CentraleSupélec

université PARIS-SACLAY

Non-Gaussian Factor Decomposition

Student

Tomas España

CentraleSupélec, Université Paris-Saclay

Supervisor

Rudy Morel

Research Fellow, Flatiron Institute

June 14, 2025

Contents

Abstract

The price movements of a financial index, such as the S&P500, form a high-dimensional multivariate time series involving hundreds of stocks. Despite this complexity, one may seek to identify a small set of underlying linear *factors* that drive these movements. A *factorization* refers to a method for extracting such factors. The most widely used factorization is Principal Component Analysis (PCA), which identifies directions in the data along which price movements exhibit the highest variance, and uses them to define the factors. However, PCA is limited by its exclusive focus on second-order statistics, which, for example, makes it highly sensitive to heavy-tailed distributions. In this work, we propose an alternative decomposition method based on non-Gaussianity of factors rather than on their variance: the more non-Gaussian a factor is, the more it is assumed to drive the market. More precisely, we iteratively extract the directions whose removal makes the remaining time series (i.e. the residuals) as close to Gaussian noise as possible. We demonstrate the effectiveness of our method on synthetic and financial datasets, showing that, at the cost of increased computation, it uncovers interpretable non-Gaussian factors overlooked by PCA, offering a novel approach to multivariate time series factorization.

Introduction

In many applications, one is faced with the challenge of understanding and simplifying the behavior of high-dimensional time series. A central problem is to extract a small number of latent factors that capture the essential dynamics of the data. Such decompositions are useful for interpretation, forecasting, and downstream modeling, and have become standard tools in fields like econometrics and finance, where they help uncover the directions that drive market movements or macroeconomic indicators.

A widely used method for this task is Principal Component Analysis (PCA), which identifies orthogonal directions in the data that successively explain the most variance (see [19, 11] for an overview). Thanks to its simplicity, efficiency, and theoretical guarantees, PCA has become a cornerstone of multivariate analysis. In financial modeling, it forms the basis of many factor models used for risk management, asset pricing, and portfolio optimization (see, e.g., [22, 14, 5, 7, 18, 8, 1, 10], among many others). However, relying solely on variance might not always be the most relevant metric for factor decomposition, especially for multiscale processes often encountered in finance which exhibit, e.g., volatility clustering. Moreover, the orthogonality constraint in PCA can be unnecessarily rigid and may hinder the discovery of more subtle dependencies in the data.

In this work, we propose an alternative decomposition approach based on non-Gaussianity of time series rather than on their variance. In order to measure the non-Gaussianity of a time series, we use the Scattering Spectra introduced in [17, 16, 15] as sufficient statistics to characterize the underlying price process. Based on these statistics, we derive a measure of *non-Gaussianity* that can be seen as capturing skewness and kurtosis of a trajectory, but in a far more expressive way than the standard measures applied to log-returns. Building on this idea, we develop a procedure for multivariate time series that iteratively identifies directions whose removal makes the residual as Gaussian (i.e. noise) as possible. By ranking components according to their contribution to non-Gaussianity, we obtain a novel form of factor decomposition. We compare our approach to PCA on both synthetic datasets and real financial time series, illustrating scenarios where our method recovers meaningful factors that PCA fails to detect. However, this gain in interpretability and sensitivity to higher-order structure comes with a higher computational cost.

The report is organized as follows. Section 1 provides a brief overview of the mathematical framework of Principal Component Analysis (PCA) and outlines its common applications in finan-

cial settings. Section 2 introduces tools for measuring the non-Gaussianity of multivariate time series. Building on these foundations, Section 3 presents our proposed factorization method based on non-Gaussianity with empirical comparisons to PCA on both synthetic and financial datasets.

0 Notations and Definitions

- Let N and T be two integers. Any $w \in \mathbb{R}^N$ is called a *direction*. Time-series $x \in \mathbb{R}^T$ are viewed as row vectors.
- Given N time series of length T, we define a multivariate time series $X \in \mathbb{R}^{N \times T}$.
- The scalar product of $(v, A) \in \mathbb{R}^N \times \mathbb{R}^{N \times T}$ is denoted by $\langle v, A \rangle := v^\top A$. Two vectors $(v, w) \in \mathbb{R}^N$ are said orthogonal if $\langle v, w \rangle = 0$ and we denote $v \perp w$.
- The norm of a matrix or vector is the Frobenius norm: $\|A\|^2 := \mathrm{Tr}(AA^\top)$.
- The *factor* associated with a direction w and a multivariate time series X is $\langle w, X \rangle$ and its *projection* is:
$$P_X(w) := w\langle w, X \rangle.$$
- The corresponding *residual* is:
$$R_X(w) := X - P_X(w).$$
- For a collection of directions $w_1, \ldots, w_n$, the residual is defined as:
$$R_X(w_1, \ldots, w_n) := X - \sum_{i=1}^{n} P_X(w_i).$$
- For $x \in \mathbb{R}^N$ and temperature parameter $\tau > 0$, we define the softmax function:
$$\mathrm{softmax}_\tau(x)_i := \frac{\exp(x_i/\tau)}{\sum_j \exp(x_j/\tau)}.$$
- The random vector $X \in \mathbb{R}^N$ is said to be non-Gaussian if at least one of its components is not Gaussian.
- Unless stated otherwise, X is assumed to be standardized, i.e., each row is centered with unit variance.

1 Principal Component Analysis: Factor Decomposition via Variance Maximization

1.1 The Principles of PCA

Principal Component Analysis (PCA) is a widely used method for extracting factors from multivariate time series [19, 11]. It achieves this by finding uncorrelated linear combinations of the original variables that explain as much variance as possible: the first component corresponds to the direction along which the data varies the most, the second captures the largest remaining variance orthogonal to the first, and so on. Thus, this decomposition relies solely on the second-order statistics of the data, as encoded in its covariance matrix.

Formally, given a multivariate time series represented as a data matrix $X \in \mathbb{R}^{N\times T}$ (where N is the number of series and T is the number of time points, assuming data is standardized), PCA determines the orthonormal factors $(v_i)_i \in \mathbb{R}^N$ such that the corresponding factors $(\langle v_i, X\rangle)_i$ have maximal variance $\big(\|\langle v_i, X\rangle\|^2\big)_i$. This is mathematically achieved through the eigenvalue decomposition of the sample covariance matrix $\tilde{X}\tilde{X}^\top$, with $\tilde{X} = \frac{1}{\sqrt{T}}X$,

$$\tilde{X}\tilde{X}^\top := \frac{1}{T}XX^\top = \sum_{i=1}^{N} \lambda_i w_i w_i^\top \tag{1}$$

where $\lambda_1 \geq \cdots \geq \lambda_N \geq 0$ are the eigenvalues (representing the variance explained by each factor) of $\tilde{X}\tilde{X}^\top$ and w_i are the corresponding orthonormal eigenvectors (the PCA factor vectors or principal directions). We can show that $v_i = \pm w_i$ (see the proof for $i = 1$ in what follows) since the eigenvectors w_i have the same characterization than the PCA factors, i.e.,

$$w_1 = \underset{\|w\|=1}{\arg\max} \|\langle w, X\rangle\|^2 \quad \text{and iteratively} \quad w_k = \underset{\|w\|=1, w\perp w_1,\ldots,w_{k-1}}{\arg\max} \|\langle w, X\rangle\|^2 \tag{2}$$

Equivalently, the $(w_i)_i$ provide the best linear reconstruction of the original data in a least-squares sense (for centered data):

$$w_1 = \underset{\|w\|=1}{\arg\min} \|R_X(w)\|^2 \quad \text{and iteratively} \quad w_k = \underset{\|w\|=1, w\perp w_1,\ldots,w_{k-1}}{\arg\min} \|R_X(w_1,\ldots,w_{k-1},w)\|^2 \tag{3}$$

This highlights PCA's dual interpretation: identifying directions of maximal variance and providing the optimal linear approximation in the least-squares sense. The second interpretation is particularly interesting since it tells that PCA can be achieved by minimizing a function $\langle\Phi\rangle$ of the residuals,

$$\langle\Phi\rangle\big(R_X(w)\big) = \|R_X(w)\|^2, \tag{4}$$

a characterization we build upon in Section 3 to introduce non-Gaussian factor decomposition.

