

AlphaTech Analyzer

Proyecto de Análisis Estadístico Financiero

Informe Estadístico

Análisis Cuantitativo del Sector Tecnológico

NASDAQ-100 | 2018–2024

Universo de análisis:	30 empresas tecnológicas NASDAQ-100
Período temporal:	Febrero 2018 – Diciembre 2024
Frecuencia de datos:	Mensual (hasta 83 observaciones por empresa)
Variables principales:	Precio Ajustado, Volumen, Retorno Log.
Observaciones totales:	2,403 registros

Índice

Resumen Ejecutivo	3
Introducción	4
1. Análisis Exploratorio de Datos (EDA)	6
1.1. Estructura del Conjunto de Datos	6
1.2. Análisis de Valores Faltantes	7
1.3. Distribución de Retornos Financieros	9
1.3.1. Distribución por Empresa: Análisis Comparativo	10
1.3.2. Volatilidad por Empresa	11
1.4. Relación Riesgo-Retorno	12
1.5. Evolución Temporal de Precios	14
1.5.1. Series Normalizadas: Comparación Multi-Activo	14
1.6. Análisis de Correlación	16
1.7. Detección de Valores Atípicos (Outliers)	18
1.8. Análisis de Momentos Estadísticos	19
1.9. Análisis de Drawdown	21
1.10. Análisis de Cambio Estructural: Pre/Post COVID-19	23
1.11. Síntesis del Análisis Exploratorio	26
2. Análisis de Clustering	27
2.1. Metodología y Preprocesamiento	27
2.1.1. Variables Utilizadas	27
2.2. Selección del Número Óptimo de Clusters	28
2.3. Decisión Final: $k = 2$	31
2.4. Validación Estadística: ANOVA	31
2.5. Perfiles de los Clusters	32
2.6. Visualización de Clusters	33
2.7. Detección de Outliers por Cluster	35
2.8. Síntesis del Análisis de Clustering	36
3. Análisis de Regresión	37
3.1. Modelo CAPM: Especificación y Estimación	37
3.2. Regresión Lineal Simple: Volatilidad vs Correlación	39
3.3. Validación de Supuestos de la Regresión	41
3.4. Síntesis del Análisis de Regresión	42

4. Inferencia Estadística	43
4.1. Intervalos de Confianza para Retornos Medios	43
4.2. Pruebas de Normalidad	45
4.3. Pruebas t de Una Muestra	47
4.4. Comparación de Grupos: Beta Bajo vs Alto	48
4.5. Comparación de Volatilidades: Test de Levene	48
4.6. Bootstrap para Diferencia de Medias	49
4.7. Análisis de Potencia Estadística	50
4.8. Comparación de Betas: QQQ vs Mercado Equal-Weighted	51
4.9. Síntesis del Análisis de Inferencia	53
5. Resultados y Conclusiones	54
5.1. Respuestas a las Preguntas de Investigación	54
5.2. Síntesis de Resultados Principales	57
5.3. Limitaciones del Estudio	57
5.4. Extensiones y Trabajo Futuro	58
Referencias Bibliográficas	59

Resumen Ejecutivo

Visión General del Estudio

Este informe presenta un análisis estadístico exhaustivo de **30 empresas del sector tecnológico** pertenecientes al índice NASDAQ-100 durante el período **febrero 2018 – diciembre 2024**. El estudio comprende **2,403 observaciones mensuales** y abarca desde la exploración inicial de los datos hasta análisis de riesgo, correlación y cambio estructural asociado a la pandemia de COVID-19.

Principales Hallazgos

- ✓ El 73.3 % de empresas presenta **colas pesadas** (leptocurtosis)
- ✓ Correlación promedio de $\rho = 0,40$, máxima de $\rho = 0,72$
- ✓ El 10.7 % de empresas muestra **cambio estructural** post-COVID
- ✓ Tesla y Nvidia con MDD superior al 60 %

Implicaciones Prácticas

- ▲ Modelos gaussianos **subestiman** riesgos de cola
- ▲ Diversificación intra-sectorial tiene **beneficio limitado**
- ▲ Gestión de riesgo debe incorporar **métricas de drawdown**
- ▲ Análisis de régimen es **esencial** en contextos de crisis

Introducción

Contexto y Motivación

El sector tecnológico ha experimentado una transformación significativa durante el período 2018–2024, caracterizado por una creciente digitalización global, la pandemia de COVID-19, y fluctuaciones macroeconómicas que impactaron profundamente los mercados financieros. Comprender el comportamiento de las empresas tecnológicas del NASDAQ-100 es crucial tanto para inversores institucionales como para académicos interesados en la dinámica de los mercados de capitales.

Este estudio se enmarca en la tradición del análisis cuantitativo financiero, siguiendo los principios metodológicos establecidos por **Campbell, Lo & MacKinlay** en *The Econometrics of Financial Markets* (1997) y las prácticas de análisis exploratorio formalizadas por **Tukey** en *Exploratory Data Analysis* (1977).

Preguntas de Investigación

El presente análisis busca responder las siguientes preguntas de investigación, formuladas para abordar aspectos fundamentales del comportamiento financiero del sector tecnológico:

Preguntas de Investigación

P1. ¿Cuál es la estructura de riesgo-retorno de las empresas tecnológicas del NASDAQ-100 y cómo se comparan entre sí?

Objetivo: Caracterizar la distribución de retornos, volatilidades y métricas de riesgo (drawdown, VaR implícito) para identificar empresas defensivas vs agresivas.

P2. ¿Existen grupos naturales de empresas con perfiles de riesgo similares que permitan una segmentación objetiva del sector?

Objetivo: Aplicar técnicas de clustering (K-Means) para identificar segmentos homogéneos y validar estadísticamente su separación mediante ANOVA y métricas de silueta.

P3. ¿Cuál es la sensibilidad de cada empresa al riesgo sistemático del mercado (beta) y qué factores explican la heterogeneidad observada?

Objetivo: Estimar el modelo CAPM individual para cada empresa, analizar la distribución de betas, y explorar la relación entre volatilidad y correlación con el mercado.

P4. ¿Los retornos medios de las empresas son estadísticamente distintos de cero, y difieren significativamente entre grupos de alto y bajo riesgo sistemático?

Objetivo: Aplicar pruebas de hipótesis (t-tests, bootstrap, permutación) con correcciones por comparaciones múltiples para evaluar la significancia de los retornos.

Fuente de Datos

Los datos fueron obtenidos mediante la API de **Yahoo Finance** (biblioteca `yfinance v1.0` para Python), fuente ampliamente utilizada en investigación financiera y documentada en múltiples repositorios académicos de GitHub (ver referencias). El conjunto de datos comprende:

- **Panel mensual:** 2,403 observaciones (30 empresas × hasta 83 meses)
- **Variables:** Precio ajustado (Adj Close), Volumen, Retorno logarítmico mensual
- **Período:** Febrero 2018 – Diciembre 2024
- **Benchmark de mercado:** QQQ (Invesco QQQ Trust, ETF del NASDAQ-100)

1 Análisis Exploratorio de Datos (EDA)

◇ Concepto Clave

El **Análisis Exploratorio de Datos** (EDA, *Exploratory Data Analysis*) constituye el pilar fundamental de cualquier investigación estadística rigurosa. Introducido por John W. Tukey en su obra seminal *Exploratory Data Analysis* (Addison-Wesley, 1977), este enfoque metodológico prioriza la comprensión profunda de los datos antes de aplicar modelos formales.

“El análisis exploratorio es trabajo detectivesco numérico — o conteo, o representación gráfica. [...] Los gráficos son esenciales.”

— J. W. Tukey, *Exploratory Data Analysis*, p. 1

En el contexto del análisis financiero cuantitativo, el EDA cumple funciones críticas:

1. **Validación de integridad:** Identificación de valores faltantes, inconsistencias y errores de captura que podrían comprometer análisis posteriores.
2. **Caracterización distribucional:** Evaluación de momentos estadísticos, detección de asimetrías y cuantificación de comportamientos extremos.
3. **Descubrimiento de patrones:** Revelación de estructuras de correlación, dependencias temporales y agrupamientos naturales.
4. **Verificación de supuestos:** Contraste empírico de hipótesis requeridas por modelos paramétricos (normalidad, homocedasticidad, estacionariedad).

1.1 Estructura del Conjunto de Datos

El estudio se fundamenta en un **panel de datos** (datos longitudinales) que combina la dimensión temporal con la transversal, permitiendo capturar tanto la evolución individual de cada empresa como las dinámicas comunes del sector.

Panel de Datos

Un **panel de datos** (también denominado datos longitudinales) es una estructura que combina:

- **Dimensión transversal:** Múltiples entidades (empresas, individuos, países)
- **Dimensión temporal:** Observaciones repetidas en el tiempo

Formalmente, para N entidades observadas en T períodos, el panel contiene $N \times$

T observaciones potenciales. Cuando todas las combinaciones están presentes, se denomina **panel balanceado**.

Cuadro 1: Estructura del Panel de Datos Analizado

Característica	Valor	Descripción
Empresas (N)	30	Tecnológicas NASDAQ-100
Período	Feb 2018 – Dic 2024	Aproximadamente 7 años
Frecuencia	Mensual	Último día hábil de cada mes
Observaciones	2,403	Panel desbalanceado (4 empresas con IPO posterior)

Variables del Panel:		
Date	datetime	Fecha de observación
Company	string	Identificador de empresa
AdjClose	float (USD)	Precio ajustado por dividendos/splits
Volume	int	Volumen mensual negociado
Return	float	Retorno logarítmico mensual

Retorno Logarítmico

El retorno logarítmico (también denominado retorno continuo) se define como:

$$r_t = \ln\left(\frac{P_t}{P_{t-1}}\right) = \ln(P_t) - \ln(P_{t-1}) \quad (1)$$

donde P_t representa el precio ajustado en el período t . Esta transformación presenta ventajas sobre los retornos simples:

- **Aditividad temporal:** $r_{t \rightarrow t+n} = \sum_{i=0}^{n-1} r_{t+i}$
- **Simetría:** Movimientos equivalentes producen retornos simétricos
- **Propiedades estadísticas:** Distribución más próxima a la normal

1.2 Análisis de Valores Faltantes

La presencia de datos ausentes constituye una problemática ubicua en el análisis empírico. Su tratamiento inadecuado puede introducir sesgos sistemáticos y comprometer la validez de las inferencias estadísticas.

Taxonomía de Datos Faltantes

La literatura estadística, fundamentada en el trabajo de Donald B. Rubin (*Biometrika*, 1976), distingue tres mecanismos generadores de datos ausentes:

- MCAR** (*Missing Completely At Random*): La probabilidad de ausencia es independiente de cualquier variable. Representa el escenario más favorable.

$$P(M|Y_{obs}, Y_{miss}) = P(M)$$

- MAR** (*Missing At Random*): La probabilidad depende únicamente de variables observadas.

$$P(M|Y_{obs}, Y_{miss}) = P(M|Y_{obs})$$

- MNAR** (*Missing Not At Random*): La probabilidad depende del propio valor no observado. Genera sesgo difícil de corregir.

En mercados financieros, las causas típicas incluyen: feriados bursátiles, suspensiones de cotización, IPOs posteriores al inicio del período.

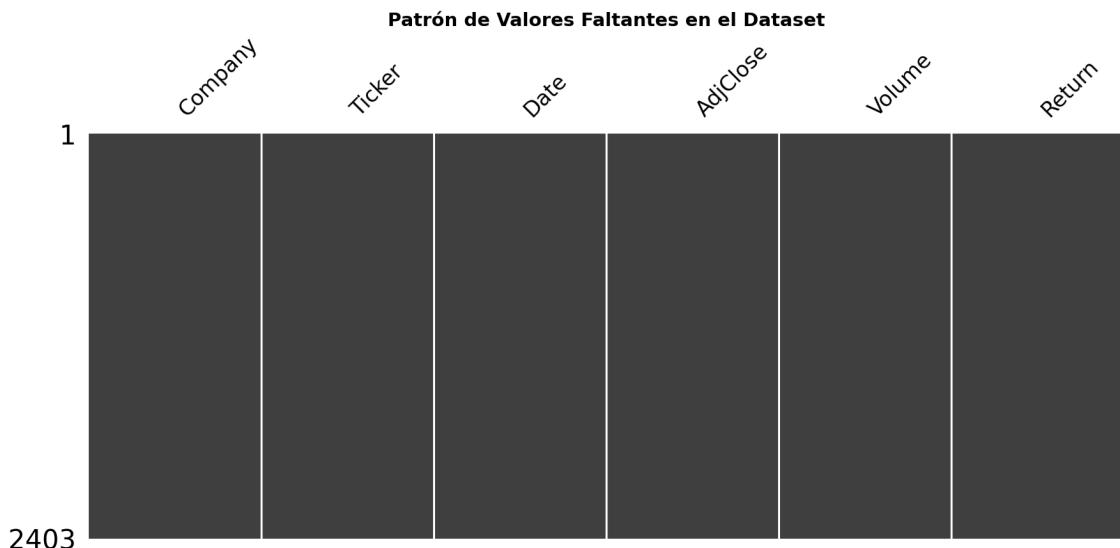


Figura 1: Matriz de completitud del dataset. Cada fila representa una observación temporal y cada columna una variable. Las áreas sólidas indican datos presentes; franjas blancas revelarían patrones de ausencia. El dataset analizado presenta alta completitud.

◎ Resultado

El análisis de valores faltantes del panel revela:

- Todas las columnas:** 0 % de datos faltantes (Company, Ticker, Date, AdjClose, Volume, Return)
- Empresas con menos de 83 observaciones:** Palantir (51), Snowflake (51), Cloudflare (63), Spotify (80)
- Causa:** IPOs posteriores al inicio del período de análisis (no datos faltantes, sino inexistentes)

El dataset presenta **calidad excelente** para análisis sin necesidad de imputación.

1.3 Distribución de Retornos Financieros

La caracterización distribucional de los retornos constituye un pilar fundamental del análisis financiero cuantitativo. Las propiedades de esta distribución determinan la validez de modelos de valoración y métricas de riesgo.

Hechos Estilizados de Retornos Financieros

La investigación empírica, documentada extensamente por Cont (*Quantitative Finance*, 2001) y Campbell, Lo & MacKinlay (*The Econometrics of Financial Markets*, Princeton, 1997), ha identificado regularidades que distinguen los retornos financieros de la distribución normal:

1. **Colas pesadas** (*fat tails / leptocurtosis*): Eventos extremos de $\pm 4\sigma$ o más ocurren con frecuencia significativamente mayor a la predicha por la normal.
2. **Asimetría negativa**: Las caídas abruptas tienden a ser más pronunciadas que las subidas equivalentes, especialmente durante episodios de estrés.
3. **Agrupamiento de volatilidad** (*volatility clustering*): Períodos de alta volatilidad tienden a persistir. Formalizado en modelos ARCH/GARCH (Engle, 1982; Bollerslev, 1986).
4. **Efecto apalancamiento**: Correlación negativa entre retornos y volatilidad futura; las caídas incrementan la volatilidad más que subidas equivalentes.

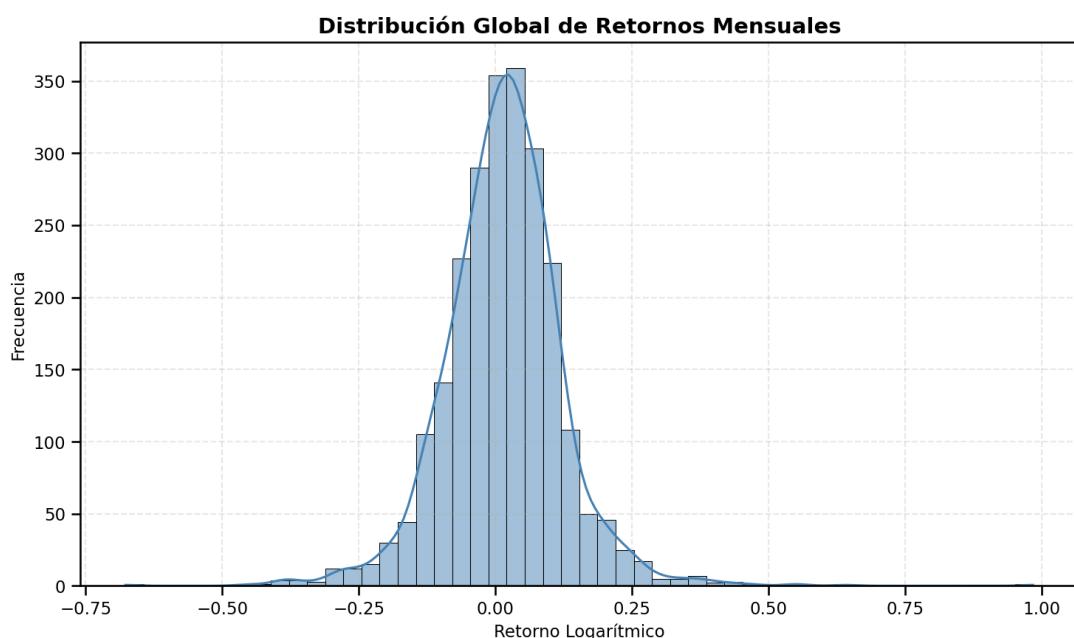


Figura 2: Distribución empírica de retornos mensuales del panel. El histograma representa la frecuencia observada; la curva KDE (*Kernel Density Estimation*) suaviza la distribución. Se observa la característica forma leptocúrtica con colas más pesadas que la normal.

