



Universidad de Valencia y Universidad del País Vasco

Máster en Banca y Finanzas Cuantitativas

Práctica 1: Estimación de modelos factoriales para los precios de los activos

Gestión de Riesgos

Gonzalo Moll Acha
Guillem Borràs Espert

Profesora: (*María Dolores Robles*)

27 de octubre de 2025

Resumen

El presente trabajo analiza y compara dos metodologías de modelización factorial aplicadas a rendimientos financieros: el modelo basado en Componentes Principales (PCA) y el modelo BARRA. A partir de una muestra de 100 empresas con datos diarios entre 2011 y 2025, se estiman ambos enfoques con el objetivo de descomponer la varianza total de los rendimientos en sus componentes sistemáticos e idiosincráticos, evaluar su capacidad explicativa y analizar su utilidad práctica en la gestión del riesgo. Los resultados muestran que el modelo PCA logra capturar las principales fuentes de variabilidad mediante un número reducido de factores ortogonales, aunque su interpretación económica es limitada. En cambio, el modelo BARRA, al incorporar características observables como beta, tamaño, valor y país, ofrece una explicación más estructurada del riesgo, incrementando significativamente el porcentaje de varianza explicada y mejorando el ajuste según el criterio AIC. En conjunto, los resultados del trabajo confirman que el modelo BARRA resulta más adecuado para la caracterización del riesgo sistemático, mientras que el PCA constituye una herramienta complementaria para la identificación de patrones estadísticos y la reducción de dimensionalidad.

Abstract

This study analyzes and compares two factor modeling approaches applied to financial returns: the Principal Component Analysis (PCA) model and the BARRA model. Using a sample of 100 firms with daily data from 2011 to 2025, both frameworks are estimated to decompose the total variance of returns into systematic and idiosyncratic components, assess their explanatory power, and evaluate their practical relevance for risk management. The results indicate that PCA effectively captures the main sources of variability through a reduced set of orthogonal factors, although its economic interpretation remains limited. In contrast, the BARRA model, by incorporating observable firm characteristics such as beta, size, value, and country, provides a more structured explanation of risk, significantly increasing the proportion of explained variance and improving model fit according to the AIC criterion. Overall, the evidence suggests that the BARRA model offers a superior representation of systematic risk, while PCA serves as a complementary statistical tool for pattern detection and dimensionality reduction in portfolio analysis.

Índice

| | |
|---|----------|
| Resumen / Abstract | 1 |
| 1. Introducción | 1 |
| 2. Análisis descriptivo | 1 |
| 3. Estimación de modelos | 2 |
| 3.1. Modelo factorial con componentes principales | 2 |
| 3.1.1. Descripción del modelo factorial con componentes principales | 2 |
| 3.2. Modelo BARRA con características de las empresas | 3 |
| 3.2.1. Descripción del modelo BARRA | 3 |
| 4. Resultados | 4 |
| 4.1. Resultados estimación modelo PCA | 4 |
| 4.2. Resultados estimación modelo BARRA | 5 |
| 4.3. Comparación resultados modelo PCA - modelo BARRA | 6 |
| 5. Conclusiones | 8 |
| 6. Apéndices | 9 |
| 6.1. Apéndice I: Análisis Descriptivo | 9 |
| 6.2. Apéndice II: PCA | 12 |
| 6.3. Apéndice III: Modelo BARRA | 16 |

1. Introducción

En un entorno empresarial cada vez más competitivo y volátil, la gestión de riesgos lleva décadas manteniendo un papel fundamental en cualquier sector organizacional. Obviando su practicidad para anticiparse a cambios no deseados o su utilidad a la hora de llevar a cabo coberturas, la gestión de riesgos se evalúa también como herramienta estratégica para evaluar oportunidades o detectar patrones.

Por otro lado, desde el punto de vista del inversor puramente financiero, a menudo se han considerado el análisis técnico y fundamental como los pilares fundamentales para el *asset allocation* y *security selection*. No obstante, desde mediados del siglo pasado, y motivados por el auge tecnológico y la mejora computacional, se comienzan a popularizar las técnicas matemático/estadístico-fundamentadas, como el PCA o el Análisis Discriminante Lineal (LDA).

Por ello, en la presente práctica, desde un punto de vista econométrico/estadístico, se tiene por objetivo estimar una serie de modelos factoriales de valoración de activos, incluyendo metodologías de reducción de la varianza y modelos de tipo BARRA. En otras palabras, el objetivo del trabajo reside en la modelización de los rendimientos bajo la estructura:

$$R_{i,t} = \alpha_i + \sum_{k=1}^K \beta_{i,k} F_{k,t} + \varepsilon_{i,t}, \quad \mathbb{E}[\varepsilon_{i,t}] = 0$$

donde $R_{i,t}$ representa el rendimiento del activo i en el periodo t ; α_i el rendimiento no explicado por los factores sistemáticos; $\beta_{i,k}$ las sensibilidades del activo frente a cada factor k ; $F_{k,t}$ los factores comunes; y $\varepsilon_{i,t}$ la perturbación o componente idiosincrático.

De este modo, el trabajo pretende no solo estimar y comparar diferentes modelos factoriales, sino también valorar la utilidad de las metodologías de reducción de varianza y dimensionalidad así como la implementación de modelos tipo BARRA como herramientas para la gestión del riesgo. Asimismo, se busca también obtener una comprensión más desarrollada de las interrelaciones entre los activos, tratando de identificar las fuentes principales de riesgo sistemático y evaluando su impacto en la estructura de los rendimientos de las empresas, contribuyendo así a la puesta en marcha de estrategias de inversión más eficientes.

2. Análisis descriptivo

Respecto a los datos, cabe mencionar que el *dataset* original contaba con cotizaciones diarias (desde el 26/09/2011 hasta el 26/09/2025) de 600 empresas.

La estructura temporal de los datos proporciona una base estadística suficientemente amplia para capturar distintas fases del mercado: periodos de expansión, recesión, *shocks* exógenos (como la crisis del Covid-19 o conflictos geopolíticos como la Guerra de Ucrania o el conflicto Palestino-Israelí), y fases de elevada volatilidad financiera. Esta diversidad temporal proporciona robustez al análisis y permite evaluar la estabilidad de los factores estimados en entornos cambiantes.

Como se ha comentado anteriormente, se tenían datos de 600 empresas, pero el trabajo se ha desarrollado sobre 100. El criterio de selección ha sido meramente práctico: las primeras 100 empresas que contenían todos los datos de cotizaciones y características, eliminando de esta manera los potenciales problemas a la hora de tratar con *NA*'s y similares. Además, este criterio ha permitido elegir una muestra considerablemente heterogénea, lo cual enriquece el análisis y evita conclusiones "miopes" o sesgadas. La base de datos original contaba también con una hoja de características (cuantitativas y cualitativas) de las empresas, que con el objetivo de facilitar

el filtrado, se ha traspuesto y duplicado en otra hoja con la misma información, denominada "CaracT".

Respecto al primer análisis en sí mismo de los datos, en el apéndice se pueden encontrar gráficos de precios (Figura 1.) y rendimientos logarítmicos (Figura 2.), así como una tabla con los principales estadísticos descriptivos de las series (Tabla 2.), y la matriz de correlaciones entre las 100 empresas en formato *heatmap* (Figura 3). Sobre estos, se puede confirmar que las series de rendimientos financieras no son normales, ya que presentan *outliers*, asimetrías y colas (curtosis) distintas de la normal.

Como se puede observar en la Figura 1, a nivel econométrico los precios de las empresas son claramente no estacionarios, y en algunos casos, han mostrado incluso tendencias explosivas en el periodo analizado. Teniendo en cuenta que la muestra temporal comprende la salida de la crisis del 2008, la crisis del euro de 2012 y la reciente pandemia y crisis sanitaria del Covid-19 en 2020, lo más probable es que dichas empresas "explosivas" pertenezcan al sector de las farmaceuticas, debido a que no se puede tratar de ninguna empresa "unicornio" pues se disponen de datos desde 2012. Respecto a los rendimientos (Figura 2), se observan series estacionarias de manera general (como cabe esperar por la construcción de los rendimientos logarítmicos), pues estos oscilan entre unas bandas de confianza entre -0,2 y 0,2, salvo contadas excepciones en 2015, 2016 o 2021. Sin embargo, un *outlier* o valor extremo llama la atención más de lo habitual en 2012, y se trata de la empresa Natwest Group, que obtuvo un extraordinario rendimiento de casi un 1 % en un solo día.

Finalmente, como se ha comentado anteriormente, en la Tabla 3 se puede observar cómo las empresas muestran momentos estadísticos muy distintos a los de la distribución normal, principalmente asimetrías en las colas y curtosis tremendamente elevadas. Por otro lado, en la Figura 3 se verifica que el criterio de elección (relativamente arbitrario) ha permitido escoger una muestra heterogénea de empresas, pues la matriz de correlaciones muestra que existen pocas altas correlaciones entre empresas.

3. Estimación de modelos

3.1. Modelo factorial con componentes principales

3.1.1. Descripción del modelo factorial con componentes principales

La técnica estadística de Análisis de Componentes Principales es una herramienta que permite reducir la variabilidad de un conjunto de datos manteniendo la mayor cantidad de información posible. Para ello, se lleva a cabo una descomposición de la matriz de correlaciones de los rendimientos (tal y como se explica al final del presente apartado).

