



**FACULTAD DE CIENCIAS
ECONÓMICAS Y EMPRESARIALES**

**Análisis de la crisis energética europea y las
energías renovables: Comparativa de un modelo
ARIMA, GARCH y GRNN para la predicción del
RENIXX World Index.**

Autor: Alfonso Muñoz Izquierdo

Tutor: María Coronado Vaca

Madrid

Junio 2023

Resumen

La caducidad que suponen los combustibles fósiles, su volatilidad y su nexo con variables económicas que afectan a la sociedad implican la formulación de medidas gubernamentales para paliar sus efectos negativos. Mediante la estimación del índice de energías renovables RENIXX World se pretende proporcionar un método complementario al análisis de este sector para la anticipación a periodos de auge de las energías renovables como alternativa al sector energético tradicional.

Además, uno de los problemas más importantes y con mayor variedad de enfoques es el análisis y la predicción de series financieras. Como resultado, este campo de estudio está en constante expansión y la forma en que se aborda el problema está en constante evolución. Por ello, se ha realizado un estudio de las energías renovables y no renovables, del delicado contexto energético actual y una comparación de un modelo de enfoque más simple como el ARIMA, pasando por un GARCH y hasta algoritmos más evolucionados como un GRNN.

Tras la comparación bajo criterios como el RMSE o la capacidad de ajuste y extrapolación del mismo, se ha concluido que el modelo que mejor cumple estos baremos es la red neuronal.

Palabras clave

Predicciones, series temporales, modelos, machine learning, red neuronal, energías renovables, índice bursátil, RENIXX World, sector energético, crisis energética, ARIMA, GARCH, GRNN.

Abstract

The expiration of fossil fuels, their volatility and their link with economic variables that affect society imply the formulation of governmental measures to mitigate their negative effects. The estimation of the RENIXX World renewable energy index is intended to provide a complementary method for the analysis of this sector in order to anticipate boom periods for renewable energies as an alternative to the traditional energy sector.

In addition, one of the most important problems with the greatest variety of approaches is the analysis and prediction of financial series. As a result, this field of study is constantly expanding and the way in which the problem is approached is constantly evolving. Therefore, a study of renewable and non-renewable energies, the current

delicate energy context and a comparison of a simpler approach model such as ARIMA, going through a GARCH and up to more evolved algorithms such as a GRNN has been carried out.

After the comparison under criteria such as RMSE or the capacity of adjustment and extrapolation of the model, it has been concluded that the model that best meets these criteria is the neural network.

Keywords

Forecast, time series, models, machine learning, neural network, renewable energies, stock index, RENIXX World, energy sector, energy crisis, ARIMA, GARCH, GRNN.

Índice

1	<i>Introducción</i>	4
1.1	Objetivo	4
1.2	Justificación del tema objeto de estudio	4
1.3	Metodología.....	5
1.4	Estructura.....	5
2	<i>Marco teórico</i>	6
2.1	Las energías tradicionales: la energía no renovable	6
2.1.1	El petróleo	6
2.1.2	El carbón	9
2.1.3	El gas natural.....	11
2.1.4	La crisis energética: el Gas Natural	12
2.2	Las energías renovables.....	14
2.2.1	RENIXX WORLD INDEX	17
2.3	La predicción de series financieras.....	18
3	<i>Estudio empírico</i>	21
3.1	Datos: Las series temporales	21
3.2	Modelo ARIMA	24
3.3	Modelo GARCH.....	28
3.4	Modelo GRNN	29
4	<i>Aplicación de los modelos</i>	31
4.1	Preparación de los datos	31
4.2	Ajuste del modelo ARIMA.....	33
4.3	Ajuste del modelo GARCH.....	40
4.4	Ajuste Red Neuronal	41
5	<i>Conclusiones</i>	44
6	<i>Bibliografía</i>	45
7	<i>Anexo</i>	48
7.1	Código implementado.....	48

1 Introducción

1.1 Objetivo

El objetivo de este trabajo es predecir el comportamiento del valor bursátil del índice RENIXX WORLD a lo largo del próximo año, haciendo uso de un modelo ARIMA (Autoregressive Integrated Moving Average), GARCH (Generalized Autoregressive Conditional Heteroscedasticity) y GRNN (General Regression Neural Network) con el fin de proporcionar una herramienta complementaria a la estimación de los precios de combustibles fósiles y, extendiendo más allá su aplicación, anticipar las crisis energéticas. Además, se pretende analizar la situación actual de las energías renovables y no renovables y examinar el contexto energético europeo para comprender la relevancia de las últimas en un futuro próximo.

1.2 Justificación del tema objeto de estudio

Después de la época de volatilidad e incertidumbre alrededor de los precios de la energía desde 2021 hasta 2022 y las consecuencias que este auge de los precios tuvo en miles de familias, entre las que se encontraron prescindir del aire acondicionado, la necesidad de ayudas públicas para hacer frente a los gastos o el cese del uso del coche (EFE, 2022), no pude sino plantearme la razón última de esa situación precaria. El descubrimiento del origen provocó en mí una impotencia derivada de la dependencia energética de Europa, que a su vez me llevó a plantearme posibles soluciones. En este momento aparecieron las energías renovables entre otras alternativas.

Mediante la predicción del precio de las empresas punteras en el sector de energías renovables, los gobiernos podrían contar con una herramienta complementaria para ser capaces de anticiparse a las épocas de alza de precios de los combustibles fósiles (gas, petróleo y carbón) que conlleven una crisis energética y emplear fondos para paliar los efectos que esta volatilidad tiene sobre la población y en la economía. Aunque no se ahondará extensamente en la correlación entre ambos sectores, se espera que un aumento en los precios de las fuentes de energía no renovables se traduzca en un buen desempeño de las empresas componentes del RENIXX World index debido al empleo prioritario de este tipo de energías en sustitución o complemento de las energías tradicionales por ser menos costosas y tener mejores expectativas. Gracias al uso cada vez más difundido de las energías renovables, las compañías que las proveen mejorarían su tecnología y

abaratarían aún más costes, provocando una mejora en sus perspectivas futuras y por tanto en su desempeño bursátil. Además, se podrían prever problemas de dependencia energética y evitar conflictos entre naciones puesto que, si se utilizan cada vez más este tipo de energías, se dejará de adquirir energías como el gas natural a países como Rusia (causante de tensiones internacionales) y los gobiernos podrán destinar estos fondos públicos a la ayuda a la adaptación a estas nuevas energías y a paliar las consecuencias del aumento de precios de los combustibles fósiles.

1.3 Metodología

Para la consecución de los objetivos propuestos, se empleará un método deductivo y se combinarán el análisis cualitativo y cuantitativo. En primer lugar, se comenzará analizando de manera teórica el contexto energético actual mediante la revisión de la literatura existente sobre las energías tradicionales. Se ahondará en la reciente crisis energética, sus causas y consecuencias, para tratar de explicar la posible relación de causalidad y su relevancia, con el asentamiento de las energías renovables como solución. Además, se explicará la naturaleza de estas últimas y en particular, del RENIXX World Index, con especial foco en la predicción de índices bursátiles.

Posteriormente, se hará uso de la herramienta R para aplicar un modelo ARIMA, un modelo GARCH y un algoritmo GRNN a la serie temporal del precio del RENIXX World Index y predecir sus valores futuros. En este proceso, se han utilizado datos extraídos de Bloomberg.

Finalmente, se compararán las métricas de desempeño de ambos modelos y se realizará una selección del modelo con mejores cualidades para la predicción del índice de energías renovables. Esta conclusión aspira a ser extrapolable a otras series temporales relacionadas con las energías renovables para, a modo de complemento del análisis del desempeño de energías tradicionales, adelantarse a sus subidas o bajadas de precio, tanto en el mercado de renovables como de materias primas.

1.4 Estructura

El trabajo consta de 3 partes; una primera contextualización de los tipos de energía y la evolución de ambas, en la que se incluye un análisis de la crisis energética actual y la revisión de modelos predictivos aplicados, para facilitar la comprensión del trabajo; una

segunda parte, donde se aplicarán tres modelos predictivos al índice de energías renovables más reconocido y finalmente, un análisis mediante la comparación de los modelos y conclusiones.

2 Marco teórico

2.1 Las energías tradicionales: la energía no renovable

El uso de la energía por el ser humano se remonta a tiempos muy pasados, ya que esta es clave para la satisfacción de las necesidades básicas del mismo. De modo que, precisamente son energías tradicionales aquellas de las que el hombre se ha servido durante siglos y en las que, en consecuencia, tiene amplia experiencia. Esta experiencia y el desarrollo de técnicas optimizadas para su aprovechamiento han hecho que su coste sea inferior a las nuevas energías. Algunas de estas fuentes de energía son la combustión de leña y carbón vegetal o la energía hidráulica (Kuyper & Ramirez Morales, 2014). Pero dentro de las mismas, podemos diferenciar energías renovables de las no renovables. En este caso, la combustión de estos materiales convertiría a la leña y el carbón en fuentes no renovables, ya que esta actividad puede generar secuelas en la salud del ser humano y el planeta. Además, su cantidad es finita.

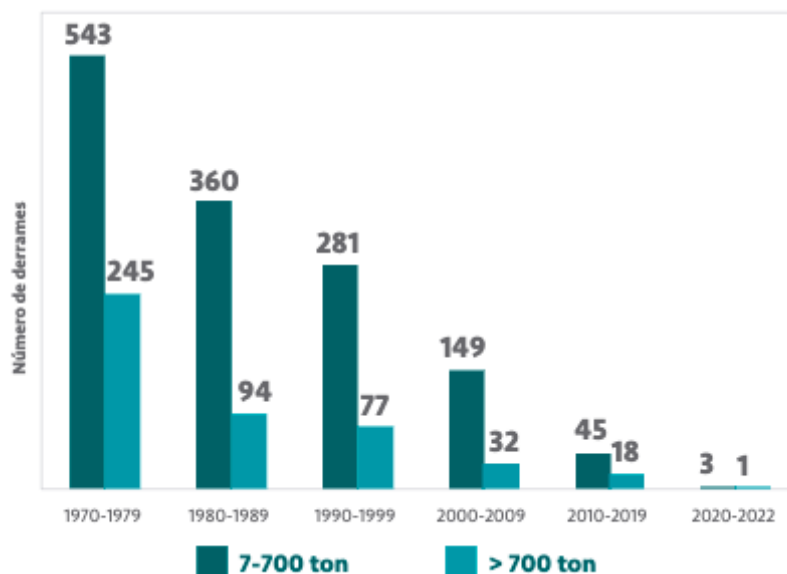
Dentro de los recursos energéticos, las energías no renovables son aquellas cuyos recursos son susceptibles de agotarse en el medio plazo o bien porque provengan de depósitos limitados o bien porque su proceso regenerativo abarque tanto tiempo que una vez agotados, sean casi imposible de obtenerse de nuevo (Kuyper & Ramirez Morales, 2014). Estos son el petróleo, el carbón y el gas natural.

2.1.1 El petróleo

El petróleo es el fruto de la transformación de seres vivos animales y vegetales que quedaron atrapados en el interior de la tierra. En consecuencia, para extraerlo son necesarias perforaciones en la corteza terrestre y marina (Kuyper & Ramirez Morales, 2014). El hecho de que el petróleo sea producto de materia ocluida durante muchos años hace que su disponibilidad sea finita y que su regeneración lleve largos periodos de tiempo. Además, su consumo es mucho más rápido que su producción, lo que lleva al agotamiento de esta fuente de energía.

Por si su extracción no fuera suficientemente laboriosa, su transporte también lo es, a la par que arriesgado. Este puede llevarse a cabo mediante mar o tierra y realizarse en grandes volúmenes, con el fin de minimizar costes. Pero al mismo tiempo que se minimizan, aumentan los riesgos asociados. Numerosos han sido los derrames, de mayor y menor magnitud, durante la historia. En el siguiente gráfico se puede observar como el número de derrames ha ido descendiendo a lo largo del tiempo, por la aparición de alternativas a esta fuente de energía y por el aumento de las precauciones tomadas a la hora de transportar el petróleo. No obstante, aunque 4 catástrofes recientes parezcan pocas, el nivel de contaminación y el impacto que tienen sobre el ecosistema marítimo de su alrededor es mucho mayor que el que pueda tener cualquier otra fuente energética.

Gráfico 1
Número de derrames globales de petróleo por década



Fuente: (Aponte, Torrejon-Managallanes, & Perez, 2022)

Uno de los últimos tuvo lugar el 15 de enero de 2022, cuando durante la descarga de la refinería La Pampilla se produjo una fuga de crudo por una cantidad equivalente a 6000 barriles (Aponte, Torrejon-Managallanes, & Perez, 2022). Esta inmensa catástrofe es el evento que nos sirve como prueba de las consecuencias medioambientales y para la salud en los organismos que habitan en todo el ecosistema costero, pues el petróleo pasa a distribuirse por toda la superficie marina hasta llegar a las costas. Estos efectos pueden tardar en difuminarse decenas de años y son el mayor impacto de esta energía no renovable. Otros peligros están relacionados con su extracción, pues el petróleo incluye

en su composición hidrocarburos gaseosos que son altamente inflamables y pueden provocar desastres ambientales como el ocurrido en 2010, cuando un escape provocó la explosión de una plataforma petrolífera estadounidense y la muerte de 11 trabajadores.

El petróleo es la fuente más utilizada para la producción de energía mundialmente, en 2022 ayudó a la generación del 30% de la energía global (Bp, 2023). La demanda de petróleo siempre ha sido muy elevada y la oferta muy austera, de ahí los altos precios de los barriles, convirtiendo al crudo en un activo cotizado y usado para la especulación. Observando la cotización de los futuros de petróleo Brent en los últimos 5 años, podemos identificar una caída pronunciada en el precio del crudo en 2020 debido a la prácticamente inexistente demanda de energía durante el confinamiento por el COVID. A partir de esta fecha, el precio del petróleo no ha dejado de aumentar hasta alcanzar su máximo en mayo de 2022, debido al conflicto entre Europa y Rusia por el que se extendieron sanciones que limitaban la importación de petróleo ruso y aumentaba así su precio global (Cinco Días, 2022) .

Gráfico 2

Evolución del precio del futuro sobre el petróleo Brent desde abril de 2019 a abril 2023



Fuente: (Investing, 2023)

La cotización del petróleo sirve como marcador de la economía y la inflación ya que es una materia prima en muchos sectores productivos y una subida de su precio implica la subida de costes en muchas industrias, que terminan en el precio del consumidor.

En el contexto actual y con foco en la demanda, el COVID-19 provocó una caída significativa; aunque la demanda regional se ha recuperado en gran medida, se prevé que los niveles anteriores a la pandemia no se alcancen hasta este año, debido principalmente

al bajo tráfico aéreo internacional. El máximo histórico del precio del crudo se alcanzó hace aproximadamente un año, sin embargo, se espera que la demanda mundial de crudo (uno de los principales factores de su precio) alcance su nivel máximo en los próximos dos a cinco años, impulsada principalmente por la electrificación y la eficiencia en todos los sectores, lo que puede llevar al precio del petróleo a batir récord histórico de nuevo. A partir de ese momento, se pronostica que esta demanda decaiga, llegando en 2050 a ser de un 35% a un 50% menor que los niveles actuales y hasta un 85% menor en su uso para la combustión (McKinsey & Company, 2022).

