



Universidad
Zaragoza

Trabajo Fin de Grado en
Administración y Dirección de
Empresas

**ESTUDIO DEL RIESGO DE MERCADO EN
DOS ÍNDICES: IBEX 35 Y S&P 500**

Autor

Iván Coloma Morte

Directores

Jesús Ángel Miguel Álvarez y Pilar Gargallo Valero

Facultad de Economía y Empresa

Universidad de Zaragoza

Enero de 2026

ÍNDICE

1. INTRODUCCIÓN.....	3
2. MÉTODOS DE ESTIMACIÓN DEL VALOR EN RIESGO (VaR)	4
2.1. Método de Varianza-Covarianza	4
2.2. Método Estadístico Paramétrico	4
2.3. Método Bootstrap.....	5
3. MODELOS HETEROCEDÁSTICOS PARA ESTUDIAR LAS RENTABILIDADES DE MERCADOS FINANCIEROS	5
3.1 Modelización de la rentabilidad media: Modelos ARMA	6
3.2 Modelización de la varianza condicional: Modelos GARCH	6
3.3 Aplicación conjunta: Modelos ARMA-GARCH.....	7
4. ANÁLISIS EMPÍRICO	7
4.1 Estudio univariante de cada serie de rentabilidades	8
4.1.1 Modelo ARMA para el índice IBEX 35	8
4.1.2 Modelo GARCH para el índice IBEX 35	9
4.1.3 Modelo ARMA para el índice S&P 500.....	11
4.1.4 Modelo GARCH para el índice S.....	13
4.2 Estimación y evaluación del VAR.....	15
4.2.1 Evaluación gráfica del VaR estimado.....	17
4.2.2 Evaluación numérica del VaR estimado	18
5. CONCLUSIONES.....	21
BIBLIOGRAFÍA.....	22

1. INTRODUCCIÓN

En un contexto de creciente volatilidad e incertidumbre en los mercados financieros globales, el estudio y la gestión del riesgo se han convertido en pilares fundamentales para las entidades financieras, inversores y organismos reguladores. Las crisis económicas del siglo XXI han evidenciado la necesidad de herramientas robustas que permitan cuantificar y anticipar posibles pérdidas, garantizando una adecuada solvencia y capacidad de respuesta ante eventos extremos. En este sentido, el análisis del riesgo de mercado se presenta como una disciplina esencial dentro del ámbito financiero, y particularmente, la medida conocida como "Valor en Riesgo" (VaR) ha adquirido una relevancia destacada.

El VaR se ha consolidado como uno de los instrumentos más utilizados para estimar la pérdida máxima esperada de una posición o cartera en condiciones normales de mercado y dentro de un determinado horizonte temporal y nivel de confianza

El objetivo principal de este trabajo es realizar un análisis comparativo del riesgo de mercado en dos de los principales índices bursátiles a nivel internacional: el IBEX 35, representativo del mercado español, y el S&P 500, que aglutina las 500 compañías más representativas del mercado estadounidense.

En una primera parte del trabajo se presentan tres métodos de estimación del VaR: el método de varianza-covarianza, el enfoque estadístico paramétrico y la técnica Bootstrap. Posteriormente, se introducen modelos econométricos más avanzados como los modelos ARMA y GARCH, que permiten tener en cuenta la dinámica temporal de la media y la varianza condicional de las rentabilidades, aspectos especialmente importantes cuando se trabaja con series financieras diarias.

En la parte empírica se realiza un análisis univariante de ambas series, se calcula el VaR con los distintos métodos propuestos y se representan gráficamente los resultados. Finalmente, se evalúa la calidad de cada estimación mediante pruebas de cubrimiento, lo que permite identificar cuál de los modelos es más adecuado en función del comportamiento real de los mercados.

El trabajo concluye con una valoración de los resultados obtenidos, reflexionando sobre las ventajas e inconvenientes de cada enfoque y su aplicabilidad en la práctica financiera. Con ello, se pretende aportar una visión clara y crítica sobre el uso del VaR como herramienta de gestión del riesgo.

2. MÉTODOS DE ESTIMACIÓN DEL VALOR EN RIESGO (VaR)

En este apartado se abordan tres enfoques distintos para la estimación del Valor en Riesgo (VaR): el método de varianza-covarianza, el método estadístico paramétrico y el método Bootstrap. Cada uno presenta características propias, tanto en su formulación como en sus supuestos y aplicabilidad a diferentes tipos de datos financieros.

2.1. Método de Varianza-Covarianza

Este método es uno de los más tradicionales y sencillos para estimar el VaR. Se basa en el supuesto de que las rentabilidades de los activos siguen una distribución normal, y que, por tanto, el VaR puede calcularse directamente a partir de la media y la desviación típica de las rentabilidades.

La expresión general del VaR bajo este método es:

$$VaR_{\{t+h,\alpha\}} = \mu_t - z_\alpha \cdot \sigma_t$$

donde z_α representa el cuantil de la distribución normal estándar para un nivel de confianza, y μ_t es la rentabilidad media calculada hasta el instante t y σ_t , la desviación típica de las rentabilidades calculada hasta el instante t . En muchas aplicaciones prácticas se considera la media nula para simplificar el cálculo, especialmente cuando se trabaja con horizontes de tiempo cortos como un día. Sin embargo, en el estudio que vamos a realizar mas adelante la temporalidad será de 1,5 y 20 días

Este método es útil cuando se desea obtener una estimación rápida del riesgo, aunque su principal limitación es la fuerte dependencia del supuesto de normalidad, lo cual no siempre es válido en las series de rentabilidades financieras.

2.2. Método Estadístico Paramétrico

Este método generaliza el enfoque varianza-covarianza al permitir que las rentabilidades no tengan por qué seguir una distribución normal, sino que puedan ajustarse a distribuciones más flexibles (como la t de Student). Para implementarlo correctamente, no basta con estimar una desviación típica constante, sino que es necesario emplear modelos que capten la dinámica temporal de la serie de rentabilidades.

En este contexto, se recurre a los modelos ARMA-GARCH:

Modelos ARMA:

Se utilizan para modelar la rentabilidad media condicional. Permiten capturar dependencias lineales en la serie temporal, de modo que la rentabilidad actual puede depender de rentabilidades pasadas y de innovaciones previas. Con esto se elimina cualquier estructura sistemática en la media antes de analizar la varianza.

Modelos GARCH:

Se aplican a los residuos del ARMA y permiten modelar la varianza condicional. Estos modelos reflejan cómo la volatilidad cambia a lo largo del tiempo, recogiendo fenómenos típicos de los mercados financieros como la volatilidad agrupada, la persistencia en el riesgo y la sensibilidad a choques extremos.

2.3. Método Bootstrap

El método bootstrap pertenece al conjunto de enfoques no paramétricos. Consiste en generar un alto número de muestras aleatorias con reemplazo a partir del conjunto original de rentabilidades.

El VaR se obtiene como el cuantil correspondiente a esta distribución muestreada artificialmente. Por ejemplo, si se generan 10.000 muestras Bootstrap, el percentil 5 (es decir, la observación número 500 ordenada de menor a mayor) proporcionaría el VaR al 95% de confianza.

Este método no requiere asumir ninguna forma funcional para la distribución de rentabilidades, por lo que resulta especialmente útil cuando se observan asimetrías o colas pesadas. Como contrapartida, requiere mayor capacidad computacional y depende mucho de la calidad de los datos históricos.

