

TRABAJO FINAL DE MÁSTER

Título: Análisis del riesgo climático y su impacto en el mercado

Autoría: Sofia Torrubiano Saurina

Tutoría: Luis Ortiz Gracia

Curso académico: 2023-2024



UNIVERSITAT DE
BARCELONA

Facultat d'Economia
i Empresa

Màster
de Ciències
Actuarials
i Financeres

Facultad de Economía y Empresa

Universidad de Barcelona

Trabajo Final de Máster

Máster en Ciencias Actuariales y Financieras

Análisis del riesgo climático y su impacto en el mercado

Autoría: Sofia Torrubiano Saurina

Tutoría: Luis Ortiz Gracia

El contenido de este documento es de exclusiva responsabilidad del autor, quien declara que no ha incurrido en plagio y que la totalidad de referencias a otros autores han sido expresadas en el texto

Resumen

El cambio climático cada vez va cobrando más visibilidad en el mundo. En el entorno empresarial y de las finanzas toma importancia el riesgo climático como un factor a tener en cuenta a la hora de medir los riesgos. En este trabajo se analiza la evolución de las anomalías de la temperatura. Después se intenta crear un modelo que intente explicar la evolución de los valores de cotización de una empresa real agrícola y ver si la temperatura del planeta tiene algún impacto en las cotizaciones. Por último, se hace una proyección del valor de las cotizaciones con los modelos ajustados.

Palabras clave: Riesgo climático, regresión, econométrico, polinómico, Gamma.

Abstract

Climate change is increasingly gaining visibility worldwide. In the business and financial environment, climate risk is becoming an important factor to consider when assessing risks. This paper analyses the evolution of temperature anomalies. Then, it attempts to create a model to explain the evolution of the stock values of a real agricultural company and to see if the planet's temperature has any impact on stock prices. Finally, a projection of the stock values is made using the adjusted models.

Keywords: Climate risk, regression, econometric, polynomial, Gamma.

Índice

1. Introducción.....	4
2. Objetivos e hipótesis.....	6
3. Metodología.....	7
4. Modelización de la anomalía de la temperatura	8
4.1. Datos y descripción de los datos	8
4.2. Estudio de la normalidad.....	9
4.3. Regresión lineal.....	10
4.4. Regresión cuadrática	12
4.5. Regresión segmentada lineal.....	13
5. Modelo econométrico	15
5.1. Descripción de los datos	15
5.2. Modelización.....	16
5.2.1. Modelo lineal múltiple.....	16
5.2.2. Modelo Gamma con función enlace identidad	18
5.2.3. Modelo polinómico.....	19
5.2.4. ARIMA (p, d, q).....	20
5.3. Resultados	21
5.4. Capacidad de predicción del modelo	23
6. Proyección de las cotizaciones	25
7. Modelización con una muestra mayor.....	28
8. Conclusiones.....	30
9. Anexos.....	31
9.1. Código hecho en RStudio:	31
10. Bibliografía	40

1. Introducción

El cambio climático lleva siendo durante muchos años un problema del cual se está concienciando a la población. Desde la industrialización, el impacto de los humanos en el medio ambiente ha crecido significativamente. Desde entonces se han podido ver en la naturaleza indicios del cambio climático y los expertos advierten de las consecuencias del cambio climático, como pueden ser un aumento del nivel del mar, deshielo de los glaciares, mayores temporadas de sequías y olas de calor, etc. Sin ir tan lejos, en España se está notando cada vez más las consecuencias del cambio climático. Los periodos de sequía y de incendios forestales en la península son cada vez más largos, causando escasez de agua que la población nota cada vez más.

El medioambiente es estudiado constantemente por organizaciones como *National Aeronautics and Space Administration* (NASA), *National Oceanic and Atmospheric Administration* (NOAA), *Intergovernmental Panel on Climate Change* (IPCC) o *World Meteorological Organization* (WMO) para ver la evolución de los componentes del clima y estudiar qué efectos puede provocar a largo plazo. El aumento de factores medioambientales adversos como inundaciones, subida del nivel del mar y sequías han derivado a que muchos gobiernos del mundo hayan tomado medidas para disminuir el impacto humano en la naturaleza y buscar una neutralidad ecológica.

Se anima a las empresas e inversores a llevar a cabo lo que se denominan “finanzas sostenibles”. Según la CNMV¹ las finanzas sostenibles “implican que en el proceso de toma de decisiones de inversión se tengan en cuenta los factores medioambientales, sociales y de buen gobierno. Todos estos criterios se conocen como criterios ASG”. Los criterios ASG se refieren a criterios que miden el desempeño de una empresa en base a tres factores: Ambientales, Sociales y Gobernanza. No obstante, las empresas no solo están afectadas a nivel de los criterios ASG. Los fenómenos meteorológicos tienen efectos directa e indirectamente en la actividad empresarial y el sector financiero. Por ejemplo, las inundaciones perjudican por un lado a las empresas, las cuales pueden verse obligadas a parar su actividad, pero por otro lado afecta indirectamente al sector financiero, ya que la población se ve obligada a destinar sus ahorros en la reparación de daños materiales.

Los efectos negativos financieros a los que se puede ver afectada una empresa puede ser debido a riesgos físicos o riesgos de transición (BCE, 2020). En primer lugar, el riesgo físico se refiere al impacto financiero directo del clima y medio ambiente como consecuencia de fenómenos meteorológicos, como pueden ser inundaciones y sequías. En segundo lugar, los riesgos de transición hacen referencia a las consecuencias financieras que se derivan del proceso de descarbonización. Dentro de los riesgos de transición pueden verse reflejados en cambios en las políticas medioambientales o avances tecnológicos que puedan afectar al modelo de negocio de la entidad. El impacto negativo que ejercen los riesgos físicos en el mercado financiero es lo que motiva a realizar este estudio sobre cómo afecta el riesgo de cambio climático a la cotización de una empresa del sector primario.

¹ Comisión Nacional del Mercado de Valores

El clima está conformado por muchas variables, las cuales puede ser complicadas de medir numéricamente. Es por eso por lo que se debe definir una variable numérica que sea representativa del medioambiente para hacer el estudio. Se propone el cambio de la temperatura global del planeta como una variable representativa del cambio climático, ya que es fácil de medir y existen suficientes registros históricos como para que los resultados sean significativos. Los datos que se utilizarán para hacer el estudio es la serie temporal univariante de las anomalías de la temperatura global del *National Oceanic and Atmospheric Administration (NOAA)* desde el 1850 al 2024, con la cual se analizará su comportamiento.

A simple vista puede parecer obvio que las empresas que más se pueden ver afectadas por el cambio climático son las empresas del sector primario, ya que su actividad depende principalmente del medio ambiente. Por este razonamiento, la segunda parte del estudio consistirá en crear un modelo econométrico con el que se intentará explicar el comportamiento de la cotización de una empresa agrícola con la evolución del índice de referencia S&P500 y el cambio de temperatura.

2. Objetivos e hipótesis

El cambio climático es un tema que toma cada vez más importancia a nivel personal y empresarial. El objetivo general del trabajo es estudiar cómo afecta el riesgo climático al mercado financiero. Para este trabajo se considerará la variable de la temperatura como un factor representativo del riesgo climático. En primer lugar, se pretende modelizar los cambios de la temperatura y seguidamente crear un modelo que explique la evolución de la cotización de una empresa real y que tenga capacidad de predecir valores de cotización futuros. Finalmente se quiere hacer una proyección de los datos de anomalías de temperatura con el primer modelo ajustado para después predecir con el modelo econométrico creados cuáles serían los valores de las cotizaciones.

Antes de comenzar con el estudio, se plantean ciertas hipótesis que se pretenden comprobar. Los expertos llevan años advirtiendo de que la situación climática empeora año tras año. Por este motivo, se espera, por un lado, que las temperaturas hayan subido a lo largo de los últimos 50 años y, por otro lado, que haya una gran anomalía de las temperaturas.

Además, también se espera que la temperatura tenga un impacto significativo en la cotización de la empresa agrícola, puesto que su actividad está directamente relacionada e influenciada por la temperatura. Cambios bruscos de la temperatura pueden provocar que el cultivo se vea influenciado negativamente y la empresa tenga menos beneficios y que, a causa de esto, el valor de cotización de la empresa sea menos estable. También se espera que el modelo creado puede tener cierta capacidad a la hora de predecir la tendencia que tendrán las cotizaciones si no ocurre alguna circunstancia inesperada como la pandemia o una guerra, que haga que los mercados se desplomen de manera generalizada.

3. Metodología

El trabajo se divide en tres procesos de estudio: la primera parte consiste en el estudio del comportamiento histórico de la anomalía de la temperatura, la segunda parte consiste en crear un modelo econométrico que consiga explicar cómo se ha comportado la cotización de una empresa y la tercera parte consiste en hacer una proyección del modelo creado anteriormente. Para realizar el estudio se utilizará como herramienta de cálculo estadístico el programa de RStudio.

En primer lugar, se realiza un trabajo de búsqueda de información relacionada con el estudio que se va a hacer. Con la información obtenida de otros autores, se listan los modelos propuestos para hacer la modelización de las anomalías de la temperatura. Seguidamente, se descargan los datos mensuales de las anomalías de la temperatura desde la página oficial del *National Oceanic and Atmospheric Administration*. Una vez obtenido lo anterior, se realiza el ajuste de la muestra con todos los modelos propuestos y se escoge qué modelo es el mejor según el criterio del coeficiente de determinación y de la raíz del error cuadrático medio (RECM).

En segundo lugar, se obtienen desde *Yahoo Finance* los valores de las cotizaciones de los últimos 5 años de una empresa agrícola y del índice bursátil S&P500. Se prueban varios modelos estadísticos para crear un modelo que explique mejor el comportamiento de las cotizaciones de la empresa. Una vez ejecutados los modelos se escoge el mejor modelo en base a los criterios de la Raíz del Error Cuadrático Medio y el Criterio de Información de Akaike (en adelante AIC). Por último, con el modelo escogido en el punto anterior se prueba de hacer una proyección y se compara con los datos reales para ver la capacidad de predicción del modelo.

Por último, se hará una proyección a dos años de las cotizaciones de la empresa. Para ello primero se predicen cuáles serían los valores de anomalía de temperatura con el mejor modelo ajustado en la primera parte. A continuación, con el modelo econométrico creado en la segunda parte se predicen los valores de las cotizaciones para los dos siguientes años. Además, para poder ver más claramente el efecto de las anomalías de la temperatura, se utilizan los mismos valores de los dos últimos años para la variable S&P500. De esta manera se puede hacer una comparación entre los valores proyectados y los valores de los dos últimos años.

4. Modelización de la anomalía de la temperatura

El clima es una variable difícil de medir, ya que comprende muchos factores. Por ello es esencial definir una variable climática numérica para poder medir el clima. Es por esta razón que se ha considerado la temperatura como una variable representativa del clima. El clima se puede representar de la siguiente manera (Mudelsee M, 2010):

$$Y(T) = Y_{tendencia}(T) + S(T) * Y_{ruido}(T)$$

Donde:

- Y es la variable que describe el clima
- T es la variable tiempo
- $Y_{tendencia}(T)$ es la tendencia de la variable climática
- $S(T)$ es la desviación estándar
- $Y_{ruido}(i)$ es el ruido o la variabilidad que no puede ser explicada por el modelo.

El componente $Y_{ruido}(T)$ es una variable de media 0 y varianza 1 y su distribución se define con la variable $S(T)$, la cual representa la variabilidad de $Y_{ruido}(T)$. Es la variable $Y_{tendencia}$ la que da a la variable $Y(T)$ unidades físicas. De aquí en adelante se define con el símbolo ε a los componentes $S(T) * Y_{ruido}(T)$.

Mediante la regresión se intenta estimar la tendencia de la variable climática, es decir, la variable $Y_{tendencia}$. Para ello se utilizarán los modelos de regresión lineal, regresión cuadrática y regresión segmentada.

4.1.Datos y descripción de los datos

Para la modelización del aumento de la temperatura se parte de los datos del NOAA. Los datos representan la anomalía de la temperatura terrestre y del océano global, la cual se calcula como la diferencia de la temperatura respecto a la media del siglo XX. La muestra se compone de 2090 observaciones mensuales desde enero de 1850 a febrero de 2024.

