución mensual de dióxido de nitrógeno en Madrid (Arroyo, 2008).....	80
Figura 26. Representación de la regresión lineal con datos clásicos, datos de intervalo (método MinMax) y datos de histograma (Irpino, 2010), (de los Santos, 2021) .....	81
Figura 27. Gráfico de caja de la distribución de precios del oro (01/01/2018-20/02/2022)	86
Figura 28. Gráfico de caja de la distribución de precios del petróleo (01/01/2018-20/02/2022) .....	86
Figura 29. Representación serie temporal de histogramas de precios del oro (1º Semestre 2018-1º Semestre 2022).....	87
Figura 30. Representación serie temporal de histogramas de precios del petróleo (1º Semestre 2018-1º Semestre 2022) .....	88
Figura 31. Representación de los cuartiles 25, 50 y 75 de la serie temporal de histogramas del oro (2016-2022).....	89
Figura 32. Representación de los cuartiles 25, 50 y 75 de la serie temporal de histogramas del petróleo (2016-2022) .....	90

## *Índice de tablas*

Tabla 1. Datos de intervalo en la medición de resultados de laboratorio .....	28
Tabla 2. Resultados Causalidad de Granger ITS del oro y el petróleo. Primer semestre del año 2019 .....	60
Tabla 3. Resultados Causalidad de Granger ITS del oro y el petróleo. Segundo semestre del año 2019 .....	61
Tabla 4. Resultados Causalidad de Granger ITS del oro y el petróleo. Primer semestre del año 2020 .....	62
Tabla 5. Resultados Causalidad de Granger ITS del oro y el petróleo. Segundo semestre del año 2020 .....	63
Tabla 6. Resultados Causalidad de Granger ITS del oro y el petróleo. Primer semestre del año 2021 .....	64
Tabla 7. Resultados Causalidad de Granger ITS del oro y el petróleo. Segundo semestre del año 2021 .....	65
Tabla 8. Resultados Causalidad de Granger ITS del oro y el petróleo. Primer semestre del año 2022 .....	66
Tabla 9. Resultados iARV y CER para cada uno de los métodos de regresión .....	74
Tabla 10. Resultados Causalidad de Granger HTS del oro y el petróleo (2010-2022) .....	92
Tabla 11. Resultados Causalidad de Granger HTS del oro y el petróleo (2019-2022) .....	92

## Capítulo 1. INTRODUCCIÓN

En la realización de este proyecto se va a llevar a cabo un estudio de la causalidad de Granger en la predicción de las series temporales. Según la causalidad de Granger, una serie temporal  $X_t$  causa otra serie temporal  $Y_t$  si es posible predecir mejor  $Y_t$  utilizando toda la información disponible en lugar de utilizar toda la información disponible sin  $X_t$ . La causalidad de Granger se refiere a la disminución de la varianza del error de previsión.

Resulta bien conocida la aplicación de este concepto en las series temporales nítidas, pero este proyecto pretende ser una extensión de lo ya estudiado, por lo que se trabajará también con series temporales de intervalos y de histogramas. Las series temporales clásicas describen cada instante con un número real, sin reflejar la variabilidad asociada. Sin embargo, la información en cada periodo de tiempo puede venir dada como un intervalo o un histograma, reflejando dicha variabilidad a lo largo del tiempo.

A lo largo del proyecto, el cual incluirá aplicaciones a los mercados financieros, se investigará la causalidad de Granger para realizar modelos de predicción de series temporales con datos clásicos, datos de intervalos (ITS) e histogramas (HTS). Se creará un sistema original, básico y visual para el análisis y la previsión de precios diarios, semanales y mensuales de los mínimos, máximos, centros y radios, considerando o no el concepto anterior.

## ***MOTIVACIÓN DEL PROYECTO***

La predicción consiste en obtener conocimientos anticipados de hechos en el futuro con el objetivo de poder prepararse de forma adecuada para hacerle frente de la mejor manera posible.

La predicción es una actividad fundamental hoy en día, y se aplica en áreas muy diversas como la ingeniería, la economía, las finanzas, el data mining, la meteorología, las ciencias sociales, etc. Por ejemplo, si se desea invertir una gran cantidad de dinero en acciones con el objetivo de obtener una rentabilidad a lo largo del tiempo, es preciso realizar previamente un análisis de predicción de los precios de dichas acciones. El mercado bursátil es un mercado volátil, esto es, cambia de forma imprevisible. Si se analiza la evolución futura de los precios de los activos financieros resulta más fácil diseñar un portafolio de acciones que haga muy probable obtener cierta rentabilidad sin riesgo asociado.

El contenido de este TFG se centrará en la predicción de magnitudes cuantitativas, y la aproximación se realizará a partir de series temporales, esto es, del análisis matemático de los datos históricos de las magnitudes. Las series temporales clásicas tienen unos 100 años de historia, y durante este tiempo también se han investigado las series temporales de intervalo y de histograma para aportar una mayor precisión.

La predicción con series temporales se ha estudiado en numerosas ocasiones, pero este TFG pretende incorporar la aplicación del concepto de la causalidad de Granger para comprobar si mejoran los resultados al predecir. Tal y como se ha explicado anteriormente, la predicción puede resultar muy relevante en determinadas ocasiones y sería de gran utilidad aportar los conocimientos que proporciona la causalidad de Granger con el objetivo de obtener resultados más precisos y, por tanto, poder anticiparse de una mejor manera al futuro.

## ***OBJETIVOS DEL PROYECTO***

Se describen los objetivos de este TFG desde los puntos de vista teórico y práctico.

### **1.1.1 TEÓRICO**

- Conocer el concepto de la causalidad de Granger. Dominar su análisis estadístico y contrastes de hipótesis.
- Conocer el tratamiento y análisis de las series temporales, especialmente las series de datos de intervalo (ITS) y de histogramas (HTS).
- Profundizar en la regresión lineal simple y múltiple para los datos de intervalo. Conocer los modelos propuestos para llevar a cabo las predicciones.

### **1.1.2 PRÁCTICO**

- Comprender la causalidad de Granger en ITS y HTS.
- Programar en MATLAB sistemas de análisis causal para la aplicación de la causalidad de Granger a series temporales de intervalos y de histogramas.
- Realizar un sistema original, básico y visual para el análisis y la previsión de los precios diarios, semanales y mensuales de los mínimos y máximos, teniendo en cuenta la causalidad de Granger.

## ***METODOLOGÍA DEL PROYECTO***

El primer paso para desarrollar cualquier trabajo es definir el proyecto a realizar. Posteriormente se deben determinar los objetivos que se persiguen (descritos en el apartado anterior), así como los pasos a seguir para la realización de las tareas, incluyendo la planificación de estas.

Para llevar a cabo este proyecto, es necesario hacer previamente una amplia investigación sobre algunos conceptos básicos relacionados con el tema del trabajo, entre los que se encuentran: los fundamentos de la causalidad de Granger, los conceptos y métodos básicos de la predicción y regresión, y las series temporales clásicas, de intervalos y de histogramas.

Además, es preciso aprender una serie de conceptos estadísticos muy útiles que se aplicarán en los diferentes capítulos del TFG. Se tomarán ejemplos reales para incorporarlos a dicho estudio. Toda esta investigación corresponde a los primeros capítulos.

Una vez se hayan estudiado con profundidad las diferentes series temporales, se aplicará la causalidad de Granger en cada una de ellas (clásicas, de intervalos y de histogramas). Se hará uso del lenguaje de programación MATLAB para crear funciones que expliquen el sistema de análisis causal según Granger para los diferentes tipos de series temporales. Esta parte es fundamental en el proyecto, en ella se aporta información relevante y de elaboración propia, y corresponde a los últimos capítulos del proyecto.

## **Capítulo 2. CAUSALIDAD DE GRANGER**

A lo largo de este apartado se va a introducir al lector en el tema del que va a tratar el resto del proyecto, la causalidad de Granger. Se explican los conceptos generales junto con su correspondiente análisis estadístico y contrastes de hipótesis. Además, se incorporan ejemplos de la aplicación de la causalidad de Granger en los mercados financieros según diferentes autores.

### ***CONCEPTOS GENERALES***

#### **2.1.1 SERIES TEMPORALES UNIVARIANTES Y MULTIVARIANTES**

Las series temporales pueden ser univariantes o multivariantes. Las series univariantes se caracterizan, como su nombre indica, por depender únicamente de una variable para predecir los valores futuros. Por ejemplo, en el caso de Google, se podría predecir el precio de sus acciones en el futuro conociendo únicamente dichos precios en los meses anteriores.

Muchas otras veces necesitamos de más de una variable para predecir de una mejor manera los valores de la serie en el futuro, además de para entender la relación entre sus diferentes componentes. Por ejemplo, si queremos predecir el IPC (Índice de Precios al Consumo) de Estados Unidos, necesitaríamos los valores de este en el tiempo pero también sería razonable contar con el tipo de interés nominal de EEUU en el mismo periodo de tiempo, ya que se sabe que ambas variables están relacionadas entre sí.

Otro ejemplo sería en la predicción del PIB per cápita de cualquier país de la OECD. Si se analizan múltiples países de la OECD se observa que todos ellos siguen prácticamente la misma tendencia ascendente, por lo que se podría predecir con cierta seguridad el mismo crecimiento en cualquier otro país de la OCDE que se quiera analizar.

Podemos afirmar que conocer la relación y dependencias existentes entre las variables es importante para obtener una mejor predicción. Estos movimientos simultáneos entre variables suelen ser muy comunes en series financieras, afectando por ejemplo la fluctuación

de la bolsa de Londres en la bolsa de Nueva York y siendo entonces imprescindible conocer dicha relación para obtener mejores resultados.

### 2.1.2 MODELOS VAR

El Modelo Vectorial Auto Regresivo es un modelo de serie de tiempo multivariante que relaciona las observaciones actuales de una variable con observaciones pasadas de la misma, además de con observaciones pasadas de otras variables (Eric, 2021).

A diferencia de los modelos autorregresivos univariados, los modelos VAR permiten que se produzca una retroalimentación entre las diferentes variables que componen el modelo. Por ejemplo, podríamos usar un modelo VAR para estudiar el precio de los *forfaits* de una estación de esquí en una determinada temporada (2021-2022). Para ello utilizamos como variables explicativas el precio de los *forfaits* y el número de pistas abiertas de las temporadas anteriores (2019-2020 y 2020-2021). Sin embargo, si se empleara un modelo Auto Regresivo (AR) clásico, las variables explicativas sólo comprenderían los valores del precio de los *forfaits* en temporadas pasadas, sin tener en cuenta la variable del número de pistas abiertas, que podría aportar mucha información relevante.

Este modelo, por tanto, tiene ciertas ventajas frente a otros modelos:

- Presenta un mejor rendimiento en el pronóstico.
- Ofrece un enfoque sistemático pero flexible con el objetivo de capturar el comportamiento complejo del mundo real.
- Otorga la capacidad de capturar la dinámica entrelazada de datos de las series temporales.

Como cualquier otro modelo, los Modelos VAR tienen numerosas aplicaciones en la vida real y se usan habitualmente en numerosos campos tales como las finanzas, biología, epidemiología, medicina, econometría, etc. A continuación se muestran algunos ejemplos:

- **Economía:** los modelos VAR se aplican para modelar la relación existente entre el ingreso personal y el gasto de consumo a lo largo del tiempo. Para ello se utiliza un sistema VAR de dos ecuaciones (dos variables explicativas).
- **Medicina:** los modelos VAR se utilizan en la medicina para modelar las relaciones pasadas y actuales entre la frecuencia cardíaca, la frecuencia respiratoria, la presión arterial y la SpO<sub>2</sub>.
- **Epidemiología:** Se utilizan para modelar las relaciones entre las tasas de infección de diferentes grupos de edad. Para ello realizan un conteo de las infecciones pasadas en los diferentes grupos de edad.

Estas son algunas de las aplicaciones en la vida real aunque existen muchos más, pudiéndose aplicar prácticamente a cualquier campo de la ciencia.

Un modelo VAR se compone de un sistema de ecuaciones que representa las relaciones entre las múltiples variables del modelo. El lenguaje con el que especificamos cada modelo hace referencia al número de variables endógenas de las que se compone el modelo y al número de términos autorregresivos que incluye. Un modelo VAR se compone de  $n$  ecuaciones (siendo  $n$  el número de variables endógenas, explicativas) e incluye  $p$  retrasos de las variables.

Por ejemplo, si tenemos dos variables endógenas y dos términos autorregresivos, decimos que el modelo es un modelo bivariado VAR(2). Si tenemos tres variables endógenas y cuatro términos autorregresivos, decimos que el modelo es un modelo Trivariado VAR(4).

Es fundamental elegir el número de variables endógenas que incluir en el modelo. Para ello, nos podemos ayudar de herramientas adicionales como la causalidad de Granger o la causalidad de Sims, con el objetivo de probar la relevancia de pronóstico de las variables. En este sentido, la causalidad de Granger prueba si una variable es "útil" para pronosticar el comportamiento de otra variable. Sus estadísticas indican si una variable es significativa al predecir otra variable.

### 2.1.3 CONCEPTO DE LA CAUSALIDAD DE GRANGER

La causalidad de Granger se define con la siguiente afirmación: “Una serie temporal  $X_t$  causa otra serie temporal  $Y_t$  si los valores pasados de  $X_t$  ayudan a predecir los valores futuros de  $Y_t$ ”. Esto quiere decir que  $Y_t$  no sólo sería una función que depende de los valores pasados de  $Y_t$ , es decir,  $Y_{t-p}$ , sino que también dependería de los valores pasados de la serie temporal  $X_t$ , es decir,  $X_{t-p}$ , al aportar esta información relevante.

La causalidad de Granger se da siempre teniendo en cuenta dos condiciones. La primera de ellas implica que la causa se da antes que el efecto, esto es, el efecto  $Y_t$  es una función que depende siempre de valores pasados de la causa  $X_t$ , nunca de valores simultáneos ni futuros. La segunda de ellas establece que la causa tiene información única para predecir los valores futuros de  $Y_t$ .

### ***ANÁLISIS ESTADÍSTICO (CONTRASTES DE HIPÓTESIS)***

Matemáticamente podemos definir la causalidad de Granger. En primer lugar, escribimos un modelo sencillo, en el que salida  $Y_t$  tan solo depende de valores anteriores de la misma serie temporal (retardos) y de valores anteriores de otra serie temporal, a la que denominamos  $X_t$ .

La ecuación resultante queda definida de la siguiente manera:

$$Y_t = \alpha + \sum_{i=1}^p \gamma_i Y_{t-i} + \sum_{j=1}^q \beta_j X_{t-j} + \varepsilon_t$$

donde  $\gamma_i$  son los coeficientes a estimar en la regresión de la primera serie temporal  $Y_t$ . Por otra parte,  $\beta_j$  son los coeficientes correspondientes a la segunda serie temporal  $X_t$ . Como cualquier serie temporal incluye un error,  $\varepsilon_t$ . El parámetro  $\alpha$  es un término constante.

A continuación, planteamos la hipótesis de que  $X_t$  causa  $Y_t$  o de que esto no ocurra. En el caso de que sí que lo hiciera, la serie temporal  $Y_t$  dependería de los valores pasados de la misma además de valores pasados de otra serie temporal  $X_t$ , ya que esta segunda añadiría valor y mejoraría la predicción de la salida.

Para comprobar si se cumple o no la causalidad de Granger realizamos el t-test para cada coeficiente, preguntándonos lo significativos que son los coeficientes  $\beta_j$ . En el caso de que ninguno de ellos sea significativo, es decir, sean todos iguales a cero (se cumple la hipótesis nula,  $H_0$ ) quiere decir que la serie temporal  $X_t$  no causa  $Y_t$ . Por el contrario, basta con que sólo uno de ellos no sea cero (se cumple la hipótesis alternativa) para determinar que dicho coeficiente es significativo y se cumple la causalidad de Granger, afirmando entonces que  $X_t$  causa  $Y_t$ .

$$H_0 : \beta_0 = \beta_1 = \beta_2 = \dots = \beta_q = 0$$

Tenemos que ser conscientes de que en el caso de que se cumpla la causalidad de Granger, la introducción de la serie temporal  $x(t)$  en el modelo mejora la predicción del modelo, aportando valor.

La causalidad de Granger tiene ciertas limitaciones entre las que se encuentran las dos siguientes: no es necesariamente una causalidad verdadera, además de que puede ser complicado encontrar la causalidad existente si la serie temporal  $X_t$  afecta a  $Y_t$  a través de una tercera variable  $Z_t$ .

Los tests de causalidad de Granger que se han realizado a lo largo del proyecto han utilizado la función *gctest* de MATLAB. Esta función realiza pruebas de causalidad de Granger de tipo *leave-one-out*, *exclude-all* y *block-wise* para las variables de respuesta de un modelo de autoregresión vectorial (VAR) completamente especificado (representado por *varm model*). Toda la información relativa a esta función se encuentra en la documentación de MATLAB<sup>1</sup>.

A continuación se explica el funcionamiento de cada uno de los tests de la causalidad de Granger en los modelos VAR. En la explicación se tienen en cuenta únicamente los

---

<sup>1</sup> Links a la documentación de la función: (1) [https://es.mathworks.com/help/econ/varm.gctest.html#mw\\_62f43143-b162-443c-ac5e-4b1c19233a32](https://es.mathworks.com/help/econ/varm.gctest.html#mw_62f43143-b162-443c-ac5e-4b1c19233a32), (2) <https://es.mathworks.com/help/econ/gctest.html>

parámetros básicos y los valores por defecto para simplificar, pero se podrían añadir otros parámetros con el objetivo de cambiar algunas condiciones del test. Por ejemplo:

- **Alpha:** parámetro que modifica el nivel de confianza al que se realiza el test (por defecto es un 5% de error). Puede tomar valores entre 0 y 1.
- **Test:** distribución estadística que actúa sobre la hipótesis nula. Puede utilizarse la distribución chi-cuadrado o el estadístico F.
- **Display:** muestra por pantalla una tabla de resumen de test. Puede tomar valores true o false.

Estos son algunos de los parámetros de input, pero hay varios más. Si se quieren conocer todos se debe recurrir a la documentación de MATLAB.

#### 2.1.4 LEAVE-ONE-OUT TEST

Este tipo de prueba realiza  $m*(m-1)$  tests, siendo  $m$  el número de variables explicativas del modelo (series temporales). Por ejemplo, si se realiza un modelo de tres series temporales VAR(2), tendremos un total de 6 tests ya que  $3*(3-1)=3*2=6$ .

Para  $j=1..m$ ,  $k=1..m$  y  $j \neq k$ , la función *gctest* prueba la hipótesis nula de que la variable  $j$  no es causa de Granger de la variable  $k$ , condicionada por todas las demás variables del modelo.

La función va probando de esta manera una a una y después se puede mostrar una tabla con todos los resultados. La hipótesis nula indica que todos los coeficientes de esa variable en la otra variable son cero, por tanto significa que la variable que estamos analizando no causa a la otra. Sin embargo, si se cumple la hipótesis alternativa quiere decir que hay al menos un coeficiente que no es cero, por tanto la variable causa la otra.

Este modelo nos muestra para cada una de las posibles combinaciones si se rechaza o no la hipótesis nula. En el caso de que se tenga suficiente evidencia y se pueda rechazar la hipótesis nula indica que esa variable causa la otra.

Adjuntamos una captura de pantalla (Figura 1) de una simulación de este tipo de prueba en MATLAB. El ejemplo está realizado sobre un modelo VAR(3) y con 3 series temporales.

```
>> h = gctest(BestMdl)
```

H0	Decision	Distribution	Statistic	PValue	CriticalValue
"Exclude lagged inflation in mslrate equation"	"Cannot reject H0"	"Chi2(3)"	7.0674	0.069782	7.8147
"Exclude lagged rgdprate in mslrate equation"	"Cannot reject H0"	"Chi2(3)"	2.5585	0.4648	7.8147
"Exclude lagged mslrate in inflation equation"	"Cannot reject H0"	"Chi2(3)"	2.7025	0.4398	7.8147
"Exclude lagged rgdprate in inflation equation"	"Reject H0"	"Chi2(3)"	14.338	0.0024796	7.8147
"Exclude lagged mslrate in rgdprate equation"	"Cannot reject H0"	"Chi2(3)"	7.0352	0.070785	7.8147
"Exclude lagged inflation in rgdprate equation"	"Reject H0"	"Chi2(3)"	12.006	0.0073619	7.8147

```
h =
6x1 logical array
0
0
0
1
0
1
```

Figura 1. Resultados ejemplo Leave-One-Out Test

Comprobamos que el número de tests es 6. En cuanto a los resultados, solo dos de las variables cumplen la causalidad de Granger. La variable *rgdprate* causa a *inflation* y viceversa. Hay que mencionar que en los tests en los que hemos rechazado la hipótesis nula son aquellos que tenían el estadístico más alto y el pvalor más bajo, como resulta lógico. Los estadísticos que podemos utilizar con la función `gctest` son chi-cuadrado o F-test, y en este caso hemos utilizado la distribución chi-cuadrado (por defecto).

Este tipo de prueba se denomina *Leave-One-Out* porque el test de una variable cualquiera se lleva a cabo teniendo en cuenta el efecto de todas las demás variables por separado pero excluyéndose a sí misma.

### 2.1.5 EXCLUDE-ALL TEST

Este tipo de prueba funciona de una manera diferente a al anterior. En primer lugar, el número de tests que se realizan internamente es diferente. En este caso únicamente se realizan  $m$  tests, es decir, una por cada variable explicativa (serie temporal) que contenga el modelo VAR con el que estemos trabajando.

Para  $j=1, \dots, m$  (cada serie temporal) la función `gctest` prueba la hipótesis nula de que todas las demás variables del modelo no causan conjuntamente la variable de respuesta  $j$ . En el

caso de que no haya evidencia suficiente para rechazar la hipótesis nula entendemos que se cumple, por lo que las variables del modelo en su conjunto NO causan la variable respuesta. Sin embargo, si podemos decir que se rechaza la hipótesis nula y por tanto se cumple la alternativa, entonces tenemos certeza de que el resto de variables en su conjunto SÍ causan la variable respuesta.

Todas estas respuestas siempre se dan con un 95% de nivel de confianza (como parámetro de *default*), pudiéndose cambiar actualizando uno de los parámetros de la función.

En la Figura 2 se muestra una captura de pantalla de una simulación de este tipo de prueba en MATLAB. El ejemplo es el mismo que el anterior y se ha realizado sobre un modelo VAR(3) y con 3 series temporales.

```
>> h_2 = gctest(EstMdl_2, 'Type', 'exclude-all')
      H0
```

	Decision	Distribution	Statistic	PValue	CriticalValue
"Exclude all but lagged mslrate in mslrate equation"	"Cannot reject H0"	"Chi2 (6)"	9.477	0.14847	12.592
"Exclude all but lagged inflation in inflation equation"	"Reject H0"	"Chi2 (6)"	19.475	0.0034327	12.592
"Exclude all but lagged rgdprate in rgdprate equation"	"Reject H0"	"Chi2 (6)"	19.16	0.0039014	12.592

```
h_2 =
3x1 logical array
0
1
1
```

Figura 2. Resultados ejemplo Exclude-All Test

Comprobamos que, en este caso, el número de tests que se han realizado es únicamente 3. En la primera línea se ve como no hay evidencia suficiente para rechazar la hipótesis nula mientras que sí que la hay en los otros dos casos.

### 2.1.6 BLOCK EXOGENEITY TEST

Una variable es *block exogenous* si no causa ninguna otra variable del modelo VAR. Este test prueba la hipótesis nula de que las variables de respuesta especificadas en la Causa no causan conjuntamente las variables de respuesta especificadas en el Efecto. Esta configuración por tanto realiza únicamente 1 test, en el que se comprueba si la Causa causa al Efecto. En el caso de que no se especifiquen variables de Causa ni Efectos gctest probará

la hipótesis nula de que todos los coeficientes de rezago, incluyendo los rezagos propios, son cero.

Si no existe evidencia suficiente para rechazar la hipótesis nula quiere decir que las variables que hayan sido definidas por causas no causan conjuntamente a aquellas que han sido definidas como efectos. Sin embargo, si se puede rechazar la hipótesis nula (y por tanto aceptar la hipótesis alternativa), se confirma que las variables de causa causan conjuntamente aquellas definidas como respuesta.

