



Facultad de Ciencias Económicas y Empresariales | ICADE

IMPACTO DE LAS NOTICIAS SOSTENIBLES EN LA COTIZACIÓN BURSÁTIL: UN ANÁLISIS DE EVENTOS SOBRE ENDESA

Clave: 202010463

Alumno: Genoveva Alférez Aragón

Director: Leandro Sergio Escobar Torres

MADRID | JUNIO 2025

Resumen

Este trabajo analiza el impacto de las noticias relacionadas con la sostenibilidad (ESG) en la cotización bursátil de una empresa española de referencia en el sector energético: Endesa. A través de una metodología que combina técnicas de *web scraping*, procesamiento del lenguaje natural (NLP) y modelos econométricos como ARIMA, CAPM y los Retornos Anormales Acumulados (CAR), se examina cómo reacciona el mercado ante eventos informativos con contenido ESG.

El análisis se basa en una base de datos compuesta por más de 2.000 noticias extraídas de fuentes digitales, clasificadas tanto por dimensión ESG (ambiental, social y de gobernanza) como por tonalidad (positiva, negativa o neutra). Los resultados muestran que las noticias con contenido ESG, especialmente aquellas con tono negativo, generan reacciones más intensas en el precio de las acciones, evidenciando una respuesta asimétrica del mercado. Además, el impacto acumulado de las noticias ESG tiende a ser más estable y generalmente favorable en comparación con las noticias no relacionadas con sostenibilidad.

Este estudio aporta a la literatura existente al aplicar herramientas analíticas avanzadas sobre datos históricos de una empresa española concreta, ofreciendo evidencia empírica valiosa tanto para investigadores académicos como para profesionales financieros interesados en la intersección entre sostenibilidad y valoración de activos.

Palabras clave

Sostenibilidad, ESG, Endesa, IBEX 35, Event Study, Series temporales, ARIMA, CAPM, Retornos anómalos, Scraping de noticias, Procesamiento del Lenguaje Natural, NLP

Abstract

This study examines the impact of sustainability-related (ESG) news on the stock price of Endesa, a leading company in Spain's energy sector. The study employs a methodology that integrates *web scraping*, natural language processing (NLP), and econometric models such as ARIMA, CAPM, and Cumulative Abnormal Returns (CAR) to analyze how the market responds to ESG-related news events.

The analysis is based on a dataset of over 2,000 news articles gathered from digital sources, thematically categorized by ESG dimensions (environmental, social, and governance) and sentiment (positive, negative, or neutral). The findings reveal that ESG news—especially those with a negative tone—tend to trigger stronger price reactions, indicating an asymmetric market response. Moreover, the cumulative impact of ESG news over time is generally more stable and positive compared to non-ESG news.

This work contributes to the existing literature by applying advanced analytical techniques to historical data from a specific Spanish company, providing empirical evidence of interest to both academic researchers and financial professionals focused on the intersection between ESG and asset valuation.

Key words

Sustainability, ESG, Endesa, IBEX 35, Event Study, Time Series, ARIMA, CAPM, Abnormal Returns, News Scraping, Natural Language Processing, NLP

Índice

1.	Introducción.....	7
2.	Revisión de la Literatura.....	8
2.1.	Teorías clásicas y modernas sobre sostenibilidad y mercado	8
2.2.	Eventos ESG y modelos de series temporales.....	9
2.3.	Natural Language Processing (NLP) y clasificación por tono	11
2.4.	Limitaciones	12
2.5.	Conclusiones de la revisión de la literatura	13
3.	Metodología.....	16
3.1.	Datos.....	16
3.2.	Scraping de noticias.....	16
3.3.	Natural Language Processing (NLP).....	16
3.4.	Análisis Descriptivo y Estadístico de Precios	17
3.5.	Evaluación de la serie temporal, test de estacionariedad, modelado con ARIMA y ruido blanco	18
3.6.	Estimación de retornos anómalos con CAPM.....	20
3.7.	CAR (Retorno anómalo acumulado).....	21
3.8.	Contrastes de significación estadística	22
4.	Estudio Empírico	23
4.1.	Datos.....	23
4.2.	Scraping de noticias.....	23
4.3.	Natural Language Processing (NLP).....	26
4.3.1.	Clasificador NLP por categoría ESG.....	26
4.3.2.	Clasificador NLP por tono	26
4.4.	Análisis descriptivo y estadístico de precios	27
4.5.	Evaluación de la serie temporal y test de estacionariedad.....	31
4.6.	Modelado con ARIMA y ruido blanco	32
4.7.	Estimación de retornos anómalos con CAPM.....	35
4.8.	CAR (Retorno anómalo acumulado).....	37
4.9.	Contrastes de significación estadística	41
5.	Conclusiones.....	43
6.	Valor del estudio y futuras direcciones.....	43
7.	Declaración de Uso IA	44
8.	Bibliografía.....	45

9.	Apéndices y Ilustraciones.....	47
9.1.	Explicación API:.....	47
	<i>Ilustración Anexo 1: Capturas de proceso de descarga API.....</i>	48
9.2.	Otros Ilustraciones.....	49
	<i>Ilustración Anexo 2: Contador de noticias por tono</i>	49
	49
	<i>Ilustración Anexo 3: Evolución de score ESG a lo largo del tiempo por tono</i>	49
	<i>Ilustración Anexo 4: Evolución de retornos anómalos a lo largo del tiempo</i>	50
	<i>Ilustración Anexo 5: Boxplot sobre el cálculo de los retornos anómalos</i>	50
9.3.	Código scraping noticias y NLP	51
8.4.	Código análisis precios, Modelos ARIMA, White Noise, CAPM, CAR y test de contrastes.....	56

Tabla de Figuras

Ilustración 1. Evolución de noticias - Endesa	25
Ilustración 2. Tabla resumen precios de cotización Endesa y estadísticas	27
Ilustración 3. Distribución de precios de Endesa	28
Ilustración 4. Cálculo de retornos simples y logarítmicos.....	28
Ilustración 5. Evolución de los precios y retornos diarios de Endesa (2015-2025)	29
Ilustración 6. Distribución de retornos simples de Endesa.....	30
Ilustración 7. Autocorrelación y autocorrelación parcial de los retornos simples.....	31
Ilustración 8. Tabla de mejores modelos según optimizador de parámetros	32
Ilustración 9. Resultados modelo ARIMA (1,0,1).....	33
Ilustración 10. Resultados modelo ARIMA (0,0,2).....	33
Ilustración 11. Resultados modelo White Noise.....	33
Ilustración 12. Métricas modelos series temporales.....	34
Ilustración 13. Evolución y autocorrelación de residuos.....	35
Ilustración 14. Distribución de los retornos anómalos diarios mediante modelo CAPM	36
Ilustración 15. Evolución temporal de los retornos anómalos con media móvil de 30 días.....	36
Ilustración 16. Distribución general del CAR por ventana de evento (± 1 , ± 3 , ± 5 días) 38	
Ilustración 17. Distribución del CAR (± 3 días) por categoría ESG	38
Ilustración 18. Distribución del CAR (± 3 días) según el tono de la noticia.....	39
Ilustración 19. Interacción entre categoría ESG y tono de la noticia en el CAR (± 3 días)	40
Ilustración 20. Evolución acumulada CAR (± 5 días) ESG vs No ESG	41
Ilustración 21. Tabla de significación - Tres contrastes.....	42

1. Introducción

A lo largo de los años se ha planteado cuál es la verdadera función de las empresas y el sector privado. Friedman en 1970 declaró en *The New York Times* que la única responsabilidad social de la empresa es aumentar beneficios. Por el contrario, otros como Freeman (1984) defienden que, más allá del beneficio puramente económico, una empresa tiene el deber de considerar los intereses de todos los stakeholders, desde trabajadores, clientes, sociedad y medio ambiente entre otros. En una línea complementaria, Porter y Kramer (2006) consideraban que las empresas eran capaces de crear valor económico y a la vez generar beneficios sociales.

Ajustando nuestro foco a la sostenibilidad, Serafeim en 2020 publicó un estudio en el que demostraba que las empresas que generan iniciativas vinculadas a ESG resultaban en mayores cotizaciones bursátiles. De esta forma, los accionistas (principales stakeholders) valoraban positivamente las iniciativas sostenibles. Con esto, el autor argumenta que las empresas se benefician al ser sostenibles, incluso financieramente. Esta idea se aleja de la teoría puramente relacionada a beneficio económico propuesta por Friedman.

Cabe destacar que, desde 1970 el mundo ha cambiado sustancialmente en muchos sentidos, desde las pautas de consumo influidas por cambios culturales, la regulación, a los vehículos de inversión y productos financieros, que han evolucionado hacia mayor sensibilidad ambiental y social. Existen estudios como el de Friede et al. (2015) en el que analizaron más de 2,000 estudios para concluir que la sostenibilidad y las prácticas ESG tienen de media un impacto positivo en los rendimientos financieros, refutando la idea de que ESG implique sacrificar rentabilidad.

Estas tensiones entre beneficio económico, sostenibilidad y creación de valor han trascendido la teoría, y hoy ocupan un lugar central en el análisis de mercados financieros. La irrupción de los criterios ESG ha transformado tanto las decisiones de inversión como la percepción del riesgo y la valoración bursátil. En este contexto, surgen preguntas clave: ¿Cómo reacciona el mercado ante eventos sostenibles? ¿Se premian las buenas prácticas ESG? ¿Todas las empresas son igualmente sensibles a estas noticias?

El objetivo de nuestra investigación es analizar cómo los eventos y prácticas relacionadas con la sostenibilidad (ESG) afectan el precio y la volatilidad de la acción de Endesa, empresa clave del sector energético en España. Para ello, extraeremos noticias relacionadas a la empresa en la que se mencione la práctica de iniciativas ESG y mediremos el impacto en el precio de cotización tras el evento a través de distintas ventanas temporales (análisis de eventos o *event study*). Además, utilizaremos modelos de series temporales como ARIMA y plantearemos el uso de GARCH. Partimos de la hipótesis de que las noticias sobre sostenibilidad con tono positivo tienen un impacto positivo en la valoración de los activos, en nuestro caso Endesa.

Esta investigación no solo pretende contribuir al debate académico, sino también ofrecer evidencia empírica sobre cómo los mercados valoran la sostenibilidad a través de una empresa concreta y representativa del sector energético español; Una de las empresas más influyentes en el panorama español, líder en la transformación energética que cumple ya 80 años operando en nuestro país. Unido a esto, mi especial conexión a Endesa, tras haber sido becaria de la Fundación Endesa durante mis años de estudios universitarios, me impulsan con gran cariño a hacer este análisis. Soy consciente del gran valor que esta compañía aporta al tejido empresarial, laboral y social español, y es por esto, que parto de la premisa de que las acciones llevadas a cabo por una empresa de estos valores deben y han de verse reflejados en su desempeño financiero y de valoración.

2. Revisión de la Literatura

2.1. Teorías clásicas y modernas sobre sostenibilidad y mercado

Durante el siglo pasado, se podía llegar a pensar que la función primordial de toda empresa era maximizar el valor generado para el accionista. Esta visión se puede ver plasmada con la idea de Friedman en el 1970 con su publicación de ensayo en el periódico The New York Times. En este, Friedman defendía que la única responsabilidad social de la empresa era crear beneficios para el accionista sin salirse del marco legal. Poco a poco esta visión ha sido cada vez más cuestionada, sobre todo en los últimos años con latente preocupación en términos medioambientales, sociales y de gobernanza.

Autores como Freeman (1984) y Porter y Kramer (2006), junto con muchos otros autores, cuestionan esta idea de Friedman. Ellos defienden modelos de creación de valor distintos, donde se da mayor importancia, ya no solo al accionista, sino a los demás stakeholders involucrados. Esta evolución en la literatura se ha visto impulsada o va de la mano de los cambios producidos en los mercados. Consumidores, empleados, inversores y gobiernos cada vez más preocupados e involucrados en la causa climática. Las empresas, ya sea filantrópicamente o por intereses específicos, deben adaptarse a lo que el consumidor e inversor del siglo XXI demanda. Esto ha hecho que los problemas de sostenibilidad (ESG), estén adquiriendo una creciente relevancia en la gestión empresarial y estratégica como en su valoración financiera y, por ende, en la literatura como objeto de estudio y análisis. Estudios como el de Hoepner y Schopohl (2018) afirman que efectivamente inversores, en este caso europeos de tipo institucional (fondos de pensiones, bancos, compañías de seguros y otros fondos de inversión) pueden aplicar sus prácticas morales a la hora de invertir sin necesariamente tener que renunciar a rendimientos financieros.

2.2.Eventos ESG y modelos de series temporales

Un enfoque de la literatura que puede resultar útil para el estudio de la sostenibilidad es considerar los eventos ESG como shocks o choques susceptibles de generar reacción en los mercados financieros a su vez generados por la reacción de los inversores. Desde esta perspectiva, los eventos sostenibles ya no solo representarían iniciativas éticas o de impacto positivo de manera desinteresada, sino un conjunto de señales que pueden alterar la percepción de valor de una empresa cotizada en el mercado. Capelle-Blancard y Petit analizaron más de 33,000 noticias relacionadas con sostenibilidad a lo que aplicaron un análisis de eventos o event study. Con ello buscaban medir el impacto de dichas publicaciones en el precio de cotización bursátil a través de la reacción de los inversores (2019). En sus conclusiones se explica que el inversor tiende a castigar más las noticias negativas vinculadas a sostenibilidad que premiar las positivas. Esto sugiere una asimetría en el mercado ante ESG. En 2015, Krüger ya había analizado esta asimetría y concluía algo muy similar; que los eventos negativos de prácticas relacionadas a sostenibilidad se veían más castigados en el precio de cotización que premiados ante eventos positivos. Estos estudios nos servirán como referencia en el análisis de eventos (event study) que queremos realizar y servirán como punto de partida.

Teniendo en cuenta que analizaremos precios históricos y rentabilidades, no debemos olvidar las dependencias que pueden existir entre estos datos, influyendo los precios y rentabilidades pasadas a los nuevos precios y variaciones de estos. Es por esto que, fuera del impacto en los precios de las practicas ESG, otro componente que debemos tener en mente son los modelos de análisis de series temporales. Estos modelos serán fundamentales para analizar las reacciones bursátiles y la volatilidad después de eventos ESG. Los modelos econométricos que contrastaremos serán progresivamente *White Noise* (WN), *Autoregressive* (AR), *Moving Average* (MA), *Autoregressive Moving Average* (ARMA), *Autoregressive Integrated Moving Average* (ARIMA) y potencialmente *Generalized Autoregressive Conditional Heteroscedasticity* (GARCH). Estos modelos estadísticos de series temporales nos permitirán modelar y comparar la capacidad predictiva sobre las series de precios de Endesa. Además, con otros modelos clásicos como CAPM podremos comprobar si el precio esperado coincide con el que resulta del mercado tras el anuncio de eventos ESG. De esta manera, compararemos el precio predicho vs el precio resultante, midiendo el impacto de las noticias de sostenibilidad en rentabilidad y su volatilidad.

Autores como Ghani et al. (2024) han empleado modelos GARCH para estudiar cómo los índices de sostenibilidad (ESG) y de incertidumbre sobre políticas climáticas (CPU) afectan a la volatilidad del mercado bursátil estadounidense. Sus resultados muestran que el índice ESG es un predictor más potente que el CPU, y que las empresas con mayor desempeño en sostenibilidad tienden a experimentar menor volatilidad, actuando así como un amortiguador del riesgo financiero. Por otro lado, Jiang (2023) analizaron la volatilidad de un activo concreto, la

acción de Apple. Para su análisis de una empresa concreta utilizaron un modelo ARIMA-GARCH. Inspirados en estos autores, y siguiendo una metodología similar al analizar un único activo, plantaremos la posibilidad de utilizar un modelo similar. Su estudio reveló que esta combinación de modelos captura eficazmente las características de volatilidad de las acciones, proporcionando predicciones precisas a corto plazo. Del mismo modo, en el contexto español también se ha hecho uso de este tipo de modelos en distintos estudios, no necesariamente asociados a un activo financiero. López et al. (2017) publicaron un estudio donde destacaban la importancia del uso de predicciones con modelos de series temporales para anticiparse a la toma de decisiones y conocimiento de valores futuros. En su caso, ellos abordaron el tráfico aeroportuario de pasajeros en España haciendo uso de modelos SARIMAX. Otro ejemplo relevante de aplicación de modelos es el estudio de McKnight (2019), quien analizó el comportamiento de cuatro acciones individuales, Harmony Gold Mining Co., Hewlett-Packard, Target Corporation y Tiffany & Co., empleando modelos clásicos como CAPM, junto con alternativas más sofisticadas como modelos autoregresivos de media móvil y modelos aditivos generalizados. Su investigación concluye que, aunque algunos modelos avanzados ofrecían mejoras marginales, el CAPM seguía proporcionando una base sólida y comparablemente efectiva para estimar el comportamiento de los retornos en determinadas condiciones de mercado. Este hallazgo refuerza la utilidad del CAPM como punto de partida válido en contextos como el presente trabajo, donde también buscaremos analizar el impacto de eventos sobre el retorno esperado de una acción concreta (Endesa) en relación con un índice de mercado (en nuestro caso el índice de referencia español, IBEX 35).

En el contexto asiático, Wang et al. (2023) llevaron a cabo un análisis de eventos aplicando modelos GARCH a empresas chinas, examinando cómo las noticias ESG afectan tanto al precio como a la volatilidad diaria de las acciones. Sus hallazgos indican que las noticias sobre ESG tienen un impacto mayor y más significativo en el mercado de valores que las noticias no relacionadas con el ESG. Dentro de las categorías que ellos estudiaron, Environment, Social y Governance, comprobaron que las noticias medioambientales tienen un impacto mayor y más significativo. Este estudio ofrece nuevas ideas para una evaluación completa y objetiva del rendimiento ESG de las empresas y la gestión de la imagen mediática de las empresas a partir del uso de las noticias ESG.

Otro ejemplo sería el de Albanese et al. en el 2023, quienes analizaron cómo respondían los mercados financieros de cinco países, Canadá, India, Japón, Inglaterra y Estados Unidos, ante noticias y eventos vinculados a energías renovables. Aunque su enfoque no se limitaba a una única empresa, el estudio muestra como los eventos ESG generan reacciones en el mercado bursátil de distinto impacto, según si la empresa tenía un perfil de mayor o menor sostenibilidad, diferenciadas como *green* o *brown* stocks. Utilizaron modelos GARCH-BEKK que les sirvieron

para concluir que los activos con perfil de mayor sostenibilidad o “green” ofrecen menor volatilidad y mejores oportunidades de diversificación, especialmente en contextos de alta incertidumbre climática. Estos hallazgos respaldan el interés por analizar cómo eventos ESG afectan el comportamiento de activos financieros, sean estos un grupo de activos o una única empresa, como es nuestro caso.

Por último, un estudio nacional de la Universitat Politècnica de Catalunya analizó cómo las noticias positivas, neutras y negativas sobre sostenibilidad afectaban a los precios de las acciones de las empresas energéticas estadounidenses y europeas, centrándose en las reacciones del mercado a corto plazo (Barroso del Toro et al. 2022). Utilizaron una metodología de análisis de eventos para analizar los rendimientos anormales medios acumulados (CAAR), ambas serán técnicas que veremos en nuestro estudio. Recopilaron 2,134 estudios de eventos según el tipo de fuente de energía (renovable, fósil o nuclear) y el tono de las noticias (clasificación que en nuestro caso será por categoría ESG y tono). Analizaron cómo las noticias positivas, neutras y negativas sobre sostenibilidad afectaron a los precios de las acciones de las empresas energéticas estadounidenses y europeas, centrándose en las reacciones del mercado a corto plazo (de igual modo, nuestro análisis analizará ventanas temporales de corto plazo). Los resultados muestran diferencias significativas entre ambos mercados. El estadounidense reaccionó de manera más optimista, incluso durante la pandemia, mientras que el europeo presentó una sensibilidad negativa mayor, registrando caídas en los precios incluso tras noticias positivas relacionadas con la energía nuclear o renovables. No obstante, en ambos contextos se observó que la energía nuclear mantenía relevancia en la agenda de los inversores. Además, los combustibles fósiles fueron menos penalizados tras noticias negativas o neutras, y mostraron reacciones iguales o más favorables ante noticias positivas, en comparación con otras fuentes de energía.

Por el hecho de ser un estudio nacional y las similitudes en las técnicas empleadas y sector (energético) nos resulta importante destacarlo.

