

UNIVERSIDAD PONTIFICIA COMILLAS

Facultad de Ciencias Económicas y Empresariales (ICADE)

TRABAJO FIN DE GRADO

Grado en Administración y Dirección de Empresas (Mención Internacional)

**Modelos cuantitativos de arbitraje entre activos de
riesgo y volatilidad: aplicación con SPY y VXX**

Autor: Íñigo García-Fuertes

Tutor: Pedro Manuel Mirete Ferrer

Madrid, mayo de 2026

Resumen

El presente Trabajo Fin de Grado desarrolla un análisis cuantitativo de la dinámica conjunta del ETF SPY (réplica del S&P 500) y del ETN VXX (exposición a la volatilidad implícita de corto plazo del VIX) durante el periodo 2018-2026, con el objetivo de diseñar y evaluar una estrategia sistemática de pair trading entre ambos activos. Mediante un marco econométrico riguroso —tests de cointegración de Engle-Granger y Johansen, modelos ARIMA y VAR, modelos de volatilidad condicional GARCH y EGARCH, correlación dinámica condicional DCC-GARCH y calibración de un proceso de Ornstein-Uhlenbeck— se caracteriza la relación entre los dos activos y se construye un spread sintético mediante una hedge ratio dinámica. Los resultados confirman la existencia de una correlación negativa fuerte y persistente ($-0,75$) pero la ausencia de cointegración estricta en niveles, debido a la deriva descendente estructural del VXX. El spread construido con beta dinámica resulta, no obstante, estacionario, lo que habilita una estrategia de reversión a la media. El backtesting walk-forward muestra un desempeño out-of-sample favorable (ratio de Sharpe de $+1,21$ y ratio de Calmar de $+1,28$, superior al del SPY), aunque fuertemente dependiente del régimen de volatilidad: la estrategia es rentable en regímenes intermedios pero pierde en condiciones de alta volatilidad, refutando una de las hipótesis iniciales. El trabajo concluye con la propuesta de extensiones metodológicas, en particular la incorporación de un filtro de régimen.

Palabras clave: arbitraje estadístico, pair trading, volatilidad, VXX, SPY, cointegración, GARCH, Ornstein-Uhlenbeck, backtesting.

Abstract

This Bachelor's Thesis develops a quantitative analysis of the joint dynamics of the SPY ETF (tracking the S&P 500) and the VXX ETN (short-term VIX implied-volatility exposure) over the 2018-2026 period, aiming to design and evaluate a systematic pair trading strategy between the two assets. Through a rigorous econometric framework — Engle-Granger and Johansen cointegration tests, ARIMA and VAR models, GARCH and EGARCH conditional volatility models, DCC-GARCH dynamic conditional correlation, and calibration of an Ornstein-Uhlenbeck process— the relationship between the two assets is characterized and a synthetic spread is constructed using a dynamic hedge ratio.

The results confirm a strong, persistent negative correlation (-0.75) but the absence of strict cointegration in levels, owing to the structural downward drift of the VXX. The spread built with a dynamic beta is nonetheless stationary, enabling a mean-reversion strategy. Walk-forward backtesting shows favourable out-of-sample performance (Sharpe ratio of $+1.21$ and Calmar ratio of $+1.28$, exceeding that of the SPY), although strongly dependent on the volatility regime: the strategy is profitable in intermediate regimes but loses money under high volatility, refuting one of the initial hypotheses. The thesis concludes by proposing methodological extensions, notably the incorporation of a regime filter.

Keywords: statistical arbitrage, pair trading, volatility, VXX, SPY, cointegration, GARCH, Ornstein-Uhlenbeck, backtesting.

Índice de contenidos

Capítulo 1. Introducción

- 1.1. Contexto y planteamiento del problema
- 1.2. Motivación e interés del estudio
- 1.3. Justificación académica y profesional
- 1.4. Delimitación del periodo de estudio
- 1.5. Objetivos
 - 1.5.1. Objetivo general
 - 1.5.2. Objetivos específicos
- 1.6. Hipótesis de investigación
- 1.7. Metodología (visión general)
- 1.8. Estructura del trabajo

Capítulo 2. Marco teórico y revisión de la literatura

- 2.1. La hipótesis de los mercados eficientes y sus límites
 - 2.1.1. Formulación clásica
 - 2.1.2. Anomalías y críticas
 - 2.1.3. Los límites del arbitraje
 - 2.1.4. Implicaciones para el presente trabajo
- 2.2. Volatilidad implícita y el índice VIX
 - 2.2.1. Volatilidad histórica y volatilidad implícita
 - 2.2.2. El índice VIX
 - 2.2.3. La prima de riesgo de la volatilidad
- 2.3. Productos cotizados sobre volatilidad: el caso VXX
 - 2.3.1. Futuros del VIX y la curva forward
 - 2.3.2. Roll yield y deriva de los productos sobre volatilidad
 - 2.3.3. Mecánica del VXX (iPath Series B)

- 2.3.4. Otros productos sobre volatilidad: contexto
- 2.4. Arbitraje estadístico y pair trading
 - 2.4.1. Concepto y fundamento del arbitraje estadístico
 - 2.4.2. Pair trading clásico: el enfoque de distancia
 - 2.4.3. Pair trading basado en cointegración
 - 2.4.4. Pair trading basado en modelos estocásticos: Ornstein-Uhlenbeck
 - 2.4.5. Métricas de evaluación
- 2.5. Modelos econométricos relevantes
 - 2.5.1. Estacionariedad y raíz unitaria
 - 2.5.2. Cointegración: Engle-Granger y Johansen
 - 2.5.3. Modelos ARIMA y VAR
 - 2.5.4. Modelos GARCH y EGARCH
 - 2.5.5. Correlación dinámica condicional: DCC-GARCH
 - 2.5.6. El proceso de Ornstein-Uhlenbeck
- 2.6. Estado del arte: el par SPY-VXX en la literatura
- 2.7. Síntesis y posicionamiento del trabajo

Capítulo 3. Metodología

- 3.1. Visión general del proceso de investigación
- 3.2. Fuentes y depuración de los datos
 - 3.2.1. Fuente
 - 3.2.2. Periodo de análisis
 - 3.2.3. Depuración
- 3.3. Análisis exploratorio y caracterización estadística
 - 3.3.1. Estadísticos descriptivos y momentos
 - 3.3.2. Tests de normalidad
 - 3.3.3. Tests de estacionariedad: ADF y KPSS
 - 3.3.4. Análisis de correlación

3.4. Modelización econométrica

3.4.1. Tests de cointegración

3.4.2. Modelos ARIMA univariantes

3.4.3. Modelo VAR y causalidad de Granger

3.4.4. Modelos GARCH y EGARCH

3.4.5. Modelo DCC-GARCH

3.5. Diseño de la estrategia de pair trading

3.5.1. Construcción del spread dinámico

3.5.2. Verificación de la estacionariedad del spread

3.5.3. Calibración del proceso de Ornstein-Uhlenbeck

3.5.4. Generación de señales

3.5.5. Dimensionado y gestión del riesgo

3.6. Backtesting y validación

3.6.1. División de la muestra

3.6.2. Métricas de evaluación

3.6.3. Estrategias de referencia

3.6.4. Análisis de sensibilidad y robustez

3.7. Soporte computacional y reproducibilidad

3.7.1. Entorno técnico

3.7.2. Replicación en Excel

3.7.3. Criterios de reproducibilidad

3.8. Limitaciones metodológicas

Capítulo 4. Datos y análisis exploratorio

4.1. Datos utilizados

4.1.1. Fuente y periodo

4.1.2. Variables y transformaciones

4.2. Análisis univariante de los activos

- 4.2.1. Evolución temporal de los precios
- 4.2.2. Estadísticos descriptivos de los retornos
- 4.2.3. Distribución empírica de los retornos
- 4.2.4. Tests de estacionariedad
- 4.2.5. Autocorrelación y heterocedasticidad
- 4.3. Análisis bivalente: la relación SPY-VXX
 - 4.3.1. Correlaciones contemporáneas
 - 4.3.2. Evolución temporal de la correlación
 - 4.3.3. Eventos extremos
 - 4.3.4. Asimetría de la respuesta del VXX
- 4.4. Regímenes de volatilidad y comparación con el VIX
 - 4.4.1. El VIX como variable de contexto
 - 4.4.2. Correlación condicional al régimen de volatilidad
- 4.5. Hechos estilizados y síntesis

Capítulo 5. Modelización econométrica

- 5.1. Tests de cointegración
 - 5.1.1. Test de Engle-Granger
 - 5.1.2. Test de Johansen
 - 5.1.3. Interpretación económica
- 5.2. Modelos ARIMA univariantes
- 5.3. Modelo VAR y causalidad de Granger
 - 5.3.1. Selección del orden y estimación
 - 5.3.2. Causalidad de Granger
- 5.4. Modelos de volatilidad condicional
 - 5.4.1. GARCH(1,1)
 - 5.4.2. EGARCH(1,1) y leverage effect
 - 5.4.3. Comparación GARCH vs EGARCH

- 5.4.4. Volatilidades condicionales estimadas
- 5.4.5. Diagnóstico de los residuos estandarizados
- 5.5. Correlación dinámica condicional: DCC-GARCH
 - 5.5.1. Parámetros estimados
 - 5.5.2. Evolución de la correlación condicional
 - 5.5.3. Comparación con la correlación rolling
- 5.6. Construcción del spread dinámico SPY-VXX
 - 5.6.1. Hedge ratio dinámica
 - 5.6.2. Spread sintético y test de estacionariedad
 - 5.6.3. Calibración del proceso de Ornstein-Uhlenbeck
 - 5.6.4. Z-score del spread y umbrales operativos
- 5.7. Síntesis de la modelización econométrica

Capítulo 6. Diseño y calibración de la estrategia de pair trading

- 6.1. Estructura de las posiciones
 - 6.1.1. Definición de las dos piernas
 - 6.1.2. Normalización del capital y P&L diario
 - 6.1.3. Beta fija a la entrada vs. beta dinámica
- 6.2. Reglas de operación
 - 6.2.1. Señal de entrada
 - 6.2.2. Señales de cierre
 - 6.2.3. Convención de ejecución
 - 6.2.4. Visualización de las reglas operativas
- 6.3. Dimensionado y gestión del riesgo
 - 6.3.1. Asignación de capital
 - 6.3.2. Limitación a una posición simultánea
 - 6.3.3. Riesgo residual y consideraciones prácticas
- 6.4. Modelización de los costes de transacción

- 6.4.1. Componentes del coste
- 6.4.2. Calibración del coste agregado
- 6.4.3. Sensibilidad de los resultados al supuesto de costes
- 6.5. Calibración de la estrategia: muestra in-sample (2018–2022)
 - 6.5.1. División de la muestra
 - 6.5.2. Parámetros calibrados
 - 6.5.3. Resultados in-sample
 - 6.5.4. Estadísticos operativos in-sample

Capítulo 7. Backtesting y resultados

- 7.1. Desempeño global de la estrategia
 - 7.1.1. Curva de equity
 - 7.1.2. Métricas globales
 - 7.1.3. Drawdown
- 7.2. Resultados out-of-sample: la prueba clave
 - 7.2.1. Importancia metodológica
 - 7.2.2. Comparación de los desempeños in-sample y out-of-sample
 - 7.2.3. Evolución anual del desempeño
- 7.3. Análisis operativo de las posiciones
 - 7.3.1. Estadísticos agregados de las operaciones
 - 7.3.2. Distribución del PnL por operación
 - 7.3.3. Episodios destacados
- 7.4. Análisis de sensibilidad y robustez
 - 7.4.1. Sensibilidad al umbral de entrada
 - 7.4.2. Sensibilidad a los costes de transacción
 - 7.4.3. Sensibilidad a la ventana de estimación
 - 7.4.4. Síntesis de robustez
- 7.5. Análisis por régimen de volatilidad

7.6. Contrastación de las hipótesis

7.6.1. Lectura integrada de los resultados

Capítulo 8. Conclusiones

8.1. Síntesis de los hallazgos principales

8.2. Contrastación de las hipótesis y aportaciones empíricas

8.3. Limitaciones del trabajo

8.4. Líneas de investigación futura

8.4.1. Incorporación de un filtro de régimen

8.4.2. Optimización dinámica de los umbrales operativos

8.4.3. Extensión a otros pares activo-volatilidad

8.4.4. Modelización de la dinámica intradiaria

8.4.5. Combinación con otros factores cuantitativos

8.5. Implicaciones académicas y profesionales

8.6. Reflexión final

Capítulo 1. Introducción

1.1. Contexto y planteamiento del problema

Los mercados financieros contemporáneos se caracterizan por una creciente complejidad estructural, una velocidad de ejecución cada vez mayor y un grado de interconexión entre activos que apenas era imaginable hace dos décadas. La proliferación de la negociación algorítmica, la fragmentación de la liquidez entre centros de negociación y la consolidación de los productos cotizados que replican factores de riesgo —volatilidad, momentum, calidad o tamaño— han transformado de manera profunda la microestructura del mercado y la forma en que se forman los precios (López de Prado, 2018; Hasbrouck, 2007).

En este entorno, las estrategias cuantitativas de arbitraje estadístico han adquirido una relevancia creciente. A diferencia del arbitraje clásico, que explota desviaciones puntuales y sin riesgo respecto de una relación de paridad, el arbitraje estadístico se fundamenta en relaciones de equilibrio probabilísticas entre activos, identificadas

mediante herramientas econométricas, y asume un riesgo residual que se gestiona estadísticamente (Avellaneda & Lee, 2010; Vidyamurthy, 2004). Su lógica descansa en la hipótesis de que, ante desequilibrios temporales en la relación entre dos o más activos correlacionados, los precios tenderán a converger nuevamente hacia su senda de equilibrio de largo plazo, ofreciendo así oportunidades de rentabilidad ajustada al riesgo.

Entre los pares de activos que han concentrado mayor atención académica y profesional en este campo destacan aquellos que enfrentan a un activo de riesgo direccional con un activo de volatilidad. El caso paradigmático lo constituye la pareja formada por el SPY —el ETF *SPDR S&P 500 Trust*, gestionado por State Street, que replica el comportamiento del índice S&P 500— y el VXX —el ETN *iPath Series B S&P 500 VIX Short-Term Futures*, emitido por Barclays Bank PLC desde el 18 de enero de 2018, que ofrece exposición continua a los futuros del VIX con vencimiento en los dos meses más próximos—. Ambos instrumentos cotizan en NYSE Arca y son, respectivamente, el ETF más líquido del mundo sobre renta variable estadounidense y uno de los productos cotizados de mayor volumen sobre volatilidad implícita (CBOE, 2023; State Street Global Advisors, 2024).

La relación entre el SPY y el VXX presenta una característica económica fundamental: una correlación negativa estructural que se intensifica en episodios de estrés financiero. Cuando el mercado de renta variable estadounidense experimenta caídas significativas, la volatilidad implícita —medida por el VIX y, por extensión, por los futuros que subyacen al VXX— tiende a repuntar de forma desproporcionada, en un fenómeno conocido como *leverage effect* o *asymmetric volatility* (Black, 1976; Bekaert & Wu, 2000). Este efecto, ampliamente documentado desde las primeras formulaciones del modelo EGARCH de Nelson (1991), implica que las innovaciones negativas sobre los retornos del subyacente generan un incremento mayor de la volatilidad esperada que las innovaciones positivas de igual magnitud.

Ahora bien, la correlación negativa entre el SPY y el VXX no constituye una relación determinista ni temporalmente estable. En el corto plazo, la dinámica del par puede desviarse de su patrón de largo plazo por múltiples factores. En primer lugar, la microestructura específica de los futuros del VIX, que cotizan habitualmente en *contango* —es decir, con precios más altos en vencimientos lejanos que en cercanos—, lo que implica que la cartera teórica subyacente al VXX incurre en una pérdida sistemática al

rebalancear continuamente las posiciones (efecto *roll yield* negativo). En segundo lugar, los choques idiosincrásicos sobre la liquidez de los futuros del VIX y las distorsiones puntuales generadas por flujos de cobertura institucionales (Alexander & Korovilas, 2013; Whaley, 2013). La combinación del *leverage effect* con la pérdida estructural de valor del VXX por el efecto del *contango* genera precisamente la asimetría de la que se nutren las estrategias cuantitativas: episodios de desviación temporal respecto del equilibrio que pueden ser explotados de manera sistemática.

A este respecto, los acontecimientos de los últimos años proporcionan abundante evidencia empírica. El episodio conocido como *Volmageddon* del 5 de febrero de 2018 que provocó la liquidación del XIV de Credit Suisse y un cambio estructural en la industria de productos cotizados sobre volatilidad—, la pandemia de COVID-19 en marzo de 2020, las tensiones inflacionarias y monetarias de 2022, la crisis bancaria regional estadounidense de marzo de 2023 y los episodios de volatilidad inducida por las medidas arancelarias del primer semestre de 2025, han sido escenarios en los que la dinámica entre el S&P 500 y la volatilidad implícita ha mostrado patrones extraordinarios, con expansiones del VIX por encima del cuantil 95 % de su distribución histórica y caídas correspondientes del SPY (CBOE, 2025). La pregunta relevante para un investigador en finanzas cuantitativas es si estas dinámicas pueden ser modelizadas de forma rigurosa y, en su caso, si las desviaciones detectadas pueden ser explotadas de manera sistemática mediante una estrategia de inversión replicable y robusta.

1.2. Motivación e interés del estudio

La motivación del presente Trabajo Fin de Grado nace de la confluencia entre tres dimensiones complementarias.

En primer lugar, una dimensión académica. Si bien la literatura sobre cointegración (Engle & Granger, 1987; Johansen, 1991), modelos de volatilidad condicional (Bollerslev, 1986; Nelson, 1991) y estrategias de pairs trading (Gatev, Goetzmann & Rouwenhorst, 2006) es extensa, la aplicación conjunta de estas herramientas al par SPY-VXX ha recibido una atención comparativamente menor en el ámbito académico, pese a su relevancia profesional. Existe, por tanto, un espacio para contribuir con evidencia empírica reciente sobre los mecanismos de transmisión entre la renta variable estadounidense y la volatilidad implícita, así como sobre la viabilidad práctica de explotar estas relaciones mediante modelos cuantitativos.

En segundo lugar, una dimensión metodológica. La integración de herramientas econométricas avanzadas —cointegración, modelos ARIMA, GARCH, EGARCH, procesos de Ornstein-Uhlenbeck— con técnicas modernas de programación científica (Python, en particular las librerías *statsmodels*, *arch*, *pandas* y *numpy*) y con instrumentos de validación empresarial habituales (Microsoft Excel) refleja con fidelidad el flujo de trabajo de los equipos de investigación cuantitativa en bancos de inversión, *hedge funds* y casas de gestión de activos. El presente trabajo busca demostrar que dicha integración es posible y productiva incluso desde un entorno académico.

En tercer lugar, una dimensión profesional. Los activos seleccionados no son arbitrarios: el SPY y el VXX se encuentran entre los productos cotizados de mayor volumen del mundo y forman parte del universo operativo cotidiano de las mesas de negociación de equity y de volatilidad de cualquier banco de inversión global. Comprender la dinámica conjunta de ambos activos, las propiedades estadísticas de su spread y la viabilidad de las estrategias que pueden construirse sobre dicha relación constituye, en consecuencia, una herramienta directamente aplicable al ejercicio profesional en *trading*, *risk management* y *portfolio construction*.

1.3. Justificación académica y profesional

La elección de este tema se fundamenta en una trayectoria académica orientada a las finanzas cuantitativas y en una experiencia profesional progresivamente especializada en mercados de capitales.

Desde el punto de vista académico, el Grado en Administración y Dirección de Empresas con Mención Internacional cursado en Universidad Pontificia Comillas (ICADE), con estancia académica en Dublín, ha proporcionado los fundamentos en economía financiera, econometría, teoría de la cartera y derivados que sustentan este trabajo. La formación recibida en asignaturas como *Análisis Financiero*, *Mercados de Capitales* y *Matemáticas Financieras* aporta el armazón teórico sobre el que se desarrolla la modelización empírica.

Desde el punto de vista profesional, la experiencia adquirida en Euronext —en sus oficinas de Dublín y Milán— me ha permitido adentrarme en la microestructura de los mercados financieros, particularmente en la formación de precios y en la mecánica operativa de los mercados de renta fija. Posteriormente, la creación y gestión de Saint

James Capital, vehículo en el que se desarrolló un primer algoritmo propio de arbitraje estadístico, ha consolidado las habilidades de modelización financiera, análisis de series temporales y gestión del riesgo de cartera. Finalmente, la incorporación como junior trader en BBVA en enero de 2026 ofrece un marco de aplicación profesional que permite contrastar las hipótesis y los resultados de este trabajo con la operativa real de una entidad financiera de primer nivel.

Este Trabajo Fin de Grado pretende, en consecuencia, integrar la formación recibida con la experiencia profesional adquirida, generando un producto académico que no solo cumpla con los estándares de rigor exigibles a una investigación de fin de grado, sino que constituya además una herramienta útil para la práctica profesional en mercados financieros.

1.4. Delimitación del periodo de estudio

El periodo de análisis se extiende desde el 18 de enero de 2018 hasta el 28 de abril de 2026, comprendiendo aproximadamente 2.080 sesiones bursátiles. La elección de la fecha de inicio responde a una motivación metodológica explícita: el VXX original (iPath S&P 500 VIX Short-Term Futures ETN, ticker original VXX) fue lanzado en 2009 y maduró el 30 de enero de 2019. El VXX actualmente disponible en mercado, sobre el que se construye este trabajo, es la *iPath Series B S&P 500 VIX Short-Term Futures ETN*, emitida por Barclays Bank PLC e iniciada el 17 de enero de 2018. Empalmar artificialmente las dos series mediante factores de ajuste introduciría un sesgo metodológico evitable, dado que las dos emisiones, pese a replicar el mismo subyacente, presentan particularidades técnicas (régimen fiscal, comisiones, mecanismo de redención) que afectan al precio observado.

El periodo seleccionado abarca, además, episodios de mercado especialmente relevantes para el análisis de la dinámica volatilidad-renta variable: el *Volmageddon* del 5 de febrero de 2018, la crisis del COVID-19 de marzo de 2020, las tensiones monetarias e inflacionarias de 2022, la crisis bancaria regional estadounidense de marzo de 2023, los episodios de volatilidad inducida por la política arancelaria estadounidense del primer semestre de 2025 y la fase de normalización posterior. La heterogeneidad de regímenes contenida en estos algo más de ocho años proporciona una base empírica sólida para evaluar la robustez de los modelos propuestos.

1.5. Objetivos

1.5.1. Objetivo general

El presente Trabajo Fin de Grado persigue analizar cuantitativamente la relación dinámica entre el SPY y el VXX durante el periodo 2018-2026, con el fin de identificar la existencia de oportunidades de arbitraje estadístico y de evaluar el desempeño de una estrategia de pair trading construida sobre dicha relación, mediante la aplicación de modelos econométricos contrastados y de técnicas de validación fuera de muestra.

1.5.2. Objetivos específicos

Para alcanzar el objetivo general, se plantean los siguientes objetivos específicos:

OE1. Examinar la relación teórica entre los retornos del S&P 500 y la volatilidad implícita, así como la mecánica específica del VXX como vehículo de exposición a futuros del VIX, identificando los principales factores estructurales (efecto *contango*, *roll yield* negativo) que afectan a su dinámica conjunta.

OE2. Caracterizar empíricamente las propiedades estadísticas de las series del SPY y el VXX —estacionariedad, distribución de retornos, momentos de orden superior, *clustering* de volatilidad— durante el periodo de estudio, así como la evolución temporal de su correlación.

OE3. Aplicar y comparar distintos modelos econométricos para describir la relación entre los dos activos: (i) tests de cointegración de Engle-Granger y Johansen sobre niveles y log-niveles; (ii) modelos ARIMA y VAR sobre los retornos; (iii) modelos GARCH y EGARCH para la transmisión de volatilidad; (iv) modelos de correlación dinámica condicional (DCC-GARCH).

OE4. Modelizar la dinámica del *spread* entre el SPY y el VXX mediante un proceso de Ornstein-Uhlenbeck, calibrar sus parámetros mediante máxima verosimilitud y derivar señales de entrada y salida basadas en el Z-score del *spread* respecto de su media de equilibrio.

OE5. Diseñar e implementar una estrategia de pair trading sistemática sobre la pareja SPY-VXX, definiendo de manera explícita los criterios de apertura, cierre, dimensionado y gestión del riesgo de las posiciones, así como una *hedge ratio* (*beta*) dinámica estimada por ventana móvil.

OE6. Evaluar la rentabilidad y el riesgo de la estrategia mediante un proceso de *backtesting* dividido en una fase de entrenamiento (*in-sample*) y una fase de validación (*out-of-sample*), utilizando métricas estándar del análisis cuantitativo: ratio de Sharpe, ratio de Sortino, ratio de Calmar, *maximum drawdown*, *hit ratio* y duración media de las operaciones.

OE7. Contrastar la robustez de la estrategia frente a regímenes de mercado diferenciados (alta y baja volatilidad), realizando análisis de sensibilidad respecto de los principales parámetros del modelo y comparando los resultados con estrategias de referencia (*buy-and-hold* del SPY, posición *short* sobre el VXX).

OE8. Documentar el proceso completo de investigación de manera reproducible, integrando código en Python (Jupyter Notebook) y modelos en Microsoft Excel, de modo que los resultados puedan ser verificados, replicados y extendidos.

1.6. Hipótesis de investigación

A partir de los objetivos planteados y de la revisión preliminar de la literatura, se formulan las siguientes hipótesis de investigación, que serán contrastadas a lo largo del trabajo:

H1. Existe una relación de correlación negativa estadísticamente significativa entre los retornos diarios del SPY y los retornos diarios del VXX durante el periodo 2018-2026, con magnitud variable a lo largo del tiempo y reforzada en episodios de estrés de mercado.

H2. Las series de precios del SPY y del VXX, pese a su elevada correlación negativa, no presentan una relación de cointegración en el sentido de Engle-Granger ni de Johansen sobre el periodo 2018-2026, debido a la deriva determinista descendente del VXX inducida por el efecto *contango*. Esta hipótesis, contrariamente al planteamiento clásico del *pair trading*, anticipa un resultado negativo para los tests de cointegración estándar.

H3. El *spread* entre el SPY y el VXX, una vez ajustado por una *hedge ratio* (*beta*) dinámica estimada por ventana móvil sobre los retornos, exhibe propiedades de reversión a la media compatibles con un proceso de Ornstein-Uhlenbeck con velocidad de reversión positiva y estadísticamente significativa.

H4. Una estrategia de *pair trading* sistemática construida sobre las desviaciones del *spread* dinámico respecto de su valor de equilibrio genera, durante el periodo de validación *out-of-sample*, retornos ajustados al riesgo (medidos por el ratio de Sharpe)

superiores a los de una estrategia *buy-and-hold* sobre el SPY, una vez descontados los costes de transacción razonables.

H5. La rentabilidad de la estrategia presenta una sensibilidad positiva al régimen de volatilidad: los retornos ajustados al riesgo de la estrategia son superiores en periodos de alta volatilidad implícita que en periodos de baja volatilidad.

1.7. Metodología (visión general)

La metodología seguida en este trabajo se articula en cinco fases secuenciales, que serán desarrolladas en detalle en el Capítulo 3:

1. **Recopilación y depuración de datos.** Se utilizarán precios diarios del SPY y el VXX procedentes de FactSet, así como del índice VIX procedente del CBOE, para el periodo 2018-01-18 a 2026-04-28. Las series se ajustarán por dividendos (*Total Return Gross*) para asegurar consistencia temporal.

2. **Análisis exploratorio y caracterización estadística.** Se calcularán estadísticos descriptivos, momentos de orden superior, tests de normalidad (Jarque-Bera) y de estacionariedad (ADF y KPSS), y se examinará la evolución temporal de la correlación entre ambos activos.

3. **Modelización econométrica.** Se aplicarán los tests de cointegración de Engle-Granger y Johansen, modelos ARIMA y VAR sobre los retornos, modelos GARCH y EGARCH para describir la dinámica de la volatilidad, y un modelo de correlación dinámica condicional (DCC-GARCH) para capturar la evolución temporal de la dependencia.

4. **Diseño y calibración de la estrategia.** Se construirá un *spread* sintético entre el SPY y el VXX a partir de una *beta* dinámica estimada por OLS sobre ventana móvil, y se modelizará dicho *spread* como un proceso de Ornstein-Uhlenbeck. Las señales de operación se definirán a partir del Z-score del *spread* respecto de su media de equilibrio.

5. **Backtesting y validación.** La estrategia se calibrará en una muestra de entrenamiento (*in-sample*: 2018-2022) y se evaluará en una muestra de validación (*out-of-sample*: 2023-2026). El desempeño se medirá con un conjunto estándar de métricas ajustadas al riesgo y se contrastará la robustez frente a variaciones de los parámetros del modelo.

El soporte computacional combina Python —principal herramienta para el procesamiento de datos, estimación econométrica y *backtesting*, mediante las *librerías pandas, numpy, statsmodels, arch, scipy y matplotlib*, con Microsoft Excel, utilizado para la replicación independiente del *backtesting* y la verificación de los principales cálculos, así como para la presentación visual de los resultados a un público no técnico.

1.8. Estructura del trabajo

El presente Trabajo Fin de Grado se organiza en ocho capítulos que recorren, de manera secuencial, las dimensiones teórica, empírica y aplicada del problema planteado:

- **Capítulo 1. Introducción.** Contextualiza el tema, justifica la elección, define los objetivos, formula las hipótesis y resume la metodología y la estructura del trabajo.
- **Capítulo 2. Marco teórico y revisión de la literatura.** Expone los fundamentos conceptuales del estudio: la teoría de los mercados eficientes y sus límites, la naturaleza de la volatilidad implícita y el papel del VIX, la mecánica específica del VXX y los productos sobre volatilidad, los principios del arbitraje estadístico y del pair trading, y los principales modelos econométricos relevantes.
- **Capítulo 3. Metodología.** Describe en detalle las técnicas estadísticas y econométricas empleadas: tests de estacionariedad y cointegración, modelos de series temporales (ARIMA, VAR), modelos de volatilidad condicional (GARCH, EGARCH, DCC-GARCH), procesos de Ornstein-Uhlenbeck y métricas de evaluación del desempeño.
- **Capítulo 4. Datos y análisis exploratorio.** Presenta las fuentes de datos, las series utilizadas, el proceso de depuración y los resultados del análisis exploratorio inicial: estadísticos descriptivos, distribuciones, correlaciones y representación gráfica.
- **Capítulo 5. Modelización econométrica.** Reporta los resultados de los tests de cointegración, los modelos ARIMA/VAR ajustados, las estimaciones GARCH/EGARCH y el análisis de la correlación dinámica.
- **Capítulo 6. Diseño y calibración de la estrategia de pair trading.** Construye el *spread* dinámico entre el SPY y el VXX, modeliza su dinámica de reversión a la media mediante un proceso de Ornstein-Uhlenbeck, calibra sus parámetros y define las reglas operativas de la estrategia.

- **Capítulo 7. Backtesting y resultados.** Presenta los resultados del *backtesting* en las fases *in-sample* y *out-of-sample*, las métricas de rendimiento y riesgo, los análisis de sensibilidad y la comparación con estrategias de referencia.
- **Capítulo 8. Conclusiones.** Sintetiza los principales hallazgos, contrasta las hipótesis formuladas, discute las limitaciones del estudio y propone líneas de investigación futura.

El trabajo se cierra con la bibliografía completa y un conjunto de anexos que incluyen el código Python desarrollado (Jupyter Notebook), el modelo Excel de *backtesting* replicado, las tablas auxiliares y los resultados detallados de robustez.

Capítulo 2. Marco teórico y revisión de la literatura

El presente capítulo establece el armazón conceptual sobre el que se asienta el análisis empírico desarrollado en los capítulos posteriores. Se estructura en seis bloques: en primer lugar (§2.1) se examina la hipótesis de los mercados eficientes y sus principales limitaciones, en tanto que constituyen el contexto teórico dentro del cual cobra sentido la búsqueda de oportunidades de arbitraje estadístico; a continuación (§2.2) se abordan los fundamentos de la volatilidad implícita y el índice VIX; el §2.3 describe la mecánica específica del VXX y los productos cotizados sobre volatilidad; el §2.4 sintetiza los principios del arbitraje estadístico y del pair trading; el §2.5 presenta los modelos econométricos que serán empleados en la parte empírica del trabajo; y, finalmente, el §2.6 revisa el estado del arte de la literatura aplicada al par SPY-VXX, posicionando este Trabajo Fin de Grado dentro del conjunto de contribuciones existentes.