De esta manera, tras aplicar el Análisis de Componentes Principales (PCA), la matriz de covarianzas (o correlaciones) de los rendimientos puede descomponerse espectralmente como:

$$\Sigma = W\Lambda W^T$$

donde:

- $W = [w_1, w_2, \dots, w_N]$ es la matriz de autovectores (*loadings*),
- $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_N)$ contiene los autovalores asociados a cada componente principal.

Cada componente principal se define como una combinación lineal de los rendimientos originales:

$$F_k = Xw_k$$

donde F_k es el vector temporal de realizaciones del componente k , y w_k son los pesos de los activos en dicha combinación. Por construcción, los componentes F_k son ortogonales y su varianza está dada por el autovalor correspondiente:

$$\text{Var}(F_k) = \lambda_k$$

La cartera réplica del componente principal k se obtiene directamente a partir del vector de pesos w_k . En la práctica, w_k define la proporción (no necesariamente normalizada a 1) en la que cada activo contribuye al factor k . Dado que los autovectores están normalizados en norma euclídea, puede ser conveniente reescalar los pesos para interpretar la cartera en términos financieros:

$$\tilde{w}_k = \frac{w_k}{\sum_i |w_{ik}|}$$

de modo que los pesos absolutos sumen la unidad. La serie temporal de la cartera réplica es entonces:

$$R_t^{(k)} = \sum_{i=1}^N \tilde{w}_{ik} r_{i,t}$$

donde $r_{i,t}$ es el rendimiento del activo i en el instante t . Esta cartera $R_t^{(k)}$ reproduce las fluctuaciones del componente principal F_k , constituyendo una versión económicamente interpretable del factor estadístico extraído por el PCA.

Por ello, el PCA puede entenderse como una técnica que genera un conjunto de carteras ortogonales que resumen las principales fuentes de variabilidad del conjunto de rendimientos.

Una vez explicado cómo funciona la técnica de *PCA*, se procede a estimar los componentes principales sobre las correlaciones de los rendimientos, ya que los valores de las desviaciones estándar difieren en más de un orden de magnitud, es decir:

$$\frac{\max(\sigma_i)}{\min(\sigma_i)} = 2,96$$

por lo que queda justificado utilizar las correlaciones en lugar de las covarianzas.

3.2. Modelo BARRA con características de las empresas

3.2.1. Descripción del modelo BARRA

Estimaremos ahora un modelo BARRA para capturar riesgo sistemático asociado a bloques de características (p. ej., beta, tamaño, valor y país) y separarlo del riesgo idiosincrático propio de cada activo. Las exposiciones $\beta_{k,i}$ (dummies por grupo, conocidas) fijan cómo cada activo influye en los factores $f_{k,t}$, cuyos rendimientos miden el comportamiento medio de cada grupo frente a su categoría de referencia. El término $\varepsilon_{i,t}$ recoge shocks específicos no explicados por los factores.

Para $i = 1, \dots, N$ activos y $t = 1, \dots, T$ fechas, el modelo BARRA dicta:

$$R_{i,t} = \alpha_i + \sum_{k=1}^K \beta_{k,i} f_{k,t} + \varepsilon_{i,t},$$

con $E[\varepsilon_{i,t}] = 0$, $\text{Var}(\varepsilon_{i,t}) = \sigma_i^2$, $\text{Cov}(\varepsilon_{i,t}, f_{k,t}) = 0$ y $\text{Cov}(f_{k,t}, f_{k',t}) = [\Omega_f]_{k,k'}$.

En forma matricial ($R_t = \alpha + B f_t + \varepsilon_t$), la descomposición de la varianza de los activos se puede escribir como (parte sistemática asociada a factores + parte idiosincrática) :

$$\text{Var}(R_t) = B \Omega_f B' + \Psi, \quad \Psi = \text{diag}(\sigma_1^2, \dots, \sigma_N^2).$$

Para una cartera con pesos w ,

$$\text{Var}(w' R_t) = \underbrace{(B'w)' \Omega_f (B'w)}_{\text{riesgo por factores}} + \underbrace{w' \Psi w}_{\text{riesgo idiosincrático}},$$

donde $B'w$ son las exposiciones de la cartera a cada factor. Esta identidad permite atribuir el riesgo total entre bloques (p.ej., tamaño, valor, país) y entre activos, y es la base para el control de concentraciones en construcción de carteras. A la identidad anterior nos referiremos como *la descomposición de la varianza en el modelo BARRA*

En nuestro caso, usamos cuatro bloques: Beta, Tamaño (market cap), Valor (book o BM) y País.

- Beta, Tamaño y Valor: 5 quintiles \Rightarrow dummies por quintil; se elimina una categoría para evitar colinealidad $\Rightarrow 3 \times (5 - 1) = 12$ columnas.
- País: 11 categorías mutuamente excluyentes; una de referencia, de nuevo para evitar colinealidad, entonces $\Rightarrow 10$ columnas.

Total de factores $K = 12 + 10 = 22$. Sea $B = [B^{(\beta)} \ B^{(\text{Size})} \ B^{(\text{Value})} \ B^{(\text{País})}]$ con $B^{(\beta)}, B^{(\text{Size})}, B^{(\text{Value})} \in \{0, 1\}^{N \times 4}$ y $B^{(\text{País})} \in \{0, 1\}^{N \times 10}$. Cada componente $f_{k,t}$ es el rendimiento del grupo frente a su categoría de referencia en t .

Estimamos los factores por fecha mediante mínimos cuadrados generalizados (pesos $\hat{\Psi}^{-1}$), ya que el modelo presenta heterocedasticidad:

$$\hat{f}_t = (B' \hat{\Psi}^{-1} B)^{-1} B' \hat{\Psi}^{-1} (R_t - \alpha), \quad \hat{\Psi} = \text{diag}(\hat{\sigma}_1^2, \dots, \hat{\sigma}_N^2).$$

Además, tenemos, para las estimaciones de los factores y de los residuos:

$$\hat{f} = \frac{1}{T} \sum_{t=1}^T \hat{f}_t, \quad \hat{\Omega}_f = \frac{1}{T-1} \sum_{t=1}^T (\hat{f}_t - \hat{f})(\hat{f}_t - \hat{f})'.$$

$$\hat{\varepsilon}_t = R_t - \hat{B} \hat{f}_t - \alpha, \quad \hat{\varepsilon} = \frac{1}{T} \sum_{t=1}^T \hat{\varepsilon}_t, \quad \hat{\sigma}_i^2 = \frac{1}{T-1} \sum_{t=1}^T (\hat{\varepsilon}_{i,t} - \hat{\varepsilon}_i)^2.$$

4. Resultados

4.1. Resultados estimación modelo PCA

Los resultados del análisis del PCA y de la estimación del modelo factorial utilizando los 3 primeros componentes mediante Python se encuentran en el Apéndice II. En este apartado se

discuten los puntos principales acerca del PCA, descomposición de la varianza con el modelo PCA, la bondad del ajuste y las distintas ratios de performance para cada una de las carteras analizadas.

Los resultados del PCA se encuentran en la Tabla 4. Con el objetivo de explicar el 70 % de la varianza total, se deciden utilizar un total de 21 componentes principales (Figura 5). Además, se verifica la propiedad de que los componentes están incorrelacionados entre sí (Figura 6). A nivel interpretativo, se puede entender el PC1 como el factor de mercado, ya que captura la mayor proporción de la varianza, y por tanto, la tendencia general del mercado. En otras palabras, se podría decir que el PC1 es análogo al factor de mercado del modelo CAPM. En segundo lugar, los factores 2 y 3 agregan un 6 % y un 4 % adicional de la varianza, y representan factores de sector, tipología de industria (o en general, grupos de acciones que se comportan de manera diferente al mercado) o diferencias de estilo de inversión (*growth*, *value*, *momentum*, etc.). Del PC4 en adelante resulta más difícil encontrar una interpretación directa, ya que al sumar menos de un 3 % cada uno automáticamente se convierten en factores demasiado específicos e idiosincráticos como para atribuirles una interpretación más general.

De manera adicional, se ha decidido calcular una serie de correlaciones entre los componentes y las variables del *dataset* original (Figura 4) concluyendo que algunas variables como "Market Cap" pueden estar correlacionadas con varios componentes.

Antes de comentar los resultados de la descomposición de la varianza, es necesario explicar cómo se han construido las carteras de este apartado. Por un lado, la cartera equiponderada se construye ponderando todas las empresas por igual, como su propio nombre indica, y actúa por tanto como un índice o cesta de acciones. No obstante, por otro lado, las carteras Beta y Size se construyen de manera diferente, siguiendo el estilo de Fama y French. Primero, se clasifican las acciones en *Small*, *Medium* o *Big* según su tamaño y en *LowBeta*, *MidBeta* o *HighBeta* según su beta. A continuación, se construyen los pesos *long* y *short* y *high* y *low* de tal manera que la cartera Size está larga en acciones pequeñas y corta en grandes, y la cartera Beta está larga en acciones con una beta elevada y corta en acciones con beta reducida. Con esta construcción, se consigue que las carteras estén balanceadas y sean neutrales en media.