Proof of the dual characterization $v_1 = w_1$. For $v \in \mathbb{R}^N$, the eigendecomposition of the sample covariance matrix yields,

$$\|\langle v, \tilde{X}\rangle\|^2 = \sum_{k=1}^{N} \lambda_k |\langle v, w_k\rangle|^2$$

Assuming $\|v\| = 1$ and decomposing v on the PCA basis $v = \sum_{k=1}^{N} \mu_k w_k$, one has $\sum_k \mu_k^2 = 1$ and thus $\|\langle v, \tilde{X}\rangle\|^2 \leq \max_k \lambda_k = \lambda_1$. This maximum is reached if and only if $v = \pm w_1$ which proves the first characterization regarding variance maximization. For the second characterization, for $v \in \mathbb{R}^N$ satisfying $\|v\| = 1$, one has

$$\begin{aligned}\|\tilde{X} - v \times \langle v, \tilde{X}\rangle\|^2 &= \|\tilde{X}\|^2 - 2\mathrm{Tr}(\tilde{X}\tilde{X}^T vv^T) + \mathrm{Tr}(vv^T\tilde{X}\tilde{X}^T vv^T) \\ &= \|\tilde{X}\|^2 - \mathrm{Tr}(v^T\tilde{X}\tilde{X}^T v) \\ &= \|\tilde{X}\|^2 - \|\langle v, \tilde{X}\rangle\|^2.\end{aligned}$$

Thus the equality between the argmin and the argmax. □

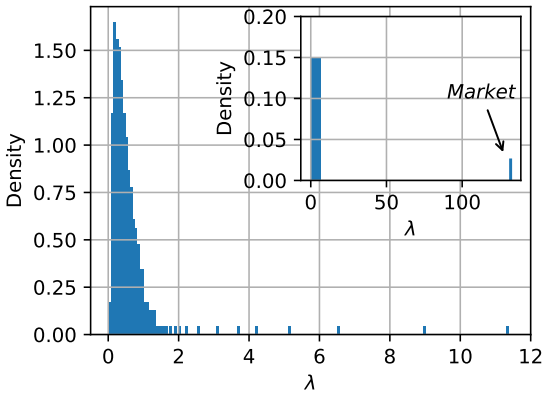
1.2 Illustrations on Financial Data

To illustrate the behavior of PCA on financial data, we apply it to a set of asset returns and highlight key stylized facts that will later guide our comparison with our own factoring approach. Specifically, we analyze the standardized daily log returns X of $N = 349$ stocks from the S&P 500 index, covering the period from 2002 to 2018. Our analysis focuses on several important features of the eigenvalues and eigenvectors of the sample correlation matrix $\mathbf{C}$, and investigates how frequently and with what intensity the PCA factors are activated over time in the data.

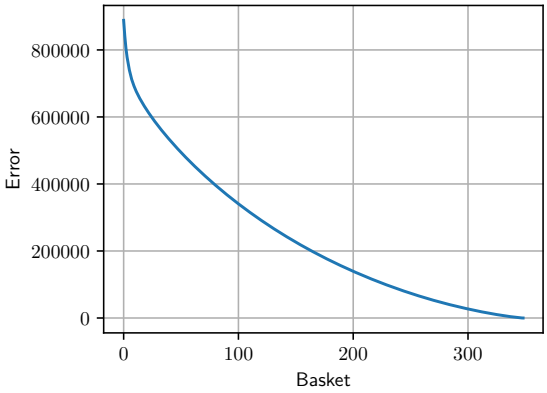
Fig. 1a shows the distribution of the eigenvalues of $\mathbf{C}$. As previously discussed, each eigenvalue λ_i quantifies the variance explained along the direction of its associated eigenvector w_i, via the relation $\lambda_i = w_i^\top \mathbf{C} w_i$. The plot reveals that the first eigenvalue is significantly larger than the rest, indicating that the leading principal component captures a disproportionately large amount of the total variance. This leading direction is commonly known as the *market mode*, as it reflects the dominant factor influencing collective movements across assets. Another well-established stylized fact in empirical finance (not shown here) is that the eigenvector corresponding to the market mode has nearly uniform weights across all stocks. This means that an equal-weighted portfolio closely aligns with the direction of maximum variance.

We also illustrate the fact that PCA seeks the best linear reconstruction of the data in the least square sense. Fig. 1b shows that the Frobenius norm of the residuals decreases monotonically as more principal components are added. Note that the curve decreases extremely fast at the very beginning of the plot which confirms the previous observation that only a very few directions 'drive' the market in terms of variance. To better understand with what intensity and frequency the PCA directions contribute to the market, we plot the activation map $P^\top X(t)$ across different time steps t in Fig. 1c for the first 50 principal components – P being the matrix which columns are the eigenvectors of $\mathbf{C}$. Positive activations are shown in red, negative in blue, and near-zero (inactive) values in white. The plot reveals that all 50 directions are globally activated, with particularly strong contributions from the first three components. We observe a gradual decline in activation intensity as we move from the first to the fiftieth direction, which is consistent with the fact that factors are ranked in descending order of their contribution to variance.

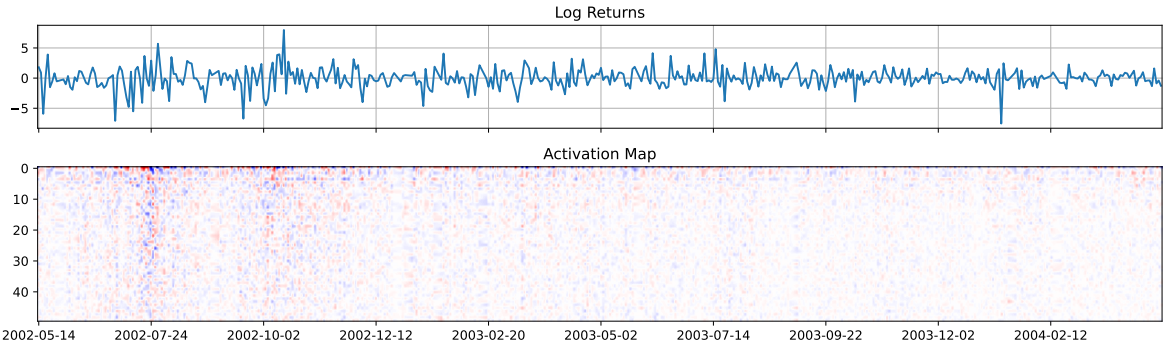
Pros and Cons of PCA PCA provides a fast and effective method for extracting factors of multivariate time-series, relying on standard linear algebra to identify orthogonal directions that maximize the variance of the data via the sample covariance matrix. Its interpretability and computational efficiency, has made PCA a widely adopted and foundational tool in multivariate analysis. However, PCA is limited by its exclusive focus on second-order statistics, making it highly sensitive, e.g., to



(a) Distribution of the eigenvalues of $\mathbf{C}$.



(b) Norm of the residuals: $i \mapsto \|R_X(w_1, \ldots, w_i)\|^2$.



(c) First row: log-returns of one stock. Second row: activation map $P^\top X$ of the first 50 PCA directions.

Figure 1: PCA analysis of daily log-returns of 349 stocks of the S&P500 from 2002 to 2018.

outliers and heavy-tailed distributions. Additionally, the assumptions of orthogonality imposed on the factors can be overly restrictive, in the sense that it may overlook meaningful correlated factors. These limitations motivate the study of alternative criteria that can reveal relevant directions beyond what variance alone can capture (see, e.g., [8, 1, 10] for enhanced PCA-based factor models).

2 Non-Gaussian Metrics of Time Series

The limitations of Principal Component Analysis (PCA) motivate the exploration of alternative criteria for identifying factors in multivariate time series. This work investigates the hypothesis that non-Gaussianity can serve as a meaningful criterion for uncovering hidden factors that may be overlooked by PCA. The underlying idea is that deviations from Gaussianity often reflect higher-order statistical dependencies—such as skewness or kurtosis—that are not captured by second-order methods like PCA. Just as PCA seeks the direction w that minimizes the mean squared error $\langle \Phi \rangle(R_X(w)) = \|R_X(w)\|^2$, the method we propose searches for the direction w that minimizes a non-Gaussianity-based func-

tion $\langle\Phi\rangle$ of the residual $R_X(w)$. We begin by constructing such a non-Gaussianity metric $\langle\Phi\rangle$ in the univariate setting, and then discuss how it can be extended and applied to the multivariate case.

2.1 Measuring non-Gaussianity of Univariate Time Series

We aim to quantify how much a univariate time series deviates from Gaussian behavior. The idea is to assign a numerical score that reflects the degree of non-Gaussianity in the data. Formally, we define a Gaussianity metric as a non-negative function that takes a univariate time series as input and returns a value representing its "distance" from Gaussianity.