En la Figura 3 que se muestra a continuación se puede visualizar lo explicado anteriormente. En este caso la variable *cause* era *dfedfunds* mientras que las variables *effect* eran: *inflation*, *money supply* y *real GDP rate*. Se realiza el test de causalidad de Granger y como resultado muestra que, a un 95% de nivel de confianza, se puede rechazar la hipótesis nula. Esto implica que la variable establecida como causa causa al resto de variables definidas como efectos.

```
gctest(EstMdl, 'Cause', cause, 'Effect', effects);
```

H0	Decision	Distribution	Statistic	PValue	CriticalValue
"Exclude lagged dfedfunds in inflation equation"	"Reject H0"	"Chi2(3)"	26.157	8.8433e-06	7.8147
"Exclude lagged dfedfunds in pcerate equation"	"Reject H0"	"Chi2(3)"	10.151	0.017325	7.8147
"Exclude lagged dfedfunds in rgdprate equation"	"Reject H0"	"Chi2(3)"	10.651	0.013772	7.8147

```
>> h_3 = adftest(DataTable.FEDFUNDS, 'Model', "ard")
h_3 =
logical
0
```

Figura 3. Resultados ejemplo Block Exogeneity Test

Una vez más se comprueba que los estadísticos toman valores altos en los casos en los que se haya rechazado la hipótesis nula. Además, los p-valores son muy pequeños, muy cercanos a 0. La distribución estadística que se ha empleado para realizar este test es chi-cuadrado.

## ***EJEMPLOS EN LOS MERCADOS FINANCIEROS***

Actualmente existen varios artículos en los se han realizado análisis acerca de la posible relación existente entre el precio del petróleo y algunos metales preciosos. Se suelen denominar metales preciosos a aquellos metales que se encuentran en estado libre en la naturaleza, es decir, que no se encuentran combinados con otros elementos formando compuestos. El oro es un ejemplo de metal precioso, y se ha considerado uno de los más importantes durante muchos siglos dado que juega un papel especial como reserva de valor en tiempos de incertidumbre política y económica. A continuación, se muestran las ideas principales en las que se basan estos artículos, junto con las conclusiones que se han obtenido.

En el informe de (Zhang & Wei, 2010) se dice que el mercado del oro y del petróleo son dos de los principales representantes de los mercados de materias primas. Es por ello que advierte de la crucial importancia de analizar su relación de cointegración y causalidad, así como investigar su respectiva contribución con el objetivo de interpretar la dinámica del mercado de las materias y primas y prever la fluctuación en los precios.

Es importante destacar que la cointegración es una característica estadística de las variables en las series temporales las cuales indican que se comparte una tendencia estocástica común, es decir, comparten un cambio aleatorio de la serie a lo largo del tiempo. La cointegración se ha convertido en una propiedad importante en el análisis contemporáneo de las series temporales.

El análisis empírico de dicho estudio (Zhang & Wei, 2010) indica que, en primer lugar, existen tendencias coherentes entre el precio del petróleo y el precio del oro, presentando un coeficiente de correlación positivo y significativo del 0.9295 durante el período de muestreo, desde enero del año 2000 hasta marzo del 2008. Esta alta correlación positiva indica que hubo una tendencia similar de precios a lo largo de los años de estudio, tal y como se puede observar en la siguiente Figura 4.

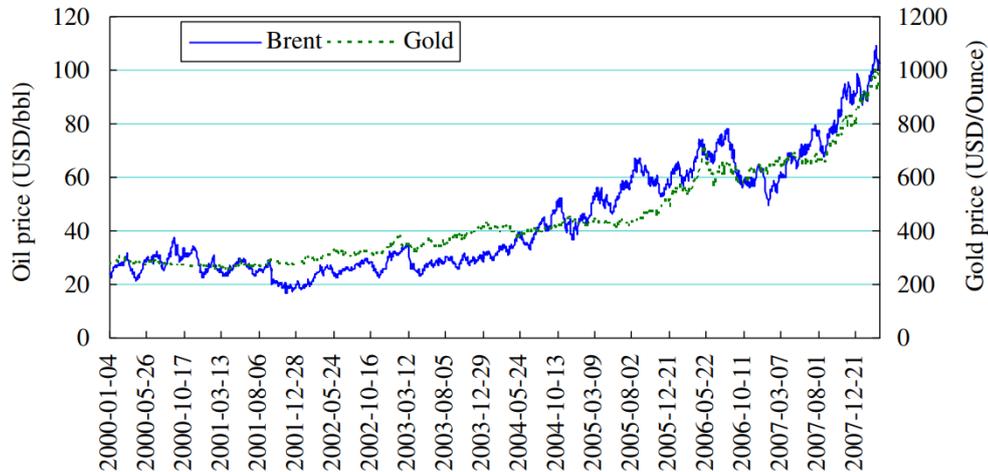


Figura 4. Alta correlación positiva entre las series temporales de precios del oro y del petróleo (2000-2007),  
(Zhang & Wei, 2010)

En segundo lugar, en el estudio se encuentra una relación de cointegración que indica que ninguna de las dos series temporales son estacionarias, y que existe un equilibrio a largo plazo entre los dos mercados, lo que significa que ambos mercados tienden a ser influenciados por factores externos comunes, como el tipo de cambio del US dólar, eventos geopolíticos y diversos factores económicos. Además, se encuentra una relación unilateral y lineal de causalidad de Granger, ya que se establece que el precio del petróleo causa el movimiento del precio del oro, pero no a la inversa. Sin embargo, no presentan ninguna causalidad no lineal, por lo que se sugiere que la interacción entre ambos activos es directa.

Por ejemplo, en el año 2002, debido a una depreciación del dólar, la inflación mundial, la manipulación de la oferta del petróleo por parte de la OPEP y algunos acontecimientos geopolíticos, tanto los precios del crudo como del oro entraron en una época de auge, y empezaron a dispararse prácticamente al mismo tiempo, manteniendo un impulso ascendente casi constante hasta el primer semestre de 2008.

Sin embargo, en la segunda mitad del 2008, por la grave crisis financiera y económica mundial, los grandes mercados de materias primas se enfrentaron a una prueba de austeridad. Como resultado, el precio del crudo descendió de 147\$ por barril a casi 30\$ a finales de 2008. De la misma manera, el oro también experimentó un fuerte colapso en el precio,

disminuyendo de 1000\$ por onza a 700\$, aproximadamente. Fue en el año 2009, con la expectativa de recuperación económica mundial, cuando la demanda del mercado de las materias primas comenzó a subir de nuevo y ambos activos comenzaron a subir de nuevo.

Este es un claro ejemplo de la relación directa que hubo entre el petróleo y el oro en los años del estudio y su clara influencia marcada por la situación económica y política del momento.

La última conclusión a la que se llega en este artículo está relacionada con el precio efectivo común entre los dos mercados. Se concluye que la contribución del precio del crudo parece mayor que la del precio del oro, ya sea con el modelo transitorio permanente (PT) o con el modelo de *information share* (IS). Esto implica que la influencia del crudo en el desarrollo económico mundial es mucho mayor, y su papel en los grandes mercados de materias primas ha sido de gran relevancia en los últimos años.

El segundo artículo que se va a discutir es de (Kanjilal & Ghosh, 2017), el cual estudia la relación dinámica del precio mundial del petróleo crudo y del oro en un modelo de corrección de errores vectorial de dos regímenes con un único vector de cointegración y un efecto de umbral en los términos de corrección de errores. Este análisis se ha realizado con los precios semanales del crudo y del oro en el periodo comprendido entre el 2 de enero del 2009 y el 20 de noviembre del 2015, es decir, después de la crisis financiera del año 2008. En la Figura 5 se muestra la evolución de ambas series temporales:

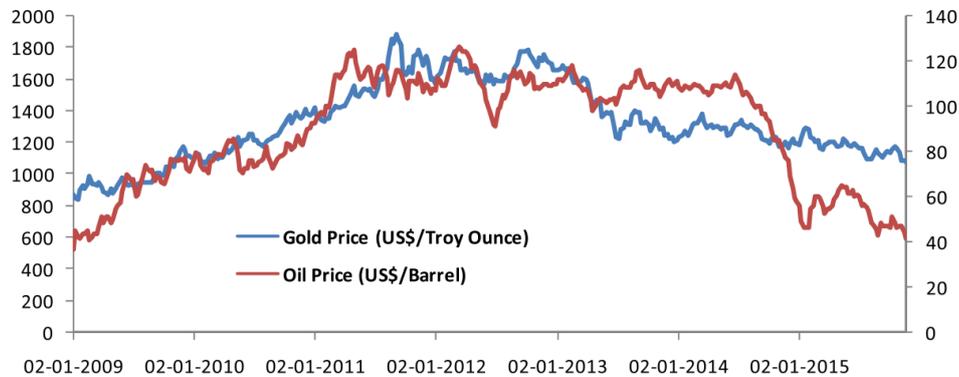


Figura 5. Evolución de las series temporales de precios del oro y del petróleo (2009-2015),  
(Kanjilal & Ghosh, 2017)

Los resultados empíricos del estudio sugieren la eficacia del modelo de cointegración umbral sobre un modelo de cointegración lineal, lo que significa la no linealidad en la relación a largo plazo entre el precio del oro y el del petróleo.

Este estudio también establece diferentes dinámicas a largo y corto plazo del oro y del petróleo en dos regímenes denominados “típico” y “extremo”, basados en el parámetro de umbral. Se revela una relación de retraso entre el oro y el precio del petróleo que sugiere que el oro es un refugio para inversores contra la inflación. Este fenómeno, sin embargo, solo ocurre en un régimen típico. Por el contrario, en un régimen extremo, en el que el precio del oro domina el mercado, los inversores cambian entre el oro y el petróleo para diversificar el riesgo de su cartera. Todo esto indica que la relación entre el precio del oro y del petróleo es no lineal y asimétrica.

Este estudio supera los realizados anteriormente al establecer que la relación entre el oro y el petróleo depende del régimen y, por tanto, no se mantiene constante durante el periodo de estudio. Este análisis demuestra solidez ya que los resultados empíricos mantienen una profunda congruencia estadística.

En tercer lugar, el estudio realizado por (Troster, 2018) propone un test paramétrico consistente de la causalidad de Granger en cuartiles. Aunque el concepto de causalidad de Granger se denota en términos de la distribución condicional, la mayoría de los artículos han

probado esta causalidad utilizando modelos de regresión de media condicional en los que las relaciones causales son lineales.

El autor concluye en este estudio que muchos análisis políticos y financieros importantes se investigan mediante tests de causalidad de Granger entre series temporales económicas. En este artículo se estudian las posibles relaciones causales en todos los cuartiles, en lugar de centrarse únicamente en una sola parte de la distribución condicional. Se realizan pruebas de Monte Carlo y se aplican a relaciones causales entre los precios del oro, el tipo de cambio USD/GBP y el precio del petróleo.

Por último, el estudio de (Balcilar et al., 2019) analiza los vínculos dinámicos entre los precios del petróleo y del oro en los mercados spot y de futuros de 1 a 12 meses utilizando datos mensuales durante el periodo 1983-2016. Para ello se utiliza la causalidad de Granger de manera recursiva. La característica fundamental que distingue a este estudio de los anteriores es que este estudio se trata del primero que investiga los vínculos causales entre el petróleo y el oro utilizando pruebas de causalidad variables en el tiempo.

El petróleo y el oro son dos activos que son frecuentemente negociables y tienen una alta liquidez y sincronización en sus movimientos. Son dos materias primas con un papel insustituible en la economía. El petróleo es la mercancía más comercializada del mundo y tiene un precio muy volátil. El oro, en cambio, con la menor volatilidad, es el líder de todos los metales preciosos. Teniendo en cuenta los riesgos bursátiles, el oro es conocido como un refugio seguro y se utiliza como un instrumento eficaz de cobertura. Un análisis a largo plazo del comportamiento del petróleo y del oro demuestra que ambas materias primas tienden históricamente a moverse simultáneamente al alza y a la baja, y un cambio en el precio de una parece inducir un cambio en el precio de la otra. La observación de las dos materias primas durante los últimos 50 años, por ejemplo, ha demostrado que tienden a comportarse de manera similar con respecto al precio, con una correlación positiva del 80%.

En este artículo se concluye, en primer lugar, que no existe poder predictivo entre los precios del petróleo y del oro. En general, los resultados de este estudio muestran que las relaciones dinámicas entre las series de precios del petróleo y del oro serán sensibles a la frecuencia de

las series temporales, al lapso del tiempo y al método utilizado. Por lo tanto, la sugerencia de este informe es que es muy importante que se lleven a cabo más estudios sobre las relaciones causales entre los precios del petróleo y del oro, así como otros casos, que permitan cambios de política y cambios significativos en el petróleo y del oro y períodos volátiles causados por recesiones y crisis financieras.

El petróleo, como materia prima, se ha ido introduciendo en el mundo financiero en las últimas tres décadas. Al igual que su papel como instrumento de cobertura contra los riesgos del mercado bursátil, el oro podría ser considerado para cubrir los riesgos del precio del petróleo. Los resultados de este artículo indican que asumir la no causalidad entre el petróleo y el oro podría llevar a peligros en el uso del oro como instrumento de cobertura contra el riesgo del precio del petróleo.

## **Capítulo 3. ANÁLISIS DE DATOS VALORADOS POR INTERVALOS**

El Análisis de Intervalos (AI) tiene sus orígenes en el mundo clásico hace 2.200 años, ante la propuesta de un método por parte de Arquímedes para proporcionar una sucesión de intervalos que contuviesen el número  $\pi$ . Pero tal y como se explica en (Jiménez, El análisis de intervalos: aplicaciones en la ingeniería, 2012), no fue hasta la década de 1960 cuando Ramón Moore empezó a investigar sobre ello, dando importancia a los métodos y aplicaciones con datos de intervalo, centrándose inicialmente en los resultados computacionales surgidos con los datos de intervalo.

Desde entonces, el análisis de intervalos ha adquirido mucha relevancia y hoy en día su estudio es de gran interés, produciéndose de esta manera muchos avances en la década actual. Aunque este tipo de datos nos hayan acompañado durante mucho tiempo y estén presentes en nuestro día a día, su análisis todavía está en desarrollo y quedan muchos aspectos por profundizar.

Los datos numéricos únicos han sido el paradigma de la estadística, el Machine Learning y la Inteligencia Artificial hasta ahora, con sus respectivas herramientas y softwares para el análisis de datos simples. Sin embargo, en muchas ocasiones no se puede proporcionar la información completa con un simple número, y es por eso por lo que surge la necesidad de trabajar con datos de intervalo. Estos datos, aunque también presentan sus inconvenientes y complejidad, aportan una información mucho más amplia que lo que se puede llegar a explicar con un solo número.

En este sentido, los datos de intervalo son una medida que se relaciona directamente con el concepto de precisión. Son muy comúnmente utilizados en aplicaciones prácticas de la ciencia ya que definen los límites de los errores cometidos en las mediciones con el instrumento correspondiente. Como bien sabemos, los aparatos de medición no presentan una precisión absoluta, y es por ello por lo que debemos contemplar un margen de error para no perder información y no afectar en la consecuente toma de decisiones. Supongamos el

caso en el que nos encontramos en un hospital y utilizamos un termómetro para tomar la temperatura de un paciente y así obtener información relevante para hacer un diagnóstico acertado. Cuando obtenemos la medición  $\bar{x}$  debemos tener en cuenta que este valor será generalmente diferente al valor real y desconocido  $x$ , luego debemos procesar los datos considerando el error cometido:  $\Delta x = \bar{x} - x$ , contemplando de esta manera la incertidumbre asociada.

Pongamos otro ejemplo relativo a las mediciones que se realizan con frecuencia en un laboratorio químico. Las conclusiones que vamos a obtener se podrían extrapolar a cualquier campo de la ciencia.

Todas las medidas experimentales vienen afectadas de una imprecisión inherente al proceso de medida, y es inevitable que tengamos un error intrínseco en la medida que puede provenir de diferentes causas. La primera de ellas se refiere al error instrumental del aparato de medida, en relación con su sensibilidad. Además, hay que tener en cuenta el error personal, esto es, las limitaciones de carácter personal como por ejemplo los errores de paralaje o los problemas de tipo visual. Por último, este error intrínseco se debe también a los errores del método de medida, que corresponden a una elección inadecuada del aparato de medida o del observador. Aunque intentemos ser lo más exactos posibles, es necesario que siempre contemos con un error absoluto (diferencia entre la medida y el valor real de la magnitud), por muy pequeño que sea.

En un laboratorio se toman diariamente cientos de medidas: masa, volumen, densidad, temperatura... y la Tabla 1 pretende mostrar las diferencias entre los datos numéricos observados y el dato de intervalo que se podría considerar si las medidas se realizan con un 5% de error total.

Instrumento laboratorio	Dato numérico (medido)	Dato de intervalo
<b>Probeta</b> (volúmenes de pequeños líquidos)	25.4 ml	$25.4 \pm 1.27 \text{ ml} = [24.13, 26.27] \text{ ml}$
<b>Densímetro</b> (densidad de los líquidos)	11.32 g/ml	$11.32 \pm 0.56 \text{ g/ml} = [10.76, 11.88] \text{ g/ml}$
<b>Balanza</b> (masa de un cuerpo)	120 g	$120 \pm 6 \text{ g} = [114, 126] \text{ g}$
<b>Termómetro</b> (temperatura de un cuerpo)	4.5°C	$4.5 \pm 0.22 \text{ °C} = [4.28, 4.72] \text{ °C}$

Tabla 1. Datos de intervalo en la medición de resultados de laboratorio

Tradicionalmente en el manejo de errores se ha supuesto que la distribución de la probabilidad de error de una medida sigue una distribución gaussiana y se podrían aplicar por tanto las técnicas estadísticas estándares para tratar dicha incertidumbre. Pero en la práctica, no siempre podemos asegurar que esa suposición sea real, y lo único que podemos afirmar es que existe una cota de incertidumbre absoluta ( $|\Delta x| \leq \Delta$ ) asociada a la medida realizada.

## CARACTERÍSTICAS Y NOTACIÓN DE LOS DATOS DE INTERVALO

Los datos de intervalo pueden definirse de diferentes maneras, y una de ellas es con su límite inferior y superior. Considerando el error de medición, únicamente podemos asegurar que el valor real  $x$  estará comprendido entre un límite inferior (*lower bound*)  $x^L = \bar{x} - \Delta$  y un límite superior (*upper bound*)  $x^U = \bar{x} + \Delta$ . Existen numerosas formas de notar los intervalos, que difieren según autores, pero para evitar el mayor número de controversias vamos a denotarlos de la siguiente manera:  $[x] = [x^L, x^U]$ , con corchetes y letras minúsculas.

Otra representación del intervalo  $[x]$  considera su centro  $x^C$  y su radio  $x^R$ . Es decir,  $[x] = \{x^C, x^R\}$  donde  $x_c = (x^L + x^U)/2$  y  $x^R = (x^U - x^L)/2$ , siendo siempre  $x^R > 0$ . Ambas representaciones

para intervalos reales son las más comunes, aunque se podría utilizar cualquier otro par de puntos mientras se puedan calcular los extremos del intervalo.

### ***SERIES TEMPORALES DE INTERVALO (ITS)***

El área más investigada de los datos de intervalo son probablemente las series temporales por sus aplicaciones directas en diversos campos: negocios, finanzas, medicina, turismo, energía... entre otros. Los índices de precios, los valores de la bolsa o la fluctuación de la moneda se pueden representar como series temporales para analizar posteriormente métodos de previsión. Hoy en día observamos un mundo mucho más incierto y complejo que hace unos años, por lo que la precisión de los pronósticos es algo crucial, siendo un requisito creciente el de proporcionar otro tipo de pronósticos más allá de los puntuales.

Existen gran cantidad de métodos de previsión. En todos ellos se utilizan los valores pasados de una variable con el objetivo de predecir los valores futuros. Así, la entrada y la salida de un método se puede modelar como una serie temporal, o lo que es lo mismo, una secuencia de observaciones de la misma variable en períodos de tiempo uniformes. Las previsiones puntuales son utilizadas en numerosas ocasiones y pueden resultar de gran eficacia, sin embargo, se ha demostrado que los pronósticos de intervalo obtienen una información mucho más relevante, mejorando la precisión de las anteriores. Los datos con valores de intervalo son un caso particular de datos simbólicos, tal y como se consideran en el campo de análisis simbólico de datos (ASD). Las variables simbólicas, como pueden ser las listas, intervalos o distribuciones de frecuencia, son más adecuadas que las variables únicas para describir situaciones complejas de la vida real.

Una serie temporal de intervalo asigna a cada período de tiempo un intervalo que abarca los valores tomados por la variable observada. Los intervalos vienen definidos por cuatro características: límite inferior, límite superior, centro y radio. Un ejemplo de serie temporal de intervalo en el campo de las finanzas podría ser la evolución del precio de las acciones de una compañía en un determinado mes, contemplando sus mínimos y máximos.

El análisis y previsión de las series temporales de intervalos ha adquirido su foco de investigación hace pocos años, por lo que todavía hay muchas cuestiones por dar forma y resolver. Por el contrario, la combinación de pronósticos es un campo de investigación más maduro en el tiempo. A modo de resumen, tal y como se explica en (Maté, 2021) hay dos consensos en relación con las previsiones de STI, “combinar pronósticos reduce el error del pronóstico final” y “un promedio simple de varios pronósticos a menudo supera complicados esquemas de ponderación”, que se denominó “rompecabezas de combinación de pronósticos (FCP)”.

### 3.1.1 CONCEPTOS ESTADÍSTICOS DE LAS ITS

A continuación, se van a definir algunos de los conceptos esenciales de las series temporales de intervalos, considerando cuestiones estocásticas. Para ello se van a definir los conceptos importantes de variable aleatoria de intervalo y proceso estocástico de intervalo, y en consecuencia se definirá el concepto de STI.

Sea  $(\Omega, F, P)$  un espacio de probabilidad donde  $\Omega$  es el conjunto de sucesos elementales,  $F$  es el  $\Omega$ -campo de sucesos y  $P : F \rightarrow [0, 1]$  la medida de probabilidad de  $\Omega$ -aditiva. Definimos una partición de  $\Omega$  en conjuntos  $A_X(x)$  tal que  $A_X(x) = \{\omega \in \Omega | X(\omega) = x\}$ , donde  $x \in [x^L, x^U]$ .

- **Variable aleatoria de intervalo:** Un mapeo  $[X] : F \rightarrow K_C(\mathbb{R}) \subset \mathbb{R}$  tal que para cada  $x \in [x^L, x^U]$ , existe un conjunto  $A_X(x) \in F$  que se denomina variable aleatoria de intervalo (iRV). Por ejemplo, las temperaturas máximas y mínimas de un lugar cada día en grados Celsius.
- **Proceso estocástico con valores de intervalo:** es una colección de valores aleatorias de intervalo indexadas por el tiempo, es decir,  $\{[X_t]\}$  for  $t \in T \subset \mathbb{R}$  con cada  $X_t$  siguiendo la definición anterior. Por ejemplo, la evolución de las temperaturas mínimas y máximas es un proceso estocástico con valores de intervalo.
- **Series temporales con valores de intervalo (STI):** es una realización de un proceso estocástico con valores de intervalo. Se puede denotar como  $\{[x_t]\} = [x^L, x^U] = [x^C, x^R]$  para  $t=1,2,\dots,T$ .

- **Paseo aleatorio con valor de intervalo:** una serie temporal de intervalo  $[Y_t]$  es un paseo aleatorio valorado por intervalos (iRW) con deriva si:

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

y como consecuencia,

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

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

Es decir, bajo una iRW con deriva, la variable con valor de intervalo en el tiempo  $t$  es el resultado de un intervalo constante más la variable con valor de intervalo en tiempo  $t-1$  más un error en el tiempo  $t$  en forma de intervalo. El caso de  $[\mu] = [0]$  es el iRW o iRW sin deriva.

- **Previsión ingenua con valor de intervalo:** el pronóstico ingenuo de valor de intervalo para una serie temporal de intervalo, para  $t=2, 3, \dots, T$ ; viene dado por:

$$[\hat{Y}_t] = [Y_{t-1}]$$

y como consecuencia,

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

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

A veces, la previsión ingenua con valor de intervalo se denomina método de previsión iRW.