La aplicación conjunta de estos tres métodos en el análisis empírico permitirá comparar su comportamiento y precisión, así como determinar cuál de ellos ofrece mejores resultados según el contexto de cada mercado analizado (IBEX 35 y S&P 500).

3. MODELOS HETEROCEDÁSTICOS PARA ESTUDIAR LAS RENTABILIDADES DE MERCADOS FINANCIEROS

Los mercados financieros presentan una serie de características empíricas que dificultan su modelización mediante técnicas estadísticas tradicionales. Una de las más relevantes es la presencia de heterocedasticidad condicional, es decir, la varianza de las rentabilidades no es constante a lo largo del tiempo, sino que tiende a agruparse en períodos de alta y baja volatilidad. Para abordar este comportamiento se emplean modelos econométricos específicos, entre los que destacan los modelos ARCH (Autoregressive Conditional Heteroskedasticity) y GARCH (Generalized ARCH).

Este tipo de modelos permiten capturar la naturaleza cambiante de la volatilidad en series temporales financieras y han demostrado ser herramientas eficaces para la estimación de riesgo y pronóstico del comportamiento de los activos financieros. El uso combinado de modelos ARMA-GARCH proporciona una representación robusta de las dinámicas tanto en la media como en la varianza de las series de rentabilidades.

3.1 Modelización de la rentabilidad media: Modelos ARMA

Antes de modelar la varianza, es necesario eliminar cualquier estructura sistemática en la media de la serie de rentabilidades. Para ello, se utilizan los modelos ARMA (Autoregressive Moving Average), que combinan componentes autorregresivos (AR) y de medias móviles (MA). Estos modelos permiten capturar la dependencia temporal entre observaciones sucesivas, lo que es habitual en los mercados financieros a corto plazo.

Un modelo ARMA(p,q) se expresa de la siguiente forma

$$R_t = \mu + \sum_{i=1}^p \varphi_i R_{t-i} + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \varepsilon_t$$

donde:

R_t es la rentabilidad en el tiempo t.

μ es la media constante.

φ_i son los coeficientes autorregresivos (AR).

θ_j son los coeficientes de medias móviles (MA).

ε_t es la rentabilidad en el tiempo t.

p es el número de rentabilidades pasadas que influyen en la actual.

q es el número de errores pasados que afectan a la rentabilidad actual.

El ajuste de un modelo ARMA adecuado se basa en la inspección de las funciones de autocorrelación (ACF) y autocorrelación parcial (PACF), así como en el uso de criterios de información como AIC y BIC. Una vez modelada correctamente la media, los residuos resultantes pueden ser analizados para modelar su varianza condicional.

3.2 Modelización de la varianza condicional: Modelos GARCH

Aunque los modelos ARMA suponen que los errores son homocedásticos (varianza constante), esta hipótesis no se ajusta bien a la realidad de los mercados financieros. Para modelar correctamente la variabilidad temporal de la varianza, se introducen los modelos ARCH y, más comúnmente, su generalización: el modelo GARCH.

El modelo ARCH(q), propuesto por Engle (1982), asume que la varianza condicional depende de los cuadrados pasados:

$$\sigma_t^2 = \omega + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2$$

donde σ_t^2 es la varianza condicional en el tiempo t , $\omega > 0$, y los coeficientes $\alpha_i \geq 0$.

Bollerslev (1986) generalizó este modelo dando lugar al GARCH(p,q), que incluye también retardos de la propia varianza condicional:

$$\sigma_t^2 = \omega + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2$$

Este modelo capta de forma más eficiente la persistencia de la volatilidad observada en los mercados. Los parámetros α_i miden el impacto de las innovaciones pasadas (choques), mientras que β_j reflejan la influencia de la volatilidad pasada. La suma $\sum \alpha_i + \sum \beta_j$ indica persistencia de las informaciones pasadas en la volatilidad actual. Cuánto más cerca sea su valor de 1, indica que el impacto de las noticias pasadas se mantiene más tiempo en la volatilidad actual.

En la práctica, el modelo GARCH (1,1) suele ser suficiente para capturar la mayoría de los patrones observados en las series financieras. Este modelo ha demostrado un buen rendimiento empírico y es ampliamente utilizado en estudios aplicados de riesgo financiero.

3.3 Aplicación conjunta: Modelos ARMA-GARCH

La combinación de un modelo ARMA para la media y un modelo GARCH para la varianza permite capturar simultáneamente las dependencias lineales y no lineales presentes en las rentabilidades financieras. Esta combinación resulta especialmente adecuada para el cálculo del VaR, ya que permite estimar la distribución condicional de la rentabilidad y, por tanto, obtener predicciones más precisas del riesgo.

En resumen, los modelos heterocedásticos como GARCH representan una mejora sustancial frente a enfoques más simples para estimar la volatilidad. Su integración con modelos ARMA ofrece una herramienta completa para la modelización de series financieras, lo que resulta de gran utilidad en contextos donde la gestión del riesgo es prioritaria, como es el caso del presente estudio sobre los índices IBEX 35 y S&P 500.

4. ANÁLISIS EMPÍRICO

Este apartado recoge el análisis aplicado sobre las series de rentabilidades diarias del IBEX 35 y del S&P 500, utilizando distintos métodos estadísticos y econométricos. El objetivo principal es evaluar y comparar el riesgo de mercado en ambos índices mediante el cálculo del Valor en Riesgo (VaR) y su posterior validación empírica. El periodo analizado comienza el 1 de enero de 2015 y finaliza el 31 de marzo del 2025, que supone 2575 observaciones. Estos datos se han obtenido a través de la página web investing.com.

4.1 Estudio univariante de cada serie de rentabilidades

4.1.1 Modelo ARMA para el índice IBEX 35

Primero realizamos un análisis de las rentabilidades diarias, modelizando un proceso ARMA sobre cada una de ellas. Así, en la Figura 1 presentamos el gráfico de las rentabilidades diarias en el panel de la izquierda. Se pueden extraer varias observaciones importantes como que la serie presenta una estructura claramente volátil, con períodos de alta y baja variabilidad. También, se observan picos extremos negativos (caídas pronunciadas), especialmente a mitad de 2016 y hacia inicios de 2020, lo que puede asociarse con eventos exógenos como el Brexit o la crisis del COVID-19 que afectó a este mercado español al igual que al resto de Europa. En esos momentos, las rentabilidades presentan movimientos de gran magnitud tanto en sentido positivo como negativo. Tras ese periodo de alta turbulencia, la volatilidad parece reducirse, aunque nunca desaparece del todo, lo que es característico de los mercados financieros.

