

#66

Observatori de Divulgació Financera



Setembre 2026

Dipòsit legal: B121662-2012

NOTA TÈCNICA

Identificació dels factors determinants del preu de l'Eurostoxx 50 mitjançant Anàlisi de Components Principals (PCA) i Regressió sobre Components Principals (PCR), 2005-2025

Carles Ruiz Rovira

Abstract

L'objectiu d'aquest estudi és analitzar l'evolució temporal dels factors econòmics, financers i operatius que condicionen el comportament de les empreses de l'Euro stoxx 50 durant el període 2005-2025, identificant patrons latents en les dades i modelitzant la seva relació amb el preu de cotització. Per assolir aquest propòsit, s'apliquen l'Anàlisi de Components Principals (PCA), una tècnica estadística de reducció de la dimensionalitat, i la Regressió sobre Components Principals (PCR), una regressió lineal múltiple que utilitza els components obtinguts com a variables independents, sobre un conjunt de més de 300 variables. Els resultats permeten identificar patrons temporals clars, tot i que no s'observa un comportament homogeni entre sectors. Així mateix, el model predictiu desenvolupat per explicar i predir l'evolució del preu de les accions de les empreses integrants de l'índex revela que l'entorn macroeconòmic i l'eficiència financera constitueixen els factors amb major incidència sobre la valoració borsària. En conseqüència, la metodologia PCA-PCR es valida com una eina quantitativa multivariable eficaç per donar suport a la presa de decisions d'inversió, gestió de carteres i anàlisi financera en els mercats de renda variable europeus.

Paraules clau: PCA · PCR · Euro stoxx 50 · comportament borsari · reducció de dimensionalitat · mercat europeu.

1. Introducció i contextualització

La predicció del comportament dels mercats borsaris ha constituït, des de fa dècades, un dels reptes centrals de la recerca financera contemporània. La necessitat d'anticipar els moviments del preu de les accions respon, d'una banda, a la voluntat dels inversors d'assignar els seus recursos de manera eficient i, de l'altra, a l'interès dels gestors corporatius per comprendre els mecanismes que articulen la valoració de les seves empreses en l'àmbit dels mercats de capitals. Tradicionalment, les aproximacions al problema han combinat dues grans famílies de variables: les endògenes de l'empresa —indicadors de liquiditat, rendibilitat, solvència, eficiència i endeutament— i les exògenes al seu perímetre comptable —condicions macroeconòmiques, evolució del marc legal, dinàmiques tecnològiques i factors de sentiment de mercat.

El panorama financer ha experimentat, durant les dues últimes dècades, una transformació profunda impulsada per dos grans motors: la proliferació d'informació financera accessible en temps real i el progrés exponencial en la capacitat de càlcul, recentment reforçat per l'avenç de la intel·ligència artificial. Aquesta transformació ha obert la porta a metodologies d'anàlisi cada cop més sofisticades, que han passat dels models de regressió lineal clàssics als sistemes híbrids que combinen tècniques estadístiques de reducció de la dimensionalitat. El resultat és un camp de recerca on la capacitat d'integrar grans volums de variables —de naturalesa heterogènia— s'ha convertit en un avantatge competitiu decisiu, tant per als participants professionals dels mercats com per a la recerca acadèmica.

En aquest context, el present treball s'inscriu en la línia de la recerca quantitativa aplicada al mercat borsari europeu. L'objectiu principal és analitzar l'evolució temporal de les variables econòmiques, financeres i operatives de les empreses que formen part de l'índex Eurostoxx 50 durant el període 2005-2025, a partir d'observacions anuals agregades per empresa, amb la finalitat d'identificar patrons latents en les dades, reduir la dimensionalitat del conjunt de variables i modelitzar la relació entre aquestes variables i el preu de cotització.

L'ús d'una freqüència anual respon a la necessitat d'homogeneïtzar les dades comptables, financeres i macroeconòmiques disponibles, evitant els problemes derivats de la manca de comparabilitat temporal entre fonts amb freqüències diferents i garantint la coherència dels indicadors analitzats al llarg del període d'estudi. Així mateix, el conjunt de dades supera les 300 variables explicatives, moltes de les quals presenten elevats nivells de correlació i redundància informativa. Aquesta situació genera problemes de multicol·linealitat, incrementa la complexitat computacional i dificulta la interpretació econòmica dels resultats, fet que justifica l'aplicació de tècniques de reducció de dimensionalitat.

Per assolir aquest propòsit, s'apliquen l'Anàlisi de Components Principals (PCA) i la Regressió sobre Components Principals (PCR), que permeten condensar la informació continguda en un nombre elevat de variables en un conjunt reduït de dimensions latents i utilitzar-les posteriorment en la modelització del preu de les accions. Tot i que existeixen alternatives metodològiques per al tractament de dades d'alta dimensionalitat, com ara la regressió LASSO¹, s'ha optat per l'enfocament PCA-PCR perquè facilita simultàniament la reducció de dimensionalitat, la mitigació dels problemes de multicol·linealitat i la interpretació econòmica dels factors subjacents. L'anàlisi s'ha desenvolupat utilitzant el llenguatge de programació R i l'entorn integrat de desenvolupament RStudio² com a plataforma de computació estadística. L'elecció de l'Eurostoxx 50 com a univers de referència no és casual. Es tracta de l'índex de referència per excel·lència del mercat d'accions de la zona euro, integrat per cinquanta empreses de gran capitalització borsària distribuïdes en onze països i en sectors tan diversos com la banca, l'energia, la indústria, el consum i la tecnologia. La seva composició garanteix una diversitat estructural significativa i una representativitat suficient per a l'anàlisi comparativa entre empreses, sectors i cicles econòmics. La cobertura temporal de dues dècades —de 2005 a 2025— permet, a més, incorporar múltiples fases del cicle econòmic: l'expansió prèvia a la crisi financera global (2005-2008), la recessió i la recuperació (2008-2019), l'impacte de la pandèmia de la COVID-19 (2020) i el context inflacionari post pandèmia (2021-2025).

¹LASSO (Least Absolute Shrinkage and Selection Operator) és una tècnica de regressió regularitzada proposada per Tibshirani (1996) que realitza simultàniament estimació i selecció de variables mitjançant una penalització L1 sobre els coeficients, especialment útil en contextos amb un elevat nombre de variables explicatives.

²RStudio és un entorn integrat de desenvolupament (IDE) per al llenguatge de programació R que facilita la gestió de dades, l'anàlisi estadística, la programació i la visualització de resultats (Posit Team, 2024).