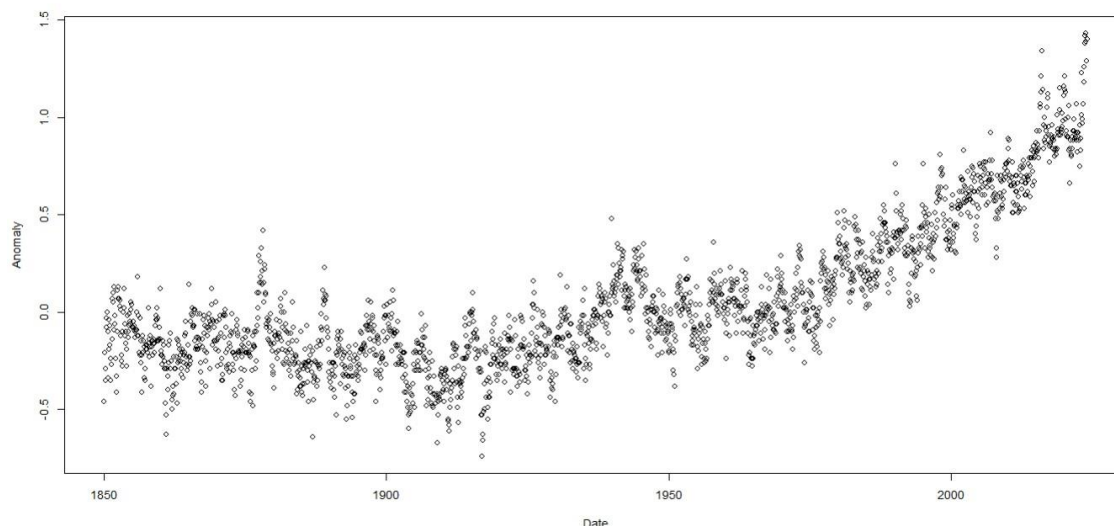


Ilustración 1: Serie temporal de las anomalías de la temperatura terrestre y del océano. Elaboración propia. Datos obtenidos del NOAA.

En la ilustración 1 se observa que la anomalía de la temperatura disminuye ligeramente desde el 1850 hasta aproximadamente 1920, donde se llega al mínimo. Cabe destacar que hay investigaciones que a principios de la primera guerra mundial hubo una anomalía climática (1914 – 1919) que causó que las temperaturas fueran más frías de lo normal y hubiera lluvias torrenciales. Este suceso se puede apreciar en el gráfico.

A partir de 1920 las temperaturas tienen un aumento abrupto, teniendo un máximo en 1940, cuando se batieron récords de temperaturas altas registradas. Del 1940 hasta alrededor de 1950 la temperatura se estabiliza para luego aumentar continuamente hasta la actualidad.

Se aprecia que en términos generales la muestra tiene una tendencia positiva. Con una aceleración en aumento a partir del 1950 aproximadamente y 1920-1940.

Tabla 1: Descripción de la serie de anomalías de la temperatura. Elaboración propia.

	Anomalías temperatura
Mínimo	-0.74
Máximo	1.43
Media	0.05
Desviación estándar	0.37

La anomalía de la temperatura tiene el mínimo absoluto de -0.74 en diciembre de 1916 y el máximo absoluto de 1.43 el noviembre de 2023. La media de las anomalías es de 0.05, de lo que se puede deducir que la temperatura ha aumentado respecto a la temperatura media del siglo XX.

4.2. Estudio de la normalidad

Antes de empezar con la modelización se realiza un estudio de normalidad de la muestra. Se realiza un QQ-Plot (ilustración 2) en el que se puede concluir que la

muestra no sigue una distribución normal. Se aprecia que la muestra se aleja significativamente de la normalidad en los primeros y últimos cuantiles.

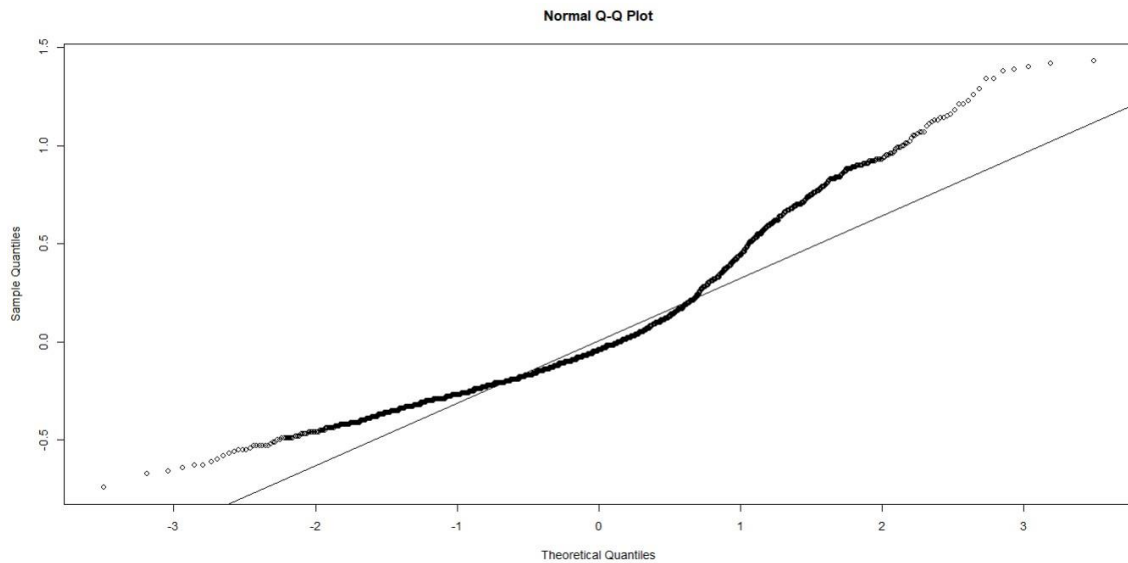


Ilustración 2: Q-Q Plot de la serie de anomalías de la temperatura. Elaboración propia.

Adicionalmente se realizan las pruebas de normalidad para comprobar matemáticamente que la muestra no sigue una distribución normal. Los tests de Pearson chi-cuadrado y Shapiro – Wilk son dos pruebas que miden la normalidad de la muestra. En ambas pruebas se fija como hipótesis nula que la muestra sigue una distribución normal. En ambos casos se ha considerado un nivel de significación del 5%. En ambos casos el p-valor es menor a 0.05, por lo que, en ambos tests, se rechazan las hipótesis nulas de normalidad. Por lo tanto, se concluye que las anomalías de la temperatura no siguen una distribución normal.

4.3.Regresión lineal

En primer lugar, se utiliza la regresión lineal como un primer intento de modelización de la muestra. El modelo de regresión lineal relaciona una variable dependiente (Y) con una serie de variables independientes o regresoras (X), de tal manera que las variables regresoras expliquen el comportamiento de la variable dependiente. La parte de comportamiento que no es posible medir con las variables independientes se representa con la variable aleatoria ε .

$$Y(i) = \beta_0 + \beta_1 * X(i) + \varepsilon$$

Donde

- Y es la variable dependiente
- β_0 indica la intercepción cuando la variable independiente es 0
- β_1 indica la tendencia de la regresión
- X es la variable independiente
- ε es el error del modelo

Con la muestra obtenida se define la variable tiempo como la variable independiente X y a la anomalía de la temperatura como la variable dependiente Y. Con RStudio se estima los coeficientes beta, resultando en los coeficientes siguientes:

$$Y(i) = -0.44191 + 0.000473 * X(i) + \varepsilon$$

Donde:

Tabla 2: Coeficientes estimados de la regresión lineal. Elaboración propia.

β_0	β_1
-0.44191	0.000473

Con los resultados anteriores se deduce que la anomalía de la temperatura tiene una relación directa (β_1) con el paso del tiempo. Es decir, la temperatura media mensual respecto a la temperatura media del siglo XX va aumentando a medida que pasan los meses.

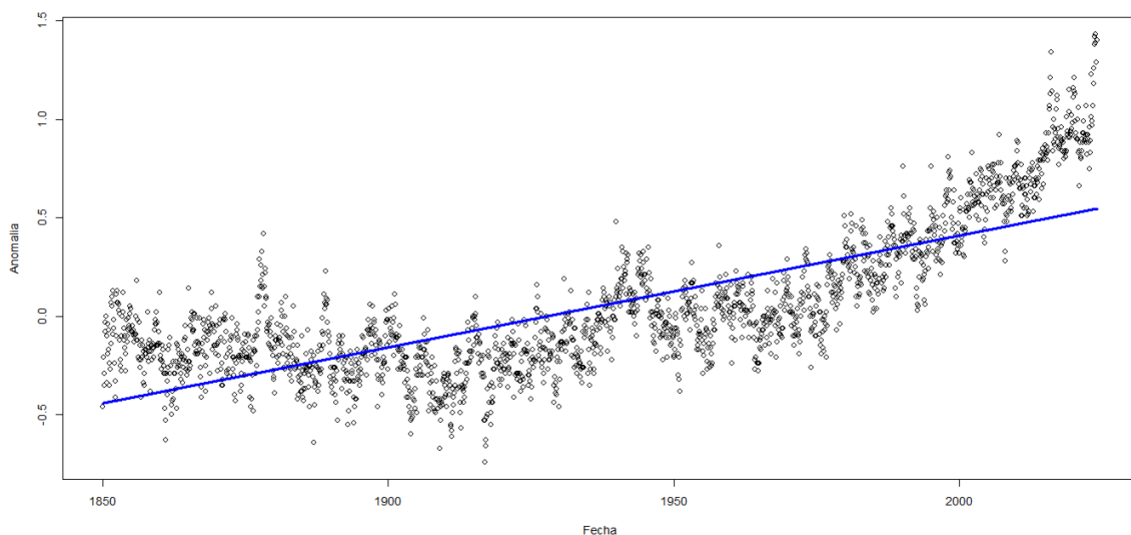


Ilustración 3: Representación del modelo lineal ajustado (azul) sobre la muestra. Elaboración propia.

Una vez obtenido la estimación de las betas se comprueba la validez del ajuste mirando la significatividad de los coeficientes y el coeficiente de determinación (R^2). El ajuste R^2 indica el porcentaje de la variable dependiente explicada por las variables independientes.

La conclusión que se obtiene es que todos los coeficientes estimados son significativos, ya que el p-valor obtenido es menor al 5% de nivel de significancia. En cuanto al coeficiente de determinación, se ha obtenido que un 60.06% de la anomalía de la temperatura viene explicada por el paso del tiempo. En principio el resultado obtenido no es satisfactorio puesto que indica que casi la mitad de la variabilidad de la dependiente no está explicada. No obstante, el ajuste R^2 no siempre es un buen indicador y a veces el ajuste da un resultado bajo cuando un modelo se ajusta bien.

Existen varias medidas que se pueden utilizar para medir la precisión de un modelo. Uno de ellos es la Raíz del Error Cuadrático Medio (RECM). Con esta medida se puede calcular la media de los errores. Al comparar varios modelos se puede concluir que el modelo que se ajusta más es el que tiene menos error, es decir, cuanto menor sea el RECM mejor será el modelo. Se calcula de la siguiente manera:

$$\text{RECM} = \sqrt{\frac{\sum_{i=1}^n (\text{predicción}_i - \text{observado}_i)^2}{n}}$$

El RECM resultante de la regresión lineal es de 0.1880761. Este resultado se puede interpretar como que la desviación media de las anomalías de la temperatura predichas por la regresión lineal y los valores observados es de 0.1880761. Este valor no indica de manera directa si es un buen modelo o no, sino que se comparará con el RECM de otros modelos para ver qué modelo se ajusta mejor.

4.4.Regresión cuadrática

Dado el gráfico de las observaciones (ilustración 1) se aprecia que los datos pueden formar una parábola. La regresión cuadrática es un modelo que ajusta los datos a una parábola. La función de la regresión cuadrática se representa de la siguiente manera:

$$Y(i) = \beta_0 + \beta_1 * X(i) + \beta_2 * X(i)^2 + \varepsilon$$

Donde Y es la variable dependiente y representa la anomalía de la temperatura y la variable independiente X es el paso del tiempo. Se realiza la estimación de los coeficientes con RStudio. La función de la parábola que mejor se ajusta la muestra es la siguiente:

$$Y(i) = -0.0517 - 0.0006461 * X(i) + 5.352 * 10^{-7} * X(i)^2 + \varepsilon$$

Donde:

Tabla 3: Valores de los coeficientes estimados del modelo cuadrático. Elaboración propia.

