Financial Innovation

RESEARCH

Open Access

# Bridging inverse participation ratio and portfolio theory

Laura Molero González[1,2*], Roy Cerqueti[2,3], Juan E. Trinidad Segovia[1] and Miguel A. Sánchez Granero[4]

*Correspondence: laura.molerogonzalez@uniroma1.it

[1] Department of Economics and Business, University of Almer´ıa, 04120 Almer´ıa, Spain
[2] Department of Social and Economic Sciences, Sapienza University of Rome, 00185 Rome, Italy
[3] GRANEM, University of Angers, 49100 Angers, France
[4] Department of Mathematics, University of Almer´ıa, 04120 Almer´ıa, Spain

## Abstract

This research seeks to delve deeper into the financial significance of the eigenvector associated with the second-largest eigenvalue of the correlation matrix of returns. The objective is not only to gain a better understanding of this value, but also of the components that determine its behavior. To this end, we rely on daily logarithmic returns of the S&P's 500 index constituents over the period ranging from 2005 to 2024. We propose the use of the Inverse Participation Ratio (IPR), which is a measure extensively used in Physics, and we introduce the $N_{eff}$ metric to identify the effective stocks within a portfolio; i.e., those capable of preserving the return dynamics of the full portfolio by themselves. Our empirical analysis shows that, in general, only long positions in the utilities sector are effective within this portfolio, which is consistent with the counter behaviour already observed and the traditional defensive nature of this economic sector. We also find that during periods of financial distress, however, short sales in the materials sector and the financial sector also emerge as effective.

**Keywords:** Inverse participation ratio, Eigenvalue, Portfolio theory

## Introduction

The study of correlation matrices of financial asset returns is one of the most relevant topics in finance (Moura et al. 2020; Baruník and Cech 2021; De Nard et al. 2022; Paulusch and Schlütter 2022; Faias 2023). Correlation matrices play a crucial role in the selection of the optimal portfolio in Modern Portfolio Theory. However, its estimation is quite challenging considering the documented presence of time dependency, noisy and extreme events (Stolfi et al. 2022). It is well known that the longer the series and the greater the number of assets to be selected, the lower the reliability of the estimation from historical return series data (Pafka and Kondor 2004).

The introduction of structural changes in the covariance matrix, such as the diagonalisation proposed by Sharpe (1964), has been one of the research lines followed by the financial literature (Dom et al. 2025). The use of Random Matrix Theory (RMT) has provided a new insight into the approach to the problem of noise in correlation matrix estimation. Several applications (see, for example, Laloux et al. 1999; Bouchaud and Potters 2003) have shown that empirical correlation matrices exhibit a few large eigenvalues and corresponding eigenvectors that capture the most significant collective effects. The remaining eigenvalues form the so-called bulk, which is primarily dominated by noise but may also contain non-negligible correlations (Kwapień and Drożdż 2012; Bun et al. 2017). Biely and Thurner (2008) study lagged correlation matrices and conclude that the largest eigenvalue is real, which confirms that the correlational structure is symmetric. However, the second and third-largest eigenvalues result to be a complex conjugate pair, which have opposite imaginary parts but the same real value. This fact indicates a strong directionality or cyclical dynamics, as one sector leads the others over time. The eigenvector associated with the second-largest eigenvalue has often been interpreted in the literature as a sectoral structure carrier (see, for instance, Plerou et al. (2002); Coronnello et al. (2005), or Jiang and Zheng (2012)).

In the specific context of finance, this means that the second-highest eigenvalue is related to the way assets co-behave. Also, the eigenvectors associated with the highest eigenvalues have a clear financial meaning: the highest eigenvalue represents the market mode (Plerou et al. 2002; Tang et al. 2023; Molero González et al. 2023; Domínguez-Monterroza et al. 2025; Molero González et al. 2025). In general, Song et al. (2011) and Tang et al. (2023) affirm that important information is contained in the first two largest eigenvalues and their dynamics. In the financial literature, it has been widely acknowledged that the largest eigenvalue of the correlation matrix of returns corresponds to the market factor, also called market mode (Plerou et al. 2002; Tang et al. 2023; Molero González et al. 2023; Domínguez-Monterroza et al. 2025; Molero González et al. 2025). Previous works of Song et al. (2011) and Tang et al. (2023) have emphasised the informational relevance of the first two largest eigenvalues and their dynamics. More specifically, Molero González et al. (2025) highlighted that the second-largest eigenvalue exhibits a counter behaviour with respect to the market spillover, which can make it be considered as a "safe haven" during periods of financial distress. Building on this, this paper seeks a better understanding of this second-largest eigenvalue and its associated eigenvector, and the components determining its behaviour.

To this end, we propose the use of the Inverse Participation Ratio (IPR), which is a measure extensively used in scattering theory, quantum mechanics, and particle physics (see e.g., Plerou et al. (2002); Liu et al. (2025)). The IPR is computed through the eigenvectors of a suitably defined reference matrix. It is used as an indicator of the spatial distribution of wave functions in quantum systems and reflects how localised a wave function is in a given system (Clark and Del Maestro 2018). Indeed, the space in which the wave function is localised is the principal responsible for the properties of the quantum system and is considered to be the set of positions for which the probability density is significant (Murphy et al. 2011; Calixto and Romera 2015; Perera and Wortis 2015). If the wave function is completely localised, it means that the particle is in a very small region of space and, consequently, the IPR value will be high. In contrast, if the wave function is extended, the IPR will be low, because the particle will be distributed at many positions in the system (Murphy et al. 2011; Singh and Xu 2016). The distinction between these cases is driven by the components of the considered eigenvectors. The IPR has been one of the main tools in assessing the localisation properties of single-particle wave functions throughout the Anderson localisation transition (see Thouless (1974); Kramer and MacKinnon (1993); Evers and Mirlin (2008); Rodriguez et al. (2009, 2010) and, more recently, Liu et al. (2025)).

Measures of eigenvector localization based on information entropy have been previously used in the analysis of financial correlation matrices (see, e.g., Drożdż et al. (2000)). Entropy-based indicators and the IPR are closely related and provide complementary perspectives on the degree of localization. In this study, however, we focus on the IPR primarily because it admits a direct interpretation in terms of the effective number of contributing assets, with well-defined limiting cases corresponding to fully localized or fully extended eigenvectors.

Within the field of finance, the IPR has been previously used to examine the eigenvectors of the correlation matrices of the stock returns. Utsugi et al. (2004) apply the IPR to data from the Tokyo Stock Exchange and show that, while the bulk of the IPR values is consistent with the predictions of RMT, significant deviations emerge outside the RMT bounds. García-Medina (2016) use the IPR to provide evidence about the emergence of common factors. The temporal analysis of this metric allowed the authors García-Medina et al. (2018) to confirm the fact that each financial index participates significantly in the eigenvector associated with the largest eigenvalue. Subsequently, García-Medina and González Farías (2020) used the IPR, in the specific context of cryptocurrencies, to distinguish between localized factors (those that have dominant weights related to specific assets) and non-localized factors (those that have uniform contributions). Zhao et al. (2025) uses the IPR to quantitatively measure the extent to which an eigenvector deviates from the predictions made by RMT and to assess how many significant elements contribute to the eigenvector. They show that there is a significant deviation in the eigenvector associated with the largest eigenvalue, which indicates how certain global events or policies can create widespread uncertainty. More recently, this measure has been used in portfolio theory to evaluate portfolio diversification (García-Medina 2025).

When interpreting the IPR in the financial field, the particle can be viewed as the factor that determines the components of the eigenvector/portfolio. This implies that the IPR lives between two corner cases: if the factor is fully extended, all stocks contribute equally to the portfolio; if the factor is concentrated, then only one asset contributes to the portfolio (García-Medina et al. 2023). This evidence suggests that the number of effective components plays a crucial role and depends on the IPR. We will present this concept in Sect. Reading the inverse participation ratio in finance.

Building on the previous contributions, this paper aims to strengthen the connection between the IPR and portfolio theory. Specifically, we consider the financial case of a set of stocks and select the correlation matrix among their returns as the starting matrix of IPR. When normalised, the eigenvectors of this correlation matrix represent portfolios of assets. Then, IPR can describe the spatial distribution of the shares of the portfolios (García-Medina 2025). This is exactly the approach followed in the present paper. For the sake of simplicity, we will refer hereafter simply to eigenvectors to describe their normalised version.

Using the logarithmic returns of the S&P's 500 stock constituents over the period 2005–2024, this study focuses on formalizing a metric, denoted $N_{\text{eff}}$, for identifying the effective components within a portfolio of assets. This metric is grounded in the IPR measure and is designed to quantify the extent to which individual stocks contribute meaningfully to the structure and behaviour of the portfolio. Building on the observation by Molero González et al. (2025) that the second-largest eigenvalue exhibits a counter behaviour with respect to market spillovers, thus acting as a potential "safe haven" during periods of financial distress, our approach aims to deepen the understanding of this eigenvalue, its associated eigenvector, and the components driving its dynamics. The $N_{\text{eff}}$ metric formally links the IPR to portfolio theory, allowing a rigorous assessment of how individual assets contribute to portfolio structure and risk.

Our empirical analysis shows that the proposed $N_{\text{eff}}$ metric can accurately identify the stocks that are effective within the portfolio, that is, those capable of preserving the return dynamics of the full portfolio by themselves, and reveals sectoral patterns that vary across market regimes. Furthermore, through this metric, we are able to provide a clearer interpretation of the second-largest eigenvalue. Specifically, we demonstrate that, in general, only long positions in the utilities sector are effective within this portfolio, which is consistent with the counter behaviour previously observed and with the traditionally defensive nature of this sector. During periods of financial distress, however, short positions in the materials sector (during the 2007 financial crisis) and in the financial sector (during the Covid-19 pandemic) also emerge as effective. Finally, by analysing the hyperspherical angles, we assess the temporal stability of the direction of the eigenvectors involved in the present analysis.

Overall, this study contributes to the literature by providing both methodological and empirical insights: it introduces a new tool for portfolio analysis, offers a clearer interpretation of the second-largest eigenvalue, and documents novel sectoral dynamics and patterns of portfolio effectiveness that have not been previously reported. While the empirical analysis relies on the S&P's 500 constituents over the period 2005–2024, it nonetheless imposes certain limitations. In particular, the focus on a single market and on large-cap firms may restrict the generalisability of the findings to other asset classes or market structures. These limitations, however, also open avenues for future research exploring whether the proposed metric behaves similarly across different markets, periods, or levels of market capitalisation.

The rest of the paper is organized as follows. Sect. Reading the inverse participation ratio in finance introduces the IPR measure in the specific field of Finance. Sect. Data presents the dataset used in this study. Section Methodology outlines the methodological instruments and procedures followed in this manuscript. Sect. Results and discussion presents and discusses the empirical findings. Section Conclusions offers concluding remarks.

## Reading the inverse participation ratio in finance

Considering a correlation matrix $\mathbf{C}$ of size $p \times p$, we will denote as $\lambda_k$ the generic eigenvalues of $\mathbf{C}$ and as $u^k = (u_1^k, \ldots, u_p^k)$ the eigenvector associated with $\lambda_k$, with $k = 1, ..., p$. The IPR $I^k$ is defined as (see e.g., Plerou et al. (2002); Liu et al. (2025)):

$$I^k = \sum_{l=1}^{p} \left(u_l^k\right)^4. \tag{1}$$

By construction, the term $I^k$ in (1) ranges in $\left[\frac{1}{p^3}, 1\right]$. Indeed, if we consider an equally weighted portfolio, the weight of the generic $l$-th stock is $u_l^k = \frac{1}{p}$, for $l = 1, \ldots, p$. By substituting in (1), we have:

$$I^k = \sum_{l=1}^{p} \left(\frac{1}{p}\right)^4 = p \cdot \left(\frac{1}{p}\right)^4 = \frac{1}{p^3} \tag{2}$$

This corresponds to the case in which the factor in the eigenvector is completely extended, that is, the portfolio is equally weighted across all assets, and it represents the lower bound of the range of values that the IPR can attain (Liu et al. 2025). At the opposite extreme, the factor in the eigenvector is completely localized, meaning that the portfolio is fully concentrated in a single asset. In this case, there exists $\bar{l} = 1, \ldots, p$ such that $u_l^k = \mathbf{1}(l = \bar{l})$, where $\mathbf{1}(\cdot)$ denotes the indicator function. The eigenvector, therefore, has one unit component while all remaining components are zero, which yields $I^k = 1$.