Respecto a la descomposición de la varianza, la cartera equiponderada mantiene una proporción casi idéntica entre riesgo sistemático y riesgo idiosincrático, lo cual indica que la cartera se puede cubrir casi en ambas proporciones con herramientas basadas en factores y con herramientas específicas de los activos. Sin embargo, las carteras Size y Beta presentan un elevado porcentaje (superior al 90 %) de riesgo sistemático, mostrando así que los factores sistemáticos del mercado no serían de gran ayuda de cara a realizar coberturas.

Por otro lado, los resultados de bondad de ajuste y ratios de performance se encuentran en la Tabla 6. La cartera equiponderada arroja los mejores valores para las métricas analizadas: un ratio de Sharpe y de Treynor del 3,4 %, un alpha de Jensen del 0,0357 % y el menor AIC de las 3 carteras, -45.744.65. A esta cartera le sigue la cartera Beta, con ratios de performance inferiores pero positivos, y un AIC de -37797.74. Por último, la cartera Size muestra ratios de performance negativos y un AIC de -35889.99. Estos resultados verifican los obtenidos en el párrafo anterior, ya que concuerdan con la descomposición de la varianza y las proporciones de riesgo que mantiene cada cartera.

4.2. Resultados estimación modelo BARRA

Los resultados de la estimación del modelo BARRA mediante Python se encuentran en Apéndice 6.3. En este apartado discutiremos los puntos principales acerca de la descomposición de la varianza con el modelo BARRA, la bondad del ajuste y las distintas ratios de performance para

cada una de las carteras analizadas.

Empezamos comentando brevemente el gráfico de correlación (Figura 10) de las estimaciones de los diferentes factores \hat{f}_k del modelo BARRA: observamos un clúster de alta correlación entre factores del bloque país y del bloque beta, mientras que el resto presentan correlaciones más bajas. Este resultado es coherente con la estructura del modelo, ya que los factores dentro de un mismo bloque suelen moverse conjuntamente, mientras que entre bloques distintos las correlaciones son cercanas a cero, lo que indica una adecuada separación de las fuentes de riesgo.

La Tabla 8 resume cómo se reparte la varianza de cada cartera entre *riesgo sistemático* (explicado por factores) y *riesgo idiosincrático* (específico y no explicado). Cuanto mayor es el porcentaje sistemático, más medible y cubrible es el riesgo con herramientas basadas en factores; cuanto mayor es el idiosincrático, más dependemos de shocks propios de los activos.

Las carteras *Equiponderada*, *PC1* y *PC2* presentan un perfil claramente factorial, con proporciones de riesgo sistemático elevadas (96,34 %, 88,37 % y 99,17 %), y una SRC dominada por factores geográficos (`country_UK`, `FR`, `SW`, `NL`, `SD`). *PC3* muestra 104,16 % sistemático y varianza idiosincrática nula, cifra $>100\%$ atribuible a errores de redondeo sobre varianzas diarias muy pequeñas, sin cambiar la lectura cualitativa: también queda dentro del espacio de factores. En contraste, *Size* (34,26 % sist.) y *Beta* (3,57 % sist.) están dominadas por riesgo idiosincrático y exhiben contaminación estilo-país (p. ej., `book_small` y dummies de país en la SRC).

El AIC es mucho más negativo (mejor) en *Equiponderada* y *PC3*, razonablemente bueno en *PC1*–*PC2* y débil en *Size*, *Beta*, reforzando que las carteras diversificadas están bien explicadas por BARRA. En ratios de performance, *Equiponderada* y *PC3* concentran los mayores Sharpe/Treynor (coherentes con su alto % sistemático), *PC1* es muy similar, *PC2* es algo menor con $\alpha < 0$, y las *long-short* muestran eficiencia modesta, confirmando la necesidad de cautela en su medición y cobertura del riesgo.

4.3. Comparación resultados modelo PCA - modelo BARRA

El objetivo de este apartado es comparar los dos enfoques factoriales estimados sobre las mismas carteras: el modelo BARRA, y el modelo de regresión PCA. La evaluación se centra en la descomposición de la varianza entre componente sistemática e idiosincrática, en el criterio de información de Akaike, $AIC = 2k - 2 \ln \mathcal{L}$ y en tres ratios de performance: Sharpe, Treynor y Alpha de Jensen. En la Tabla 1 se presentan de forma esquemática las diferentes métricas para las carteras consideradas.

En la cartera equiponderada, BARRA explica prácticamente la totalidad del riesgo con un 96.34 % de varianza sistemática frente al 46.78 % del modelo PCA. Esta superioridad se refleja de manera contundente en el AIC, donde BARRA alcanza $-51\,206,40$ frente a $-45\,744,65$ del modelo con PCs, es decir, una mejora de aproximadamente 5,462 puntos AIC. La lectura es clara: los factores económicos capturan dimensiones de riesgo—especialmente geográficas y de estilo—que no están completamente contenidas en los tres primeros modos ortogonales de variación del conjunto de activos.

En la estrategia *SizeLS* el contraste es todavía más marcado. Con solo tres PCs, el componente sistemático es prácticamente nulo (1.03 %), lo que indica que el spread de tamaño no está bien representado por los primeros modos globales de las correlaciones de los activos. BARRA eleva la parte explicada hasta el (34.26 %), y además reduce el AIC de $-35\,889,99$ a $-40\,125,61$, una ganancia de alrededor de 4,236 puntos. Esta evidencia sugiere que el tilt de tamaño se alinea mejor con características económicas explícitas que con las direcciones de máxima varianza agregada que prioriza PCA.

La estrategia Beta LS presenta un patrón distinto y, en cierto modo, ilustrativo de los límites de ambos enfoques. El riesgo idiosincrático domina con claridad: BARRA solo atribuye un (3.57%) a factores, mientras que PCA(3) asciende levemente al (6.66%). En términos de AIC, el modelo con PCs obtiene aquí una ligera ventaja ($-37\,797,74$ frente a $-37\,078,87$), si bien el bajo porcentaje de riesgo sistemático en ambos casos revela que buena parte de la varianza procede de componentes específicos no recogidos por estos conjuntos de factores. En suma, el spread por beta, tal como está implementado, no queda bien capturado ni por el esqueleto BARRA estándar ni por un espacio de solo tres PCs. No obstante, en este caso el modelo PCA supera al BARRA en ambas métricas con una ligera ventaja.

En cuanto a los ratios de performance (Tabla), las diferencias entre BARRA y PCA se concentran en las estrategias long-short. En la cartera Equiponderada los ratios son muy próximos: el Sharpe pasa de 0,034378 con PCA a 0,0356 con BARRA y el Treynor de 0,034412 a 0,0345; la única discrepancia visible es el α de Jensen, positivo con PCA ($\approx 3,57 \times 10^{-4}$) y nulo con BARRA, lo que sugiere un pequeño exceso de retorno no explicado cuando el *benchmark* son los PCs. En Size LS el contraste es claro: con BARRA los ratios son positivas (Sharpe 0,0120, Treynor 0,0098), mientras que con PCA resultan negativas (Sharpe $-0,011697$, Treynor $-0,019131$); el α es negativo en ambos casos, de mayor magnitud en BARRA ($-1,0 \times 10^{-4}$ frente a $-2,6 \times 10^{-5}$), indicando menor compensación por riesgo y que usar solo tres PCs penaliza especialmente esta estrategia. En Beta LS los resultados son moderados para ambos modelos: el Sharpe es algo mayor con PCA (0,017865 vs. 0,0138), pero BARRA presenta un Treynor y un α ligeramente superiores (0,0206 y $1,0 \times 10^{-4}$ frente a 0,019095 y $6,9 \times 10^{-5}$).

Tabla 1: Mejora de BARRA frente a PCA(3) en ajuste (% riesgo sistemático y AIC)

| Cartera | % Riesgo sistemático | | | AIC | | |
|---------------|----------------------|--------|---------------|------------|------------|----------|
| | BARRA | PCA(3) | Δ p.p. | BARRA | PCA(3) | Δ |
| Equiponderada | 96,34 | 46,78 | 49,56 | -51 206,40 | -45 744,65 | -5461,75 |
| Size_LS | 34,26 | 1,03 | 33,23 | -40 125,61 | -35 889,99 | -4235,62 |
| Beta_LS | 3,57 | 6,66 | -3,09 | -37 078,87 | -37 797,74 | 718,87 |

Nota: Δ p.p. = BARRA - PCA(3) en puntos porcentuales; Δ AIC = BARRA - PCA(3). Valores de Δ AIC más negativos favorecen a BARRA.

A modo de resumen, el balance global favorece a BARRA en las carteras amplias y representativas del espacio de riesgo, tanto por la proporción de varianza explicada como por la mejora sustantiva del AIC, a pesar de la mayor complejidad paramétrica del modelo (22 factores en BARRA vs 3 en PCA). La razón es doble: por un lado, BARRA incorpora información económica y sectorial que estructura el riesgo de forma interpretable; por otro, no fuerza la ortogonalidad entre factores, lo que permite recoger co-movimientos reales entre estilos y geografías que un espacio PCA reducido puede dejar fuera. Para estrategias long-short muy específicas, como en Beta LS, ninguno de los dos marcos explica suficientemente la variación, lo que sugiere ampliar el número de factores en el enfoque PCA (p.ej., $K > 3$), enriquecer el set BARRA con características adicionales o considerar otras técnicas de estimación como por ejemplo empleando algoritmos de Machine Learning.