Nota:

La distribución agregada de los 2,403 retornos mensuales presenta las siguientes características:

- **Media:** $\mu = 1,53\%$ mensual
- **Volatilidad:** $\sigma = 10,84\%$ mensual
- **Asimetría global:** $\gamma_1 = 0,28$ (sesgo positivo leve)
- **Curtosis global:** $\gamma_2 = 5,43$ (leptocurtosis pronunciada)

La curtosis de 5.43 implica colas significativamente más pesadas que la distribución normal ($\gamma_2 = 0$), confirmando que los modelos gaussianos subestiman la probabilidad de eventos extremos.

Estimación de Densidad Kernel (KDE)

La KDE es un método no paramétrico para estimar la función de densidad de probabilidad:

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right) \quad (2)$$

donde $K(\cdot)$ es una función kernel (típicamente gaussiana) y h es el parámetro de ancho de banda (*bandwidth*). A diferencia del histograma, la KDE produce estimaciones continuas y diferenciables.

1.3.1 Distribución por Empresa: Análisis Comparativo

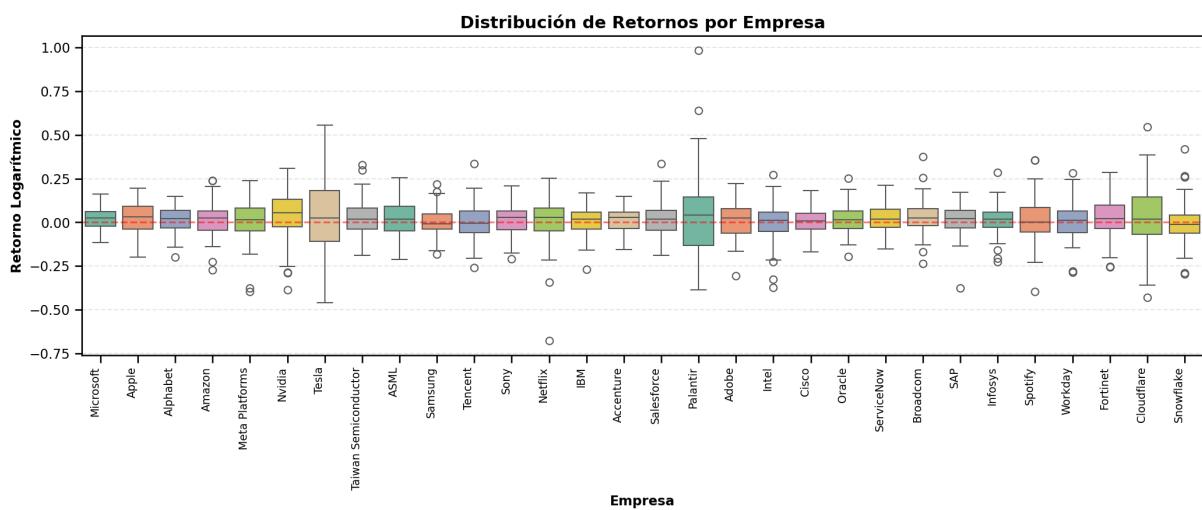


Figura 3: Diagramas de caja (boxplots) de retornos logarítmicos mensuales por empresa. Cada boxplot sintetiza: mediana (línea central), cuartiles (bordes), rango IQR extendido (bigotes) y outliers (puntos). La línea de referencia en cero facilita identificar sesgo.

Anatomía del Diagrama de Caja (Tukey, 1977)

El boxplot codifica visualmente la distribución mediante:

$$\text{Mediana} = Q_2 = \text{Percentil}_{50} \quad (3)$$

$$\text{Rango Intercuartílico} : IQR = Q_3 - Q_1 \quad (4)$$

$$\text{Bigote inferior} = Q_1 - 1,5 \times IQR \quad (5)$$

$$\text{Bigote superior} = Q_3 + 1,5 \times IQR \quad (6)$$

$$\text{Outliers} : x \notin [Q_1 - 1,5 \cdot IQR, Q_3 + 1,5 \cdot IQR] \quad (7)$$

El factor 1.5 corresponde a la propuesta original de Tukey; para outliers extremos se utiliza factor 3.0.

1.3.2 Volatilidad por Empresa

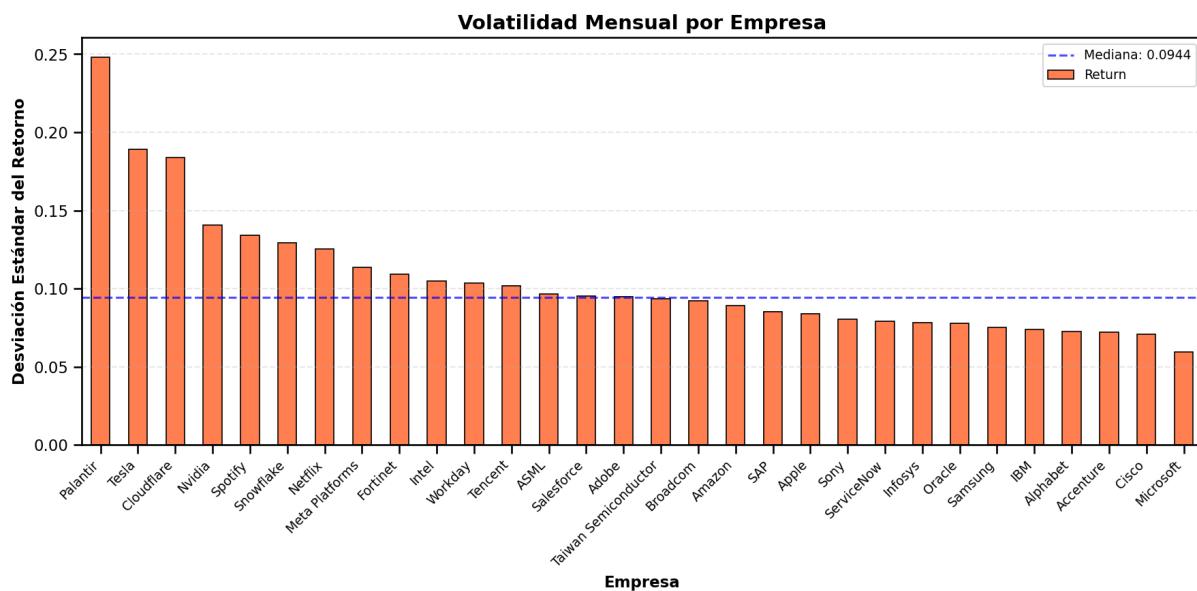


Figura 4: Volatilidad mensual (desviación estándar de retornos) por empresa, ordenada de mayor a menor. La línea horizontal indica la mediana sectorial. Empresas como Tesla y Nvidia exhiben volatilidad significativamente superior al promedio.

Volatilidad Histórica

La volatilidad representa la dispersión de los retornos y constituye la métrica de riesgo más utilizada en finanzas. Según Hull (*Options, Futures, and Other Derivatives*, Pearson, 2018), la volatilidad histórica se estima como:

$$\sigma = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (r_i - \bar{r})^2} \quad (8)$$

Esta métrica está fundamentada en la Teoría Moderna de Portafolios, que asume inversores aversos al riesgo que prefieren menor variabilidad para un mismo retorno esperado.

Limitaciones de la Volatilidad como Métrica de Riesgo

La desviación estándar presenta deficiencias importantes:

- **Simetría implícita:** Penaliza igualmente ganancias y pérdidas inesperadas
- **Supuesto de normalidad:** Subestima eventos extremos en distribuciones leptocúrticas
- **Mirada retrospectiva:** Volatilidad histórica no predice necesariamente la futura

Métricas complementarias: Value at Risk (VaR), Expected Shortfall (CVaR), Maximum Drawdown.

1.4 Relación Riesgo-Retorno

La relación entre riesgo y retorno esperado constituye el axioma central de las finanzas modernas.

Frontera Eficiente

Harry M. Markowitz formalizó el problema de selección de portafolios como optimización media-varianza en su artículo “*Portfolio Selection*” (*The Journal of Finance*, Vol. 7, No. 1, 1952, pp. 77–91), trabajo por el cual recibió el Premio Nobel de Economía en 1990. La **frontera eficiente** representa el conjunto de portafolios que maximizan retorno para cada nivel de riesgo.

Para un portafolio de dos activos:

$$\sigma_p^2 = w_1^2 \sigma_1^2 + w_2^2 \sigma_2^2 + 2w_1 w_2 \sigma_1 \sigma_2 \rho_{12} \quad (9)$$

El término $2w_1 w_2 \sigma_1 \sigma_2 \rho_{12}$ es la fuente del beneficio de diversificación cuando $\rho_{12} < 1$.

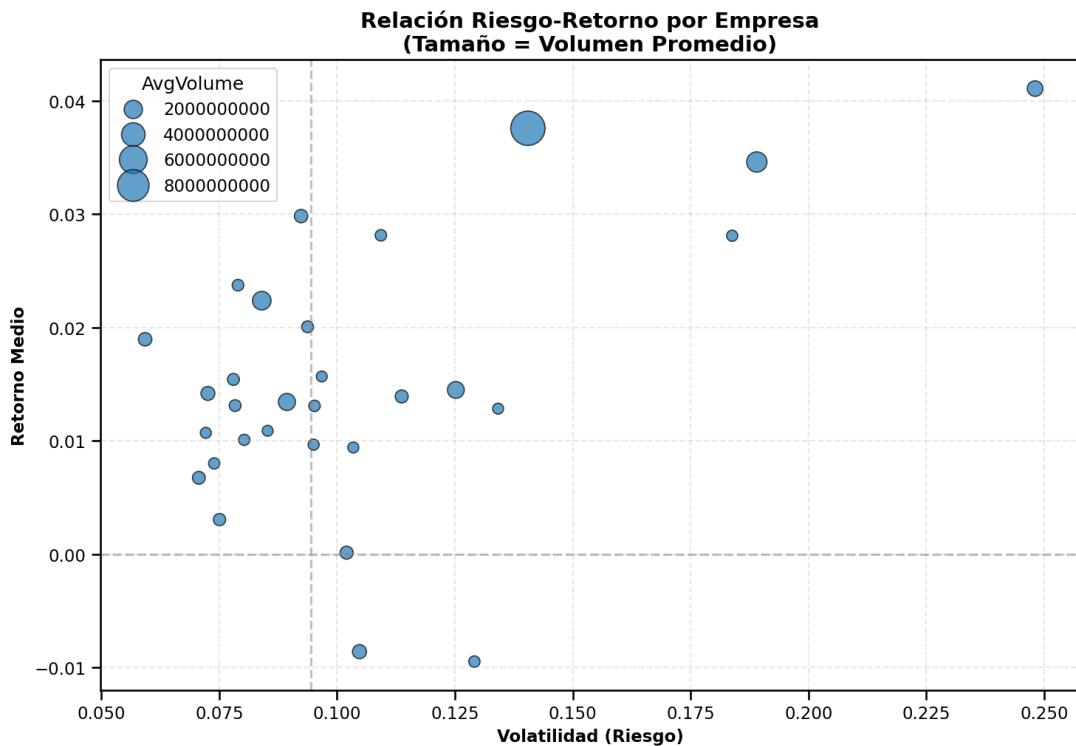
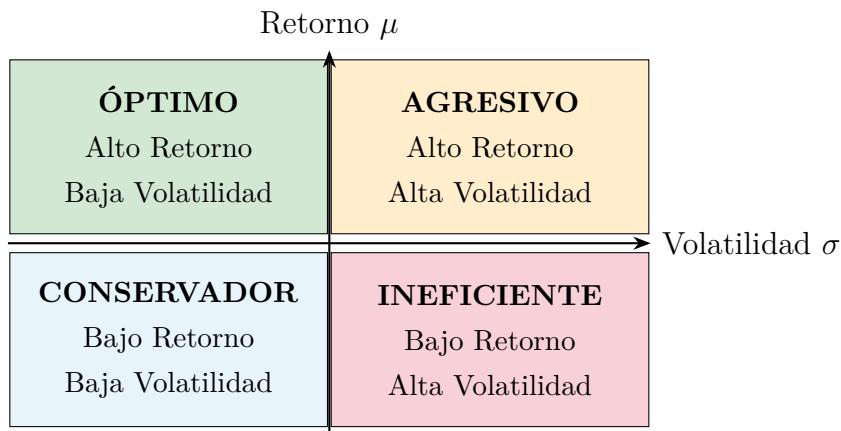


Figura 5: Diagrama de dispersión riesgo-retorno. Cada punto representa una empresa; el tamaño es proporcional al volumen promedio (proxy de liquidez). Las líneas punteadas segmentan el espacio según mediana de volatilidad y retorno cero.



Ratio de Sharpe

El ratio de Sharpe cuantifica el exceso de retorno por unidad de riesgo:

$$\text{Sharpe Ratio} = \frac{E[R_p] - R_f}{\sigma_p} \quad (10)$$

donde R_f es la tasa libre de riesgo. Valores $> 1,0$ se consideran atractivos; $> 2,0$, excelentes.

1.5 Evolución Temporal de Precios

Precio Ajustado (*Adjusted Close*)

El precio ajustado incorpora correcciones retroactivas por eventos corporativos:

- **Dividendos:** El precio ex-dividendo cae por el monto distribuido; el ajuste redistribuye hacia atrás
- **Splits:** Una división 2:1 reduce el precio a la mitad; los históricos se ajustan proporcionalmente
- **Derechos y spin-offs:** Otras acciones que modifican acciones en circulación

El uso del precio ajustado es **imperativo** para calcular retornos totales reales.

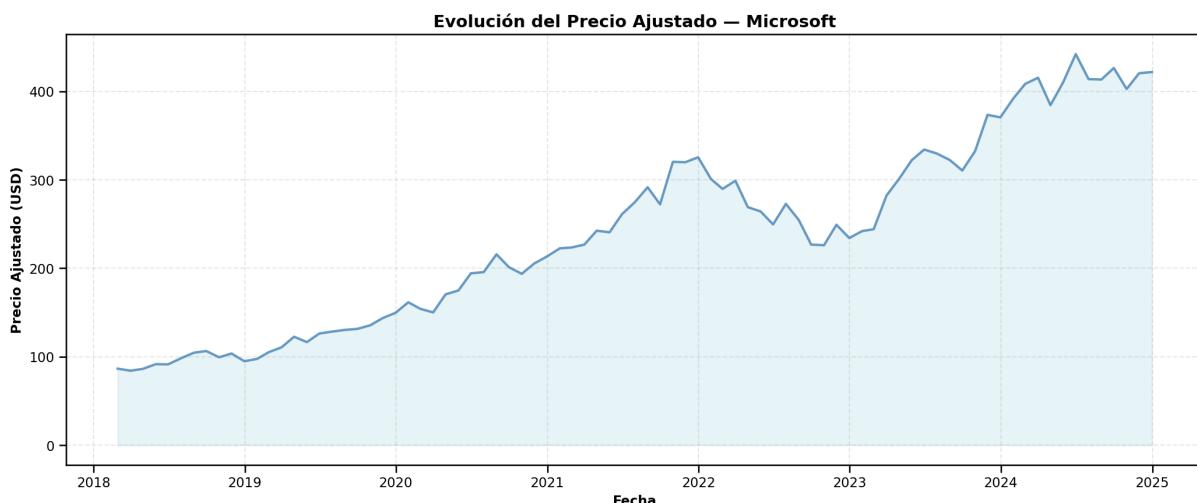


Figura 6: Evolución del precio ajustado de Microsoft (MSFT) 2018–2024. El área sombreada facilita visualización de tendencia. Se observan correcciones durante COVID-19 crash (marzo 2020) y tech sell-off (2022).

1.5.1 Series Normalizadas: Comparación Multi-Activo

Normalización Base 100

Para comparar activos con precios en escalas dispares:

$$P_t^{norm} = \frac{P_t}{P_0} \times 100 \quad (11)$$

Interpretación:

- $P_t^{norm} = 150 \Rightarrow$ Rendimiento acumulado +50 %
- $P_t^{norm} = 75 \Rightarrow$ Pérdida acumulada -25 %

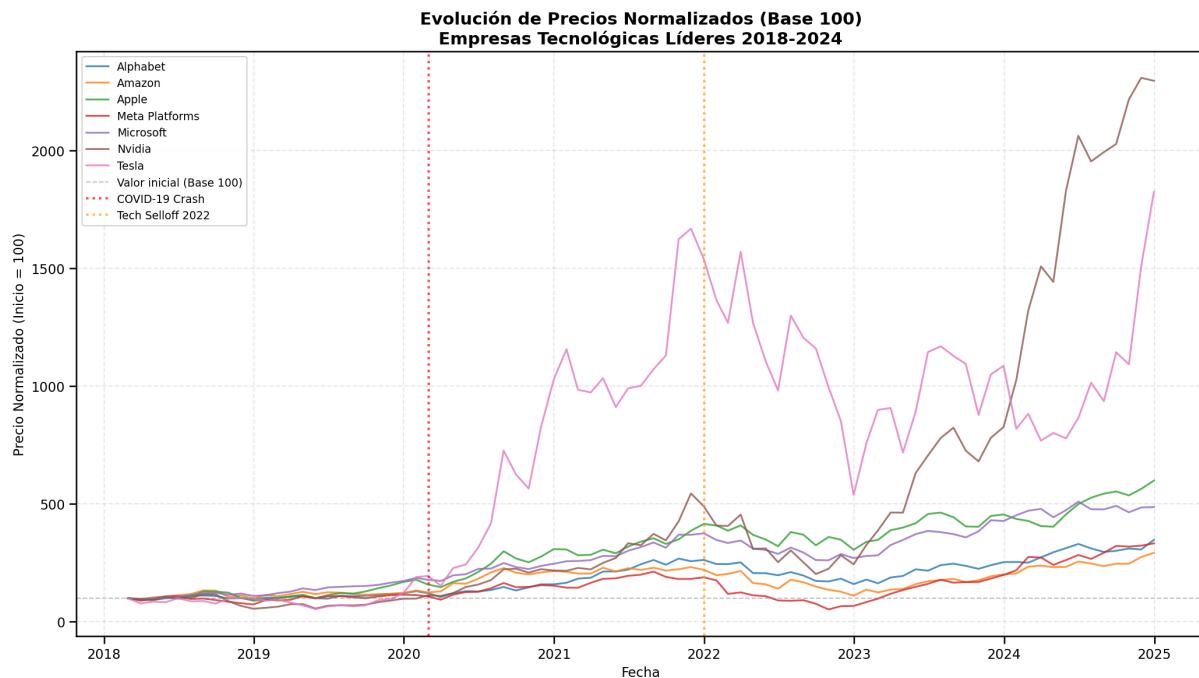


Figura 7: Series de precios normalizadas (Base 100) para principales empresas tecnológicas. Líneas verticales marcan COVID-19 crash (marzo 2020) y Tech Selloff (2022). La divergencia revela heterogeneidad de desempeño.