2.1. La hipótesis de los mercados eficientes y sus límites

2.1.1. Formulación clásica

La hipótesis de los mercados eficientes (*Efficient Market Hypothesis*, EMH) es uno de los pilares de la moderna teoría de las finanzas. Formulada en su versión más influyente por Fama (1965, 1970), establece que los precios de los activos financieros incorporan plenamente toda la información relevante disponible, de modo que ningún inversor puede obtener de forma sistemática rendimientos ajustados al riesgo superiores a los del mercado mediante el análisis de dicha información. Fama (1970) distingue tres formas de eficiencia, atendiendo al conjunto informativo considerado: la eficiencia débil (los

precios incorporan toda la información histórica de precios), la semifuerte (incorporan toda la información pública disponible) y la fuerte (incorporan incluso la información privada).

Bajo cualquiera de las tres formas de la EMH, las desviaciones respecto del valor fundamental son aleatorias e impredecibles, y los precios siguen aproximadamente un *random walk* (Samuelson, 1965). En este marco, las estrategias de inversión activa carecerían de fundamento racional, y la gestión pasiva sería, en términos esperados, óptima.

2.1.2. Anomalías y críticas

Desde finales de los años setenta, una creciente literatura ha documentado anomalías sistemáticas que cuestionan la validez práctica de la EMH. Entre las más relevantes cabe destacar el efecto tamaño (Banz, 1981), el efecto valor (Fama & French, 1992), el momentum (Jegadeesh & Titman, 1993), las reversiones de largo plazo (De Bondt & Thaler, 1985) y la *post-earnings announcement drift* (Bernard & Thomas, 1989). Estas anomalías, lejos de desaparecer tras su descubrimiento, han mostrado en muchos casos persistencia temporal —si bien con magnitudes variables— y han dado origen a familias enteras de estrategias de inversión sistemática.

La aparición de estas anomalías ha impulsado el desarrollo de la denominada *behavioral finance* (Shleifer, 2000; Barberis & Thaler, 2003), que incorpora elementos de la psicología cognitiva para explicar desviaciones de la racionalidad agregada de los mercados, así como del programa de investigación sobre los límites del arbitraje (Shleifer & Vishny, 1997). Este último resulta particularmente relevante para el presente trabajo.

2.1.3. Los límites del arbitraje

El argumento clásico que sustenta la EMH descansa en la actuación de los arbitrajistas: aunque algunos inversores se comporten de manera irracional, los arbitrajistas profesionales corregirían rápidamente cualquier desviación del precio respecto del valor fundamental, restableciendo la eficiencia. Shleifer y Vishny (1997) cuestionan empíricamente esta visión. Argumentan que el arbitraje en el mundo real es realizado por un número limitado de agentes especializados que gestionan capital ajeno, y que se enfrentan a restricciones de capital, riesgos de financiación y horizontes finitos. En presencia de estas fricciones, los arbitrajistas pueden no estar en condiciones de mantener

una posición lo suficientemente prolongada como para que se materialice la convergencia, lo que permite la persistencia de desviaciones del precio respecto del fundamental.

De Long, Shleifer, Summers y Waldmann (1990) introducen además el concepto de *noise trader risk*: en presencia de inversores no informados cuyas decisiones afectan a los precios, los arbitrajistas asumen un riesgo adicional —el de que los precios se alejen aún más del fundamental antes de converger—, lo que limita su agresividad operativa.

Estas limitaciones son particularmente relevantes en mercados donde existen barreras técnicas, costes de financiación elevados o restricciones a las posiciones cortas, como es el caso de los productos cotizados sobre volatilidad. La existencia de un *contango* sistemático en los futuros del VIX (analizada en §2.3) constituye, por ejemplo, un caso paradigmático de desviación persistente que no es plenamente arbitrada por los participantes del mercado, pese a su predictibilidad estadística.

2.1.4. Implicaciones para el presente trabajo

La hipótesis subyacente a la estrategia desarrollada en este Trabajo Fin de Grado no es la negación radical de la EMH, sino la afirmación más matizada —y empíricamente sustentada— de que existen ineficiencias temporales y persistentes en la relación entre activos correlacionados que pueden ser detectadas mediante herramientas econométricas y explotadas dentro de los límites impuestos por los costes de transacción y los riesgos de financiación. La cuestión empírica que el trabajo trata de resolver es si dichas ineficiencias son lo suficientemente robustas como para sostener una estrategia sistemática con rendimientos ajustados al riesgo positivos en condiciones de mercado realistas.

2.2. Volatilidad implícita y el índice VIX

2.2.1. Volatilidad histórica y volatilidad implícita

En el ámbito financiero, el término volatilidad se refiere comúnmente a la dispersión de los retornos de un activo en torno a su media. Conviene distinguir, sin embargo, dos conceptos claramente diferenciados.

La volatilidad histórica (o realizada) es una medida estadística calculada *ex post* a partir de la serie temporal de retornos observados. Habitualmente se estima como la desviación

típica muestral de los retornos logarítmicos diarios, anualizada mediante el factor $\sqrt{252}$ (siendo 252 el número aproximado de días bursátiles en un año):

$$\hat{\sigma}_{anual} = \sqrt{252} \cdot \sqrt{\frac{1}{T-1} \sum_{t=1}^T (r_t - \bar{r})^2}$$

donde $r_t = \ln(P_t/P_{t-1})$ y $\bar{r}$ es la media muestral de los retornos.

La volatilidad implícita (*implied volatility*, IV), por el contrario, es una medida *forward-looking* extraída del precio de las opciones cotizadas. Dado un precio observado de una opción y conocidos el resto de parámetros del modelo de valoración (precio del subyacente, precio de ejercicio, tipo de interés libre de riesgo, tiempo a vencimiento y dividendos), la volatilidad implícita es el valor que, introducido en la fórmula de Black-Scholes-Merton (Black & Scholes, 1973; Merton, 1973), reproduce exactamente el precio de mercado de la opción. Constituye, por tanto, la expectativa del mercado sobre la volatilidad futura del subyacente, condicionada a la información disponible y a las preferencias por el riesgo de los participantes.

A diferencia de la volatilidad histórica, la implícita no se observa directamente: debe inferirse mediante un proceso numérico de inversión (típicamente el método de Newton-Raphson). En consecuencia, la volatilidad implícita es un objeto teórico que depende del modelo de valoración utilizado. La constatación empírica de que la volatilidad implícita varía sistemáticamente en función del precio de ejercicio y del vencimiento de la opción —fenómeno conocido como *volatility smile* o *volatility surface*— constituye, de hecho, una de las principales evidencias en contra de los supuestos del modelo de Black-Scholes (Rubinstein, 1985; Derman & Kani, 1994).

2.2.2. El índice VIX

El VIX (*CBOE Volatility Index*) es el indicador de volatilidad implícita más relevante de los mercados financieros mundiales. Lanzado por el *Chicago Board Options Exchange* en 1993 con una metodología basada en las ocho opciones *at-the-money* sobre el S&P 100, fue redefinido en 2003 mediante una metodología puramente *model-free*, basada en una replicación discretizada del precio de un swap de varianza sobre el S&P 500 (CBOE, 2003, 2014; Demeterfi et al., 1999). Desde entonces, el VIX representa la volatilidad implícita anualizada esperada del S&P 500 a 30 días vista.

La fórmula de cálculo del VIX, en su formulación post-2003, es:

$$\sigma^2 = \frac{2}{T} \sum_i \frac{\Delta K_i}{K_i^2} e^{rT} Q(K_i) - \frac{1}{T} \left[\frac{F}{K_0} - 1 \right]^2$$

donde T es el tiempo a vencimiento, K_i son los precios de ejercicio, ΔK_i es el espaciado entre ejercicios consecutivos, r es el tipo libre de riesgo, $Q(K_i)$ es el precio medio de las opciones con ejercicio K_i , F es el precio forward del subyacente y K_0 es el primer ejercicio inferior al *forward*. El VIX se obtiene como $VIX = 100 \cdot \sigma$.

Conocido popularmente como el "índice del miedo" (*fear gauge*), el VIX exhibe varias propiedades estadísticas distintivas (Whaley, 2000, 2009): (i) presenta una correlación negativa elevada con los retornos del S&P 500, especialmente acentuada en episodios de caídas; (ii) muestra reversión a la media de largo plazo en torno a niveles del 15–20 %; (iii) exhibe *clustering* de volatilidad —periodos de alta volatilidad seguidos de alta volatilidad y viceversa—; y (iv) presenta saltos discontinuos al alza en presencia de eventos sistémicos.

2.2.3. La prima de riesgo de la volatilidad

Una de las contribuciones más relevantes de la literatura reciente al estudio de la volatilidad implícita es la documentación sistemática de la denominada variance risk premium (VRP) o prima de riesgo de la varianza. Bollerslev, Tauchen y Zhou (2009), entre otros, muestran que el cuadrado del VIX excede sistemáticamente —en términos esperados— al valor realizado posteriormente de la varianza del S&P 500. Es decir:

$$E[VIX_t^2] > E[RV_{t, t+30}]$$

donde RV denota la varianza realizada. Esta brecha sistemática se interpreta económicamente como una prima que los participantes del mercado están dispuestos a pagar por protegerse frente al riesgo de incrementos no anticipados de la volatilidad. Dicha prima es, por construcción, negativa para el comprador de volatilidad y positiva para el vendedor: en términos esperados, **comprar volatilidad implícita y mantenerla hasta vencimiento es una operación con esperanza negativa**.

Este resultado tiene implicaciones directas y profundas para la valoración del VXX, como se desarrolla en el apartado siguiente.

2.3. Productos cotizados sobre volatilidad: el caso VXX

2.3.1. Futuros del VIX y la curva *forward*

El VIX es un índice teórico no negociable directamente. La forma operativa habitual de tomar exposición a la volatilidad implícita es a través de los futuros del VIX, listados en el *Cboe Futures Exchange* (CFE) desde 2004. Cada contrato de futuro del VIX establece la obligación de liquidar, en una fecha de vencimiento específica (tercer miércoles de cada mes), el valor que el VIX presente en ese momento.

Una propiedad fundamental de los futuros del VIX, derivada de la propia naturaleza del subyacente, es la estructura temporal típicamente en contango del mercado. En aproximadamente el 80 % de los días de cotización desde 2005 (Eraker & Wu, 2017), los futuros del VIX con vencimientos más lejanos cotizan a precios superiores a los de vencimientos cercanos. Económicamente, esta estructura refleja la combinación de tres factores: (i) la reversión a la media del VIX hacia niveles superiores a los del momento (cuando el VIX está bajo); (ii) la prima de riesgo de la varianza descrita en §2.2.3; y (iii) la convexidad de la varianza respecto del VIX al cuadrado. Cuando los mercados experimentan tensiones, la curva puede invertirse y mostrar *backwardation* (precios decrecientes con el vencimiento), aunque este régimen suele ser temporal.

2.3.2. *Roll yield* y deriva de los productos sobre volatilidad

Cualquier producto que mantenga una exposición continua a futuros del VIX —como el VXX— debe rebalancear su cartera periódicamente, vendiendo los contratos próximos a vencer y comprando los siguientes. Cuando la curva está en *contango*, esta operación implica vender contratos cotizados a un precio inferior y comprar contratos cotizados a un precio superior, generando un rendimiento de *roll* negativo (*negative roll yield*).

El impacto de esta deriva sobre el valor del producto es notable. Whaley (2013) y Alexander y Korovilas (2013) documentan, a partir de los datos disponibles, pérdidas medias del orden del 60–80 % anuales para una cartera teórica replicante del VXX en regímenes de *contango* normal. Esta erosión sistemática del valor explica que, pese a presentar repuntes muy intensos en episodios de estrés (*rallies* del 50–100 % en pocas semanas), el VXX exhiba una tendencia descendente estructural en horizontes largos, perdiendo de manera consistente más del 90 % de su valor cada cinco a siete años.

2.3.3. Mecánica del VXX (*iPath Series B*)

El VXX es un *Exchange-Traded Note* (ETN) emitido por Barclays Bank PLC bajo la marca *iPath*, que replica diariamente el rendimiento del S&P 500 VIX Short-Term Futures Index Total Return (SPVXSTR). Este índice mantiene una posición sintética en los dos futuros del VIX más cercanos al vencimiento, ponderada de tal forma que la duración media (en días hasta vencimiento) se mantenga constante en treinta días naturales. El proceso es continuo: cada día se transfiere un porcentaje pequeño ($1/N$, siendo N el número de días hábiles del mes) desde la posición en el primer vencimiento hacia la posición en el segundo vencimiento.

Es importante destacar dos cuestiones técnicas relevantes para el análisis empírico:

Primero, el VXX original (con la misma denominación pero CUSIP distinto) fue emitido en enero de 2009 con una fecha de vencimiento estipulada de 30 de enero de 2019. Para asegurar la continuidad del producto en el mercado, Barclays emitió en enero de 2018 una nueva serie del producto, conocida como VXX *iPath Series B* (CUSIP 06746P165), con vencimiento en 2048. Aunque la serie B replica el mismo índice subyacente y mantiene una continuidad económica respecto del producto original, las dos emisiones son contractualmente distintas, presentan estructuras fiscales y de comisiones ligeramente diferentes y, sobre todo, constituyen series temporales no continuas para fines estadísticos. Esta circunstancia justifica la elección del periodo 2018-2026 en el presente trabajo, evitando los problemas metodológicos derivados de empalmar las dos series mediante factores de ajuste.

Segundo, el VXX ha experimentado a lo largo de su historia varios reverse splits (consolidaciones de acciones), motivados por la deriva descendente del producto. La *iPath Series B* ha realizado al menos un reverse split 4-a-1 en agosto de 2022, lo que reduce su precio nominal pero no su valor económico. Estos eventos están plenamente recogidos en los datos ajustados utilizados en este trabajo, procedentes de FactSet.

2.3.4. Otros productos sobre volatilidad: contexto

El VXX no es el único producto cotizado sobre volatilidad, aunque sí el más líquido en su categoría. El universo de ETPs sobre volatilidad incluye, entre otros, el **VIXY** (ProShares VIX Short-Term Futures ETF, ETF de estructura distinta al ETN pero metodológicamente análogo), el **UVXY** (ETF apalancado 1.5x sobre el mismo índice

subyacente), el SVXY (ETF inverso 0.5x sobre la volatilidad de corto plazo) y el extinto XIV, cuya implosión durante el episodio del *Volmageddon* (5 de febrero de 2018, con una pérdida intradía superior al 90 %) supuso un punto de inflexión en la industria. Para una revisión sistemática del comportamiento de estos productos, véase Whaley (2013) y Bordonado, Molnár y Samdal (2017).

2.4. Arbitraje estadístico y *pair trading*

2.4.1. Concepto y fundamento del arbitraje estadístico

El **arbitraje estadístico** se diferencia del arbitraje clásico por su naturaleza probabilística. Mientras que el arbitraje clásico se basa en la detección de oportunidades sin riesgo — típicamente relaciones de paridad incumplidas, como el clásico arbitraje *cash-and-carry* o las desviaciones del *put-call parity*—, el arbitraje estadístico identifica relaciones de equilibrio entre activos que se cumplen *en términos esperados* y explota las desviaciones temporales respecto de dichas relaciones, asumiendo un riesgo residual.

Avellaneda y Lee (2010) caracterizan formalmente el arbitraje estadístico como una estrategia de inversión con esperanza positiva y varianza acotada, construida a partir de una cartera de activos cuyos retornos residuales —tras descontar la exposición común a factores sistémicos— exhiben propiedades estacionarias y ergódicas. La condición operativa para la viabilidad de la estrategia es que la velocidad de reversión a la media del residuo sea suficientemente elevada en relación con su volatilidad, de modo que la duración media de las posiciones permanezca acotada.

2.4.2. *Pair trading* clásico: el enfoque de distancia

El *pair trading* es la versión más simple y directa del arbitraje estadístico. Consiste en seleccionar dos activos que, sobre la base del análisis fundamental, técnico o estadístico, mantengan una relación de equilibrio estable, y construir una posición *market-neutral* (larga en uno de los activos y corta en el otro) cuando el spread entre ambos se desvíe significativamente de su valor de equilibrio. Cuando dicho spread retorne a su valor de equilibrio, las posiciones se cierran realizando el beneficio.

Gatev, Goetzmann y Rouwenhorst (2006) realizan el primer estudio empírico amplio sobre la rentabilidad del *pair trading* mediante el denominado enfoque de distancia: para cada par de activos candidatos, se calcula la distancia euclídea entre los precios normalizados durante una ventana de formación, y se seleccionan los pares con menor

distancia. La estrategia abre posiciones cuando el spread normalizado supera dos desviaciones típicas y las cierra al retornar al equilibrio. El estudio documenta retornos anuales en exceso del 11 % entre 1962 y 2002 en el mercado estadounidense, si bien con desempeño decreciente a lo largo del tiempo.

2.4.3. *Pair trading* basado en cointegración

El segundo gran enfoque metodológico, asociado a Vidyamurthy (2004), Caldeira y Moura (2013) y Krauss (2017) entre otros, formaliza el *pair trading* como una estrategia construida sobre una relación de cointegración entre los activos. Bajo este enfoque, los activos se seleccionan no por la similitud histórica de sus precios, sino por el cumplimiento estadístico de la condición de cointegración (definida formalmente en §2.5.2): si dos series no estacionarias presentan una combinación lineal estacionaria, dicha combinación constituye el spread de equilibrio del par, y las desviaciones respecto del mismo serán transitorias por construcción.

El enfoque cointegracional aporta varias ventajas teóricas: (i) proporciona una base estadística rigurosa para la selección de pares y la definición del spread; (ii) permite estimar la velocidad de convergencia y, con ello, calibrar parámetros operativos como umbrales de entrada y salida; y (iii) ofrece un marco natural para incorporar dinámicas de error-corrección (modelos VECM) que enriquecen el modelo de comportamiento del par.

2.4.4. *Pair trading* basado en modelos estocásticos: Ornstein-Uhlenbeck

Una tercera vía, complementaria a las anteriores, modela explícitamente la dinámica del spread como un proceso estocástico de reversión a la media. La elección natural es el proceso de **Ornstein-Uhlenbeck** (Vasicek, 1977; Bertram, 2010), descrito por la ecuación diferencial estocástica:

$$dX_t = \theta(\mu - X_t)dt + \sigma dW_t$$

donde X_t es el spread, μ es su nivel de equilibrio, $\theta > 0$ es la velocidad de reversión a la media, σ es la volatilidad y W_t es un proceso de Wiener estándar. El proceso de Ornstein-Uhlenbeck es la versión continua del modelo AR(1) y proporciona un marco analítico tratable para derivar reglas de operación óptimas, calcular probabilidades de tocar barreras y estimar la duración esperada de las posiciones.

Este enfoque presenta dos ventajas operativas relevantes para el caso SPY-VXX. **Primero**, no requiere la cointegración estricta entre los precios: basta con que el spread —construido a partir de una *hedge ratio* estimada— exhiba propiedades de reversión a la media. **Segundo**, los parámetros del proceso pueden estimarse mediante máxima verosimilitud sobre datos discretos a partir de fórmulas cerradas, proporcionando una calibración robusta y reproducible.

2.4.5. Métricas de evaluación

La evaluación del desempeño de una estrategia de *pair trading*, como la de cualquier estrategia cuantitativa, exige el uso de métricas ajustadas al riesgo, dado que la rentabilidad bruta es generalmente engañosa. Las métricas estándar (Sharpe, 1966; Sortino & Price, 1994; Pardo, 2008) incluyen:

- **Ratio de Sharpe:** $SR = (R - R_f)/\sigma_R$, donde R es el rendimiento anualizado, R_f el tipo libre de riesgo y σ_R la volatilidad anualizada. Mide la prima por unidad de riesgo total.
- **Ratio de Sortino:** análogo al Sharpe pero utilizando únicamente la desviación típica de los retornos negativos (*downside deviation*), reconociendo que los inversores se preocupan asimétricamente por los retornos.
- **Ratio de Calmar:** cociente entre el rendimiento anualizado y el *maximum drawdown*, medida directamente relacionada con la pérdida máxima histórica de la estrategia.
- ***Maximum drawdown* (MDD):** mayor caída relativa desde un máximo histórico hasta un mínimo posterior antes de un nuevo máximo.
- ***Hit ratio*:** porcentaje de operaciones cerradas con beneficio.
- **Duración media de las operaciones y rotación de cartera:** indicadores operativos que afectan a los costes de transacción.

Estas métricas serán empleadas de forma sistemática en el Capítulo 7 para evaluar el desempeño de la estrategia desarrollada.

2.5. Modelos econométricos relevantes

Esta sección presenta los principales modelos econométricos que serán empleados en la parte empírica del trabajo. La presentación es deliberadamente concisa: se enfatizan los conceptos esenciales, las hipótesis subyacentes y la interpretación económica, dejando la

implementación operativa para el Capítulo 3 (Metodología) y los Capítulos 5 y 6 (resultados empíricos).

2.5.1. Estacionariedad y raíz unitaria

Una serie temporal $\{X_t\}$ es estacionaria en sentido débil cuando su media, varianza y autocovarianzas no dependen del tiempo. Formalmente, $E[X_t] = \mu$, $\text{Var}(X_t) = \sigma^2$ y $\text{Cov}(X_t, X_{t-s}) = \gamma_s$ para todo t y s . La estacionariedad es una condición necesaria para la aplicación válida de la mayoría de las técnicas de inferencia estadística sobre series temporales.

Una serie no estacionaria que se vuelve estacionaria al diferenciarla d veces se dice integrada de orden d , denotada $I(d)$. La mayoría de las series financieras de precios son $I(1)$: sus niveles son no estacionarios pero sus diferencias (o, equivalentemente, sus retornos logarítmicos) sí lo son.

Los tests de raíz unitaria permiten contrastar empíricamente la estacionariedad de una serie. Los más empleados son el test ampliado de Dickey-Fuller (ADF; Dickey & Fuller, 1979; Said & Dickey, 1984), cuya hipótesis nula es la presencia de raíz unitaria (no estacionariedad), y el **test KPSS** (Kwiatkowski, Phillips, Schmidt & Shin, 1992), cuya hipótesis nula es la estacionariedad. La aplicación conjunta de ambos tests, con hipótesis nulas opuestas, ofrece evidencia más robusta sobre la naturaleza estocástica de la serie.

2.5.2. Cointegración: Engle-Granger y Johansen

Dos series X_t e Y_t , ambas $I(1)$, están **cointegradas** si existe una combinación lineal $Z_t = Y_t - \beta X_t$ que es estacionaria, $I(0)$. Económicamente, la cointegración indica la existencia de una relación de equilibrio de largo plazo entre las dos series: aunque cada una de ellas se desvíe estocásticamente sin tendencia clara, las desviaciones respecto del equilibrio son transitorias y tienden a corregirse con el tiempo.

El **test de Engle-Granger** (Engle & Granger, 1987) procede en dos etapas: (i) se estima la regresión cointegrante por OLS, $Y_t = \alpha + \beta X_t + u_t$; y (ii) se aplica el test ADF sobre los residuos $\hat{u}_t$. Si los residuos rechazan la hipótesis de raíz unitaria con valores críticos específicos (más exigentes que los estándar, debido a que se aplican sobre estimadores), las series son cointegradas. Una limitación importante del test de Engle-Granger es que

su resultado puede depender de qué variable se considera dependiente y cuál independiente.

El test de Johansen (Johansen, 1988, 1991) aborda esta limitación mediante un planteamiento multivariante basado en el modelo VECM (Vector Error-Correction Model). Estima el rango r de la matriz de cointegración mediante dos estadísticos: el estadístico de la traza y el estadístico del valor propio máximo. Permite la detección simultánea de múltiples vectores de cointegración en sistemas de más de dos variables.

2.5.3. Modelos ARIMA y VAR

Los modelos ARIMA(p,d,q) (Box & Jenkins, 1976) constituyen el marco clásico para la descripción univariante de series temporales. Combinan un componente autorregresivo de orden p , una integración de orden d y un componente de medias móviles de orden q :

$$\phi(L)(1 - L)^d X_t = \theta(L)\varepsilon_t$$

donde $\phi(L)$ y $\theta(L)$ son polinomios en el operador de retardo L , y ε_t es ruido blanco. La selección del orden óptimo (p, d, q) se realiza habitualmente mediante criterios de información (AIC, BIC) y diagnósticos de los residuos.

Los modelos VAR(p) (Sims, 1980) extienden el enfoque ARMA al caso multivariante. En su forma reducida:

$$\mathbf{X}_t = \mathbf{c} + \mathbf{A}_1 \mathbf{X}_{t-1} + \dots + \mathbf{A}_p \mathbf{X}_{t-p} + \boldsymbol{\varepsilon}_t$$

donde $\mathbf{X}_t$ es un vector de variables endógenas. Permiten analizar relaciones dinámicas entre variables, incluyendo la causalidad en el sentido de Granger (Granger, 1969) y las funciones de impulso-respuesta.

2.5.4. Modelos GARCH y EGARCH

La constatación empírica de que las series financieras presentan **heterocedasticidad condicional** —*clustering* de volatilidad: periodos de alta volatilidad seguidos por periodos de alta volatilidad— motivó el desarrollo de los modelos ARCH (Engle, 1982) y, posteriormente, de los modelos GARCH (Bollerslev, 1986).

El modelo **GARCH(p,q)** especifica la varianza condicional como:

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

La especificación más popular en la práctica es el GARCH(1,1), que captura adecuadamente las propiedades estilizadas de la volatilidad financiera (*clustering*, persistencia, reversión a la media de largo plazo) con un número reducido de parámetros.

El modelo **EGARCH** (Nelson, 1991) introduce dos modificaciones relevantes: (i) modeliza el logaritmo de la varianza condicional, garantizando la positividad sin necesidad de restricciones sobre los parámetros; y (ii) permite que las innovaciones positivas y negativas tengan efectos asimétricos sobre la volatilidad futura, capturando el fenómeno del *leverage effect*:

$$\ln(\sigma_t^2) = \omega + \sum_{i=1}^q \left[\alpha_i \frac{|\varepsilon_{t-i}|}{\sigma_{t-i}} + \gamma_i \frac{\varepsilon_{t-i}}{\sigma_{t-i}} \right] + \sum_{j=1}^p \beta_j \ln(\sigma_{t-j}^2)$$

El parámetro γ captura la asimetría: un valor negativo y significativo de γ indica que las innovaciones negativas incrementan más la volatilidad que las positivas de igual magnitud.

2.5.5. Correlación dinámica condicional: DCC-GARCH

La extensión multivariante de los modelos GARCH permite estudiar la dependencia condicional entre dos o más activos. El modelo DCC-GARCH (*Dynamic Conditional Correlation*; Engle, 2002) descompone la modelización en dos etapas: en la primera, se ajusta un modelo GARCH univariante a cada una de las series; en la segunda, se modeliza la matriz de correlación condicional como un proceso GARCH sobre los residuos estandarizados.

La especificación DCC permite capturar la evolución temporal de la correlación entre los activos —en lugar de asumirla constante—, lo que resulta particularmente relevante para el par SPY-VXX, cuya correlación varía sustancialmente entre regímenes de mercado, como muestra el análisis exploratorio del Capítulo 4.

2.5.6. El proceso de Ornstein-Uhlenbeck

El **proceso de Ornstein-Uhlenbeck** (OU) es el modelo estocástico de reversión a la media en tiempo continuo más sencillo y extendido. Originalmente introducido en física

por Uhlenbeck y Ornstein (1930), fue adaptado al ámbito financiero por Vasicek (1977) en el contexto de la modelización de tipos de interés. Su ecuación diferencial estocástica es:

$$dX_t = \theta(\mu - X_t)dt + \sigma dW_t$$

donde $\theta > 0$ es la velocidad de reversión, μ el nivel de equilibrio, σ la volatilidad y W_t un movimiento browniano estándar. La solución analítica del proceso, condicionada a un valor inicial X_0 , es:

$$X_t = X_0 e^{-\theta t} + \mu(1 - e^{-\theta t}) + \sigma \int_0^t e^{-\theta(t-s)} dW_s$$

con distribución condicional gaussiana de media $\mu + (X_0 - \mu)e^{-\theta t}$ y varianza $\frac{\sigma^2}{2\theta}(1 - e^{-2\theta t})$. La **vida media** del proceso —tiempo esperado para que una desviación inicial se reduzca a la mitad— viene dada por $\tau_{1/2} = \ln(2)/\theta$.

Para datos discretos muestreados a intervalos Δt , la versión exacta del OU es un AR(1) gaussiano, y los parámetros pueden estimarse mediante máxima verosimilitud a partir de la regresión:

$$X_{t+1} = a + bX_t + \eta_t$$

con las relaciones $\theta = -\ln(b)/\Delta t$, $\mu = a/(1 - b)$ y $\sigma = \text{std}(\eta)\sqrt{\frac{2\theta}{1-b^2}}$. Esta calibración será aplicada al *spread* SPY-VXX en el Capítulo 6.

2.6. Estado del arte: el par SPY-VXX en la literatura

La aplicación específica de modelos de arbitraje estadístico al par SPY-VXX ha recibido una atención académica relativamente limitada, pese a su evidente relevancia profesional. Los principales antecedentes pueden agruparse en tres líneas.

Estudios sobre la dinámica volatilidad-renta variable. La literatura sobre el *leverage effect* y la *asymmetric volatility* —Black (1976), Christie (1982), Bekaert y Wu (2000), entre otros— constituye el sustrato teórico sobre el que se asienta cualquier estrategia que explote la correlación negativa entre el SPY y el VXX. Más recientemente, Bandi y Renò (2012) y Aït-Sahalia, Fan y Li (2013) han profundizado en la modelización de saltos comunes a precios y volatilidad, mostrando que una fracción significativa de los

movimientos extremos en los retornos del S&P 500 se explica por saltos contemporáneos en la volatilidad.

Estudios sobre el desempeño de los productos cotizados sobre volatilidad. La literatura más directamente relevante para la mecánica del VXX incluye los trabajos de Whaley (2013) —"*Trading volatility: At what cost?*"—, Alexander y Korovilas (2013), Bordonado, Molnár y Samdal (2017) y Eraker y Wu (2017), que documentan sistemáticamente la deriva descendente de los productos *long volatility* y caracterizan la prima de riesgo de los futuros del VIX. Estos trabajos constituyen el fundamento empírico para la formulación de estrategias *short-volatility*.

Estudios sobre estrategias de cobertura y *trading* con volatilidad. Daigler y Rossi (2006), Szado (2009) y Hill, Balasubramanian, Tiernan y Schmidt (2010), entre otros, exploran la incorporación de productos sobre volatilidad como cobertura de carteras de renta variable, documentando ganancias significativas en el ratio de Sharpe ajustado al riesgo en presencia de eventos extremos. Más específicamente, Konstantinidi, Skiadopoulos y Tzagkaraki (2008) y posteriores han examinado la predictibilidad de la dinámica del VIX y los futuros sobre el VIX, identificando regularidades estadísticas explotables.

Aunque la literatura ha abordado por separado la dinámica del SPY, la del VIX, la del VXX y los principios del *pair trading*, la integración sistemática de estos componentes en una estrategia cuantitativa explícita sobre el par SPY-VXX, con un enfoque metodológico riguroso (cointegración, GARCH dinámico, Ornstein-Uhlenbeck) y un *backtesting* exhaustivo en regímenes diferenciados, no constituye un objeto de estudio frecuente en la literatura académica revisada. El presente Trabajo Fin de Grado se propone, precisamente, contribuir a llenar este espacio con un análisis empírico actualizado al periodo 2018-2026.

2.7. Síntesis y posicionamiento del trabajo

El marco teórico expuesto a lo largo de este capítulo permite delimitar con precisión el espacio conceptual en el que se inscribe el presente Trabajo Fin de Grado. La hipótesis subyacente es que, dentro de los límites del arbitraje, existen ineficiencias temporales en la dinámica conjunta del SPY y el VXX —cuya correlación negativa estructural se ve modulada por el *leverage effect*, el *contango* y la prima de riesgo de la varianza— que

pueden ser detectadas mediante herramientas econométricas y explotadas mediante estrategias sistemáticas de *pair trading*.