5. Conclusiones

En conclusión, el presente trabajo ha permitido contrastar de manera empírica la capacidad explicativa de dos modelos factoriales utilizados en la modelización de rendimientos financieros (al menos de manera teórica): el modelo de Componentes Principales (PCA) y el modelo de tipo BARRA. A lo largo del análisis, se ha comprobado que aunque ambos métodos parten de una estructura común de descomposición del rendimiento en factores sistemáticos e idiosincráticos, sus bases conceptuales y su alcance explicativo son relativamente distintos.

La herramienta de PCA, al basarse en una descomposición puramente estadística/matemática, logra capturar las principales fuentes de variabilidad del conjunto de activos sin necesidad de información externa; sin embargo, su interpretación económica resulta limitada y su capacidad de explicar la varianza total se reduce considerablemente cuando se restringe el número de componentes. Además, se sabe que añadir un mayor número de componentes tampoco habría garantizado un modelo con mejores resultados. Por el contrario, el modelo BARRA al incorporar características observables de las empresas y estructurales como el país, el tamaño o el estilo, ofrece una descomposición de la varianza más coherente con la realidad económica, alcanzando niveles de riesgo sistemático explicados superiores.

Los resultados empíricos corroboran esta diferencia: mientras el PCA explica de forma adecuada la dinámica global del mercado a través de sus primeros componentes (particularmente el PC1, análogo al factor de mercado del CAPM), el modelo BARRA logra capturar con mayor precisión las fuentes específicas de riesgo, reduciendo de manera sustancial el AIC y mejorando la calidad del ajuste. En carteras como la equiponderada, el modelo BARRA ha mostrado una clara mejora tanto en la descomposición de la varianza como en las métricas utilizadas, mostrando que los factores económicos aportan una interpretación más estable del comportamiento de los activos. En cambio, en estrategias más específicas, como las carteras long-short de tamaño o beta, ambos modelos encuentran limitaciones para recoger plenamente la estructura de los rendimientos, lo que pone de evidencia la necesidad de ampliar el número de factores o introducir nuevas dimensiones.

En resumen, la evidencia empírica sugiere que aunque el PCA constituye una herramienta útil para identificar patrones y reducir la dimensionalidad del problema, el modelo BARRA proporciona una representación más interpretable y económicamente fundamentada del riesgo sistemático. La combinación de ambos enfoques (con el objetivo de combinar la potencia descriptiva del PCA y la interpretabilidad estructural del BARRA) se podría mostrar como un camino prometedor para futuras investigaciones o extensiones del trabajo.

6. Apéndices

6.1. Apéndice I: Análisis Descriptivo

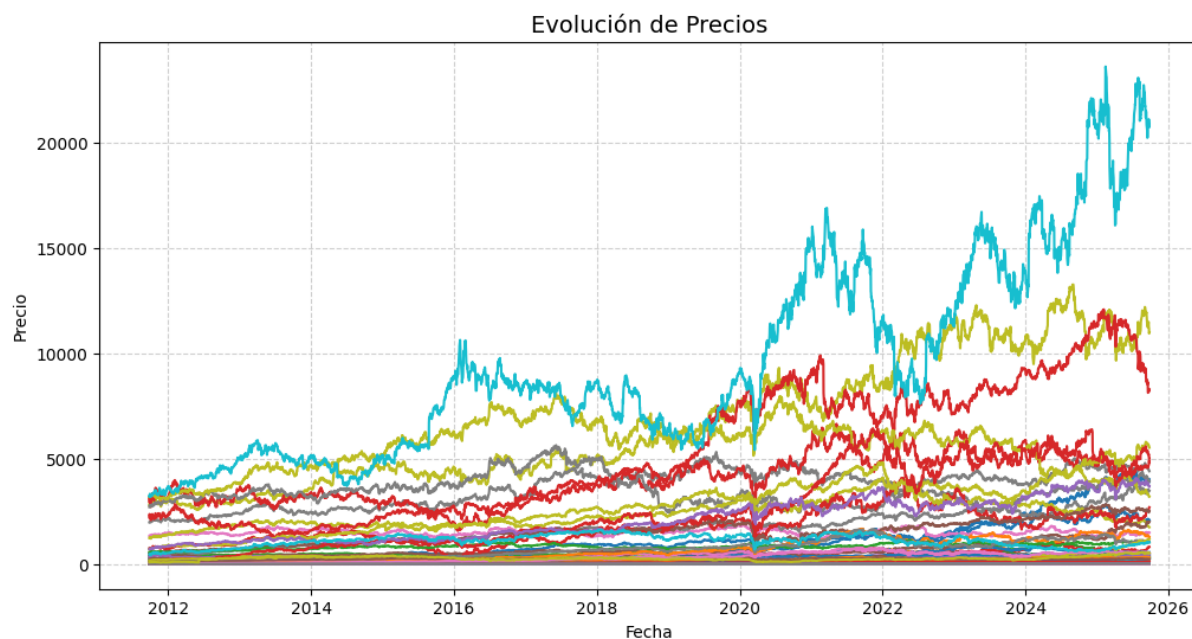


Figura 1: Evolución de precios de la muestra de 100 empresas.

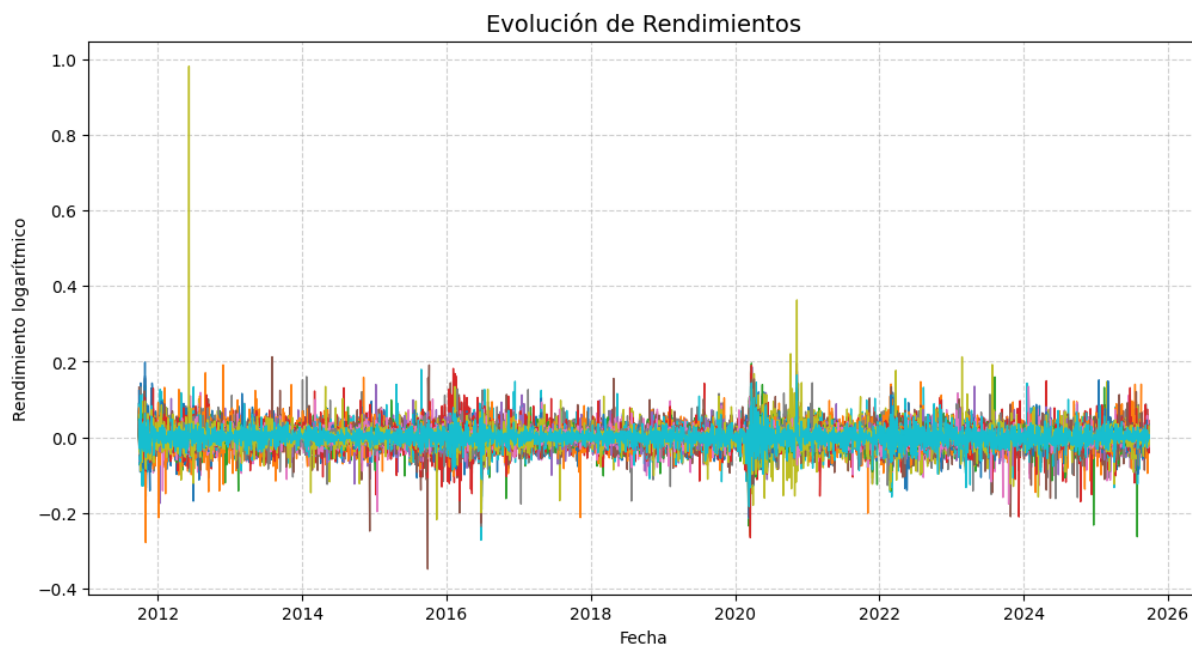


Figura 2: Evolución de rendimientos logarítmicos de la muestra de 100 empresas
Destaca un alto rdto. de la empresa Natwest Group en 2012.