β_0	β_1	β_2
-0.0517	- 0.0006461	$5.352 * 10^{-7}$

De los coeficientes anteriores se puede observar que el coeficiente lineal (β_1) es negativo, lo que se interpretaría como que anomalía de la temperatura tiene una relación inversa con el tiempo. El coeficiente cuadrático (β_2) es positivo y muy cercano a 0, lo que indica que la curvatura de la parábola tiene una ligera forma de U (ilustración 3).

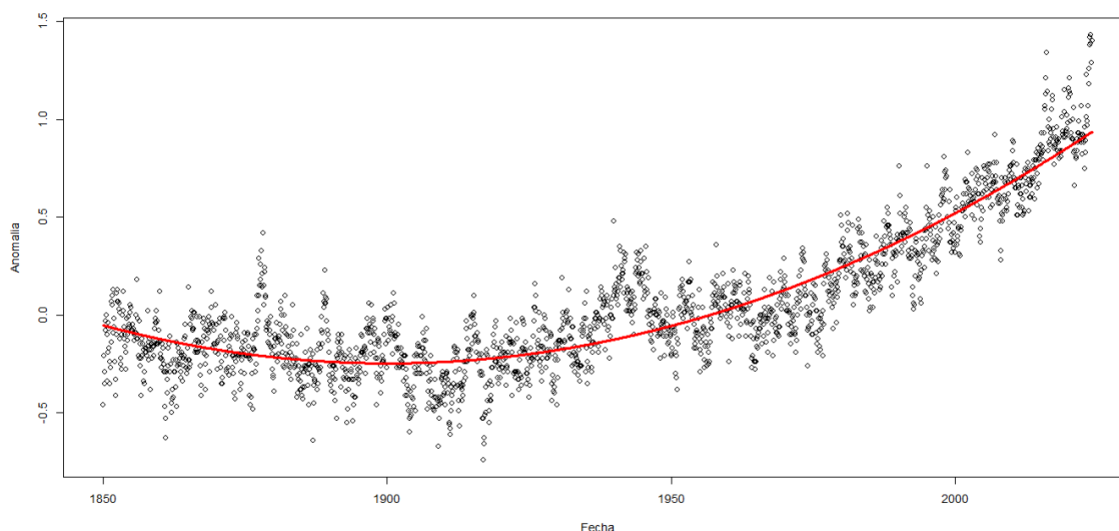


Ilustración 4: Representación de la regresión cuadrática estimada (rojo) encima de la muestra. Elaboración propia.

Al igual que en el modelo de regresión lineal, se comprueba que los coeficientes beta estimados sean significativos y se calcula el coeficiente de determinación. El resultado obtenido es que los coeficientes son significativos con un nivel de confianza del 5%. Además, el ajuste R^2 es de 0.8245, lo que indica que 82.45% de la anomalía de la temperatura viene explicada por el paso del tiempo.

El ajuste R^2 obtenido con la regresión cuadrática es mejor que con la regresión lineal. Sin embargo, como se ha explicado anteriormente, se calcula el RECM como otra medida para comparar qué modelo se ajusta mejor. El resultado del RECM es de 0.1196751, el cual es menor que el de la regresión lineal. En definitiva, los resultados obtenidos con la regresión cuadrática son mejores que los de la regresión lineal, con lo que se puede concluir que la regresión cuadrática se ajusta mejor a la muestra que la regresión lineal. No obstante, el ajuste aún es mejorable.

4.5.Regresión segmentada lineal

En los apartados anteriores ya se ha podido comprobar que la anomalía de la temperatura no sigue un modelo lineal. Esto se puede deber a que la temperatura presenta cambios bruscos repentinos que con modelos lineales no se pueden ajustar. Una alternativa a la regresión lineal es el modelo de regresión segmentada lineal (Mudelsee, 2009). Este modelo es una extensión del modelo lineal, con el cual se puede captar los cambios bruscos de la tendencia. De tal manera que el modelo lo conforman dos o más relaciones lineales distintas. A los puntos donde hay un cambio de tendencia se les llama puntos de ruptura. Por lo tanto, lo primordial de este modelo es la identificación de los puntos de ruptura.

En la ilustración 1 se pueden apreciar visualmente que hay varios puntos donde la tendencia cambia. Del 1850 al 1920 la tendencia es ligeramente negativa, del 1920 al 1940 la tendencia es claramente positiva, a partir de entonces se estabiliza hasta aproximadamente el 1970 que vuelve a tener una tendencia positiva significativa. Es decir, en total existen tres puntos de cambio de tendencia claros. Por este motivo se

realizan varias pruebas con diferentes modelos con uno, dos y tres puntos de ruptura para ver qué modelo se ajusta más.

Se realizan varias pruebas con diferentes puntos de rupturas resumidas en la tabla siguiente:

Tabla 4: Resultados del ajuste R^2 y RECM. Elaboración propia.

Puntos de ruptura estimados	Coefficiente R^2	RECM
1969-01	81.51	0.123744
1909-04, 1977-10	84.20	0.113769
1909-04, 1977-11	84.20	0.113770
1918-03, 1941-10, 1970-12	85.52	0.109642
1909-05, 1975-10, 2011-11	84.52	0.113275

Los resultados que se obtienen indican que el modelo segmentado que mejor se ajusta es el que tiene 3 puntos de ruptura en marzo de 1918, octubre 1941 y diciembre de 1970, ya que es la regresión que tiene mayor coeficiente R^2 y menor RECM. Además, también se puede ver que con menos puntos de ruptura se consiguen peores resultados. Esto es debido a que con menos puntos de ruptura el modelo no se ajusta a los cambios bruscos de tendencia.

5. Modelo econométrico

Se pretende crear un modelo econométrico en el que la cotización de una empresa agrícola sea explicada por el índice de referencia S&P500 y la serie de anomalías de la temperatura modelizada anteriormente. Con este modelo se investigará qué tanto ha podido afectar la temperatura en la cotización. La temperatura puede afectar a las operaciones de las empresas y, como consecuencia, al valor de las acciones.

Se obtienen datos de cotización de los últimos 5 años de una empresa real a través de *Yahoo Finance*. Esta empresa se dedica a la actividad de agricultura, por lo que es lógico pensar que los cambios en la temperatura afecten directamente a los resultados de la empresa y su cotización. Además, esta empresa forma parte del índice S&P500, por lo tanto, también se utilizarán los valores del índice de los últimos 5 años para explicar el comportamiento de las cotizaciones.

De tal manera que la variable dependiente a explicar son los valores de cotización de la empresa agrícola y las variables independientes que intentan explicar el comportamiento de la dependiente son el índice S&P500 y la serie de anomalías de la temperatura modelizado anteriormente.

5.1.Descripción de los datos

En los gráficos siguientes se presentan de manera visual los datos de cotización de la empresa (izquierda) y los valores del S&P500 (derecha) para las mismas fechas. Ambas series de datos van de febrero de 2019 a febrero de 2024.

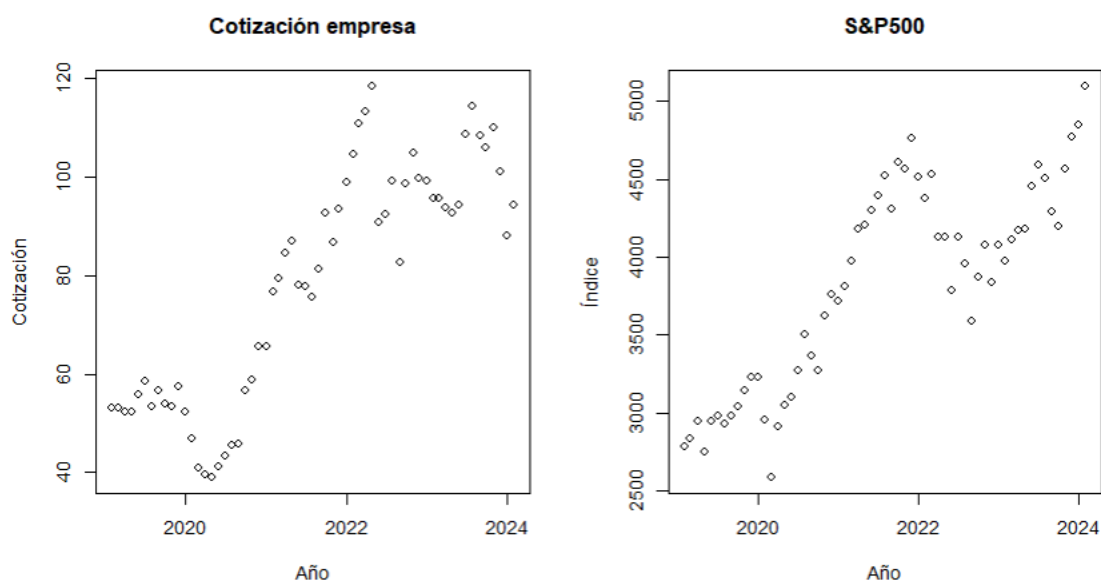


Ilustración 5: Representación gráfica de los valores de cotización de una empresa real (izquierda) y de los valores del índice S&P500. Elaboración propia.

Se puede observar que ambas series tienen una tendencia general al alza. En el 2020 ambas series sufren una bajada, siendo el mínimo absoluto, a causa de la pandemia del COVID. El mayor impacto de la pandemia fue en marzo de 2020, cuando muchos países decidieron aplicar el confinamiento como medida para frenar la pandemia. A partir de entonces los valores suben de manera continuada hasta principios del 2022, cuando el inicio del conflicto Rusia-Ucrania provoca un impacto negativo en el sector financiero por la incertidumbre.

Cabe destacar que el efecto adverso de la pandemia tuvo un impacto más claro en las cotizaciones de la empresa. Es decir, a esta empresa en concreto le costó más volver a subir sus valores de cotización. Por otro lado, el efecto negativo del conflicto Rusia-Ucrania parece que afectó más al índice S&P500, puesto que en ese año la caída que sufre es más grande. No obstante, el índice S&P500 tiene una tendencia mucho más clara.

5.2.Modelización

A partir de los datos mostrados anteriormente se crea un modelo donde la variable dependiente es los valores de cotización de la empresa y las independientes son el índice S&P500 y las anomalías de la temperatura. Se prueban algunos de los modelos que se suelen utilizar a nivel económico. (William H., 2002) cubre una amplia gama de modelos econométricos.

5.2.1. Modelo lineal múltiple

Anteriormente se ha utilizado el modelo lineal simple para modelizar las anomalías de la temperatura. En este caso, hay dos variables independientes:

$$Y(i) = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \varepsilon$$

Donde:

- Y es la variable dependiente de cotización de la empresa
- β_0 es el coeficiente de intercepción cuando las X son 0
- β_1 y β_2 son los coeficientes de regresión de X_1 y X_2 respectivamente
- X_1 es la variable independiente S&P500
- X_2 es la variable independiente de anomalía de la temperatura
- ε es el error del modelo

Los coeficientes resultantes del modelo son:

Tabla 5: Resultados de los coeficientes del modelo lineal múltiple. Elaboración propia.

β_0	β_1	β_2
-36.82	0.030	-0.95

Se hace un test para ver si los coeficientes son significativos y se obtiene que el p-valor de las anomalías de la temperatura (β_2) de 0.9239. Esto se interpreta como que, bajo un nivel de significatividad del 5%, las anomalías de la temperatura no son un factor influyente en el valor de las cotizaciones. El resto de variables son claramente significativas para el modelo.