2.3. Natural Language Processing (NLP) y clasificación por tono

Como ya hemos comentado a lo largo de estas líneas, haremos un análisis de eventos y analizaremos una serie de noticias que se contrastarán con precios de cotización. Para el análisis de estas noticias la metodología que llevaremos a cabo es una denominada Natural Language Processing (NLP) o Procesamiento del Lenguaje Natural. Esta metodología, cada vez más abordada en la literatura reciente gracias a los avances de la programación, big data y machine learning, nos permitirá extraer el mensaje principal de noticias relacionadas a sostenibilidad, categorizándolas y situándolas en nuestra línea del tiempo a analizar (10 años). Gracias a NLP podremos medir su tono, énfasis o tipo de lenguaje, cosas que influyen en la magnitud de la reacción de los inversores y del mercado. Cahan et al. (2015) estudiaron qué relación existía entre

la cobertura mediática y la responsabilidad social corporativa (RSC). Con su estudio explicaron que tanto el canal, como el estilo y la cobertura tienen un impacto en los inversores. Similar a esto llevaremos a cabo esa recopilación de noticias (en nuestro caso el canal es internet a través de Google links) y se aplicará scraping, analizando los textos de las noticias y clasificándolas según lo que contengan. Asimismo, el uso de tecnologías como el procesamiento del lenguaje natural (NLP) ha demostrado ser una herramienta eficaz para analizar grandes volúmenes de información no estructurada, como las noticias financieras. En este sentido, el estudio de Dorfleitner y Zhang (2024) aplica modelos de NLP basados en BERT para construir un conjunto de datos con más de 10.000 noticias relacionadas con ESG, procedentes de diversas fuentes. Su análisis revela que el mercado responde significativamente al tono de estas noticias, con reacciones más intensas ante noticias negativas. Además, observan que las empresas con mejor historial ESG tienden a sufrir impactos menores frente a eventos negativos. Estos hallazgos refuerzan el enfoque adoptado en este trabajo, donde el uso que haremos de scraping y análisis de sentimiento nos servirá para evaluar cómo el contenido textual de noticias sostenibles puede explicar movimientos anómalos en los precios de cotización. Este uso de tecnologías de análisis textual nos servirá como puente entre el contenido que encontremos que contenga menciones a la sostenibilidad y la valoración del mercado.

Otro punto por considerar que ya ha sido cubierto por algunos autores es la importancia que toma el tono y la narrativa con la que se transmite el mensaje. La forma en la que las noticias son comunicadas, aparte de su contenido, es fundamental en el estudio del impacto ESG. La literatura ha evidenciado que la forma de utilizar el lenguaje, el tono emocional percibido en la noticia y la cobertura mediática pueden amplificar o moderar la reacción que tiene el mercado ante la misma. Recordemos el trabajo seminal de Cahan et al. (2015), donde demuestran cómo la percepción de los inversores también está medida por el canal de comunicación. En ese caso no nos bastaría que la empresa tenga un comportamiento de sostenibilidad ejemplar si el mensaje que se quiere transmitir al público (más concretamente al inversor) no está comunicado adecuadamente. Gracias al uso de NLP, ya no solo podremos clasificar las diferentes noticias extraídas por temática (elementos individuales aislados, Environment, Social y Governance), sino también por el tono (positivas, negativas o neutras). Estos estudios justifican y refuerzan nuestra idea de automatizar la clasificación de las noticias de Endesa con NLP en relación a su contenido ESG y tono. La combinación de NLP y el análisis de modelos econométricos ofrece una visión íntegra del impacto de estas noticias en nuestra empresa cotizada a analizar.

2.4.Limitaciones

Aunque a lo largo de esta revisión de la literatura hayamos encontrado estudios que se sitúan en diferentes regiones, la literatura no especifica sobre empresas individuales españolas,

creando una laguna en el campo de los estudios seminales. Esto representa una buena oportunidad para contribuir de manera empírica con nuestro estudio, cubriendo uno de los principales actores en el campo de las empresas energéticas españolas con gran influencia en nuestro país.

Aparte de esta falta de estudios en el espacio nacional, existen otras limitaciones que hemos encontrado. Existen muchos trabajos que utilizan puntuaciones de ESG como los ratings elaborados por Bloomberg, MSCI o Sustainalytics. Estas métricas son útiles para medir cómo de sostenible es una empresa tras una evaluación extensa y valiosa. No obstante, estos ratings se suelen publicar con una frecuencia mensual o anual, no capturando las dinámicas temporales que noticias si pueden capturar. Esto limita a los ratings ESG para su aplicabilidad a la hora de realizar event studies. Otra limitación encontrada es la falta de integración entre métodos. Aunque encontremos estudios que realicen event study y otros que utilicen modelos econométricos, son pocos los que combinan ambos métodos. Por último, los estudios revisados se centran en paneles de empresas en vez de series temporales sobre un activo concreto único. Son escasos los estudios que pongan foco en una sola compañía, especialmente en nuestro país. Unido a estas limitaciones, los modelos NLP han avanzado mucho en el idioma inglés, estando poco desarrollado en español. Esta limitación representa una dificultad técnica, así como una oportunidad para mejorarlo.

2.5. Conclusiones de la revisión de la literatura

Teniendo en cuenta todos los estudios mencionados hasta ahora, vemos sensato el enfoque propuesto, proponiendo una clasificación automática de noticias sobre Endesa y sostenibilidad midiendo la tonalidad y el contenido como medio para analizar el impacto de estas. La combinación entre Scraping de noticias y el uso de modelos NLP para la clasificación de estas y modelos econométricos como ARMA y GARCH nos permitirá ofrecer una visión integrada del impacto en el precio de cotización de nuestra empresa cotizada a estudiar.

Esta revisión de la literatura nos trae a la convergencia de tres enfoques que crean diversas sinergias y que respaldan las técnicas metodológicas que aplicaremos en este análisis:

1. Event Study: autores como Krüger (2015), Capelle-Blancard y Petit (2019) o Friede et al. (2015) ya comprobaron con sus respectivos estudios el impacto de eventos vinculados a sostenibilidad sobre activos y mercados encontrando diversas conclusiones muy relevantes. Esto nos da la llave para medir el impacto puntual en los retornos tras un evento de tipo ESG y sus conclusiones nos servirán para comparar los resultados finales de nuestro análisis.
2. Modelos de series temporales: autores como Jiang (2023) que aplicó modelos como ARIMA-GARCH para analizar el comportamiento de una acción concreta (Apple) o Wang et al. (2023) que aplicaron modelos GARCH nos sirven para respaldar el uso de estos modelos de series temporales para predecir futuros precios y retornos, así

como volatilidades. McKnight (2019) en su estudio nos da la confianza para hacer uso del modelo clásico vinculado al mercado de CAPM al comprobar sus habilidades predictivas robustas frente a nuevos modelos más complejos.

3. NLP: sugerido y empleado por autores como Cahan et al. (2015) o Dorfleitner y Zhang (2024), que analizaron el tono y estilo de noticias vinculadas a sostenibilidad usando este método de clasificación. El uso de NLP aplicado a noticias sostenibles, nos permitirá clasificar y cuantificar contenido valioso dentro de las distintas noticias y que servirán para poner fecha a los eventos a estudiar.

La unión y articulación de estos enfoques nos sirve como base para la elaboración de nuestro TFG. Al aplicar esta triangulación sobre Endesa, buscamos ofrecer evidencia microeconómica sobre cómo los inversores valoran la sostenibilidad reflejada en mercados financieros y en qué condiciones se produce una reacción significativa en el precio.

De los estudios mencionados y otros revisados, la mayoría de estos tienen como escenario mercados internacionales (una de las limitaciones comentadas), siendo Estados Unidos y China protagonistas. En Europa también existen trabajos relevantes que relacionan sostenibilidad con valoración bursátil. Wanday y Ajour El Zein (2022) analizaron cómo la integración de estrategias ESG (ambientales, sociales y de gobernanza) influye en los rendimientos financieros de las empresas del sector energético en Europa, lo que encaja a la perfección con nuestro estudio por el sector, energético, y región, Europa, materia de estudio. Aunque hicieron uso de indicadores financieros como el retorno sobre el patrimonio (ROE), retorno sobre activos (ROA), retorno sobre ventas (ROS) y retorno sobre la inversión (ROI) de empresas cotizadas, lo complementaron con modelos SARIMA. Los resultados revelan una relación positiva entre las prácticas ESG y los rendimientos para los inversores, destacando que las empresas con estrategias ESG sólidas tienden a ofrecer mayores retornos. Estos hallazgos respaldan la hipótesis de que el mercado valora positivamente las iniciativas sostenibles, lo cual es relevante para el presente estudio. Veremos con nuestro análisis si se cumple efectivamente esto que Wanday y Ajour El Zein afirmaron en 2022 y si los inversores valoran, lo que bajo mi humilde opinión debería ser altamente valorado. No obstante, tener una buena estrategia sostenible que los inversores valoren no es nada sencillo. Así lo describieron Wanday y Ajour El Zein:

Implementar una buena estrategia ESG es una de las tareas más difíciles a las que se enfrentan las empresas energéticas. Los líderes tienen más probabilidades de lograr resultados cuando las organizaciones adoptan un enfoque estratégico de ESG vinculado a los objetivos y valores de la empresa. Dado que ESG abarca actividades sociales y de gobernanza, además del medio ambiente, las empresas de petróleo y gas deben analizar sus propias demandas y las de sus empleados. Para satisfacer a una variedad de partes interesadas, una empresa de petróleo y gas debe desarrollar una estrategia ESG. Los

programas ESG también son importantes para promover la innovación en la industria energética y reducir los riesgos.

Con la publicación del informe *Who Cares Wins* (Quien se preocupa, gana) de la Iniciativa del Pacto Mundial de las Naciones Unidas en 2004, se acuñó oficialmente el término ESG (ONU, 2004). Este informe estableció el ambicioso objetivo de reunir tres de los pilares más importantes de las finanzas éticas: el medio ambiente, lo social y la gobernanza. Todos ellos abordan retos diferentes y tienen un objetivo de evaluación distinto (Billio et al., 2021).

La mayoría de las empresas que se han centrado en la maximización de los beneficios han ignorado durante décadas las responsabilidades medioambientales, sociales y de gobernanza (ESG). Las obligaciones ESG no solo se consideraban poco influyentes en el éxito financiero, sino que también se veían como una posible carga para este, ya que estaban relacionadas con el aumento de los costes (Billio et al., 2021). El objetivo de este artículo es investigar la correlación entre la rentabilidad para los inversores y las actividades ESG en el sector energético. (Wanday & Ajour El Zein, 2022, pp. 1-2).

Veremos en este trabajo si efectivamente *Who cares* realmente *wins* tomando a Endesa como empresa comprometida con estas prácticas y que integra esta estrategia sostenible como parte de su estrategia global.

A modo de conclusión, la literatura existente nos ofrece una base robusta, aunque dispersa para abordar el impacto bursátil de eventos sostenibles. Encontramos grandes herramientas desde estudios de asimetría en las reacciones del mercado, análisis de tono con NLP hasta modelos como GARCH. A pesar de ello, siguen existiendo vacíos importantes especialmente en contextos no anglosajones y centrados en un único activo. Este trabajo busca aportar evidencia nueva al debate sobre la sostenibilidad, a través de una metodología innovadora. Lo hacemos en un contexto temporal reciente y en línea con los cambios nuevos del mercado con datos no estructurados en español. Esperamos que este estudio sirva para rellenar ese vacío encontrado y sirva como base para nuevos estudios.

Aunque lo que encontramos en la literatura relacionado con este tema suele ser con datos agregados (esto son índices bursátiles, fondos de inversión con varios activos, carteras, etc), en nuestro caso hemos optado por hacerlo con una empresa individual. Esto nos permite llevar a cabo un análisis más detallado y profundo sobre la que es una de las empresas más influyentes en el panorama español, líder en la transformación energética que cumple ya 80 años operando en nuestro país. Unido a esto, mi especial conexión a Endesa, tras haber sido becaria de la Fundación Endesa durante mis años de estudios universitarios, me impulsan con gran cariño a hacer este análisis. Soy consciente del gran valor que esta compañía aporta al tejido empresarial, laboral y social español, y es por esto, que parto de la premisa de que las acciones llevadas a cabo por una empresa de estos valores deben y han de verse reflejados en su desempeño financiero y de valoración.

3. Metodología

3.1. Datos

Para llevar a cabo el análisis se utilizan dos tipos de datos: series financieras históricas de cotización y noticias relacionadas con sostenibilidad. La combinación de ambos permite vincular eventos informativos (ESG) con el comportamiento del precio de la acción.

En primer lugar, se emplean datos históricos diarios del precio de la acción de Endesa, empresa cotizada en el IBEX 35. Se consideran variables financieras estándar como precio de apertura, precio de cierre, máximos, mínimos y volumen de negociación, que permiten analizar retornos y volatilidad. Estos datos se obtienen de fuentes públicas y fiables como *Investing.com* y se estructuran en un *DataFrame* para facilitar su tratamiento.

En segundo lugar, se recogen noticias que contienen simultáneamente el término “Endesa” y palabras clave vinculadas a sostenibilidad como “ESG”, “medioambiente”, “gobernanza”, “transición energética” o “social”. Esta recopilación forma la base de eventos a analizar en el estudio de eventos posterior. Los detalles de la técnica de extracción se desarrollan en el apartado siguiente.

3.2. Scraping de noticias

Para la obtención de eventos informativos relevantes relacionados con sostenibilidad, se exploraron dos enfoques metodológicos distintos de scraping de noticias, con el fin de seleccionar la solución más eficaz y reproducible:

- *Google RSS*: sistema basado en feedparser, especializado en la lectura de fuentes RSS. Esta metodología permitió obtener noticias recientes bien ordenadas cronológicamente. Su ventaja principal es la velocidad de ejecución y la posibilidad de automatizar múltiples consultas a través de un bucle en Python, pese a su limitación de 100 resultados por query.
- *Google Custom Search API*: alternativa con mayor cobertura histórica. A pesar de ello, presentó varias limitaciones prácticas: la búsqueda es más lenta, se obtienen resultados duplicados al iterar por fechas cercanas, y requiere un proceso de instalación más complejo, explicado en detalle en el Apéndice. Por estos motivos, se optó por descartar este método en favor del anterior.

Finalmente, se diseñó un sistema de scraping utilizando RSS con una arquitectura flexible basada en múltiples queries (más de 60), que abarcan:

- Dimensiones ESG: sostenibilidad, transición energética, medio ambiente, energía renovable, responsabilidad social, etc
- Eventos negativos: sanciones, denuncias, escándalos
- Fechas históricas: queries forzadas por año

La estrategia fue escogida por su equilibrio entre cobertura temática y eficiencia computacional. La implementación técnica se detalla en el Apéndice, y los resultados obtenidos se describen empíricamente en el capítulo siguiente.

3.3. Natural Language Processing (NLP)

El análisis manual de noticias presenta importantes limitaciones en cuanto a escalabilidad, objetividad y reproducibilidad. Dado el elevado volumen de noticias recopiladas (más de 2,000

artículos en un rango temporal amplio), un enfoque basado en lectura y etiquetado humano habría requerido una inversión de tiempo considerable y habría estado sujeto a sesgos personales e inconsistencias en la clasificación.

Es por esto que utilizaremos técnicas de procesamiento del lenguaje natural (NLP) para automatizar este proceso, asegurando criterios homogéneos de clasificación, menor tiempo de procesamiento y replicabilidad del análisis. En particular, el uso de modelos preentrenados de clasificación temática y de sentimiento posibilita evaluar de forma sistemática el contenido de cada noticia, maximizando así la objetividad del estudio y permitiendo el tratamiento de grandes volúmenes de datos en tiempos razonables.

En el apartado empírico siguiente se describe la implementación concreta de los modelos utilizados para llevar a cabo estas tareas.

3.4. Análisis Descriptivo y Estadístico de Precios

Antes de clasificar, cruzar y medir los cambios en rentabilidad causados por las noticias sobre sostenibilidad, es necesario realizar un análisis descriptivo de la evolución histórica del precio de las acciones y calcular los retornos diarios. Para ello, se utilizarán datos de cotización diaria de Endesa desde enero de 2015 hasta abril de 2025, que incluyen columnas como:

- Fecha (Date)
- Precio de cierre (Close)
- Precio de apertura (Open)
- Precio máximo del día (High)
- Precio mínimo del día (Low)
- Volumen negociado (Vol.)
- Cambio de precio diario (Change)

Sin embargo, para calcular las rentabilidades, prescindiremos del campo “Change” y realizaremos nuestros propios cálculos a partir del precio de cierre.

En el análisis financiero es habitual utilizar dos tipos de rentabilidad: retorno simple y retorno logarítmico.

- a) El retorno simple diario se calcula como el cambio porcentual entre el precio de cierre de un día y el del día anterior:

$$\text{Retorno simple}_t = \frac{P_t - P_{t-1}}{P_{t-1}}$$

donde:

- R_t es el retorno simple en el día t
- P_t es el precio de cierre en el día t
- P_{t-1} es el precio de cierre del día anterior.

- b) El retorno logarítmico transforma el cociente de precios mediante logaritmos naturales:

$$\text{Retorno logarítmico}_t = \ln \left(\frac{P_t}{P_{t-1}} \right)$$

Este tipo de retorno es especialmente útil cuando se trabaja con supuestos estadísticos que requieren estacionariedad o cuando se pretende modelizar series temporales con modelos como ARIMA o GARCH. Además, los retornos logarítmicos son aditivos en el tiempo, lo cual facilita algunos cálculos acumulativos.

Ambos tipos de retorno son muy similares para variaciones pequeñas (como las diarias). Sin embargo, en este trabajo nos centraremos en el uso del retorno simple porque su interpretación es más directa y clara desde el punto de vista económico: refleja el cambio porcentual día a día en el precio de la acción. Los retornos logarítmicos suelen estar mejor distribuidos que los simples y son muy útiles para supuestos en modelos estadísticos. Además, hace que los retornos sean más estacionarios facilitando el uso de modelos ARIMA o GARCH que asumen estacionariedad. No obstante, como nosotros queremos analizar e interpretar el impacto porcentual de las noticias sobre sostenibilidad en el precio de la acción buscamos retornos que sean más explicativos. Es por esto que vamos a hacer uso de los retornos simples.

El cálculo de estos retornos nos permitirá, en el capítulo empírico, analizar la evolución diaria del precio, detectar eventos atípicos y vincular la aparición de noticias ESG con movimientos anómalos del mercado.

3.5. Evaluación de la serie temporal, test de estacionariedad, modelado con ARIMA y ruido blanco

Para medir el impacto de un evento (noticias sostenibles) sobre el precio de cotización, primero necesitamos entender cómo se suele comportar la acción y así compararla con el comportamiento real después del evento. Debemos plantearnos preguntas como ¿cuál es la estructura temporal que tienen los precios y los retornos? ¿sigue una secuencia aleatoria o más bien depende de precios pasados mostrando cierta memoria a corto plazo?

Una serie temporal es una secuencia de observaciones ordenadas en el tiempo (ej. precios diarios de acciones). En nuestro caso, el retorno sobre el precio de cierre de Endesa cada día desde 2015.

Los modelos econométricos y estadísticos como ARIMA suelen asumir:

- Que el comportamiento actual de la serie de precios depende de manera lineal de valores pasados (AR, MA).
- Que la relación entre el pasado y el presente es estable en el tiempo (estacionariedad).

Por tanto, antes de modelar, es necesario estudiar la estacionariedad de los retornos, a través del **Test de Dickey-Fuller (ADF)**. Este test permite comprobar si los retornos tienen propiedades estadísticas estables.

Los modelos posibles que se plantean dentro del análisis de series temporales son:

- AR (p): “En este modelo, el valor actual del proceso se expresa como una combinación lineal finita de los valores pasados del proceso y una perturbación aleatoria a_t ” (Box et al., 2015, p. 8). Captura la dependencia de retornos pasados, es decir, cómo influye el retorno anterior en el lag t-n al retorno en t:

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \phi_2 \tilde{z}_{t-2} + \dots + \phi_p \tilde{z}_{t-p} + a_t$$

- MA (q): Los modelos MA suponen que el valor observado depende linealmente de un número finito de perturbaciones aleatorias anteriores (Box et al., 2015). Captura la dependencia de errores pasados, es decir, cómo influye el término de error en el lag t-n al retorno en t. Esto también se conoce como shocks:

$$\tilde{z}_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}$$

- ARMA (p,q): “Para lograr una mayor flexibilidad en el ajuste de series temporales reales, a veces es ventajoso incluir tanto términos autorregresivos como de media móvil en el modelo. Esto lleva al modelo mixto autorregresivo–media móvil (ARMA)” (Box et al., 2015, p.10). Captura ambas dependencias, tanto cómo influyen los retornos como los errores pasados en el retorno actual:

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \dots + \phi_p \tilde{z}_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}$$

- ARIMA (p,d,q): Captura lo mismo que el modelo ARMA, pero permite que haya tendencias. Los modelos ARIMA (Autorregresivos Integrados de Media Móvil) permiten realizar series temporales que no son estacionarias (no cumplirían el Test ADF) pero que pueden hacerse estacionarias mediante diferenciación. Según Box et al. (2015), este tipo de modelos generaliza los procesos ARMA, pero incorporando un parámetro de diferenciación d , que representa el número de diferencias necesarias para estabilizar la media de la serie.

