



Facultad de Ciencias Económicas y Empresariales.

Análisis Fractal, ARFIMA y PCA: Aproximación a los mercados de futuros del petróleo.

Autor: Ángel González de Vega.

Tutor: José Portela González.

MADRID | Abril, 2023

Contenido

Índice de figuras y tablas.....	3
Resumen	5
Abstract.....	6
1. Introducción	7
1.1. Motivación	7
1.2. Objetivos.....	7
1.3. Metodología.....	8
2. Estado del Arte.....	10
2.1. Aproximación teórica a los mercados de futuros del petróleo.....	10
2.2. Aproximación al análisis y trading con futuros del petróleo.....	13
2.3. Aproximación teórica a las técnicas de Machine Learning utilizadas.....	14
3. Datos y metodología.....	18
3.1. Dataset utilizado:	18
3.2. Construcción de las curvas.....	19
3.3. Construcción de los spreads.....	21
4. Principales resultados.....	27
4.1. Estudio de las curvas: Análisis de componentes principales.....	33
4.2. Modelización con ARFIMA de los <i>spreads</i> de futuros del petróleo.....	43
4.3. Dimensión fractal de los <i>spreads</i>	51
5. Conclusiones.....	56
6. Referencias.....	57
7. Anexo I: Código.....	59

Índice de figuras y tablas.

Ilustración 1- Tipos de curvas según precios de vencimientos

Ilustración 2 - Estructura de las curvas de futuros.

Ilustración 3 - Curvas construidas.

Ilustración 4 - Ejemplo de futuros negociados a 20/11/2022.

Ilustración 5 - Contratos construidos en la elaboración de los *butterflies*.

Ilustración 6 - Contratos utilizados en la elaboración de los *condors*.

Ilustración 7 - Variables que forman la curva de futuros.

Ilustración 8 - Precio por vencimiento del contrato.

Ilustración 9 - Boxplots de los *spreads*.

Ilustración 10 - *Butterflies* de 3 meses.

Ilustración 11 - *Butterflies* de 6 meses

Ilustración 12 - *Butterflies* de 12 meses

Ilustración 13 - *Butterflies*: 2 contratos de 6 meses y un contrato de 12 meses.

Ilustración 14 - *Butterflies*: 2 contratos de 3 meses y un contrato de 6 meses

Ilustración 15 - 2 *condors* de 6 meses y uno de 12 meses.

Ilustración 16 - *Condors* de 12 meses.

Ilustración 17 - *Condors* de 6 meses.

Ilustración 18 - Correlación entre las diferentes posiciones de la curva.

Ilustración 19 - Importancia de cada componente principal.

Ilustración 20 - Correlación entre componentes principales.

Ilustración 21 - Loading vectors del análisis PCA.

Ilustración 22 - Proporción de la varianza explicada.

Ilustración 23 - Contribución de las variables a la componente principal.

Ilustración 24 - Biplot con las dos primeras componentes principales.

Ilustración 25 - Observaciones y componentes principales.

Ilustración 26 - Efectos del cambio de vencimiento.

Ilustración 27 - *Frames* de los gráficos animados.

Ilustración 28 - Gráfico original (izquierda) y reconstruido (derecha)

Ilustración 29 - Reconstrucción (línea roja) de la primera y última curva.

Ilustración 30 - Reconstrucción de *spreads* de una estrategia simple (izquierda) y una compleja (derecha).

Ilustración 31 - Reconstrucción del *spread* M2Z2M3 (rojo).

Ilustración 32 - Reconstrucción con dos y cuatro componentes principales.

Ilustración 33 - Reconstrucción con dos y cuatro componentes principales.

Ilustración 34 - Gráficos y FAS de las estrategias simples

Ilustración 35 - Modelos

Ilustración 36 - FAS y FAP de la serie y de la serie simulada

Ilustración 37 - Modelización a través de los parámetros obtenidos.

Ilustración 38 - Serie temporal y FAS

Ilustración 39 - Modelización ARIMA

Ilustración 40 - FAS y FAP de la serie y de la serie simulada

Ilustración 41 - Simulación basada en los parámetros

Ilustración 42 - Serie temporal y coeficiente de los *butterflies* de tres meses

Ilustración 43 – Series temporales y coeficiente de los *butterflies* de 6 meses

Ilustración 44 – Series temporales y coeficiente de los *butterflies* de 12 meses

Ilustración 45 - Series temporales y coeficiente de los 2 *butterflies* de 3 meses y un *butterfly* de 6 meses

Ilustración 46 -Series temporales y coeficiente de los 2 *butterflies* de 6 meses y un *butterfly* de 12 meses

Ilustración 47 - Series temporales y coeficiente de los *condors* de 6 meses

Ilustración 48 - Series temporales y coeficiente de los *condors* de 12 meses

Ilustración 49 - Series temporales y coeficientes de los dos *condors* de 6 meses y un *condor* de 12 meses

Resumen.

El petróleo tiene una importancia estratégica y fundamental en el mundo actual. No solo por la importancia estratégica en términos de países productores, sino que también por las funciones que tiene en diferentes campos, como el industrial o el energético, o las características propias. Por ello, la inversión en este producto ha crecido de manera exponencial y, por lo tanto, las técnicas para intentar predecir su precio o modelar su comportamiento también se han visto aumentadas. En consecuencia, este Trabajo de Fin de grado presenta un análisis de los mercados de futuros del petróleo usando el análisis de componentes principales, los modelos ARFIMA y el análisis fractal. Se ha examinado la relación entre las curvas de futuros del petróleo y los *spreads*, proporcionando además una introducción a las técnicas utilizadas, revisando la literatura existente al respecto en diferentes aspectos: Los futuros del petróleo y las técnicas de *machine learning*.

Además, se muestran también los resultados obtenidos en los distintos análisis, obteniendo que el análisis PCA es útil para determinar la estructura de los mercados, que los modelos ARFIMA son efectivos para modelar las dinámicas y que el análisis fractal puede proporcionar información sobre la naturaleza, en algunos sentidos, caótica de los mercados.

Palabras clave: Petróleo, Futuros del petróleo, derivados, *spreads*, análisis fractal, ARFIMA, análisis de componentes principales.

Abstract.

Oil has an strategic and fundamental significance nowadays. Its importance comes from not only its strategic value in terms of supply countries, but also from the different purposes it has in the industrial or the energetic field. Therefore, the investment in this commodity has increasingly grown up, so the techniques to predict price or modelling its behavior have risen too. Consequently, this dissertation presents an analysis of the oil futures markets using Principal Component Analysis (PCA), ARFIMA models and fractal analysis. The relationship among oil futures curves and spreads is examined, giving, in addition, an introduction of the techniques used and checking the existing state-of-the-art in terms of oil futures and machine learning techniques.

Furthermore, the results obtained in the different analysis are shown. The principal consequences demonstrate us that PCA is useful to determine the structure in the markets, that ARFIMA models can effectively model the different dynamics and that fractal analysis can provide us information about the chaotic nature of the markets.

Key words: Oil, oil futures, derivatives, spreads, fractal analysis, ARFIMA, principal component analysis.

1. Introducción.

1.1. Motivación.

Las materias primas, últimamente, han cobrado mucho protagonismo como activo de inversión, convirtiéndose en una parte más de las carteras de muchos inversores, ya sean personales o institucionales. Además, las materias primas cobran especial relevancia actualmente, debido a la guerra existente entre Rusia y Ucrania y al escenario de elevada inflación en el que nos encontramos.

Dentro de estas materias primas, las hay que se negocian en mercado con mucha liquidez, donde los volúmenes de contratación son muy elevados. Este crecimiento viene dado por la formación de precios en el mercado del producto físico y por el aumento de participantes en los mercados de futuros:

- Por un lado, la importancia histórica de los mercados de futuros como formadores de precio ha crecido, haciendo que otros tipos de contratos de abastecimiento estén siendo sustituidos por futuros.
- Por el otro, el aumento de participantes en el mercado, que hace que se asuman riesgos y permiten el aumento de liquidez en el mercado.

Entre estos mercados encontramos a los mercados de futuros del petróleo. Debido a la liquidez de este mercado, los volúmenes de negociación existentes y la importancia estratégica del petróleo, es clave realizar una aproximación a este tipo de mercados, su evolución a lo largo del tiempo y las técnicas de inversión que se han empleado en el mismo. Además, entre estas técnicas de inversión, y debido al aumento del uso de las nuevas tecnologías y el Big Data por parte de fondos y gestoras, se han ido empleando diferentes técnicas estadísticas y de Machine Learning que han tenido relevancia en los mercados de futuros del petróleo. Es por ello que, desde este Trabajo de Fin de Grado, se va a llevar a cabo un análisis de los mercados de futuros del petróleo usando técnicas de Machine Learning.

1.2. Objetivos.

El objetivo principal del TFG es, mediante el análisis de los mercados de futuros del petróleo y usando el precio de diferentes contratos y diferentes técnicas de Machine Learning, poder encontrar una forma de modelado de las dinámicas temporales existentes en los mercados de futuros, con el fin de alcanzar estrategias óptimas de inversión en diferentes aspectos:

1. Analizar los spreads de futuros del petróleo, haciendo hincapié en si son eficientes y estudiando la existencia de comportamiento persistente para, en caso de que de que existiese, modelar esta característica.

2. Estudio del modelado de los spreads con el fin de estudiar la gestión de riesgos. Si es posible estudiar las dinámicas que siguen los spreads, podremos aplicar esta dinámica a la gestión de riesgos.
3. Estudiar las curvas de futuros del petróleo mediante el estudio de los factores que lideran el precio del petróleo, investigando estrategias que puedan ofrecernos rentabilidad.

1.3. Metodología.

Vamos a dividir la metodología a utilizar en dos aspectos diferentes: Las técnicas usadas y los datos.

1.3.1. Datos.

Entre los datos empleados, cabe destacar que no se han empleado series temporales del precio de los futuros del petróleo en sí, pero el conjunto de datos empleado sí deriva directamente del precio de estos futuros:

- Curvas de futuros: Usando los precios de cierre de diferentes contratos de futuros del petróleo, se han construido 763 curvas con 11 puntos cada una. Estas 763 curvas corresponden a una curva por cada día de negociación (*trading day*). Los 11 puntos corresponden a diferentes contratos de futuros del petróleo. Más adelante, en el apartado 3 del índice (Datos y metodología), veremos los problemas obtenidos en la construcción de las curvas y la solución de estos, especialmente de aquellos relacionados con el vencimiento y liquidez de los contratos.
- Spreads: Un spread es, básicamente, un diferencial de precios. Es una técnica en la que se abren, simultáneamente, una posición larga y una posición corta para aprovechar este diferencial de precios. El objetivo de los spreads es captar la diferencia de precios existentes entre futuros con diferentes fechas de vencimientos. Podemos encontrar diferentes tipos de spreads, en nuestro caso, nos centraremos en estrategias complejas: *Butterflies* y *Condors*.

1.3.2. Técnicas.

Para cumplir los objetivos anteriormente presentados, las técnicas de Machine Learning que vamos a emplear son las siguientes:

- Análisis fractal: Para el estudio de la persistencia de los *spreads* se va a emplear el coeficiente de Hurst. Este coeficiente es una medida de dependencia de las series temporales. Hay series temporales que siguen un comportamiento de dependencia a largo plazo, es decir, que los eventos de un periodo influyen en los siguientes.

- Modelos ARFIMA: Para la gestión del riesgo y simulación de series temporales, utilizaremos modelos ARFIMA (*Autorregresive fractionally integrated moving average*). Los modelos ARFIMA son diferentes modelos que generalizan los modelos ARIMA al permitir valores no enteros en el parámetro de diferenciación d .
- Análisis de componentes principales (PCA): Para el estudio de las curvas de futuros del petróleo vamos a utilizar el análisis PCA. Este método consiste en la simplificación de un conjunto de datos con diferentes dimensiones y donde existe el problema de la multicolinealidad, donde la relación entre variables es fuerte.

2. Estado del Arte.

2.1. Aproximación teórica a los mercados de futuros del petróleo.

Los futuros son derivados financieros. Un derivado es un producto cuyo valor deriva del precio de otro producto, denominado activo subyacente. El activo subyacente puede ser cualquier instrumento financiero: Una acción, un índice bursátil, materias primas... Son productos que están sujetos al efecto apalancamiento, ya que la inversión inicial es muy menor en comparación con la exposición al subyacente, y los retornos pueden multiplicarse en relación con la inversión inicial, ya sean positivos o negativos.

Los derivados pueden negociarse en mercados organizados o no organizados. La principal diferencia que existe entre los mercados es la estandarización existente en los mercados organizados. Dentro de estos mercados nos interesan especialmente un tipo de derivado: Los futuros.

Los futuros son contratos por los que se acuerda el intercambio de una cantidad concreta de un activo subyacente en una fecha determinada, a un precio determinado, al que llamaremos *strike*. En los contratos de futuros podemos tomar dos tipos de posiciones:

- Posición larga: Es la posición adoptada por el comprador. En el vencimiento del contrato, el comprador tiene derecho a recibir el activo subyacente, en caso de que se liquidara por entrega física.
- Posición corta: Es la posición adoptada por el vendedor, que se compromete a entregar el activo subyacente en caso de liquidación por entrega física.

Los futuros se negocian en mercados bursátiles. Entre los beneficios encontrados en estos mercados podemos destacar la estandarización o la liquidez. Entre los mercados de futuros a destacar, encontramos el CME o el NYSE Euronext.

Anteriormente comentábamos que podría haber futuros con distintos tipos de activo subyacente. En este caso, nos interesan los futuros del petróleo, concretamente los futuros del Brent.

Futuros del petróleo: Brent.

Uno de los contratos de futuros más importantes son los futuros del petróleo. Dentro de los futuros del petróleo, podemos destacar dos: los futuros del Brent y los futuros del *West Texas Intermediate* (WTI, en adelante). Estos son los dos a destacar porque forman parte de la canasta de referencia de la Organización de Países Productores de Petróleo.

De entre estos dos, nos vamos a centrar en los futuros del Brent. Este tipo de petróleo se extrae de diferentes yacimientos en el Mar del Norte, y sirve de petróleo de referencia en los mercados europeos. Su ligereza lo hace significativamente útil para la producción de dos de los derivados del petróleo más demandados: Gasolina y gasóleo. Los futuros sobre el Brent se han convertido en uno de los productos financieros que más se

negocian en la actualidad, debido a la influencia que tienen en la economía real y la volatilidad que últimamente han venido ofreciendo.

Como hemos venido comentando, los futuros se negocian en mercados regulados. En este sentido, los futuros sobre el Brent se comercializan en el *Intercontinental Exchange* (ICE). Anteriormente decíamos que una de las características de estos mercados es la estandarización, es decir, todos los contratos negociados tienen una serie de características comunes que los ayudan a ser estandarizados. Entre ellas, podemos destacar:

- Símbolo del contrato: B
- Tamaño del contrato: 1.000 barriles por contrato.
- Unidad de negocio: Cualquier múltiplo de 1.000 barriles.
- Divisa: Dólares americanos.
- Fecha de vencimiento: El último día de negociación del contrato es aquel correspondiente al último día laborable del segundo mes que precede el vencimiento. (Por tanto, si el contrato vence en marzo, este se liquida y expira el último día laborable de enero).
- Opciones de liquidación: Efectivo, por diferencia de precios (la diferencia existente entre el precio de compra y el precio al vencimiento del contrato), o en especie.

En estos tipos de contratos la liquidez es muy importante, ya que afecta al precio del futuro y nos indican qué contratos están negociándose, haciendo referencia al volumen. Además, la liquidez afecta al precio, tiene relación con la volatilidad y nos proporciona signos importantes de información (Haugom y Ray, 2017).

Es por ello por lo que, debido a los diferentes usos del petróleo, sus múltiples usos y productos derivados, y su influencia en la vida real, se lleva mucho tiempo queriendo modelarlos y predecirlos, siendo este el objetivo de este TFG.

Curvas:

Una curva de futuros es una serie temporal formada por el conjunto de precios de distintos contratos de futuro con diferente vencimiento. En nuestro caso, tenemos 763 curvas de futuro, una para cada día de negociación existente. Cada curva está formada por 11 puntos, es decir, 11 contratos de distinto vencimiento. Aquí hay dos temas que podemos introducir: *Contango* y *backwardation*.

Estos términos son muy utilizados en mercados de materia primas para indicar la estructura que tiene la curva. Si la curva está en *contango*, los precios de vencimientos más alejados son mayores que los precios de vencimiento más cercano. Por el contrario, si la curva está en *backwardation*, ocurre el suceso opuesto al *contango*, los precios con vencimientos más cercanos son menores que los precios con vencimientos más alejados.

Ilustración 1- Tipos de curvas según precios de vencimientos



En el apartado 3, donde estudiaremos el conjunto de datos, explicaremos las particularidades propias de las curvas y su proceso de creación.

Spreads.

Un spread de futuros es una técnica en la que, simultáneamente, se toman una posición larga y una posición corta sobre un producto. El objetivo es obtener rendimiento de un cambio en precio entre estas dos posiciones.

Hay distintos tipos de spreads, pero nos vamos a centrar en aquellos realizados en el mismo mercado, es decir, los *intra-commodity* spreads. Las diferentes posiciones en un spread son consideradas “patas”. En este sentido, y siguiendo una estrategia sencilla de dos patas, un inversor podría tomar una posición en largo en un contrato de futuros del Brent en marzo, y tomar una posición en corto en un contrato de futuros del Brent en septiembre, negociando ese diferencial de precios entre contratos.

Hay diferentes estrategias que se pueden llevar a cabo con spreads, ya sean simples o complejas. Nos vamos a centrar en estrategias complejas, en las que son conocidas como *butterflies* y *condors*:

- *Butterflies*: Esta técnica se da cuando, simultáneamente, se compran y se venden distintos contratos de futuros en 3 fechas de vencimiento diferentes. Un ejemplo podría ser tomar una posición en largo en junio de 2022, 2 posiciones cortas en septiembre de 2022 y otra posición larga en diciembre de 2022. Es una técnica que se puede emplear cuando se cree que en el corto plazo los precios van a permanecer estables, pero van a subir en el largo plazo. Las diferencias temporales pueden o no ser iguales.
- *Condors*: Esta estrategia es muy parecida a la *butterfly*, pero teniendo 4 patas, es decir, 4 contratos de diferente vencimiento. Al igual que en la anterior estrategia, las diferencias temporales pueden o no ser iguales.

En el apartado 3, donde estudiaremos el conjunto de datos, veremos los distintos *spreads* con los que contamos.

Hemos mencionado anteriormente que los mercados de futuros del petróleo son muy importantes por la importancia estratégica que tiene el producto y es por ello que se han intentado modelar de diferentes maneras. Por tanto, una vez hemos introducido estos mercados, vamos a aproximarnos a diferentes técnicas que se han aplicado con estos productos financieros centrándonos en nuestra metodología.

2.2. Aproximación al análisis y trading con futuros del petróleo.

En este apartado se hace referencia a las publicaciones que se han realizado sobre el Machine Learning en los futuros del petróleo, centrándonos en las técnicas de reversión a la media, análisis de componentes principales y modelos ARFIMA, durante los últimos 10 años. Además, intentaremos también ver que se hace con el *rollover* y los *spreads* de los contratos en dichas publicaciones.

En cuanto a la reversión a la media, Lubnau y Todorova (2015) estudian los *spreads* entre distintos productos con técnicas de reversión a la media y bandas de Bollinger, usando distintos futuros de energía en un periodo de tiempo desde 1992 hasta 2013, concluyendo que hay estrategias en el corto plazo con capacidad de obtener rentabilidad. Además, y aunque esté fuera de nuestro horizonte temporal, cabe destacar que Tabak y Cajueiro (2007) analizaron los mercados del petróleo usando coeficientes de Hurst, encontrando evidencias que el mercado ha ido siendo más eficiente a lo largo del tiempo.