A conceptually straightforward approach to constructing such a metric is to compare the statistical moments of the observed time series to those of a Gaussian distribution. While the first two moments capture location and scale, higher-order moments reveal deviations from Gaussian shape: for example, nonzero skewness indicates asymmetry, and kurtosis different from 3 suggests heavier or lighter tails. Thus, the metric can be built as follows: for a standardized time series $X \in \mathbb{R}^T$, we compare empirical moments to those of a standard normal variable $Z \sim \mathcal{N}(0,1)$. Specifically, consider the moments $m_1 = \mathbb{E}\{|Z|\} = \sqrt{2/\pi}$, $m_2 = \mathbb{E}\{Z|Z|\} = 0$, $m_3 = \mathbb{E}\{Z^3\} = 0$, and $m_4 = \mathbb{E}\{Z^4\} = 3$. A simple Gaussianity metric $\langle\Phi\rangle(X)$ can then be defined by summing the absolute differences between the empirical estimates $\hat{m}_i(X)$ and their theoretical Gaussian counterparts:

$$\langle\Phi\rangle(X) = \sum_{i=1}^{4} |\hat{m}_i(X) - m_i|. \tag{5}$$

However, while intuitive, this metric relies on moments computed only from marginal distributions which provides a limited characterization of non-Gaussianity, especially for complex time series dynamics (i.e., features that manifest across different time scales). This insufficiency motivates the development of more sophisticated descriptive statistics capable of providing a richer picture of non-Gaussian behavior, such as multiscale analyses.

The scattering spectra introduced in [17] are a set of sufficient statistics for financial price time-series [16], which capture non-Gaussian properties such as intermittency [13, 12], time-reversal asymmetries (i.e., leverage and Zumbach effect [23, 6]), and also volatility roughness [9]. These statistics are based on a cascade of wavelet decompositions to analyze patterns at different time scales. At each stage, a modulus operator is applied to the wavelet coefficients, allowing subsequent layers to capture the evolution of energy across scales (see [2] for an application to the classification of price jumps). Considering non-linear correlations of wavelet coefficients across times and scales yields a set of statistics $\Phi \in \mathbb{R}^M$, the *scattering spectra*, where $M \approx \log_2^3(T)$ for T the length of the time series. A key advantage highlighted by [17, 16] is that these Scattering Spectra provide a more refined characterization of multiscale processes compared to traditional methods based on simple moments or wavelet coefficients alone, while being designed for robust estimation from finite, single realizations. The effectiveness of their method has been demonstrated on both synthetic processes (e.g., fractional Brownian motion, multifractal random walks) and empirical data ranging from turbulence [17], through seismology [21, 20], to finance [17, 16]. This multiscale representation in turn enables the construction of a powerful non-Gaussianity metric, denoted $\langle\Phi\rangle$, whose precise formulation will be detailed in the following section.

To provide an initial assessment of the effectiveness of the function $\langle\Phi\rangle$, we compute non-Gaussianity scores for different types of time series. Fig. 2 shows a realization of a Gaussian process, serving as a baseline for comparison. Figs. 3 and 4 display two realizations of Multifractal Random Walks (MRWs),

which are well-known for their non-Gaussian behavior and characteristic volatility clustering. Note that the MRW of Fig. 4 has stronger volatility clustering than the one of Fig. 3. As expected, the score for the Gaussian sample is the lowest with $\langle\Phi\rangle(X) \approx 0.076$, whereas the MRW samples yield markedly higher values: $\langle\Phi\rangle(X) = 0.177$ and $\langle\Phi\rangle(X) = 0.277$, respectively. This illustrates the capacity of the scattering spectra to capture effectively structural differences in time series. Note that, due to implementation-specific details, the baseline score for a Gaussian process is slightly above zero, around 0.075. Higher values should be interpreted as evidence of non-Gaussianity.

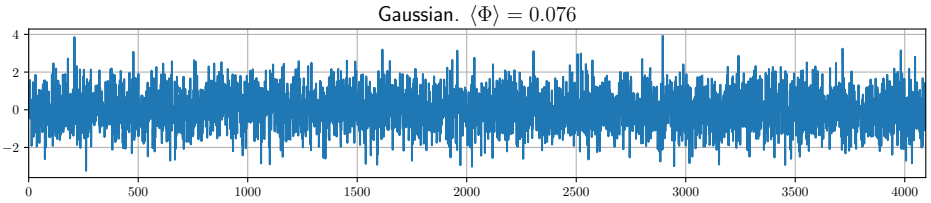


Figure 2: Gaussian Time Series

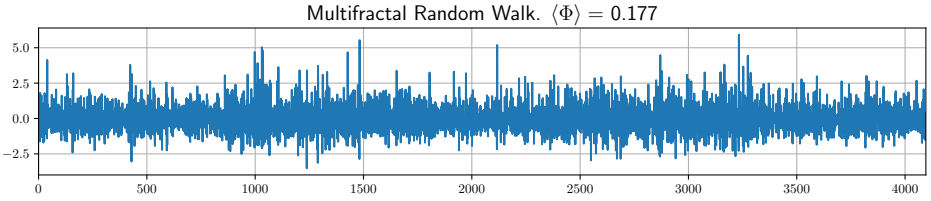


Figure 3: Multifractal Random Walk - 1

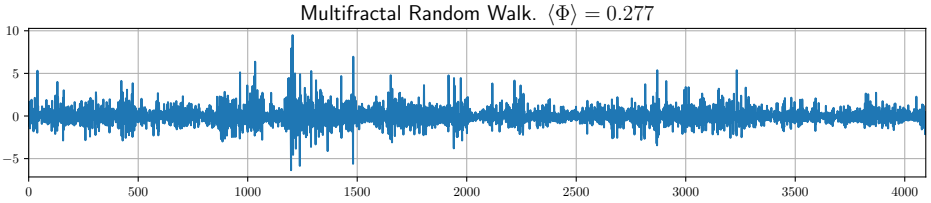


Figure 4: Multifractal Random Walk - 2

2.2 Measuring non-Gaussianity of Multivariate Time Series

Henceforth, we denote by $\Phi : \mathbb{R}^T \to \mathbb{R}^M$ the function that computes the M-dimensional scattering spectra (SS) vector for any given univariate time series of length T, as explained previously. It is important to recall that the scattering spectra are a multi-scale analogue of the standard skewness and kurtosis. We now explain how to use Φ to build non-Gaussianity scores in the univariate and multivariate setting.

Definition 1 (Non-Gaussian metric (NGM))**.** *Let X be a univariate time series with T realizations. We denote $\Phi(X) = (\Phi_1, \ldots, \Phi_M)^\top \in \mathbb{R}^M$ its scattering spectra vector and define its non-Gaussian metric by*

$$\langle \Phi \rangle(X) = \frac{1}{M} \sum_{i=1}^{M} \Phi_i(X) \in \mathbb{R}_+ \tag{6}$$

We extend this definition to a multivarite times series $X = (X_1, \ldots, X_N)^\top \in \mathbb{R}^{N \times T}$ in three possible ways and note $\langle \Phi \rangle(X) = (\langle \Phi \rangle(X_1), \ldots, \langle \Phi \rangle(X_N))^\top \in \mathbb{R}^N$,

- *average:* $\langle \Phi \rangle^{avg}(X) = \frac{1}{N} \sum_{i=1}^{N} \langle \Phi \rangle(X_i)$
- *maximum:* $\langle \Phi \rangle^{max}(X) = \max\{\langle \Phi \rangle(X)\}$
- *softmax:* $\langle \Phi \rangle^{softmax}(X) = \langle \mathrm{softmax}_\tau(\langle \Phi \rangle(X)), \langle \Phi \rangle(X) \rangle, \quad \tau > 0$

While the average metric $\langle \Phi \rangle^{avg}$ provides an overall assessment of non-Gaussianity across all component series, the maximum metric $\langle \Phi \rangle^{max}$ specifically isolates the contribution of the single most non-Gaussian component. The motivation for considering this maximum variant relates to practical aspects and specific use-cases that will be elaborated upon later. Complementary to the maximum, the softmax metric $\langle \Phi \rangle^{softmax}$ provides a smooth and differentiable approximation of the maximum function. This differentiability is a crucial property, making the softmax variant particularly suitable for scenarios where the non-Gaussian metric needs to be integrated into objective functions optimized via gradient-based algorithms. Next, we show that the scattering spectra Φ and the non-Gaussian metric are intricately related.

Proposition 1. *Let $X \in \mathbb{R}^N$ be a random vector with Gaussian components. Then, $\langle \Phi \rangle(X) = 0$.*

This result, established in [17], shows that the non-Gaussianity metric (NGM) vanishes for purely Gaussian time series. Although the converse has not been formally proven, we adopt the working assumption that the NGM provides a reliable proxy for deviations from Gaussianity. Based on this property, the interpretation of $\langle \Phi \rangle$ is straightforward: the smaller the NGM value, the closer the time series is to Gaussian behavior; higher values indicate increasingly pronounced non-Gaussian features.