Una vez explicados los aspectos relacionados con cuestiones estocásticas, vamos a definir la medida de distancia entre la previsión obtenida y el valor real, con el objetivo de poder evaluar la previsión de cualquier pronóstico. En las series temporales clásicas, la distancia es simplemente el valor absoluto (o al cuadrado) de la diferencia entre la observación y la previsión, pero necesitamos también una distancia que se pueda aplicar a intervalos.

Limitándonos a la distancia euclídea obtenemos que:

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

Cada distancia se expresa en términos de los límites inferior y superior, y en términos de centros y radios.

## ***VISUALIZACIÓN***

Los datos únicos presentan numerosas visualizaciones estándar que se usan muy a menudo y son fácilmente representables y analizables por todos. En cambio, las visualizaciones con datos de intervalo son más complejas y es un área que todavía está abierta a muchas mejoras. Los intervalos necesitan muchos puntos para ser representados (como mínimo dos) y se convierte por tanto en una tarea más difícil, especialmente si queremos incluir muchas variables o un gráfico más analítico. Aun así, se pueden crear algunas representaciones de intervalos como diagramas de dispersión o de cajas.

Los tipos de visualizaciones los podemos dividir en dos (Maceda, 2018): de variables transversales, de series temporales de intervalo. Vamos a describir brevemente cada uno de ellos:

### **3.1.2 VISUALIZACIÓN DE INTERVALOS TRANSVERSALES**

Los intervalos transversales incluyen datos de diferentes entidades (regiones, personas, empresas...) en un mismo momento. Estos datos no tienen un orden en sí mismos, y por tanto no hay ninguna exigencia en el orden, sino que más bien se pueden reordenar de la manera que más convenga proporcionando cierta libertad.

Este tipo de visualizaciones es diferente en función del número de variables que se desean representar. Podemos representar todos los intervalos de una sola variable, creando un gráfico bastante intuitivo, o podemos incluir intervalos de varias variables, lo que lo hace más complejo porque muchos datos tienen que ser explicados.

En general las visualizaciones más comunes son las que incluyen una única variable, tanto en los gráficos transversales como en los de series temporales de intervalos, pero cabe destacar que si el número de datos a representar es excesivamente grande podríamos tener un pequeño problema ya que habría demasiadas líneas y no se distinguirían entre ellas, haciendo difícil su interpretación. En las siguientes figuras (Figura 6 y Figura 7) se muestra ejemplos de visualizaciones con una o varias variables.

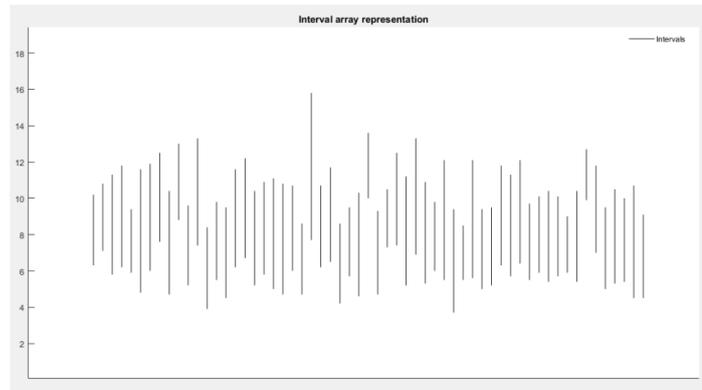


Figura 6. Representación gráfica de los intervalos de una variable (Maceda, 2018)

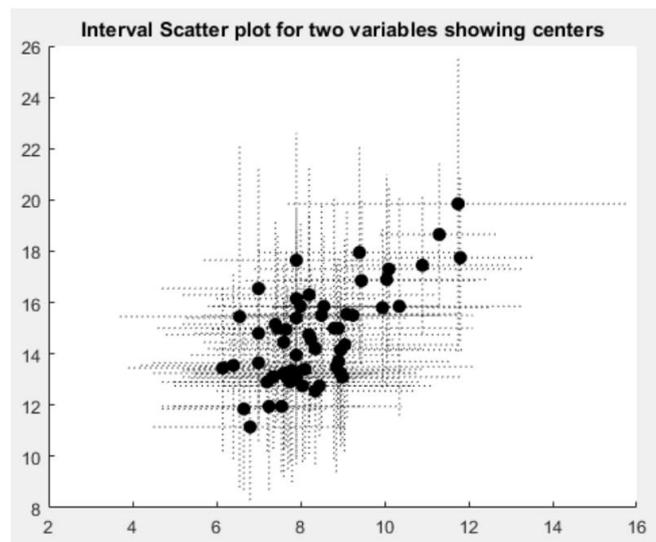
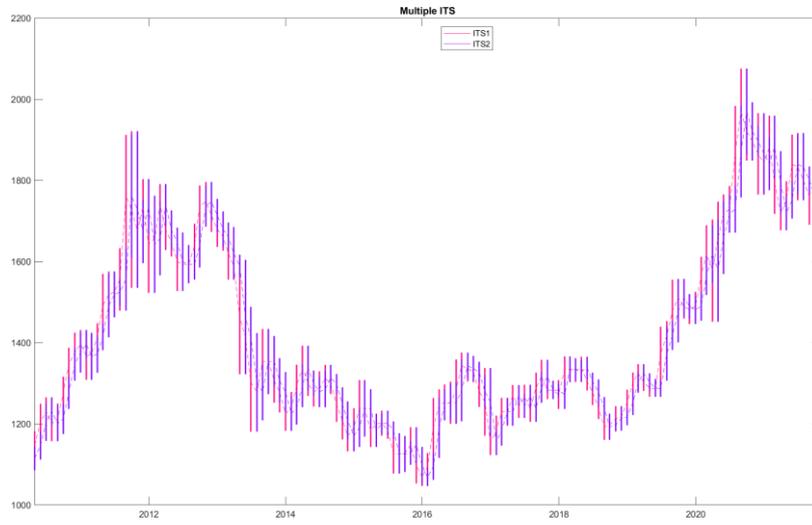


Figura 7. Representación gráfica de los intervalos de dos variables (Maceda, 2018)

### 3.1.3 VISUALIZACIÓN DE SERIES TEMPORALES DE INTERVALOS

Las series temporales de intervalos son los datos de intervalo más utilizados, por lo que no resulta sorprendente que también lo sean sus visualizaciones. Hay representaciones para una única serie temporal de intervalo o para varias. A continuación, se muestra un ejemplo gráfico (Figura 8) de dos series temporales.



*Figura 8. Representación gráfica de dos ITS*

## **Capítulo 4. REGRESIÓN LINEAL SIMPLE Y MÚLTIPLE PARA DATOS DE INTERVALO**

Los métodos estadísticos y de minería de datos se han desarrollado principalmente en casos en los que las variables toman un único valor. En la vida real, hay muchas ocasiones en las que el uso de este tipo de variables puede suponer una pérdida importante de información o dar lugar a cálculos que requieren mucho tiempo.

En el caso de las variables cuantitativas, se puede conseguir una descripción más completa en términos de datos de intervalo, es decir, el valor que toma una variable es un intervalo cerrado en los números reales. Es especialmente útil cuando es conveniente resumir grandes conjuntos de datos de tal manera que el resumen resultante tenga un tamaño manejable y mantenga toda la información posible de los datos originales.

Por ejemplo (Jose Emmanuel Chacón and Oldemar Rodríguez, 2021), supongamos que queremos sustituir la información de todas las transacciones realizadas por el propietario de una tarjeta de crédito por una única “transacción” que resuma todas las transacciones originales. Esto se consigue gracias a una nueva transacción que tendrá en sus campos no sólo números sino también intervalos definidos por el mínimo y el máximo de compra. En este sentido no se pierde nada de información.

La regresión lineal es un modelo matemático usado para aproximar la relación de dependencia entre una variable dependiente “ $y$ ” (variable respuesta) y una o varias variables independientes “ $x_i$ ” (variables explicativas). Se denomina regresión lineal simple cuando hay una única variable explicativa mientras que la regresión lineal múltiple hace referencia a varias variables independientes. La regresión lineal se puede aplicar en numerosas situaciones por muy sencillas que sean, como por ejemplo para estimar el precio de la vivienda en función de su superficie en metros cuadrados y su luminosidad, o bien para aproximar la calificación obtenida en una asignatura según el número de horas de estudio semanal.

La idea de la regresión es que, dado un vector de entrada  $x^t = (x_1, x_2, x_3, \dots, x_p)$ , donde  $p$  es el número de variables y  $x^t$  representa la transposición de  $x$ , queremos predecir la variable de respuesta mediante el siguiente modelo lineal:

$$\hat{y} = \hat{\beta}_0 + \sum_{j=1}^p x_j \hat{\beta}_j$$

donde  $\beta_0$  es el intercepto. Si se incluye la constante uno en el vector  $x$  y  $\beta_0$  se incluye en el vector de coeficientes  $\beta$ , entonces el modelo lineal se puede escribir como un producto vectorial de la siguiente forma:

$$\hat{y} = x^t \beta$$

Para ajustar el modelo lineal con los datos de entrenamiento, el método de estimación más popular es el método de los mínimos cuadrados. Se eligen los coeficientes  $\beta$  para minimizar la suma de los cuadrados de los residuos.

$$S = \sum_{i=1}^n (y_i - x_i^t \beta)^2$$

donde  $n$  es el número de observaciones del conjunto de datos. La función de la suma de los cuadrados de los residuos  $S$  es una función cuadrática, esto es, el mínimo siempre existe. Se puede escribir también de la siguiente manera:

$$S = (y - X\beta)^t (y - X\beta)$$

donde  $X$  es una matriz  $n \times p$  donde cada fila es un vector del conjunto de entrenamiento, y  $y$  es un vector de tamaño  $n$  (el vector de salida en el conjunto de entrenamiento). Se conoce que si  $X^t X$  es una matriz no singular, la solución entonces viene dada por:

$$\hat{\beta} = (X^t X)^{-1} X^t y$$

El valor aproximado de este modelo para la componente  $x_i$  se puede aproximar como  $\hat{y} = x_i^t \hat{\beta}$  y los valores ajustados para el nuevo caso  $x^t = (1, x_1, x_2, x_3, \dots, x_p)$ , vienen dados por  $\hat{y} = x^t \hat{\beta}$ .

En la práctica existen varios métodos para encontrar los coeficientes  $\beta$ , pero los métodos existentes, y en particular el método de los mínimos cuadrados, requieren mucho trabajo y tiempo para grandes conjuntos de datos, y otros no son lo suficientemente precisos con este tipo de conjunto de datos. Es por ello que, se han creado nuevos algoritmos no iterativos para identificar los coeficientes de regresión múltiple cuando se procesan grandes volúmenes de datos. Estos algoritmos resultan de gran eficiencia en cuanto a la precisión y velocidad, en comparación con los métodos existentes. Aun así, los modelos posteriores se van a explicar teniendo en cuenta el método de optimización tradicional de los mínimos cuadrados.

Dicho esto, la regresión lineal es una técnica muy comúnmente utilizada para datos únicos y durante el siglo XX para la predicción y muchos autores han profundizado sobre los métodos de regresión, escribiendo artículos, libros y manuales. Un ejemplo es el libro (Rencher, 1995).

## ***PARADIGMAS***

Sin embargo, no ocurre lo mismo para datos de intervalo, y todavía no se ha estudiado con detalle la aplicación de algo tan básico como la regresión lineal. La investigación de los métodos para regresiones con intervalos se ha empezado a desarrollar recientemente, aproximadamente a partir del año 2000. Para poder tratar la regresión de intervalos es importante que diferenciamos cuatro paradigmas:

### **4.1.1 PARADIGMA 1**

$\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n = y$ , donde las variables independientes  $x_n$  y la variable independiente  $y$  corresponden a datos únicos.

### **4.1.2 PARADIGMA 2**

$\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n = [Y]$ , donde las variables independientes  $x_n$  son variables de datos únicos mientras que la variable dependiente  $[Y]$  es una variable de intervalo.

### 4.1.3 PARADIGMA 3

$\beta_0 + \beta_1[X_1] + \dots + \beta_n[X_n] = y$ , donde las variables independientes  $[X_n]$  son variables de datos de intervalo mientras que la variable dependiente  $y$  es una variable de dato único (no de intervalo).

### 4.1.4 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  $[Y]$  corresponden a variables con datos de intervalos.

Los modelos de regresión que vamos a explicar a continuación han sido creados para trabajar con el paradigma número 4 (excepto el modelo MG), y aunque puede ser que también funcionen para otros paradigmas, no han sido creados para ellos y se deberían utilizar diferentes modelos de regresión.

## *MÉTODOS DE REGRESIÓN LINEAL PARA INTERVALOS*

### 4.1.5 MÉTODO DE CENTRO (CM-CENTRE METHOD)

Este método fue introducido por Billard y Diday en el año 2002 (Billard, L and Diday, E., 2000). Se obtiene un modelo de regresión a partir de los centros de los intervalos y aplica dicho modelo a los límites superior e inferior para obtener la variable de salida  $[\hat{Y}]$ .

Considerando:

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

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

Por tanto, la suma de los cuadrados de los errores de este modelo se puede definir como:

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

donde  $\varepsilon_{li}$  y  $\varepsilon_{ui}$  son (despejando de las ecuaciones anteriores):

$$\varepsilon_{li} = y_{li} - \beta_0 - \beta_1 x_{l1} - \dots - \beta_n x_{ln}$$

$$\varepsilon_{ui} = y_{ui} - \beta_0 - \beta_1 x_{u1} - \dots - \beta_n x_{un}$$

Con una regresión lineal de los centros de  $[Y]$  y  $[X_1, X_2, \dots, X_n]$ , es posible obtener los valores  $\beta_0, \beta_1, \dots, \beta_n$  que minimicen la suma de los cuadrados de los residuos  $S$ .

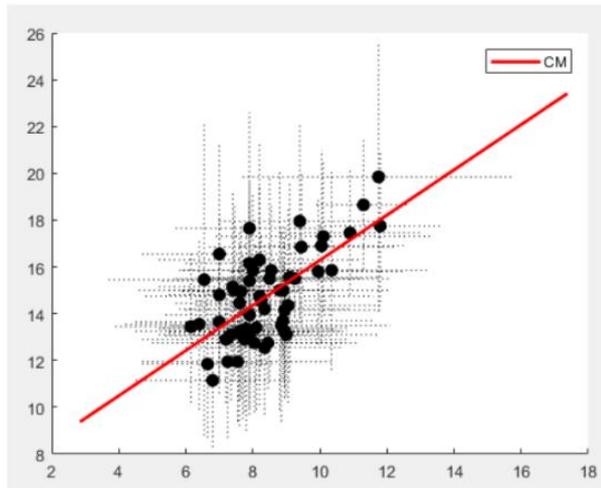
El resultado del valor esperado de  $\beta$ :  $\beta = (\beta_1, \beta_2, \dots, \beta_n)$  predecirá la variable de salida del método de regresión  $\hat{Y}_{CM}$  de la siguiente manera:

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

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

En la Figura 9 mostramos un ejemplo simple de una regresión lineal de intervalo. La línea roja indica la línea de ajuste de la regresión realizada para los centros de  $[Y]$  y de  $[X]$ . Este ajuste se utiliza para cada uno de los límites (superior e inferior) para obtener  $\hat{Y}_{CM}$ .

Este modelo es el primero de todos en desarrollarse y por eso es un poco rudimentario, no tiene en cuenta los rangos de los intervalos sino únicamente los centros de estos. Es por ello por lo que las predicciones de variables que utilizan este método no son del todo precisas y es mejor que se utilicen otros métodos más elaborados, en el caso de que sea posible. Este método sentó las bases y la idea de otros métodos que vinieron después con un enfoque similar.



*Figura 9. Ejemplo de Método del Centro (Maceda, 2018)*

#### 4.1.6 MÉTODO DE LOS MÍNIMOS Y MÁXIMOS (MM-MINMAX METHOD)

El método MinMax fue introducido por Billard y Dilay en el año 2002 (Billard, L and Diday, E., 2000) como una alternativa de mejora al método del centro que fue desarrollado por los mismos autores unos años antes. Este método utiliza diferentes parámetros para obtener los límites inferiores y superiores de  $[\hat{Y}]$ . Esto es lo mismo que suponer la independencia entre los valores de los límites superior e inferior y tratarlos como si se tratasen de variables diferentes. En este caso la regresión lineal funciona de la siguiente manera:

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

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

Ahora escribimos la suma de los cuadrados de los residuos, que tendrá la siguiente ecuación:

$$S_{\text{MinMax}} = \sum_{i=1}^n (\varepsilon_{li})^2 + \sum_{i=1}^n (\varepsilon_{ui})^2$$

Donde los errores de los límites inferior y superior son (despejando de las ecuaciones anteriores):

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

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

Minimizando, y teniendo en cuenta los límites superiores  $[Y_u]$  y las variables independientes  $[X_{u1}, X_{u2}, \dots, X_{un}]$  de una manera separada que los límites inferiores  $[Y_l]$  y  $[X_{l1}, X_{l2}, \dots, X_{ln}]$  por dos regresiones lineales independientes, podemos obtener los valores  $\beta_0^l, \beta_1^l, \dots, \beta_n^l$  y  $\beta_0^u, \beta_1^u, \dots, \beta_n^u$  que minimizan la suma de los cuadrados de los residuos.

Los valores resultantes  $\hat{\beta}^l = (\hat{\beta}_0^l, \hat{\beta}_1^l, \dots, \hat{\beta}_n^l)$  y  $\hat{\beta}^u = (\hat{\beta}_0^u, \hat{\beta}_1^u, \dots, \hat{\beta}_n^u)$  predecirán  $\hat{Y}_{\text{MinMax}}$  mediante la siguiente ecuación:

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

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

En la Figura 10 se muestra un ejemplo simple de una regresión lineal de intervalo utilizando el método MinMax que acabamos de definir. Ambas líneas verdes indican el ajuste de dos diferentes regresiones llevadas a cabo por los límites superiores e inferiores de  $[X]$  para obtener  $[\hat{Y}_L]$  y  $[\hat{Y}_U]$ .

Este modelo es una clara mejora del método del centro, ya que tiene en cuenta las singularidades que tienen los datos con valor de intervalo, al considerar los límites superior e inferior de forma independiente, haciendo por tanto una estimación más fiable para la predicción de la variable de salida  $[\hat{Y}]$ . Este modelo fue el primer modelo por intervalos que realmente proporcionó una estimación decente con la que trabajar y poder obtener predicciones precisas.

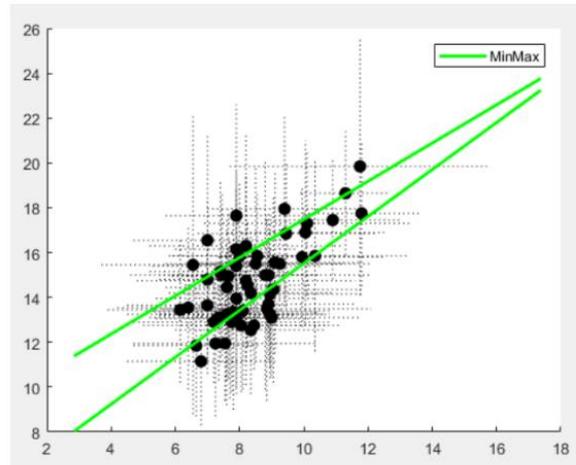


Figura 10. Ejemplo del Método MinMax (Maceda, 2018)

#### 4.1.7 MÉTODO DEL CENTRO Y RANGO (CRM-CENTRE AND RANGE METHOD)

El método del centro y del rango fue introducido por Lima Neto y De Carvalho en el año 2004 (Carvalho, Eufrásio de A. Lima Neto and Francisco de A.T. de, 2008), también como una alternativa de mejora al método del centro desarrollado por Billard y Diday. Este método toma la idea del método del centro y decide añadir los rangos (o radios) de los intervalos para mejorar el rendimiento de la predicción. Este método sigue la idea del método MinMax de añadir información complementaria a la del método del centro con el objetivo de satisfacer las necesidades de la complejidad de los datos de intervalo.

En este caso, el intervalo  $W$  no estará representado por sus límites inferior y superior  $[W_L, W_U]$  sino que se va a tener en cuenta el centro y el radio  $[W_c, W_r]$ , siendo el centro  $W_c = (W_L + W_U)/2$  y el radio  $W_r = (W_U - W_L)/2$ .

Las ecuaciones de la regresión lineal en este modelo son las siguientes:

La primera de ellas es relativa al centro y la segunda al radio.

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

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

Luego, la suma de los cuadrados de los residuos se escribe así:

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

donde los errores  $\varepsilon_{ci}$  y  $\varepsilon_{ri}$  son (despejando):

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

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

Minimizando y teniendo en cuenta por una parte los centros  $[Y_c]$  y las variables  $[X_{c1}, X_{c2}, \dots, X_{cn}]$  como en el método del centro, y de una manera separada el radio centros  $[Y_r]$  y las variables  $[X_{r1}, X_{r2}, \dots, X_{rn}]$ , por dos regresiones lineales independientes, podemos obtener 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 minimicen la suma de los cuadrados de los residuos,  $S_{CRM}$ .

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

$$\hat{y}_{ci} = \hat{\beta}_0^c + \hat{\beta}_1^c x_{i1} + \dots + \hat{\beta}_n^c x_{in}$$

$$\hat{y}_{ri} = \hat{\beta}_0^r + \hat{\beta}_1^r x_{i1} + \dots + \hat{\beta}_n^r x_{in}$$

En la Figura 11 se muestra un sencillo ejemplo de una regresión lineal con datos de intervalo haciendo uso del método del centro y rango. La línea de ajuste de este método no puede representarse en un único gráfico de dispersión, por lo que se muestran dos gráficos de dispersión, el de arriba con los centros y el de abajo con los radios. Ambos incorporan las líneas de ajuste en rojo utilizando el método CRM.

Este modelo emplea la misma idea que el método MinMax, aportando más información que el método de centros (CM) para poder añadir precisión al resultado. El rendimiento obtenido

es similar al del método MinMax, explicado anteriormente, pero obviamente este modelo es más preciso que el método de centros, por el simple hecho de que añade más información.

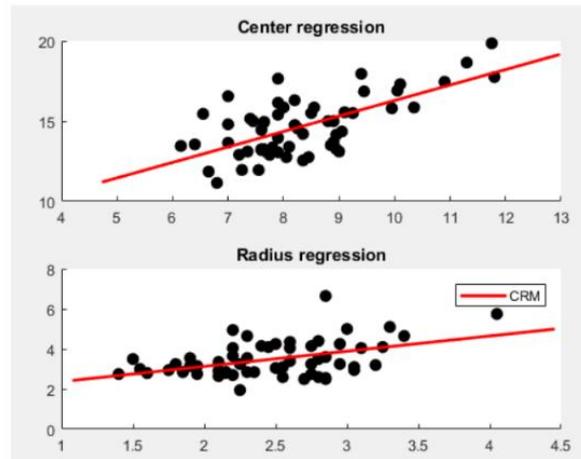


Figura 11. Ejemplo del Método del Centro y Rango (Maceda, 2018)

#### 4.1.8 MÉTODO DEL CENTRO Y RANGO RESTRINGIDO (CCRM-CONSTRAINED CENTRE AND RANGE METHOD)

Este método fue introducido por Lima Neto y De Carvalho en el año 2010 (Carvalho, Eufrásio de A. Lima Neto and Francisco de A.T. de, 2008) como una mejora al método de centros y rangos que el mismo pronunció unos años antes. Este método utiliza el mismo principio que el método de centros y radios de utilizar el centro y el radio de las variables  $[X_1, X_2, \dots, X_n]$  para predecir la variable de salida  $[\hat{Y}]$ . Sin embargo, este método añade la restricción de que el radio no puede ser negativo, algo que podía ocurrir en el método CRM y que matemáticamente no tiene ningún sentido.

La ecuación de la regresión lineal de CCRM es la misma que la del método CRM, pero la suma de los cuadrados de la desviación de este modelo viene determinada como:

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

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

Formalizando, y con una notación matricial, el resultado del problema de optimización viene determinado por:

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

$$\text{s.t. } \beta_i^r \geq 0$$

Los valores óptimos de  $\beta^c$  se pueden encontrar resolviendo un problema de regresión ordinario teniendo en cuenta que la segunda parte de la ecuación,  $\|Y_r - X_r\beta^r\|^2$ , no depende de  $\beta^c$ . Por tanto,  $\beta^c$  será el mismo que utilizando el método de centro y rango.