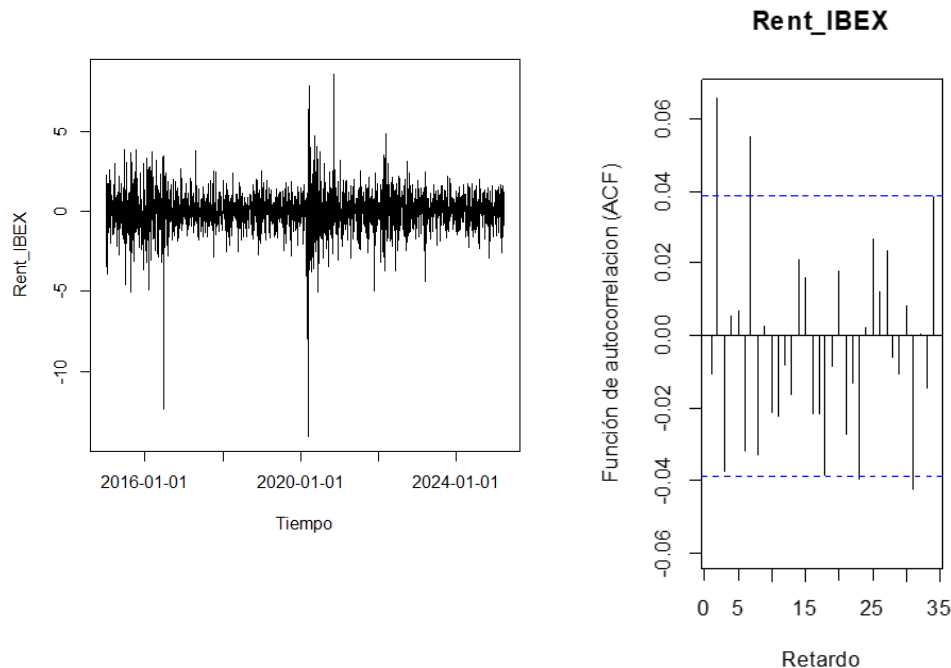


Figura 1 Rentabilidades diarias del IBEX 35 y su función de autocorrelación (ACF)

En el panel de la derecha, está dibujada la función de autocorrelación junto a sus bandas de confianza, mostrando que el segundo retardo es significativo, por lo tanto, probaremos modelos de orden 2. Este resultado refuerza la idoneidad del modelo ARMA como punto de partida, aunque también pone en evidencia la necesidad de considerar modelos adicionales de volatilidad como GARCH para completar el análisis.

Tras comentar los gráficos pasamos a analizar las estimaciones de los diferentes parámetros que hemos extraído. En particular, el modelo AR (2) es seleccionado empleando los usuales criterios AIC, BIC, ...

- El parámetro **AR (2)** es estadísticamente significativo ($t = 3.35$, $p < 0.01$), indicando una autocorrelación de segundo orden relevante en la serie.
- El parámetro **AR (1)** ha sido estimado como cero, lo cual sugiere que no aporta capacidad explicativa adicional
- La desviación estándar de los errores (σ) es de aproximadamente 1.217, con una significación estadística muy elevada ($p < 0.001$), lo cual confirma la presencia de una variabilidad importante en la serie.

Los errores estándar robustos muestran que la significación estadística de AR (2) se reduce ($p \approx 0.087$), situándose cerca del umbral del 10%, lo que podría sugerir sensibilidad ante la presencia de heterocedasticidad o estructura de varianza no constante en los residuos.

En cuanto a los residuos estandarizados fueron evaluados mediante el test de Ljung-Box ponderado, donde los p-valores asociados a distintos retardos (Lag [1], Lag [5], Lag [9]) indican que no se puede rechazar la hipótesis nula de ausencia de autocorrelación. Esto sugiere que los residuos del modelo no presentan dependencias lineales adicionales que no haya sido capturada por el ARMA planteado.

Sin embargo, el test de Ljung-Box aplicado a los residuos al cuadrado muestra valores altamente significativos ($p < 0.0001$), lo que evidencia la presencia de heterocedasticidad condicional, un fenómeno común en series financieras. Este hallazgo es corroborado por los resultados del test ARCH-LM, que indica con claridad la existencia de efectos ARCH de orden 2, 5 y 10

La matriz de ranking confirma que la especificación ARMA (2,0) figura entre las mejores según el criterio BIC. No obstante, los problemas de heterocedasticidad y posible inestabilidad de las estimaciones de los parámetros sugieren que sería adecuado complementar el análisis con un modelo GARCH.

4.1.2 Modelo GARCH para el índice IBEX 35

Tras haber realizado la estimación del modelo GARCH suponiendo una distribución normal en los errores, hemos podido contrastar que es necesario emplear una distribución t-student. Esto es debido a la presencia de una elevada curtosis y de eventos extremos en la serie de rentabilidades del IBEX 35, observados tanto visualmente como en los residuos del modelo ARCH, lo cual justifica el uso de una distribución con colas más pesadas. En este contexto, la distribución t-Student es preferible a la normal. Esto mejora tanto la calidad del ajuste como la robustez de las predicciones de riesgo financiero.

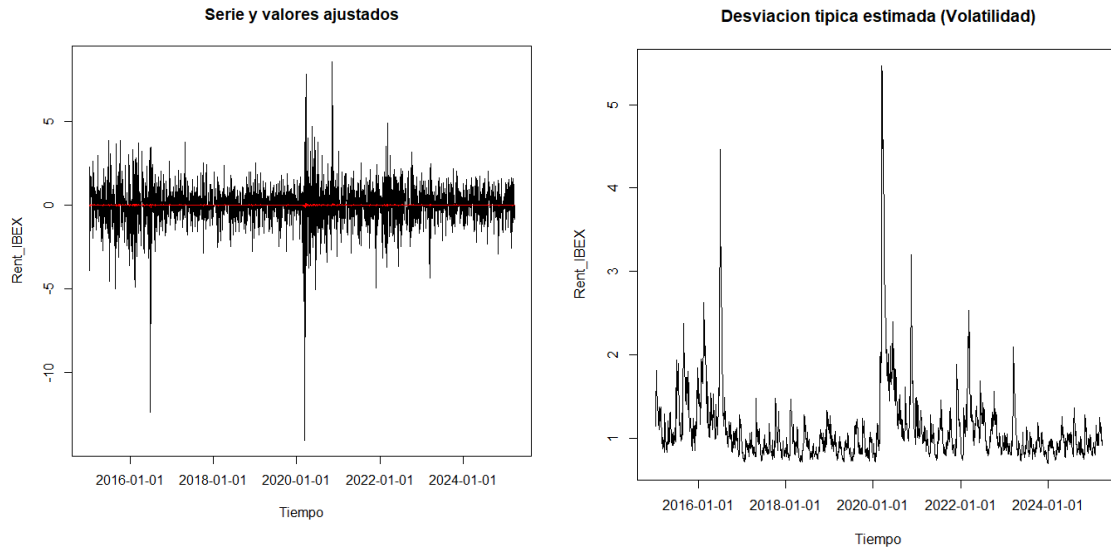


Figura 2: Estimaciones de las rentabilidades y de la volatilidad diaria

El panel de la derecha presenta la serie de rentabilidades observadas (línea negra) con los valores ajustados por el modelo AR (2) (línea roja). La línea roja representa la media condicional estimada, que se mantiene cerca de cero a lo largo de todo el periodo, lo cual es esperable en una serie de rentabilidades financieras, donde la media es típicamente baja o nula. Mientras que en el panel de la izquierda se presenta la evolución temporal de la desviación típica condicional, es decir, la volatilidad estimada por el modelo GARCH a lo largo del tiempo. Se observa con claridad el fenómeno de volatilidad agrupada: hay períodos prolongados de baja volatilidad seguidos de aumentos bruscos.

Los picos más altos de volatilidad coinciden con eventos de mercado altamente disruptivos, especialmente alrededor del 23 de junio de 2016 coincidiendo con el referéndum del Brexit y en torno al 13 de marzo de 2020, inicio de la pandemia del COVID-19. Esto confirma que el modelo GARCH es capaz de capturar correctamente la dinámica de la volatilidad, algo que no sería posible con modelos que asumen varianza constante.