El enfoque metodológico combina los tres pilares descritos: (i) la construcción de un *spread* dinámico mediante una *hedge ratio* estimada por ventana móvil; (ii) la modelización del *spread* como un proceso de Ornstein-Uhlenbeck y la verificación de sus propiedades estadísticas mediante tests de cointegración y estacionariedad; y (iii) la generación de señales operativas a partir del Z-score del *spread* respecto de su media de equilibrio, complementada con un análisis de la dinámica de volatilidad mediante modelos GARCH/EGARCH y DCC-GARCH.

A diferencia del enfoque clásico de *pair trading* basado en cointegración estricta —que, como anticipa la H2 y como se confirmará empíricamente en el Capítulo 5, no se cumple para el par SPY-VXX en el periodo analizado—, el presente trabajo adopta un enfoque más flexible basado en la reversión a la media del *spread* dinámico. Esta elección metodológica refleja la realidad económica del par: la deriva descendente del VXX impide la existencia de una relación de cointegración en niveles, pero no anula la posibilidad de explotar la dinámica de corto plazo del *spread* ajustado por una *hedge ratio* dinámica.

El siguiente capítulo desarrolla en detalle la metodología empleada, especificando tanto las técnicas econométricas como los criterios de diseño y validación de la estrategia.

Capítulo 3. Metodología

3.1. Visión general del proceso de investigación

La investigación empírica desarrollada en este Trabajo Fin de Grado se articula en cinco fases secuenciales, encadenadas de modo que cada una alimenta a la siguiente. El esquema es el siguiente: (1) recopilación y depuración de datos; (2) análisis exploratorio y caracterización estadística de las series; (3) modelización econométrica de la dinámica univariante y conjunta del SPY y el VXX; (4) diseño y calibración de la estrategia de *pair trading*; y (5) *backtesting* y validación del desempeño en regímenes diferenciados de mercado. Las técnicas estadísticas aplicadas en cada fase son las descritas en el Capítulo 2; el presente capítulo se centra en su implementación operativa, en las decisiones

metodológicas adoptadas y en los criterios empleados para garantizar la robustez y la reproducibilidad de los resultados.

La metodología se inspira en el flujo de trabajo estándar de los equipos de investigación cuantitativa en bancos de inversión y *hedge funds*, adaptado a las restricciones propias de un trabajo académico. En particular, se ha priorizado la transparencia operativa: todos los parámetros del modelo se calibran exclusivamente sobre la muestra de entrenamiento, los resultados *out-of-sample* se reportan sin recalibración, y el código fuente está disponible íntegramente en los anexos para permitir la replicación independiente.

3.2. Fuentes y depuración de los datos

3.2.1. Fuente

Los precios diarios del SPY y del VXX se han obtenido de FactSet, plataforma profesional de datos financieros utilizada habitualmente en la industria. Para cada activo se han descargado las siguientes variables: precio de cierre, precio de apertura, máximo y mínimo diarios, volumen negociado y, como variable principal, el Total Return Gross —serie ajustada por dividendos y acciones corporativas (*splits*), que representa fielmente el rendimiento total acumulable que obtendría un inversor con reinversión de dividendos—. La serie del Total Return es la utilizada para el cálculo de los retornos en todo el análisis posterior.

Adicionalmente, se han incorporado los datos diarios del índice VIX procedentes del *Cboe Volatility Index Historical Data*, en su versión publicada por el propio CBOE. El VIX se utiliza únicamente como variable de contexto para identificar regímenes de volatilidad y no entra en la modelización principal del par.

3.2.2. Periodo de análisis

El periodo de estudio se extiende desde el 18 de enero de 2018, fecha de lanzamiento de la *iPath Series B S&P 500 VIX Short-Term Futures ETN* (CUSIP 06746P165), hasta el 28 de abril de 2026. Esta delimitación temporal responde a una motivación metodológica explícita: como se argumentó en el Capítulo 1, el VXX original maduró el 30 de enero de 2019, y la actual iPath Series B —pese a replicar el mismo índice subyacente— constituye una emisión contractualmente distinta. Empalmar las dos series introduciría un sesgo metodológico evitable. La fecha de cierre se establece en la última sesión bursátil completa disponible en el momento de elaboración del trabajo.

3.2.3. Depuración

El proceso de depuración aplicado a las series ha incluido los siguientes pasos:

- **Sincronización de fechas.** Se han mantenido únicamente las sesiones en las que ambos activos cuentan con observación válida (intersección de las series). El resultado son 2.080 observaciones diarias, correspondientes a 2.079 retornos logarítmicos.
- **Verificación de consistencia.** Se han comprobado los rangos de precios, la coherencia de las variables OHLC (Open, High, Low, Close) y la ausencia de huecos no justificados por días no bursátiles.
- **Tratamiento de eventos corporativos.** La serie *Total Return Gross* incorpora ya los ajustes por dividendos y *splits*. En particular, el VXX experimentó un *reverse split* 4-a-1 en agosto de 2022 que se encuentra plenamente reflejado en la serie ajustada.
- **Cálculo de los retornos logarítmicos.** Sobre la serie *Total Return Gross* se han calculado los retornos logarítmicos diarios mediante la transformación $r_t = \ln(P_t/P_{t-1})$, que presentan mejores propiedades estadísticas (aditividad temporal, simetría asintótica) que los retornos aritméticos.

3.3. Análisis exploratorio y caracterización estadística

El análisis exploratorio constituye una fase imprescindible previa a cualquier modelización econométrica formal. Su objetivo es caracterizar las propiedades estilizadas de las series y detectar los rasgos que guiarán las elecciones metodológicas posteriores. Las técnicas aplicadas son las siguientes.

3.3.1. Estadísticos descriptivos y momentos

Para cada serie de retornos logarítmicos diarios se calculan: media, desviación típica, mediana, mínimo, máximo, asimetría (*skewness*) y exceso de curtosis (*kurtosis*). Adicionalmente, se reportan los retornos anualizados (mediante el factor 252) y la volatilidad anualizada (factor $\sqrt{252}$), así como el ratio de Sharpe ex post asumiendo tipo libre de riesgo cero.

3.3.2. Tests de normalidad

La hipótesis de normalidad se contrasta mediante el **test de Jarque-Bera** (Jarque & Bera, 1980), basado en la asimetría y la curtosis muestrales:

$$JB = \frac{T}{6} \left[S^2 + \frac{(K - 3)^2}{4} \right]$$

donde T es el tamaño muestral, S la asimetría y K la curtosis. Bajo la hipótesis nula de normalidad, JB se distribuye asintóticamente como una χ^2 con dos grados de libertad. El rechazo de la normalidad implica que las inferencias basadas en supuestos gaussianos — en particular, el ratio de Sharpe estándar y los intervalos de confianza simétricos— deben interpretarse con precaución, y motiva el uso de distribuciones t de Student en los modelos GARCH posteriores.

3.3.3. Tests de estacionariedad: ADF y KPSS

La estacionariedad de las series se contrasta mediante dos tests complementarios:

- **Test ADF de Dickey-Fuller ampliado** (Dickey & Fuller, 1979; Said & Dickey, 1984), con hipótesis nula H_0 : presencia de raíz unitaria (no estacionariedad). El rechazo de H_0 —p-valor inferior al 5 %— evidencia estacionariedad.
- **Test KPSS** (Kwiatkowski et al., 1992), con hipótesis nula opuesta H_0 : estacionariedad. El no rechazo de H_0 —p-valor superior al 5 %— es compatible con la estacionariedad de la serie.

La aplicación conjunta de ambos tests aporta robustez: cuando ambos tests apuntan en la misma dirección, la conclusión sobre la naturaleza estocástica de la serie es sólida. Se aplican sobre los niveles, los logaritmos y los retornos de cada activo.

3.3.4. Análisis de correlación

La correlación entre los activos se examina en tres dimensiones:

- **Correlación contemporánea** (Pearson) entre niveles, log-niveles y retornos.
- **Correlación móvil** (*rolling correlation*) sobre ventanas de 60 y 120 días para describir la evolución temporal de la dependencia.
- **Función de correlación cruzada** (*cross-correlation function*, CCF) hasta diez retardos, para detectar posibles relaciones de adelanto-retardo entre los retornos.

3.4. Modelización econométrica

3.4.1. Tests de cointegración

Pese a que la H2 anticipa un resultado negativo —es decir, que las series del SPY y el VXX no están cointegradas en el periodo 2018-2026 debido a la deriva descendente estructural del VXX—, los tests de cointegración se aplican sistemáticamente para documentar empíricamente este resultado, en línea con la propuesta original del trabajo. Se emplean dos enfoques complementarios:

Test de Engle-Granger (Engle & Granger, 1987), aplicado en dos etapas: (i) estimación por OLS de la regresión cointegrante $Y_t = \alpha + \beta X_t + u_t$, tanto sobre los niveles como sobre los logaritmos; (ii) test ADF sobre los residuos $\hat{u}_t$ con valores críticos específicos de Engle-Granger. El test se aplica además en ambas direcciones (SPY $\rightarrow$ VXX y VXX $\rightarrow$ SPY) para verificar la robustez del resultado.

Test de Johansen (Johansen, 1988, 1991), aplicado sobre el sistema bivalente (log SPY, log VXX). Se reportan tanto el estadístico de la traza (*trace statistic*) como el estadístico del valor propio máximo (*max-eigenvalue statistic*) para los valores hipotetizados $r = 0$ y $r \leq 1$, contrastándolos con los valores críticos al 95 % de significación.

3.4.2. Modelos ARIMA univariantes

La dinámica univariante de los retornos se modeliza mediante modelos ARIMA(p,0,q) sobre las series ya estacionarias de retornos logarítmicos. La selección del orden óptimo (p, q) se realiza mediante una búsqueda exhaustiva en el espacio $\{0, \dots, 3\} \times \{0, \dots, 3\}$, eligiendo el modelo que minimiza el criterio de información bayesiano (BIC) de Schwarz (1978). Se prefiere el BIC sobre el AIC para evitar la sobreparametrización en muestras grandes. La adecuación del modelo seleccionado se verifica posteriormente mediante el test de Ljung-Box sobre los residuos.

3.4.3. Modelo VAR y causalidad de Granger

La dinámica conjunta de los retornos se modeliza mediante un modelo VAR(p) (Sims, 1980) sobre el vector $(r_{SPY,t}, r_{VXX,t})$. La selección del orden p se realiza mediante los criterios AIC, BIC, FPE (*final prediction error*) y HQIC (Hannan-Quinn). Sobre el VAR estimado se aplican los tests de causalidad de Granger (Granger, 1969) en ambas direcciones, mediante el estadístico F sobre la restricción de exclusión de los coeficientes

correspondientes a los retardos del activo causante. El rechazo de la hipótesis nula de no-causalidad se interpreta como evidencia de que la información contenida en el pasado de un activo aporta capacidad predictiva sobre el otro, más allá de la información contenida en su propio pasado.

3.4.4. Modelos GARCH y EGARCH

La dinámica de la volatilidad condicional de cada activo se modeliza mediante:

GARCH(1,1) (Bollerslev, 1986):

$$\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2$$

con $\omega > 0$, $\alpha, \beta \geq 0$, $\alpha + \beta < 1$ (condición de estacionariedad).

EGARCH(1,1) (Nelson, 1991):

$$\ln(\sigma_t^2) = \omega + \alpha \left(\frac{|\varepsilon_{t-1}|}{\sigma_{t-1}} \right) + \gamma \frac{\varepsilon_{t-1}}{\sigma_{t-1}} + \beta \ln(\sigma_{t-1}^2)$$

donde el coeficiente γ captura la asimetría: un valor negativo y significativo de γ es coherente con el *leverage effect* —innovaciones negativas incrementan más la volatilidad que innovaciones positivas de igual magnitud—.

Ambos modelos se estiman por máxima verosimilitud asumiendo distribución *t* de Student estandarizada para las innovaciones, dado que la evidencia previa (test de Jarque-Bera) descarta la normalidad. La elección entre GARCH y EGARCH se realiza mediante el criterio de información de Akaike (AIC): se prefiere el modelo con menor AIC. El orden (1, 1) se justifica por ser la especificación dominante en la literatura aplicada y por presentar el mejor compromiso parsimonia-ajuste en los modelos GARCH (Hansen & Lunde, 2005).

3.4.5. Modelo DCC-GARCH

La evolución temporal de la correlación entre los activos se modeliza mediante el **DCC-GARCH** de Engle (2002), siguiendo el enfoque en dos etapas:

Etapas 1: ajuste de modelos EGARCH univariantes a cada serie, obteniendo los residuos estandarizados $z_{i,t} = \varepsilon_{i,t} / \sigma_{i,t}$, $i \in \{SPY, VXX\}$.

Etapas 2: estimación de la dinámica DCC sobre los residuos estandarizados:

$$Q_t = (1 - a - b)\bar{Q} + a z_{t-1} z'_{t-1} + b Q_{t-1}$$

donde $\bar{Q}$ es la matriz de covarianza incondicional de los residuos estandarizados, y $a, b > 0$ con $a + b < 1$. La matriz de correlación condicional se obtiene como $R_t = D_t^{-1} Q_t D_t^{-1}$, con $D_t = \text{diag}(\sqrt{Q_{ii,t}})$.

La estimación de los parámetros (a, b) se realiza por máxima verosimilitud condicional sobre la log-verosimilitud gaussiana de los residuos estandarizados, partiendo de valores iniciales $(a_0, b_0) = (0, 05; 0, 90)$ y optimizando mediante el algoritmo L-BFGS-B con restricciones de positividad y persistencia.

3.5. Diseño de la estrategia de *pair trading*

3.5.1. Construcción del *spread* dinámico

Dado que la H2 anticipa la ausencia de cointegración estricta en niveles, la estrategia se construye sobre una versión más flexible del *spread*: una combinación lineal con coeficiente de cobertura (*hedge ratio*) dinámico estimado por OLS sobre los retornos en una ventana móvil. Específicamente, en cada instante t se estima:

$$r_{VXX,s} = \alpha_t + \beta_t r_{SPY,s} + \varepsilon_s, \quad s \in [t - W + 1, t]$$

donde $W = 60$ días bursátiles. El valor $\hat{\beta}_t = \text{Cov}(r_{SPY}, r_{VXX}) / \text{Var}(r_{SPY})$ obtenido en cada ventana se utiliza para construir el *spread* sobre los logaritmos de los precios:

$$S_t = \ln(P_{VXX,t}) - \hat{\beta}_t \ln(P_{SPY,t})$$

La elección de $W = 60$ días responde a un compromiso entre dos objetivos contrapuestos: ventanas más cortas captan rápidamente cambios de régimen pero introducen ruido; ventanas más largas aportan estabilidad estimacional pero pueden suavizar transiciones reales. El análisis de sensibilidad al parámetro W se reporta en el Capítulo 7.

3.5.2. Verificación de la estacionariedad del *spread*

Una vez construido el *spread* dinámico, se verifica su carácter estacionario mediante el test ADF. La estacionariedad del *spread* —pese a la no cointegración de los precios

subyacentes— constituye la condición técnica para la viabilidad estadística de la estrategia de reversión a la media.

3.5.3. Calibración del proceso de Ornstein-Uhlenbeck

Sobre el *spread* estacionario se ajusta un proceso de Ornstein-Uhlenbeck mediante el procedimiento clásico AR(1) sobre datos discretos:

$$S_{t+1} = a + b S_t + \eta_t, \quad \eta_t \sim N(0, \sigma_\eta^2)$$

Los parámetros del proceso continuo se obtienen a partir de las estimaciones AR(1) según las relaciones:

$$\hat{\theta} = -\frac{\ln \hat{b}}{\Delta t}, \quad \hat{\mu} = \frac{\hat{a}}{1 - \hat{b}}, \quad \hat{\sigma} = \hat{\sigma}_\eta \sqrt{\frac{2\hat{\theta}}{1 - \hat{b}^2}}$$

donde $\Delta t = 1/252$ corresponde a la frecuencia diaria. La **vida media** del proceso, definida como el tiempo esperado para que una desviación inicial se reduzca a la mitad, viene dada por $\tau_{1/2} = \ln(2)/\hat{\theta}$, y se reporta en días bursátiles para facilitar la interpretación operativa.

3.5.4. Generación de señales

Las señales de operación se generan a partir del Z-score del *spread* respecto de su media móvil de equilibrio. Concretamente:

$$Z_t = \frac{S_t - \bar{S}_t^{(W)}}{\sigma_{S,t}^{(W)}}$$

donde $\bar{S}_t^{(W)}$ y $\sigma_{S,t}^{(W)}$ son la media y desviación típica móviles del *spread* sobre una ventana de $W = 60$ días.

Las reglas operativas son:

- **Entrada en posición *short spread*** (vender VXX, comprar SPY ponderado por $|\beta|$) cuando $Z_t > z_{in}$, con $z_{in} = +2.0$.
- **Entrada en posición *long spread*** (comprar VXX, vender SPY ponderado por $|\beta|$) cuando $Z_t < -z_{in}$, con $-z_{in} = -2.0$.
- **Cierre de posiciones** cuando $|Z_t| < z_{out}$, con $z_{out} = 0.5$ (*mean reversion partial*).

- ***Stop-loss*** cuando $|Z_t| > Z_{stop}$, con $Z_{stop} = 4.0$, para limitar pérdidas en escenarios de divergencia extrema.

La elección de los umbrales $(Z_{in}, Z_{out}, Z_{stop}) = (2.0, 0.5, 4.0)$ se justifica por la práctica habitual en la literatura de *pair trading* (Gatev et al., 2006; Krauss, 2017) y se somete a análisis de sensibilidad en el Capítulo 7. Las posiciones se construyen *market-neutral*: cada posición en VXX se compensa con una posición en SPY de signo opuesto, ponderada por $|\hat{\beta}_t|$, de modo que la exposición al factor de mercado sea aproximadamente nula.

3.5.5. Dimensionado y gestión del riesgo

El dimensionado de las posiciones se realiza mediante una asignación de capital fijo a cada operación, equivalente al 100 % del capital disponible en posiciones *market-neutral*. Esta aproximación —simple pero estándar en el *backtesting* académico— se justifica por la naturaleza del trabajo: el objetivo no es la optimización del *sizing* —que requeriría modelos adicionales (Kelly, *risk parity*, etc.)—, sino la evaluación de la calidad de las señales generadas.

Los costes de transacción se modelan mediante un *spread* combinado de comisiones, *bid-ask spread* y *slippage* equivalente al 5 puntos básicos (0,05 %) por operación de apertura o cierre, aplicado sobre el nominal de cada pierna. Este nivel es coherente con los costes habituales para inversores institucionales que operan estos instrumentos en NYSE Arca con liquidez profunda (Frazzini, Israel & Moskowitz, 2018). La sensibilidad de los resultados a los costes se evalúa en el Capítulo 7 contemplando escenarios alternativos (1, 5, 10 y 20 puntos básicos).

3.6. Backtesting y validación

3.6.1. División de la muestra

La muestra se divide en dos subperiodos no solapados:

- **Muestra de entrenamiento (*in-sample*):** del 18 de enero de 2018 al 31 de diciembre de 2022. Este subperiodo se utiliza exclusivamente para la calibración de los parámetros del modelo: orden de los modelos ARIMA y VAR, parámetros GARCH y EGARCH, parámetros OU y, en su caso, optimización de los umbrales operativos.

- **Muestra de validación (*out-of-sample*):** del 1 de enero de 2023 al 28 de abril de 2026. Este subperiodo se utiliza únicamente para la evaluación del desempeño de la estrategia previamente calibrada, sin recalibración adicional.

Esta separación, conocida como **walk-forward simple**, evita el principal sesgo del *backtesting* académico: la sobreoptimización de parámetros sobre la muestra completa (*overfitting*).

3.6.2. Métricas de evaluación

El desempeño de la estrategia se mide con un conjunto estándar de métricas ajustadas al riesgo:

- **Rendimiento total y anualizado**, calculados sobre la curva de equity de la estrategia.
- **Volatilidad anualizada**, $\sigma_{anual} = \sigma_{diaria} \cdot \sqrt{252}$.
- **Ratio de Sharpe** (Sharpe, 1966): $SR = (\bar{R} - R_f)/\sigma$, con $R_f = 0$.
- **Ratio de Sortino** (Sortino & Price, 1994): análogo al Sharpe pero usando la *downside deviation* $\sigma_- = \sqrt{\frac{1}{T} \sum \min(R_t, 0)^2}$.
- ***Maximum drawdown*** (MDD): mayor caída relativa $-\min_t(V_t - \max_{s \leq t} V_s) / \max_{s \leq t} V_s$ desde un máximo histórico de la curva de equity hasta un mínimo posterior.
- **Ratio de Calmar**: $\text{Calmar} = R_{anual} / |MDD|$.
- ***Hit ratio***: porcentaje de operaciones cerradas con beneficio neto.
- **Duración media de las operaciones** (en días bursátiles).
- **Número total de operaciones y rotación anual de cartera.**

3.6.3. Estrategias de referencia

Para evaluar el valor añadido de la estrategia desarrollada, se comparan sus resultados con dos estrategias de referencia (*benchmarks*):

- **Buy-and-hold SPY**: posición larga continua en el SPY desde el inicio del periodo de validación.
- **Posición short VXX continua**: estrategia que toma una posición corta permanente en el VXX, capturando la deriva estructural negativa del producto. Esta es una estrategia de referencia particularmente exigente, dado que históricamente ha

generado retornos elevados en términos absolutos pero con un perfil de cola muy adverso.

3.6.4. Análisis de sensibilidad y robustez

La robustez de la estrategia se evalúa mediante los siguientes ejercicios:

- **Sensibilidad a los umbrales operativos:** variación de Z_{in} entre 1.0 y 3.0 en pasos de 0.5, manteniendo constante la relación Z_{in}/Z_{out} .
- **Sensibilidad a la ventana de estimación:** variación de W entre 30, 60, 90 y 120 días.
- **Sensibilidad a los costes de transacción:** escenarios de 1, 5, 10 y 20 puntos básicos.
- **Análisis por régimen:** cálculo separado de las métricas sobre los subperiodos en los que el VIX se encuentra en el cuartil superior (alta volatilidad) frente a los subperiodos en el cuartil inferior (baja volatilidad), para contrastar la H5.

3.7. Soporte computacional y reproducibilidad

3.7.1. Entorno técnico

El análisis se ha implementado en **Python 3.12**, utilizando las siguientes librerías especializadas:

- `pandas` y `numpy` para la manipulación de datos y el cálculo numérico.
- `statsmodels` para los tests de raíz unitaria, cointegración, modelos ARIMA y VAR.
- `arch` (paquete de Kevin Sheppard) para los modelos GARCH y EGARCH.
- `scipy` para las optimizaciones (L-BFGS-B en la calibración del DCC-GARCH y del proceso OU) y los tests estadísticos.
- `matplotlib` para la generación de gráficos.

El código completo está organizado en un **Jupyter Notebook** estructurado por capítulos, disponible en el Anexo A del trabajo. Cada celda incluye comentarios que explican el propósito y la lógica de la operación realizada.

3.7.2. Replicación en Excel

Como complemento al análisis en Python, se ha replicado el *backtesting* completo en un modelo de Microsoft Excel que reproduce los principales cálculos de la estrategia. Este modelo —disponible en el Anexo B— cumple dos funciones: (i) verificación cruzada de los resultados, mediante la comparación numérica con los obtenidos en Python; y (ii) presentación visual de los resultados a un público no técnico, lo que es habitualmente requerido en la práctica profesional.

3.7.3. Criterios de reproducibilidad

Para garantizar la reproducibilidad de los resultados se han adoptado las siguientes medidas:

- Las semillas aleatorias (`random_state`) se fijan explícitamente en todos los procedimientos que las utilicen.
- Las dependencias de Python están especificadas en un archivo `requirements.txt` con versiones congeladas.
- Los datos brutos se almacenan en formato CSV sin transformaciones, y todas las transformaciones se documentan en el código.
- Los resultados intermedios (parámetros estimados, residuos, *spread* construido) se guardan en archivos CSV/JSON para inspección y verificación independiente.

3.8. Limitaciones metodológicas

Es necesario reconocer las principales limitaciones de la metodología empleada, que se exploran con mayor detalle en el Capítulo 8:

- **Frecuencia diaria.** El análisis se realiza sobre datos de cierre diario. Una implementación operativa real requeriría información intradiaria (al menos horaria) y una modelización explícita de la microestructura del mercado, lo que excede el alcance de un Trabajo Fin de Grado.
- **Modelo de costes simplificado.** El supuesto de costes lineales ignora componentes no lineales relevantes en operativa real (impacto de mercado, efectos de tamaño, costes de financiación de las posiciones cortas).
- **Ausencia de modelización del riesgo de financiación.** La operativa con posiciones cortas sobre el VXX implica costes de préstamo (*borrow*) que pueden ser materiales

en periodos de alta demanda. Este coste se aproxima dentro del *spread* combinado pero no se modeliza explícitamente.

- **Backtesting único *out-of-sample.** El procedimiento *walk-forward simple* es estándar pero menos exigente que un *cross-validation de series temporales* o un *walk-forward** con múltiples ventanas. Esta limitación se reconoce y se sugieren extensiones en el Capítulo 8.
- **Ausencia de datos de futuros del VIX.** La modelización de la *term structure* completa de la volatilidad —que enriquecería la comprensión del *contango* y del *roll yield*— exigiría datos de futuros del VIX que excedían las posibilidades del trabajo. Se ha optado por trabajar exclusivamente con el VXX, que ya incorpora estos efectos en su precio observado.

A pesar de estas limitaciones, la metodología propuesta constituye un marco de análisis riguroso y reproducible, suficiente para contrastar las hipótesis formuladas y extraer conclusiones empíricamente fundamentadas sobre la dinámica del par SPY-VXX y la viabilidad de las estrategias de arbitraje estadístico construidas sobre dicho par.

Capítulo 4. Datos y análisis exploratorio

El presente capítulo presenta el conjunto de datos utilizado en este Trabajo Fin de Grado y desarrolla el análisis exploratorio que constituye el punto de partida de la modelización econométrica. La estructura del capítulo refleja el flujo lógico habitual del *Exploratory Data Analysis* (Tukey, 1977): comenzando por la descripción de las fuentes y la naturaleza de las series (§4.1), se procede a la caracterización estadística univariante (§4.2), al análisis de la estructura de dependencia entre activos (§4.3) y al estudio de los regímenes de mercado (§4.4). El capítulo concluye (§4.5) con una síntesis de los hechos estilizados identificados, que motivarán las elecciones de modelización del Capítulo 5.

4.1. Datos utilizados

4.1.1. Fuente y periodo

Los datos diarios del SPY y del VXX proceden de FactSet, descargados con la herramienta *Price History* en la modalidad Total Return Gross, que ajusta los precios por dividendos y por todas las acciones corporativas relevantes (incluido el *reverse split* 4-a-

1 del VXX en agosto de 2022). Adicionalmente, se han incorporado los datos diarios del índice VIX procedentes del histórico oficial del CBOE.

El periodo de análisis comprende desde el 18 de enero de 2018 —fecha de lanzamiento de la *iPath Series B S&P 500 VIX Short-Term Futures ETN*— hasta el 28 de abril de 2026. Tras la sincronización de fechas y la depuración descrita en el Capítulo 3, el conjunto de datos final contiene 2.080 observaciones diarias por activo —lo que equivale a aproximadamente 8,28 años naturales y 2.079 retornos logarítmicos diarios—.

4.1.2. Variables y transformaciones

Para cada activo se han obtenido las siguientes variables: precio de cierre ajustado, precio de apertura, precio máximo, precio mínimo, volumen negociado y serie *Total Return Gross*. La modelización posterior se realiza sobre los retornos logarítmicos diarios calculados sobre la serie *Total Return Gross*:

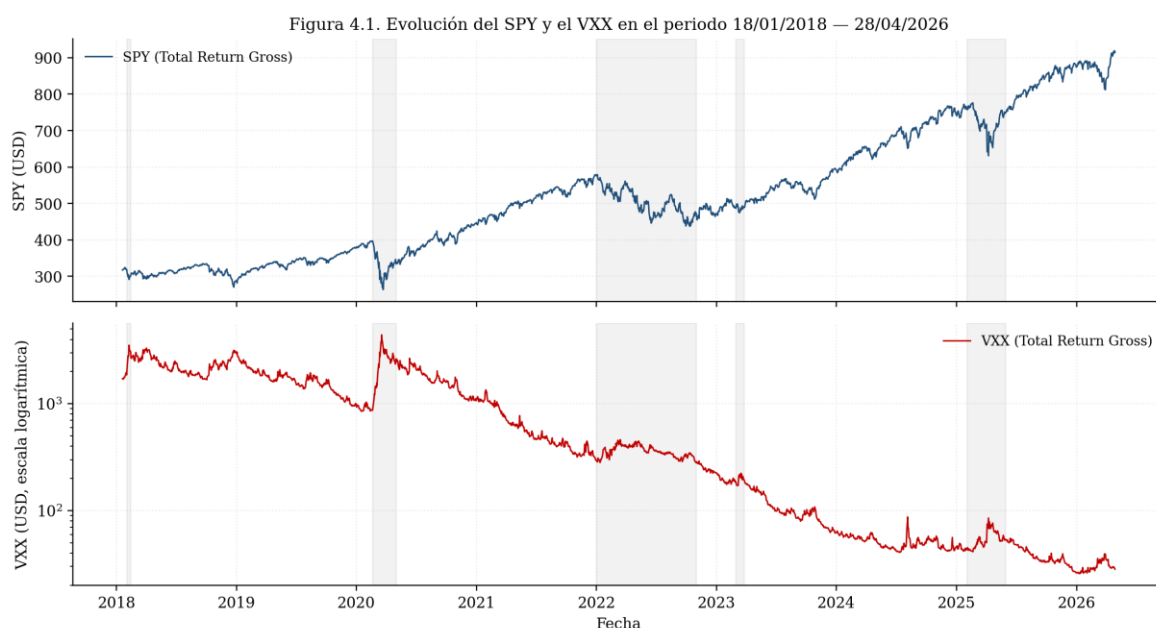
$$r_t = \ln\left(\frac{TR_t}{TR_{t-1}}\right)$$

La elección de la serie *Total Return* —en lugar de la serie de precios desnudos— es metodológicamente relevante: garantiza que los retornos representan el rendimiento total efectivamente acumulable por un inversor con reinversión de dividendos, y elimina los saltos artificiales asociados a las fechas de pago de dividendos (especialmente relevantes en el SPY) y al *reverse split* del VXX.

4.2. Análisis univariante de los activos

4.2.1. Evolución temporal de los precios

La Figura 4.1 representa la evolución temporal de los precios de cierre ajustados del SPY (panel superior) y del VXX (panel inferior, en escala logarítmica). Las áreas sombreadas marcan los principales episodios de tensión durante el periodo: el *Volmageddon* de febrero de 2018, la pandemia de COVID-19 (febrero-abril de 2020), las tensiones inflacionarias y monetarias de 2022, la crisis bancaria regional estadounidense de marzo de 2023 y la fase de tensiones arancelarias del primer semestre de 2025.



El comportamiento divergente de los dos activos a largo plazo es evidente y estadísticamente abrumador. Mientras el SPY ha experimentado una apreciación sostenida —pasando de aproximadamente 280 USD en enero de 2018 a alrededor de 712 USD a finales de abril de 2026—, el VXX ha exhibido una pérdida de valor estructural casi monótonica, con caídas extremadamente pronunciadas en los periodos de calma de mercado. La escala logarítmica del panel inferior es necesaria precisamente para visualizar la magnitud de esta deriva descendente.

La **Figura 4.2** presenta esta divergencia de forma más nítida, normalizando ambos activos a base 100 en la fecha inicial:



La rentabilidad acumulada del SPY en el periodo asciende a **+189,9%**, equivalente a una rentabilidad anualizada del **+13,0%**, mientras que la del VXX se sitúa en **-98,4%**

(pérdida acumulada del 98,4% del valor inicial), equivalente a una rentabilidad anualizada del -49,9%. Esta divergencia, lejos de ser un accidente histórico del periodo concreto, es la manifestación directa de los mecanismos económicos descritos en el Capítulo 2: el *contango* sistemático de los futuros del VIX y el *roll yield* negativo asociado al rebalanceo continuo del VXX. El producto, en su forma actual, está estructuralmente diseñado para perder valor en horizontes largos, y debe entenderse como un instrumento de cobertura táctica más que como una inversión sostenible en el tiempo.