2.1.2 El carbón

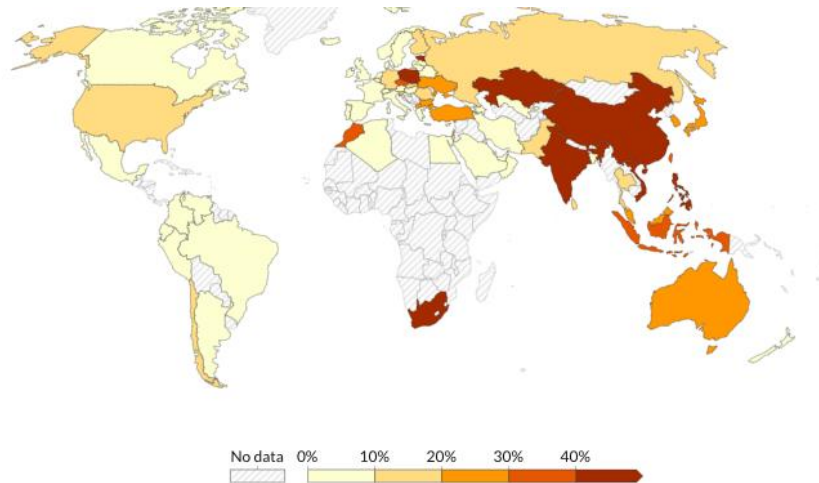
A diferencia del petróleo, el carbón se forma gracias a procesos que implican altas temperaturas y a partir de restos de plantas acumulados en el subsuelo. Es una fuente de energía mucho más abundante que la anterior, aunque también finita.

Su uso comenzó a extenderse con la revolución industrial y dio pie al desarrollo de nuevas tecnologías. Durante los siguientes siglos, el carbón era mayormente utilizado para la generación de electricidad y el uso del ferrocarril. En la actualidad, el carbón ha perdido importancia relativa a la que tenía tiempo atrás debido a los impactos negativos que genera sobre el medio ambiente y la aparición de fuentes más limpias, que analizaré con detenimiento posteriormente. En concreto, al ser el carbón el material más utilizado para la generación de energía por combustión resulta lógico analizar la cantidad de materias nocivas para la atmósfera que se expulsan. La más importante en comparación con las demás es el dióxido de carbono (CO₂). Después de su expulsión, el CO₂ llega a la troposfera desde donde emite radiación infrarroja de vuelta a la superficie terrestre, lo que hace que aumente la temperatura global generando lo que se conoce como efecto invernadero. De manera dosificada, este efecto no supone un peligro para la Tierra ya que ayuda al mantenimiento de una temperatura cálida, sin embargo, el exceso hace que esta aumente a mayor escala. Para controlar este efecto, se han tomado medidas que tienen que ver con la restricción de la emisión de dióxido de carbono y la descarbonización progresiva de las industrias (Aponte, Torrejon-Managallanes, & Perez, 2022).

No obstante, el carbón representaba un 26,9% de la producción energética global en 2021 (Bp, 2023). Entre los países en los que el carbón supone una mayor parte de la energía

total, se encuentran China, India o Sudáfrica donde el porcentaje que el carbón representa es superior al 40%, como se puede observar en el gráfico 3.

Gráfico 3
Porcentaje que representaba el carbón de la energía total de cada país en 2021.



Fuente: (Ritchie, Roser, & Rosado, 2022)

Cabe ahondar, en este apartado en concreto, en los objetivos de descarbonización y de impedir que la temperatura global no aumente más de 1,5 grados centígrados este siglo. Puesto que la transición a las energías renovables no es suficiente para alcanzarlos, un nuevo grupo de tecnologías llamadas CCUS (captura uso y almacenamiento de carbono) es capaz de almacenar de manera permanente el CO₂ expulsado, reduciendo en gran medida las emisiones de CO₂ (Repsol, 2023). Además, se están desarrollando posibles usos futuros que se podría dar a esta gran cantidad de CO₂ capturado.

El CCUS es hoy un nicho, pero se prevé que aumente notablemente. Para lograr las metas propuestas, es probable que esto requiera una aceleración significativa en comparación con las previsiones de crecimiento actuales. Estas tecnologías podrían actuar como una importante palanca de descarbonización para los sectores cuyo impacto sea más difícil de eliminar y como motor de arranque para el hidrógeno azul.

Para 2050, se espera que el 80 % de la adopción de CCUS se despliegue en los sectores del cemento, la siderurgia y la producción de H₂, aunque se espera que esté sujeta a incertidumbres sectoriales específicas y que muestre una diferenciación regional. Entre

los segmentos industriales, se prevé que los sectores de producción de hidrógeno azul, hierro, acero y cemento sean los que más utilicen el CCUS (juntos representan el 85% del total mundial en el escenario de mayor aceleración) (McKinsey & Company, 2022).

La adopción de CCUS variará según la región, principalmente en función de las ambiciones de descarbonización, los aspectos económicos del hidrógeno azul y la disponibilidad de palancas alternativas de descarbonización.

2.1.3 El gas natural

El gas natural es una mezcla de gases, compuesta principalmente por metano, que se han formado por descomposición en procesos que abarcan millones de años y que se encuentra en pozos, donde es naturalmente almacenada. (Aponte, Torrejon-Managallanes, & Perez, 2022).

Como el petróleo, para su extracción es necesaria la perforación de la tierra, aunque a diferencia de él, el transporte del gas natural resulta más seguro ya que no es posible usar el medio marítimo, al tener que almacenarse en estado líquido a muy bajas temperaturas y en grandes cantidades. En concreto, uno de los riesgos asociados a esta fuente fósil es precisamente este. Además, existe peligro de daño ambiental como es las filtraciones de gas natural a la atmósfera junto con su consiguiente efecto invernadero, la contaminación de acuíferos y el gran consumo de energía y de agua para su extracción, demasiado grande en comparación con la cantidad de energía que se puede crear a partir de la materia extraída.

El gas ha aumentado gradualmente su participación en la combinación energética y se espera que desempeñe un papel clave durante la transición gracias a su amplia gama de aplicaciones y a su uso como sustitutivo del petróleo. De hecho, su participación en la energía mundial fue de un 24,42% en 2021 (Bp, 2023).

Se prevé que la demanda de gas crezca un 10% en la próxima década en todos los escenarios. Sin embargo, después de 2030, las previsiones de gas se inclinarán hacia su declive impulsadas por la creciente presión de descarbonización en los edificios y la industria, aunque se espera que la demanda de gas sea más resistente que la de otros combustibles fósiles. Este declive provocaría que su cuota en la demanda de energía

primaria disminuya del 24% actual al 15% en 2050. Además, se producirá un desplazamiento progresivo de la oferta hacia China (McKinsey & Company, 2022).

La incertidumbre sobre el ritmo de la transición energética puede influir también en la volatilidad de los precios del gas y provocar ciclos de inversión aún más pronunciados que los observados en 2021 y 2022.

En el futuro, el gas podría desempeñar un nuevo papel en la producción de hidrógeno azul y amoníaco, y la infraestructura gasística podría reutilizarse para combustibles bajos en carbono, como el hidrógeno y el biogás, o el transporte de CO₂ para el CCUS (Captura Almacenamiento y Uso del Carbono).

2.1.4 La crisis energética: el Gas Natural

Desde hace más de un año, la mayoría de los hogares españoles y europeos han notado el encarecimiento de la energía, detonante de máximos históricos en cuanto a inflación (Jorriñ, 2022). En el contexto de la crisis energética europea, aunque todas las fuentes fósiles sufren estragos por el agotamiento de recursos y el camino hacia la descarbonización, es de mayor relevancia la que está relacionada con el gas por los matices políticos implicados. Cabe destacar dos factores clave como causantes de la situación. En primer lugar, después del COVID-19 se produjo una reactivación repentina de la economía a nivel global. Esto hizo que, de un día para otro, empresas e individuos volvieran a consumir energía, aunque a niveles inferiores a prepandemia. Este aumento de la demanda de gas sorprendió a los principales suministradores europeos y, en consecuencia, abrió un hueco entre la demanda y la oferta de gas. Este hueco se dilató debido al añadido aumento de la demanda a causa de un verano con temperaturas más cálidas de lo normal y a un invierno gélido. El segundo factor clave detonante de la crisis energética europea es la oferta de gas. Rusia, desde una posición de independencia energética, pues cuenta con fuentes de gas y suministros propios, decide restringir el abastecimiento con fines políticos y por el aumento de costes y la escasez de inversiones. El conflicto con Ucrania, país por donde discurren conductos que comunican Rusia con otros países europeos, incrementa la tensión energética debido a la relación de interdependencia entre ambos países. Para entender mejor la posición de Rusia, es conveniente analizar la historia en materia de energía (Alvarez, 2022).

En 1973 se produce un embargo petrolero por parte de los países árabes productores de crudo a Occidente, con el fin de presionar a Estados Unidos y a sus países aliados en la defensa de Israel. En este momento, Rusia que, por entonces, era la segunda productora de gas a nivel mundial y la tercera de petróleo, contaba con independencia, altas reservas energéticas y capacidad productora, aprovecha a hacerse con la cuota de mercado antes dominada por los países árabes, mediante la construcción de gaseoductos que la conectaban con Alemania. De esta manera, Rusia pasaba a ser el primer país suministrador y del que, desde ese momento en adelante, dependería buena parte de Occidente para cubrir sus necesidades energéticas. De hecho, al cabo de los años, Rusia no tardó en desarrollar su propia política y estrategia energéticas, introduciendo el concepto de seguridad energética tanto para importadores como para exportadores y por el cual se daba prioridad a acuerdos bilaterales y se imponían cláusulas restrictivas en los contratos.

La posición de la Unión Europea en este panorama energético internacional no es otra que la de dependencia. Actualmente, Rusia suministra el 40% del gas a la unión europea, el 30% del petróleo y el 25% del carbón, aunque hay países dentro de la UE cuya dependencia es todavía mayor. El 40% del gas suministrado se realiza a través de 3 principales vías, entre las que se encuentra la ruta de Ucrania y el Nord Stream, utilizado por Rusia en diferentes momentos de la historia para su beneficio político (Leon, 2021). Desde esta posición, países como Alemania, Francia, España o Italia han puesto en marcha planes de choque sobre todo para controlar la inflación como efecto secundario de esta situación. Entre estas medidas se encuentran la reducción de impuestos en materia de energía, la concesión de subvenciones relacionadas con la energía o la modificación de los procedimientos para establecer el precio de la electricidad. A pesar de los esfuerzos de Europa por controlar esta situación, la opinión popular duda de la efectividad de estas medidas (Torres, 2022).

Cabe destacar la relación entre Rusia y Ucrania en torno al gas, pues la primera suministra gas a la segunda, pero a su vez utiliza su territorio para transportar el gas a los demás países occidentales a los que abastece, por lo que Ucrania cobra una tasa a Rusia que alcanza un 3% de su PIB (Leon, 2021). En este sentido, siempre han existido tensiones entre ambos países por no ponerse de acuerdo en el precio y otros términos contractuales, que han llevado a repetidos cortes en el suministro de gas.

En cuanto al Nord Stream y Nord Stream 2, son la principal vía de suministro energético a Alemania. El primero, se terminó de construir en 2012, momento en el que Alemania concedió los permisos para la puesta en marcha de la segunda versión, con el doble de capacidad que la primera. Las razones por las que Alemania aprobó el Nord Stream 2 son, en primer lugar, porque esperaba una creciente demanda de gas y futuro riesgo de desabastecimiento y, en segundo lugar, porque de esta manera Alemania se convertiría en el segundo distribuidor de gas en Europa, actividad de la que obtendría altos beneficios.

Sin embargo, organizaciones como la Agencia Internacional de la Energía (AIE) y países como Estados Unidos y Reino Unido, no dejan de alertar de lo delicada que es la situación de dependencia y de los riesgos que conlleva para Occidente, además de acusar a Rusia de la falta de voluntad para acabar con la crisis energética.

Por otro lado, Rusia argumenta que, como consecuencia de la construcción del nuevo gasoducto, Europa reduciría la dependencia energética de Ucrania aumentando así su seguridad energética y la accesibilidad al gas para más países miembros. Detrás de estos argumentos se esconde un claro beneficio para Rusia, pues una de las situaciones que genera conflictos en la actualidad es su lucha para reducir su dependencia de Ucrania, país al que paga una tasa estratosférica por el uso de su territorio para el suministro de gas. Además, aumentaría el número de países abastecidos, así como su dependencia, reafirmando su posición como principal fuente de energía.

Este contexto pone de manifiesto la discordia dentro de la Unión Europea, donde países como Alemania están divididos en torno a la idea del Nord Stream 2 y donde no se han conseguido mejoras unificadas en el ámbito energético.

2.2 Las energías renovables

Mientras los gobiernos y las empresas se comprometen cada vez más con objetivos de descarbonización estrictos, los mercados energéticos se enfrentan a una volatilidad extrema impulsada por las tensiones geopolíticas y el repunte de la demanda de energía, que deriva en una crisis energética global. En este contexto surgen las energías renovables como alternativa e impulsoras de los objetivos mundiales en torno a energía.

Sin embargo, aunque las industrias estén más concienciadas con lograr las *net-zero emissions*, se espera que la demanda de energía no renovables alcance su máximo antes de 2030 como se ha expuesto anteriormente.