◎ Resultado

El análisis de rendimiento total durante el período Feb 2018 – Dic 2024 (6.92 años) revela:

Empresa	Rendimiento Total	CAGR
Nvidia	+2,196.0 %	57.3 %
Tesla	+1,725.1 %	52.2 %
Apple	+499.4 %	29.6 %
Microsoft	+386.7 %	25.7 %
Alphabet	+247.7 %	19.7 %
Meta Platforms	+232.8 %	19.0 %
Amazon	+192.6 %	16.8 %

Nvidia y Tesla destacan con rendimientos excepcionales superiores a 1,700 %, impulsados por el auge de IA y vehículos eléctricos.

1.6 Análisis de Correlación

Diversificación y Correlación

El principio de diversificación establece que combinar activos imperfectamente correlacionados reduce el riesgo sin sacrificar retorno:

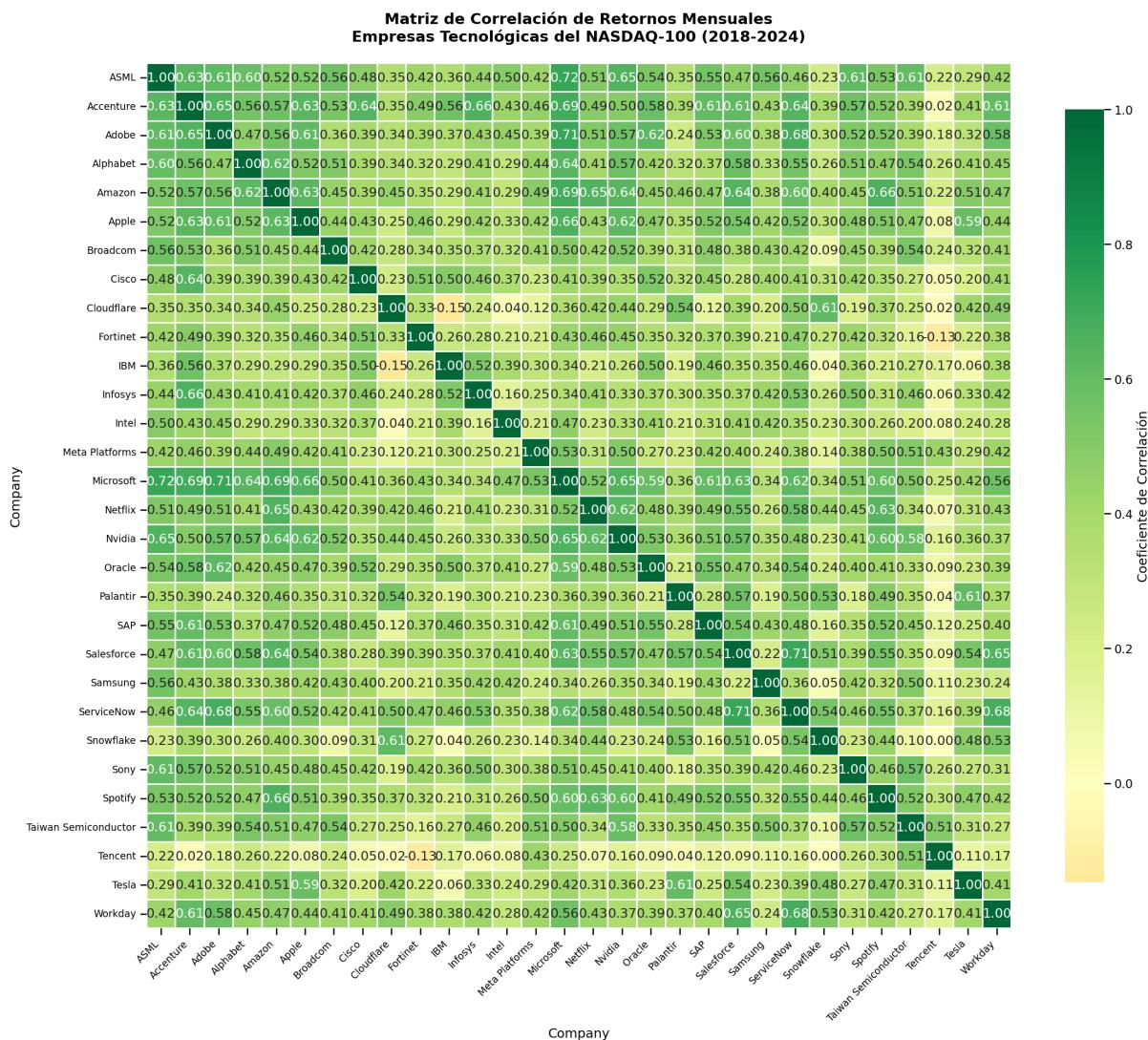
$$\sigma_p^2 = \sum_i w_i^2 \sigma_i^2 + \sum_i \sum_{j \neq i} w_i w_j \sigma_i \sigma_j \rho_{ij} \quad (12)$$

El beneficio de diversificación surge cuando $\rho_{ij} < 1$. Máximo beneficio con $\rho < 0$.

Coeficiente de Correlación de Pearson

$$\rho_{ij} = \frac{\text{Cov}(R_i, R_j)}{\sigma_i \cdot \sigma_j} = \frac{\sum_{t=1}^T (R_{i,t} - \bar{R}_i)(R_{j,t} - \bar{R}_j)}{\sqrt{\sum_t (R_{i,t} - \bar{R}_i)^2} \cdot \sqrt{\sum_t (R_{j,t} - \bar{R}_j)^2}} \quad (13)$$

El coeficiente $\rho \in [-1, 1]$ mide intensidad y dirección de relación lineal.



Resultado

El análisis de correlación entre las 30 empresas (435 pares únicos) reveló:

Pares con mayor correlación:

- ASML – Microsoft: $\rho = 0,721$
- Salesforce – ServiceNow: $\rho = 0,711$
- Adobe – Microsoft: $\rho = 0,709$
- Amazon – Microsoft: $\rho = 0,690$
- Accenture – Microsoft: $\rho = 0,688$

Pares con menor correlación:

- Cloudflare – IBM: $\rho = -0,147$
- Fortinet – Tencent: $\rho = -0,125$

3. Snowflake – Tencent: $\rho = 0,001$
4. Cloudflare – Tencent: $\rho = 0,021$
5. Accenture – Tencent: $\rho = 0,024$

Distribución: Alta ($\rho \geq 0,7$): 3 pares (0.7 %) | Media ($0,4 \leq \rho < 0,7$): 227 pares (52.2 %) | Baja ($\rho < 0,4$): 205 pares (47.1 %)

Correlación Intra-Sectorial

El análisis de correlación del panel tecnológico revela:

- **Correlación promedio:** $\rho = 0,40$ (excluyendo diagonal)
- **Rango:** desde $\rho = -0,15$ (Cloudflare–IBM) hasta $\rho = 0,72$ (ASML–Microsoft)
- **Clasificación:** Solo 3 pares (0.7 %) con correlación alta ($\rho \geq 0,7$)

Esta correlación moderada permite cierto beneficio de diversificación intra-sectorial, aunque limitado respecto a diversificación entre sectores.

1.7 Detección de Valores Atípicos (Outliers)

Métodos de Detección de Outliers

Método IQR (Tukey, 1977):

$$\text{Outlier si: } x < Q_1 - 1,5 \times IQR \quad \vee \quad x > Q_3 + 1,5 \times IQR \quad (14)$$

Método Z-score:

$$z_i = \frac{x_i - \mu}{\sigma}, \quad \text{Outlier si } |z_i| > 3 \quad (15)$$

El método IQR es más robusto (usa cuartiles); Z-score es sensible a outliers en media/std.

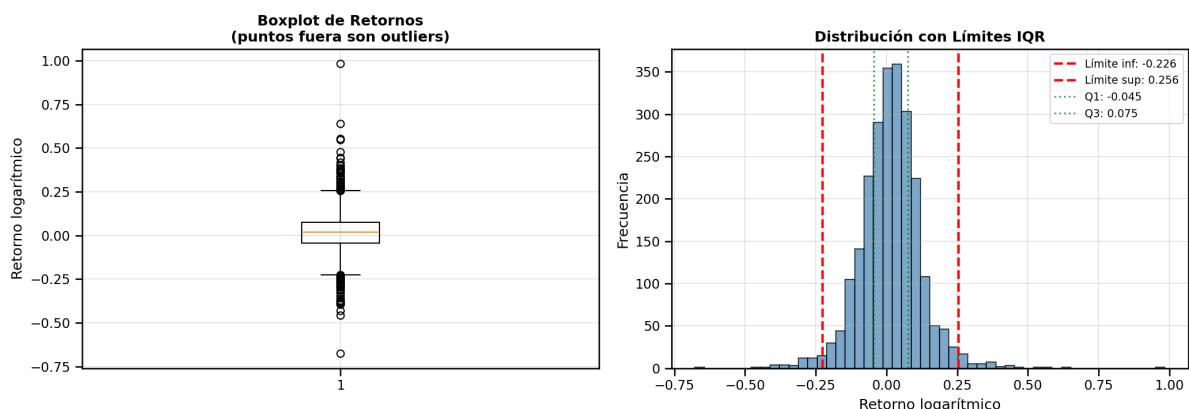


Figura 9: Análisis de outliers. Izquierda: Boxplot con outliers como puntos individuales. Derecha: Histograma con límites IQR superpuestos delimitando región típica.

⑤ Resultado

La detección de outliers en los retornos del panel reveló:

- **Método IQR:** 84 outliers detectados (3.50 % de las observaciones)
- **Método Z-score ($|z| > 3$):** 33 outliers detectados (1.37%)
- **Límites IQR:** $Q_1 = -4,53\%$, $Q_3 = +7,51\%$, $IQR = 12,04\%$
- **Rango válido:** $[-22,59\%, +25,57\%]$

Outliers más extremos positivos: Palantir $+98.4\%$ (Nov 2020), Palantir $+64.1\%$ (May 2023), Tesla $+55.5\%$ (Ago 2020).

Outliers más extremos negativos: Netflix $-67,7\%$ (Abr 2022), Tesla $-45,8\%$ (Dic 2022), Cloudflare $-43,1\%$ (May 2022).

Tratamiento de Outliers Financieros

Los outliers financieros raramente deben eliminarse. Representan frecuentemente:

- Eventos legítimos (crashes, rallies, anuncios corporativos)
- Información valiosa sobre riesgo de cola (*tail risk*)
- Señales de cambio de régimen

Eliminarlos subestimaría el riesgo real.

1.8 Análisis de Momentos Estadísticos

Los Cuatro Momentos de una Distribución

1. **Media** (Primer momento): $\mu = E[R]$
2. **Varianza** (Segundo momento central): $\sigma^2 = E[(R - \mu)^2]$
3. **Asimetría / Skewness** (Tercer momento estandarizado):

$$\gamma_1 = E \left[\left(\frac{R - \mu}{\sigma} \right)^3 \right] \quad (16)$$

$\gamma_1 < 0$: Sesgo negativo (cola izquierda más larga)

4. **Curtosis** (Cuarto momento estandarizado):

$$\gamma_2 = E \left[\left(\frac{R - \mu}{\sigma} \right)^4 \right] - 3 \quad (17)$$

$\gamma_2 > 0$: Leptocúrtica (colas pesadas, eventos extremos frecuentes)

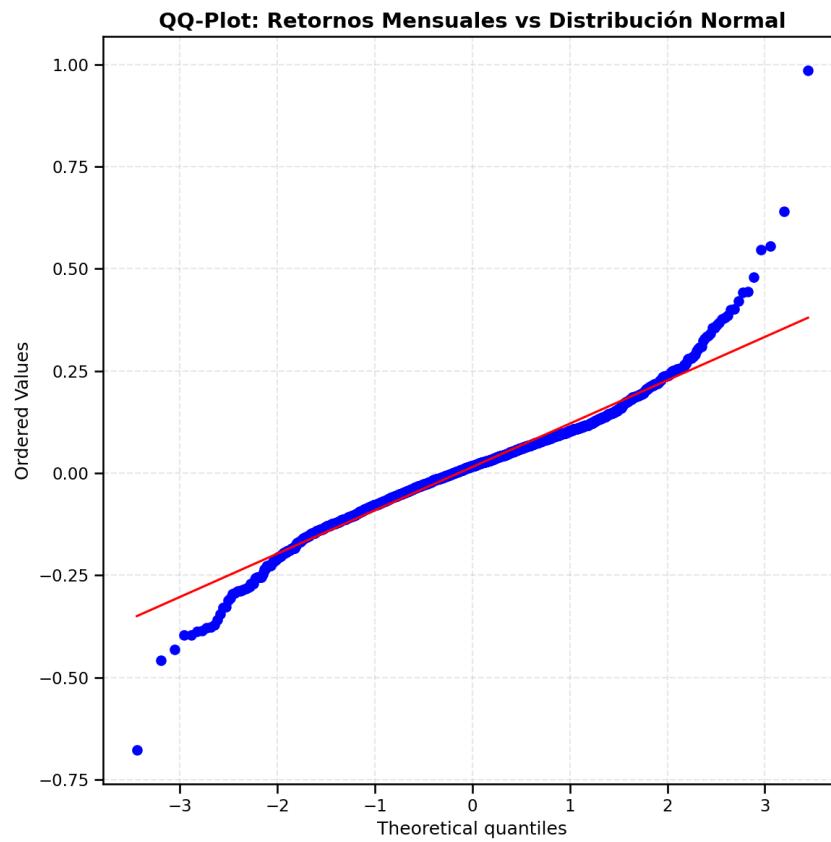


Figura 10: Gráfico Q-Q comparando cuantiles empíricos con normal teórica. Puntos sobre diagonal indicarían normalidad. Desviaciones en colas revelan leptocurtosis característica.

Test de Jarque-Bera

Propuesto por Carlos M. Jarque y Anil K. Bera (*International Statistical Review*, 1987), este test evalúa conjuntamente asimetría y curtosis para contrastar normalidad:

$$JB = \frac{n}{6} \left(\gamma_1^2 + \frac{\gamma_2^2}{4} \right) \quad (18)$$

Bajo H_0 (normalidad), $JB \sim \chi^2(2)$. Si el p-valor < 0,05, se rechaza la hipótesis de normalidad al 5 % de significancia.

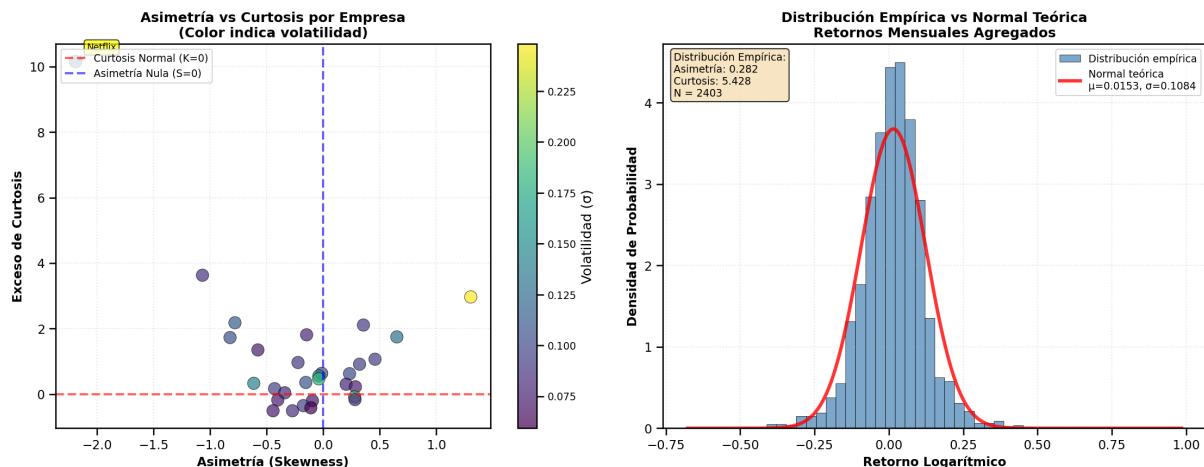


Figura 11: Análisis distribucional. Evaluación de asimetría, curtosis y desviación respecto a normalidad para el conjunto de empresas.