Si hablamos del análisis de componentes principales, Ferreira y Nadkarni (2018) estudiaron cuándo se debía entrar en una posición de mercado haciendo un análisis de componentes principales, comparándolo después con otros tipos de estrategias más simples. Establecieron que es fundamental el análisis PCA para el rendimiento del sistema. Por otro lado, Jacobson (2015) estudió la capacidad de predicción de los mercados de futuros utilizando PCA, estableciendo que podían generarse técnicas de inversión que obtuviesen beneficio, aunque con un Sharpe ratio bajo¹

Usando modelos ARFIMA, Sang y Seong-Min (2013) analizaron y predijeron la volatilidad en los mercados de futuros del petróleo usando modelos ARFIMA, estudiando simultáneamente la volatilidad y la consistencia de las rentabilidades. Cavaliere et al. (2015) estudiaron los modelos ARFIMA aplicándolos a las dinámicas de precio entre los precios de futuros y los precios spot, encontrando que hay modelos parcialmente integrados que apoyan la idea de la eficiencia en los mercados.

¹ El Ratio de Sharpe es una medida de rendimiento, midiendo la rentabilidad de una inversión por unidad de riesgo.

En cuanto al *rollover*, hay autores que solucionan los problemas de vencimiento de los contratos haciendo *rollover*², como Lubnau y Todorova (2015).

Por último, en cuanto a los spreads, hay autores como Bucca y Cummins (2011) que establecen 3 tipos de spreads: *Calendar spreads* (diferencia de precios entre contratos con diferente vencimiento), *Crack spreads* (diferencia de precios entre un producto y sus derivados, como el Brent y la gasolina) y *Locational spreads* (diferencia de precios entre el mismo producto, o un producto muy similar, que se negocia en diferentes áreas geográficas) (Buca y Cummins, 2012). Nosotros nos centraremos en el conocido como *calendar spread*, consistente en el diferencial de precios de dos contratos del mismo producto con distinto vencimiento. Además, estos autores usan técnicas de reversión a la media y encuentran estrategias rentables, con ratios de Sharpe mayores a 2 en muchas ocasiones.

2.3. Aproximación teórica a las técnicas de Machine Learning utilizadas.

Coefficientes de Hurst.

El coeficiente de Hurst fue empleado por primera vez por un hidrólogo británico llamado Harold Edwin Hurst, que estudió la capacidad de almacenamiento de las reservas del río Nilo, documentando la presencia de dependencia de largo alcance en la hidrología. Por ello, en el año 1951 propuso un método para el cálculo de procesos con memoria larga, basándose en un conjunto de fluctuaciones a lo largo del tiempo.

El método propuesto estimaba el coeficiente en el rango $[0,1]$. Dependiendo del valor tomado por el coeficiente, las series temporales tendrían un comportamiento u otro. Por ello, si el coeficiente es $H = 0.5$, los valores actuales de la serie no dependen de valores pasados, es decir, los valores son independientes. Cuando el coeficiente H se encuentra en el rango $[0,0.5)$, la serie temporal demuestra un comportamiento antipersistente. Este tipo de series desarrollan comportamientos de reversión a la media: Si un valor estaba muy alto en periodos anteriores de la serie, los valores posteriores tienen a ir reduciéndose para acercarse a la media. La fortaleza del proceso de reversión a la media es mayor a medida que el coeficiente H se va acercando a 0. Por último, cuando el coeficiente H se encuentra en el rango $(0.5, 1]$, los valores de la serie aumentan y disminuyen siguiendo una dirección, no siguiendo un proceso de ruido blanco. Estas series suelen seguir una tendencia, aunque pueden ser interrumpidas por alguna discontinuidad. A medida que el valor de H se acerca a uno, puede ser más clara la existencia de una tendencia.

² Lo veremos más adelante, pero el rollover hace referencia a la prolongación de un contrato una vez ha llegado el vencimiento comprando el contrato con el vencimiento inmediatamente superior.

La aplicación del coeficiente de Hurst en series temporales financieras se popularizó a finales del siglo pasado, cuando Peters (1991, 1994) introdujo la Hipótesis Fractal de los Mercados (FMH, de sus siglas en inglés *Fractal Market Hypothesis*).

Para llevar a cabo la estimación del coeficiente de Hurst, hay que llevar a cabo un análisis de rango reescalado. Este análisis es el rango de las sumas parciales de desviaciones de la media, que se vuelve a reescalar por sus desviaciones típicas. Los pasos para llevar a cabo este proceso son:

1. Dividimos la serie con un periodo L en un total de N subperiodos. Cada subperiodo s tiene igual tamaño, por lo que $L = L_s * N_s$.
2. Para cada subperiodo, calculamos la media (M_s) y la desviación típica (S_s)
3. Calculamos una serie sin media ($D_{i,s}$) extrayendo la media de cada periodo a la serie de precios original.

$$D_{i,s} = X_{i,s} - M_s$$

4. Calculamos una serie acumulada de cada serie sin media.

$$Y_{i,s} = \sum_{j=1}^i D_{j,s}$$

5. Calculamos el rango de cada subperiodo.

$$R_s = \max(Y_{1,s}, \dots, Y_{L_s,s}) - \min(Y_{1,s}, \dots, Y_{L_s,s})$$

6. Calculamos la media del rango reescalado R_s/S_s para cada subperiodo.

$$(R, S)_{L_s} = \frac{1}{N_s} \sum_{s=1}^{N_s} \frac{R_s}{S_s}$$

Este estadístico sigue la siguiente relación:

$$E[(R/S)_L] = c * L_s^H$$

En la anterior ecuación, c es una constante independiente de L_s y H , que es el coeficiente de Hurst. El coeficiente de Hurst puede estimarse usando una regresión lineal simple sobre una muestra de horizontes temporales que van incrementándose.

Modelos ARFIMA.

Los modelos ARFIMA (del inglés *autorregresive fractionally integrated moving average*) son una derivación de los modelos ARIMA donde el parámetro de diferenciación d es un número no entero.

Un modelo ARIMA(p,d,q) puede ser definido como,

$$\phi_p(B) \nabla^d z_t = \theta_q(B) a_t$$

donde el parámetro “nabla” ∇ es un proceso estacionario, siendo d un número entero positivo, B son los retardos, a_t es un proceso de ruido blanco y los términos ϕ_p y θ_q corresponden a la parte autorregresiva y de media móvil, respectivamente.

Sin embargo, los modelos ARFIMA permiten una variación del orden de diferenciación d en el intervalo $(-0.5, 0.5)$, Es decir, el modelo ARFIMA generaliza los modelos ARIMA permitiendo grados de diferenciación fraccionales.

ARFIMA trata los procesos de memoria larga. Estos procesos de memoria larga son procesos estacionarios cuyas funciones de autocorrelación decaen lentamente. Los modelos ARFIMA nos ofrecen modelos de parametrización de procesos de memoria larga encajando con los modelos ARMA, que son usados para procesos de memoria corta. Los procesos de memoria larga se pueden obtener integrando fraccionalmente procesos de memoria corta, que es lo que hace ARFIMA.

Hemos usado los modelos ARFIMA para estimar la serie temporal del spread a partir de un punto determinado. Estudiando y determinando un modelo ARFIMA para los spreads, hemos usado los parámetros del modelo para estimar la serie temporal a partir de un punto determinado y ver si la dinámica que se sigue es similar a la del spread.

Análisis de Componentes Principales (PCA).

Los spreads de futuros del petróleo son calculados, como hemos visto anteriormente, comprando un contrato en un vencimiento y vendiendo otro con un vencimiento superior. En nuestro caso trabajamos con spreads especiales, como los *butterflies* o los *condors*, que son estrategias más complejas, pero la idea es muy similar. Por lo tanto, el subyacente del spread son los contratos de futuros del petróleo, pudiendo analizar diferentes aspectos: La serie temporal del spread y la serie temporal del subyacente.

Si analizamos la dinámica del subyacente (los contratos de futuros en este caso), podemos modelar el spread. Es por ello que hemos creado diferentes vectores de 11 puntos, formado por 11 contratos de futuros, para modelar esta dinámica. Este modelado se lleva a cabo a través del análisis PCA.

El análisis PCA es un proceso por el cual transformamos un conjunto de variables originales en nuevo conjunto de variables (las componentes principales) que son combinación lineal de las variables de partida. Estas nuevas variables se caracterizan por tener correlación. Además, la información de cada variable se puede ordenar de acuerdo con la información que llevan incorporada.

Por lo tanto, supongamos que tenemos un conjunto de n observaciones y p variables. Llamemos a estas variables X_1, X_2, \dots, X_p . Con este análisis buscamos un conjunto nuevo de variables Z_1, Z_2, \dots, Z_p de tal manera que se maximice la varianza explicada por estas variables y, por tanto, la información. Por ello, cada Z_i será una combinación lineal de las variables originales. Para la primera componente principal tendremos:

$$z_1 = a_{11}X_1 + a_{12}X_{12} + \dots + a_{1p}X_p,$$

Siendo a_{1p} los “*loading vectors*” para la primera componente principal y X_p las variables que teníamos inicialmente. Z_{1p} hace referencia a los scores de la primera componente principal. Un score es una proyección de los datos sobre la dirección definida por el *loading* vector de la primera componente principal.

A la hora de estimar la segunda de las componentes principales, se estiman de igual forma que la primera, pero añadiendo una restricción. Anteriormente decíamos que las nuevas variables tendrían correlación igual a cero, por lo que la restricción añadida es que la dirección de la segunda componente principal sea ortogonal a la dirección de la primera componente.

De la misma forma se van estimando las demás componentes principales, pero teniendo siempre en cuenta que la primera componente recogerá la mayor variabilidad, después se busca entre todas las direcciones ortogonales aquella con la segunda mayor variabilidad y así sucesivamente con las posteriores componentes principales.

La idea de fondo de este análisis es quedarnos con un número de variables menor al número original de variables, pero teniendo en cuenta que esas componentes recojan la mayor varianza posible.

Si partimos de una matriz con los scores de las primeras componentes principales, podemos llegar a reconstruir de forma más o menos aproximada los datos originales. Este es el proceso que vamos a seguir: Vamos a obtener las componentes principales de las curvas de futuros, y usando la matriz de los scores, vamos a intentar reconstruir los spreads utilizando un número menos de variables.

3. Datos y metodología.

3.1. Dataset utilizado:

En este primer apartado, explicaremos los conjuntos de datos utilizados y sus implicaciones básicas, mientras que en apartados posteriores explicaremos cómo han sido calculados estos conjuntos de datos:

Curvas de futuros:

En primer lugar, se ha utilizado un conjunto de datos que contiene los precios de cierre de diferentes contratos de futuros del petróleo. Los contratos que tratamos son los de los meses de marzo, junio, septiembre y diciembre. El motivo del uso de estos contratos es bastante sencillo, ya que son aquellos que tienen más liquidez. El rango temporal incluye desde enero del año 2019 hasta enero del 2022, teniendo datos de 763 días de negociación. Por ello, para el estudio de las dinámicas de las curvas, contamos con 763 curvas diferentes, una para cada día de negociación, Cada curva está formada por 11 puntos, que conforman distintos futuros del petróleo de distintos vencimientos.

Más adelante, cuando veamos el proceso de construcción de las curvas, veremos cómo se han solucionado diferentes problemas que nos han surgido, como la liquidez de los contratos y el vencimiento de estos.

Spreads.

Para tratar de estudiar y/o modelar las dinámicas que tienen los spreads, contamos con un conjunto de datos que tiene 28 spreads diferentes. Entre estos spreads existen diferentes tipos de estrategias, algunas más simples y otras más complejas, para poder determinar diferentes dinámicas que pueden llegar a existir en el mercado.

En cuanto a fechas, contamos con datos de 537 días de negociación, siendo el tamaño de las series temporales, por tanto, de 537 puntos. El inicio de la serie temporal es el día 02/01/2020, y termina el día 28/01/2022.

Cuando tratamos con los datos de futuros hay que tener muy en cuenta un aspecto fundamental de estos: El vencimiento. Por ello, y para tener diferentes spreads que tengan el mismo número de puntos, el primer contrato de futuros con el que contamos para el cálculo de los spreads tiene un vencimiento posterior al último día de negociación del que tenemos datos.

Ambos conjuntos de datos son de elaboración propia, cuya elaboración relato a continuación. Sin embargo, se han construido a través de un conjunto de datos inicial que contiene los precios de distintos contratos de futuros del petróleo. Este conjunto de datos ha sido obtenido de un profesional en el campo del *trading* con futuros del petróleo. La colaboración que se llevó a cabo se realizó en el contexto de la *Research Community* de la Universidad Pontificia Comillas, en la que se trabajaron diferentes aspectos relacionados con los futuros del petróleo y el mundo que les rodea. En ella se realizó un estudio sobre el *Machine Learning* aplicado a estos mercados, realizando un

estudio en el que se ahondaba sobre las diferentes técnicas aplicadas, los resultados obtenidos, las conclusiones y las implicaciones existentes para posteriores estudios.

3.2. Construcción de las curvas.

En este apartado vamos a exponer cómo se han construido las curvas de futuros del petróleo. Una curva está formada por una serie temporal de 11 puntos, constituyendo cada punto un contrato de futuros. Contamos con un total de 763 curvas, entre las fechas de 01/02/2019 y 14/01/2022. Aunque la liquidez varíe a lo largo del tiempo (Haumond y Ray, 2017), hay una serie de contratos que mantiene unas cotas de liquidez y volumen más elevada, como son los contratos de marzo, junio, septiembre y diciembre. Especialmente, los contratos de junio y diciembre son los que más volumen de contratación tienen.³

Una vez introducida esta construcción, poniendo en contexto diferentes contratos y fechas, se va a exponer la realización de las curvas. Lo más sencillo e intuitivo hubiese sido coger los 11 primeros puntos, pero surge el primer problema: Los contratos de futuros, al ser derivados financieros, tienen una fecha de vencimiento, por lo que para cada conjunto de 3 meses que se va sucediendo, tenemos un contrato menos disponible, por lo que esta posibilidad no es factible.

Otra posibilidad que se tenía era coger los 11 primeros puntos, aunque teniendo en cuenta las fechas de vencimiento y el conocido *rollover* no es una opción viable. Además, hay ciertos contratos de futuros con vencimiento superior a 2 años, cuyos volúmenes de negociación son muy pequeños, por lo que no se utilizan ni en bases de datos ni en representaciones de las curvas.

Por ello, hemos realizado el siguiente modelo (siguiente página) para representar las curvas, cuya explicación se relata a continuación:

- Las sombras amarillas son aquellos contratos que tienen liquidez, y aquellos espacios sin sombra son los contratos que no. Cada punto negro representa un punto para la curva.
- Aunque no nos supone un problema, ya que lo que nos importa es la posición, no se puede replicar la misma estructura para todas las curvas, ya que hay contratos que no están representados por no tener liquidez. A pesar de ello, se ha intentado replicar una estructura más o menos similar para todos los contratos.
- Los contratos de junio y diciembre han tenido especial relevancia por dos motivos: Realizar una mejor extrapolación de los datos y los contratos con vencimientos superiores a los 2 años tienen volúmenes de negociación mucho más bajos.

Por ello, la estructura resultante es la siguiente:

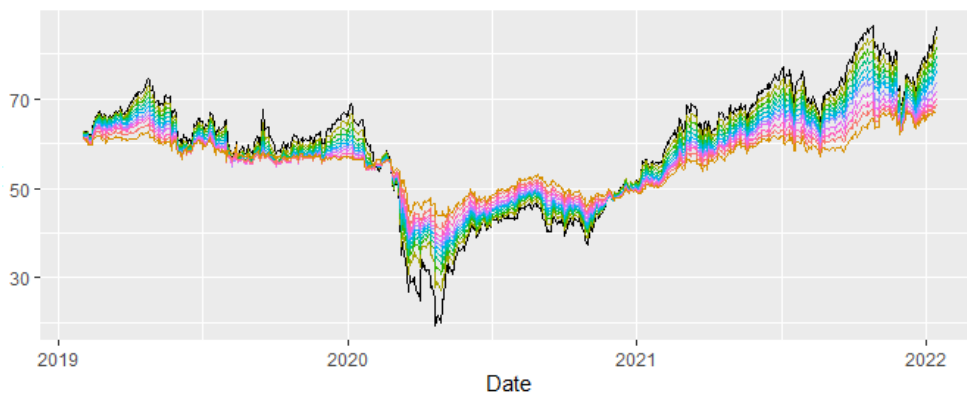
³ Aunque no se incluya bibliografía al respecto, este caso es conocimiento experto de una persona que trabaja con futuros del petróleo.

Ilustración 2 - Estructura de las curvas de futuros.

timestamp	2019				2020				2021				2022				2023				2024				2025				
	H19	M19	U19	Z19	H20	M20	U20	Z20	H21	M21	U21	Z21	H22	M22	U22	Z22	H23	M23	U23	Z23	H24	M24	U24	Z24	H25	M25	U25	Z25	
ene-19																													
feb-19		*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
mar-19		*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
abr-19		*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
may-19			*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
jun-19			*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
jul-19			*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
ago-19				*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
sep-19				*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
oct-19				*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
nov-19				*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
dic-19				*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
ene-20				*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
feb-20				*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
mar-20				*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
abr-20				*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
may-20				*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
jun-20				*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
jul-20				*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
ago-20				*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
sep-20				*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
oct-20				*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
nov-20				*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
dic-20				*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
ene-21				*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
feb-21				*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
mar-21				*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
abr-21				*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
may-21				*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
jun-21				*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
jul-21				*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
ago-21				*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
sep-21				*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
oct-21				*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
nov-21				*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
dic-21				*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
ene-22				*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*

Aunque la estructura sea la que se aprecia en la tabla y sean estos los contratos seleccionados, no se van a modelar de esta forma. Con esto quiero referirme a que vamos a seleccionar los 11 puntos para cada fecha en cuestión, para después hacer el análisis de las componentes principales por posición: La primera posición para cada fecha será el primer punto disponible, independientemente del contrato que sea, y así, sucesivamente. Con esto, intentaremos ver las dinámicas que se siguen por la posición que corresponda.

Ilustración 3 - Curvas construidas.








3.3. Construcción de los spreads.

En este apartado vamos a exponer como se han construido los diferentes spreads que tenemos. Como comentábamos en apartados anteriores, los spreads se calculan directamente usando los contratos de futuros. Por ello, contamos con los contratos de marzo (representado con la letra H), junio (representado con la letra M), septiembre (representado con la letra U) y diciembre (representado con la letra Z) de los años 2022, 2023 y 2024, los contratos de junio (M) y diciembre (Z) del año 2025 y el contrato de diciembre (Z) de 2026.

Tenemos estos datos por asuntos de volumen de contratos negociados y liquidez: Normalmente, los contratos de futuros del petróleo tienen vencimiento mensual y suelen tener muchos vencimientos abiertos, pero en el caso de los futuros, lo normal es que se negocie siempre el futuro de primer vencimiento. Sin embargo, los contratos citados al inicio de este apartado suelen tener mucho volumen de negociación y, por tanto, más liquidez. Especialmente en nuestro caso, los contratos de junio (M) y diciembre (Z) son los que suelen tener un volumen más elevado de negociación. Por este motivo tenemos los datos de estos contratos con vencimientos más alejados, como es el caso de los futuros de 2025 y 2026.

Brent Crude Futures

CONTRACT	LAST	TIME(GMT)	% CHANGE	VOLUME
 JAN23	87.740	11/18/2022 10:59 PM	0.137	374364
 FEB23	87.320	11/18/2022 10:53 PM	0.195	366969
 MAR23	86.760	11/18/2022 10:21 PM	0.231	272603
 APR23	86.130	11/18/2022 10:21 PM	0.233	122640
 MAY23	85.470	11/18/2022 10:21 PM	0.235	74844

4

Además, a pesar de tener datos desde enero del año 2019, este conjunto de datos empieza un año después, en enero del año 2020. Esto se debe al deseo de tener un mismo número de puntos en todas las series temporales de los spreads. Si hubiésemos usado contratos del año 2019 para calcular nuestros spreads, tendríamos series temporales con 196 puntos, hasta la fecha del 01/11/2019, día que el último contrato de dicho año vence.

Una vez explicadas algunas características de la liquidez y el volumen, vamos a dar paso a la explicación del cálculo de los diferentes tipos de spread, comentando las diferentes estrategias:

- Butterflies.