4.2.2. Estadísticos descriptivos de los retornos

La Tabla 4.1 reporta los principales estadísticos descriptivos de los retornos logarítmicos diarios de ambos activos.

Tabla 4.1. Estadísticos descriptivos de los retornos logarítmicos diarios (18/01/2018 — 28/04/2026, N = 2.079)

Estadístico	SPY	VXX
Media diaria (%)	0,0512	-0,1982
Desviación típica diaria (%)	1,2200	4,5137
Asimetría (Skewness)	-0,5605	+1,3239
Exceso de curtosis	13,41	7,09
Mínimo diario (%)	-11,59	-23,12
Máximo diario (%)	+9,99	+33,09
Rentabilidad anualizada (%)	+12,90	-49,94
Volatilidad anualizada (%)	19,37	71,65
Ratio de Sharpe (rf=0)	+0,67	-0,70
Estadístico Jarque-Bera	15.687	4.966
p-valor JB	< 0,0001	< 0,0001

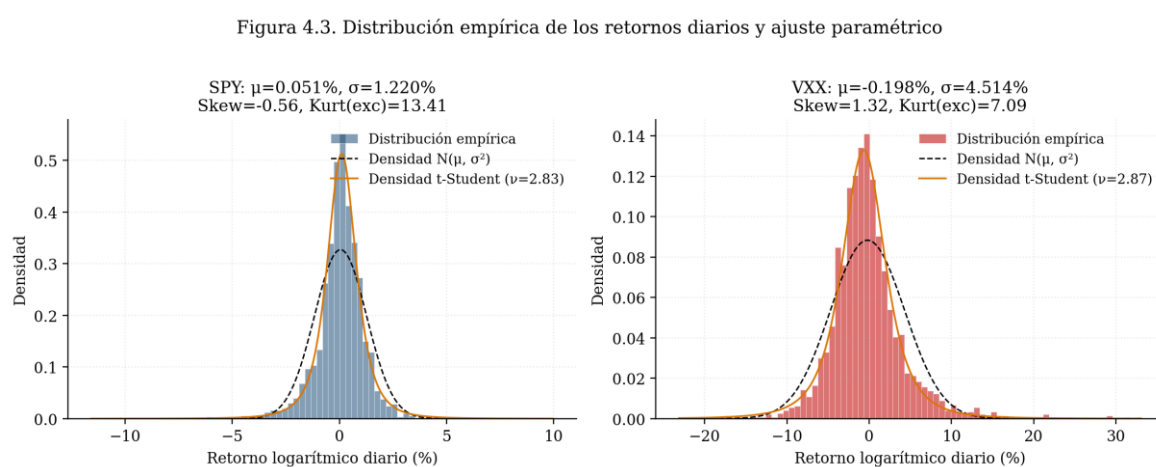
Los datos de la Tabla 4.1 son muy informativos. La volatilidad del VXX es 3,7 veces superior a la del SPY —71,7 % frente a 19,4 % anualizadas—, una diferencia que no es trivial: refleja que el VXX es un instrumento de exposición a la volatilidad implícita de corto plazo, cuya naturaleza misma exige un comportamiento mucho más volátil que el subyacente. La asimetría es de signo opuesto en ambos activos: negativa para el SPY (más caídas extremas que subidas extremas) y positiva para el VXX (más subidas extremas que caídas extremas), lo que es consistente con su carácter de cobertura. La curtosis es muy elevada en ambos casos —excesos de 13,4 y 7,1 respectivamente, frente

al valor 0 de una distribución normal—, indicando colas mucho más pesadas que las gaussianas.

El test de Jarque-Bera rechaza contundentemente la hipótesis de normalidad para ambos activos (estadísticos de 15.687 y 4.966, frente al valor crítico de 5,99 al 5 %), lo que motiva el uso posterior de distribuciones t de Student en los modelos GARCH.

4.2.3. Distribución empírica de los retornos

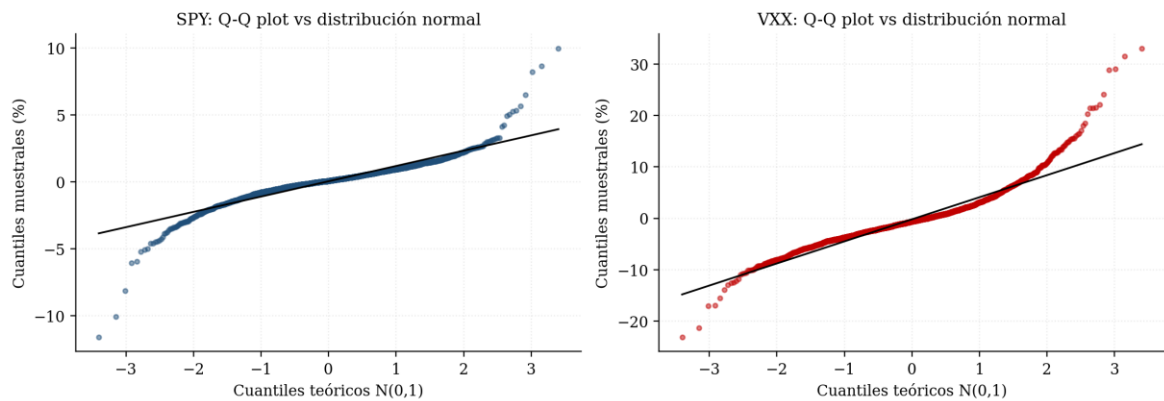
La Figura 4.3 representa la distribución empírica de los retornos diarios junto con los ajustes paramétricos correspondientes a las distribuciones normal y t de Student.



La inadecuación de la distribución normal es manifiesta en ambos casos: la densidad gaussiana subestima sistemáticamente la probabilidad de los retornos cercanos a cero (kurtosis por exceso) y, sobre todo, la probabilidad de los retornos extremos (colas pesadas). La distribución t de Student, en cambio, ofrece un ajuste visualmente excelente, con grados de libertad estimados de aproximadamente 2,8 en ambos activos —un valor extremadamente bajo, indicativo de colas particularmente pesadas, en línea con los estadísticos de curtosis reportados—.

La Figura 4.4 confirma estas observaciones mediante los gráficos cuantil-cuantil (Q-Q plots) frente a la distribución normal. Las desviaciones sistemáticas de la línea diagonal en las dos colas son inequívocas: ambas series presentan colas más pesadas que la distribución normal, especialmente la cola izquierda en el SPY (consistente con la asimetría negativa) y la cola derecha en el VXX (consistente con la asimetría positiva).

Figura 4.4. Gráficos cuantil-cuantil frente a la distribución normal



4.2.4. Tests de estacionariedad

La Tabla 4.2 reporta los resultados de los tests ADF y KPSS aplicados a los niveles, los logaritmos y los retornos de cada activo.

Tabla 4.2. Tests de estacionariedad ADF y KPSS

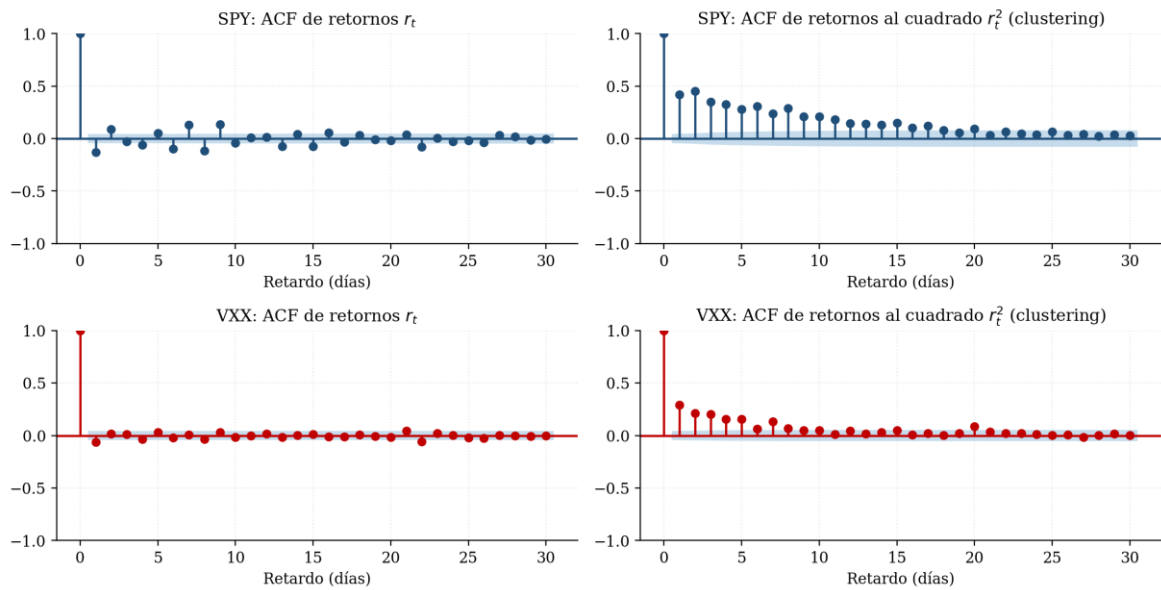
Serie	ADF stat	ADF p-valor	KPSS stat	KPSS p-valor	Conclusión
SPY (niveles)	0,427	0,983	6,435	< 0,01	I(1)
VXX (niveles)	-2,091	0,248	5,982	< 0,01	I(1)
SPY (log)	-0,339	0,920	6,622	< 0,01	I(1)
VXX (log)	-0,083	0,951	7,110	< 0,01	I(1)
SPY (retornos log)	-14,431	< 0,001	0,054	> 0,10	I(0)
VXX (retornos log)	-48,461	< 0,001	0,098	> 0,10	I(0)

Los resultados son inequívocos y mutuamente coherentes: las series de niveles y log-niveles son no estacionarias (no se rechaza la hipótesis nula de raíz unitaria en ADF y se rechaza la hipótesis nula de estacionariedad en KPSS), mientras que las series de retornos son claramente estacionarias (se rechaza ADF al 1 % y no se rechaza KPSS). Las series son por tanto integradas de orden 1, $I(1)$, lo que constituye la condición previa necesaria —aunque no suficiente— para la posterior aplicación de los tests de cointegración.

4.2.5. Autocorrelación y heterocedasticidad

La Figura 4.8 representa las funciones de autocorrelación (ACF) de los retornos y de los retornos al cuadrado.

Figura 4.8. Funciones de autocorrelación de retornos y retornos al cuadrado



Los resultados son característicos de las series financieras de alta frecuencia. La ACF de los retornos en niveles muestra autocorrelaciones moderadas pero persistentes en el SPY y prácticamente nulas en el VXX más allá del primer retardo: el test de Ljung-Box rechaza la independencia serial al 5 % en ambos activos para los primeros retardos. La ACF de los retornos al cuadrado, en cambio, exhibe autocorrelaciones positivas, significativas y persistentes durante varias decenas de retardos en ambos activos.

Esta diferencia entre la ACF de los retornos y la de los retornos al cuadrado es la firma estadística canónica del fenómeno de **volatility clustering: los retornos son aproximadamente independientes en términos lineales, pero su magnitud presenta una fuerte estructura temporal. Los tests formales confirman esta observación: el test ARCH-LM** de Engle (1982) rechaza la hipótesis nula de ausencia de efectos ARCH con niveles de significación inferiores al 0,001 % en ambos activos (estadístico LM de 588 para el SPY y 255 para el VXX, con cinco retardos).

La presencia de heterocedasticidad condicional motiva, como se desarrollará en el Capítulo 5, la aplicación de modelos GARCH y EGARCH para la modelización de la dinámica de la volatilidad.

4.3. Análisis bivariante: la relación SPY-VXX

4.3.1. Correlaciones contemporáneas

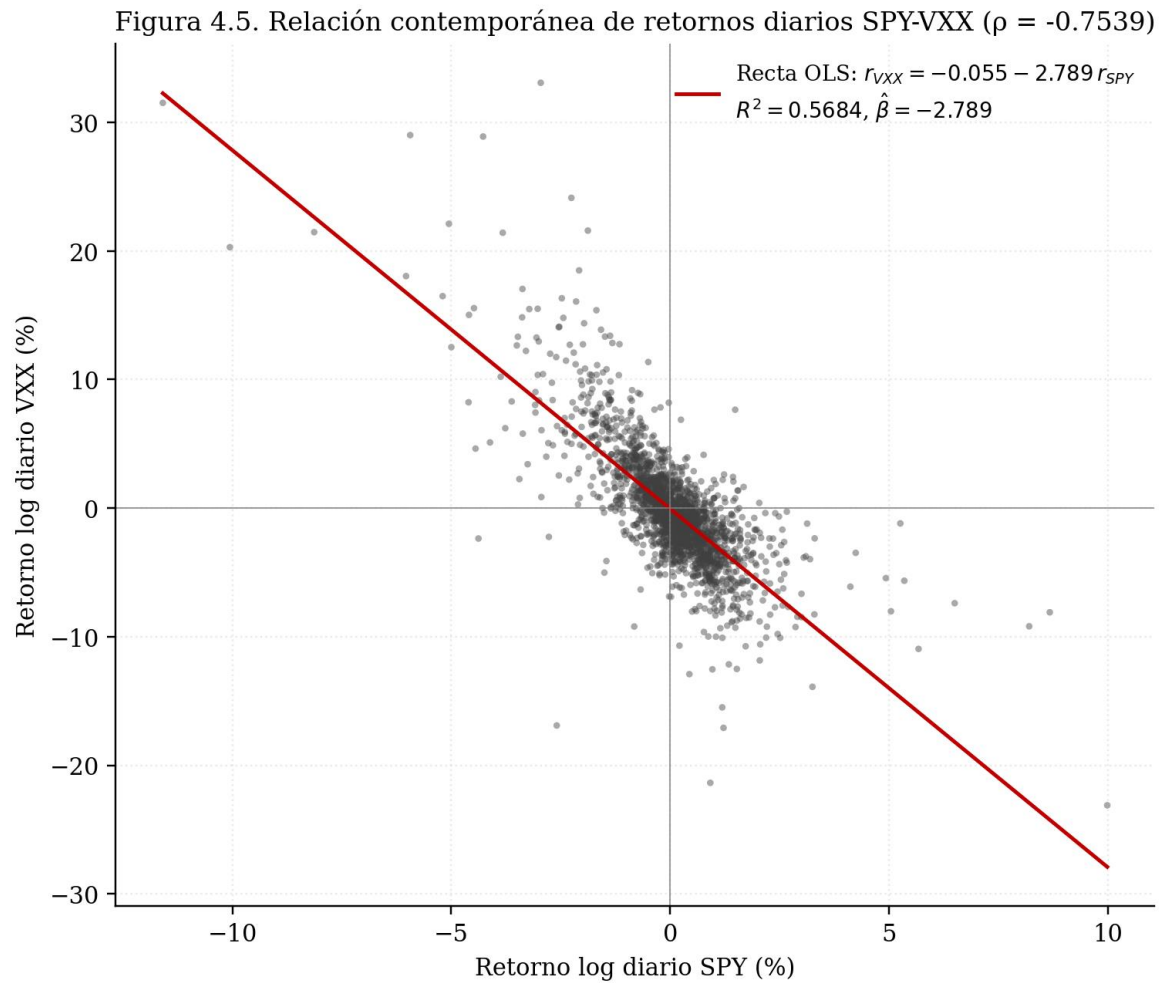
La Tabla 4.3 reporta las correlaciones lineales (Pearson) entre los dos activos calculadas sobre tres transformaciones distintas: niveles, logaritmos y retornos logarítmicos.

Tabla 4.3. Correlaciones contemporáneas SPY-VXX

Transformación	Correlación de Pearson
Niveles (precios)	−0,8125
Log-niveles	−0,9500
Retornos logarítmicos	−0,7539

Las tres correlaciones son negativas y de magnitud sustancial, en línea con la hipótesis H1. La correlación más elevada en log-niveles (−0,95) refleja en gran medida la deriva determinística común de las dos series —el SPY tendiendo al alza y el VXX tendiendo a la baja— y, por tanto, no constituye una medida limpia de la dependencia económica entre los activos. La correlación más informativa es la de los retornos logarítmicos (−0,75), que mide la dependencia contemporánea de las innovaciones diarias.

La Figura 4.5 representa gráficamente esta relación mediante un diagrama de dispersión y la recta de regresión por mínimos cuadrados ordinarios.



La regresión OLS de los retornos del VXX sobre los del SPY proporciona la siguiente relación contemporánea:

$$r_{VXX,t} = -0,055 - 2,789 r_{SPY,t} + \hat{\varepsilon}_t, \quad R^2 = 0,5684$$

El coeficiente $\hat{\beta} = -2,789$ admite una interpretación económica directa: por cada punto porcentual de variación del SPY, el VXX se mueve 2,79 puntos porcentuales en sentido contrario, en términos esperados. Esta magnitud refleja el carácter apalancado del VXX como producto de exposición a volatilidad: la volatilidad implícita es, por construcción, una función convexa del precio del subyacente, y los productos de corto plazo sobre futuros del VIX amplifican esta no linealidad. El coeficiente de determinación $R^2 = 0,57$ indica que aproximadamente el 57 % de la varianza de los retornos diarios del VXX es explicado por los retornos contemporáneos del SPY, dejando un residuo del 43 % asociado a factores idiosincrásicos —principalmente, cambios en la estructura temporal de los futuros del VIX, primas de riesgo y *flow* específico—.

4.3.2. Evolución temporal de la correlación

La correlación contemporánea entre los retornos no es, sin embargo, una constante temporal. La Figura 4.6 representa la correlación móvil (*rolling*) en ventanas de 60 y 120 días.



La correlación móvil oscila entre $-0,40$ y $-0,96$ durante el periodo de estudio, con una media en torno a $-0,79$. Esta variabilidad sustancial es de gran relevancia operativa: una correlación menos negativa indica una desconexión temporal entre los dos activos, que puede ser una oportunidad o una señal de alerta para la estrategia de *pair trading*. Los episodios más notables de correlación menos negativa se concentran en periodos de volatilidad moderada con flujos de cobertura específicos (mediados de 2022, mediados de 2023), mientras que los episodios de correlación más negativa (cercana a $-0,95$) coinciden con periodos de tensión generalizada (marzo de 2020, primavera de 2025).

La Tabla 4.4 desagrega la correlación por subperiodos anuales, junto con la volatilidad anualizada de cada activo en el correspondiente subperiodo.

Tabla 4.4. Correlación y volatilidad por subperiodo anual

Periodo	N obs	ρ (SPY, VXX)	σ SPY anual.	σ VXX anual.
2018	239	$-0,8128$	17,4 %	77,2 %
2019	252	$-0,8706$	12,5 %	58,3 %
2020 (COVID)	253	$-0,7722$	33,6 %	94,9 %
2021	252	$-0,8433$	13,0 %	75,3 %
2022 (Inflación)	251	$-0,6873$	24,3 %	57,8 %
2023	250	$-0,7265$	13,1 %	55,6 %
2024	252	$-0,7434$	12,6 %	74,9 %
2025 (Aranceles)	250	$-0,8889$	19,3 %	71,8 %
2026 (parcial)	80	$-0,8512$	14,2 %	69,2 %

La correlación más débil —en términos absolutos— se observa en 2022, durante las tensiones inflacionarias y el endurecimiento monetario de la Reserva Federal. Este resultado es coherente con la literatura: en regímenes inflacionarios, los movimientos de la renta variable tienen una componente significativa atribuible al tipo de descuento, lo que altera la dinámica habitual entre acciones y volatilidad. Por el contrario, en 2025 — con la volatilidad inducida por la política arancelaria— la correlación se sitúa en su mínimo histórico del periodo (−0,89), reflejando un comportamiento más prototípico de "risk-off".

4.3.3. Eventos extremos

La Tabla 4.5 reporta los cinco mayores movimientos diarios del SPY (al alza y a la baja) y los retornos contemporáneos del VXX en esos días.

Tabla 4.5. Mayores movimientos diarios del SPY y retornos contemporáneos del VXX

Fecha	Retorno SPY	Retorno VXX
Caídas		
16/03/2020	−11,59 %	+31,52 %
12/03/2020	−10,06 %	+20,30 %
09/03/2020	−8,13 %	+21,47 %
04/04/2025	−6,03 %	+18,05 %
11/06/2020	−5,94 %	+29,02 %
Subidas		
09/04/2025	+9,99 %	−23,12 %
24/03/2020	+8,67 %	−8,10 %
13/03/2020	+8,20 %	−9,19 %
06/04/2020	+6,50 %	−7,40 %
26/03/2020	+5,67 %	−10,95 %

Cuatro de las cinco mayores caídas del SPY se concentran en marzo de 2020 (crisis del COVID-19), mientras que cuatro de las cinco mayores subidas también ocurren en ese mismo mes —reflejando la elevada volatilidad bidireccional del periodo—. La quinta caída más fuerte se produce el 4 de abril de 2025, día en que el VXX se aprecia un +18 %, mientras que la mayor subida del SPY del periodo (+10,0 %) ocurre cinco días

después, el 9 de abril de 2025, con una caída del VXX del -23% . Estos episodios ilustran el fenómeno descrito en la H1: la correlación negativa entre los activos no solo está presente en los momentos de tensión, sino que se intensifica precisamente en los días de mayor magnitud absoluta de los movimientos.

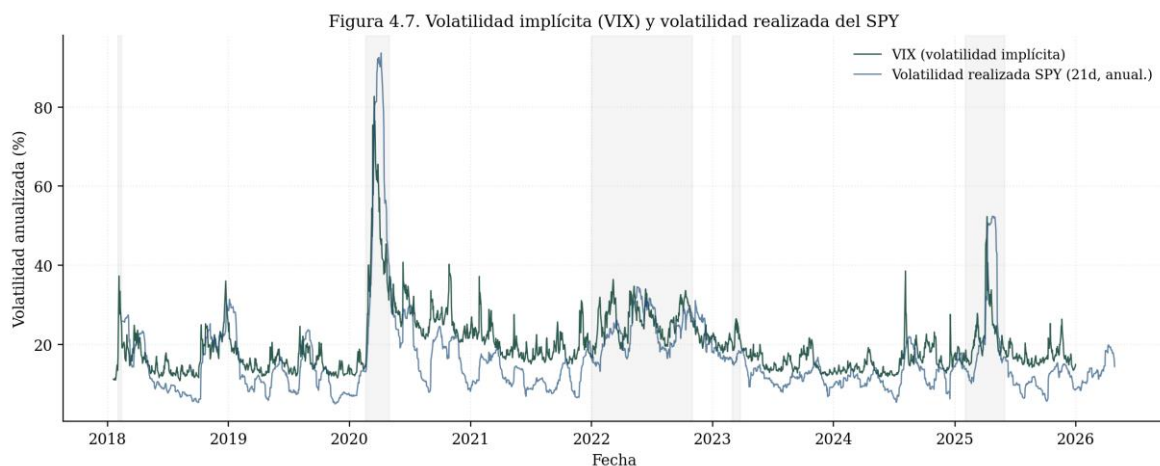
4.3.4. Asimetría de la respuesta del VXX

Una observación adicional relevante puede extraerse de la Tabla 4.5: la magnitud absoluta del movimiento del VXX en respuesta a una caída del SPY tiende a ser **mayor** que la magnitud absoluta de su respuesta a una subida del SPY de igual cuantía. Por ejemplo, frente a la caída del SPY del $-11,6\%$ del 16/03/2020 el VXX subió un $+31,5\%$; pero frente a la subida del SPY del $+8,7\%$ del 24/03/2020, el VXX cayó solo un $-8,1\%$. Este patrón es la manifestación bilateral del *leverage effect* clásico (Black, 1976) y motiva la utilización de modelos asimétricos como el EGARCH en el Capítulo 5.

4.4. Regímenes de volatilidad y comparación con el VIX

4.4.1. El VIX como variable de contexto

La Figura 4.7 representa la evolución del VIX junto con la volatilidad realizada del SPY (calculada sobre ventanas móviles de 21 días, anualizada).



El VIX se mantiene durante la mayor parte del periodo dentro del rango histórico de los 12-25 puntos, con tres episodios notables de superación del nivel 40: febrero de 2018 (Volmageddon), marzo de 2020 (COVID) y abril de 2025 (aranceles). La diferencia sistemática entre el VIX y la volatilidad realizada subsiguiente —el primero supera, en términos esperados, a la segunda— constituye la *variance risk premium* descrita en el

Capítulo 2: una manifestación empírica de la prima que los participantes del mercado pagan por la cobertura frente a incrementos no anticipados de la volatilidad.

4.4.2. Correlación condicional al régimen de volatilidad

Para evaluar la sensibilidad de la correlación al régimen de volatilidad imperante, se segmenta la muestra en cuartiles del VIX y se calcula la correlación entre los retornos del SPY y del VXX en cada cuartil. La Tabla 4.6 reporta los resultados.

Tabla 4.6. Correlación SPY-VXX por cuartil del VIX

Régimen	Rango VIX	N	ρ (SPY, VXX)
Q1 (calma)	$VIX < 14,87$	500	-0,6415
Q2	$14,87 \leq VIX < 17,90$	497	-0,7844
Q3	$17,90 \leq VIX < 22,62$	502	-0,7996
Q4 (estrés)	$VIX \geq 22,62$	500	-0,7513

El patrón es interesante. La correlación es menos negativa en los regímenes extremos —tanto en el cuartil de calma (-0,64) como en el cuartil de estrés (-0,75)— y más negativa en los regímenes intermedios (-0,78 y -0,80). Una explicación plausible es la siguiente: en regímenes de calma, el VXX está dominado por la deriva determinista del *contango* (que es relativamente independiente de los movimientos diarios del SPY), debilitando la correlación; en regímenes de estrés extremo, las dinámicas de cobertura institucional, los gaps de liquidez y los flujos especulativos pueden generar movimientos del VXX que no se corresponden estrictamente con los del SPY contemporáneo, debilitando también la correlación. Es en los regímenes "normales" (Q2 y Q3) donde la dinámica clásica subyacente domina y la correlación alcanza valores extremos.

Esta observación tiene implicaciones operativas no triviales para el diseño de la estrategia: el régimen ideal para una estrategia de *pair trading* basada en reversión a la media —donde los activos están más fuertemente vinculados— corresponde a los cuartiles intermedios de volatilidad, no a los extremos.

4.5. Hechos estilizados y síntesis

El análisis exploratorio desarrollado en este capítulo ha permitido caracterizar empíricamente las propiedades estadísticas del SPY y el VXX durante el periodo 2018-

2026 y la estructura de dependencia entre ambos activos. Los principales **hechos estilizados** identificados se resumen en los puntos siguientes:

HE-1. Divergencia estructural a largo plazo. El SPY ha generado una rentabilidad acumulada del +189,9 % (anualizada del +13,0 %) frente a una pérdida acumulada del VXX del -98,4 % (anualizada del -49,9 %). Esta divergencia es la manifestación directa del *contango* sistemático y del *roll yield* negativo del VXX.

HE-2. Heterogeneidad de los momentos de orden superior. Ambas series rechazan claramente la normalidad. Presentan asimetrías de signo opuesto (SPY negativa, VXX positiva) y curtosis muy elevadas (excesos de 13,4 y 7,1), compatibles con la elección de distribuciones *t* de Student en los modelos paramétricos posteriores.

HE-3. Volatility clustering significativo. Los tests de Ljung-Box sobre retornos al cuadrado y los tests ARCH-LM rechazan contundentemente la independencia y la homocedasticidad serial. Las funciones de autocorrelación de los retornos al cuadrado exhiben persistencia visible hasta los 30 retardos. Estas propiedades motivan la modelización GARCH/EGARCH del Capítulo 5.

HE-4. Series $I(1)$ en niveles y log-niveles, $I(0)$ en retornos. Los tests ADF y KPSS coinciden en este resultado, condición necesaria para la aplicación de los tests de cointegración.

HE-5. Correlación negativa estructural pero variable. La correlación entre los retornos diarios es persistentemente negativa (-0,75 en media para todo el periodo, entre -0,40 y -0,96 en ventanas móviles), pero presenta variabilidad temporal sustancial.

HE-6. Beta apalancada. La regresión OLS contemporánea muestra una *beta* de los retornos del VXX respecto del SPY de aproximadamente -2,79, reflejando el apalancamiento implícito del producto sobre la volatilidad de corto plazo.

HE-7. Asimetría en la respuesta del VXX. Las caídas del SPY generan respuestas en magnitud absoluta superior del VXX que las subidas equivalentes, manifestación del *leverage effect* canónico.

HE-8. Modulación por régimen. La intensidad de la correlación entre activos varía con el régimen de volatilidad imperante, con valores extremos en los regímenes intermedios y valores menos extremos en los regímenes de calma o de estrés agudo.

Estos ocho hechos estilizados serán los que guíen las decisiones metodológicas del Capítulo 5 (modelización econométrica) y del Capítulo 6 (diseño de la estrategia). En particular, motivan: (i) la aplicación de los tests de cointegración pese a la baja probabilidad *a priori* de cointegración estricta (HE-1); (ii) la utilización de distribuciones *t* de Student en los modelos GARCH (HE-2); (iii) la elección del EGARCH frente al GARCH simétrico (HE-3, HE-7); (iv) la aplicación del DCC-GARCH para capturar la variabilidad temporal de la correlación (HE-5, HE-8); y (v) la construcción del *spread* mediante una *hedge ratio* dinámica en lugar de un coeficiente fijo (HE-5, HE-6).

Capítulo 5. Modelización econométrica

El presente capítulo desarrolla la modelización econométrica de la dinámica univariante y conjunta de los retornos del SPY y del VXX a partir de las técnicas descritas en los Capítulos 2 y 3. La estructura del capítulo refleja la lógica progresiva de la modelización: comenzando por los tests de cointegración (§5.1), pasando por los modelos univariantes ARIMA (§5.2), el modelo bivalente VAR con análisis de causalidad de Granger (§5.3), los modelos de volatilidad condicional GARCH y EGARCH (§5.4), el modelo de correlación dinámica DCC-GARCH (§5.5) y, finalmente, la construcción y caracterización del *spread* sintético entre los dos activos (§5.6). El capítulo cierra (§5.7) con un balance integrado de los resultados, que constituye el puente hacia el diseño operativo de la estrategia desarrollado en el Capítulo 6.

5.1. Tests de cointegración

5.1.1. Test de Engle-Granger

El test de cointegración de Engle-Granger se aplica en dos etapas. En la primera, se estima por mínimos cuadrados ordinarios la regresión cointegrante entre los logaritmos de los precios, considerando alternativamente cada activo como variable dependiente:

$$\ln(VXX_t) = \alpha + \beta \ln(SPY_t) + u_t$$

La estimación arroja $\hat{\beta} = -2,1247$ con un coeficiente de determinación $R^2 = 0,9023$. El coeficiente negativo es consistente con la relación inversa esperada: una variación porcentual del SPY se asocia, en términos de niveles logarítmicos, a una variación porcentual del VXX en sentido contrario y de mayor magnitud.

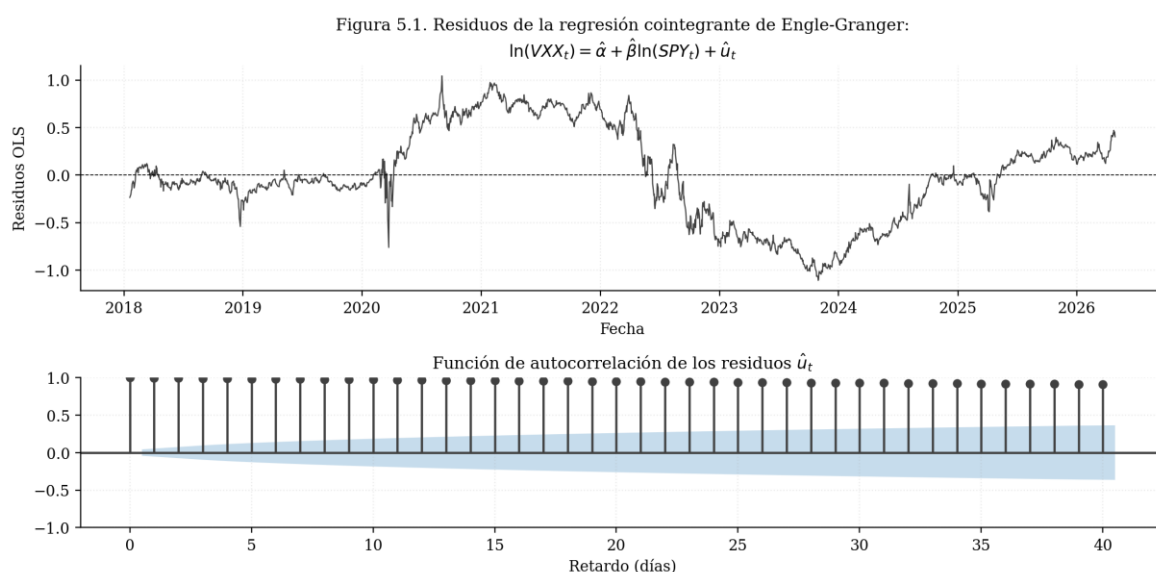
En la segunda etapa, se aplica el test ADF a los residuos $\hat{u}_t$ con los valores críticos específicos de Engle-Granger. Los resultados se reportan en la Tabla 5.1.