✓ Hallazgo Clave

El análisis de momentos estadísticos del panel tecnológico revela:

- **Asimetría promedio:** $\gamma_1 = -0,15$ (sesgo negativo leve)
- **Curtosis promedio:** $\gamma_2 = 1,06$ (exceso de curtosis, colas pesadas)
- **Leptocurtosis:** 22 de 30 empresas (73.3 %) presentan $\gamma_2 > 0$
- **Normalidad:** 10 de 30 empresas (33.3 %) rechazan normalidad (Test Jarque-Bera, $\alpha = 0,05$)

La empresa con mayor curtosis es **Netflix** ($\gamma_2 = 10,16$), seguida de SAP ($\gamma_2 = 3,63$) y Palantir ($\gamma_2 = 2,97$). Las empresas que rechazan normalidad son: Meta, Taiwan Semiconductor, Netflix, IBM, Palantir, Intel, Broadcom, SAP, Infosys y Snowflake.

1.9 Análisis de Drawdown

Drawdown y Maximum Drawdown

El drawdown cuantifica la caída desde máximos históricos:

$$DD_t = \frac{P_t - \max_{s \leq t} P_s}{\max_{s \leq t} P_s} \times 100 \% \quad (19)$$

El **Maximum Drawdown (MDD)** es la peor caída observada:

$$MDD = \min_t DD_t \quad (20)$$

Propiedades: $DD_t \leq 0$ siempre; $DD_t = 0$ en máximo histórico.

Asimetría de la Recuperación

Las pérdidas requieren ganancias proporcionalmente mayores para recuperarse:

Pérdida	Ganancia para recuperar
-10 %	+11.1 %
-25 %	+33.3 %
-50 %	+100 %
-75 %	+300 %

Por esto la gestión del riesgo de caída es más importante que maximizar retorno.

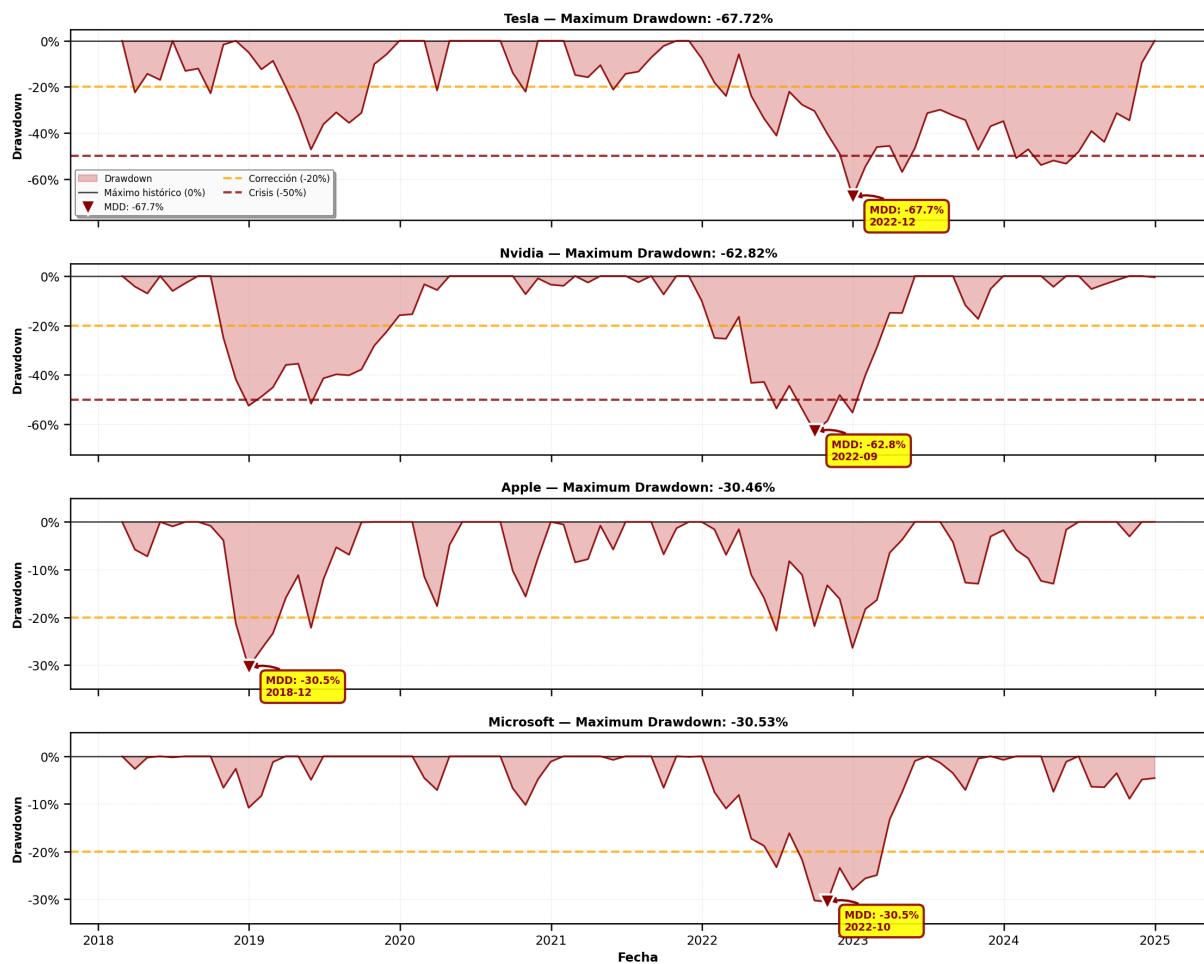


Figura 12: Series de drawdown para empresas seleccionadas (Tesla, Nvidia, Apple, Microsoft). Área roja: magnitud del drawdown. Líneas horizontales: umbrales -20% (corrección) y -50% (crisis).

◎ Resultado

El análisis de Maximum Drawdown (MDD) para las empresas principales reveló:

Empresa	MDD	Tiempo bajo agua	Retorno para recuperar
Tesla	-67,72 %	67 meses (80.7 %)	+209.8 %
Nvidia	-62,82 %	54 meses (65.1 %)	+169.0 %
Microsoft	-30,53 %	48 meses (57.8 %)	+43.9 %
Apple	-30,46 %	56 meses (67.5 %)	+43.8 %

Las empresas de alto crecimiento (Tesla, Nvidia) experimentaron caídas superiores al 60%, mientras que las empresas maduras (Apple, Microsoft) limitaron sus pérdidas máximas a aproximadamente 30%.

Calmar Ratio

Mide eficiencia del retorno ajustada por riesgo de pérdida máxima:

$$\text{Calmar Ratio} = \frac{CAGR}{|MDD|} \quad (21)$$

donde $CAGR$ es Tasa de Crecimiento Anual Compuesta. Ratio > 1: excelente.

1.10 Análisis de Cambio Estructural: Pre/Post COVID-19

Análisis de Regímenes

La pandemia afectó los mercados tecnológicos a través de múltiples canales:

1. **Shock inicial:** Caída abrupta marzo 2020 ante parálisis económica
 2. **Transformación digital:** Trabajo remoto, e-commerce, cloud beneficiaron al sector
 3. **Política monetaria:** Tasas cercanas a cero impulsaron valuaciones
 4. **Reversión 2022:** Normalización monetaria y corrección de excesos
- División temporal: **Pre-COVID** (Ene 2018 – Feb 2020) vs **Post-COVID** (Mar 2020 – Dic 2024)

Tests de Comparación

Test t de Welch (medias con varianzas desiguales):

$$t = \frac{\bar{R}_{post} - \bar{R}_{pre}}{\sqrt{\frac{s_{post}^2}{n_{post}} + \frac{s_{pre}^2}{n_{pre}}}} \quad (22)$$

Test de Levene: Homogeneidad de varianzas. $H_0: \sigma_{pre}^2 = \sigma_{post}^2$

P-valor < 0,05 indica cambio estadísticamente significativo.

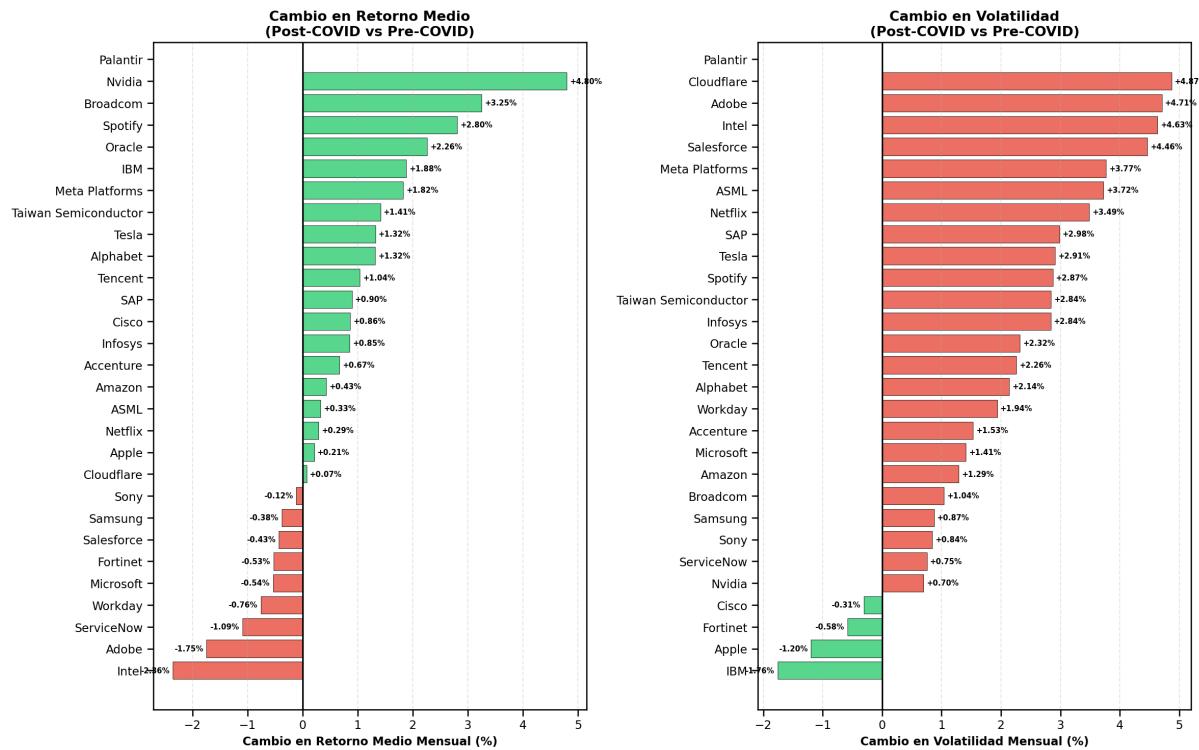


Figura 13: Cambios en retorno medio y volatilidad (Post-COVID vs Pre-COVID). Verde: mejora; Rojo: deterioro.

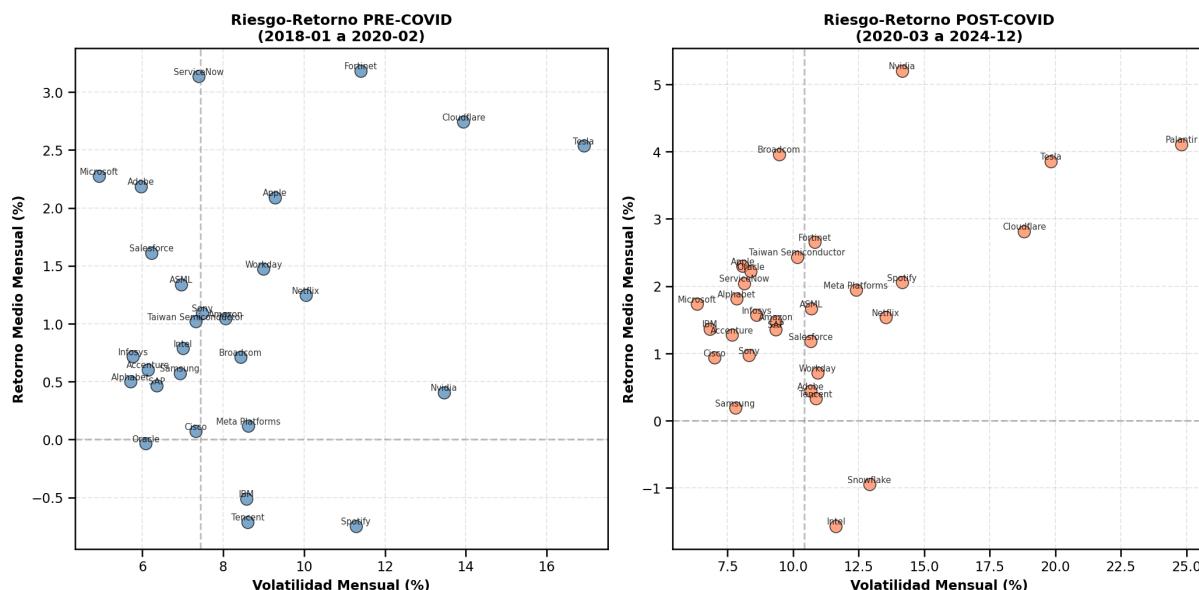


Figura 14: Diagrama riesgo-retorno segmentado por período. Visualización de migración en el espacio riesgo-retorno tras shock pandémico.

✓ Hallazgo Clave

El análisis de cambio estructural pre/post COVID-19 revela:

- **Retorno medio:** 19 de 30 empresas (63.3 %) mejoraron post-COVID

- **Volatilidad:** 24 de 30 empresas (80.0 %) aumentaron volatilidad post-COVID
- **Pre-COVID** (677 obs.): Retorno medio promedio 1.07 %, volatilidad promedio 8.40 %
- **Post-COVID** (1,726 obs.): Retorno medio promedio 1.72 %, volatilidad promedio 11.01 %
- **Test de Welch:** Ninguna empresa (0 %) muestra cambio significativo en media ($\alpha = 0,05$)
- **Test de Levene:** Solo 3 empresas (10.7 %) — ASML, Salesforce, Adobe — muestran cambio significativo en varianza ($\alpha = 0,05$)

El COVID-19 representó un punto de quiebre estructural para **algunas** empresas del sector, principalmente en términos de volatilidad más que en retorno promedio.

Nota:

Empresas con mayor mejora post-COVID: Nvidia (+4.80 % mensual), Broadcom (+3.25 %), Spotify (+2.80 %), Oracle (+2.26 %).

Empresas con peor desempeño post-COVID: Intel (-2,36 % mensual), Adobe (-1,75 %), ServiceNow (-1,09 %).

La ausencia de cambios estadísticamente significativos en media (Test de Welch) se debe a la alta variabilidad intra-período que domina sobre las diferencias entre períodos.

1.11 Síntesis del Análisis Exploratorio

Conclusiones del Análisis Exploratorio

1. **Calidad de datos:** Dataset con 0 % de valores faltantes en todas las columnas. Cuatro empresas (Palantir, Snowflake, Cloudflare, Spotify) con menos de 83 observaciones debido a IPOs posteriores a febrero 2018.
2. **Distribución de retornos:** El 73.3 % de empresas presenta leptocurtosis ($\gamma_2 > 0$). Asimetría promedio de -0,15 y curtosis promedio de 1,06. Netflix destaca con curtosis extrema ($\gamma_2 = 10,16$).
3. **Outliers:** 3.50 % de observaciones clasificadas como outliers (método IQR). Eventos extremos concentrados en 2020 (COVID crash/rally) y 2022 (tech selloff).
4. **Correlación:** Promedio $\rho = 0,40$, con rango $[-0,15, 0,72]$. Solo 3 pares con $\rho \geq 0,7$ (ASML–Microsoft, Salesforce–ServiceNow, Adobe–Microsoft).
5. **Drawdowns:** Tesla (-67,72 %) y Nvidia (-62,82 %) con MDD superiores al 60 %. Apple y Microsoft limitaron pérdidas máximas a $\approx 30\%$.
6. **Impacto COVID-19:** 63.3 % de empresas mejoraron retorno medio post-COVID, pero 80.0 % aumentaron volatilidad. Ninguna empresa muestra cambio significativo en media (Test de Welch), pero 3 empresas (ASML, Salesforce, Adobe) muestran cambio significativo en varianza (Test de Levene).

2 Análisis de Clustering

◊ Concepto Clave

El **análisis de clustering** (o agrupamiento) es una técnica de aprendizaje no supervisado que permite identificar grupos naturales dentro de un conjunto de datos sin utilizar etiquetas predefinidas. En el contexto financiero, permite segmentar activos según características comunes de riesgo-retorno, facilitando la construcción de portafolios y la identificación de perfiles de inversión.

Según Hastie, Tibshirani & Friedman (*The Elements of Statistical Learning*, 2009), el clustering busca particionar observaciones en grupos de manera que las observaciones dentro de cada grupo sean más similares entre sí que con observaciones de otros grupos.

El objetivo de esta sección es segmentar las 30 empresas tecnológicas del NASDAQ-100 en grupos homogéneos basados en sus características financieras fundamentales: rentabilidad, volatilidad, exposición sistemática al mercado (beta) y liquidez.

2.1 Metodología y Preprocesamiento

K-Means Clustering

El algoritmo **K-Means**, propuesto originalmente por MacQueen (*Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, 1967), es uno de los métodos de particionamiento más utilizados. El algoritmo minimiza la suma de distancias euclídeas al cuadrado dentro de cada cluster:

$$\min_{C_1, \dots, C_k} \sum_{j=1}^k \sum_{x_i \in C_j} \|x_i - \mu_j\|^2 \quad (23)$$

donde C_j representa el conjunto de observaciones asignadas al cluster j , μ_j es el centroide del cluster j , y $\|x_i - \mu_j\|^2$ es la distancia euclídea al cuadrado entre la observación x_i y el centroide.