Los butterflies son estrategias de spread que se construyen usando tres fechas de vencimiento. Los butterflies se emplean cuando esperamos movimientos en el activo subyacente del derivado en cuestión. Tenemos distintos tipos de butterflies, teniendo en cuenta, además, que se pueden combinar entre ellos, como veremos ahora:

1. *Butterflies* de 3 meses: Los contratos que se negocian tienen diferencias de vencimiento de 3 meses. Contamos con 4 spreads de este tipo.
2. *Butterflies* de 6 meses: Los contratos que se negocian tienen diferencias de vencimiento de 6 meses. Contamos con 6 spreads de este tipo.
3. *Butterflies* de 12 meses: Los contratos que se negocian tienen diferencias de vencimiento de doce meses. Contamos con 3 spreads de este tipo.
4. 2 *butterflies* de 3 meses y un *butterfly* de 6 meses: Comentábamos antes que hay estrategias dentro de los *butterflies* donde se podían combinar. Este es un ejemplo de ello, donde se compran diferentes contratos para cumplir las diferencias temporales de 3 y 6 meses, respectivamente. Contamos con 4 *spreads* de este tipo.

⁴ El precio mostrado es del día 18 de noviembre, ya que fue el último día de negociación.

5. Dos *butterflies* de 6 meses y un *butterfly* de 12 meses: Es otra de las estrategias donde se combinan diferentes tipos de *butterflies* para cumplir las diferencias temporales, de 6 y 12 meses, respectivamente. Contamos con 3 spreads de este tipo.

En la primera tabla se puede observar con mayor detalle los spreads construidos y los contratos utilizados para ello.

- *Condors*:

Los *condors* son estrategias para el cálculo de spreads en las que empleamos cuatro fechas de vencimiento diferentes. Tenemos diferentes tipos de *condors*, teniendo en cuenta que, además, se pueden combinar entre ellos.

1. *Condors* de 6 meses: Los contratos negociados tienen diferencias de vencimiento de 6 meses. Contamos con 4 spreads de este tipo.
2. *Condors* de 12 meses: Los contratos negociados tienen diferencias de vencimiento de 12 meses. Contamos con 2 spreads de este tipo.
3. 2 *condors* de 6 meses y un *condor* de 12 meses: Al igual que ocurre con los *butterflies*, hay estrategias con los *condors* que combinan diferentes tipos de contratos. Contamos con 2 spreads de este tipo.

En la segunda tabla se puede observar con mayor detalle los spreads construidos y los contratos utilizados para ello.

Ilustración 5 - Contratos construidos en la elaboración de los butterflies.

	H2	M2	U2	Z2	H3	M3	U3	Z3	H4	M4	U4	Z4	M5	Z5	Z6
FLY 3 Meses															
M2U2Z2		1	-2	1											
Z2H3M3				1	-2	1									
M3U3Z3						1	-2	1							
Z3H4M4								1	-2	1					
FLY 6 Meses															
M2Z2M3		1		-2		1									
Z2M3Z3				1		-2		1							
M3Z3M4						1		-2		1					
Z3M4Z4								1		-2		1			
FLY 12 Meses															
Z2Z3Z4				1				-2				1			
Z3Z4Z5								1				-2		1	
Z4Z5Z6												1		-2	1
2*FLY 3 Meses - FLY 6 Meses															
M2U2Z2M3		1	-4	4		-1									
Z2H3M3Z3				1	-4	4		-1							
M3U3Z3M4						1	-4	4		-1					
Z3H4M4Z4								1	-4	4		-1			
2*FLY 6 Meses - FLY 12 Meses															
Z2M3Z3Z4				1		-4		4				-1			
Z3M4Z4Z5								1		-4		4		-1	
Z4M5Z5Z6												1	-4	4	-1

Ilustración 6 - Contratos utilizados en la elaboración de los condors.

	H2	M2	U2	Z2	H3	M3	U3	Z3	H4	M4	U4	Z4	M5	Z5	Z6
CONDOR 6M															
M2ZM3Z3		1		-3		3		-1							
Z2M3Z3M4				1		-3		3		-1					
M3Z3M4Z4						1		-3		3		-1			
Z3M4Z4M5								1		-3		3	-1		
CONDOR 12M															
Z2Z3Z4Z5				1				-3				3		-1	
Z3Z4Z5Z6								1				-3		3	-1
2 CONDOR 6 Meses - CONDOR 12 Meses															
Z2M3Z3M4Z4Z5				1		-6		9		-2		-3		1	
Z3M4Z4M5Z5Z6								1		-6		9	-2	-3	1

Una vez introducida la temática y los problemas a resolver, la metodología utilizada, la aproximación teórica a la misma y la construcción del conjunto de datos, vamos a ver los principales resultados que hemos obtenido.

Antes de ver los resultados, vamos a repasar lo que se va a estudiar:

- Estudio de las curvas de futuros del petróleo con el fin de analizar los factores que lideran el precio del petróleo a través del análisis de componentes principales. Además, en este apartado, se van a usar las componentes principales para realizar una reconstrucción de las curvas.
- Modelado de los *spreads* con el fin de estudiar la gestión de riesgos a través de los modelos fraccionalmente integrados (ARFIMA).
- Por último, vamos a estudiar si los *spreads* tienen un comportamiento persistente mediante el análisis fractal. Este estudio se va a llevar a cabo midiendo la dependencia de las series temporales, a través del coeficiente de Hurst.

4. Principales resultados.

Como bien comentábamos en apartados anteriores, el petróleo en el mundo actual es un recurso de gran importancia debido a su uso en diferentes sectores de la economía: Desde la producción de combustible hasta la fabricación de productos químicos y plásticos. Es por ello que los futuros del petróleo son uno de los mercados más seguidos por inversores, institucionales o minoristas, y analistas financieros.

Por ello, en este análisis, vamos a tratar de estudiar diferentes aspectos sobre estos productos usando el programa R Studio, un software muy útil en el campo del análisis estadístico y la programación. Creo importante mencionar que se ha usado esta herramienta por diversos motivos: En primer lugar, es una plataforma muy utilizada por analistas de datos al permitir programar de manera eficiente y rápida. Además, ofrece una variedad amplia de librerías y herramientas para realizar gráficos y visualizaciones de datos. Por último, es uno de los programas que más se ha usado a lo largo del grado académico, por lo que la familiaridad con este es elevada.

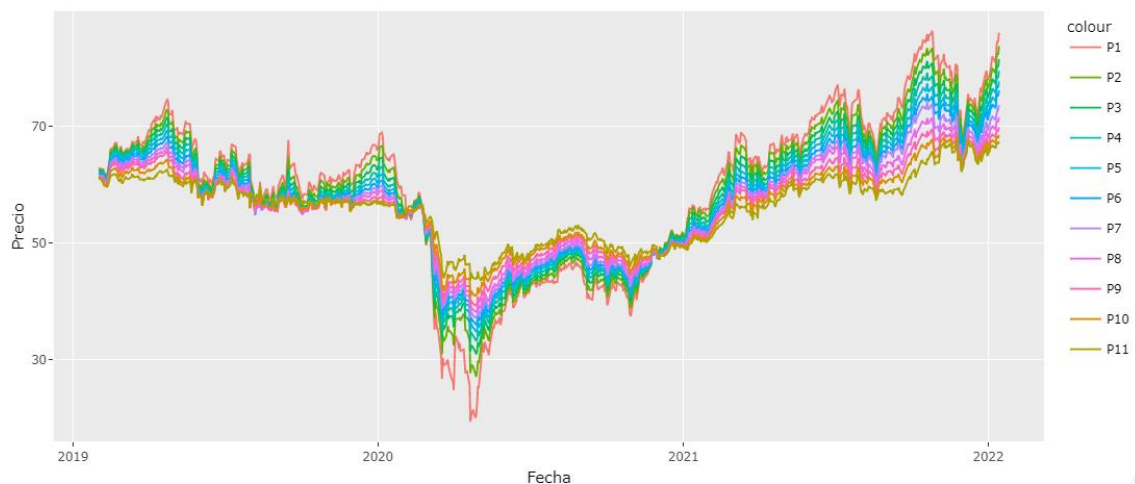
Antes de profundizar en los resultados principales del estudio, se va a realizar un análisis introductorio de algunos estadísticos importantes para tratar de entender el comportamiento que tienen estos productos financieros. Vamos a realizar dicho análisis usando los dos conjuntos de datos de los que disponemos: Por un lado, se verán los principales estadísticos de las curvas del petróleo. Por otro lado, un análisis estadístico introductorio de los spreads será realizado.

- Análisis estadístico de las curvas.

Este conjunto de datos cuenta con un total de 9.156 observaciones: Tenemos 12 columnas, siendo once las posiciones y la fecha. Por otro lado, el conjunto de datos tiene 763 filas, por lo que tenemos información para un total de 763 días de negociación de futuros del petróleo.

El gráfico de las 11 variables que forman la curva es el siguiente:

Ilustración 7 - Variables que forman la curva de futuros.

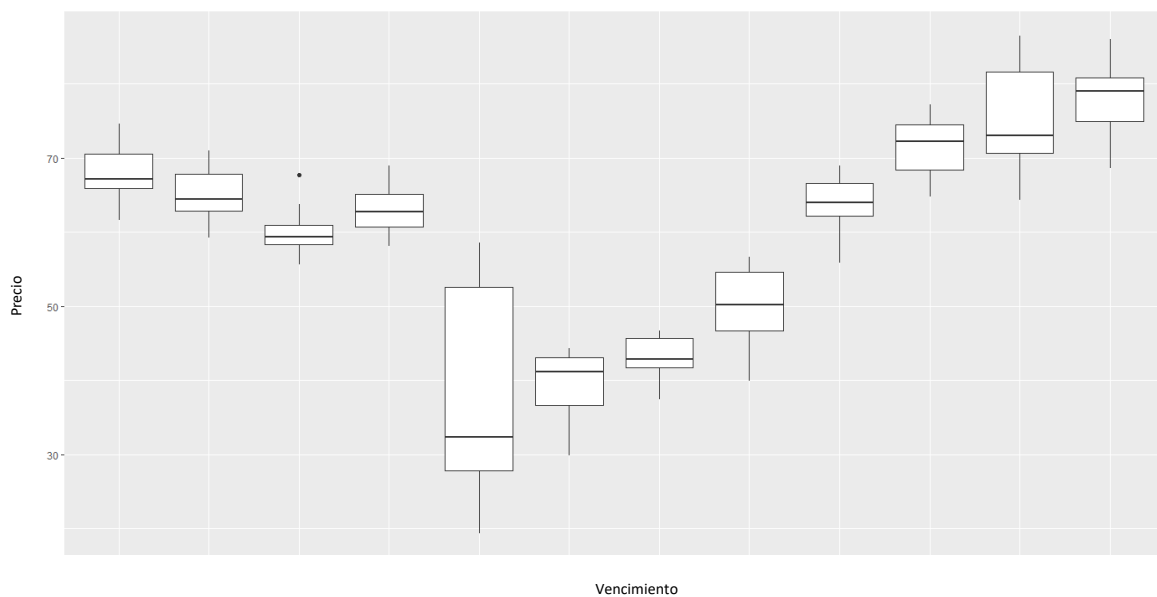


La figura se caracteriza por una serie de aspectos que son importantes destacar. Por un lado, todas las posiciones siguen una estructura similar, aunque los movimientos en las primeras posiciones son más destacados que los movimientos en las últimas posiciones. Es por ello por lo que la primera posición lidera a las demás, a excepción de algunos momentos de convergencia. Lo podemos ver en los momentos iniciales y finales de 2020, donde esos momentos de convergencia dan lugar a que, posteriormente, la primera posición lidere a las demás.

Por otro lado, también se puede destacar como los cambios de precio son pronunciados. Aunque empieza con cierta estabilidad en torno a los 60€, a inicios del año 2020 hay una gran bajada por la pandemia del COVID-19. Cuando llega a su punto más bajo, empieza una escalada hasta llegar a los más de 70€ finales. Aunque haya movimientos bruscos, intentaremos captar las dinámicas de las series usando el análisis de componentes principales. El objetivo final es sencillo: Si podemos obtener buenos resultados modelando la dinámica de las series, podremos obtener resultados muy cercanos a los spreads, ya que estos últimos derivan de las primeras.

Por otro lado, también podemos observar los datos por vencimiento: Esto significa ver el conjunto de datos que hace referencia a los contratos dependiendo del momento en el que venzan.

Ilustración 8 - Precio por vencimiento del contrato.



En este caso, si tenemos en cuenta la evolución de los *boxplots* en el eje Y, podemos observar que se sigue una estructura muy similar a la del gráfico anterior. Sin embargo, en este gráfico se pueden observar diferentes aspectos. Por un lado, cabe añadir que la presencia de *outliers* es residual. Además, cabe añadir que hay valores cuyo rango de precios es muy pequeño. Este aspecto es la tónica general, únicamente obviado por los vencimientos quinto y undécimo. Estos momentos coinciden justamente con los

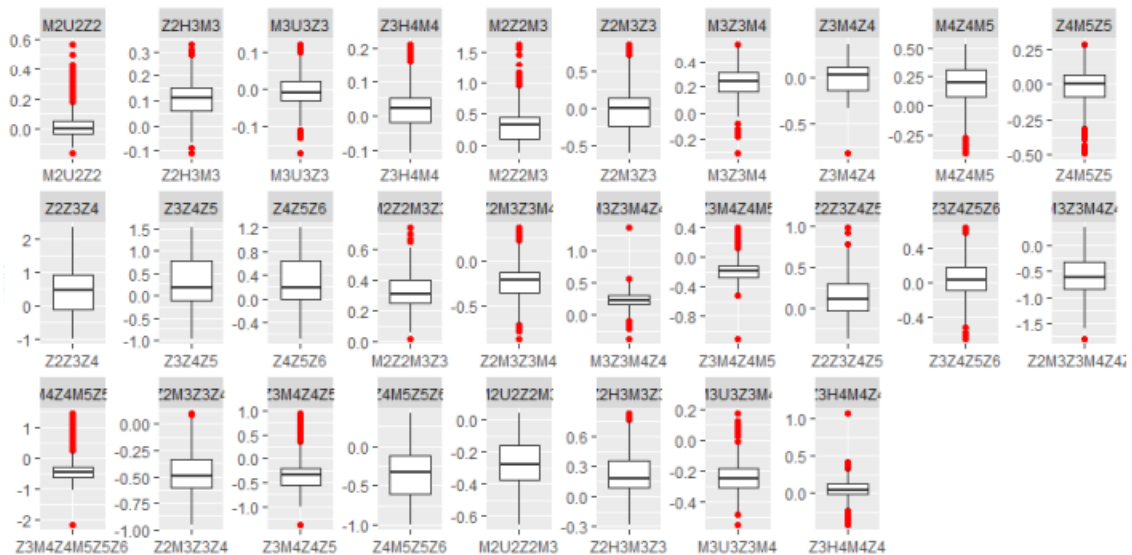
momentos donde la variación existente entre la primera posición y la última son de mayor rango.

- Análisis estadístico de los *spreads*.

Aunque lo hemos mencionado en anteriores apartados, nuestro conjunto de datos cuenta con un total de 15.573 observaciones, donde tenemos 29 columnas diferentes, siendo 28 *spreads* y la fecha. Además, el conjunto de datos tiene 537 filas, lo que supone tener datos para un total de 537 días hábiles, es decir, de algo más de dos años. Si nos fijamos en la tabla X, el primer vencimiento haría referencia a la primera fecha disponible, a partir del 01/02/2019, hasta la fecha de vencimiento del contrato más cercano, esto es, 01/05/2019.

Los *boxplots* de los 28 *spreads* los podemos encontrar en la siguiente imagen:

Ilustración 9 - *Boxplots* de los *spreads*.



Los *boxplots* son una herramienta gráfica comúnmente utilizada para representar la distribución de un conjunto de datos numéricos. Estos diagramas muestran la mediana, los cuartiles y los valores extremos de los datos en un solo gráfico, lo que permite una fácil comparación entre múltiples conjuntos de datos.

Hay un aspecto muy importante en los anteriores *boxplots* que podemos destacar: La presencia de *outliers* o valores atípicos en, prácticamente, todos los *spreads*, y la concentración que presentan los datos, por la estrechez que presenta el rango intercuartílico en muchos *spreads*. Cuando los datos tienen un rango pequeño y la presencia de valores atípicos es significativa, los *boxplots* pueden proporcionar información valiosa sobre la distribución de los datos. Los valores atípicos son valores que se encuentran muy por encima o por debajo del rango esperado de los datos, y pueden indicar la presencia de eventos raros que afectan al *dataset*. Uno de los motivos

de la presencia de estos valores atípicos puede ser el impacto que el COVID-19 tuvo en estos mercados, como podemos ver en imágenes anteriores en la gráfica de las curvas. Y, como los spreads provienen de las curvas de futuros del petróleo, los eventos afectados en estos mercados, por tanto, también afectan en este apartado.

Por otro lado, el gráfico de la serie temporal de los diferentes spreads los podemos encontrar a continuación, dividido por la estrategia de los contratos:

Ilustración 10 - Butterflies de 3 meses.

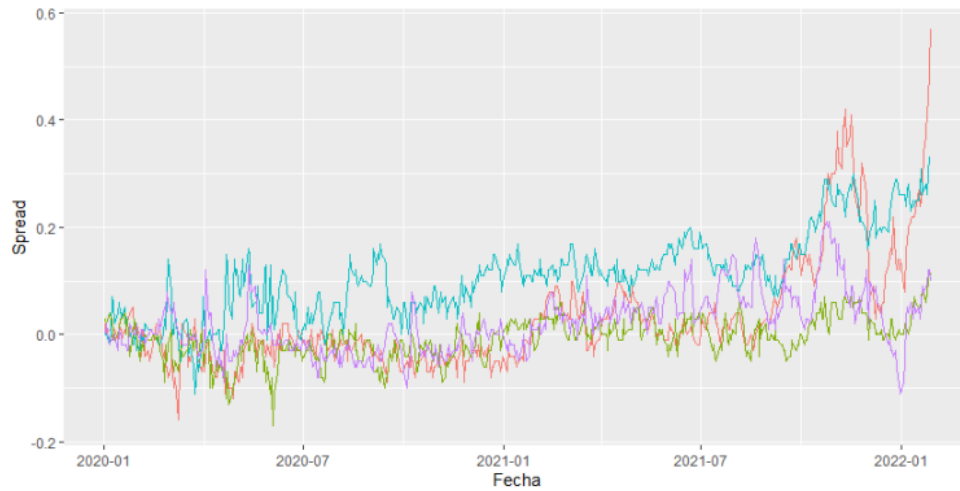


Ilustración 11 - Butterflies de 6 meses.



Ilustración 12 - Butterflies de 12 meses



Ilustración 14 - Spreads: 2 contratos de 3 meses y un contrato de 6 meses

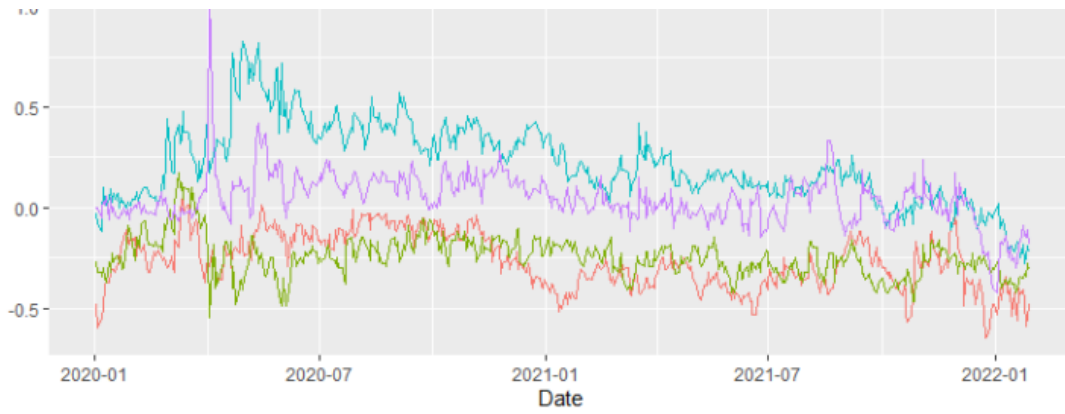


Ilustración 13 - Spreads: 2 contratos de 6 meses y un contrato de 12 meses.