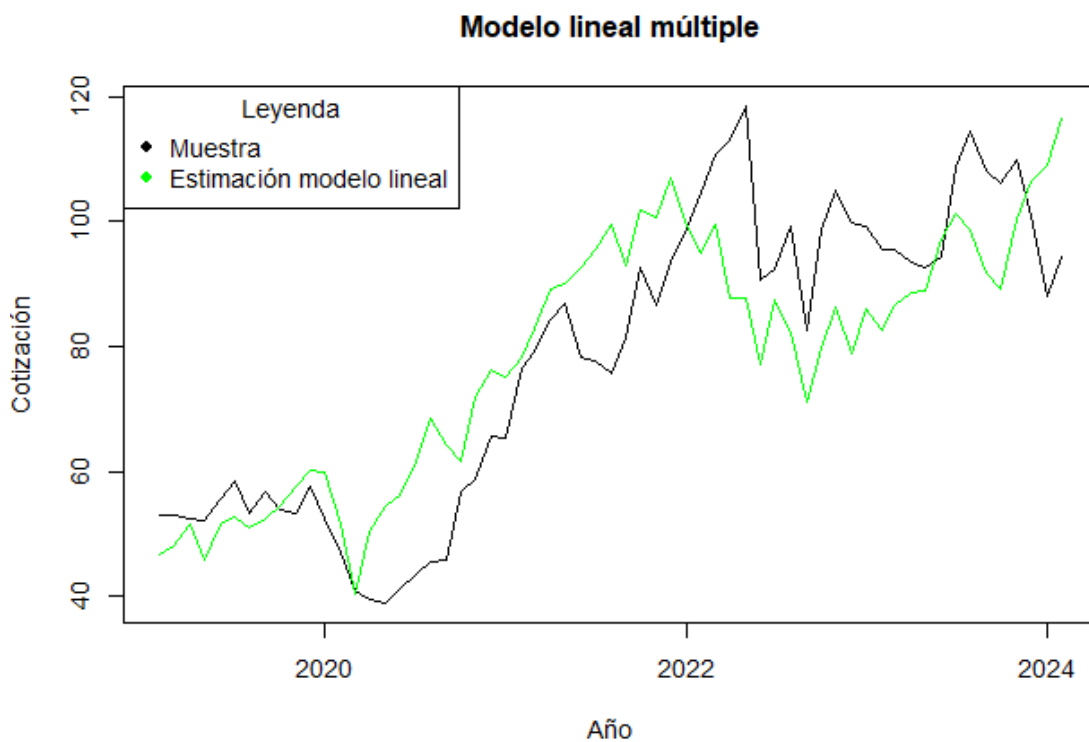


Ilustración 6: Gráfico de la evolución del valor de las cotizaciones siguiendo el modelo lineal múltiple. Elaboración propia.

En el anterior gráfico (ilustración 6) se puede apreciar que el modelo no consigue estimar bien los valores de la muestra, sobre todo los máximos y mínimos. A simple vista, ya se puede ver que este modelo no se ajusta bien. Además, el ajuste R^2 es de 69.87%, lo que indica que el modelo no se ajusta bien a los datos.

5.2.2. Modelo Gamma con función enlace identidad

Otro modelo lineal es utilizar la distribución Gamma con la función de enlace identidad, es decir, sin transformación de la variable dependiente. El modelo Gamma puede ser adecuado para el modelo porque es un modelo útil para series con variables positivas y que presentan asimetría, como es el caso de los valores de cotización.

Al tener una función de enlace identidad, la función del modelo tiene la misma forma que la regresión lineal múltiple, con la diferencia de que en la regresión lineal múltiple la variable dependiente se modeló como una combinación lineal de las variables independientes, mientras que en el modelo Gamma la variable dependiente sigue una distribución gamma. Además, los errores también se distribuyen siguiendo una distribución gamma.

Los coeficientes resultantes del modelo son:

Tabla 6: Resultados de los coeficientes del modelo Gamma. Elaboración propia.

β_0	β_1	β_2
-38.71	0.031	-2.84

A diferencia del modelo lineal, en este caso el p-valor de la variable de anomalías de la temperatura (β_2) es de 0.78005 y, por lo tanto, no se considera una variable significativa para el modelo. El resto de variables son claramente significativas para el modelo.

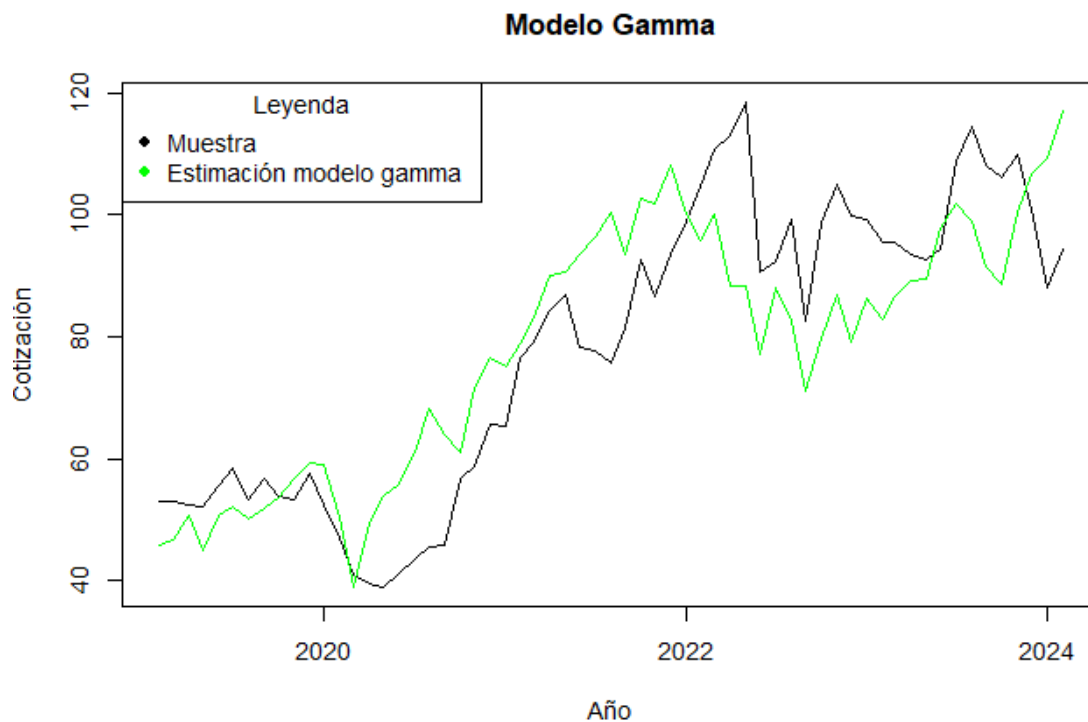


Ilustración 7: Representación gráfica de los valores de cotización estimados por el modelo Gamma. Elaboración propia.

En la ilustración 7 se puede apreciar que el gráfico se parece mucho al del modelo lineal. Al igual que en el modelo lineal el modelo no consigue estimar bien los valores reales de cotización.

5.2.3. Modelo polinómico

El modelo polinómico es un modelo de regresión no lineal. Esto permite que haya más flexibilidad para ajustarse al modelo cuando haya volatilidad. Es decir, cuando haya muchos cambios de la tendencia, de manera que un modelo lineal no sería capaz de representar.

Anteriormente se ha visto que el modelo cuadrático se ha ajustado a la serie de anomalías de la temperatura mejor que el modelo lineal, ya que la regresión lineal no era capaz de modelizar la curvatura de la serie. El modelo cuadrático es un caso específico del modelo polinómico, más concretamente cuando el modelo es de grado 2. La forma del modelo polinómico en este caso es:

$$Y(i) = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \beta_3 * X_{12} + \beta_4 * X_{22} + \beta_5 * X_1 * X_2$$

Donde:

- β_0 es el coeficiente de intercepción
- β_1 y β_2 son los coeficientes lineales de X_1 y X_2 respectivamente
- β_3 y β_4 son los coeficientes cuadráticos de X_1 y X_2 respectivamente
- β_5 es con coeficiente de interacción entre X_1 y X_2

Los coeficientes estimados con el modelo polinómico están resumidos en la siguiente tabla:

Tabla 7: Coeficientes estimados del modelo polinómico. Elaboración propia.

Coeficiente	Valor estimado
β_0	$-1.767*10^2$
β_1	$1.282*10^{(-1)}$
β_2	$-8.246*10^1$
β_3	$-1.289*10^{(-5)}$
β_4	$4.934 *10^1$
β_5	$-2.566*10^{(-3)}$

Se comprueba que los coeficientes sean significativos en el modelo con un nivel del 5%. Los coeficientes β_1 y β_3 son significativos, mientras que el resto de coeficientes no lo son. El coeficiente β_0 tiene un p-valor de 0.09691, lo cual no le hace significativos con un nivel del 5%, pero sí que es significativo con un nivel de significatividad del 10%. Coincide que los coeficientes que no son significativos son los coeficientes relacionados

con la variable anomalías de la temperatura. Por lo tanto, se interpreta como que el modelo considera que la variable anomalías de la temperatura no es significativa.

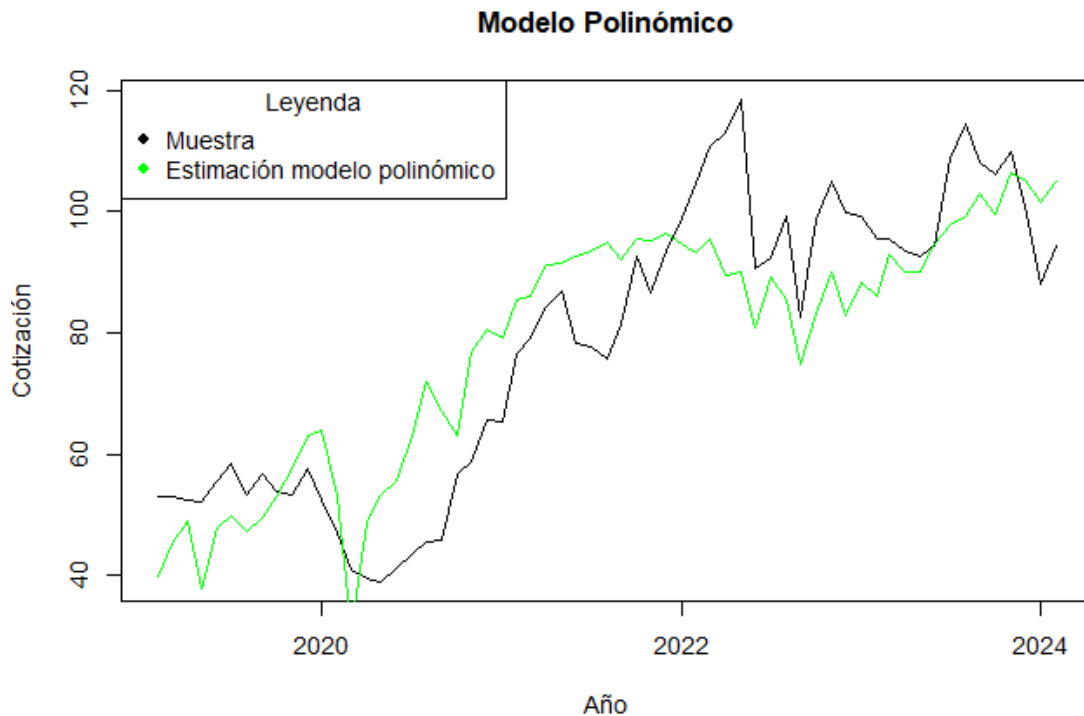


Ilustración 8: Representación gráfica del modelo polinómico estimado. Elaboración propia.

En la ilustración 8 se puede ver que el modelo polinómico parece ajustar mejor los valores más recientes que el modelo lineal o el Gamma. No obstante, el modelo polinómico no consigue estimar bien el pico que hay en el año 2022. A pesar de esto, el ajuste R^2 del modelo polinómico es de 73.77%, lo cual lo convierte en una mejor opción al modelo lineal.

5.2.4. ARIMA (p, d, q)

El modelo ARIMA es útil para el análisis y predicción de series temporales. El modelo tiene tres componentes: autorregresivo (AR), integrado (I) y media móvil (MA). Para identificar qué modelo ARIMA se ajusta mejor a los datos se deben encontrar los siguientes datos:

- El valor **p** indica el orden del proceso autorregresivo
- El valor **d** indica cuantas veces se debe diferenciar la serie para hacerla estacionaria
- El valor **q** indica el número de términos de media móvil

Una herramienta para encontrar estos valores es la función `auto.arima()` de RStudio (Hyndman & Khandakar, 2008), la cual encuentra los valores óptimos minimizando el AIC. El resultado que da esta función como el mejor modelo es el ARIMA (0,1,0). Esto quiere decir que el componente p es 0, el componente d es 1 y el componente q es 0.

Esto se interpreta como que la serie temporal es un proceso puramente aleatorio, ya que la serie se modela como una diferencia de primer orden de un proceso de ruido blanco. Los coeficientes resultantes del modelo son de -4.61 para la variable anomalía de la temperatura y de 0.015 para la variable S&P500.