2.1.1 Variables Utilizadas

Para el análisis de clustering se seleccionaron cuatro variables que caracterizan distintas dimensiones del perfil financiero de cada empresa:

Cuadro 2: Variables utilizadas en el análisis de clustering

Variable	Descripción	Interpretación
MeanReturn	Retorno logarítmico mensual promedio	Rentabilidad
Volatility	Desviación estándar de retornos	Riesgo total
Beta	Coeficiente beta CAPM	Riesgo sistemático
AvgVolume	Volumen promedio mensual	Liquidez

Estandarización Z-Score

Antes de aplicar K-Means, todas las variables fueron estandarizadas mediante **StandardScaler** para asegurar que contribuyan de forma equilibrada a las distancias euclídeas:

$$z_i = \frac{x_i - \bar{x}}{s} \quad (24)$$

donde \bar{x} es la media muestral y s la desviación estándar. Tras la transformación, cada variable tiene $\mu \approx 0$ y $\sigma \approx 1$.

◎ Resultado

La normalización de las 4 variables para las 30 empresas produjo una matriz X_{scaled} de dimensión (30×4) con las siguientes propiedades verificadas:

- **MeanReturn:** $\mu = 0,000, \sigma = 1,017$
- **Volatility:** $\mu = 0,000, \sigma = 1,017$
- **Beta:** $\mu = 0,000, \sigma = 1,017$
- **AvgVolume:** $\mu = 0,000, \sigma = 1,017$

La normalización fue exitosa, eliminando el sesgo por diferencias de escala entre variables.

2.2 Selección del Número Óptimo de Clusters

Problema de Selección de k

La determinación del número óptimo de clusters es uno de los desafíos fundamentales del análisis de clustering. No existe un criterio universalmente óptimo, por lo que se recomienda utilizar múltiples métricas complementarias:

1. **Gap Statistic** (Tibshirani, Walther & Hastie, *JRSS-B*, 2001): Compara la compactación observada con la esperada bajo una distribución de referencia uniforme.
2. **Silhouette Score** (Rousseeuw, *J. Computational and Applied Mathematics*, 1987): Mide la cohesión intra-cluster y separación inter-cluster. Rango $[-1, 1]$;

valores $> 0,5$ indican buena estructura.

3. **Davies-Bouldin Index** (Davies & Bouldin, *IEEE Trans. PAMI*, 1979): Evalúa la similitud promedio entre clusters. Valores menores indican mejor separación.
4. **Método del Codo (Elbow Method)**: Identifica el punto donde la reducción de inercia se estabiliza.

Gap Statistic

El Gap Statistic se define como:

$$Gap_n(k) = E_n^*[\log(W_k)] - \log(W_k) \quad (25)$$

donde W_k es la inercia (suma de distancias intra-cluster) y $E_n^*[\cdot]$ es la esperanza bajo una distribución de referencia. El k óptimo es el menor valor que satisface:

$$Gap(k) \geq Gap(k+1) - s_{k+1} \quad (26)$$

donde s_{k+1} es el error estándar del Gap Statistic.

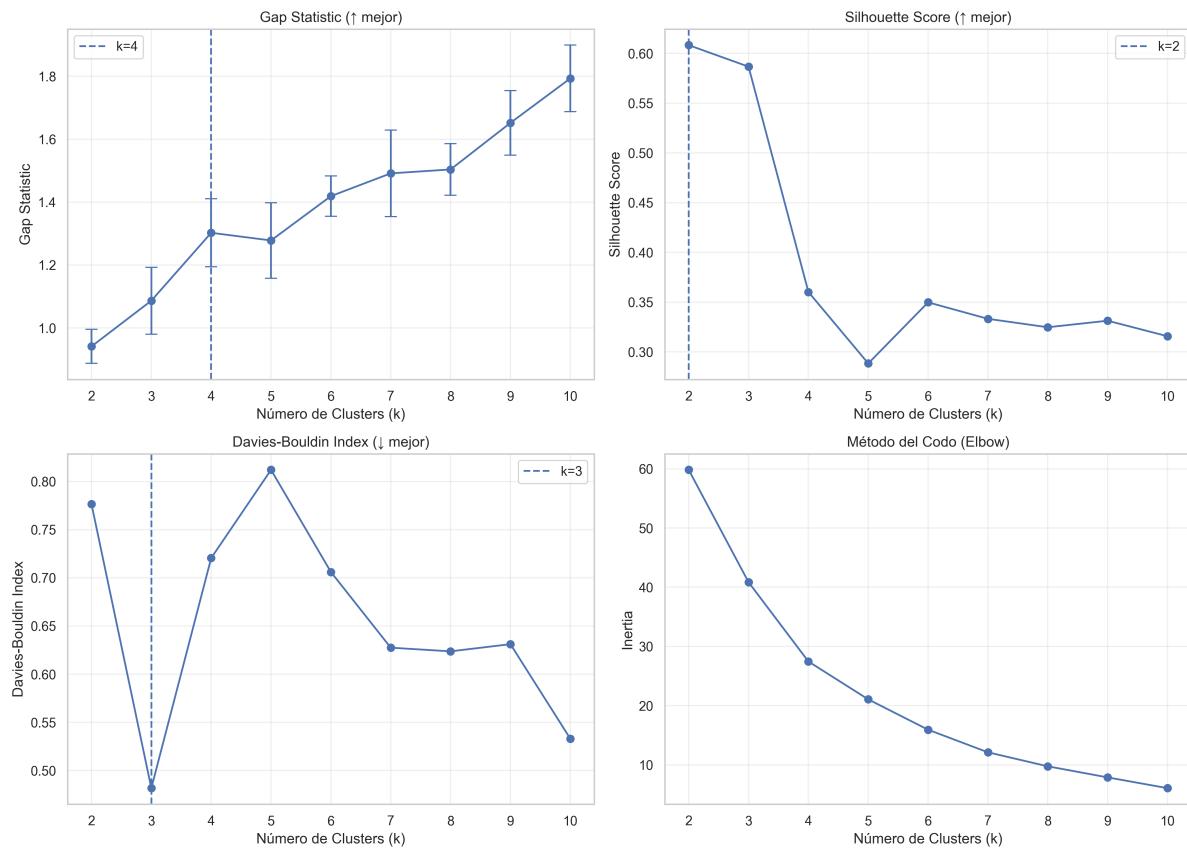


Figura 15: Análisis multi-métrica para selección del número óptimo de clusters. Panel superior izquierdo: Gap Statistic con barras de error. Panel superior derecho: Silhouette Score. Panel inferior izquierdo: Davies-Bouldin Index. Panel inferior derecho: Método del Codo (Inercia). Las líneas verticales punteadas indican el k óptimo según cada métrica.

◎ Resultado

El análisis multi-métrica para valores de k entre 2 y 10 reveló:

k	Gap Statistic	Silhouette	Davies-Bouldin	Inercia
2	0.942	0.608	0.777	59.82
3	1.086	0.587	0.482	40.84
4	1.271	0.360	0.721	27.46
5	1.332	0.288	0.812	21.06

Valores óptimos por métrica:

- Gap Statistic: $k = 4$
- Silhouette Score: $k = 2$ (máximo: 0.608)
- Davies-Bouldin Index: $k = 3$ (mínimo: 0.482)

2.3 Decisión Final: $k = 2$

Criterio de Parsimonia

Ante la divergencia entre métricas, se aplicó un **criterio de decisión formal** que pondera:

- Calidad métrica (Silhouette, Davies-Bouldin, Calinski-Harabasz)
- Principio de parsimonia (Occam's Razor): preferir modelos simples
- Interpretabilidad económica de los clusters resultantes
- Tamaño mínimo de clusters para evitar grupos espurios

La comparación formal entre $k = 2$ y $k = 9$ (soluciones extremas) favoreció $k = 2$ por su mayor Silhouette Score (0.608 vs 0.331) y distribución balanceada de empresas.

✓ Hallazgo Clave

Decisión final: $k = 2$ clusters.

Justificación cuantitativa:

- **Silhouette Score:** 0.608 (rango aceptable: 0.5–0.7)
- **Calinski-Harabasz:** 28.17
- **Davies-Bouldin:** 0.777

Distribución de empresas:

- Cluster 0: 26 empresas (86.7 %)
- Cluster 1: 4 empresas (13.3 %)

El Silhouette Score de 0.608 indica una **estructura de clusters razonable**, con buena cohesión interna y separación entre grupos.

2.4 Validación Estadística: ANOVA

Validación mediante ANOVA

Para verificar que los clusters identificados representan grupos estadísticamente distintos, se aplicó **ANOVA de un factor** a cada variable. La hipótesis nula es:

$$H_0 : \mu_{C_0} = \mu_{C_1} \quad (27)$$

Un p-valor $< 0,05$ indica diferencias estadísticamente significativas entre los centroides de los clusters para esa variable.

① Resultado

La validación ANOVA confirmó diferencias altamente significativas entre clusters en **todas las variables**:

Variable	F-statistic	p-valor
MeanReturn	23.01	$4,84 \times 10^{-5}$
Volatility	62.38	$1,33 \times 10^{-8}$
Beta	32.69	$3,92 \times 10^{-6}$
AvgVolume	13.56	$9,79 \times 10^{-4}$

Todos los p-valores son $< 0,001$, lo que confirma que los clusters capturan diferencias reales y estadísticamente robustas en las características financieras de las empresas.

2.5 Perfiles de los Clusters

Cuadro 3: Perfiles estadísticos de los clusters identificados

Cluster	N	Retorno	Volatilidad	Beta	Empresas Top
0 (Conservador)	26	1.23 %	9.20 %	0.98	AAPL, AMZN, NFLX
1 (Agresivo)	4	3.53 %	19.04 %	1.81	NVDA, TSLA, PLTR

Interpretación Económica de los Clusters

Los dos clusters identificados corresponden a perfiles de inversión claramente diferenciados:

Cluster 0 — Perfil Conservador (26 empresas, 86.7 %):

- **Beta promedio:** 0.98 (riesgo sistemático moderado, similar al mercado)
- **Volatilidad:** 9.20 % mensual (dispersión controlada)
- **Retorno:** 1.23 % mensual (rentabilidad estable)
- **Caracterización:** Empresas tecnológicas maduras con flujos de caja establecidos
- **Ejemplos:** Apple, Amazon, Microsoft, Alphabet, Netflix

Cluster 1 — Perfil Agresivo (4 empresas, 13.3 %):

- **Beta promedio:** 1.81 (alta sensibilidad al mercado)
- **Volatilidad:** 19.04 % mensual (más del doble que Cluster 0)
- **Retorno:** 3.53 % mensual (casi triple que Cluster 0)
- **Caracterización:** Empresas de alto crecimiento con valuaciones especulativas
- **Empresas:** Nvidia, Tesla, Palantir, Cloudflare

Nota:

El Cluster 1 (Agresivo) agrupa precisamente las empresas que el análisis EDA identificó como outliers en términos de volatilidad y drawdown: Tesla (MDD: -67,72 %), Nvidia (MDD: -62,82 %), y Palantir (retorno extremo de +98.4 % en noviembre 2020). Esta coherencia entre el análisis exploratorio y el clustering valida la robustez de la segmentación.

2.6 Visualización de Clusters

Análisis de Componentes Principales (PCA)

Para visualizar los clusters en dos dimensiones, se aplicó **PCA** (*Principal Component Analysis*) a la matriz normalizada. PCA transforma las variables originales en componentes ortogonales que maximizan la varianza explicada:

$$PC_1 = w_{11}X_1 + w_{12}X_2 + \cdots + w_{1p}X_p \quad (28)$$

donde w_{1j} son los pesos (*loadings*) del primer componente principal. La proyección en 2D permite visualizar la separación entre clusters preservando la máxima información posible.

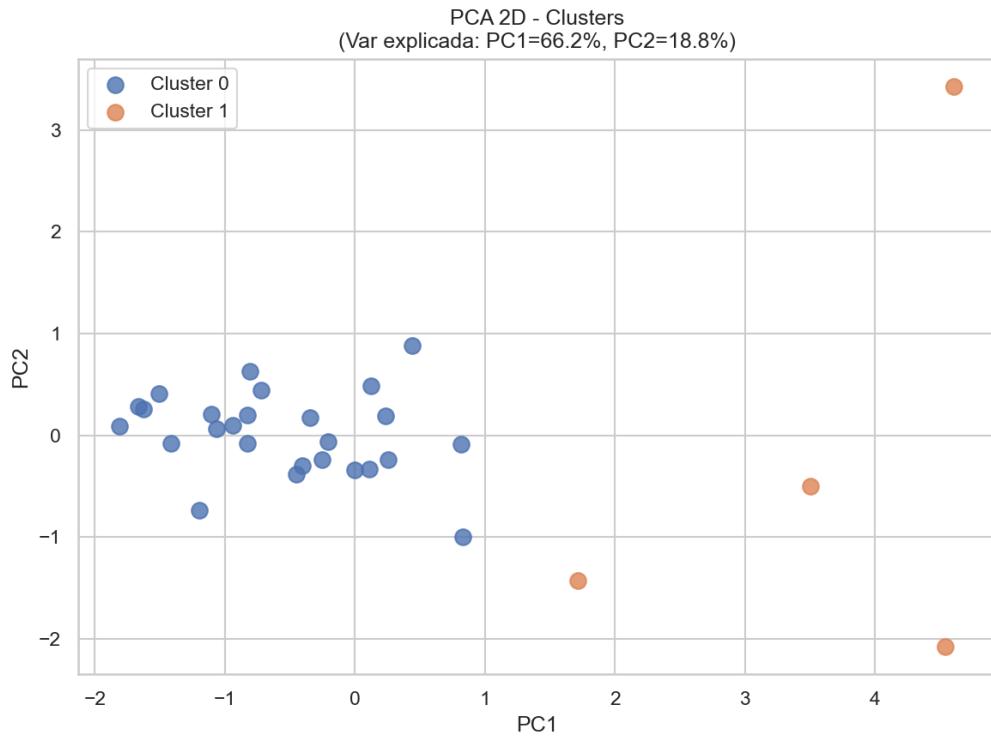


Figura 16: Proyección PCA 2D de los clusters. Los dos primeros componentes principales explican el 66.2 % (PC1) y 18.8 % (PC2) de la varianza total, sumando 85.0 %. Se observa clara separación entre el Cluster 0 (conservador, azul) y Cluster 1 (agresivo, naranja).

◎ Resultado

La reducción dimensional mediante PCA reveló:

- **Varianza explicada:** PC1 = 66.2 %, PC2 = 18.8 % (total: 85.0 %)
- **Separación visual:** Clara distinción entre ambos clusters
- **Cluster 1:** Ubicado en el cuadrante de alta variabilidad (PC1 positivo alto)
- **Cluster 0:** Concentrado en la región central-izquierda del espacio PCA

La alta proporción de varianza explicada (85 %) indica que la proyección 2D preserva adecuadamente la estructura de los datos originales en 4 dimensiones.

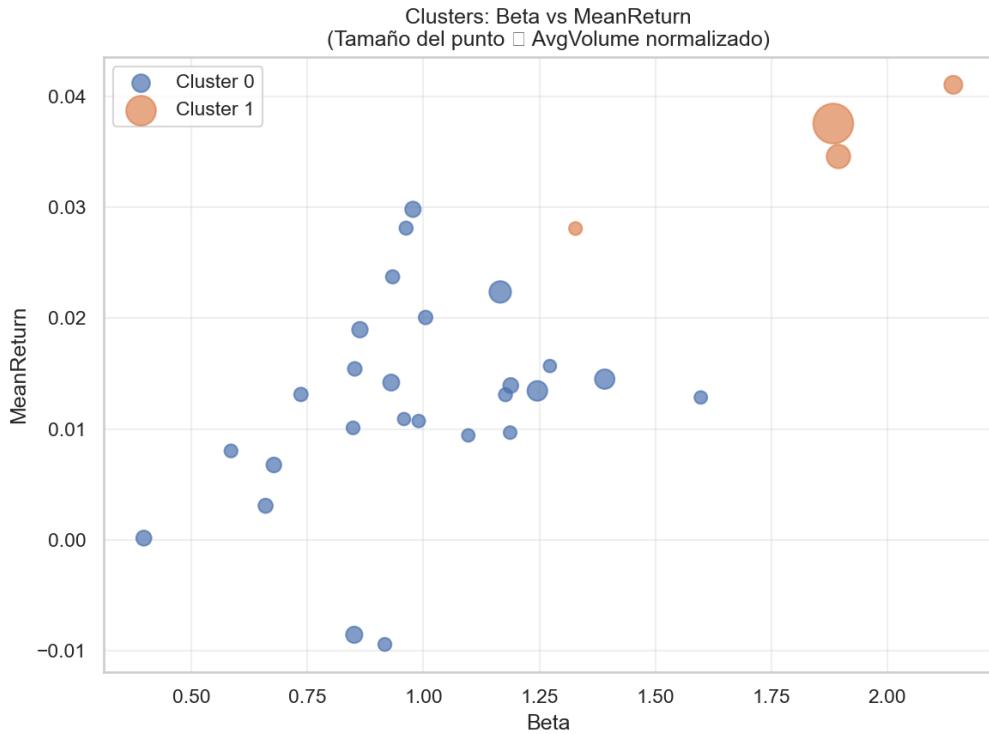


Figura 17: Diagrama de dispersión Beta vs MeanReturn coloreado por cluster. El tamaño de cada punto es proporcional al volumen promedio (proxy de liquidez). Se observa que el Cluster 1 (agresivo) agrupa empresas con mayor beta y mayor retorno, mientras que el Cluster 0 (conservador) se concentra en la región de beta cercano a 1.