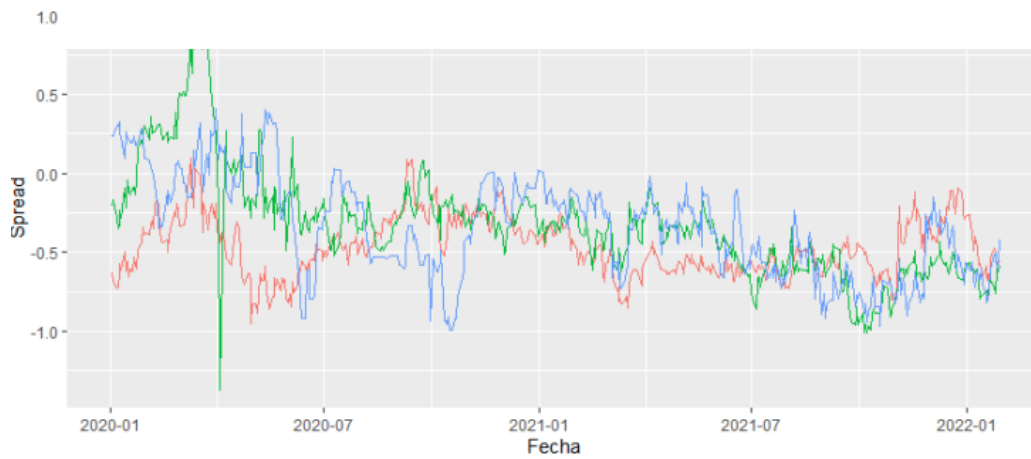


Ilustración 17 - Condors de 6 meses.

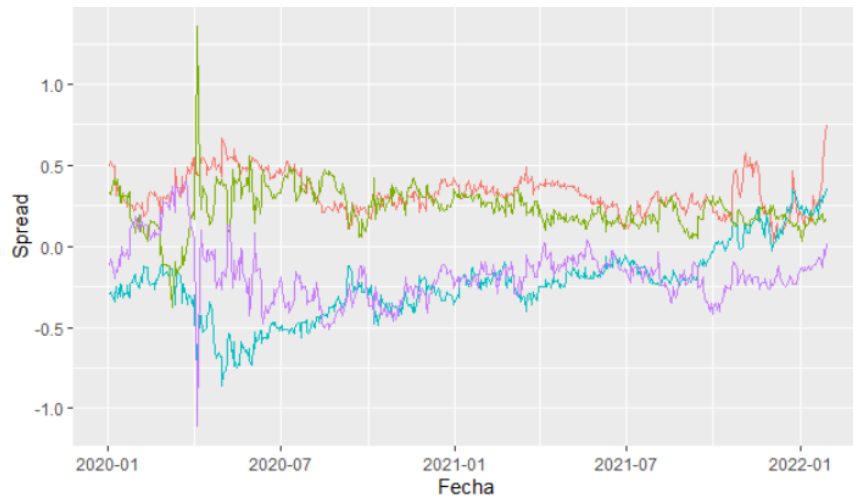


Ilustración 16 - Condors de 12 meses.

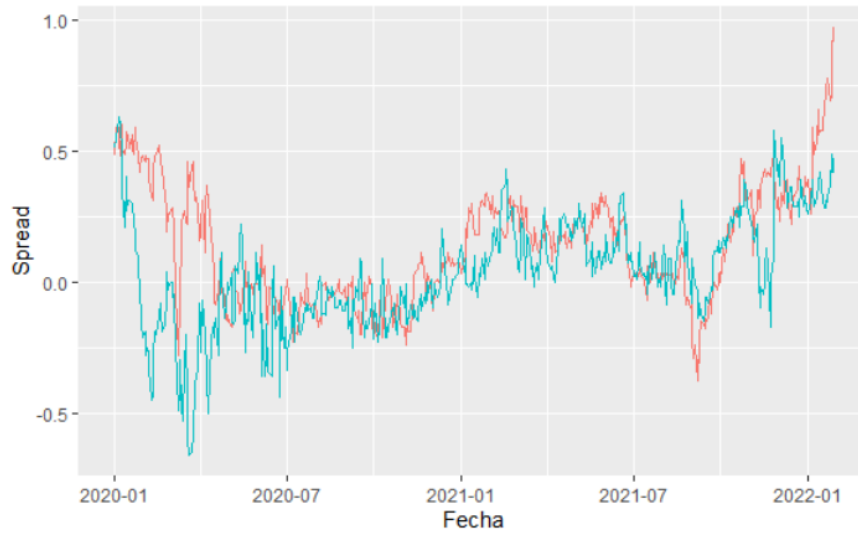
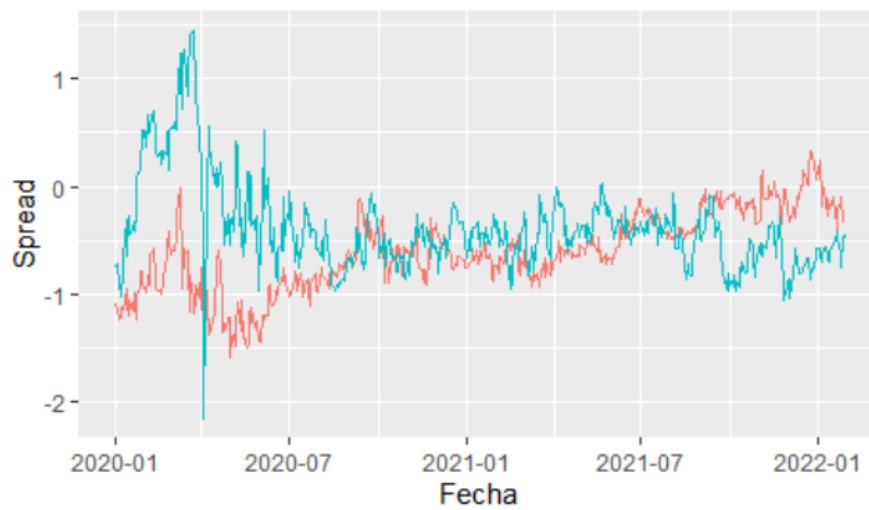


Ilustración 15 - 2 condors de 6 meses y uno de 12 meses.



Como podemos observar a lo largo de los últimos gráficos, las estructuras varían de manera especialmente relevante en los diferentes tipos de contratos encontrados, teniendo comportamientos muy diferentes, incluso si un mismo contrato forma parte de dos spreads diferentes. Según se vayan estudiando los contratos más a fondo, iremos observando diferentes aspectos de las series y cómo afectan a esta.

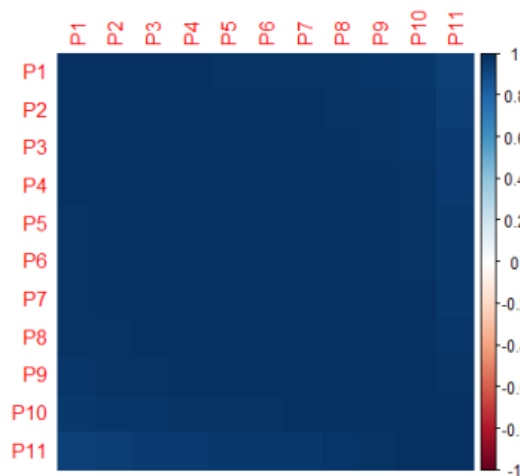
4.1. Estudio de las curvas: Análisis de componentes principales.

Como se ha mencionado en apartados anteriores, los spreads de futuros del petróleo son calculados, simplificando de forma extrema, comprando un contrato en un vencimiento y vendiendo otro de un vencimiento superior. En el caso de los *butterflies* o *condors*, aunque sean más contratos los que intervienen, la dinámica es muy similar. De aquí extraemos que el subyacente del *spread* es la curva de futuros, por lo que se puede analizar la serie temporal del *spread* y la serie temporal del subyacente. Analizando las dinámicas del subyacente, podremos modelar el *spread*.

Por ello, para llevar a cabo el análisis del subyacente, se va a realizar un análisis de componentes principales (PCA). El análisis PCA nos va a ayudar en este caso a ver qué factores, entre los que tenemos disponibles, son los que lideran el precio de los futuros del petróleo.

Como mencionábamos en apartados anteriores, el análisis PCA es un proceso por el cual se transforma un conjunto de variables originales (en este caso, las posiciones) en un conjunto nuevo de variables que son combinación lineal de estas variables de partida. Las nuevas variables están caracterizadas por no tener correlación y por poder ordenarse siguiente el criterio de la información incorporada⁵. También cabe destacar que este análisis es bueno en aquellos casos en los que hay mucha información redundante, como vemos en el siguiente gráfico observando la correlación existente entre las posiciones.

Ilustración 18 - Correlación entre las diferentes posiciones de la curva.



⁵ En PCA podemos entender la información recogida en el sentido de variabilidad: A mayor varianza, más información se recoge.

Como vemos en la ilustración 17, la correlación entre los diferentes vencimientos es muy alta y positiva, cercana a uno (correlación total) en mucho de los casos, por lo que la relación es muy fuerte y el análisis de componentes principales puede ser útil e idóneo en estos casos.

Para el análisis PCA, en caso del uso de variables cuantificadas en diferentes medidas, es necesario centrar y estandarizar estas variables. En nuestro caso, las medidas están en la misma escala. Sin embargo, la librería usada para realizar el análisis tiene una función llamada *prcomp*, que centra las variables por defecto, por lo que también estandarizaremos.

Ilustración 19 - Importancia de cada componente principal.

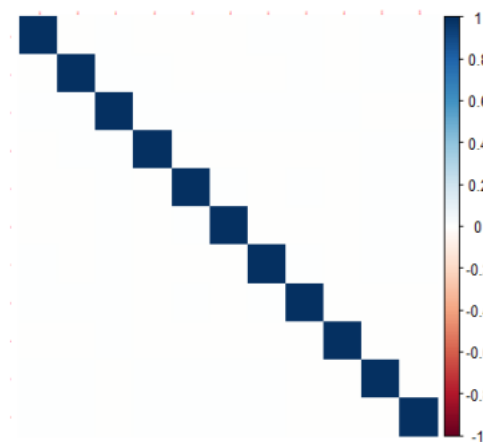
```
Importance of components:
Standard deviation      PC1    PC2    PC3    PC4    PC5    PC6    PC7    PC8    PC9    PC10
Proportion of variance 0.988 0.01034 0.00116 0.00038 0.00005 0.00005 0.00001 0.000010 0.000000 0.0000
Cumulative Proportion 0.988 0.99833 0.99949 0.99987 0.99992 0.99997 0.99999 0.999990 1.000000 1.0000

PC11
Standard deviation      0.00165
Proportion of variance 0.00000
Cumulative Proportion 1.00000
```

En la ilustración 18 podemos ver los primeros resultados del análisis. Al haber once variables, existen once componentes principales. De cada componente principal podemos obtener la matriz de los *scores*, o matriz de rotación, que determina la dirección sobre la que se proyecta. Como podemos observar, usando únicamente las dos primeras componentes principales se recoge un 99,83% de la información, valor muy cercano al 100% de la información total, por lo que podemos determinar que las dos primeras componentes principales explican casi toda la variabilidad existente en nuestro conjunto de datos.

Además, y como punto especialmente reseñable, podemos añadir que no existe correlación entre las componentes principales, como bien se demuestra en la ilustración 19.

Ilustración 20 - Correlación entre componentes principales.



En la siguiente imagen se pueden observar los *loading vectors* del análisis realizado. Se puede ver como la primera componente principal tiene asociaciones negativas

prácticamente constantes con todas las variables, sabiendo que el 98,8% de la información viene aquí recogida. En la segunda componente, vemos como hay relaciones negativas fuertes con la primera y segunda posición, y relaciones positivas con la décima y undécima, siendo especialmente relevante la última.

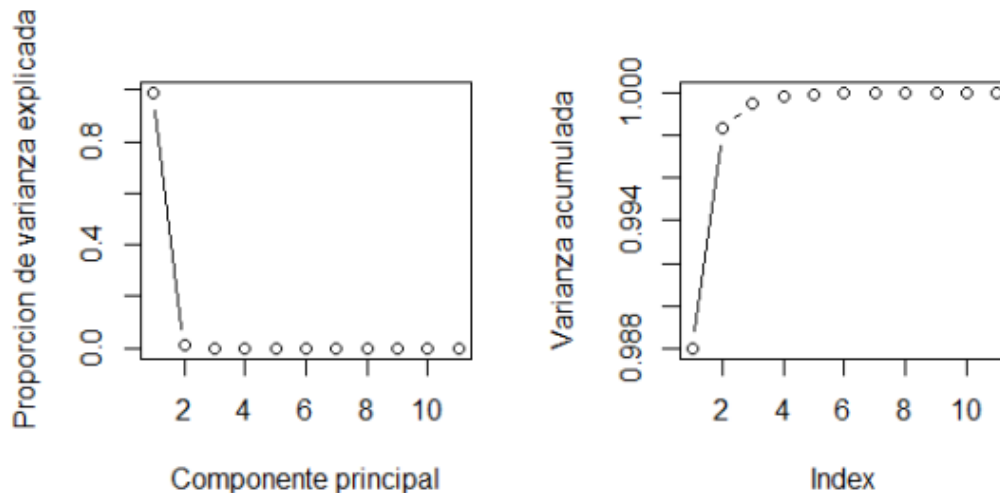
Ilustración 21 - Loading vectors del análisis PCA.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
P1	-0.2995101	-0.404137041	0.68766392	-0.30577778	-0.21931785	-0.31057747	0.1506565	0.0399862971
P2	-0.3014176	-0.322532496	0.19394517	0.19492214	0.19669564	0.42126077	-0.3590659	-0.0131295260
P3	-0.3022556	-0.243549515	-0.03474589	0.26191168	0.19779472	0.27732769	-0.1063746	-0.0008878006
P4	-0.3027424	-0.171537121	-0.16475699	0.21881044	0.11951268	0.02339568	0.1671644	-0.1442858505
P5	-0.3029858	-0.111498199	-0.24217171	0.16667182	0.05931973	-0.18526379	0.3225056	-0.1949939483
P6	-0.3031154	-0.052129799	-0.27350974	0.14127323	0.08937432	-0.36418382	0.3774288	0.3305724157
P7	-0.3031216	-0.006351473	-0.30656275	-0.03004515	-0.34558028	-0.37562201	-0.6910369	0.2316841293
P8	-0.3030295	0.088300296	-0.21557276	-0.24471346	-0.60456403	0.33263038	0.1459331	-0.4775137336
P9	-0.3024070	0.203999694	-0.10049566	-0.52176044	0.13522858	0.40429230	0.1493803	0.5722812427
P10	-0.3009043	0.364879758	0.05968191	-0.36302224	0.55691153	-0.25584732	-0.1992437	-0.4561132635
P11	-0.2950381	0.667955784	0.41663306	0.48937252	-0.18691513	0.03202331	0.0425603	0.1134315805

	PC9	PC10	PC11
P1	-0.102504140	0.02729619	0.021865063
P2	0.537559567	-0.26496188	-0.158383356
P3	-0.341527325	0.57941187	0.447002711
P4	-0.516658119	-0.14622132	-0.672301947
P5	-0.005881312	-0.58465565	0.536293118
P6	0.494642893	0.37879069	-0.173596074
P7	-0.121446408	-0.07992702	0.018131784
P8	0.175478364	0.18883851	-0.045943415
P9	-0.146919368	-0.17237110	0.041765824
P10	0.062100895	0.11473635	-0.024609975
P11	-0.035442887	-0.04149909	0.009990019

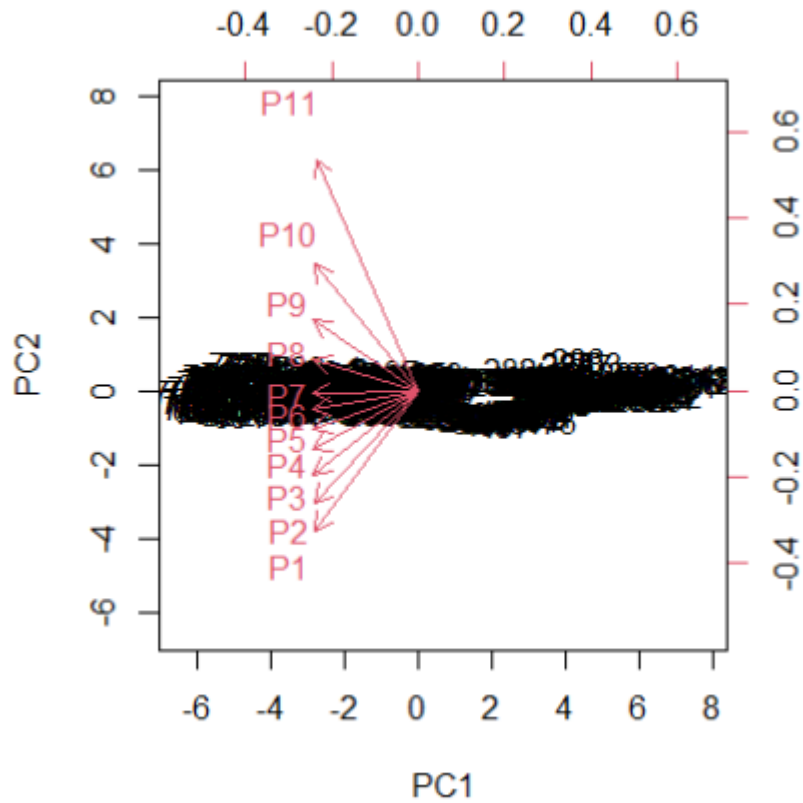
Como se mencionaba anteriormente, las primeras componentes principales recogen la mayoría de la información. En el siguiente conjunto de gráficos podemos observar como la principal cantidad de varianza viene explicado en la primera componente principal, aportando una información muchas veces menor en las posteriores componentes.

Ilustración 22 - Proporción de la varianza explicada.



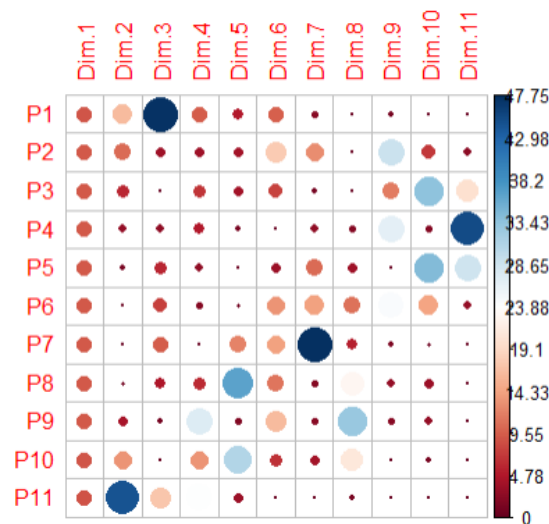
A continuación, se muestra un *biplot* con las dos primeras componentes principales. El uso de este gráfico se ha realizado para realizar posibles conglomerados.

Ilustración 24 - Biplot con las dos primeras componentes principales.



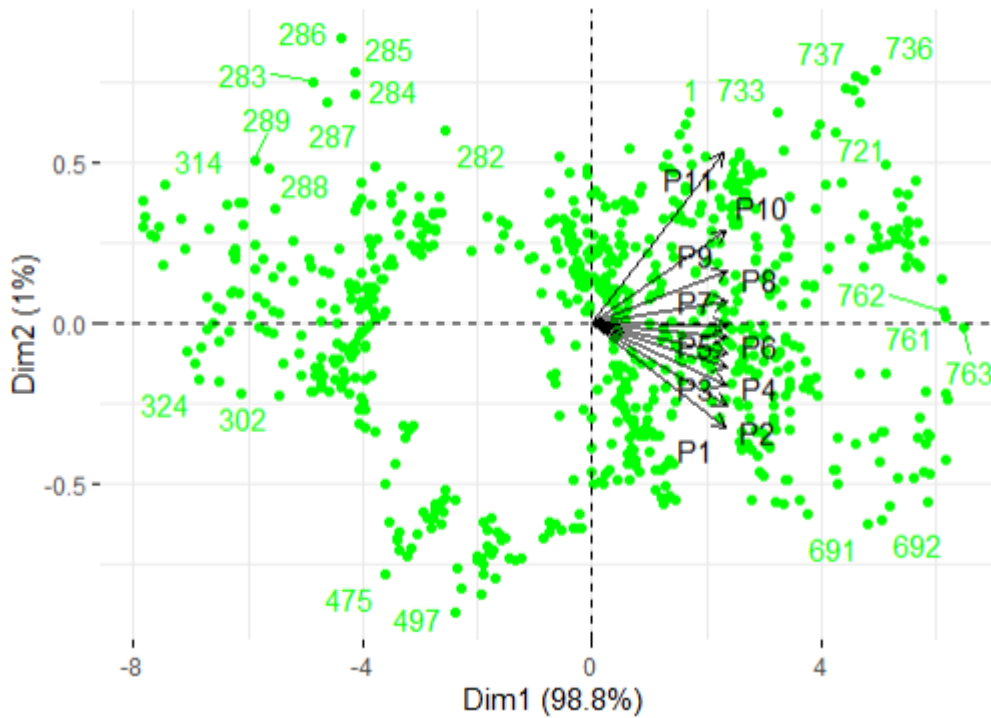
En el gráfico anterior, los puntos negros hacen referencia a las diferentes series temporales que tenemos, mientras que los símbolos rojos hacen referencia a los *loading vectors*. Apoyando lo que veníamos mencionando, la primera componente explica mucha más información que la segunda componente principal, viendo como todas las variables tienen influencia negativa sobre la primera componente principal. Por otro lado, las variables P1 y P2 tienen influencias negativas grandes sobre la segunda componente, mientras que P10 y P11 tienen influencias positivas significativas, indicando que esta componente pone foco en lo que pasa en estas posiciones. Esto, además, viene apoyado por el siguiente gráfico, donde podemos observar la contribución de cada variable a la componente principal.