De una manera similar, la primera parte de la ecuación, esto es  $\|Y_c - X_c\beta^c\|^2$ , puede considerarse una constante para  $\beta^r$ . La estimación de  $\beta^r$  puede obtenerse resolviendo un problema de regresión restringido. Este problema puede reconocerse como un problema particular de mínimos cuadrados (NNLS). Por lo tanto, CCRM consiste en resolver dos problemas de regresión distintos.

Los valores resultantes de los parámetros a estimar presentan las siguientes ecuaciones:

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

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

En la Figura 12 que se muestra a continuación se muestra un ejemplo de una regresión lineal que utiliza el método del centro y el rango restringido (CCRM). La línea azul muestra CCRM mientras que la línea roja muestra CRM. Como es esperado, la línea de ajuste del centro es la misma para ambos métodos (CRM y CCRM), pero la línea de ajuste del radio en cambio es ligeramente distinta por la restricción que se incorpora.

Este modelo es un paso adelante respecto a CRM, ya que no sólo proporciona una buena predicción, sino que también tiene en cuenta casos en los que CRM o incluso MinMax crearían un resultado ilógico, como intervalos con radios negativos o límites inferiores mayores que los superiores.

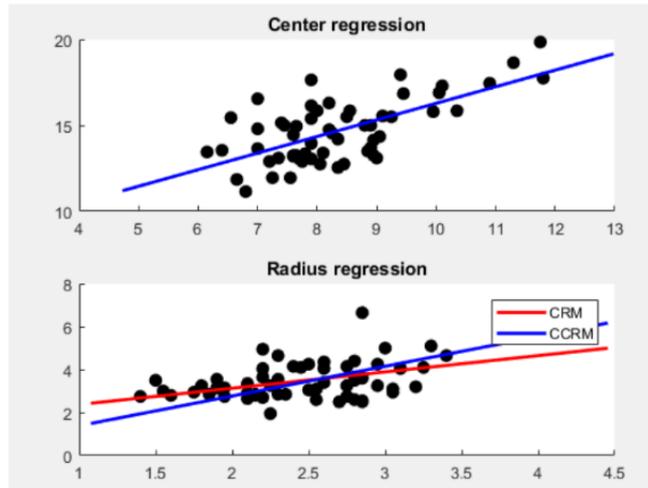


Figura 12. Ejemplo del Método del Centro y Rango Restringido (Maceda, 2018)

#### 4.1.9 MODELO MG

Este modelo se va a explicar teniendo en base al Paradigma 2, anteriormente mencionado.

Los procedimientos descritos de los métodos de regresión anteriores hacen uso de las variables relacionadas con los intervalos, mínimo-máximo, centro-radio, de forma independiente, esto es, cada una de estas variables se trata como una variable aleatoria independiente, asumiendo que dichas variables siempre se encuentran completa y adecuadamente definidas.

Sin embargo, esta premisa no se cumple en muchas ocasiones. Por esta razón, a continuación se va a desarrollar un modelo (modelo MG(k)) basado en la aritmética de intervalos, en el que se toma el intervalo en su conjunto para realizar una regresión lineal múltiple. El modelo propuesto es una extensión de los modelos anteriores que tiene en cuenta las relaciones entre los centros y los radios de los intervalos en una única ecuación.

La estimación de este problema, que se puede escribir como un problema de minimización, se analiza teóricamente y después se prueba experimentalmente.

La diferencia fundamental entre los modelos anteriores y los nuevos radica en la aplicación a los problemas reales. Este último modelo proporciona una información mucho más precisa frente a los tradicionales, preservando la coherencia con respecto a la información que proporcionan los datos de intervalo.

El modelo MG es la extensión natural del modelo M que fue desarrollado por Blanco Fernández (Blanco-Fernández et al., 2011), transformando la regresión lineal simple a múltiple en un escenario más flexible. El modelo MG(k) fue propuesto por García-Bárzana (García-Bárzana et al., 2020) y se muestra a continuación. El parámetro k hace referencia al número de variables independientes. En las próximas páginas se va a desarrollar el modelo en función del parámetro k y después se especificará tomando k los valores 1 y 2, es decir, haciendo una regresión con únicamente 1 y 2 variables independientes.

Sea  $y$  un intervalo aleatorio y sean  $x_1, x_2, x_3, \dots, x_k$  k intervalos aleatorios explicativos, el modelo MG propuesto es el siguiente:

$$y = [(b_1 \text{ Centre}(x)^t + b_4 \text{ Radii}(x)^t) \pm (b_2 \text{ Radii}(x)^t + b_3 |\text{Centre}(x)^t|)] + \varepsilon$$

donde  $b_1, b_4 \in \mathbb{R}^k$  y  $b_2, b_3 \in \mathbb{R}^{k+}$ ,  $\text{Centre}(x) = (x_1^C, x_2^C, \dots, x_k^C) \in \mathbb{R}^k$  y  $\text{Radii}(x) = (x_1^R, x_2^R, \dots, x_k^R)$ . Por último,  $\varepsilon$  es un error de intervalo aleatorio tal que  $E[\varepsilon|x] = \Delta \in K_C(\mathbb{R})$ . A partir de esta condición, es sencillo ver que  $x$  y  $\varepsilon$  no están correlacionados, es decir,  $\sigma_{\varepsilon, x_i} = 0$  para todo  $i=1, \dots, k$ . Las relaciones lineales separadas para las componentes de centro y radio se muestran a continuación:

$$y^C = (x^C)^t b_1 + (x^R)^t b_1 + \varepsilon^C,$$

$$y^R = (x^R)^t b_2 + |(x^C)^t| b_3 + \varepsilon^R.$$

Las dos variables,  $y^C, y^R$  se pueden obtener a partir de la información proporcionada por los intervalos aleatorios independientes en  $x$ , caracterizados por el vector aleatorio  $(x^C, x^R)$ . De esta propiedad se puede concluir que el modelo MG permite más flexibilidad en los intervalos aleatorios frente a otros modelos basados en la aritmética de conjuntos. Sin

embargo, la dificultad aumenta al incluir más coeficientes en el proceso de estimación, tal y como ocurre en los problemas clásicos de regresión lineal.

Con el objetivo de facilitar la notación vamos a definir los siguientes intervalos:  $x^M = [\text{Centre}(x)^t, \text{Centre}(x)^t]$ ,  $x^S = [-\text{Radii}(x)^t, \text{Radii}(x)^t]$ ,  $x^C = [|\text{Centre}(x)^t|, |\text{Centre}(x)^t|]$ , y  $x^R = [\text{Radii}(x)^t, \text{Radii}(x)^t]$ . Utilizando los anteriores intervalos expresamos el modelo en forma matricial.

$$y = X^{Bl} B + \varepsilon,$$

donde  $X^{Bl} = (x^M | x^S | x^C | x^R) \in Kc(\mathbb{R})^{1 \times 4k}$  y  $B = (b_1 | b_2 | b_3 | b_4)^t$ .

Dada una muestra aleatoria simple obtenida de los intervalos aleatorios  $(y, x_1, x_2, x_3, \dots, x_k)$  obtenemos:

$$y = X^{eBl} B + \varepsilon$$

donde  $y = ([y]_1, \dots, [y]_n)^t$ ,  $X^{eBl} = (X^M | X^S | X^C | X^R) \in Kc(\mathbb{R})^{n \times 4k}$ ,  $B = (b_1 | b_2 | b_3 | b_4)^t$  y  $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^t$ , de forma que  $E(\varepsilon|x) = 1^n \Delta$ .

#### 4.1.9.1 Modelo Autorregresivo (AR) de MG(k)

Si se utiliza la implementación autorregresiva (AR) del modelo MG(k) con el objetivo de relacionar el periodo actual con el valor de los k periodos anteriores, obtenemos la siguiente ecuación:

$$y = [(b_1 \text{Centre}(x)^t + b_4 \text{Radii}(x)^t) \pm (b_2 \text{Radii}(x)^t + b_3 |\text{Centre}(x)^t|)] + \varepsilon$$

siendo  $\text{Centre}(x) = (y_{t-1}^C, y_{t-2}^C, \dots, y_{t-k}^C)$ , y  $\text{Radii}(x) = (y_{t-1}^R, y_{t-2}^R, \dots, y_{t-k}^R)$ .

Carlos de los Santos creó en su TFG (de los Santos, 2021) una función en MATLAB para hacer una predicción del modelo MG, basándose en series temporales. En su TFG se detallan los parámetros de entrada y salida y se explica el funcionamiento de la función, pero a continuación se va a mostrar la visualización de salida para entenderlo a grandes rasgos. En la Figura 13 se utilizaron el precio de las acciones de Amazon.

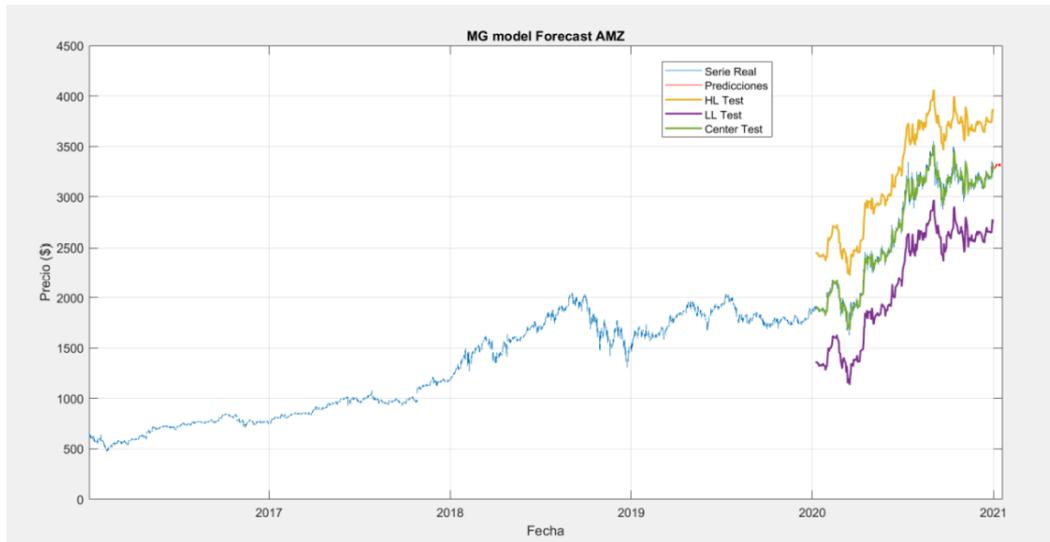


Figura 13. Representación del Modelo MG AR (de los Santos, 2021)

Para la realización de la función se creó un bucle que iteraba sobre los distintos valores de  $k$ , y después de dividir los datos en conjuntos de training y test, se estimó el modelo y se computaron los errores cuadráticos medios de ambos conjuntos, así como el valor de  $R^2$  para cada uno de los valores de  $k$ . A continuación, se muestran algunas de las representaciones descritas anteriormente.

La Figura 14 muestra la evolución del error cuadrático medio de ambos conjuntos de datos, en función del número de variables independientes  $k$ .

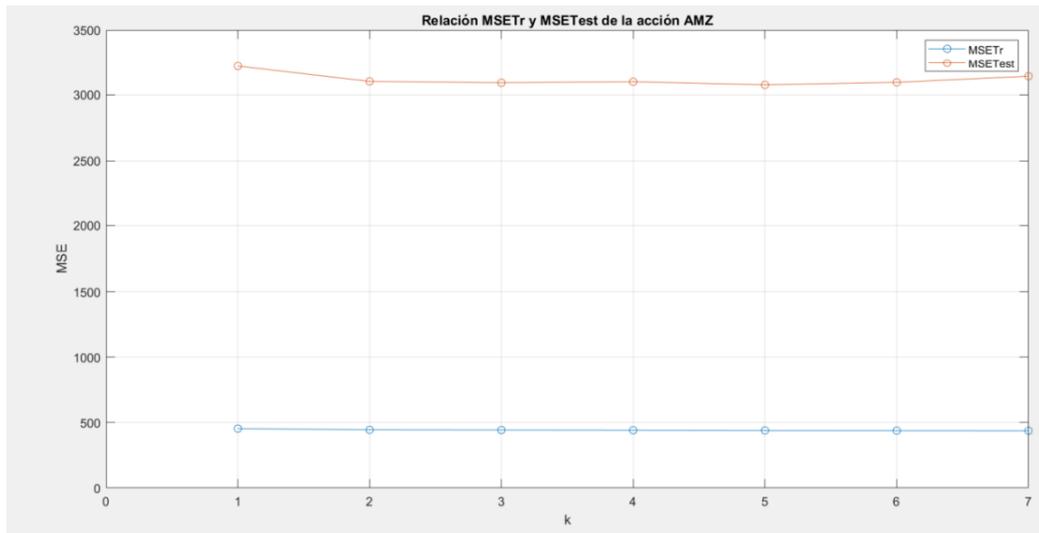


Figura 14. Evolución del error del Modelo MG (de los Santos, 2021)

La Figura 15 sin embargo muestra la evolución del  $R^2$  del modelo, también en función del parámetro k. Se comprueba que el  $R^2$  es superior a medida que aumenta el parámetro k.

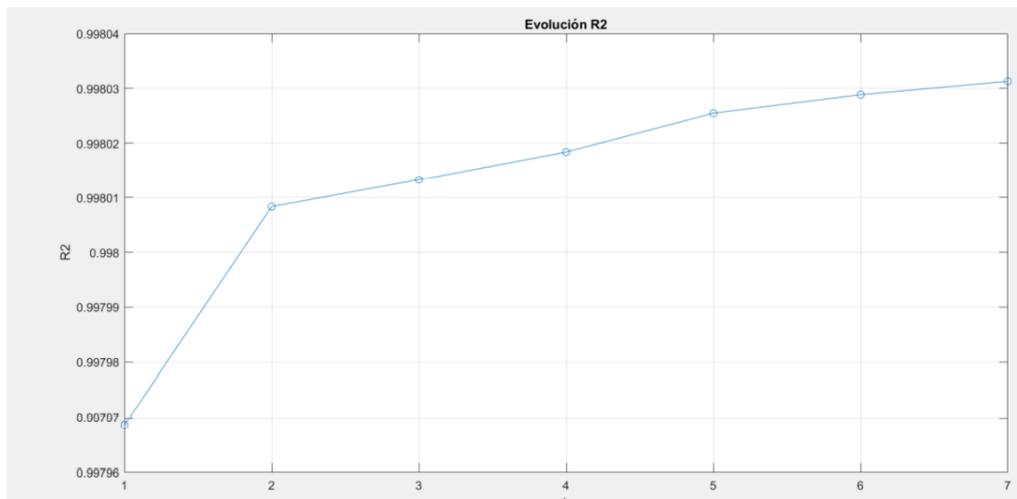


Figura 15. Evolución del  $R^2$  del Modelo MG (de los Santos, 2021)

#### 4.1.9.2 Modelos MG(1) y MG(2)

Una vez ya hemos descrito el modelo MG en base a k variables independientes, especificamos el mismo con tan sólo 1 variable independiente, es decir, siendo  $k=1$ .

$$y = [(b_1 \text{Centre}(x)^I + b_4 \text{Radii}(x)^I) \pm (b_2 \text{Radii}(x)^I + b_3 |\text{Centre}(x)^I|)] + \varepsilon$$

En este caso particular,  $b_1, b_4 \in \mathbb{R}^k$  y  $b_2, b_3 \in \mathbb{R}^{k+}$ ,  $\text{Centre}(x) = (x_1^C)$  y  $\text{Radii}(x) = (x_1^R)$ .

Lo mismo ocurre con el modelo MG(2). Se parte del modelo general MG(k) y se establece que el número de intervalos aleatorios explicativos es 2. La ecuación por tanto es la misma, pero habría que especificar que  $\text{Centre}(x) = (x_1^C, x_2^C)$  mientras que  $\text{Radii}(x) = (x_1^R, x_2^R)$ .

#### 4.1.9.3 MÉTODO DE ESTIMACIÓN DE LOS MÍNIMOS CUADRADOS

El método de estimación de los mínimos cuadrados (LS - Least Squares Estimation Model), trata de buscar aquellos parámetros  $\hat{B}$  y  $\hat{\Delta}$  que minimizan  $d^2_{\theta}(y, X^{eBl} A + 1^n C)$  con  $A \in \mathbb{R}^{4k}$   $x^{1 \cdot} C \in Kc(\mathbb{R})$  y garantizar la existencia de los residuos  $\varepsilon = y - H^{XeBl} A$ . Resulta fácil observar que  $\text{Radii}(X^{eBl} A) = \text{Radii} X a_2 + |\text{Centre} X| a_3$ , por lo que la siguiente condición se debe incluir en el problema de optimización:

$$\text{Radii} X a_2 + |\text{Centre} X| a_3 \leq \text{Radii} y$$

Teniendo esto en cuenta, podemos obtener la estimación de  $\Delta$  de la siguiente forma:

$$\hat{\Delta} = \bar{y} - H^{\overline{XeBl}} \hat{B}$$

Entonces, la estimación del parámetro B se obtiene resolviendo lo siguiente:

$$\text{minimizar } d^2_{\theta}(y - H^{XeBl} A, \bar{y} - H^{\overline{XeBl}} A)$$

sujeto a la condición anterior

$$\text{Radii} X a_2 + |\text{Centre} X| a_3 \leq \text{Radii} y$$

con  $A = (a_1 | a_2 | a_3 | a_4)$  tal que  $a_1, a_4 \in \mathbb{R}_k$  y  $a_2, a_3 \in \mathbb{R}^{k+}$ .

Este problema puede ser transformado en un problema de optimización cuadrática para el cálculo de los estimadores. Este problema presenta la siguiente estructura:

$$\min \frac{1}{2} x^t H x - f^t x$$

sujeto a

$$Ax \geq b$$

Para profundizar sobre este modelo consultar (García-Bárzana et al., 2020).

## Capítulo 5. CAUSALIDAD DE GRANGER EN ITS

### *ESTUDIO DE LA CAUSALIDAD DE GRANGER EN LAS SERIES TEMPORALES DE INTERVALO DEL ORO Y EL PETRÓLEO (2010-2021) EN FUNCIÓN DE LOS DIFERENTES MÉTODOS DE REGRESIÓN*

En este apartado se estudia la causalidad de Granger en dos series temporales de intervalo (ITS): los precios del oro y del petróleo. Se utiliza la información de datos mensuales desde el 30 de abril del año 2010 hasta el 30 de octubre del año 2021.

A continuación se muestran dos imágenes. En la primera se muestra la evolución de ambas series temporales de intervalo en el periodo de tiempo estudiado (el eje izquierdo de la gráfica corresponde a los valores del oro y el derecho del petróleo), mientras que en la segunda se muestra sus respectivas mediana y media. Se puede observar a simple vista que en la serie temporal del oro dichos valores son mucho más altos y con un intervalo más amplio.

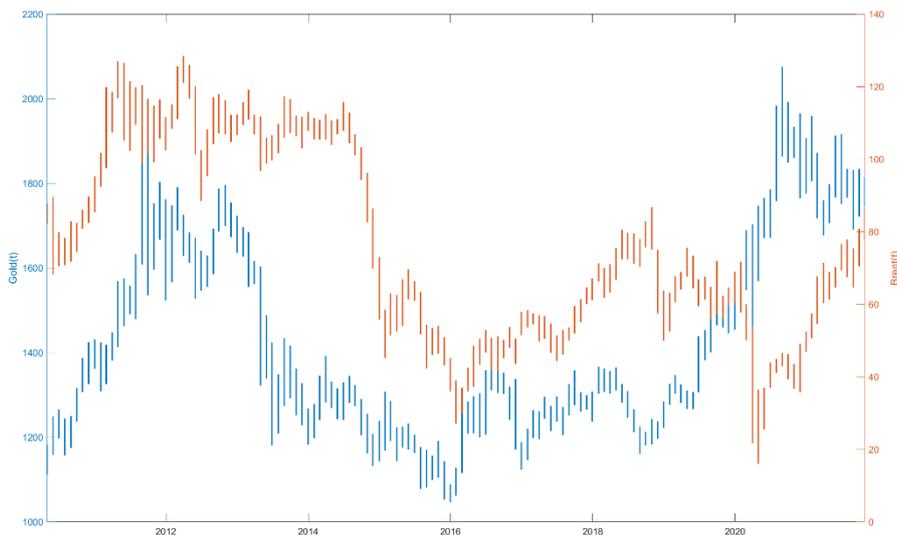


Figura 16. Evolución de los precios de las ITS del oro y del petróleo (2010-2021)

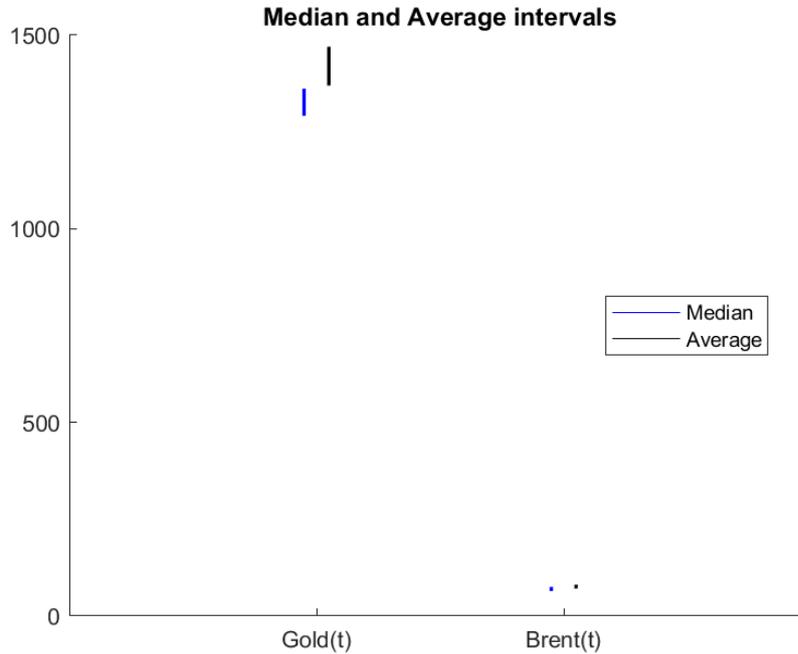


Figura 17. Media y mediana de las ITS del oro y del petróleo (2010-2021)

### 5.1.1 ANÁLISIS

En primer lugar se plantea la siguiente ecuación de la causalidad de Granger (Min Jang et al., 2019):

$$Y_t = \alpha + \sum_{i=1}^p \gamma_i Y_{t-i} + \sum_{j=1}^q \beta_j X_{t-j} + \varepsilon_t$$

donde  $Y_t$  corresponde al tiempo actual de la serie temporal del oro,  $Y_{t-i}$  son los valores pasados de esta misma serie y  $\gamma_i$  son los coeficientes a estimar en la regresión. Por otra parte,  $X_{t-j}$  son los valores pasados de la serie temporal del petróleo, y  $\beta_j$  son sus respectivos coeficientes. Como cualquier serie temporal incluye un error,  $\varepsilon_t$ . El parámetro  $\alpha$  es un término constante.

Una vez tenemos formulada la ecuación, procedemos a realizar el contraste de hipótesis de la causalidad de Granger. La hipótesis nula ( $H_0$ ) indica que todos los coeficientes  $\beta_j$  de la ecuación anterior son 0, es decir, la serie temporal  $X_{t-j}$ , en nuestro caso la serie temporal del petróleo, no ayudaría en la predicción de los precios del oro, por lo que el petróleo no causaría el oro.

$$H_0 : \beta_0 = \beta_1 = \beta_2 = \dots = \beta_q = 0$$

El test de causalidad de Granger, sin embargo, puede dar el resultado contrario, es decir, puede darse la situación de que se tenga que rechazar la hipótesis nula (confirmándose por tanto la hipótesis alternativa) porque haya al menos un coeficiente  $\beta_j$  que sea distinto de 0, y por tanto la serie temporal del petróleo ayudaría en la predicción de precios del oro.

Este test de causalidad de Granger se puede dar en ambas direcciones, es decir, se estudia si la serie temporal del petróleo causa o no causa la serie temporal del oro y viceversa.

