

#66



Observatorio de Divulgación Financiera

Septiembre 2026

Depósito legal: B121662-2012

NOTA TÉCNICA

Identificación de los factores determinantes del precio del Eurostoxx 50 mediante Análisis de Componentes Principales (PCA) y Regresión sobre Componentes Principales (PCR), 2005-2025

Carles Ruiz Rovira

Resumen

El objetivo de este estudio es analizar la evolución temporal de los factores económicos, financieros y operativos que condicionan el comportamiento de las empresas del Euro Stoxx 50 durante el período 2005-2025, identificando patrones latentes en los datos y modelizando su relación con el precio de cotización. Para lograr este propósito, se aplican el Análisis de Componentes Principales (PCA), una técnica estadística de reducción de la dimensionalidad, y la Regresión sobre Componentes Principales (PCR), una regresión lineal múltiple que utiliza los componentes obtenidos como variables independientes, sobre un conjunto de más de 300 variables. Los resultados permiten identificar patrones temporales claros, aunque no se observa un comportamiento homogéneo entre sectores. Asimismo, el modelo predictivo desarrollado para explicar y predecir la evolución del precio de las acciones de las empresas integrantes del índice revela que el entorno macroeconómico y la eficiencia financiera constituyen los factores con mayor incidencia sobre la valoración bursátil. En consecuencia, la metodología PCA-PCR se valida como una herramienta cuantitativa multivariable eficaz para dar soporte a la toma de decisiones de inversión, gestión de carteras y análisis financiero en los mercados de renta variable europeos.

Palabras clave: PCA · PCR · Euro Stoxx 50 · comportamiento bursátil · reducción de dimensionalidad · mercado europeo.

1. Introducción y contextualización

La predicción del comportamiento de los mercados bursátiles ha constituido, desde hace décadas, uno de los retos centrales de la investigación financiera contemporánea. La necesidad de anticipar los movimientos del precio de las acciones responde, por un lado, a la voluntad de los inversores de asignar sus recursos de manera eficiente y, por otro, al interés de los gestores corporativos por comprender los mecanismos que articulan la valoración de sus empresas en el ámbito de los mercados de capitales. Tradicionalmente, las aproximaciones al problema han combinado dos grandes familias de variables: las endógenas de la empresa —indicadores de liquidez, rentabilidad, solvencia, eficiencia y endeudamiento— y las exógenas a su perímetro contable —condiciones macroeconómicas, evolución del marco legal, dinámicas tecnológicas y factores de sentimiento de mercado.

El panorama financiero ha experimentado, durante las dos últimas décadas, una transformación profunda impulsada por dos grandes motores: la proliferación de información financiera accesible en tiempo real y el progreso exponencial en la capacidad de cálculo, recientemente reforzado por el avance de la inteligencia artificial. Esta transformación ha abierto la puerta a metodologías de análisis cada vez más sofisticadas, que han pasado de los modelos de regresión lineal clásicos a los sistemas híbridos que combinan técnicas estadísticas de reducción de la dimensionalidad. El resultado es un campo de investigación en el que la capacidad de integrar grandes volúmenes de variables —de naturaleza heterogénea— se ha convertido en una ventaja competitiva decisiva, tanto para los participantes profesionales de los mercados como para la investigación académica.

En este contexto, el presente trabajo se inscribe en la línea de la investigación cuantitativa aplicada al mercado bursátil europeo. El objetivo principal es analizar la evolución temporal de las variables económicas, financieras y operativas de las empresas que forman parte del índice Eurostoxx 50 durante el período 2005-2025, a partir de observaciones anuales agregadas por empresa, con la finalidad de identificar patrones latentes en los datos, reducir la

dimensionalidad del conjunto de variables y modelizar la relación entre estas variables y el precio de cotización.

El uso de una frecuencia anual responde a la necesidad de homogeneizar los datos contables, financieros y macroeconómicos disponibles, evitando los problemas derivados de la falta de comparabilidad temporal entre fuentes con frecuencias diferentes y garantizando la coherencia de los indicadores analizados a lo largo del período de estudio. Asimismo, el conjunto de datos supera las 300 variables explicativas, muchas de las cuales presentan elevados niveles de correlación y redundancia informativa. Esta situación genera problemas de multicolinealidad, incrementa la complejidad computacional y dificulta la interpretación económica de los resultados, hecho que justifica la aplicación de técnicas de reducción de dimensionalidad.

Para lograr este propósito, se aplican el Análisis de Componentes Principales (PCA) y la Regresión sobre Componentes Principales (PCR), que permiten condensar la información contenida en un número elevado de variables en un conjunto reducido de dimensiones latentes y utilizarlas posteriormente en la modelización del precio de las acciones. Aunque existen alternativas metodológicas para el tratamiento de datos de alta dimensionalidad, como la regresión LASSO¹, se ha optado por el enfoque PCA-PCR porque facilita simultáneamente la reducción de dimensionalidad, la mitigación de los problemas de multicolinealidad y la interpretación económica de los factores subyacentes. El análisis se ha desarrollado utilizando el lenguaje de programación R y el entorno integrado de desarrollo RStudio² como plataforma de computación estadística. La elección del Eurostoxx 50 como universo de referencia no es casual. Se trata del índice de referencia por excelencia del mercado de acciones de la zona euro, integrado por cincuenta empresas de gran capitalización bursátil distribuidas en once países y en sectores tan diversos como la banca, la energía, la industria, el consumo y la tecnología. Su composición garantiza una diversidad estructural significativa y una representatividad suficiente para el análisis comparativo entre empresas, sectores y ciclos económicos. La cobertura temporal de dos décadas —de 2005 a 2025— permite, además, incorporar múltiples fases del ciclo económico: la expansión previa a la crisis financiera global (2005-2008), la recesión y la recuperación (2008-2019), el impacto de la pandemia de la COVID-19 (2020) y el contexto inflacionario post pandemia (2021-2025).

¹LASSO (Least Absolute Shrinkage and Selection Operator) es una técnica de regresión regularizada propuesta por Tibshirani (1996) que realiza simultáneamente estimación y selección de variables mediante una penalización L1 sobre los coeficientes, especialmente útil en contextos con un elevado número de variables explicativas.

²RStudio es un entorno integrado de desarrollo (IDE) para el lenguaje de programación R que facilita la gestión de datos, el análisis estadístico, la programación y la visualización de resultados (Posit Team, 2024).

2. Marco teórico: indicadores, técnicas y literatura de referencia

2.1. Indicadores financieros endógenos

Los indicadores internos de las empresas siguen siendo la base sobre la que la mayoría de inversores y analistas construyen sus modelos de predicción de rentabilidad y crecimiento bursátil. La literatura clásica, con Fama y French a la cabeza, sentó las bases de los modelos multifactoriales, que han sido progresivamente ampliados para integrar nuevas categorías de variables. En el ámbito nacional, diversas investigaciones han puesto de relieve que los ratios de rotación de activos y los márgenes operativos presentan la mayor consistencia predictiva en períodos de alta volatilidad. Paralelamente, los indicadores basados en flujos de caja libre ajustados y la volatilidad de los beneficios han ofrecido resultados superiores a la rentabilidad contable cuando se incorpora el componente de riesgo de liquidez.

Más recientemente, el auge de la inteligencia artificial ha permitido derivar características sintéticas a partir de datos contables y de mercado, hasta el punto de proponer enfoques basados en redes neuronales de grafos ("Graph Neural Networks")³ para modelar las relaciones interempresariales, con la conclusión de que la conectividad dentro de las redes sectoriales mejora la predicción de precios por encima de las métricas contables aisladas. En todo caso, la literatura converge en una premisa central: los indicadores endógenos no deben interpretarse de forma aislada, sino como parte de sistemas de interacción entre eficiencia operativa, innovación y resiliencia financiera.

2.2. Factores macroeconómicos exógenos

El estudio de los factores macroeconómicos exógenos sigue siendo central para comprender la dinámica de los mercados financieros globales. La investigación reciente enfatiza la relación entre los ciclos económicos, la política monetaria y el comportamiento bursátil,

destacando cómo los cambios en la liquidez y en los tipos de interés condicionan las trayectorias sectoriales y geográficas. En el ámbito europeo, se ha evidenciado que el tipo de cambio efectivo del euro y los índices de confianza económica ("Economic Sentiment Index")⁴ resultan más predictivos de los rendimientos bursátiles que las tasas de paro o la inflación, especialmente en contextos de incertidumbre política y social.