The portfolio weights associated with an eigenvector can be interpreted as measuring the relative exposure of each asset to a given factor. When a factor is concentrated on only a small number of stocks, those assets exhibit a dominant exposure to the factor, while the remaining assets contribute negligibly. In the extreme case in which the factor is fully localized in a single asset, only one stock effectively contributes (Plerou et al. 2002). Conversely, if the factor is completely extended, all $p$ assets contribute equally, and the number of effective components equals $p$. This interpretation naturally leads to a formal definition of the number of effective components $N_{\text{eff}}$ of the $k$-th eigenvector based on the inverse participation ratio, as follows:

$$N_{\text{eff}}^k = \frac{1}{\sqrt[3]{I^k}}. \tag{3}$$

According to (3) and in agreement with the arguments above, we have that $N_{\text{eff}} = p$ for the equally weighted vector while $N_{\text{eff}} = 1$ for the completely localised case.

In the following, we need to deal with ordered eigenvalues. Without loss of generality and for the sake of simplicity, we assume that:

$$\lambda_1 > \lambda_2, \cdots > \lambda_p.$$

**Table 1** Descriptive statistics of the vector $R_m$. As said in the main text, the value $N$ represents the total number of observation days

| Statistic | Value |
|---|---|
| $N$ | 5029 |
| Mean | 0.000390 |
| Standard deviation | 0.012919 |
| Minimum | -0.135831 |
| Q1 | -0.004675 |
| Q2 | 0.000871 |
| Q3 | 0.006249 |
| Maximum | 0.107203 |
| Skewness | -0.605556 |

## Data

This section focuses on describing the dataset employed in the analysis. To achieve the objectives of this study, we have taken the daily closing prices of the stocks belonging to the S&P's 500 stock market. The study period ranges from January 6th, 2005, to November 21st, 2024. Therefore, the total number of trading days considered is $N = 5029$.

Since not all firms remain continuously listed or liquid throughout the entire sample, a data-quality criterion has been adopted to construct the sample. Concretely, we have established a selection criterion whereby only companies with less than 5% of missing trading days are taken, to clean the sample and avoid problems in later stages. This process allows removing the stocks with insufficient price histories (e.g., due to recent index entry, delisting, low liquidity, or shorter trading records), which ensures a stable and high-quality panel for correlation estimation. Thus, the sample consists of a total number of $p = 484$ companies. Then, we computed the daily logarithmic returns.

The constituent universe of the S&P 500 employed in this study is static, based on the index composition as of the end of 2024. Analyzing a fixed set of firms retrospectively from 2005 to 2024 introduces a potential survivorship bias. However, this selection ensures a continuous and high-quality panel of data, essential for the stability of correlation matrix estimators. Furthermore, since our study focuses on dominant market collective modes and broad sectoral structures, previous literature (Plerou et al. 2002) suggests these modes are robust to the idiosyncratic behavior of the firm subset.

Since some stocks are more volatile than others, and for the sake of comparison, we conveniently work with standardised returns. The matrix under consideration is the correlation matrix $\mathbf{C}$ for the daily logarithmic returns of the $p$ stocks belonging to the S&P's 500 market.

Just to provide a piece of information about the considered dataset, we present an illustrative paradigmatic example based on the highest eigenvalue $\lambda_1$ of $\mathbf{C}$. Specifically, we consider the equally weighted eigenvector/portfolio $u^1$, which corresponds to $IPR = \frac{1}{p^3}$. To this aim, we consider for each day a naive portfolio with equally weighted shares of $1/p$ on all the considered stocks, i.e. $u^1 = (1/p, \ldots, 1/p)$, where the eigenvector has been suitably normalised. We obtain the series of the logarithmic returns of this portfolio. Authoritative literature explains that it represents the market factor (see

**Table 2** Global industry classification standard activity sector, jointly with the identification number given to each one to facilitate further analysis

| ID number | Sector of activity |
|---|---|
| 1 | Industrials |
| 2 | Healthcare |
| 3 | Information technology |
| 4 | Utilities |
| 5 | Financials |
| 6 | Materials |
| 7 | Consumer |
| 8 | Real estate |
| 9 | Communication services |
| 10 | Consumer staples |
| 11 | Energy |

Plerou et al. (2002); Tang et al. (2023); Molero González et al. (2023); Domínguez-Monterroza et al. (2025) or Molero González et al. (2025)). We denote this series by $R_m = (R_m(1), \ldots, R_m(N))$, where $m$ stands for *market*.

Table 1 shows the main descriptive statistics of the components of $R_m$. The mean daily logarithmic return of $R_m$ is 0.0390%, while the standard deviation is $\sigma_m = 1.2919\%$. The lowest log-return in $R_m$ is -13.5831% and the highest is 10.7203%. During half of the period under analysis, the log-returns in $R_m$ remained below 0.0871%. Only 25% of the days took values of the market factor greater than 0.6249%. The distribution of the elements of $R_m$ exhibits a slightly negative asymmetry, which suggests that there are more small values away from the mean than large ones.

## Methodology

This section is devoted to the presentation of the methodological instruments and procedures employed in the present analysis. To achieve the objective of our study, a sliding window experiment has been designed, with windows of $n = 504$ days, ensuring $n \geq p$ to avoid high-dimensionality issues. Although the condition $n \geq p$ does not fully eliminate high-dimensionality problems, it allows preventing the covariance matrix from becoming singular. When $n < p$, the sample covariance matrix $\Sigma = n^{-1}X^TX$ ranks, at most, $n$, resulting in $(p - n)$ exact zero eigenvalues. Thus, the matrix becomes singular and non-invertible.

While the MP distribution is mathematically well defined for any aspect ratio $q = \frac{p}{n} > 0$, including $q > 1$, in that regime it necessarily includes a point mass at zero of weight $1 - \frac{1}{q}$, reflecting the $(p - n)$ zero eigenvalues produced by the rank deficiency of the sample covariance matrix. Since the present analysis focuses on the continuous part of the MP spectrum and requires a non-singular empirical covariance matrix for subsequent procedures, we restrict attention to windows satisfying $n \geq p$.

Windows move $\Delta = 60$ days successively, that is, roughly one earnings cycle. The number of analysed windows is $J = 76$ and the generic window is $j = 1, \ldots, J$.

For each window $j$, we compute the eigenvalues $\lambda_1(j), \ldots, \lambda_p(j)$ of $\mathbf{C}_j$, jointly with their associated eigenvectors, $u^1(j), \ldots, u^p(j)$, respectively. Then, we compute the $I^k(j)$ and $N^k_{eff}(j)$ of Sect. Reading the inverse participation ratio in finance for each window $j$, using Eqs. (1) and (3), respectively.

As an initial robustness check, for each window $j$, we take the eigenvectors associated with the largest and second-largest eigenvalues of $\mathbf{C}_j$, e.g., $u^1(j)$ and $u^2(j)$, respectively. As usual, we consider a normalised version of them as portfolios. Then, we compute the expected (mean) return of each portfolio given by the eigenvector by considering all stocks, i.e. $E\left(R_{u^1}(j)\right) = (r^1_1(j), \ldots, r^1_p(j))$ and $E\left(R_{u^2}(j)\right) = (r^2_1(j), \ldots, r^2_p(j))$.

Then, we restrict the procedure to the effective components $N^k_{eff}(j)$, with $k = 1, 2$. To implement this step, we first we consider an order statistic of the components of $R_{u^1}(j)$ and $R_{u^2}(j)$, by sorting them in decreasing order. We obtain the vectors $R_{u^{(1)}}(j) = (r^1_{(1)}(j), \ldots, r^1_{(p)}(j))$ and $R_{u^{(2)}}(j) = (r^2_{(1)}(j), \ldots, r^2_{(p)}(j))$, where $r^k_{(1)}(j) > \cdots > r^k_{(p)}(j)$, for $k = 1, 2$. Once the number $N^k_{\text{eff}}(j)$ is found, then we consider the first $N^k_{\text{eff}}(j)$ elements of $R_{u^{(k)}}(j)$. Since the elements of vector $R_{u^k}(j)$ can be positive and negative and they are ranked in decreasing order, $N^k_{\text{eff}}(j)$ induces two sets of effective components – one of $N_{\text{eff}}^{k,+}(j)$ positive and one of $N_{\text{eff}}^{k,-}(j)$ negative – so that $N_{\text{eff}}^{k}(j) = N_{\mathit{eff}}^{k,+}(j) + N_{\mathit{eff}}^{k,-}(j)$ and we consider the vector of the expected effective returns

$$E\left(R'_{u^k}(j)\right) = \left(r_s^k(j) : s \in \{1, \ldots, N_{\mathit{eff}}^{k,+}(j)\} \cup \{N_{\mathit{eff}}^{k,-}(j) + 1, \ldots, p\}\right).$$

We focus our attention on the second-largest eigenvalue of $\mathbf{C}$ and its associated eigenvector. However, we find a limitation in interpreting its meaning, derived from the fact that the largest eigenvalue is an order of magnitude larger than the others. Thus, to analyse $\lambda_2$, we must remove the effect of $\lambda_1$ (Plerou et al. 2002). To do so, we computed the correlation matrix of the residual $\varepsilon_{i,j}$ of the Sharpe (1964) Single Index Model. Thus, for each window $j$ and for each asset $i$, we estimate the model of the following expression using Ordinary Least Squares (OLS) applied within each window:

$$R_{i,j} = \alpha_i + \beta_i R_{m,j} + \varepsilon_{i,j} \tag{4}$$

The resulting residuals $\hat{\varepsilon}_{i,j}$ are then used to construct the correlation matrices for the subsequent RMT analysis.

Let us denote by $\mathbf{R}$ the correlation matrix of the residual. We shall hereinafter represent its eigenvalues and their associated eigenvectors as $\lambda'_k$ and $v^k$, respectively, with $k = 1, ..., p$.

As previously done with $\lambda_1$ and $\lambda_2$, we continue with the sliding window experiment, and we calculate the $I^k(j)$ and $N_{\mathit{eff}}^{k}(j)$ for the largest eigenvalue $\lambda'_1(j)$ of $\mathbf{R}(j)$. The robustness check presented above is also performed for this eigenvalue and its associated eigenvector.

To analyse the composition of the eigenvector associated with the second-largest eigenvalue and facilitate further experiments, the stocks were classified into sectors, and each sector was given a number. Table 2 shows the different activity sectors and their associated identification number. This classification was made based on the Global Industry Classification Standard (GICS).

The sectoral classification follows the GICS structure as of the first quarter of 2025. A static approach, beyond being more affordable, is also particularly suitable in our context. Indeed, while individual firms may have changed sectors over time, these reclassifications do not interfere with the overall findings. Given that our study identifies broad defensive versus cyclical clusters, these localized shifts do not significantly affect the dominant components.

Finally, we focus on the portfolio associated with $\lambda'_1(j)$ when considering only the effective stocks. On each window, we analyse the sector distribution of the companies included in it by examining their frequency, e.g., by counting how many companies belong to each sector in the portfolio.

## Results and discussion

This section presents the results obtained in the study. We start by checking if the elimination of the market that has been performed – i.e., taking the correlation matrix of the residuals of the Sharpe Single Index Model in (4) – was done correctly. To do so, we compute the Pearson correlation coefficient between $u^2(j)$ and $v^1(j)$. Figure 1 shows the results obtained. For the great majority of the period, $\rho_{u^2,v^1}$ remained above 0.95,

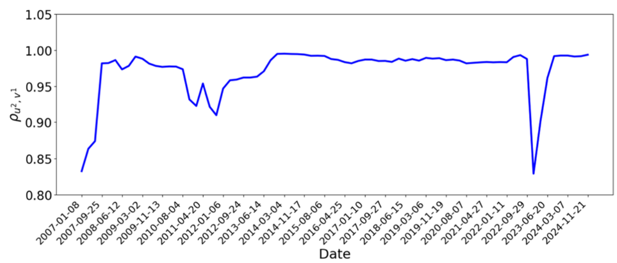



**Fig. 1** Pearson correlation coefficient between $u^2(j)$ and $v^1(j)$ as a function of time

meaning that both eigenvectors correspond to the same linear combination of assets. From August 2010 to September 2012, the correlation coefficient moved between 0.9 and 0.95. Just from January to September 2007 and from September 2022 to June 2023, the correlation coefficient between the two variables fell below 0.85.