Los siguientes análisis se han realizado teniendo en cuenta tres retardos de cada una de las dos series temporales y utilizando retornos. Teniendo en cuenta que  $P_t$  es el precio de un activo en un tiempo  $t$ , un retorno simple  $R_t(\tau)$  es el ratio de cambio de precio al mantener el activo desde el tiempo  $t - \tau$  hasta  $t$  (Maté, 2022).

Se ha definido la ecuación anterior con su correspondiente contraste para cada uno de los cinco métodos estudiados (CM, Mm, CRM, CCRM y MG). Además, en cada uno de ellos se han tenido en cuenta las cuatro características de los intervalos (mínimos, máximos, centros y radios). De esta manera, se han estudiado un total de veinte casos, cuatro por cada uno de los cinco métodos.

Por ejemplo, al estudiar los mínimos con el método de los centros la ecuación quedaría formulada de la siguiente manera:

$$Y^{CM,L}_t = \alpha + \sum_{i=1}^p \gamma_i Y_{t-i}^{CM,L} + \sum_{j=1}^q \beta_j X_{t-j}^{CM,L} + \varepsilon_t$$

De la misma manera, al estudiar los máximos con el método del Centro y Rango Restringido, la ecuación quedaría así:

$$Y^{CCRM,H}_t = \alpha + \sum_{i=1}^p \gamma_i Y_{t-i}^{CCRM,H} + \sum_{j=1}^q \beta_j X_{t-j}^{CCRM,H} + \varepsilon_t$$

En este último caso se calcula la predicción del oro teniendo en cuenta tres instantes pasados de la misma serie temporal y tres instantes pasados de la serie temporal de intervalos de precios del petróleo. Se hace la regresión a partir de los centros y radios de los intervalos (Método del Centro y Rango Restringido, CCRM).

Después, se realizan los test de causalidad de Granger en MATLAB, ejecutando el contraste de hipótesis. Existen tres maneras de hacer dichos tests, tal y como se han explicado en el Capítulo 2, pero solo se ha realizado una de ellas ya que al estar analizando únicamente el efecto de dos series temporales carece de sentido realizar la prueba con el resto de los tests. El test que se ha utilizado es el de *Leave-One-Out*.

### 5.1.2 RESULTADOS

Tras haber realizado los tests para cada uno de los veinte casos expuestos anteriormente, se han obtenido los siguientes resultados. La flecha verde indica causalidad, mientras que la flecha roja indica que no existe ninguna causalidad. Además, también vienen escritos los correspondientes p-valores.

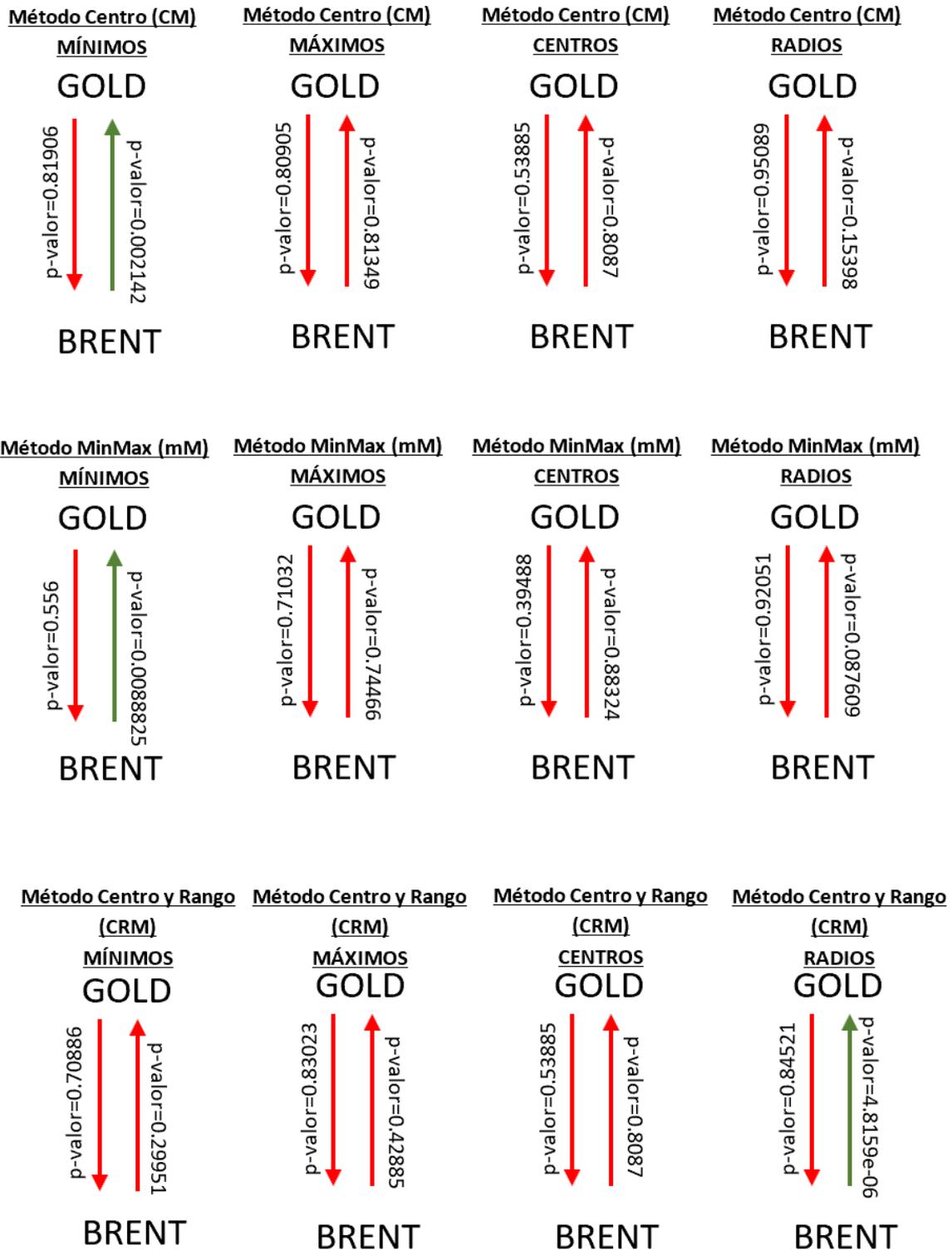


Figura 18. Gráficos de causalidad ITS del oro y del petróleo (2010-2021). Métodos CM, mM y CRM

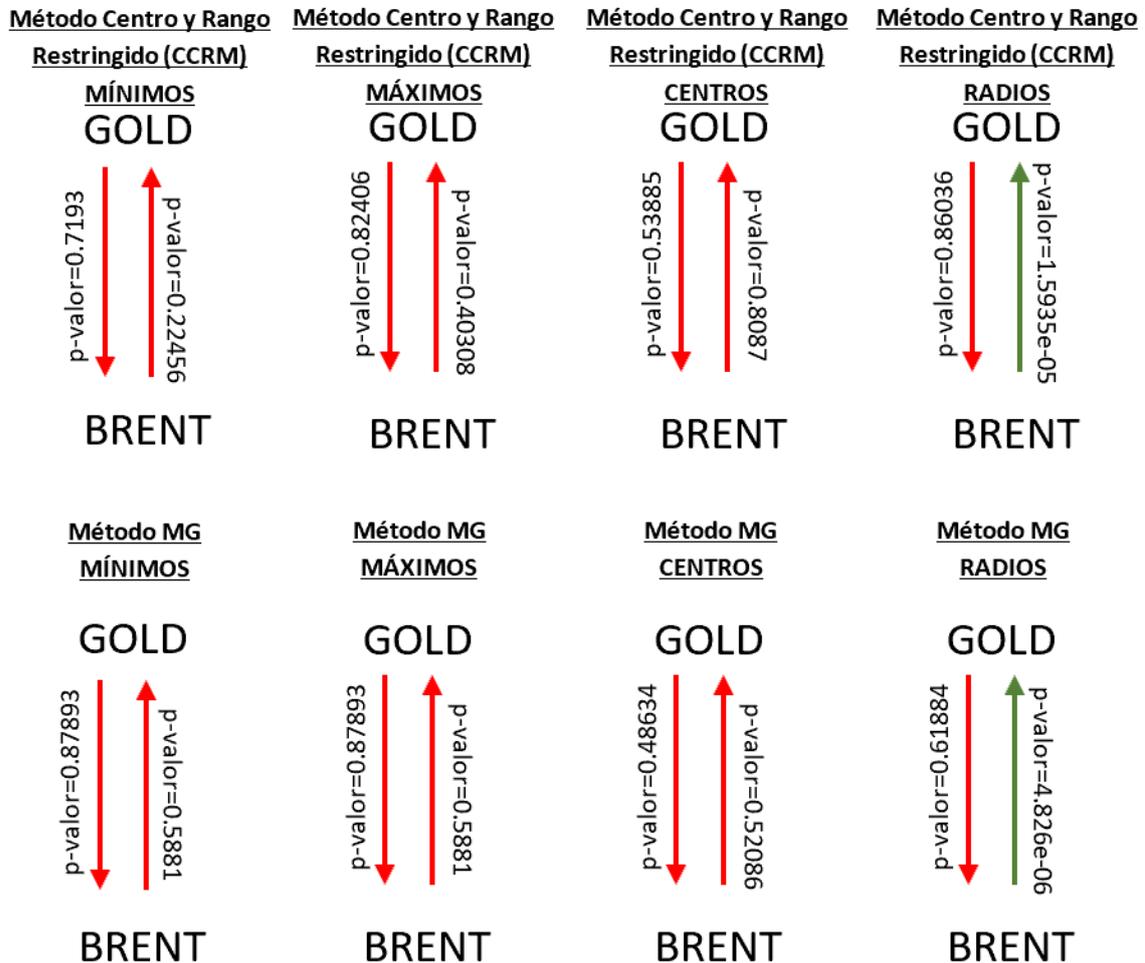


Figura 19. Gráficos de causalidad ITS del oro y del petróleo (2010-2021). Métodos CCRM y MG

Los esquemas de causalidad expuestos anteriormente indican que, en la mayoría de los casos, el oro no causa al petróleo, ni tampoco en sentido contrario. Solo se puede apreciar causalidad en algún caso, siendo siempre esta del petróleo hacia el oro.

En todos los métodos de regresión se obtienen los mismos resultados de no causalidad entre las dos series temporales, por lo que podemos afirmar con cierta robustez que, durante el periodo de tiempo estudiado (30/04/2010 - 30/10/2021) las series temporales de intervalo de frecuencia mensual del oro y el petróleo no presentan causalidad de Granger, o lo que es lo mismo, ninguna de ellas ayuda a predecir la otra.

## ***LA CAUSALIDAD DE GRANGER BAJO EL ENFOQUE DE SERIES TEMPORALES DE INTERVALO. EL CASO DEL ORO Y EL PETRÓLEO ANTES, DURANTE Y DESPUÉS DE LA CRISIS DE LA PANDEMIA DEL COVID-19***

El estudio llevado a cabo anteriormente aborda más de 10 años, desde el primer semestre de 2010 hasta el segundo semestre de 2021. Los resultados implican que, analizando un amplio periodo de tiempo, se puede afirmar que no existe una causalidad entre el oro y el petróleo.

Sin embargo, un estudio similar se ha realizado con datos de antes, durante y después de la crisis de la pandemia del Covid-19, obteniendo otro tipo de conclusiones.

### **5.1.3 ANÁLISIS**

En este análisis se ha investigado la causalidad de Granger de las series temporales de intervalo de los precios del oro y del petróleo con valores semanales recogidos en semestres, desde el 4 de enero de 2019 hasta el 20 de mayo de 2022 (5 semestres en total). Se han utilizado para ellos los cinco métodos de regresión de intervalos que hemos estudiado: Método del Centro, Método de los Mínimos y Máximos, Método del Centro y Rango, Método del Centro y Rango Restringido y Modelo MG.

La investigación se ha realizado para cada uno de los atributos del intervalo: mínimos, máximos, centros y radios. Los resultados obtenidos se muestran en las siguientes tablas:

## 5.1.4 RESULTADOS

### 5.1.4.1 Primer semestre 2019

MÉTODO DE LOS CENTROS (CM)				
	MÍNIMO	MÁXIMO	CENTRO	RADIO
<b>Brent -&gt; Gold</b>	No Causa	No Causa	No Causa	No Causa
<b>Gold -&gt; Brent</b>	No Causa	No Causa	No Causa	No Causa

MÉTODO DE LOS MÍNIMOS Y MÁXIMOS (mM)				
	MÍNIMO	MÁXIMO	CENTRO	RADIO
<b>Brent -&gt; Gold</b>	No Causa	No Causa	Causa	No Causa
<b>Gold -&gt; Brent</b>	No Causa	No Causa	No Causa	No Causa

MÉTODO DEL CENTRO Y EL RANGO (CRM)				
	MÍNIMO	MÁXIMO	CENTRO	RADIO
<b>Brent -&gt; Gold</b>	Causa	No Causa	No Causa	No Causa
<b>Gold -&gt; Brent</b>	No Causa	No Causa	No Causa	No Causa

MÉTODO DEL CENTRO Y EL RANGO RESTRINGIDO (CCRM)				
	MÍNIMO	MÁXIMO	CENTRO	RADIO
<b>Brent -&gt; Gold</b>	No Causa	No Causa	No Causa	No Causa
<b>Gold -&gt; Brent</b>	No Causa	No Causa	No Causa	No Causa

MÉTODO MG (MG)				
	MÍNIMO	MÁXIMO	CENTRO	RADIO
<b>Brent -&gt; Gold</b>	No Causa	No Causa	Causa	No Causa
<b>Gold -&gt; Brent</b>	No Causa	No Causa	No Causa	No Causa

Tabla 2. Resultados Causalidad de Granger ITS del oro y el petróleo. Primer semestre del año 2019

### 5.1.4.2 Segundo semestre 2019

<b>MÉTODO DE LOS CENTROS (CM)</b>				
	<b>MÍNIMO</b>	<b>MÁXIMO</b>	<b>CENTRO</b>	<b>RADIO</b>
<b>Brent -&gt; Gold</b>	No Causa	No Causa	No Causa	No Causa
<b>Gold -&gt; Brent</b>	No Causa	No Causa	No Causa	No Causa

<b>MÉTODO DE LOS MÍNIMOS Y MÁXIMOS (mM)</b>				
	<b>MÍNIMO</b>	<b>MÁXIMO</b>	<b>CENTRO</b>	<b>RADIO</b>
<b>Brent -&gt; Gold</b>	No Causa	No Causa	Causa	No Causa
<b>Gold -&gt; Brent</b>	No Causa	No Causa	No Causa	No Causa

<b>MÉTODO DEL CENTRO Y EL RANGO (CRM)</b>				
	<b>MÍNIMO</b>	<b>MÁXIMO</b>	<b>CENTRO</b>	<b>RADIO</b>
<b>Brent -&gt; Gold</b>	Causa	No Causa	No Causa	No Causa
<b>Gold -&gt; Brent</b>	No Causa	No Causa	No Causa	No Causa

<b>MÉTODO DEL CENTRO Y EL RANGO RESTRINGIDO (CCRM)</b>				
	<b>MÍNIMO</b>	<b>MÁXIMO</b>	<b>CENTRO</b>	<b>RADIO</b>
<b>Brent -&gt; Gold</b>	Causa	No Causa	No Causa	No Causa
<b>Gold -&gt; Brent</b>	No Causa	No Causa	No Causa	No Causa

<b>MÉTODO MG (MG)</b>				
	<b>MÍNIMO</b>	<b>MÁXIMO</b>	<b>CENTRO</b>	<b>RADIO</b>
<b>Brent -&gt; Gold</b>	No Causa	No Causa	No Causa	Causa
<b>Gold -&gt; Brent</b>	No Causa	No Causa	No Causa	No Causa

Tabla 3. Resultados Causalidad de Granger ITS del oro y el petróleo. Segundo semestre del año 2019

### 5.1.4.3 Primer semestre 2020

<b>MÉTODO DE LOS CENTROS (CM)</b>				
	<b>MÍNIMO</b>	<b>MÁXIMO</b>	<b>CENTRO</b>	<b>RADIO</b>
<b>Brent -&gt; Gold</b>	Causa	No Causa	Causa	No Causa
<b>Gold -&gt; Brent</b>	No Causa	No Causa	No Causa	No Causa

<b>MÉTODO DE LOS MÍNIMOS Y MÁXIMOS (mM)</b>				
	<b>MÍNIMO</b>	<b>MÁXIMO</b>	<b>CENTRO</b>	<b>RADIO</b>
<b>Brent -&gt; Gold</b>	Causa	No Causa	Causa	No Causa
<b>Gold -&gt; Brent</b>	No Causa	No Causa	No Causa	No Causa

<b>MÉTODO DEL CENTRO Y EL RANGO (CRM)</b>				
	<b>MÍNIMO</b>	<b>MÁXIMO</b>	<b>CENTRO</b>	<b>RADIO</b>
<b>Brent -&gt; Gold</b>	Causa	Causa	Causa	No Causa
<b>Gold -&gt; Brent</b>	No Causa	No Causa	No Causa	No Causa

<b>MÉTODO DEL CENTRO Y EL RANGO RESTRINGIDO (CCRM)</b>				
	<b>MÍNIMO</b>	<b>MÁXIMO</b>	<b>CENTRO</b>	<b>RADIO</b>
<b>Brent -&gt; Gold</b>	Causa	Causa	Causa	No Causa
<b>Gold -&gt; Brent</b>	No Causa	No Causa	No Causa	No Causa

<b>MÉTODO MG (MG)</b>				
	<b>MÍNIMO</b>	<b>MÁXIMO</b>	<b>CENTRO</b>	<b>RADIO</b>
<b>Brent -&gt; Gold</b>	No Causa	Causa	Causa	No Causa
<b>Gold -&gt; Brent</b>	No Causa	No Causa	No Causa	No Causa

Tabla 4. Resultados Causalidad de Granger ITS del oro y el petróleo. Primer semestre del año 2020

**5.1.4.4 Segundo semestre 2020**

<b>MÉTODO DE LOS CENTROS (CM)</b>				
	<b>MÍNIMO</b>	<b>MÁXIMO</b>	<b>CENTRO</b>	<b>RADIO</b>
<b>Brent -&gt; Gold</b>	No Causa	Causa	No Causa	No Causa
<b>Gold -&gt; Brent</b>	No Causa	No Causa	No Causa	No Causa

<b>MÉTODO DE LOS MÍNIMOS Y MÁXIMOS (mM)</b>				
	<b>MÍNIMO</b>	<b>MÁXIMO</b>	<b>CENTRO</b>	<b>RADIO</b>
<b>Brent -&gt; Gold</b>	Causa	Causa	Causa	No Causa
<b>Gold -&gt; Brent</b>	No Causa	No Causa	No Causa	No Causa

<b>MÉTODO DEL CENTRO Y EL RANGO (CRM)</b>				
	<b>MÍNIMO</b>	<b>MÁXIMO</b>	<b>CENTRO</b>	<b>RADIO</b>
<b>Brent -&gt; Gold</b>	Causa	Causa	Causa	No Causa
<b>Gold -&gt; Brent</b>	No Causa	No Causa	No Causa	No Causa

<b>MÉTODO DEL CENTRO Y EL RANGO RESTRINGIDO (CCRM)</b>				
	<b>MÍNIMO</b>	<b>MÁXIMO</b>	<b>CENTRO</b>	<b>RADIO</b>
<b>Brent -&gt; Gold</b>	Causa	Causa	Causa	No Causa
<b>Gold -&gt; Brent</b>	No Causa	No Causa	No Causa	No Causa

<b>MÉTODO MG (MG)</b>				
	<b>MÍNIMO</b>	<b>MÁXIMO</b>	<b>CENTRO</b>	<b>RADIO</b>
<b>Brent -&gt; Gold</b>	Causa	No Causa	Causa	No Causa
<b>Gold -&gt; Brent</b>	No Causa	No Causa	No Causa	No Causa

Tabla 5. Resultados Causalidad de Granger ITS del oro y el petróleo. Segundo semestre del año 2020

**5.1.4.5 Primer semestre 2021**

<b>MÉTODO DE LOS CENTROS (CM)</b>				
	<b>MÍNIMO</b>	<b>MÁXIMO</b>	<b>CENTRO</b>	<b>RADIO</b>
<b>Brent -&gt; Gold</b>	No Causa	No Causa	No Causa	No Causa
<b>Gold -&gt; Brent</b>	No Causa	No Causa	No Causa	No Causa

<b>MÉTODO DE LOS MÍNIMOS Y MÁXIMOS (mM)</b>				
	<b>MÍNIMO</b>	<b>MÁXIMO</b>	<b>CENTRO</b>	<b>RADIO</b>
<b>Brent -&gt; Gold</b>	No Causa	Causa	Causa	Causa
<b>Gold -&gt; Brent</b>	No Causa	No Causa	No Causa	No Causa

<b>MÉTODO DEL CENTRO Y EL RANGO (CRM)</b>				
	<b>MÍNIMO</b>	<b>MÁXIMO</b>	<b>CENTRO</b>	<b>RADIO</b>
<b>Brent -&gt; Gold</b>	No Causa	Causa	No Causa	Causa
<b>Gold -&gt; Brent</b>	No Causa	No Causa	No Causa	No Causa

<b>MÉTODO DEL CENTRO Y EL RANGO RESTRINGIDO (CCRM)</b>				
	<b>MÍNIMO</b>	<b>MÁXIMO</b>	<b>CENTRO</b>	<b>RADIO</b>
<b>Brent -&gt; Gold</b>	Causa	Causa	Causa	No Causa
<b>Gold -&gt; Brent</b>	No Causa	No Causa	No Causa	No Causa

<b>MÉTODO MG (MG)</b>				
	<b>MÍNIMO</b>	<b>MÁXIMO</b>	<b>CENTRO</b>	<b>RADIO</b>
<b>Brent -&gt; Gold</b>	No Causa	No Causa	No Causa	No Causa
<b>Gold -&gt; Brent</b>	No Causa	No Causa	No Causa	No Causa

*Tabla 6. Resultados Causalidad de Granger ITS del oro y el petróleo. Primer semestre del año 2021*

*5.1.4.6 Segundo semestre 2021*

<b>MÉTODO DE LOS CENTROS (CM)</b>				
	<b>MÍNIMO</b>	<b>MÁXIMO</b>	<b>CENTRO</b>	<b>RADIO</b>
<b>Brent -&gt; Gold</b>	No Causa	No Causa	No Causa	No Causa
<b>Gold -&gt; Brent</b>	No Causa	No Causa	No Causa	No Causa

<b>MÉTODO DE LOS MÍNIMOS Y MÁXIMOS (mM)</b>				
	<b>MÍNIMO</b>	<b>MÁXIMO</b>	<b>CENTRO</b>	<b>RADIO</b>
<b>Brent -&gt; Gold</b>	No Causa	Causa	Causa	No Causa
<b>Gold -&gt; Brent</b>	No Causa	No Causa	No Causa	No Causa

<b>MÉTODO DEL CENTRO Y EL RANGO (CRM)</b>				
	<b>MÍNIMO</b>	<b>MÁXIMO</b>	<b>CENTRO</b>	<b>RADIO</b>
<b>Brent -&gt; Gold</b>	No Causa	No Causa	No Causa	Causa
<b>Gold -&gt; Brent</b>	No Causa	No Causa	No Causa	No Causa

<b>MÉTODO DEL CENTRO Y EL RANGO RESTRINGIDO (CCRM)</b>				
	<b>MÍNIMO</b>	<b>MÁXIMO</b>	<b>CENTRO</b>	<b>RADIO</b>
<b>Brent -&gt; Gold</b>	No Causa	No Causa	No Causa	Causa
<b>Gold -&gt; Brent</b>	No Causa	No Causa	No Causa	No Causa

<b>MÉTODO MG (MG)</b>				
	<b>MÍNIMO</b>	<b>MÁXIMO</b>	<b>CENTRO</b>	<b>RADIO</b>
<b>Brent -&gt; Gold</b>	No Causa	No Causa	No Causa	No Causa
<b>Gold -&gt; Brent</b>	No Causa	No Causa	No Causa	No Causa

*Tabla 7. Resultados Causalidad de Granger ITS del oro y el petróleo. Segundo semestre del año 2021*

5.1.4.7 Primer semestre 2022

<b>MÉTODO DE LOS CENTROS (CM)</b>				
	<b>MÍNIMO</b>	<b>MÁXIMO</b>	<b>CENTRO</b>	<b>RADIO</b>
<b>Brent -&gt; Gold</b>	No Causa	No Causa	No Causa	No Causa
<b>Gold -&gt; Brent</b>	No Causa	No Causa	No Causa	No Causa

<b>MÉTODO DE LOS MÍNIMOS Y MÁXIMOS (mM)</b>				
	<b>MÍNIMO</b>	<b>MÁXIMO</b>	<b>CENTRO</b>	<b>RADIO</b>
<b>Brent -&gt; Gold</b>	No Causa	Causa	No Causa	No Causa
<b>Gold -&gt; Brent</b>	No Causa	No Causa	No Causa	No Causa

<b>MÉTODO DEL CENTRO Y EL RANGO (CRM)</b>				
	<b>MÍNIMO</b>	<b>MÁXIMO</b>	<b>CENTRO</b>	<b>RADIO</b>
<b>Brent -&gt; Gold</b>	No Causa	No Causa	No Causa	Causa
<b>Gold -&gt; Brent</b>	No Causa	No Causa	No Causa	No Causa

<b>MÉTODO DEL CENTRO Y EL RANGO RESTRINGIDO (CCRM)</b>				
	<b>MÍNIMO</b>	<b>MÁXIMO</b>	<b>CENTRO</b>	<b>RADIO</b>
<b>Brent -&gt; Gold</b>	No Causa	No Causa	No Causa	No Causa
<b>Gold -&gt; Brent</b>	No Causa	No Causa	No Causa	No Causa

<b>MÉTODO MG (MG)</b>				
	<b>MÍNIMO</b>	<b>MÁXIMO</b>	<b>CENTRO</b>	<b>RADIO</b>
<b>Brent -&gt; Gold</b>	No Causa	No Causa	No Causa	No Causa
<b>Gold -&gt; Brent</b>	No Causa	No Causa	No Causa	No Causa

Tabla 8. Resultados Causalidad de Granger ITS del oro y el petróleo. Primer semestre del año 2022

Como conclusión, los resultados empíricos muestran que el análisis de causalidad es del petróleo al oro, mientras que no se ha encontrado causalidad en sentido contrario.