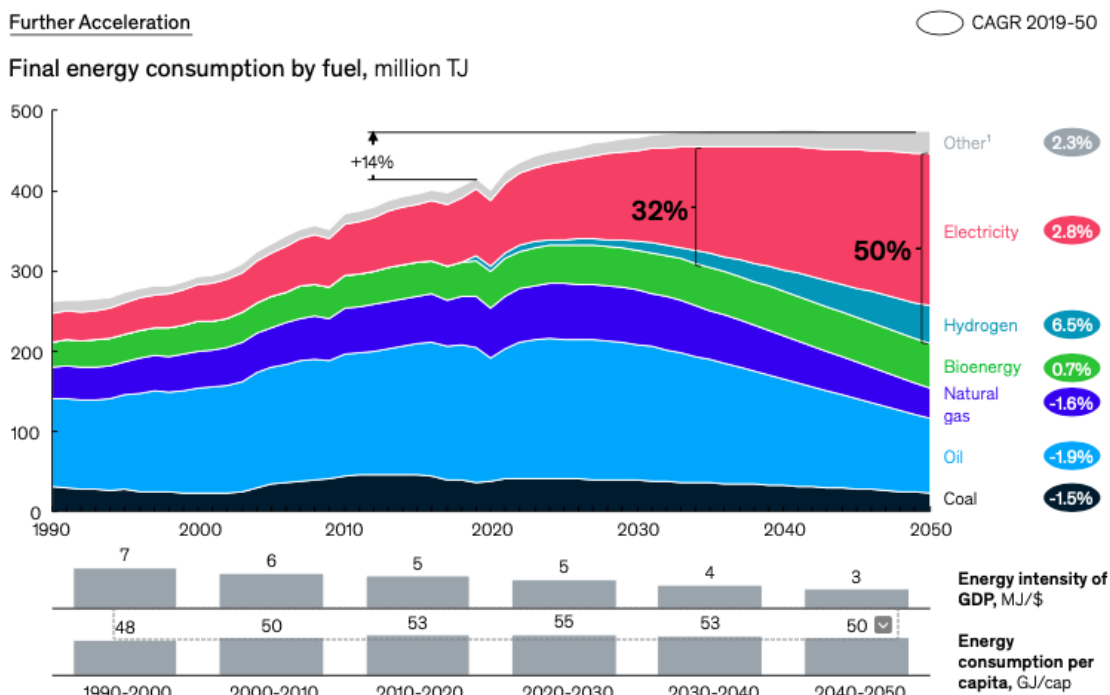
Se entiende como energías renovables aquellas que se obtienen de recursos naturales, se regeneran de manera constante y no se agotan con su uso. Estos dos últimos atributos son los que más las diferencian de los combustibles fósiles o energías no renovables. Su origen se remonta a la antigüedad, cuando se utilizaba la energía solar para calentar el agua en los baños romanos y la energía eólica para mover molinos de viento. Sin embargo, el desarrollo moderno de las energías renovables comenzó a partir del siglo XIX con la invención de la célula fotovoltaica y con el primer molino de viento eléctrico (García, 2017). Este desarrollo ha venido impulsado tanto por disminuir la dependencia de las fuentes de energía finitas como por la necesidad de producir energía de manera limpia para el planeta.

A pesar del rápido crecimiento de la economía mundial y del aumento de la población en 2.000 millones de personas, se prevé que el consumo de energía crezca sólo un 14% y después se estabilice a lo largo del tiempo (McKinsey & Company, 2022). La continua reducción de la intensidad energética del PIB es un factor clave, impulsado por una mayor eficiencia en el uso final de los edificios, el transporte y la industria. La electrificación desempeña un papel importante, ya que el cambio a soluciones eléctricas suele ir acompañado de un aumento de la eficiencia en muchos segmentos, como la calefacción y los turismos.

Además, el papel de la electricidad en el consumo final pasará del 20% actual al 40% en 2050 (Sanborn & Chatterjee, 2023). En línea, es de esperar que la correspondiente duplicación del consumo de electricidad, combinada con la adopción del hidrógeno, compense el consumo de combustibles fósiles (que excluye la demanda primaria de carbón y gas para la generación de electricidad), que podría ser un 40 % inferior en 2050 en comparación con 2020. En la siguiente gráfica se puede observar la evolución de la distribución del consumo energético.

Gráfico 4

Evolución del consumo de energía final por combustible de 1990 a 2050.



Fuente: (McKinsey & Company, 2022)

Según McKinsey & Company (2022), las energías renovables se convertirán en la nueva carga básica, representando el 50% del mix energético en 2030 y el 85% en 2050. La energía solar y la eólica ya tienen un coste inferior al de los combustibles fósiles en la mayoría de los países y se prevé que sean cada vez más competitivas a nivel mundial. Aunque las energías renovables se expandirán, no lo harán de manera homogénea, sino que existirán diferencias regionales en las vías de descarbonización del mix energético, impulsadas por la política, los factores económicos y la disponibilidad de tierras y recursos.

Gracias al gráfico 4 podemos concluir que a medida que pase el tiempo, cada vez se hará más uso de las energías renovables, apostando menos por energías tradicionales que son finitas y suponen un coste mayor para la sociedad. En consecuencia, la relación entre ambas energías es indirecta, ya que a medida que el uso de las energías renovables esté más extendido, se desarrollarán mejores tecnologías alrededor y su coste se abaratará. Sin embargo, y debido al menor uso de los combustibles fósiles, su coste aumentará al existir una menor oferta por la disminución de la demanda. Esto quiere decir que las primeras cada vez representan mayor amenaza para las segundas y en situaciones de crisis y

volatilidad de las energías no renovables, el cambio a las renovables como alternativa será más notable.

2.2.1 RENIXX WORLD INDEX

El RENIXX World Index es un índice bursátil compuesto por las 30 compañías más relevantes, en términos de capitalización bursátil, del sector de las energías renovables cuyo porcentaje del total de su facturación proveniente de este sector sea mayor al 50%. Incluye acciones de sectores como la energía eólica, solar o hidroeléctrica y abarca un amplio espectro de actividades, productos y servicios dentro de cada industria. Por ejemplo, en el sector de la energía eólica, la gama abarca desde fabricantes, promotores de proyectos de energía eólica y proveedores hasta empresas de servicios públicos que han invertido en proyectos de energía eólica en tierra y en el mar y operan plantas. En cuanto a las empresas solares, RENIXX World incluye empresas como fabricantes de células y módulos fotovoltaicos, fabricantes de equipos fotovoltaicos, proveedores, promotores de proyectos o empresas de energías renovables. Debido a que el sector de la bioenergía es más lento que el de la energía solar, en el RENIXX World están representadas relativamente pocas empresas.

Desde el 2006, fecha en la que fue constituido, el RENIXX World es el índice de referencia de la industria. Comenzó contando con las 20 empresas más grandes del sector y en 2007 el número se amplió a 30. Por aquel entonces, la capitalización bursátil de las empresas que lo conformaban ascendía a 65.000 millones de euros, anticipando la importancia que el sector cobraría en los años venideros. A fecha 1 de enero de 2023, la capitalización bursátil del índice es 611.013 millones de euros, por lo que se ha producido un aumento del 840% (Fondo Económico Internacional para las energías renovables, 2023). Esto se debe a que el índice se creó cuando este sector era muy joven y las fuentes de las que se nutren las empresas de energía renovable apenas tenían cabida en el mix energético internacional. Conforme estas se han ido considerando cada vez más como una alternativa robusta frente a los combustibles fósiles, su capitalización bursátil ha aumentado.

2.3 La predicción de series financieras

Una vez analizada la situación actual del panorama energético europeo, necesaria para entender la relación inversa entre los dos principales tipos de energías, es imprescindible la revisión de literatura acerca del uso de modelos predictivos aplicados a series bursátiles para poder aplicar uno de los modelos como complemento a la predicción del contexto energético.

Numerosos son los estudios dedicados a la predicción de series temporales, específicamente aquellos que se centran en la predicción del precio de instrumentos de financieros o de índices bursátiles. Este campo despierta tal interés debido a que, mediante la ayuda de métodos econométricos como modelos ARIMA o de procesos de *deep learning*, de ser posible la aproximación del precio futuro de estos activos al que realmente será, los inversores pueden obtener rentabilidades casi aseguradas en sus inversiones. Una vez más, el dinero ha movilizad a expertos de todo el mundo hacia la búsqueda de un modelo que aproxime a la perfección dicha variable. Sin embargo, es comúnmente conocido que la predicción del precio de cualquier activo en el mercado bursátil es un problema difícil de resolver (Berat Sezer, Ugur Gudelek, & Murat Ozbayoglu, 2020), pues a parte de verse influido por la oferta y la demanda como cualquier precio, aquí intervienen también las perspectivas de futuro del activo, que a su vez recogen cualquier información a tiempo real susceptible de afectarlo.

Ya que en el presente trabajo se pretende predecir un índice bursátil, la mayor parte de la literatura revisada gira alrededor de este tema. Aunque en estos estudios el fin último sea la rentabilidad económica, el objetivo es diversificar su uso para también complementar la predicción de los precios de las energías tradicionales, que son más comunes en la sociedad actual, y ayudar a los colectivos que más puedan ser afectados por una crisis energética mediante la promulgación de medidas preventivas por parte del gobierno como pueden ser ayudas o subvenciones previas a épocas de crisis.

En 2013 se publicó un estudio en el que se realizó una comparativa de la mayor parte de los modelos para la predicción de índice bursátiles, en la que concluyeron que los métodos lineales presentan unas ventajas relacionadas con la interpretabilidad a tener en cuenta en este caso, ya que conocer el funcionamiento interno y la razón de nuestras predicciones resulta muy práctico a la hora de darle un uso más allá que la rentabilidad financiera

(Garcia, Jalal, Garzon, & Lopez, 2013). De manera opuesta, los métodos no lineales mejoran las predicciones, pero se pierde la claridad del flujo, así como la interpretabilidad y aumenta la complejidad del modelo.

Wang, Wang, Zhang & Guo (2012) plantearon también modelos híbridos en los que se combinan modelos lineales y no lineales para la predicción. El uso de ambos de manera conjunta es una alternativa que mejora los inconvenientes y la capacidad predictiva de cada uno por separado, pero incrementa mucho el riesgo computacional.

Otros estudios que investigan el mismo campo como el realizado por Zhao & Guo (2014), se intenta pronosticar el consumo anual de energía mediante un modelo PSOCA (Particle Swarm Optimization Cultural Algorithm) y GRNN (Generalized Regression Neural Network). Este último será un tipo de red neuronal que se implementará en el presente trabajo. En el estudio, se comprueba cómo el modelo GRNN es un buen estimador de series temporales debido a su gran habilidad para solventar problemas no lineales. El modelo PSOCA es empleado para el cálculo de la sigma, elemento indispensable en el modelo GRNN, por lo tanto, la predicción se logra mediante un modelo integrado, que arroja mejores resultados que cada uno por separado. Además, aunque no en el mismo campo, un estudio combina los modelos ARIMA y GRNN para el pronóstico de la incidencia de Hepatitis en China (Wei, Jiang, Liang, & Gao, 2016) y en él se observa cómo ambos tienen buenas cualidades predictoras juntos y por separado. De igual manera, Li, Liu, Le & Wang (2005) proponen también un estudio comparativo entre un modelo ARMA (Autoregressive Moving Average) y un GRNN para la predicción de series temporales, el que se prueba el algoritmo GRNN como una alternativa más actualizada a los modelos ARMA y también se realiza una predicción conjunta para observar los beneficios que ambos modelos aportan.

En algunos estudios, como el realizado por Bermejo, Torres & Masferrer (2009), se han probado modelos ARIMA en mercados continuos obteniendo resultados con una precisión elevada, sobre todo para el corto plazo y con un ratio de coste/eficiencia muy bajo. Este modelo suele ser utilizado como base de comparación para otros modelos más complejos, que intentan corregir alguna de sus carencias. En otro estudio, se comparó la precisión en las estimaciones de un modelo auto arima y otro ARIMA customizado (Khan & Alghulaiakh, 2020), observándose que en ambos casos tanto el MAPE (Mean Absolute

Percentage Error) cómo otras métricas comparativas eran favorables. Por último, se ha aplicado también un SARIMA (Seasonal Autoregressive Moving Average) en comparación con otros modelos a series temporales (Kumar Dubey, Kumar, García-Díaz, Kumar Sharma, & Kanhaiya, 2021), en particular al consumo de energía diario, y esta ha resultado ser desfavorable para el SARIMA en favor de modelos más complejos como redes neuronales como LSTM (Long Short-Term Memory).

Puesto que la predicción a realizar involucra un episodio de crisis energética, es interesante analizar cómo el modelo ARIMA se comporta en estos ecosistemas de alta volatilidad. En 2021, Pellicer llevó a cabo un estudio en el que se aplicaba el mencionado modelo a la crisis sanitaria de la COVID-19 en el que demuestra cómo es útil para la estimación del precio de acciones dentro de un intervalo al 95% de confianza. Debido a la alta volatilidad de los últimos periodos en las series temporales relacionadas con la energía, se ha decidido utilizar un modelo que en principio sea capaz de captar la misma, como es el GARCH. Francq & Zakonian (2019) ofrecen en su libro un enfoque exhaustivo y sistemático para comprender los modelos GARCH de series temporales y sus aplicaciones, al tiempo que presentan los resultados más avanzados sobre la teoría y los aspectos prácticos de GARCH. Se estudia en detalle la estructura probabilística de los modelos GARCH estándar, así como la inferencia estadística y se examina las aplicaciones financieras. Por otro lado, se estudia cómo la inclusión de modelos GARCH dentro de otros modelos mejoran las predicciones al ajustar mejor al volatilidad (Chandramouli Rayadurgam & Mangalagiri, 2023) y se analizan aplicaciones de modelos GARCH a las series financieras concluyendo que el GARCH(1,2) es el que mejor se ajusta a las mismas (Casas & Cepeda, 2008). Además, teniendo en cuenta que las energías renovables están en pleno apogeo, su desempeño en los mercados bursátiles puede asemejarse al del mercado de las criptomonedas, eliminando los periodos más recientes, donde el precio de los activos se dispara y después se caracteriza por su elevada volatilidad. Un estudio concluye que uno de los modelos que mejor capta esta volatilidad es el TGARCH y que los efectos positivos en el precio de los activos provocan mayor volatilidad que los negativos (Fakhfekh & Jeribi, 2020).

Sin embargo, no existe literatura en la que se aplique cualquiera de estos modelos a un índice bursátil relacionado con el sector energético, ni existe tampoco una comparación de tres tipos de modelos diferentes entre sí, y menos, artículos en los que se intente relacionar el resultado con la predicción del desempeño de las energías tradicionales y

por ende, de posibles crisis energéticas venideras. Este vacío en la investigación será en el que intentaré cubrir con mi posterior análisis.

3 Estudio empírico

3.1 Datos: Las series temporales

Una serie temporal es una secuencia de observaciones o mediciones realizadas en diferentes momentos a lo largo del tiempo. En otras palabras, se trata de una secuencia de datos ordenados cronológicamente. Estas observaciones se recopilan a intervalos regulares o irregulares, lo que permite analizar cómo una variable o fenómeno cambia y se desarrolla en el tiempo (Dama & Sinoquet, 2021).

Las series temporales se encuentran en una amplia variedad de disciplinas, desde la economía y las finanzas hasta la meteorología, la epidemiología, la ingeniería y más. Algunos ejemplos comunes de series temporales incluyen datos económicos como el crecimiento del PIB, el precio de las acciones en la bolsa de valores, la temperatura diaria registrada en una ubicación específica, el número de casos de una enfermedad a lo largo del tiempo, las ventas mensuales de un producto, entre otros.

El análisis de series temporales busca comprender y modelar los patrones y las características inherentes a los datos a lo largo del tiempo. Al estudiar una serie temporal, se pueden identificar tendencias, patrones estacionales, ciclos, efectos de regresión, cambios estructurales y propiedades de autocorrelación, entre otros aspectos.

El objetivo principal del análisis de series temporales es predecir o pronosticar los valores futuros de la variable en estudio, basándose en la información histórica disponible (Dama & Sinoquet, 2021). Para lograrlo, se utilizan técnicas y modelos estadísticos específicos, como los modelos ARIMA (Autorregresivos Integrados de Medias Móviles), modelos de suavizado exponencial, modelos de regresión con componentes de series temporales, entre otros.