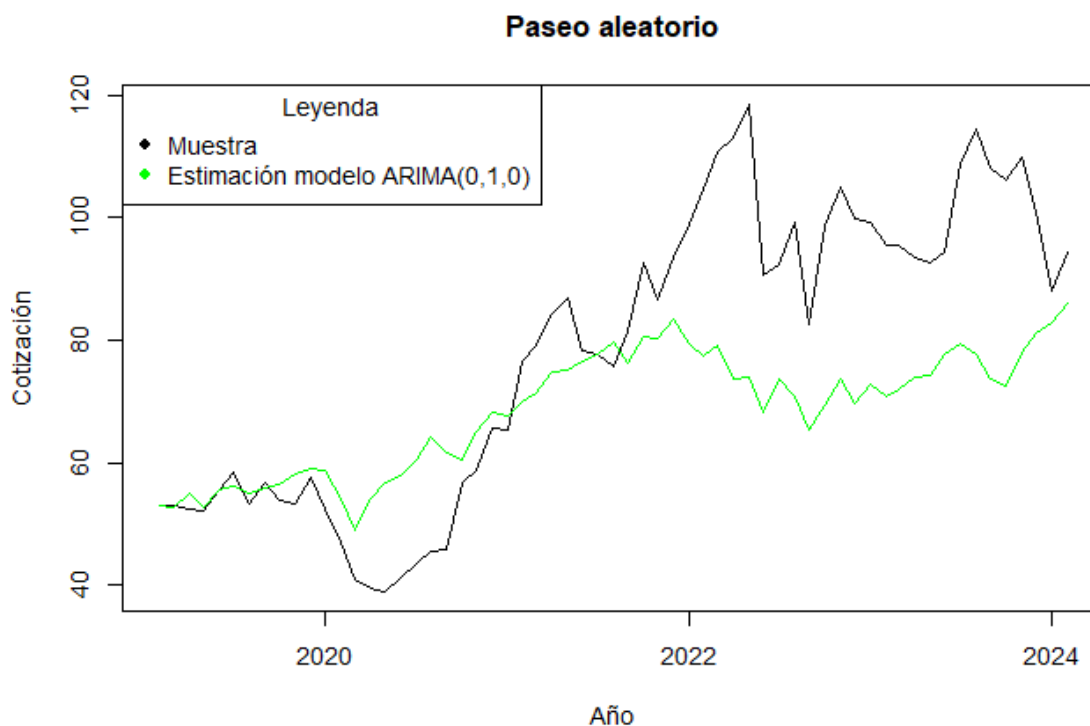


Ilustración 9: Representación gráfica del Paseo aleatorio. Elaboración propia.

En la ilustración 9 está representado en verde el modelo ARIMA (0,1,0) sobre las observaciones de la serie temporal. Se puede apreciar que este modelo no se ajusta correctamente, por lo que se puede esperar un alto valor del RCEM.

5.3.Resultados

El AIC es un indicador que se utiliza para comparar modelos estadísticos (Shumway & Stoffer, 2017). A diferencia de otros indicadores, el AIC tiene en cuenta tanto la bondad de ajuste del modelo como la dificultad de este. Es decir, el AIC evalúa el modelo para que se ajuste lo mejor posible, pero minimizando el número de parámetros del modelo. La fórmula del AIC es:

$$AIC = 2 * k - 2 * \ln(L)$$

Donde:

- k es el número de parámetros
- L es el valor maximizado de la función de verosimilitud

Por lo tanto, se busca que el valor del AIC sea lo menor posible. No obstante, si los resultados del AIC de varios modelos son muy similares, es conveniente utilizar otro indicador para decidir qué modelo es mejor. Por este motivo, también se calcula el RECM de los modelos.

Tabla 8: Tabla resumen del RECM y AIC de los diferentes modelos. Elaboración propia.

Modelo	RECM	AIC
Lineal múltiple	10.74	493.10
Gamma	10.84	487.30
Polinómico	10.32	490.66
ARIMA (0,1,0)	14.68	398.69

Siguiendo el criterio del RECM, el mejor modelo es el polinómico y, seguidamente, el modelo lineal y el Gamma, puesto que tienen un menor error. Por otro lado, según el indicador AIC, el mejor modelo es el ARIMA y, seguidamente el modelo Gamma.

El modelo lineal y el modelo Gamma tienen un RECM muy parecido. Por ello, se sigue el criterio de AIC para concluir que el modelo Gamma es mejor que el modelo lineal. El modelo polinómico tiene el mejor resultado de RECM y un AIC ligeramente peor que el modelo Gamma, por lo que también se puede considerar como un buen modelo para esta muestra. Por último, el modelo ARIMA tiene el peor resultado de RECM y el mejor resultado de AIC. Como se ha visto anteriormente, el modelo ARIMA (0,1,0) es un modelo puramente aleatorio. Por este motivo, a pesar de su buen AIC, se descarta como un modelo capaz de ajustar la muestra.

En la ilustración 10 se puede apreciar que visualmente el gráfico es bastante parecido, sobre todo en los primeros años. Como es difícil justificar cual es el mejor modelo entre el Gamma y el polinómico se trabajará con ambos modelos para ver la comparación entre un modelo y otro.

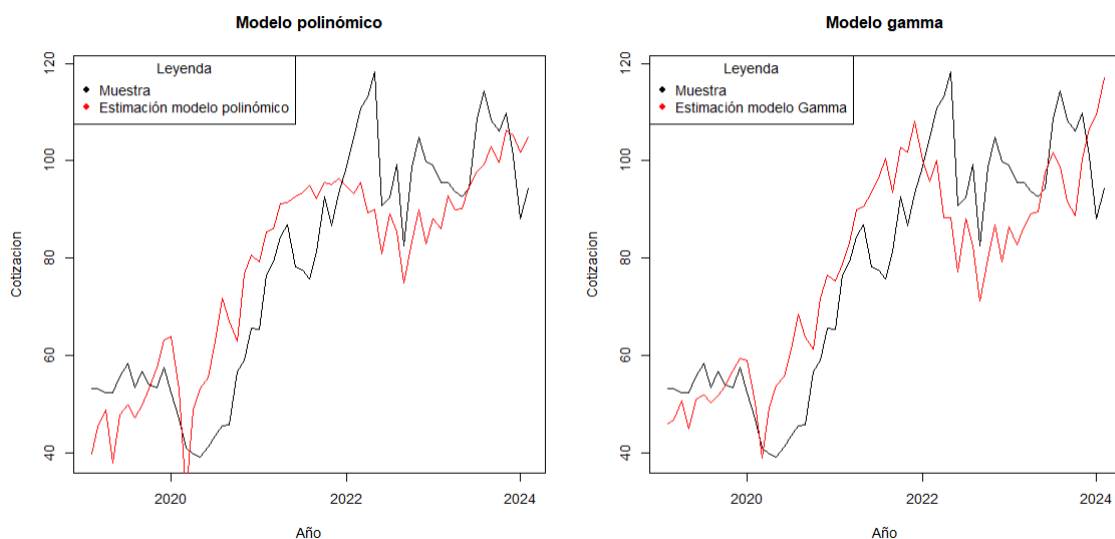


Ilustración 10: Gráfico del modelo polinómico (izquierda) y del modelo Gamma (derecha). Elaboración propia.

5.4.Capacidad de predicción del modelo

Se realiza una predicción a un año utilizando tanto el modelo polinómico como el Gamma. Para ello se divide la muestra en dos partes: una primera parte llamada *data_train* con los datos de los 4 primeros años y una segunda parte llamada *data_test* el cual abarca los datos del último año. Con los datos *data_train* se crea el modelo, primero para el modelo polinómico y luego para el Gamma. A continuación, se hace una predicción del modelo a un año. Por último, se hace una comparación de los datos predichos con los datos reales *data_test*.

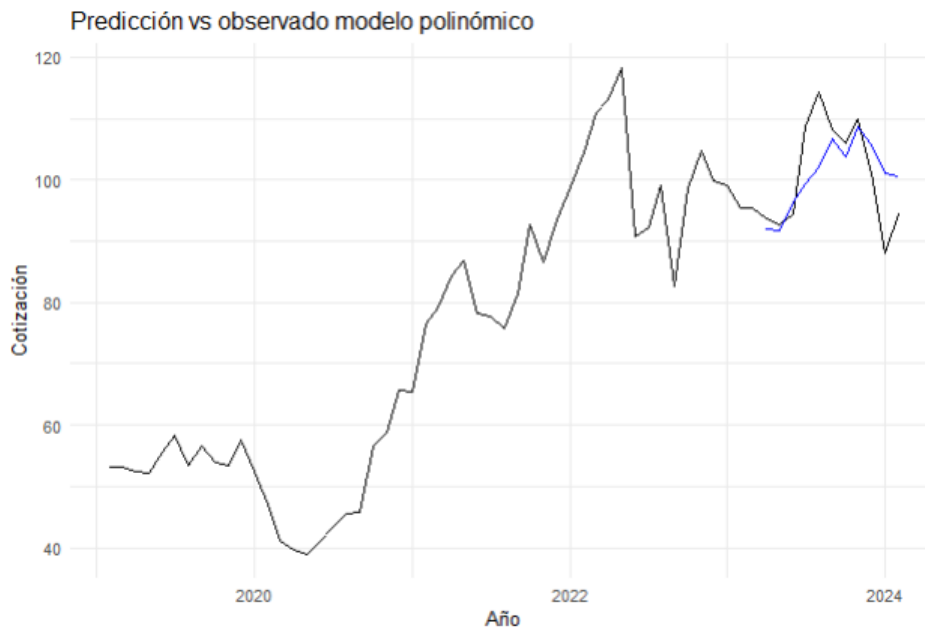


Ilustración 11: Representación gráfica de la predicción del modelo polinómico (azul) sobre los datos observados (gris). Elaboración propia.

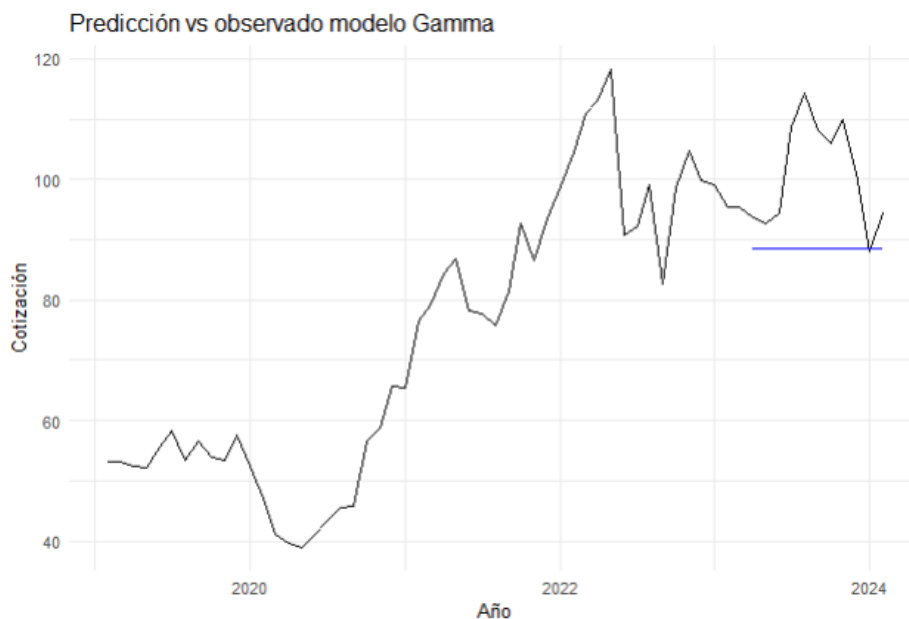


Ilustración 12: Representación gráfica de la predicción del modelo gamma (azul) sobre los datos observados (gris). Elaboración propia.

Por un lado, la predicción realizada con el modelo polinómico (ilustración 11) muestra que se ajusta relativamente bien, ya que puede predecir las subidas y bajadas de las cotizaciones, aunque no de manera exacta. Por otro lado, la predicción realizada con el modelo Gamma muestra que no es capaz de predecir subidas y bajadas, ya que es un modelo lineal. A pesar de esto, sí que consigue predecir mejor la tendencia global de los valores. Por lo tanto, se puede concluir que el modelo polinómico tiene una mayor capacidad que el modelo Gamma a la hora de predecir los valores de cotización de la empresa.

6. Proyección de las cotizaciones

Finalmente, se procede a hacer una proyección a dos años del valor de las cotizaciones de la empresa con los modelos que se ajustan mejor: el modelo polinómico y el modelo gamma. Para ello se necesitan saber los valores proyectados de las variables independientes, que son las anomalías de la temperatura y el S&P500.