2. Marc teòric: indicadors, tècniques i literatura de referència

2.1. Indicadors financers endògens

Els indicadors interns de les empreses continuen sent la base sobre la qual la majoria d'inversors i analistes construeixen els seus models de predicció de rendibilitat i creixement borsari. La literatura clàssica, amb Fama i French al capdavant, va assentar les bases dels models multifactorials, que han estat progressivament ampliat per integrar noves categories de variables. En el camp nacional, diverses investigacions han posat en relleu que els ràtios de rotació d'actius i els marges operatius presenten la major consistència predictiva en períodes d'alta volatilitat. Paral·lelament, els indicadors basats en fluxos de caixa lliure ajustats i la volatilitat dels beneficis han ofert resultats superiors a la rendibilitat comptable quan s'incorpora el component de risc de liquiditat.

Més recentment, l'auge de la intel·ligència artificial ha permès derivar característiques sintètiques a partir de dades comptables i de mercat, fins al punt de proposar enfocaments basats en xarxes neuronals de grafs ("Graph Neural Networks")³ per modelar les relacions inter empresarials, amb la conclusió que la connectivitat dins de les xarxes sectorials millora la predicció de preus per sobre de les mètriques comptables aïllades. En tot cas, la literatura convergeix en una premissa central: els indicadors endògens no s'han d'interpretar de forma aïllada, sinó com a part de sistemes d'interacció entre eficiència operativa, innovació i resiliència financera.

2.2. Factors macroeconòmics exògens

L'estudi dels factors macroeconòmics exògens segueix essent central per comprendre la dinàmica dels mercats financers globals. La recerca recent emfatitza la relació entre els cicles econòmics, la política monetària i el comportament borsari, destacant com els canvis en la

liquiditat i en els tipus d'interès condicionen les trajectòries sectorials i geogràfiques. En l'àmbit europeu, s'ha evidenciat que el tipus de canvi efectiu de l'euro i els índexs de confiança econòmica ("Economic Sentiment Index")⁴ resulten més predictius dels rendiments borsaris que les taxes d'atur o la inflació, especialment en contextos d'incertesa política i social.

Una línia emergent és la incorporació de factors climàtics i de sostenibilitat ambiental entre les variables exògenes. Determinades investigacions han mostrat que els xocs climàtics extrems afecten tant la volatilitat de les sèries financeres com el dinamisme dels inversors en sectors com l'energia i les assegurances. En paral·lel, els índexs d'incertesa de política econòmica (EPU)⁵ i la volatilitat implícita (VIX) s'han consolidat com a poderosos explicadors dels rendiments a curt termini, reforçant la noció que els factors exògens no únicament influeixen en les empreses a través dels mecanismes de transmissió macroeconòmica, sinó també mitjançant canvis en la percepció del risc sistèmic.

2.3. Tècniques estadístiques i aprenentatge automàtic

La predicció borsària ha transitat dels models estadístics tradicionals —regressions lineals, models ARIMA i GARCH⁶— cap a sistemes híbrids potenciat per l'aprenentatge automàtic i el deep learning. Estudis recents han posat en relleu com les dinàmiques dels mercats financers presenten patrons no lineals que escapen a l'anàlisi convencional, i que requereixen mètodes capaços d'aprendre dependències complexes i adaptatives. Dins d'aquest paradigma, l'enfocament multimodal ha guanyat preeminència: els models incorporen variables financeres històriques, indicadors macroeconòmics, mètriques ESG, indicadors de sentiment i factors tècnics.

En el camp de la reducció de la dimensionalitat, el PCA⁷ i els autoencodificadors⁸ han demostrat la seva utilitat per condensar la informació continguda en grans volums de característiques financeres sense una pèrdua significativa de capacitat predictiva. La Regressió sobre

³Graph Neural Networks (GNN): models d'aprenentatge profund dissenyats per processar dades estructurades en forma de graf, permetent capturar les relacions i interdependències existents entre diferents entitats connectades. En l'àmbit financer, s'utilitzen per modelitzar les connexions entre empreses, sectors o mercats i millorar la precisió predictiva dels models tradicionals.

⁴Economic Sentiment Index (ESI): indicador compost elaborat per la Comissió Europea que sintetitza les expectatives i percepcions d'empreses i consumidors sobre l'evolució de l'activitat econòmica de la zona euro a partir d'enquestes de confiança sectorials i de consum.

⁵Economic Policy Uncertainty (EPU): índex que mesura el grau d'incertesa associat a la política econòmica mitjançant l'anàlisi de contingut de mitjans de comunicació i altres indicadors relacionats amb decisions governamentals, regulació i política fiscal o monetària.

⁶Els models ARIMA (AutoRegressive Integrated Moving Average) modelitzen sèries temporals a partir dels seus retards i errors, mentre

que els models GARCH (Generalized AutoRegressive Conditional Heteroskedasticity) capturen la volatilitat condicional de les sèries financeres (Bollerslev, 1986; Box i Jenkins, 1976).

⁷L'Anàlisi de Components Principals (PCA) és una tècnica estadística de reducció de la dimensionalitat que transforma un conjunt de variables correlacionades en components ortogonals (no correlacionats) ordenats per la variància que expliquen. Referència fonamental: Jolliffe, I.T. (2002). Principal Component Analysis. Springer. Aplicada en finances per: Gu, Kelly i Xiu (2024); Carbó-Valverde et al. (2020).

⁸Autoencodificadors (Autoencoders): xarxes neuronals no supervisades que aprenen representacions comprimides de les dades mitjançant una capa latent de menor dimensió, utilitzades habitualment per a la reducció de dimensionalitat i l'extracció automàtica de característiques complexes en conjunts de dades d'elevada dimensió.

Components Principals (PCR) constitueix, precisament, una tècnica híbrida que combina el PCA amb la regressió lineal múltiple, amb l'objectiu de modelitzar la relació entre una variable dependent —en aquest cas, el preu de l'acció— i un conjunt de variables explicatives, evitant els problemes derivats de la multicollinearitat i l'excés de dimensionalitat. En el present treball, la combinació PCA-PCR es justifica, d'una banda, per la naturalesa de les dades —més de 300 variables amb un elevat grau de correlació entre elles— i, de l'altra, per la conveniència d'obtenir components latents interpretables des d'un punt de vista econòmic.

3. Disseny de la recerca i metodologia

3.1. Mostra, dades i variables

La mostra de l'estudi comprèn les cinquanta companyies que integren l'índex Eurostoxx 50 a data 21 de setembre de 2025, tal com han estat identificades en les plataformes TIKR Terminal⁹ i Morningstar. El període d'anàlisi abasta vint anys, de 2005 a 2025, la qual cosa permet incorporar múltiples cicles econòmics rellevants i garanteix la robustesa dels models predictius, en línia amb les recomanacions metodològiques de la literatura especialitzada que subratlla la importància d'incloure diverses fases del cicle econòmic en la validació de models financers, les dades s'agreguen anualment.