El modelo ARIMA de orden (p,d,q) está formado por un componente autorregresivo de orden p, una diferenciación de orden d y un componente de media móvil de orden q (Box et al., 2015):

$$w_t = \phi_1 w_{t-1} + \dots + \phi_p w_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}$$

Antes de aplicar un modelo, debemos entender cuál es la estructura interna de la serie temporal (como es su "memoria"). ACF y PACF son herramientas que analizamos a través de gráficos con las que podemos detectar patrones de autocorrelación en la serie:

- ACF: Mide cuántos valores de la variable o lags anteriores explican el valor actual. Esto nos sirve para determinar el orden q en modelos $MA(q)$.
- PACF: Mide cuánto influye un lag específico, pero eliminando los efectos de otros lags anteriores. Te dice cuál es la influencia real y directa de un lag, sin que esté “contaminado” por lags intermedios. El PACF elimina esa dependencia "heredada". Este nos da la pista del orden p en modelos $AR(p)$.

3.6. Estimación de retornos anómalos con CAPM

Hasta este punto, hemos estimado distintos modelos para explicar el comportamiento de los retornos diarios de Endesa en ausencia de eventos externos, como el modelo de ruido blanco, $ARIMA(1,0,1)$ y $ARIMA(0,0,2)$. Estos modelos nos permiten aproximar cómo se comportan los retornos de la acción bajo condiciones normales, es decir, sin que exista información nueva que afecte su cotización.

Bajo este supuesto, cualquier desviación significativa respecto al retorno esperado podría deberse a la incorporación de nueva información al mercado, como por ejemplo una noticia ESG. No obstante, también es razonable pensar que una parte de estas variaciones en los retornos puede estar explicada por movimientos generales del mercado, especialmente si consideramos días en los que hay una tendencia clara en el IBEX 35, índice de referencia en España.

Por este motivo, se incluye un modelo adicional basado en una regresión lineal entre los retornos de Endesa y los retornos del IBEX 35. Este enfoque parte de la idea fundamental del modelo CAPM (Capital Asset Pricing Model), en el que:

$$R_{i,t} = \alpha_i + \beta_i \cdot R_{m,t} + \varepsilon_{i,t}$$

donde:

- $R_{i,t}$ es el retorno de la acción (Endesa) en el día t
- $R_{m,t}$ es el retorno del mercado (IBEX 35) en el mismo día
- α_i representa la rentabilidad no explicada por el mercado
- β_i mide la sensibilidad de la acción respecto al índice
- $\varepsilon_{i,t}$ es el término de error (residual)

Este modelo nos permite estimar retornos esperados ajustados por el mercado y calcular los retornos anómalos (AR) como la diferencia entre el retorno real y el retorno estimado por la regresión:

$$AR_{i,t} = R_{i,t} - (\hat{\alpha}_i + \hat{\beta}_i \cdot R_{m,t})$$

Este enfoque se utiliza ampliamente en la literatura de finanzas por su simplicidad y porque captura de forma directa la parte del comportamiento de una acción que no puede explicarse por el mercado, lo cual es especialmente relevante en estudios de eventos (event studies). De este modo, si una noticia ESG provoca una reacción del precio, pero esta no coincide con una reacción similar en el índice, esa diferencia puede atribuirse al evento en cuestión. Por tanto, este modelo se convierte en una herramienta clave para aislar el impacto específico de las noticias ESG.

Una vez obtenidos los retornos diarios de Endesa y del índice de referencia (IBEX 35), el siguiente paso consiste en estimar el **retorno esperado para cada día**, con el objetivo de aislar los movimientos del precio que no pueden explicarse únicamente por el comportamiento general del mercado.

Para ello, se implementa una **regresión móvil de tipo CAPM**, que permite estimar dinámicamente la sensibilidad de Endesa respecto al mercado en distintos momentos del tiempo. Esta estimación móvil se realiza utilizando una **ventana deslizante de 120 días**: para cada día tt , se toman los 120 días anteriores para estimar los parámetros de la regresión.

El modelo utilizado para cada ventana de 120 días es:

$$R_{i,t} = \alpha_t + \beta_t \cdot R_{m,t} + \varepsilon_t$$

donde:

- $R_{i,t}$ es el retorno de Endesa en el día tt
- $R_{m,t}$ es el retorno del IBEX 35 en el mismo día
- α_t, β_t son los coeficientes estimados mediante regresión lineal móvil

A partir de esta regresión, el **retorno esperado** se calcula como:

$$\widehat{R}_{i,t} = \alpha_t + \beta_t \cdot R_{m,t}$$

Posteriormente, el **retorno anómalo (Abnormal Return, AR)** se calcula como la diferencia entre el retorno real y el retorno esperado:

$$AR_{i,t} = R_{i,t} - \widehat{R}_{i,t}$$

Estos valores permiten detectar desviaciones sistemáticas o significativas que no se explican por el comportamiento general del mercado, y por tanto pueden atribuirse a eventos específicos como las noticias ESG.

3.7. CAR (Retorno anómalo acumulado)

Para analizar el efecto total que puede ejercer una noticia ESG sobre el precio de la acción de Endesa, se utiliza la métrica CAR (Cumulative Abnormal Return). El CAR representa la suma

de los retornos anómalos (AR) durante una ventana temporal en torno a un evento, en nuestro caso, la publicación de una noticia ESG:

$$CAR_{[t_0-w, t_0+w]} = \sum_{t=t_0-w}^{t_0+w} AR_t$$

donde:

- t_0 es el día de la noticia,
- w es el número de días antes/después del evento.

Vamos a calcular esto para tres ventanas: ± 1 , ± 3 y ± 5 días:

- ± 1 día (impacto inmediato)
- ± 3 días (reacción corta)
- ± 5 días (efectos rezagados)

Estas ventanas permiten capturar tanto el shock inmediato como posibles reacciones del mercado que se manifiestan con cierto retraso.

En términos computacionales, para cada observación (noticia ESG clasificada), se recorrió el histórico de retornos anómalos, se identificaron los valores en la ventana correspondiente y se realizó la suma. De este modo, se obtuvieron tres nuevas métricas por evento: CAR_1, CAR_3 y CAR_5.

Cabe destacar que, dado que el cálculo del retorno esperado se basó en una regresión móvil de 120 días, las noticias publicadas durante los primeros tres meses del periodo de análisis no cuentan con estimaciones válidas de AR. Por ello, dichas observaciones fueron excluidas de los análisis estadísticos posteriores.

3.8. Contrastes de significación estadística

Una vez calculados los retornos anormales acumulados (CAR) para cada evento ESG, se procedió a evaluar su significancia estadística mediante tres contrastes clásicos en la literatura de estudios de eventos: el test de Patell, el test de Boehmer, Musumeci y Poulsen (BMP) y el test de signos. Todos los contrastes se aplicaron sobre los valores de CAR utilizando ventanas simétricas de ± 1 , ± 3 y ± 5 días en torno a la fecha de publicación de cada noticia ESG.

- Test de Patell: Consiste en estandarizar cada CAR con su desviación típica durante el periodo de estimación previo y luego calcular un estadístico Z agregado. Este test evalúa si, en conjunto, los retornos anormales acumulados se desvían significativamente de cero, siendo útil para observar efectos agregados en múltiples eventos.
- Test BMP: Aplica un contraste t clásico sobre la media de los CAR frente a cero, bajo los supuestos de normalidad y homocedasticidad. Es ampliamente utilizado por su simplicidad y utilidad para detectar impactos promedio de los eventos.

- Test de signos: De naturaleza no paramétrica, se limita a comparar el número de CAR positivos frente a negativos. Bajo la hipótesis nula, se espera una distribución 50-50. Este test es especialmente útil cuando se sospecha de outliers o asimetrías, ya que no exige normalidad ni varianzas homogéneas.

Estos contrastes permiten valorar el impacto estadístico de las noticias sostenibles sin necesidad de recurrir a modelos más complejos como GARCH, ya que los propios tests incluyen ajustes que mitigan los efectos de heteroscedasticidad.

4. Estudio Empírico

4.1. Datos

En este estudio vamos a utilizar los datos sobre el precio de las acciones de Endesa como conjunto de datos experimental, que contiene información sobre la cotización diaria de las acciones de Endesa a lo largo de una década, desde 2015 hasta 2025. Los principales campos del conjunto de datos incluyen indicadores financieros comunes como el precio de apertura, el precio de cierre, el precio más alto, el precio más bajo y el volumen de negociación, que reflejan plenamente las fluctuaciones del precio de las acciones de Endesa. Nosotros haremos uso principalmente del precio a cierre, “Close” y la fecha, “Date”. Estos datos los hemos obtenido a través de plataformas de datos financieros, Investing en nuestro caso.

Además del uso de precios históricos disponemos de noticias en las que se menciona “Endesa” junto con alguna palabra relacionada con sostenibilidad (como “ESG”, “medioambiente”, “gobernanza”, “social”, etc).

4.2. Scraping de noticias

Para la extracción de las noticias hemos probado con dos metodologías de búsqueda:

- Google RSS: noticias recientes bien ordenadas en el tiempo. Además, es una fuente rápida (el código para extraer las noticias tarda poco tiempo en ejecutarse). Tiene un límite de 100 noticias por query, pero se pueden hacer muchas queries simultáneas con un bucle de búsqueda en el código utilizado (bucle for).
- API Google Search: búsqueda limitada que da acceso a noticias históricas y antiguas de Google. Su cobertura es más amplia, aunque su búsqueda es más lenta en la ejecución. Está limitado a 10 noticias por cada iteración, pero al iterar por día no supone un problema. Como limitaciones en el proceso de búsqueda hemos encontrado que las noticias guardadas para las distintas fechas tienden a duplicarse, guardando información similar, ya sea porque se trata de la misma noticia con distinta fuente o porque se haya guardado una misma noticia para varias fechas. Al filtrar por día, suele extraer las mismas noticias para fechas cercanas (Ej: durante todos los días durante 6 meses extrae las mismas noticias. Esto se soluciona haciendo una eliminación de los links duplicados). Es por este motivo que hemos decidido descartar

esta opción de búsqueda, puesto que las ventajas que nos ofrecía (antigüedad de noticias) no compensaba con el valor final de las noticias conseguidas una vez depuradas. Además, nos gustaría advertir del proceso tan complejo que supone la instalación de la API (explicada en Apéndice 5.1)

Se utilizó la librería `feedparser`, especializada en la lectura de fuentes RSS, para conectarse directamente con el agregador de noticias de Google. La clave del sistema es la construcción dinámica del enlace de búsqueda (url generalizado que se adapta a cada query introducida):

```
"https://news.google.com/rss/search?q={query.replace(' ','+')}&hl=es&gl=ES&ceid=ES:es"
```

La función desarrollada en Python realiza los siguientes pasos:

1. Lanza la consulta RSS con la palabra clave deseada (query).
2. Extrae para cada resultado:
 - a. El título (title)
 - b. Un resumen (summary)
 - c. El enlace original (link)
 - d. La fecha de publicación (published)
3. Devuelve un *DataFrame* con todas las noticias obtenidas para esa búsqueda.

Para superar la limitación de 100 resultados por consulta impuesta por Google, se diseñó una lista de más de 60 queries que cubren tanto:

- Dimensiones ESG: sostenibilidad, transición energética, medio ambiente, energía renovable, responsabilidad social, etc.
- Eventos negativos: protestas, sanciones, denuncias ambientales, vertidos, escándalos.
- Fechas históricas: búsquedas forzadas con años desde 2010 a 2021 para capturar eventos más antiguos.

Una vez recopilados los resultados de todas las búsquedas, se realizó una consolidación de los datos:

- Se eliminaron duplicados utilizando el campo `link` como identificador único.
- Se almacenó el conjunto final en el archivo `df_news.csv` para su posterior clasificación y análisis.

Estas queries se ejecutaron automáticamente en un `bucle for`, generando un conjunto muy amplio y diverso de noticias. Conseguimos extraer con esta herramienta un volumen de 2,306 noticias (véase en Ilustración 1) en las que se menciona la empresa Endesa, así como otras palabras clave vinculadas a sostenibilidad tanto positivas como negativas.

Ilustración 1. Evolución de noticias - Endesa



Como podemos observar en el Ilustración 1 la evolución en la publicación de noticias ha aumentado a lo largo de los años por dos motivos especialmente, (i) la creciente preocupación y referencia a los conceptos de sostenibilidad, especialmente desde 2015 (aparición de ratings ESG en mercados financieros), agenda 2030, etc. (ii) la naturaleza de nuestro medio de descarga de noticias, Google RSS, que conserva y extrae en mayor volumen noticias recientes en comparación a antiguas.

Guardamos el contenido de las noticias en torno a cuatro columnas almacenadas en una nueva base de datos `df_news_clean`, que una vez reorganizada contiene:

- ‘published’: fecha de publicación de la noticia que nos servirá para cruzar por día con los precios y retornos de la acción y así medir el impacto de estas
- ‘title_summary’: título de la noticia publicada en Google que nos sirve para descargarla y clasificarla, unido a un breve resumen de la noticia, más extenso que el título y complementará a este a la hora de clasificar la naturaleza de la noticia
- ‘link’: nos permitirá recuperar la noticia y conocer la fuente, pero servirá como identificador (puesto que los links son únicos y permite que eliminemos los duplicados a través de este)

Este paso es esencial porque permite construir una base de datos de eventos no financieros que potencialmente pueden impactar el precio de la acción, combinando eficiencia en la ejecución, cobertura temática amplia y flexibilidad temporal. La metodología aplicada fue clave para poder identificar estos eventos, base sobre la cual se construirá posteriormente el Event Study y que se usarán más adelante para definir ventanas de análisis y calcular los retornos anómalos.

4.3. Natural Language Processing (NLP)

4.3.1. Clasificador NLP por categoría ESG

Una vez recopiladas y limpiadas las noticias relevantes sobre Endesa, el siguiente paso consistió en clasificarlas según su contenido temático: si hacían referencia a aspectos medioambientales, sociales o de gobernanza (ESG) o si de lo contrario, su contenido no estaba relacionado a ESG (No ESG). Para ello se utilizó un modelo de procesamiento del lenguaje natural (NLP) basado en inteligencia artificial.

Se empleó un modelo preentrenado de tipo [xlm-roberta-large-xnli](#), que permite hacer clasificación "zero-shot", es decir, sin necesidad de volver a entrenar el modelo con nuevos datos. Esta capacidad resulta especialmente útil para tareas donde no se dispone de un conjunto de entrenamiento etiquetado previamente. El modelo fue implementado a través de la librería transformers de [Hugging Face](#), utilizando el [pipeline zero-shot-classification](#). Además, se procesaron los titulares y resúmenes por lotes para mejorar la eficiencia ([batch_size=8](#)).

Se definieron las siguientes categorías como etiquetas de clasificación:

(i) Medioambiente, (ii) Social, (iii) Gobernanza, (iv) Finanzas (v) Energía (vi) Política

A cada noticia (concretamente al campo [title_summary](#)) se le aplicó esta clasificación, y el modelo asignó la categoría más probable y una puntuación entre 0 y 1. Para aumentar la fiabilidad del resultado, se aplicó un umbral de confianza del 0.55. Si ninguna etiqueta superaba este valor, se asignaba la categoría "No ESG".

Finalmente, se exportó el conjunto completo a un nuevo archivo CSV ([noticias_endesa_esg_roberta.csv](#)) sin filtrar ninguna noticia. Este paso permitió etiquetar cada noticia con una categoría temática ESG de forma automática y escalable. Esto facilitó después el filtrado de eventos relevantes para el estudio del impacto sobre la cotización de Endesa.

4.3.2. Clasificador NLP por tono

Después de clasificar cada noticia según su temática ESG, se aplicó un análisis de sentimiento o tono con el objetivo de identificar si la noticia era percibida como positiva, negativa o neutra. Esta dimensión es clave para valorar no solo el contenido, sino también la posible dirección del impacto sobre el mercado. Se utilizó el modelo preentrenado [nlptown/bert-base-multilingual-uncased-sentiment](#), que está diseñado para análisis de sentimiento en múltiples idiomas, incluido el español. Este modelo devuelve una valoración que va de 1 a 5 estrellas, donde:

- 1–2: Sentimiento negativo
- 3: Sentimiento neutro
- 4–5: Sentimiento positivo

Este modelo fue aplicado mediante el pipeline [sentiment-analysis](#) de la librería [transformers](#) de [Hugging Face](#). En casos donde el modelo no pudo analizar el texto (por ejemplo,

si estaba vacío o mal formateado), se asignó el valor por defecto “Neutro”. Finalmente, la clasificación de tono se añadió como una nueva columna llamada tono y se guardó el conjunto final en un nuevo archivo: noticias_endesa_esg_completo.csv. Este paso añadió una capa más de información útil a cada evento ESG, permitiendo más adelante distinguir si las noticias que afectan el precio de la acción son de carácter positivo, negativo o neutral. Esta diferenciación resulta muy valiosa en la interpretación de los retornos anómalos.

4.4. Análisis descriptivo y estadístico de precios

Como ya hemos mencionado, descargamos la serie histórica de precios de Endesa de la fuente Investing. Antes de clasificar, cruzar y medir los cambios en rentabilidad causados por las noticias sobre sostenibilidad, vamos a hacer un resumen y análisis estadístico de los precios. Disponemos de precios diarios desde enero 2015 hasta fecha de hoy (abril 2025), lo que representan 2,631 observaciones (véase en Ilustración 2)

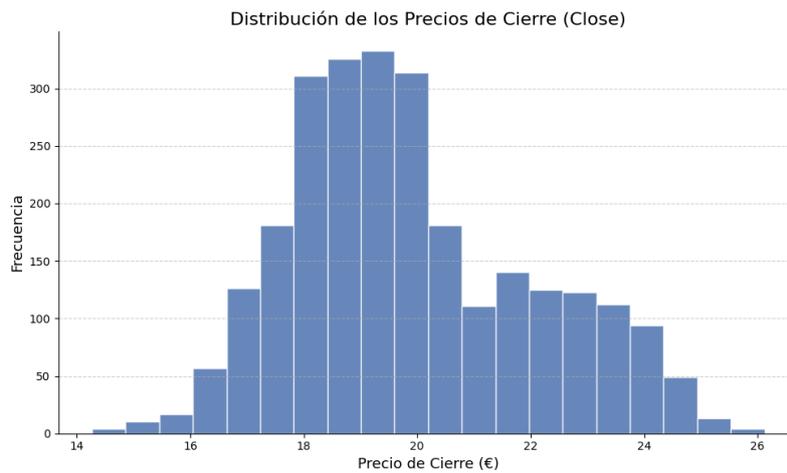
Ilustración 2. Tabla resumen precios de cotización Endesa y estadísticas

	Date	Price	Open	High	Low	Vol.	Change %	Close
0	04/11/2025	24.23	23.92	24.25	23.79	3.72K	1.89%	count 2631.00
								mean 19.89
1	04/10/2025	23.78	24.77	24.77	23.53	1.73M	2.99%	std 2.14
2	04/09/2025	23.09	23.19	23.38	22.89	1.79M	-2.08%	min 14.27
								25% 18.32
3	04/08/2025	23.58	23.10	23.78	22.90	1.90M	2.52%	50% 19.52
								75% 21.39
4	04/07/2025	23.00	21.70	23.70	21.00	0.47K	-5.93%	max 26.12

Vemos en Ilustración 2 que la descarga tiene como columnas la fecha “Date”, precio de cierre “Price” el cual cambiamos a partir de ahora como “Close”, el precio de apertura “Open”, el precio máximo que alcanza en el día “High” y el precio mínimo del día “Low”. Otras columnas relevantes que pueden ser útiles en nuestro análisis son “Vol.” que representa el volumen de negociación contando por transacción realizada y “Change” que es la rentabilidad o variación entre precios (no obstante, calcularemos nosotros mismos la rentabilidad sobre los precios de cierre “Close”). Además, tratamos los datos ordenándolos de mayor antigüedad a menor para posteriormente calcular los retornos.