Tabla 2: Estadísticos descriptivos de los rendimientos diarios de las empresas (muestra parcial)

| | 3I GROUP | LVMH | NOVO NORDISK 'B' | ASML HOLDING | NESTLE 'N' | NOVARTIS 'R' | L'OREAL | SSAB B ... |
|-------|----------|---------|------------------|--------------|------------|--------------|---------|-------------|
| count | 3654 | 3654 | 3654 | 3654 | 3654 | 3654 | 3654 | 3654 |
| mean | 0.0008 | 0.0005 | 0.0005 | 0.0009 | 0.0001 | 0.0002 | 0.0004 | 0.0001 ... |
| std | 0.0171 | 0.0168 | 0.0181 | 0.0199 | 0.0097 | 0.0114 | 0.0137 | 0.0237 ... |
| min | -0.1519 | -0.0908 | -0.2628 | -0.1701 | -0.0707 | -0.1030 | -0.0788 | -0.1518 ... |
| 25 % | -0.0072 | -0.0084 | -0.0080 | -0.0092 | -0.0047 | -0.0053 | -0.0070 | -0.0122 ... |
| 50 % | 0.0008 | 0.0003 | 0.0000 | 0.0008 | 0.0000 | 0.0000 | 0.0003 | 0.0000 ... |
| 75 % | 0.0097 | 0.0092 | 0.0097 | 0.0115 | 0.0053 | 0.0062 | 0.0077 | 0.0125 ... |
| max | 0.1572 | 0.1206 | 0.1592 | 0.1306 | 0.0603 | 0.0743 | 0.0810 | 0.1604 ... |

Tabla 3: Otros estadísticos descriptivos de los rendimientos diarios (muestra parcial)

| Activo | Media | Desv. típica | Asimetría | Curtosis |
|------------------|-----------|--------------|------------|------------|
| 3I GROUP | 0,000 832 | 0,017 077 | −0,344 251 | 7,475 467 |
| LVMH | 0,000 464 | 0,016 847 | 0,150 596 | 3,586 417 |
| NOVO NORDISK 'B' | 0,000 509 | 0,018 134 | −1,708 645 | 26,182 030 |
| ASML HOLDING | 0,000 930 | 0,019 935 | −0,240 661 | 5,271 180 |
| NESTLE 'N' | 0,000 102 | 0,009 729 | −0,424 887 | 4,864 625 |



6.2. Apéndice II: PCA

Tabla 4: Varianza explicada y acumulada por las componentes principales

| Componente | Var. explicada (%) | Var. acumulada (%) |
|------------|--------------------|--------------------|
| PC1 | 37,6300 | 37,6300 |
| PC2 | 5,9809 | 43,6109 |
| PC3 | 4,0292 | 47,6401 |
| PC4 | 2,5735 | 50,2136 |
| PC5 | 2,3073 | 52,5209 |
| PC6 | 1,9415 | 54,4625 |
| PC7 | 1,5809 | 56,0434 |
| PC8 | 1,4447 | 57,4881 |
| PC9 | 1,2670 | 58,7551 |
| PC10 | 1,2286 | 59,9837 |
| PC11 | 1,1233 | 61,1069 |
| PC12 | 1,0443 | 62,1512 |
| PC13 | 0,9639 | 63,1151 |
| PC14 | 0,9540 | 64,0691 |
| PC15 | 0,9075 | 64,9767 |
| PC16 | 0,8691 | 65,8458 |
| PC17 | 0,8280 | 66,6738 |
| PC18 | 0,7875 | 67,4613 |
| PC19 | 0,7691 | 68,2304 |
| PC20 | 0,7528 | 68,9832 |
| PC21 | 0,7454 | 69,7286 |

Tabla 5: Ratios de performance de las carteras

| Cartera | Sharpe | Treynor | Alfa Jensen |
|--------------|---------|--------------------------|--------------------------|
| Cartera_Beta | 1,6900 | $4,4900 \times 10^{-5}$ | $-2,8100 \times 10^{-5}$ |
| Cartera_Size | -9,1000 | $-5,9100 \times 10^{-5}$ | $-8,4700 \times 10^{-5}$ |
| Cartera_PC3 | -2,8500 | $3,6900 \times 10^{-5}$ | $-1,4100 \times 10^{-5}$ |
| Cartera_PC2 | 1,2000 | $-4,3600 \times 10^{-5}$ | $1,8100 \times 10^{-5}$ |
| Cartera_PC1 | 3,3300 | $6,0900 \times 10^{-5}$ | 0,0 |
| Cartera_Eq | 3,4400 | $6,2700 \times 10^{-5}$ | $1,0600 \times 10^{-5}$ |

Tabla 6: Ratios de performance de las carteras tras ajustar modelo PCA

| Cartera | Sharpe | Treynor | Alfa Jensen |
|---------------|-----------|-----------|------------------------|
| Equiponderada | 0,034 38 | 0,034 41 | $3,57 \times 10^{-4}$ |
| Size LS | -0,011 70 | -0,019 13 | $-2,60 \times 10^{-5}$ |
| Beta LS | 0,017 87 | 0,019 09 | $6,90 \times 10^{-5}$ |

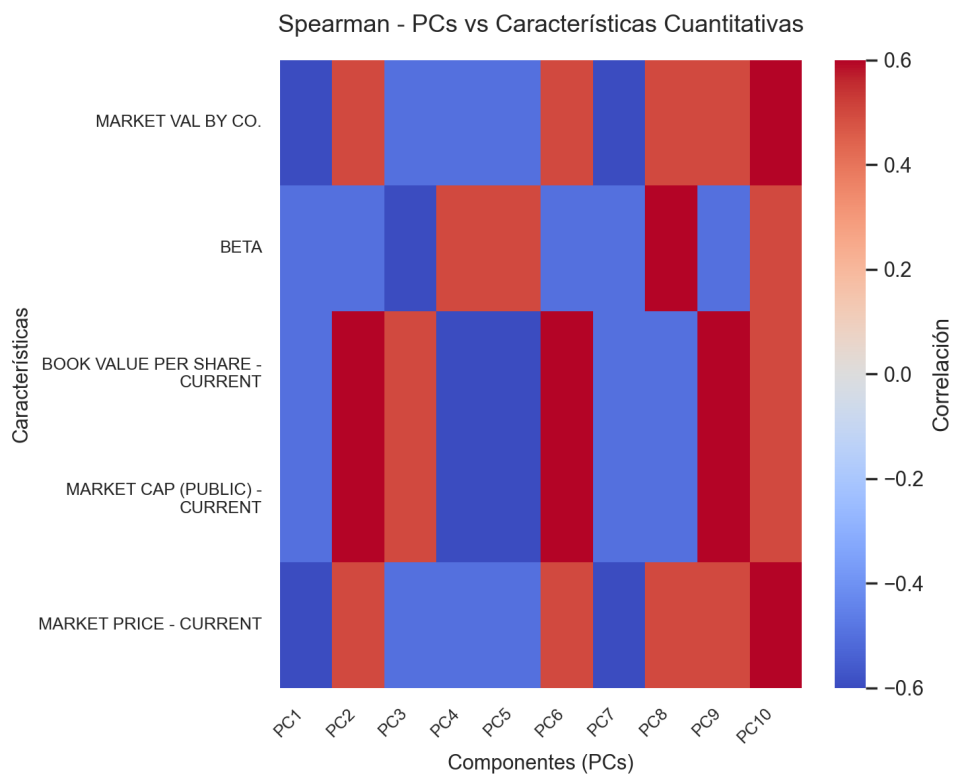


Figura 4: Correlaciones de spearman entre componentes principales y características cuantitativas.