Además de la predicción, el análisis de series temporales también se utiliza para entender y describir el comportamiento de una variable a lo largo del tiempo, identificar cambios o anomalías, evaluar el impacto de eventos o políticas, realizar comparaciones entre

diferentes períodos y apoyar la toma de decisiones informadas en diversas áreas. Es decir, que además de realizar predicciones se pueden crear modelos explicativos.

Uno de los tipos de series temporales más interesantes de analizar, pues encontrar un modelo que estimara casi a la perfección supondría ganancias millonarias, son las series financieras. Estas son conjuntos de datos secuenciales que registran la evolución de variables financieras a lo largo del tiempo. Son fundamentales en el análisis y la predicción de los mercados financieros, ya que proporcionan información valiosa sobre la dinámica de los precios, volatilidad y tendencias de los activos financieros (Berat Sezer, Ugur Gudelek, & Murat Ozbayoglu, 2020).

Una de las características principales de las series financieras es su temporalidad. Los datos se registran en intervalos regulares, como días, semanas, meses o años, lo que permite analizar la evolución de la variable a lo largo del tiempo. Esta temporalidad facilita la identificación de patrones y tendencias significativas, lo que a su vez ayuda a los analistas a tomar decisiones informadas sobre inversiones y estrategias comerciales.

Otra característica importante de las series financieras es la presencia de estacionalidad. Algunas variables financieras exhiben patrones estacionales, lo que implica que ciertos eventos o condiciones específicas se repiten en momentos particulares del año (Berat Sezer, Ugur Gudelek, & Murat Ozbayoglu, 2020). Identificar estos patrones estacionales es crucial para comprender la dinámica del mercado y aprovechar las oportunidades comerciales.

Además, las series financieras suelen presentar tendencias, es decir, un movimiento general de incremento o decremento a lo largo del tiempo. Estas tendencias pueden ser ascendentes (indicando un aumento sostenido en el valor de la variable), descendentes (indicando una disminución sostenida) o laterales (sin una dirección clara). El análisis de tendencias permite a los analistas pronosticar el comportamiento futuro de la variable y tomar decisiones en consecuencia.

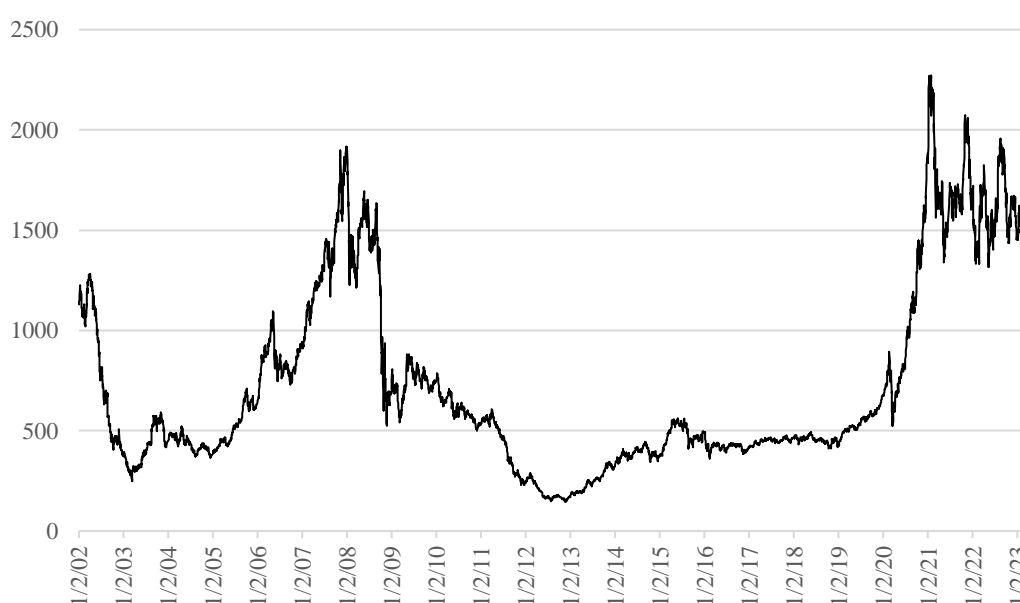
La volatilidad es otra característica importante de las series financieras. Representa la magnitud de los cambios en los precios o valores de los activos financieros a lo largo del

tiempo. Una alta volatilidad indica una mayor fluctuación y riesgo en el mercado, mientras que una baja volatilidad sugiere estabilidad.

Por último, las series financieras suelen mostrar autocorrelación, lo que significa que las observaciones están correlacionadas entre sí en el tiempo. Es decir, el valor de una observación está relacionado con los valores anteriores o posteriores en la serie (Jenkins, Box, Reinsel, & Ljung, 2015). El análisis de la autocorrelación permite identificar patrones repetitivos en los datos y utilizar modelos estadísticos para pronosticar valores futuros. De hecho, estos análisis se llevarán a cabo en posteriores capítulos del presente trabajo.

En este caso, la serie financiera a analizar consistirá en la cotización de índice RENIXX, dotado de todas las características citadas anteriormente. La captura de los datos con los que se trabajará en el proyecto ha sido realizada a través de Bloomberg. Mediante el uso de su terminal se ha obtenido el nivel de cierre del índice desde su apertura, el 2 de enero de 2002, hasta la fecha en la que se produjo la extracción, el 5 de mayo de 2023. No obstante, aunque en el siguiente gráfico se reproduzca la serie completa, para la estimación se ha utilizado únicamente los 5 últimos años, desde el 1 de enero de 2018 hasta la fecha de extracción.

Gráfico 5
Evolución diaria desde enero de 2002 hasta mayo de 2023 del índice
RENIXX WORLD.



Fuente: (Bloomberg, 2023) Elaboración propia

3.2 Modelo ARIMA

En 1970, Box y Jenkins presentaron el modelo ARIMA. Se utiliza para reconocer, estimar y diagnosticar modelos ARIMA con datos de series temporales. Este modelo es igualmente conocido como metodología Box-Jenkins. Debido a que necesita un esfuerzo computacional muy bajo en comparación con los resultados que produce, este modelo es uno de los más populares en la previsión financiera. Además, han demostrado ser precisos en sus predicciones a corto plazo y habitualmente superan a los modelos estructurales complejos en predicciones con estas características (Adebiyi & Ayo, 2014).

En el modelo ARIMA, el valor futuro de una variable es una combinación lineal de valores y errores pasados, expresada de la siguiente manera:

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$$

donde, Y_t es el valor real y ε_t es el error aleatorio en t , ϕ_i y θ_j son los coeficientes, p y q son componentes que se suelen denominar autorregresivo y media móvil respectivamente (Adebiyi & Ayo, 2014). Es necesaria la diferencia (d) para que los predictores sean independientes y así la serie pueda llegar a ser estacionaria. Esta variable toma valor 0 cuando la serie ya es estacionaria (Kumar Dubey, Kumar, García-Díaz, Kumar Sharma, & Kanhaiya, 2021).

En este caso, utilizaremos lo más conocido como SARIMA, un proceso similar, pero con la diferencia de que contiene una parte regular, como se explicó anteriormente, asociada a observaciones consecutivas, y una parte estacional que representa la dependencia asociada a observaciones separadas por s periodos (Barrigüete, 2021). Por lo tanto, la forma de llamar a este proceso sería ARIMA (p,d,q) x (P,D,Q)[s]. Para detectar la estacionalidad presente en los datos, este proyecto utilizará el modelo SARIMA.

En cualquier caso, en ambos modelos es necesaria la explicación de los componentes AR y MA.

En primer lugar, el análisis y pronóstico de series de tiempo dependen del componente AR. La sección autorregresiva captura la relación lineal entre los valores anteriores y actuales de una serie de tiempo. En otras palabras, significa que el valor actual de una serie de tiempo depende de una combinación ponderada de sus valores pasados, donde

los coeficientes de ponderación determinan la influencia relativa de cada valor pasado en el valor actual (de Arce & Mahia, 2001). Esto permite la captura de patrones y tendencias históricas en los datos y su uso en la predicción del comportamiento futuro de la serie.

La fórmula general del componente AR en un modelo ARIMA se expresa de la siguiente manera:

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p}$$

En esta fórmula Y_t representa el valor de la serie de tiempo en el momento t . $\phi_0, \phi_1, \phi_2, \phi_p$ son los coeficientes autorregresivos. Cada coeficiente ϕ_i indica la influencia o peso que el valor pasado Y_{t-1} tiene sobre el valor actual Y_t . Estos coeficientes se estiman a partir de los datos y reflejan la estructura temporal de la serie (Khan & Alghulaiakh, 2020). De manera que el orden de este componente (p) indica cuántas observaciones pasadas influyen y un proceso AR(1) quedaría expresado de la siguiente forma:

$$Y_t = \phi_0 + \phi_1 Y_{t-1}$$

El componente AR en un modelo ARIMA es crucial. Puede analizar y modelar las dependencias temporales en los datos con una herramienta que permite una comprensión más profunda de los patrones y comportamientos de la serie de tiempo.

La estimación de máxima verosimilitud o el método de mínimos cuadrados se utilizan para estimar los coeficientes autorregresivos. Estos métodos permiten ajustar el modelo ARIMA a los datos observados y determinar los valores ideales de los coeficientes autorregresivos que mejor se ajusten a los patrones de la serie.

En un modelo ARIMA, el componente MA es el promedio móvil de los errores o residuos anteriores en una serie de tiempo. El componente MA se concentra en el componente aleatorio o no sistemático de la serie, a diferencia del componente AR, que captura las relaciones lineales entre los valores pasados y actuales de la serie. Este proceso también se conoce como modelo de medias móviles, y su abreviación en inglés es *moving averages* (de Arce & Mahia, 2001). La fórmula general del componente MA en un modelo ARIMA es:

$$Y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$$

En esta fórmula: ε_t es el término de error en el momento t , que representa el componente aleatorio o no sistemático de la serie. Estos errores se consideran ruido blanco con media cero y varianza constante. θ_1 , θ_2 y θ_q son los coeficientes de promedio móvil. Cada coeficiente θ_i indica la influencia o peso que el error pasado ε_{t-1} tiene sobre el valor actual Y_t .

El componente MA permite capturar los efectos a corto plazo o las perturbaciones aleatorias que afectan la serie de tiempo. Estos efectos pueden ser el resultado de eventos impredecibles o factores temporales no modelados (Kumar Dubey, Kumar, García-Díaz, Kumar Sharma, & Kanhaiya, 2021). Al incorporar los términos de promedio móvil en el modelo ARIMA, se busca eliminar o reducir la autocorrelación en los errores de serie y capturar cualquier patrón sistemático en estos errores.

Una manera de determinar el orden del componente AR y MA es mediante el uso de la función de autocorrelación simple y la función de autocorrelación parcial.

La FAS es una herramienta estadística que permite evaluar la presencia de autocorrelación en una serie de tiempo, especialmente útil en el análisis de modelos ARIMA, mediante la cuantificación de la relación entre los valores pasados y actuales de una serie de tiempo. Proporciona información sobre la dependencia lineal entre observaciones separadas por varios desfases o retrasos (Kumar Dubey, Kumar, García-Díaz, Kumar Sharma, & Kanhaiya, 2021). El retraso es la cantidad de tiempo que transcurre entre dos observaciones. Por ejemplo, un retraso de uno indica la relación entre un valor y su valor anterior, mientras que un retraso de dos indica la relación entre un valor y su valor dos períodos anteriores.

La FAS es una herramienta estadística que permite evaluar la presencia de autocorrelación en una serie de tiempo, especialmente útil en el análisis de modelos ARIMA, mediante la cuantificación de la relación entre los valores pasados y actuales de una serie de tiempo. Proporciona información sobre la dependencia lineal entre observaciones separadas por varios desfases o retrasos. El término "retraso" se refiere a la cantidad de tiempo que transcurre entre dos observaciones (Kumar Dubey, Kumar, García-Díaz, Kumar Sharma, & Kanhaiya, 2021). Por ejemplo, un retraso de uno indica

la relación entre un valor y su valor anterior, mientras que un retraso de dos indica la relación entre un valor y su valor dos períodos anteriores.

De esta manera, si los residuos están auto correlacionados, esto indica que aún hay información no explotada o patrones sistemáticos en los datos que no han sido capturados por el modelo.

El uso de un gráfico, a menudo acompañado de un segundo en el que se representa la FAP, es muy útil para el análisis de la FAS. La autocorrelación perfecta entre los valores actuales y ellos mismos, o varianza, se indica con un valor de 1 en el retraso 0 (Barrigüete, 2021). En el gráfico, se usan líneas o barras para representar los valores de autocorrelación; las líneas que se extienden más allá de los límites de confianza indican que hay una autocorrelación significativa.

La FAS se calcula mediante el coeficiente de correlación de Pearson entre las observaciones de la serie de tiempo a diferentes retrasos, expresada en la siguiente fórmula:

$$FAS(k) = \frac{Cov(X_t, X_{t-k})}{Var(X_t)}$$

Donde $FAS(k)$ es el valor de la FAS en el retraso k , $Cov(X_t, X_{t-k})$ es la covarianza entre la observación en el tiempo t y la observación en el tiempo $t-k$ y $Var(X_t)$ es la varianza de la observación en el tiempo t (Kumar Dubey, Kumar, García-Díaz, Kumar Sharma, & Kanhaiya, 2021).

Además, el análisis de la FAS puede sugerir que es necesario diferenciar la serie temporal (Khan & Alghulaiakh, 2020). La serie necesita ser diferenciada para lograr estacionariedad porque la FAS proporciona información sobre la dependencia lineal entre los valores pasados y presentes.

- Si la FAS muestra una autocorrelación significativa en el retraso 1 (o incluso en varios retrasos), esto indica que los valores de la serie están correlacionados y la serie no es estacionaria. En este caso, la diferenciación es necesaria para eliminar la tendencia y la dependencia temporal de los datos.
- Si la FAS muestra una autocorrelación insignificante en cada retraso, esto indica que la serie es estacionaria y no requiere diferenciación.

En cuanto a la FAP, mide la correlación parcial entre una observación en un tiempo dado y su valor rezagado en un tiempo anterior, controlando los efectos de las observaciones intermedias, ayudando a identificar una correlación directa entre dos puntos en el tiempo y excluyendo la influencia de las observaciones intermedias.

Los términos AR (autorregresivos) y MA (medias móviles) en el modelo se ordenan utilizando FAP. Si su FAP se encuentra fuera de los límites de confianza establecidos, un retraso es considerado significativo, al igual que en la FAS.

3.3 Modelo GARCH

En econometría y finanzas, un tipo de modelo conocido como modelo GARCH (Generalized Autoregressive Conditional Heteroskedasticity) se utiliza para modelar y predecir la volatilidad de una serie temporal, particularmente cuando se trata de los precios de los activos financieros. En 1986, Bollerslev lo propuso como una extensión del modelo ARCH (Autoregressive Conditional Heteroskedasticity) (Francq & Zakoian, 2019).