Además, conviene subrayar que la causalidad que se obtiene es mucho más fuerte en aquellos semestres en los que la crisis mundial provocada por la pandemia ha sido más pronunciada.

El primer caso diagnosticado de coronavirus fue el 17 de noviembre de 2019 en Wuhan, China, pero fue en el primer semestre de 2020 cuando este virus se transmitió a muchos países a nivel mundial y empezaron a aumentar drásticamente el número de contagios.

La Organización Mundial de la Salud (OMS) la declaró una emergencia de salud pública de importancia internacional el 30 de enero de 2020 y la reconoció como una pandemia el 11 de marzo de 2020. En consecuencia, muchos gobiernos optaron por aplicar políticas de confinamiento domiciliario y distanciamiento social y es por ello que más allá del impacto sobre la vida de millones de personas en todo el mundo, la pandemia supuso un duro golpe sobre la evolución de la economía a nivel global. Muchos países registraron una caída histórica de sus valores del PIB, concretamente en España el Producto Interior Bruto descendió un 11% en el año 2020. En la Figura 20 se puede observar la caída abismal con respecto al año anterior.

Evolución del PIB de España

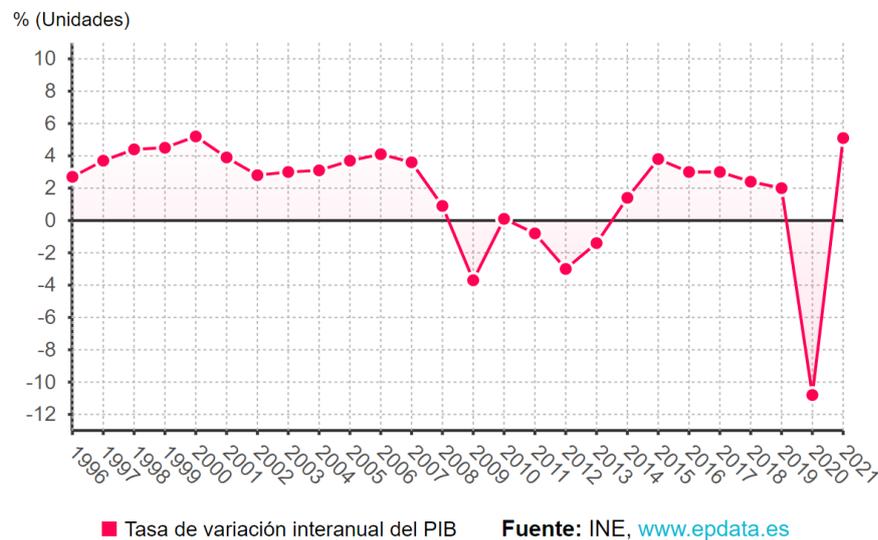


Figura 20. Evolución del PIB de España (1996-2021) (Instituto Nacional de Estadística, 2022)

En el año 2021 se empezó a notar una ligera reactivación de la economía. Se suavizaron las medidas impuestas por el gobierno y las olas de coronavirus comenzaron a estar más controladas por los programas de vacunación masiva que se realizaron mundialmente.

La recuperación económica aún continúa, con algunos casos diagnosticados hoy en día, aunque la incidencia es mucho más baja y ya no existe el colapso sanitario. Según las cifras confirmadas por el Banco Mundial (Banco Mundial, 2022), la perspectiva de crecimiento a nivel mundial para el año 2022 es del 4.1%, mientras que la del 2023 es de un 2.3%. En el año 2020 fue de un -3.4%. Esta previsión era mayor pero se ha desacelerado como reflejo de los brotes de Covid-19, la disminución del apoyo fiscal y las persistentes dificultades en las cadenas de suministro, como consecuencia de la actual guerra de Ucrania. Todos estos datos indican que la economía se está viendo reactivada por la pandemia, pero no quiere decir que esta haya terminado por completo.

Después de explicar cronológicamente la crisis del Covid-19 se destaca la relación con los resultados obtenidos de la causalidad de Granger en las ITS. En el año previo a la pandemia,

2019, se detecta que hay una causalidad muy débil del petróleo hacia el oro en cada uno de los cinco métodos de regresión estudiados. Sin embargo, en el primer semestre de 2020, cuando desató la crisis con el inicio de las restricciones, podemos ver una causalidad mucho más fuerte. El petróleo causa al oro en todos los métodos presentando robustez, y teniendo en cuenta por lo menos la mitad de las cuatro características de los intervalos (mínimos, máximos, centros y radios), por lo que podemos afirmar que es una causalidad fuerte.

En el segundo semestre de 2020 obtenemos resultados similares a los del primero, con una causalidad unidireccional fuerte. Los primeros seis meses del año siguiente, 2021, también muestran cierta causalidad, aunque no tan fuerte como en el año 2020. Esto se debe a que la crisis del Covid-19 no estaba atravesando su época más complicada por los inicios de la vacunación. En el segundo semestre de 2021 ya se presenta una causalidad mucho más débil (casi nula) por la pronta recuperación económica, al igual que ocurre en lo que llevamos de este año.

Con todo ello se concluye que la causalidad de Granger estudiada en las series temporales de intervalo del oro y el petróleo es mucho más fuerte en los momentos de crisis (valle) frente a los momentos de auge (pico), siempre teniendo en cuenta que la causalidad es unidireccional del petróleo hacia el oro. Estos resultados cobran sentido dado que las épocas de crisis se caracterizan por la incertidumbre y consecuente cambio en los precios del oro (activo sin riesgo), por lo que los cambios de precio en el petróleo perjudican a la estabilidad del oro y de la economía.

Con respecto al análisis de la causalidad de Granger en un periodo de 10 años podemos afirmar que no se puede observar una causalidad clara entre el oro y el petróleo, pero sin embargo si reducimos el periodo de análisis y nos centramos en un ciclo económico es entonces cuando podemos ver las diferencias.

## ***ESTUDIO DE LAS MÉTRICAS DE PRECISIÓN IARV Y CER PARA CADA UNO DE LOS MÉTODOS DE REGRESIÓN DE INTERVALOS***

Se han analizado las métricas de precisión de iARV, CR, y ER para cada uno de los cinco métodos de regresión de intervalos que se han aplicado. Para el análisis se han utilizado datos mensuales comprendidos en algo más de 10 años (30/04/2010-30/10/2021), las mismas fechas que en el estudio realizado con las ITS del oro y del petróleo.

A una serie temporal de intervalo  $\{[x_t]\}$  le corresponde una predicción  $\{[\hat{x}_t]\}$  donde  $t = 1, 2 \dots T$ .

Hay diferentes maneras para evaluar la precisión de las series temporales de intervalo. Entre ellas se encuentran (Carlos Maté, 2021):

### **5.1.5 VARIANZA RELATIVA MEDIA DEL INTERVALO (INTERVAL AVERAGE RELATIVE VARIANCE - IARV)**

El coeficiente se expresa con la siguiente notación:

$$iARV = \frac{\sum_{i=1}^n (\widehat{y}_i^L - y_i^L)^2 + \sum_{i=1}^n (\widehat{y}_i^U - y_i^U)^2}{\sum_{i=1}^n (\overline{y}_i^L - y_i^L)^2 + \sum_{i=1}^n (\overline{y}_i^U - y_i^U)^2}$$

Esta medida de precisión compara las predicciones del modelo con las predicciones dadas por los valores medios de la variable de intervalo.

Menores valores de iARV indican mejores estimaciones, convergiendo a 0 cuando el modelo de regresión es perfecto.

A parte de las medidas de error iARV, existen algunos ratios que evalúan la precisión del modelo. Son conceptos muy sencillos, pero aun así consiguen evaluar correctamente los modelos de regresión.

La idea principal que hay detrás de estas dos medidas (CR y ER) es comparar la intersección entre dos rangos (el real y el predicho) para cada una de las amplitudes individuales (real y predicha).

### 5.1.6 TASA DE CUBRIMIENTO (COVERAGE RATE - CR)

Siendo  $w([x]_t \cap [\hat{x}]_t)$  la longitud de la intersección existente entre el intervalo real y el intervalo predicho con el método M, y  $w([x]_t)$  la amplitud del intervalo real, entonces la coverage rate en un tiempo t se define como el ratio entre las dos longitudes, que se expresa de la siguiente manera:

$$CR_M = \frac{1}{T} \sum_{t=1}^T \frac{w([x]_t \cap [\hat{x}]_t)}{w([x]_t)}$$

Cuanto más cerca se encuentre este ratio del 100%, más próxima estará la intersección entre los dos intervalos con respecto al intervalo real, por lo que significa que el intervalo predicho cubre una gran parte del intervalo actual.

De todas maneras, este ratio no tiene en cuenta la posibilidad de que el intervalo predicho sea más amplio que el intervalo real. Por ejemplo, imaginemos que el intervalo a predecir es [10,12], entonces el intervalo [9,13] cubre el 100% del intervalo, pero es peor resultado (menos eficiente) que el intervalo [10.3, 12.5]. Este segundo constituiría un mejor resultado de predicción.

### 5.1.7 TASA DE EFICIENCIA (EFFICIENCY RATE - ER)

Teniendo en cuenta, al igual que anteriormente, que  $w([x]_t \cap [\hat{x}]_t)$  es la longitud de la intersección existente entre el intervalo real y el intervalo predicho con el método M, y  $w([\hat{x}]_t)$  corresponde a la amplitud del intervalo predicho. En este caso, el ratio de eficiencia en un tiempo t viene determinada por el cociente entre las dos longitudes.

$$ER_M = \frac{1}{T} \sum_{t=1}^T \frac{w([x]_t \cap [\hat{x}]_t)}{w([\hat{x}]_t)}$$

Cuanto más cerca se encuentre este coeficiente del 100% indica que mayor es la proporción en la que el intervalo predicho intersecciona con el intervalo real. Esta medida de eficiencia, como su nombre indica, penaliza las situaciones en las que el intervalo predicho es innecesariamente más amplio que el intervalo real. Sin embargo, en este caso no se tienen en cuenta aquellos casos en los que el intervalo real es más amplio que el intervalo predicho. Por ejemplo, imaginemos que el intervalo real que vamos a tratar de predecir es el [9,13], entonces el intervalo predicho [9.5, 12.5] es más eficiente en términos de cobertura que el intervalo [10,12].

### 5.1.8 RATIO DE CUBRIMIENTO Y EFICIENCIA (COVERAGE AND EFFICIENCY RATE – CER)

Para predecir series temporales es importante que ambos ratios sean lo más altos posibles. Una medida global de las medidas de cobertura y eficiencia es la media entre ambos ratios. Esto es,

$$CER_M = \frac{1}{2} (CR_M + ER_M)$$

A continuación se muestra una imagen en la que se muestra un ejemplo de sus correspondientes ratios de cobertura y eficiencia (Maceda, 2018).

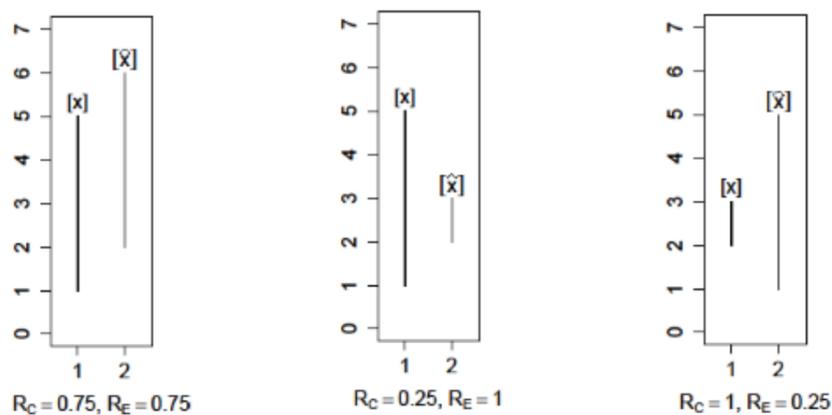


Figura 21. Ejemplos de Ratios de Cobertura y Eficiencia (Maceda, 2018)

### 5.1.9 RESULTADOS

Teniendo en cuenta todas estas métricas de precisión, se han obtenido estos resultados para cada uno de los métodos:

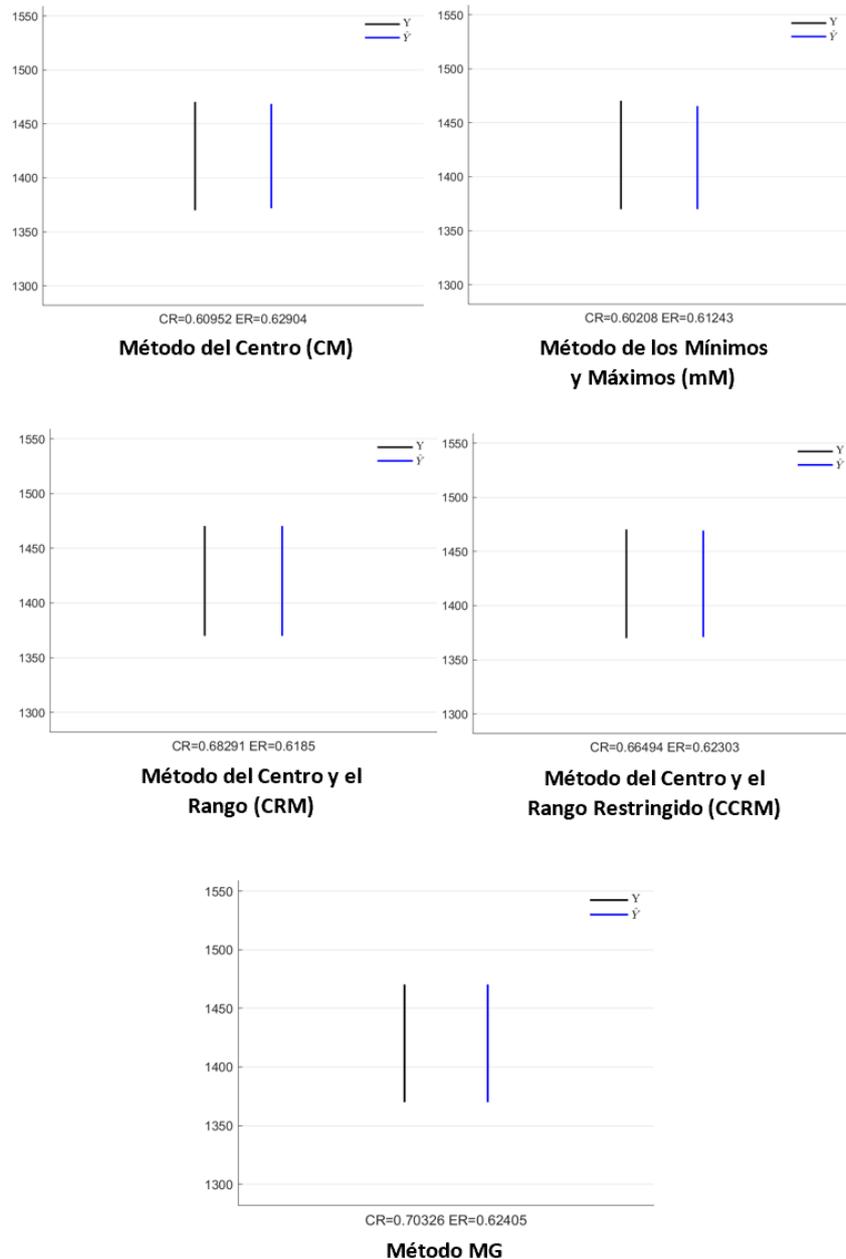


Figura 22. Resultados CER para cada uno de los métodos

A modo de resumen:

MÉTRICAS	CM	mM	CRM	CCRM	MG
<b>iARV</b>	0.0603	0.0603	0.0517	0.0519	<b>0.0497</b>
<b>CER</b>	0.6193	0.6073	0.6507	0.6440	<b>0.6637</b>

*Tabla 9. Resultados iARV y CER para cada uno de los métodos de regresión*

El método que mejores métricas presenta, como era de esperar al ser el más reciente y complejo, es el modelo MG, presentando el valor CER más alto, y el valor iARV más bajo de entre todos los modelos. Aun así, los análisis anteriores se han llevado a cabo con todos los métodos para presentar una mayor robustez en los resultados.

## **Capítulo 6. ANÁLISIS DE DATOS VALORADOS POR HISTOGRAMAS**

Las series temporales en las que las observaciones son datos únicos son muy útiles para representar muchas situaciones teóricas y prácticas. De hecho, son prácticamente el único tipo de serie temporal que se suele considerar. Sin embargo, las series temporales clásicas no describen fielmente los fenómenos en los que se dispone de un conjunto de realizaciones de la variable observada en cada punto del tiempo, o cuando la variable observada tiene cierto tipo de variabilidad. Hay dos situaciones típicas en las que esto ocurre:

- 1) Si una variable se mide a lo largo del tiempo para cada individuo de un grupo, y el interés no reside en los individuos sino en el grupo en su conjunto. En este caso, una serie temporal de la media muestral de la variable observada a lo largo del tiempo sería una débil representación.
- 2) Cuando una variable se observa con una frecuencia determinada (por ejemplo, minutos), pero debe analizarse con una frecuencia inferior (por ejemplo, días). En este caso, si la variable sólo se muestrea en la frecuencia más baja, o si sólo se muestran los valores más altos y más bajos, se descuida mucha información.

Estas dos situaciones que se han explicado anteriormente son claros ejemplos de la necesidad de las series temporales de histograma (HTS), ofreciendo una mayor información que una serie temporal con información agregada.

En la Figura 23 se muestran las dos típicas situaciones en las que se emplean HTS. En las dos representaciones de la parte inferior de la gráfica se muestran datos de alta frecuencia. En este caso la serie temporal de histograma proporciona una forma de resumir dichos datos, teniendo en cuenta la variabilidad entre cada periodo. Tal y como dicen (Engle & Russell, 2004) la alta frecuencia de datos tienen algunas características que plantean nuevos retos: espaciamiento temporal irregular, patrones diurnos, discrepancia de precios y dependencia compleja. Estas características dificultan el hecho de predecir estos datos. Sin embargo, representarlos en forma de HTS ofrece una forma de abordar estas dificultades. El objetivo

en este caso no es la obtención de previsiones para el siguiente instante, sino el de obtener una previsión de la distribución global para el siguiente periodo. La información proporcionada por una previsión en forma de histograma complementa la información proporcionada por una previsión del valor cercano, y puede ser muy útil a la hora de establecer estrategias de negociación.

En la representación de la parte superior de la Figura 23 se muestran datos contemporáneos y espaciales, y en este caso la serie temporal de histograma parece una forma natural de gestionar este tipo de información, especialmente si se dispone de un gran número de observaciones para cada instante de tiempo. Las HTS son preferibles de utilizar frente a las medias o los totales ya que en este último caso se estaría perdiendo mucha información.

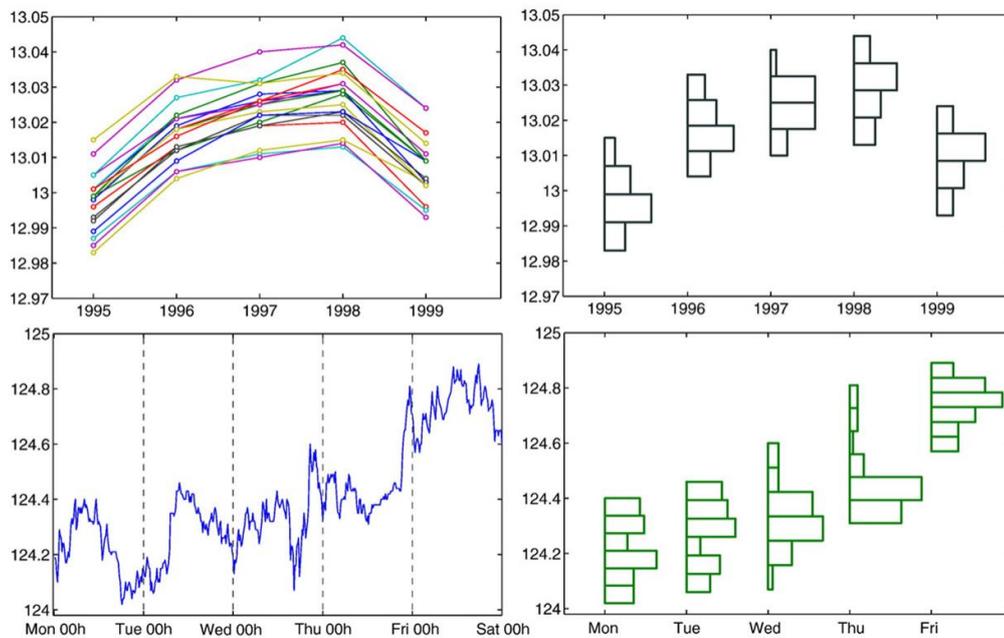


Figura 23. Ejemplos de HTS (Arroyo & Maté, 2009)

## ***VENTAJAS DE LOS HISTOGRAMAS***

En este trabajo se propone el uso de los histogramas como una herramienta para representar distribuciones. Sin embargo, cabe preguntarse por qué razón se utilizan los histogramas en lugar de cualquier otro método que represente la densidad subyacente. Es por ello que se van a mencionar algunas de las ventajas a favor del uso de los histogramas (Arroyo, 2008).

### **6.1.1 REPRESENTACIÓN CERCANA A LOS DATOS ORIGINALES**

Un histograma es una representación cercana a los datos originales y que no precisa de la imposición sobre los mismos de ninguna distribución *a priori*.

### **6.1.2 VERSATILIDAD**

La versatilidad de los histogramas permite al analista centrarse en las características que sean de un mayor interés, por ejemplo, en un conjunto de cuartiles o en una parte del rango de la variable. La definición del histograma es suficientemente versátil y flexible como para adaptarse a distintos requerimientos.

### **6.1.3 PRECISIÓN**

Un histograma permite describir las características esenciales de los datos con una precisión razonable.

### **6.1.4 FACILIDAD EN EL TRATAMIENTO COMPUTACIONAL**

La sencilla estructura de los histogramas simplifica su tratamiento computacional. El hecho de manejar histogramas disminuye drásticamente el tiempo de computación requerido por operaciones aritméticas, cálculo de distancias y estimación del baricentro de un conjunto de histogramas.