Figure 2 shows the evolution over time of the dynamics of $\lambda_2(j)$ and $\lambda_1'(j)$. It can be observed that, in general, both eigenvalues have similar dynamics, with some remarkable deviations. Indeed, two exceptions to this generally similar behaviour are found in the 2007 financial crisis and the Covid-19 pandemic, where, while a decrease is observed for $\lambda_2(j)$, an increase can be identified for $\lambda_1'(j)$. Interestingly, the decreasing behaviour of $\lambda_2(j)$ in occasion of a growth of $\lambda_1(j)$ was already observed by Molero González et al. (2025), where the authors show that during crises, e.g., during periods of high volatility, $\lambda_1(j)$ increases, while $\lambda_2(j)$ exhibits a counter behaviour. By eliminating the market, the residual correlation matrix presents a comparable dynamic. In particular, it can be observed that $\lambda_1'(j)$ tends to increase during crisis periods while $\lambda_2(j)$ decreases. This phenomenon indicates that, by eliminating the market mode, a new form of organisation appears in the system: certain groups of assets continue to show a coordinated behaviour among themselves. In other words, what we see is which assets continue to move together, but not because of the market. These assets are effective, i.e., relevant in this sense. Specifically, eliminating the market leaves a pattern that is not explained by the market but by the group itself. From an economic point of view, knowing who makes up this group can be of great interest.

To validate the worthiness of the proposed metric in identifying the truly effective components of the system, we compare the return of full portfolios with that of the ones composed only of the assets identified as effective. This analysis tests whether the assets identified by our methodology genuinely capture the return dynamics of

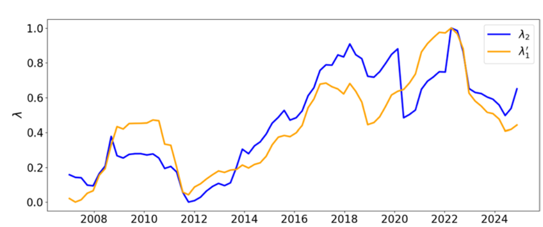
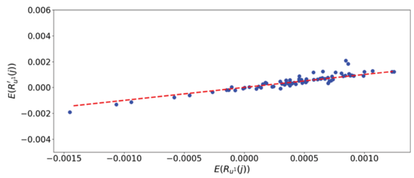
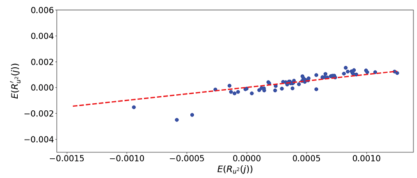


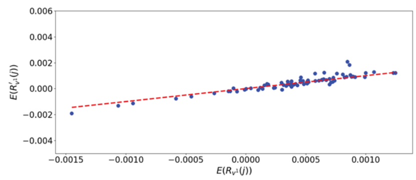

**Fig. 2** Evolution over time of the second-highest eigenvalue $\lambda_2$ of the correlation matrix of returns **C** and the largest eigenvalue $\lambda_1'$ of the residual correlation matrix **R**. Both variables were standardised to allow comparisons on a common scale and avoid scaling problems in the analysis




(a) $u^1$




(b) $u^2$

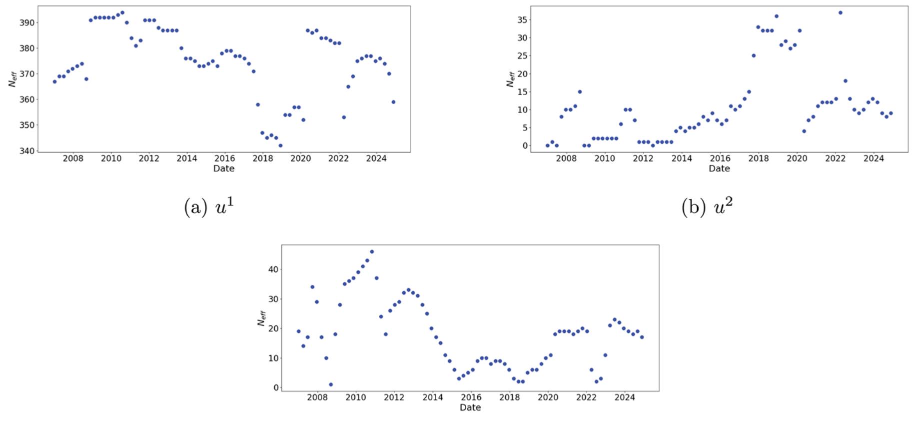



(c) $v^1$

**Fig. 3** Scatter plot showing the correlation between the return of the portfolio considering all assets, $E\left(R_{u^k}(j)\right)$, and the return of the portfolio considering only the effective ones, $E\left(R'_{u^k}(j)\right)$, for the eigenvectors associated with **a** the largest and **b** the second-largest eigenvalues of **C**, and **c** the largest eigenvalue of **R**. The red dashed line represents the linear regression fit, showing a positive relationship between the two return measures

the complete portfolio, thereby confirming the representational power of the selected components. Thus, Fig. 3 shows the correlations between the expected return of the portfolios containing all stocks ($E\left(R_{u^k}(j)\right)$) and the expected return of the portfolios containing only the stocks identified as effective ($E\left(R'_{u^k}(j)\right)$), for (a) the eigenvector associated with the largest eigenvalue of **C**, (b) the eigenvector associated with the second-largest eigenvalue of **C** and (c) the eigenvector associated with the largest eigenvalue of **R**. The red dashed line represents the linear regression fit. For the three cases, a positive linear relationship is observed, indicating that the stocks considered as effective are able to preserve the return dynamics of the full portfolio. The dispersion of the points around the regression line provides insights into the quality of the approximation. Lower dispersion suggests that the effective assets closely replicate the behaviour of the full portfolio, implying a strong representativeness of the selected direction. Conversely, higher dispersion indicates that some information is lost when restricting the portfolio to only the effective components. In our case, a statistically satisfactory fit is observed in the three cases in both cases of the visual appealing of the fit and the goodness-of-fit parameter, with an $R^2$ given by 0.79, 0.82, and 0.86 for panel (a), (b), and (c), respectively. Consequently, these results validate the efficacy of the proposed metric to identify the stocks determining the behaviour of the portfolios; e.g., what we have termed as effective components.

Once proved the validity of the $N_{eff}$ metric, we compute $N^k_{eff}(j)$ for $u^1$, $u^2$ and $v^1$. Figure 4 shows the evolution over time of the number of effective components for (a) the eigenvector associated with the largest eigenvalue of **C**, (b) the eigenvector associated with the second-largest eigenvalue of **C** and (c) the eigenvector associated with the largest eigenvalue of **R**. As can be seen, the number of effective components is never constant for none of the three eigenvectors. For $u^2$ and $v^1$, the $N_{eff}$ increases during periods




(a) $u^1$ (b) $u^2$

(c) $v^1$

**Fig. 4** Number of effective components $N^k_{eff}(j)$ as a function of time, for the eigenvectors associated with **a** the largest and **b** the second-largest eigenvalues of **C**, and **c** the largest eigenvalue of **R**

of crisis. To illustrate this, see panels (a) and (c) in 2008-2010 and panel (b) in 2020-2022.

From Fig. 4 it is striking the fact that after removing the market mode, it is observed a general increase in the number of effective components in the eigenvector associated with the second-largest eigenvalue of the correlation matrix (see Subfigures 4b and c). Only the period 2018-2020 exhibits a deviation from this relation. This result may suggest that the market factor was obscuring a latent, collective structure that becomes more widespread and interpretable once the market influence is projected out.

Based on the results previously obtained, and to investigate non-market related dependencies, we will focus our attention especially on $u^1$ and $v^1$. Since we aim to identify effective components that drive portfolio behaviour beyond market-wide dynamics, $\lambda_1$ and $\lambda'_1$, and, thus, $u^2$ and $v^1$ provide a natural direction to analyse, since they have proved to reflect a strong structure that, in the case of $v^2$, even remains after removing the market mode. This makes it especially valuable for understanding the counter or defensive behaviour during times of financial stress.

Importantly, the focus on the second eigenvector is not arbitrary. While previous literature (Plerou et al. 2002; Jiang and Zheng 2012) shows that this eigenvector captures meaningful sectoral structure beyond the market mode, recent evidence (Molero González et al. 2025) indicates that it is the only one that consistently exhibits countercyclical behavior across markets.

We have also explored higher-order eigenvectors (third and fourth) and find that they do not display a stable sectoral structure nor a consistent economic interpretation. Instead, they exhibit fragmented compositions with simultaneous long and short positions across sectors, without a persistent pattern over time (see Appendix B for more details). This lack of robustness further supports focusing on the second eigenvector.

As mentioned in Sect. Methodology, the stocks belonging to the S&P's 500 selected for this analysis have been classified by sector to analyse the sector distribution of the

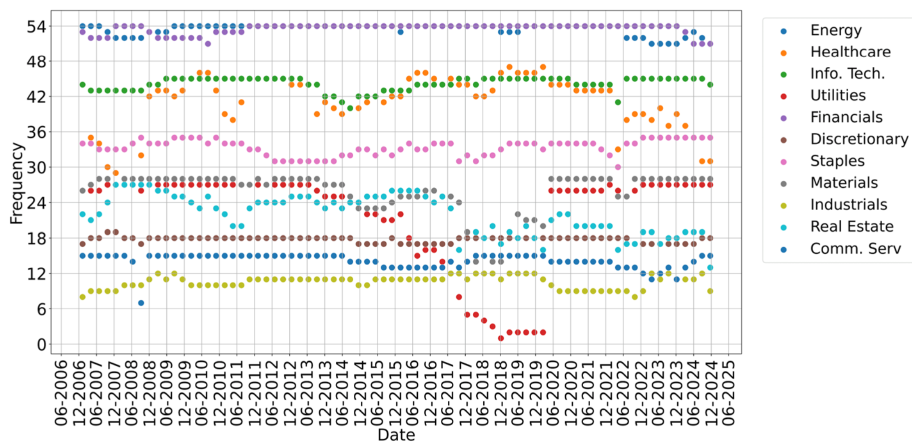



**Fig. 5** Evolution over time of the sector distribution of the effective stocks in $u^1$. As it is the market mode, no short positions exist within this eigenvector. The *y*-axis indicates the number of effective stocks per sector contributing to the eigenvector

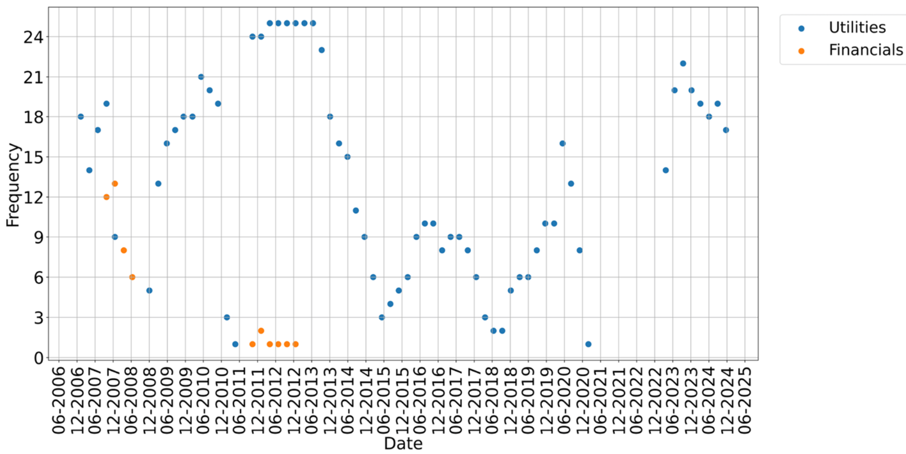
effective components of this portfolio, i.e., the eigenvector. As a further step, we have done this analysis by differentiating between long and short positions[1]

Figure 5 shows the evolution over time of the effective stocks in the portfolio associated to the highest eigenvalue. As can be seen, the number of effective stocks by economic sector remains highly stable over time. An exception to this general trend is found for the utilities sector, which from mid-2017 until the end of 2020 experienced a decline in the number of effective stocks within the market mode. From the end 2020 onwards, the trend recovers and returns to the frequencies seen before the aforementioned period.