Tabla 5.1. Test de cointegración de Engle-Granger

Especificación	Estadístico	p-valor	CV 5 %	Conclusión
Niveles, SPY $\rightarrow$ VXX	-2,4163	0,3260	-3,3617	No cointegradas
Niveles, VXX $\rightarrow$ SPY	-1,7531	0,6664	-3,3617	No cointegradas
Log-niveles, SPY $\rightarrow$ VXX	-1,3625	0,8200	-3,3617	No cointegradas
Log-niveles, VXX $\rightarrow$ SPY	-1,3489	0,8252	-3,3617	No cointegradas

Los cuatro tests coinciden en no rechazar la hipótesis nula de presencia de raíz unitaria en los residuos: los p-valores oscilan entre 0,326 y 0,825, todos sustancialmente superiores al nivel de significación del 5 %. La conclusión es inequívoca: las series del SPY y del VXX no presentan una relación de cointegración en sentido de Engle-Granger durante el periodo 2018-2026.

La Figura 5.1 ilustra gráficamente la causa de este resultado. El panel superior representa los residuos $\hat{u}_t$ de la regresión cointegrante log-lineal: su evolución temporal exhibe una persistencia evidente, con desviaciones que tardan meses o años en revertirse. El panel inferior, que reporta la función de autocorrelación de los residuos, confirma esta persistencia: las autocorrelaciones decaen lentamente, manteniendo valores superiores a 0,8 incluso en retardos de 40 días, comportamiento típico de un proceso $I(1)$.



5.1.2. Test de Johansen

Para verificar la robustez del resultado, se aplica adicionalmente el test multivariante de Johansen (1988, 1991) sobre el sistema bivalente $(\ln SPY_t, \ln VXX_t)$, con un retardo de diferencia y especificación con constante. Los resultados se reportan en la Tabla 5.2.

Tabla 5.2. Test de cointegración de Johansen

H ₀	Trace stat	CV 5 %	Max-Eigen stat	CV 5 %	Conclusión
$r = 0$	13,4180	15,4943	10,5295	14,2646	No rechaza
$r \leq 1$	2,8885	3,8415	2,8885	3,8415	No rechaza

Tanto el estadístico de la traza como el del valor propio máximo se sitúan por debajo de los respectivos valores críticos al 5 %, en ambas hipótesis. La conclusión es coherente con la del test de Engle-Granger: el rango de cointegración del sistema es cero, es decir, no existe ninguna combinación lineal estacionaria de los logaritmos del SPY y del VXX.

5.1.3. Interpretación económica

El resultado anterior, lejos de constituir una limitación del análisis, aporta una conclusión empírica relevante que confirma la H2 formulada en el Capítulo 1. La razón económica del rechazo de la cointegración es directa: el VXX presenta una deriva determinística descendente estructural —debida al *contango* sistemático y al *roll yield* negativo—, mientras que el SPY presenta una deriva ascendente. Aunque ambas series se mueven contemporáneamente con elevada correlación negativa en los retornos, sus medias

incondicionales evolucionan en sentidos opuestos a velocidades distintas. Una combinación lineal $\ln(VXX_t) - \beta \ln(SPY_t)$ con β constante hereda inevitablemente alguna de estas dos derivas y, por tanto, no puede ser estacionaria en niveles.

Esta conclusión motiva la transición metodológica del enfoque clásico de cointegración estricta hacia el enfoque más flexible de reversión a la media sobre un *spread* dinámico, desarrollado en §5.6 y formalizado operativamente en el Capítulo 6.

5.2. Modelos ARIMA univariantes

La selección del orden óptimo de los modelos ARIMA se realiza mediante una búsqueda exhaustiva en el espacio $\{0, 1, 2, 3\}^2$ sobre los retornos logarítmicos de cada activo, eligiendo el modelo que minimiza el criterio de información bayesiano (BIC). Los resultados se resumen en la Tabla 5.3.

Tabla 5.3. Modelos ARIMA óptimos por BIC

Activo	Orden óptimo	AIC	BIC	Log-verosimilitud
SPY	ARIMA(2,0,2)	6.616,25	6.650,09	-3.302,13
VXX	ARIMA(1,0,0)	12.163,63	12.180,55	-6.078,82

Los órdenes seleccionados son notablemente distintos. Para los retornos del SPY, el modelo óptimo es un ARIMA(2,0,2) —es decir, un ARMA(2,2)—, lo que indica una estructura dinámica relativamente rica con dos retardos autorregresivos y dos retardos de medias móviles. La presencia de componentes MA significativos es coherente con la microestructura de un activo ampliamente arbitrado como el SPY, donde los flujos institucionales y los efectos de cierre sesional generan correlaciones de corto plazo que se compensan rápidamente.

Para los retornos del VXX, el modelo óptimo es un ARIMA(1,0,0) —un AR(1) puro— con un coeficiente autorregresivo $\hat{\phi}_1 = -0,0616$ que, aunque estadísticamente significativo, es de magnitud reducida. Esta parsimonia es coherente con el comportamiento del producto: el VXX, al replicar futuros del VIX a corto plazo, presenta una dinámica fuertemente influenciada por los choques contemporáneos del subyacente, dejando poco margen para estructuras autocorrelativas complejas.

Es importante señalar que las componentes ARMA capturan únicamente la dinámica lineal de los retornos. Los hechos estilizados HE-3 y HE-7 del Capítulo 4 —*volatility clustering* y asimetría— motivan la modelización adicional de la varianza condicional, desarrollada en §5.4.

5.3. Modelo VAR y causalidad de Granger

5.3.1. Selección del orden y estimación

La dinámica conjunta de los retornos se modeliza mediante un VAR sobre el vector $(r_{SPY,t}, r_{VXX,t})$. La selección del orden p se realiza utilizando los criterios AIC, BIC, FPE y HQIC, evaluados sobre retardos de 1 a 10. Los resultados se sintetizan en la Tabla 5.4.

Tabla 5.4. Selección del orden del modelo VAR (criterios de información)

Retardos	AIC	BIC	FPE	HQIC
1	2,5589	2,5752	12,9223	2,5648
2	2,5544	2,5817	12,8643	2,5644
5	2,5384	2,5984	12,6592	2,5602
9	2,5050	2,6086	12,2399	2,5429
10	2,5072	2,6217	12,2683	2,5491

El AIC, FPE y HQIC seleccionan un orden $p = 9$, mientras que el BIC selecciona $p = 1$. La discrepancia es coherente: el BIC penaliza con mayor intensidad la inclusión de parámetros adicionales (penalización proporcional a $\ln T$), mientras que el AIC y el HQIC son más permisivos con la complejidad. Para preservar la parsimonia y facilitar la interpretación de los efectos dinámicos, se selecciona el VAR(1) como especificación principal, en línea con la elección estándar en aplicaciones financieras de alta frecuencia. Como verificación, los principales contrastes (causalidad de Granger) se replican con el VAR(9), obteniéndose conclusiones idénticas en signo y significación.

5.3.2. Causalidad de Granger

Los tests F de causalidad de Granger sobre el VAR(1) arrojan los resultados de la Tabla 5.5.

Tabla 5.5. Test de causalidad de Granger sobre VAR(1)

Hipótesis nula	Estadístico F	p-valor	Conclusión
VXX no causa-Granger SPY	6,7908	0,0092	Rechaza H_0
SPY no causa-Granger VXX	15,9875	0,0001	Rechaza H_0

Los resultados son inequívocos: existe causalidad de Granger bidireccional entre los retornos del SPY y los del VXX. Ambas hipótesis nulas se rechazan con niveles de significación inferiores al 1 %.

La interpretación económica de este resultado merece ser matizada. La causalidad de Granger es un concepto de predictibilidad estadística, no de causalidad estructural: indica que el pasado de una variable contiene información útil para predecir la otra, más allá de su propio pasado. La direccionalidad bidireccional refleja, por tanto, dos canales complementarios de transmisión informativa entre los mercados de renta variable y de volatilidad:

- **SPY $\rightarrow$ VXX** ($F = 15,99$): los movimientos del SPY anticipan parcialmente los del VXX, lo que es coherente con la idea de que la volatilidad implícita responde a los movimientos de precios subyacentes con un cierto retardo —típicamente atribuible a los procesos de revaluación y rebalanceo de las opciones implicadas en el cálculo del VIX—.
- **VXX $\rightarrow$ SPY** ($F = 6,79$): la dirección inversa también es estadísticamente significativa, aunque con un estadístico menor. Una interpretación plausible es que las dinámicas de los flujos de cobertura sobre el VXX —especialmente en regímenes de tensión— preceden ocasionalmente a los movimientos del SPY, particularmente cuando las posiciones de cobertura son lo suficientemente grandes como para generar impacto de mercado.

La existencia de causalidad bidireccional implica que los dos activos forman un sistema dinámico genuinamente acoplado, en el que la información fluye en ambas direcciones.

Esta observación refuerza la decisión metodológica de modelizar el sistema mediante VAR y, posteriormente, mediante DCC-GARCH —que capturan precisamente esta interdependencia—.

5.4. Modelos de volatilidad condicional

5.4.1. GARCH(1,1)

La Tabla 5.6 reporta los resultados de la estimación por máxima verosimilitud del modelo GARCH(1,1) con distribución t de Student estandarizada para los retornos del SPY y del VXX.

Tabla 5.6. Estimaciones GARCH(1,1) con distribución t de Student

Parámetro	SPY (coef.)	SPY (p-valor)	VXX (coef.)	VXX (p-valor)
μ	0,1162	0,000	-0,5784	0,000
ω	0,0331	0,000	1,9998	0,000
α	0,1641	0,000	0,2861	0,000
β	0,8205	0,000	0,6572	0,000
ν	5,7416	0,000	4,0480	0,000
$\alpha + \beta$	0,9846	—	0,9433	—
Log-verosimilitud	-2.812,70	—	-5.682,39	—
AIC	5.635,40	—	11.374,79	—
BIC	5.663,60	—	11.402,98	—

Los resultados son coherentes con la teoría y con los hechos estilizados identificados en el Capítulo 4. La suma $\alpha + \beta$ —indicador de la persistencia de los choques en la volatilidad— es muy elevada en ambos casos (0,985 para el SPY, 0,943 para el VXX), inferior a la unidad y por tanto compatible con la estacionariedad débil del proceso, pero suficientemente próxima a la unidad como para indicar que los choques sobre la volatilidad se disipan lentamente.

Los grados de libertad estimados de la t de Student ($\nu = 5,74$ para el SPY y $\nu = 4,05$ para el VXX) confirman la presencia de colas pesadas en los residuos estandarizados, en particular para el VXX —donde un valor tan bajo de ν es indicativo de eventos extremos relativamente frecuentes—.

5.4.2. EGARCH(1,1) y *leverage effect*

El modelo EGARCH(1,1) introduce el parámetro de asimetría γ , que mide la respuesta diferencial de la volatilidad ante choques positivos y negativos. La Tabla 5.7 reporta las estimaciones.

Tabla 5.7. Estimaciones EGARCH(1,1) con distribución t de Student

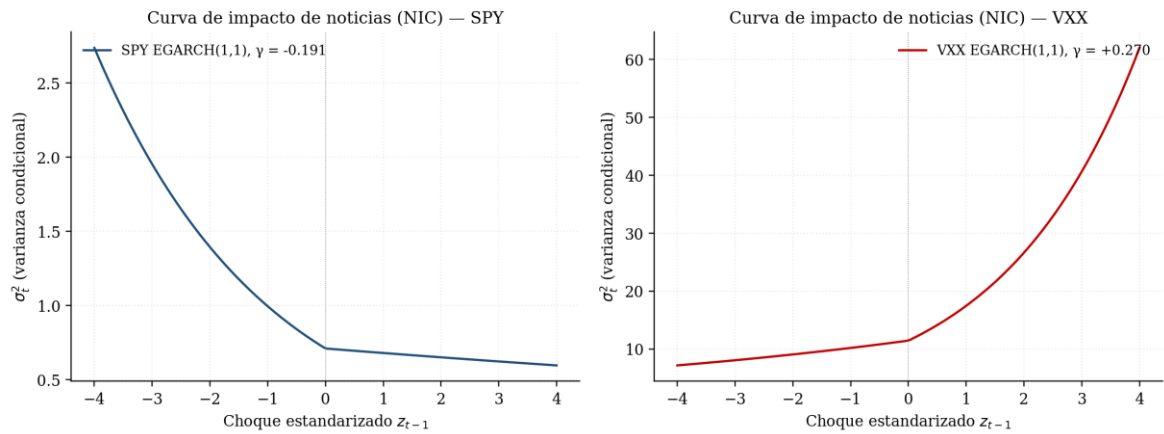
Parámetro	SPY (coef.)	SPY (p-valor)	VXX (coef.)	VXX (p-valor)
μ	0,0787	0,000	-0,4106	0,000
ω	-0,0068	0,163	0,1957	0,000
α	0,1469	0,000	0,1527	0,000
γ (asimetría)	-0,1906	0,000	+0,2698	0,000
β	0,9699	0,000	0,9235	0,000
ν	6,1852	0,000	4,5552	0,000
Log-verosimilitud	-2.767,03	—	-5.645,00	—
AIC	5.546,06	—	11.302,00	—
BIC	5.579,90	—	11.335,84	—

El coeficiente de asimetría γ es estadísticamente significativo al 0,1 % en ambos casos, con signos opuestos:

- Para el **SPY**, $\hat{\gamma} = -0,191$. El signo negativo confirma el **leverage effect** clásico de Black (1976): un choque negativo de magnitud unitaria sobre los retornos del SPY incrementa más la varianza condicional siguiente que un choque positivo de igual magnitud. Específicamente, el efecto sobre $\ln(\sigma_t^2)$ es $\alpha \cdot |z| + \gamma \cdot z = (0,147 + 0,191) \cdot |z| = 0,338 \cdot |z|$ para choques negativos, frente a $(0,147 - 0,191) \cdot |z| = -0,044 \cdot |z|$ para choques positivos —es decir, los choques positivos reducen ligeramente la volatilidad esperada—.
- Para el **VXX**, $\hat{\gamma} = +0,270$. El signo es opuesto al del SPY, lo que es plenamente coherente con la naturaleza inversa del producto: el VXX se aprecia ante caídas del subyacente (correlacionadas con incrementos de la volatilidad), de modo que sus propios choques positivos (subidas de precio) están asociados a contextos de tensión generalizada y, por tanto, a mayores volatilidades subsiguientes.

La Figura 5.3 ilustra gráficamente esta asimetría mediante las curvas de impacto de noticias (NIC; Engle & Ng, 1993): la varianza condicional σ_t^2 como función del choque estandarizado anterior z_{t-1} .

Figura 5.3. Asimetría de la respuesta de la volatilidad: curvas de impacto de noticias EGARCH



La curva del SPY presenta una asimetría marcada a la izquierda: un choque negativo de magnitud 4 induce una varianza condicional cinco veces superior a la inducida por un choque positivo de igual magnitud. La curva del VXX presenta una asimetría especular, con un brazo derecho mucho más pronunciado.

5.4.3. Comparación GARCH vs EGARCH

La elección entre los dos modelos se realiza mediante el criterio de Akaike (AIC). Los resultados son inequívocos:

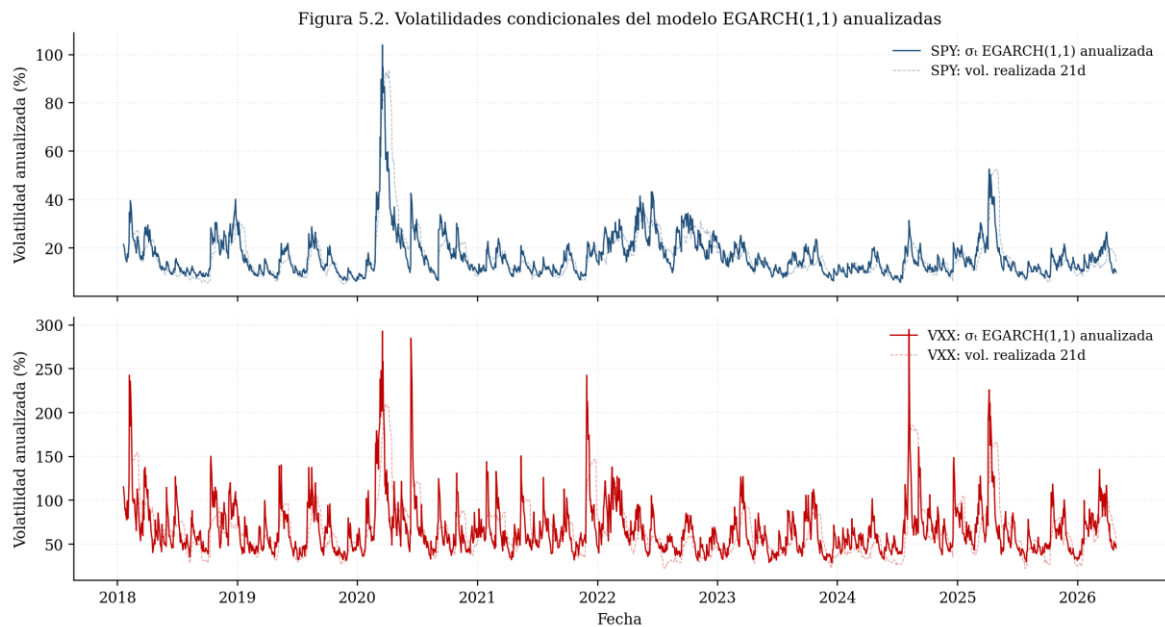
Tabla 5.8. Comparación GARCH vs EGARCH (AIC)

Activo	GARCH AIC	EGARCH AIC	Diferencia	Modelo preferido
SPY	5.635,40	5.546,06	89,34	EGARCH
VXX	11.374,79	11.302,00	72,79	EGARCH

El EGARCH es preferido en ambos casos con márgenes de AIC muy amplios, en línea con la significación estadística del parámetro de asimetría reportada arriba. Esta evidencia justifica la utilización del EGARCH como especificación univariante de referencia en el resto del análisis.

5.4.4. Volatilidades condicionales estimadas

La Figura 5.2 representa las volatilidades condicionales anualizadas obtenidas del modelo EGARCH(1,1) para ambos activos, junto con la volatilidad realizada de 21 días como referencia.

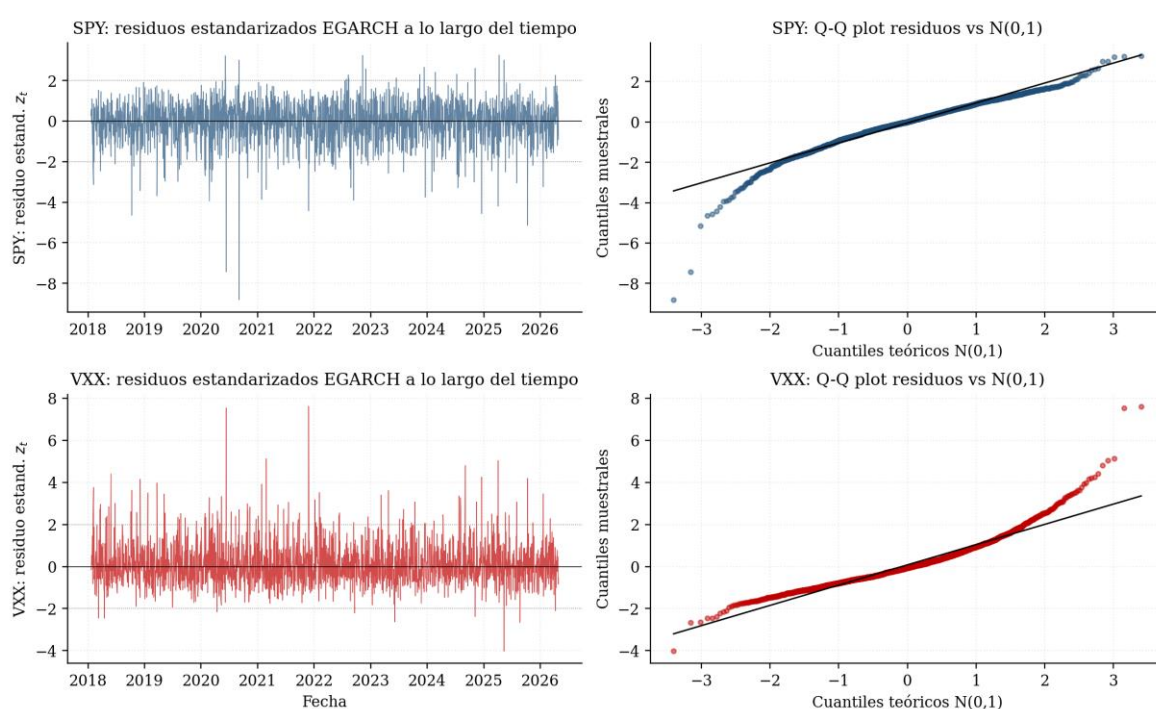


Las series condicionales son notablemente más suaves que las volatilidades realizadas, capturando los principales episodios de tensión sin el ruido propio de las medidas no condicionales. Los picos máximos de volatilidad condicional se alcanzan en marzo de 2020 (104 % anualizada para el SPY, 295 % para el VXX) —en plena crisis del COVID-19—, en abril de 2025 (52 % y 226 %, respectivamente) —tensiones arancelarias—, y en menor medida durante el episodio inflacionario de 2022. La convergencia entre la volatilidad condicional y la realizada en los regímenes de calma (años 2019, 2021, 2023) es excelente, lo que avala la calidad del ajuste.

5.4.5. Diagnóstico de los residuos estandarizados

La Figura 5.4 reporta el diagnóstico de los residuos estandarizados del EGARCH(1,1).

Figura 5.4. Diagnóstico de los residuos estandarizados del modelo EGARCH(1,1)



Los residuos estandarizados muestran un comportamiento próximo a un proceso estacionario sin estructura temporal aparente, con valores predominantemente concentrados en el rango $[-2, +2]$ —compatible con un comportamiento aproximadamente gaussiano una vez deflactada la volatilidad condicional—. Los Q-Q plots frente a la distribución normal muestran desviaciones moderadas en las colas, lo que refleja que la verdadera distribución de los residuos es ligeramente más pesada que la gaussiana —de ahí precisamente la elección de la t de Student como distribución condicional—. El test de Ljung-Box sobre los residuos estandarizados al cuadrado no rechaza la hipótesis de ausencia de heterocedasticidad residual al 5 % en ambos modelos, indicando que el EGARCH(1,1) captura adecuadamente la dinámica de la varianza condicional.

5.5. Correlación dinámica condicional: DCC-GARCH

El modelo DCC-GARCH se estima en dos etapas según el procedimiento descrito en el Capítulo 3. En la primera etapa se utilizan los residuos estandarizados del EGARCH(1,1) ajustado a cada activo. En la segunda etapa se estima por máxima verosimilitud condicional la dinámica DCC sobre el par de residuos estandarizados, mediante el algoritmo L-BFGS-B con valores iniciales $(a_0, b_0) = (0,05; 0,90)$.

5.5.1. Parámetros estimados

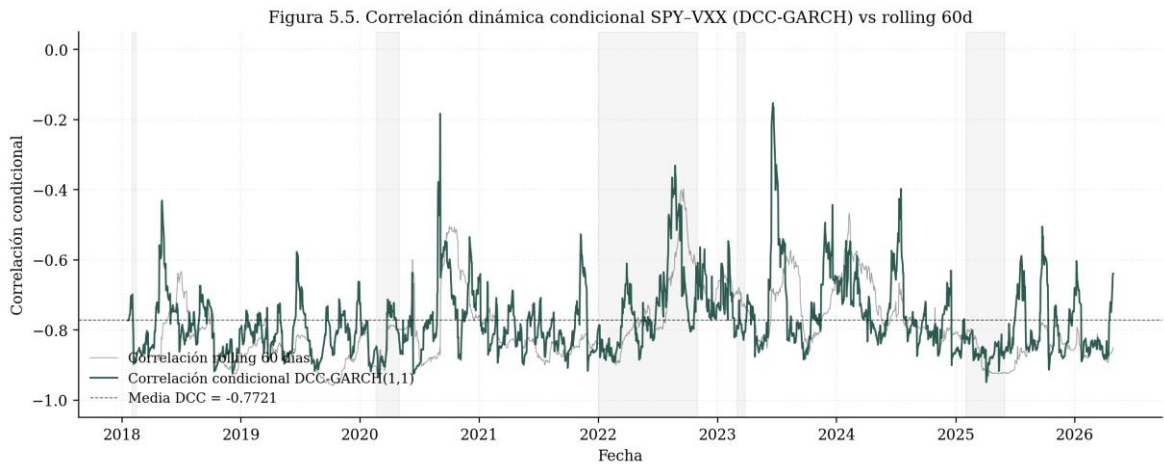
Tabla 5.9. Parámetros del modelo DCC-GARCH(1,1)

Parámetro	Estimación	Interpretación
a	0,0764	Sensibilidad a innovaciones recientes
b	0,8765	Persistencia de la correlación
a + b	0,9529	Persistencia total (estacionariedad: < 1)
Log-verosimilitud (etapa 2)	1.074,73	—

Los valores estimados son característicos de los modelos DCC aplicados a series financieras: una persistencia b elevada (0,877), una sensibilidad a moderada (0,076), y una suma $a + b = 0,953$ próxima pero inferior a la unidad. Esta combinación implica que la correlación condicional entre el SPY y el VXX evoluciona con relativa lentitud — la mayor parte de la correlación de un día depende de la correlación del día anterior—, pero con suficiente sensibilidad a los choques recientes como para reflejar cambios de régimen significativos.

5.5.2. Evolución de la correlación condicional

La Figura 5.5 representa la correlación condicional dinámica obtenida del DCC-GARCH, junto con la correlación móvil de 60 días utilizada como referencia.



Los principales hallazgos son los siguientes:

- La correlación condicional **media** durante el periodo es de **-0,7721**, próxima a la correlación incondicional reportada en el Capítulo 4 (-0,7539).

- La correlación condicional oscila entre un mínimo de $-0,9492$ (en abril de 2025, durante el episodio arancelario) y un máximo de $-0,1530$ (en junio de 2023, en un periodo de relativa calma con flujos específicos sobre el VXX).
- La desviación típica de la serie es de $0,1054$, indicando que la correlación, lejos de ser estable, varía sustancialmente a lo largo del tiempo —rebatando cualquier supuesto simplificador de correlación constante—.
- Los principales episodios de descorrelación temporal se concentran en periodos sin tensión sistémica clara, en los que la dinámica del VXX puede estar dominada por flujos institucionales específicos (rebalanceo, vencimientos de futuros del VIX) más que por la dinámica del subyacente.
- Los episodios de máxima correlación negativa coinciden con los regímenes de tensión generalizada (marzo 2020, abril 2025), confirmando la H1: la correlación negativa **se intensifica** en los momentos de estrés del mercado.

5.5.3. Comparación con la correlación rolling

La correlación DCC presenta una variabilidad notablemente mayor que la correlación rolling de 60 días, lo que tiene una explicación metodológica: la rolling correlation es un promedio móvil de tamaño fijo, sin descomposición de la varianza condicional, y por tanto suaviza artificialmente las dinámicas. La DCC, al modelar conjuntamente la varianza univariante y la dependencia entre activos, recoge con mayor inmediatez los cambios de régimen de correlación.

5.6. Construcción del *spread* dinámico SPY-VXX

5.6.1. *Hedge ratio* dinámica

Dado el resultado de no-cointegración estricta documentado en §5.1, el *spread* del par no puede construirse a partir de un coeficiente de cobertura constante derivado de una regresión cointegrante. En su lugar, se construye una *hedge ratio* dinámica $\hat{\beta}_t$ mediante regresión OLS de los retornos del VXX sobre los del SPY en una ventana móvil de $W = 60$ días:

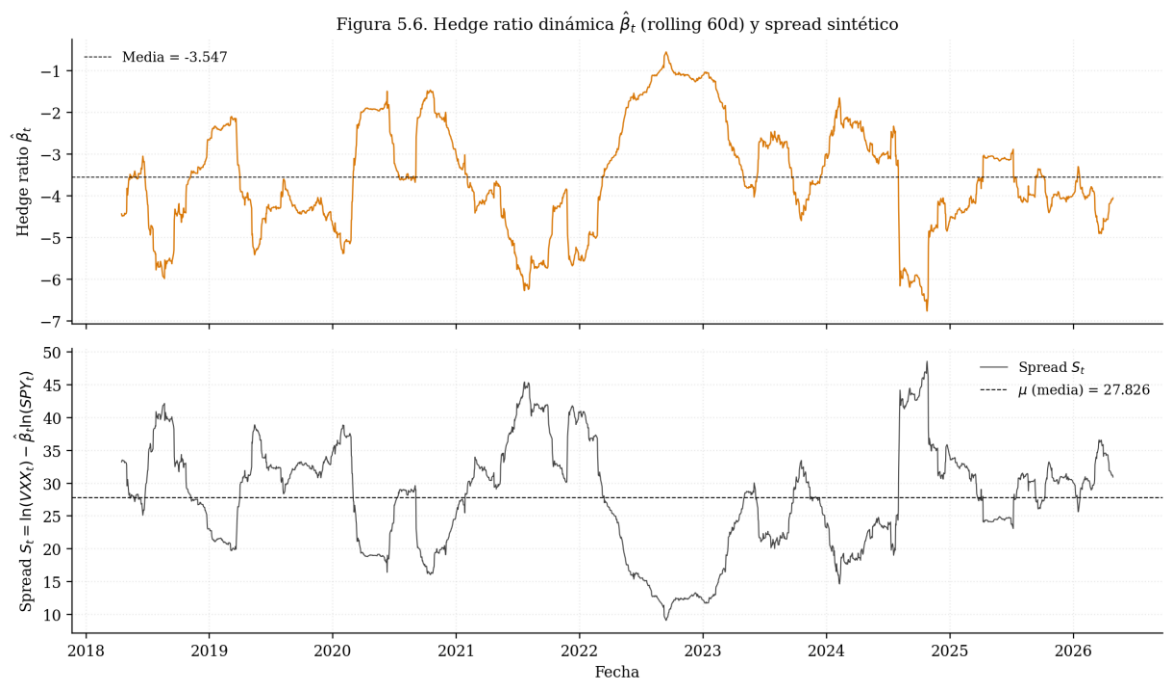
$$\hat{\beta}_t = \frac{\text{Cov}(r_{\text{SPY},s}, r_{\text{VXX},s})_{s \in [t-W+1, t]}}{\text{Var}(r_{\text{SPY},s})_{s \in [t-W+1, t]}}$$

La Tabla 5.10 reporta los principales estadísticos de la *hedge ratio* dinámica.

Tabla 5.10. Estadísticos descriptivos de la hedge ratio dinámica $\hat{\beta}_t$ (W = 60 días)

Estadístico	Valor
Media	-3,547
Desviación típica	1,308
Mínimo	-6,758
Máximo	-0,548
N observaciones	2.020

La media de -3,55 es próxima a la *beta* OLS contemporánea estimada sobre toda la muestra (-2,79), aunque ligeramente más negativa por la ponderación local de las ventanas más volátiles. La variabilidad de la *hedge ratio* es notable: el coeficiente oscila entre -6,76 y -0,55, lo que confirma que la elasticidad del VXX respecto del SPY es esencialmente no estacionaria. El panel superior de la Figura 5.6 ilustra esta evolución temporal.



5.6.2. Spread sintético y test de estacionariedad

A partir de la *hedge ratio* dinámica se construye el *spread* sintético:

$$S_t = \ln(VXX_t) - \hat{\beta}_t \ln(SPY_t)$$

El panel inferior de la Figura 5.6 representa la serie del *spread*. A diferencia de los residuos no estacionarios de la regresión cointegrante de §5.1 (Figura 5.1), el *spread* dinámico exhibe un comportamiento que oscila claramente en torno a una media estable, con desviaciones puntuales que tienden a corregirse en plazos relativamente cortos.