El objetivo principal de un modelo GARCH es capturar y modelar la heteroscedasticidad condicional, es decir, la variación de la volatilidad a lo largo del tiempo en función de datos previos. La idea fundamental detrás de un modelo GARCH es que la variabilidad de un proceso estocástico está influenciada por su propia trayectoria y errores previos. Esto significa que la volatilidad actual depende de los errores previos y de la volatilidad anterior (Francq & Zakoian, 2019). Esta es la principal diferencia con un modelo ARIMA, que se centra en modelar la tendencia y autocorrelación de una serie temporal en lugar de su volatilidad. El paquete que se utilizará en la estimación contiene una herramienta que permite calcular los valores reales de la serie a partir de la volatilidad estimada por el modelo.

El componente autorregresivo (AR) y el componente de media móvil (MA) son los dos elementos principales del modelo GARCH. El componente AR captura la relación autorregresiva de la volatilidad, que es cómo la volatilidad en un momento dado se relaciona con la volatilidad en períodos anteriores. El componente MA captura la relación de dependencia entre la volatilidad actual y los errores pasados.

La especificación exacta de un modelo GARCH puede variar dependiendo de la forma funcional y los parámetros utilizados. Sin embargo, en general, un modelo GARCH se puede expresar de la siguiente manera:

$$\sigma_t^2 = \omega + \sum_{i=1}^p \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^q \beta_i \sigma_{t-i}^2$$

Donde σ_t^2 es la varianza condicional de la serie en el tiempo t, ω es el término de constante que representa la varianza constante, α_i son los coeficientes del componente autorregresivo (AR) que capturan la dependencia de los errores pasados al cuadrado, ϵ_{t-i}^2 son los errores pasados al cuadrado, β_i son los coeficientes del componente de media móvil (MA) que capturan la dependencia de la volatilidad pasada y σ_{t-j}^2 son las volatilidades pasadas (Francq & Zakoian, 2019).

El modelo GARCH se estima utilizando métodos de máxima verosimilitud, donde se busca encontrar los valores óptimos de los coeficientes α y β que maximicen la función de verosimilitud.

Este modelo se utiliza ampliamente en entornos econométricos y financieros para predecir la volatilidad futura de los activos financieros; también es esencial para la gestión de riesgos y la valoración de derivados financieros. Además, aplicar este modelo en intervalos temporales mejora la comprensión de las tendencias de volatilidad y los patrones de comportamiento de los precios de los activos (Francq & Zakoian, 2019).

Como restricciones o desventajas de este modelo destaca la asunción de una determinada distribución para los errores que puede que no sea la que mejor refleje su naturaleza real así como la no captura de cambios abruptos en la volatilidad.

3.4 Modelo GRNN

Un modelo GRNN (Red neuronal generalizada de regresión) es un tipo de red neuronal utilizado en problemas de regresión. El modelo GRNN tiene un enfoque analítico y puede realizar cálculos directos en lugar de usar un proceso de entrenamiento iterativo, elementos que lo distinguen de otras redes neuronales (Martinez, Frias, Conde, & Martinez, 2019).

Un modelo GRNN depende de los datos de entrenamiento, que son pares de entradas y salidas correspondientes. Durante la fase de entrenamiento, el modelo almacena estos

patrones de entrada y salida en una matriz de gramática, de la que depende la predicción. Para predecir, se utiliza una función de distancia, como la distancia euclidiana, para calcular la similitud entre el patrón de entrada y los patrones almacenados:

$$\| dist \|_j = \sqrt{\sum_{i=1}^R (x_i - IW_{ji})^2}$$

Los patrones almacenados en la matriz de gramian reciben pesos en función de su similitud con el patrón de entrada y estos pesos se utilizan para ponderar las salidas correspondientes a cada patrón almacenado (Zhao & Guo, 2014). Una vez que se han asignado los pesos a los patrones almacenados, se realiza una combinación lineal de las salidas ponderadas para obtener la predicción final del modelo. Esta combinación lineal se utiliza para estimar el valor objetivo o la salida deseada. La función gaussiana es la que generalmente se selecciona como función de transmisión de unas capas a otras dentro de este modelo:

$$G(x, x_i) = \exp\left(-\frac{\|x - x_i\|^2}{2\sigma^2}\right)$$

Donde x_i y σ son el centro y el parámetro de suavizado respectivamente.

Los modelos GRNN tienen uso en una amplia gama de campos, particularmente en la predicción de series temporales. Son útiles para predecir valores en una serie de tiempo, como las ventas, la demanda de productos o el comportamiento financiero (Wei, Jiang, Liang, & Gao, 2016). También se usan en el análisis de datos financieros, donde se pueden modelar y predecir variables como los precios de las acciones, las tasas de interés o el riesgo crediticio. Los modelos GRNN también se usan para estimar y aproximar funciones en problemas de regresión no lineal.

La capacidad de cálculo analítico directo de los modelos GRNN los hace más rápidos y eficientes en comparación con otros modelos que requieren procesos de entrenamiento iterativos. También tienen una mejor capacidad de generalización porque almacenan toda la información de entrenamiento en una matriz de gramian (Li, Liu, Le, & Waing, 2005). Además, en comparación con otras redes neuronales, son menos sensibles a la inicialización de pesos y parámetros.

Sin embargo, los modelos GRNN tienen ciertas restricciones. Debido a que almacenan toda la información de entrenamiento en la matriz de gramática, necesitan más memoria. Además, pueden funcionar menos bien en tareas más complejas, como la clasificación de imágenes o el procesamiento del lenguaje natural, donde otros modelos pueden ser más apropiados (Wei, Jiang, Liang, & Gao, 2016).

4 Aplicación de los modelos

4.1 Preparación de los datos

En el presente apartado se llevará a cabo la explicación del proceso del ajuste del modelo a los datos de entrenamiento en R Studio, así como la posterior predicción y el contraste con los datos de validación para comprobar la calidad de esta.

Tras el proceso de extracción de los datos mencionado en el apartado 3.1, se ha procedido a la importación de estos, una vez han sido ordenados en Excel, a la herramienta R Studio.

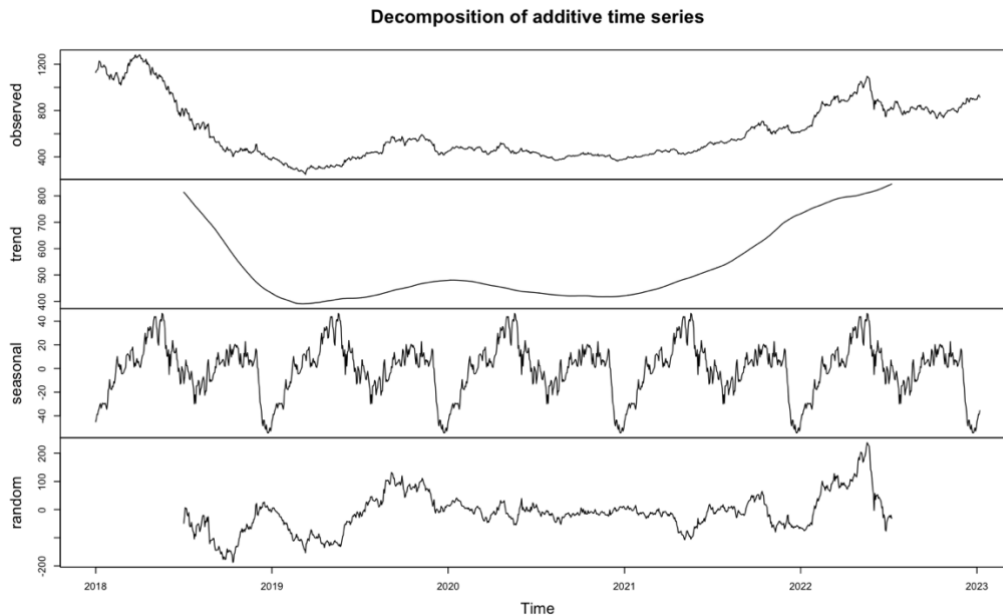
Antes de cualquier tratamiento, se han ejecutado una serie de funciones para obtener una idea inicial de la naturaleza de los datos.

```
class(data)
str(data)
dim(data)
colnames(data)
summary(data)
head(data)
tail(data)
```

De este pequeño análisis concluimos que el dataset contiene 5378 observaciones y dos variables, el precio de cierre y la fecha. Además, sabemos que la media histórica del índice son 725.5 así como otros datos de su distribución.

Después de dar formato a la fecha, convertimos el dataset a una serie temporal que comienza en enero de 2018 y termina en mayo de 2023, con una frecuencia diaria, en este caso el año laboral tiene aproximadamente 252 días. Se ha elegido esta frecuencia ya que la bolsa únicamente permanece abierta los días laborables. A continuación, mediante la función *decompose()* se puede observar la serie temporal bruta, su tendencia, estacionalidad y el componente aleatorio.


```
ST = ts(data$Close,start=c(2018,1), end=c(2023,5),frequency = 252)
componentes=decompose(ST)
plot(componentes)
```



Tal y como se aprecia en el gráfico elaborado por la función, existe tendencia en la serie, que parece ser cuadrática por la forma de curva convexa, así como estacionalidad anual al repetirse el mismo patrón cada año. Por lo tanto, de primera mano podemos concluir que la serie no es estacionaria en media y presenta estacionalidad. Estas conclusiones son importantes para la toma de decisiones posteriores y se respaldarán con otros tests.

En el siguiente paso, se ha procedido a dividir la muestra entre entrenamiento y validación, siendo esta última el último año, sobre el que se harán las predicciones. Esta escisión es muy útil a la hora de evaluar el poder predictivo del modelo ya que nos permite comparar las predicciones que este realiza con los valores reales de la serie y calcular métricas como el RMSE. Al mismo tiempo, se calculan los outliers en el set de validación con el fin de tratarlos de alguna forma, ya sea mediante su eliminación o el reemplazo de su valor por la media de los datos inmediatamente circundantes. En este caso, no se observa ningún dato atípico y por ello no se realiza ningún tratamiento más. Cabe mencionar que la óptima fórmula para el estudio de outliers es *tso()* del paquete *tsoutliers*, pero en este proyecto no se ha utilizado por el alto coste computacional que suponía.

```
data.train <- subset(data, "2018-01-02"<=Date & Date<"2023-04-05")
data.train.ts <- as.ts(data.train$Close, frequency = 252)

Q1 <- quantile(data.train.ts, 0.25)
Q3 <- quantile(data.train.ts, 0.75)
IQR <- Q3 - Q1
```

```

umbral <- 1.5 * IQR

outliers <- which(data.train.ts < Q1 - umbral | datos.train.ts > Q3 + umbral)

data.validate <- subset(data, "2023-04-05" <= Date & Date <= "2023-05-05")
data.validate.ts <- as.ts(data.validate$Close, frequency=252)

```

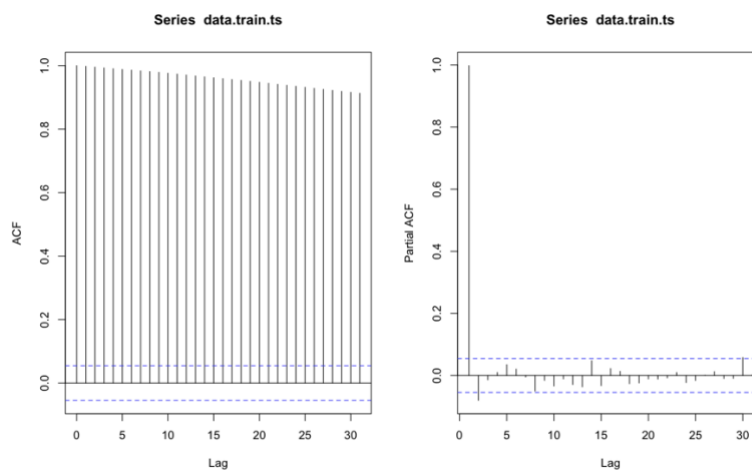
4.2 Ajuste del modelo ARIMA

Para comenzar a ajustar el modelo es necesario conocer el orden de los componentes AR y MA. Para ello, en un primer instante graficaremos las FAS y FAP de la serie sin ninguna transformación para llevar a cabo un análisis.

```

par(mfrow=c(1,2))
acf(data.train.ts)
pacf(data.train.ts)

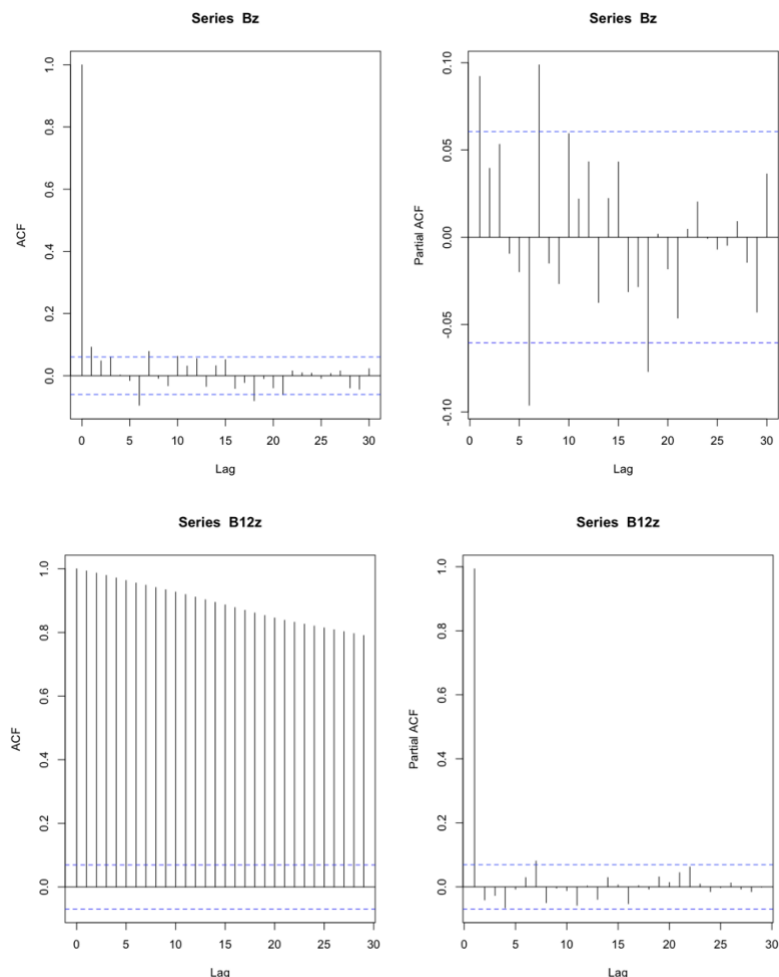
```



A partir de los anteriores gráficos podemos concluir la necesidad de diferenciación por el decrecimiento prolongado de las FAS, que es lo que se realizará a continuación mediante el test de Dickey-Fuller. Este se basa en la idea de que, si una serie de tiempo es no estacionaria, es probable que la serie de tiempo y su primera diferencia (la diferencia entre cada observación y la observación anterior) sean correlacionadas. El test calcula un estadístico de prueba que compara la correlación entre la serie de tiempo y su primera diferencia con un valor crítico. Si el estadístico de prueba es menor que el valor crítico, se puede rechazar la hipótesis nula de que la serie de tiempo es no estacionaria y se concluye que la serie de tiempo es estacionaria. Por otro lado, si el estadístico de prueba es mayor que el valor crítico, no se puede rechazar la hipótesis nula y se concluye que la serie de tiempo es no estacionaria. En este caso, el p-valor es 0.4758 por lo que rechazamos la hipótesis nula y concluimos que la serie no es estacionaria. En consecuencia, se diferenciará la serie y se obtendrá la serie Bz. Por otro lado, se diferencia

la serie en la parte estacional (resulta en la serie B12z) con el periodo observado en la gráfica anterior (anual) para comprobar si solo queda una FAP significativa, si es así, implicará que la serie efectivamente tiene estacionalidad anual (Barrigüete, 2021). Además de este análisis, se concibe la presencia de heterocedasticidad, para lo que se calcula el parámetro Lambda de la transformación Box-Cox, técnica utiliza para eliminar la heterocedasticidad mediante la estabilización de la varianza. Si este valor es cercano a 0, indicará que existe heterocedasticidad, por lo que se aplicará la función BoxCox() para obtener la serie “z”, con varianza constante.