En cuanto a los resultados numéricos, el modelo GARCH (1,1) da las siguientes estimaciones junto a sus p-valores:

- La constante $\omega = 0.0657$ ($p < 0.001$),
- El coeficiente de shocks recientes $\alpha_1 = 0.1154$ ($p < 0.01$)
- El parámetro de persistencia $\beta_1 = 0.8336$ ($p < 0.01$).

Los tres parámetros son significativos y, como suele ser habitual, la suma de $\alpha_1 + \beta_1 = 0,949$ indica una alta persistencia, aunque inferior a uno, lo cual nos asegura la estacionariedad del modelo. Es decir, es un patrón típico de los mercados bursátiles donde los shocks de volatilidad tienden a extenderse en el tiempo, pero acaban disipándose gradualmente.

El parámetro de forma (shape) de la distribución t-Student, con un valor de 6.65, $p < 0,001$ confirma la existencia de colas más gruesas que en la distribución normal, lo que justifica plenamente su uso. Esto implica que los eventos extremos como grandes caídas o subidas son más probables de lo que predeciría una distribución normal.

Las pruebas de autocorrelación a través del test de Ljung-Box sobre los residuos y sobre los residuos al cuadrado arrojan p-valores elevados, indicando que no quedan dependencias lineales o no lineales sin capturar. Del mismo modo, los tests ARCH no son significativos, lo que sugiere que la heterocedasticidad condicional ha sido correctamente modelada.

El test de estabilidad de Nyblom presenta un valor conjunto de 1.0569, inferior al valor crítico, lo que confirma la estabilidad de los parámetros a lo largo del tiempo. Además, la prueba de sesgo de signo no detecta asimetrías en la respuesta de la volatilidad ante shocks positivos o negativos, por lo que no sería necesario recurrir a modelos asimétricos como EGARCH o GJR-GARCH.

Por último, el test de bondad de ajuste de Pearson rechaza la hipótesis nula, lo que indica que, aunque el modelo reproduce bien la varianza y el comportamiento general de la serie, puede haber diferencias en los extremos de la distribución empírica. Esto es razonable, dado que los movimientos extremos del mercado suelen ser difíciles de predecir con exactitud.

Así pues, el modelo ARMA–GARCH con distribución t-Student ofrece un ajuste sólido y realista para las rentabilidades del IBEX 35. Logra capturar la persistencia de la volatilidad, la presencia de colas pesadas y la estabilidad de los parámetros, aportando una representación fiel del comportamiento del mercado español durante el periodo 2015–2025. En definitiva, se trata de un modelo parsimonioso pero potente, capaz de reflejar tanto los momentos de calma como los de alta turbulencia financiera.

4.1.3 Modelo ARMA para el índice S&P 500

Comenzamos realizando un análisis gráfico de la serie de rentabilidades diarias del S&P 500 y de las funciones de autocorrelación y autocorrelación parcial para estudiar las posibles dependencias lineales. La Figura 3, en su panel de la izquierda, se presenta la serie de rentabilidades del índice en el periodo de estudio. Se observa que la rentabilidad media del índice oscila en torno a cero, sin tendencia visible a largo plazo. Este comportamiento es característico de los mercados bursátiles desarrollados, donde los precios reflejan de manera eficiente la información disponible y, por tanto, las rentabilidades tienden a ser impredecibles. Se aprecia un episodio de alta volatilidad, particularmente en marzo de 2020, correspondientes a la crisis por la pandemia de COVID-19. También hay tramos de volatilidad reducida, pero nunca completamente ausente, un fenómeno típico de los mercados financieros. La existencia de picos extremos (positivos y negativos) refuerza la idea de colas pesadas y posible no normalidad en los residuos. También, se confirma la presencia de volatilidad agrupada, reforzando la necesidad de usar un modelo ARCH o GARCH.

En el panel de la derecha se muestra la función de autocorrelación simple de las rentabilidades hasta el retardo 35. Se observa que solo algunos retardos iniciales muestran autocorrelaciones marginalmente significativas, todas de baja magnitud ($< 10\%$), lo cual indica una débil dependencia lineal a corto plazo. Este comportamiento es típico en rentabilidades financieras, donde no suele haber autocorrelación en la serie original, pero sí en los residuos al cuadrado, señalando la posible presencia de agrupamiento de volatilidad.

Esto muestra que el patrón observado justifica el uso de modelos ARMA para capturar la dinámica en media, pero también indica la necesidad de modelizar la varianza condicional, motivando el uso posterior del modelo ARCH o GARCH.

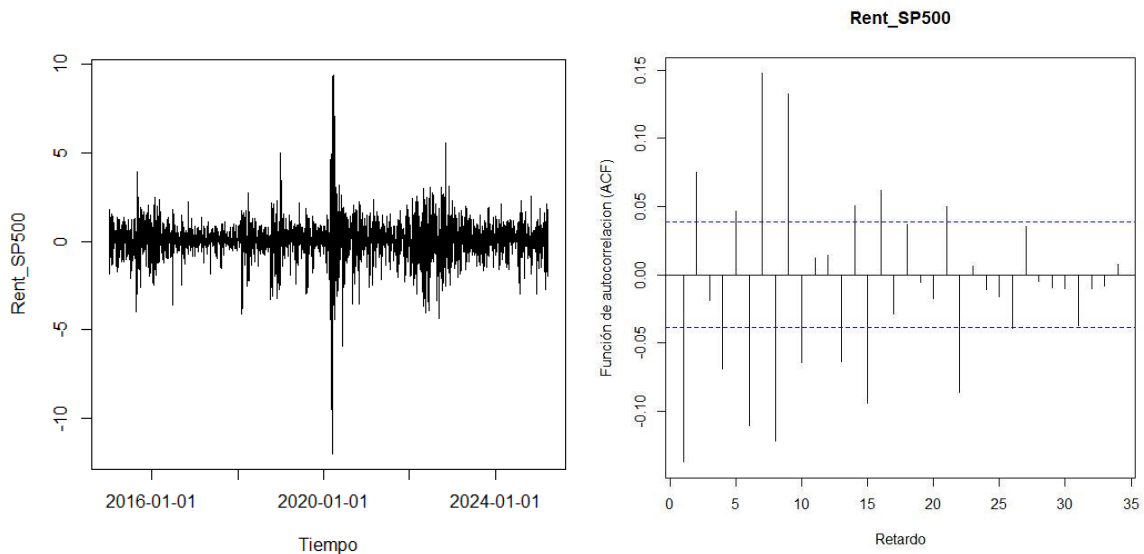


Figura 3: Rentabilidad diarias del S&P 500 y su función de autocorrelación (ACF)

Tras comentar los gráficos pasamos a analizar los diferentes parámetros que hemos estimado al analizar un modelo ARMA (2,2).

- El parámetro AR (1) presenta un valor estimado de -1.739 , con una significación estadística muy elevada ($t = -62.87$, $p < 0.001$), lo que confirma su importancia en la explicación de la dinámica de la serie.
- El parámetro AR (2) también resulta altamente significativo ($t = -25.68$, $p < 0.001$), indicando la existencia de autocorrelación de segundo orden en las rentabilidades del S&P 500.
- El término de medias móviles **MA (1)** es significativo ($t=45.72$, $p<0.001$)
- El coeficiente **MA (2)** es estadísticamente significativo ($t = 31.34$, $p<0.001$).