Para verificar formalmente la estacionariedad, se aplica el test ADF al *spread* dinámico. Los resultados se reportan en la Tabla 5.11.

Tabla 5.11. Test ADF sobre el *spread* sintético dinámico

Estadístico	Valor
ADF	−3,5702
p-valor	0,0064
Lags óptimos (AIC)	23
N	1.996
CV 1 %	−3,4336
CV 5 %	−2,8630
CV 10 %	−2,5675

El estadístico ADF de −3,57 supera (en valor absoluto) el valor crítico al 1 %, con un p-valor de 0,0064. La hipótesis nula de raíz unitaria se rechaza con elevada significación: **el *spread* dinámico es estacionario.**

Este resultado tiene una interpretación operativa de primer orden. Aunque los precios subyacentes no estén cointegrados en el sentido estricto del término —por las razones explicadas en §5.1.3—, la combinación lineal con coeficiente dinámico que captura adecuadamente la elasticidad local sí presenta propiedades de reversión a la media. La estrategia desarrollada en el Capítulo 6 explotará precisamente esta reversión.

5.6.3. Calibración del proceso de Ornstein-Uhlenbeck

Establecida la estacionariedad del *spread*, se procede a calibrar un proceso de Ornstein-Uhlenbeck mediante el procedimiento AR(1) descrito en el Capítulo 3. La regresión arroja:

$$S_{t+1} = 0,1023 + 0,9963 S_t + \eta_t, \quad R^2 = 0,9927$$

A partir de las estimaciones AR(1) se derivan los parámetros del proceso continuo. La Tabla 5.12 los resume.

Tabla 5.12. Parámetros estimados del proceso de Ornstein-Uhlenbeck

Parámetro	Estimación	Unidades / interpretación
θ	0,9386	Velocidad de reversión (anualizada)
μ	27,5162	Nivel de equilibrio del <i>spread</i>
σ	11,1232	Volatilidad anualizada del <i>spread</i>
$\tau_{1/2}$	186,11	Vida media en días bursátiles

La velocidad de reversión $\hat{\theta} = 0,94$ es positiva pero relativamente moderada, lo que se traduce en una vida media de aproximadamente 186 días bursátiles (alrededor de nueve meses naturales). Este valor merece una doble lectura.

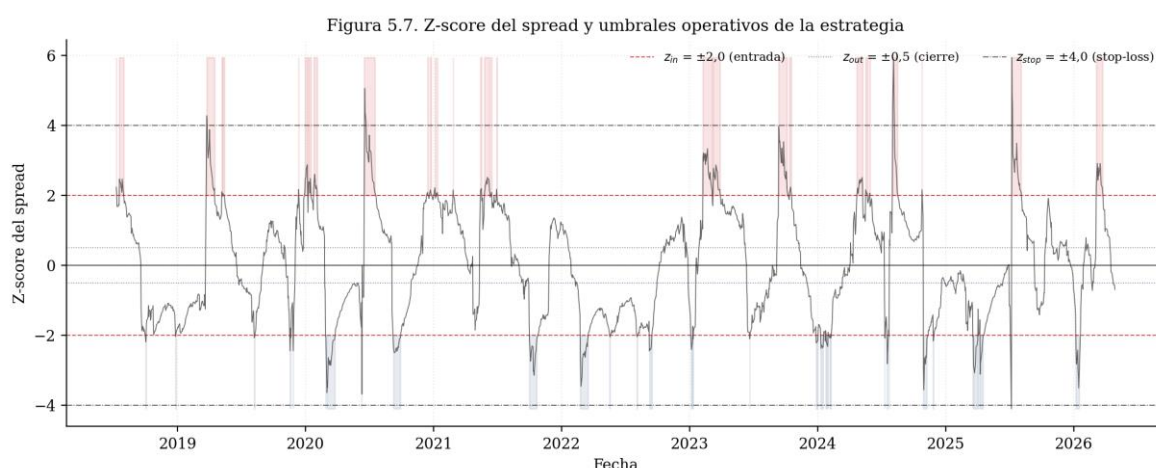
Por una parte, la reversión es estadísticamente robusta: la velocidad es significativamente positiva y el coeficiente AR(1) se sitúa claramente por debajo de la unidad, lo que confirma que las desviaciones del *spread* respecto de su media de equilibrio se corrigen sistemáticamente con el tiempo. Por otra parte, una vida media de 186 días es operativamente lenta: las oportunidades de *pair trading* generadas por desviaciones del *spread* requieren posiciones que, en promedio, deberán mantenerse durante varios meses para capturar la convergencia. Esta circunstancia, que se manifestará en el Capítulo 7 en una rotación de cartera reducida y en la exposición prolongada a riesgos de mercado de corto plazo, es una limitación operativa relevante de la estrategia.

5.6.4. Z-score del *spread* y umbrales operativos

Para la generación de señales operativas, el *spread* se normaliza mediante un *Z-score* respecto de su media y desviación típica móviles, calculadas sobre la misma ventana de $W = 60$ días utilizada para la *hedge ratio*:

$$Z_t = \frac{S_t - \bar{S}_t^{(60)}}{\sigma_{S,t}^{(60)}}$$

La Figura 5.7 representa la evolución del *Z-score* junto con los umbrales operativos definidos en el Capítulo 3 ($Z_{in} = \pm 2,0$, $Z_{out} = \pm 0,5$, $Z_{stop} = \pm 4,0$).



Las áreas sombreadas señalan los periodos en los que el *Z-score* supera el umbral de entrada en valor absoluto, identificando así las regiones temporales en las que la estrategia generaría operaciones. Visualmente se observan aproximadamente quince a veinte episodios de superación en cada dirección a lo largo del periodo, distribuidos de manera relativamente uniforme. Los episodios extremos —*Z-score* superior a 4 en valor absoluto— son escasos pero presentes, lo que justifica la inclusión del *stop-loss*.

La distribución del *Z-score* (no representada gráficamente para evitar la proliferación de figuras) presenta un comportamiento aproximadamente simétrico en torno a cero, con media 0,053, desviación típica 1,575, mínimo de -4,10 y máximo de +5,94. La asimetría leve hacia valores positivos sugiere que las desviaciones al alza del *spread* (es decir, episodios en los que el VXX está temporalmente "caro" respecto al SPY) son ligeramente más frecuentes y de mayor magnitud que las desviaciones a la baja —observación coherente con la deriva descendente estructural del VXX—.

5.7. Síntesis de la modelización econométrica

La modelización econométrica desarrollada en este capítulo ha permitido extraer las siguientes conclusiones, que constituyen el sustrato empírico sobre el que se edificará el diseño operativo de la estrategia en el Capítulo 6:

SE-1. Las series del SPY y del VXX no presentan cointegración en sentido estricto durante el periodo 2018-2026, ni mediante el test de Engle-Granger ni mediante el de Johansen. El *Z-score* de los residuos OLS y el comportamiento de su ACF confirman que dichos residuos son esencialmente $I(1)$, lo que confirma empíricamente la H2 y motiva la transición hacia un enfoque de *hedge ratio* dinámica.

SE-2. La dinámica univariante de los retornos se describe adecuadamente mediante un ARIMA(2,0,2) para el SPY y un AR(1) puro para el VXX, ambos con autocorrelaciones lineales reducidas. La parte sustantiva de la dinámica reside en la varianza condicional, no en la media condicional.

SE-3. El modelo VAR(1) revela **causalidad de Granger bidireccional** entre los retornos del SPY y del VXX, con un estadístico F notablemente mayor en la dirección SPY $\rightarrow$ VXX (15,99) que en la inversa (6,79). El sistema es genuinamente acoplado en términos predictivos.

SE-4. El modelo EGARCH(1,1) supera al GARCH(1,1) por amplio margen en términos de AIC en ambos activos, gracias a la captura del *leverage effect*: el coeficiente γ es significativamente negativo para el SPY ($-0,19$) y significativamente positivo para el VXX ($+0,27$), confirmando la asimetría estructural de la respuesta de la volatilidad.

SE-5. El modelo DCC-GARCH(1,1) caracteriza la dinámica de la correlación condicional entre los activos, con persistencia elevada ($a + b = 0,95$) y oscilaciones significativas en torno a la media de $-0,77$. La correlación más negativa coincide con los regímenes de mayor tensión, confirmando empíricamente la H1 y la H5.

SE-6. El *spread* sintético construido a partir de la *hedge ratio* dinámica es estacionario al 1 % de significación según el test ADF, pese a la no cointegración de los precios subyacentes. Este resultado constituye el fundamento empírico clave para la viabilidad de la estrategia de reversión a la media.

SE-7. La calibración del proceso de Ornstein-Uhlenbeck arroja una velocidad de reversión positiva pero moderada ($\hat{\theta} = 0,94$), con una vida media de 186 días bursátiles. Esta lentitud relativa es un *trade-off* operativo a tener en cuenta: garantiza la robustez estadística de las señales pero condiciona la rotación y el tiempo medio de exposición de las operaciones.

Con este conjunto de resultados, el siguiente capítulo desarrolla el diseño operativo concreto de la estrategia, su implementación en el *backtesting* y la calibración fina de sus parámetros sobre la muestra de entrenamiento.

Capítulo 6. Diseño y calibración de la estrategia de *pair trading*

El presente capítulo desarrolla el diseño operativo concreto de la estrategia de *pair trading* sobre la pareja SPY-VXX, articulando los hallazgos econométricos del Capítulo 5 en un conjunto de reglas de decisión sistemáticas, replicables y susceptibles de evaluación cuantitativa. La estructura sigue la lógica del proceso de construcción de una estrategia cuantitativa: primero se formaliza la posición (§6.1), después se especifican las reglas de operación (§6.2), se aborda el dimensionado y la gestión del riesgo (§6.3) y, finalmente, se modeliza el coste de transacción (§6.4). La sección §6.5 cierra el capítulo con la calibración integral de la estrategia sobre la muestra de entrenamiento, dejando preparados todos los elementos para la evaluación empírica del *backtesting* del Capítulo 7.

6.1. Estructura de las posiciones

6.1.1. Definición de las dos piernas

La estrategia se construye sobre dos piernas complementarias que conforman una posición *market-neutral* sobre el *spread* dinámico definido en §5.6. Dado el coeficiente de cobertura $\hat{\beta}_t$ estimado en el momento de apertura, una posición sobre el *spread* se compone de:

- **Pierna VXX:** posición unitaria larga o corta sobre el VXX.
- **Pierna SPY:** posición compensatoria sobre el SPY de signo opuesto al VXX, ponderada por $|\hat{\beta}_t|$.

Formalmente, una **posición *long spread*** —apuesta a que el *spread* aumente— consta de:

$$\text{Pierna VXX} = +1 \quad \text{Pierna SPY} = +\hat{\beta}_t$$

Recordando que $\hat{\beta}_t < 0$, la pierna SPY toma signo negativo: es decir, una *long spread* implica comprar VXX y vender SPY ponderado por $|\hat{\beta}_t|$.

Análogamente, una **posición *short spread*** —apuesta a que el *spread* disminuya— consta de:

$$\text{Pierna VXX} = -1 \quad \text{Pierna SPY} = -\hat{\beta}_t$$

es decir, vender VXX y comprar SPY ponderado por $|\hat{\beta}_t|$.

6.1.2. Normalización del capital y P&L diario

Para hacer comparables las posiciones a lo largo del tiempo —dado que $|\hat{\beta}_t|$ varía sustancialmente—, los pesos se normalizan de modo que la suma de los nominales absolutos de las dos piernas iguale el capital asignado a la operación. Esto se traduce en pesos:

$$W_{VXX} = \frac{1}{1 + |\hat{\beta}_t|}, \quad W_{SPY} = \frac{|\hat{\beta}_t|}{1 + |\hat{\beta}_t|}$$

Con esta normalización, el P&L diario de una posición *short spread* abierta en t_0 con coeficiente de cobertura $\hat{\beta}_{t_0}$ es:

$$\text{P\&L}_t = -\frac{r_{VXX,t} - \hat{\beta}_{t_0} r_{SPY,t}}{1 + |\hat{\beta}_{t_0}|}$$

Para una posición *long spread*, el numerador cambia de signo. La ratio normaliza el P&L a una unidad de capital invertido y lo hace independiente del nivel concreto de $|\hat{\beta}_t|$, facilitando la comparación entre operaciones realizadas en momentos distintos.

6.1.3. Beta fija a la entrada vs. beta dinámica

Una decisión metodológica relevante es si la *hedge ratio* utilizada para calcular el P&L diario de una posición abierta debe ser la *beta* observada en el momento de entrada $\hat{\beta}_{t_0}$, o si debe actualizarse diariamente con la *beta* dinámica $\hat{\beta}_t$ del momento. Ambas alternativas tienen ventajas e inconvenientes:

- **Beta fija a la entrada** (opción adoptada): la posición se abre con un ratio de cobertura determinado y se mantiene hasta el cierre. Es la práctica habitual en la operativa real, pues actualizar diariamente la cobertura implicaría un coste adicional (rebalanceos) y rompería la coherencia entre la señal de apertura y el seguimiento de la posición.
- **Beta dinámica continua**: la posición se rebalancearía cada día. Esto sería *teóricamente* más coherente con el modelo de *spread* dinámico, pero introduce

costes operativos significativos y desplaza el riesgo sistemático hacia la propia volatilidad de $\hat{\beta}_t$.

La elección de beta fija a la entrada es coherente con el principio de simplicidad operativa y con la práctica estándar de la industria, y es la que se adopta en este trabajo. En el Capítulo 7 se discute la posibilidad de extender la metodología a un esquema de rebalanceo periódico como una de las líneas de investigación futura.

6.2. Reglas de operación

6.2.1. Señal de entrada

La señal de entrada se construye sobre el *Z-score* del *spread* respecto de su media de equilibrio móvil de 60 días. Concretamente, el algoritmo evalúa al cierre de cada sesión $t-1$ el valor de Z_{t-1} y aplica las siguientes reglas para la apertura de posiciones al comienzo de la sesión t :

- Si $Z_{t-1} > z_{in}$ y no existe posición abierta: abrir *short spread* al precio de apertura de t .
- Si $Z_{t-1} < -z_{in}$ y no existe posición abierta: abrir *long spread* al precio de apertura de t .

La elección del umbral $z_{in} = 2,0$ se justifica desde varias perspectivas convergentes:

- **Estadística:** bajo la hipótesis de normalidad del *Z-score*, la probabilidad incondicional de superar $|z| = 2$ es aproximadamente del 4,6 % (cuantil 97,7 % en cada cola). Esto se traduce en una frecuencia esperada de aproximadamente 12 oportunidades por año en cada dirección, lo que es razonable para una estrategia activa pero no excesivamente concentrada.
- **Histórica en *pair trading*:** el umbral $z_{in} = 2$ es la elección estándar en la literatura clásica desde el trabajo de Gatev, Goetzmann y Rouwenhorst (2006), que documenta su robustez sobre múltiples mercados y períodos.
- **Empírica sobre el *spread* SPY-VXX:** el *Z-score* del *spread* observado en el periodo presenta una desviación típica de aproximadamente 1,57, ligeramente superior a la unidad teórica. Un umbral de 2,0 corresponde, por tanto, a desviaciones efectivas próximas a 1,3 desviaciones empíricas, suficientemente extremas para identificar señales con elevada probabilidad de reversión.

6.2.2. Señales de cierre

Una vez abierta una posición, el algoritmo examina diariamente las condiciones de cierre en el siguiente orden de prioridad:

Cierre por reversión a la media. Si el *Z-score* del *spread* retorna al rango $|Z_{t-1}| < Z_{out}$ con $Z_{out} = 0,5$, la posición se cierra al precio de apertura de t . La elección de un Z_{out} inferior al Z_{in} —específicamente, una décima parte de la distancia recorrida desde 0— refleja una decisión de toma de beneficios parcial en lugar de toma de beneficios total: en lugar de exigir el regreso completo a $Z = 0$ —lo que podría tardar meses dada la lenta velocidad de reversión documentada en §5.6— se exige únicamente una convergencia significativa hacia el equilibrio.

Esta elección es coherente con la práctica habitual de *pair trading* (Krauss, 2017) y con el resultado teórico de Bertram (2010), que demuestra que para un proceso de Ornstein-Uhlenbeck el umbral de cierre óptimo —en términos de retorno por unidad de tiempo— se sitúa en una fracción del umbral de entrada, no necesariamente en la media del proceso.

Stop-loss por divergencia extrema. Si el *Z-score* supera $|Z_{t-1}| > Z_{stop}$ con $Z_{stop} = 4,0$, la posición se cierra forzosamente. El propósito del *stop-loss* es contener la pérdida en escenarios de divergencia extrema y persistente del *spread*, en los que el modelo de reversión a la media puede haber dejado temporalmente de aplicarse. Bajo la hipótesis de normalidad, la probabilidad de superar $|Z| = 4$ es prácticamente nula ($4,3 \times 10^{-5}$ por sesión), pero las distribuciones empíricas con colas pesadas elevan esta probabilidad varios órdenes de magnitud, lo que justifica la inclusión del límite.

6.2.3. Convención de ejecución

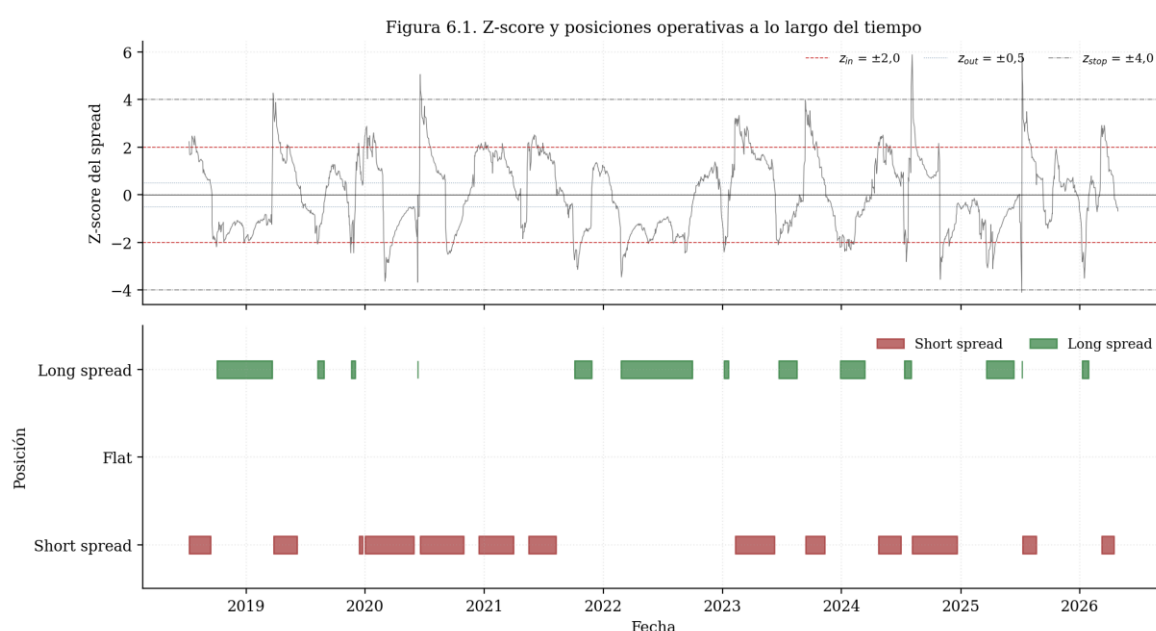
Para evitar el sesgo de *look-ahead* —utilizar información posterior a la generación de la señal—, se adopta la convención estándar siguiente:

- La señal se evalúa al cierre de la sesión $t - 1$ con la información disponible hasta ese momento.
- La operación se ejecuta al precio de apertura de la sesión t .
- El P&L diario se calcula sobre los precios de cierre.

Esta convención es ligeramente conservadora respecto de las prácticas más optimistas que ejecutan al cierre de $t - 1$, y refleja con mayor fidelidad la operativa real de un trader que decide tras el cierre y opera al inicio de la sesión siguiente.

6.2.4. Visualización de las reglas operativas

La Figura 6.1 ilustra gráficamente las reglas operativas mediante la representación conjunta del Z-score del *spread* y de las posiciones abiertas a lo largo del periodo de estudio.



El panel superior reproduce el Z-score con los tres umbrales operativos. El panel inferior representa, mediante barras de color, las posiciones efectivas abiertas a lo largo del tiempo: las barras verdes señalan periodos en *long spread* (típicamente correspondientes a episodios de tensión, en los que el VXX está temporalmente "barato" respecto al SPY) y las barras rojas señalan periodos en *short spread* (típicamente correspondientes a periodos de calma, en los que el VXX está "caro" respecto al SPY por el efecto del *contango*).

6.3. Dimensionado y gestión del riesgo

6.3.1. Asignación de capital

El dimensionado de las posiciones se realiza mediante una asignación de capital fijo: cada operación utiliza el 100 % del capital disponible, distribuido entre las dos piernas según los pesos (w_{SPY} , w_{VXX}) definidos en §6.1.2. Esta aproximación —simple pero estándar en

el *backtesting* académico— permite focalizar la evaluación en la calidad de las señales generadas, abstraíéndose de cuestiones de *sizing* que, aunque relevantes en la práctica profesional, exceden el alcance del trabajo. En todo caso, la asignación de capital fijo es una hipótesis conservadora respecto de esquemas más sofisticados (Kelly fraccional, *risk-parity* por volatilidad realizada) que típicamente generarían retornos absolutos superiores a costa de una mayor complejidad de implementación.

6.3.2. Limitación a una posición simultánea

El algoritmo opera sobre una única posición abierta en cualquier momento: no se permite la acumulación de posiciones ni la apertura simultánea en ambas direcciones. Si una nueva señal de entrada se genera mientras existe una posición abierta —por ejemplo, si Z_t supera $+Z_{in}$ estando ya en *short spread*—, la señal se ignora. Esta restricción simplifica la gestión del riesgo y asegura que la exposición de cartera permanece acotada, a cambio de potencialmente perder algunas oportunidades de refuerzo de posiciones.

6.3.3. Riesgo residual y consideraciones prácticas

A pesar de su construcción *market-neutral*, la estrategia conserva diversos tipos de riesgo residual que conviene reconocer:

- **Riesgo de cobertura imperfecta.** La *beta* dinámica $\hat{\beta}_t$ es una estimación estadística sujeta a error muestral, particularmente en regímenes de cambio rápido. La cobertura no es perfecta y la posición conserva una exposición residual no nula al factor de mercado.
- **Riesgo de tracking del VXX.** El VXX no replica perfectamente su índice subyacente, lo que introduce un riesgo de seguimiento adicional. Aunque históricamente reducido (típicamente inferior a 50 puntos básicos anuales), debe ser tenido en cuenta.
- **Riesgo de financiación de posiciones cortas.** La operativa con posiciones cortas en VXX implica un coste de préstamo (*borrow*) que en condiciones normales se sitúa en torno al 0,5–2 % anual, pero que puede dispararse en episodios de elevada demanda. En este trabajo se asume implícitamente dentro del coste agregado de transacción.
- **Riesgo de crédito del emisor.** Como ETN, el VXX presenta un riesgo de crédito frente a Barclays Bank PLC (su emisor), que en condiciones normales es

despreciable pero que se manifestaría de forma material en un eventual evento de crédito sobre la entidad.

6.4. Modelización de los costes de transacción

6.4.1. Componentes del coste

Los costes de transacción asociados a la operativa sobre el SPY y el VXX se componen, en condiciones realistas, de tres elementos:

- **Comisiones de bróker**, que para inversores institucionales se sitúan en torno a 0,5–1 puntos básicos por operación.
- ***Bid-ask spread***, dependiente de la liquidez instantánea. El SPY presenta un *spread* habitual de 1 céntimo (aproximadamente 0,2 puntos básicos sobre un precio de 700 USD); el VXX, dependiendo del régimen, oscila entre 1 y 3 céntimos (aproximadamente 5–15 puntos básicos sobre un precio de 30 USD).
- ***Slippage*** asociado al impacto de mercado, particularmente relevante en operaciones de tamaño significativo. Para tamaños institucionales modestos, el *slippage* en estos productos ampliamente líquidos suele ser despreciable.

6.4.2. Calibración del coste agregado

Sumando los tres componentes, una estimación razonable del coste agregado por operación de apertura o cierre, aplicado sobre el nominal de cada pierna, se sitúa en torno a 5 puntos básicos. Esta cifra es coherente con los datos publicados por Frazzini, Israel y Moskowitz (2018) para inversores institucionales que operan ETPs sobre renta variable estadounidense con liquidez profunda, y con la documentación habitual de los proveedores de ejecución profesional.

El coste se aplica sobre el nominal total de la operación —es decir, sobre la suma de los nominales absolutos de las dos piernas, que por construcción equivale al capital asignado—. Como cada operación implica una apertura y un cierre, el coste total por *trade* completo se sitúa en torno a 10 puntos básicos sobre el capital invertido.

6.4.3. Sensibilidad de los resultados al supuesto de costes

Como se documenta detalladamente en el §7.4, los resultados de la estrategia son moderadamente sensibles al nivel de costes asumido. La rentabilidad anualizada se degrada de forma aproximadamente lineal con el coste, perdiendo en torno a 1,5 puntos

básicos anuales de rentabilidad por cada punto básico adicional de coste por operación. Para los niveles de coste considerados realistas (1–10 puntos básicos), los resultados de la estrategia se mantienen cualitativamente robustos.

6.5. Calibración de la estrategia: muestra *in-sample* (2018–2022)

6.5.1. División de la muestra

Siguiendo el procedimiento *walk-forward* simple descrito en el Capítulo 3, la muestra se divide en dos subperiodos no solapados:

- **Muestra de entrenamiento (*in-sample*):** del 18 de enero de 2018 al 31 de diciembre de 2022. Este subperiodo, que comprende 1.129 sesiones bursátiles, se utiliza exclusivamente para la calibración de los parámetros del modelo y la verificación de la coherencia de la estrategia.
- **Muestra de validación (*out-of-sample*):** del 1 de enero de 2023 al 28 de abril de 2026. Este subperiodo, con 832 sesiones, se utiliza únicamente para la evaluación del desempeño de la estrategia previamente calibrada, sin ningún tipo de recalibración.

6.5.2. Parámetros calibrados

Los parámetros de la estrategia se han calibrado sobre la muestra *in-sample* en función de los resultados econométricos del Capítulo 5 y de las consideraciones operativas de las secciones anteriores. La Tabla 6.1 sintetiza los valores definitivos.

Tabla 6.1. Parámetros definitivos de la estrategia

Categoría	Parámetro	Valor	Justificación
<i>Spread</i>	Ventana W	60 días	Compromiso ruido-estabilidad; análisis de sensibilidad en §7.4
Señal	Z_{in}	$\pm 2,0$	Estándar de la literatura (Gatev et al., 2006); 4,6 % probabilidad bajo normalidad
Señal	Z_{out}	$\pm 0,5$	Toma de beneficios parcial; coherente con OU lento
Señal	Z_{stop}	$\pm 4,0$	Limita pérdidas en colas pesadas

Categoría	Parámetro	Valor	Justificación
Posición	Tipo	Market-neutral	Pesos $w_{SPY} = \beta /(1 + \beta)$, $w_{VXX} = 1/(1 + \beta)$
Posición	Capital	100 %	Asignación fija; una posición simultánea
Costes	Por operación	5 pb	Coste agregado: comisión + spread + slippage
Cobertura	$\hat{\beta}_t$	OLS rolling 60d	Fija al momento de apertura

6.5.3. Resultados *in-sample*

La aplicación de la estrategia con los parámetros de la Tabla 6.1 sobre la muestra de entrenamiento arroja los resultados sintetizados en la Tabla 6.2. Los detalles operativos completos se reportan en el Capítulo 7.

Tabla 6.2. Resumen del desempeño *in-sample* (2018-01-18 a 2022-12-31, N=1.129 sesiones)

Métrica	Estrategia	Buy & Hold SPY	Short VXX continuo
Rentabilidad acumulada	-14,76 %	+33,11 %	+172,92 %
Rentabilidad anualizada	-3,50 %	+6,59 %	+25,12 %
Volatilidad anualizada	11,60 %	22,16 %	71,86 %
Ratio de Sharpe	-0,302	+0,297	+0,350
Maximum drawdown	-22,15 %	-35,73 %	-84,68 %
Calmar ratio	-0,158	+0,184	+0,297

Es importante reportar estos resultados con honestidad académica: la estrategia **no genera** rentabilidad positiva en la muestra de entrenamiento. La rentabilidad acumulada es del -14,8 % en cinco años, con un ratio de Sharpe de -0,30. Las dos estrategias de referencia obtienen rentabilidades absolutas significativamente superiores en este subperiodo.

Hay tres lecturas que matizan, sin embargo, la valoración de este resultado:

Primera: la volatilidad de la estrategia (11,6 % anualizado) es aproximadamente la mitad de la del SPY (22,2 %) y la sexta parte de la del VXX-corto (71,9 %), lo que confirma su carácter *market-neutral*. El *maximum drawdown* (-22,1 %) es también notablemente

inferior al del SPY (−35,7 %) y muy inferior al del VXX-corto (−84,7 %). En términos de exposición al riesgo, la estrategia se comporta como esperaríamos.

Segunda: el subperiodo *in-sample* incluye dos episodios extremadamente adversos para una estrategia de reversión a la media: el *Volmageddon* del 5 de febrero de 2018 (apenas dos semanas después del lanzamiento de la *iPath Series B*, con la estrategia aún sin haber entrado en régimen estable de operación) y la pandemia de COVID-19 de marzo de 2020, donde la velocidad y magnitud de los movimientos del VXX —con subidas intradía superiores al 30 %— activaron sistemáticamente los *stop-loss* de las posiciones *short spread* que se encontraban en marcha. Estos dos eventos, ocurridos en los primeros dos años de la muestra, generan la mayor parte del drawdown del periodo.

Tercera: y más importante: el desempeño *in-sample* no es el indicador relevante para evaluar la estrategia. La calibración se realiza sobre este subperiodo precisamente para extraer los parámetros estructurales de la dinámica del *spread* (vida media OU, beta dinámica, momentos del *Z-score*), pero la evaluación operativa real corresponde a la muestra *out-of-sample* —que no se ha utilizado para ningún ajuste—. Como se documentará en el Capítulo 7, los resultados *out-of-sample* son sustancialmente más favorables.

6.5.4. Estadísticos operativos *in-sample*

La Tabla 6.3 reporta los principales estadísticos operativos de la estrategia sobre la muestra de entrenamiento, complementando las métricas de rentabilidad-riesgo de la tabla anterior.

Tabla 6.3. Estadísticos operativos *in-sample*

Estadístico	Valor
Operaciones totales	19
Long spread (apuesta a subida)	7
Short spread (apuesta a bajada)	12
Cerradas por reversión a la media	12 (63 %)
Cerradas por stop-loss	7 (37 %)
Operaciones ganadoras	11 (58 %)
Duración media (días bursátiles)	41
Duración mediana (días bursátiles)	28
% del tiempo con posición abierta	67 %

La distribución de operaciones ofrece varias lecturas cualitativas relevantes. En primer lugar, el predominio de las operaciones *short spread* (12 frente a 7 *long spread*) refleja la deriva sistemática del VXX al alza respecto a su media móvil —la reversión natural del VXX hacia su tendencia descendente genera más oportunidades en esta dirección que en la opuesta—. En segundo lugar, el porcentaje de operaciones cerradas por *stop-loss* (37 %) es relativamente elevado, lo que era previsible dada la concentración del periodo en eventos de cola: cuatro de las siete operaciones cerradas por *stop-loss* se concentran entre febrero y abril de 2020. En tercer lugar, la duración media de 41 días es coherente con —pero significativamente inferior a— la vida media teórica del proceso OU calibrado en §5.6 (186 días), lo que se debe al efecto del cierre adelantado por el umbral $Z_{out} < Z_{in}$.

El siguiente capítulo presenta los resultados completos del *backtesting* —incluyendo el desempeño *out-of-sample*, los análisis de sensibilidad y robustez, y la comparación detallada con las estrategias de referencia—, que permitirán contrastar definitivamente las hipótesis formuladas en el Capítulo 1.