Ilustración 23 - Contribución de las variables a la componente principal.



Otro de los factores que se ha estudiado es el gráfico de las observaciones usando la primera y segunda componente principal, dibujando, además, las componentes principales.

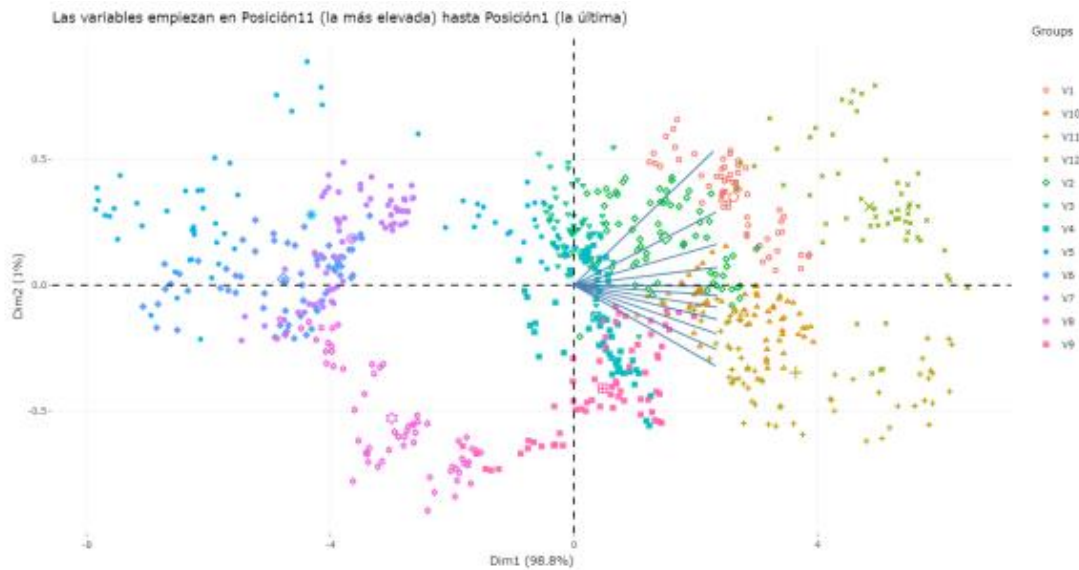
Ilustración 25 - Observaciones y componentes principales.



El gráfico es ilegible por la gran cantidad de información que hay, aunque podamos usarlo para ver las posiciones de las componentes principales y las observaciones, observando que hay datos que están en sentido opuesto a la dirección de las componentes (aquellas que están en el sector a la izquierda del cero). Sin embargo, un efecto que podemos estudiar en este gráfico es el cambio del vencimiento. Si recordamos la construcción de las curvas, podemos tener presente como se usaban diferentes vencimientos debido a las características propias de los contratos de futuros. Por ello, el siguiente gráfico muestra los contratos existentes teniendo en cuenta los diferentes vencimientos.

El gráfico muestra las componentes principales (la dirección que tienen) y las diferentes observaciones de las que disponemos.

Ilustración 26 - Efectos del cambio de vencimiento.



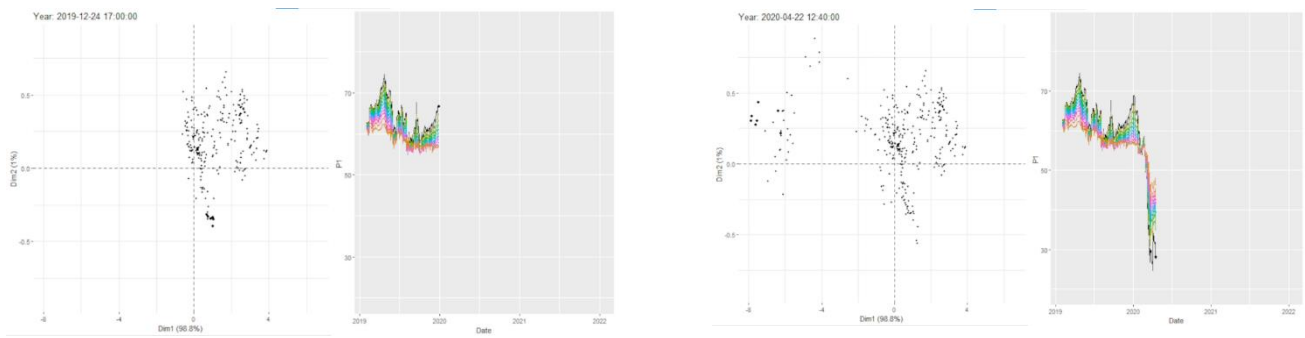
Se puede observar que todas las variables (es decir, nuestras posiciones) tienen influencias positivas en la primera componente principal. En la segunda componente hay alguna influencia, sobre todo por parte de la primera posición y la undécima.

Uno de los factores que ha ayudado a la obtención de conclusiones es la realización de un gráfico animado que consta de dos sub-gráficos: Por un lado, el gráfico anteriormente mostrado y por otro, un gráfico animado de las curvas a medida que el tiempo va pasando. Las principales conclusiones obtenidas son las siguientes:

1. En el año 2019 las observaciones van surgiendo de manera homogénea en el “primer cuadrante” habiendo bastante influencia de la primera principal en algunos momentos, mientras que en otros momentos se queda muy cercano al eje Y.
2. A finales del año 2019 y a principios del año 2020 el precio de los futuros experimenta una subida que hace que la segunda componente principal empiece a tener significación de forma negativa.
3. Con la llegada del COVID existe una fuerte bajada en los precios de futuros que hacen que la primera componente principal tome fuerza de manera negativa.
4. Una vez empieza el precio a estabilizarse la primera componente principal sigue teniendo fuerza, aún de forma negativa. Pero alrededor de octubre del 2020 vuelve a haber otra bajada de precios que hace que la segunda componente principal tome fuerza de manera negativa.
5. Desde inicios del año 2021 los precios experimentan una subida, aumentando el componente positivo de las primeras y segundas componentes principales poco a poco.

Debido a la imposibilidad de plasmar el gráfico animado en el documento, en el siguiente gráfico podemos ver unos fotogramas del gráfico animado que anteriormente mencionábamos.

Ilustración 27 - Frames de los gráficos animados.



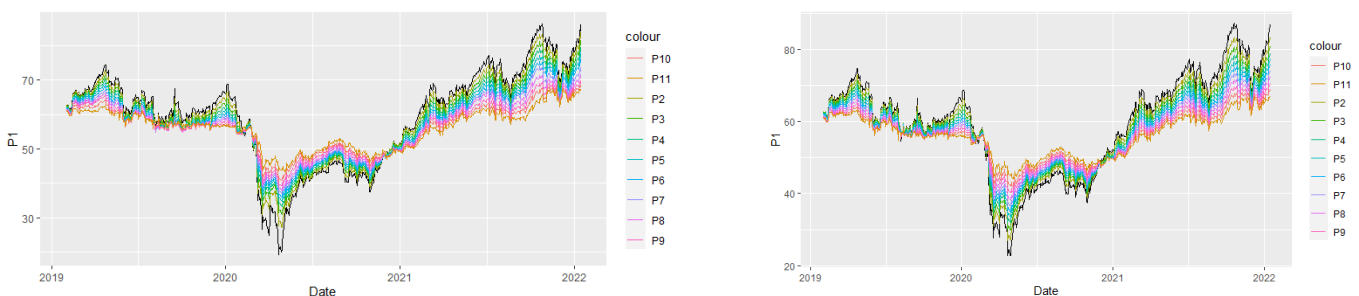
Se puede observar la diferente posición que van tomando las observaciones a lo largo del tiempo, con una evolución de los vencimientos. En este caso, las fechas tomadas hacen referencia al final del año 2019 y a la bajada de precios que hubo por el efecto del COVID-19.

Por último, y para terminar con este análisis, se ha realizado una reconstrucción de las curvas usando las componentes principales. El sentido de esta reconstrucción es sencillo: Si sabemos que los *spreads* derivan de las curvas, y podemos reconstruir las curvas con las componentes principales de una manera fiable y con una cantidad de la información decente, podremos acercarnos a los *spreads* de un modo, en cierto sentido, fidedigna.

Por ello, la reconstrucción se ha llevado a cabo usando diferentes componentes principales. Se ha empezado con dos. Una vez se ha realizado el análisis, vemos que la primera componente principal recoge la mayoría de la variabilidad y, por tanto, de la información. Sin embargo, se va a realizar la reconstrucción usando las dos primeras componentes principales, recogiendo un 99,84% de la información.

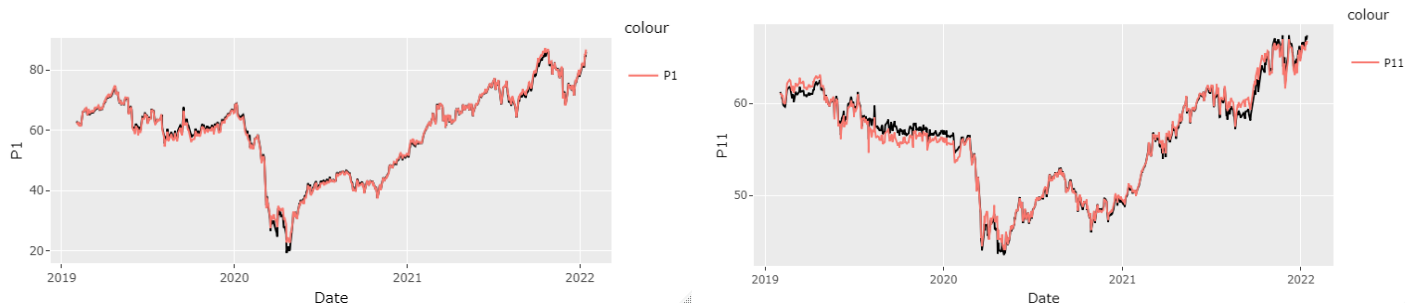
La reconstrucción de las curvas se ha realizado el proceso inverso al que se realiza en análisis de componentes principales. La media de cada variable es sumada al resultado que se obtiene de multiplicar las coordenadas de cada observación en las componentes principales por la matriz de rotación (matriz de los loading vectors). Esto se ha hecho a través de un bucle for para cada una de las observaciones, obteniendo el dataset original reconstruido con las dos primeras componentes principales. Se ha extraído en una base de datos nueva para su tratamiento y, además, ver como era la reproducción de los *spreads* con esta base de datos. Gráficamente, estos han sido los resultados obtenidos, donde el gráfico izquierdo muestra el conjunto original y el derecho, el reconstruido.

Ilustración 28 - Gráfico original (izquierda) y reconstruido (derecha)



Si entramos en el análisis curva a curva, todas siguen un patrón muy similar, a excepción de la primera y la última, como bien se muestra en el siguiente gráfico, donde encontramos en la izquierda el gráfico de la primera curva y a la derecha el gráfico de la última.

Ilustración 29 - Reconstrucción (línea roja) de la primera y última curva.



Las curvas reconstruidas siguen un patrón muy similar al de la curva original, aunque se encuentre alguna diferencia significativa a inicios del año 2020 y finales del año 2021. Sin embargo, en la última posición, vemos como el patrón es similar de la mitad del año 2020 hasta la mitad del año 2021, pero cambia en el periodo diferente.

A la hora del cálculo de los *spreads*, se pueden mencionar una serie de problemas con los que se ha tenido que lidiar:

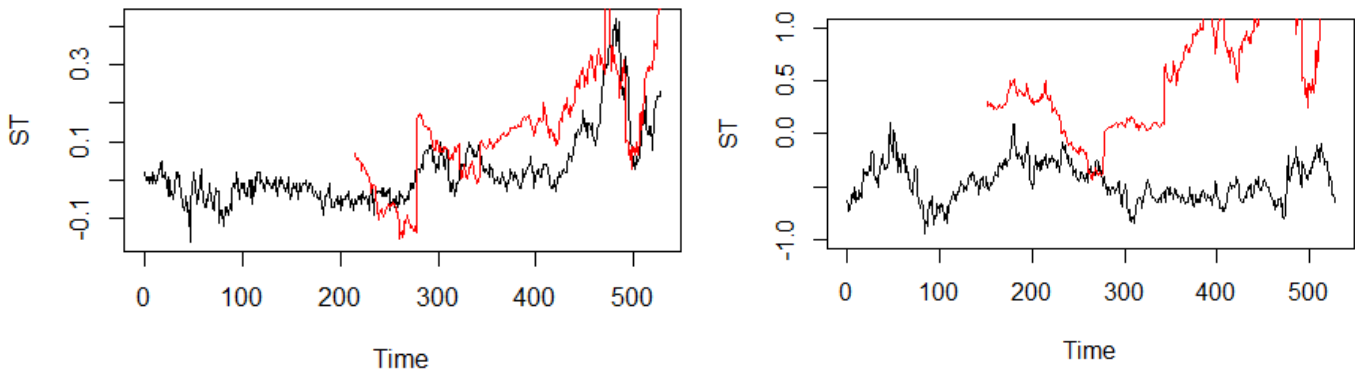
- **Liquidez:** La liquidez que tienen los diferentes contratos de futuros hacía que no se pudiesen calcular los *spreads* en su totalidad. Hay contratos que, por su lejanía en el futuro o por sus características, no son líquidos, como bien mencionábamos en apartados anteriores.
- **Datos de contratos:** Hay ciertos contratos, como los de marzo de 2024 y 2025, o los de septiembre de 2023, 2024 y 2025, de los que no se tenían datos.

Por todo ello, se han podido reproducir algunos *spreads*, pero no en su totalidad. Para ello, se ha extraído la información obtenida en la reconstrucción, se ha guardado en un fichero y se ha distribuido según el contrato al que hacía referencia. Para el cálculo de los *spreads*, el procedimiento ha seguido la misma estructura:

1. En el bucle para la reconstrucción de las curvas, se han añadido líneas de código indicando que se guardase en un fichero el resultado obtenido.
2. El fichero se ha importado a una hoja con formato .xlsx donde se han guardado las curvas por posición (en una hoja) y por fecha (en otra hoja diferente).
3. Usando la hoja de curvas por fechas y viendo la estructura de los *spreads* (Ilustraciones 4 y 5), se han calculado estos.

A la hora de la reproducción de los *spreads* con la reconstrucción de las curvas usando dos componentes principales se han encontrado diferencias importantes. En el caso que se presenta a continuación, vemos la reproducción del diferencial de precios en diferentes casos: Por un lado, una estrategia simple, con el *spread* M2U2Z2 y por otro lado, una estrategia compleja, con el *spread* Z2M3Z3Z4.

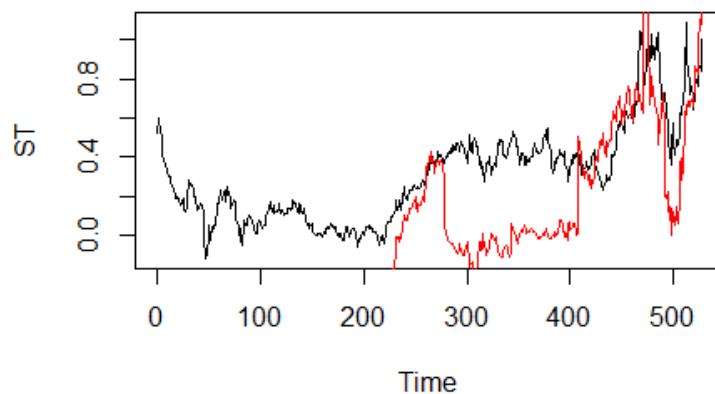
Ilustración 30 - Reconstrucción de spreads de una estrategia simple (izquierda) y una compleja (derecha).



En este caso, la estrategia más simple ha captado mejor la reconstrucción. Vemos que la estrategia más compleja tiene muchas diferencias con el spread original, aunque este aspecto no coge de sorpresa: Si nos fijamos en la reconstrucción de las curvas, la curva once es la que menos similitudes tiene con la curva original. Los contratos que vencen en diciembre (los de la “Z”), suelen ocupar la última posición. El *spread* de la estrategia compleja está formado con tres contratos que vencen en diciembre y que, por tanto, ocupan la última posición.

No hay norma general en la reproducción de los spreads, ya que los hay con muchas similitudes y los hay con diferencias significativas. Es el caso del *spread* M2Z2M3: Aunque haya diferencias que parecen importantes, vemos como capta de forma más o menos fidedigna las últimas posiciones.

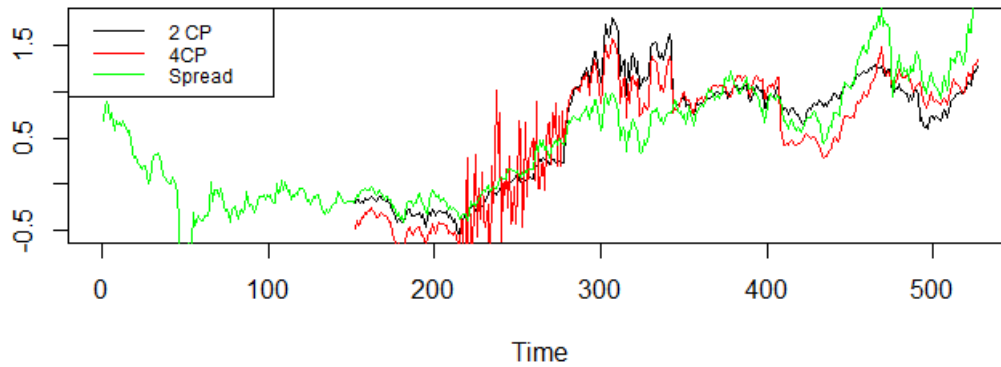
Ilustración 31 - Reconstrucción del *spread* M2Z2M3 (rojo).



Por último, podemos obtener el mismo gráfico comparando el *spread*, la reconstrucción con dos componentes principales y la reconstrucción con cuatro componentes principales. En el siguiente caso, podemos ver el *spread* Z2Z3Z4. Se observan situaciones donde la reconstrucción con 4 CP está más cerca del *spread* que la reconstrucción con 2 CP, y hay situaciones donde pasa lo contrario. Además, hay situaciones donde nos encontramos que un *spread* reconstruido está por encima del

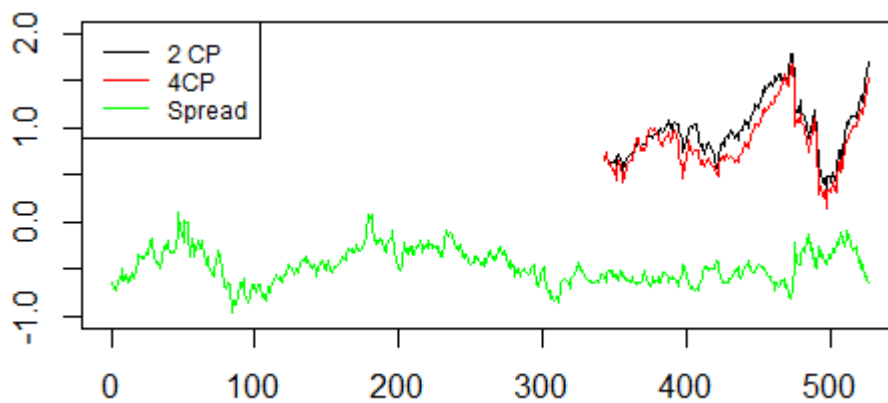
spread normal, mientras que el otro *spread* reconstruido está por debajo. Hay zonas donde el *spread* reconstruido con 4CP parece que tiene mucha volatilidad.