Una vez obtenemos la serie diferenciada en la parte regular y sin heterocedasticidad, se ha procedido a graficar las FAS y FAP de la nueva serie.



Sin embargo, como se puede observar, no se puede concluir ningún orden del modelo ARIMA a partir de las gráficas. Únicamente se puede concluir que existe estacionalidad

de periodo anual. Por ello, la mejor solución es probar un auto arima sobre la serie temporal z (después de la transformación Box-Cox) y observar que patrones es capaz de captar para así, obtener una idea previa del modelo a ajustar.

```
autoarima = auto.arima(z,seasonal=TRUE,stepwise=FALSE, approx=FALSE)
```

Con los siguientes resultados:

```
Series: z
ARIMA(3,1,2) with drift

Coefficients:
      ar1      ar2      ar3      ma1      ma2  drift
      0.1417 -0.894  0.1361 -0.0514  0.9098      0
s.e.  0.0527  0.051  0.0321  0.0446  0.0510      0

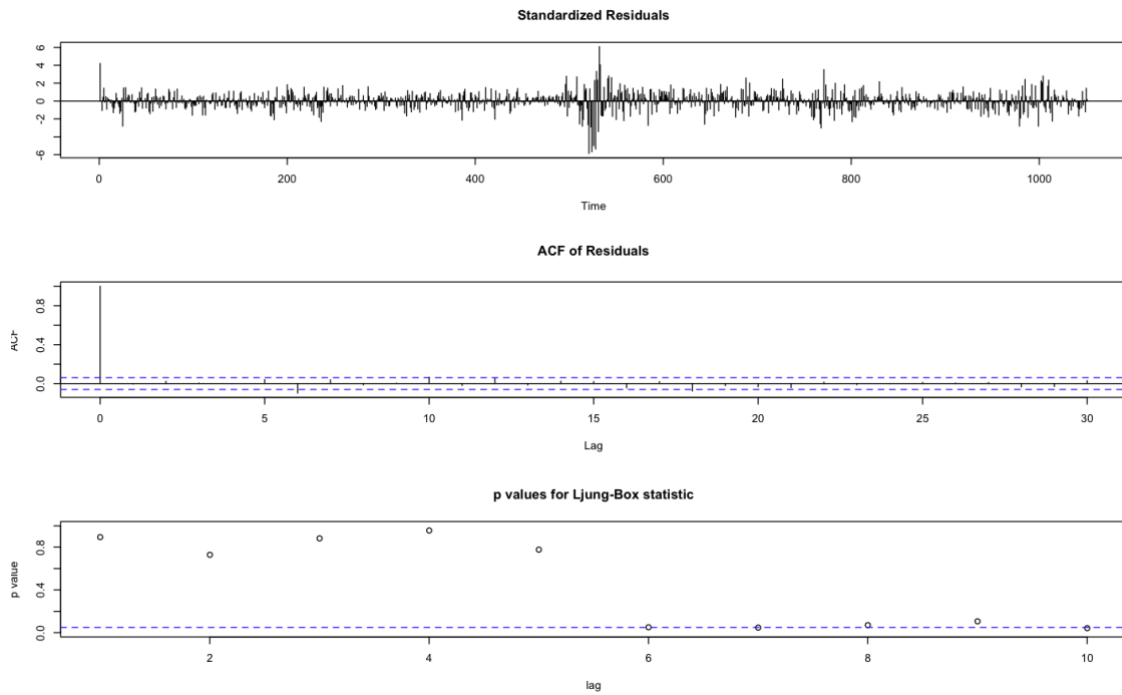
sigma^2 = 1.643e-07: log likelihood = 6723.38
AIC=-13432.77  AICc=-13432.66  BIC=-13398.07
```

Si comprobamos tanto el nivel de significación de los coeficientes como la calidad de los residuos, observamos que se trata de un buen modelo con capacidad predictiva. Los resultados muestran que la mayoría de los coeficientes llevan consigo un p-valor bajo por lo que son significativos y además, la distribución de los residuos es similar a la del ruido blanco, solo existe una FAS significativa en el retardo primero y el test de Ljung-Box presenta p-valores altos, aunque solo en los retardos más próximos, por lo que no se rechaza la hipótesis nula de que existe autocorrelación en los datos. Un modelo mejor debería de tener todos los p-valores por encima del nivel de significación.

```
library(lmtest)
coeftest(autoarima)
tsdiag(autoarima)

z test of coefficients:

      Estimate Std. Error z value Pr(>|z|)
ar1  1.4167e-01 5.2744e-02  2.6860 0.007231 **
ar2 -8.9397e-01 5.0987e-02 -17.5333 < 2.2e-16 ***
ar3  1.3609e-01 3.2128e-02  4.2359 2.276e-05 ***
ma1 -5.1446e-02 4.4624e-02 -1.1529 0.248958
ma2  9.0982e-01 5.1043e-02  17.8245 < 2.2e-16 ***
drift 2.4826e-05 3.3978e-05  0.7306 0.464998
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



Aunque los resultados obtenidos son aceptables, el modelo predicho por la función `auto.arima` no parece captar la estacionalidad de la serie, que según indica la gráfica es anual. Por ello, se ha decidido crear un bucle que itere con distintos valores para los coeficientes $(p,d,q)(P,D,Q)$ y encuentre el mejor modelo basándose en el criterio AIC (Akaike Information Criterion). Este criterio sirve para comparar modelos y mide la calidad de ajuste del modelo penalizando la complejidad del mismo. Por lo que un modelo simple con un buen ajuste tendrá un mejor AIC que un modelo complejo con sobreajuste. De esta manera, se elegirá el modelo que tenga un menor AIC, junto con otros criterios.

Para encontrar el mejor modelo mediante la ejecución del bucle, se parte de la información proporcionada por el método `auto.arima` y se parametriza teniéndolo en cuenta. Aunque se observó que la estacionalidad era de periodo anual, R no permite un periodo tan alto en la estimación del modelo por lo que se ha tenido que buscar otro periodo que tuviese sentido para la serie, siendo este el semanal (5).

```

per=5 #sabemos que la estacionalidad es anual por la gráfica decompose
ordenregular = 3 # Hasta que orden regular queremos probar
ordenestacional = 3 # Hasta que orden estacional queremos probar
difregular = 3 # Hasta que orden de diferenciación regular queremos probar
difestacional = 3 # Hasta que orden de diferenciación regular queremos probar

```

```

lista<-c()
lista2<-c()
for (p in 0:ordenregular){
  for (d in 0:difregular){
    for (q in 0:ordenregular){
      for (P in 0:ordenestacional){
        for (D in 0:difestacional){
          for (Q in 0:ordenestacional){
            prueba = try(Arima(data.train.ts, order=c(p,d,q),
                                seasonal = list(order=c(P,D,Q), period=per),
                                lambda = Lambda,
                                include.constant = TRUE),silent = TRUE)

            p1<-try(prueba$aic, silent = TRUE)
            lista<-append(lista,p1,after = length(lista))
            ORDEN<-paste(p,d,q,P,D,Q)
            lista2<-append(lista2,ORDEN,after = length(lista2))

          }}}}}

```

```

valores<-na.omit(data.frame(as.numeric(lista),lista2))
valores

```

```

dimension=dim(valores)
dimension[1]

```

```

min(valores$as.numeric.lista.)

```

```

#Partimos de un modelo naive
Modelo.naive = Arima(z, order=c(0,0,0),
                    seasonal = list(order=c(0,0,0), period=0),
                    lambda = Lambda,
                    include.constant = TRUE)

```

```

mejor.aic = Modelo.naive$aic
mejor.aic

```

```

for (i in 1:dimension[1])
{
  if(valores[i,1] < mejor.aic)
  {
    mejor.aic=valores[i,1]
    mejor.modelo=valores[i,2]
  }
}

```

```

}

```

```

mejor.modelo

```

```

arima_bucle = Arima(data.train.ts, order=c(3,2,3),
                    seasonal = list(order=c(2,1,2), period=5),
                    lambda = Lambda,
                    include.constant = FALSE)

```

```

arima_bucle
coefest(arima_bucle)
tsdiag(arima_bucle)

```

Este código nos devolverá el modelo que mejor se ajuste a nuestra serie de datos en el formato (p d q) (P D Q), así como los test para poder compararlo con el auto arima. De igual manera que en el modelo arima anterior, se ha analizado los residuos y se ha evaluado la significatividad de los coeficientes y, aunque los coeficientes de la parte estacional son muy significativos, no supone un problema grave por lo que ambos estudios han resultado ser positivos:

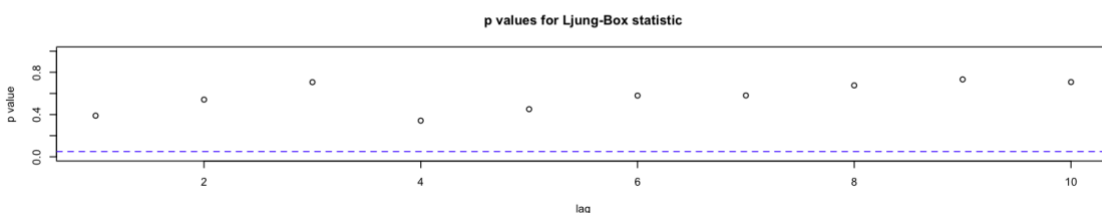
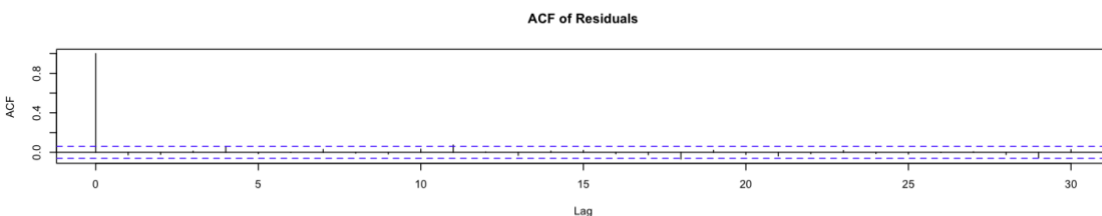
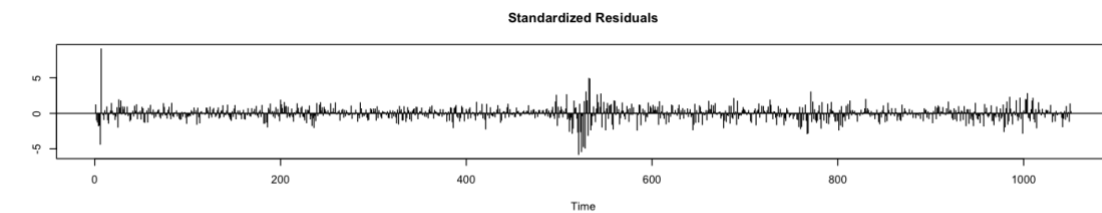
```
Series: data.train.ts
ARIMA(3,2,3)(2,1,2)[5]
Box Cox transformation: lambda= -0.5682675
```

```
Coefficients:
      ar1      ar2      ar3      ma1      ma2      ma3      sar1      sar2      sma1
-1.3992 -0.5585 0.1600 0.5020 -0.7401 -0.746 -0.8479 -0.0838 -0.2411
s.e.  0.0749  0.0941 0.0355 0.0702 0.0626 0.081 0.5600 0.0596 0.5606
      sma2
-0.7456
s.e.  0.5567
```

```
sigma^2 = 1.832e-07: log likelihood = 6616.02
AIC=-13210.03 AICc=-13209.78 BIC=-13155.57
```

z test of coefficients:

	Estimate	Std. Error	z value	Pr(> z)
ar1	-1.399183	0.074909	-18.6785	< 2.2e-16 ***
ar2	-0.558536	0.094107	-5.9351	2.937e-09 ***
ar3	0.160014	0.035480	4.5100	6.484e-06 ***
ma1	0.501955	0.070240	7.1463	8.915e-13 ***
ma2	-0.740070	0.062588	-11.8244	< 2.2e-16 ***
ma3	-0.745986	0.081019	-9.2076	< 2.2e-16 ***
sar1	-0.847927	0.559993	-1.5142	0.1300
sar2	-0.083817	0.059554	-1.4074	0.1593
sma1	-0.241100	0.560634	-0.4300	0.6672
sma2	-0.745625	0.556707	-1.3394	0.1805

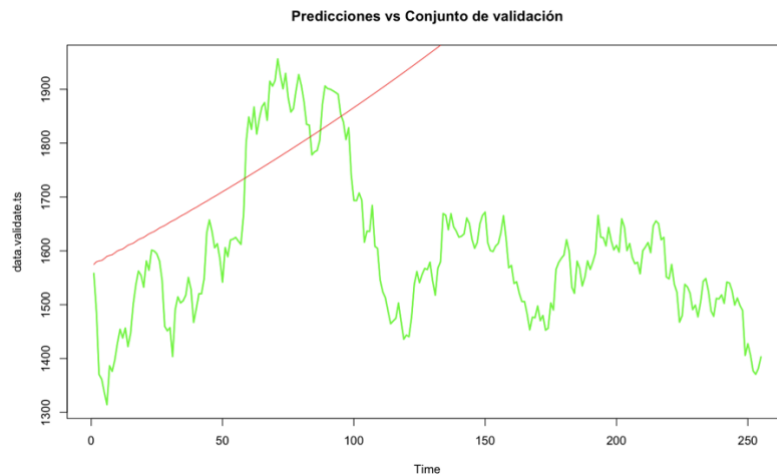


En cuanto al AIC, el AIC del modelo auto arima es -14432.77 y el del modelo del bucle es -13210.71. En este caso, no hace falta comparar ambos modelos bajo este criterio para elegir el segundo como el mejor. El hecho de que el modelo auto.arima no perciba la parte estacional ya nos hace rechazar el modelo, pues la presencia de dicho elemento es clara en la serie a analizar.

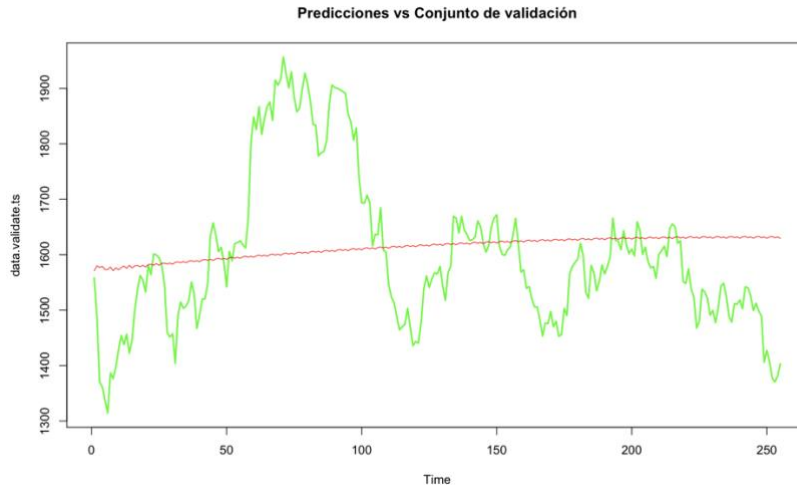
Aun así, se han realizado predicciones con ambos modelos para calcular el RMSE sobre las mismas. El RMSE (Root Mean Squared Error) es una medida comúnmente utilizada para evaluar la precisión de un modelo de pronóstico o regresión (Wang & Lu, 2018). Mide la diferencia promedio entre los valores observados y los valores pronosticados por el modelo.

Las predicciones con el primer modelo (auto arima) se han realizado de la siguiente manera, con el horizonte temporal igual al del set de validación:

```
prediccion_autoarima <- forecast(autoarima, h=(255))  
prediccion_autoarima
```



Mientras que las predicciones del modelo ajustado mediante el bucle serían:



Si R hubiera permitido estimar con un periodo acorde a la serie, la estacionalidad se habría captado mucho mejor y se podría haber obtenido unas predicciones que lo reflejasen.

Después del cálculo del RMSE, se ha obtenido que para el modelo de auto.arima este es de 515.847 y para el modelo ajustado a través del bucle es de 140.7669. La diferencia es grande debido a que el primer modelo, tal y como muestra el test de Ljung-Box, solo predice bien periodos recientes. Sin embargo, el segundo modelo, al predecir lo que es aproximadamente una media o la tendencia, los errores cometidos son menores. Se concluye que ninguno de los dos modelos consigue captar la naturaleza de la serie temporal por lo que se han probado los siguientes modelos con el fin de conseguir un modelo que aporte unas predicciones más fiables.

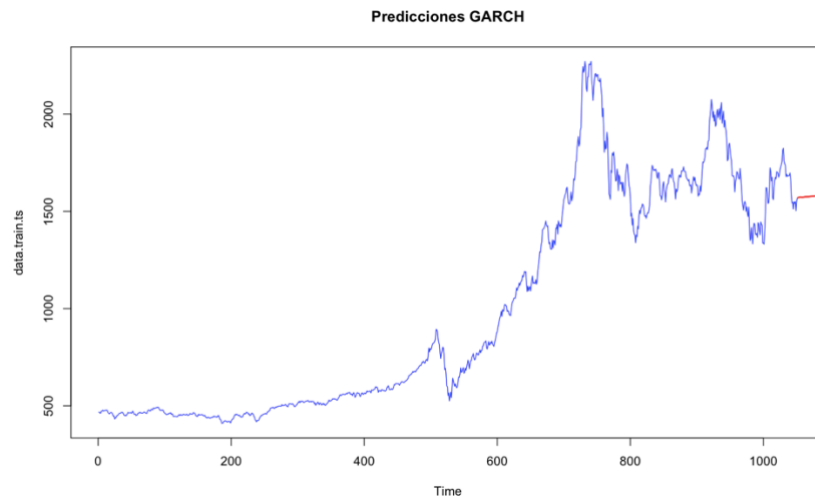
4.3 Ajuste del modelo GARCH

Para el ajuste del modelo GARCH se utilizará la misma serie temporal que en el apartado anterior, sin realizar ninguna transformación previa en ella, es decir, el train set en bruto.

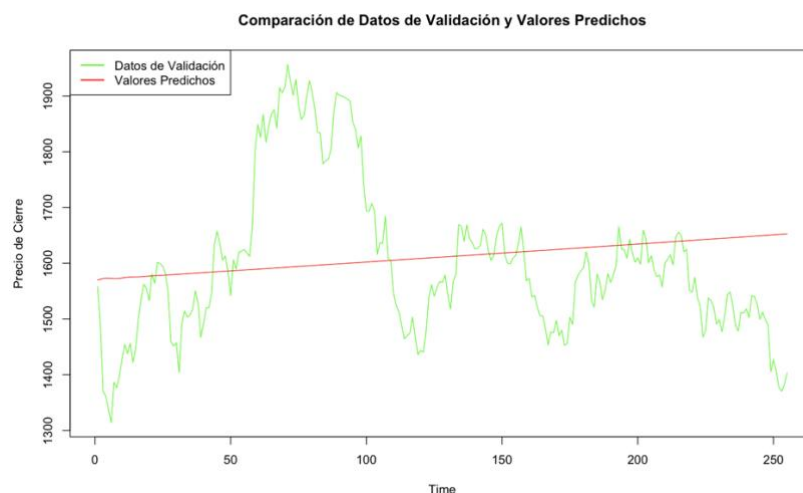
En la estimación, se ha usado la librería rugarch, que proporciona herramientas para el modelado y análisis de modelos de volatilidad financiera utilizando la metodología GARCH (Generalized Autoregressive Conditional Heteroskedasticity). Dentro de este, la función ugarchspec nos permitirá parametrizar el atributo spec, que será utilizado para estimar las predicciones del modelo.

```
spec <- ugarchspec(variance.model = list(model = "sGARCH", garchOrder = c(0, 2)),
                  mean.model = list(armaOrder = c(3, 2)))
garch_model <- ugarchfit(spec, data = data.train.ts)
plot(garch_forecast)
```

Después de probar con diferentes órdenes GARCH, debido a que el orden ARMA quedaba definido por la primera parte de nuestra estimación, se ha llegado a la conclusión de que este modelo era el que mejor RMSE reportaba. Mediante la función plot, se grafica el conjunto de entrenamiento de la serie temporal, desde la observación 1 hasta la 1051, junto con la predicción, de la 1052 hasta la 1306, como se observa a continuación.



Para comparar los datos del conjunto de validación, se ha calculado el RMSE y se han graficado las predicciones. Como se puede observar, el modelo GARCH tampoco capta bien la volatilidad de la serie temporal y pronostica unas predicciones con tendencia creciente, cuyo error es 144.98, una cifra aceptable. Debido a que parece que este modelo tampoco capta bien la naturaleza de los datos, se prueba una red neuronal.



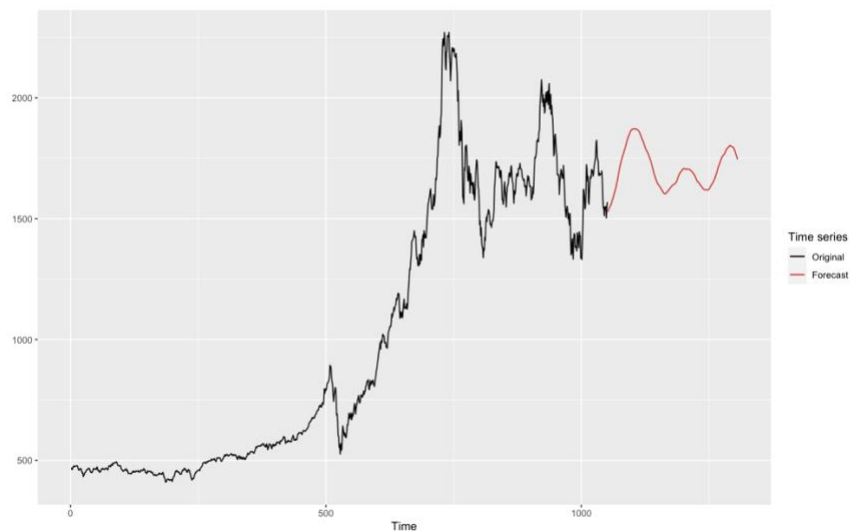
4.4 Ajuste Red Neuronal

Para el ajuste de este modelo se ha recurrido al paquete "tsfgrnn". Este es utilizado para la previsión de series temporales basada en redes neuronales de regresión generalizada (GRNN). Con la ayuda del software, se pueden crear modelos GRNN vinculados a series

temporales y utilizar dichos modelos para predecir los valores futuros de las series temporales (Martinez, Frias, Conde, & Martinez, 2019).

Para la predicción, definimos la longitud del test set y en el atributo lags indicamos el número de retardos, que después de calcular el RMSE con diferentes valores y teniendo en cuenta el orden AR de la serie, se concluye que es 3. El método MIMO hace referencia a (Multiple Input Multiple Output) y se basa en la utilización de vectores de valores sucesivos de series temporales como objetivos de entrenamiento. Estos vectores tienen la misma longitud que el horizonte de predicción (Martinez, Frias, Conde, & Martinez, 2019).

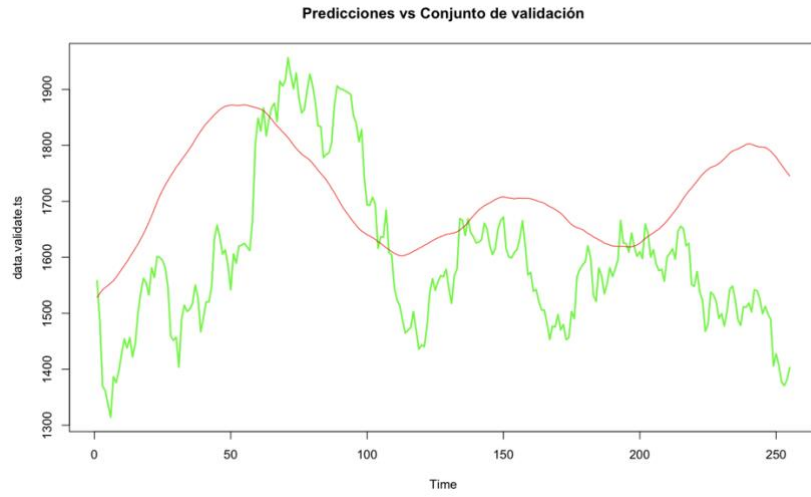
```
pred <- grnn_forecasting(timeS = data.train.ts, h = 255, lags = c(3), msas = "MIMO", transform = "none")
autoplot(pred)
```



La función autoplot nos devuelve un gráfico sencillo del conjunto de entrenamiento al que añade la predicción. Como se puede observar, independientemente del RMSE que obtengamos de este modelo, el modelo capta la presencia de tendencia y estacionalidad, a diferencia de los anteriores, y el ajuste a simple vista parece ser más elaborado y acertado.

Para comprobarlo, se calcula el RMSE obteniendo el resultado de 176.7435. Este valor es ligeramente superior al del modelo ARIMA, pero debido a que este modelo capta la estacionalidad y la tendencia, se concluye que su carácter predictivo es mayor y más extrapolable a periodos futuros u otras series temporales de tipología similar.