Les variables independents s'agrupen en quatre categories de caràcter endogen: el balanç de situació (99 variables), el compte de resultats (90 variables), l'estat de fluxos d'efectiu (79 variables) i els ràtios financers anualitzats (74 variables). A aquestes s'afegeixen deu variables macroeconòmiques exògenes de la zona euro, obtingudes d'Eurostat, de la Comissió Europea i de Trading Economics: el Producte Interior Brut anual, l'índex de productivitat, la taxa d'atur, el salari mig mensual brut, l'índex del cost laboral, la formació bruta de capital fix, l'índex de mercat borsari de la zona euro (EU50), la inflació mesurada per l'HICP, el deute públic total i el tipus de canvi EUR/USD al tancament de cada any. La variable dependent és el preu de cotització de les accions, concretament el preu de tancament de l'últim dia hàbil de cada any —o l'1 d'octubre de 2025 per a l'any en curs. El nombre total d'observacions és $n = 50 \text{ empreses} \times 21 \text{ anys} = 1.050$ observacions empresa-any, amb un conjunt de més de 300 variables

independents, fet que justifica l'aplicació del PCA per reduir la dimensionalitat.

3.2. Tractament i processat de dades

El tractament previ de les dades ha seguit un protocol rigorós per garantir la qualitat estadística del conjunt. En primer lloc, s'han seleccionat les observacions anuals per empresa i s'ha reorganitzat la base de dades en format ample (wide format), amb cada fila representant una combinació empresa-any. En segon lloc, s'han convertit les cadenes buides i els valors absents en nuls, i s'han eliminat les columnes amb més del 30% de dades mancants. Finalment, s'ha aplicat una normalització individual per a cadascuna de les variables —mitjana zero i desviació estàndard igual a u— per evitar que les diferències de magnitud distorsionin els resultats del PCA. El resultat és una matriu de dades neta, homogènia i apta per a l'anàlisi multivariable.

3.3. Anàlisi de Components principals (PCA)

El PCA és una tècnica estadística de reducció de la dimensionalitat que transforma un conjunt de variables independents altament correlacionades de forma lineal en un nou conjunt de variables no correlacionades, denominades components principals mitjançant un procés de projecció de la informació original en un nou espai de menor dimensió. Cadascun d'aquests components és una combinació lineal de les variables originals, ordenat en funció de la quantitat de variància que explica de la matriu de variàncies-covariàncies de les variables originals, denominada "S". En el cas d'aquest estudi, on es treballa amb més de 300 variables financeres i macroeconòmiques per a 50 empreses i 21 anys, el PCA permet superar tres reptes clau: la multicollinearitat entre variables, el soroll i la redundància informativa, i la impossibilitat de visualitzar les relacions en espais de dimensió superior a tres.

Tot i que existeixen variants avançades com l'Anàlisi de Components Principals No Lineal (NLPCA) —dissenyada per capturar estructures complexes mitjançant xarxes neuronals—, en aquest estudi s'ha descartat el seu ús. Aquesta decisió es justifica per l'elevat cost computacional d'aquestes tècniques, el risc de sobre ajustament (overfitting) davant l'alta dimensionalitat temporal de les dades i, principalment, per l'objectiu de preservar la màxima interpretabilitat econòmica dels factors obtinguts, una qualitat que es perdria amb la naturalesa de "caixa negra" de l'enfocament no lineal.

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

Morningstar:

El procediment d'aplicació ha seguit cinc etapes. En primer lloc, s'ha construït la matriu d'entrada excloent la variable dependent (PRICE) per evitar que influeixi en l'extracció dels components latents. En segon lloc, s'ha aplicat l'algorisme NIPALS¹⁰, que calcula els components principals de forma iterativa sense necessitat d'imputar prèviament els valors absents, especialment útil en matrius de gran mida i amb dades incompletes. En tercer lloc, per validar i complementar els resultats, s'ha realitzat una imputació de dades i una PCA clàssica¹¹ amb el nombre òptim de components¹². En quart lloc, s'ha realitzat una agrupació automàtica dels perfils anuals mitjançant l'algorisme K-means aplicat sobre les coordenades mitjanes per any. En cinquè i últim lloc, s'ha calculat el perfil mig per empresa al llarg del període, agregant les coordenades de PCA per any, amb l'objectiu de classificar les empreses per sector i detectar similituds estructurals.

3.4. Regressió sobre components principals PCR

La Regressió de Components Principals (PCR, per les seves sigles en anglès) és una tècnica híbrida que combina l'Anàlisi de Components Principals (PCA) amb la regressió lineal múltiple, permeten modelitzar la relació entre la variable dependent (preu) amb el conjunt de variables explicatives, fent ús de les components principals obtingudes amb anterioritat. D'aquesta manera s'eliminen els problemes de multicolinealitat que freqüentment s'invaliden en les estimacions de mercat.

En el marc de l'estudi, la PCR és adequada per tres punts: l'elevat nombre de variables inicials (més de 300 variables), la influència simultània de múltiples factors interns i externs sobre el preu, i la necessitat d'un enfocament robust i sintètic que faciliti la interpretació dels resultats.

L'especificació estàndard d'aquest model pren la següent forma:

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

on $PRICE_i$ és el preu de tancament anual de l'acció per a l'observació i , $Dim.k_i$ representa el valor del component principal k per a la mateixa observació, β_k són els coeficients estimats per mínims quadrats ordinaris (MQO) i ε_i és el terme d'error aleatori.

No obstant això la modelització directa dels preus en els seus nivells originals presenta limitacions severes en l'àmbit dels mercats financers, com l'asimetria en la

distribució, l'heteroscedasticitat i problemes d'omissió de dinàmiques temporals i estructurals. Per aquest motiu, s'amplia la recerca implementant una estratègia d'estimació progressiva en sis estadis (models 0 a 5). Aquest enfocament permet aïllar l'aportació marginal de cada transformació i variable de control sobre la capacitat explicativa del model (R^2 ajustat) i garanteix la robustesa dels coeficients estimats:

- Model 0 (original): Constitueix la línia de base descrita a l'equació (1), on es modelitza el preu en nivells en funció única dels components principals generats per la PCA.
- Model 1 (Log-lineal) Incorpora la transformació logarítmica de la variable dependent ($\log(Price_i)$). Esvaint l'asimetria de la distribució de preus, reduint l'impacte d'observacions extremes (outliers) i reinterpretant els coeficients en variacions percentuals.
- Model 2 (inèrcia temporal/dinàmic) Afegeix el retard de la variable dependent en escala logarítmica ($\log(Price_{i,t-1})$) com a variable explicativa. Permetent capturar l'autocorrelació i la inèrcia inherent als mercats financers, capturant la desviació del valor del preu respecte el seu propi passat.
- Model 3 (efectes fixos de temps) Incorpora un vector de variables fictícies d'any (any_factor), que absorbeix els xocs macroeconòmics globals i temporals (com ara crisis de mercat o canvis regulatoris) que afecten de manera simultània i homogènia totes les empreses de la mostra en un mateix període.
- Model 4 (efectes fixos d'empresa) Substitueix els controls temporals per un vector de variables fictícies a nivell empresarial ($Empresa_factor$). Permetent capturar l'heterogeneïtat no observada i invariant en el temps de cada companyia.
- Model 5 (Efectes fixos Bidireccionals) Integra els simultàniament els components de la PCA, el preu retardat, i els efectes fixos tant de temps com d'empresa. Permetent la neteja del biaix per variables omeses tant en la dimensió transversal com temporal, oferint un marc de contrast més robust per a la validació dels components del model PCR.