El precio medio a lo largo de estos diez años ha sido de 19,89€/acción (véase en Ilustración 2), con una desviación estándar de 2,14€. Pasando a la distribución de los precios, estos siguen una distribución normal con media entorno a casi los 20€ con colas anchas a la derecha de la distribución (Ilustración 3)

Ilustración 3. Distribución de precios de Endesa



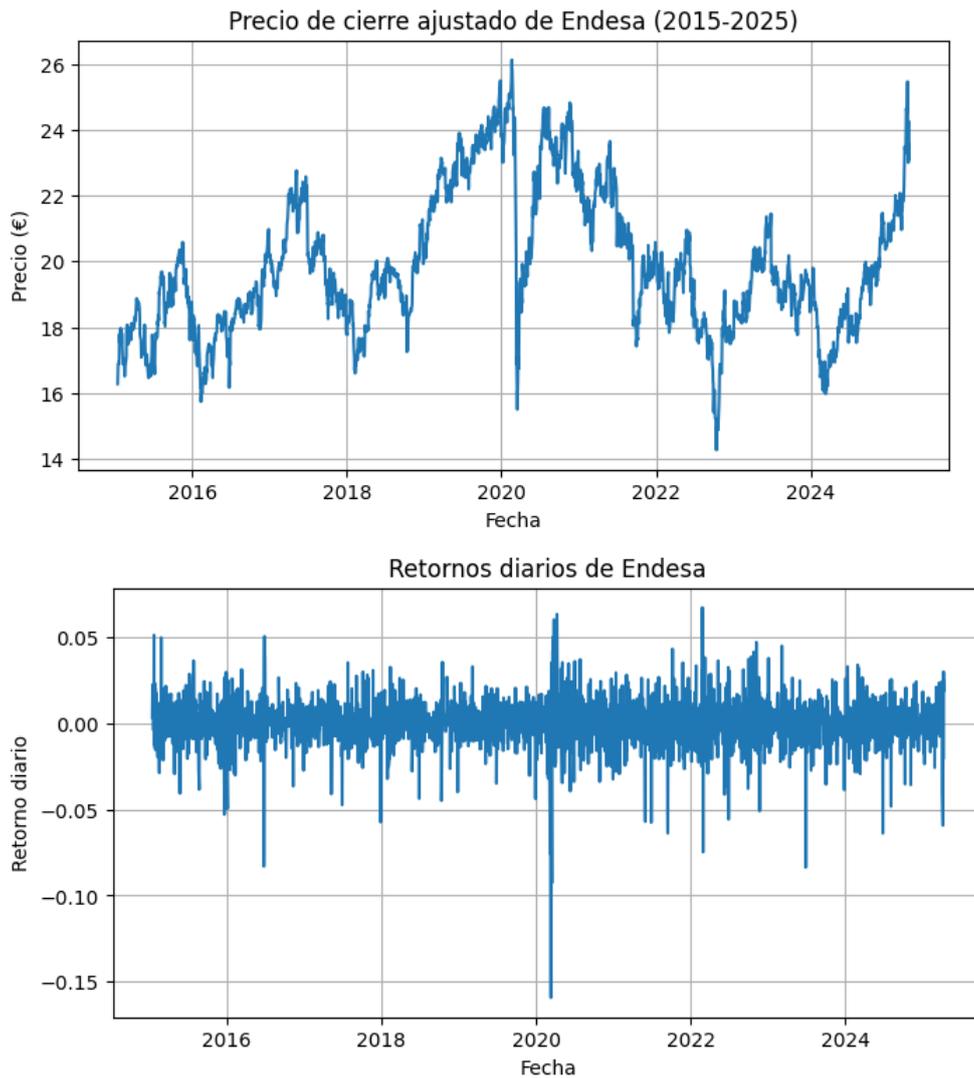
A continuación, calculamos los retornos, tanto simples como logarítmicos:

Ilustración 4. Cálculo de retornos simples y logarítmicos

	Close	Return	LogReturn
2622	16.275	0.00339	0.00339
2621	16.500	0.01382	0.01373
2620	16.870	0.02242	0.02218
2619	16.950	0.00474	0.00473
2618	16.880	-0.00413	-0.00414

En el Ilustración 4 podemos observar cómo estos apenas presentan diferencias hasta el cuarto decimal.

Ilustración 5. Evolución de los precios y retornos diarios de Endesa (2015-2025)



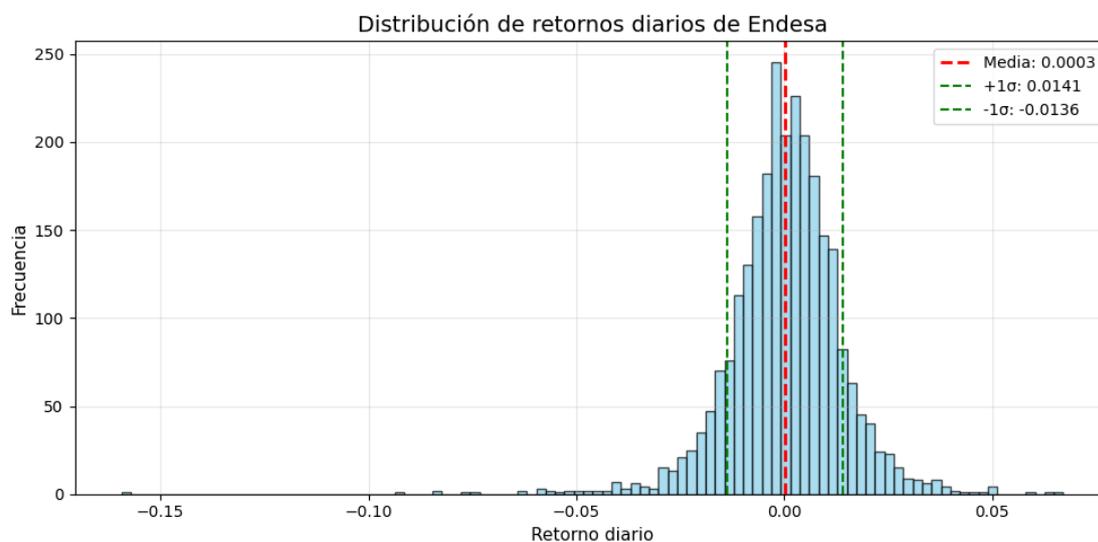
A lo largo de estos años la acción de Endesa ha ido variando en torno a unos precios dentro del rango de 14€ y 26€ por acción (Ilustración 5). Vemos en su evolución en el precio de cierre una tendencia ascendente desde el año de inicio de nuestro análisis, 2015 hasta principios del año 2020, donde alcanzaba máximos históricos. Después de la caída abrupta producida a causa del covid-19, la acción remontaba precios históricos y desde entonces ha ido cotizando de manera muy volátil hasta fecha de hoy. Fue en 2022 donde registraron su mínima cotización bursátil. Durante este año, las acciones de Endesa experimentaron una notable caída, causada por la unión de varios factores. A raíz del estallido de la guerra de Ucrania y Rusia, se desencadenó en Europa la crisis energética. Esto generó gran incertidumbre en el mercado energético europeo. A esto se sumó la implementación de medidas regulatorias en el negocio de distribución y la reducción de márgenes en el sector de energías renovables. Esto resultó en un efecto negativo sobre los resultados financieros de la compañía. A pesar de estas adversidades, Endesa logró un beneficio ordinario neto de 2,398 millones de euros y alcanzó una inversión récord de 2,343 millones de

euros. Esto nos lleva a pensar que lo que determinó en gran medida el bajo precio de cotización no fueron los resultados financieros y la buena gestión de la empresa sino las condiciones macroeconómicas y de entorno incierto (Endesa, 2023).

“Las acciones del Sector Eléctrico se vieron afectadas por la incertidumbre de los inversores ante la situación de crisis económica y social sin precedentes provocada por el conflicto bélico en Ucrania, que se reflejó en una fuerte volatilidad de los precios energéticos y una creciente inflación. Esta situación desencadenó a lo largo del año la adopción de importantes medidas regulatorias con impacto en el sector, aprobadas tanto a nivel europeo como por parte del Gobierno español, para tratar de mitigar la presión de los precios sobre la población. La elevada volatilidad del mercado en 2022, caracterizada por movimientos muy extremos en la cotización, se produjo con un volumen de contratación acumulado inferior al del año anterior” (Endesa, 2023, p. 30).

Pasando ahora a la distribución de los retornos simples, vemos como estos siguen una normal con media entorno al 0 y con desviación típica de +/-1.4% (véase en Ilustración 6)

Ilustración 6. Distribución de retornos simples de Endesa



En la distribución vemos como presenta cierta asimetría a la izquierda registrando valores más bajos en los retornos negativos que valores altos en los retornos positivos, en línea con los descrito por autores como Dorfleitner y Zhang (2024) que explicaban una reacción más extrema ante eventos negativos frente a los positivos. No obstante, podemos ver que se trata de casos aislados, siendo el extremo de la izquierda posiblemente causado por la crisis del covid o la crisis energética con el estallido de la guerra.

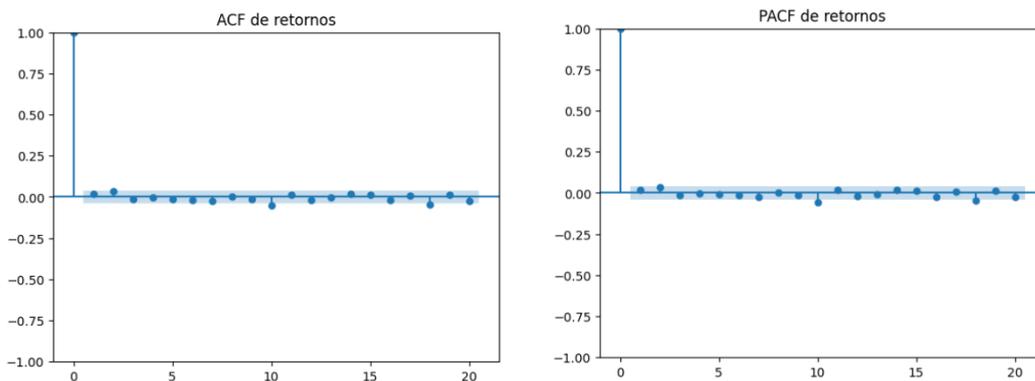
4.5. Evaluación de la serie temporal y test de estacionariedad

Siguiendo con nuestro análisis, recordemos que buscamos medir el impacto de un evento (noticias sostenibles) sobre el precio de cotización.

Antes de modelar los retornos, el siguiente paso que vamos a realizar es el estudio de la estacionariedad de los retornos a través del Test de Dickey-Fuller (ADF). Aplicamos este test sobre los retornos puesto que estos nos van a permitir medir los cambios en el valor de nuestro activo que es la acción de Endesa. Además, los precios no suelen ser estacionarios, puesto que suelen presentar tendencias (alcistas en nuestro caso) en el tiempo. Este test es necesario para comprobar si los retornos son estacionarios, lo que implica que sus propiedades estadísticas son estables. Es por esto que para realizar los modelos requieren que la serie sí sea estacionaria. Después de hacer el test con la librería statsmodels, obtenemos un resultado del estadístico ADF de -34.52 , siendo un valor muy negativo y evidenciando la posibilidad de rechazar nuestra hipótesis nula. El p-value obtenido es de 0.000 , estando muy por debajo del 0.05 a un nivel de significación del 5% . Podemos rechazar con esto la hipótesis nula con alta confianza (siendo H_0 que la serie no es estacionaria y H_a que la serie es estacionaria).

Una vez hemos observado que es estacionaria la serie de retornos, vamos a ver cómo podemos modelar la serie. Hemos de mencionar que dentro de las series temporales podemos plantear diferentes modelos que capturarán diferentes factores. Para identificar cuáles utilizaremos evaluaremos su autocorrelación y la de tipo parcial (ACF y PACF), (véase en Ilustración 7).

Ilustración 7. Autocorrelación y autocorrelación parcial de los retornos simples



A priori parece que los retornos sobre los precios de cotización de Endesa presentan lo conocido como ruido blanco o *White Noise*. En el análisis de series temporales, una serie de retornos se considera ruido blanco cuando sus valores son independientes entre sí y tienen media cero y varianza constante. En otras palabras, no presentan una estructura predecible o patrón temporal discernible. Que los retornos de Endesa se comporten como ruido blanco implica que, en ausencia de eventos externos, el precio sigue un comportamiento aleatorio y no autocorrelacionado. Esto sugiere que los shocks o variaciones significativas pueden atribuirse a

eventos externos, como la publicación de noticias ESG, dado que el comportamiento base no es sistemáticamente dependiente del pasado.

Fama introdujo con su estudio empírico publicado en 1970, la Hipótesis de los Mercados Eficientes (EMH), que establece que los precios reflejan toda la información disponible, por lo que los retornos siguen un comportamiento aleatorio, es decir, se comportan como ruido blanco si no hay nueva información relevante. Si partiésemos de esta hipótesis estaríamos asumiendo que los inversores en la bolsa española, más concretamente en el IBEX 35, con la acción de Endesa (ELE), son capaces de reflejar toda la información disponible a través de la compra y venta de la acción, impacto que se vería reflejado en el precio de cotización. Bajo esta premisa, los cambios significativos en los precios se explicarían únicamente por la incorporación de nueva información externa, como, por ejemplo, eventos macroeconómicos o factores más concretos como noticias relacionadas con factores ESG. Durante este estudio veremos si estas noticias ESG, según su categoría y tono, realmente siguen un patrón que marque el rumbo del precio de cotización y el mercado sea efectivamente eficiente.

4.6. Modelado con ARIMA y ruido blanco

Para analizar la estructura de los retornos diarios de Endesa (ELE), hemos estimado tres modelos: un modelo de ruido blanco (constante), un modelo ARIMA (1,0,1) y un modelo ARIMA (0,0,2). Este último fue el recomendado por un optimizador desarrollado en python de AIC/BIC que te sugiere aquella combinación de parámetros óptima (véase en Ilustración 8).

Ilustración 8. Tabla de mejores modelos según optimizador de parámetros

Mejores modelos ordenados por AIC:			
	order	aic	bic
2	(0, 0, 2)	-15007.669421	-14984.179600
16	(2, 0, 0)	-15007.535560	-14984.045739
0	(0, 0, 0)	-15006.985365	-14995.240455
24	(3, 0, 0)	-15006.224812	-14976.862536
3	(0, 0, 3)	-15006.204678	-14976.842402

La finalidad es evaluar si existe alguna dependencia temporal en la serie que justifique una modelización más compleja que el supuesto aleatorio básico.

A continuación, se presentan las métricas comparativas más relevantes:

1. Resultado ARIMA (1,0,1) (Ilustración 9):

Ilustración 9. Resultados modelo ARIMA (1,0,1)

```

=== ARIMA(1,0,1) ===
                                SARIMAX Results
=====
Dep. Variable:                    y      No. Observations:      2624
Model:                            ARIMA(1, 0, 1)  Log Likelihood          8396.233
Date:                            Tue, 27 May 2025  AIC                  -16784.467
Time:                            19:32:29      BIC                  -16760.977
Sample:                            0          HQIC                 -16775.960
                                - 2624
Covariance Type:                  opg
=====
              coef    std err          z      P>|z|      [0.025    0.975]
-----
const          0.0003     0.000     1.775     0.076    -3.53e-05     0.001
ar.L1         -0.0157     1.772    -0.009     0.993    -3.490     3.458
ma.L1          0.0045     1.774     0.003     0.998    -3.472     3.481
sigma2         9.733e-05    2.66e-06    36.649     0.000     9.21e-05     0.000
=====
Ljung-Box (L1) (Q):                0.00  Jarque-Bera (JB):                0.92
Prob(Q):                            0.98  Prob(JB):                        0.63
Heteroskedasticity (H):              1.01  Skew:                            0.04
Prob(H) (two-sided):                 0.87  Kurtosis:                        3.05
=====

```

2. Resultado ARIMA (0,0,2) (Ilustración 10):

Ilustración 10. Resultados modelo ARIMA (0,0,2)

```

=== ARIMA(0,0,2) ===
                                SARIMAX Results
=====
Dep. Variable:                    y      No. Observations:      2624
Model:                            ARIMA(0, 0, 2)  Log Likelihood          8397.345
Date:                            Tue, 27 May 2025  AIC                  -16786.691
Time:                            19:32:30      BIC                  -16763.201
Sample:                            0          HQIC                 -16778.184
                                - 2624
Covariance Type:                  opg
=====
              coef    std err          z      P>|z|      [0.025    0.975]
-----
const          0.0003     0.000     1.824     0.068    -2.52e-05     0.001
ma.L1         -0.0103     0.020    -0.519     0.603    -0.049     0.028
ma.L2         -0.0290     0.020    -1.464     0.143    -0.068     0.010
sigma2         9.724e-05    2.65e-06    36.682     0.000     9.2e-05     0.000
=====
Ljung-Box (L1) (Q):                0.00  Jarque-Bera (JB):                0.90
Prob(Q):                            0.96  Prob(JB):                        0.64
Heteroskedasticity (H):              1.01  Skew:                            0.04
Prob(H) (two-sided):                 0.91  Kurtosis:                        3.05
=====

```

3. Resultado White Noise (Ilustración 11):

Ilustración 11. Resultados modelo White Noise

```

=== Modelo de Ruido Blanco ===
Media estimada:                    0.000339
Varianza estimada:                 0.000097
Log-Likelihood:                    8396.072
AIC:                               -16790.143
BIC:                               -16784.271

```

A diferencia de los otros dos modelos, el ruido blanco no se estima con un proceso estadístico que tenga parámetros dinámicos como AR (autoregresivos) o MA (medias móviles). Simplemente se calcula la media y la varianza de los datos, sin estructura temporal que reportar. A continuación, se presentan las métricas AIC/BIC y Log-Likelihood de los tres modelos (véase Ilustración 12)

Ilustración 12. Métricas modelos series temporales

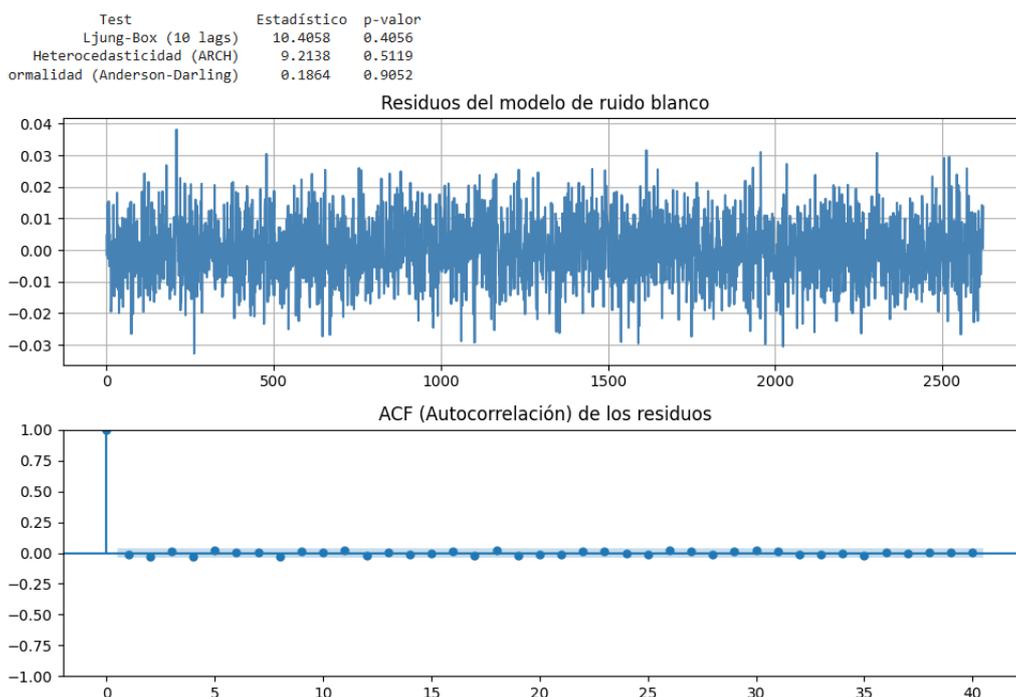
Modelo	AIC	BIC	Log-Likelihood
Ruido Blanco	-16790.143	-16784.271	8396.072
ARIMA(1,0,1)	-16784.467	-16760.977	8396.233
ARIMA(0,0,2)	-16786.691	-16763.201	8397.345

Algunas conclusiones son las siguientes (véanse en Ilustración 12):

- El modelo de ruido blanco ya presenta un ajuste bastante bueno, con un AIC inferior al del modelo ARIMA (1,0,1). Esto sugiere que la serie no contiene una estructura autoregresiva fuerte ni efectos de medias móviles que justifiquen un modelo más complejo.
- El modelo ARIMA (1,0,1) no mejora significativamente el log-likelihood (8396.23 frente a 8396.07 del ruido blanco), y sus coeficientes AR y MA son estadísticamente no significativos (p-valores de 0.993 y 0.998 respectivamente), lo cual refuerza la hipótesis de que los retornos se comportan como ruido blanco.
- El modelo ARIMA(0,0,2) ofrece la mejor puntuación de AIC y log-likelihood, aunque de manera marginal. Sin embargo, sus coeficientes tampoco son significativos al 5%, lo que implica que la mejora es más numérica que sustantiva.

En los tres modelos analizados, los resultados del test de Ljung-Box muestran que no hay autocorrelación significativa en los residuos, es decir, los errores del modelo no siguen un patrón y parecen comportarse de forma aleatoria. Por otro lado, el test de heterocedasticidad indica que la variabilidad de los errores se mantiene constante a lo largo del tiempo, lo que sugiere que no es necesario utilizar modelos más complejos como GARCH para capturar cambios en la volatilidad. Respecto a la distribución de los residuos, el test de Jarque-Bera no detecta desviaciones graves respecto a una distribución normal. Sin embargo, algunos valores apuntan a una ligera asimetría y a colas algo más pesadas de lo habitual (distribución que ya habíamos observado con representaciones gráficas anteriores), lo que es común en series financieras.

Ilustración 13. Evolución y autocorrelación de residuos



4.7. Estimación de retornos anómalos con CAPM

Una vez explicada la teoría del proceso en el apartado de metodología, explicamos los pasos realizados en nuestro entorno de python. [Hugging Face](#)

Recordemos que ya teníamos calculados los retornos sobre los precios de cotización de Endesa. No obstante, debemos hacer lo mismo con los del índice de referencia, el IBEX 35. Para ello cargamos la serie de precios descargada también de la plataforma Investing y cargados en formato `.csv`. Con estos precios calculamos los retornos simples sobre el índice en la misma ventana temporal que los de Endesa (desde principios de enero de 2015 hasta abril de 2025). Una vez cargados los datos, se ordenaron cronológicamente y se calcularon los retornos diarios simples con la función `pct_change()` de `pandas`, que aplica la fórmula estándar.