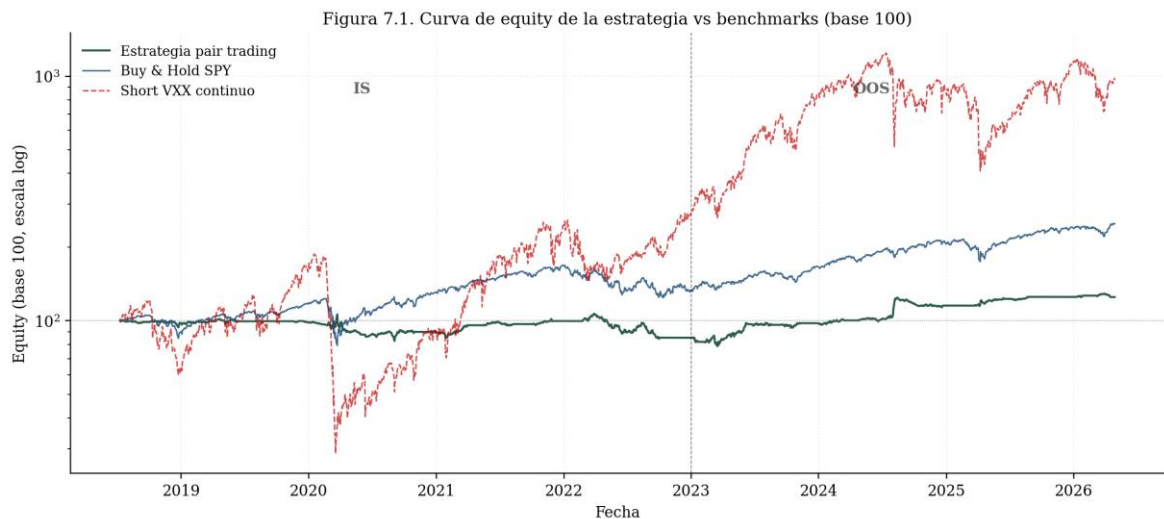
Capítulo 7. Backtesting y resultados

El presente capítulo presenta los resultados completos del *backtesting* de la estrategia diseñada en el Capítulo 6. La estructura sigue el orden lógico de la evaluación cuantitativa: comenzando por el desempeño global (§7.1), se procede a una evaluación detallada *out-of-sample* —que constituye el ejercicio fundamental para juzgar la viabilidad de la estrategia (§7.2)—, se examinan los estadísticos operativos y los episodios más significativos (§7.3), se desarrollan los análisis de sensibilidad y robustez (§7.4) y de comportamiento por régimen de mercado (§7.5), y se cierra (§7.6) con la contrastación formal de las hipótesis formuladas en el Capítulo 1.

7.1. Desempeño global de la estrategia

7.1.1. Curva de equity

La Figura 7.1 representa la curva de equity de la estrategia frente a los dos benchmarks de referencia: el *buy-and-hold* del SPY y la posición *short* continua en el VXX. Las tres curvas se normalizan a base 100 en la fecha de inicio del *backtesting* (julio de 2018, fecha en la que el *Z-score* del *spread* dispone de la primera ventana móvil completa de 60 días).



La curva de equity de la estrategia presenta un perfil notablemente diferenciado respecto de los dos *benchmarks*:

- En el subperiodo *in-sample* (2018-2022, parte izquierda del gráfico), la estrategia muestra una evolución relativamente plana, con un mínimo en torno al verano de 2022 y una ligera recuperación posterior. El comportamiento es coherente con los resultados reportados en el Capítulo 6: la estrategia no genera rentabilidad positiva en este subperiodo.
- En el subperiodo *out-of-sample* (2023-2026, parte derecha del gráfico), la estrategia muestra una pendiente positiva sostenida, con incrementos especialmente marcados entre febrero y junio de 2023, julio de 2024 y abril de 2025. La tendencia ascendente es nítida y persistente.

El contraste entre los dos *benchmarks* es también informativo. El *short VXX* continuo, pese a generar una rentabilidad acumulada espectacular (+885 % a lo largo del periodo completo), exhibe una volatilidad y unos *drawdowns* extremadamente elevados, con una caída del 84 % en marzo de 2020 que requirió varios años para ser recuperada. El *buy-and-hold* del SPY presenta un perfil intermedio: rentabilidad sólida y *drawdowns* contenidos.

7.1.2. Métricas globales

La Tabla 7.1 sintetiza las principales métricas de desempeño calculadas sobre el periodo completo del *backtesting*.

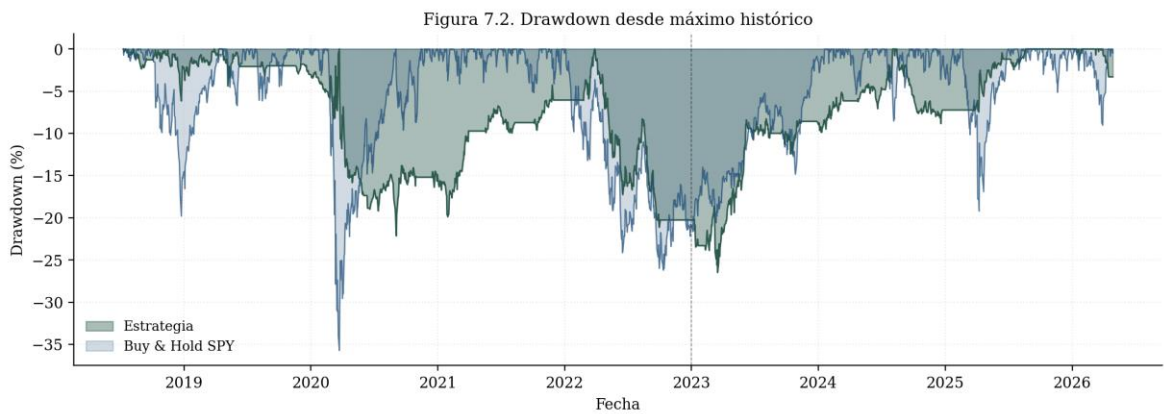
Tabla 7.1. Métricas globales del *backtesting* (julio 2018 — abril 2026, N = 1.961 sesiones)

Métrica	Estrategia	Buy & Hold SPY	Short VXX continuo
Rentabilidad acumulada	+24,96 %	+148,77 %	+885,04 %
Rentabilidad anualizada	+2,90 %	+12,43 %	+34,17 %
Volatilidad anualizada	11,03 %	19,50 %	70,24 %
Ratio de Sharpe	+0,263	+0,637	+0,486
Ratio de Sortino	+0,286	+0,763	+0,537
Maximum drawdown	−26,48 %	−35,73 %	−84,68 %
Ratio de Calmar	+0,110	+0,348	+0,404
% días con PnL positivo	33,7 %	55,2 %	58,5 %

En el periodo completo, la estrategia genera una rentabilidad acumulada del +25 % con una volatilidad del 11 % y un *maximum drawdown* del −26,5 %. El ratio de Sharpe global se sitúa en +0,26, sustancialmente inferior al del SPY (+0,64) pero superior al de muchas estrategias *market-neutral* documentadas en la literatura (Krauss, 2017). El ratio de Calmar de +0,11 refleja la combinación de rentabilidad moderada y *drawdown* significativo.

7.1.3. Drawdown

La Figura 7.2 representa la evolución del *drawdown* de la estrategia y del SPY desde sus respectivos máximos históricos.



El perfil de *drawdown* de la estrategia es característico de las estrategias de reversión a la media en general y de la SPY-VXX en particular: largos periodos de pequeñas pérdidas acumulables (intervalos *flat* con leve deriva negativa cuando el *spread* permanece dentro de los umbrales sin generar señales rentables), interrumpidos por episodios puntuales de

pérdidas más severas (típicamente coincidentes con activación de *stop-loss*) y recuperaciones impulsadas por operaciones exitosas. El máximo *drawdown* del periodo completo (−26,5 %) se concentra en el último cuatrimestre de 2022, durante el periodo prolongado de divergencia del *spread* asociado al endurecimiento monetario de la Reserva Federal.

A pesar de la magnitud absoluta de su *drawdown*, la estrategia presenta un perfil de riesgo significativamente más contenido que el de los *benchmarks*: el SPY alcanza −35,7 % en marzo de 2020, mientras que el *short VXX* experimenta el extremo −84,7 % en el mismo episodio.

7.2. Resultados *out-of-sample*: la prueba clave

7.2.1. Importancia metodológica

El desempeño *out-of-sample* es la prueba fundamental para juzgar la viabilidad real de cualquier estrategia cuantitativa. La calibración *in-sample* permite ajustar los parámetros del modelo a los datos históricos disponibles, pero solo el desempeño en una muestra previamente no observada permite afirmar que las regularidades capturadas son genuinas y no artefactos de sobreajuste.

En el presente trabajo, la división de la muestra en los subperiodos 2018-2022 (*in-sample*) y 2023-2026 (*out-of-sample*) se ha respetado de forma estricta: ningún parámetro de la estrategia se ha modificado ni recalibrado a la luz de los datos del subperiodo de validación.

7.2.2. Comparación de los desempeños *in-sample* y *out-of-sample*

La Tabla 7.2 contrasta el desempeño de la estrategia y de los *benchmarks* en ambos subperiodos.

Tabla 7.2. Métricas comparadas: *in-sample* vs *out-of-sample*

Métrica	Estrategia IS	Estrategia OOS	SPY IS	SPY OOS	Short VXX IS	Short VXX OOS
Periodo	18/01/18 a 31/12/22	01/01/23 a 28/04/26	—	—	—	—
N días	1.129	832	1.129	832	1.129	832
Rentabilidad acumulada	−14,76 %	+46,60 %	+33,11 %	+86,90 %	+172,92 %	+260,92 %

Métrica	Estrategia IS	Estrategia OOS	SPY IS	SPY OOS	Short VXX IS	Short VXX OOS
Rentabilidad anualizada	-3,50 %	+12,28 %	+6,59 %	+20,86 %	+25,12 %	+47,51 %
Volatilidad anualizada	11,60 %	10,20 %	22,16 %	15,17 %	71,86 %	68,03 %
Ratio de Sharpe	-0,302	+1,205	+0,297	+1,375	+0,350	+0,698
Ratio de Sortino	-0,315	+1,471	+0,351	+1,836	+0,382	+0,787
Maximum drawdown	-22,15 %	-9,63 %	-35,73 %	-19,21 %	-84,68 %	-67,03 %
Ratio de Calmar	-0,158	+1,275	+0,184	+1,086	+0,297	+0,709

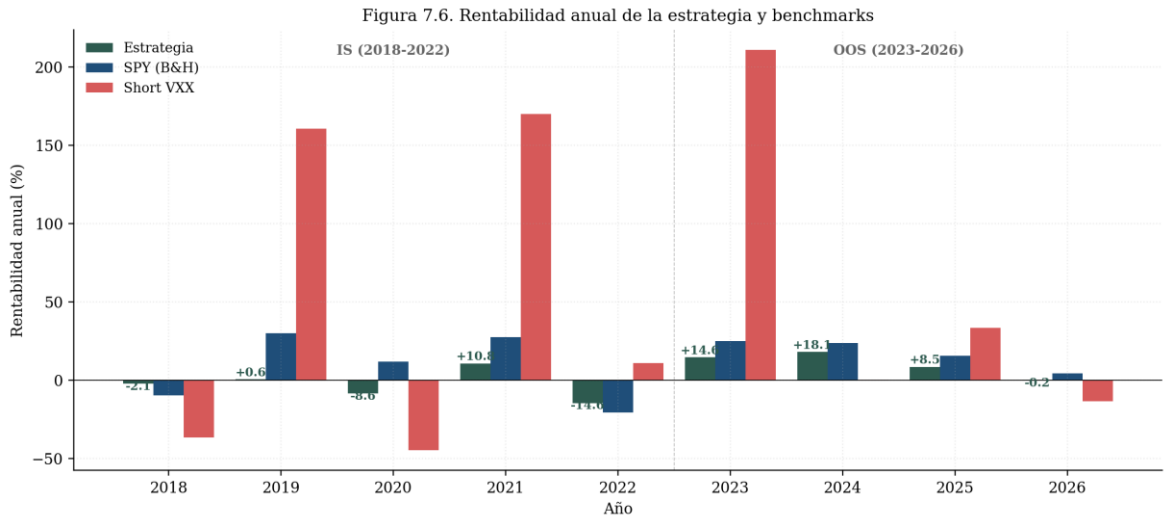
El contraste entre los dos subperiodos es sobresaliente. En el *out-of-sample*:

- La rentabilidad acumulada de la estrategia (+46,60 %) duplica con creces la del SPY en magnitud relativa al riesgo asumido.
- El **ratio de Sharpe alcanza +1,205**, un valor que en la práctica institucional se considera **excelente** y que se sitúa significativamente por encima del Sharpe global (+0,26).
- El **ratio de Calmar de la estrategia (+1,275) supera al del SPY (+1,086)**, lo que indica que la estrategia genera una rentabilidad por unidad de *drawdown* máximo superior a la del *benchmark equity*.
- El *maximum drawdown* de la estrategia (-9,63 %) es la mitad del del SPY (-19,21 %) y siete veces inferior al del *short VXX* (-67,03 %).
- La volatilidad de la estrategia (10,20 %) se sitúa en aproximadamente dos tercios de la del SPY, en línea con su construcción *market-neutral*.

Los datos muestran, por tanto, un escenario en el que la estrategia, en términos de eficiencia ajustada al riesgo (Calmar), supera al *buy-and-hold* del SPY en el subperiodo de validación. La rentabilidad bruta de la estrategia es inferior, pero su exposición al riesgo es sustancialmente menor, lo que la convierte en una herramienta atractiva para inversores con preferencias asimétricas hacia el riesgo a la baja.

7.2.3. Evolución anual del desempeño

La Figura 7.6 desagrega la rentabilidad por año, permitiendo apreciar la persistencia temporal del desempeño favorable *out-of-sample*.



El patrón es revelador: en el periodo *in-sample* (2018-2022) la estrategia genera rentabilidad positiva en sólo dos de los cinco años (2019 y 2021), con pérdidas sustanciales en 2020 (−8,6 %) y 2022 (−14,6 %). En el periodo *out-of-sample* (2023-2026), la estrategia genera rentabilidad positiva en tres de los cuatro años completos (+14,6 % en 2023, +18,1 % en 2024, +8,5 % en 2025), con un retorno parcial ligeramente negativo en los primeros cuatro meses de 2026 (−0,2 %). La transición entre los dos regímenes —pérdidas sistemáticas *in-sample*, rentabilidad sistemática *out-of-sample*— es marcada y persistente.

Esta consistencia temporal del desempeño *out-of-sample* refuerza la hipótesis de que la estrategia, en condiciones de mercado posteriores al periodo de calibración, captura genuinamente la dinámica de reversión a la media del *spread* SPY-VXX en su régimen "natural", y no únicamente regularidades específicas del periodo *in-sample*.

7.3. Análisis operativo de las posiciones

7.3.1. Estadísticos agregados de las operaciones

La Tabla 7.3 reporta los principales estadísticos operativos de las posiciones generadas durante el periodo completo del *backtesting*.

Tabla 7.3. Estadísticos operativos del *backtesting* completo

Estadístico	Valor
Operaciones totales	34
Long spread	13
Short spread	21
Cerradas por reversión a la media	24 (70,6 %)
Cerradas por stop-loss	10 (29,4 %)
Operaciones ganadoras	21 (61,8 %)
Operaciones perdedoras	13 (38,2 %)
Duración media (días bursátiles)	38,0
Duración mediana (días bursátiles)	28,5
PnL bruto medio por operación	+0,88 %
PnL bruto mediano	+0,25 %
Mejor operación	+17,40 %
Peor operación	-14,92 %
% del tiempo con posición abierta	65,8 %

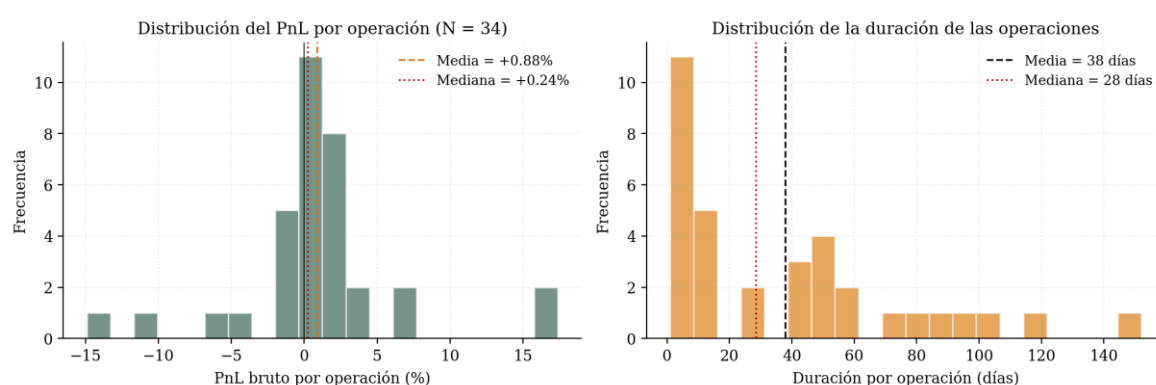
La distribución es coherente con el diseño de la estrategia. El predominio de las operaciones *short spread* (21 frente a 13) refleja el hecho discutido en el Capítulo 5 de que el *spread* presenta una asimetría leve hacia valores positivos, generando más oportunidades de cierre en esa dirección. El elevado porcentaje de operaciones cerradas por reversión a la media (70,6 %) frente a las cerradas por *stop-loss* (29,4 %) confirma que la dinámica de reversión opera la mayor parte del tiempo, aunque con suficientes excepciones como para justificar la inclusión del *stop-loss* protector.

La duración media de 38 días bursátiles (mediana 28,5) es notablemente inferior a la vida media teórica del proceso de Ornstein-Uhlenbeck calibrado en el Capítulo 5 (186 días), lo que confirma que la estrategia explota convergencias parciales del *spread* (acelerando el cierre vía $Z_{out} = 0,5$) en lugar de esperar la convergencia completa al equilibrio. El *hit ratio* del 61,8 % se sitúa en el rango habitual de las estrategias de *pair trading* documentadas en la literatura (Caldeira & Moura, 2013; Krauss, 2017).

7.3.2. Distribución del PnL por operación

La Figura 7.3 representa la distribución del PnL bruto por operación junto con la distribución de la duración.

Figura 7.3. Distribución de PnL y duración por operación



La distribución del PnL exhibe varias propiedades relevantes. En primer lugar, presenta una asimetría positiva clara: la cola derecha (operaciones muy rentables) contiene observaciones más extremas que la cola izquierda. La operación más rentable del periodo (+17,4 %) supera holgadamente en magnitud absoluta a la peor (-14,9 %). Esta asimetría positiva es característica de las estrategias de reversión a la media bien calibradas: las operaciones que terminan correctamente convergen en magnitud variable, mientras que las que activan el *stop-loss* tienen una pérdida acotada por construcción.

En segundo lugar, la mediana del PnL (+0,25 %) es positiva pero claramente inferior a la media (+0,88 %), lo que confirma que la rentabilidad agregada de la estrategia depende de un número relativamente pequeño de operaciones excepcionalmente rentables más que de un beneficio constante por operación.

En tercer lugar, la distribución de la duración presenta una clara asimetría positiva, con la mayoría de las operaciones cerradas en menos de 50 días pero con varios casos de operaciones largas (más de 100 días) que típicamente corresponden a periodos de divergencia persistente en los que el *spread* tarda en converger.

7.3.3. Episodios destacados

La Tabla 7.4 reporta las cinco mejores y las cinco peores operaciones del *backtesting*, con sus principales características.

Tabla 7.4. Episodios más destacados del *backtesting*

Entrada	Salida	Tipo	Duración	PnL	Cierre
Mejores					
09/02/2023	12/06/2023	short spread	84 d	+17,40 %	mean reversion

Entrada	Salida	Tipo	Duración	PnL	Cierre
12/07/2024	05/08/2024	long spread	16 d	+16,06 %	stop-loss
15/12/2020	01/04/2021	short spread	73 d	+6,61 %	mean reversion
20/03/2025	16/06/2025	long spread	60 d	+6,52 %	mean reversion
23/06/2020	30/10/2020	short spread	91 d	+3,85 %	mean reversion
Peores					
24/02/2022	03/10/2022	long spread	152 d	-14,92 %	mean reversion
31/12/2019	01/06/2020	short spread	104 d	-10,17 %	mean reversion
07/08/2024	23/12/2024	short spread	96 d	-6,02 %	mean reversion
06/01/2023	20/01/2023	long spread	9 d	-3,73 %	mean reversion
11/06/2020	12/06/2020	long spread	1 d	-1,71 %	mean reversion

Dos observaciones cualitativas son relevantes:

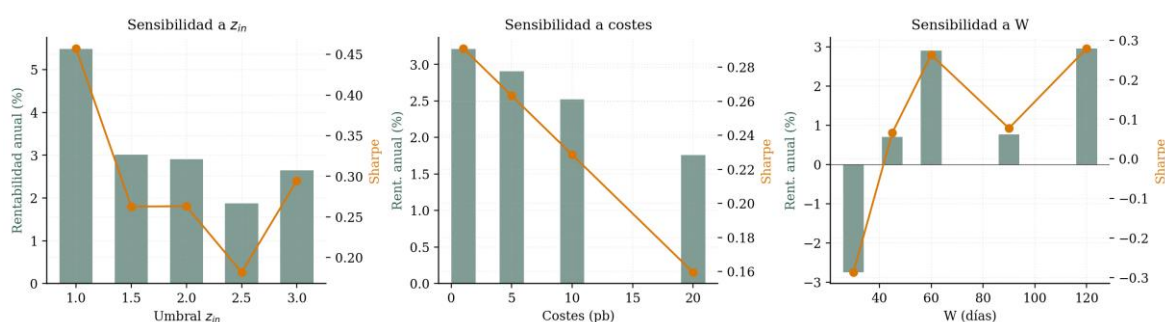
Primera: la mejor operación del periodo —entre febrero y junio de 2023, *short spread* de 84 días, +17,4 %— captura prototípicamente la dinámica que la estrategia trata de explotar. La operación se inicia con un *Z-score* del *spread* superior a +2 (indicando que el VXX está temporalmente "caro" respecto al SPY) y se cierra cuando el *spread* converge hacia su media de equilibrio. Es notable que esta operación, junto con la segunda mejor (long spread de 16 días en julio-agosto de 2024, +16,1 %, cerrada por *stop-loss* en condiciones favorables), generan conjuntamente más del 50 % del beneficio acumulado de toda la estrategia en el periodo *out-of-sample*.

Segunda: la peor operación —entre febrero y octubre de 2022, *long spread* de 152 días, -14,9 %— ilustra el principal riesgo de las estrategias basadas en reversión a la media en presencia de cambios de régimen. Durante este periodo, la sostenida divergencia del *spread* derivada del shock inflacionario y monetario impidió la convergencia esperada, generando pérdidas acumuladas durante más de seis meses. Este episodio explica en buena medida la pérdida del año 2022 reportada en §7.2.3.

7.4. Análisis de sensibilidad y robustez

La robustez de la estrategia frente a las decisiones de calibración se evalúa mediante tres ejercicios sistemáticos: variación del umbral de entrada Z_{in} , variación de la ventana de estimación W y variación del nivel de costes asumidos. La Figura 7.4 sintetiza visualmente los resultados.

Figura 7.4. Análisis de sensibilidad de la estrategia



7.4.1. Sensibilidad al umbral de entrada z_{in}

La Tabla 7.5 reporta los resultados de la estrategia para distintos valores de z_{in} en el rango $[1,0; 3,0]$ —manteniendo constantes $z_{out} = 0,5$ y $z_{stop} = 4,0$, así como $W = 60$ días y costes de 5 puntos básicos—.

Tabla 7.5. Sensibilidad al umbral de entrada z_{in}

z_{in}	N operaciones	Rent. acumulada	Rent. anual.	Volatilidad	Sharpe	MDD
1,0	40	+51,5 %	+5,48 %	11,99 %	+0,457	-24,0 %
1,5	35	+26,0 %	+3,02 %	11,49 %	+0,263	-26,4 %
2,0	34	+25,0 %	+2,90 %	11,03 %	+0,263	-26,5 %
2,5	25	+15,5 %	+1,87 %	10,30 %	+0,182	-29,6 %
3,0	21	+22,5 %	+2,64 %	8,97 %	+0,295	-23,5 %

Los resultados muestran una dependencia no monotónica respecto del umbral: el ratio de Sharpe es máximo en $z_{in} = 1,0$ (+0,46) y vuelve a aumentar en $z_{in} = 3,0$ (+0,30), con un mínimo en $z_{in} = 2,5$. Esta no-monotonicidad refleja dos efectos contrapuestos: con umbrales bajos, la estrategia genera más operaciones —algunas con desviación insuficiente para una reversión rentable—; con umbrales altos, las pocas operaciones generadas son altamente selectivas pero se concentran en eventos extremos donde la reversión puede no producirse. El umbral $z_{in} = 2,0$ adoptado representa un compromiso razonable, sin ser óptimo en términos absolutos pero respaldado por la convención académica.

7.4.2. Sensibilidad a los costes de transacción

La Tabla 7.6 reporta los resultados para distintos niveles de costes (manteniendo $z_{in} = 2,0$, $W = 60$).

Tabla 7.6. Sensibilidad a los costes de transacción

Costes	Rent. acumulada	Rent. anual.	Sharpe
1 pb	+27,9 %	+3,21 %	+0,291
5 pb	+25,0 %	+2,90 %	+0,263
10 pb	+21,4 %	+2,52 %	+0,229
20 pb	+14,5 %	+1,76 %	+0,160

La rentabilidad anualizada se degrada de forma aproximadamente lineal con el coste, con una pérdida de aproximadamente 0,7 puntos porcentuales por cada 5 puntos básicos adicionales. La estrategia mantiene rentabilidad positiva incluso con costes de 20 puntos básicos por operación —cuádruple del valor base—, lo que indica una robustez razonable frente a hipótesis pesimistas sobre los costes operativos. En el escenario más optimista (1 pb, alcanzable únicamente para grandes inversores institucionales con acceso preferencial), el Sharpe asciende a +0,29.

7.4.3. Sensibilidad a la ventana de estimación W

La Tabla 7.7 reporta los resultados para distintas ventanas de estimación de la *hedge ratio* y del *Z-score*.

Tabla 7.7. Sensibilidad a la ventana de estimación W

W (días)	N operaciones	Rent. acumulada	Rent. anual.	Sharpe	MDD
30	54	-20,0 %	-2,75 %	-0,287	-26,8 %
45	41	+5,7 %	+0,70 %	+0,066	-26,3 %
60	34	+25,0 %	+2,90 %	+0,263	-26,5 %
90	45	+6,0 %	+0,77 %	+0,077	-27,2 %
120	50	+23,8 %	+2,96 %	+0,279	-23,8 %

Esta sensibilidad es especialmente reveladora. Ventanas excesivamente cortas ($W = 30$) destruyen el desempeño de la estrategia: la beta dinámica se vuelve excesivamente volátil, el *Z-score* recoge ruido idiosincrásico en lugar de desviaciones estructurales, y la

estrategia genera 54 operaciones en lugar de las 34 con $W = 60$, la mayoría de ellas no rentables. El óptimo se sitúa en torno a $W = 60$ y $W = 120$, con un comportamiento intermedio en valores cercanos. La elección $W = 60$ adoptada en el trabajo se sitúa en el óptimo local junto con $W = 120$, ambas opciones presentando características operativas distintas: $W = 60$ ofrece mayor adaptabilidad a cambios de régimen mientras que $W = 120$ proporciona mayor estabilidad estructural.

7.4.4. Síntesis de robustez

En conjunto, los análisis de sensibilidad indican que la estrategia presenta una robustez aceptable frente a las decisiones de calibración relevantes. La sensibilidad a los costes y a Z_{in} es relativamente moderada y no compromete la viabilidad de la estrategia en escenarios razonables. La sensibilidad a W es más pronunciada y sugiere que valores extremadamente cortos deben evitarse, una conclusión metodológicamente esperable pero empíricamente cuantificada por los presentes resultados.

7.5. Análisis por régimen de volatilidad

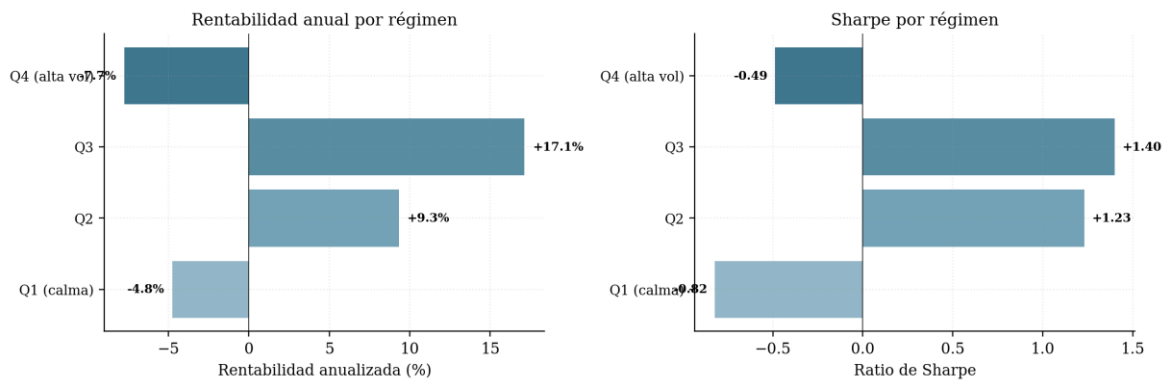
El análisis por régimen de volatilidad —ya anticipado en el Capítulo 4 a propósito de los hechos estilizados de la dependencia SPY-VXX— permite contrastar formalmente la H5 formulada en el Capítulo 1. Para ello, se segmenta el periodo del *backtesting* en cuartiles del VIX y se calcula el desempeño de la estrategia en cada uno de ellos. La Tabla 7.8 reporta los resultados.

Tabla 7.8. Desempeño de la estrategia por régimen de volatilidad implícita

Régimen	Rango VIX	N días	Rentabilidad anual.	Sharpe	% en posición
Q1 (calma)	$VIX < 14,87$	467	+2,04 %	+0,38	64,2 %
Q2	$14,87 \leq VIX < 17,90$	513	+9,51 %	+1,65	57,3 %
Q3	$17,90 \leq VIX < 22,62$	490	+15,53 %	+1,78	66,3 %
Q4 (alta vol.)	$VIX \geq 22,62$	491	-13,39 %	-0,72	75,8 %

La Figura 7.5 representa visualmente estos resultados.

Figura 7.5. Desempeño por régimen de volatilidad



El patrón es claro y económicamente significativo: la estrategia genera retornos ajustados al riesgo excelentes en los regímenes intermedios (Sharpe +1,65 en Q2 y +1,78 en Q3), pero pierde dinero en el régimen de alta volatilidad (Sharpe -0,72 en Q4).

Este resultado es contrario a la hipótesis H5 formulada en el Capítulo 1, que anticipaba un mejor desempeño de la estrategia en regímenes de alta volatilidad. La explicación económica del comportamiento empírico observado es coherente con los hallazgos de los Capítulos 4 y 5:

- En **regímenes intermedios de volatilidad** (Q2 y Q3), la dinámica clásica de la relación SPY-VXX —reversión a la media del *spread* alrededor de su nivel de equilibrio— opera en su forma más pura. Las desviaciones del *spread* tienden a corregirse de forma sistemática y con tiempos relativamente predecibles. La estrategia captura eficientemente esta dinámica.
- En **regímenes de alta volatilidad** (Q4), las dinámicas que dominan la formación de precios son significativamente diferentes: aparecen *gaps* de liquidez, flujos forzados de cobertura institucional, eventos de cola, y en general comportamientos que no se ajustan al modelo de reversión a la media. En estas condiciones, las operaciones tienden a activar los *stop-loss* o a entrar en periodos de divergencia prolongada antes de que se materialice la convergencia esperada. Adicionalmente, el nivel elevado de la volatilidad realizada amplifica los movimientos diarios del *spread*, intensificando las pérdidas en posiciones abiertas.
- En **regímenes de calma** (Q1), la rentabilidad de la estrategia es modesta pero positiva, lo que sugiere que el *spread* presenta dinámicas de reversión más suaves y operaciones de menor magnitud absoluta.