4. Anàlisi de resultats

4.1. La variància explicada i la composició dels components

L'aplicació del PCA al conjunt global de dades ha permès identificar la distribució de la variància explicada entre els components principals. El primer component

¹⁰NIPALS (Nonlinear Iterative Partial Least Squares): algorisme iteratiu per al càlcul del PCA en presència de valors absents. Wold, H. (1966). Nonlinear estimation by iterative least squares procedures.

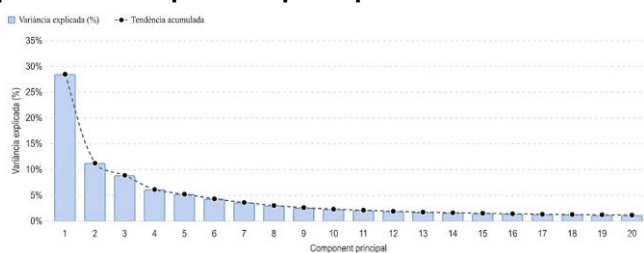
¹¹La PCA clàssica s'obté per descomposició espectral (eigendecomposition) de la matriu de covariàncies S o per

descomposició en valors singulars (SVD) de la matriu de dades centrada.

¹²El nombre òptim de components s'ha determinat combinant el criteri del gràfic de sediment (scree plot) i el llinar de variància acumulada explicada del 60%, tal com es detalla a la secció d'Anàlisi de resultats.

principal explica aproximadament el 27% de la variància total, i els següents aporten menys informació de forma decreixent. A partir del sisè component, l'increment marginal de la variància explicada s'aplanarà per sota del 3%, la qual cosa indica una pèrdua d'eficiència en la incorporació de noves dimensions. La suma acumulada dels cinc primers components supera el 60% de la variància total, lliar considerat adequat en els estudis financers per garantir la interpretabilitat sense reduir la capacitat explicativa. En conseqüència, el model es construeix a partir d'aquestes cinc dimensions latents.

Figura 1. Gràfic de la variància explicada per les 20 primeres components principals

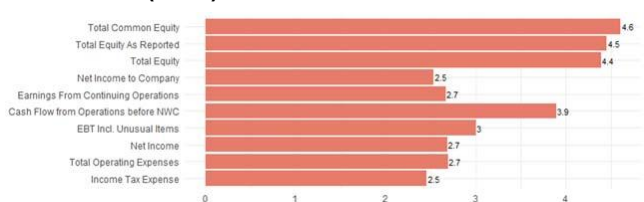


Font: Elaboració Pròpia mitjançant R

L'anàlisi de la composició interna de cadascun dels components permet identificar el seu significat econòmic i conèixer la contribució de cada una de les variables que componen la dimensió.

Dim.1 recull variables de patrimoni net, resultats comptables i flux de caixa operatiu. S'interpreta com la dimensió d'estabilitat financera i rendibilitat operativa, amb una alta correlació interna entre net income i patrimoni net acumulat.

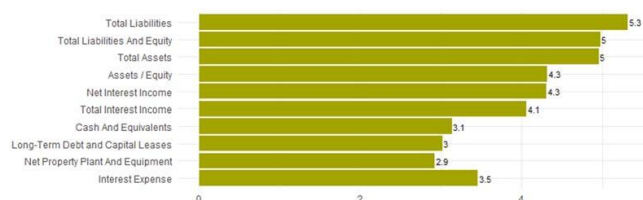
Figura 2. Contribució relativa de les variables a la Dimensió 1 (PCA).



Font: Elaboració Pròpia mitjançant R

Dim.2 agrupa variables d'escala i estructura de deute (Total Assets, Total Liabilities, Net Interest Income). S'interpreta com la dimensió d'escala operativa i exposició financera, i captura principalment les característiques del sector bancari i assegurator.

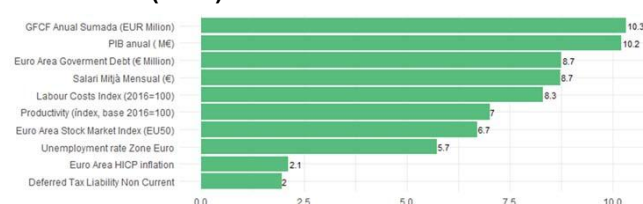
Figura 3. Contribució relativa de les variables a la Dimensió 2 (PCA).



Font: Elaboració Pròpia mitjançant R

Dim.3 està dominada per les nou variables macroeconòmiques de l'estudi (GFCF, PIB, deute públic, salari, cost laboral, productivitat, EU50, atur i HICP), que aporten més del 65% de la contribució total. S'interpreta com la dimensió de l'entorn econòmic agregat de la zona euro.

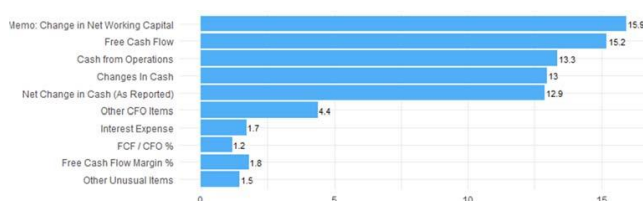
Figura 4. Contribució relativa de les variables a la Dimensió 3 (PCA).



Font: Elaboració Pròpia mitjançant R

Dim.4 recull variables de flux de caixa i liquiditat (Free Cash Flow, Cash from Operations, Changes in Cash). S'interpreta com la dimensió de dinàmica de liquiditat i eficiència operativa.

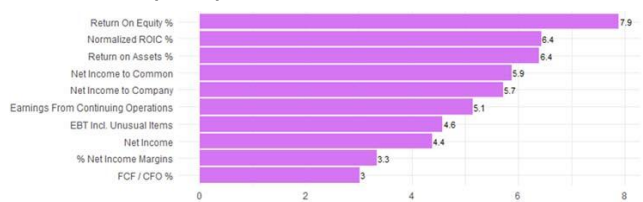
Figura 5. Contribució relativa de les variables a la Dimensió 4 (PCA).