No obstante, para la variable S&P500 se utilizarán los valores de los último dos años, es decir, de tal a tal, con el objetivo de poder apreciar el impacto de los cambios de temperatura en las cotizaciones de la empresa. En el apartado 4 se ha modelizado las anomalías de la temperatura, de tal manera que se puede hacer una proyección utilizando el modelo escogido en el apartado 4.5.

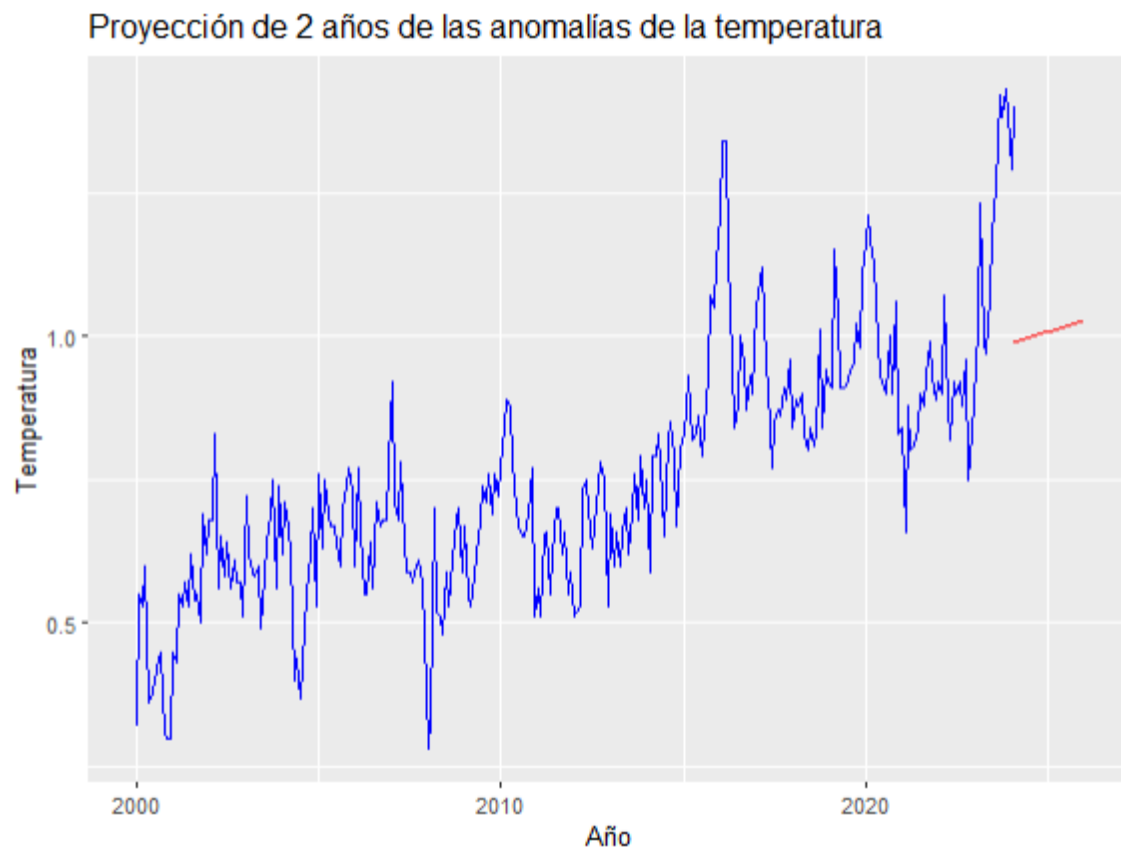


Ilustración 13: Proyección a dos años de las anomalías de la temperatura. Elaboración propia.

La ilustración 13 muestra en azul los datos de anomalía de la temperatura a partir de enero del 2000 hasta febrero de 2024 y en rojo la proyección a dos años con el modelo segmentado lineal modelizado en el apartado 4.5. Se prevé que la tendencia de las anomalías de la temperatura siga aumentando.

Una vez obtenidas las proyecciones de las anomalías de la temperatura se aplica el modelo polinómico y el modelo gamma ajustados en el apartado 5.2 para obtener los nuevos valores de cotización.

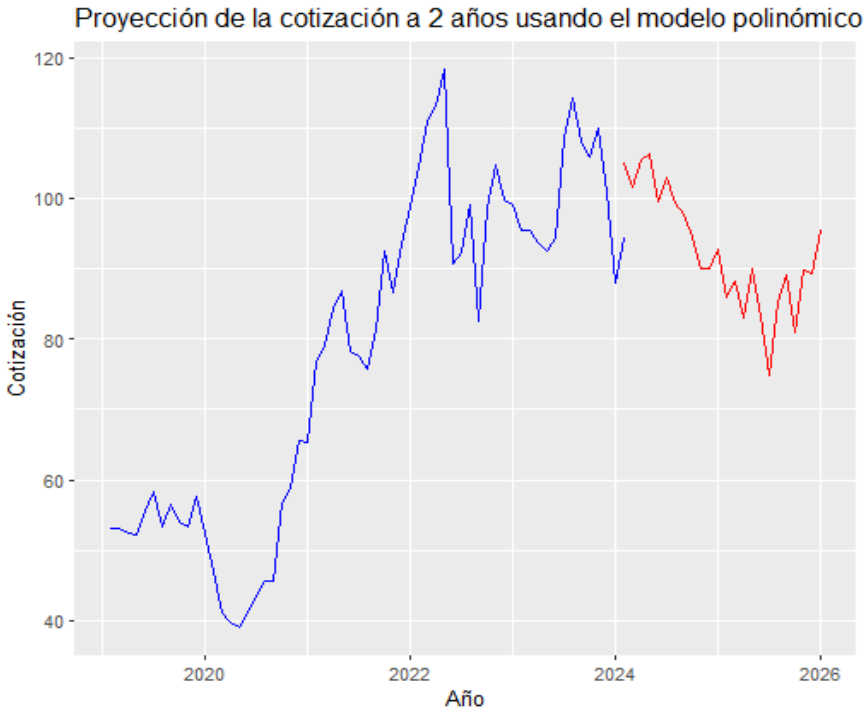


Ilustración 14: Proyección de los valores de cotización a dos años utilizando el modelo polinómico. Elaboración propia.

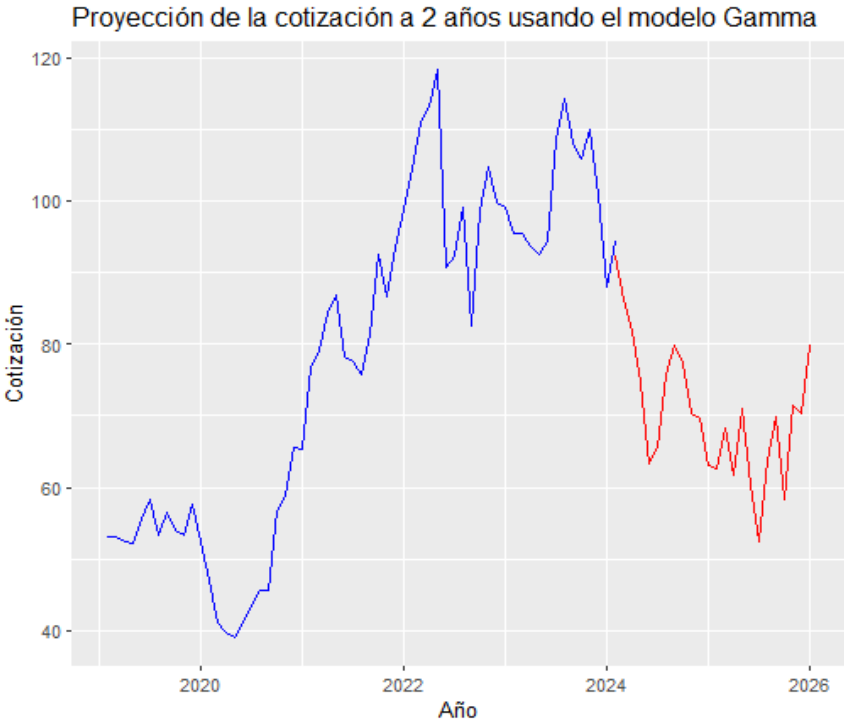


Ilustración 15: Proyección de los valores de cotización a dos años utilizando el modelo Gamma. Elaboración propia.

Se puede ver que el modelo Gamma (ilustración 15) predice unos valores mucho menores que el modelo polinómico (ilustración 14). Para comparar ambos modelos se calculan varios estadísticos. En la tabla 9 muestra los principales estadísticos del modelo polinómico y gamma para los valores de las cotizaciones proyectados.

Tabla 9: Principales estadísticos de las proyecciones de las cotizaciones con el modelo polinómico y gamma. Elaboración propia.

	Modelo polinómico	Modelo Gamma
Mínimo	74.87	52.46
Máximo	106.26	92.39
Media	92.52	70.41
Desviación estándar	8.38	9.37

Se puede ver que la principal diferencia entre ambos modelos es que el modelo Gamma es más pesimista y predice valores de cotización más bajos que el modelo polinómico. Esto también se refleja en la media, la cual la diferencia entre ambos modelos es de más de 20. Otro dato a destacar es que la desviación estándar del modelo Gamma es más alta que en el modelo polinómico. A continuación, se calculan los mismos estadísticos para los dos últimos años de los valores reales.

Tabla 10: Principales estadísticos de las cotizaciones de los últimos dos años. Elaboración propia.

	Cotización
Mínimo	82.57
Máximo	118.32
Media	100.48
Desviación estándar	9.13

Si se comparan los estadísticos de las proyecciones (tabla 9) con los estadísticos de los dos últimos años de los valores reales de las cotizaciones (tabla10) se puede apreciar que hay una gran diferencia. A excepción de la desviación estándar los demás estadísticos son mayores a los del modelo polinómico y Gamma. Teniendo en cuenta esto y que anteriormente se ha visto que el modelo polinómico tiene mayor capacidad de predicción, se puede concluir que el modelo polinómico es mejor para proyectar a futuro que el modelo Gamma.

Además, se debe tener en cuenta también que para la proyección se ha predicho que las anomalías de temperatura aumentarían. Es decir, para la proyección se han tenido en cuenta valores más altos de la variable anomalías de la temperatura. Al haber mantenido los mismos valores para la variable S&P500 se puede apreciar que un aumento de las temperaturas provoca una ligera disminución de los valores de cotización.

7. Modelización con una muestra mayor

A pesar de que los coeficientes de la variable temperatura han resultado ser no significativos, el hecho de que se haya observado un ligero impacto de las anomalías de temperatura en el valor de las cotizaciones puede significar que, la variable, aun siendo significativa, no muestra un gran impacto a causa de la pequeña muestra. Por este motivo se vuelve a realizar la modelización del modelo econométrico con una muestra más grande, que va de enero de 2006 hasta febrero de 2024.

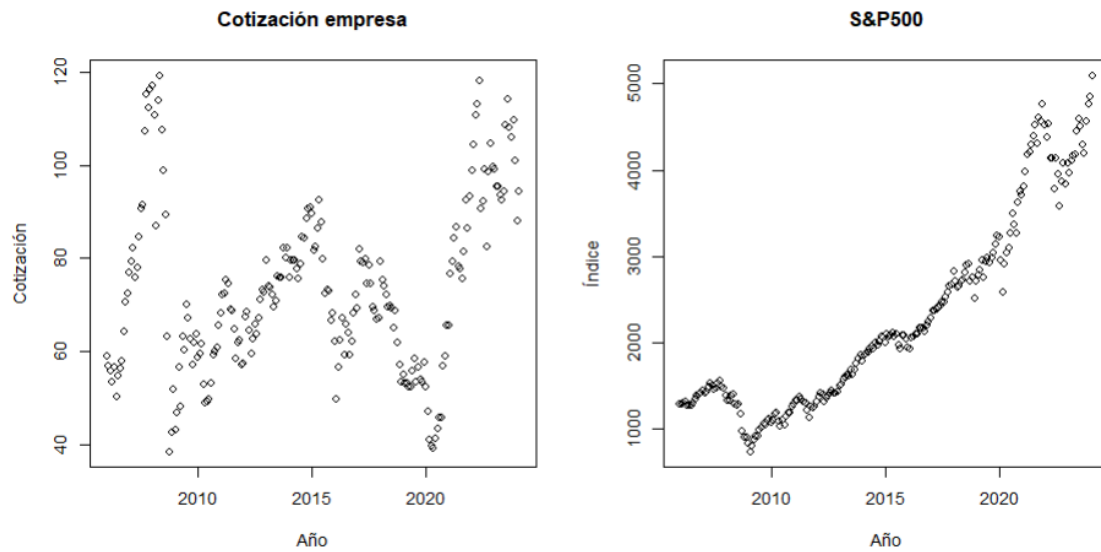


Ilustración 16: Representación gráfica de la cotización de la empresa (izquierda) y del S&P500 (derecha) para el periodo de 2006-2024. Elaboración propia.