Ilustración 32 - Reconstrucción con dos y cuatro componentes principales.



Después hay casos donde la reconstrucción no tiene relación ninguna con el valor real, como ocurre con el *spread* Z2M3Z3Z4.

Ilustración 33 - Reconstrucción con dos y cuatro componentes principales.



Para finalizar, y con el fin de recapitular, en este apartado se han mostrado diferentes resultados relacionados con el análisis de componentes principales, donde hemos visto las implicaciones que tiene para las curvas y las diferentes posiciones en los contratos. Además, se ha visto la reconstrucción de las curvas usando un número determinado de componentes principales y la construcción de los *spreads* a partir de estas componentes principales.

4.2. Modelización con ARFIMA de los *spreads* de futuros del petróleo.

Como mencionábamos en el apartado dos, los modelos ARFIMA son una derivación de los modelos ARIMA en los que el parámetro de diferenciación d no es un número entero. Con el estudio de la modelización ARFIMA en el ámbito de los *spreads* de los futuros del petróleo estudiamos la gestión de los riesgos, ya que, si tenemos la posibilidad de estudiar las dinámicas de estos diferenciales, podremos aplicar esta dinámica a la gestión de los riesgos. Para modelizarlos, hemos utilizado las diferentes funcionalidades que ofrecen los paquetes *fracdiff*, *arfima*, *forecast*, *lmtest* y *pracma* de R Studio.

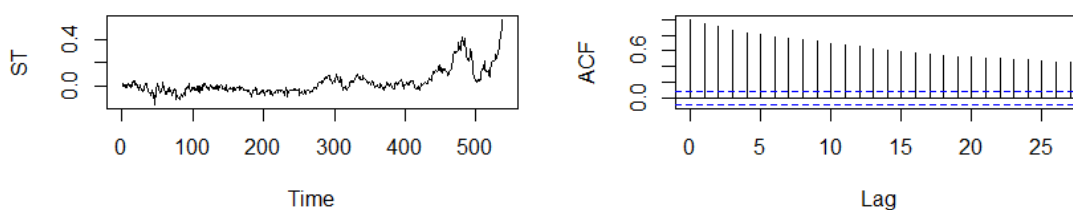
La elección de estos modelos viene dada por una serie de factores, entre los que se pueden destacar aquellos señalados por Darmawan y Oktavianus (2017). Los modelos ARFIMA pueden gestionar de una mejor forma la no estacionariedad de los datos, además de tener mejor *accuracy* en el manejo de los datos con memoria larga. Esto se debe a que los diferentes valores que puede tomar el parámetro d puede explicar la correlación en datos con memoria corta, memoria larga y estructuras simultáneas (Darmawan y Oktavianus, 2017).

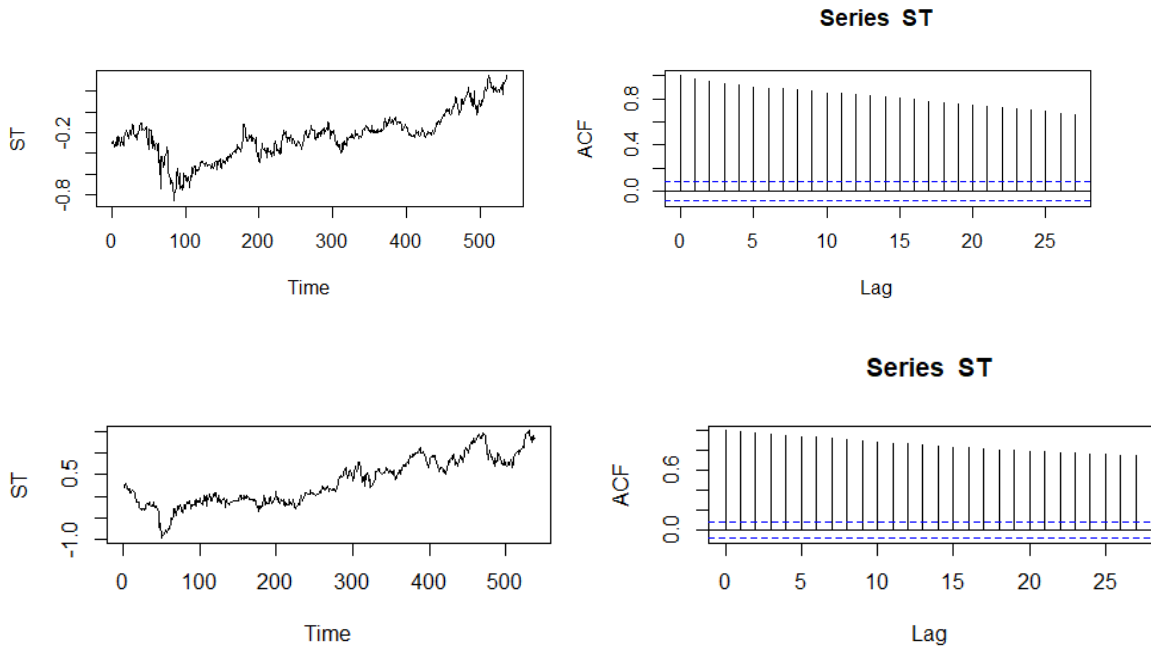
El procedimiento seguido es rutinario, obteniendo la estructura de la serie temporal usando la función *arfima* del paquete *arfima*. Estudiamos el modelo, su estructura y sus residuos, con el fin de ver si es interesante para el estudio. A continuación, se estima el parámetro de diferenciación. Por último, indicando los parámetros del modelo, simulamos la serie temporal a partir de un punto de la serie temporal original. Con el fin de ver los resultados, vamos a analizar, de las estrategias expuestas anteriormente, tres que pueden considerarse como sencillas y tres que son complejas. Entre las sencillas, se han escogido un *butterfly* (M2U2Z2) de tres meses, uno de doce (Z3Z4Z5) y un *condor* de seis meses (Z2M3Z3M4). Entre las complejas, se han escogido 2 *butterflies* de tres meses y uno de seis (M2U2Z3M3), dos *butterflies* de 6 meses y uno de 12 (Z2M3Z3Z4) y dos *condors* de 6 meses y uno de 12 (Z2M3Z3M4Z4Z5).

- Estrategias simples:

De entre las estrategias simples, se han escogido los siguientes *spreads*: M2U2Z2, Z3Z4Z5 y Z2M3Z3M4. Como mencionábamos en la introducción del apartado, ARFIMA trabaja mejor con procesos de memoria larga. Para ello, vamos a ver la estructura que tiene la serie temporal de los diferentes *spreads*, junto a su función de autocorrelación simple.

Ilustración 34 - Gráficos y FAS de las estrategias simples





Como se puede observar, todos son procesos de memoria larga, aunque los hay algunos más pronunciados que otros. En el tercer caso vemos como el decaimiento de los retardos es mucho menor que en los casos anteriores, por lo que podremos testear la hipótesis si en este tipo de contrato específico la modelización ARFIMA es mejor.

Para la estimación del modelo, se ha usado la función *arfima*, cuyos parámetros son la serie temporal, el orden del modelo (para estimarlo, hemos hecho una simulación sencilla, usando la función *auto.arima*) y el número de inicios de cada parámetro.

En el primer *spread*, el modelo era un ARIMA(1,1,0), un (2,1,1) para el segundo y (3,1,1) para el último.

Estos han sido los coeficientes de los modelos realizados, siguiendo el orden anteriormente mencionado:

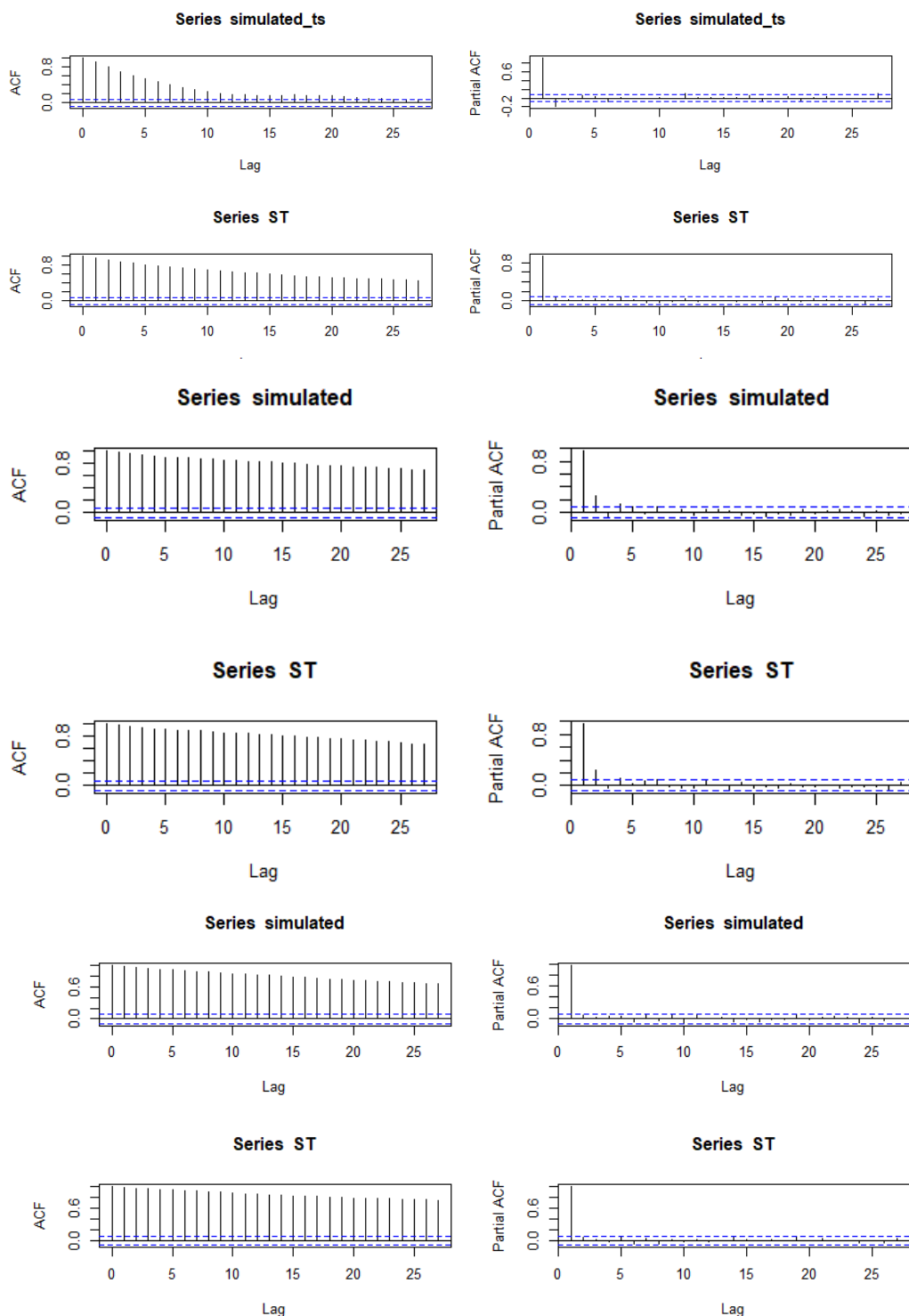
Ilustración 35 - Modelos

	Estimate	Std. Error	Th.	Std. Err.	z-value	Pr(> z)
phi(1)	0.5952041	0.0547540		0.1132696	10.87051	< 2e-16 ***
d.f	0.4830348	0.0267114		0.1099091	18.08346	< 2e-16 ***
Fitted mean	0.0711587	0.1763236		NA	0.40357	0.68653
	Estimate	Std. Error	Th.	Std. Err.	z-value	Pr(> z)
phi(1)	1.232469956	0.214872055		0.237970586	5.73583	9.7035e-09 ***
phi(2)	-0.275814632	0.176238475		0.192431806	-1.56501	0.117581
theta(1)	0.554533436	0.230173847		0.192093727	2.40919	0.015988 *
d.f	-0.765222757	0.324653662		0.305466856	-2.35704	0.018421 *
Fitted mean	0.002801253	0.000596657		NA	4.69491	2.6672e-06 ***
	Estimate	Std. Error	Th.	Std. Err.	z-value	Pr(> z)
phi(1)	0.36443249	0.12557866		0.14973310	2.90203	0.00370758 **
phi(2)	0.17384481	0.06453681		0.07242237	2.69373	0.00706572 **
phi(3)	-0.15739322	0.04516636		0.04649731	-3.48474	0.00049261 ***
theta(1)	0.65725100	0.20270668		0.23415807	3.24237	0.00118538 **
d.f	-0.01471867	0.14926487		0.13762267	-0.09861	0.92144972
Fitted mean	0.00115977	0.00103682		NA	1.11858	0.26331852

En los gráficos anteriores, podemos observar los diferentes parámetros de los modelos ARFIMA, donde el parámetro ϕ hace referencia a los parámetros de AR, θ a los parámetros de MA y $d.f$ es el parámetro de diferenciación. Como podemos observar, prácticamente la mayoría de los parámetros de la parte AR implican que el valor de la serie temporal en un instante del tiempo tienen un relación lineal positiva con el valor anterior. Con los parámetros de la parte MA pasa exactamente lo mismo, pero con el error, en lugar del valor de la serie en un instante del tiempo.

También se han realizado las funciones de autocorrelación simple y autocorrelación parcial de la serie simulada y la serie original. Estos son los resultados de cada contrato:

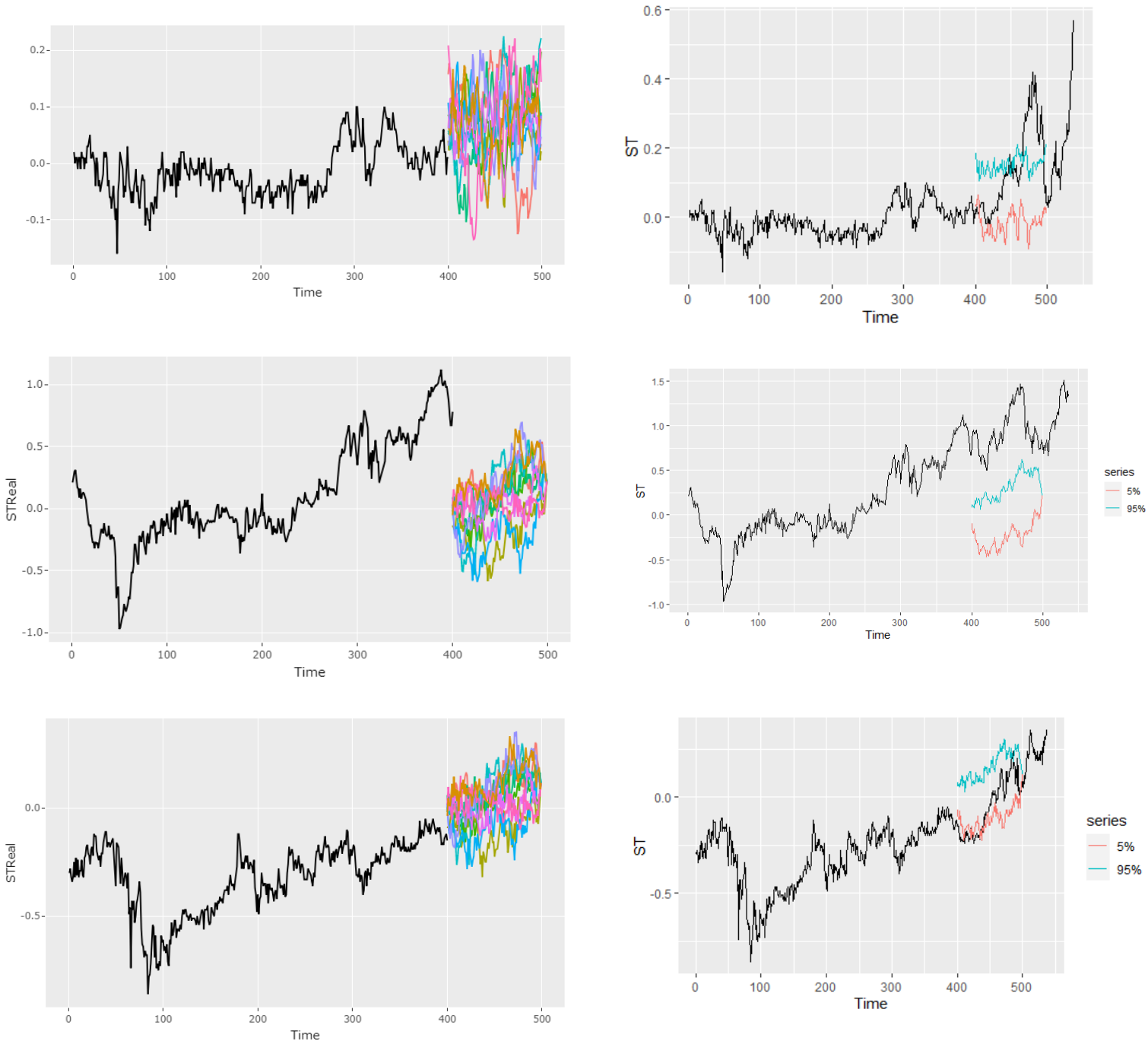
Ilustración 36 - FAS y FAP de la serie y de la serie simulada



A excepción de la primera serie temporal, observamos como las funciones de autocorrelación simple y autocorrelación parcial del segundo y tercer *spread* son muy similares a las de la serie simulada. Sin embargo, la serie temporal del primer *spread* guarda diferencias con la serie temporal simulada, sobre todo en términos de estas funciones.

Por último, se ha creado un conjunto de series temporales simuladas, que parten de un punto de la serie temporal real, para ver si el modelo verdaderamente capta las dinámicas, junto al dibujo de los intervalos de confianza (al 5% y al 95%) de por dónde irían las series. Estos son los resultados para cada *spread*, en el orden anteriormente indicado.

Ilustración 37 - Modelización a través de los parámetros obtenidos.

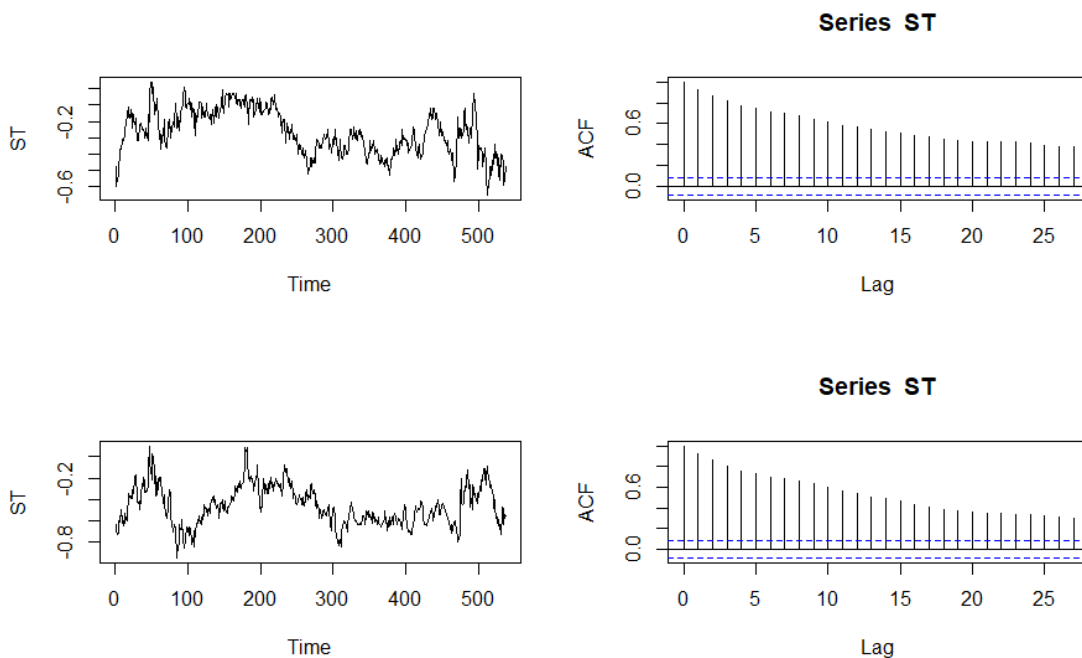


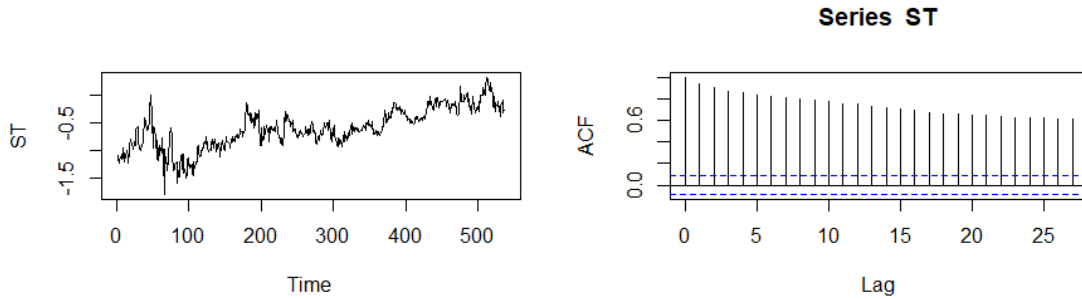
En los dos primeros spreads observamos con claridad como el conjunto de series simuladas no capta de la mejor manera las dinámicas de estos *spreads*. En el primero, vemos como los valores de la simulación parece que tienen dispersión. Además, en este aspecto, los intervalos de confianza no capta las subidas del último spread, aunque hay que mencionar que es justo en el momento de la simulación cuando empieza la fuerte subida. Caso similar tenemos en el segundo *spread*, donde no se capta nada bien la dinámica o tendencia de la serie a lo largo de todo el proceso. El tercer caso es diferente, ya que sí que parece que los modelos captan las dinámicas, idea reforzada en el gráfico de los intervalos de confianza, donde se recoge la tendencia del diferencial de precios, tanto en la serie original como en el conjunto de series simuladas.