Font: Elaboració Pròpia mitjançant R

Dim.5 recull indicadors de rendibilitat (ROE, ROIC, ROA, Net Income Margin). S'interpreta com la dimensió de rendibilitat i eficiència financera.

Figura 6. Contribució relativa de les variables a la Dimensió 5 (PCA).



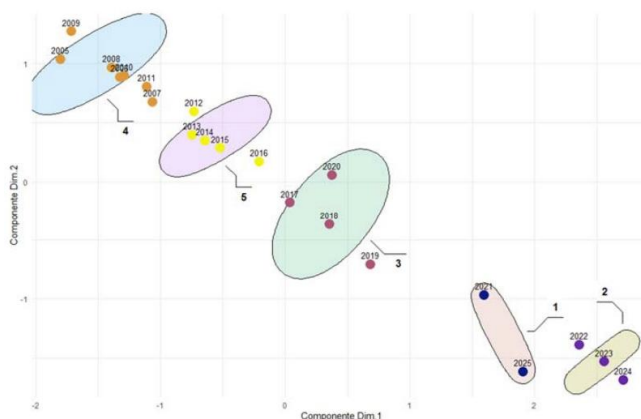
Font: Elaboració Pròpia mitjançant R

4.2. Agrupació temporal: l'empremta dels cicles econòmics

L'anàlisi gràfica dels perfils anuals en l'espai bidimensional dels components principals revela una

evolució temporal clara i coherent amb els grans cicles econòmics de la zona euro durant el període estudiat. Quan es representa la distribució dels anys en el pla Dim.1 versus Dim.2, s'observen quatre grans agrupaments seqüencials: els anys 2005-2011 es concentren a la zona superior esquerra del gràfic, reflectint una etapa marcada per balanços tensionats i un alt apalancament, amb la crisi financera global com a punt d'inflexió central; els anys 2012-2016 se situen al centre esquerre, mostrant una etapa de recuperació i consolidació; els anys 2017-2020 apareixen al centre del gràfic, amb estructures patrimonials estables i una activitat operativa moderada que caracteritza la fase de maduresa; i els anys 2021-2025 es desplacen cap a l'extrem inferior dret, suggerint una transformació estructural amb menys exposició financera però amb una sòlida solidesa comptable.

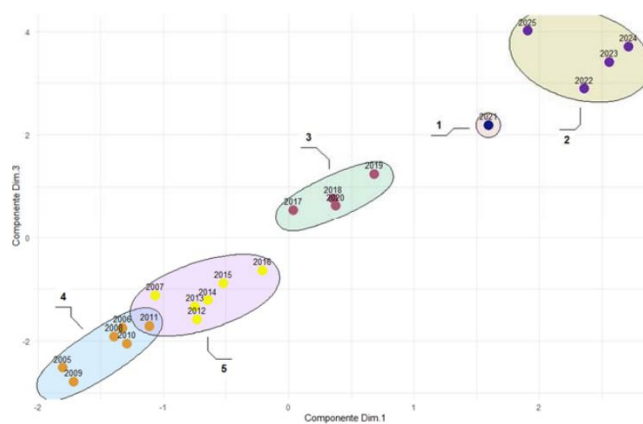
Figura 7. Gràfic agrupat per perfils anuals en el espai PCA (Dim. 1 vs. Dim. 2.)



Font: Elaboració Pròpia mitjançant R

La projecció en el pla Dim.1 versus Dim.3 —que confronta l'estabilitat comptable amb el context macroeconòmic— és particularment reveladora. Els anys previs i immediatament posteriors a la crisi del 2008 es concentren a la zona inferior esquerra, combinant fragilitat comptable i entorn macroeconòmic desfavorable. L'any 2020, marcat per la pandèmia de la COVID-19, apareix sistemàticament com un extrem en la majoria dels gràfics construïts, confirmant l'impacte disruptiu i singular d'aquest any sobre l'eficiència operativa i els fluxos de caixa agregats. Els anys 2022-2025, en canvi, se situen a l'extrem superior dret, amb valors elevats en ambdues dimensions, evidenciant un context de dinàmica macroeconòmica intensa —parcialment vinculat als processos inflacionaris post pandèmia— i d'expansió estructural de les empreses de l'índex.

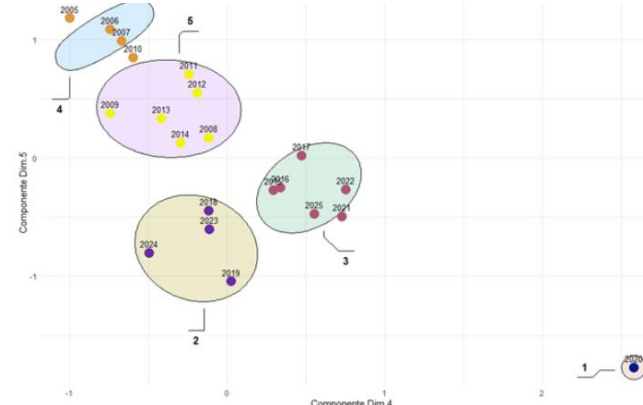
Figura 8. Gràfic agrupat per perfils anuals en el espai PCA (Dim. 1 vs. Dim. 3.)



Font: Elaboració Pròpia mitjançant R

Un dels resultats més rellevants de l'anàlisi temporal és que les dimensions vinculades a factors estructurals i macroeconòmics —especialment Dim.1, Dim.2 i Dim.3— generen agrupaments clarament seqüencials i lineals al llarg del temps. En contrast, les dimensions relacionades amb la gestió interna —Dim.4 i Dim.5— mostren agrupaments menys lineals, reflectint que l'evolució de la liquiditat i la rendibilitat no segueix sempre una trajectòria temporal contínua, sinó que respon a canvis interns en la gestió, decisions estratègiques puntuals o xocs específics de cada empresa o sector.

Figura 9. Gràfic agrupat per perfils anuals en el espai PCA (Dim. 4 vs. Dim. 5.)