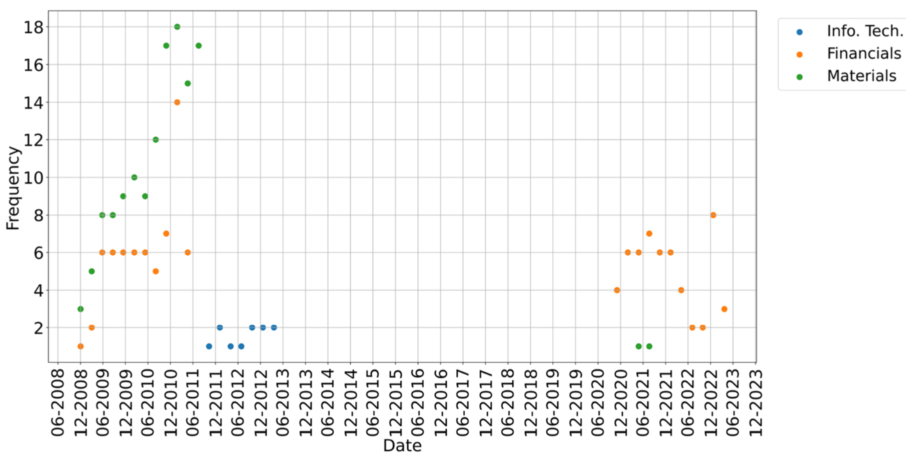
Figure 6 presents the evolution over time of the sector distribution of the stocks identified as effective in the portfolio associated to the second-highest eigenvalue of the correlation matrices of returns, differentiating between long positions and short ones. Results show that, in general, the stocks identified as effective in the long portfolio correspond to companies in the utilities sector, consistent with their traditionally defensive nature. However, this pattern changes during periods of financial stress. During both the 2007–2008 financial crisis and the Covid-19 pandemic, short positions also became effective, indicating that downside movements in specific sectors played a key role in shaping the system's dynamics. From June 2008 to June 2011, effective short positions were primarily concentrated in the materials sector, with a smaller presence in the financial sector. In contrast, between 2020 and 2023, during the Covid-19 crisis and its aftermath, effective short positions were almost entirely concentrated in the financial sector. These observations suggest that while utilities consistently act as safe havens on the long side, the sectoral composition of effective short positions varies across crises, possibly reflecting sector-specific vulnerabilities or differential investor expectations.

In financial markets, a safe-haven asset is one that retains or increases its value during adverse market conditions, enabling investors to protect their portfolios (Baur and Mcdermott 2010). Utilities are widely regarded as a defensive sector because the services

[1] The analysis procedure to select the effective components is performed, as mentioned in Section Methodology, by taking the stocks based on their weight in absolute value. However, for constructing $R'_{u^k}(j)$, we take the stocks with their original sign; e.g., as long or short positions, as applicable.




(a) Long positions




(b) Short positions

**Fig. 6** Evolution over time of the sector distribution of the effective stocks in $v^1$, differentiating between **a** stocks that originally correspond to long positions and **b** those that are short sales. The *y*-axis indicates the number of effective stocks per sector contributing to the eigenvector

they provide are essential and characterised by highly inelastic demand. Their stable, often regulated cash flows and low sensitivity to the business cycle lead utilities to preserve value and even outperform the market during downturns (Zulfiqar Ali et al. 2024; Tronzano 2025). These features explain why utilities frequently exhibit safe-haven-like behaviour during episodes of financial turmoil.

In line with this interpretation, the sectoral composition of the effective short positions is not stable across crises, but instead reflects the specific vulnerabilities associated with each episode. In particular, the prominence of the materials sector during the 2007–2011 period is in line with the reversal of the commodity boom and the contraction in global industrial and construction activity following the global financial crisis (Tang and Xiong 2012; Erten and Ocampo 2013). By contrast, during the Covid-19 episode, financials became dominant on the short side, which supports the heightened systemic sensitivity of the sector under conditions of severe uncertainty and pressure on financial intermediation. Overall, these results indicate that the proposed $N_{\text{eff}}$ metric captures not only statistical regularities in the eigenvector structure, but also economically interpretable shifts in the sectors through which financial stress is transmitted.

In Appendix A we report the robustness tests conducted. Specifically, we explore the sensitivity of the results to alternative specifications of the sliding-window parameters. First, we fix the step size at $\Delta = 60$ days and vary the window length using $n = 630, 756, 882$ days, corresponding to 36 months, 42 months, and 48 months of data. Second, we fix the window size at $n = 504$ days and assess robustness with respect to different step sizes, considering $\Delta = 5, 20, 40$ days, which represent one week, one month, and two months of trading days, respectively. These analyses confirm that a particular choice of window length or step size does not drive the main results.

As a final step in this paper, we have studied the stability in the direction of the eigenvector studied. To analyse how the direction of the eigenvector changes over time, we focus on the hyperspherical angles, which are scale-invariant. Thus, we transform the eigenvector to angular coordinates and measure how much those angles change over time to quantify their directional stability. In practical terms, what we do is to convert the vector in Cartesian coordinates of arbitrary dimension $\mathbb{R}^d$, with $d = p$ in our particular case, to hyperspherical coordinates.

Let $\mathbf{x} = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$ be a non-zero vector in Cartesian coordinates. The hyperspherical coordinates generalize spherical coordinates to $d$-dimensional Euclidean space and consist of a radius $r \in \mathbb{R}^+$ and a set of $d-1$ angles $(\theta_1, \theta_2, \ldots, \theta_{d-1})$, where:

- $\theta_1, \ldots, \theta_{d-2} \in [0, \pi]$ (elevation angles),
- $\theta_{d-1} \in [0, 2\pi)$ (azimuthal angle).

The radius is defined as:

$$r = ||\mathbf{x}|| = \sqrt{\sum_{i=1}^{d} x_i^2} \tag{5}$$

The angles are computed recursively as:

$$\theta_i = \begin{cases} arccos\left(\frac{x_i}{r}\right) \text{ for } i = 1, \ldots, d-2 \\ atan2(x_d, d_{d-1}) \text{ for } i = d-1 \end{cases} \tag{6}$$

where:

- *arccos* is applied to the ratio of the current component to the norm of the remaining components, ensuring $\theta_i \in [0, \pi]$ for $i < d-1$.
- *atan*2 function computes the azimuthal angle in the range $[0, 2\pi)$.

For practical use and visualization, angles $\theta_i$ are converted from radians to degrees via:

$$\theta_i^\circ = \theta_i \frac{180}{\pi}$$

This coordinate system is useful to represent directions in higher-dimensional spaces. In this way, we measure how the direction changes and, therefore, how stable the eigenvector is, by calculating the component-by-component angular difference between two angle vectors, and then we obtain the angular differences between consecutive windows. Finally, we use the Student's t-test to check whether the differences are statistically significant or not. Since the aim of the analysis is to identify the temporal windows where the angular changes are statistically significant, the problem can be framed as a multiple testing task focused on localisation rather than global hypothesis rejection. Accordingly, we control the False Discovery Rate (FDR) and adopt the Benjamini and Hochberg (1995) procedure, which remains valid under positive dependence.

In order to be able to illustrate the results as clearly as possible, we have created a scatter plot, where 1 represents the existence of statistically significant differences between the angles from one window to another, and 0 in the opposite case. For the case of $u^1$, e.g., the largest eigenvector of $\mathbf{C}$, it is expected that it should be stable, as it is the market factor. In fact, this is the result that we obtained and is what is shown in Fig. 7. As can be seen, no statistically significant differences are found from one window to the other for this eigenvector, with the exception of very small periods.

Figure 8 shows the same results, but for $v^1$. For this eigenvector, we find some periods for which the differences in the angles become statistically significant and, thus, the portfolio becomes unstable, that is, changes its direction. These periods are from 2009 to 2011, the beginning of 2018, and from 2022 to 2023. These periods where the differences become statistically significant indicate structural shifts in the portfolio composition. Notably, they partially overlap with those in which short positions become effective, particularly in the materials sector (2008–2011) and the financial sector (2020–2023). This suggests that the instability in $v^1$ captures moments when the system reorganises, possibly due to stress or revaluation of risk across specific sectors, which in turn are reflected in the effective components detected by our methodology.

## Conclusions

The Inverse Participation Ratio (IPR), originally developed in Physics to characterise the degree of localisation of wave functions in quantum systems (Clark and Del Maestro 2018), has increasingly proven useful for analysing structural and concentration

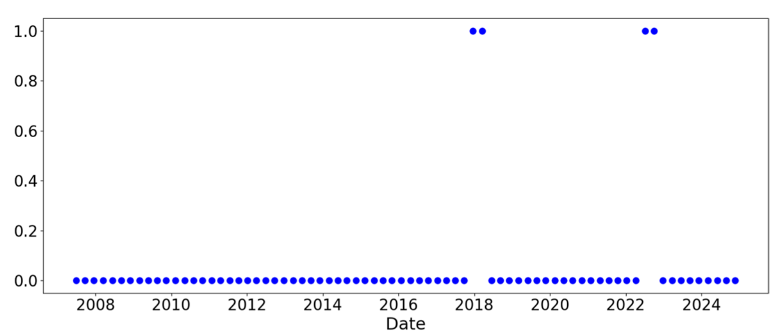



**Fig. 7** Evolution over time of the angular differences between consecutive windows for $u^1$. The *y*-axis indicates whether statistically significant differences exist between the angles across consecutive windows. A value of 1 denotes significance, while a value of 0 indicates no significant difference

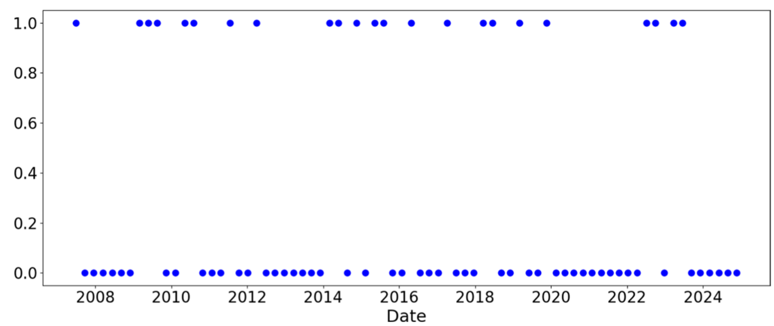



**Fig. 8** Evolution over time of the angular differences between consecutive windows for $v^1$. The *y*-axis indicates whether statistically significant differences exist between the angles across consecutive windows. A value of 1 denotes significance, while a value of 0 indicates no significant difference

phenomena in financial markets. Building on the observation by Molero González et al. (2025) that the second-largest eigenvalue displays a counter behaviour with respect to market spillovers, thus acting as a potential "safe haven" during periods of financial distress, this study aims to strenghten the integration of the IPR within the framework of portfolio theory to deepen the understanding of this eigenvalue, its associated eigenvector, and the components driving its dynamics. To this end, we formalise the $N_{eff}$ metric as a means of identifying the effective components within a portfolio of assets. Our empirical analysis, based on the daily logarithmic returns of S&P's 500 constituents from January 6th, 2005, to November 21st, 2024, provides new insights into the market structure underlying the second-largest eigenvalue. The main conclusions of the study are summarised below.

First, we focus on identifying the effective stocks and show that a reduced portfolio composed solely of these assets is able to preserve the return dynamics of the full portfolio. Moreover, the number of effective components associated with the considered eigenvectors is never constant over time, exhibiting a general increase during crises. Our analysis shows that, in general, only long positions in the utilities sector emerge as effective, a finding consistent with the sector's traditional defensive nature.

Second, downside movements in specific sectors play a key role in shaping the system's dynamics during high-volatility episodes. While utilities consistently act as long-term safe havens in the long term, the sectoral composition of effective short positions varies across crises, reflecting sector-specific vulnerabilities and shifts in investor expectations.

Third, the analysis of temporal stability in the direction of the eigenvectors studied reveals no statistically significant differences for the market factor. In contrast, periods of change in the eigenvector associated with the second-largest eigenvalue coincide with those in which short positions become effective within the portfolio.

Overall, the findings of this paper provide valuable insights into the internal structure of financial markets through the lens of the IPR, highlighting its potential as a universal tool to explore the financial markets' dynamics. The analysis performed shows that identifying effective components using the proposed $N_{eff}$ metric allows portfolio managers to focus only on the assets that significantly influence the returns dynamics of the portfolios. In practical terms, this can allow a more efficient portfolio construction by reducing the number of assets required without compromising exposure to systemic risk, and improving transaction costs efficiency.

The persistent defensive behavior of the utilities sector, which acts as a safe haven, highlights its potential role in risk mitigation strategies during periods of high volatility. By monitoring the temporal stability of the assets identified as effective and the evolution of the eigenvalues and their associated eigenvectors, investors can anticipate shifts in portfolio risk structure, which allows them to adjust their positions dynamically in response to the changing market conditions.

Our empirical analysis is based exclusively on data from companies listed on the S&P's 500 market. Future research may reveal different dynamics in emerging markets or other developed economies with distinct structural characteristics. Furthermore, the $N_{eff}$ metric could be integrated in a portfolio optimization framework, allowing the testing of strategies that overweight the components identified as effective outperform traditional diversification approaches in terms of risk-adjusted returns.