- Estrategias complejas:

De entre las estrategias simples, se han escogido los siguientes *spreads*: M2U2Z2M3, Z2M3Z3Z4 y Z2M3Z3M4Z4Z5. Como mencionábamos en la introducción del apartado, ARFIMA trabaja mejor con procesos de memoria larga. Para ello, vamos a ver la estructura que tiene la serie temporal de los diferentes *spreads*, junto a su función de autocorrelación simple.

Ilustración 38 - Serie temporal y FAS





Al igual que en el caso anterior, tenemos diferentes estructuras en estos tipos de estrategias. Hay series donde la tendencia se ve claramente, como es el caso de la última, mientras que hay otros *spreads* donde no hay una estructura tan clara y las fluctuaciones son más evidentes. Vamos a continuar siguiendo la misma estructura que seguimos en el caso anterior: Indicando los modelos, mostrando las funciones de autocorrelación de la serie y, por último, simulando un conjunto de series.

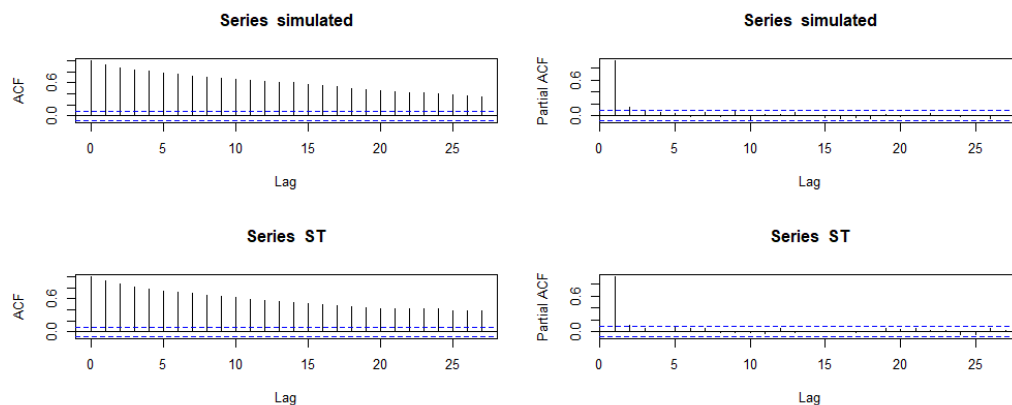
En cuanto a modelos, estos son los parámetros: Para el primer *spread* (M2U2Z2M3), tenemos un ARIMA(2,1,2), un ARIMA(1,1,1) para el segundo *spread* y un ARIMA(0,1,3) para el tercer y último *spread*.

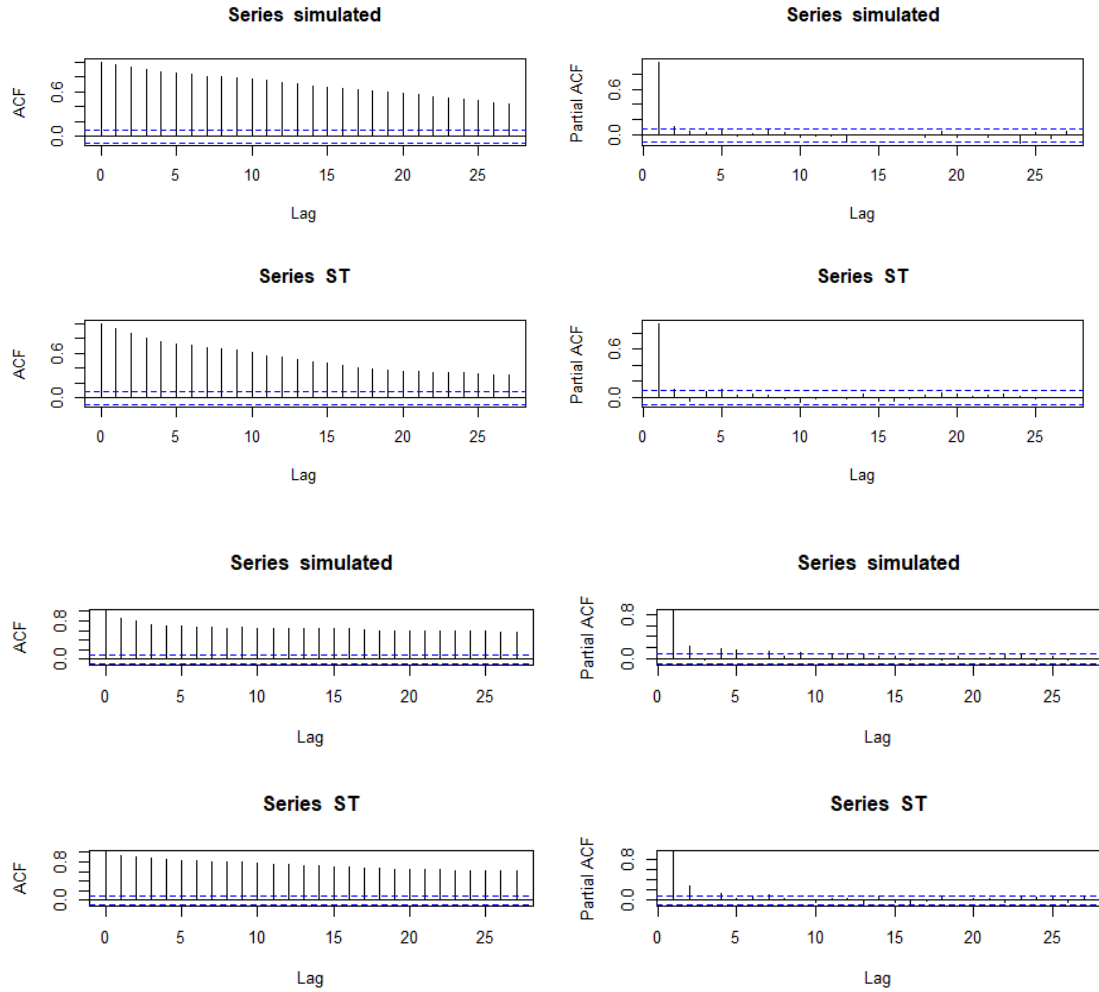
Ilustración 39 - Modelización ARIMA

	Estimate	Std. Error	Th.	Std. Err.	z-value	Pr(> z)
phi(1)	0.544225691	0.060959003	7.783957862	8.92773	< 2.22e-16	***
phi(2)	0.439246045	0.055118055	7.685701009	7.96919	1.5972e-15	***
theta(1)	0.547904964	0.057332364	7.697402159	9.55664	< 2.22e-16	***
theta(2)	0.451505076	0.057332364	7.694153625	7.87522	3.4014e-15	***
d.f	-0.197243345	0.052858301	0.093504011	-3.73155	0.00019031	***
Fitted mean	-0.000337372	0.000332566	NA	-1.01445	0.31036651	
phi(1)	9.68875e-01	1.51954e-02	1.50499e-02	6.37611e+01	< 2.22e-16	***
theta(1)	9.99572e-01	6.32376e-07	1.29769e-03	1.58066e+06	< 2.22e-16	***
d.f	-1.40839e-01	4.61086e-02	4.65833e-02	-3.05452e+00	0.0022542	**
Fitted mean	-1.34720e-04	2.83531e-04	NA	-4.75150e-01	0.6346808	

Las funciones de autocorrelación simple y parcial son las siguientes:

Ilustración 40 - FAS y FAP de la serie y de la serie simulada

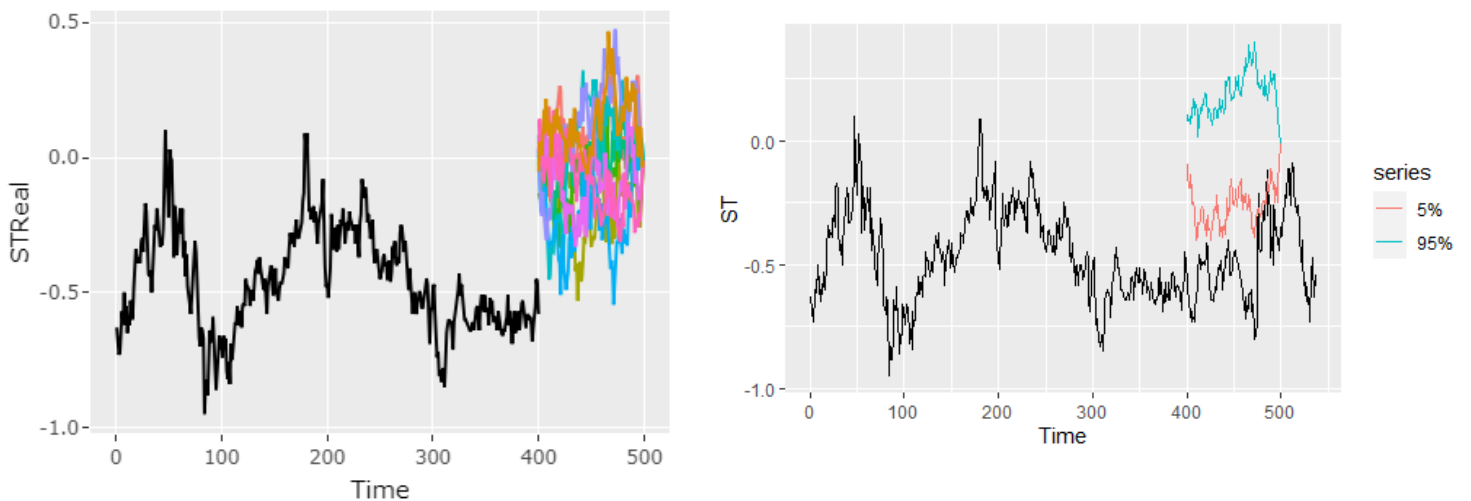


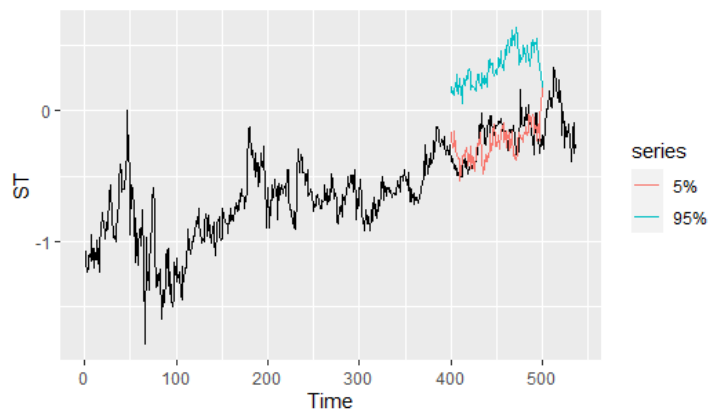
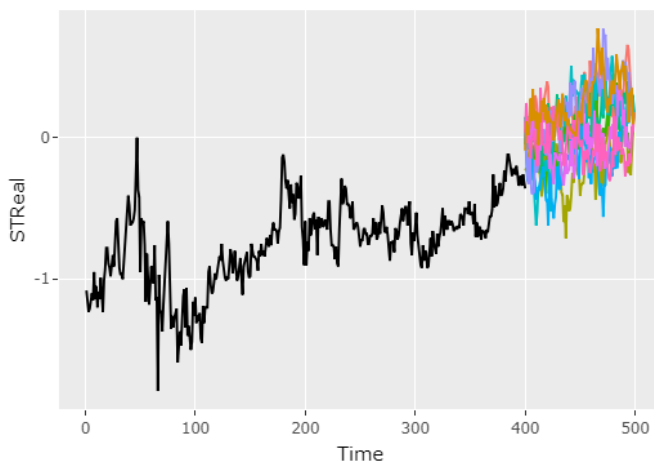
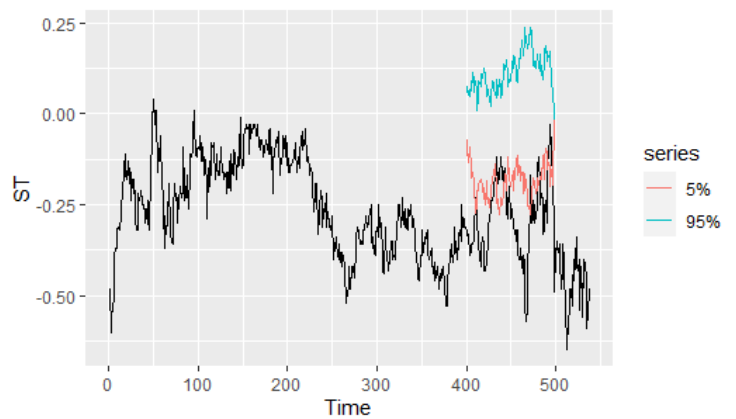
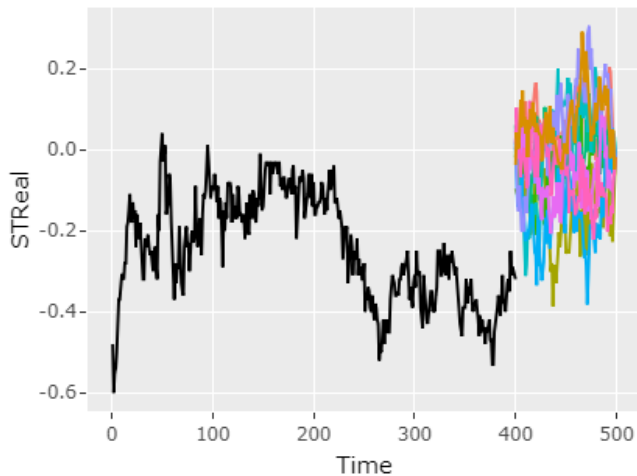


En este caso, podemos observar que, tanto las funciones de autocorrelación simple como las funciones de autocorrelación parcial de la serie simulada y la serie real son muy similares, no mostrándose apenas diferencias.

Por último, aquí se muestran las simulaciones y los intervalos de confianza.

Ilustración 41 - Simulación basada en los parámetros





Como podemos observar, y como también ocurría en aquellos casos de las estrategias simples, en casos en los que hay fluctuaciones más severas, sin existir una tendencia clara y concisa, el modelo no capta las dinámicas de la mejor manera. Sin embargo, en el último caso, donde se observa una tendencia clara, se capturan mejor las dinámicas. Lo podemos ver no solo en todas las series que se simulan y su similar estructura, si no que se puede captar también en los intervalos de confianza.

Para concluir, podemos destacar una serie de puntos en este apartado: Por un lado, y un aspecto que es importante, que una estrategia sea sencilla o compleja (en cuanto al número de contratos se refiere), no implica que sea más sencillas de modelizar, o que sus dinámicas sean más fáciles de captar. Hemos visto como estrategias complejas donde las dinámicas han sido captadas por la simulación de las series de una forma en cierto modo buena y también hemos observado estrategias sencillas en las que las dinámicas eran difíciles de captar. Y en segundo lugar, hemos visto como las estrategias con una tendencia clara y continuada muestran las mejores simulaciones en cuando a series, y su serie temporal se encuentra dentro del intervalo de confianza marcado.

4.3. Dimensión fractal de los *spreads*.

El objetivo que tenemos en este apartado es estudiar el coeficiente de Hurst de cada serie temporal, los *spreads* en este caso, y ver sus tendencias en el sentido de la aleatoriedad o la reversión a la media. Para ello, usaremos el coeficiente de Hurst. Este coeficiente es una medida de independencia de las series temporales que sirve como elemento para distinguir series fractales. Posteriormente a este análisis, surgió el análisis de rango reescalado (R/S), definido como un método para evaluar la ocurrencia de eventos poco comunes, dando origen a una herramienta ideal en procesos financieros.

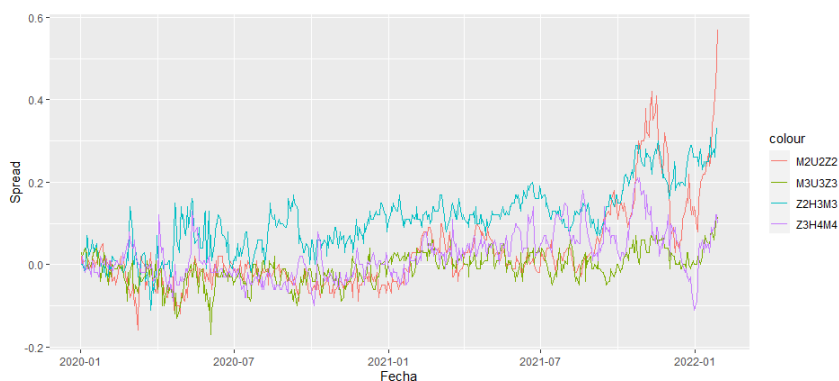
El coeficiente de Hurst (en adelante, coeficiente H), puede tomar valores en el rango $[0,1]$, teniendo diferentes interpretaciones.

- $H=0.5$: Ruido blanco. Proceso aleatorio e independiente.
- $0.5 < H \leq 1$: El proceso es persistente y correlacionado. Hay efectos de memoria a largo plazo.
- $0 \leq H < 0.5$: Comportamiento antipersistente o anti-correlacional. Este tipo de sistemas tiende a regresar constantemente al lugar de procedencia.

Vamos a ir viendo los diferentes *spreads*, graficando todas sus series temporales, y sus coeficientes de Hurst, teniendo así una versión integral del proceso.

- *Butterflies* de 3 meses.

Ilustración 42 - Serie temporal y coeficiente de los *butterflies* de tres meses

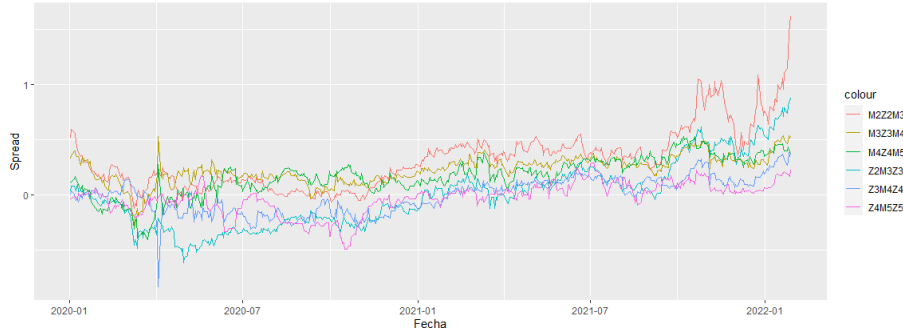


Spread	H
M2U2Z2	0,365
Z2H3M3	0,263
M3U3Z3	0,217
Z3H4M4	0,351

Todos los coeficientes son menores que 0,5, indicando comportamientos antipersistentes en la serie, especialmente en el segundo y el tercer spread.