Por lo tanto, parece adecuado explicar las rentabilidades diarias ajustando un modelo ARMA (2,2). La desviación estándar de los errores (sigma constante) se estima en 1.086, con un valor $t = 12.64$ y $p < 0.001$, lo que indica una fuerte variabilidad en la serie.

Los errores estándar robustos permiten evaluar la sensibilidad del modelo ante la presencia de heterocedasticidad. Con esta corrección, todos los coeficientes se mantienen significativos, aunque con ligeras reducciones en sus t-valores. Esto sugiere que la estructura de media está bien especificada, pero puede estar sujeta a cambios en la varianza no explicados por el modelo ARMA.

En cuanto a los residuos estandarizados, fueron evaluados mediante el test de Ljung-Box ponderado. Los p-valores obtenidos para distintos retardos (Lag [1] = 0.857; Lag [19] = 0.110) indican que no se puede rechazar la hipótesis nula de ausencia de autocorrelación, excepto en el retardo Lag [11] ($p \approx 0.0006$), lo que sugiere que aún persiste cierta dependencia temporal que el modelo no ha capturado del todo. No obstante, en general, el ajuste en media puede considerarse adecuado.

Por otro lado, el test de Ljung-Box para los residuos estandarizados cuadráticos muestra resultados altamente significativos ($p < 0.0001$ en todos los retardos evaluados), lo que pone de manifiesto la existencia de heterocedasticidad condicional, es decir, varianza no constante a lo largo del tiempo. Este hallazgo se ve reforzado por los resultados del test ARCH-LM, que detecta efectos ARCH significativos en los retardos 2, 5 y 10 (todos con $p < 0.001$).

Adicionalmente, el test de estabilidad de Nyblom devuelve un estadístico conjunto de 2.061, superior al valor crítico del 1% (1.88), lo que indica inestabilidad en los parámetros del modelo, especialmente en la varianza (σ). Este comportamiento es típico en mercados financieros, donde la volatilidad suele evolucionar con el tiempo.

La matriz de ranking confirma que la especificación ARMA (2,2) figura entre las mejores según el criterio BIC, lo cual valida su elección para modelar la estructura de media. Sin embargo, los problemas evidentes de heterocedasticidad y posible inestabilidad sugieren que sería adecuado complementar el análisis con un modelo GARCH, con el fin de capturar la dinámica de la varianza condicional y permitir una estimación más precisa del riesgo financiero a través del cálculo del Valor en Riesgo (VaR).

4.1.4 Modelo GARCH para el índice S&P 500

Hemos realizado un ajuste del modelo empleando un modelo GARCH (1,1) para la varianza condicional y, como se ha comentado previamente, hemos detectado una alta curtosis que se explica mejor utilizando una distribución t de Student para los residuos estandarizados. La distribución t-Student es preferible a la normal, ya que permite modelar adecuadamente los shocks extremos sin subestimar su probabilidad. Esto mejora tanto la calidad del ajuste como la robustez de las predicciones de riesgo financiero.

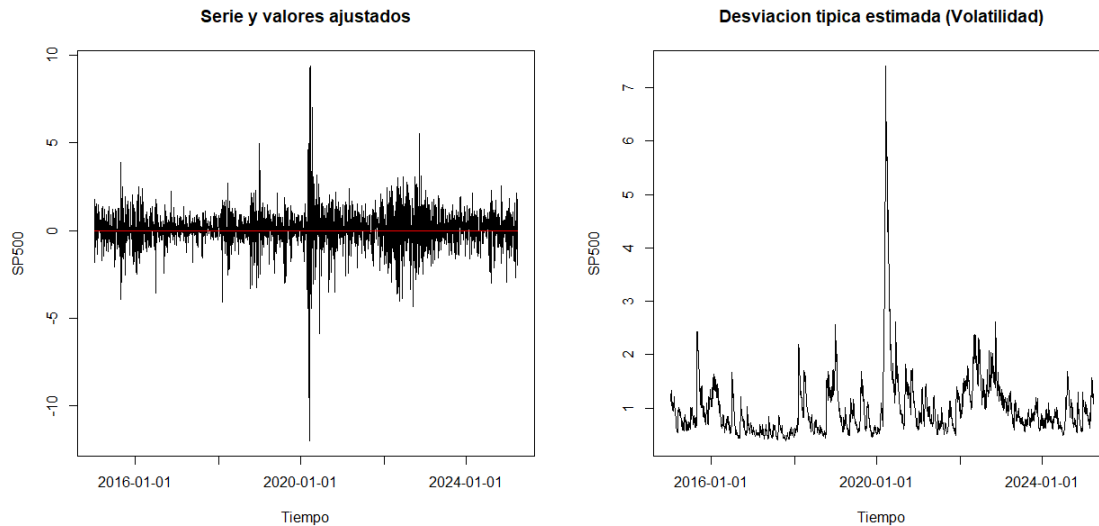


Figura 4. Estimaciones de las volatilidades y rentabilidades diarias

La Figura 4, en su panel de la izquierda, muestra la evolución de las rentabilidades del índice S&P 500 (línea negra), junto con la media condicional estimada por el modelo (línea roja), durante el período 2015–2025. Por otro lado, en el panel de la derecha se representa la evolución de la desviación estándar condicional (volatilidad estimada) generada por el modelo GARCH a lo largo del tiempo. El gráfico de desviación típica condicional muestra que el modelo GARCH (1,1) es capaz de replicar con precisión la evolución de la volatilidad del índice S&P 500. Los picos de riesgo observados coinciden con eventos de alta incertidumbre financiera, como la crisis del COVID-19 en 2020. Además, se confirma la existencia de agrupamiento de volatilidad, lo que valida empíricamente la elección de este tipo de modelos para cuantificar el riesgo financiero.

En cuanto a los resultados numéricos, el ajuste del modelo GARCH (1,1) proporciona las siguientes estimaciones:

- La constante $\omega = 0,0191 (p < 0,001)$
- El coeficiente que mide el impacto del shock reciente $\alpha_1 = 0,1614 (p < 0,01)$
- El parámetro que mide la influencia de la volatilidad prevista $\beta_1 = 0,8336 (p < 0,01)$.

La suma $\alpha_1 + \beta_1 = 0,995$ muestra una altísima persistencia en la volatilidad. Esto implica que los shocks en la varianza tienden a perdurar durante largos periodos, lo que refleja la inercia típica de los mercados bursátiles globales: cuando la volatilidad aumenta bruscamente, tarda bastante tiempo en volver a niveles normales. Este resultado, aunque teóricamente cercano al límite de la estacionariedad, es muy representativo del comportamiento real del S&P 500.

En cuanto a la distribución de los errores, el parámetro de forma (shape) toma un valor de 6.04 ($p < 0.001$), confirmando la existencia de colas más gruesas que las de la distribución

normal. Esto significa que los eventos extremos, como fuertes caídas o subidas del índice, son más frecuentes de lo que la teoría clásica predeciría. En términos prácticos, este hallazgo valida la elección de la t-Student, ya que mejora el ajuste del modelo y proporciona estimaciones de riesgo más realistas, especialmente al calcular medidas como el Valor en Riesgo (VaR).

Los tests de diagnóstico de autocorrelación en residuos muestran que el modelo se comporta correctamente. El test de Ljung-Box aplicado tanto a los residuos estandarizados como a los residuos al cuadrado arroja p-valores altos (todos > 0.3), indicando que no queda autocorrelación residual. Asimismo, los tests ARCH no resultan significativos ($p > 0.4$ en todos los lags), lo que sugiere que la heterocedasticidad condicional ha sido capturada adecuadamente.