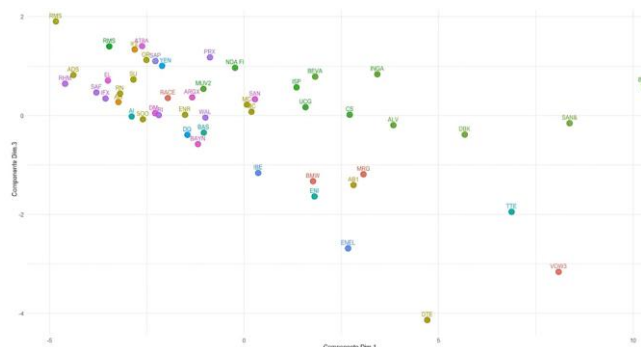
Font: Elaboració Pròpia mitjançant R

4.3. Perfils empresarials: la limitació de la segmentació sectorial

A diferència dels patrons temporals, l'anàlisi dels perfils empresarials no ha permès identificar agrupaments sectorials consistents. Malgrat l'aplicació de les mateixes tècniques de reducció de la dimensionalitat i de clustering, els resultats mostren una dispersió significativa entre sectors, sense una segmentació clara que permeti diferenciar els comportaments financers agregats per activitat econòmica. L'única excepció parcial és el sector financer, que en alguns gràfics tendeix a concentrar-se en zones específiques de l'espai

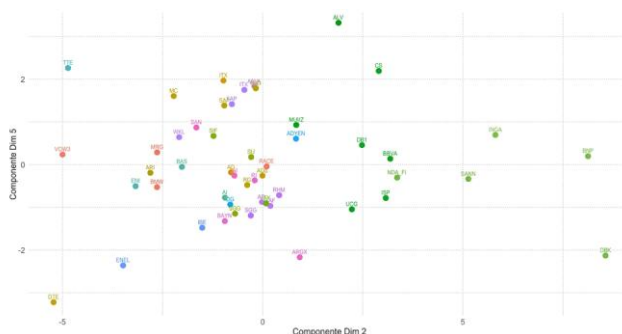
PCA, però amb una elevada variació interna que no permet establir-ne una diferenciació robusta.

Figura 10. Gràfic d'agrupament per perfils empresarials en el espai PCA (Dim. 1 vs. Dim 3.)



Font: Elaboració Pròpia mitjançant R

Figura 11. Gràfic d'agrupament per perfils empresarials en el espai PCA (Dim. 2 vs. Dim 5.)



Font: Elaboració Pròpia mitjançant R

Aquesta limitació té una lectura dual. D'una banda, posa de manifest que les variables estructurals capturades pel PCA no són suficients per discriminar els comportaments financers per activitat econòmica; dit altrament, les similituds entre empreses del mateix sector es dilueixen quan s'analitzen a través de dimensions tan àmplies com les considerades en aquest estudi. De l'altra, suggereix que els factors interns i idiosincràtics de cada empresa —la seva estratègia, la seva posició competitiva, la seva història— exerceixen una influència superior sobre els perfils financers agregats que la mera pertinença sectorial. Aquesta és, en si mateixa, una conclusió d'interès per a la gestió de carteres i per al disseny de models de valoració empresarial.

4.4. Resultats de la PCR: coeficients i capacitat predictiva

Els resultats obtinguts en l'estimació de la línia de base (model 0), on es modelitza el preu en nivells en funció exclusiva dels cinc components principals seleccionats anteriorment de la PCA, revelen una capacitat explicativa feble, amb un coeficient de determinació ajustat (R^2 ajustat) de 0,1599. Aquesta feblesa inicial evidencia que les relacions directe en els mercats

finances estan distorsionades per fenòmens d'heteroscedasticitats i d'asimetria latents en les series lineals originals.

Figura 12. Resum del model lineal 0 de regressió de components principals

```
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
```

Font: Elaboració Pròpia mitjançant R

Amb la introducció de la transformació logarítmica sobre la variable dependent (model 1), s'aconsegueix un correcció immediata de l'asimetria distributiva i de l'impacte dels valors extrems. En aquest estadi, l'ajust assoleix un increment significatiu al 0,2662 (R^2 ajustat). Des del punt de vista de la significació estadística, destaquen que quatre dels cinc components principals generats per la PCA, resulten altament significatius a un nivell de confiança superior al 99,9% ($p_value < 0,001$), amb l'única excepció de la quarta dimensió (flux de caixa i liquiditat) on $p_value = 0,026$. Aquests resultants demostra l'eficàcia de la reducció de la variància per capturar els vectors latents del mercat un cop estabilitzada la variància de la variable dependent.

Figura 13. Resum del model lineal 1 (Log-lineal) de regressió de components principals

```
Call:
lm(formula = log_preco ~ 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
```

Font: Elaboració Pròpia mitjançant R

La transició cap al model 2 (especificació dinàmica) suposa el canvi estructural de major profunditat del procés d'estimació en incorporar el retard de la variable dependent en escala logarítmica ($\log_price_lag1$). Els resultats obtinguts mostren un salt dràstic en la capacitat explicativa, on el R^2 ajustat se situa en un 0,8975.

Aquesta magnitud confirma la dependència estocàstica altament persistent i una forta memòria de curt termini en el procés de formació de preus, un comportament canònic en l'anàlisi de series temporals financeres.

Figura 14. Resum del model 2 (inèrcia temporal) de regressió de components principals

```
Call:
lm(formula = log_preco ~ 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
```

Font: Elaboració Pròpia mitjançant

L'avaluació de l'heterogeneïtat no observada mitjançant efectes fixos aïllats s'ha articulats en els models 3 i 4. El model 3 incorpora el vector de variables fictícies temporals (any_factor), per contra el model 4 i fa èmfasi en l'heterogeneïtat sectorial i empresarial mitjançant l'esquema d'efectes fixos d'empresa (empresa_factor).

Figura 15. Resum del model 3 (efectes fixos de temp¹³) de regressió de components principals

```
Call:
lm(formula = log_preco ~ 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
```

Font: Elaboració Pròpia mitjançant

El model 4 assoleix un R^2 ajustat del 0,9125 i reflecteix de primera mà l'impacte de xocs macroeconòmics globals i transversals. Respecte el model 4 reporta un R^2 ajustat del 0,9106 i un error residual del 0,3661.

¹³L'any de referència utilitzat en el model és 2005, corresponent al primer any disponible de la mostra, respecte el qual s'interpreten els coeficients de les dummies temporals.

Figura 16. Resum del model 4 (efectes fixos d'empresa¹⁴) de regressió de components principals

```
Call:
lm(formula = log_preco ~ 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
```

Font: Elaboració pròpia mitjançant R

Finalment, l'estimació del model 5 (efectes fixos bidireccionals) tanca la seqüència mitjançant una estructura completa que integra els components de la PCA, la inèrcia temporal i els controls tant transversals com longitudinals. Els resultats assoleixen el màxim nivell d'ajust dels sis models presentats, amb un R^2 ajustat del 0,9258 i l'error residual més baixa de la sèrie (0,3335). El comportament d'aquest model és robust, atès que l'efecte net dels components de la PCR, malgrat la inclusió d'una matriu exigent de controls bitemporals.

Figura 17. Resum del model 5 (efectes fixos bidireccionals) de regressió de components principals

```
Call:
lm(formula = log_preco ~ 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
```

Font: Elaboració pròpia mitjançant R

L'anàlisi dels coeficients estimats és reveladora quant al pes relatiu de cada una de les dimensions latents en la formació del preu. La Dim.5 —rendibilitat i eficiència financera— presenta el coeficient positiu més elevat i estadísticament significatiu de tot el bloc de components

¹⁴L'empresa de referència utilitzada en el model és ABI - Anheuser-Busch InBev, corresponent a la primera unitat en ordre alfanumèric de la base de dades utilitzada, respecte a la qual s'interpreten els coeficients de les dummies d'empresa.