Este paso generó dos series de tiempo: una con los retornos diarios de Endesa y otra con los del IBEX 35, ambas sincronizadas por fecha.

Para capturar esta relación dinámica entre los retornos de Endesa y del IBEX 35, se implementó una **regresión lineal móvil** con una ventana deslizante de 120 días. Para cada día t , se extrajo la información de los 120 días previos y se aplicó una regresión lineal mediante la función `LinearRegression()` de `sklearn`, con el modelo CAPM descrito en la sección metodológica.

Este proceso se automatizó mediante un `bucle for`, en el que se guardaron en cada iteración:

- los coeficientes estimados α_t y β_t

- el retorno esperado $R_{i,t}$
- y finalmente el retorno anómalo $AR_{i,t}$, que se calculó como:

$$AR_{i,t} = R_{i,t} - (\alpha_t + \beta_t \cdot R_{m,t})$$

Este valor fue añadido como una nueva columna en el *DataFrame* principal, permitiendo analizar día a día cuándo se producen desviaciones relevantes respecto al comportamiento que cabría esperar según el mercado.

Para interpretar mejor el comportamiento de los retornos anómalos, se elaboraron varios gráficos utilizando [matplotlib](#) y [seaborn](#), representando tanto por la evolución temporal de los $AR_{i,t}$ como su distribución estadística.

Ilustración 14. Distribución de los retornos anómalos diarios mediante modelo CAPM

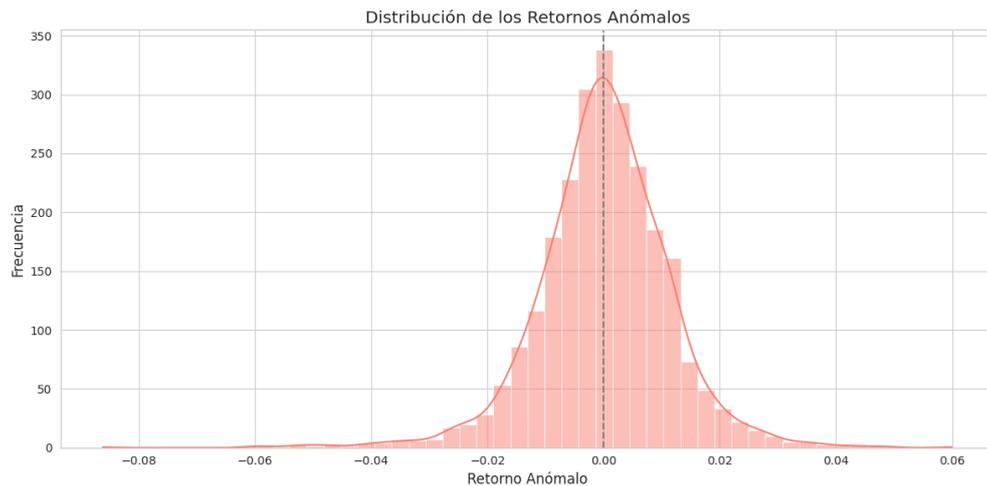
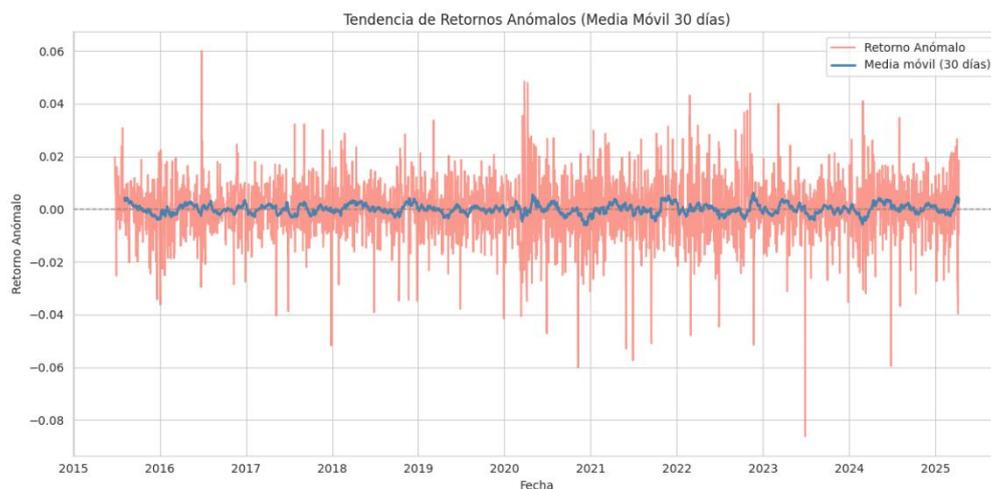


Ilustración 15. Evolución temporal de los retornos anómalos con media móvil de 30 días



En la distribución de los retornos anómalos obtenida tras aplicar el modelo CAPM con ventana móvil de 120 días (véase Ilustración 14), observamos que los datos siguen una forma aproximadamente normal, centrada en torno al cero. Esta simetría refuerza la hipótesis de que, en ausencia de eventos informativos, los retornos anómalos se distribuyen de manera aleatoria, sin tendencia sistemática, tal como predice la Hipótesis de los Mercados Eficientes en su versión débil. Sin embargo, también se detecta una cola más pronunciada hacia la izquierda, lo que sugiere que, aunque infrecuentes, los retornos negativos extremos han sido más intensos que los positivos. Este hallazgo es coherente con lo documentado por Krüger (2015), quien señala que los inversores tienden a reaccionar más intensamente ante noticias ESG negativas que ante noticias positivas o neutras.

Por otro lado, la evolución temporal de los retornos anómalos, acompañada por su media móvil de 30 días (véase Ilustración 15), revela que, aunque los valores individuales presentan alta volatilidad diaria, la tendencia general se mantiene próxima al cero a lo largo de la serie histórica. Esto sugiere que, si bien hay reacciones puntuales significativas atribuibles a eventos, no hay un sesgo estructural persistente ni una acumulación sistemática de retornos anómalos. La ausencia de tendencias prolongadas también apoya la idea de que el mercado puede estar incorporando la información ESG en el precio, sin permitir arbitrajes prolongados, lo cual también es compatible con una hipótesis de mercado relativamente eficiente.

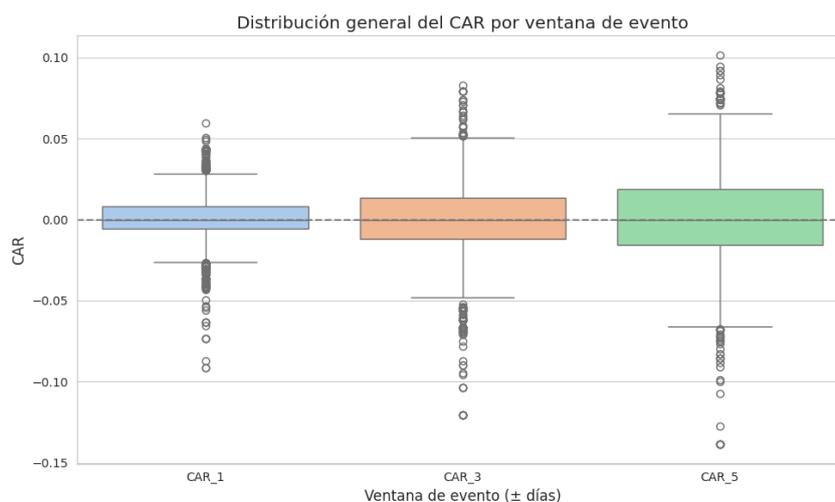
Finalmente, los picos negativos aislados y recurrentes vuelven a destacar la importancia de analizar el tono y la categoría ESG de cada evento informativo, pues el contenido de las noticias puede ser clave en la magnitud del impacto observado en el precio. Estas diferencias en tono y categoría de noticias las veremos con mayor profundidad en los siguientes apartados.

4.8. CAR (Retorno anómalo acumulado)

Una vez calculados los retornos anómalos diarios (AR), se procedió a estimar el impacto acumulado de las noticias ESG sobre la cotización de Endesa mediante el CAR en las tres ventanas temporales (± 1 , ± 3 y ± 5 días).

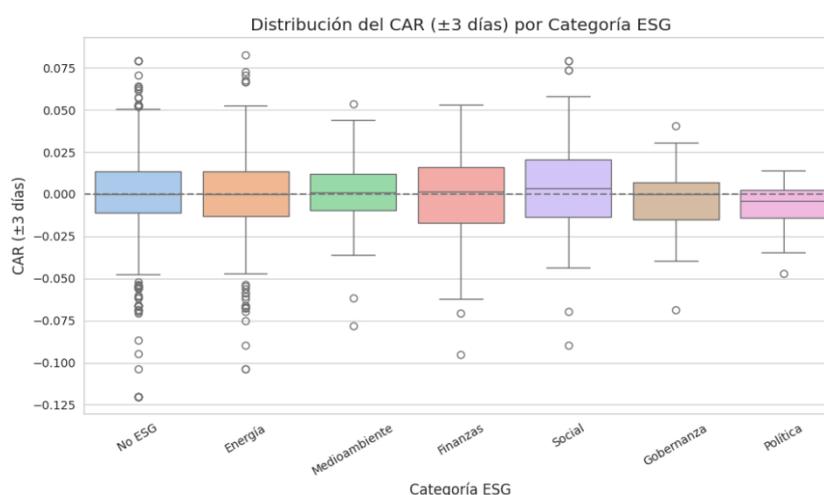
En el gráfico general (véase Ilustración 16) se visualiza la distribución del CAR para las tres ventanas. Se observa que, en todos los casos, la mayoría de los valores se agrupan en torno a cero, lo que indica que muchas noticias no generan una reacción significativa en el mercado. Sin embargo, también se aprecia una dispersión mayor en las ventanas más amplias, especialmente en ± 5 días, lo que sugiere que algunos eventos generan efectos rezagados o acumulativos.

Ilustración 16. Distribución general del CAR por ventana de evento (± 1 , ± 3 , ± 5 días)



A continuación, se exploró cómo varía la reacción del mercado en función del **tipo de noticia ESG**. En el Ilustración 17 se representa la distribución del CAR (± 3 días) por categoría ESG (No ESG, Energía, Medioambiente, Finanzas, Social, Gobernanza y Política).

Ilustración 17. Distribución del CAR (± 3 días) por categoría ESG

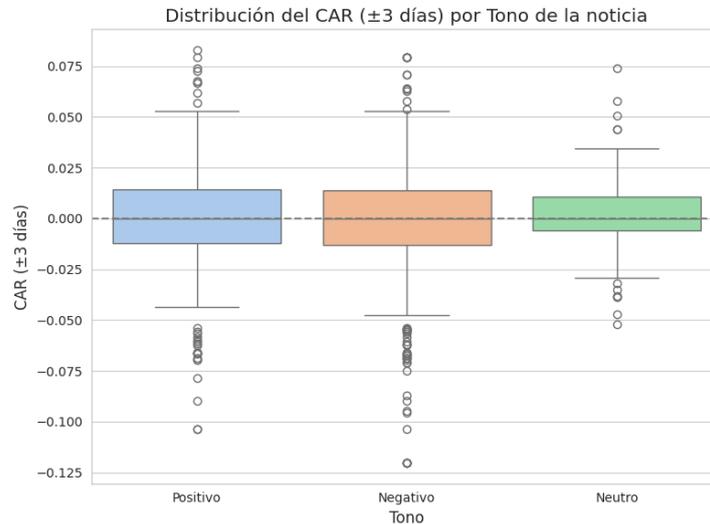


En el gráfico (véase Ilustración 17) se analiza la distribución del CAR (± 3 días) según la categoría temática ESG de cada noticia. Aunque la mediana se mantiene próxima a cero en todas ellas, lo que sugiere una ausencia de efecto sistemático, se observan diferencias relevantes en la dispersión y en la aparición de valores extremos. Las categorías de Gobernanza y Política concentran más retornos negativos, lo que podría indicar que ciertos temas regulatorios o institucionales generan reacciones adversas en el mercado. Por su parte, las noticias clasificadas como “No ESG” muestran una elevada frecuencia de outliers negativos, lo que sugiere que no todos los eventos relevantes para el mercado están ligados a sostenibilidad. En contraste, las categorías Social y Finanzas exhiben una mayor variabilidad en los valores positivos, aunque sin un desplazamiento claro en su centro. Las categorías de Energía y Medioambiente presentan una distribución más contenida y simétrica, lo que podría reflejar que, aunque estos temas están muy presentes en el discurso sostenible, su impacto directo sobre la acción no siempre es inmediato o extremo. En conjunto, estos resultados reflejan que ciertos tipos de contenido ESG sí despiertan

respuestas diferenciadas en el precio de la acción, y que el enfoque metodológico empleado permite identificar estos matices de forma estructurada y replicable.

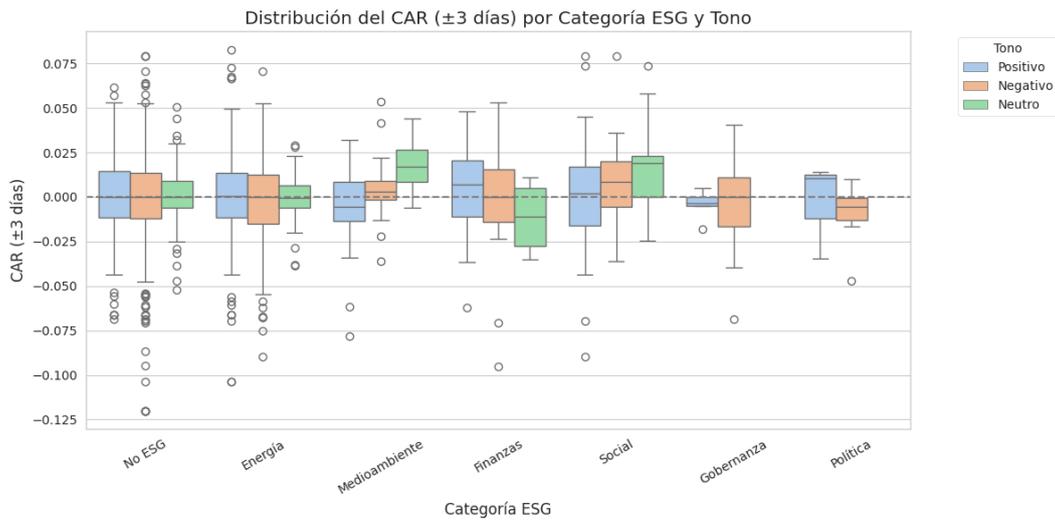
Del mismo modo, se evaluaron las diferencias en el CAR (± 3 días) por tono.

Ilustración 18. Distribución del CAR (± 3 días) según el tono de la noticia



En el gráfico (véase Ilustración 18) se analiza cómo varía el CAR (± 3 días) en función del tono de la noticia ESG (positivo, negativo o neutro). Las tres distribuciones presentan medianas cercanas a cero, lo que indica que en términos generales no se detecta un sesgo sistemático claro en la reacción del mercado. No obstante, se aprecian diferencias significativas en la dispersión y en la frecuencia de valores extremos. Las noticias de tono negativo concentran una mayor cantidad de outliers hacia la izquierda, lo que sugiere que este tipo de eventos tienen más probabilidades de generar caídas notables en el precio. En cambio, el tono positivo, si bien no produce una media desplazada hacia arriba, presenta más valores extremos en el lado superior, lo que indica que en algunos casos concretos puede haber reacciones alcistas destacables. Por su parte, las noticias de tono neutro muestran una distribución más contenida y simétrica. Estos resultados son coherentes con lo señalado en la literatura (por ejemplo, Krüger, 2015), donde se destaca que las malas noticias ESG suelen ser más penalizadas por el mercado que las positivas son recompensadas. En conjunto, el análisis confirma que el tono informativo es una dimensión relevante a considerar, y que la metodología empleada permite capturar estos efectos de forma automatizada y escalable.

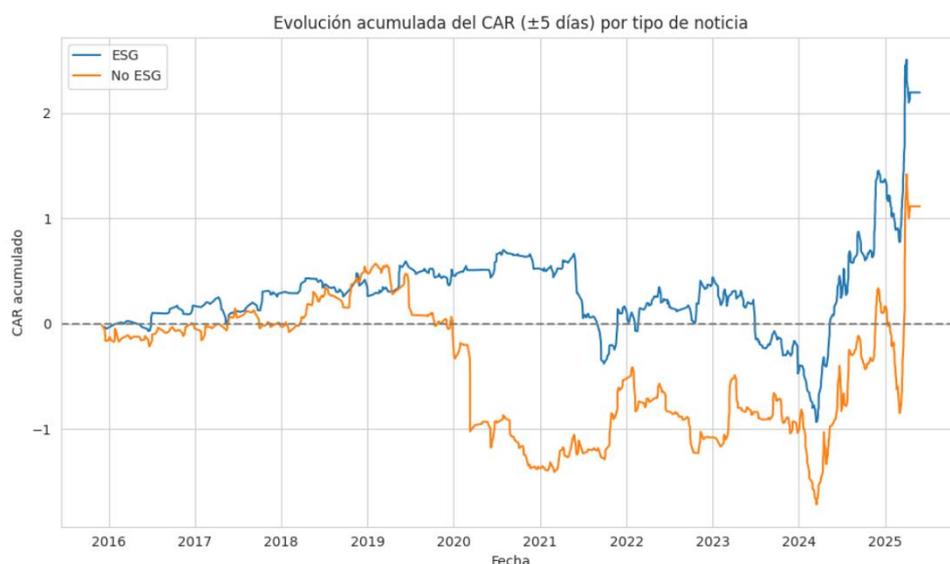
Ilustración 19. Interacción entre categoría ESG y tono de la noticia en el CAR (± 3 días)



En el gráfico (véase Ilustración 19) se representa la distribución del CAR (± 3 días) combinando categoría temática ESG y tono de la noticia. Aunque la mayoría de las medianas se mantienen cercanas a cero, se observan patrones relevantes cuando se cruzan ambas dimensiones. En particular, el tono negativo genera una mayor dispersión hacia retornos negativos en varias categorías, especialmente en Gobernanza y Política, lo que sugiere que este tipo de contenidos puede estar asociado a caídas más acusadas en el precio. En contraste, los tonos positivos en categorías como Energía, Finanzas y Social presentan una mayor amplitud hacia valores positivos, reflejando que algunas noticias bien recibidas por el mercado pueden traducirse en reacciones alcistas concretas. También es destacable que las noticias No ESG, encontramos outliers y valores extremos para las noticias de tono negativo.

Por último y como cierre a este análisis del CAR, hemos querido medir la evolución acumulada del impacto de las noticias ESG frente a las No ESG a lo largo del tiempo.

Ilustración 20. Evolución acumulada CAR (± 5 días) ESG vs No ESG



En el gráfico (véase Ilustración 20) se presenta la evolución acumulada del CAR (± 5 días) según el tipo de noticia (ESG o No ESG). La curva asociada a las noticias ESG muestra una trayectoria predominantemente positiva, con una tendencia ascendente sostenida desde 2016, que se acentúa especialmente a partir de 2023. Esto sugiere que, en conjunto, las noticias ESG han estado asociadas a reacciones de mercado generalmente favorables o neutras, contribuyendo de forma acumulada a una apreciación del valor de la acción en los días posteriores a su publicación. Por el contrario, las noticias No ESG exhiben un comportamiento más volátil y, en gran parte del periodo, especialmente entre 2020 y 2024, acumulan un impacto negativo, con caídas sostenidas en el CAR total. Este contraste sugiere que la información ESG no solo es relevante para los inversores, sino que tiende a generar una narrativa más estable y potencialmente positiva en los mercados, mientras que otro tipo de noticias tienden a producir reacciones más negativas o impredecibles. Esta última visualización refuerza la idea de que la sostenibilidad no solo es un criterio ético o reputacional, sino que también puede tener un componente financiero tangible en la evolución del valor bursátil.

4.9. Contrastes de significación estadística

Una vez calculados los retornos anormales acumulados (CAR) para cada evento ESG, se procedió a evaluar si estos resultados eran estadísticamente significativos. Para ello, se aplicaron tres contrastes comúnmente utilizados en la literatura de estudios de eventos: el test de Patell, el test BMP (Boehmer, Musumeci y Poulsen) y el test de signos. Los contrastes se aplicaron sobre los valores de CAR utilizando una ventana de ± 3 días en torno a la fecha de publicación de cada noticia ESG (es decir, desde $t-3$ hasta $t+3$).