This study is not without limitations, which naturally point to several promising avenues for future research. First, the analysis focuses exclusively on the S&P's 500 constituents, thereby reflecting the behavior of a large and mature equity market. Extending the methodology to other asset classes or international markets would help assess the broader applicability of the proposed $N_{eff}$ metric. Second, although the present paper highlights the potential of the IPR-based methodology to identify effective assets, we do not formally evaluate how this information can be incorporated into real-world portfolio construction or risk management. Future research should examine whether selecting assets according to their effectiveness leads to improved out-of-sample performance, lower portfolio volatility, or enhanced risk-management outcomes, and assess the temporal stability of these assets to determine their practical utility for investors and policymakers. Finally, while the present study focuses on the second-largest eigenvector due to its unique counter-cyclical and sectoral properties, the analysis of higher-order eigenvectors remains an open and promising line of research. In particular, future work could explore whether the $N_{eff}$ metric is able to detect transient or sub-sectoral structures in these eigenvectors, even if they lack the stability and cross-market consistency observed for the second eigenvector.

## Appendix A

### Sliding window robustness check

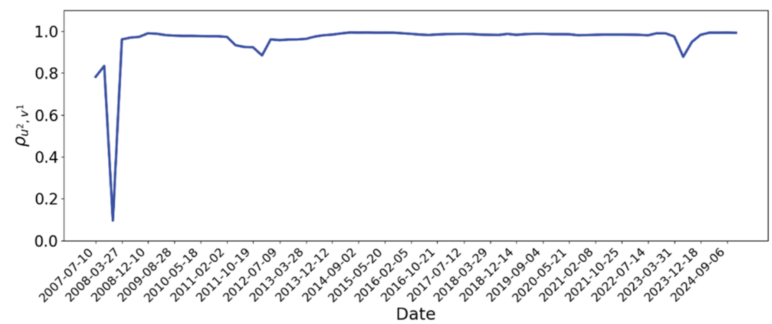
This section is devoted to presenting the sensitivity analysis performed to test the robustness of the empirical results to alternative parameter choices. Concretely, we have considered the following cases:

I) First, we establish a fixed value of $\Delta = 60$ days, and we change the window length, considering $n = 630, 756, 882$ days; this is three years, three years and a half, and four years, respectively.

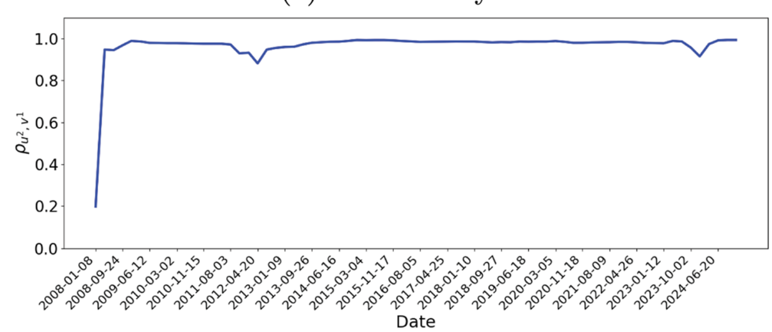
II) Second, we select as fixed parameter the window length, taking $n = 504$ days, and we check the robustness of the results for the step sizes $\Delta = 5, 20, 40$ days, which correspond to one week, one month, and two months of trading days, respectively.

#### Considering a fixed step size value of $\Delta = 60$ days

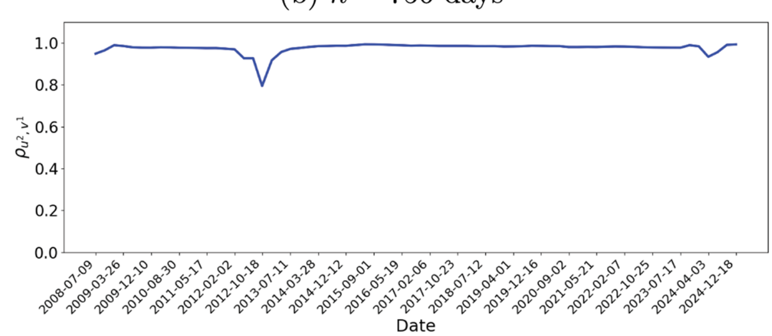
See Figs. 9, 10, 11, 12, 13, 14, 15, 16.




(a) $n = 630$ days

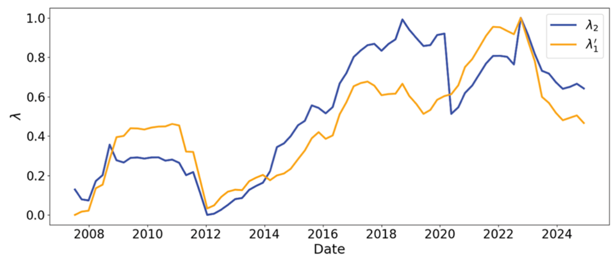



(b) $n = 756$ days

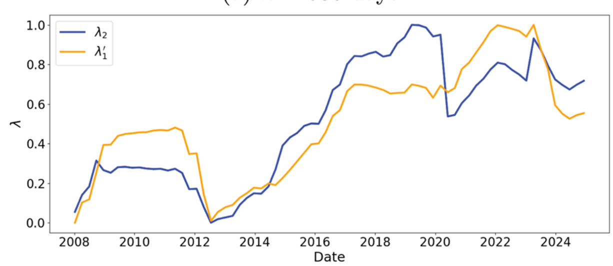



(c) $n = 882$ days

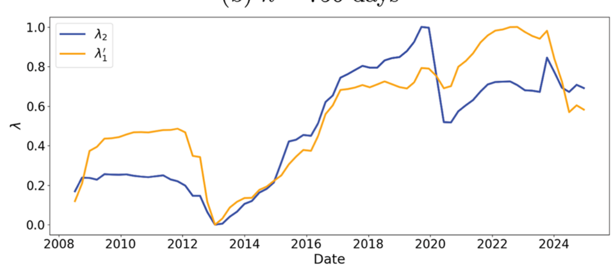
**Fig. 9** Pearson correlation coefficient between $u^2(j)$ and $v^1(j)$ as a function of time for the different window lengths **a** $n = 630$ days, **b** $n = 756$ days, and **c** $n = 882$ days




(a) $n = 630$ days

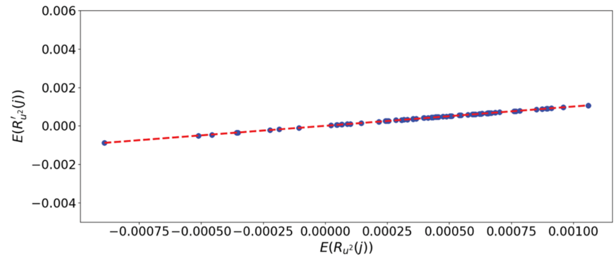



(b) $n = 756$ days

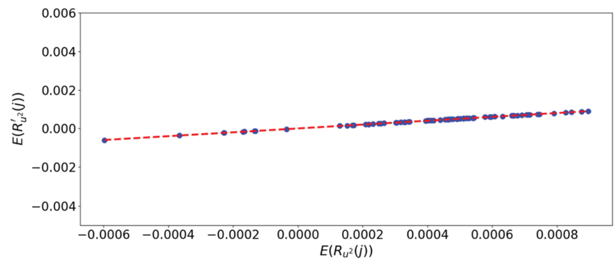



(c) $n = 882$ days

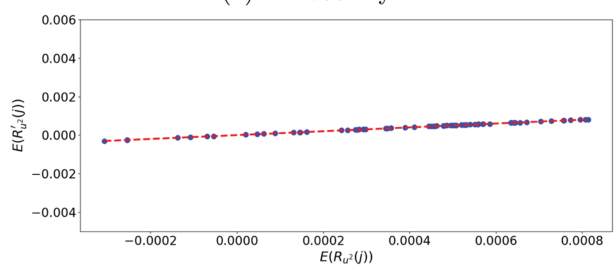
**Fig. 10** Evolution over time of the second-highest eigenvalue $\lambda_2$ of the correlation matrix of returns **C** and the largest eigenvalue $\lambda_1'$ of the residual correlation matrix **R**, for the different window lengths **a** $n = 630$ days, **b** $n = 756$ days, and **c** $n = 882$ days. Both variables were standardised to allow comparisons on a common scale and avoid scaling problems




(a) $n = 630$ days

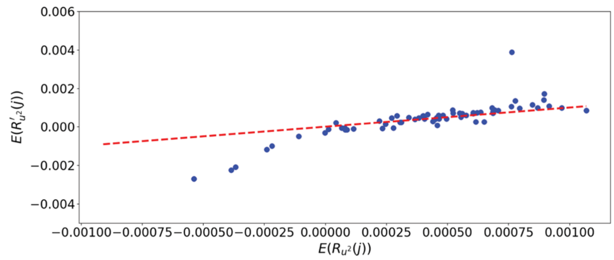



(b) $n = 756$ days

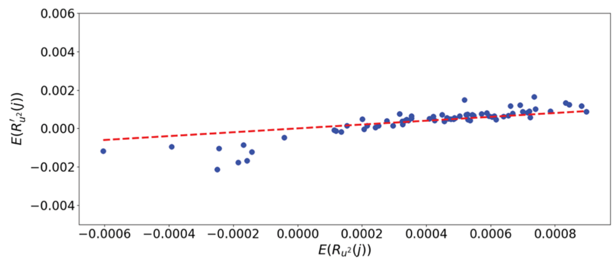



(c) $n = 882$ days

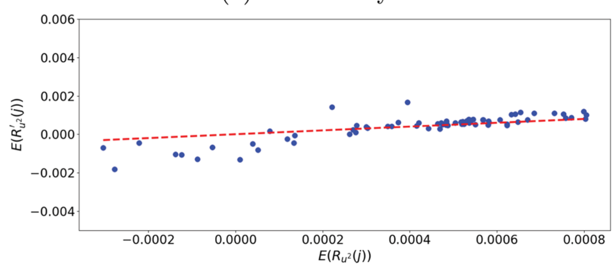
**Fig. 11** Scatter plot showing the correlation between the return of the portfolio considering all assets, $E\left(R_{u^1}(j)\right)$, and the return of the portfolio considering only the effective ones $E\left(R'_{u^1}(j)\right)$ for the eigenvector associated with the largest eigenvalue of the correlation matrix **C**, and for the different window lengths **a** $n = 630$ days, **b** $n = 756$ days, and **c** $n = 882$ days. The red dashed line represents the linear regression fit, showing a positive relationship between the two return measures




(a) $n = 630$ days

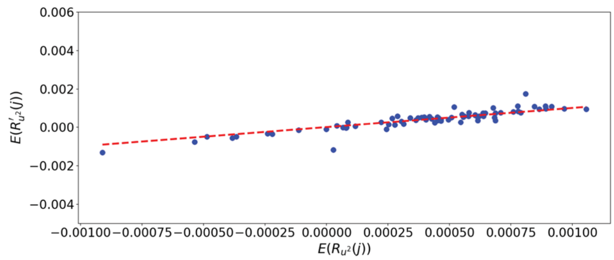



(b) $n = 756$ days

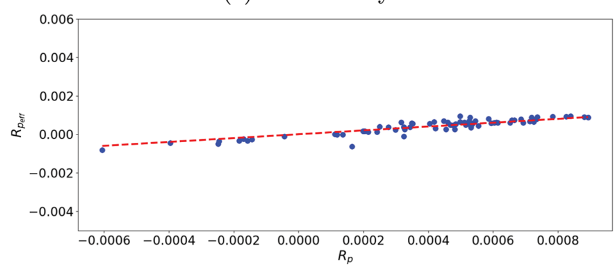



(c) $n = 882$ days

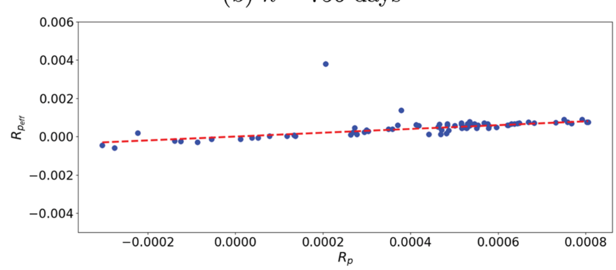
**Fig. 12** Scatter plot showing the correlation between the return of the portfolio considering all assets, $E\left(R_{u^2}(j)\right)$, and the return of the portfolio considering only the effective ones $E\left(R'_{u^2}(j)\right)$ for the eigenvector associated with the second-largest eigenavalue of the correlation matrix **C**, and for the different window lengths **a** $n = 630$ days, **b** $n = 756$ days, and **c** $n = 882$ days. The red dashed line represents the linear regression fit, showing a positive relationship between the two return measures