Una línea emergente es la incorporación de factores climáticos y de sostenibilidad ambiental entre las variables exógenas. Determinadas investigaciones han mostrado que los choques climáticos extremos afectan tanto a la volatilidad de las series financieras como al dinamismo de los inversores en sectores como la energía y los seguros. En paralelo, los índices de incertidumbre de política económica (EPU)⁵ y la volatilidad implícita (VIX) se han consolidado como poderosos explicadores de los rendimientos a corto plazo, reforzando la noción de que los factores exógenos no únicamente influyen en las empresas a través de los mecanismos de transmisión macroeconómica, sino también mediante cambios en la percepción del riesgo sistémico.

2.3. Técnicas estadísticas y aprendizaje automático

La predicción bursátil ha transitado de los modelos estadísticos tradicionales —regresiones lineales, modelos ARIMA y GARCH⁶— hacia sistemas híbridos potenciados por el aprendizaje automático y el deep learning. Estudios recientes han puesto de relieve cómo las dinámicas de los mercados financieros presentan patrones no lineales que escapan al análisis convencional, y que requieren métodos capaces de aprender dependencias complejas y adaptativas. Dentro de este paradigma, el enfoque multimodal ha ganado preeminencia: los modelos incorporan variables financieras históricas, indicadores macroeconómicos, métricas ESG, indicadores de sentimiento y factores técnicos.

³Graph Neural Networks (GNN): modelos de aprendizaje profundo diseñados para procesar datos estructurados en forma de grafo, permitiendo capturar las relaciones e interdependencias existentes entre diferentes entidades conectadas. En el ámbito financiero, se utilizan para modelizar las conexiones entre empresas, sectores o mercados y mejorar la precisión predictiva de los modelos tradicionales.

⁴Economic Sentiment Index (ESI): indicador compuesto elaborado por la Comisión Europea que sintetiza las expectativas y percepciones de empresas y consumidores sobre la evolución de la actividad económica de la zona euro a partir de encuestas de confianza sectoriales y de consumo.

⁵Economic Policy Uncertainty (EPU): índice que mide el grado de incertidumbre asociado a la política económica mediante el análisis de contenido de medios de comunicación y otros indicadores relacionados con decisiones gubernamentales, regulación y política fiscal o monetaria.

⁶Los modelos ARIMA (AutoRegressive Integrated Moving Average) modelizan series temporales a partir de sus retardos y errores, mientras que los modelos GARCH (Generalized AutoRegressive Conditional Heteroskedasticity) capturan la volatilidad condicional de las series financieras (Bollerslev, 1986; Box y Jenkins, 1976).

En el ámbito de la reducción de la dimensionalidad, el PCA⁷ y los autoencodificadores⁸ han demostrado su utilidad para condensar la información contenida en grandes volúmenes de características financieras sin una pérdida significativa de capacidad predictiva. La Regresión sobre Componentes Principales (PCR) constituye, precisamente, una técnica híbrida que combina el PCA con la regresión lineal múltiple, con el objetivo de modelizar la relación entre una variable dependiente —en este caso, el precio de la acción— y un conjunto de variables explicativas, evitando los problemas derivados de la multicolinealidad y el exceso de dimensionalidad. En el presente trabajo, la combinación PCA-PCR se justifica, por un lado, por la naturaleza de los datos —más de 300 variables con un elevado grado de correlación entre ellas— y, por otro, por la conveniencia de obtener componentes latentes interpretables desde un punto de vista económico.

3. Diseño de la investigación y metodología

3.1. Muestra, datos y variables

La muestra del estudio comprende las cincuenta compañías que integran el índice Eurostoxx 50 a fecha 21 de septiembre de 2025, tal como han sido identificadas en las plataformas TIKR Terminal⁹ y Morningstar. El período de análisis abarca veinte años, de 2005 a 2025, lo que permite incorporar múltiples ciclos económicos relevantes y garantiza la robustez de los modelos predictivos, en línea con las recomendaciones metodológicas de la literatura especializada que subraya la importancia de incluir diversas fases del ciclo económico en la validación de modelos financieros; los datos se agregan anualmente. Las variables independientes se agrupan en cuatro categorías de carácter endógeno: el balance de situación (99 variables), la cuenta de resultados (90 variables), el estado de flujos de efectivo (79 variables) y los ratios financieros anualizados (74 variables). A estas se añaden diez variables macroeconómicas exógenas de la zona euro, obtenidas de Eurostat, de la

Comisión Europea y de Trading Economics: el Producto Interior Bruto anual, el índice de productividad, la tasa de paro, el salario medio mensual bruto, el índice del coste laboral, la formación bruta de capital fijo, el índice de mercado bursátil de la zona euro (EU50), la inflación medida por el HICP, la deuda pública total y el tipo de cambio EUR/USD al cierre de cada año. La variable dependiente es el precio de cotización de las acciones, concretamente el precio de cierre del último día hábil de cada año —o el 1 de octubre de 2025 para el año en curso. El número total de observaciones es $n = 50 \text{ empresas} \times 21 \text{ años} = 1.050$ observaciones empresa-año, con un conjunto de más de 300 variables independientes, hecho que justifica la aplicación del PCA para reducir la dimensionalidad.

3.2. Tratamiento y procesamiento de datos

El tratamiento previo de los datos ha seguido un protocolo riguroso para garantizar la calidad estadística del conjunto. En primer lugar, se han seleccionado las observaciones anuales por empresa y se ha reorganizado la base de datos en formato ancho (wide format), con cada fila representando una combinación empresa-año. En segundo lugar, se han convertido las cadenas vacías y los valores ausentes en nulos, y se han eliminado las columnas con más del 30% de datos faltantes. Finalmente, se ha aplicado una normalización individual para cada una de las variables —media cero y desviación estándar igual a uno— para evitar que las diferencias de magnitud distorsionen los resultados del PCA. El resultado es una matriz de datos limpia, homogénea y apta para el análisis multivariable.

3.3. Análisis de Componentes Principales (PCA)

El PCA es una técnica estadística de reducción de la dimensionalidad que transforma un conjunto de variables independientes altamente correlacionadas de forma lineal en un nuevo conjunto de variables no correlacionadas, denominadas componentes principales, mediante un proceso de proyección de la información original en un nuevo espacio de menor dimensión. Cada uno de estos componentes es una combinación lineal de las variables originales, ordenado en función de la cantidad de varianza que explica de la matriz de varianzas-covarianzas de las variables originales, denominada "S". En el caso de este estudio,

datos mediante una capa latente de menor dimensión, utilizadas habitualmente para la reducción de dimensionalidad y la extracción automática de características complejas en conjuntos de datos de elevada dimensión.

⁹TIKR Terminal: <https://tikr.com/> Morningstar: <https://www.morningstar.com/>

⁷El Análisis de Componentes Principales (PCA) es una técnica estadística de reducción de la dimensionalidad que transforma un conjunto de variables correlacionadas en componentes ortogonales (no correlacionados) ordenados por la varianza que explican. Referencia fundamental: Jolliffe, I.T. (2002). *Principal Component Analysis*. Springer. Aplicada en finanzas por: Gu, Kelly y Xiu (2024); Carbó-Valverde et al. (2020).

⁸Autoencodificadores (Autoencoders): redes neuronales no supervisadas que aprenden representaciones comprimidas de los

donde se trabaja con más de 300 variables financieras y macroeconómicas para 50 empresas y 21 años, el PCA permite superar tres retos clave: la multicolinealidad entre variables, el ruido y la redundancia informativa, y la imposibilidad de visualizar las relaciones en espacios de dimensión superior a tres.

Aunque existen variantes avanzadas como el Análisis de Componentes Principales No Lineal (NLPCA) —diseñada para capturar estructuras complejas mediante redes neuronales—, en este estudio se ha descartado su uso. Esta decisión se justifica por el elevado coste computacional de estas técnicas, el riesgo de sobreajuste (overfitting) ante la alta dimensionalidad temporal de los datos y, principalmente, por el objetivo de preservar la máxima interpretabilidad económica de los factores obtenidos, una cualidad que se perdería con la naturaleza de "caja negra" del enfoque no lineal.