3 Non-Gaussian Factor Decomposition

3.1 Minimizing the Non-Gaussianity of Residuals: an Iterative Algorithm

We now introduce the core methodology proposed in this work for factoring multivariate time series using the non-Gaussian metric $\langle \Phi \rangle$ introduced earlier. The guiding principle is analogous to that of PCA: whereas PCA seeks the direction w that minimizes the mean squared error of the residual $R_X(w)$ (see Section 1), our approach searches for the direction (or *basket* in financial terms) w that minimizes the non-Gaussianity of the residual, measured by $\langle \Phi \rangle(R_X(w))$. In other words, we aim to identify directions that, when removed from the data, leave a residual time series that is as close to Gaussian as possible. This leads us to the notion of an *exact non-Gaussian direction*, which formalizes the ideal case in which the extracted direction leaves behind a (perfectly) Gaussian time series. In this sense, our approach aims to isolate the non-Gaussian components while treating the remaining Gaussian residual as noise, implying that the extracted direction concentrates all the informative signal.

Definition 2 (Exact Non-Gaussian Direction)**.** *Let $X \in \mathbb{R}^N$ be a non-Gaussian process. A direction $w \in \mathbb{R}^N$ with $\|w\| = 1$ is an exact non-Gaussian direction if the components of the residual $R_X(w)$ are Gaussian processes. Note that this implies $\langle\Phi\rangle(R_X(w)) = 0$.*

While this definition provides a theoretical ideal, real-world time series rarely admit such perfectly isolating directions. To address this, we introduce a practical iterative algorithm designed to extract a set of K non-Gaussian directions $\{w_1, w_2, \ldots, w_K\}$. At each step k, the algorithm identifies the direction w_k that minimizes the non-Gaussianity of the residual after removing the previously extracted factors $w_1, \ldots, w_{k-1}$. Let $X \in \mathbb{R}^{N\times T}$ be a standardized multivariate time series. The iterative extraction process is defined as follows:

$$w_1 = \underset{\|w\|=1, w\in\mathbb{R}^N}{\arg\min} \langle\Phi\rangle^{avg}\left(R_X(w)\right) \tag{7}$$

$$\vdots$$

$$w_k = \underset{\|w\|=1, w\in\mathbb{R}^N}{\arg\min} \langle\Phi\rangle^{avg}\left(R_X(w_1, \ldots, w_{k-1}, w)\right) \tag{8}$$

This iterative minimization ranks the extracted directions by their ability to reduce the non-Gaussian content of the signal, offering a fundamentally different decomposition compared to variance-based methods such as PCA. Note that the non-Gaussian scores $\langle\Phi\rangle^{\max}$ and $\langle\Phi\rangle^{\text{softmax}}$ defined in Def. 1 can also be used.

Another key distinction between this NGM-based factorization and PCA lies in the constraints imposed on the extracted factors. Unlike PCA, our methodology does not require the directions $\{w_k\}$ to be orthogonal. This relaxation broadens the search space and allows for the discovery of correlated directions of non-Gaussianity that may be overlooked by PCA. Moreover, whereas PCA always produces exactly N orthogonal factors (that form a basis of $\mathbb{R}^N$), the number of non-Gaussian directions $\{w_k\}$ is not predetermined. This raises the practical question of how many factors K should be extracted using the iterative procedure of Eqs. (7–8). For real-world data, we do not expect the residual $\langle\Phi\rangle^{\text{avg}}(R_X(w_1, \ldots, w_K))$ to vanish entirely for any finite K and thus, a stopping criterion is required. A solution could consist in monitoring the decrease of the non-Gaussian metric at each step and terminate the algorithm once the improvement from adding a new factor w_{k+1} falls below a given tolerance ϵ. In other words, we stop when

$$\langle\Phi\rangle^{\text{avg}}(X_k) - \langle\Phi\rangle^{\text{avg}}(X_{k+1}) < \epsilon,$$

indicating that additional extraction yields negligible progress in making the residual more Gaussian.

3.2 Principal Component Analysis vs non-Gaussian Directions: Examples

We now give a series of simple examples to better grasp the notions introduced previously and see how PCA and NGM differ. The first example builds an explicit random vector which has a unique exact non-Gaussian direction.

Example 1. *Let $X = \epsilon + w_1 X_1$ be a random vector with ϵ a multivariate Gaussian and X_1 a non-Gaussian process such that ϵ and X_1 are independent. Then, X admits a unique exact non-Gaussian direction which is w_1.*

Proof. First, we show that w_1 is an exact non-Gaussian direction. This is clear since

$$X - w_1\langle w_1, X\rangle = (I - w_1 w_1^\top)\epsilon \tag{9}$$

is composed of univariate gaussian processes only. To prove unicity, we use the following intermediary result which will be proved at the end. Let Y, Z be two independent random variables such that $G = Y + Z$ is Gaussian and Y is Gaussian. Then Z is also Gaussian. Coming back to our proof, suppose there exist two solutions $w_1 \neq w_2$. Then,

$$X - w_2 w_2^\top \epsilon - w_2 X_1 \tag{10}$$

is composed of univariate Gaussian processes. The intermediary result implies that X_1 is composed of univariate gaussian processes. This is absurd and yields the result.

We now prove the intermediary result. Since Y and Z are independent, we can write the characteristic function $\varphi_G(t)$ of G as follows

$$\varphi_G(t) = \varphi_Y(t) \cdot \varphi_Z(t) \tag{11}$$

Since Y is Gaussian, we know that φ_Y is invertible which yields

$$\varphi_Z(t) = \varphi_G(t) \cdot \varphi_Y^{-1}(t) \tag{12}$$

We recognize the characteristic function of the normal distribution $\mathcal{N}(\mu - \mu_1, \sqrt{\sigma^2 - \sigma_1^2})$. □

The next example illustrates how the orthogonality constraint in PCA can limit the extraction of important components. In the two-dimensional case, PCA always returns the same pair of orthogonal directions, regardless of the underlying data distribution—a direct consequence of enforcing orthogonality. In contrast, our non-Gaussian approach is free from this constraint and can more flexibly adapt to the structure in the data.

Example 2. *If $N = 2$, and $Var(X_1) = Var(X_2)$, then the PCA eigenvectors of $X = (X_1, X_2)^\top$ are always $\frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix}$ and $\frac{1}{\sqrt{2}}\begin{bmatrix}1\\-1\end{bmatrix}$.*

Proof. The covariance matrix of X is of the form $\begin{bmatrix}a & b\\ b & a\end{bmatrix}$ with $a = \mathrm{Var}(X_1) = \mathrm{Var}(X_2)$. We notice that the sum of its rows is constant equal to $a+b$, thus the vector $v_1 = \begin{bmatrix}1\\1\end{bmatrix}$ is an eigenvector of $\mathrm{Cov}(x)$. Since $\mathrm{Cov}(x)$ is symmetric we know that any non-zero vector v_2 orthogonal to v_1 is an eigenvector. □

Building on the two previous examples, we now build a process X that admits an exact non-Gaussian direction which does not coincide with any PCA eigenvector, highlighting a fundamental difference between the two methodologies.

Example 3. *Let $X = \epsilon + w_1 X_1$ be a random vector in $\mathbb{R}^2$ where $w_1 = (w_{11}\, w_{12})^\top$ such that $|w_{11}| \neq |w_{12}|$ and X_1 is a non-gaussian centered random variable with finite order two moment. Choose $\epsilon =$*

$(\alpha_1\epsilon_1\, \alpha_2\epsilon_1)^\top$ *where* ϵ_1 *is a standard gaussian independent from* X_1 *and choose* (α_1, α_2) *such that the components of* X *have unit variance. Then, Expl. 1 assures that the unique exact non-gaussian direction is* w_1 *which, by construction, does not coincide neither with* $\frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix}$ *or with* $\frac{1}{\sqrt{2}}\begin{bmatrix}1\\-1\end{bmatrix}$*, which are the PCA directions according to Expl. 2.*

3.3 Synthetic Data: Retrieving the non-Gaussian Directions

The goal of this section is to evaluate our methodology on synthetic data by generating multivariate time series with known distributions and known non-Gaussian directions. This controlled setting allows us to assess whether the procedure introduced earlier is capable of correctly recovering the true non-Gaussian directions. We focus on a simple low-dimensional example where the time series has $N = 2$ components and admits an exact non-Gaussian direction w^*, meaning that the residual obtained after removing this direction is purely Gaussian (see Def. 2). In this context, we also compare the performance and robustness of the different multivariate non-Gaussianity metrics $\langle\Phi\rangle^{\text{avg}}$, $\langle\Phi\rangle^{\max}$, and $\langle\Phi\rangle^{\text{softmax}}$ introduced in Def. 1.