(a) $n = 630$ days

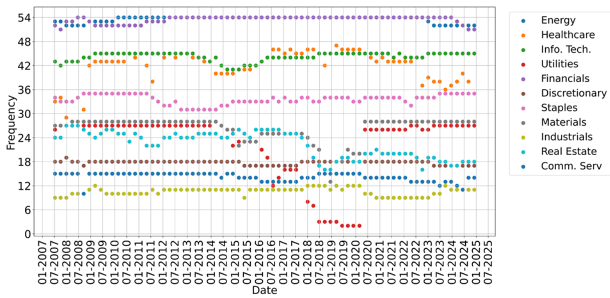



(b) $n = 756$ days

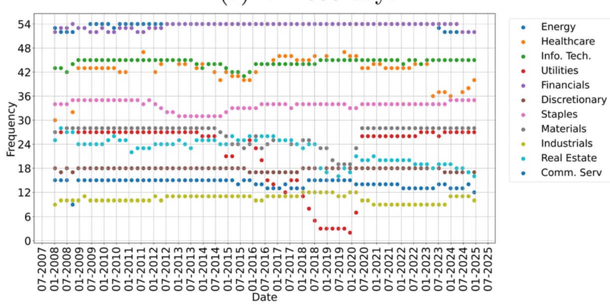



(c) $n = 882$ days

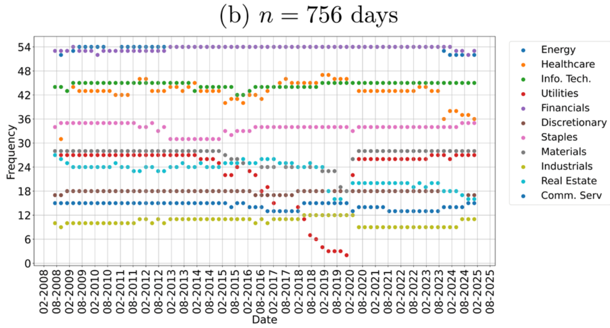
**Fig. 13** Scatter plot showing the correlation between the return of the portfolio considering all assets, $E\left(R_{v^1}(j)\right)$, and the return of the portfolio considering only the effective ones $E\left(R'_{v^1}(j)\right)$ for the eigenvector associated with the largest eigenvalue of the residual correlation matrix **R**, and for the different window lengths **a** $n = 630$ days, **b** $n = 756$ days, and **c** $n = 882$ days. The red dashed line represents the linear regression fit, showing a positive relationship between the two return measures




(a) $n = 630$ days

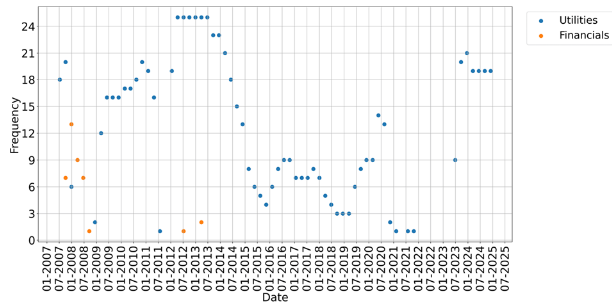



(b) $n = 756$ days

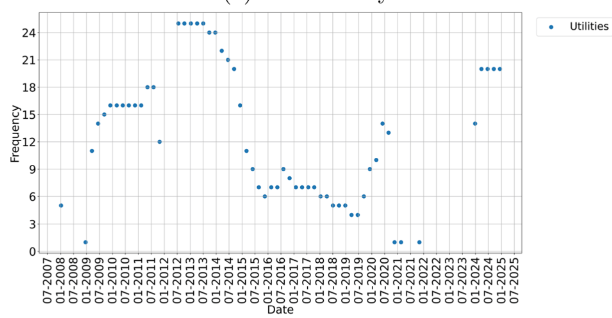



(c) $n = 882$ days

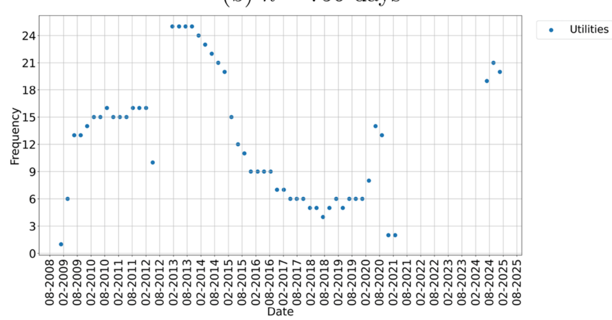
**Fig. 14** Evolution over time of the sector distribution of the effective stocks in $u^1$ for the different window lengths **a** $n = 630$ days, **b** $n = 756$ days, and **c** $n = 882$ days. As it is the market factor, no short positions exist within this eigenvector




(a) $n = 630$ days

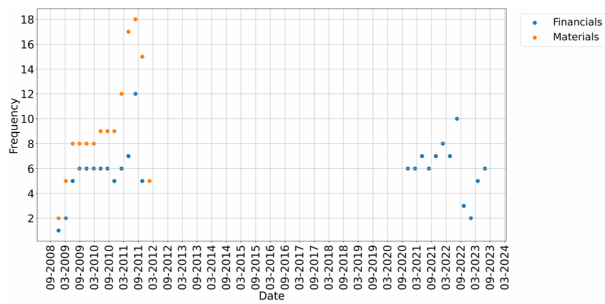



(b) $n = 756$ days

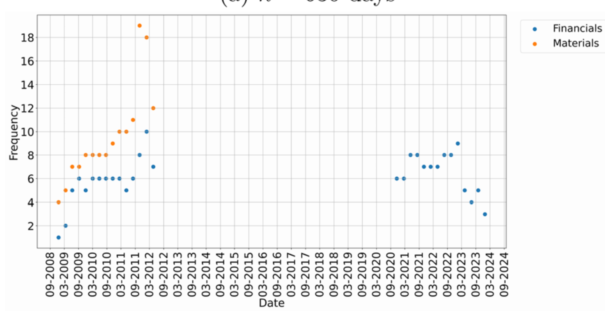



(c) $n = 882$ days

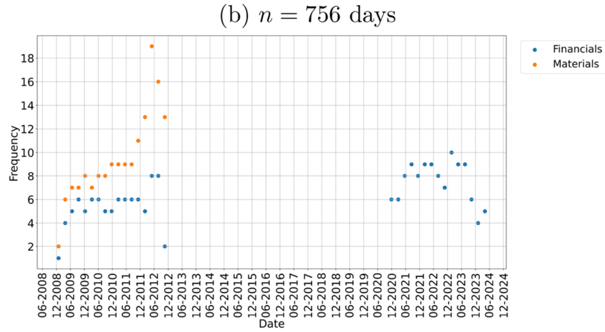
**Fig. 15** Evolution over time of the sector distribution of the effective stocks that originally correspond to long positions in $v^1$ for the different window lengths **a** $n = 630$ days, **b** $n = 756$ days, and **c** $n = 882$ days




(a) $n = 630$ days




(b) $n = 756$ days

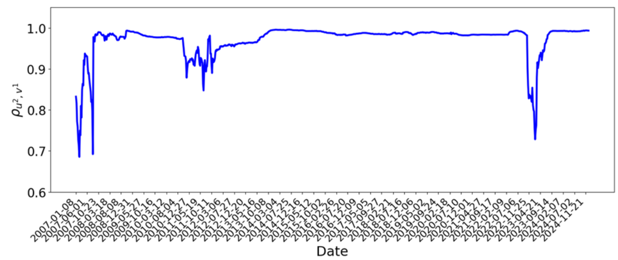



(c) $n = 882$ days

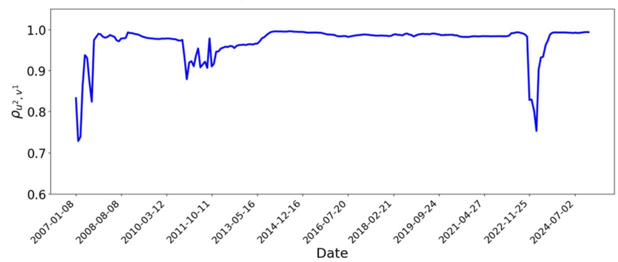
**Fig. 16** Evolution over time of the sector distribution of the effective stocks that originally correspond to short positions in $v^1$ for the different window lengths **a** $n = 630$ days, **b** $n = 756$ days, and **c** $n = 882$ days

### Considering a fixed window length $n = 504$ days

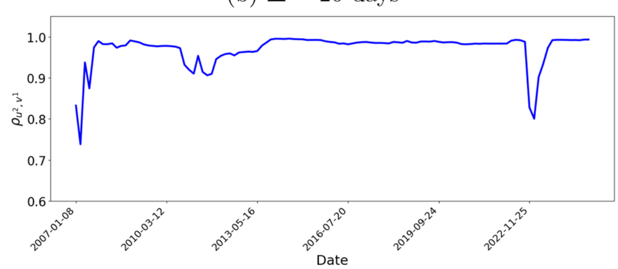
See Figs. 17, 18, 19, 20, 21, 22, 23, 24




(a) $\Delta = 5$ days

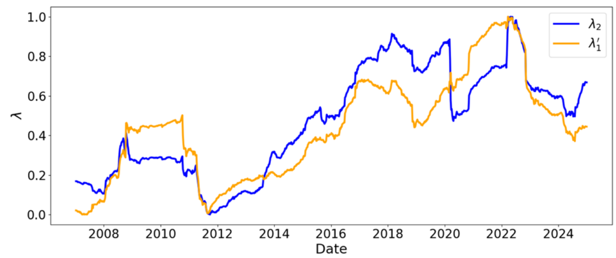



(b) $\Delta = 20$ days

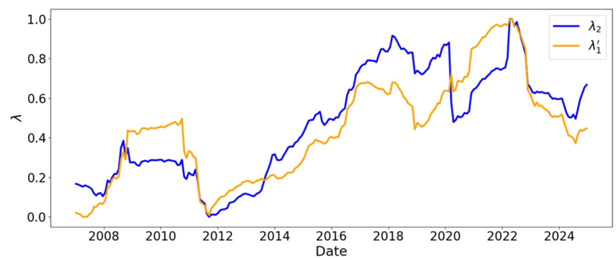



(c) $\Delta = 40$ days

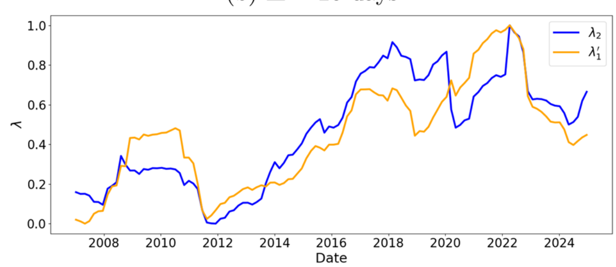
**Fig. 17** Pearson correlation coefficient between $u^2(j)$ and $v^1(j)$ as a function of time for the different step sizes **a** $\Delta = 5$ days, **b** $\Delta = 20$ days, and **c** $\Delta = 40$ days.