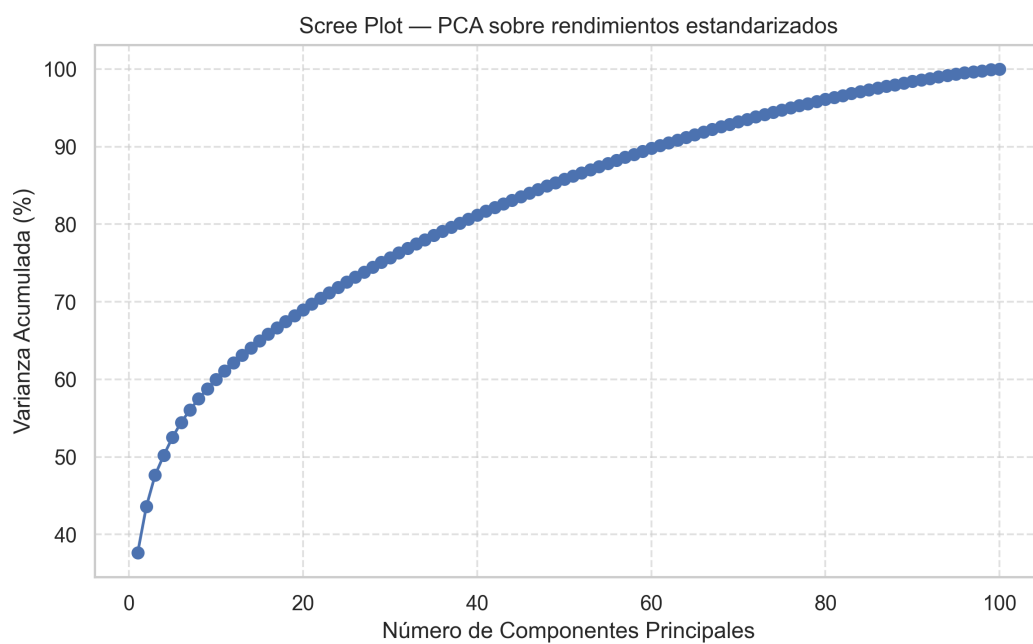


Figura 5: Gráfico de PC's respecto a la varianza acumulada

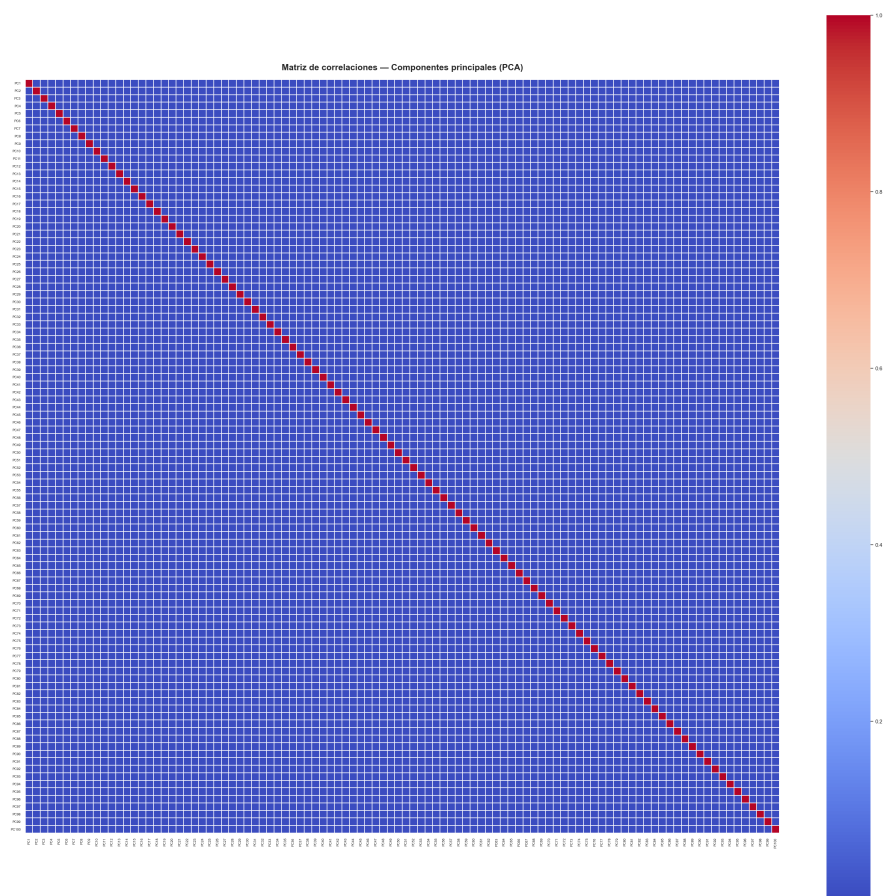


Figura 6: Matriz de correlaciones de los PC's

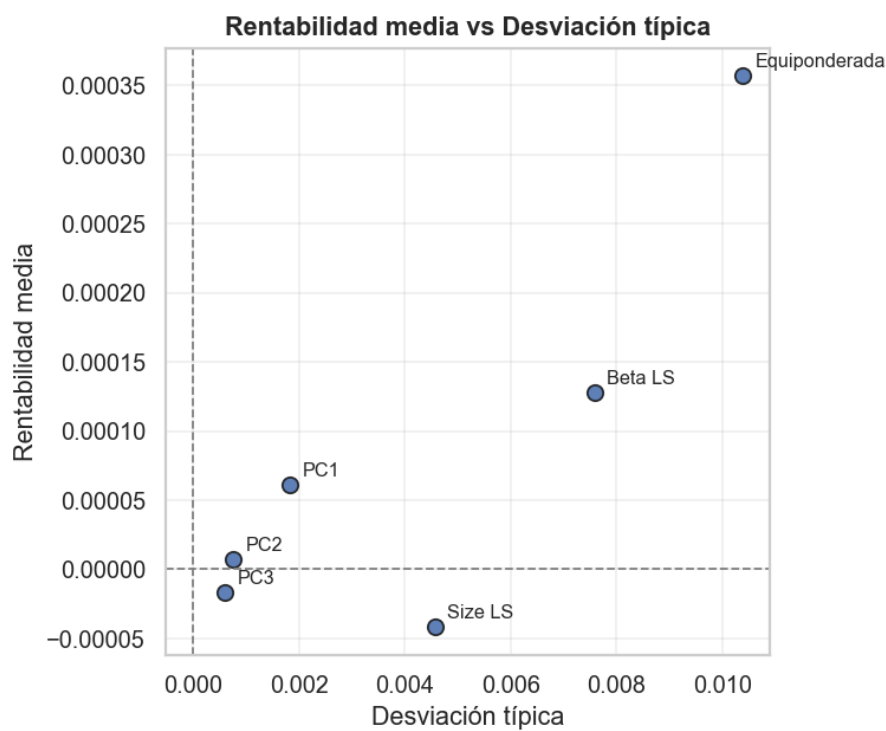


Figura 7: Gráfico rentabilidad - desv. típica para las carteras consideradas

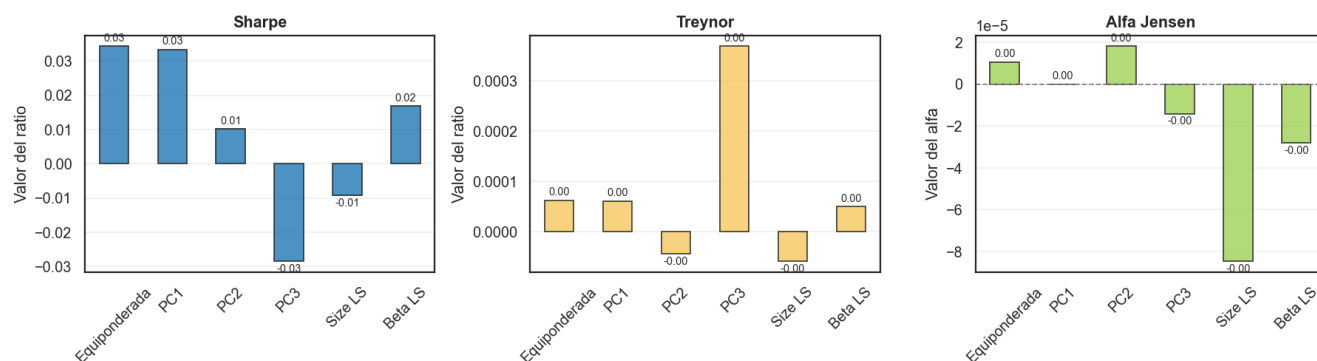


Figura 8: Ratios de performance de las carteras Equiponderada, PC1, PC2, PC3, Size_LS y Beta_LS

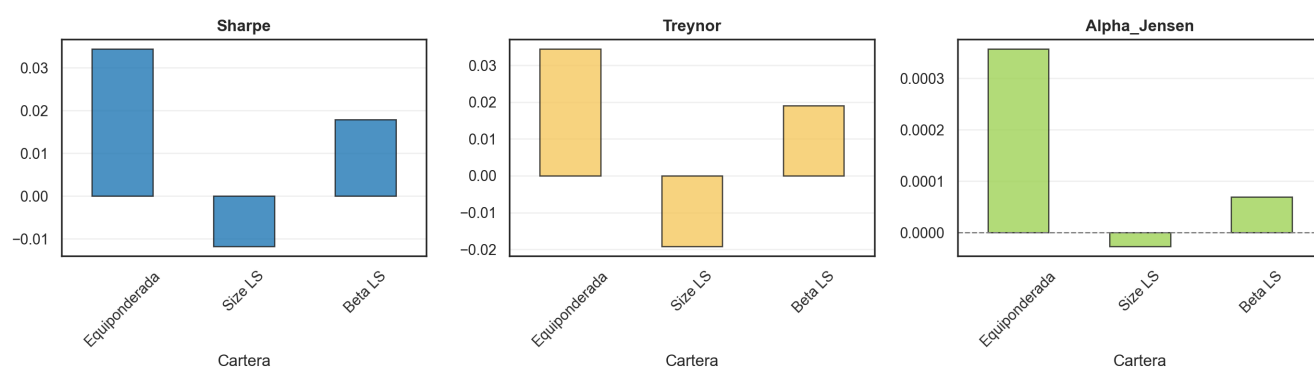


Figura 9: Ratios de performance de las carteras Equiponderada, Size_LS y Beta_LS tras sus ajustes mediante el modelo PCA

6.3. Apéndice III: Modelo BARRA

Tabla 7: Medias temporales de los factores estimados en el modelo BARRA

| Factor | Media temporal |
|---------------|----------------|
| country_DK | 0,000 331 |
| country_ES | 0,000 268 |
| country_FN | 0,000 353 |
| country_FR | 0,000 454 |
| country_IR | 0,000 777 |
| country_IT | 0,000 415 |
| country_NL | 0,000 265 |
| country_NW | 0,000 285 |
| country_SD | 0,000 391 |
| country_SW | 0,000 331 |
| country_UK | 0,000 347 |
| beta_def | −0,000 057 |
| beta_muy_agr | 0,000 019 |
| beta_muy_def | −0,000 064 |
| beta_neutra | −0,000 036 |
| book_med | 0,000 004 |
| book_mega | 0,000 059 |
| book_micro | −0,000 015 |
| book_small | 0,000 010 |
| size_MegaCap | −0,000 149 |
| size_MicroCap | 0,000 006 |
| size_MidCap | 0,000 039 |
| size_SmallCap | −0,000 004 |