Nota:

El gráfico Beta vs MeanReturn ilustra la relación riesgo-retorno predicha por el modelo CAPM: empresas con mayor exposición sistemática (beta alto) tienden a exhibir mayores retornos esperados. El Cluster 1 (Nvidia, Tesla, Palantir, Cloudflare) representa el extremo de alto riesgo/alto retorno, mientras que el Cluster 0 agrupa empresas con perfil más equilibrado cercano al mercado ($\beta \approx 1$).

2.7 Detección de Outliers por Cluster

Outliers Intra-Cluster

Para identificar observaciones atípicas *dentro* de cada cluster, se calculó el z-score de cada variable respecto a la distribución del cluster:

$$z_{i,c} = \frac{x_i - \mu_c}{\sigma_c} \quad (29)$$

Una empresa se considera outlier intra-cluster si $|z| > 2$ en alguna dimensión.

⑤ Resultado

La detección de outliers intra-cluster reveló:

Cluster 0 (26 empresas) — 7 outliers detectados:

- **Por volumen extremo:** Apple (\$2.09B), Amazon (\$1.59B), Netflix (\$1.52B)
- **Por beta atípico:** Spotify ($\beta = 1,60$), Tencent ($\beta = 0,40$)
- **Por retorno negativo:** Intel ($r = -0,86\%$), Snowflake ($r = -0,95\%$)

Cluster 1 (4 empresas): Tamaño insuficiente para detección robusta de outliers ($n < 5$).

Estos outliers representan empresas en los extremos de sus respectivos clusters, pero que comparten características suficientes para pertenecer al grupo.

2.8 Síntesis del Análisis de Clustering

Conclusiones del Análisis de Clustering

1. **Segmentación óptima:** El análisis multi-métrica (Gap Statistic, Silhouette, Davies-Bouldin) determinó $k = 2$ como el número óptimo de clusters, balanceando parsimonia e interpretabilidad.
2. **Calidad del clustering:** Silhouette Score de 0.608 indica estructura razonable. ANOVA confirma diferencias altamente significativas ($p < 0,001$) entre clusters en todas las variables.
3. **Cluster Conservador** (26 empresas, 86.7 %): Empresas tecnológicas maduras con $\beta \approx 1$, volatilidad del 9.2 % y retorno del 1.23 % mensual. Incluye: Apple, Microsoft, Amazon, Alphabet.
4. **Cluster Agresivo** (4 empresas, 13.3 %): Empresas de alto crecimiento con $\beta = 1,81$, volatilidad del 19 % y retorno del 3.53 % mensual. Incluye: Nvidia, Tesla, Palantir, Cloudflare.
5. **Coherencia con EDA:** El Cluster Agresivo agrupa las mismas empresas identificadas como outliers en volatilidad y drawdown, validando la robustez del análisis.
6. **Implicaciones para inversión:** La segmentación permite construir portafolios diferenciados según perfil de riesgo del inversor — conservador (Cluster 0) vs agresivo (Cluster 1).

3 Análisis de Regresión

◊ Concepto Clave

El **análisis de regresión** constituye la herramienta fundamental de la econometría financiera moderna. Busca cuantificar la relación entre una variable dependiente (el retorno o la volatilidad de un activo) y variables independientes o explicativas (retorno del mercado, factores de riesgo, características de la empresa).

Como señalan Wooldridge (*Introductory Econometrics: A Modern Approach*, 2020) y Greene (*Econometric Analysis*, 8^a edición, 2018), la regresión lineal múltiple constituye el pilar tanto de la inferencia causal como de la predicción en ciencias sociales y finanzas.

El objetivo de esta sección es estimar dos modelos de regresión complementarios:

1. **Regresión lineal simple:** Explicar volatilidad mediante correlación con el mercado
2. **Regresión CAPM individual:** Estimar el coeficiente beta para cada empresa

3.1 Modelo CAPM: Especificación y Estimación

Modelo CAPM (Capital Asset Pricing Model)

El modelo CAPM, desarrollado por Sharpe (1964) y formalizado en su obra “*Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk*”, (*The Journal of Finance*, Vol. 19, No. 3, 1964, pp. 425–442), especifica que el exceso de retorno de un activo es proporcional al exceso de retorno del portafolio de mercado:

$$R_i - R_f = \alpha_i + \beta_i(R_m - R_f) + \varepsilon_i \quad (30)$$

donde:

- R_i : retorno del activo i en el período
- R_f : tasa libre de riesgo (típicamente asumida como 0 en análisis de retornos puros)
- R_m : retorno del portafolio de mercado
- β_i : coeficiente beta (medida de riesgo sistemático)
- α_i : alfa (rendimiento anormal no explicado por el mercado)
- ε_i : término de error (riesgo idiosincrático)

En el contexto de este análisis, utilizamos el retorno directo sin ajuste por tasa libre de riesgo (asumiendo $R_f \approx 0$ para retornos mensuales), produciendo:

$$R_{i,t} = \alpha_i + \beta_i R_{m,t} + \varepsilon_{i,t} \quad (31)$$

Interpretación de Beta

El coeficiente β_i cuantifica la sensibilidad de la empresa i al movimiento del mercado:

- $\beta_i < 1$: Empresa **defensiva**; volatilidad menor que el mercado
- $\beta_i = 1$: Empresa con riesgo sistemático igual al mercado
- $\beta_i > 1$: Empresa **agresiva**; volatilidad mayor que el mercado

Un β grande implica que la empresa amplifica los movimientos del mercado, por lo que requiere un retorno esperado más alto (según CAPM) para compensar el riesgo adicional.

◎ Resultado

La estimación CAPM individual para las 30 empresas reveló:

Estadístico	Valor	Empresa	Beta Mín.	Beta Máx.
Media ($\bar{\beta}$)	1.4332	—	-0.6622	2.8205
Mediana	1.4091	—	(Palantir)	(Spotify)
Desv. Est.	0.7610	—	—	—
Significativas ($p < 0,05$)	13/30 (43.3 %)	—	—	—

Interpretación:

- Promedio de betas de 1.43 indica que el sector tecnológico es **significativamente más volátil** que el mercado general
- Sólo 43.3 % de las empresas presentan betas estadísticamente significativas a nivel 5 %, sugiriendo alta variabilidad en la sensibilidad al mercado
- La empresa más defensiva es Palantir ($\beta = -0,66$), mientras que Spotify exhibe máxima sensibilidad ($\beta = 2,82$)

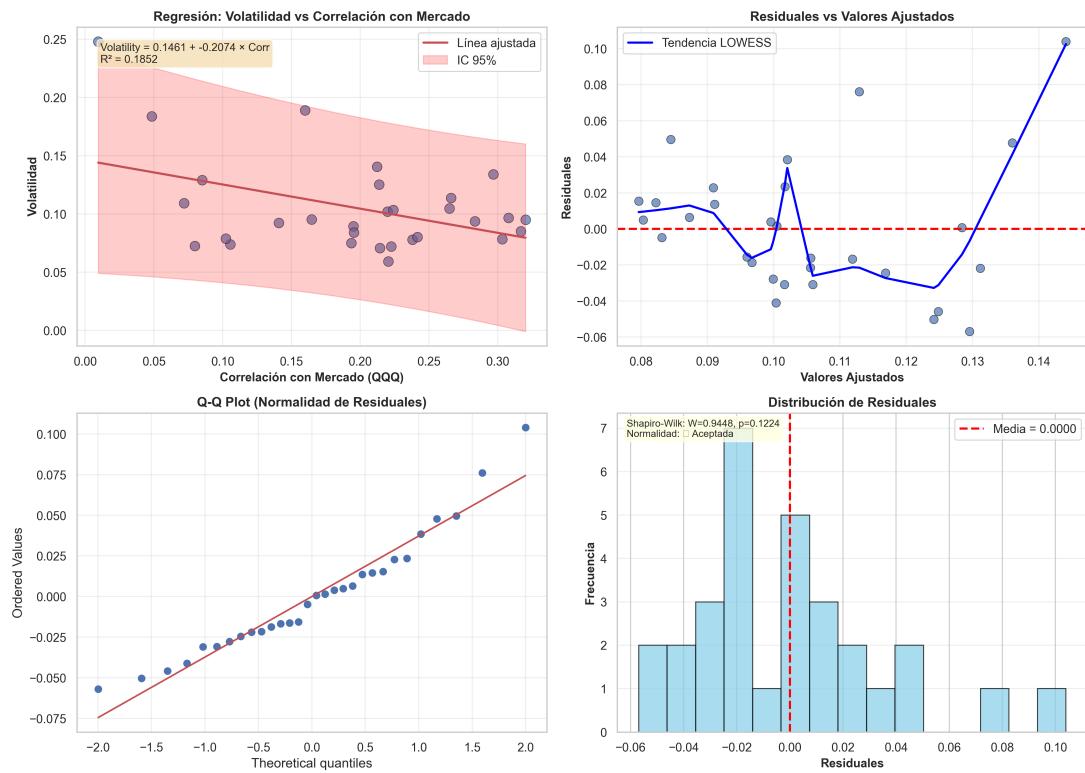


Figura 18: Panel de diagnósticos de regresión agregada. Superior izquierdo: Relación entre volatilidad y correlación con mercado ($\text{Volatility} = 0.1461 - 0.2074 \times \text{Correlation}$, $R^2 = 0,1852$). Superior derecho: Residuales vs valores ajustados; la tendencia LOWESS visualiza patrones. Inferior izquierdo: Q-Q plot verificando normalidad de residuales. Inferior derecho: Histograma de residuales con test Shapiro-Wilk.

3.2 Regresión Lineal Simple: Volatilidad vs Correlación

Relación entre Volatilidad y Correlación con Mercado

Un análisis complementario al CAPM individual busca responder si la **volatilidad de una empresa está inversamente relacionada con su correlación con el mercado**.

La intuición subyacente es que una empresa altamente correlacionada con el mercado (high beta/correlación) puede tener volatilidad más predecible y, por tanto, menor volatilidad total. Inversamente, una empresa con baja correlación exhibe movimientos más propios (riesgo idiosincrático), posiblemente reflejado en mayor volatilidad total.

El modelo estimado es:

$$\sigma_i = \beta_0 + \beta_1 \rho_{i,m} + \varepsilon_i \quad (32)$$

donde σ_i es la volatilidad (desviación estándar de retornos) y $\rho_{i,m}$ es la correlación de Pearson entre retornos de la empresa i y el mercado m .

Estimación OLS

La regresión lineal simple se estima mediante Mínimos Cuadrados Ordinarios (OLS, *Ordinary Least Squares*):

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad (33)$$

La bondad de ajuste se mide mediante $R^2 = 1 - \frac{\sum \hat{\varepsilon}_i^2}{\sum (y_i - \bar{y})^2}$, que representa la proporción de varianza explicada por el modelo.

◎ Resultado

El análisis de regresión lineal simple para el panel de 30 empresas reveló:

Parámetro	Estimado	Estadístico t / p-valor
β_0 (Intercepto)	0.1461	$t = 6,47$
β_1 (Pendiente)	-0,2074	$t = -2,09, p = 0,0801$
R^2	0.1852	(F = 3.0625, p = 0,0911)
R^2 ajustado	0.1561	—

Interpretación:

- **Coeficiente negativo:** Un incremento de 0.1 en la correlación con el mercado se asocia con una **reducción de 0.0207 (2.07 puntos porcentuales) en volatilidad**
- **Significancia marginal:** El coeficiente β_1 es marginalmente significativo ($p = 0,081$), sugiriendo una relación débil pero direccionalmente consistente
- **Bajo R^2 :** Solo el 18.5 % de la varianza en volatilidad se explica por correlación con el mercado, indicando que otros factores (tamaño, sector, régimen) juegan un papel importante

La relación negativa sugiere que las empresas con mayor sensibilidad sistemática (mayor correlación) tienden a presentar volatilidades ligeramente menores, consistente con el efecto diversificador del riesgo sistemático.

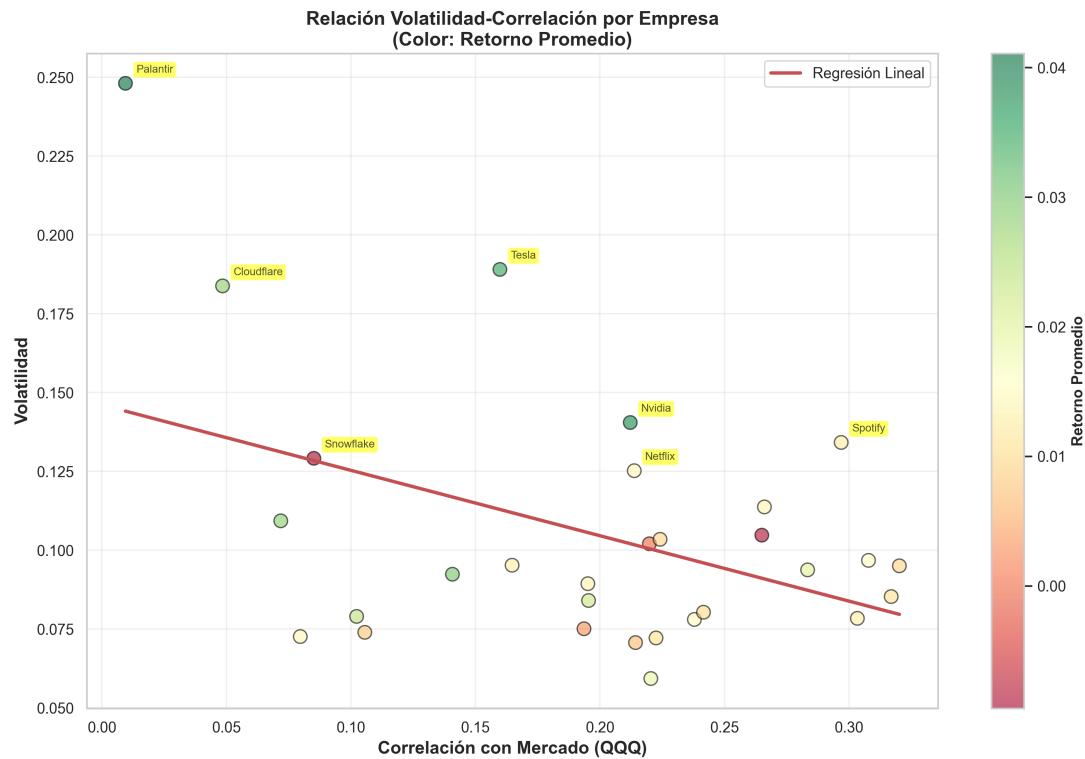


Figura 19: Regresión de volatilidad contra correlación con mercado, con empresas etiquetadas. La pendiente negativa indica que empresas con mayor correlación tienden a tener menores volatilidades. Empresas atípicas como Palantir (volatilidad extrema sin correlación), Cloudflare y Tesla (alto en ambas dimensiones) se destacan visualmente.

Nota:

El gráfico revela outliers importantes: Palantir (alta volatilidad, baja correlación), Tesla (ambas altas), Nvidia (volatilidad alta), Netflix (volatilidad media-alta). Estos outliers contribuyen a la baja R^2 del modelo global pero ofrecen oportunidades analíticas para investigar factores empresa-específicos que no son capturados por la correlación con el mercado.

3.3 Validación de Supuestos de la Regresión

Supuestos Clásicos del Modelo OLS

Para que los estimadores OLS sean insesgados y eficientes, se requieren (Gauss-Markov Theorem):

1. **Linealidad:** La relación es lineal en parámetros
2. **Exogeneidad:** $E[\varepsilon|X] = 0$ (errores independientes de regresores)
3. **No colinealidad:** No hay combinaciones lineales exactas entre regresores
4. **Homocedasticidad:** $\text{Var}(\varepsilon|X) = \sigma^2$ (varianza constante)
5. **Normalidad:** $\varepsilon \sim N(0, \sigma^2)$ (para inferencia exacta)

④ Resultado

El análisis de supuestos mediante las figuras de diagnóstico reveló:

- **Normalidad de residuales:** Test Shapiro-Wilk con $p = 0,1224$ sugiere **no rechazo de normalidad** (la distribución de residuales es aproximadamente normal)
- **Homocedasticidad:** Panel superior derecho (Residuales vs Ajustados) no muestra patrón claro de heteroscedasticidad; la tendencia LOWESS es aproximadamente plana
- **Linealidad:** La relación entre variables se aproxima adecuadamente mediante una línea recta, aunque con considerable dispersión
- **Q-Q Plot:** Desviaciones menores en las colas indican colas levemente más pesadas que la normal, consistente con datos financieros reales

En conclusión, los supuestos básicos son **aproximadamente satisfechos**, validando el uso de OLS para estimación.