En cuanto al test de estabilidad de Nyblom, el valor conjunto de 2.31 supera ligeramente el valor crítico al 1% (2.12), lo que sugiere una leve inestabilidad en algunos parámetros, especialmente en β_1 y en el parámetro de forma. Este resultado puede deberse a cambios estructurales en el mercado estadounidense durante el periodo analizado, como la crisis del Covid 19. Aun así, el modelo mantiene una buena capacidad predictiva y no presenta problemas de especificación graves.

Por último, el test de bondad de ajuste de Pearson rechaza la hipótesis nula en todos los grupos ($p < 0,0001$), lo que indica que, aunque el modelo reproduce correctamente la estructura general de la varianza, no encaja perfectamente con toda la distribución empírica, especialmente en las colas. Sin embargo, esto no invalida los resultados; simplemente refleja la complejidad inherente de los mercados financieros y la dificultad de modelar eventos extremos con total precisión.

De esta forma, el modelo ARMA–GARCH con distribución t-Student proporciona una representación sólida y coherente del comportamiento de las rentabilidades del S&P 500. Captura la alta persistencia de la volatilidad, la presencia de colas gruesas y la asimetría ante los choques, ofreciendo una visión precisa y realista del riesgo de mercado en el índice estadounidense. Aunque podrían explorarse modelos asimétricos para perfeccionar el ajuste, los resultados actuales son plenamente satisfactorios y coherentes con la teoría y la evidencia empírica disponible.

4.2 Estimación y evaluación del VAR

Tras el análisis de los modelos que se ajustan a las rentabilidades diarias de ambos índices, nos proponemos estimar el VaR para varios horizontes y empleando dos métodos diferentes de cálculo:

- **Método paramétrico:** Se basa en la predicción de los valores futuros de las rentabilidades r_{t+h} empleando la estimación del modelo hasta el instante t y con la distribución t de Student para los residuos estandarizados.
- **Método Bootstrap:** Las rentabilidades futuras r_{t+h} se calculan con el modelo estimado hasta el instante t , pero empleando un método bootstrap para generar los

residuos estandarizados. Con este método permitimos que la distribución residual no esté prefijada de antemano y sean los propios datos los que determinen su comportamiento.

En ambos métodos se generan 10.000 valores futuros de las rentabilidades y se calcula el VaR, por ejemplo, con un nivel del $(1-\alpha)100\%$ tomando el cuantil de los valores futuros que deja por debajo el $\alpha 100\%$ de ellos:

$$\text{VaR}_{t+h;1-\alpha} \text{ tal que } \frac{\#\{r_{t+h} < \text{VaR}_{t+h;1-\alpha}\}}{N} = \alpha$$

Se han considerado tres horizontes con los que hemos estimado el VaR (1 día ($h=1$), 1 semana ($h=5$) y 1 mes ($h=20$)) y hemos considerado el nivel de confianza usual del 95% ($\alpha=0.05$). Para cada horizonte e índice bursátil se ha estimado el VaR con ambos métodos de predicción: paramétrica y Bootstrap.

Para la realización de este análisis debemos contrastar algunas hipótesis sobre la bondad del VaR a través de diferentes pruebas, con el objetivo de analizar las diferencias entre la utilización del método bootstrap y paramétrico.

Estos cuatro test nos permiten comprobar si estamos estimando correctamente las predicciones de riesgo calculadas con el modelo VaR, así como nos ayuda a ver si las pérdidas que se producen en la práctica son consistentes con el nivel de confianza asumido.

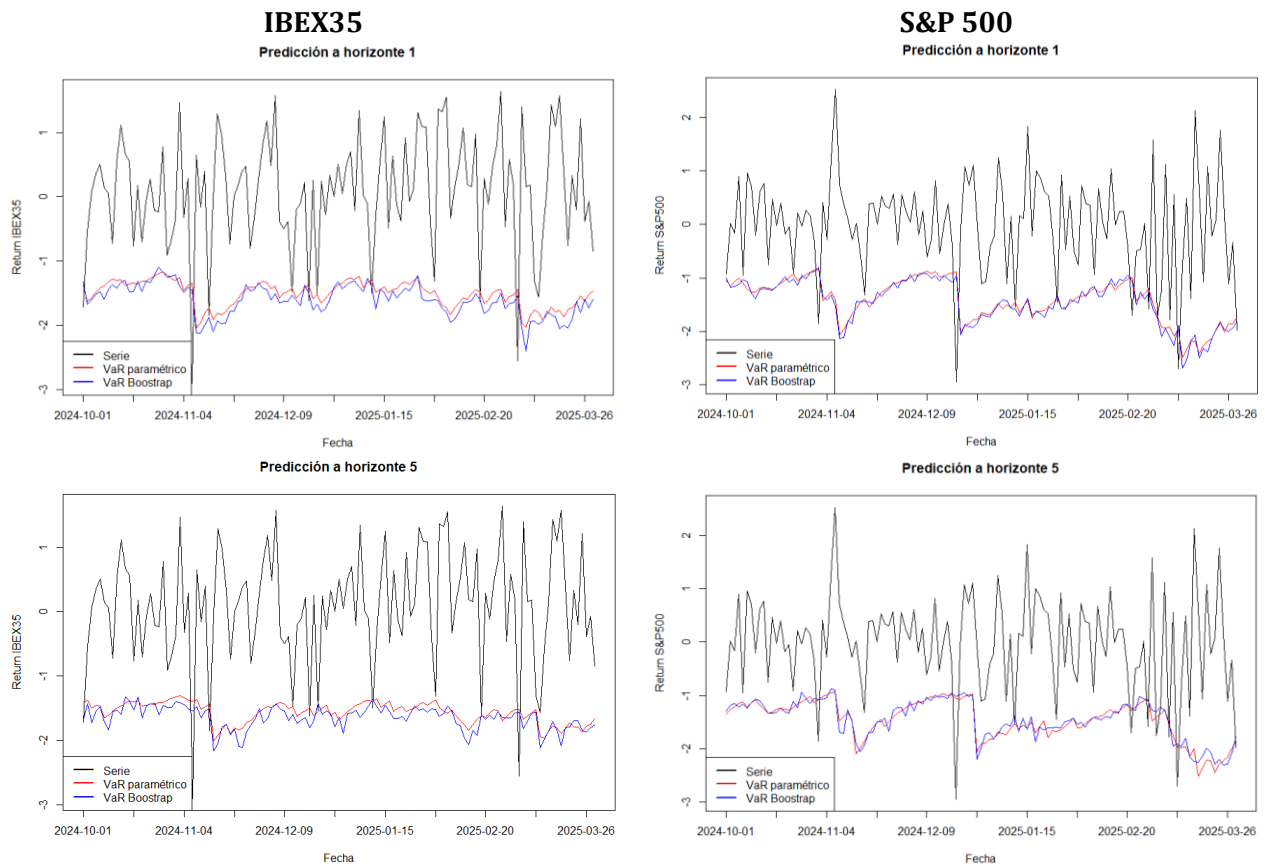
Cada uno de ellos mide una propiedad o característica diferente del VaR y nos da información sobre su ajuste a la serie real.