Synthetic Data We sample a non-Gaussian process $X \in \mathbb{R}^{N\times T}$ with $N = 2$ and $T = 4096$ such that,

$$X = \begin{bmatrix}\epsilon_1 + y\\ \epsilon_2\end{bmatrix} = \epsilon + w^* Y \in \mathbb{R}^2$$

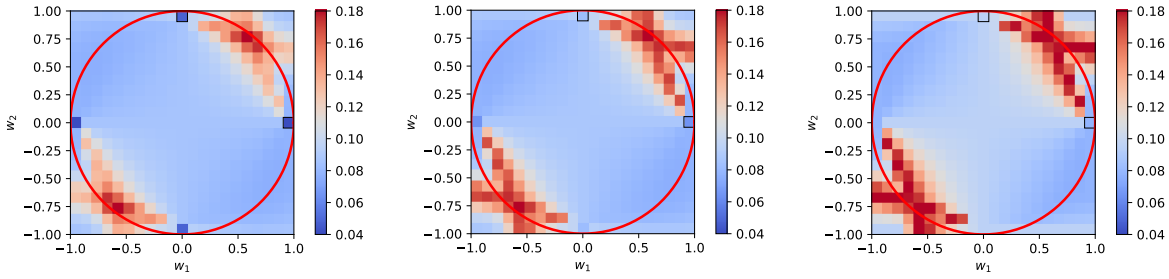
where $\epsilon = (\epsilon_1, \epsilon_2)^\top$, $Y = (y, 0)^\top$ with y a non-Gaussian process and $w^* = (1, 0)^\top$. From Expl. 1 we know that w^* is the unique exact non-Gaussian direction of X.

We implement the iterative algorithm described in Eqs.(7–8) by minimizing the non-Gaussian metric of the consecutive residuals. We use the `L-BFGS-B` gradient-descent based algorithm with an initial guess w_0 uniformly sampled on the sphere, $w_0 \sim \mathscr{U}(\mathbb{S}^{N-1})$.

Loss Landscape Visualizing the loss landscape is a common tool for gaining intuition about the structure of an objective function, particularly when optimizing via gradient descent. It helps identify symmetries, flat regions, local minima, and how well the loss function distinguishes between meaningful solutions. Fig. 5 shows the loss landscapes corresponding to the three different non-Gaussianity metrics. The red circle represents the unit ℓ_2-sphere $\mathbb{S}$, which defines the constraint set for admissible directions. The two sparse directions $w^* = (1, 0)^\top$ and $(0, 1)^\top$ are highlighted by thin black rectangles.

We first observe a central symmetry in all three landscapes, which is expected since $R_X(w) = R_X(-w)$. This symmetry allows us to only focus on the half-space where $w_1 \geq 0$. Ideally, we want the point corresponding to the true non-Gaussian direction w^* to be a clear minimum—appearing as a dark blue region in the heatmap. In Fig. 5a, corresponding to the average non-Gaussianity metric, we indeed observe a low loss at w^*. However, we also notice that the other sparse direction $(0, 1)^\top$ yields a similarly low value. This indicates that the optimization could converge to either sparse direction, depending on initialization. This behavior can be explained by the sparsity of the corresponding residuals: $R_X(w^*) = (0, \epsilon_2)^\top$ and $R_X((0, 1)^\top) = (\epsilon_1 + y, 0)^\top$. Since the non-Gaussianity score of a constant time series is zero, both residuals have a very low average NGM, even though only one of them is the true non-Gaussian direction. To address this limitation, we use the maximum and softmax alternatives to the NGM. As shown in Figs. 5b and 5c, these versions significantly penalize the

spurious sparse direction $(0,1)^\top$, producing higher loss values there while retaining a clear minimum at w^*. The softmax version, in particular, offers a smooth and differentiable approximation of the maximum, making it well-suited for use with gradient-based optimization.



(a) $\langle\Phi\rangle^{avg}$ (b) $\langle\Phi\rangle^{softmax}$ (c) $\langle\Phi\rangle^{max}$

Figure 5: Heatmap of the different NGMs. The superimposed red circle is the unit sphere (feasible region).

Convergence Results We give the convergence results of the three non-Gaussian metrics for different initial guesses uniformly sampled on the sphere $w_0 \sim \mathscr{U}(\mathbb{S}^{N-1})$, including the sparse direction $(0,1)^\top$. For the softmax version, we chose the temperature parameter $\tau = 0.2$ — note that when $\tau \to 0$, the metric behaves as the maximum non-Gaussian metric. For both the average and softmax versions, regardless of the initial guess, the algorithm converged to w^*. However, for the maximum score, the algorithm converged to $(0.98, 0.20)^\top$ (regardless of the initial guess). This showcases the importance of having a differentiable objective function.

Bias of the Average Non-Gaussian Metric toward Sparse Directions We previously noted that the average non-Gaussianity metric can exhibit a bias toward sparse directions. This issue arises when the residual $R_X(w)$ contains many zero components: the average $\langle\Phi\rangle^{\text{avg}}(X)$ is then artificially lowered, as it averages several zero values with a few non-zero ones. As a result, sparse directions can yield scores even lower than the Gaussian reference level (see Fig. 5a), which is not the intended behavior of the metric. However, one can show that, almost surely, at most one component of $R_X(w)$ is exactly zero (see Prop. 2 below). This suggests that the bias introduced by sparsity is primarily a concern in low-dimensional settings ($N \approx 1$) and becomes negligible as the dimension increases ($N \gg 1$). To mitigate this issue when necessary, the softmax version of the non-Gaussianity score offers a more robust alternative. It downweights inactive components without completely ignoring them, avoiding the pitfalls of sparse directions. However, its performance depends on the choice of the temperature parameter, which may not always be straightforward to tune.

Proposition 2. *Let X be a random vector in $\mathbb{R}^N$ with a density. Let $w \in \mathbb{R}^N$ such that $\|w\| = 1$. Then at most one component of $R_X(w)$ is zero, almost surely.*

Proof. Lets find a necessary and sufficient condition for the first component $(R_X(w))_1$ of the residual to be zero almost surely. We denote $e_1 = (1, 0, \ldots, 0)^\top \in \mathbb{R}^N$. Then,

$$(R_X(w))_1 = (e_1^\top - w_1 w^\top)X \tag{13}$$

Since $\|w\| = 1$, $(e_1^\top - w_1 w^\top) \in \mathbb{R}^{1\times N}$ is of rank one if and only if $w \neq e_1$ in which case $\mathrm{Ker}((e_1^\top - w_1 w^\top)) = 0$. Thus, $(R_X(w))_1 = 0$ almost surely if and only if $w = e_1$. Thus, it is impossible that more than two components of the residual are zero. □

Should We Optimize Iteratively or Simultaneously over Directions? An alternative to the iterative extraction of non-Gaussian directions is to optimize over multiple directions simultaneously—for instance, optimizing K_0 directions in parallel using gradient descent. While this is an appealing idea that could improve precision, it remains an open direction for future research. It is important to point out that this approach introduces a notable drawback when $K_0 \approx N$. In such cases, the average non-Gaussianity metric fails systematically. Indeed, simultaneous optimization encourages the selection of sparse directions—namely, the canonical basis vectors e_i—corresponding to the K_0 components of X with the highest individual non-Gaussianity scores. By picking these vectors, the residual $R_X(e_1, \ldots, e_{K_0})$ ends up having (at least) K_0 zero components, and the average is taken over these zeros and the remaining $N - K_0$ components—which typically exhibit the lowest non-Gaussianity score. As a result, the average score is artificially minimized, not because of meaningful structure, but due to this sparsity artifact. Thus, one must be cautious when applying simultaneous optimization, as it can artificially favor sparse solutions, once again highlighting the softmax formulation as a more robust –and furthermore differentiable– alternative.

3.4 Financial Data: Uncovering New Factors

In this section, we analyze the daily log returns X of financial assets from the S&P500 index over the period 2009 to 2025. Our objective is to compare the factors extracted by PCA with those obtained through our non-Gaussian methodology. The analysis presented here focuses exclusively on the average non-Gaussianity metric. A thorough evaluation of the maximum and softmax variants is left for future work, as these could further improve the robustness and interpretability of the results. It is important to point out that our experiments were constrained by computational limitations. In particular, the computation of scattering spectra is time-intensive, which makes gradient descent computationally demanding, especially in high dimensions. For reference, extracting a single non-Gaussian direction from a dataset of 349 stocks and 4096 time points takes approximately 5.5 hours —we leave a more thorough study in high dimensions for future research. Thus, we limit our empirical analysis to lower-dimensional subsets of the data.

3.4.1 Low-dimensional Study with 10 Stocks

In this section, we focus on six distinct sectors of the S&P500, selecting 10 stocks from each among Consumer Staples, Financials, Industrials, Information Technology, Materials, and Utilities. We also lead a cross-sector analysis by picking one representative stock from each of these six sectors, along with four additional stocks from the Communication Services, Consumer Discretionary, Energy, and Health Care sectors.