($\beta = 0,10683$; $p_value = 1,31 \times 10^{-10}$), la qual cosa confirma que les millores en la gestió financera eficient exerceixen un impacte positiu i de gran intensitat sobre el preu de les accions, consolidant-se com un factor clau en la valoració empresarial. La Dim.2 —escala operativa i exposició financera— presenta el segon coeficient positiu del model ($\beta = 0,03593$; $p_value = 0,025101$), tot i que amb una significació estadística més moderada, suggerint que la seva influència en el preu, tot i ser rellevant, és la menys robusta de les dimensions positives.

En sentit contrari, la Dim.3 —l'entorn macroeconòmic— presenta l'impacte negatiu més intens i significatiu ($\beta = -0,09700$; $p_value = 0,005641$), la qual cosa suggereix que determinades configuracions d'aquest entorn —un cop controlat l'efecte inercial del propi mercat— s'associen a preus menys elevats. Seguint aquesta tendència negativa, les dimensions Dim.4 —dinàmica de liquiditat i eficiència operativa— i Dim.1 —estabilitat financera i rendibilitat comptable— també es mostren robustament significatives i inverses ($\beta = -0,04598$ amb $p_value = 6,70 \times 10^{-6}$ i $\beta = -0,04090$ amb $p_value = 0,000134$, respectivament), evidenciant com el pes dels passius o determinats increments d'escala operen com a factors de descompte en la valoració final una vegada aïllat el fort component autoregressiu de la variable temporitzada ($\log(\text{PRICE_lag1})$: $\beta = 0,69002$; $p_value < 2 \times 10^{-16}$).

Per avaluar la consistència de l'estratègia seqüencial implementada, la següent taula X detalla de forma agregada l'evolució de les mètriques de rendiment i el guany incremental respecte a la formulació original del model 0 fins arribar al model 5.

Figura 18. Comparativa del R2 ajustat i millora respecte al model base (M0)

	Model	R2_ajustat	Millora_vs_M0
M0: original (PRICE, 5 comp)		0.1599	+0%
M1: $\log(\text{PRICE}) + 5$ 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%

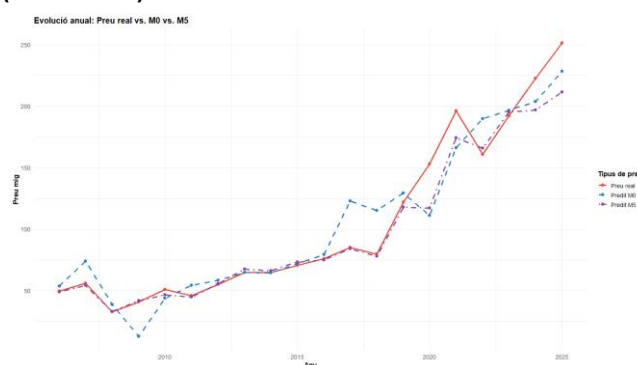
Font: Elaboració pròpia mitjançant R

L'anàlisi comparativa de les mètriques del model permet extraure conclusions metodològiques clau. L'evolució del R² ajustat demostra que, més enllà de la millora elemental que suposa el tractament logarítmic de la variable dependent en el model 1, el verdader sal qualitatiu (+73,8%) es deriva de modelitzar la naturalesa autoregressiva del mercat model 2, fet que redueix simultàniament l'errada estàndard residual d'1,049 a 0,392. D'altra banda, l'avaluació dels efectes fixos

suggereix que les fluctuacions temporals de caràcter macroeconòmic (model 3) exerceixen una tracció lleugerament superior sobre la variància dels preus de l'Eurostoxx 50 en el període estudiat que no pas l'efecte de les empreses (model 4).

Finalment, la convergència formal en el model 5 bidireccional demostra la complementarietat dels dos vectors de control. En assolir un guany net del 76,6% d'explicabilitat i minimitzar l'errada residual al seu nivell òptim (0,3335), validant empíricament l'estratègia progressiva adoptada. Aquesta especificació final no només optimitza l'ajust numèric, sinó que garanteix que els coeficients dels components de la PCR es trobin lliures d'interferències estructurals i conjunturals, consolidant-se com a la solució òptima per a la inferència d'aquest cas.

Figura 19. Gràfic de l'evolució anual del preu mig real vs. preu mig predit¹⁵ pel model 0 i el model 5 (2005-2025)



Font: Elaboració Pròpia mitjançant R

L'anàlisi visual de la capacitat predictiva a través de l'evolució del preu mig real i preus predits per als models 0 i 5 durant el període estudiat revela patrons altament diferenciats segons el nivell d'especificació. Es visualitza clarament com el model final (M5) exhibeix un ajust superior a l'hora de reproduir la tendència general del mercat, esmorteint de manera eficient les fortes desviacions que pateix el model base (M0) en moments d'alta volatilitat o disruptions externes. Tal i com es detecta en el cicle 2007 a 2009 o el xoc del 2017. Concretament el model 0 pateix severes sobreestimacions i pics d'error relatiu respecte el preu real. En contrast a la línia predictiva de M5 que traça gairebé a la perfecció la trajectòria real en aquests subperíodes, demostrant la correcció de la matriu bitemporal d'efectes fixos.

No obstant això, cap a la recta final de la sèrie històrica, emergeixen desajustos derivats de la pròpia naturalesa dels components. El model final subestima notablement

¹⁵El preu predit correspon als valors ajustats del model PCR sobre la mostra analitzada, calculats com a mitjana anual dels valors estimats per empresa.

el preu real durant la forta escalada alcista del final del període, sobretot l'any 2025, tram on el model de base M0 manté una trajectòria ascendent una mica més propera a la realitat. Aquestes divergències reflecteixen la complexitat d'aïllar els components latents en cicles de mercat post pandèmics altament expansius. Tot i que el model 5 es consolida com una especificació de major robustesa i menys erràtica en el còmput global del panell, les desviacions observades a l'últim tram posen de manifest la pressió de factors exògens que escapen a la descomposició de la PCA, tals com la innovació estratègica o ruptures abruptes en les expectatives dels inversors.

Taula 1. Diferències de preus predits entre els models M0 i M5 amb els preus mig reals del EursoToxx 50 per al període 2006 a 2025

Any	Preu mig real	Predicció M0	Diferència Real amb M0	Predicció M5	Diferència Real amb 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%

Font: Elaboració pròpia amb R

5. Discussió i conclusions

5.1. Principals aportacions

La recerca presentada en aquest estudi amplia substancialment l'evidència empírica sobre l'aplicabilitat de la combinació PCA-PCR per a l'anàlisi del comportament borsari europeu mitjançant una

estratègia metodològica progressiva i iterativa. Mitjançant la incorporació de cinc models addicionals a la PCR clàssic que han permès un refinament economètric en la significació estadística dels coeficients com a la robustesa estructura del model.