- El **test de exceso** comprueba si la frecuencia con la que las pérdidas reales supera el VaR estimado coincide con el nivel de significación teórico. Por ejemplo, si tenemos un VaR al 95%, se espera que aproximadamente el 5% de las observaciones superen ese umbral. Este contraste evalúa el nivel de cobertura global del VaR, pero no considera la dependencia temporal de violaciones.
- El **test de duración** analiza si las excedencias del VaR pasan de forma aleatoria a lo largo del tiempo o si, por el contrario, tienden a agruparse. Este contraste complementa al test de exceso ya que evalúa no solo la cantidad, sino también como se distribuye los excesos a lo largo del tiempo.
- El **test de independencia** combina las dos pruebas anteriores en un único contraste global. Comprueba al mismo tiempo si el número de violaciones es el esperado y si estas se distribuyen de manera aleatoria. Este contraste es especialmente útil cuando se va a comparar varios modelos, ya que ofrece una medida integral de la calidad del VaR.
- La **función de pérdida** evalúa el rendimiento del modelo VaR teniendo en consideración no solo si se producen violaciones, sino también la magnitud de las pérdidas cuando estas ocurren. Esta se calcula como el promedio de una función cuadrática o exponencial que penaliza las desviaciones entre las pérdidas reales y el VaR estimado. Cuanto menor sea el valor de la función de pérdida, mejor será el

ajuste del modelo. Si la función de pérdida es elevada, nos quiere decir que las diferencias entre el riesgo predicho y las pérdidas observadas son grandes, lo cual reduce la fiabilidad del VaR.

4.2.1 Evaluación gráfica del VaR estimado

Los gráficos se muestran en la Figura 5, siendo la primera columna para el IBEX35 y la segunda para el S&P500 mientras que la primera fila corresponde para horizonte 1 día, la segunda línea para horizonte 1 semana y, por último, la tercera línea para el horizonte 1 mes. En todos ellos, la línea roja representa el VaR con el método paramétrico y la línea azul representaría el VaR calculado a través del método Bootstrap, reflejando así la diferencia que cada uno de estos métodos.



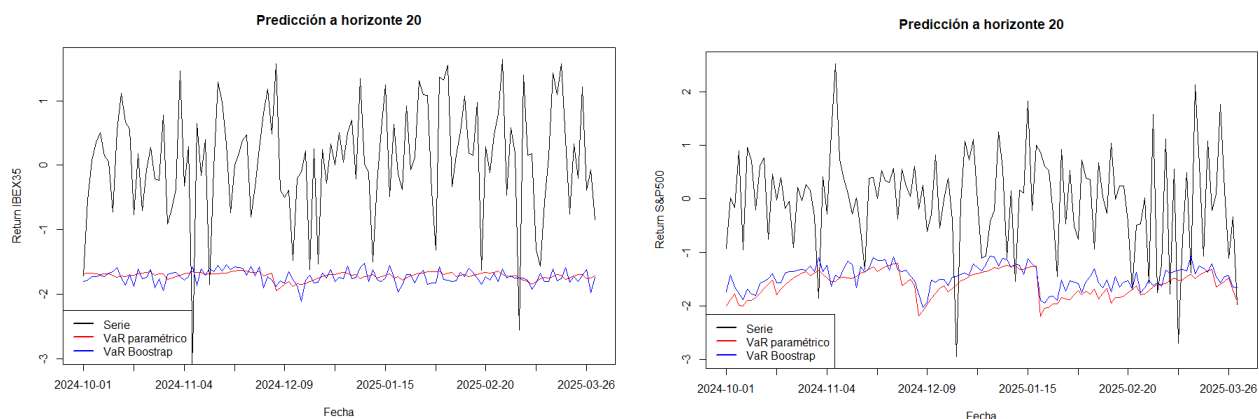


Figura 5. VaR estimado para horizontes $h=1, 5$ y 20 con un nivel del 95%

En general, podemos ver como los gráficos analizados reflejan de manera clara como la volatilidad y el riesgo de mercado evolucionan de forma distinta según el método de estimación empleado.

En el caso del método bootstrap podemos ver cómo se adapta mejor a la dinámica empírica del mercado y capta la realidad de un mercado que se expande con mayor intensidad durante periodos de tensión financiera. Esto es debido a que emplea directamente la función empírica de los residuos estandarizados, por lo tanto, no suaviza tanto como el caso paramétrico. Por otro lado, podemos ver como el método paramétrico refleja una realidad mucho más conservadora que nos lleva en algunos casos a infravalorar la magnitud del riesgo que hay en estos mercados. Su estructura teórica limita la respuesta ante variaciones muy extremas del mercado.

En cuanto al horizonte temporal de predicción, podemos observar que a medida que este se amplía, habrá un incremento general del VaR, lo cual es esperable ya que el riesgo acumulado crece con el tiempo. Sin embargo, las diferencias entre métodos se reducen para horizontes largos como los 20 días, donde la volatilidad promedio se regula de una forma más clara. En los horizontes cortos, el método Bootstrap tiende a generar valores de VaR más elevados, especialmente durante los periodos de crisis, lo que sugiere una mejor capacidad para capturar los choques de volatilidad de alta frecuencia.

4.2.2 Evaluación numérica del VaR estimado

A continuación, vamos a evaluar el comportamiento del VaR para ambos índices y métodos de cálculo. Se emplearán los contrastes y medidas enumerados anteriormente.

La Tabla 1 muestra los resultados para el índice IBEX 35. El test de exceso muestra en todos los horizontes, tanto en el método paramétrico como el bootstrap, p-valores superiores a $0,05$, lo cual nos dice que el número de violaciones del VaR se ajusta razonablemente bien al esperado. A horizontes de 1 y 5 días ambos métodos presentan el mismo p-valor (0.7305), esto sugiere que hay convergencia en la cobertura. Hay que

destacar que a 1 día el método bootstrap presenta el mejor ajuste, lo cual nos muestra que este reproduce mejor la frecuencia real de pérdidas extremas. A 20 días el método paramétrico obtiene un p-valor ligeramente mayor, mostrando buen desempeño en horizontes largos, aunque no se detectan problemas de cobertura, el método bootstrap destaca en el horizonte más corto.

El segundo test realizado de duración revela información importante sobre la estructura temporal de las violaciones. Destaca el p-valor a 1 día en el método bootstrap que nos está sugiriendo una posible dependencia temporal de las violaciones, que se producen con mayor intensidad en periodos de alta volatilidad. Por el contrario, el método paramétrico muestra un comportamiento más estable. En cuanto a los horizontes de 5 y 20 días, ambos métodos presentan p-valores altos, que nos indica que las violaciones se distribuyen aleatoriamente. Podemos ver cómo se concentran más violaciones en horizontes cortos, una señal típica de mercados sensibles a eventos inesperados como el Brexit.

Tabla 1. Test de bondad de ajuste del VaR para el IBEX 35

ÍNDICE IBEX 35					
Horizonte	Método	Test de exceso	Test de duración	Test de independencia	Función de pérdida
h=1	Paramétrico	1.2239 (0.2686)	1.37 (0.3200)	2.4938 (0.2874)	11.0468
	Bootstrap	0.1187 (0.7305)	2.21 (0.0500)	0.8462 (0.6550)	11.1134
h=5	Paramétrico	0.1187 (0.7305)	2.21 (0.0500)	0.8462 (0.6550)	11.1051
	Bootstrap	0.0039 (0.9503)	2.14 (0.0800)	0.5191 (0.7714)	11.0448
h=20	Paramétrico	0.8981 (0.3433)	1.15 (0.76)	1.1005 (0.5768)	10.8052
	Bootstrap	2.0773 (0.1495)	1.04 (0.9400)	2.2283 (0.3282)	10.0082

En el test de independencia podemos observar cómo ninguno de los métodos presenta deficiencias graves. Para horizontes de 1 y 5 días, los p-valores son todos mayores a 0.28, lo que significa que la cobertura y distribución temporal son correctas. A 20 días, los valores siguen siendo altos, lo que consolida la idea de que los modelos funcionan mejor en horizontes largos.