El procedimiento de aplicación ha seguido cinco etapas. En primer lugar, se ha construido la matriz de entrada excluyendo la variable dependiente (PRICE) para evitar que influya en la extracción de los componentes latentes. En segundo lugar, se ha aplicado el algoritmo NIPALS¹⁰, que calcula los componentes principales de forma iterativa sin necesidad de imputar previamente los valores ausentes, especialmente útil en matrices de gran tamaño y con datos incompletos. En tercer lugar, para validar y complementar los resultados, se ha realizado una imputación de datos y un PCA clásico¹¹ con el número óptimo de componentes¹². En cuarto lugar, se ha realizado una agrupación automática de los perfiles anuales mediante el algoritmo K-means aplicado sobre las coordenadas medias por año. En quinto y último lugar, se ha calculado el perfil medio por empresa a lo largo del período, agregando las coordenadas de PCA por año, con el objetivo de clasificar las empresas por sector y detectar similitudes estructurales.

3.4. Regresión sobre componentes principales PCR

La Regresión de Componentes Principales (PCR, por sus siglas en inglés) es una técnica híbrida que combina el Análisis de Componentes Principales (PCA) con la regresión lineal múltiple, permitiendo modelizar la relación entre la variable dependiente (precio) y el conjunto de variables explicativas, haciendo uso de los componentes principales obtenidos con anterioridad. De este modo se eliminan los problemas de

multicolinealidad que frecuentemente invalidan las estimaciones de mercado.

En el marco del estudio, la PCR es adecuada por tres motivos: el elevado número de variables iniciales (más de 300 variables), la influencia simultánea de múltiples factores internos y externos sobre el precio, y la necesidad de un enfoque robusto y sintético que facilite la interpretación de los resultados.

La especificación estándar de este modelo toma la siguiente forma:

$$PRICE_i = \beta_0 + \beta_1 Dim.1_i + \dots + \beta_5 Dim.5_i + \varepsilon_i \quad (1)$$

donde $PRICE_i$ es el precio de cierre anual de la acción para la observación i , $Dim.k_i$ representa el valor del componente principal k para la misma observación, β_k son los coeficientes estimados por mínimos cuadrados ordinarios (MCO) y ε_i es el término de error aleatorio.

No obstante, la modelización directa de los precios en sus niveles originales presenta limitaciones severas en el ámbito de los mercados financieros, como la asimetría en la distribución, la heteroscedasticidad y problemas de omisión de dinámicas temporales y estructurales. Por este motivo, se amplía la investigación implementando una estrategia de estimación progresiva en seis estadios (modelos 0 a 5). Este enfoque permite aislar la aportación marginal de cada transformación y variable de control sobre la capacidad explicativa del modelo (R^2 ajustado) y garantiza la robustez de los coeficientes estimados:

- **Modelo 0 (original):** Constituye la línea base descrita en la ecuación (1), donde se modeliza el precio en niveles en función única de los componentes principales generados por el PCA.
- **Modelo 1 (Log-lineal):** Incorpora la transformación logarítmica de la variable dependiente ($\log(Price_i)$). Suavizando la asimetría de la distribución de precios, reduciendo el impacto de observaciones extremas (outliers) y reinterpretando los coeficientes en variaciones porcentuales.
- **Modelo 2 (inercia temporal/dinámico):** Añade el retardo de la variable dependiente en escala logarítmica ($\log(Price_{i,t-1})$) como variable explicativa. Permitiendo capturar la autocorrelación y la inercia inherente a los mercados financieros, capturando la desviación del valor del precio respecto a su propio pasado.
- **Modelo 3 (efectos fijos de tiempo):** Incorpora un vector de variables ficticias de año (any_factor), que absorbe los choques macroeconómicos globales y temporales (como crisis de mercado o cambios regulatorios) que

¹⁰NIPALS (Nonlinear Iterative Partial Least Squares): algoritmo iterativo para el cálculo del PCA en presencia de valores ausentes. Wold, H. (1966). Nonlinear estimation by iterative least squares procedures.

¹¹La PCA clásica se obtiene por descomposición espectral (eigendecomposition) de la matriz de covarianzas S o por

descomposición en valores singulares (SVD) de la matriz de datos centrada.

¹²El número óptimo de componentes se ha determinado combinando el criterio del gráfico de sedimentación (scree plot) y el umbral de varianza acumulada explicada del 60%, tal como se detalla en la sección de Análisis de resultados.

afectan de manera simultánea y homogénea a todas las empresas de la muestra en un mismo período.

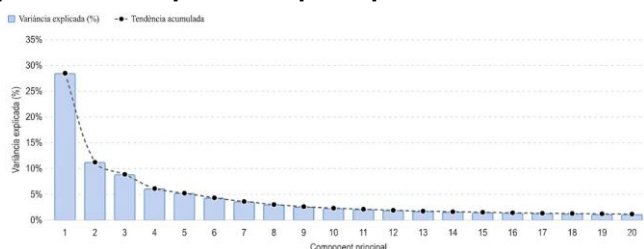
- **Modelo 4 (efectos fijos de empresa):** Sustituye los controles temporales por un vector de variables ficticias a nivel empresarial (Empresa_factor). Permitiendo capturar la heterogeneidad no observada e invariante en el tiempo de cada compañía.
- **Modelo 5 (Efectos fijos bidireccionales):** Integra simultáneamente los componentes de la PCA, el precio retardado, y los efectos fijos tanto de tiempo como de empresa. Permitiendo la limpieza del sesgo por variables omitidas tanto en la dimensión transversal como temporal, ofreciendo un marco de contraste más robusto para la validación de los componentes del modelo PCR.

4. Análisis de resultados

4.1. La varianza explicada y la composición de los componentes

La aplicación del PCA al conjunto global de datos ha permitido identificar la distribución de la varianza explicada entre los componentes principales. El primer componente principal explica aproximadamente el 27% de la varianza total, y los siguientes aportan menos información de forma decreciente. A partir del sexto componente, el incremento marginal de la varianza explicada se aplanan por debajo del 3%, lo que indica una pérdida de eficiencia en la incorporación de nuevas dimensiones. La suma acumulada de los cinco primeros componentes supera el 60% de la varianza total, umbral considerado adecuado en los estudios financieros para garantizar la interpretabilidad sin reducir la capacidad explicativa. En consecuencia, el modelo se construye a partir de estas cinco dimensiones latentes.

Figura 1. Gráfico de la varianza explicada por los 20 primeros componentes principales



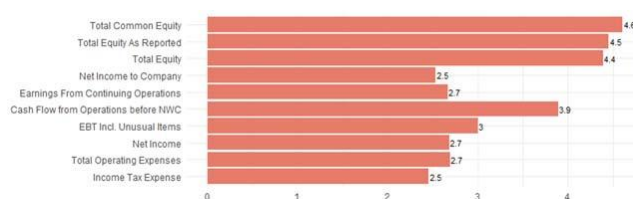
Fuente: Elaboración propia mediante R

El análisis de la composición interna de cada uno de los componentes permite identificar su significado económico y conocer la contribución de cada una de las variables que componen la dimensión.

Dim.1 recoge variables de patrimonio neto, resultados contables y flujo de caja operativo. Se interpreta como

la dimensión de estabilidad financiera y rentabilidad operativa, con una alta correlación interna entre net income y patrimonio neto acumulado.

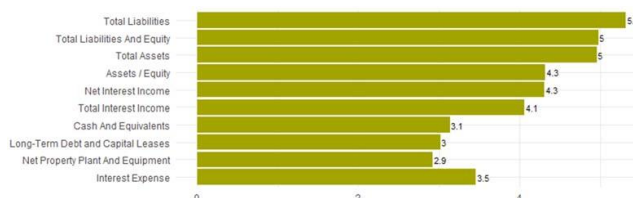
Figura 2. Contribución relativa de las variables a la Dimensión 1 (PCA).



Fuente: Elaboración propia mediante R

Dim.2 agrupa variables de escala y estructura de deuda (Total Assets, Total Liabilities, Net Interest Income). Se interpreta como la dimensión de escala operativa y exposición financiera, y captura principalmente las características del sector bancario y asegurador.

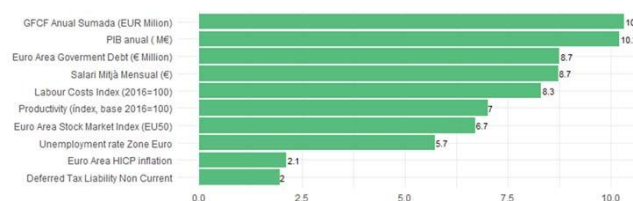
Figura 3. Contribución relativa de las variables a la Dimensión 2 (PCA).