En la ilustración 16 quedan representadas las variables de cotización de la empresa y del S&P500 para el periodo de 2006 a 2024. Se puede apreciar que los valores de S&P500 han tenido una tendencia al alza constante, mientras que el valor de las cotizaciones de la empresa ha sido más volátil durante este periodo. El procedimiento que se sigue para calcular los importes de RECM, AIC y la significatividad de los coeficientes es el mismo que se sigue en el apartado 5.2. En este caso, se ha obviado el modelo ARIMA, ya que se ha visto que no es un buen modelo para esta serie. En la tabla 11 se muestran los resultados obtenidos con el modelo lineal, el modelo Gamma y el modelo polinómico.

Tabla 11: Resumen de los resultados del RECM, AIC y significatividad. Elaboración propia

Modelo	RECM	AIC	Significatividad (5%)
Lineal múltiple	12.44	1833.23	Todas las variables son significativas
Gamma	12.41	1819.80	Todas las variables son significativas
Polinómico	12.16	1809.42	β_0 , β_2 y β_4 son significativas

Se puede apreciar que, al haber aumentado la muestra, ahora la variable de anomalía de la temperatura es significativa para todos los modelos. Por lo tanto, se puede concluir que los cambios de temperatura tienen un impacto negativo significativo en el valor de las cotizaciones de la empresa agrícola en el que, a mayor anomalía de la temperatura, menor es el valor de las cotizaciones. Por otro lado, observando la tabla también se puede concluir que, según el criterio de AIC y el RECM, el mejor modelo es el polinómico, seguido del modelo Gamma y, por último, el modelo lineal múltiple.

8. Conclusiones

La temperatura del planeta ha aumentado y, en consecuencia, las anomalías de la temperatura respecto a la media del siglo XX. Se ha apreciado un constante crecimiento de las anomalías desde el 1950. Tras hacer varias pruebas con diferentes modelos, tanto lineales como no lineales, se ha podido comprobar que la regresión lineal segmentada es un modelo con el que se ajusta bien, teniendo los mejores resultados del coeficiente R^2 y RECM.

Por un lado, se han probado tanto modelos lineales como no lineales para ajustar los datos de las anomalías de temperatura. La regresión lineal segmentada ha demostrado ser el modelo más eficaz, ya que permite modelizar los cambios bruscos de tendencia, obteniendo los mejores resultados en términos de coeficiente R^2 y RECM.

Por otro lado, se ha desarrollado un modelo econométrico para investigar si existe alguna relación significativa entre la variable anomalías de la temperatura y las cotizaciones de una empresa agrícola. A pesar de no haber ningún modelo que se ajuste claramente mejor que los demás, se concluye que los modelos que mejor pueden encajar, siguiendo los criterios del AIC y RECM son el modelo polinómico y el modelo Gama. Se ha visto que, para un periodo de 5 años, los coeficientes de la variable no resultan significativos.

No obstante, al hacer una proyección a futuro se ha visto que, dejando la variable S&P500 igual, ante un aumento de la temperatura las cotizaciones se prevén que sean ligeramente más bajas. Esto tiene especial importancia porque se ha comprobado que, para ambos modelos, el coeficiente de la variable anomalías de la temperatura (β_2) ha salido siempre negativa. Es decir, que existe una relación inversa entre los valores de cotización y las anomalías de temperatura.

Por este motivo se decide volver calcular la significatividad de los coeficientes de las anomalías de temperatura con una muestra mucho más grande. Esta vez los coeficientes de los modelos lineal y Gamma son todos significativos. Para el modelo polinómico no todos los coeficientes son significativos. No obstante, el coeficiente que va directamente ligado a la anomalía de la temperatura es significativo. Por lo tanto, se concluye que el hecho que la temperatura global del planeta esté aumentando constantemente y, que la anomalía de la temperatura respecto al siglo XX incrementa, tiene efectos negativos adversos en la cotización de la empresa agrícola.

9. Anexos

9.1. Código hecho en RStudio:

```
library(readxl)
library(ggplot2)
library(nortest)
library(fBasics)
library(univariateML)
library(fUnitRoots)
library(lmtest)
library(forecast)
library(LSTS)
library(stats)
library(lubridate)
library(dplyr)
library(tidyverse)
library(fpp2)
library(lmtest)
library(ggplot2)
library(psych)
library(segmented)
library(effects)
library(e1071)

anomalias_temperatura <- read_excel("C:/Users/sofia/Desktop/TFM/anomalias
temperatura.xlsx",
                                sheet = "Hoja1")

attach(anomalias_temperatura)
anomalias_temperatura<-cbind(anomalias_temperatura,time(Anomaly))
colnames(anomalias_temperatura)<-c("Date", "Anomaly", "Time")

## Convertir a tipo fecha
class(Date)
anomalias_temperatura$Date <- as.Date(anomalias_temperatura$Date, format =
"%d/%m/%Y")
class(anomalias_temperatura$Date)

##Descripción variables
summary(anomalias_temperatura)
#La mínima anomalía es de -0.74 y la máxima es de 1.43. La media es de 0.5262
#Año inicial enero 1850 hasta febrero 2024
sd(Anomaly)

##Representación gráfica
```

```

plot(anomalias_temperatura$Date, anomalias_temperatura$Anomaly)
hist(Anomaly)
qqnorm(anomalias_temperatura$Anomaly)
qqline(anomalias_temperatura$Anomaly)
hist(Anomaly)

##Descartar normalidad
pearson.test(Anomaly)

#considerando un criterio de significanza del 5%, se rechaza la hipotesis nula de
normalidad ya que pvalor es menor a 0.05
normalTest(Anomaly) #se rechaza la hipotesis de normalidad

##Lineal Regression
#el tiempo se considera como una variable independiente y se utiliza para explicar y
predecir la variabilidad de la variable dependiente a lo largo del tiempo
data2<-na.omit(anomalias_temperatura)
datafin<-as.data.frame(data2)
colnames(datafin)<-c("Fecha", "Anomalia", "Time")
attach(datafin)
linealmodel<-lm(Anomalia~Time)
print(linealmodel)
summary(linealmodel) #El ajuste R-squared es de 0.6 lo que dice que el 60% de la
variabilidad de la anomalia de la temperatura viene explicada con el paso del tiempo
plot(linealmodel$fitted.values, linealmodel$residuals)
qqnorm(linealmodel$residuals)
qqline(linealmodel$residuals)
shapiro.test(linealmodel$residuals) #se rechaza hipotesis de normalidad
plot(Fecha,Anomalia)
lines(Fecha,predict(linealmodel),col="blue",lwd=3)
predicciónlineal <- predict(linealmodel)
RMSElineal<-mean(sqrt((predicciónlineal-Anomalia)^2));RMSElineal #0.1880761

###Regresion cuadratica
cuadratica<-lm(Anomalia~Time+I(Time^2))
print(cuadratica)
summary(cuadratica) #El ajuste R-squared es de 0.8245 lo que dice que el 82% de la
variabilidad de la anomalia de la temperatura viene explicada con el paso del tiempo
plot(cuadratica$fitted.values, cuadratica$residuals)
qqnorm(cuadratica$residuals)
qqline(cuadratica$residuals)
shapiro.test(cuadratica$residuals) #0.4989 se acepta la hipotesis con 5%
plot(Fecha,Anomalia)
lines(Fecha,predict(cuadratica),col="red",lwd=3)
prediccióncuadratica <- predict(cuadratica)

```

```
RMSEcuadratica<-mean(sqrt((prediccióncuadratica-Anomalia)^2));RMSEcuadratica
#0.1196751
```

```
##Regresión segmentada
```

```
segmentada1<-segmented(linealmodel,seg.Z = ~ Time)
```

```
summary(segmentada1) #81.51
```

```
predicción1 <- predict(segmentada1)
```

```
RMSE1<-mean(sqrt((predicción1-Anomalia)^2));RMSE1 #0.1237441
```

```
segmentada2<-segmented(linealmodel,seg.Z = ~ Time,psi = c(804,1715) )
```

```
summary(segmentada2) #84.2
```

```
lines(Fecha,predict(segmentada2),col="green",lwd=3)
```

```
predicción2 <- predict(segmentada2)
```

```
RMSE2<-mean(sqrt((predicción2-Anomalia)^2));RMSE2 #0.1137693
```

```
segmentada2.2<-segmented(linealmodel,seg.Z = ~ Time, psi = c(804,1080) )
```

```
summary(segmentada2.2) #84.2
```

```
lines(Fecha,predict(segmentada2.2),col="green",lwd=3)
```

```
predicción2.2 <- predict(segmentada2.2)
```

```
RMSE2.2<-mean(sqrt((predicción2.2-Anomalia)^2));RMSE2.2 #0.1137705
```

```
segmentada3<-segmented(linealmodel,seg.Z = ~ Time,psi = c(804,1080,1715) )
```

```
summary(segmentada3) #85.35
```

```
lines(Fecha,predict(segmentada3),col="green",lwd=3)
```

```
predicción3 <- predict(segmentada3)
```

```
RMSE3<-mean(sqrt((predicción3-Anomalia)^2));RMSE3 #0.1096422
```

```
segmentada3.2<-segmented(linealmodel,seg.Z = ~ Time,psi = c(522,1044,1566) )
```

```
summary(segmentada3.2) #84.52
```

```
lines(Fecha,predict(segmentada3.2),col="green",lwd=3)
```

```
predicción3.2 <- predict(segmentada3.2)
```

```
RMSE3.2<-mean(sqrt((predicción3.2-Anomalia)^2));RMSE3.2 #0.113275
```

```
plot(Fecha,Anomalia)
```

```
lines(Fecha,predict(segmentada3),col="green",lwd=3)
```

```
lines(Fecha,predict(cuadratica),col="red",lwd=3)
```

```
lines(Fecha,predict(linealmodel),col="blue",lwd=3)
```

```
#####
#####
```

```
SP500 <- read_excel("C:/Users/sofia/Desktop/TFM/SP500.xlsx",sheet = "5Y")
```

```
Cotizacion <- read_excel("C:/Users/sofia/Desktop/TFM/Cotización.xlsx",sheet = "5Y")
```

```
par(mfrow = c(1, 2))
```

```
plot(Cotizacion, xlab="Año",ylab="Cotización",main="Cotización empresa")
```

```

plot(SP500, xlab="Año",ylab="Índice",main="S&P500")

summary(Cotizacion)
sd(Cotizacion$Close)
qqnorm(Cotizacion$Close)
qqline(Cotizacion$Close)
normalTest(Cotizacion$Close)
pearson.test(Cotizacion$Close)

summary(SP500)
sd(SP500$Close)
qqnorm(SP500$Close)
qqline(SP500$Close)
normalTest(SP500$Close)
pearson.test(SP500$Close)

Temperatura <-
data.frame(anomalias_temperatura$Date[1:61],anomalias_temperatura$Anomaly[1:61])
;Temperatura
colnames(Temperatura)<-c("Fecha", "Anomalia")
data<-
data.frame(Temperatura$Fecha,Cotizacion$Close,SP500$Close, Temperatura$Anomalia
)
colnames(data)<-c("Fecha", "Cotización", "SP500", "Anomalia")

skewness(data$Cotización) #negativo -> la cola izquierda es más pesada
kurtosis(data$Cotización)

par(mfrow = c(1, 1))
linealmodel2<-lm(data$Cotización~data$SP500+data$Anomalia)
print(linealmodel2)
summary(linealmodel2)
colors<- c("black","green")
plot(data$Fecha,data$Cotización,type = "l",xlab="Año", ylab = "Cotización",main =
"Modelo lineal múltiple")
lines(data$Fecha,predict(linealmodel2),col="green")
legend("topleft", legend = c("Muestra","Estimación modelo lineal"), col = colors, pch =
19, title = "Leyenda")
AIC(linealmodel2)
predicción4<- predict(linealmodel2)
RMSE4<-mean(sqrt((predicción4-data$Cotización)^2));RMSE4 #9.829889

glm1<-
glm(data$Cotización~data$SP500+data$Anomalia,family=Gamma(link=identity))