Tabla 8: Descomposición de varianza por cartera en el modelo BARRA

| Cartera | Var. total | Var. sist. | Var. idio. | % sist. | % idio. | AIC |
|---------------|------------------------|------------------------|------------------------|---------|---------|-----------|
| Eq. ponderada | $1,079 \times 10^{-4}$ | $1,039 \times 10^{-4}$ | $3,953 \times 10^{-6}$ | 96.34 | 3.66 | -51206.40 |
| PC1 | $1,185 \times 10^{-4}$ | $1,047 \times 10^{-4}$ | $1,379 \times 10^{-5}$ | 88.37 | 11.63 | -48567.71 |
| PC2 | $1,041 \times 10^{-4}$ | $1,037 \times 10^{-4}$ | $8,703 \times 10^{-7}$ | 99.17 | 0.83 | -47979.01 |
| PC3 | $9,781 \times 10^{-5}$ | $1,019 \times 10^{-4}$ | 0.0 | 104.16 | 0.00 | -50412.94 |
| Size_LS | $2,084 \times 10^{-5}$ | $7,136 \times 10^{-6}$ | $1,370 \times 10^{-5}$ | 34.26 | 65.75 | -40125.61 |
| Beta_LS | $5,749 \times 10^{-5}$ | $2,053 \times 10^{-6}$ | $5,544 \times 10^{-5}$ | 3.57 | 96.43 | -37078.87 |

Tabla 9: Contribución al riesgo sistemático (SRC) por bloque de factores para cada cartera

| Cartera | Factor | SRC [$\times 10^{-2}$] | % SRC |
|---------------|---------------|--------------------------|-------|
| Eq. ponderada | country_UK | 23,090 | 23,09 |
| | country_FR | 21,686 | 21,69 |
| | country_SW | 15,393 | 15,39 |
| | country_NL | 8,466 | 8,47 |
| | country_SD | 7,273 | 7,27 |
| PC1 | country_FR | 22,031 | 22,03 |
| | country_UK | 21,703 | 21,70 |
| | country_SW | 15,224 | 15,22 |
| | country_NL | 9,842 | 9,84 |
| | country_SD | 6,623 | 6,62 |
| Size_LS | book_small | 77,003 | 77,00 |
| | country_UK | 24,465 | 24,47 |
| | country_ES | 9,917 | 9,92 |
| | country_IT | 3,527 | 3,53 |
| | country_SD | 2,765 | 2,77 |
| Beta_LS | book_small | 46,189 | 46,19 |
| | country_IR | 18,514 | 18,51 |
| | size_MicroCap | 13,917 | 13,92 |
| | country_FR | 13,662 | 13,66 |
| | country_IT | 8,048 | 8,05 |

Tabla 10: Ratios de performance por cartera en el modelo BARRA

| Cartera | Sharpe | Treynor | Alpha |
|---------------|--------|---------|---------|
| Eq. ponderada | 0.0356 | 0.0345 | 0.0000 |
| PC1 | 0.0354 | 0.0345 | 0.0000 |
| PC2 | 0.0340 | 0.0323 | -0.0000 |
| PC3 | 0.0361 | 0.0344 | -0.0000 |
| Size_LS | 0.0120 | 0.0098 | -0.0001 |
| Beta_LS | 0.0138 | 0.0206 | 0.0001 |

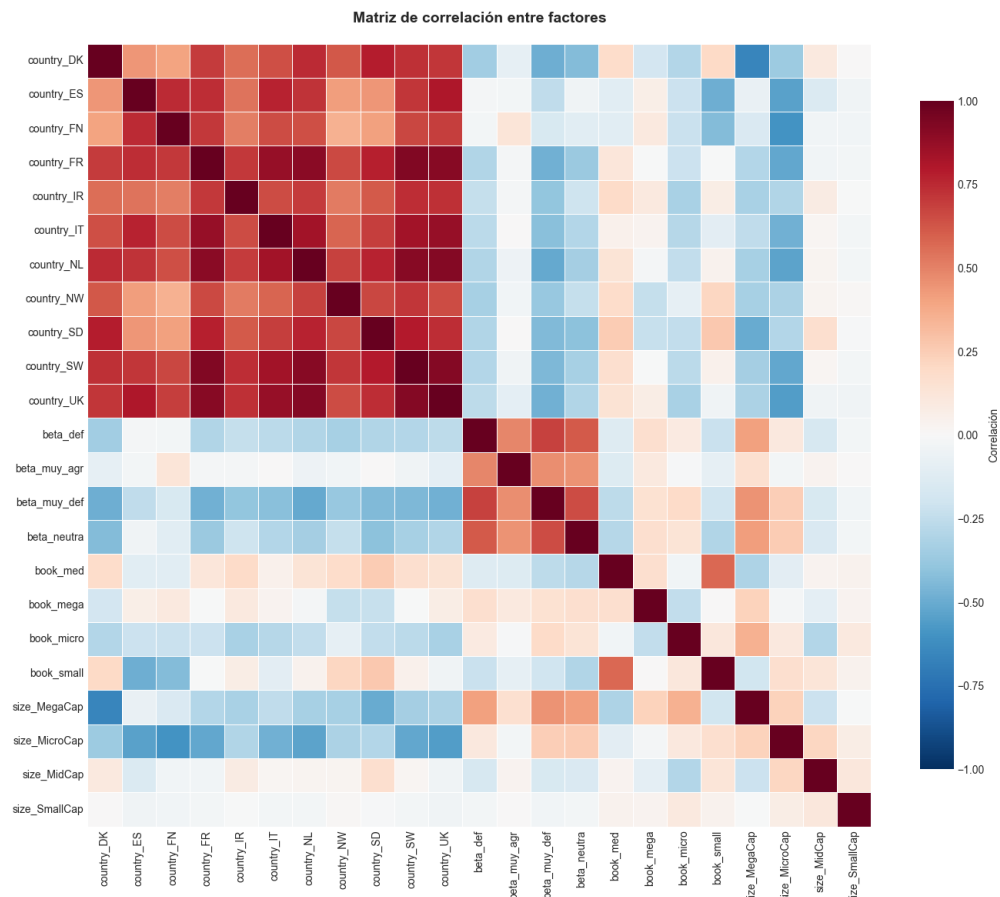


Figura 10: Gráfico de correlaciones de los factores ajustados en el modelo BARRA.

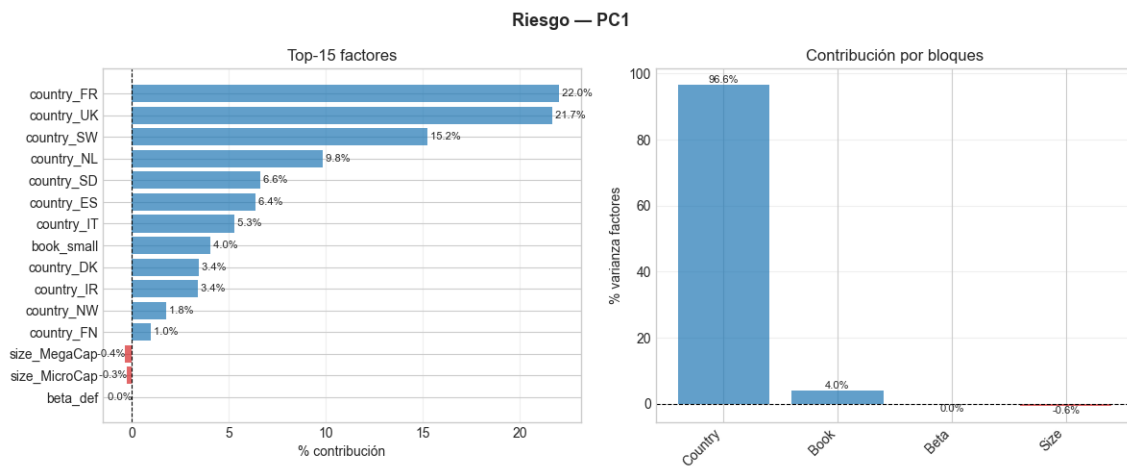


Figura 11: Descomposición de la varianza sistemática para cada bloque de factores para la Cartera PC1.

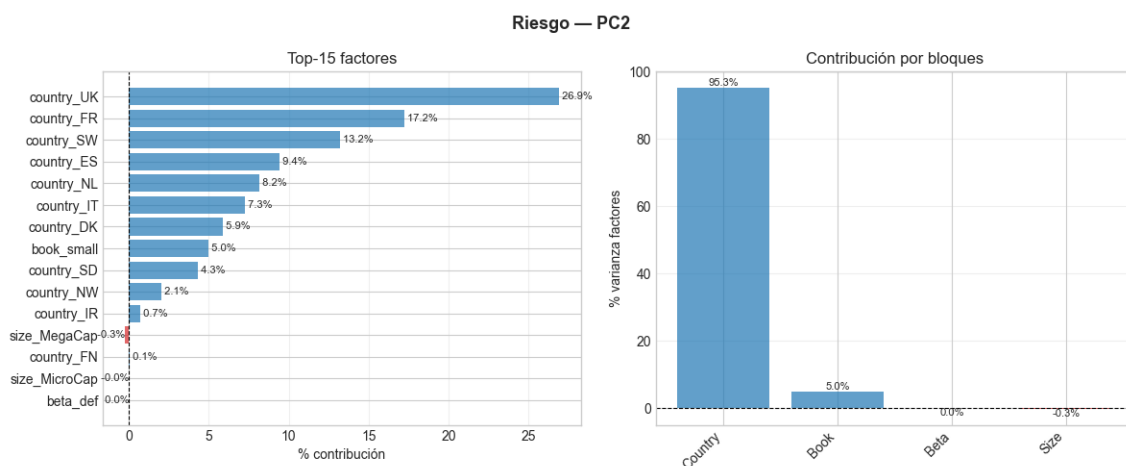


Figura 12: Descomposición de la varianza sistemática para cada bloque de factores para la Cartera PC2.