Fuente: Elaboración propia mediante R

Dim.3 está dominada por las nueve variables macroeconómicas del estudio (GFCF, PIB, deuda pública, salario, coste laboral, productividad, EU50, paro e HICP), que aportan más del 65% de la contribución total. Se interpreta como la dimensión del entorno económico agregado de la zona euro.

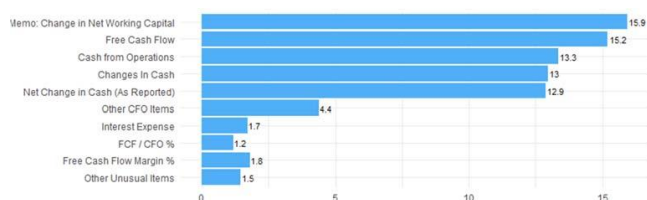
Figura 4. Contribución relativa de las variables a la Dimensión 3 (PCA).



Fuente: Elaboración propia mediante R

Dim.4 recoge variables de flujo de caja y liquidez (Free Cash Flow, Cash from Operations, Changes in Cash). Se interpreta como la dimensión de dinámica de liquidez y eficiencia operativa.

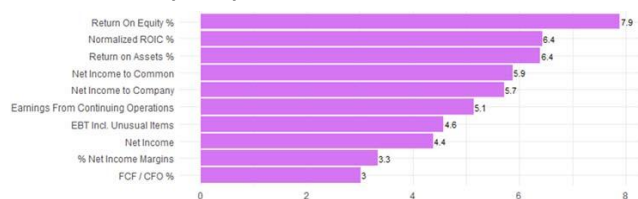
Figura 5. Contribución relativa de las variables a la Dimensión 4 (PCA).



Fuente: Elaboración propia mediante R

Dim.5 recoge indicadores de rentabilidad (ROE, ROIC, ROA, Net Income Margin). Se interpreta como la dimensión de rentabilidad y eficiencia financiera.

Figura 6. Contribución relativa de las variables a la Dimensión 5 (PCA).

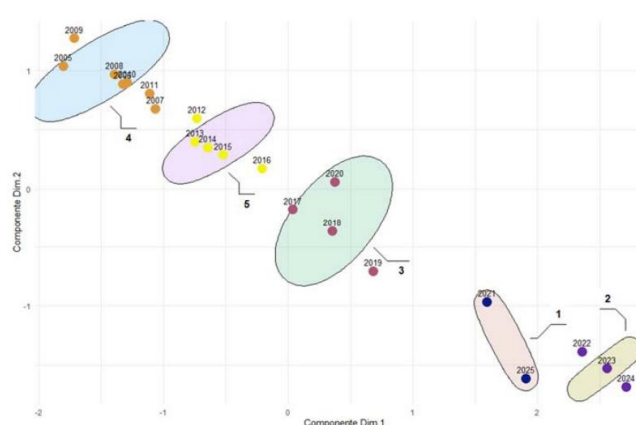


Fuente: Elaboración propia mediante R

4.2. Agrupación temporal: la huella de los ciclos económicos

El análisis gráfico de los perfiles anuales en el espacio bidimensional de los componentes principales revela una evolución temporal clara y coherente con los grandes ciclos económicos de la zona euro durante el período estudiado. Cuando se representa la distribución de los años en el plano Dim.1 versus Dim.2, se observan cuatro grandes agrupamientos secuenciales: los años 2005-2011 se concentran en la zona superior izquierda del gráfico, reflejando una etapa marcada por balances tensionados y un alto apalancamiento, con la crisis financiera global como punto de inflexión central; los años 2012-2016 se sitúan en el centro izquierdo, mostrando una etapa de recuperación y consolidación; los años 2017-2020 aparecen en el centro del gráfico, con estructuras patrimoniales estables y una actividad operativa moderada que caracteriza la fase de madurez; y los años 2021-2025 se desplazan hacia el extremo inferior derecho, sugiriendo una transformación estructural con menos exposición financiera pero con una sólida solidez contable.

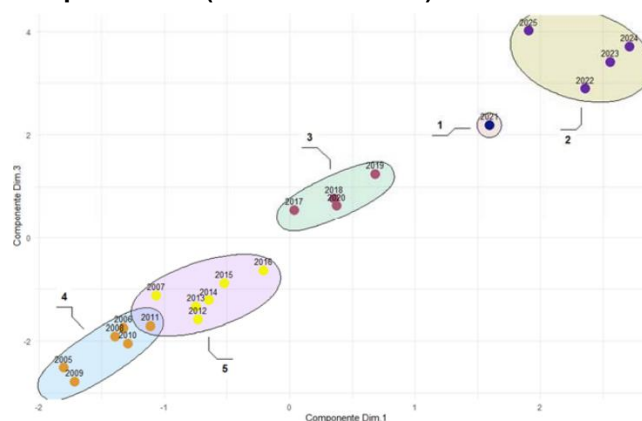
Figura 7. Gráfico agrupado por perfiles anuales en el espacio PCA (Dim. 1 vs. Dim. 2.)



Fuente: Elaboración propia mediante R

La proyección en el plano Dim.1 versus Dim.3 —que confronta la estabilidad contable con el contexto macroeconómico— es particularmente reveladora. Los años previos e inmediatamente posteriores a la crisis de 2008 se concentran en la zona inferior izquierda, combinando fragilidad contable y entorno macroeconómico desfavorable. El año 2020, marcado por la pandemia de la COVID-19, aparece sistemáticamente como un extremo en la mayoría de los gráficos construidos, confirmando el impacto disruptivo y singular de ese año sobre la eficiencia operativa y los flujos de caja agregados. Los años 2022-2025, en cambio, se sitúan en el extremo superior derecho, con valores elevados en ambas dimensiones, evidenciando un contexto de dinámica macroeconómica intensa —parcialmente vinculado a los procesos inflacionarios post pandemia— y de expansión estructural de las empresas del índice.

Figura 8. Gráfico agrupado por perfiles anuales en el espacio PCA (Dim. 1 vs. Dim. 3.)

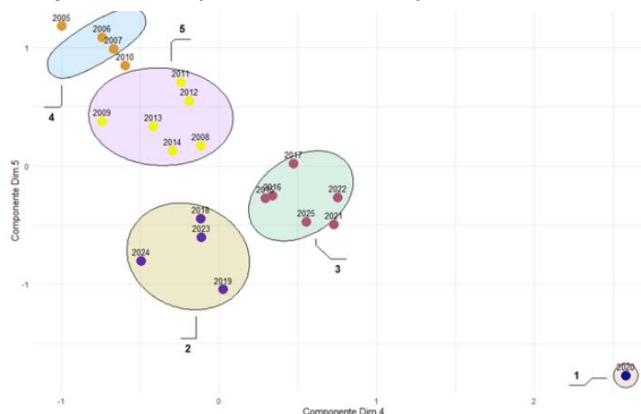


Fuente: Elaboración propia mediante R

Uno de los resultados más relevantes del análisis temporal es que las dimensiones vinculadas a factores estructurales y macroeconómicos —especialmente Dim.1, Dim.2 y Dim.3— generan agrupamientos claramente secuenciales y lineales a lo largo del tiempo. En contraste, las dimensiones relacionadas con la

gestión interna —Dim.4 y Dim.5— muestran agrupamientos menos lineales, reflejando que la evolución de la liquidez y la rentabilidad no sigue siempre una trayectoria temporal continua, sino que responde a cambios internos en la gestión, decisiones estratégicas puntuales o choques específicos de cada empresa o sector.

Figura 9. Gráfico agrupado por perfiles anuales en el espacio PCA (Dim. 4 vs. Dim 5.)