Por último, cuando vemos la función de pérdida podemos destacar que el mejor valor lo obtiene el método bootstrap a 20 días, lo que indica alta precisión en entornos más largos. En los horizontes cortos los valores son muy similares.

En la Tabla 2 podemos ver los resultados de la bondad de ajuste del VaR para el índice S&P 500. Se observa un comportamiento mucho más estable, con resultados numéricos que reflejan un mercado sólido, creciente y estable que pocos mercados financieros consiguen.

Con el primer test de excesos podemos observar cómo en todos los horizontes, los p-valores de ambos métodos son elevados. Esto nos hace comprobar ya esa estabilidad que hemos comentado anteriormente, viendo que las violaciones observadas están en los niveles esperados. Destaca que en el método paramétrico los horizontes cortos muestran una cobertura muy precisa mientras que en el bootstrap mejora en horizontes largos.

En el test de duración podemos ver como los p-valores son altos para cualquier horizonte. No se detecta agrupamiento de violaciones, no existen indicios de dependencia temporal significativa y tanto el método paramétrico como el bootstrap superan el test. Una vez más, vemos como este mercado estadounidense refleja solidez al corto y largo plazo.

Con el test de independencia podemos ver que los valores son altos en todos los horizontes y que conforme el horizonte de predicción es mayor el p-valor aumenta. Con esto podemos entender que en este mercado cuanto más a largo plazo vayamos, más estabilidad y mejores resultados vamos a obtener. Esto nos indica una buena precisión global, siendo así un mercado más predecible y menos severo. También cabe destacar que en el método bootstrap se vuelven a obtener mejores valores en horizontes largos.

Tabla 2. Test de bondad de ajuste del VaR para el S&P 500

ÍNDICE S&P 500					
Horizonte	Método	Test de exceso	Test de duración	Test de independencia	Función de pérdida
h=1	Paramétrico	1.2239 (0.2686)	1.32 (0.3600)	2.4938 (0.2874)	11.3900
	Bootstrap	0.5372 (0.4636)	1.26 (0.49)	5.99 (0.4686)	11.3160
h=5	Paramétrico	1.2239 (0.2686)	1.01 (0.9600)	2.4938 (0.2874)	11.2937
	Bootstrap	0.5372 (0.4636)	0.71 (0.2300)	0.9406 (0.6248)	10.8194
h=20	Paramétrico	0.5372 (0.4636)	1.2 (0.5800)	1.1005 (0.5768)	10.9706
	Bootstrap	2.0773 (0.1495)	1.04 (0.9400)	2.2283 (0.3282)	11.0082

Estos resultados obtenidos a partir de las gráficas anteriores nos dan una información donde podemos identificar importantes diferencias entre el IBEX 35 y el S&P 500. En el mercado

español podemos observar que los valores presentados tienden a ser más irregulares en las pruebas de duración e independencia, algo que refleja su mayor exposición a episodios de volatilidad intensa. Podemos afirmar que el mercado español tiene periodos de riesgo más acusados y concentrados. Por otro lado, el S&P 500 muestra una estructura de riesgo mucho más estable. En casi todos los casos los p-valores son altos, lo que como ya se ha explicado anteriormente nos aporta una cobertura más adecuada destacando la independencia temporal de las violaciones. Esto se debe a una mayor diversificación y profundidad del mercado estadounidense, donde participan las 500 empresas más importantes del país, a diferencia del mercado español, donde tan solo participan 35.

Analizando la función de pérdida podemos observar como los valores del S&P 500 son inferiores a los del mercado español. Esto nos muestra que los modelos son capaces de reproducir con una mayor precisión el comportamiento del riesgo en el mercado americano. Sin embargo, aunque este mercado sea más robusto y estable, cabe destacar que el mercado español goza de unos números muy favorables y que depositan una gran confianza en la mayoría de los inversores a nivel mundial.

5. CONCLUSIONES

El objetivo principal de este trabajo ha sido analizar el comportamiento del riesgo del IBEX 35 y S&P 500, a través de la estimación del VaR con los métodos bootstrap y paramétrico en diferentes esto, hemos horizontes temporales. Para lograr utilizado modelos ARMA-GARCH con distribución t-student para modelizar la volatilidad condicional. La estimación del VaR se ha realizado con las predicciones empleando los métodos mencionados anteriormente. La bondad de ajuste del VaR ha sido validada mediante contrastes estadísticos, como el test de exceso de cobertura, el test de duración, el test de independencia y el cálculo de la función de pérdida.

La primera conclusión que se ha podido obtener es que ambos mercados tienen periodos de relativa tranquilidad intercalados con periodos de alta incertidumbre. Sin embargo, conforme el horizonte temporal se amplía, vemos como ambos mercados permanecen estables y sin grandes perturbaciones que los desvíen de su tendencia alcista.

En segundo lugar, en cuanto al análisis comparativo entre métodos revela algunas diferencias sustanciales. El VaR paramétrico presenta un comportamiento más estable y suavizado, especialmente en horizontes cortos, debido a la estructura teórica de la distribución t-Student. Por otro lado, el método bootstrap destaca por su flexibilidad y capacidad de reproducir con gran precisión los fenómenos de cola y episodios de volatilidad extrema.

En tercer lugar, el análisis numérico confirma que ambos métodos superan los contrastes de bondad de ajuste en la mayoría de los casos, lo que indica que las estimaciones del VaR

son consistentes y fiables. Sin embargo, el método bootstrap muestra p-valores mas ajustados en los horizontes cortos del IBEX 35, lo que sugiere cierta dependencia temporal en las violaciones. Esta diferencia refuerza la idea de que los mercados como el español, los cuales son menos líquidos o más sensibles a choques externos como el Brexit o el Covid 19, presentan mayor concentración de episodios de riesgo extremo.

Por último, quiero comentar que los resultados indican que el VaR es una herramienta útil y válida para la gestión del riesgo de mercado, especialmente cuando se complementa con las pruebas de validación adecuadas y se tienen en cuenta las limitaciones inherentes a cada método. Así hemos podido comprobar que estos mercados tienen un riesgo moderado y que conforme el horizonte temporal va aumentando la estabilidad lo hace simultáneamente.

BIBLIOGRAFÍA

Fox, J., y Bouchet-Valat, M. (2024). Rcmdr: R Commander. R package version 2.9-5.

Miguel, J.A. y Gargallo, P. (2024). STUZ; Librería para series temporales. R package version 1.0.0

Novalés, A. (2013). Modelos ARCH univariantes y multivariantes. Dpto. Economía Cuantitativa. Universidad Complutense de Madrid.
<https://www.ucm.es/data/cont/media/www/pag-41460/Arch.pdf>

Skoog, J. y Enocksson, D., 2012. Evaluating VaR with the ARCH/GARCH Family. s.l.:Uppsala University.

Tsay, R. S., 2010. Analysis of Financial Time Series. Wiley.

ENLACES PARA EXTRACCIÓN DE DATOS

<https://es.finance.yahoo.com>