```
plot(data.validate.ts, lw = 2, col = "green", main = "Predicciones vs Conjunto de validación")
lines(predictions, col = "red", lw = 1)
mse <- mean((data.validate$Close - predictions)^2)
rmse <- sqrt(mse)
rmse
```



El código completo se encuentra en el apartado Anexo.

5 Conclusiones

Tras el estudio de los tipos de energías renovables y la alternativa que estas van a suponer en un futuro próximo a los combustibles fósiles debido a su disponibilidad finita, se ha expuesto la importancia que este análisis puede tener como herramienta complementaria al estudio de venideras crisis energéticas, sobre todo en la etapa de transición entre ambas. El primer modelo, obtenido mediante la función `auto.arima()`, ha arrojado resultados mejorables debido a que, por la complejidad de la serie, no ha conseguido captar bien los patrones estacionales observados en la misma. Es por ello por lo que se ha probado un modelo ARIMA personalizado teniendo en cuenta la estacionalidad de la serie, cuyos resultados no han sido mucho más esperanzadores que el modelo anterior, aunque ambos predecían valores en el corto plazo con errores bajos. Por consiguiente, un modelo GARCH se ajustaría mejor a la varianza de la serie y proporcionaría predicciones más precisas. Sin embargo, aunque las últimas fueron acertadas, se consideró que el modelo no era extrapolable a otros periodos o series porque tampoco lograba percibir la estacionalidad y por ende, ajustarse bien al índice de energías renovables, por lo que se pasó a probar un modelo más complejo; el GRNN. Gracias a la aplicación de este último modelo, se ha comprobado que las estimaciones de este índice son posibles y que el mismo capta con bastante robustez los patrones observados en el análisis de la serie financiera.

Para estudios posteriores, se propone realizar un estudio de correlaciones entre este índice y los precios de los combustibles fósiles, así como de predicciones realizadas con este modelo sobre los mismos. De esta manera se podrá complementar y proporcionar mayor robustez a la predicción de crisis energéticas.

6 Bibliografía

- Adebiyi, A., & Ayo, C. (2014). Stock Price Prediction Using the ARIMA Model . *16th International Conference on Computer Modelling and Simulation*.
- Alvarez, G. G. (2022). Se acerca el invierno: claves para entender la pobreza energética en el contexto de la crisis energética europea. *Panorama Social*(36), 65-77.
- Aponte, H., Torrejon-Managallanes, J., & Perez, A. (2022). Marea Negra en el Peru: reflexiones sobre un derrame de petróleo en el Pacífico sudamericano. *South Sustainability*, 3(1), 1-3.
- Barrigüete, J. L. (2021). Machine Learning II: Series Temporales. *Universidad Pontificia de Comillas*.
- Berat Sezer, O., Ugur Gudelek, M., & Murat Ozbayoglu, A. (2020). Financial time series forecasting with deep learning : A systematic literature review: 2005–2019. *Applied Soft Computing Journal*, 1-10.
- Bermejo Iglesias, A., Torres Pruñonosa, J., & Masferrer Llavínés , N. (2009). La utilización de los modelos ARIMA en la estimación del precio de acciones: una aplicación preliminar. *Administrando en entornos inciertos. XXIII Congreso Anual AEDEM*, 1-13.
- Bp. (2023). *Statistical Review of World Energy to pass to new home at the Energy Institute*.
- Casas, M., & Cepeda, E. (2008). Modelos ARCH, GARCH y EGARCH: Aplicaciones a las series financieras. *Cuadernos de Economía*, 27, 287-319.
- Chandramouli Rayadurgam, V., & Mangalagiri, J. (2023). Does inclusion of GARCH variance in deep learning models improve financial contagion prediction? *Finance Research Letters*, 54, 1-9.
- Cinco Días. (20 de Mayo de 2022). ¿Qué está pasando con el precio del petróleo? *Cinco Días El País*.
- Dama, F., & Sinoquet, C. (26 de septiembre de 2021). Time Series Analysis and Modeling to Forecast: a Survey. *Universidad de Nantes*.
- De Arce, R., & Mahia, R. (2001). Modelos ARIMA. *U.D.I. Econometría e Informática*.
- EFE. (21 de Agosto de 2022). Los precios desbordan a las familias: "Tenemos que hacer números a diario". *20minutos*.

- Fakhfekh, M., & Jeribi, A. (2020). Volatility dynamics of crypto-currencies' returns: Evidence from asymmetric and long memory GARCH models. *Research in International Business and Finance*, 51, 1-10.
- Fondo Económico Internacional para las energías renovables. (2023). Obtenido de <https://www.iwr.de/renixx/>
- Francq, C., & Zakoian, J.-M. (2019). *GARCH Models: Structure, Statistical Inference and Financial Applications*. John Wiley & Sons.
- Garcia, M. C., Jalal, M. A., Garzon, L. A., & Lopez, J. M. (2013). Métodos para predecir índices bursátiles. *Ecos de economía*(37), 51-82.
- Investing. (23 de Marzo de 2023). *Commodities*. Obtenido de Investing.com: <https://es.investing.com/commodities/brent-oil-historical-data>
- Jenkins, G. M., Box, G. E., Reinsel, G. C., & Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control*. John Wiley & Sons.
- Khan, S., & Alghulaiakh, H. (2020). ARIMA Model for Accurate Time Series Stocks Forecasting. *International Journal of Advanced Computer Science and Applications*, 11(7), 25-32.
- Kumar Dubey, A., Kumar, A., García-Díaz, V., Kumar Sharma, A., & Kanhaiya, K. (2021). Study and analysis of SARIMA and LSTM in forecasting time series data. *Sustainable Energy Technologies and Assessments*, 1-11.
- Kuyper, J. C., & Ramirez Morales, S. (2014). *Fuentes de energía renovables y no renovables. Aplicaciones*. Mexico: Alfaomega.
- Leon, S. V. (2021). El Nord Stream 2, la política energética rusa frente a Europa y las alternativas para la UE. *Instituto Español de Estudios Estratégicos*(135), 890-903.
- Li, W.-M., Liu, J.-W., Le, J.-J., & Waing, X.-R. (2005). The financial time series forecasting bases on proposed ARMA -GRNN model. *Fourth International Conference on Machine Learning and Cybernetics*, (págs. 18-21). Guangzhou.
- Martinez, F., Frias, M., Conde, A., & Martinez, A. (2019). *Time Series Forecasting with GRNN in R: the tsfgrnn Package*. Obtenido de <https://cran.r-project.org/web/packages/tsfgrnn/vignettes/tsfgrnn.html>
- McKinsey & Company. (2022). *Global Energy Persepective 2022*.
- Pellicer, J. G. (2021). Aplicación del modelo ARIMA en el mercado continuo español durante la crisis sanitaria de la Covid-19.

- Repsol. (2023). *La tecnología, una de las principales vías para frenar el cambio climático*. Obtenido de <https://www.repsol.com/es/sostenibilidad/cambio-climatico/nuevos-desarrollos-tecnologicos/index.cshtml>
- Ritchie, H., Roser, M., & Rosado, P. (2022). Energy. *Our World in Data*.
- Sanborn, S., & Chatterjee, A. (2023). *Deloitte 2023 renewable energy industry outlook*.
- Taskaya-Temizel, T., & Casey, M. (2005). A comparative study of autoregressive neural network hybrids. *Science Direct*, 781–789.
- Torres, R. (2022). Crisis energética: las respuestas de Alemania, España, Francia e Italia. *Cuadernos de información económica*(288), 11-18.
- Velilla-Avilez, D. A., Velez-Jaramillo, I. E., & Franco Sepulveda, G. (2014). Analisis del precio del carbón mediante redes neuronales artificiales (RNA). *Boletin de ciencias de la tierra*(35), 31-35.
- Wang, J. J., Wang, J. Z., Zhang, Z. G., & Guo, S. P. (2012). Stock index forecasting based on a hybrid model. *Omega*, 40(6), 758-766.
- Wei, W., Jiang, J., Liang, H., & Gao, L. (2016). Application of a Combined Model with Autoregressive Integrated Moving Average (ARIMA) and Generalized Regression Neural Network (GRNN) in Forecasting Hepatitis Incidence in Heng County, China. *PLOS*, 1-13.
- Zhao, H., & Guo, S. (2014). Annual Energy Consumption Forecasting Based on PSOCA-GRNN Model. *Hindawi Publishing Corporation*, 1-12.

7 Anexo

7.1 Código implementado

```
#limpiamos el espacio de trabajo
rm(list = ls())

library(readxl)
###[1] ANALIZAR LOS DATOS ###
#Cargamos los datos
setwd(dirname(rstudioapi::getActiveDocumentContext())$path)
data = read_excel("renixx.xlsx")

#Realizamos un breve estudio de los datos cargados
class(data)
str(data)
dim(data)
colnames(data)
summary(data)
head(data)
tail(data)

#Damos formato a la fecha como objeto de tipo fecha
data$Date <- as.Date(data$Date, "%m/%d/%Y")

#Creamos la serie temporal con periodicidad diaria
ST = ts(data$Close,start=c(2015,1), end=c(2023,5),frequency = 252)
#252 porque son los días laborables de un año

#descomponemos la serie en tendencia, estacionalidad y componente aleatorio
componentes=decompose(ST)
plot(componentes)
#observamos que hay tanto tendencia como estacionalidad

#dividimos la muestra en train y validacion
data.train <- subset(data, "2018-01-02"<=Date & Date<"2022-05-05")
data.train.ts <- as.ts(data.train$Close, frequency = 252)
plot(data.train.ts)
#Tratamiento de outliers

# Paso 1: Calcular el rango intercuartílico (IQR)
Q1 <- quantile(data.train.ts, 0.25)
Q3 <- quantile(data.train.ts, 0.75)
IQR <- Q3 - Q1

# Paso 2: Definir el umbral para identificar los outliers
umbral <- 1.5 * IQR

# Paso 3: Identificar los índices de los outliers
outliers <- which(data.train.ts < Q1 - umbral | data.train.ts > Q3 + umbral)
```

```

# Paso 4: Sustituir los outliers por la media de los datos circundantes

data.validate <- subset(data, "2022-05-05"<=Date & Date<="2023-05-05")
data.validate.ts <- as.ts(data.validate$Close, frequency=252)
plot(data.validate.ts)

#tratamiento de outliers
library(tsoutliers)
#modelo.outliers = auto.arima(data.train.ts, trace=TRUE)
#outliers=tso(ST)
#plot(outliers)
#despues cambiaremos el modelo arima, este es un primer ajuste

#dibujamos y analizamos las FAS y FAP de la serie sin ninguna transformación
par(mfrow=c(1,2))
acf(data.train.ts)
pacf(data.train.ts)
par(mfrow=c(1,1))

#vemos si es necesario estabilizar la media (diferenciar)
# Test de Dickey-Fuller
#install.packages("tseries")
library(tseries)
adf.test(data.train.ts, alternative="stationary", k=30)
# p-valor =0.375 -> rechazamos h0
#con este p valor puede hacer falta diferencias - la metemos en el modelo

#vemos si es necesario estabilizar la varianza (heterocedasticidad)
library(forecast)

Lambda=BoxCox.lambda(data.train.ts)
Lambda
#como el valor de lambda es cercano a 0 es cuando se obtiene la mayor verosimilitud
#hay heterocedasticidad y hay que eliminarla
z = BoxCox(data.train.ts,Lambda)
plot(z)

#diferenciamos la serie y analizamos las FAS Y FAP para identificar modelos posibles
ggtsdisplay(z, lag.max = 20)

Bz <- diff(z,differences = 1)
ggtsdisplay(Bz, lag.max = 20)

par(mfrow=c(1,2))
acf(Bz)
pacf(Bz)
par(mfrow=c(1,1))

# Seasonal Differentiation
B12z<- diff(z, lag =252 , differences = 1)
ggtsdisplay(B12z,lag.max = 200) #differences contains the order of differentiation
par(mfrow=c(1,2))

```

```

acf(B12z)
pacf(B12z)
par(mfrow=c(1,1))
###[2] AJUSTE DEL MODELO ARIMA ###

autoarima = auto.arima(z,seasonal=TRUE,stepwise=FALSE, approx=FALSE)
autoarima

#(3,1,2) with drift

library(lmtest)
coeftest(autoarima)
tsdiag(autoarima)

prediccion_autoarima <- forecast(autoarima, h=(255))
prediccion_autoarima

pred_transformed <- InvBoxCox(prediccion_autoarima$mean, lambda = Lambda)

#RMSE
valores_predautoarima <- as.vector(pred_transformed)
# Graficar las predicciones y el conjunto de validación
plot(data.validate.ts, lw = 2, col = "green", main = "Predicciones vs Conjunto de validación")
lines(valores_predautoarima, col = "red", lw = 1)
mse_autoarima <- mean((data.validate$Close - valores_predautoarima)^2)
rmse_autoarima <- sqrt(mse_autoarima)
rmse_autoarima

per=5 #sabemos que la estacionalidad es anual por la gráfica decompose
ordenregular = 3 # Hasta que orden regular queremos probar
ordenestacional = 3 # Hasta que orden estacional queremos probar
difregular = 3 # Hasta que orden de diferenciación regular queremos probar
difestacional = 3 # Hasta que orden de diferenciación regular queremos probar

lista<-c()
lista2<-c()
for (p in 0:ordenregular){
  for (d in 0:difregular){
    for (q in 0:ordenregular){
      for (P in 0:ordenestacional){
        for (D in 0:difestacional){
          for (Q in 0:ordenestacional){
            prueba = try(Arima(data.train.ts, order=c(p,d,q),
                          seasonal = list(order=c(P,D,Q), period=per),
                          lambda = Lambda,
                          include.constant = TRUE),silent = TRUE)

            p1<-try(prueba$aic, silent = TRUE)
            lista<-append(lista,p1,after = length(lista))
          }
        }
      }
    }
  }
}

```

```

ORDEN<-paste(p,d,q,P,D,Q)
lista2<-append(lista2,ORDEN,after = length(lista2))

}}}}

valores<-na.omit(data.frame(as.numeric(lista),lista2))
valores

dimension=dim(valores)
dimension[1]

min(valores$as.numeric.lista.)

#Partimos de un modelo naive
Modelo.naive = Arima(data.train.ts, order=c(0,0,0),
                    seasonal = list(order=c(0,0,0), period=5),
                    lambda = Lambda,
                    include.constant = TRUE)

mejor.aic = Modelo.naive$aic
mejor.aic

for (i in 1:dimension[1])
{
  if(valores[i,1] < mejor.aic)
  {
    mejor.aic=valores[i,1]
    mejor.modelo=valores[i,2]
  }
}

mejor.modelo

arima_bucle = Arima(data.train.ts, order=c(3,2,3),
                  seasonal = list(order=c(2,1,2), period=5),
                  lambda = Lambda,
                  include.constant = FALSE)
arima_bucle
coefest(arima_bucle)
tsdiag(arima_bucle)

#realizamos las predicciones
prediccion_arimabucle <- forecast(arima_bucle, h=(255))
prediccion_arimabucle
#ERROR OUT OF SAMPLE
valores_predarimabucle <- as.vector(prediccion_arimabucle$mean)
# Graficar las predicciones y el conjunto de validación
plot(data.validate.ts, lw = 2, col = "green", main = "Predicciones vs Conjunto de validación")
lines(valores_predarimabucle, col = "red", lw = 1)
mse_arimabucle <- mean((data.validate$Close - valores_predarimabucle)^2)
rmse_arimabucle <- sqrt(mse_arimabucle)

```

```
rmse_arimabucle
```

```
#no se ha encontrado un modelo ARIMA que ajuste y prediga bien  
#se van a probar otros modelos: GARCH y RNA
```

```
#####  
##### RNA #####  
#####
```

```
#install.packages("tsfgrnn")  
library(tsfgrnn)  
pred <- grnn_forecasting(timeS = data.train.ts, h = 255, lags = c(3), msas = "MIMO", transform  
= "none")  
autoplot(pred)  
predictions <- as.vector(pred$prediction)
```

```
#ERROR OUT OF SAMPLE  
# Graficar las predicciones y el conjunto de validación  
plot(data.validate.ts, lw = 2, col = "green", main = "Predicciones vs Conjunto de validación")  
lines(predictions, col = "red", lw = 1)  
mse <- mean((data.validate$Close - predictions)^2)  
rmse <- sqrt(mse)  
rmse
```

```
#####  
##### GARCH #####  
#####
```

```
library(rugarch)  
#Definir el objeto spec para el modelo GARCH  
spec <- ugarchspec(variance.model = list(model = "sGARCH", garchOrder = c(0, 2)),  
                  mean.model = list(armaOrder = c(3, 2)))
```

```
#Ajustar el modelo GARCH a la serie temporal  
garch_model <- ugarchfit(spec, data = data.train.ts)
```

```
#Obtener las predicciones condicionales y la volatilidad pronosticada  
garch_forecast <- ugarchforecast(garch_model, n.ahead = 255)
```

```
#Visualizar las predicciones y la volatilidad pronosticada  
plot(garch_forecast)
```

```
# Obtener las predicciones condicionales  
pred <- fitted(garch_model)
```

```
# Establecer los límites del eje Y  
ylim <- c(min(data.train.ts, pred), max(data.train.ts, pred))
```

```
# Graficar los datos y las predicciones  
plot(data.train.ts, ylim = ylim, type = "l", col = "blue", main = "Predicciones GARCH")
```

```

# Obtener la longitud del conjunto de entrenamiento
n_train <- length(data.train.ts)
indices_pred <- n_train + seq_along(garch_forecast@forecast$seriesFor)
# Graficar las predicciones
lines(indices_pred, garch_forecast@forecast$seriesFor, col = "red", lw = 2)

#ERROR OUT OF SAMPLE
# Obtener los valores predichos del modelo GARCH para la serie de validación
predictions <- as.numeric(garch_forecast@forecast$seriesFor)

# Crear un gráfico que compare los datos de validación y los valores predichos
plot(data.validate.ts, main = "Comparación de Datos de Validación y Valores Predichos",
      ylab = "Precio de Cierre", col = "green", lwd = 1)
lines(predictions, col = "red", lwd = 1)
legend("topleft", legend = c("Datos de Validación", "Valores Predichos"),
      col = c("green", "red"), lwd = 2)

mse <- mean((data.validate$Close - predictions)^2)
rmse <- sqrt(mse)
rmse

```