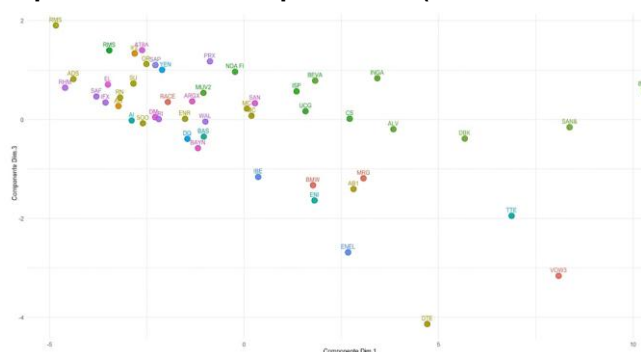


Fuente: Elaboración propia mediante R

4.3. Perfiles empresariales: la limitación de la segmentación sectorial

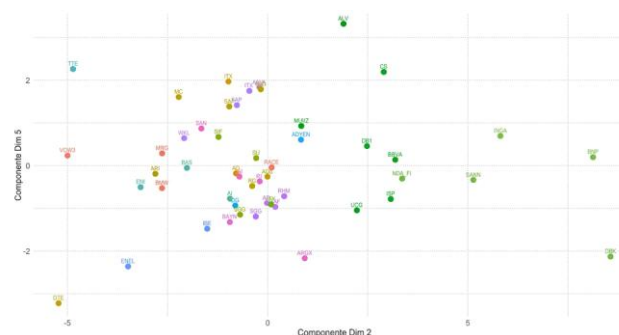
A diferencia de los patrones temporales, el análisis de los perfiles empresariales no ha permitido identificar agrupamientos sectoriales consistentes. Pese a la aplicación de las mismas técnicas de reducción de la dimensionalidad y de clustering, los resultados muestran una dispersión significativa entre sectores, sin una segmentación clara que permita diferenciar los comportamientos financieros agregados por actividad económica. La única excepción parcial es el sector financiero, que en algunos gráficos tiende a concentrarse en zonas específicas del espacio PCA, pero con una elevada variación interna que no permite establecer una diferenciación robusta.

Figura 10. Gráfico de agrupamiento por perfiles empresariales en el espacio PCA (Dim. 1 vs. Dim 3.)



Fuente: Elaboración propia mediante R

Figura 11. Gráfico de agrupamiento por perfiles empresariales en el espacio PCA (Dim. 2 vs. Dim 5.)



Fuente: Elaboración propia mediante R

Esta limitación tiene una lectura dual. Por un lado, pone de manifiesto que las variables estructurales capturadas por el PCA no son suficientes para discriminar los comportamientos financieros por actividad económica; dicho de otro modo, las similitudes entre empresas del mismo sector se diluyen cuando se analizan a través de dimensiones tan amplias como las consideradas en este estudio. Por otro, sugiere que los factores internos e idiosincráticos de cada empresa —su estrategia, su posición competitiva, su historia— ejercen una influencia superior sobre los perfiles financieros agregados que la mera pertenencia sectorial. Esta es, en sí misma, una conclusión de interés para la gestión de carteras y para el diseño de modelos de valoración empresarial.

4.4. Resultados de la PCR: coeficientes y capacidad predictiva

Los resultados obtenidos en la estimación de la línea base (modelo 0), donde se modeliza el precio en niveles en función exclusiva de los cinco componentes principales seleccionados anteriormente de la PCA, revelan una capacidad explicativa débil, con un coeficiente de determinación ajustado (R^2 ajustado) de 0,1599. Esta debilidad inicial evidencia que las relaciones directas en los mercados financieros están distorsionadas por fenómenos de heteroscedasticidad y de asimetría latentes en las series lineales originales.

Figura 12. Resumen del modelo lineal 0 de regresión de componentes principales

```
Call:
lm(formula = PRICE ~ Dim.1 + Dim.2 + Dim.3 + Dim.4 + Dim.5, data = datos_pcr)

Residuals:
    Min       1Q   Median       3Q      Max
-342.07  -81.05  -28.42   29.75  2067.74

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  101.559      7.275   13.961 < 2e-16 ***
Dim.1         -4.393      1.633   -2.690 0.007288 ***
Dim.2         -7.292      2.147   -3.396 0.000716 ***
Dim.3         31.271      2.745  11.391 < 2e-16 ***
Dim.4          7.333      3.037   2.415 0.015966 *
Dim.5         13.464      3.319   4.057 5.43e-05 ***

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 209.7 on 849 degrees of freedom
Multiple R-squared:  0.1648,    Adjusted R-squared:  0.1599
F-statistic: 33.52 on 5 and 849 DF,  p-value: < 2.2e-16
```

Fuente: Elaboración propia mediante R

Con la introducción de la transformación logarítmica sobre la variable dependiente (modelo 1), se consigue una corrección inmediata de la asimetría distributiva y del impacto de los valores extremos. En este estadio, el ajuste alcanza un incremento significativo al 0,2662 (R^2 ajustado). Desde el punto de vista de la significación estadística, destaca que cuatro de los cinco componentes principales generados por la PCA resultan altamente significativos a un nivel de confianza superior al 99,9% ($p_value < 0,001$), con la única excepción de la cuarta dimensión (flujo de caja y liquidez) donde $p_value = 0,026$. Estos resultados demuestran la eficacia de la reducción de la varianza para capturar los vectores latentes del mercado una vez estabilizada la varianza de la variable dependiente.

Figura 13. Resumen del modelo lineal 1 (Log-lineal) de regresión de componentes principales

```
call:
lm(formula = log_precio ~ Dim.1 + Dim.2 + Dim.3 + Dim.4 + Dim.5,
    data = datos_pcr)
```

Residuals:					
Min	1Q	Median	3Q	Max	
-4.0597	-0.7035	0.1405	0.7061	3.2924	

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.860869   0.036381 106.123 < 2e-16 ***
Dim.1        -0.034194   0.008168  -4.186 3.13e-05 ***
Dim.2        -0.081406   0.010740  -7.580 9.05e-14 ***
Dim.3         0.172865   0.013729 12.591 < 2e-16 ***
Dim.4         0.033866   0.015189  2.230  0.026 *
Dim.5         0.150499   0.016596  9.068 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.049 on 849 degrees of freedom
Multiple R-squared:  0.2705,    Adjusted R-squared:  0.2662
F-statistic: 62.95 on 5 and 849 DF,  p-value: < 2.2e-16
```

Fuente: Elaboración propia mediante

La transición hacia el modelo 2 (especificación dinámica) supone el cambio estructural de mayor profundidad del proceso de estimación al incorporar el retardo de la variable dependiente en escala logarítmica ($\log_price_lag1$). Los resultados obtenidos muestran un salto drástico en la capacidad explicativa, donde el R^2 ajustado se sitúa en un 0,8975. Esta magnitud confirma la dependencia estocástica altamente persistente y una fuerte memoria de corto plazo en el proceso de formación de precios, un comportamiento canónico en el análisis de series temporales financieras.

Figura 14. Resumen del modelo 2 (inercia temporal) de regresión de componentes principales

```
call:
lm(formula = log_precio ~ Dim.1 + Dim.2 + Dim.3 + Dim.4 + Dim.5 +
    log(PRICE_lag1), data = datos_pcr)
```

Residuals:					
Min	1Q	Median	3Q	Max	
-4.1586	-0.1510	0.0149	0.1767	2.4087	

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.373460   0.050099   7.454 2.23e-13 ***
Dim.1        -0.010881   0.003070  -3.545 0.000414 ***
Dim.2        -0.015469   0.004116  -3.758 0.000183 ***
Dim.3         0.034779   0.005475   6.353 3.45e-10 ***
Dim.4        -0.010938   0.005710  -1.916 0.055748 .
Dim.5         0.033631   0.006409   5.247 1.95e-07 ***
log(PRICE_lag1) 0.916787   0.012676  72.324 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.392 on 848 degrees of freedom
Multiple R-squared:  0.8982,    Adjusted R-squared:  0.8975
F-statistic: 1247 on 6 and 848 DF,  p-value: < 2.2e-16
```

Fuente: Elaboración propia mediante

La evaluación de la heterogeneidad no observada mediante efectos fijos aislados se ha articulado en los modelos 3 y 4. El modelo 3 incorpora el vector de variables ficticias temporales (any_factor), mientras que el modelo 4 hace énfasis en la heterogeneidad sectorial y empresarial mediante el esquema de efectos fijos de empresa ($empresa_factor$).