Todas estas ventajas justifican la utilización de histogramas para este proyecto.

## ***TIPOS DE HISTOGRAMAS***

Una vez se han descrito las ventajas de la utilización de los histogramas, se exponen los distintos tipos de histogramas que puede haber. Se debe elegir el tipo de histograma para representar la HTS de acuerdo con el objetivo del análisis que se quiera realizar. Hay 4 tipos diferentes de histogramas (Arroyo, 2008):

### **6.1.5 HISTOGRAMAS EQUISPACIADOS**

Estos histogramas son los más populares. Se usa un número de intervalos que permita describir los datos de una manera precisa pero sin enmascarar las características de la distribución.

### **6.1.6 HISTOGRAMAS EQUIFRECUENCIALES**

Permiten construir histogramas a partir de los deciles de la distribución. Dentro de este tipo de histogramas se clasificarían los diagramas de caja (*boxplot*).

### **6.1.7 HISTOGRAMAS CONSTRUIDOS SOBRE UNA PARTICIÓN DEL DOMINIO DE LA VARIABLE**

Son histogramas que permiten estudiar el comportamiento de la variable en unos determinados intervalos, que serán elegidos por el analista según lo que considere interesante.

### **6.1.8 HISTOGRAMAS DEFINIDOS A PARTIR DE UNA SECUENCIA DE CUARTILES**

Este tipo de histogramas permiten estudiar el comportamiento de la distribución en cuartiles, pudiendo hacer énfasis en aquellas partes que supongan un mayor interés.

## ***SERIES TEMPORALES DE HISTOGRAMA (HTS)***

Las series temporales de intervalo (ITS) también reflejan la variabilidad de una observación en un determinado instante mediante un rango de valores, sin embargo, no informan de lo que sucede entre los extremos considerados, es decir, no indican cómo se distribuyen las observaciones dentro de dicho rango.

Esta carencia es cubierta por los histogramas, que permiten representar de una manera más completa la distribución de datos cuantitativos.

Una serie temporal de histogramas  $\{h_{X_t}\}$  puede definirse como una secuencia de distribuciones observadas en instantes sucesivos en el tiempo denotados por  $t=1, \dots, n$  donde cada distribución es representada mediante un histograma  $h_{X_t}$  que viene dado por:

$$h_{X_t} = \{([I]_{t,1}, \pi_{t,1}), \dots, ([I]_{t,p_t}, \pi_{t,p_t})\}, \text{ con } t=1, \dots, n,$$

donde  $\{\pi_{t,i}\}$  con  $i=1, \dots, p_t$  es una distribución de frecuencia o de probabilidad en el dominio considerado que cumple que  $\pi_{t,i} \geq 0$  y que  $\sum_{i=1}^{p_t} \pi_{t,i} = 1$ ; y donde  $[I]_{t,i} = [[\underline{I}_{t,i}, \bar{I}_{t,i})$  con  $-\infty < \underline{I}_{t,i} \leq \bar{I}_{t,i} < \infty$  e  $\underline{I}_{t,i} \leq \bar{I}_{t-1}$ , con  $i \geq 2$ .

Es importante observar que, a pesar de su simplicidad, la ecuación anterior cubre todos los posibles estimadores de densidad.

Un ejemplo de serie temporal de histogramas se muestra en la Figura 24, reflejando la distribución mensual de las precipitaciones en China.

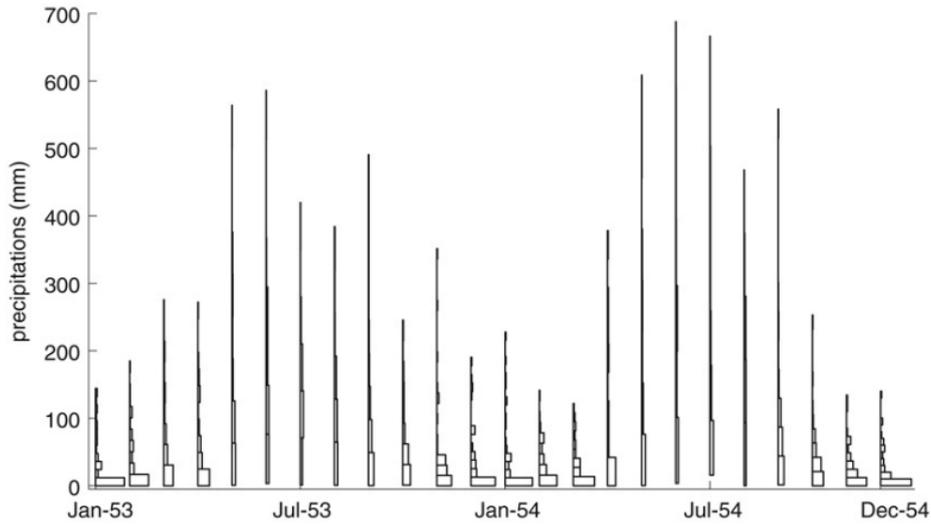


Figura 24. HTS de las precipitaciones mensuales en China (1953-1954) (Arroyo & Maté, 2009)

Otro ejemplo podría ser esta representación gráfica (Figura 25), en la que se muestra la evolución de los niveles de dióxido de nitrógeno en Madrid, desde enero de 1994 hasta diciembre de 1995. Cada histograma representa la distribución de la media mensual del nivel de  $\text{NO}_2$  en Madrid.

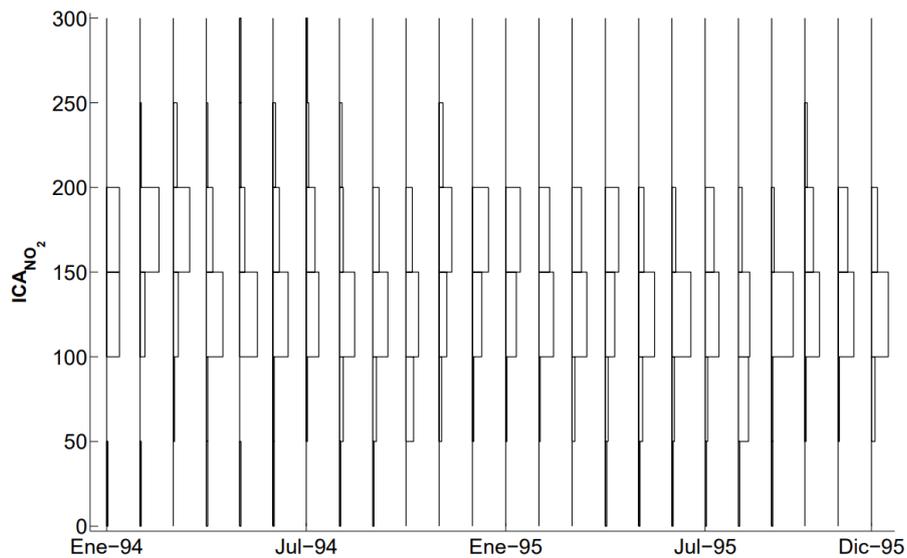


Figura 25. HTS de la distribución mensual de dióxido de nitrógeno en Madrid (Arroyo, 2008)

## MÉTODOS DE REGRESIÓN LINEAL DE LAS HTS

La regresión lineal es un método general para estimar/describir la asociación entre una variable resultado (variable dependiente) y uno o varios predictores (variables independientes) en una ecuación, dando como resultado lo siguiente:

$$\text{datos} = \text{ajuste del modelo} + \text{residuos}$$

La regresión lineal es un concepto sencillo de aplicar en los datos clásicos. La complejidad con los datos de intervalo es un poco mayor, pero aun así hemos estudiado varios métodos de regresión que tratan de ajustar al máximo las predicciones. Sin embargo, el concepto de regresión cuando se trata de histogramas no es tan sencillo, ya que se intenta buscar una transformación lineal que nos permita predecir el histograma de la variable resultado.

A continuación se muestra una imagen (Figura 26) en la que se muestra claramente la diferencia entre la complejidad de la regresión con datos clásicos, de intervalo y de histograma.

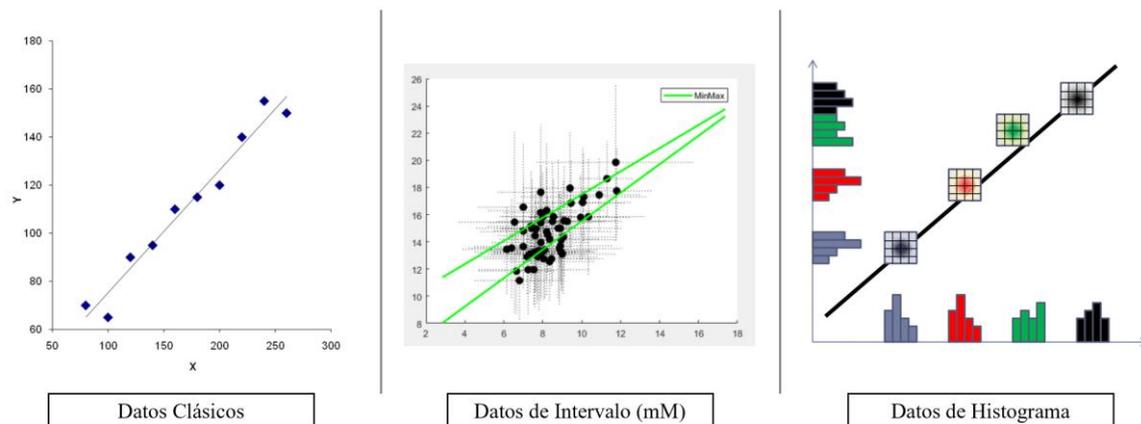


Figura 26. Representación de la regresión lineal con datos clásicos, datos de intervalo (método MinMax) y datos de histograma (Irpino, 2010), (de los Santos, 2021)

## ***MEDIDAS DE PRECISIÓN DE LAS HTS***

En las series temporales clásicas, las medidas de precisión se basan en la diferencia entre los valores observados y los previstos, es decir,  $\Delta x = x - \hat{x}$ , donde  $x$  es la observación y  $\hat{x}$  es el valor previsto. Sin embargo, debido a la complejidad de los histogramas, las series temporales de histograma requieren un enfoque diferente para cuantificar las diferencias entre histogramas.

En la previsión de la densidad, (Diebold et al., 1998) proponen el uso de pruebas de bondad de ajuste para evaluar si la predicción de la densidad se corresponde con la densidad real o no. Por desgracia, este enfoque no es útil para la medición de la precisión de HTS, ya que no cuantifica la desviación entre las dos densidades.

(Hall & Mitchell, 2007) proponen un enfoque diferente que utilizan una medida de disimilitud, el criterio de información de Kullback-Leibler (KLIC), para combinar las previsiones de densidad y cuantificar las diferencias entre las densidades previstas y las densidades reales. Sin embargo, el KLIC no es apropiado para HTS ya que requiere que el soporte de una de las densidades consideradas sea el mismo o esté contenido en el soporte de la otra, lo cual es una condición que se incumple con frecuencia en HTS.

Por lo tanto, deben considerarse otras medidas de disimilitud. Este artículo va a considerar las distancias de (Mallows, 1972). Dadas dos distribuciones de densidad,  $f(x)$  y  $g(x)$  comprendidas en  $\mathbb{R}$ , las distancias de Wasserstein (o Kantorovich) y de Mallows entre  $f(x)$  y  $g(x)$  se definen como:

$$D_W(f, g) = \int_0^1 |F^{-1}(t) - G^{-1}(t)| dt$$

$$D_M(f, g) = \int_0^1 (F^{-1}(t) - G^{-1}(t))^2 dt$$

donde  $F^{-1}(t)$  y  $G^{-1}(t)$  están  $\in [0,1]$  y las funciones de distribución cumulativas inversas son  $f(x)$  y  $g(x)$ , respectivamente.

La distancia entre dos histogramas se puede desglosar de la siguiente forma (Irpino, 2010):

$$D_M(X, Y) = \int_0^1 (F^{-1}(t) - G^{-1}(t))^2 dt =$$

$$(\overline{h_{X_t}} - \overline{h_{Y_t}})^2 + (\sigma h_{X_t} - \sigma h_{Y_t})^2 + 2\sigma h_{X_t} \sigma h_{Y_t} (1 - \varphi(h_{X_t}, h_{Y_t}))$$

El primer término hace referencia a la diferencia de localización entre los dos histogramas (diferencia de sus medias). Con el segundo término se obtiene la diferencia entre los tamaños (diferencia entre sus desviaciones típicas) y en este último término,  $2\sigma h_{X_t} \sigma h_{Y_t} (1 - \varphi(h_{X_t}, h_{Y_t}))$  se obtiene la diferencia entre las formas de los dos histogramas que se están evaluando,  $h_{X_t}$  y  $h_{Y_t}$ . La suma de estos tres términos da como resultado la distancia total entre los dos histogramas.

Cabe destacar que si ambos histogramas tienen la misma forma, entonces la distancia quedará reducida a esta ecuación:

$$D_M(X, Y) = (\overline{h_{X_t}} - \overline{h_{Y_t}})^2 + (\sigma h_{X_t} - \sigma h_{Y_t})^2$$

De la misma forma, si los histogramas tienen la misma forma y tamaño, la ecuación a resolver para obtener la distancia será:

$$D_M(X, Y) = (\overline{h_{X_t}} - \overline{h_{Y_t}})^2$$

Las ecuaciones se han propuesto para calcular la distancia entre dos histogramas, pero podría calcularse la distancia de dos series temporales de histograma  $\{h_{X_t}\}$  y  $\{h_{Y_t}\}$  únicamente calculando la distancia entre sus respectivos histogramas dos a dos y haciendo una media.

(Arroyo et al., 2008) propuso la siguiente medida de distancia. Siendo  $\{h_{X_t}\}$  la serie temporal de histograma observada y  $\{\hat{h}_{X_t}\}$  su correspondiente predicción donde  $t=1, 2, \dots, n$ , entonces el error medio de distancia (MDE por sus siglas en inglés) está representado por la siguiente ecuación:

$$MDE(\{h_{x_t}\}, \{\hat{h}_{x_t}\}) = \frac{1}{n} \sum_{t=1}^n D(h_{x_t}, \hat{h}_{x_t})$$

donde  $D(h_{x_t}, \hat{h}_{x_t})$  es la medida de Wasserstein y Mallows explicada anteriormente. Como las distancias no pueden ser negativas, el MDE simplemente las promedia en el tiempo.

## Capítulo 7. CAUSALIDAD DE GRANGER EN HTS

### *LA CAUSALIDAD DE GRANGER BAJO EL ENFOQUE DE SERIES TEMPORALES DE HISTOGRAMA. EL CASO DEL ORO Y EL PETRÓLEO ANTES, DURANTE Y DESPUÉS DE LA CRISIS DE LA PANDEMIA DEL COVID-19*

En este apartado se va a llevar a cabo un análisis en el que se estudia la causalidad de Granger en las series temporales de histograma (HTS) de los precios del oro y del petróleo. Para ello, al igual que en el caso anterior de series temporales de intervalo (ITS), se han tenido en cuenta datos semanales recogidos en semestres. El estudio se ha realizado con datos de antes, durante y después de la pandemia del Covid-19. Este estudio será similar al ya realizado, con el objetivo de contrastar las conclusiones que se obtienen en ambos casos (intervalos e histogramas).

Antes de explicar el análisis que se ha llevado a cabo, se muestran algunas figuras de ejemplo con la representación gráfica de los histogramas del oro y del petróleo, con sus respectivas series temporales. Se han utilizado gráficos de caja (*boxplots*) para la representación, ya que en ellos se visualizan de una mejor manera los cuartiles, que utilizaremos a continuación para el análisis.

La Figura 27 muestra un diagrama de caja o gráfico de cajas y bigotes de los precios del oro, entre los años 2018 y el primer semestre del 2022. La visualización se ha realizado a partir de los datos semanales de precios incluidos en ese periodo de tiempo.

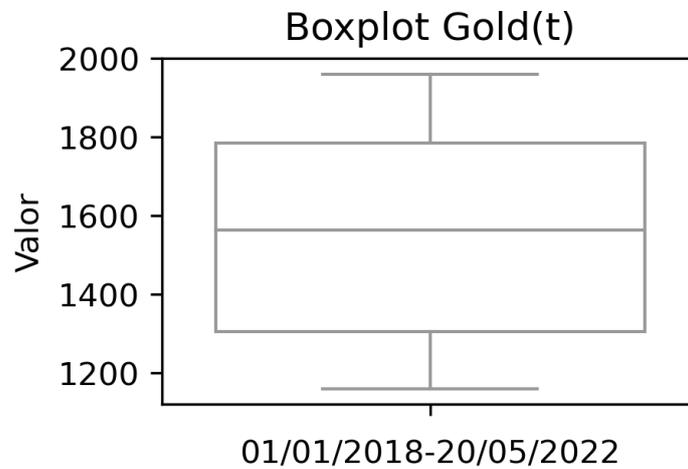


Figura 27. Gráfico de caja de la distribución de precios del oro (01/01/2018-20/02/2022)

La Figura 28 muestra un gráfico de caja similar al anterior, con datos obtenidos en el mismo periodo de tiempo y con la misma frecuencia semanal. Sin embargo, este gráfico de caja muestra la distribución de los precios del petróleo.

Se puede observar que la distribución de precios del oro es bastante más dispersa que la del petróleo, reflejando una mayor variabilidad en los precios.

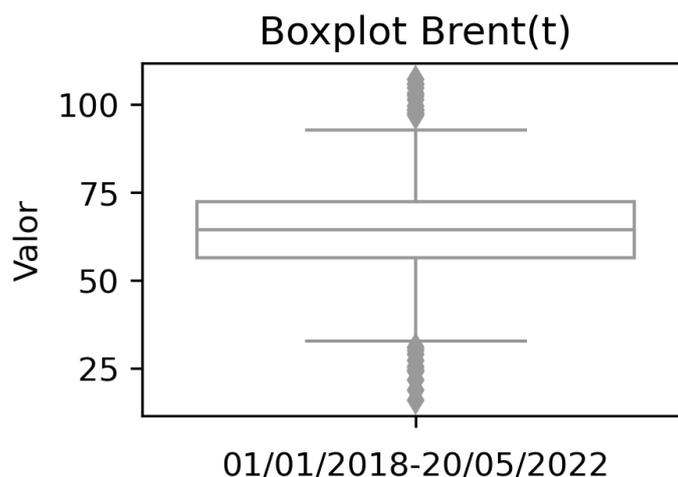
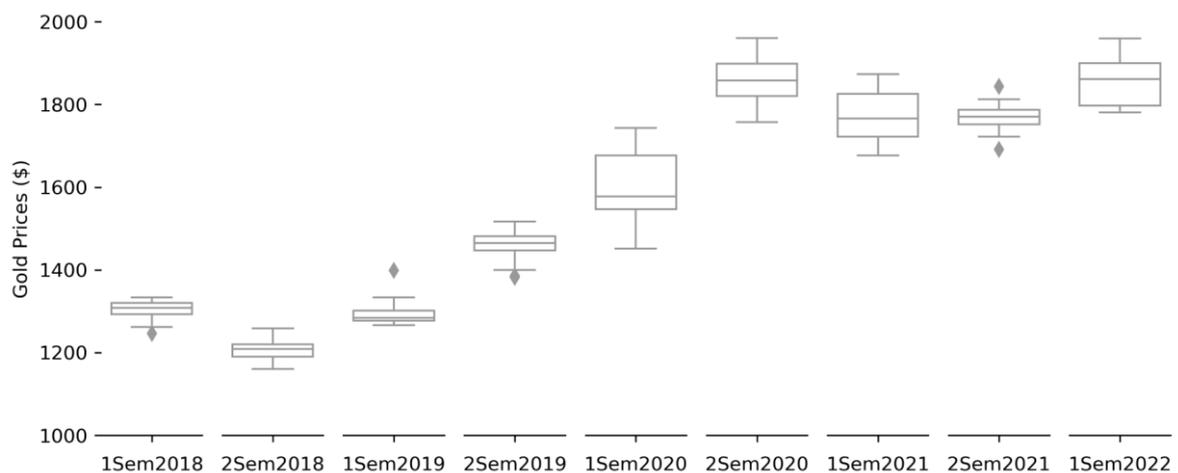


Figura 28. Gráfico de caja de la distribución de precios del petróleo (01/01/2018-20/02/2022)

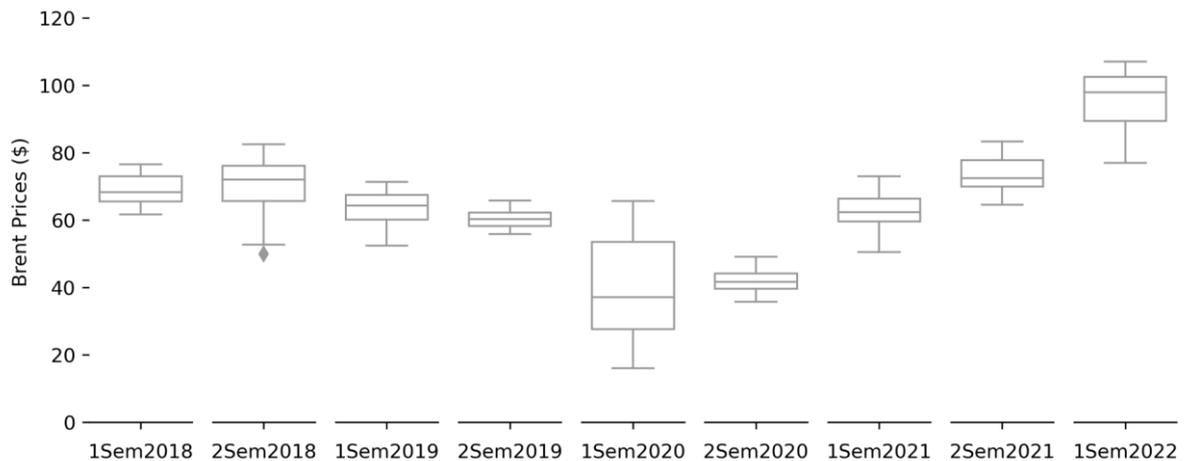
Las dos figuras anteriores (Figura 27 y Figura 28) son ejemplos de dos diagramas de cajas que muestran la distribución de los precios del oro y del petróleo en cierto periodo de tiempo. Sin embargo, a continuación se muestran dos series temporales de histograma de los mismos activos (oro y petróleo), con datos semanales recogidos en semestres a lo largo del tiempo.

Las series temporales podrían ser de datos clásicos, de intervalos (reflejando el mínimo y el máximo) o de histogramas (reflejando toda la distribución de precios en cada semestre). Estas últimas son las más complejas pero también las más completas. En las HTS que se muestran a continuación se pueden visualizar claramente los diferentes cuartiles de las distribuciones gracias a los gráficos de cajas.

Hay que destacar que la pandemia comenzó en el primer semestre de 2020, y es a partir de este momento cuando se empieza a notar un incremento en los precios. Esta subida se visualiza más claramente en la serie temporal de histogramas del oro con respecto a la del petróleo.



*Figura 29. Representación serie temporal de histogramas de precios del oro  
(1º Semestre 2018-1º Semestre 2022)*



*Figura 30. Representación serie temporal de histogramas de precios del petróleo  
(1º Semestre 2018-1º Semestre 2022)*

La tendencia de estas series temporales de histograma es la misma que obteníamos con intervalos. Sin embargo, de esta forma obtenemos mucha más información dado que no solo conocemos dónde se encuentran los valores más extremos de cada conjunto de datos, sino que también obtenemos la información de aquellos tramos en los que obtenemos una mayor cantidad de datos.

### 7.1.1 ANÁLISIS

Se va a estudiar la causalidad de Granger con las series temporales de histograma teniendo en cuenta los cuartiles de cada distribución.

Las siguientes Figuras 31 y 32 muestran los cuartiles principales (25%, 50% y 75%) de cada una de las dos series temporales de histogramas del oro y del petróleo. Los cuartiles son los tres elementos de un conjunto de datos ordenados que dividen el conjunto en cuatro partes iguales en cuanto a frecuencia.

Cabe destacar que cuanto más distancia exista entre los cuartiles significa que más dispersión existe en la distribución de ese semestre en concreto. Sin embargo, si los líneas de los cuartiles se encuentran muy juntas quiere decir que la distribución de los precios en ese periodo de tiempo es muy compacta, habiendo muy poca variabilidad entre sus valores.