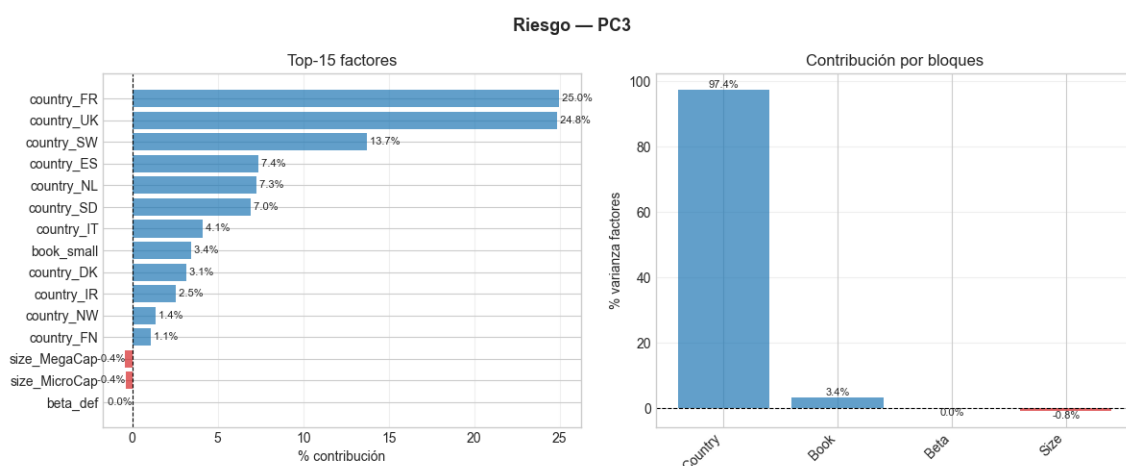


Figura 13: Descomposición de la varianza sistemática para cada bloque de factores para la Cartera PC3.

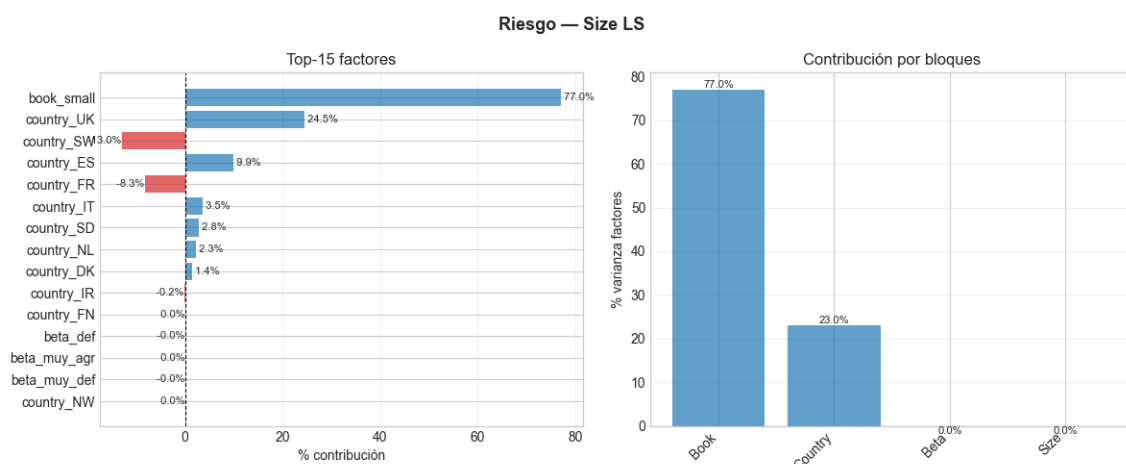


Figura 14: Descomposición de la varianza sistemática para cada bloque de factores para la Cartera Size.

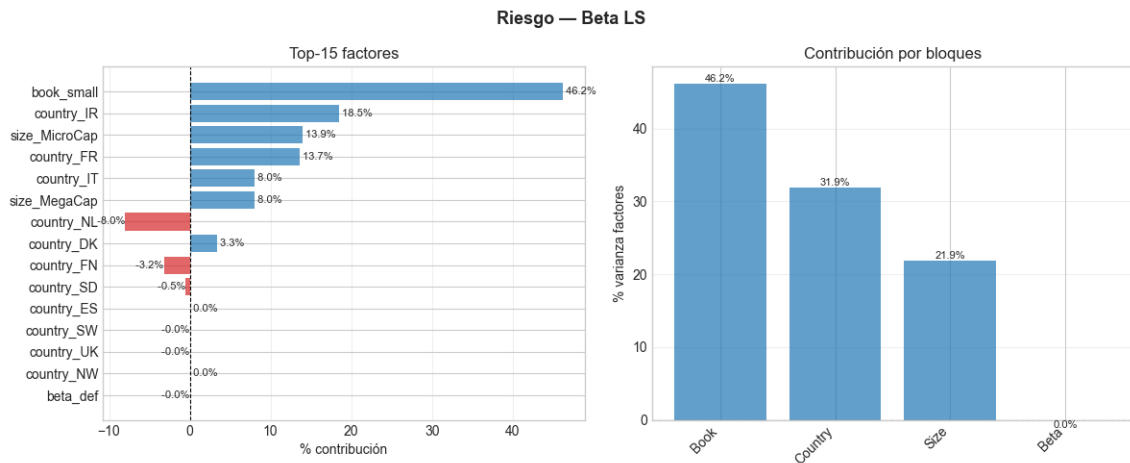


Figura 15: Descomposición de la varianza sistemática para cada bloque de factores para la Cartera Beta.

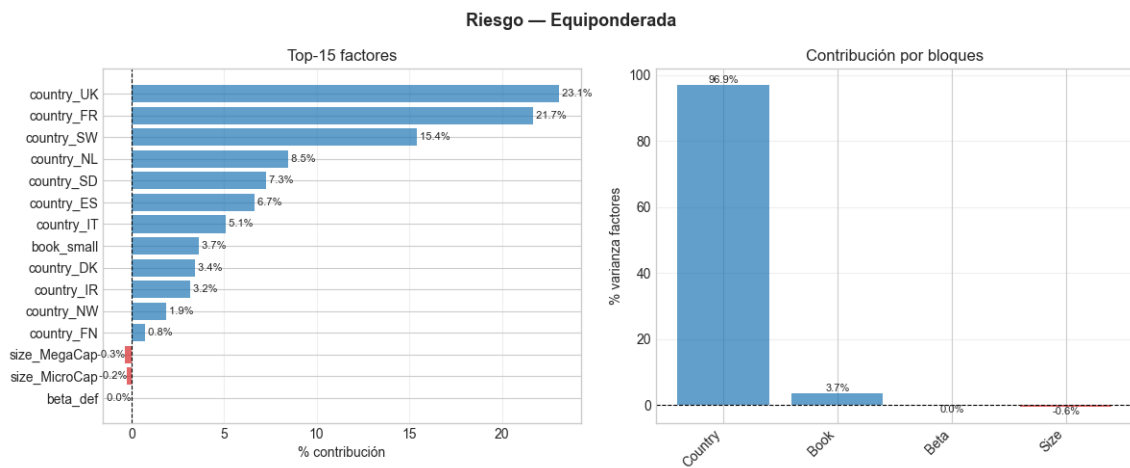


Figura 16: Descomposición de la varianza sistemática para cada bloque de factores para la Cartera Equiponderada.

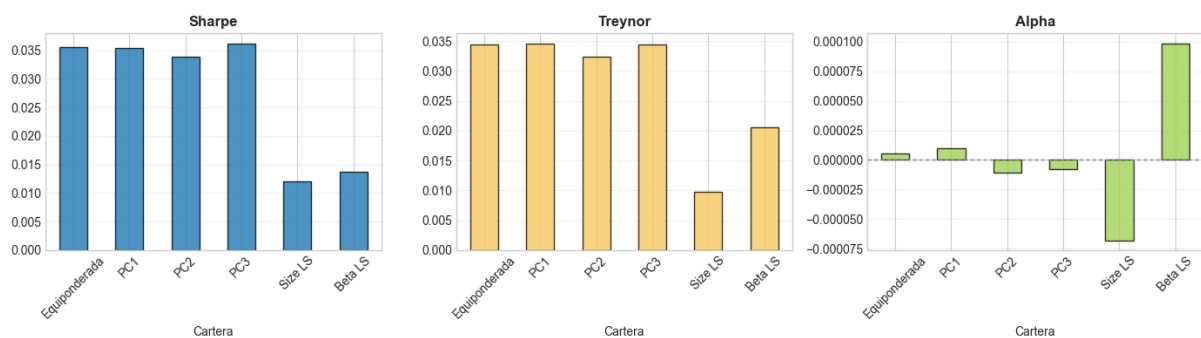


Figura 17: Ratios de performance para cada una de las carteras según el modelo BARRA