Figura 15. Resumen del modelo 3 (efectos fijos de tiempo¹³) de regresión de componentes principales

```
Call:
lm(formula = log_precio ~ Dim.1 + Dim.2 + Dim.3 + Dim.4 + Dim.5 +
    log(PRICE_lag1) + Any_factor, data = datos_pcr)
```

Residuals:					
Min	1Q	Median	3Q	Max	
-4.1763	-0.1273	-0.0023	0.1274	2.3111	

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.398187   0.111177   3.582 0.000361 ***
Dim.1        -0.012589   0.007463  -1.687 0.091997 .
Dim.2        -0.011867   0.008801  -1.348 0.177909 .
Dim.3         0.025176   0.028807   0.874 0.382391
Dim.4        -0.009278   0.009023  -1.028 0.304134
Dim.5         0.033412   0.013423   2.489 0.013002 *
log(PRICE_lag1) 0.928708   0.011902  78.032 < 2e-16 ***
---
Any_factor2007 -0.091693   0.088388  -1.037 0.299854
Any_factor2008 -0.610135   0.085475  -7.138 2.07e-12 ***
...
Any_factor2024 -0.065334   0.246398  -0.265 0.790955
Any_factor2025 -0.041679   0.251950  -0.165 0.868650
...
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3622 on 829 degrees of freedom
Multiple R-squared:  0.9151,    Adjusted R-squared:  0.9125
F-statistic: 357.3 on 25 and 829 DF,  p-value: < 2.2e-16
```

Fuente: Elaboración propia mediante

El modelo 4 alcanza un R^2 ajustado del 0,9125 y refleja de primera mano el impacto de choques macroeconómicos globales y transversales. Respecto al modelo 4, reporta un R^2 ajustado del 0,9106 y un error residual del 0,3661.

¹³El año de referencia utilizado en el modelo es 2005, correspondiente al primer año disponible de la muestra, respecto al cual se interpretan los coeficientes de las dummies temporales.

Figura 16. Resumen del modelo 4 (efectos fijos de empresa¹⁴) de regresión de componentes principales

```
Call:
lm(formula = log_precio ~ Dim.1 + Dim.2 + Dim.3 + Dim.4 + Dim.5 +
    log(PRICE_lag1) + Empresa_factor, data = datos_pcr)

Residuals:
    Min       1Q   Median       3Q      Max
-2.96952  -0.15297   0.01687   0.18014   2.37933

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.280282    0.127582   10.035 < 2e-16 ***
Dim.1        -0.005537    0.007773   -0.712  0.476506
Dim.2        -0.022640    0.012772   -1.773  0.076678
Dim.3         0.060779    0.007674    7.921  7.92e-15 ***
Dim.4        -0.007655    0.005742   -1.333  0.182810
Dim.5         0.040165    0.007883    5.095  4.35e-07 ***
log(PRICE_lag1) 0.698385    0.023124   30.201 < 2e-16 ***
---
Empresa_factorAD  -0.353791    0.155266   -2.279  0.022953 *
Empresa_factorADYEN 0.721806    0.190039    3.798  0.000157 ***
---
Empresa_factorBBVA -0.668939    0.161901   -4.132  3.98e-05 ***
Empresa_factorENEL -0.511521    0.128079   -3.994  7.10e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3661 on 799 degrees of freedom
Multiple R-squared:  0.9164,    Adjusted R-squared:  0.9106
F-statistic: 159.2 on 55 and 799 DF,  p-value: < 2.2e-16
```

Fuente: Elaboración propia mediante R

Finalmente, la estimación del modelo 5 (efectos fijos bidireccionales) cierra la secuencia mediante una estructura completa que integra los componentes de la PCA, la inercia temporal y los controles tanto transversales como longitudinales. Los resultados alcanzan el máximo nivel de ajuste de los seis modelos presentados, con un R^2 ajustado del 0,9258 y el error residual más bajo de la serie (0,3335). El comportamiento de este modelo es robusto, dado que el efecto neto de los componentes de la PCR se mantiene pese a la inclusión de una matriz exigente de controles bitemporales.

Figura 17. Resumen del modelo 5 (efectos fijos bidireccionales) de regresión de componentes principales

```
Call:
lm(formula = log_precio ~ Dim.1 + Dim.2 + Dim.3 + Dim.4 + Dim.5 +
    log(PRICE_lag1) + Any_factor + Empresa_factor, data = datos_pcr)

Residuals:
    Min       1Q   Median       3Q      Max
-2.86914  -0.13466   0.00853   0.12962   2.28137