Los resultados se resumen en la siguiente tabla (véase Ilustración 21):

Ilustración 21. Tabla de significación - Tres contrastes

Ventana	Test	Estadístico	p-value	Resultado
±1	Patell	1.6017	NaN	No significativo
±1	BMP (t-student)	1.6017	0.1094	No significativo
±1	Signos	NaN	0.0083	Significativo
±3	Patell	0.3649	NaN	No significativo
±3	BMP (t-student)	0.3649	0.7152	No significativo
±3	Signos	NaN	0.0157	Significativo
±5	Patell	2.4770	NaN	Significativo
±5	BMP (t-student)	2.4770	0.0133	Significativo
±5	Signos	NaN	0.0258	Significativo

Los resultados de los tres tests aplicados a las ventanas simétricas de ± 1 , ± 3 y ± 5 días alrededor del evento se resumen a continuación:

- En las ventanas de ± 1 y ± 3 días, ni el test de Patell ni el BMP mostraron significancia estadística ($Z \approx 0.36$; $p \approx 0.715$), lo que indica que, en promedio, los retornos anormales acumulados no se desvían significativamente de cero.
- El test de signos, en cambio, sí arrojó resultados significativos en ambas ventanas ($p \approx 0.008$ y $p \approx 0.016$, respectivamente), lo cual sugiere que existe una asimetría sistemática en la dirección de los retornos, es decir, que hubo más eventos con CAR positivo (o negativo) de lo que se esperaría bajo la hipótesis nula.
- En la ventana de ± 5 días, todos los tests aplicados (Patell, BMP y signos) resultaron estadísticamente significativos, lo que confirma que el efecto de las noticias se acumula con el tiempo y se refleja con mayor claridad en horizontes temporales más amplios.

Aunque hayamos abordado diversos estudios que hacían uso de modelos GARCH, que han sido utilizados para capturar la volatilidad condicional en series financieras, su uso no se considera necesario en este análisis por varias razones:

1. Los contrastes aplicados (Patell y BMP) ya incluyen ajustes estándar que mitigan el impacto de la heteroscedasticidad, ofreciendo resultados válidos incluso en presencia de volatilidad no constante.
2. El objetivo principal de este trabajo no es modelar la dinámica completa de la varianza condicional, sino evaluar el impacto medio de noticias puntuales sobre los retornos de Endesa.
3. Los resultados obtenidos con los contrastes clásicos son estadísticamente significativos en varias ventanas, lo cual ofrece evidencia robusta del efecto sin necesidad de recurrir a modelos más complejos.
4. Por último, modelos como Scholes-Williams o M-GARCH, pensados para mercados poco líquidos o con asincronía de precios, no son necesarios en este contexto, dada la liquidez del activo analizado (Endesa) y la frecuencia diaria de los datos.

En resumen, se opta por una metodología contrastada en la literatura que permite captar el impacto de las noticias sobre los retornos anormales sin necesidad de introducir modelos adicionales que podrían aumentar innecesariamente la complejidad del análisis sin aportar una mejora significativa en los resultados.

5. Conclusiones

Este estudio demuestra que las noticias relacionadas con la sostenibilidad tienen la capacidad de provocar reacciones identificables en los mercados financieros, incluso cuando se analizan desde el prisma de una sola empresa cotizada. Al centrar el análisis en Endesa, compañía emblemática del sector energético español, y aplicar una metodología robusta que combina scraping de noticias, técnicas de NLP y modelos econométricos de series temporales, se logra capturar cómo el mercado incorpora este tipo de información en la evolución diaria del precio de la acción. Esto confirma la hipótesis y premisa de la que partíamos: Las noticias sobre sostenibilidad con tono positivo tienen un impacto positivo en la valoración de los activos, en nuestro caso Endesa.

Los resultados evidencian que el tono de las noticias es un factor determinante en la magnitud y dirección de la reacción bursátil. Las noticias percibidas como negativas tienden a provocar caídas más pronunciadas en el precio que las positivas logran impulsar al alza. Esta asimetría refleja una sensibilidad del inversor hacia los riesgos ESG y sugiere que los impactos reputacionales o regulatorios adversos son rápidamente descontados por el mercado.

Además, se observan diferencias relevantes según el tipo de contenido. Las noticias que abordan temas de gobernanza o decisiones políticas suelen estar más asociadas a movimientos negativos, mientras que las relacionadas con sostenibilidad medioambiental, transición energética o iniciativas sociales, si bien no siempre generan reacciones significativas, sí han mostrado cierta capacidad para construir una narrativa de valor en el tiempo. Esta trayectoria acumulada, visible en los retornos anómalos agregados, muestra que las noticias vinculadas a sostenibilidad tienden a contribuir de forma más estable y positiva al comportamiento de la acción frente a otro tipo de noticias más volátiles o especulativas.

La implementación de modelos como CAPM y ARIMA, junto con el análisis empírico de retornos anómalos, permite aislar estos impactos y distinguir entre movimientos normales de mercado y los atribuibles a eventos informativos. Esto refuerza la idea de que el mercado español, en cierto grado, ya incorpora en sus precios aspectos relacionados con sostenibilidad, especialmente cuando estos son comunicados de forma clara y pública.

6. Valor del estudio y futuras direcciones

Este trabajo aporta evidencia concreta sobre el vínculo entre sostenibilidad y valoración bursátil en un contexto poco explorado: una empresa española individual con una estrategia activa en ESG. Además, demuestra la potencia de unir herramientas tecnológicas como NLP y scraping con teoría financiera para abordar temas emergentes en la economía actual.

Para futuras investigaciones, sería valioso ampliar el análisis a otras compañías del IBEX 35 o explorar sectores distintos para contrastar sensibilidades. También sería interesante aplicar modelos más avanzados de clasificación por categoría y tono o datos de mayor frecuencia. Del mismo modo, también se podrían aplicar otros modelos para capturar posibles cambios en la volatilidad tras eventos ESG o estudiar cómo varía la respuesta del mercado en función del canal de la noticia.

7. Declaración de Uso IA

Declaración de Uso de Herramientas de Inteligencia Artificial Generativa en Trabajos Fin de Grado

ADVERTENCIA: Desde la Universidad consideramos que ChatGPT u otras herramientas similares son herramientas muy útiles en la vida académica, aunque su uso queda siempre bajo la responsabilidad del alumno, puesto que las respuestas que proporciona pueden no ser veraces. En este sentido, NO está permitido su uso en la elaboración del Trabajo fin de Grado para generar código porque estas herramientas no son fiables en esa tarea. Aunque el código funcione, no hay garantías de que metodológicamente sea correcto, y es altamente probable que no lo sea. Por la presente, yo, Genoveva Alférez Aragón, estudiante de E2 + Analytics de la Universidad Pontificia Comillas al presentar mi Trabajo Fin de Grado titulado " IMPACTO DE LAS NOTICIAS SOSTENIBLES EN LA COTIZACIÓN BURSÁTIL: UN ANÁLISIS DE EVENTOS SOBRE ENDESA", declaro que he utilizado la herramienta de Inteligencia Artificial Generativa ChatGPT u otras similares de IAG de código sólo en el contexto de las actividades descritas a continuación: 1. Brainstorming de ideas de investigación: Utilizado para idear y esbozar posibles áreas de investigación. 2. Crítico: Para encontrar contra-argumentos a una tesis específica que pretendo defender. 3. Referencias: Usado conjuntamente con otras herramientas, como Science, para identificar referencias preliminares que luego he contrastado y validado. 5. Interpretador de código: Para realizar análisis de datos preliminares. 6. Estudios multidisciplinares: Para comprender perspectivas de otras comunidades sobre temas de naturaleza multidisciplinar. 8. Corrector de estilo literario y de lenguaje: Para mejorar la calidad lingüística y estilística del texto. 10. Sintetizador y divulgador de libros complicados: Para resumir y comprender literatura compleja. 13. Revisor: Para recibir sugerencias sobre cómo mejorar y perfeccionar el trabajo con diferentes niveles de exigencia. 15. Traductor: Para traducir textos de un lenguaje a otro.

Afirmo que toda la información y contenido presentados en este trabajo son producto de mi investigación y esfuerzo individual, excepto donde se ha indicado lo contrario y se han dado los créditos correspondientes (he incluido las referencias adecuadas en el TFG y he explicitado para que se ha usado ChatGPT u otras herramientas similares). Soy consciente de las implicaciones académicas y éticas de presentar un trabajo no original y acepto las consecuencias de cualquier violación a esta declaración.

Fecha: 03/06/2025

Firma: _____



8. Bibliografía

- Albanese, M., Caporale, G. M., Colella, I., & Spagnolo, N. (2025). Climate policies, energy shocks and spillovers between green and brown stock price indices (CESifo Working Paper No. 11747). *CESifo Working Paper Series*. <https://doi.org/10.2139/ssrn.5195750>
- Barroso del Toro, A., Vivas Crisol, L., & Tort-Martorell, X. (2022). *Comparing the impacts of sustainability narratives on American and European energy shareholders: A multi-event study analysing reactions to news before and during COVID-19*. *Sustainability*, 14(23), 15836. <https://doi.org/10.3390/su142315836>
- Billio, M., Costola, M., Hristova, I., Latino, C., & Pelizzon, L. (2021). Inside the ESG ratings: (Dis)agreement and performance. *Corporate Social Responsibility and Environmental Management*, 28(5), 1426–1445. <https://doi.org/10.1002/csr.2177>
- Box, G.E.P., Jenkins, G.M. and Reinsel, G.C. (2015) *Time Series Analysis: Forecasting and Control*. 5th edn. Hoboken, New Jersey: John Wiley & Sons. Available at: https://books.google.es/books?id=rNt5CgAAQBAJ&redir_esc=y (Accessed: 29 April 2025).
- Cahan, S. F., Chen, C., Chen, L., & Nguyen, N. H. (2015). Corporate social responsibility and media coverage. *Journal of Banking & Finance*, 59, 409–422. <https://doi.org/10.1016/j.jbankfin.2015.07.004>
- Capelle-Blancard, G., & Petit, A. (2019). *Every little helps? ESG news and stock market reaction*. *Journal of Business Ethics*, 157(2), 543–565. <https://doi.org/10.1007/s10551-017-3667-3>
- Dickey, D. A., & Fuller, W. A. (1979). *Distribution of the estimators for autoregressive time series with a unit root*. *Journal of the American Statistical Association*, 74(366), 427–431. <https://doi.org/10.2307/2286348>
- Dorfleitner, G., & Zhang, R. (2024). *ESG News Sentiment and Stock Price Reactions: A Comprehensive Analysis*. *Financial Innovation*, 10(1), 1–25. <https://doi.org/10.1007/s41471-024-00185-3>
- Endesa. (2023). *Estado de información no financiera y sostenibilidad 2022*. <https://www.endesa.com/content/dam/enel-es/home/inversores/registrosoficiales/hechosrelevantes/documentos/2022/estado-de-informacion-no-financiera-y-sostenibilidad-2022.pdf>
- Freeman, R. E. (1984). *Strategic management: A stakeholder approach*. Pitman Publishing.
- Friede, G., Busch, T., & Bassen, A. (2015). *ESG and financial performance: Aggregated evidence from more than 2000 empirical studies*. *Journal of Sustainable Finance & Investment*, 5(4), 210–233. <https://doi.org/10.1080/20430795.2015.1118917>
- Friedman, M. (1970, September 13). *The social responsibility of business is to increase its profits*. *The New York Times Magazine*.

<https://www.nytimes.com/1970/09/13/archives/a-friedman-doctrine-the-social-responsibility-of-business-is-to.html>

- Ghani, U., Zhu, B., Qin, Q., & Ghani, M. (2024). *Forecasting US stock market volatility: Evidence from ESG and CPU indices*. *Finance Research Letters*, 59, 104811. <https://doi.org/10.1016/j.frl.2023.104811>
- Hoepner, A.G.F., Schopohl, L. On the Price of Morals in Markets: An Empirical Study of the Swedish AP-Funds and the Norwegian Government Pension Fund. *J Bus Ethics* 151, 665–692 (2018). <https://doi.org/10.1007/s10551-016-3261-0>
- Jiang, H. Q. (2023). *The application of the ARIMA-GARCH hybrid model for forecasting the Apple stock price*. *Journal of Intelligence and Knowledge Exploration*, 1(1), 1–10. <https://doi.org/10.62517/jike.202304102>
- Krüger, P. (2015). *Corporate goodness and shareholder wealth*. *Journal of Financial Economics*, 115(2), 304–329. <https://doi.org/10.1016/j.jfineco.2014.09.008>
- López, A. M., Flores, M. A., & Sánchez, J. I. (2017). Modelos de series temporales aplicados a la predicción del tráfico aeroportuario español de pasajeros: Un enfoque agregado y desagregado. *Estudios de Economía Aplicada*, 35(2), 395–418. <https://doi.org/10.25115/eea.v35i2.2478>
- McKnight, S. (2019). *Statistical Analysis for Financial Data: A Case Study of Four Stocks*. Montana State University. https://math.montana.edu/grad_students/writing-projects/2019/McKnight2019.pdf
- ONU (2004). *Who cares wins: Connecting financial markets to a changing world*. Tech. rep. United Nations Global Compact. <https://books.google.es/books?hl=en&lr=&id=7Zx2LAs4wMwC>
- Porter, M. E., & Kramer, M. R. (2006). *Strategy and society: The link between competitive advantage and corporate social responsibility*. *Harvard Business Review*, 84(12), 78–92. <https://hbr.org/2006/12/strategy-and-society-the-link-between-competitive-advantage-and-corporate-social-responsibility>
- Serafeim, G. (2020, September). *Social-impact efforts that create real value*. *Harvard Business Review*. <https://hbr.org/2020/09/social-impact-efforts-that-create-real-value>
- Wanday, J., & Ajour El Zein, S. (2022). Higher expected returns for investors in the energy sector in Europe using an ESG strategy. *Frontiers in Environmental Science*, 10, 1031827. <https://doi.org/10.3389/fenvs.2022.1031827>
- Wang, Y., Chen, Y., Zhao, J., & Xu, X. (2023). The effect of ESG news on the Chinese stock market. *Journal of Global Information Management*, 31(4), 1–20. <https://doi.org/10.4018/JGIM.313916>

9. Apéndices y Ilustraciones

9.1. Explicación API:

En el caso de la API el proceso que hemos seguido ha sido más complejo pues requería la creación de una API propia desde Google (Google Custom Search API, también conocida como Programmable Search):

1. Paso 1: Creación del motor de búsqueda personalizado. Se accedió a la plataforma Google Programmable Search Engine y se creó un nuevo motor de búsqueda, seleccionando la opción de búsqueda en toda la web (modo abierto). Esta configuración asegura la máxima cobertura de resultados, evitando limitar la búsqueda a un subconjunto específico de fuentes. El motor fue configurado para priorizar resultados en idioma español (hl=es) y orientado geográficamente a España (gl=ES). El identificador del motor de búsqueda, conocido como cx, fue generado automáticamente y registrado para su uso posterior.

<https://programmablesearchengine.google.com/controlpanel/create/congrats?cx=b640e6d2b90d847b1>

```
<script      async      src="https://cse.google.com/cse.js?cx=b640e6d2b90d847b1">
</script>
<div class="gcse-search"></div>
```

2. Paso 2: Obtención del identificador del motor (cx). Al completar la configuración, se obtuvo un identificador único del motor de búsqueda (cx), en este caso:

[b640e6d2b90d847b1](https://cse.google.com/cse?cx=b640e6d2b90d847b1)

Este código permite vincular las peticiones que se realicen mediante la API con el motor de búsqueda personalizado creado en el paso anterior.

<https://cse.google.com/cse?cx=b640e6d2b90d847b1>

3. Paso 3: Activación de la API y generación de la clave (API Key). Desde la consola de Google Cloud Console, se creó un proyecto específico para el TFG, y se habilitó el servicio Custom Search API. Posteriormente, se generó una clave de autenticación de API (API Key) que permite realizar llamadas al servicio. Esta clave se utiliza junto con el parámetro cx en las solicitudes HTTP para recuperar resultados de búsqueda.

La clave generada fue la siguiente:

[AIzaSyDAHpE9sJoCCYAB4dFfZ9vMgmoNgsJBCOU](https://programmablesearchengine.google.com/controlpanel/overview?cx=b640e6d2b90d847b1)

<https://programmablesearchengine.google.com/controlpanel/overview?cx=b640e6d2b90d847b1>

Una vez obtenidas las credenciales necesarias, se implementó un script en Python que realiza peticiones diarias a la API para buscar noticias relacionadas con "Endesa ESG OR sostenibilidad OR medioambiente OR gobernanza OR social".

Ilustración Anexo 1: Capturas de proceso de descarga API

Las imágenes muestran el proceso de configuración de un buscador personalizado en Google. La primera captura, titulada "Vista general", muestra los datos básicos del buscador: "Nombre del buscador" (Google Noticias ESG Endesa), "Descripción" (Añadir descripción), "Código" (Obtener código), "ID de buscador" (b640e6d2b90d847b1) y "URL pública" (https://cse.google.com/cse?cx=b640e6d2b90d847b1). La segunda captura, titulada "Funciones de búsqueda", muestra la configuración de búsqueda, incluyendo opciones como "Búsqueda por imágenes", "Búsqueda Segura" y "Buscar en toda la Web", así como la selección de la "Región" (Todas las region...). La tercera captura muestra un mensaje de éxito: "Tu nuevo buscador se ha creado", con instrucciones para copiar el código y pegarlo en el sitio web, y un código de ejemplo:

```
<script async src="https://cse.google.com/cse.js?cx=b640e6d2b90d847b1">
</script>
<div class="gsc-search"></div>
```

 y botones para "Vista previa" y "Personalizar".