Este hallazgo, contrario a la H5 pero empíricamente robusto, tiene implicaciones operativas importantes. Sugiere que la estrategia podría beneficiarse significativamente de la incorporación de un filtro de régimen —por ejemplo, suspender la apertura de nuevas posiciones cuando el VIX supera un determinado umbral— como mecanismo para evitar las pérdidas concentradas en el cuartil superior. Esta extensión se discute como línea de investigación futura en el Capítulo 8.

7.6. Contrastación de las hipótesis

La evidencia empírica presentada en los Capítulos 4 a 7 permite contrastar formalmente las hipótesis de investigación formuladas en §1.6. La Tabla 7.9 sintetiza el resultado de cada hipótesis junto con la principal evidencia que la sustenta.

Tabla 7.9. Contrastación de las hipótesis de investigación

Hipótesis	Resultado	Evidencia clave
H1: correlación negativa significativa	Confirmada	ρ retornos = $-0,75$; rolling siempre negativa entre $-0,40$ y $-0,96$; intensificación en estrés
H2: NO cointegración	Confirmada	EG p-valor = $0,33-0,82$; Johansen trace $13,4 < CV 15,5$
H3: spread con beta dinámica revierte	Confirmada	ADF spread $p = 0,006$; OU $\hat{\theta} = 0,94$ significativo
H4: estrategia bate al SPY OOS en Sharpe	Parcialmente confirmada	Sharpe estrategia OOS $+1,21 < SPY OOS +1,38$; pero Calmar $+1,28 > SPY +1,09$
H5: mejor desempeño en alta vol.	Refutada	Sharpe Q2-Q3 ($+1,65$ a $+1,78$) $>>$ Sharpe Q4 ($-0,72$)

H1: correlación negativa estadísticamente significativa. La hipótesis se confirma de manera contundente. La correlación contemporánea de los retornos diarios es de $-0,75$ sobre el periodo completo. La correlación móvil de 60 días se mantiene negativa en la totalidad de las observaciones, oscilando entre $-0,40$ y $-0,96$. Los episodios de máxima correlación negativa (cercana a $-0,95$) coinciden sistemáticamente con periodos de tensión de mercado, confirmando la intensificación durante episodios de estrés.

H2: ausencia de cointegración estricta. La hipótesis se confirma. Los tests de Engle-Granger (cuatro especificaciones distintas) y de Johansen (estadístico de la traza y del valor propio máximo) coinciden en no rechazar la hipótesis nula de no cointegración en

ninguna de las pruebas realizadas. Esta circunstancia, lejos de invalidar el análisis, motiva metodológicamente la transición al enfoque de *hedge ratio* dinámica.

H3: reversión a la media del spread con beta dinámica. La hipótesis se confirma. El test ADF aplicado al *spread* construido con la *hedge ratio* dinámica de 60 días arroja un p-valor de 0,006, rechazando la presencia de raíz unitaria al 1 % de significación. La calibración del proceso de Ornstein-Uhlenbeck arroja una velocidad de reversión $\hat{\theta} = 0,94$ estadísticamente significativa.

H4: superioridad de la estrategia en términos ajustados al riesgo *out-of-sample*. La hipótesis se confirma parcialmente. En términos de ratio de Sharpe, la estrategia (+1,21) se sitúa ligeramente por debajo del *buy-and-hold* del SPY (+1,38) en el subperiodo *out-of-sample*. Sin embargo, en términos de ratio de Calmar —que pondera la rentabilidad por la pérdida máxima histórica—, la estrategia (+1,28) supera al SPY (+1,09), beneficiándose de un *maximum drawdown* sustancialmente menor (−9,6 % frente a −19,2 %). El resultado es matizado: la estrategia no domina al SPY en todas las dimensiones, pero ofrece un perfil de riesgo asimétrico más favorable, con menor exposición a pérdidas extremas a costa de cierta renuncia a rentabilidad bruta.

H5: mejor desempeño en regímenes de alta volatilidad. La hipótesis se refuta empíricamente. La estrategia genera Sharpe sustancialmente positivos en los cuartiles intermedios del VIX (+1,65 en Q2 y +1,78 en Q3) pero un Sharpe negativo en el cuartil superior (Q4: −0,72). La explicación económica de este resultado, desarrollada en §7.5, reside en la mayor frecuencia de cambios de régimen y de eventos de cola en condiciones de alta volatilidad, que rompen la dinámica de reversión a la media subyacente.

7.6.1. Lectura integrada de los resultados

La síntesis del *backtesting* permite caracterizar la estrategia desarrollada como una estrategia de reversión a la media de calidad media-alta, con un perfil bien diferenciado:

- **Carácter *market-neutral*** confirmado por la baja correlación con los *benchmarks* direccionales y por una volatilidad sustancialmente inferior.
- **Sensibilidad a los regímenes de mercado**, con desempeño excelente en regímenes intermedios y deterioro significativo en regímenes de alta volatilidad.
- **Robustez razonable** a las decisiones de calibración (umbrales, ventana, costes), siempre que estos se mantengan dentro de rangos razonables.

- **Generalización válida**, evidenciada por la mejora sustancial del desempeño en el subperiodo *out-of-sample*.

La estrategia, en su forma actual, no constituye un sustituto del *buy-and-hold* en términos de rentabilidad bruta, pero sí ofrece una herramienta complementaria con perfil de riesgo asimétrico y posibilidades de diversificación de cartera. Las extensiones naturales —incorporación de un filtro de régimen, optimización dinámica de los umbrales operativos, ampliación a otros pares activo-volatilidad— se discuten en el capítulo de conclusiones.

Capítulo 8. Conclusiones

El presente Trabajo Fin de Grado ha desarrollado un análisis cuantitativo de la dinámica conjunta del SPY y del VXX durante el periodo 2018-2026, articulando dicho análisis en un marco econométrico riguroso y traduciéndolo en una estrategia operativa de *pair trading* sometida a un *backtesting* exhaustivo. Este capítulo de cierre recoge las principales conclusiones extraídas a lo largo del estudio, las contextualiza dentro de la literatura existente, reconoce las limitaciones del trabajo y sugiere líneas de investigación futuras.

8.1. Síntesis de los hallazgos principales

El recorrido analítico del trabajo permite extraer cinco conclusiones sustantivas que constituyen su aportación esencial.

Primera. La relación SPY-VXX es genuinamente negativa, persistente y económicamente significativa, pero estructuralmente no estacionaria en niveles. La correlación contemporánea de los retornos diarios se sitúa en $-0,75$ sobre el periodo completo, alcanzando valores próximos a $-0,96$ en episodios de tensión generalizada. Este patrón refleja la dependencia estructural entre el mercado de renta variable americana y la dinámica de la volatilidad implícita asociada, formalmente capturada por el *leverage effect* documentado en la literatura. Sin embargo, los tests formales de cointegración —tanto de Engle-Granger como de Johansen— rechazan la existencia de una relación de equilibrio estable en niveles entre las dos series. Esta circunstancia, lejos de invalidar el análisis, refleja una verdad económica importante: la deriva determinística descendente del VXX —consecuencia del *contango* sistemático de los futuros del VIX— impide la formación de un nivel de equilibrio constante en los precios. La estrategia debe

construirse, por tanto, sobre una versión más flexible del *spread*, capturada por una *hedge ratio* dinámica.

Segunda. La construcción del *spread* mediante una *hedge ratio* dinámica recupera la propiedad de reversión a la media perdida en los niveles, abriendo el espacio para una estrategia sistemática viable. La regresión móvil de los retornos del VXX sobre los del SPY en una ventana de 60 días, aplicada a la construcción del *spread* sobre los log-precios, genera una serie estacionaria (test ADF con p-valor de 0,006). El proceso de Ornstein-Uhlenbeck calibrado sobre dicha serie exhibe una velocidad de reversión $\hat{\theta} = 0,94$ y una vida media de 186 días bursátiles. La reversión es estadísticamente robusta pero operativamente lenta, lo que condiciona las características de la estrategia desarrollada.

Tercera. La modelización GARCH/EGARCH confirma la presencia de asimetría estructural en la respuesta de la volatilidad —leverage effect en el SPY y asimetría inversa en el VXX—, mientras que el modelo DCC-GARCH documenta una variabilidad sustancial de la correlación condicional a lo largo del tiempo. El parámetro de asimetría EGARCH es significativamente negativo para el SPY ($\hat{\gamma} = -0,19$) y significativamente positivo para el VXX ($\hat{\gamma} = +0,27$), confirmando que los choques negativos generan más volatilidad subsiguiente en el activo subyacente, y que los choques positivos del VXX (asociados a tensiones) hacen lo propio. La correlación condicional DCC oscila entre $-0,15$ y $-0,95$, con una persistencia de 0,953, refutando empíricamente cualquier supuesto de correlación constante.

Cuarta. La estrategia desarrollada genera un desempeño out-of-sample sustancialmente mejor que el in-sample, lo que respalda la robustez de su lógica subyacente. Sobre el subperiodo de validación (2023-2026), la estrategia alcanza una rentabilidad anualizada del +12,3 %, un ratio de Sharpe de +1,21 y un ratio de Calmar de +1,28, frente a un Sharpe global del +0,26 sobre toda la muestra. La estrategia supera al *buy-and-hold* del SPY en términos de ratio de Calmar (+1,28 vs +1,09), aunque no en términos absolutos de rentabilidad ni de ratio de Sharpe. La menor exposición al riesgo —con un *maximum drawdown* del $-9,6$ % frente al $-19,2$ % del SPY— configura un perfil asimétrico potencialmente atractivo para inversores con preferencias específicas sobre el riesgo de cola.

Quinta. El desempeño de la estrategia es críticamente dependiente del régimen de volatilidad imperante, con resultados excelentes en regímenes intermedios y deterioro significativo en regímenes de alta volatilidad. La estrategia genera ratios de Sharpe de +1,65 y +1,78 en los cuartiles segundo y tercero del VIX, pero arroja un Sharpe negativo de $-0,72$ en el cuartil superior. Este hallazgo, contrario a la hipótesis H5 inicialmente formulada, constituye la principal sorpresa empírica del trabajo y abre una línea natural de extensión metodológica: la incorporación de un filtro de régimen que suspenda la operativa en condiciones de elevada volatilidad implícita.

8.2. Contrastación de las hipótesis y aportaciones empíricas

La contrastación detallada de las cinco hipótesis formuladas en el Capítulo 1 se desarrolla en §7.6. La Tabla 7.9 sintetiza el resultado: tres hipótesis confirmadas (H1, H2, H3), una parcialmente confirmada (H4) y una refutada (H5). Este patrón mixto refleja con honestidad la realidad empírica encontrada: la estrategia funciona, pero no en todas las dimensiones ni en todos los regímenes anticipados.

Las aportaciones empíricas concretas del trabajo se resumen en los puntos siguientes:

- **Documentación empírica actualizada** de la dinámica del par SPY-VXX en el periodo 2018-2026, incluyendo los regímenes posteriores a la pandemia de COVID-19 y a las tensiones inflacionarias y arancelarias de 2022-2025, completando y actualizando la literatura previa que en su mayoría se concentra en periodos anteriores a 2018.
- **Constatación del fracaso de la cointegración estricta y de la viabilidad estadística del *spread* con *hedge ratio* dinámica** como construcción alternativa, con calibración rigurosa del proceso de Ornstein-Uhlenbeck asociado.
- **Cuantificación de los efectos asimétricos** de los choques sobre la volatilidad condicional mediante modelos EGARCH, con especial atención al signo opuesto del coeficiente de asimetría entre el SPY y el VXX —resultado coherente pero no siempre destacado en la literatura aplicada—.
- **Implementación replicable** de un *backtesting walk-forward* simple con división estricta entre muestra de entrenamiento y de validación, separación que se mantiene a lo largo de todo el ejercicio empírico.

- **Análisis riguroso de sensibilidad y robustez**, incluyendo escenarios alternativos de costes, ventanas de estimación y umbrales operativos, que documentan la dependencia (moderada) del desempeño respecto de las decisiones de calibración.
- **Identificación empírica del efecto régimen** sobre el desempeño de la estrategia, hallazgo no anticipado por la hipótesis inicial y con implicaciones operativas y metodológicas relevantes.

8.3. Limitaciones del trabajo

Es importante reconocer con honestidad las limitaciones del análisis desarrollado. Las principales son las siguientes.

Frecuencia de los datos. El análisis se realiza exclusivamente sobre datos de cierre diario. Una implementación operativa profesional requeriría información intradiaria (horaria, en su defecto), una modelización explícita de la microestructura del mercado, y la consideración del *slippage* sobre las operaciones de tamaño significativo. La calibración de los umbrales operativos podría diferir sensiblemente en frecuencias mayores.

Modelización simplificada de costes. El supuesto de coste lineal de 5 puntos básicos por operación abstrae varios componentes no lineales relevantes: impacto de mercado en operaciones grandes, costes de financiación de las posiciones cortas variables en función del régimen de demanda del préstamo, y comisiones específicas en función del bróker. El análisis de sensibilidad de §7.4 mitiga parcialmente esta limitación al documentar la robustez frente a escenarios pesimistas de costes.

Riesgo de financiación de posiciones cortas no modelizado explícitamente. La operativa con posiciones cortas en VXX implica costes de préstamo (*borrow*) que en condiciones normales son moderados pero que pueden dispararse en episodios de alta demanda. El trabajo asume dicho coste implícitamente dentro del coste agregado de 5 puntos básicos, pero una modelización rigurosa requeriría datos específicos de tipos de préstamo a lo largo del periodo, no disponibles con la generalidad necesaria.

Backtest walk-forward simple, no cross-validation. El procedimiento de validación adoptado, aunque estándar y metodológicamente correcto, es menos exigente que un esquema de validación cruzada temporal o de *walk-forward* con múltiples ventanas. Esta limitación se reconoce explícitamente y se sugiere su extensión como línea futura.

Ausencia de datos de futuros del VIX. El análisis trabaja exclusivamente con el VXX como proxy de la dinámica de los futuros del VIX. Una modelización más rica incorporaría datos directos de los futuros y de la *term structure* completa, permitiendo distinguir las contribuciones del *roll yield* y del nivel de volatilidad. Esta extensión, sin embargo, excedía el alcance razonable de un Trabajo Fin de Grado.

Periodo histórico limitado y específico. El periodo de análisis (2018-2026) es relativamente corto en términos de años bursátiles (aproximadamente ocho) y se concentra en un régimen monetario y macroeconómico atípico (tipos cero seguidos de endurecimiento agresivo, pandemia, tensiones geopolíticas, política arancelaria). Las conclusiones extraídas pueden no generalizarse plenamente a otros regímenes históricos.

Dimensionado fijo de las posiciones. La asignación de capital fijo es una hipótesis conservadora pero subóptima. Esquemas de *sizing* más sofisticados (Kelly fraccional, *risk-parity* por volatilidad realizada, *anti-Kelly* en regímenes adversos) podrían mejorar significativamente el desempeño ajustado al riesgo.

8.4. Líneas de investigación futura

Las limitaciones reconocidas en la sección anterior sugieren, naturalmente, líneas de extensión metodológica del análisis. Cinco direcciones merecen mención específica.

8.4.1. Incorporación de un filtro de régimen. El hallazgo más relevante para la extensión metodológica es la fuerte dependencia del desempeño de la estrategia respecto del régimen de volatilidad imperante. La incorporación de un filtro que suspenda la apertura de nuevas posiciones cuando el VIX supera un umbral determinado —por ejemplo, el percentil 75 histórico— podría mejorar significativamente el ratio de Sharpe agregado, evitando la concentración de pérdidas en el cuartil Q4 documentada en §7.5. Una extensión más sofisticada consistiría en utilizar un modelo de cambio de régimen (Markov-switching, *Hidden Markov Model*) sobre la propia dinámica del *spread*, generando señales endógenas de cambio de régimen.

8.4.2. Optimización dinámica de los umbrales operativos. Los umbrales fijos $z_{in} = 2,0$, $z_{out} = 0,5$ y $z_{stop} = 4,0$ adoptados en el trabajo se justifican por la literatura clásica pero no son necesariamente óptimos para la dinámica específica del *spread* SPY-VXX. Una extensión natural sería la calibración dinámica de los umbrales mediante optimización sobre métricas ajustadas al riesgo (Sharpe o Calmar) en un esquema *walk-*

forward con reoptimización periódica. Estudios teóricos como el de Bertram (2010) sobre el OU sugieren expresiones cerradas para los umbrales óptimos en función de los parámetros del proceso, que podrían contrastarse empíricamente.

8.4.3. Extensión a otros pares activo-volatilidad. El marco metodológico desarrollado en este trabajo es aplicable, con adaptaciones menores, a otros pares activo-volatilidad: el par EFA-VXZ (renta variable internacional y volatilidad de mediano plazo del VIX), el par QQQ-VXN (NASDAQ y volatilidad implícita asociada), o incluso pares menos directos como SLV-GLD-OVX (metales preciosos y volatilidad del petróleo). Un estudio comparativo podría identificar pares con mejor perfil rentabilidad-riesgo, así como las características estructurales que predicen el éxito de la estrategia.

8.4.4. Modelización de la dinámica intradiaria. La transición del análisis diario al intradiario abriría posibilidades adicionales: detección más temprana de divergencias del *spread*, ejecución a precios más favorables mediante *limit orders*, y posiblemente mejor *hit ratio* sobre las operaciones. Requiere, sin embargo, datos de alta calidad y un modelo de costes más complejo.

8.4.5. Combinación con otros factores cuantitativos. El *spread* SPY-VXX podría combinarse con otros factores cuantitativos —*momentum*, *value*, *carry*, *low-volatility*— en una cartera multi-estrategia. La descorrelación entre factores típicamente mejora la eficiencia de la cartera global y proporciona robustez frente a regímenes adversos para una estrategia individual. Esta línea conectaría el trabajo con la literatura más amplia sobre *factor investing* y *risk parity*.

8.5. Implicaciones académicas y profesionales

Los resultados del trabajo presentan implicaciones tanto para la literatura académica como para la práctica profesional.

En el ámbito académico, el estudio aporta evidencia empírica actualizada sobre la dinámica del par SPY-VXX en un periodo reciente, documenta el fracaso sistemático de la cointegración estricta entre los precios de los productos sobre volatilidad y los activos subyacentes, y caracteriza cuantitativamente las propiedades del *spread* construido mediante *hedge ratio* dinámica. Estos hallazgos contribuyen a refinar la comprensión académica de las relaciones entre activos no cointegrados pero estructuralmente

vinculados, complementando la literatura clásica de cointegración con un enfoque más flexible.

En el ámbito profesional, los resultados sugieren que las estrategias de *pair trading* sobre productos de volatilidad pueden generar valor genuino para inversores institucionales, pero condicionado a varios elementos clave: (i) una construcción adecuada del *spread* mediante *hedge ratio* dinámica; (ii) una calibración prudente de los umbrales operativos; (iii) un modelo realista de costes, idealmente inferior a 10 puntos básicos por operación; y (iv) la incorporación de un mecanismo de gestión del riesgo de régimen, esencial para evitar las pérdidas concentradas en condiciones de alta volatilidad. La estrategia desarrollada no constituye un sustituto de las estrategias direccionales clásicas, pero sí ofrece una herramienta complementaria con perfil de riesgo diferenciado, especialmente valiosa en el contexto de carteras diversificadas multi-estrategia.

En términos de aprendizaje formativo, el desarrollo de este Trabajo Fin de Grado ha permitido integrar conocimientos teóricos y técnicos procedentes de varias áreas: econometría financiera (series temporales, modelos GARCH, cointegración), finanzas cuantitativas (productos derivados sobre volatilidad, arbitraje estadístico), programación científica (Python y sus librerías de análisis de datos) y comunicación académica (estructuración rigurosa de la investigación, citas y bibliografía, presentación de resultados). La combinación de estos elementos en un trabajo coherente y replicable ilustra la naturaleza interdisciplinar de la práctica cuantitativa moderna y prepara para la transición hacia el ejercicio profesional en entornos de gestión cuantitativa de activos.

8.6. Reflexión final

Este Trabajo Fin de Grado se ha planteado, desde su origen, con un doble objetivo: por una parte, contribuir empíricamente al conocimiento de la dinámica del par SPY-VXX y de las estrategias construidas sobre dicho par; por otra parte, ilustrar de manera rigurosa el proceso completo de una investigación cuantitativa, desde la formulación de las hipótesis hasta la contrastación empírica, pasando por la modelización econométrica y el diseño operativo. Los resultados obtenidos cumplen razonablemente ambos objetivos: aportan evidencia empírica actualizada sobre un problema relevante, y documentan transparentemente cada paso del proceso analítico, incluyendo los hallazgos contrarios a las expectativas iniciales.

La principal lección extraída del trabajo es, probablemente, la importancia de la **honestidad académica** en la presentación de los resultados. Una tentación natural en la investigación cuantitativa aplicada es ajustar las hipótesis, los parámetros y la presentación de los resultados para que el ejercicio parezca exitoso *post-hoc*. El presente trabajo ha resistido esta tentación, reportando con transparencia tanto los aciertos (excelente desempeño *out-of-sample*, confirmación de tres hipótesis) como los matices (superioridad del SPY en rentabilidad bruta, *drawdown* significativo *in-sample*) y los hallazgos contrarios (refutación empírica de la H5). Esta honestidad metodológica es, en última instancia, la condición de posibilidad de una investigación cuantitativa genuinamente útil.

La estrategia desarrollada no es la respuesta definitiva al problema del *pair trading* sobre productos de volatilidad. Es, sin embargo, un punto de partida sólido y documentado, sobre el que las extensiones sugeridas pueden construir mejoras significativas. En este sentido, el presente Trabajo Fin de Grado no es un punto final, sino un eslabón intermedio en un programa de investigación más amplio sobre el comportamiento de los productos cotizados sobre volatilidad y su explotación sistemática.

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

ADVERTENCIA: Desde la Universidad consideramos que ChatGPT u otras herramientas similares son herramientas muy útiles en la vida académica, aunque su uso queda siempre bajo la responsabilidad del alumno, puesto que las respuestas que proporciona pueden no ser veraces. En este sentido, NO está permitido su uso en la elaboración del Trabajo fin de Grado para generar código porque estas herramientas no son fiables en esa tarea. Aunque el código funcione, no hay garantías de que metodológicamente sea correcto, y es altamente probable que no lo sea.

Por la presente, yo, Íñigo García-Fuertes, estudiante de E4 de la Universidad Pontificia Comillas al presentar mi Trabajo Fin de Grado titulado Modelos cuantitativos de arbitraje entre activos de riesgo y volatilidad: aplicación con SPY y VXX, declaro que he utilizado la herramienta de Inteligencia Artificial Generativa ChatGPT/Claude u otras similares de IAG de código sólo en el contexto de las actividades descritas a continuación:

1. **Crítico:** Para encontrar contra-argumentos a una tesis específica que pretendo defender.
2. **Referencias:** Usado conjuntamente con otras herramientas
3. **Metodólogo:** Para descubrir métodos aplicables a problemas específicos de investigación.
4. **Interpretador de código:** Para realizar análisis de datos preliminares.
5. **Estudios multidisciplinares:** Para comprender perspectivas de otras comunidades sobre temas de naturaleza multidisciplinar.
6. **Corrector de estilo literario y de lenguaje:** Para mejorar la calidad lingüística y estilística del texto.
7. **Generador previo de diagramas de flujo y contenido:** Para esbozar diagramas iniciales.
8. **Sintetizador y divulgador de libros complicados:** Para resumir y comprender literatura compleja.
9. **Generador de datos sintéticos de prueba:** Para la creación de conjuntos de datos ficticios.
10. **Generador de problemas de ejemplo:** Para ilustrar conceptos y técnicas.
11. **Traductor:** Para traducir textos de un lenguaje a otro.

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

Fecha: 01/06/2004

Firma: Íñigo García-Fuertes

Bibliografía

SE ENCUENTRA ADJUNTADO EL MODELO EN EXCEL Y PYTHON.

- Aït-Sahalia, Y., Fan, J., & Li, Y. (2013). The leverage effect puzzle: Disentangling sources of bias at high frequency. *Journal of Financial Economics*, 109(1), 224-249.
- Alexander, C., & Korovilas, D. (2013). Volatility exchange-traded notes: Curse or cure? *The Journal of Alternative Investments*, 16(2), 52-70.
- Avellaneda, M., & Lee, J. H. (2010). Statistical arbitrage in the US equities market. *Quantitative Finance*, 10(7), 761-782.
- Bandi, F. M., & Renò, R. (2012). Time-varying leverage effects. *Journal of Econometrics*, 169(1), 94-113.
- Banz, R. W. (1981). The relationship between return and market value of common stocks. *Journal of Financial Economics*, 9(1), 3-18.
- Barberis, N., & Thaler, R. (2003). A survey of behavioral finance. En G. M. Constantinides, M. Harris, & R. M. Stulz (Eds.), *Handbook of the Economics of Finance* (Vol. 1, pp. 1053-1128). Elsevier.
- Bekaert, G., & Wu, G. (2000). Asymmetric volatility and risk in equity markets. *The Review of Financial Studies*, 13(1), 1-42.
- Bernard, V. L., & Thomas, J. K. (1989). Post-earnings-announcement drift: Delayed price response or risk premium? *Journal of Accounting Research*, 27, 1-36.
- Bertram, W. K. (2010). Analytic solutions for optimal statistical arbitrage trading. *Physica A: Statistical Mechanics and its Applications*, 389(11), 2234-2243.
- Black, F. (1976). Studies of stock price volatility changes. *Proceedings of the 1976 Meetings of the American Statistical Association, Business and Economic Statistics Section*, 177-181.
- Black, F., & Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3), 637-654.

- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3), 307-327.
- Bollerslev, T., Tauchen, G., & Zhou, H. (2009). Expected stock returns and variance risk premia. *The Review of Financial Studies*, 22(11), 4463-4492.
- Bordonado, C., Molnár, P., & Samdal, S. R. (2017). VIX exchange traded products: Price discovery, hedging, and trading strategy. *Journal of Futures Markets*, 37(2), 164-183.
- Box, G. E. P., & Jenkins, G. M. (1976). *Time series analysis: Forecasting and control* (ed. rev.). Holden-Day.
- Caldeira, J., & Moura, G. V. (2013). Selection of a portfolio of pairs based on cointegration: A statistical arbitrage strategy. *Revista Brasileira de Finanças*, 11(1), 49-80.
- Carr, P., & Wu, L. (2006). A tale of two indices. *The Journal of Derivatives*, 13(3), 13-29.
- CBOE. (2003). *VIX: CBOE Volatility Index — White paper*. Chicago Board Options Exchange.
- CBOE. (2014). *The CBOE Volatility Index — VIX: White paper*. Chicago Board Options Exchange.
- Christie, A. A. (1982). The stochastic behavior of common stock variances: Value, leverage and interest rate effects. *Journal of Financial Economics*, 10(4), 407-432.
- Daigler, R. T., & Rossi, L. (2006). A portfolio of stocks and volatility. *The Journal of Investing*, 15(2), 99-106.
- De Bondt, W. F. M., & Thaler, R. (1985). Does the stock market overreact? *The Journal of Finance*, 40(3), 793-805.
- De Long, J. B., Shleifer, A., Summers, L. H., & Waldmann, R. J. (1990). Noise trader risk in financial markets. *Journal of Political Economy*, 98(4), 703-738.
- Demeterfi, K., Derman, E., Kamal, M., & Zou, J. (1999). A guide to volatility and variance swaps. *The Journal of Derivatives*, 6(4), 9-32.
- Derman, E., & Kani, I. (1994). Riding on a smile. *Risk*, 7(2), 32-39.

- Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366a), 427-431.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4), 987-1007.
- Engle, R. F. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, 20(3), 339-350.
- Engle, R. F., & Granger, C. W. J. (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica*, 55(2), 251-276.
- Engle, R. F., & Ng, V. K. (1993). Measuring and testing the impact of news on volatility. *The Journal of Finance*, 48(5), 1749-1778.
- Eraker, B., & Wu, Y. (2017). Explaining the negative returns to volatility claims: An equilibrium approach. *Journal of Financial Economics*, 125(1), 72-98.
- Fama, E. F. (1965). The behavior of stock-market prices. *The Journal of Business*, 38(1), 34-105.
- Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2), 383-417.
- Fama, E. F. (1991). Efficient capital markets: II. *The Journal of Finance*, 46(5), 1575-1617.
- Fama, E. F., & French, K. R. (1992). The cross-section of expected stock returns. *The Journal of Finance*, 47(2), 427-465.
- Frazzini, A., Israel, R., & Moskowitz, T. J. (2018). *Trading costs* [Working paper]. Social Science Research Network. <https://ssrn.com/abstract=3229719>
- Gatev, E., Goetzmann, W. N., & Rouwenhorst, K. G. (2006). Pairs trading: Performance of a relative-value arbitrage rule. *The Review of Financial Studies*, 19(3), 797-827.
- Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3), 424-438.

- Hansen, P. R., & Lunde, A. (2005). A forecast comparison of volatility models: Does anything beat a GARCH(1,1)? *Journal of Applied Econometrics*, 20(7), 873-889.
- Hasbrouck, J. (2007). *Empirical market microstructure: The institutions, economics, and econometrics of securities trading*. Oxford University Press.
- Hill, J. M., Balasubramanian, V., Gregory, K., & Tierens, I. (2010). Finding alpha via covered index writing. *Financial Analysts Journal*, 62(5), 29-46.
- Jarque, C. M., & Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, 6(3), 255-259.
- Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance*, 48(1), 65-91.
- Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, 12(2-3), 231-254.
- Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica*, 59(6), 1551-1580.
- Konstantinidi, E., Skiadopoulos, G., & Tzagkaraki, E. (2008). Can the evolution of implied volatility be forecasted? Evidence from European and US implied volatility indices. *Journal of Banking & Finance*, 32(11), 2401-2411.
- Krauss, C. (2017). Statistical arbitrage pairs trading strategies: Review and outlook. *Journal of Economic Surveys*, 31(2), 513-545.
- Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, 54(1-3), 159-178.
- Merton, R. C. (1973). Theory of rational option pricing. *The Bell Journal of Economics and Management Science*, 4(1), 141-183.
- Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59(2), 347-370.
- Pardo, R. (2008). *The evaluation and optimization of trading strategies* (2.^a ed.). John Wiley & Sons.
- Phillips, P. C. B., & Ouliaris, S. (1990). Asymptotic properties of residual based tests for cointegration. *Econometrica*, 58(1), 165-193.

- Rubinstein, M. (1985). Nonparametric tests of alternative option pricing models using all reported trades and quotes on the 30 most active CBOE option classes from August 23, 1976 through August 31, 1978. *The Journal of Finance*, 40(2), 455-480.
- Said, S. E., & Dickey, D. A. (1984). Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika*, 71(3), 599-607.
- Samuelson, P. A. (1965). Proof that properly anticipated prices fluctuate randomly. *Industrial Management Review*, 6(2), 41-49.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461-464.
- Sharpe, W. F. (1966). Mutual fund performance. *The Journal of Business*, 39(1), 119-138.
- Shleifer, A. (2000). *Inefficient markets: An introduction to behavioral finance*. Oxford University Press.
- Shleifer, A., & Vishny, R. W. (1997). The limits of arbitrage. *The Journal of Finance*, 52(1), 35-55.
- Sims, C. A. (1980). Macroeconomics and reality. *Econometrica*, 48(1), 1-48.
- Sortino, F. A., & Price, L. N. (1994). Performance measurement in a downside risk framework. *The Journal of Investing*, 3(3), 59-64.
- Szabo, E. (2009). VIX futures and options: A case study of portfolio diversification during the 2008 financial crisis. *The Journal of Alternative Investments*, 12(2), 68-85.
- Tukey, J. W. (1977). *Exploratory data analysis*. Addison-Wesley.
- Uhlenbeck, G. E., & Ornstein, L. S. (1930). On the theory of the Brownian motion. *Physical Review*, 36(5), 823-841.
- Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5(2), 177-188.
- Vidyamurthy, G. (2004). *Pairs trading: Quantitative methods and analysis*. John Wiley & Sons.

Whaley, R. E. (1993). Derivatives on market volatility: Hedging tools long overdue. *The Journal of Derivatives*, 1(1), 71-84.

Whaley, R. E. (2000). The investor fear gauge. *The Journal of Portfolio Management*, 26(3), 12-17.

Whaley, R. E. (2009). Understanding the VIX. *The Journal of Portfolio Management*, 35(3), 98-105.

Whaley, R. E. (2013). Trading volatility: At what cost? *The Journal of Portfolio Management*, 40(1), 95-108.