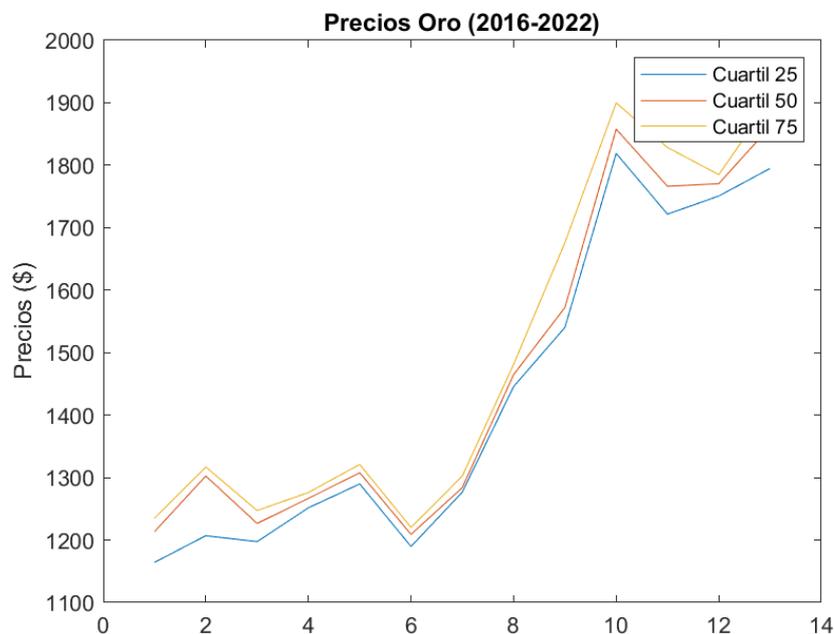


Figura 31. Representación de los cuartiles 25, 50 y 75 de la serie temporal de histogramas del oro (2016-2022)

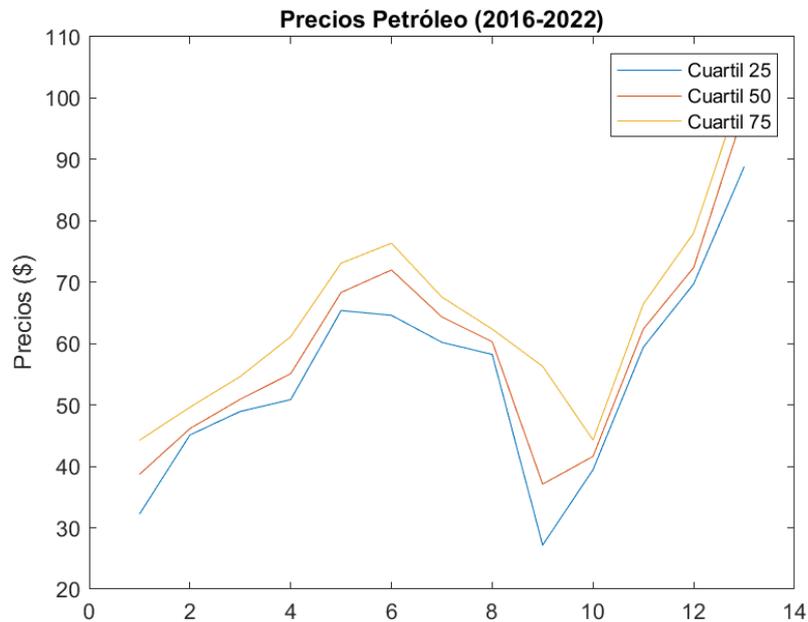


Figura 32. Representación de los cuartiles 25, 50 y 75 de la serie temporal de histogramas del petróleo (2016-2022)

En este sentido, podemos decir que las series temporales de histograma las podemos simplificar, a grandes rasgos, en tres series temporales clásicas. La primera serie temporal recogerá los valores correspondientes al cuartil 25 (Q1) de las distribuciones de datos. La segunda serie temporal recogerá los valores de las medianas, cuartil 50 (Q2), de cada una de las distribuciones. En último lugar, se recogerá una serie temporal clásica con los valores del tercer cuartil, cuartil 75 (Q3).

Es cierto que los tres cuartiles principales de la distribución no son un reflejo completo del comportamiento de la distribución, pero si recogen gran parte de la información.

Una vez tenemos cada una de las dos series temporales de histogramas reducidas a series temporales clásicas con sus tres cuartiles, procedemos a realizar el análisis de causalidad de Granger a lo largo del tiempo.

Este análisis se va realizar con los tests de causalidad de Granger que incorpora el lenguaje de programación MATLAB, tal y como llevamos haciendo a lo largo de este proyecto.

Los pasos para realizar este análisis se describen a continuación. En primer lugar, se selecciona el periodo de tiempo del que se quieren extraer resultados y posteriores conclusiones. Se descargan los datos que hay en ese periodo de tiempo. Los datos que se han empleado en este TFG han sido descargados de Bloomberg, una compañía estadounidense de asesoría financiera, software, data y media bursátil.

Una vez se tienen los datos en la frecuencia deseada, en este análisis, datos semanales, se realiza en MATLAB la serie temporal de histogramas de los datos, creando un gráfico de caja/histograma/distribución para cada uno de los semestres. De esta manera, tal y como se puede observar en las Figuras 29 y 30, se puede visualizar la evolución de la distribución de precios a lo largo de diferentes semestres.

Acto seguido, se procede a dividir la HTS en tres series temporales clásicas, correspondientes a los cuartiles. En último lugar, se realizan los tests de causalidad de Granger para cada una de las series temporales de cuartiles. Es en este caso cuando obtendremos los resultados de causalidad o no causalidad entre las series temporales de precios entre el oro y el petróleo.

## 7.1.2 RESULTADOS

En este apartado se muestran los resultados obtenidos con las funciones de MATLAB de causalidad de Granger, después de haber descrito el proceso mencionado anteriormente.

Se analizan las series temporales de histograma de dos periodos de tiempo. En primer lugar, se tiene en cuenta un periodo de tiempo más amplio, de aproximadamente 12 años. Se mostrarán los resultados obtenidos de la causalidad de Granger en los diferentes cuartiles de la distribución.

Después se analizará la HTS de los años de antes, durante y después de la crisis de la pandemia del Covid-19. Se mostrarán también los resultados obtenidos en los cuartiles.

### 7.1.2.1 1º Semestre 2010 - 1º Semestre 2022

	<b>CUARTIL 25 (Q1)</b>	<b>CUARTIL 50 (Q2)</b>	<b>CUARTIL 75 (Q3)</b>
<b>Brent -&gt; Gold</b>	Causa	No Causa	No Causa
<b>Gold -&gt; Brent</b>	No Causa	No Causa	No Causa

*Tabla 10. Resultados Causalidad de Granger HTS del oro y el petróleo (2010-2022)*

### 7.1.2.2 1º Semestre 2020 - 1º Semestre 2022

	<b>CUARTIL 25 (Q1)</b>	<b>CUARTIL 50 (Q2)</b>	<b>CUARTIL 75 (Q3)</b>
<b>Brent -&gt; Gold</b>	Causa	Causa	No Causa
<b>Gold -&gt; Brent</b>	Causa	Causa	No Causa

*Tabla 11. Resultados Causalidad de Granger HTS del oro y el petróleo (2019-2022)*

En los resultados anteriores podemos observar que, en el análisis realizado en un periodo de tiempo amplio, prácticamente no apreciamos causalidad en ninguno de los sentidos. Solo hemos encontrado causalidad en las series temporales clásicas del primer cuartil y únicamente del petróleo hacia el oro, pero este resultado no es nada significativo en comparación con todo el conjunto.

Sin embargo, si reducimos este periodo de tiempo a los años de la pandemia, aproximadamente los dos años más recientes, encontramos resultados diferentes. En este caso sí que obtenemos resultados de causalidad de Granger entre las dos series temporales del oro y del petróleo. Concretamente, la causalidad viene determinada por los dos primeros cuartiles de la distribución de los datos. Aunque el tercer cuartil no muestre resultados de causalidad, podemos afirmar que la causalidad es bastante robusta al darse en los otros dos cuartiles de la distribución.

Estos resultados no son sorprendentes ya que, al igual que habíamos observado en el estudio de la causalidad de Granger con ITS, la causalidad es mucho más fuerte en aquellos momentos en los que se haya experimentado una crisis. Es decir, en este análisis no encontramos causalidad cuando estudiamos un periodo extenso en el tiempo, en el que ha habido momentos de auge y momentos de crisis, sin embargo la causalidad es mucho más fuerte en cuanto acortamos el estudio a un periodo de tiempo menor, en el que nos encontrábamos sometidos a una grave crisis económica mundial por causa de la pandemia.

Además, podemos observar que, a diferencia del análisis con series temporales de intervalo, la causalidad en este caso es bidireccional. Con este estudio se puede afirmar que en este periodo de crisis, el petróleo causa al oro tal y como habíamos observado con las ITS, pero además se afirma que el oro también causa al petróleo.

Estos resultados tienen sentido dado que las series temporales de histograma ofrecen mayor información que las de intervalo al reflejar la distribución de los datos. Es por ello que, al dar una información más completa del fenómeno en su conjunto, es capaz de apreciar la causalidad en ambos sentidos.

## **Capítulo 8. CONCLUSIONES Y TRABAJOS FUTUROS**

### ***CONCLUSIONES***

Una vez se ha desarrollado el proyecto se proceden a mencionar las conclusiones que se han obtenido.

#### **8.1.1 CONCLUSIONES RESPECTO A LOS OBJETIVOS**

En relación a los objetivos propuestos al principio del trabajo, se puede afirmar que se han conseguido alcanzar todos ellos.

En primer lugar, se ha obtenido un mayor conocimiento de las series temporales tanto clásicas como de intervalo y de histograma.

Además, se ha estudiado en profundidad el concepto de la causalidad de Granger con su posterior aplicación a un ejemplo práctico de dichas series temporales. De esta manera se han perfeccionado los modelos de regresión propuestos para obtener mejores predicciones.

Por último, destacar que todo lo anteriormente mencionado se ha programado en MATLAB, por lo que se ha mejorado el manejo de este lenguaje de programación.

#### **8.1.2 CONCLUSIONES RESPECTO A LOS RESULTADOS**

##### ***8.1.2.1 Series Temporales de Intervalo (ITS)***

En relación con los métodos de regresión lineal múltiple descritos y utilizados en este trabajo, la primera conclusión que obtenemos es sobre las medidas de precisión.

En el estudio se ha llevado a cabo un análisis amplio en el tiempo (más de 10 años) y se concluye que el método de regresión que mejores métricas presenta es el modelo MG. Estos

resultados tienen sentido ya que es el modelo más reciente y complejo y tiene en cuenta todo el intervalo en su conjunto para realizar la predicción, al contrario que los métodos anteriores que únicamente consideran las características del intervalo (mínimo, máximo, centro y/o radio) de manera independiente. Como conclusión, aunque todo el análisis de ITS se ha realizado teniendo en cuenta los cinco métodos de regresión, cabe destacar que el modelo MG es el que presenta resultados más precisos.

Por otra parte, después de estudiar la causalidad de Granger en las ITS del oro y del petróleo a lo largo del tiempo se ha concluido lo siguiente.

En primer lugar, la relación causal es unidireccional en todo momento, siendo siempre esta del petróleo hacia el oro. Además, podemos afirmar que si estudiamos un largo periodo de tiempo (30/04/2010 - 30/10/2021), la relación de causalidad existente es muy débil, prácticamente nula.

Sin embargo, después de estudiar dicha causalidad en un periodo más corto (04/01/2019 – 20/05/2022), se pueden observar resultados más concretos. Es importante destacar que la causalidad es mucho más fuerte en aquellos momentos en los que la crisis de la pandemia se ha visto más pronunciada, debilitándose en el momento de la recuperación económica. Gracias a este estudio se puede concluir que la causalidad de Granger está ligada al ciclo económico, siendo más fuerte en momentos de “valle” y más débil en los “picos” o épocas de auge económico.

#### ***8.1.2.2 Series Temporales de Histograma (HTS)***

El mismo estudio que se ha realizado con series temporales de intervalo se ha replicado con las series temporales de histograma.

Los resultados que se han obtenido ayudan a reafirmar la conclusión obtenida anteriormente con las ITS de que la causalidad de Granger es mucho más fuerte en aquellos momentos “valle”, es decir, de crisis económica. Se ha estudiado la causalidad de Granger en un periodo extenso en el tiempo (más de 10 años), que como es evidente, incluía las diferentes partes

del ciclo económico, momentos de auge y prosperidad y momentos de recesión y crisis. Al contemplar este largo periodo de tiempo hemos concluido que no existía causalidad.

Sin embargo, después hemos acotado estos años a los de la pandemia (2020-2022) y los resultados reflejaban causalidad. Esto nos lleva a reafirmar nuestra conclusión anterior: la causalidad de Granger entre las series temporales del oro y el petróleo es mucho más fuerte en momentos de recesión y crisis económica.

Por otra parte, y a diferencia del estudio con ITS, la causalidad de Granger que se ha obtenido es bidireccional, del oro hacia el petróleo y del petróleo hacia el oro. Anteriormente la causalidad que habíamos obtenido estaba solo presente en uno de los sentidos, del petróleo hacia el oro.

Tal y como hemos comentado a lo largo del proyecto, las HTS representan una información más completa de toda la distribución frente a las ITS. Al tener más flexibilidad y adaptabilidad a las circunstancias del mercado, recogen una información más completa del fenómeno en su conjunto, y por tanto es capaz de detectar la causalidad de Granger también en sentido contrario.

Por tanto, como reflexión general, las conclusiones obtenidas en el estudio de la causalidad de Granger en HTS reafirman y complementan a las ya obtenidas en el análisis de la causalidad de Granger con las ITS.

## ***DESARROLLOS FUTUROS***

Finalmente, una vez se ha desarrollado el proyecto al completo y se han mostrado los resultados y conclusiones, se mencionan los posibles desarrollos futuros que se pueden llevar a cabo.

En primer lugar, en relación con el análisis de causalidad realizado con las series temporales de los precios del oro y del petróleo, se pretende hacer una extensión con las criptomonedas. Hoy en día las criptomonedas son de gran relevancia y son comúnmente utilizadas como un activo de inversión. Es por ello que sería muy interesante estudiar la posible relación de diferentes tipos de criptomonedas (Bitcoin, Ethereum, Cardano...) con otros activos financieros: divisas, oro, petróleo, índices, etc. con el objetivo de estudiar su influencia y ser de ayuda para inversores.

Adicionalmente, sería muy interesante escribir un artículo en la revista *Economics Letters* de la mano del tutor del TFG, Carlos Maté. Este artículo ya está pensado y será sobre el estudio de la causalidad de Granger del oro y del petróleo antes, durante y después de la crisis del Covid-19, todo ello bajo el punto de vista de las ITS. El desarrollo y conclusiones de este tema se han aportado en este proyecto, y en el futuro se quiere enviar a la revista dado que se trata de un tema novedoso aplicado a una reciente crisis económica.

En último lugar, como desarrollo futuro de este proyecto sería interesante ampliar los tests de causalidad de Granger a todo el intervalo e histograma en su conjunto. En este proyecto se han realizado las pruebas de causalidad teniendo en cuenta series temporales clásicas obtenidas a partir de las ITS e HTS, contemplando únicamente los aspectos más relevantes de cada una de dichas series temporales (mínimos, máximos, centros y radios en ITS y cuartiles en HTS).

Para el futuro, se plantea la posibilidad de desarrollar un test de causalidad de Granger que tenga en cuenta todo el intervalo y toda la distribución en su conjunto. Este desarrollo puede ser bastante complejo pero seguro que aportará información más precisa sobre el estudio de causalidad de Granger en series temporales de intervalo e histogramas.

## Capítulo 9. BIBLIOGRAFÍA

*Aptech*. Obtenido de <https://www.aptech.com/blog/introduction-to-the-fundamentals-of-vector-autoregressive-models/>

Arroyo, J., & Maté, C. (2009). Forecasting histogram time series with k-nearest neighbours methods. *International Journal of Forecasting*, 25(1), 192-207.

Arroyo, J., Maté, C., Munoz San Roque, A., & Sarabia, A. (2008). Exponential smoothing methods for histogram time series based on histogram arithmetic. *Madrid: Universidad Complutense de Madrid*.

Balcilar, M., Ozdemir, Z. A., & Shahbaz, M. (2019). On the time-varying links between brent and gold: New insights from the rolling and recursive rolling approaches. *International Journal of Finance & Economics*, 24(3), 1047-1065.

*Banco Mundial*. (2022). Obtenido de Perspectivas Económicas Mundiales: <https://www.bancomundial.org/es/publication/global-economic-prospects>.

Billard, L., & Diday, E. (2000). Regression analysis for interval-valued data. In *Data analysis, classification, and related methods* (pp. 369-374). Springer, Berlin, Heidelberg.

Blanco-Fernández, A., Corral, N., & González-Rodríguez, G. (2011). Estimation of a flexible simple linear model for interval data based on set arithmetic. *Computational Statistics & Data Analysis*, 55(9), 2568-2578.

Chacón, J. E., & Rodríguez, O. (2021). Regression models for symbolic interval-valued variables. *Entropy*, 23(4), 429.

Diebold, F. X., Gunther, T. A., & Tay, A. S. (1998). Reevaluating density forecasts with applications to financial risk management. *International Economic Review*, vol. 39, no. 4.

In *Symposium on Forecasting and Empirical Methods in Macroeconomics and Finance* (p. 863).

Engle, R. F., & Russell, J. R. (2004). Analysis of high frequency and transaction data. *forthcoming in LP Hansen and Y. Ait-Sahalia (eds.), Handbook of Financial Econometrics. Amsterdam: North-Holland.*

Gallardo, J. A. (2008). Métodos de predicción para series temporales de intervalos e histogramas. *Departamento de Organización Industrial, ICAI, Universidad Pontificia Comillas, 41.*

García-Bárzana, M., Ramos-Guajardo, A. B., Colubi, A., & Kontoghiorghes, E. J. (2020). Multiple linear regression models for random intervals: a set arithmetic approach. *Computational Statistics, 35(2), 755-773.*

Hall, S. G., & Mitchell, J. (2007). Combining density forecasts. *International Journal of Forecasting, 23(1), 1-13.*

*Instituto Nacional de Estadística* . (2022). Obtenido de [www.epdata.es](http://www.epdata.es)

Jang, S. M., Yi, E., Kim, W. C., & Ahn, K. (2019). Information flow between Bitcoin and other investment assets. *Entropy, 21(11), 1116.*

Kanjilal, K., & Ghosh, S. (2017). Dynamics of crude brent and gold price post 2008 global financial crisis–New evidence from threshold vector error-correction model. *Resources Policy, 52, 358-365.*

Maceda Dal-Re, P. (2018). Analytics system with interval-valued data.

Mallows, C. L. (1972). A note on asymptotic joint normality. *The Annals of Mathematical Statistics, 508-515.*

Maté, C. (2012). El análisis de intervalos. Aplicaciones en ingeniería. In *Anales de mecánica y electricidad* (Vol. 89, pp. 20-27).

Maté, C. G. (2021). Combining Interval Time Series Forecasts. A First Step in a Long Way (Research Agenda). *Revista Colombiana de Estadística*, 44(1), 123-157.

Maté, C. (2022). *Interval Returns*. Madrid.

Neto, E. D. A. L., & De Carvalho, F. D. A. (2008). Centre and range method for fitting a linear regression model to symbolic interval data. *Computational Statistics & Data Analysis*, 52(3), 1500-1515.

Rencher, A. C. (2005). A review of “Methods of Multivariate Analysis”.

Santos Jiménez, C. d. (2021). TFG Grado de Ingeniería de Telecomunicación: Predicción de intervalos de precios en los mercados de valores. Comparación entre paradigma de datos clásicos y datos de intervalo.

Troster, V. (2018). Testing for Granger-causality in quantiles. *Econometric Reviews*, 37(8), 850-866.

Verde, R., & Irpino, A. (2010). Ordinary least squares for histogram data based on Wasserstein distance. In *Proceedings of COMPSTAT'2010* (pp. 581-588). Physica-Verlag HD.

Zhang, Y. J., & Wei, Y. M. (2010). The crude brent market and the gold market: Evidence for cointegration, causality and price discovery. *Resources Policy*, 35(3), 168-177.

## **ANEXO I: ALINEACIÓN DEL PROYECTO CON LOS ODS**

Los Objetivos de Desarrollo Sostenible (ODS) u Objetivos Globales son 17 objetivos interconectados diseñados para ser un plan para lograr un futuro mejor y más sostenible para todos. Fueron elaborados en 2015 por los jefes de Gobierno y de Estado 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 con este año límite.

La idea principal que persiguen estos objetivos es que, independientemente de su nivel de riqueza y desarrollo, cada país deberá ser 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.

Atendiendo a los objetivos de desarrollo sostenible promulgados por la ONU, este trabajo se enmarca principalmente en el contexto de tres de ellos, los números 4, 8 y 10, que se explican más a detalle a continuación:

### ***OBJETIVO 4: GARANTIZAR UNA EDUCACIÓN INCLUSIVA, EQUITATIVA Y DE CALIDAD Y PROMOVER OPORTUNIDADES DE APRENDIZAJE DURANTE TODA LA VIDA PARA TODOS:***

Este objetivo se centra en la educación de todas las personas, fundamentando que la educación permite una movilidad socioeconómica ascendente y es esencial para el desarrollo económico de los países, permitiendo salir de situaciones de pobreza.

Este Trabajo de Fin de Grado pretende colaborar con este objetivo en la medida en que proporciona herramientas útiles para mejorar la predicción de series temporales, aportando el concepto de la causalidad de Granger. Además, aporta diversos conceptos estadísticos que pueden resultar muy interesantes.

Hoy en día, muchas personas invierten gran parte de su dinero con la ilusión de duplicar esta cantidad pasados un años, arriesgando todos sus ahorros y terminando por arruinarse. Esto ocurre por el afán de conseguir un ascenso económico y por tener poca formación al respecto. Este estudio pretende ser de gran ayuda para colaborar en la educación financiera de todas estas personas, tratando de mejorar las previsiones económicas y minimizando su riesgo.

***OBJETIVO 8: PROMOVER EL CRECIMIENTO ECONÓMICO INCLUSIVO Y SOSTENIBLE, EL EMPLEO Y EL TRABAJO DECENTE PARA TODOS:***

Este objetivo está íntegramente relacionado con el contenido de este estudio. La idea principal consiste en que un crecimiento económico inclusivo y sostenido puede impulsar el progreso, crear empleos decentes para todos y mejorar los estándares de vida.

Este trabajo ofrece una serie de métodos de regresión para obtener predicciones futuras de distintos activos e instrumentos financieros. Además, incorpora el concepto de la causalidad de Granger para mejorar dichas predicciones. Muchas personas deciden invertir en bolsa parte de su dinero, y está demostrado que más del 90% de las personas que deciden invertir su dinero acaban perdiéndolo. Este proyecto puede ser utilizado para mejorar las estrategias de inversión de las personas con el fin de mejorar el rendimiento financiero y disminuir el riesgo asociado.

Seguramente las inversiones en bolsa no sean la principal fuente de ingresos del individuo medio, pero todo lo descrito a lo largo de este proyecto puede servir como ayuda para formarse y recibir de esta manera un suplemento económico adicional, favoreciendo así al crecimiento económico.

***OBJETIVO 10: REDUCIR LA DESIGUALDAD EN Y ENTRE TODOS LOS PAÍSES:***

La desigualdad dentro de los países y entre estos es un continuo motivo de preocupación, de hecho, la reciente pandemia del Covid-19 ha intensificado las desigualdades existentes y ha

afectado más que nadie a los pobres y las comunidades más vulnerables. Además, en el frente económico, la crisis provocada por el Covid-19 ha aumentado significativamente el desempleo mundial y ha recortado drásticamente los ingresos de los trabajadores.

En este sentido el proyecto también puede servir como aportación a este objetivo, ayudando a actuar a las personas con más seguridad y certeza en el mundo financiero, sobre todo a aquellas que hayan visto reducidos sus ingresos en estos últimos años. Este proyecto reduce las diferencias y oportunidades de inversión entre aquellos que han tenido la suerte de formarse y optar por educación de calidad, y aquellos que no han sido educados y desconocen las diferentes teorías y modelos financieros.