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.97649    0.14322    6.818  1.85e-11 ***
Dim.1        -0.04090    0.01065   -3.838  0.000134 ***
Dim.2         0.03593    0.01601    2.244  0.025101 *
Dim.3        -0.09700    0.03495   -2.776  0.005641 **
Dim.4        -0.04598    0.01014   -4.534  6.70e-06 ***
Dim.5         0.10683    0.01640    6.515  1.31e-10 ***
log(PRICE_lag1) 0.69002    0.02306   29.920 < 2e-16 ***
---
Any_factor2007  0.04981    0.08409    0.592  0.553766
Any_factor2008 -0.50235    0.07979   -6.296  5.10e-10 ***
---
Any_factor2024  1.20974    0.29670    4.077  5.02e-05 ***
Any_factor2025  1.27597    0.30447    4.191  3.10e-05 ***
---
Empresa_factorAD  -0.47149    0.14488   -3.254  0.001185 **
Empresa_factorADYEN 0.62781    0.17474    3.593  0.000348 **
---
Empresa_factorBBVA -0.82534    0.15207   -5.427  7.64e-08 ***
Empresa_factorENEL -0.57312    0.12059   -4.752  2.39e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3335 on 780 degrees of freedom
Multiple R-squared:  0.9322,    Adjusted R-squared:  0.9258
F-statistic: 145 on 74 and 780 DF,  p-value: < 2.2e-16
```

Fuente: Elaboración propia mediante R

El análisis de los coeficientes estimados es revelador en cuanto al peso relativo de cada una de las dimensiones latentes en la formación del precio. La Dim.5 —rentabilidad y eficiencia financiera— presenta el

coeficiente positivo más elevado y estadísticamente significativo de todo el bloque de componentes ($\beta = 0,10683$; $p_value = 1,31 \times 10^{-10}$), lo que confirma que las mejoras en la gestión financiera eficiente ejercen un impacto positivo y de gran intensidad sobre el precio de las acciones, consolidándose como un factor clave en la valoración empresarial. La Dim.2 —escala operativa y exposición financiera— presenta el segundo coeficiente positivo del modelo ($\beta = 0,03593$; $p_value = 0,025101$), aunque con una significación estadística más moderada, sugiriendo que su influencia en el precio, aun siendo relevante, es la menos robusta de las dimensiones positivas.

En sentido contrario, la Dim.3 —el entorno macroeconómico— presenta el impacto negativo más intenso y significativo ($\beta = -0,09700$; $p_value = 0,005641$), lo que sugiere que determinadas configuraciones de este entorno —una vez controlado el efecto inercial del propio mercado— se asocian a precios menos elevados. Siguiendo esta tendencia negativa, las dimensiones Dim.4 —dinámica de liquidez y eficiencia operativa— y Dim.1 —estabilidad financiera y rentabilidad contable— también se muestran robustamente significativas e inversas ($\beta = -0,04598$ con $p_value = 6,70 \times 10^{-6}$ y $\beta = -0,04090$ con $p_value = 0,000134$, respectivamente), evidenciando cómo el peso de los pasivos o determinados incrementos de escala operan como factores de descuento en la valoración final una vez aislado el fuerte componente autorregresivo de la variable temporizada ($\log(PRICE_lag1)$: $\beta = 0,69002$; $p_value < 2 \times 10^{-16}$).

Para evaluar la consistencia de la estrategia secuencial implementada, la siguiente tabla X detalla de forma agregada la evolución de las métricas de rendimiento y la ganancia incremental respecto a la formulación original del modelo 0 hasta llegar al modelo 5.

Figura 18. Comparativa del R^2 ajustado y mejora respecto al modelo base (M0)

	Model	$R^2_ajustat$	Millora_vs_M0
M0: original (PRICE, 5 comp)		0.1599	+0%
M1: log(PRICE) + 5 comp		0.2662	+10.6%
M2: M1 + lag preu		0.8975	+73.8%
M3: M2 + dummies any		0.9125	+75.3%
M4: M2 + dummies empresa		0.9106	+75.1%
M5: M2 + dummies any + empresa		0.9258	+76.6%

Fuente: Elaboración propia mediante R

El análisis comparativo de las métricas del modelo permite extraer conclusiones metodológicas clave. La evolución del R^2 ajustado demuestra que, más allá de la mejora elemental que supone el tratamiento logarítmico de la variable dependiente en el modelo 1, el verdadero salto cualitativo (+73,8%) se deriva de modelizar la

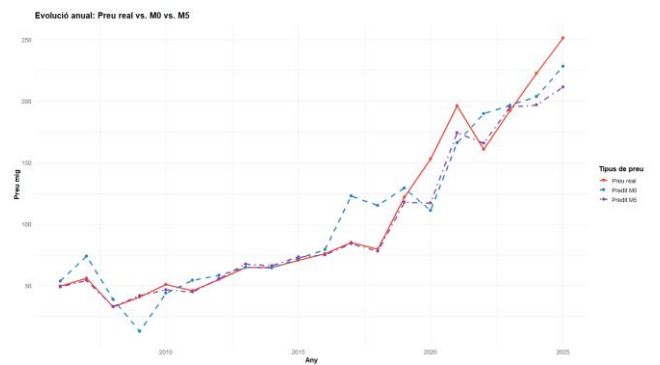
¹⁴La empresa de referencia utilizada en el modelo es ABI - Anheuser-Busch InBev, correspondiente a la primera unidad en orden

alfanumérico de la base de datos utilizada, respecto a la cual se interpretan los coeficientes de las dummies de empresa.

naturaleza autorregresiva del mercado en el modelo 2, hecho que reduce simultáneamente el error estándar residual de 1,049 a 0,392. Por otra parte, la evaluación de los efectos fijos sugiere que las fluctuaciones temporales de carácter macroeconómico (modelo 3) ejercen una tracción ligeramente superior sobre la varianza de los precios del Eurostoxx 50 en el período estudiado que el efecto de las empresas (modelo 4).

Finalmente, la convergencia formal en el modelo 5 bidireccional demuestra la complementariedad de los dos vectores de control. Al alcanzar una ganancia neta del 76,6% de explicabilidad y minimizar el error residual a su nivel óptimo (0,3335), se valida empíricamente la estrategia progresiva adoptada. Esta especificación final no solo optimiza el ajuste numérico, sino que garantiza que los coeficientes de los componentes de la PCR se encuentren libres de interferencias estructurales y coyunturales, consolidándose como la solución óptima para la inferencia de este caso.

Figura 19. Gráfico de la evolución anual del precio medio real vs. precio medio predicho¹⁵ por el modelo 0 y el modelo 5 (2005-2025)



Fuente: Elaboración propia mediante R

El análisis visual de la capacidad predictiva a través de la evolución del precio medio real y los precios predichos por los modelos 0 y 5 durante el período estudiado revela patrones altamente diferenciados según el nivel de especificación. Se visualiza claramente cómo el modelo final (M5) exhibe un ajuste superior a la hora de reproducir la tendencia general del mercado, amortiguando de manera eficiente las fuertes desviaciones que sufre el modelo base (M0) en momentos de alta volatilidad o disrupciones externas. Tal y como se detecta en el ciclo 2007 a 2009 o en el choque de 2017. Concretamente, el modelo 0 sufre severas sobreestimaciones y picos de error relativo respecto al precio real. En contraste, la línea predictiva de M5 traza casi a la perfección la trayectoria real en

estos subperíodos, demostrando la corrección de la matriz bitemporal de efectos fijos.

No obstante, hacia el tramo final de la serie histórica, emergen desajustes derivados de la propia naturaleza de los componentes. El modelo final subestima notablemente el precio real durante la fuerte escalada alcista del final del período, sobre todo el año 2025, tramo en el que el modelo base M0 mantiene una trayectoria ascendente algo más cercana a la realidad. Estas divergencias reflejan la complejidad de aislar los componentes latentes en ciclos de mercado post pandémicos altamente expansivos. Aunque el modelo 5 se consolida como una especificación de mayor robustez y menos errática en el cómputo global del panel, las desviaciones observadas en el último tramo ponen de manifiesto la presión de factores exógenos que escapan a la descomposición de la PCA, tales como la innovación estratégica o rupturas abruptas en las expectativas de los inversores.

Tabla 1. Diferencias de precios predichos entre los modelos M0 y M5 con los precios medios reales del Eurostoxx 50 para el período 2006 a 2025

Año	Precio medio real	Predicción M0	Diferencia Real con M0	Predicción M5	Diferencia Real con M5
2006	49,50	53,60	8,43%	49,00	-0,99%
2007	56,00	73,80	31,80%	54,20	-3,21%
2008	32,70	38,80	18,60%	33,00	0,67%
2009	40,80	12,90	-68,30%	41,70	2,19%
2010	50,90	43,80	-13,90%	46,50	-8,62%
2011	46,00	54,20	18,00%	44,80	-2,49%
2012	55,00	58,30	6,10%	55,80	1,45%
2013	64,80	64,80	-0,05%	67,50	4,19%
2014	64,90	64,40	-0,82%	66,10	1,84%
2015	70,70	71,70	1,53%	73,10	3,44%
2016	75,80	79,20	4,50%	75,10	-0,93%
2017	84,90	123,00	44,80%	84,20	-0,90%
2018	79,70	115,00	44,50%	78,10	-2,04%
2019	122,00	129,00	6,12%	118,00	-3,30%
2020	153,00	111,00	-27,40%	117,00	-23,40%
2021	196,00	166,00	-15,30%	174,00	-11,20%
2022	161,00	190,00	17,90%	166,00	3,12%
2023	192,00	196,00	2,21%	195,00	1,61%
2024	222,00	204,00	-8,49%	197,00	-11,50%
2025	251,00	228,00	-9,10%	211,00	-15,90%

Fuente: Elaboración propia con R

¹⁵El precio predicho corresponde a los valores ajustados del modelo PCR sobre la muestra analizada, calculados como media anual de los valores estimados por empresa.

5. Discusión y conclusiones

5.1. Principales aportaciones

La investigación presentada en este estudio amplía sustancialmente la evidencia empírica sobre la aplicabilidad de la combinación PCA-PCR para el análisis del comportamiento bursátil europeo mediante una estrategia metodológica progresiva e iterativa. Mediante la incorporación de cinco modelos adicionales a la PCR clásica, se ha logrado un refinamiento econométrico en la significación estadística de los coeficientes, así como en la robustez de la estructura del modelo.

Las cinco dimensiones latentes identificadas sintetizan de manera coherente los principales ejes de comportamiento empresarial —estabilidad financiera, escala operativa, entorno macroeconómico, dinámica de liquidez y eficiencia financiera—, y la arquitectura de los cinco modelos confirma con mayor solidez su capacidad explicativa cuando se integran en especificaciones dinámicas y con estructuras de control más sofisticadas.

Los resultados muestran que la mejora metodológica implementada no se limita a un incremento cuantitativo del rendimiento predictivo, sino que refuerza cualitativamente la consistencia de inferencia del modelo. La progresión desde el modelo base hasta los modelos avanzados evidencia una reducción sistemática de los errores de estimación y una estabilización notable de los coeficientes, hecho que consolida la validez de los factores latentes identificados por la PCA como vectores explicativos relevantes del comportamiento del mercado.

El análisis comparativo de los resultados anuales muestra una clara superioridad de los modelos avanzados respecto a la especificación inicial. Mientras que el modelo base presentaba desviaciones importantes en episodios de tensión de mercado, los modelos refinados consiguen reproducir con mayor precisión las oscilaciones estructurales del período analizado. Esta mejora es especialmente significativa en los episodios de alta volatilidad, donde la nueva especificación demuestra una capacidad superior para absorber choques coyunturales y minimizar errores extremos.