Gaussianity of Residuals Fig. 6 displays the evolution of the non-Gaussianity metric $\langle \Phi \rangle^{\mathrm{avg}}$ computed on the residuals $R_X(w_1, \ldots, w_i)$ for $i \in \{1, \ldots, 10\}$, comparing the results obtained with PCA and with non-Gaussian directions. As expected, the non-Gaussianity curve associated with our method decreases monotonically. This confirms that the iterative algorithm behaves as intended: each new

direction reduces the non-Gaussianity of the residual. Notably, the curve begins to plateau after approximately 5 extracted directions, suggesting that beyond this point, additional directions contribute little to reducing non-Gaussianity. Interestingly, this plateau occurs above the Gaussian reference level, implying that even after several iterations, a residual level of non-Gaussianity persists in the data. It would be interesting to investigate this behavior more thoroughly by continuing the algorithm with additional directions to see whether the non-Gaussianity continues to plateau or eventually decreases further.

Regarding the PCA results, we observe that the non-Gaussianity metric also exhibits a generally decreasing trend, albeit not monotonically. This suggests that PCA directions do tend to reduce non-Gaussianity to some extent, despite being optimized solely to minimize variance. As expected, the PCA curve remains consistently above the one from our method, confirming that PCA is suboptimal in this setting when measured by non-Gaussianity.

Basket Weights Having confirmed that the algorithm behaves as expected in terms of convergence and residual Gaussianity, we now turn to a closer analysis of the weights of the identified directions. Fig. 7 displays the weights of the 10 PCA and non-Gaussian directions for the Consumer Staples sector.

The first notable observation is that the leading direction is remarkably similar for both methods. Just like PCA, the non-Gaussian approach identifies the market mode as the most dominant direction—retrieving a substantial portion of the non-Gaussianity in the data. As shown earlier in Fig. 1a, the market mode explains a large amount of the total variance, and this also results in a high concentration of non-Gaussian structure along that direction. However, beyond the first direction, the comparison becomes less straightforward. The remaining directions from the two methods appear noticeably different, but no obvious structural pattern can be discerned.

To explore other characteristics such as sparsity, Fig. 8 shows the ℓ_1 norms of the directions across all six sectors. Overall, both PCA and non-Gaussian directions exhibit a gradual decline in ℓ_1 norm, though the differences between the two methods are generally small. For PCA, this trend is expected: sparse directions can correspond to low-variance components, especially when they exploit strong correlations between assets (e.g., buying one stock while shorting another). For the non-Gaussian directions, the interpretation is more nuanced. As previously discussed, sparsity in the direction weights does not necessarily imply sparsity in the residual, nor does it automatically lead to lower non-Gaussianity scores. Nevertheless, the observed trend suggests that non-Gaussian directions tend to become progressively sparser, which may be a stylized fact.

Similarity of the Subspaces Spanned by the Directions To better understand how the non-Gaussian directions differ from those obtained via PCA, we use a similarity measure known as the *fraction of common modes* (FCM), which quantifies the overlap between the subspaces spanned by the first n directions of each method. This metric takes values between 0 and 1, with 1 indicating that the two subspaces are identical [3, 4].

The FCM is computed as follows: let $\mathbf{v}_1, \ldots, \mathbf{v}_N$ denote the PCA directions, and $\mathbf{u}_N, \ldots, \mathbf{u}_1$ the non-Gaussian directions. Define $\mathbf{P}^{\mathrm{PCA}}$ and $\mathbf{P}^{\mathrm{NGM}}$ as the $N \times N$ matrices whose columns are the respective directions. To compute the FCM for the first n directions, we: (i) form the matrix $\left(\mathbf{P}^{\mathrm{PCA}}\right)^\top \mathbf{P}^{\mathrm{NGM}}$; (ii) extract the top-left $n \times n$ submatrix; (iii) compute its singular values $\{w_k\}_{1 \le k \le n}$; and (iv) take the geometric mean $\left(\prod_{k=1}^{n} w_k\right)^{1/n}$, which yields the FCM value for that n. Plotting this as a function of n reveals how the similarity between the two subspaces evolves as more directions are considered.

Fig. 9 shows the FCM values across all six sectors. The curves consistently start near 1, indicating that the first direction is identical for PCA and the non-Gaussian method. This confirms earlier observations that both approaches capture the same dominant direction in the data, i.e., the market mode. We note that the FCM then strongly decreases on the next two or three directions, meaning that the directions found by the methods significantly differ. This suggests that the PCA orthogonality constraint imposed between the first two or three directions might be overly restrictive and overlook important correlated factors that are captured by the non-Gaussian approach — justifying why the subspace they span is so different and why the FCM metric drops. A similar pattern is observed in Fig. 10, where we compute the FCM of the cross-sector stocks. To validate this intuition, we plot the activation maps of the directions in the next paragraph.

Activation Maps To better understand how the extracted directions contribute to the dynamics of the multivariate time series, we plot the activation maps $\mathbf{P}^{\mathrm{PCA}}X(t)$ and $\mathbf{P}^{\mathrm{NGM}}X(t)$, which represent the time-varying activations of the PCA and non-Gaussian directions, respectively.

Figs. 11 to 13 display the activation maps for three sectors: Consumer Staples, Financials, and Industrials. In each figure, the first row corresponds to PCA activations, and the second row to non-Gaussian activations. We also include the ℓ_1 norms of the activations to quantify their overall intensity in the last row. In all sectors, the first component—corresponding to the market mode—is consistently and strongly activated, as expected. However, the observed figures are similar for both methods: despite the lack of orthogonality constraints, the non-Gaussian directions do not appear to be more frequently or more strongly activated than the PCA directions, suggesting there is no real advantage to use non-Gaussian directions in these sector-specific settings.

However, the situation changes when analyzing the activation maps of the representative cross-sector stocks (Fig. 14). Here, the non-Gaussian directions show a greater number of active components. This suggests a key distinction: within a single sector, stock dynamics are often homogeneous, and a single dominant factor (e.g., the 'sector' market mode) captures most of the behavior. In contrast, cross-sector stock dynamics require a richer set of latent factors to be described. In this more heterogeneous setting, the non-Gaussian methodology reveals additional structure that PCA fails to capture. This observation highlights the potential value of non-Gaussian directions for high-dimensional applications, which we leave as a promising direction for future research. In the next section, we briefly discuss results related to PCA in the high-dimensional regime.

3.4.2 High-Dimensional Study: 349 Stocks

In this section, we analyze the daily log returns of 349 stocks from the S&P500. Due to the high computational cost of computing non-Gaussian directions in such a large-dimensional setting, our analysis is restricted to PCA. A first notable finding concerns the non-Gaussianity of the residuals produced by the PCA directions. As shown in Fig. 15, the non-Gaussianity metric of the residuals decreases in an almost perfectly monotonic manner. This means that by retrieving variance, the PCA directions also naturally retrieves non-Gaussianity, which can be expected. We also examine the non-Gaussianity of individual projections $w_i^\top X \in \mathbb{R}^{1\times T}$ in Fig. 16. Interestingly, the non-Gaussianity metric globally decreases and reaches near-Gaussian levels starting from the 150th component until the last. Understanding the underlying reasons for this behavior—and whether it generalizes to other financial datasets—is an interesting question that we leave for future research.

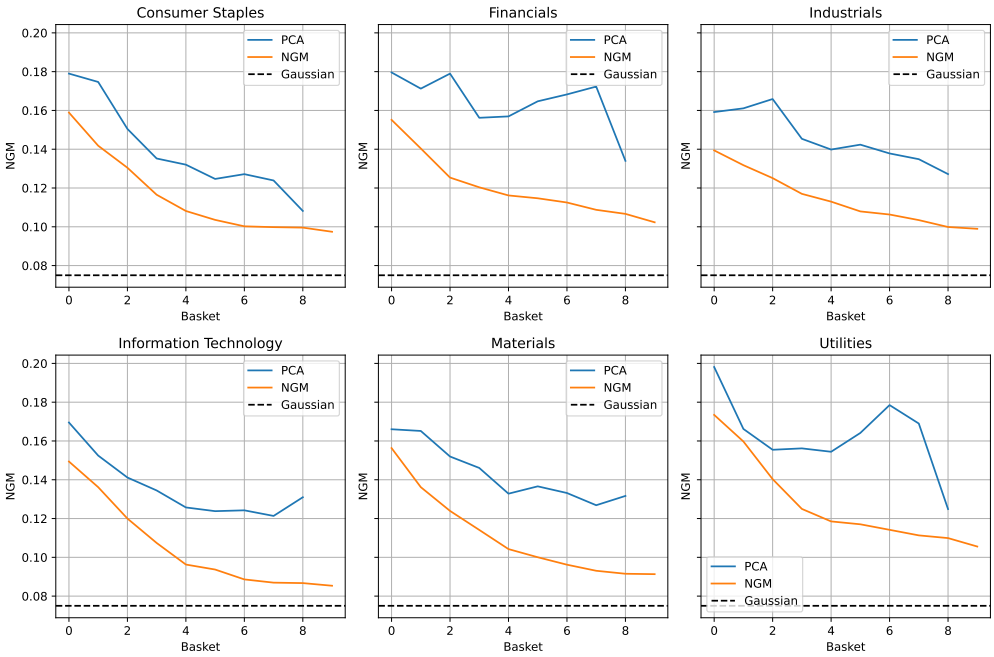


Figure 6: Average non-Gaussian metric of the residuals: $i \mapsto \langle \Phi \rangle^{avg}(R_X(w_1, \ldots, w_i))$