Les cinc dimensions latents identificades sintetitzen de manera coherent els principals eixos de comportament empresarial —estabilitat financera, escala operativa, entorn macroeconòmic, dinàmica de liquiditat i eficiència financera—, i l'arquitectura dels cinc models confirma amb major solidesa la seva capacitat explicativa quan són integrades en especificacions dinàmiques i amb estructures de control més sofisticades.

Els resultats mostren que la millora metodològica implementada no es limita a un increment quantitatiu del rendiment predictiu, sinó que reforça qualitativament la consistència d'inferència del model. La progressió des del model base fins als models avançats evidencia una reducció sistemàtica dels errors d'estimació i una estabilització notable dels coeficients, fet que consolida la validesa dels factors latents identificats per la PCA com a vectors explicatius rellevants del comportament del mercat.

L'anàlisi comparativa dels resultats anuals mostra una clara superioritat dels models avançats respecte a l'especificació inicial. Mentre que el model base presentava desviacions importants en episodis de tensió de mercat, els models refinats aconsegueixen reproduir amb major precisió les oscil·lacions estructurals del període analitzat. Aquesta millora és especialment significativa en els episodis d'alta volatilitat, on la nova especificació demostra una capacitat superior per absorbir xocs conjunturals i minimitzar errors extrems.

La incorporació seqüencial de noves especificacions també ha permès verificar empíricament que la naturalesa autoregressiva dels preus i el control simultani de l'heterogeneïtat temporal i transversal constitueixen elements essencials per assolir una modelització robusta del mercat europeu.

5.2. Implicacions per a la pràctica professional

Els resultats d'aquest estudi tenen implicacions rellevants per a diversos col·lectius professionals. Per als gestors de carteres i inversors institucionals, la confirmació que el context macroeconòmic exerceix la influència més potent sobre els preus de l'Eurostoxx 50 subratlla la importància de disposar de models de seguiment dels cicles econòmics de la zona euro que siguin àgils i actualitzats. En particular, la identificació del PIB, la formació bruta de capital fix i els índexs de productivitat com a principals variables macroeconòmiques del tercer component suggereix que els senyals d'acceleració o desacceleració econòmica

agregada han de tenir una ponderació elevada en les decisions d'asset allocation en renda variable europea.

Per als analistes d'empresa i directors financers, la dimensió de rendibilitat i eficiència financera —que recull el ROE, el ROIC i el ROA com a variables principals— confirma que les empreses que optimitzen l'ús dels seus recursos tendeixen a presentar preus més elevats i estables en el temps. Això reforça la utilitat dels ràtios d'eficiència com a indicadors avançats de la valoració borsària, especialment en un context on el mercat premia de forma creixent la creació de valor sobre la mera escala operativa.

Per als equips de recerca i els professionals de la gestió quantitativa, el treball demostra que la combinació PCA-PCR representa una eina valuosa per al diagnòstic estructural i la modelització predictiva en contextos d'alta dimensionalitat, però cal interpretar-la com una eina complementària dins d'un marc analític més ampli que incorpori tècniques no lineals, variables qualitatives i indicadors de sentiment de mercat.

5.3. Limitacions i línies de recerca futures

L'estudi presenta diverses limitacions que convé tenir presents en la interpretació dels resultats. En primer lloc, la capacitat explicativa limitada del model PCR —amb un R^2 ajustat del 15,21%— indica que una part substancial de la variabilitat del preu no queda capturada per les cinc dimensions latents, la qual cosa reflecteix la influència de factors externs i qualitius no tractats: reputació de marca, innovació, posicionament estratègic, expectatives de mercat i sentiment dels inversors. En segon lloc, tant el PCA com la PCR assumeixen relacions lineals entre les variables, la qual cosa limita la capacitat del model per capturar dinàmiques no lineals o interaccions complexes comunes en entorns econòmics reals. En tercer lloc, la comparació entre el preu real i el predit s'ha realitzat de forma agregada per any, la qual cosa pot suavitzar errors individuals per empresa i ocultar desviacions significatives en casos particulars.

A partir d'aquestes limitacions, s'identifiquen diverses línies de recerca futures de gran interès. La incorporació de variables qualitatives —indicadors ESG, mètriques de reputació i indicadors d'innovació— podria millorar substancialment la capacitat explicativa del model. L'aplicació de tècniques no lineals —Random Forest, XGBoost o xarxes neuronals recurrents de tipus LSTM— o LASSO com a alternativa de regularització en contextos d'alt nombre de variables, permetria capturar les relacions complexes entre components i preu en contextos d'alta dispersió. Finalment, la validació del model amb dades d'altres índexs borsaris internacionals i la seva extensió a nivell d'empresa individual —en lloc

del plantejament agregat aquí adoptat— constituïrien passos naturals per avaluar la robustesa i la generalització dels resultats.

5.4. Reflexió de síntesi

L'ampliació metodològica desenvolupada en aquest treball confirma que l'enfocament PCA-PCR adquireix el seu màxim potencial quan s'aplica dins d'un esquema iteratiu de refinament economètric orientat a reforçar simultàniament capacitat predictiva, significació estadística i robustesa estructural.

La incorporació dels cinc models addicionals ha permès transcendir les limitacions inicials de l'especificació base i construir una arquitectura analítica notablement més sòlida, capaç de capturar amb major precisió la naturalesa multifactorial del comportament borsari europeu.

Els resultats obtinguts evidencien que la combinació entre reducció de dimensionalitat, modelització dinàmica i controls estructurals constitueix una aproximació especialment eficaç per analitzar entorns financers d'alta complexitat. Tanmateix, també recorden que cap model lineal és capaç d'explicar íntegrament la dinàmica dels mercats de capitals.

El valor real d'aquest enfocament rau, per tant, no només en la seva capacitat predictiva, sinó en la seva utilitat com a instrument rigorós d'interpretació estructural, integrable dins d'un marc analític més ampli que combini solidesa estadística, criteri econòmic i capacitat crítica d'interpretació contextual.

Bibliografia

- 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* (2a 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 l'autor

Carles Ruiz Rovira és graduat en Economia, amb especialització en anàlisi política i màrqueting. Actualment amplia la seva formació en finances i anàlisi de dades, així com en programació estadística. És Màster en Finances per l'Institut d'Estudis Financers (IEF).

Altres publicacions

Set 2026	NT	Identificació dels factors determinants del preu de l'Eurostoxx 50 mitjançant Anàlisi de Components Principals (PCA) i Regressió sobre Components Principals (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