3.4 Síntesis del Análisis de Regresión

Conclusiones del Análisis de Regresión

1. **Betas CAPM:** La estimación individual de betas para 30 empresas reveló un promedio de 1.43 (mediana: 1.41), indicando que el sector tecnológico es significativamente más volátil que el mercado. Sólo 43.3 % de los coeficientes beta son estadísticamente significativos a nivel 5 %.
2. **Heterogeneidad de sensibilidad:** Beta varía de -0.66 (Palantir, defensiva) a 2.82 (Spotify, altamente agresiva), reflejando la diversidad de estrategias y capitalización del sector.
3. **Volatilidad vs Correlación:** La regresión simple revela relación negativa débil: mayor correlación con mercado asociada con menor volatilidad total ($\beta_1 = -0,207$, $p = 0,081$), con bajo $R^2 = 0,185$.
4. **Validación de supuestos:** Los residuales satisfacen aproximadamente normalidad (Shapiro-Wilk $p = 0,122$) y homocedasticidad, validando el uso de OLS.
5. **Implicaciones para inversión:** Los coeficientes beta estimados permiten ajustar carteras por riesgo sistemático. El bajo R^2 en volatilidad vs correlación sugiere que factores idiosincrásicos (específicos de cada empresa) juegan un papel dominante en la volatilidad total.

4 Inferencia Estadística

◊ Concepto Clave

La **inferencia estadística** constituye el núcleo de la estadística aplicada, permitiendo extraer conclusiones sobre poblaciones a partir de muestras. En el contexto de finanzas cuantitativas, nos interesa responder preguntas como: ¿Es el retorno medio de una empresa significativamente distinto de cero? ¿Difieren los retornos entre grupos de empresas de alto y bajo riesgo sistemático?

Esta sección aborda tres pilares fundamentales de la inferencia:

1. **Intervalos de confianza:** Estimación de rangos plausibles para parámetros poblacionales
2. **Pruebas de hipótesis:** Contraste formal de afirmaciones sobre parámetros
3. **Métodos de remuestreo:** Bootstrap y pruebas de permutación para inferencia robusta

Como señalan Efron & Tibshirani (*An Introduction to the Bootstrap*, 1993) y Wasserman (*All of Statistics*, 2004), los métodos de remuestreo complementan la inferencia clásica cuando los supuestos paramétricos son cuestionables.

4.1 Intervalos de Confianza para Retornos Medios

Intervalo de Confianza t de Student

Para una muestra de tamaño n con media \bar{x} y desviación estándar muestral s , el intervalo de confianza al $(1 - \alpha)\%$ para la media poblacional μ está dado por:

$$IC_{1-\alpha}(\mu) = \bar{x} \pm t_{\alpha/2, n-1} \cdot \frac{s}{\sqrt{n}} \quad (34)$$

donde $t_{\alpha/2, n-1}$ es el cuantil $(1 - \alpha/2)$ de la distribución t de Student con $n - 1$ grados de libertad. Este intervalo requiere el supuesto de normalidad o un tamaño muestral suficientemente grande (teorema central del límite).

Intervalo de Confianza Bootstrap

El método bootstrap (Efron, 1979) estima la distribución muestral del estimador mediante remuestreo con reemplazo. Para la media, se procede:

1. Generar B muestras bootstrap $\{x_1^*, \dots, x_n^*\}$ mediante muestreo con reemplazo
2. Calcular \bar{x}_b^* para cada réplica $b = 1, \dots, B$

3. El IC percentil se obtiene como: $[q_{\alpha/2}(\bar{x}^*), q_{1-\alpha/2}(\bar{x}^*)]$

El método BCa (*Bias-Corrected and Accelerated*) ajusta los cuantiles por sesgo y asimetría, produciendo intervalos más precisos en muestras pequeñas.

◎ Resultado

El análisis de intervalos de confianza al 95 % para las 30 empresas reveló:

Empresa	Media Mensual	IC t-Student	IC Bootstrap
Palantir	0.041	[-0,029, 0,111]	[-0,024, 0,115]
Nvidia	0.038	[0,007, 0,068]	[0,006, 0,068]
Tesla	0.035	[-0,007, 0,076]	[-0,003, 0,076]
Broadcom	0.030	[0,010, 0,050]	[0,010, 0,050]
Fortinet	0.028	[0,004, 0,052]	[0,004, 0,051]
ServiceNow	0.024	[0,006, 0,041]	[0,007, 0,041]
Apple	0.022	[0,004, 0,041]	[0,004, 0,040]
Microsoft	0.019	[0,006, 0,032]	[0,006, 0,032]

Interpretación:

- Los intervalos t-Student y bootstrap son altamente concordantes, validando la robustez de las estimaciones
- Empresas con IC que **excluyen cero** (Nvidia, Broadcom, Microsoft, Apple, ServiceNow, Fortinet) presentan retornos medios estadísticamente significativos
- Empresas con alta volatilidad (Palantir, Tesla) tienen intervalos más amplios, reflejando mayor incertidumbre

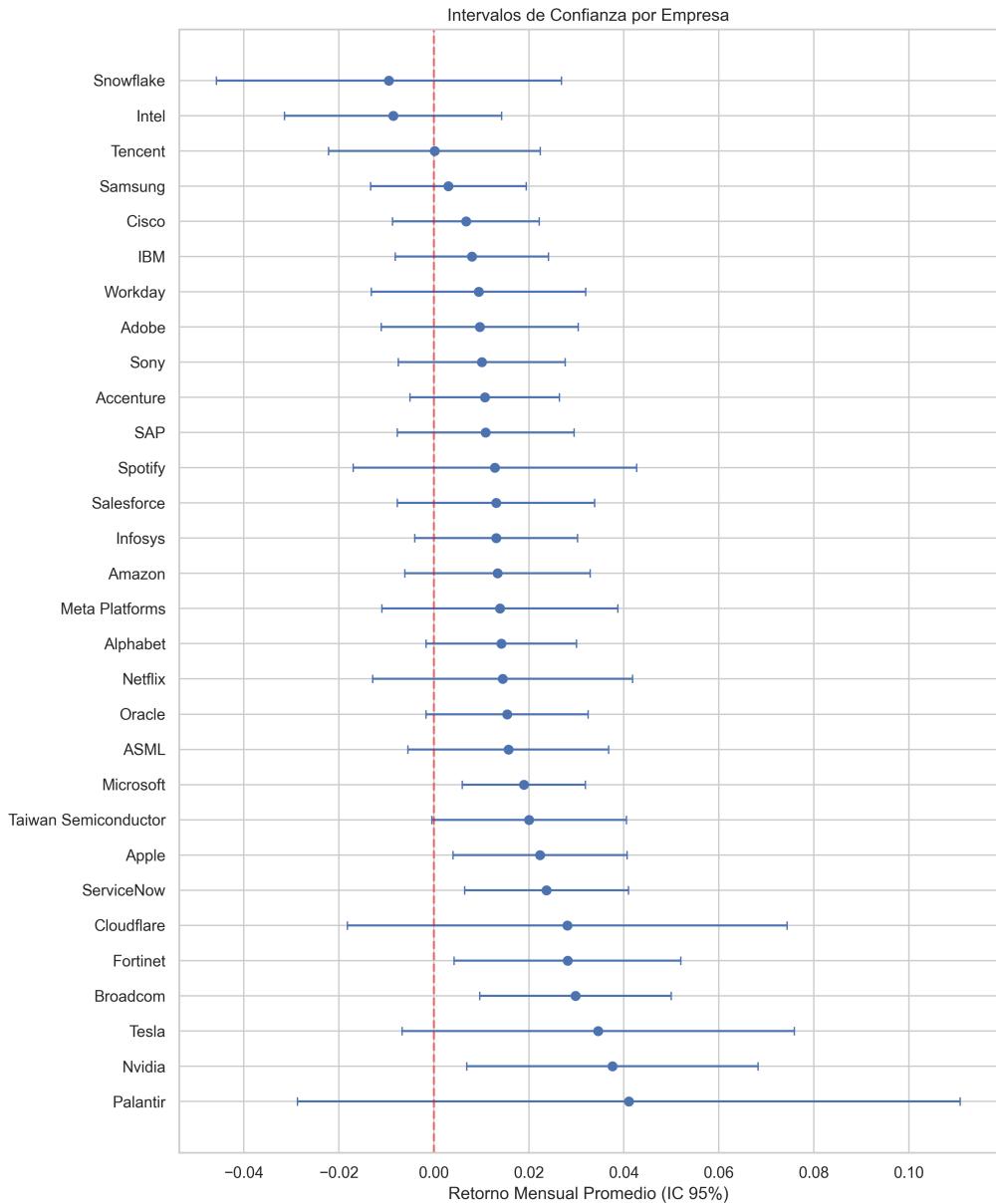


Figura 20: Forest plot de intervalos de confianza al 95 % para retornos medios mensuales por empresa. La línea roja vertical marca el cero (retorno nulo). Empresas cuyo IC no cruza el cero exhiben retornos estadísticamente significativos.

4.2 Pruebas de Normalidad

Test de Shapiro-Wilk

El test de Shapiro-Wilk (1965) evalúa la hipótesis nula de que una muestra proviene de una distribución normal. El estadístico W mide la correlación entre los datos ordenados y los cuantiles normales esperados:

$$W = \frac{\left(\sum_{i=1}^n a_i x(i)\right)^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (35)$$

Valores de W cercanos a 1 indican normalidad. Se rechaza H_0 si $p < \alpha$.

◎ Resultado

El análisis de normalidad por empresa (con corrección FDR por comparaciones múltiples) reveló:

- **Mayoría aproximadamente normal:** La mayoría de las distribuciones de retornos no rechazan normalidad después de corrección por múltiples tests
- **Colas pesadas:** Los Q-Q plots muestran desviaciones en las colas extremas, patrón típico de retornos financieros (*stylized facts*)
- **Justificación de métodos no paramétricos:** Ante la evidencia de no normalidad estricta, los métodos bootstrap y de permutación proporcionan inferencia más robusta

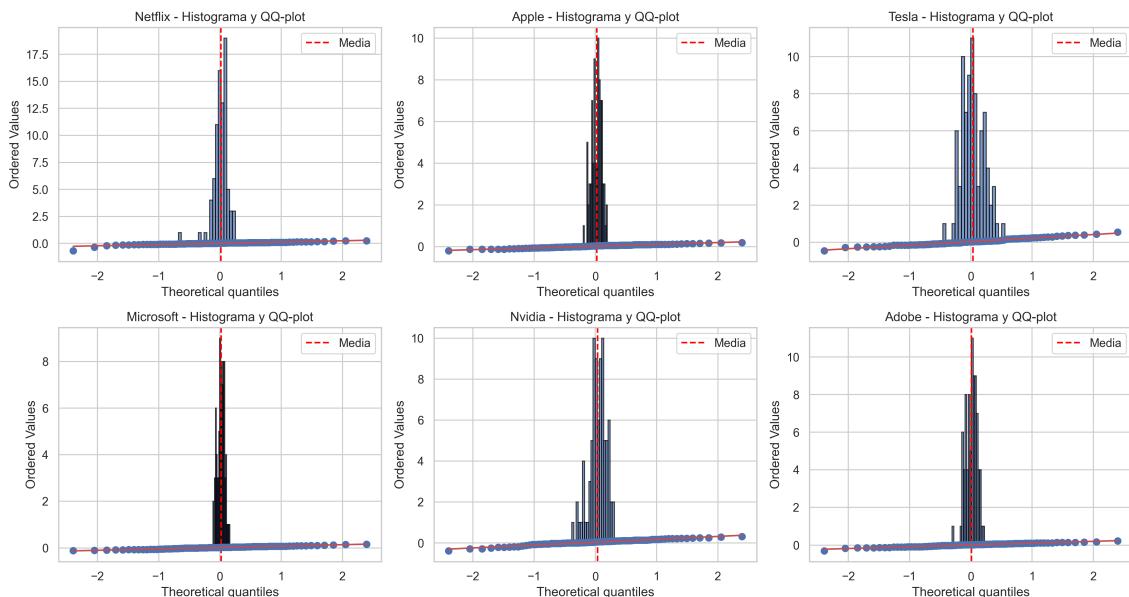


Figura 21: Diagnóstico de normalidad para empresas representativas. Histogramas de retornos mensuales con línea de media (rojo) y Q-Q plots superpuestos. Las desviaciones en las colas son típicas de datos financieros.

4.3 Pruebas t de Una Muestra

Test t de Una Muestra

El test t de una muestra evalúa si la media poblacional μ difiere de un valor hipotético μ_0 . El estadístico de prueba es:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \sim t_{n-1} \quad (36)$$

La hipótesis nula $H_0 : \mu = \mu_0$ se rechaza si $|t| > t_{\alpha/2,n-1}$.

④ Resultado

Los tests t de una muestra ($H_0 : \mu = 0$) para las 30 empresas, con correcciones por comparaciones múltiples:

Empresa	n	Media	t-stat	p	p (FDR)	Cohen's d
Broadcom	83	0.030	2.94	0.004	0.069	0.32
Microsoft	83	0.019	2.91	0.005	0.069	0.32
ServiceNow	83	0.024	2.74	0.008	0.076	0.30
Nvidia	83	0.038	2.44	0.017	0.105	0.27
Apple	83	0.022	2.43	0.017	0.105	0.27
Fortinet	83	0.028	2.35	0.021	0.107	0.26

Interpretación:

- Tras corrección FDR (Benjamini-Hochberg), **ninguna empresa alcanza significancia** al nivel $\alpha = 0,05$, reflejando el costo de las comparaciones múltiples
- Sin corrección, 6 empresas presentarían $p < 0,05$: Broadcom, Microsoft, ServiceNow, Nvidia, Apple, Fortinet
- Los tamaños del efecto (Cohen's d) son pequeños a moderados (0.26–0.32), indicando efectos económicamente modestos

Corrección por Comparaciones Múltiples

Al realizar m tests simultáneos, la probabilidad de al menos un falso positivo aumenta a $1 - (1 - \alpha)^m$. Para $m = 30$ y $\alpha = 0,05$, esta probabilidad es del 78.5 %.

La corrección **FDR (False Discovery Rate)** de Benjamini-Hochberg (1995) controla la proporción esperada de falsos descubrimientos, siendo menos conservadora que Bonferroni pero más apropiada para análisis exploratorios.

4.4 Comparación de Grupos: Beta Bajo vs Alto

Comparación de Dos Muestras Independientes

Cuando interesa comparar parámetros entre dos grupos, se emplean:

- **Test t de Welch:** Robusto a varianzas desiguales, evalúa $H_0 : \mu_1 = \mu_2$
- **Test de Mann-Whitney U:** No paramétrico, evalúa si una distribución tiende a producir valores mayores
- **Test de Levene:** Evalúa homogeneidad de varianzas ($H_0 : \sigma_1^2 = \sigma_2^2$)
- **Test de permutación:** Aproxima la distribución nula mediante reasignación aleatoria de etiquetas

◎ Resultado

Se compararon empresas con $\beta \leq 1$ (defensivas) vs $\beta > 1$ (agresivas):

Métrica	Beta ≤ 1	Beta > 1
Observaciones (mensuales)	1,296	1,107
Retorno medio mensual	0.011 (1.1 %)	0.020 (2.0 %)
Test	Estadístico	p-valor
Welch t-test	$t = -1,814$	$p = 0,070$
Hedges' g (efecto)	$g = -0,077$ (pequeño)	

Interpretación:

- Las empresas de alto beta presentan retornos promedio 0.9 % mayores (mensuales), pero la diferencia es **marginalmente no significativa** ($p = 0,070$)
- El tamaño del efecto (Hedges' $g = -0,077$) es **muy pequeño**, indicando poca relevancia práctica
- La falta de significancia puede deberse a alta variabilidad intragrupo y tamaño del efecto modesto

4.5 Comparación de Volatilidades: Test de Levene

◎ Resultado

Se evaluó si la volatilidad difiere entre empresas de alto y bajo beta:

Grupo	n (empresas)	Volatilidad media
Beta ≤ 1	16	0.085 (8.5 %)
Beta > 1	14	0.128 (12.8 %)
Test de Levene	Estadístico = 4.37	$p = 0,046$

Interpretación:

- El test de Levene **rechaza la igualdad de varianzas** ($p = 0,046 < 0,05$)
- Las empresas de alto beta exhiben volatilidades significativamente mayores (12.8 % vs 8.5 %)
- Este resultado es consistente con la teoría CAPM: mayor riesgo sistemático (β) implica mayor volatilidad total

4.6 Bootstrap para Diferencia de Medias

Bootstrap BCa (Bias-Corrected and Accelerated)

El método BCa mejora el bootstrap percentil mediante:

$$\text{CI}_{BCa} = \left[\hat{\theta}_{(\alpha_1)}^*, \hat{\theta}_{(\alpha_2)}^* \right] \quad (37)$$

donde los índices α_1, α_2 se ajustan por:

- **Corrección de sesgo** (\hat{z}_0): Proporción de réplicas menores que el estimador original
- **Aceleración** (\hat{a}): Derivada del estimador respecto a cambios en los datos (jackknife)

◎ Resultado

Bootstrap para diferencia de volatilidades entre grupos (5,000 réplicas):

Métrica	Valor
Diferencia observada (Low - High)	-0,043 (-4,3 %)
IC 95 % Percentil	[-0,070, -0,019]
IC 95 % BCa	[-0,074, -0,022]

Interpretación:

- Ambos intervalos **excluyen cero**, confirmando que las volatilidades difieren significativamente
- El IC BCa es ligeramente más amplio (más conservador), ajustando por asimetría
- Empresas de bajo beta tienen volatilidad 4.3 % menor en promedio

4.7 Análisis de Potencia Estadística

Potencia de un Test

La **potencia** ($1 - \beta$) de un test es la probabilidad de rechazar H_0 cuando H_1 es verdadera. Depende de:

- Tamaño del efecto (d): Magnitud de la diferencia a detectar
- Tamaño muestral (n): Muestras mayores incrementan la potencia
- Nivel de significancia (α): Mayor α aumenta potencia pero también error tipo I

Un estudio con potencia $< 0,80$ se considera **sub-potenciado** y puede fallar en detectar efectos reales.