4 Conclusion & Perspectives

In this report, we introduced a novel factor decomposition method for multivariate time series based on non-Gaussianity. While Principal Component Analysis is the standard approach for extracting factors that explain the most variance, it relies solely on second-order statistics and enforces orthogonality constraints that may hinder the discovery of more subtle or informative factors. Building on the scattering spectra of Morel et al. [17], we proposed an iterative procedure to identify directions whose removal makes the residual time series as close to Gaussian noise as possible. This leads to a new class of factors, called non-Gaussian directions, ranked by their contribution to non-Gaussianity rather than variance of the multivariate process. We demonstrated that this approach can recover interpretable components overlooked by PCA, both on synthetic data and financial time series.

This work opens several perspectives. First, extending the methodology to high-dimensional settings is a natural next step. Our initial experiments on 349 financial assets showed encouraging signs, but a full application of the non-Gaussian direction extraction procedure requires greater computational resources due to the high cost of computing the scattering spectra. Also, this methodology could be applied beyond finance, particularly to time series in the biomedical domain. Physiological signals such as EEG or ECG recordings are often characterized by non-Gaussian and multi-scale behavior, making them natural candidates for a non-Gaussian factor analysis framework. Finally, this approach offers interesting potential for source separation problems, where the goal is to disentangle latent components from observed mixtures. By focusing on directions that extract non-Gaussian content, the methodology could serve as a powerful alternative to classical approaches such as Independent Component Analysis (ICA), especially in cases where sources exhibit multiscale non-Gaussian

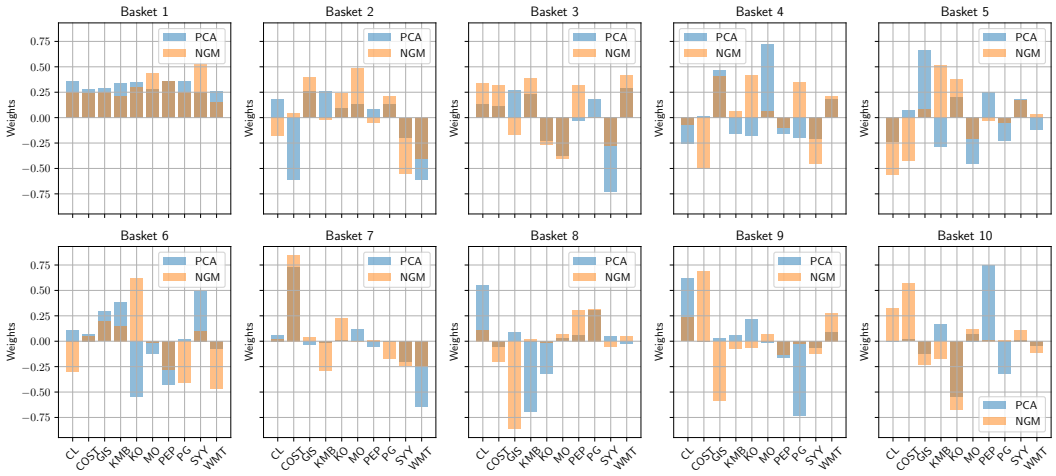


Figure 7: Weights of each of the 10 directions (*baskets*) for the Consumer Staples sector.

dynamics.

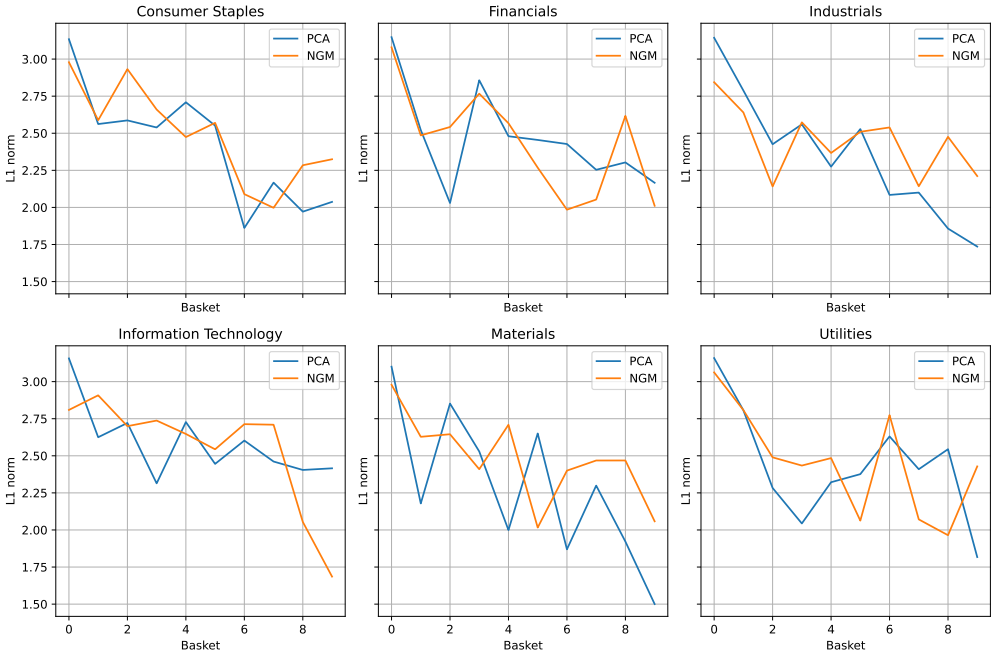


Figure 8: ℓ_1 norm of each direction and for different sectors.

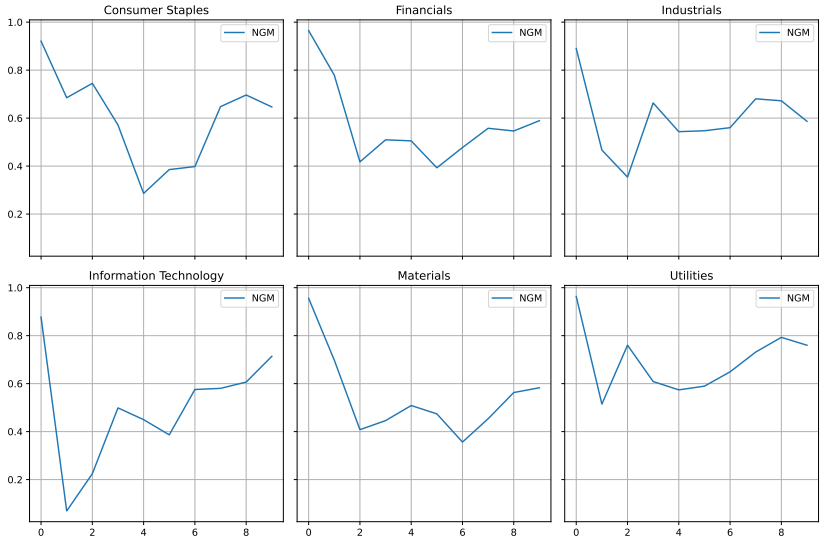


Figure 9: Fraction of Common Modes for different sectors.

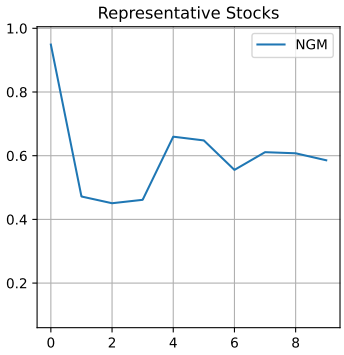


Figure 10: Fraction of Common Modes for cross-sector stocks.