9.2. Otros Ilustraciones

Ilustración Anexo 2: Contador de noticias por tono

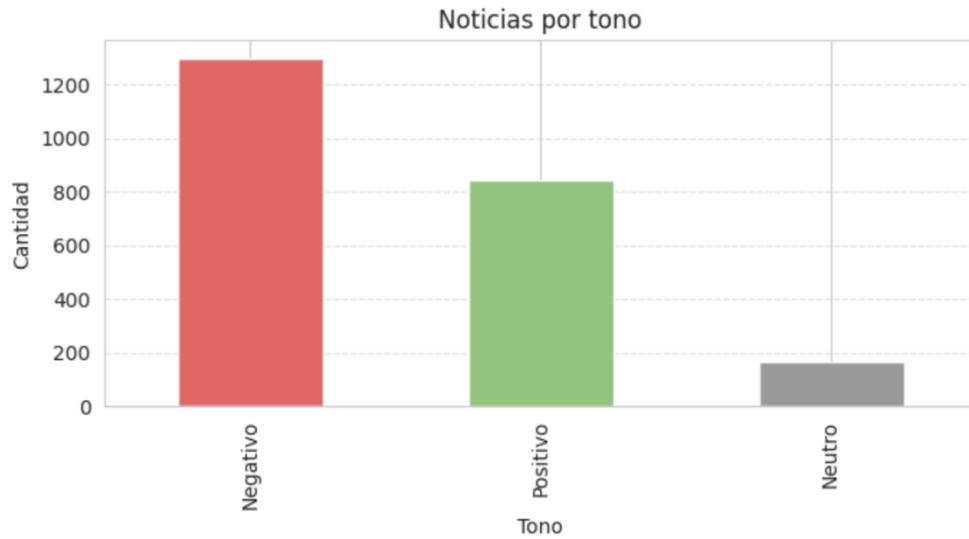


Ilustración Anexo 3: Evolución de score ESG a lo largo del tiempo por tono

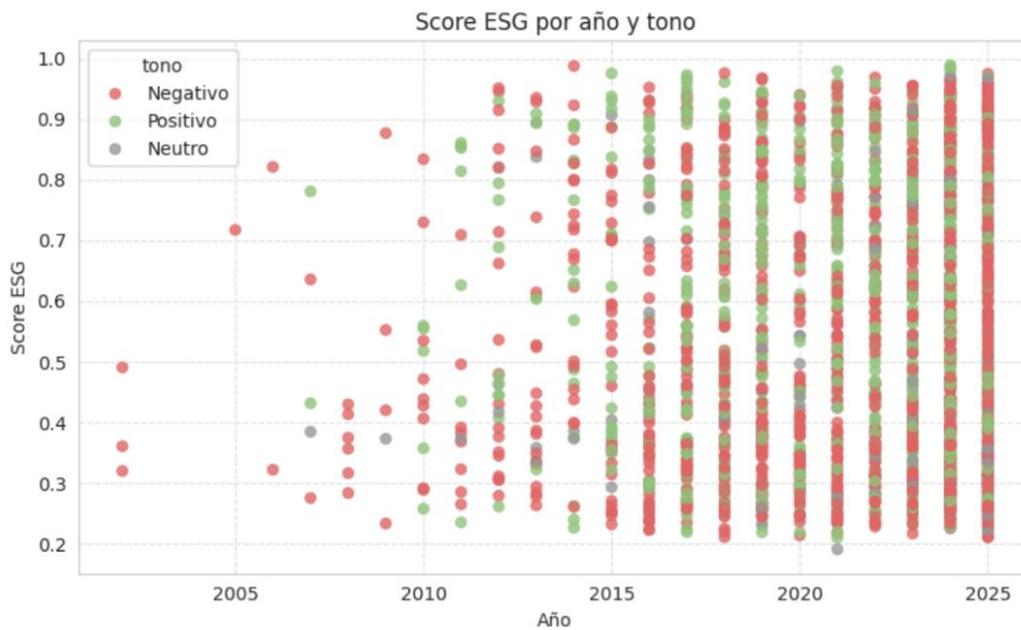


Ilustración Anexo 4: Evolución de retornos anómalos a lo largo del tiempo

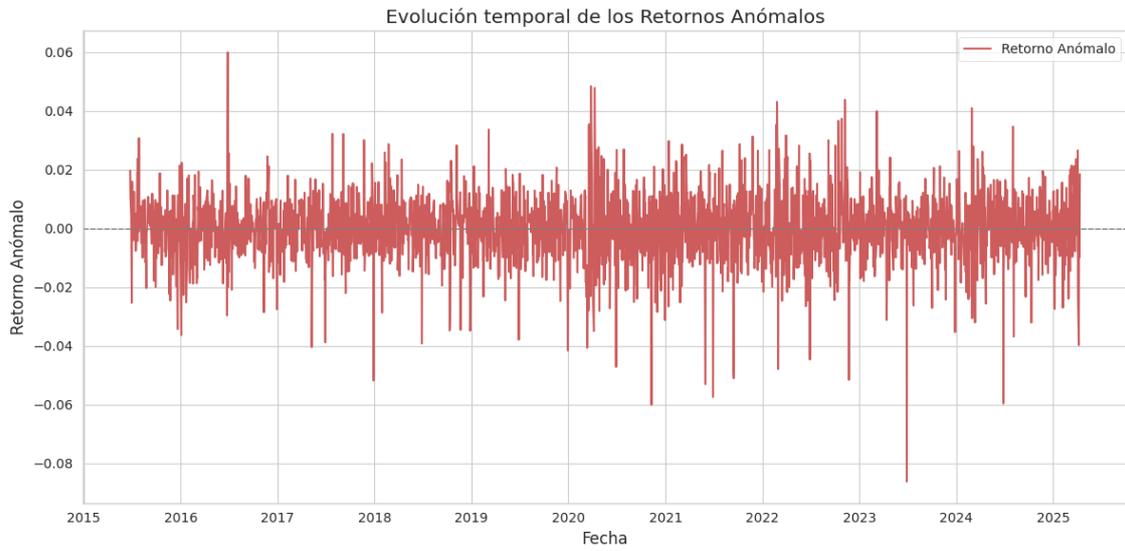
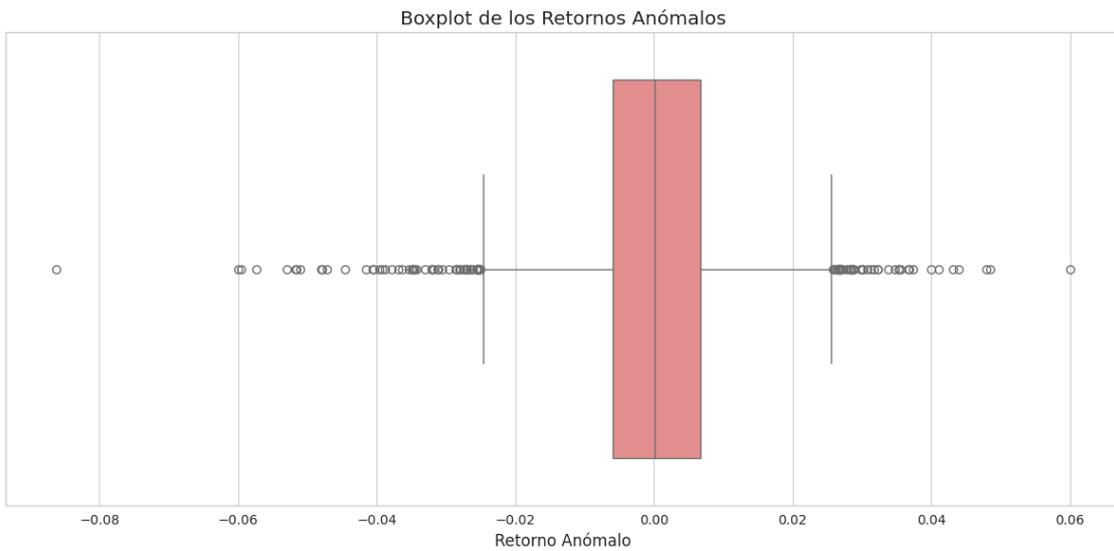


Ilustración Anexo 5: Boxplot sobre el cálculo de los retornos anómalos



9.3. Código scraping noticias y NLP

```
8. !pip install feedparser
9.
10. import feedparser
11. import pandas as pd
12. from datetime import datetime
13.
14. def get_news_feed(query="Endesa", max_articles=1000):
15.     url =
16.     f"https://news.google.com/rss/search?q={query.replace(' ',
17.     '+')}&hl=es&gl=ES&ceid=ES:es"
18.     feed = feedparser.parse(url)
19.
20.     articles = []
21.     for entry in feed.entries[:max_articles]:
22.         title = entry.title
23.         summary = entry.summary
24.         link = entry.link
25.         published = datetime(*entry.published_parsed[:6])
26.         articles.append({
27.             "title": title,
28.             "summary": summary,
29.             "link": link,
30.             "published": published
31.         })
32.     return pd.DataFrame(articles)
33.
34. # --- NUEVO BLOQUE DE BÚSQUEDAS ESPECÍFICAS ---
35. # Búsquedas orientadas a distintos años y temas para forzar
36. # antigüedad
37. queries = [
38.     "Endesa sostenibilidad",
39.     "Endesa ESG",
40.     "Endesa medio ambiente",
41.     "Endesa transición energética",
42.     "Endesa energías renovables",
43.     "Endesa responsabilidad social",
44.     "Endesa energía verde",
45.     "Endesa energía solar",
46.     "Endesa energía eólica",
47.     "Endesa energía verde",
48.     "Endesa cambio climático",
49.     "Endesa parque solar",
50.     "Endesa impacto",
51.     "Endesa ambiental",
```

50. #Connotacion negativa
51. "Endesa críticas",
52. "Endesa polémica",
53. "Endesa denuncia ambiental",
54. "Endesa contaminación",
55. "Endesa residuos",
56. "Endesa sanción",
57. "Endesa multa",
58. "Endesa protestas",
59. "Endesa apagón",
60. "Endesa vertidos",
61. "Endesa escándalo",
62. # Forzando por años
63. "Endesa sostenibilidad 2015",
64. "Endesa sostenibilidad 2016",
65. "Endesa sostenibilidad 2017",
66. "Endesa sostenibilidad 2018",
67. "Endesa sostenibilidad 2019",
68. "Endesa sostenibilidad 2020",
69. "Endesa ESG 2015",
70. "Endesa ESG 2016",
71. "Endesa ESG 2017",
72. "Endesa medio ambiente 2015",
73. "Endesa medio ambiente 2016",
74. "Endesa medio ambiente 2017",
75. "Endesa COP21",
76. "Endesa cambio climático 2015",
77. "Endesa renovables 2016",
78. "Endesa renovables 2017",
79. "Endesa informe de sostenibilidad 2010",
80. "Endesa informe de sostenibilidad 2011",
81. "Endesa informe de sostenibilidad 2012",
82. "Endesa informe de sostenibilidad 2013",
83. "Endesa informe de sostenibilidad 2014",
84. "Endesa informe de sostenibilidad 2015",
85. #Forzado por más años en antigüedad
86. "Endesa responsabilidad social corporativa 2010",
87. "Endesa responsabilidad social corporativa 2011",
88. "Endesa responsabilidad social corporativa 2012",
89. "Endesa eficiencia energética 2010",
90. "Endesa eficiencia energética 2011",
91. "Endesa renovables 2012",
92. "Endesa renovables 2013",
93. "Endesa renovables 2014",
94. "Endesa energías renovables 2015",
95. "Endesa transición energética 2016",
96. "Endesa huella de carbono 2016",
97. "Endesa COP21 2015",
98. "Endesa COP22 2016",

```

99.     "Endesa cambio climático 2017",
100.    "Endesa Plan de Descarbonización 2017",
101.    "Endesa sostenibilidad 2018",
102.    "Endesa innovación energética 2018",
103.    "Endesa medio ambiente 2018",
104.    "Endesa medio ambiente 2019",
105.    "Endesa medio ambiente 2020",
106.    "Endesa medio ambiente 2021",
107. ]
108.
109. df_list = []
110.
111. for query in queries:
112.     df_temp = get_news_feed(query)
113.     df_list.append(df_temp)
114.
115. # Combinamos todas las noticias y eliminamos duplicados
116. df_news = pd.concat(df_list,
117.                    ignore_index=True).drop_duplicates(subset=["link"])
118. # Mostramos resultado
119. print(len(df_news))
120. print(df_news.head())
121. df_news.to_csv("df_news.csv", index=False)
122. print("📄 Guardado en: df_news.csv")
123. import pandas as pd
124. from bs4 import BeautifulSoup
125.
126. # Cargar el archivo original
127. df = pd.read_csv("df_news.csv")
128.
129. # Convertir 'published' en datetime
130. df['published'] = pd.to_datetime(df['published'],
131.                                errors='coerce')
132.
133. # Limpiar campos
134. def clean_html(raw_html):
135.     if pd.isna(raw_html):
136.         return ""
137.     return BeautifulSoup(raw_html, "html.parser").get_text()
138.
139. # Limpiar summary y título
140. df['summary'] = df['summary'].apply(clean_html)
141. df['title'] = df['title'].apply(lambda x: x.strip() if
142.                                isinstance(x, str) else x)
143.
144. # Eliminar duplicados entre título y resumen
145. def clean_summary(row):
146.     if pd.isna(row['summary']):

```

```

145.         return ""
146.     if row['summary'].strip() == row['title'].strip():
147.         return ""
148.     if row['title'].strip() in row['summary'].strip():
149.         return row['summary'].replace(row['title'],
150.         '').strip()
151.     return row['summary'].strip()
152. df['summary'] = df.apply(clean_summary, axis=1)
153.
154. # Añadir columna con el año
155. df['year'] = df['published'].dt.year
156.
157. # Crear nueva columna combinada
158. def combine_title_summary(row):
159.     if row['summary']:
160.         return f"{row['title']}. {row['summary']}"
161.     else:
162.         return row['title']
163.
164. df['title_summary'] = df.apply(combine_title_summary, axis=1)
165.
166. # Seleccionar solo columnas finales
167. df_news_clean = df[['published', 'year', 'title_summary',
168.     'link']].sort_values('published')
169.
170. # Guardar resultado
171. df_news_clean.to_csv("df_news_clean.csv", index=False)
172.
173. # Vista previa
174. print(df_news_clean.head())
175.
176. # === 1. Importar librerías necesarias ===
177. from transformers import pipeline
178. import pandas as pd
179. from tqdm import tqdm
180.
181. # === 2. Cargar modelo de clasificación ESG ===
182. classifier_esg = pipeline("zero-shot-classification",
183.     model="joeddav/xlm-roberta-large-xnli")
184.
185. # === 3. Definir etiquetas ESG ===
186. etiquetas_esg = ["Medioambiente", "Social", "Gobernanza",
187.     "Finanzas", "Energía", "Política"]
188.
189. # === 4. Cargar el DataFrame limpio ===
190. df_news_clean = pd.read_csv("df_news_clean.csv")
191.
192. # === 5. Preparar los textos a clasificar ===

```

```

189. df_news_clean["esg_tag"] =
    df_news_clean["title_summary"].fillna("")
190. textos = df_news_clean["esg_tag"].tolist()
191.
192. # === 6. Clasificación por lotes ===
193. print("🔗 Clasificando noticias por categoría ESG...")
194. resultados = []
195. batch_size = 8
196.
197. for i in tqdm(range(0, len(textos), batch_size)):
198.     batch = textos[i:i+batch_size]
199.     batch_result = classifier_esg(batch,
    candidate_labels=etiquetas_esg)
200.     resultados.extend(batch_result)
201.
202. # === 7. Asignar categoría y score con umbral 0.55 ===
203. def obtener_etiqueta(result, umbral=0.55):
204.     return result["labels"][0] if result["scores"][0] >
    umbral else "No ESG"
205.
206. df_news_clean["categoria_esg"] = [obtener_etiqueta(r,
    umbral=0.55) for r in resultados]
207. df_news_clean["score_categoria"] = [r["scores"][0] for r in
    resultados]
208.
209. # === 8. Guardar TODAS las noticias (nada se filtra) ===
210. df_news_clean.to_csv("noticias_endesa_esg_roberta.csv",
    index=False)
211.
212. print("✅ Clasificación ESG completada con umbral 0.55")
213. print("📁 Guardado en: noticias_endesa_esg_roberta.csv")
214.
215. # === 1. Importar librerías necesarias ===
216. from transformers import pipeline
217. import pandas as pd
218. from tqdm import tqdm
219.
220. # === 2. Cargar el archivo clasificado por ESG ===
221. df_esg = pd.read_csv("noticias_endesa_esg_roberta.csv")
222.
223. # === 3. Preparar textos a analizar (usamos 'title_summary')
    ===
224. textos = df_esg["title_summary"].fillna("").tolist()
225.
226. # === 4. Cargar modelo de análisis de tono ===
227. classifier_sentimiento = pipeline("sentiment-analysis",
    model="nlptown/bert-base-multilingual-uncased-sentiment")
228.
229. # === 5. Analizar tono de cada texto ===

```

```

230. print("🕒 Analizando tono de las noticias...")
231. tonos = []
232.
233. for text in tqdm(textos):
234.     try:
235.         resultado = classifier_sentimiento(text[:512])[0] #
        Limita a 512 tokens
236.         score_label = resultado["label"]
237.         # Simplificar a tres categorías
238.         if "5" in score_label or "4" in score_label:
239.             tono = "Positivo"
240.         elif "3" in score_label:
241.             tono = "Neutro"
242.         else:
243.             tono = "Negativo"
244.     except:
245.         tono = "Neutro" # En caso de error, asumimos Neutro
246.     tonos.append(tono)
247.
248. # === 6. Añadir la columna de tono al DataFrame ===
249. df_esg["tono"] = tonos
250.
251. # === 7. Guardar archivo final ===
252. df_esg.to_csv("noticias_endesa_esg_completo.csv",
        index=False)
253.
254. print("✅ Análisis de tono completado.")
255. print("📁 Guardado en: noticias_endesa_esg_completo.csv")

```

8.4. Código análisis precios, Modelos ARIMA, White Noise, CAPM, CAR y test de contrastes

```

9. import pandas as pd
10.
11. #Cargamos los precios de cotización de Endesa de Investing
12. df_prices = pd.read_csv("endesa_precios.csv", sep=";")
13.
14. #Ordenamos los precios de mayor antigüedad a menor
15. df_prices["Date"] = pd.to_datetime(df_prices["Date"],
        format="%m/%d/%Y")
16. df_prices = df_prices.sort_values("Date")
17.
18. # Renombramos "Price" a algo más cómodo si quieres (opcional)
19. df_prices.rename(columns={"Price": "Close"}, inplace=True)
20.
21. df_prices.head()
22. df_prices['Close'].describe().round(2)
23. import matplotlib.pyplot as plt
24.

```

```

25. plt.figure(figsize=(8, 4))
26. plt.hist(df_prices['Close'], bins=20, color="#4C72B0",
    edgecolor='white', alpha=0.85)
27.
28. # Títulos y etiquetas
29. plt.title("Distribución de los Precios de Cierre (Close)",
    fontsize=12)
30. plt.xlabel("Precio de Cierre (€)", fontsize=10)
31. plt.ylabel("Frecuencia", fontsize=10)
32.
33. # Estética: quitar bordes superior y derecho
34. plt.gca().spines[['top', 'right']].set_visible(False)
35.
36. # Añadir grid sutil
37. plt.grid(axis='y', linestyle='--', alpha=0.6)
38.
39. # Ajustar espaciado
40. plt.tight_layout()
41.
42. # Mostrar
43. plt.show()
44.
45. #Calculamos los retornos diarios simples
46. df_prices['Return'] = df_prices['Close'].pct_change()
47.
48. # Calculamos los retornos diarios logarítmicos
49. import numpy as np
50. df_prices['LogReturn'] = np.log(df_prices['Close'] /
    df_prices['Close'].shift(1))
51.
52. # Quitamos los primeros NaN
53. df_prices = df_prices.dropna()
54.
55. df_prices[['Close', 'Return', 'LogReturn']].head().round(5)
56. import matplotlib.pyplot as plt
57.
58. # Precio
59. plt.figure(figsize=(8,4))
60. plt.plot(df_prices['Date'], df_prices['Close'])
61. plt.title('Precio de cierre ajustado de Endesa (2015-2025)')
62. plt.xlabel('Fecha')
63. plt.ylabel('Precio (€)')
64. plt.grid()
65. plt.show()
66.
67. # Retornos
68. plt.figure(figsize=(8,4))
69. plt.plot(df_prices['Date'], df_prices['Return'])
70. plt.title('Retornos diarios de Endesa')

```

```

71. plt.xlabel('Fecha')
72. plt.ylabel('Retorno diario')
73. plt.grid()
74. plt.show()
75. # Estadísticas descriptivas básicas
76. print("Sacamos la media, desviación estándar, máximos,
    mínimos de los retornos diarios:")
77. df_prices['Return'].describe()
78. import matplotlib.pyplot as plt
79.
80. # Calcular media y desviación estándar
81. mean_return = df_prices['Return'].mean()
82. std_return = df_prices['Return'].std()
83.
84. # Crear histograma
85. plt.figure(figsize=(10, 5))
86. plt.hist(df_prices['Return'], bins=100, color='skyblue',
    edgecolor='black', alpha=0.7)
87.
88. # Añadir línea de la media
89. plt.axvline(mean_return, color='red', linestyle='--',
    linewidth=2, label=f'Media: {mean_return:.4f}')
90.
91. # Añadir líneas de desviaciones típicas
92. plt.axvline(mean_return + std_return, color='green',
    linestyle='--', linewidth=1.5, label=f'+1σ: {(mean_return +
    std_return):.4f}')
93. plt.axvline(mean_return - std_return, color='green',
    linestyle='--', linewidth=1.5, label=f'-1σ: {(mean_return -
    std_return):.4f}')
94.
95. # Títulos y etiquetas
96. plt.title("Distribución de retornos diarios de Endesa",
    fontsize=14)
97. plt.xlabel("Retorno diario", fontsize=11)
98. plt.ylabel("Frecuencia", fontsize=11)
99. plt.grid(alpha=0.3)
100.
101. # Leyenda
102. plt.legend()
103.
104. # Mostrar
105. plt.tight_layout()
106. plt.show()
107.
108. from statsmodels.tsa.stattools import adfuller
109.
110. adf_result = adfuller(df_prices['Return'])
111.