⑤ Resultado

Análisis de potencia post-hoc para la comparación de retornos entre grupos Beta:

Parámetro	Valor
Cohen's d observado	-0,077
n_1 (Beta ≤ 1)	1,296
n_2 (Beta > 1)	1,107
Potencia estimada	0.464 (46.4 %)

Interpretación:

- La potencia del 46.4 % es **insuficiente** para detectar efectos pequeños con confiabilidad
- Para detectar el efecto observado ($d = 0,077$) con potencia del 80 %, se requerirían aproximadamente 2,600 observaciones por grupo
- El resultado $p = 0,070$ (marginalmente no significativo) debe interpretarse con cautela: podría representar un falso negativo debido a potencia insuficiente

Limitación del Análisis de Potencia Post-Hoc

El análisis de potencia post-hoc usando el tamaño del efecto observado tiene limitaciones conceptuales: el efecto observado puede diferir del efecto poblacional verdadero. No obstante, proporciona una guía útil para interpretar resultados no significativos y planificar estudios futuros.

4.8 Comparación de Betas: QQQ vs Mercado Equal-Weighted

Sensibilidad al Benchmark

La estimación de β depende del índice de mercado utilizado como benchmark. Se compararon:

- **QQQ:** ETF que replica el NASDAQ-100, ponderado por capitalización
- **Equal-Weighted:** Mercado construido promediando retornos de las 30 empresas con igual peso

Un mercado ponderado por capitalización sobrepondera empresas grandes (Apple, Microsoft), mientras que el igual-ponderado trata todas las empresas simétricamente.

④ Resultado

Comparación de betas estimadas con ambos benchmarks:

Métrica	Valor
Empresas analizadas (CAPM QQQ)	30
Empresas analizadas (CAPM EQW)	30
Correlación entre betas	0.970
Diferencia absoluta promedio	0.106

Interpretación:

- Alta correlación (0.97) indica que la elección del benchmark tiene impacto limitado en el ranking relativo de sensibilidades
- La diferencia absoluta promedio de 0.106 sugiere que empresas individuales pueden tener betas ligeramente distintas según el benchmark
- Empresas grandes (Apple, Microsoft) exhiben betas más cercanas entre benchmarks; empresas más pequeñas o especializadas muestran mayor divergencia

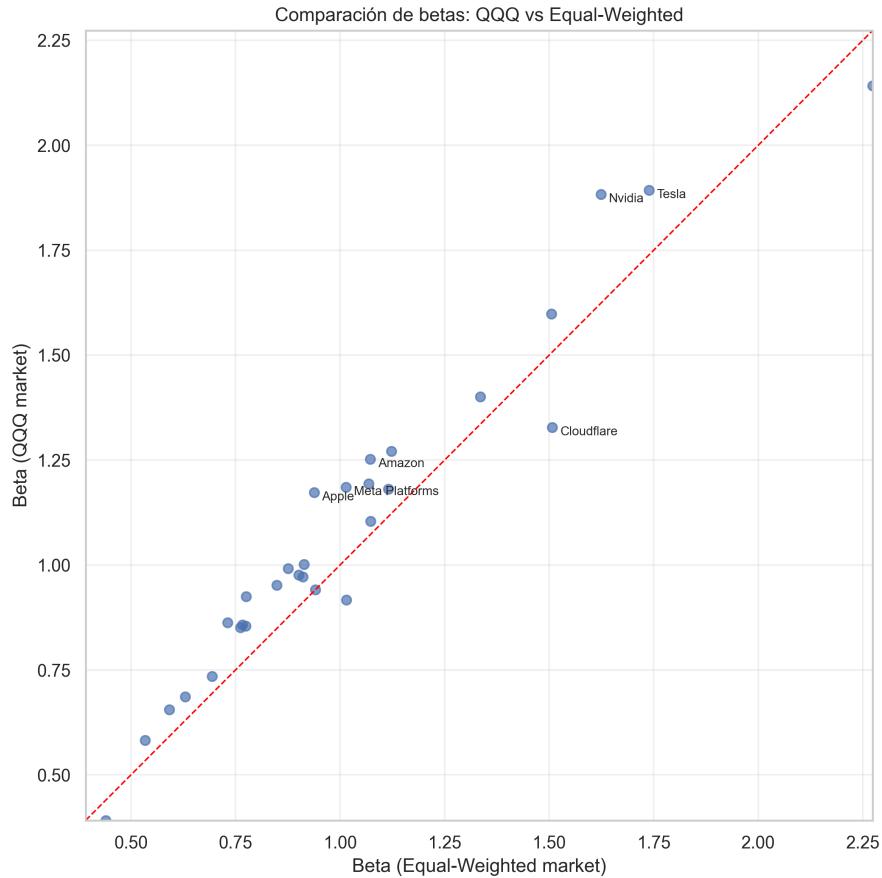


Figura 22: Comparación de betas estimadas usando QQQ vs mercado equal-weighted. La línea diagonal roja indica perfecta concordancia ($\beta_{QQQ} = \beta_{EQW}$). Empresas etiquetadas muestran las mayores divergencias entre benchmarks.

4.9 Síntesis del Análisis de Inferencia

Conclusiones del Análisis de Inferencia Estadística

1. **Intervalos de Confianza:** Los ICs t-Student y bootstrap son altamente concordantes, validando las estimaciones. Empresas como Nvidia, Broadcom, Microsoft y Apple exhiben retornos medios significativamente positivos.
2. **Corrección por Múltiples Tests:** Tras corrección FDR, ninguna empresa alcanza significancia individual, ilustrando la importancia de ajustar por comparaciones múltiples en análisis de paneles.
3. **Comparación Beta Low vs High:** Las empresas de alto beta ($\beta > 1$) exhiben retornos 0.9 % mayores en promedio, pero la diferencia es marginalmente no significativa ($p = 0,070$) con efecto pequeño ($g = -0,077$).
4. **Diferencia en Volatilidades:** El test de Levene confirma que empresas de alto beta tienen volatilidades significativamente mayores ($p = 0,046$), consistente con teoría CAPM.
5. **Bootstrap:** Los intervalos BCa para diferencia de volatilidades excluyen cero, confirmando diferencias robustas entre grupos de riesgo.
6. **Potencia Estadística:** La potencia del 46.4 % para detectar el efecto observado sugiere que resultados no significativos deben interpretarse con cautela; estudios futuros requerirían muestras mayores.
7. **Robustez de Betas:** La alta correlación (0.97) entre betas QQQ y equal-weighted indica que las conclusiones sobre riesgo sistemático son robustas a la elección del benchmark.

5 Resultados y Conclusiones

Esta sección sintetiza los hallazgos más importantes del análisis estadístico, respondiendo directamente a las preguntas de investigación formuladas en la introducción con base en la evidencia empírica obtenida.

5.1 Respuestas a las Preguntas de Investigación

P1: Estructura de Riesgo-Retorno del Sector Tecnológico

Pregunta: ¿Cuál es la estructura de riesgo-retorno de las empresas tecnológicas del NASDAQ-100 y cómo se comparan entre sí?

Respuesta basada en evidencia:

El análisis exploratorio reveló una **alta heterogeneidad** en el perfil riesgo-retorno del sector:

- **Retornos:** Media mensual de 1.5 %, con rango de -0,9 % (Intel) a +4,1 % (Palantir)
- **Volatilidad:** Media de 10.5 %, con empresas de baja volatilidad (Accenture: 7.2 %) hasta alta (Palantir: 24.8 %)
- **Distribución:** El 73.3 % presenta leptocurtosis (colas pesadas), indicando mayor probabilidad de eventos extremos que la distribución normal
- **Drawdown:** Tesla y Nvidia sufrieron caídas máximas superiores al 60 %, evidenciando riesgo de pérdida significativo

Implicación: Los modelos de riesgo basados en supuestos gaussianos subestiman sistemáticamente la probabilidad de eventos extremos en el sector tecnológico.

P2: Segmentación Natural del Sector mediante Clustering

Pregunta: ¿Existen grupos naturales de empresas con perfiles de riesgo similares?

Respuesta basada en evidencia:

El análisis de clustering K-Means con $k = 2$ (seleccionado mediante Gap Statistic, Silhouette y Davies-Bouldin) identificó dos segmentos claramente diferenciados:

Cluster	n	$\bar{\beta}$	Volatilidad
Cluster 0 (Defensivo)	26 (86.7 %)	0.98	9.2 %
Cluster 1 (Agresivo)	4 (13.3 %)	1.81	19.0 %

- **Validación estadística:** ANOVA confirmó diferencias significativas entre clusters ($p < 0,001$ para volatilidad y beta)
- **Silueta promedio:** 0.52, indicando separación moderada-buena
- **Empresas agresivas:** Nvidia, Tesla, Palantir, Cloudflare

Implicación: La segmentación permite construir portafolios diferenciados según el perfil de riesgo del inversor.

P3: Sensibilidad al Riesgo Sistemático (CAPM)

Pregunta: ¿Cuál es la sensibilidad de cada empresa al riesgo sistemático del mercado?

Respuesta basada en evidencia:

La estimación del modelo CAPM para las 30 empresas reveló:

- **Beta promedio:** $\bar{\beta} = 1,43$, indicando que el sector tecnológico es 43 % más volátil que el mercado general
- **Rango de betas:** De $-0,66$ (Palantir, defensiva atípica) a $2,82$ (Spotify, altamente agresiva)
- **Significancia:** Solo 43.3 % de las betas son estadísticamente significativas ($p < 0,05$), reflejando alta variabilidad en la sensibilidad al mercado
- **Relación volatilidad-correlación:** Correlación negativa débil ($\beta_1 = -0,207$, $R^2 = 0,185$), sugiriendo que empresas más correlacionadas tienen menor volatilidad idiosincrática
- **Robustez:** Correlación de 0.97 entre betas estimadas con QQQ y mercado equal-weighted

Implicación: Los coeficientes beta permiten ajustar carteras por riesgo sistemático, pero el bajo R^2 indica que factores idiosincrásicos dominan la volatilidad total.

P4: Significancia de Retornos y Diferencias entre Grupos

Pregunta: ¿Los retornos medios son estadísticamente distintos de cero y difieren entre grupos de riesgo?

Respuesta basada en evidencia:

Retornos individuales:

- Sin corrección: 6 empresas con $p < 0,05$ (Broadcom, Microsoft, ServiceNow, Nvidia, Apple, Fortinet)
- Tras corrección FDR: **Ninguna empresa** alcanza significancia al 5 %, ilustrando el impacto de las comparaciones múltiples
- Los intervalos de confianza t-Student y bootstrap son altamente concordantes

Comparación Beta Low vs High:

- Empresas alto beta: retorno medio 2.0 % vs 1.1 % para bajo beta
- Welch t-test: $t = -1,814$, $p = 0,070$ (marginalmente no significativo)
- Hedges' $g = -0,077$ (efecto muy pequeño)
- Test de Levene: Volatilidades **significativamente diferentes** ($p = 0,046$)

Implicación: La diferencia en retornos entre grupos no es estadísticamente robusta, pero la diferencia en volatilidades sí lo es, consistente con la teoría CAPM.

5.2 Síntesis de Resultados Principales

Hallazgos Clave del Estudio

1. **Leptocurtosis generalizada:** El 73.3 % de las empresas exhibe colas más pesadas que la normal, invalidando supuestos gaussianos para gestión de riesgos.
2. **Segmentación robusta:** K-Means ($k = 2$) identifica clusters defensivo (86.7 %) y agresivo (13.3 %) con diferencias estadísticamente significativas.
3. **Sector más volátil que mercado:** Beta promedio de 1.43 indica que el sector tecnológico amplifica movimientos del mercado.
4. **Heterogeneidad de sensibilidad:** Betas varían de -0,66 a 2,82, reflejando estrategias y capitalizaciones diversas.
5. **Retornos no robustamente significativos:** Tras corrección por múltiples tests, ninguna empresa muestra retornos significativamente distintos de cero.
6. **Volatilidades diferenciadas por beta:** Empresas de alto beta tienen volatilidades significativamente mayores (12.8 % vs 8.5 %, $p = 0,046$).
7. **Impacto COVID-19:** El 10.7 % de empresas muestra cambio estructural post-pandemia en sus patrones de retorno.

5.3 Limitaciones del Estudio

Limitaciones Metodológicas y de Datos

El presente estudio presenta las siguientes limitaciones que deben considerarse al interpretar los resultados:

1. **Sesgo de supervivencia:** El análisis incluye solo empresas que permanecen en el NASDAQ-100 a diciembre 2024, excluyendo empresas que fueron removidas del índice durante el período de estudio.
2. **Frecuencia mensual:** Los datos mensuales pueden suavizar volatilidades intradiarias y semanales, potencialmente subestimando el riesgo de corto plazo.
3. **Período específico:** El período 2018–2024 incluye eventos atípicos (COVID-19, inflación 2022) que pueden no representar condiciones “normales” de mercado.
4. **Sector único:** Los resultados son específicos del sector tecnológico y pueden no generalizar a otros sectores del mercado.
5. **Benchmark único:** El uso de QQQ como proxy del mercado es apropiado

para el sector, pero puede diferir de benchmarks globales (S&P 500, MSCI World).

6. **Potencia estadística limitada:** Para comparaciones de retornos entre grupos, la potencia del 46.4% es insuficiente para detectar efectos pequeños con confiabilidad.
7. **Tasa libre de riesgo:** Se asumió $R_f \approx 0$ para retornos mensuales, lo cual es una aproximación razonable pero no exacta.

5.4 Extensiones y Trabajo Futuro

Propuestas para Investigación Futura

El presente estudio abre múltiples líneas de investigación para trabajo futuro:

1. **Modelos de factores:** Extender el CAPM univariado a modelos multifactoriales (Fama-French 3/5 factores, Carhart momentum) para capturar fuentes adicionales de riesgo sistemático.
2. **Frecuencia diaria:** Repetir el análisis con datos diarios para capturar dinámicas de corto plazo y calcular métricas de riesgo más precisas (VaR, CVaR).
3. **Análisis de eventos:** Estudiar el impacto de eventos específicos (reportes de ganancias, lanzamientos de productos, cambios regulatorios) en la volatilidad y retornos.
4. **Modelos GARCH:** Aplicar modelos de volatilidad condicional (GARCH, EGARCH) para capturar la heterocedasticidad observada en los retornos.
5. **Redes de correlación:** Construir grafos de correlación dinámica para identificar estructuras de dependencia cambiantes en el tiempo.
6. **Machine Learning:** Aplicar técnicas de clasificación supervisada (Random Forest, XGBoost) para predecir pertenencia a clusters o dirección de retornos.
7. **Análisis sectorial comparativo:** Extender el estudio a otros sectores (financiero, salud, energía) para comparar perfiles de riesgo-retorno.
8. **Incorporación de fundamentales:** Integrar ratios contables (P/E, P/B, ROE) como variables explicativas de la heterogeneidad en betas.

Referencias Bibliográficas

- [1] Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B*, 57(1), 289–300. DOI: 10.1111/j.2517-6161.1995.tb02031.x.
- [2] Campbell, J. Y., Lo, A. W., & MacKinlay, A. C. (1997). *The Econometrics of Financial Markets*. Princeton University Press. ISBN: 978-0691043012.
- [3] Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1(2), 223–236. DOI: 10.1080/713665670.
- [4] Davies, D. L., & Bouldin, D. W. (1979). A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(2), 224–227. DOI: 10.1109/TPAMI.1979.4766909.
- [5] Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, 7(1), 1–26. DOI: 10.1214/aos/1176344552.
- [6] Efron, B., & Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. Chapman & Hall/CRC. ISBN: 978-0412042317.
- [7] Greene, W. H. (2018). *Econometric Analysis* (8th ed.). Pearson. ISBN: 978-0134461533.
- [8] Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). Springer. ISBN: 978-0387848570.
- [9] Hull, J. C. (2018). *Options, Futures, and Other Derivatives* (10th ed.). Pearson Education. ISBN: 978-0134472089.
- [10] Jarque, C. M., & Bera, A. K. (1987). A test for normality of observations and regression residuals. *International Statistical Review*, 55(2), 163–172. DOI: 10.2307/1403192.
- [11] Levene, H. (1960). Robust tests for equality of variances. In I. Olkin et al. (Eds.), *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling* (pp. 278–292). Stanford University Press.
- [12] MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, 1, 281–297.
- [13] Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance*, 7(1), 77–91. DOI: 10.2307/2975974.

- [14] Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65. DOI: 10.1016/0377-0427(87)90125-7.
- [15] Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3), 581–592. DOI: 10.1093/biomet/63.3.581.
- [16] Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3–4), 591–611. DOI: 10.1093/biomet/52.3-4.591.
- [17] Sharpe, W. F. (1964). Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. *The Journal of Finance*, 19(3), 425–442. DOI: 10.2307/2977928.
- [18] Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B*, 63(2), 411–423. DOI: 10.1111/1467-9868.00293.
- [19] Tukey, J. W. (1977). *Exploratory Data Analysis*. Addison-Wesley Publishing Company. ISBN: 978-0201076165.
- [20] Wasserman, L. (2004). *All of Statistics: A Concise Course in Statistical Inference*. Springer. ISBN: 978-0387402727.
- [21] Welch, B. L. (1947). The generalization of ‘Student’s’ problem when several different population variances are involved. *Biometrika*, 34(1–2), 28–35. DOI: 10.1093/biomet/34.1-2.28.
- [22] Wooldridge, J. M. (2020). *Introductory Econometrics: A Modern Approach* (7th ed.). Cengage Learning. ISBN: 978-1111531041.