- *Butterflies* de 6 meses.

Ilustración 43 – Series temporales y coeficiente de los *butterflies* de 6 meses

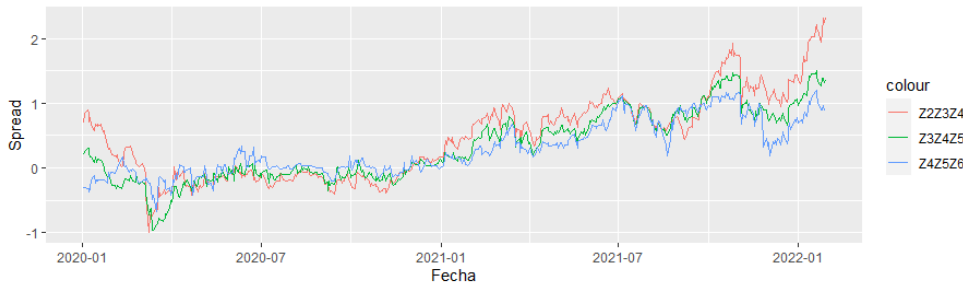


Spread	H
M2Z2M3	0,544
Z2M3Z3	0,502
M3Z3M4	0,476
Z3M4Z4	0,416
M4Z4M5	0,308
Z4M5Z5	0,415

Vemos como los comportamientos son muy similares, indicando comportamientos de ruido blanco en diferentes *spreads*, especialmente en el segundo, donde el valor es muy cercano a 0,5. En este caso no se muestran de forma clara comportamientos de memoria a largo plazo

- *Butterflies* de 12 meses.

Ilustración 44 – Series temporales y coeficiente de los *butterflies* de 12 meses

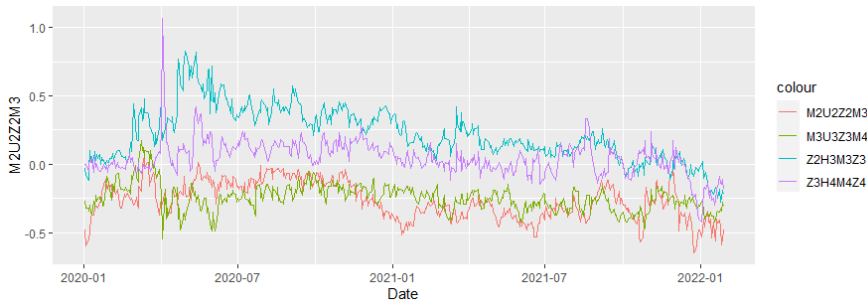


Spread	H
Z2Z3Z4	0,522
Z3Z4Z5	0,445
Z4Z5Z6	0,355

Vemos comportamientos algo diferentes entre series, aunque el primer y el segundo contrato tienen valores muy cercanos a 0,5, especialmente en el primero. El último se aleja más de dicho valor, indicando cierto comportamiento antipersistente.

- 2 *butterflies* de 3 meses y un *butterfly* de 6 meses.

Ilustración 45 - Series temporales y coeficiente de los 2 *butterflies* de 3 meses y un *butterfly* de 6 meses

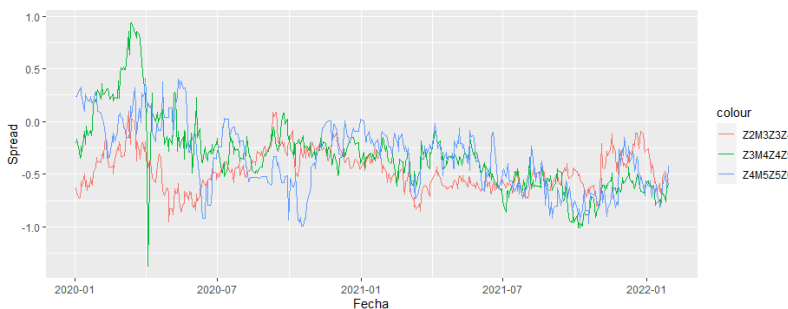


Spread	H
M2U2Z2M3	0,398
Z2H3M3Z3	0,372
M3U3Z3M4	0,316
Z3H4M4Z4	0,419

Los comportamientos de estos *spreads* nos pueden indicar que existen procesos de reversion a la media, especialmente en el tercer contrato. Además, si nos fijamos en las gráficas, podemos observar cómo los precios fluctúan alrededor de un valor, teniendo en cuenta la idiosincrasia de cada *spread*.

- 2 *butterflies* de 6 meses y un *butterfly* de 12 meses.

Ilustración 46 - Series temporales y coeficiente de los 2 *butterflies* de 6 meses y un *butterfly* de 12 meses

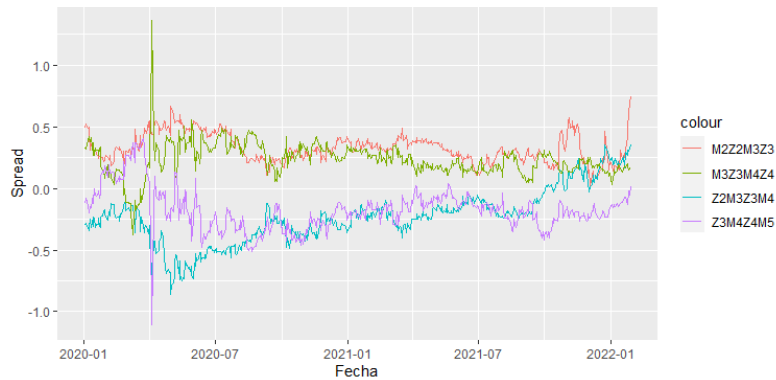


Spread	H
Z2M3Z3Z4	0,474
Z3M4Z4Z5	0,392
Z4M5Z5Z6	0,269

Los comportamientos en este caso tienen dinámicas muy similares a las que observábamos en casos anteriores, si nos fijamos en sus coeficientes. En este caso, podemos destacar el valor más bajo del coeficiente H, el del tercer contrato.

- *Condors* de 6 meses.

Ilustración 47 - Series temporales y coeficiente de los condors de 6 meses

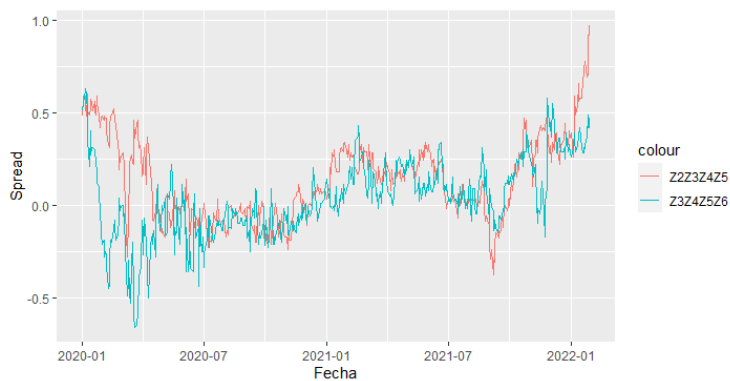


Spread	H
M2Z2M3Z3	0,387
Z2M3Z3M4	0,429
M3Z3M4Z4	0,335
Z3M4Z4M5	0,409

Aunque tengamos contratos de distinta tipología (cambiamos de estrategias, de *butterflies* a *condors*), son valores de coeficientes que están por debajo de 0,5, aunque no están muy alejados, obviando el caso del tercer contrato, que está un poco más desviado.

- *Condors* de 12 meses.

Ilustración 48 - Series temporales y coeficiente de los condors de 12 meses

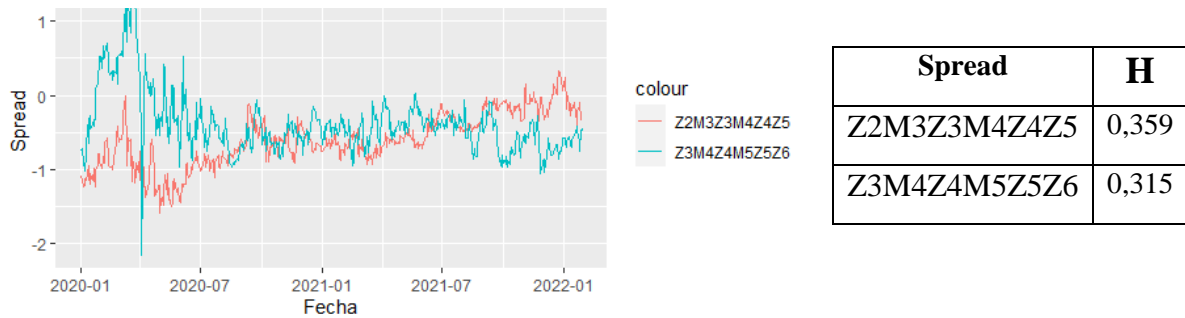


Spread	H
Z2Z3Z4Z5	0,498
Z3Z4Z5Z6	0,338

Vemos como hay un *spread* con un coeficiente más bajo, mientras que el otro tiene un comportamiento más cercano al ruido blanco, debido a su cercanía a 0,5.

- 2 *Condors* de 6 meses y un *condor* de 12 meses.

Ilustración 49 - Series temporales y coeficientes de los dos *condors* de 6 meses y un *condor* de 12 meses



Ambos *spreads* siguen las dinámicas de sus predecesores, indicando comportamientos que tienen tendencia a la reversión a la media.

Como conclusión, podemos indicar que todos los contratos, según sus coeficientes de Hurst, tienen comportamientos muy similares, con coeficientes que varían entre 0,25 y 0,5, no subiendo ningún coeficiente de 0,5, o si sube, en un grado muy leve. También podemos destacar que, aquellos *spreads* que tienen el contrato Z6, que es aquel que vence en diciembre de 2026, tienen coeficientes muy bajos. Por último, también se puede destacar que hay *spreads* que, aunque tenga contratos muy similares, no implica que tengan dinámicas y estructuras similares.

Para terminar con este TFG, vamos a exponer las conclusiones en el siguiente apartado, mostrando los *insights* obtenidos a lo largo de todo este trabajo.

5. Conclusiones.

Con el fin de terminar este Trabajo de Fin de Grado, se van a exponer las conclusiones obtenidas. El procedimiento para presentarlas va a ser el siguiente: Se va a ir realizando un recorrido por lo que se ha ido presentando en este TFG, abarcando los diferentes aspectos y ahondando en aquellos sobre los que se han obtenido las principales conclusiones.

En primer lugar, se introdujo la temática de este trabajo, aclarando los objetivos que se tenían, junto a la metodología a utilizar para conseguir los objetivos: El análisis fractal, los modelos ARFIMA y el análisis de componentes principales aplicado a los mercados de futuros del petróleo. Para continuar, se realizó una aproximación a los mercados de futuros y las técnicas utilizadas, tanto de inversión como de *machine learning*. Además, tras esta aproximación, se realizó un breve estudio sobre el *machine learning* en estos mercados, prestando especial atención a aquellas técnicas que aquí se tratan. Como primera conclusión que se puede extraer, se destaca la gran cantidad de técnicas utilizadas en estos mercados, junto al gran marco temporal sobre el que se han ido aplicando. Además, podemos destacar que los resultados son muy amplios, obteniendo desde técnicas rentables hasta resultados que nada tienen que ver con la rentabilidad, pero sí con otros aspectos de estos mercados como la liquidez.

Para continuar, se realizó una introducción de los datos utilizados y su obtención, indicando los diferentes procedimientos para llegar a los datos que han servido de alimentación al código. En este apartado, como segunda conclusión y basándonos en el análisis de las técnicas mencionado en el párrafo anterior, se pudo observar los diferentes tipos de datos que alimentan los análisis de estos mercados: Desde el precio *spot* (actual) del petróleo, hasta diferentes tipos de *spreads* que estudian diferenciales de precios entre productos de diferentes tipología (queroseno y gasolina, por ejemplo) o productos obtenidos en diferentes lugares (petróleo producido en Europa y petróleo producido en Estados Unidos).

Por último, se realizó el análisis de los datos usando RStudio y aplicando las técnicas que hemos venido mencionando. En cuanto al análisis de componentes principales, y debido a los resultados obtenidos, podemos concluir que las primeras posiciones de las curvas son aquellas que lideran el precio del petróleo. Esta idea apoya a una idea extendida en los mercados de derivados, consistente en que el precio de futuro de un producto tiende a converger con su precio *spot*. Además, se puede obtener que hay diferentes posiciones que forman grupos de información, por las posiciones en las que aparecían en los gráficos. En cuanto a modelos ARFIMA, podemos destacar que, aunque dependiendo del contrato, hay modelos que se adaptan de buena manera a las dinámicas de las series (*spreads*), según hemos visto en las gráficas que tenemos y en los gráficos de los cuantiles. Por último, viendo la dimensión fractal observada a través del exponente de Hurst, podemos concluir que, generalmente, hay series con conmutación a largo plazo entre valores altos y bajos en pares adyacentes, indicando que un valor alto probablemente vendrá seguido de un valor bajo, y viceversa. También se observan algunos casos en los que la serie es no correlacionada, con un exponente de valor extremadamente cercano a 0,5.

6. Referencias.

- Aranzadi, C., Merino, M. Á. L., & Simarro, R. P. (2011). Los Nuevos mercados energéticos.
- ARFIMA | Stata. (s. f.). <https://www.stata.com/features/overview/arfima/>
- Bucca, A. & Cummins, M. (2011). Quantitative Spread Trading on Crude Oil and Refined Products Markets. *SSRN Electronic Journal*, 12, 1857-1875. <https://doi.org/10.2139/ssrn.1932471>
- Bucca, A. & Cummins, M. (2012). Quantitative Spread Trading on Crude Oil and Refined Product Markets. *Quantitative Finance*, 12, 1857-1875. <https://doi.org/10.1080/14697688.2012.715749>
- Cajueiro, D. & Tabak, B. (2007). Are the crude oil markets becoming weakly efficient over time? A test for time-varying long-range dependence on price and volatility. *Energy Economics*. <https://doi.org/10.1016/j.eneco.2006.06.007>
- Cavaliere, G., Nielsen, M. R. & Taylor, A. R. (2015). Bootstrap score tests for fractional integration in heteroskedastic ARFIMA models, with an application to price dynamics in commodity spot and futures markets. *Journal of Econometrics*, 187, 557-579. <https://doi.org/10.1016/j.jeconom.2015.02.039>
- CNMV - Futuros. (s. f.). <https://www.cnmv.es/Portal/inversor/Futuros.aspx>
- Castellanos, E. Mercado Oficial de Opciones y Futuros Financieros en España. (2019). Manual de Futuros.
- ¿Cuál es la diferencia entre el petróleo Brent y el West Texas Intermediate (WTI)? El Orden Mundial - EOM. <https://elordenmundial.com/diferencia-petroleo-brent-wti/?nab=0>
- Ferreira Neves, R. & Nadkarni, J. (2018). Combining NeuroEvolution and Principal Component Analysis to trade in the financial markets. *Expert Systems with Applications*, 103, 184-195. <https://doi.org/10.1016/j.eswa.2018.03.012>
- Fundamentals of Futures and Options Markets Paperback. (2022b). Pearson.
- Haugom, E & Ray, R. (2017). Heterogeneous traders, liquidity and volatility in crude oil futures markets. *Journal of Commodity Markets*, 5, 36-49. <https://doi.org/10.1016/j.jcomm.2017.01.001>

- Hyndman, R. & Athanasopoulos, G. (2014). *Forecasting: principles and practice*.
<https://www.otexts.org/fpp>
- ICE Futures and Options | Product Guide. (s. f.). <https://www.theice.com/products/219>
- Jacobsson, M. (2015). Forecasting commodity futures using Principal Component Analysis and Copula.
- Kang, S. H. & Yoon, S. M. (2013b). Modeling and forecasting the volatility of petroleum futures prices. *Energy Economics*, 36, 354-362.
<https://doi.org/10.1016/j.eneco.2012.09.010>
- Lubnau, T. & Todorova, N. (2015). Trading on mean-reversion in energy futures markets. *Energy Economics*, 51, 312-319.
<https://doi.org/10.1016/j.eneco.2015.06.018>
- Mitra, S. K. (2012). Is Hurst Exponent Value Useful in Forecasting Financial Time Series? *Asian Social Science*, 8(8). <https://doi.org/10.5539/ass.v8n8p111>
- Peña, D. (2005). *Análisis de series temporales*. Alianza.

7. Anexo I: Código.

En este Anexo I se realizará una breve explicación del código utilizado, mencionando las librerías usadas y las principales características y pasos para conseguir lo que se ha realizado.

En primer lugar, cabe destacar que el programa que se ha utilizado es RStudio. RStudio es un entorno y un lenguaje de programación que tiene enfoque en el análisis estadístico. La utilización de este *software* por delante de otros como Python o Matlab viene por el uso que se ha dado a este lenguaje a lo largo de todo el grado, ya que se ha hecho hincapié en este *software* en lugar de otros como los ya mencionados, siendo por tanto RStudio un programa con el que se tenía mayor grado de familiaridad.

En cuanto a código, se realizó la carga del conjunto de datos con la librería *readxl*. De todos los *spreads* existentes, se dividieron por tipología usando *dplyr*. Tras esto, se realizaron las gráficas de todos los *spreads*, junto a las funciones de autocorrelación simple y parcial, para acercarnos a la posible estructura de nuestro conjunto de datos. Además, se obtuvieron diferentes estadísticos sobre las curvas y los *spreads*, junto a diferentes gráficos de cajas y bigotes de estos conjuntos de datos, con el fin de seguir acercándonos a la estructura ya mencionada. Para la realización de estos gráficos, se ha usado principalmente la librería *ggplot2* y *plotly*.

En cuanto al análisis PCA, se ha usado, en primer lugar, la librería *corrplot* para obtener el gráfico de correlaciones de las variables que teníamos. Se ha utilizado la función *prcomp* para obtener las componentes principales y todas las matrices (matriz de rotación, etc.) que son necesarias en este análisis. Además, se han realizado diferentes gráficos usando la función *biplot* y las librerías *factoextra* y *FactoMineR*. En cuanto a la reconstrucción de las curvas, se ha usado un bucle que expongo a continuación:

```
for (i in 1:763) {  
  plot(1:11, as.matrix(curvas_sinf[i,]), type = "l", xlab = "Date", ylab = "Price", ylim = c(55,70))  
  curvas.rec <- pcs$center+(pcs$x[i,1:2]]*%t(pcs$rotation[,1:2]))*ifelse(pcs$scale, pcs$scale, 1)  
  #write.csv(curvas.rec, gsub(".csv", "_processed.csv", i))  
  write.table(curvas.rec, file = "PCS_REC.txt", append = TRUE)  
  lines(1:11, curvas.rec[1,], col="red")  
  legend(x=5, y=70, legend = c("Real", "Reconstruccion"), col = c("black","red"), lty = 1, cex = 0.5)  
  Sys.sleep(0.5)  
}
```

En el bucle, se recrean las curvas para todas las posiciones, haciendo el proceso inverso al que se realiza en el análisis de componentes principales. La recreación se ha guardado en un fichero de texto (se ha elegido esta opción, aunque también podamos observar la extracción en un fichero .csv, por temas de formato) que ha sido exportado en un fichero Excel para su tratamiento y posterior volcado en la herramienta.

Por último, para la realización de los *frames* que anteriormente se exponen, se ha utilizado la librería *gganimate*.

En cuanto a los modelos ARFIMA, se han usado las librerías *fracdiff*, *arfima*, *forecast* y *pracma*. Se ha realizado una estimación de modelos ARIMA para cada uno de los *spreads*. Tras esto, se han chequeado los residuos del modelo y usando la función *sim_from_fitted*, se ha simulado una serie temporal usando los parámetros del modelo. Se han dibujado las funciones de autocorrelación simple y parcial de la gráfica original y la simulada con el fin de ver si las estructuras, en términos de autocorrelación, eran similares.

Tras esto, se crea una matriz con un número predeterminado de observaciones, y usando la misma función anterior para la simulación, se realiza un bucle para la obtención de las diferentes series y su posterior dibujo. Además, se extraen los intervalos de confianza al 5% y al 95%, para también graficarse junto a las series simuladas.

Por último, se ha utilizado el paquete *pracma* para tratar los coeficientes de Hurst. Este ha sido el procedimiento más sencillo, ya que se ha usado únicamente la teoría y la función *hurstexp* para obtener las conclusiones.