La incorporación secuencial de nuevas especificaciones también ha permitido verificar empíricamente que la naturaleza autorregresiva de los precios y el control simultáneo de la heterogeneidad temporal y transversal

constituyen elementos esenciales para lograr una modelización robusta del mercado europeo.

5.2. Implicaciones para la práctica profesional

Los resultados de este estudio tienen implicaciones relevantes para diversos colectivos profesionales. Para los gestores de carteras e inversores institucionales, la confirmación de que el contexto macroeconómico ejerce la influencia más potente sobre los precios del Eurostoxx 50 subraya la importancia de disponer de modelos de seguimiento de los ciclos económicos de la zona euro que sean ágiles y actualizados. En particular, la identificación del PIB, la formación bruta de capital fijo y los índices de productividad como principales variables macroeconómicas del tercer componente sugiere que las señales de aceleración o desaceleración económica agregada deben tener una ponderación elevada en las decisiones de asset allocation en renta variable europea.

Para los analistas de empresa y directores financieros, la dimensión de rentabilidad y eficiencia financiera —que recoge el ROE, el ROIC y el ROA como variables principales— confirma que las empresas que optimizan el uso de sus recursos tienden a presentar precios más elevados y estables en el tiempo. Esto refuerza la utilidad de los ratios de eficiencia como indicadores avanzados de la valoración bursátil, especialmente en un contexto en el que el mercado premia de forma creciente la creación de valor sobre la mera escala operativa.

Para los equipos de investigación y los profesionales de la gestión cuantitativa, el trabajo demuestra que la combinación PCA-PCR representa una herramienta valiosa para el diagnóstico estructural y la modelización predictiva en contextos de alta dimensionalidad, aunque debe interpretarse como una herramienta complementaria dentro de un marco analítico más amplio que incorpore técnicas no lineales, variables cualitativas e indicadores de sentimiento de mercado.

5.3. Limitaciones y líneas de investigación futuras

El estudio presenta diversas limitaciones que conviene tener presentes en la interpretación de los resultados. En primer lugar, la capacidad explicativa limitada del modelo PCR —con un R^2 ajustado del 15,21%— indica que una parte sustancial de la variabilidad del precio no queda capturada por las cinco dimensiones latentes, lo que refleja la influencia de factores externos y cualitativos no tratados: reputación de marca, innovación, posicionamiento estratégico, expectativas de mercado y sentimiento de los inversores. En segundo lugar, tanto el PCA como la PCR asumen relaciones lineales entre las variables, lo que limita la capacidad del

modelo para capturar dinámicas no lineales o interacciones complejas comunes en entornos económicos reales. En tercer lugar, la comparación entre el precio real y el predicho se ha realizado de forma agregada por año, lo que puede suavizar errores individuales por empresa y ocultar desviaciones significativas en casos particulares.

A partir de estas limitaciones, se identifican diversas líneas de investigación futuras de gran interés. La incorporación de variables cualitativas —indicadores ESG, métricas de reputación e indicadores de innovación— podría mejorar sustancialmente la capacidad explicativa del modelo. La aplicación de técnicas no lineales —Random Forest, XGBoost o redes neuronales recurrentes de tipo LSTM— o LASSO como alternativa de regularización en contextos de alto número de variables, permitiría capturar las relaciones complejas entre componentes y precio en contextos de alta dispersión. Finalmente, la validación del modelo con datos de otros índices bursátiles internacionales y su extensión a nivel de empresa individual —en lugar del planteamiento agregado aquí adoptado— constituirían pasos naturales para evaluar la robustez y la generalización de los resultados.

5.4. Reflexión de síntesis

La ampliación metodológica desarrollada en este trabajo confirma que el enfoque PCA-PCR adquiere su máximo potencial cuando se aplica dentro de un esquema iterativo de refinamiento econométrico orientado a reforzar simultáneamente capacidad predictiva, significación estadística y robustez estructural.

La incorporación de los cinco modelos adicionales ha permitido trascender las limitaciones iniciales de la especificación base y construir una arquitectura analítica notablemente más sólida, capaz de capturar con mayor precisión la naturaleza multifactorial del comportamiento bursátil europeo.

Los resultados obtenidos evidencian que la combinación entre reducción de dimensionalidad, modelización dinámica y controles estructurales constituye una aproximación especialmente eficaz para analizar entornos financieros de alta complejidad. No obstante, también recuerdan que ningún modelo lineal es capaz de explicar íntegramente la dinámica de los mercados de capitales.

El valor real de este enfoque radica, por tanto, no solo en su capacidad predictiva, sino en su utilidad como instrumento riguroso de interpretación estructural, integrable dentro de un marco analítico más amplio que combine solidez estadística, criterio económico y capacidad crítica de interpretación contextual.

Bibliografía

- Baley, R., & Turen, M. (2025). Machine learning applications in Euro Stoxx forecasting: Comparative precision analysis. *Journal of Quantitative Finance and Economics*, 32(4), 441–462.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3), 307–327.
- Box, G.E.P., & Jenkins, G.M. (1976). *Time Series Analysis: Forecasting and Control*. Holden-Day.
- Boucher, M., & Esnault, A. (2023). Prévisions boursières et économie réelle en Europe: Un modèle intégratif. *Revue d'Économie Financière*, 147(2), 55–73.
- Carbó-Valverde, S., Cuadros-Soler, J., & Rodríguez-Fernández, J. (2020). High performance computing in financial prediction: From regression to AI models. *Computational Economics*, 55(1), 23–47.
- Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1), 1–22.
- Fernández-Torres, A. (2024). Macroeconomic risk and equity volume: Evidence from European sovereign markets. *European Economic Review*, 162, 104398.
- Gómez-Pérez, J., & Cuadrado, I. (2022). Desempeño operativo y predicción bursátil en mercados ibéricos. *Estudios Financieros*, 22(7), 43–59.
- Gu, S., Kelly, B., & Xiu, D. (2024). Empirical asset pricing via machine learning: Factor models revisited. *Review of Financial Studies*, 37(1), 1–35.
- Jolliffe, I.T. (2002). *Principal Component Analysis* (2.^a ed.). Springer.
- Lemoine, C., & Dubois, F. (2024). Résilience financière et solvabilité: Nouvelles métriques de prévision dans les cycles de crise. *Finance et Développement Durable*, 29(3), 93–112.
- Liu, J., Wang, K., & Hu, Z. (2024). Graph neural networks for firm interconnectivity in financial forecasting. *Neural Computing and Applications*, 37(2), 303–324.
- Mistol, L., & Möhler, F. (2023). A benchmark study of supervised and unsupervised machine learning techniques for stock market prediction. *Quantitative Finance Letters*, 15(2), 156–175.
- Nguyen, T., Lam, H., & Chow, W. (2024). Interest rates and asymmetric responses in Asian stock markets. *Asian Economic Journal*, 38(1), 89–112.
- Nicolosi, A., & Tartaglia, C. (2023). Bayesian momentum and volatility in European technology assets. *Quantitative Finance*, 23(9), 1510–1528.
- Posit Team. (2024). *RStudio: Integrated Development Environment for R* (Version 2024.12). Posit Software, PBC. Boston, MA. <https://posit.co/download/rstudio-desktop/>
- Ruíz-García, N., & Méndez, V. (2024). Uncertainty indices and risk perception in emerging markets. *Cuadernos de Economía y Dirección de la Empresa*, 27(4), 366–382.
- Su, H. (2025). Sustainability integration in financial machine learning models: ESG and profitability alignment in Europe. *European Journal of Sustainable Finance*, 11(3), 192–218.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B*, 58(1), 267–288.
- Tridico, P., Morales, J., & Wang, D. (2024). Neural temporal models and macro volatility: Deep learning for adaptive markets. *Finance Research Letters*, 62, 104578.
- Wang, X., Chen, Y., & Zhou, J. (2025). Climate shocks and stock returns: Evidence from global energy markets. *Energy Economics*, 127, 106889.
- Wold, H. (1966). Nonlinear estimation by iterative least squares procedures. In *Research Papers in Statistics* (pp. 411–444). Wiley.
- Zhou, X., & Luan, Y. (2025). Dimensionality reduction and clustering in financial forecasting using autoencoders. *Expert Systems with Applications*, 238, 121555.

Sobre el autor

Carles Ruiz Rovira es graduado en Economía, con especialización en análisis político y marketing. Actualmente amplía su formación en finanzas y análisis de datos, así como en programación estadística. Es Máster en Finanzas por el Institut d'Estudis Financers (IEF).

Otras publicaciones

Sep 2026	NT	Identificación de los factores determinantes del precio del Eurostoxx 50 mediante Análisis de Componentes Principales (PCA) y Regresión sobre Componentes Principales (PCR), 2005-2025	Carles Ruiz Rovira
Jun 2026	NT	La Paradoja de la Calidad	
Jun 2026	DT	Bitcoin en Tesorería – Marco, métricas y Gobierno	Ferran Carreras Lavila, Eric Garcia Ribera
Jun 2026	DT	La huella de carbono de la investigación en economía en la era de la IA generativa	Andres Alonso-Robisco
Jun 2026	DT	Ahorro y riqueza intergeneracional ante el reto demográfico	Fernando Rojas, Pablo Victoria, Diego Romero, Francisco del Olmo García
May 2026	DT	Aplicación de sistemas multiagente de IA	Júlia Pou, Eric Garcia, Miquel Heredia, Carlos Salinas, Gerard Alba
	DT	Comportamiento racional en los mercados financieros	Antonio Hormigos
Mar 2026	DT	Stablecoins en Tesorería: Usos, Ventajas y Riesgos	Carlos de Fuenmayor
Dic 2025	DT	Fondos de Inversión Alternativa: Private Equity y Deuda Privada	Josep Martínez Pascual
Nov 2025	DT	La energía nuclear en el marco de la transición energética	Antonio González Jiménez
Oct 2025	DT	Artificial Intelligence in Economics and Finance: An AI-driven Literature Review	Andres Alonso-Robisco