(a) $\Delta = 5$ days

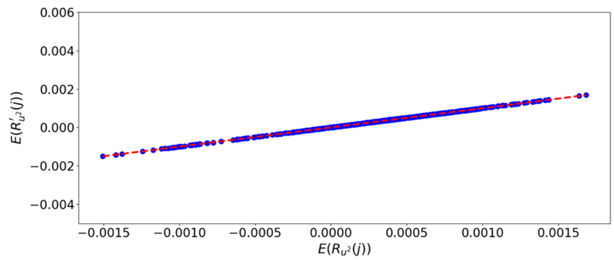



(b) $\Delta = 20$ days

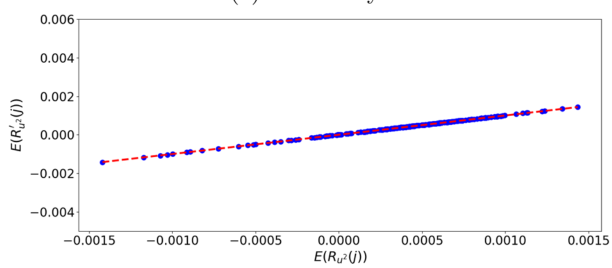



(c) $\Delta = 40$ days

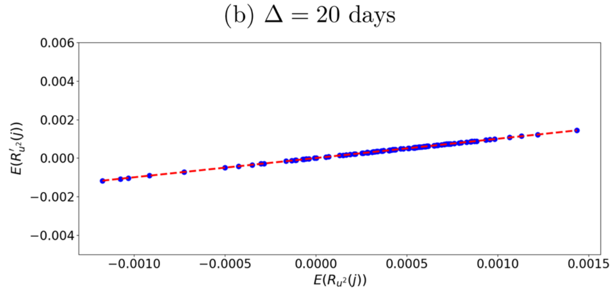
**Fig. 18** Evolution over time of the second-highest eigenvalue $\lambda_2$ of the correlation matrix of returns **C** and the largest eigenvalue $\lambda_1'$ of the residual correlation matrix **R**, for the different step sizes **a** $\Delta = 5$ days, **b** $\Delta = 20$ days, and **c** $\Delta = 40$ days. Both variables were standardised to allow comparisons on a common scale and avoid scaling problems




(a) $\Delta = 5$ days

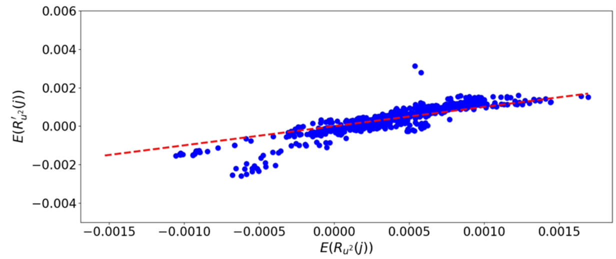



(b) $\Delta = 20$ days

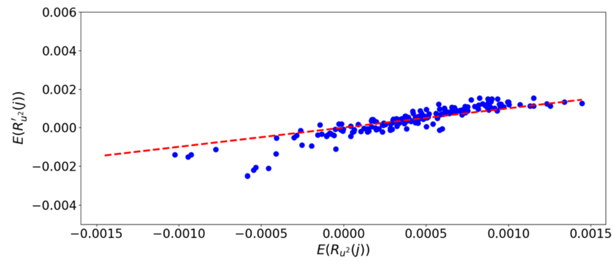



(c) $\Delta = 40$ days

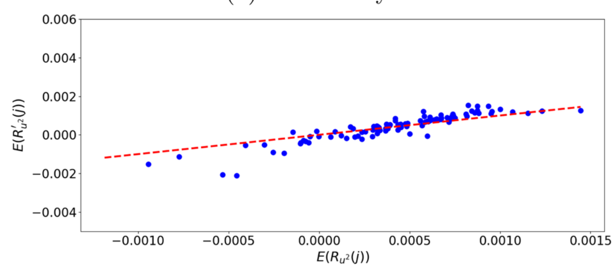
**Fig. 19** Scatter plot showing the correlation between the return of the portfolio considering all assets, $E\left(R_{u^1}(j)\right)$, and the return of the portfolio considering only the effective ones $E\left(R'_{u^1}(j)\right)$ for the eigenvector associated with the largest eigenvalue of the correlation matrix **C**, and for the different step sizes **a** $\Delta = 5$ days, **b** $\Delta = 20$ days, and **c** $\Delta = 40$ days. The red dashed line represents the linear regression fit, showing a positive relationship between the two return measures




(a) $\Delta = 5$ days

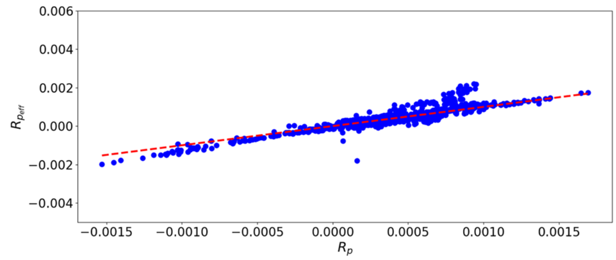



(b) $\Delta = 20$ days

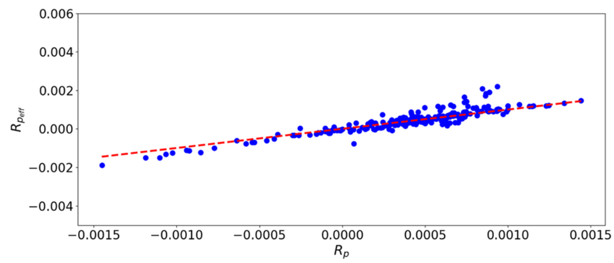



(c) $\Delta = 40$ days

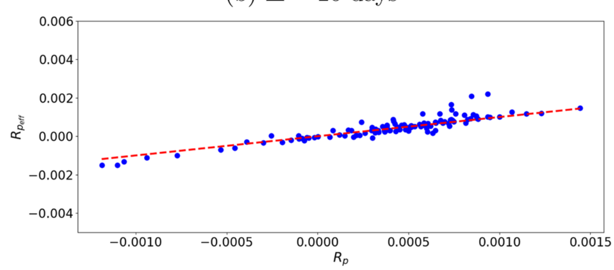
**Fig. 20** Scatter plot showing the correlation between the return of the portfolio considering all assets, $E\left(R_{u^2}(j)\right)$, and the return of the portfolio considering only the effective ones $E\left(R'_{u^2}(j)\right)$ for the eigenvector associated with the second-largest eigenvalue of the correlation matrix **C**, and for the different step sizes **a** $\Delta = 5$ days, **b** $\Delta = 20$ days, and **c** $\Delta = 40$ days. The red dashed line represents the linear regression fit, showing a positive relationship between the two return measures




(a) $\Delta = 5$ days

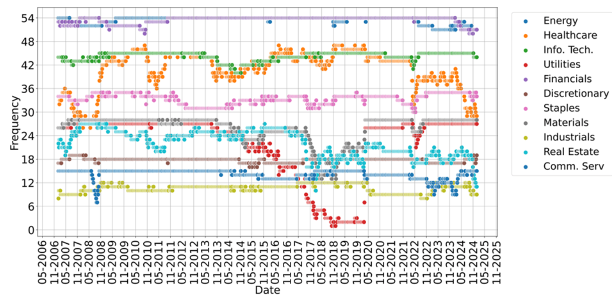



(b) $\Delta = 20$ days

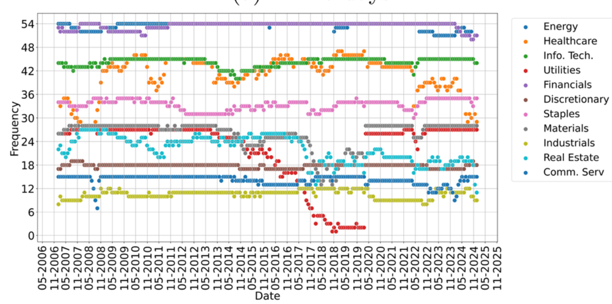



(c) $\Delta = 40$ days

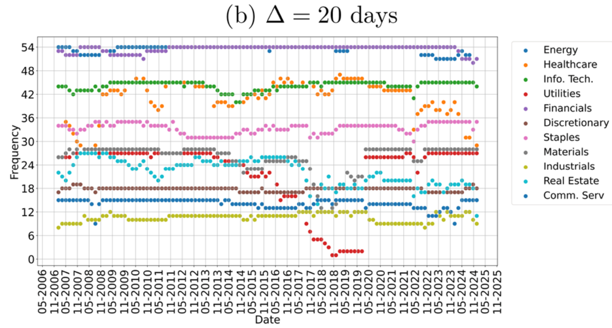
**Fig. 21** Scatter plot showing the correlation between the return of the portfolio considering all assets, $E\left(R_{v^1}(j)\right)$, and the return of the portfolio considering only the effective ones $E\left(R'_{v^1}(j)\right)$ for the eigenvector associated with the largest eigenvalue of the residual correlation matrix **R**, and for the different step sizes **a** $\Delta = 5$ days, **b** $\Delta = 20$ days, and **c** $\Delta = 40$ days. The red dashed line represents the linear regression fit, showing a positive relationship between the two return measures




(a) $\Delta = 5$ days

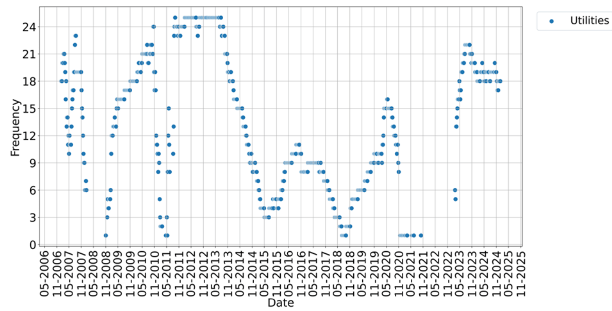



(b) $\Delta = 20$ days

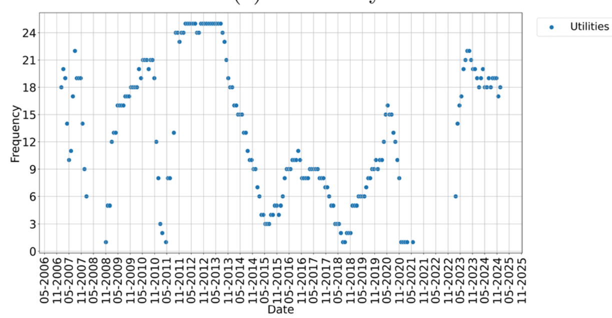



(c) $\Delta = 40$ days

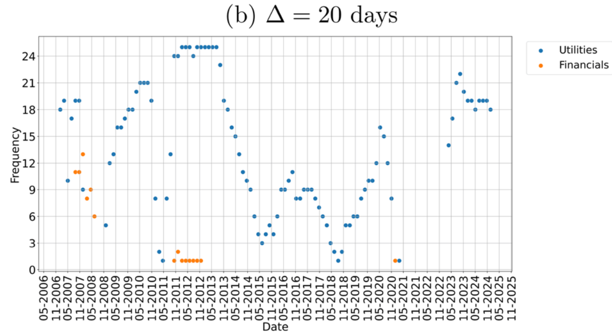
**Fig. 22** Evolution over time of the sector distribution of the effective stocks in $u^1$ for the different step sizes **a** $\Delta = 5$ days, **b** $\Delta = 20$ days, and **c** $\Delta = 40$ days. As it is the market factor, no short positions exist within this eigenvector




(a) $\Delta = 5$ days

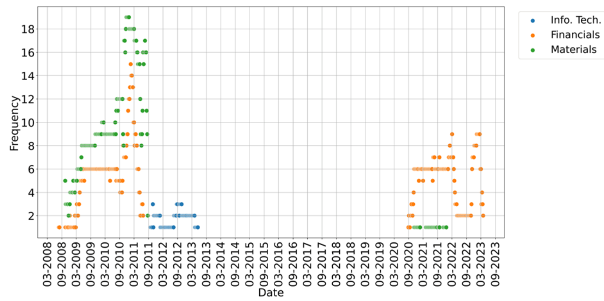



(b) $\Delta = 20$ days

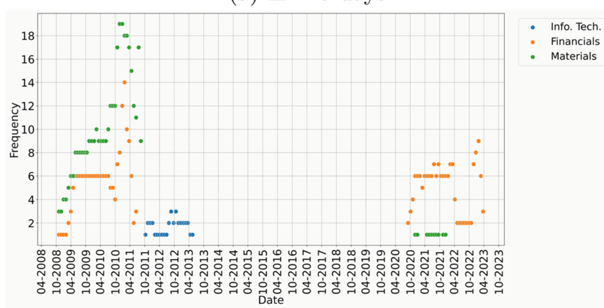



(c) $\Delta = 40$ days

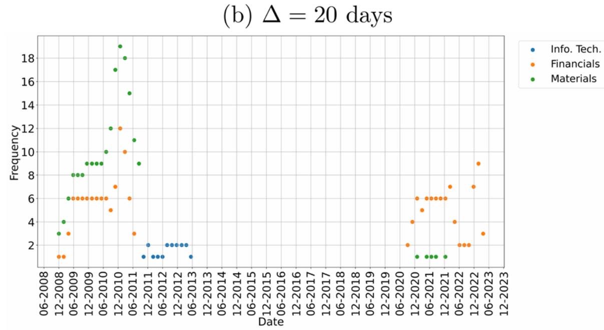
**Fig. 23** Evolution over time of the sector distribution of the effective stocks that originally correspond to long positions in $v^1$ for the different step sizes **a** $\Delta = 5$ days, **b** $\Delta = 20$ days, and **c** $\Delta = 40$ days




(a) $\Delta = 5$ days




(b) $\Delta = 20$ days



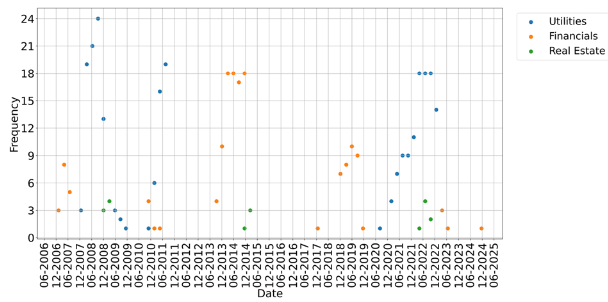

(c) $\Delta = 40$ days

**Fig. 24** Evolution over time of the sector distribution of the effective stocks that originally correspond to short positions in $v^1$ for the different step sizes **a** $\Delta = 5$ days, **b** $\Delta = 20$ days, and **c** $\Delta = 40$ days.