```

```

112. print("=== Test de Dickey-Fuller ===")
113. print(f"Estadístico ADF: {adf_result[0]:.4f}")
114. print(f"Valor p: {adf_result[1]:.4f}")
115. print("Valores críticos:", adf_result[4])
116.
117. if adf_result[1] < 0.05:
118.     print("☑ La serie de retornos es estacionaria")
119. else:
120.     print("⚠ La serie NO es estacionaria (requiere
    diferenciación)")
121.
122. from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
123.
124. plot_acf(df_prices['Return'], lags=20)
125. plt.title("ACF de retornos")
126. plt.show()
127.
128. plot_pacf(df_prices['Return'], lags=20)
129. plt.title("PACF de retornos")
130. plt.show()
131. # !pip install pandas numpy matplotlib statsmodels
132.
133. import pandas as pd
134. import numpy as np
135. from statsmodels.tsa.arima.model import ARIMA
136.
137. # --- 1. Simulación de retornos (reemplaza por tus datos
    reales) ---
138. np.random.seed(42)
139. returns = np.random.normal(0, 0.01, 2624)
140.
141. # --- 2. Modelo de Ruido Blanco (media constante) ---
142. mean_wn = np.mean(returns)
143. var_wn = np.var(returns)
144. residuals_wn = returns - mean_wn
145.
146. loglik_wn = -0.5 * len(returns) * (np.log(2 * np.pi) +
    np.log(var_wn) + 1)
147. aic_wn = -2 * loglik_wn + 2 * 1 # 1 parámetro estimado (la
    media)
148. bic_wn = -2 * loglik_wn + np.log(len(returns)) * 1
149.
150. # --- Explicación del modelo de Ruido Blanco ---
151. print("=== Modelo de Ruido Blanco ===")
152. print(f"Media estimada:           {mean_wn:.6f}")
153. print(f"Varianza estimada:        {var_wn:.6f}")
154. print(f"Log-Likelihood:           {loglik_wn:.3f}")
155. print(f"AIC:                     {aic_wn:.3f}")
156. print(f"BIC:                     {bic_wn:.3f}")

```

```

157. print("\nNota: El modelo de ruido blanco no genera un
      `summary()` porque no implica una estructura temporal compleja,
      sino simplemente la estimación de una constante (la media).")
158. print("-" * 60)
159.
160. # --- 3. Modelo ARIMA(1,0,1) ---
161. model_101 = ARIMA(returns, order=(1, 0, 1)).fit()
162. print("=== ARIMA(1,0,1) ===")
163. print(model_101.summary())
164.
165. # --- 4. Modelo ARIMA(0,0,2) ---
166. model_002 = ARIMA(returns, order=(0, 0, 2)).fit()
167. print("\n=== ARIMA(0,0,2) ===")
168. print(model_002.summary())
169.
170. # --- 5. Comparación de métricas con pandas limpio ---
171. comparison_df = pd.DataFrame({
172.     "Modelo": ["Ruido Blanco", "ARIMA(1,0,1)",
173.               "ARIMA(0,0,2)"],
174.     "AIC": [aic_wn, model_101.aic, model_002.aic],
175.     "BIC": [bic_wn, model_101.bic, model_002.bic],
176.     "Log-Likelihood": [loglik_wn, model_101.llf,
177.                        model_002.llf]
178. })
179.
180. # Mostrar tabla bonita con pandas
181. pd.set_option('display.colheader_justify', 'center')
182. pd.set_option('display.expand_frame_repr', False)
183.
184. print("\n=== Comparación de Métricas ===")
185. comparison_df
186.
187. import warnings
188. from statsmodels.tsa.arima.model import ARIMA
189.
190. warnings.filterwarnings("ignore") # Evita los mensajes
    molestos
191.
192. # Rango de modelos a evaluar
193. p_range = range(0, 4)
194. d_range = range(0, 2)
195. q_range = range(0, 4)
196.
197. results = []
198.
199. for p in p_range:
200.     for d in d_range:

```

```

201.         for q in q_range:
202.             try:
203.                 model = ARIMA(df_prices['Return'], order=(p,
204.                     d, q))
205.                 model_fit = model.fit()
206.                 results.append({
207.                     'order': (p, d, q),
208.                     'aic': model_fit.aic,
209.                     'bic': model_fit.bic
210.                 })
211.             except:
212.                 continue
213. # Convertimos a DataFrame para ordenarlo
214. df_results = pd.DataFrame(results)
215. df_results = df_results.sort_values('aic') # También puedes
216.     usar 'bic'
217. print("Mejores modelos ordenados por AIC:")
218. print(df_results.head())
219.
220. import pandas as pd
221. import matplotlib.pyplot as plt
222. import seaborn as sns
223.
224. # Cargar la base de datos
225. df = pd.read_csv("noticias_endesa_esg_completo.csv")
226.
227. # Asegurar formato datetime
228. df['published'] = pd.to_datetime(df['published'],
229.     errors='coerce')
230. df['year'] = df['published'].dt.year
231.
232. # Colores personalizados más vivos
233. tone_colors = {
234.     'Negativo': '#e06666', # rojo
235.     'Positivo': '#93c47d', # verde
236.     'Neutro': '#999999', # gris
237. }
238. # 1. Noticias por año + total
239. news_by_year = df['year'].value_counts().sort_index()
240. total_news = news_by_year.sum()
241.
242. plt.figure(figsize=(7, 4))
243. news_by_year.plot(kind='bar', color='#6fa8dc')
244. plt.title('Noticias sobre Endesa por año')
245. plt.xlabel('Año')
246. plt.ylabel('Cantidad')

```

```

247. plt.xticks(rotation=45)
248. plt.grid(axis='y', linestyle='--', alpha=0.5)
249. # Mostrar el total en la esquina superior derecha
250. plt.text(0.95, 0.95, f'Total: {total_news:,} noticias',
251.          transform=plt.gca().transAxes,
252.          ha='right', va='top', fontsize=10,
253.          bbox=dict(facecolor='white', alpha=0.8,
                edgecolor='gray'))
254. plt.tight_layout()
255. plt.show()
256.
257. # 2. Noticias por tono
258. news_by_tone = df['tono'].value_counts()
259. tone_bar_colors = [tone_colors.get(tone, '#999999') for tone
    in news_by_tone.index]
260.
261. plt.figure(figsize=(7, 4))
262. news_by_tone.plot(kind='bar', color=tone_bar_colors)
263. plt.title('Noticias por tono')
264. plt.xlabel('Tono')
265. plt.ylabel('Cantidad')
266. plt.grid(axis='y', linestyle='--', alpha=0.5)
267. plt.tight_layout()
268. plt.show()
269.
270. # 3. Noticias por categoría ESG
271. news_by_esg_cat = df['categoria_esg'].value_counts()
272.
273. plt.figure(figsize=(7, 4))
274. news_by_esg_cat.plot(kind='bar', color='#76a5af') # azul
    verdoso
275. plt.title('Noticias por categoría ESG')
276. plt.xlabel('Categoría ESG')
277. plt.ylabel('Cantidad')
278. plt.grid(axis='y', linestyle='--', alpha=0.5)
279. plt.tight_layout()
280. plt.show()
281.
282. # 4. Gráfico de puntos: score vs año con color por tono
283. plt.figure(figsize=(8, 5))
284. sns.scatterplot(
285.     data=df,
286.     x='year',
287.     y='score_categoria',
288.     hue='tono',
289.     palette=tone_colors,
290.     alpha=0.8,
291.     edgecolor=None
292. )

```

```

293. plt.title('Score ESG por año y tono')
294. plt.xlabel('Año')
295. plt.ylabel('Score ESG')
296. plt.grid(True, linestyle='--', alpha=0.5)
297. plt.tight_layout()
298. plt.show()
299.
300. import pandas as pd
301. import numpy as np
302.
303. # === 1. Cargar precios desde archivos CSV (formato
      mm/dd/yyyy) ===
304. df_endesa = pd.read_csv("endesa_precios.csv", sep=";",
      parse_dates=["Date"], dayfirst=False)
305. df_ibex = pd.read_csv("IBEX_35_precios.csv", sep=";",
      parse_dates=["Date"], dayfirst=False)
306.
307. # === 2. Limpiar formato numérico de precios ===
308. df_endesa["Price"] =
      df_endesa["Price"].astype(str).str.replace(",", "",
      "").astype(float)
309. df_ibex["Price"] =
      df_ibex["Price"].astype(str).str.replace(",", "", "").astype(float)
310.
311. # === 3. Renombrar columna Price por Close ===
312. df_endesa.rename(columns={"Price": "Close"}, inplace=True)
313. df_ibex.rename(columns={"Price": "Close"}, inplace=True)
314.
315. # === 4. Ordenar correctamente por fecha (de más antigua a
      más reciente) ===
316. df_endesa =
      df_endesa.sort_values("Date").reset_index(drop=True)
317. df_ibex = df_ibex.sort_values("Date").reset_index(drop=True)
318.
319. # === 5. Calcular retornos simples diarios ===
320. df_endesa["Return"] = df_endesa["Close"].pct_change()
321. df_ibex["Return"] = df_ibex["Close"].pct_change()
322.
323. # === 6. Eliminar filas con NaN (por el shift) ===
324. df_endesa = df_endesa.dropna(subset=["Return"])
325. df_ibex = df_ibex.dropna(subset=["Return"])
326.
327. # === 7. Combinar ambos datasets por fecha ===
328. df_merged = pd.merge(
329.     df_endesa[["Date", "Return"]],
330.     df_ibex[["Date", "Return"]],
331.     on="Date",
332.     suffixes=("_endesa", "_ibex")
333. )

```

```

334.
335. # === 8. Mostrar primeros datos para validar ===
336. print(df_merged.head())
337.
338. import pandas as pd
339. import numpy as np
340. from sklearn.linear_model import LinearRegression
341.
342. # === Cargar retornos combinados de Endesa e IBEX 35 ===
343. # Asegúrate de tener df_merged con columnas: 'Date',
    'Return_endesa', 'Return_ibex'
344.
345. # Inicializar nuevas columnas para resultados
346. df_merged["Expected_return"] = np.nan
347. df_merged["Abnormal_return"] = np.nan
348.
349. # Definir tamaño de la ventana para la regresión móvil
350. ventana = 120 # número de días anteriores a usar para
    estimar alpha y beta
351.
352. # Recorrer desde el índice 120 hasta el final (suficientes
    datos)
353. for i in range(ventana, len(df_merged)):
354.     # Datos de la ventana
355.     window_data = df_merged.iloc[i - ventana:i]
356.     X = window_data["Return_ibex"].values.reshape(-1, 1) #
    variable independiente
357.     y = window_data["Return_endesa"].values #
    variable dependiente
358.
359.     # Ajustar modelo de regresión lineal (CAPM)
360.     modelo = LinearRegression().fit(X, y)
361.
362.     # Retorno del mercado en el día actual
363.     r_m = df_merged.iloc[i]["Return_ibex"]
364.
365.     # Predecir retorno esperado según el modelo
366.     expected_r = modelo.predict([[r_m]])[0]
367.
368.     # Guardar el retorno esperado y el retorno anómalo (real
    - esperado)
369.     df_merged.at[i, "Expected_return"] = expected_r
370.     df_merged.at[i, "Abnormal_return"] =
    df_merged.iloc[i]["Return_endesa"] - expected_r
371.
372. # Eliminar las primeras filas que no tienen ventana
    suficiente
373. df_regresion = df_merged.dropna(subset=["Expected_return",
    "Abnormal_return"]).reset_index(drop=True)

```

```

374.
375. # Mostrar primeras filas para validar
376. print(df_regresion[["Date", "Return_endesa", "Return_ibex",
    "Expected_return", "Abnormal_return"]].head())
377.
378. import matplotlib.pyplot as plt
379. import seaborn as sns
380.
381. # Estilo suave
382. sns.set_style("whitegrid")
383. plt.rcParams['figure.figsize'] = (12, 6)
384.
385. # === 1. Evolución temporal de los retornos anómalos ===
386. plt.figure()
387. plt.plot(df_regresion["Date"],
    df_regresion["Abnormal_return"], label="Retorno Anómalo",
    color="indianred")
388. plt.axhline(0, color="gray", linestyle="--", linewidth=1)
389. plt.title("Evolución temporal de los Retornos Anómalos")
390. plt.xlabel("Fecha")
391. plt.ylabel("Retorno Anómalo")
392. plt.legend()
393. plt.tight_layout()
394. plt.show()
395.
396. # === 2. Histograma de distribución ===
397. plt.figure()
398. sns.histplot(df_regresion["Abnormal_return"], bins=50,
    kde=True, color="salmon")
399. plt.axvline(0, color="gray", linestyle="--")
400. plt.title("Distribución de los Retornos Anómalos")
401. plt.xlabel("Retorno Anómalo")
402. plt.ylabel("Frecuencia")
403. plt.tight_layout()
404. plt.show()
405.
406. # === 3. Boxplot para detectar outliers ===
407. plt.figure()
408. sns.boxplot(x=df_regresion["Abnormal_return"],
    color="lightcoral")
409. plt.title("Boxplot de los Retornos Anómalos")
410. plt.xlabel("Retorno Anómalo")
411. plt.tight_layout()
412. plt.show()
413.
414. # === 4. Media móvil de 30 días (tendencia general) ===
415. plt.figure()
416. df_regresion["Abnormal_return_ma30"] =
    df_regresion["Abnormal_return"].rolling(30).mean()

```

```

417. plt.plot(df_regresion["Date"],
            df_regresion["Abnormal_return"], alpha=0.8, label="Retorno
            Anómalo", color="salmon")
418. plt.plot(df_regresion["Date"],
            df_regresion["Abnormal_return_ma30"], color="steelblue",
            linewidth=2, label="Media móvil (30 días)")
419. plt.axhline(0, color="gray", linestyle="--", linewidth=1)
420. plt.title("Tendencia de Retornos Anómalos (Media Móvil 30
            días)")
421. plt.xlabel("Fecha")
422. plt.ylabel("Retorno Anómalo")
423. plt.legend()
424. plt.tight_layout()
425. plt.show()
426.
427. import pandas as pd
428. from datetime import timedelta
429.
430. # === 1. Cargar las noticias ESG clasificadas ===
431. df_news = pd.read_csv("noticias_endesa_esg_completo.csv")
432. df_news['published'] = pd.to_datetime(df_news['published'])
433.
434. # === 2. Asegurar formato del DataFrame de retornos anómalos
            ===
435. df_regresion['Date'] = pd.to_datetime(df_regresion['Date'])
436.
437. # === 3. Obtener la fecha mínima válida tras 120 días de
            datos históricos ===
438. min_fecha_valida = df_regresion['Date'].min() +
            timedelta(days=120)
439.
440. # === 4. Filtrar las noticias para que estén dentro del rango
            donde ya hay estimación de AR ===
441. df_news = df_news[df_news['published'] >=
            min_fecha_valida].copy()
442.
443. # === 5. Inicializar lista para almacenar resultados ===
444. car_resultados = []
445.
446. # === 6. Calcular CARs en tres ventanas para cada noticia ===
447. for _, fila in df_news.iterrows():
448.     fecha_evento = fila['published']
449.     resultado = {
450.         'fecha_evento': fecha_evento,
451.         'categoria_esg': fila.get('categoria_esg',
            'Desconocido'),
452.         'tono': fila.get('tono', 'Desconocido'),
453.         'titulo': fila.get('title_summary', '')[:100],
454.     }

```

```

455.
456.     for ventana in [1, 3, 5]:
457.         inicio = fecha_evento - timedelta(days=ventana)
458.         fin = fecha_evento + timedelta(days=ventana)
459.
460.         subset = df_regresion[
461.             (df_regresion['Date'] >= inicio) &
462.             (df_regresion['Date'] <= fin)
463.         ]
464.
465.         car_valor = subset['Abnormal_return'].sum()
466.         resultado[f'CAR_{ventana}'] = round(car_valor, 8)
467.
468.         car_resultados.append(resultado)
469.
470.     # === 7. Crear DataFrame final, convertir columnas CAR a
         numérico por seguridad ===
471.     df_car = pd.DataFrame(car_resultados)
472.     for col in ['CAR_1', 'CAR_3', 'CAR_5']:
473.         df_car[col] = pd.to_numeric(df_car[col], errors='coerce')
474.
475.     # === 8. Exportar y mostrar ===
476.     df_car.to_csv("resultados_car.csv", index=False,
         float_format='%.8f')
477.
478.     with pd.option_context('display.float_format',
         '{:.8f}'.format):
479.         display(df_car.head())
480.
481.     import matplotlib.pyplot as plt
482.     import seaborn as sns
483.     import pandas as pd
484.
485.     # Aseguramos que los datos están bien cargados
486.     df_car = pd.read_csv("resultados_car.csv")
487.
488.     # Convertimos columnas relevantes por seguridad
489.     df_car['CAR_1'] = pd.to_numeric(df_car['CAR_1'],
         errors='coerce')
490.     df_car['CAR_3'] = pd.to_numeric(df_car['CAR_3'],
         errors='coerce')
491.     df_car['CAR_5'] = pd.to_numeric(df_car['CAR_5'],
         errors='coerce')
492.
493.     # === 1. Gráfico descriptivo general para las tres ventanas
         ===
494.     plt.figure(figsize=(10, 6))
495.     sns.boxplot(data=df_car[['CAR_1', 'CAR_3', 'CAR_5']],
         palette="pastel")

```

```

496. plt.title("Distribución general del CAR por ventana de
evento")
497. plt.ylabel("CAR")
498. plt.xlabel("Ventana de evento (± días)")
499. plt.axhline(0, linestyle="--", color="gray")
500. plt.tight_layout()
501. plt.show()
502.
503. # === 2. Gráfico por categoría ESG (CAR_3) ===
504. plt.figure(figsize=(10, 6))
505. sns.boxplot(data=df_car, x='categoria_esg', y='CAR_3',
palette="pastel")
506. plt.title("Distribución del CAR (±3 días) por Categoría ESG")
507. plt.xlabel("Categoría ESG")
508. plt.ylabel("CAR (±3 días)")
509. plt.axhline(0, linestyle="--", color="gray")
510. plt.xticks(rotation=30)
511. plt.tight_layout()
512. plt.show()
513.
514. # === 3. Gráfico por tono (CAR_3) ===
515. plt.figure(figsize=(8, 6))
516. sns.boxplot(data=df_car, x='tono', y='CAR_3',
palette="pastel")
517. plt.title("Distribución del CAR (±3 días) por Tono de la
noticia")
518. plt.xlabel("Tono")
519. plt.ylabel("CAR (±3 días)")
520. plt.axhline(0, linestyle="--", color="gray")
521. plt.tight_layout()
522. plt.show()
523.
524. # === 4. Gráfico combinado categoría ESG y tono (CAR_3) ===
525. plt.figure(figsize=(12, 6))
526. sns.boxplot(data=df_car, x='categoria_esg', y='CAR_3',
hue='tono', palette="pastel")
527. plt.title("Distribución del CAR (±3 días) por Categoría ESG y
Tono")
528. plt.xlabel("Categoría ESG")
529. plt.ylabel("CAR (±3 días)")
530. plt.axhline(0, linestyle="--", color="gray")
531. plt.legend(title="Tono", bbox_to_anchor=(1.05, 1), loc='upper
left')
532. plt.xticks(rotation=30)
533. plt.tight_layout()
534. plt.show()
535.
536. import pandas as pd
537. import statsmodels.api as sm

```

```

538. from statsmodels.formula.api import ols
539.
540. # === 1. Cargar los resultados CAR ===
541. df_car = pd.read_csv("resultados_car.csv")
542.
543. # === 2. Asegurar que no haya valores nulos en columnas clave
    ===
544. df_test = df_car[['CAR_3', 'categoria_esg', 'tono']].dropna()
545.
546. # === 3. ANOVA por Categoría ESG ===
547. modelo_categoria = ols('CAR_3 ~ C(categoria_esg)',
    data=df_test).fit()
548. anova_categoria = sm.stats.anova_lm(modelo_categoria, typ=2)
549. print("\n=== ANOVA por Categoría ESG ===")
550. print(anova_categoria)
551.
552. # === 4. ANOVA por Tono ===
553. modelo_tono = ols('CAR_3 ~ C(tono)', data=df_test).fit()
554. anova_tono = sm.stats.anova_lm(modelo_tono, typ=2)
555. print("\n=== ANOVA por Tono ===")
556. print(anova_tono)
557.
558. # === 5. ANOVA por interacción Categoría ESG * Tono ===
559. modelo_interaccion = ols('CAR_3 ~ C(categoria_esg) *
    C(tono)', data=df_test).fit()
560. anova_interaccion = sm.stats.anova_lm(modelo_interaccion,
    typ=2)
561. print("\n=== ANOVA por Interacción ESG x Tono ===")
562. print(anova_interaccion)
563.
564. import numpy as np
565. import pandas as pd
566. from scipy import stats
567. from scipy.stats import binomtest
568.
569. # Cargar archivo si no está cargado
570. df = pd.read_csv("resultados_car.csv")
571.
572. # Ejecutar contrastes para cada ventana
573. ventanas = [1, 3, 5]
574. test_results = []
575.
576. for v in ventanas:
577.     col = f'CAR_{v}'
578.     if col not in df.columns:
579.         continue
580.
581.     std = df[col].std()
582.     z_patell = (df[col] / std).sum() / np.sqrt(len(df))

```

```

583.
584.     bmp_t_stat, bmp_p = stats.ttest_1samp(df[col], popmean=0)
585.     n_pos = np.sum(df[col] > 0)
586.     n = np.sum(df[col] != 0)
587.     sign_test_p = binomtest(n_pos, n=n, p=0.5,
    alternative='two-sided').pvalue
588.
589.     test_results.extend([
590.         {
591.             'Ventana': f'±{v}',
592.             'Test': 'Patell',
593.             'Estadístico': round(z_patell, 4),
594.             'p-value': None,
595.             'Resultado': "Significativo" if abs(z_patell) >
    1.96 else "No significativo"
596.         },
597.         {
598.             'Ventana': f'±{v}',
599.             'Test': 'BMP (t-student)',
600.             'Estadístico': round(bmp_t_stat, 4),
601.             'p-value': round(bmp_p, 4),
602.             'Resultado': "Significativo" if bmp_p < 0.05 else
    "No significativo"
603.         },
604.         {
605.             'Ventana': f'±{v}',
606.             'Test': 'Signos',
607.             'Estadístico': None,
608.             'p-value': round(sign_test_p, 4),
609.             'Resultado': "Significativo" if sign_test_p <
    0.05 else "No significativo"
610.         }
611.     ])
612.
613. # Crear tabla
614. tabla_contrastes = pd.DataFrame(test_results)
615. print(tabla_contrastes)
616.
617. import matplotlib.pyplot as plt
618.
619. # Cargar resultados de CAR
620. df_car = pd.read_csv("resultados_car.csv")
621. df_car['fecha_evento'] =
    pd.to_datetime(df_car['fecha_evento'])
622.
623. # Clasificación: ESG vs No ESG
624. df_car['tipo'] = df_car['categoria_esg'].apply(lambda x: 'No
    ESG' if x == 'No ESG' else 'ESG')
625.

```

```

626. # Crear series de tiempo medias acumuladas por tipo
627. evolucion = df_car[['fecha_evento', 'CAR_5',
    'tipo']].dropna()
628. evolucion = evolucion.sort_values('fecha_evento')
629.
630. evolucion['CAR_acum'] =
    evolucion.groupby('tipo')['CAR_5'].cumsum()
631.
632. # Graficar
633. plt.figure(figsize=(10,6))
634. for tipo, grupo in evolucion.groupby('tipo'):
635.     plt.plot(grupo['fecha_evento'], grupo['CAR_acum'],
        label=tipo)
636.
637. plt.axhline(0, linestyle='--', color='gray')
638. plt.title('Evolución acumulada del CAR (±5 días) por tipo de
    noticia')
639. plt.xlabel('Fecha')
640. plt.ylabel('CAR acumulado')
641. plt.legend()
642. plt.tight_layout()
643. plt.show()
644.

```