```

```
summary(glm1)
plot(data$Fecha,data$Cotización,type="l",xlab="Año", ylab = "Cotización",main =
"Modelo Gamma")
lines(data$Fecha,predict(glm1),col="green")
legend("topleft", legend = c("Muestra","Estimación modelo gamma"), col = colors, pch
= 19, title = "Leyenda")
predicción4.1<- predict(glm1)
RMSE4.1<-mean(sqrt((predicción4.1-data$Cotización)^2));RMSE4.1 #9.699051
```

```
modelo_polinomico <- lm(data$Cotización ~
data$SP500+data$Anomalia+I(data$SP500^2)+I(data$Anomalia^2)+I(data$Anomalia*
data$SP500))
print(modelo_polinomico)
summary(modelo_polinomico)
plot(data$Fecha,data$Cotización,type="l",xlab="Año", ylab = "Cotización",main =
"Modelo Polinómico")
lines(data$Fecha,predict(modelo_polinomico),col="green")
legend("topleft", legend = c("Muestra","Estimación modelo polinómico"), col = colors,
pch = 19, title = "Leyenda")
AIC(modelo_polinomico)
predicción4.3<- predict(modelo_polinomico)
RMSE4.3<-mean(sqrt((predicción4.3-data$Cotización)^2));RMSE4.3 #9.157216
```

```
datos<-ts(data$Cotización)
data3 <- array(c(data$Anomalia, data$SP500),dim = c(61,2))
dimnames(data3) <- list(c("row"[1:61]),c("Anomalia","SP500"))
modelo_arima<-auto.arima(data$Cotización, xreg = data3[,c("Anomalia", "SP500")])
summary(modelo_arima)
predicción4.4 <- predict(modelo_arima,newxreg = data3)
RMSE4.4<-mean(sqrt((predicción4.4$pred-data$Cotización)^2));RMSE4.4 #14.46104
plot(data$Fecha,data$Cotización,type="l",xlab="Año", ylab = "Cotización",main =
"Paseo aleatorio")
lines(data$Fecha,predicción4.4$pred,col="green")
legend("topleft", legend = c("Muestra","Estimación modelo ARIMA(0,1,0)"), col =
colors, pch = 19, title = "Leyenda")
```

```
par(mfrow = c(1, 2))
colors2<- c("black","red")
plot(data$Fecha,data$Cotización,type="l", ylab = "Cotizacion",xlab = "Año", main =
"Modelo polinómico")
lines(data$Fecha,predict(modelo_polinomico),col="red")
legend("topleft", legend = c("Muestra","Estimación modelo polinómico"), col =
colors2, pch = 19, title = "Leyenda")
plot(data$Fecha,data$Cotización,type="l", ylab = "Cotizacion",xlab = "Año",main =
"Modelo gamma")
lines(data$Fecha,predict(glm1),col="red")
```

```

legend("topleft", legend = c("Muestra","Estimación modelo Gamma"), col = colors2,
pch = 19, title = "Leyenda")

####Capacidad de predicción del modelo

data_train <-
data.frame(Temperatura$Fecha[1:50],data$Cotización[1:50],SP500$Close[1:50],Tempe
ratura$Anomalia[1:50])
colnames(data_train)<-c("Fecha","Cotizacion", "SP500", "Temperatura")
data_test <-
data.frame(Temperatura$Fecha[1:11],data$Cotización[1:11],SP500$Close[1:11],Tempe
ratura$Anomalia[1:11])
colnames(data_test)<-c("Fecha","Cotizacion", "SP500", "Temperatura")

names(data_train)
modelo_polinomico2 <- lm(Cotizacion ~ poly(Temperatura, 2) + poly(SP500, 2),data =
data_train)
forecast_poli <- forecast(modelo_polinomico2, h =
length(data_test$Cotizacion),newdata = data_test)
plot(forecast_poli$fitted)

observado <- data.frame(year=data$Fecha, cotizacion=data$Cotización)
predicho <- data.frame(year=Temperatura$Fecha[1:11],
cotizacion=forecast_poli$mean)

ggplot()+geom_line(data=observado, aes(x=year,y=cotizacion), color="black")+
geom_line(data=predicho, aes(x=year,y=cotizacion), color="blue")+labs(title =
"Predicción vs observado modelo polinómico",
x = "Año",
y = "Cotización",
color = "Legend") +
scale_color_manual(values = c("black" = "black", "blue" = "blue"),
labels = c("Observado", "Proyectado")) +
theme_minimal()

names(data_train)
modelo_gamma2 <- glm(Cotizacion~SP500+Temperatura,gaussian(link = identity),data
= data_train)
predict_gamma2 <- predict(modelo_gamma2,newdata=data_test)
ts_predict <- ts(predict_gamma2)
forecast_gamma <- forecast(ts_predict, h = length(data_test$Cotizacion))
plot(forecast_gamma$fitted)

observado <- data.frame(year=data$Fecha, cotizacion=data$Cotización)

```

```

predicho <- data.frame(year=Temperatura$Fecha[1:11],
cotizacion=forecast_gamma$mean)

ggplot()+geom_line(data=observado, aes(x=year,y=cotizacion), color="black")+
  geom_line(data=predicho, aes(x=year,y=cotizacion), color="blue")+labs(title =
"Predicción vs observado modelo Gamma",
                                x = "Año",
                                y = "Cotización",
                                color = "Legend") +
  scale_color_manual(values = c("black" = "black", "blue" = "blue"),
                    labels = c("Observado", "Proyectado")) +
  theme_minimal()

#####Proyección
nuevas_fechas <- seq(max(anomalias_temperatura$Date) + 1, by = "month", length.out
= 24)
nuevo_time<-seq(max(anomalias_temperatura$Time) + 1, length.out = 24)
nuevos_datos <- data.frame(Time = nuevo_time)
nuevos_datos$Anomalia <- 0
predicciones <- predict(segmentada3, newdata = nuevos_datos);predicciones
proyeccion <- data.frame(fecha = nuevas_fechas, prediccion = predicciones);proyeccion
muestra<-
data.frame(anomalias_temperatura$Date[1:289],anomalias_temperatura$Anomaly[1:28
9],anomalias_temperatura$Time[1:289])
colnames(muestra)<-c("Año","Temperatura","Time")
ggplot() +
  geom_line(data = muestra, aes(x = Año, y = Temperatura), color = "blue") +
  geom_line(data = proyeccion, aes(x = fecha, y = prediccion), color = "red") +
  ggtitle("Proyección de 2 años de las anomalías de la temperatura")

#polinomico
nuevo_sp500<- data.frame(SP500$Close[1:24])
colnames(nuevo_sp500)<-c("SP500")
nueva_data <-
data.frame(Time=nuevo_time,Anomalia=proyeccion$prediccion,SP500=nuevo_sp500$
SP500)
nueva_data$Cotización <- 0
Cotizacion_proyeccion <- predict(modelo_polinomico,newdata =
nueva_data);Cotizacion_proyeccion
proyeccion_cotiz <- data.frame(fecha = nuevas_fechas, prediccion_poli =
Cotizacion_proyeccion[1:24])
muestra_cotiz<-data.frame(Año=data$Fecha,Cotización=data$Cotización)
colnames(muestra)<-c("Año","Temperatura","Time")
ggplot() +

```

```

geom_line(data = muestra_cotiz, aes(x = Año, y = Cotización), color = "blue") +
geom_line(data = proyeccion_cotiz, aes(x = fecha, y = prediccion_poli), color = "red")
+
ggtitle("Proyección de la cotización a 2 años usando el modelo polinómico") +
labs(x = "Año", y = "Cotización")

summary(proyeccion_cotiz)
sd(proyeccion_cotiz$prediccion)

#Gamma
nueva_data2 <-
data.frame(Time=nuevo_time,Anomalia=proyeccion$prediccion,SP500=nuevo_sp500$
SP500)
nueva_data2$Cotización <- 0
Cotizacion_proyeccion2 <- predict(glm1,newdata =
nueva_data2);Cotizacion_proyeccion2
proyeccion_cotiz2 <- data.frame(fecha2 = nuevas_fechas, prediccion2 =
Cotizacion_proyeccion2[1:24])
muestra_cotiz2<-data.frame(Año=data$Fecha,Cotización=data$Cotización)
colnames(muestra_cotiz2)<-c("Año","Temperatura","Time")
ggplot() +
geom_line(data = muestra_cotiz2, aes(x = Año, y = Cotización), color = "blue") +
geom_line(data = proyeccion_cotiz2, aes(x = fecha2, y = prediccion2), color = "red")
+
ggtitle("Proyección de la cotización a 2 años usando el modelo Gamma") +
labs(x = "Año", y = "Cotización")

summary(proyeccion_cotiz2)
sd(proyeccion_cotiz2$prediccion2)

#Estadísticos dos últimos años cotizaciones
nuevo_cotización<- data.frame(Cotizacion$Close[1:24])
summary(nuevo_cotización$Cotizacion.Close.1.24.)
sd(nuevo_cotización$Cotizacion.Close.1.24.)

#Prueba con muestra más grande
SP500_2006 <- read_excel("C:/Users/sofia/Desktop/TFM/^SPX (2).xlsx")
Cotizacion_2006 <- read_excel("C:/Users/sofia/Desktop/TFM/BG (1).xlsx")

Temperatura <-
data.frame(anomalias_temperatura$Date[1:218],anomalias_temperatura$Anomaly[1:21
8]);Temperatura

```

```

colnames(Temperatura)<-c("Fecha","Anomalia")

data<-
data.frame(Temperatura$Fecha,Cotizacion_2006$Close,SP500_2006$Close,Temperatur
a$Anomalia)

colnames(data)<-c("Fecha","Cotización","SP500_2006","Anomalia")


par(mfrow = c(1, 1))

linealmodel2<-lm(data$Cotización~data$SP500_2006+data$Anomalia)

print(linealmodel2)

summary(linealmodel2)

AIC(linealmodel2)

predicción4<- predict(linealmodel2)

RMSE4<-mean(sqrt((predicción4-data$Cotización)^2));RMSE4


glm1<-
glm(data$Cotización~data$SP500_2006+data$Anomalia,family=Gamma(link=identity)
)

summary(glm1)

predicción4.1<- predict(glm1)

RMSE4.1<-mean(sqrt((predicción4.1-data$Cotización)^2));RMSE4.1


modelo_polinomico <- lm(data$Cotización ~
data$SP500_2006+data$Anomalia+I(data$SP500_2006^2)+I(data$Anomalia^2)+I(data
$Anomalia*data$SP500_2006))

print(modelo_polinomico)

summary(modelo_polinomico)

AIC(modelo_polinomico)

predicción4.3<- predict(modelo_polinomico)

RMSE4.3<-mean(sqrt((predicción4.3-data$Cotización)^2));RMSE4.3

```

10. Bibliografía

- Banco Central Europeo. (2020). Guía sobre riesgos relacionados con el clima y medioambientales.
- Banco de España. (2021). Riesgo climático, Transición energética, Riesgos financieros y Crecimiento económico global. 1-11.
- Comisión Nacional del Mercado de Valores. (s. f.). Las Finanzas Sostenibles.
- Garcia Jorcano, L., & Sanchis Marco, L. (2024). Forecasting the effect of extreme sea-level rise on financial market risk. *International Review of Economics & Finance*, 93, Parte B, 1-27.
- H. Greene, W. (2002). *Econometric analysis* (8.a ed.). Prentice Hall Iberia.
- Hyndman, R. J., & Khandakar, Y. (2008). Automatic Time Series Forecasting: The forecast Package for R. 27(3), 1-22.
- Mudelsee, M. (2009). Break function regression. 174, 49-63.
- Mudelsee, M. (2010). *Climate Time Series Analysis* (Vol. 51). Springer Cham.
- Mudelsee, M. (2019). Trend analysis of climate time series: A review of methods. 190, 310-322.
- National Centers for Environmental Information. (2024). Global Surface Temperature Anomalies [dataset]. <https://www.ncei.noaa.gov/access/monitoring/global-temperature-anomalies/anomalies>
- R. Solow, A. (1987). Testing for Climate Change: An Application of the Two-Phase Regression Model. 26(10), 1401-1405.
- Shumway, R. H., & Stoffer, D. S. (2017). *Time Series Analysis and Its Applications: With R Examples*. Springer Cham.