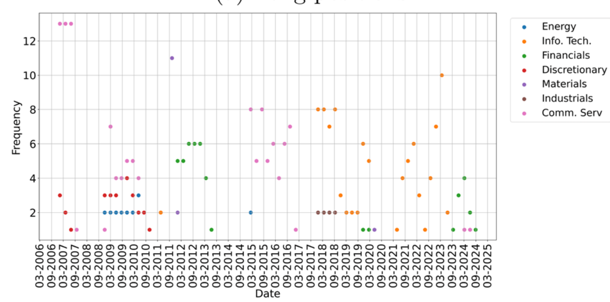
## Sector distribution of the effective stocks in the portfolios associated with the third and the fourth largest eigenvalues

Figures 25 and 26 show the evolution over time of the sector distribution of the effective stocks in the portfolios associated with the third and the fourth largest eigenvalues, respectively. As can be seen, and unlike the eigenvector associated with the second-largest eigenvalue, these higher-order eigenvectors do not exhibit a clear and persistent sectoral dominance. Instead, they display simultaneously significant long and short positions throughout the sample, with a fragmented and unstable sectoral composition over time.

Therefore, there is no clear sectoral dominance comparable to that observed for the portfolio associated with the second-largest eigenvalue. Instead, the weights are distributed across sectors, and both long and short positions remain almost simultaneously relevant throughout the sample. This contrasts with the eigenvector associated with the second-largest eigenvalue, which exhibits a much more stable and interpretable structure, typically dominated by defensive sectors on the long side.

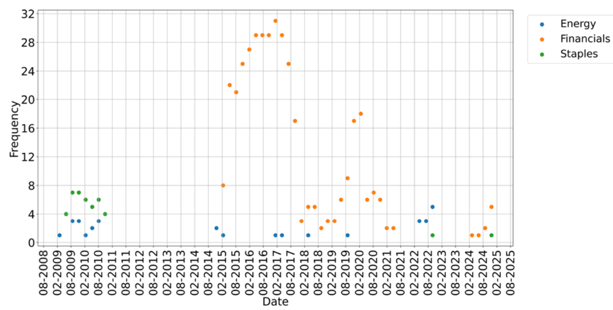



(a) Long positions

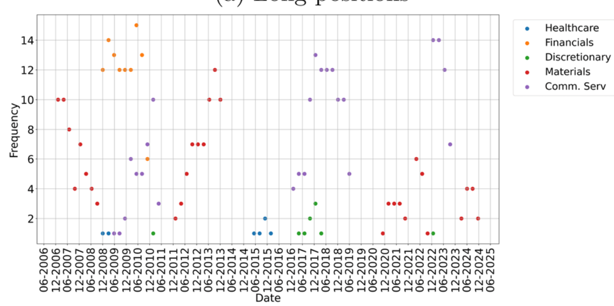



(b) Short positions

**Fig. 25** Evolution over time of the sector distribution of the effective stocks in the portfolio associated with the third-largest eigenvalue, differentiating between **a** stocks that originally correspond to long positions and **b** those that are short sales. The *y*-axis indicates the number of effective stocks per sector contributing to the eigenvector




(a) Long positions




(b) Short positions

**Fig. 26** Evolution over time of the sector distribution of the effective stocks in the portfolio associated with the fourth-largest eigenvalue, differentiating between **a** stocks that originally correspond to long positions and **b** those that are short sales. The *y*-axis indicates the number of effective stocks per sector contributing to the eigenvector

**Acknowledgements**
Not applicable.

**Author contributions**
All authors equally contributed to this study.

**Funding**
Laura Molero González is funded by the Spanish Ministerio de Ciencia, Innovación y Universidades (FPU2023) and the Spanish Ministerio de Ciencia, Innovaci´on y Universidades, Spain, through project No. PID2021-127836NBI00 (funded by MCIN/AEI/10.13039/501100011033/FEDER "A way to make Europe"). Juan E. Trinidad Segovia and Miguel A. Sánchez Granero are funded by the Spanish Ministerio de Ciencia, Innovaci´on y Universidades, Spain, through project No. PID2021-127836NBI00 (funded by MCIN/AEI/10.13039/501100011033/FEDER "A way to make Europe").

**Data availability**
The datasets used in this study are publicly available at the Yahoo Finance repository.

## Declarations

**Consent for publication**
All authors agree to the publication of the present manuscript.

**Competing interests**
The authors declare that they have no Competing interests.

Received: 30 September 2025 Revised: 19 June 2026 Accepted: 21 June 2026
Published online: 08 September 2026

**References**

Baruník J, Cech F (2021) Measurement of common risks in tails: a panel quantile regression model for financial returns. J Financ Mark 52:100562

Baur DG, McDermott TK (2010) Is gold a safe haven? international evidence. Journal of Banking & Finance, 34(8):1886–1898. New Contributions to Retail Payments: Conference at Norges Bank (Central Bank of Norway) 14–15 November 2008
Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Roy Stat Soc: Ser B (Methodol) 57(1):289–300
Biely C, Thurner S (2008) Random matrix ensembles of time-lagged correlation matrices: derivation of eigenvalue spectra and analysis of financial time-series. Quantitative Financ 8(7):705–722
Bouchaud J-P, Potters M (2003) Theory of financial risk and derivative pricing: from statistical physics to risk management. Cambridge University Press
Bun J, Bouchaud J-P, Potters M (2017) Cleaning large correlation matrices: tools from random matrix theory. Phys Rep 666:1–109
Calixto M, Romera E (2015) Inverse participation ratio and localization in topological insulator phase transitions. J Stat Mech: Theory Exp 2015(6):P06029
Clark TB, Del Maestro A (2018) Moments of the inverse participation ratio for the Laplacian on finite regular graphs. J Phys A: Math Theor 51(49):495003
Coronnello C, Tumminello M, Lillo F, Miccichè S, Mantegna RN (2005) Sector identification in a set of stock return time series traded at the London stock exchange. Acta Phys Pol, B 36(9):2653–2669
De Nard G, Engle RF, Ledoit O, Wolf M (2022) Large dynamic covariance matrices: enhancements based on intraday data. J Bank Financ 138:106426
Dom MS, Howard C, Jansen M, Lohre H (2025) Beyond gmv: the relevance of covariance matrix estimation for risk-based portfolio construction. Quantitative Financ 25(3):403–419
Domínguez-Monterroza A, Jiménez-Martín A, Mateos-Caballero A (2025) Correlation structure of the Spanish stock market around COVID-19 using random matrix theory. Comput Econ. https://doi.org/10.1007/s10614-024-10820-0
Drożdż S, Grümmer F, Górski AZ, Ruf F, Speth J (2000) Dynamics of competition between collectivity and noise in the stock market. Physica A 287(3–4):440–449
Erten B, Ocampo JA (2013) Super cycles of commodity prices since the mid-nineteenth century. World Dev 44:14–30
Evers F, Mirlin AD (2008) Anderson transitions. Rev Mod Phys 80(4):1355–1417
Faias JA (2023) Predicting the equity risk premium using the smooth cross-sectional tail risk: the importance of correlation. J Financ Mark 63:100769
García-Medina A (2016) Global financial indices and twitter sentiment: a random matrix theory approach. Physica A 461:509–522
García-Medina A (2025) High-dimensional covariance matrix estimators on simulated portfolios with complex structures. Phys Rev E 111:024316
García-Medina A, González Farías G (2020) Transfer entropy as a variable selection methodology of cryptocurrencies in the framework of a high dimensional predictive model. PLoS ONE 15(1):e0227269
García-Medina A, Miccichè S, Mantegna RN (2023) Two-step estimators of high-dimensional correlation matrices. Phys Rev E 108:044137
García-Medina A, Sandoval L, Urrutia Bañuelos E, Martínez-Argüello A (2018) Correlations and flow of information between the New York times and stock markets. Physica A 502:403–415
Jiang XF, Zheng B (2012) Anti-correlation and subsector structure in financial systems. Europhys Lett 97(4):48006
Kramer B, MacKinnon A (1993) Localization: theory and experiment. Rep Prog Phys 56(12):1469
Kwapień J, Drożdż S (2012) Physical approach to complex systems. Phys Rep 515(3–4):115–226
Laloux L, Cizeau P, Bouchaud J-P, Potters M (1999) Noise dressing of financial correlation matrices. Phys Rev Lett 83(7):1467
Liu Y, Sierant P, Stornati P, Lewenstein M, Płodzień M (2025) Quantum algorithms for inverse participation ratio estimation in multiqubit and multiqudit systems. Phys Rev A 111(5):052614
Molero González L, Cerqueti R, Mattera R, Trinidad Segovia JE (2025) The random matrix-based informative content of correlation matrices in stock markets. Chaos: An Interdiscip J of Nonlinear Sci 35(9):093111
Molero González L, Trinidad Segovia JE, Sánchez Granero MA, García Medina A (2023) Market beta is not dead: an approach from random matrix theory. Financ Res Lett 55:103816
Moura GV, Santos AA, Ruiz E (2020) Comparing high-dimensional conditional covariance matrices: implications for portfolio selection. J Bank Financ 118:105882
Murphy N, Wortis R, Atkinson W (2011) Generalized inverse participation ratio as a possible measure of localization for interacting systems. Phys Rev B-Condens Matter Mater Phys 83(18):184206
Pafka S, Kondor I (2004) Estimated correlation matrices and portfolio optimization. Physica A 343:623–634
Paulusch J, Schlütter S (2022) Sensitivity-implied tail-correlation matrices. J Bank Financ 134:106333
Perera J, Wortis R (2015) Energy dependence of localization with interactions and disorder: the generalized inverse participation ratio of an ensemble of two-site Anderson-Hubbard systems. Phys Rev B 92:085110
Plerou V, Gopikrishnan P, Rosenow B, Amaral LAN, Guhr T, Stanley HE (2002) Random matrix approach to cross correlations in financial data. Phys Rev E 65(6):066126
Rodriguez A, Vasquez LJ, Römer RA (2009) Multifractal analysis with the probability density function at the three-dimensional anderson transition. Phys Rev Lett 102(10):106406
Rodriguez A, Vasquez LJ, Slevin K, Römer RA (2010) Critical parameters from a generalized multifractal analysis at the Anderson transition. Phys Rev Lett 105:046403
Sharpe WF (1964) Capital asset prices: a theory of market equilibrium under conditions of risk. J Financ 19(3):425–442
Singh A, Xu D (2016) Random matrix application to correlations amongst the volatility of assets. Quantitative Finance 16(1):69–83
Song D-M, Tumminello M, Zhou W-X, Mantegna RN (2011) Evolution of worldwide stock markets, correlation structure, and correlation-based graphs. Phys Rev E 84:026108

Stolfi P, Bernardi M, Vergni D (2022) Robust estimation of time-dependent precision matrix with application to the cryptocurrency market. Financial Innovation 8(1):47
Tang K, Xiong W (2012) Index investment and the financialization of commodities. Financ Anal J 68(6):54–74
Tang Y, Xiong J, Cheng Z, Zhuang Y, Li K, Xie J, Zhang Y (2023) Looking into the market behaviors through the lens of correlations and eigenvalues: an investigation on the Chinese and US markets using RMT. Entropy 10:25
Thouless DJ (1974) Electrons in disordered systems and the theory of localization. Phys Rep 13(3):93–142
Tronzano M (2025) Assessing the effectiveness of some defensive assets in global stock portfolios: evidence from daily data (2021–2024). J Risk Financ Manag 18(12)
Utsugi A, Ino K, Oshikawa M (2004) Random matrix theory analysis of cross correlations in financial markets. Phys Rev E-Stat Nonlinear Soft Matter Phys 70(2):026110
Zhao X, Geng X, Huang Y, Wu Y, Zhang N (2025) The resonance effect of economic policy uncertainty worldwide: a time–frequency analysis. North American J Econ Financ 78:102437
Zulfiqar Ali I, Muhammad A, Khurram S, Mobeen A, Imran H (2024) Safe haven properties of industrial stocks against ESG in the United States: portfolio implication for sustainable investments. Energy Econ 136:107712

## Publisher's Note