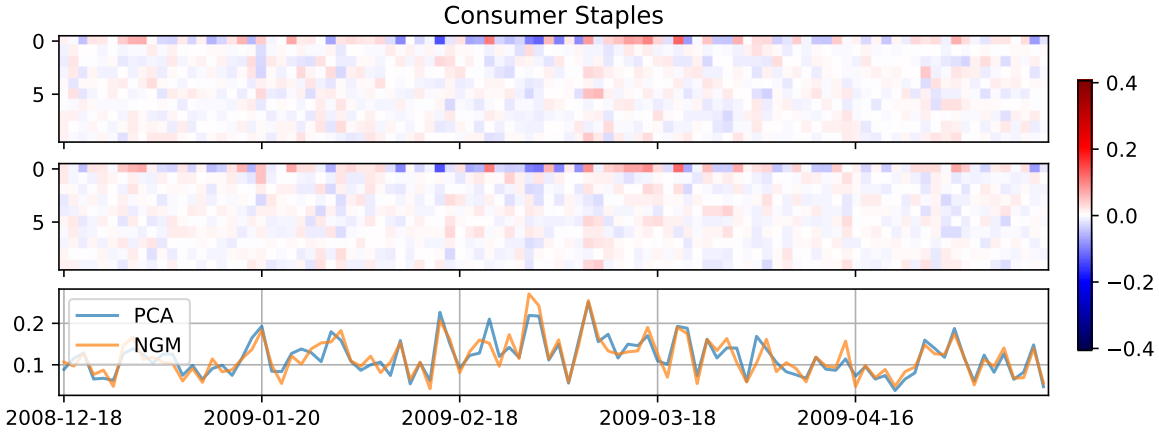


Figure 11: Activation Map for Consumer Staples. First two rows: PCA and NGM. Last row: ℓ_1 norm of the activation map.

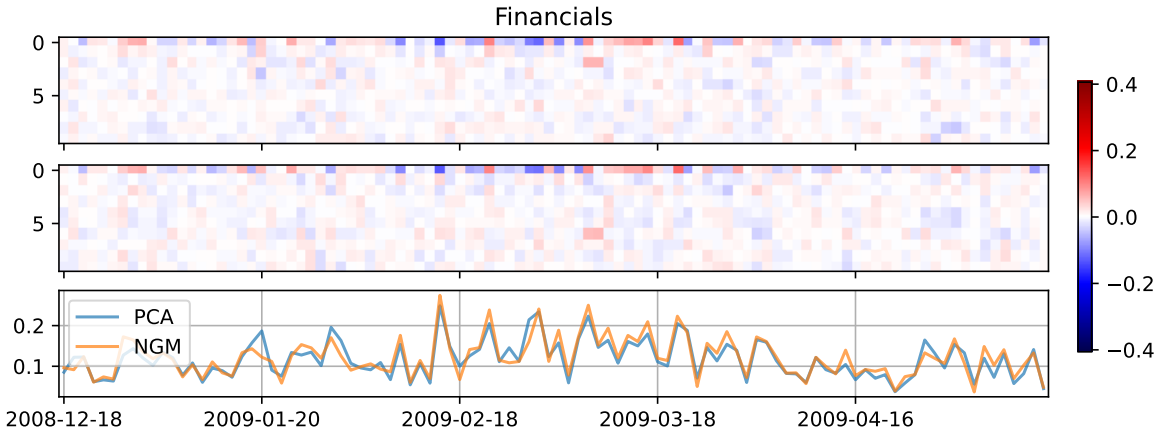


Figure 12: Activation Map for Financials. First two rows: PCA and NGM. Last row: ℓ_1 norm of the activation map.

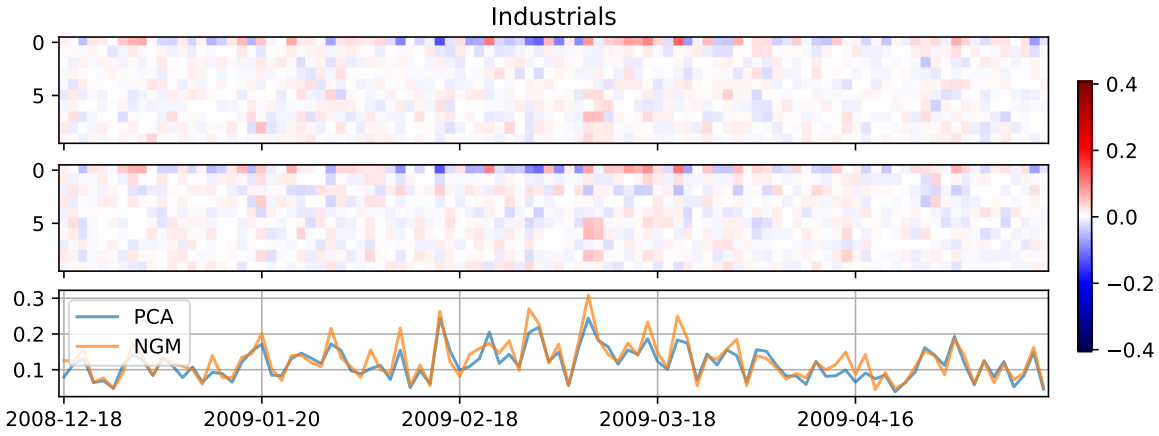


Figure 13: Activation Map for Industrials. First two rows: PCA and NGM. Last row: ℓ_1 norm of the activation map.

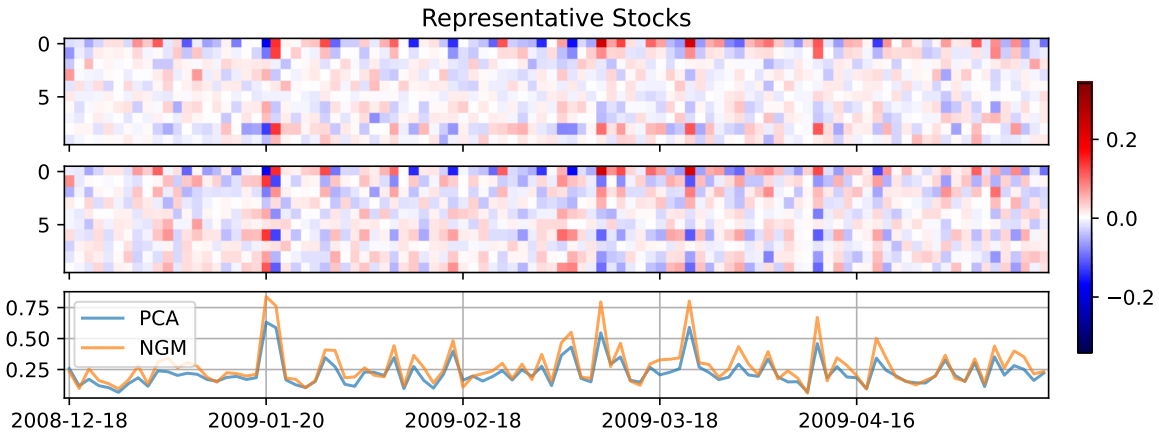


Figure 14: Activation Map for cross-sector stocks. First two rows: PCA and NGM. Last row: ℓ_1 norm of the activation map.

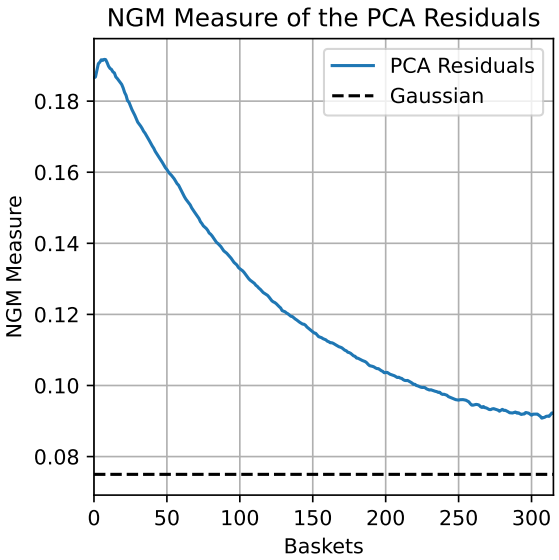


Figure 15: $i \mapsto \langle \Phi \rangle^{avg}(R_X(w_1, \ldots, w_i))$

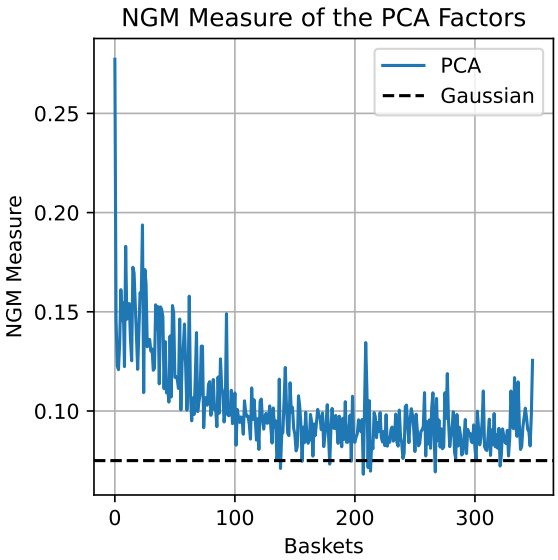


Figure 16: $i \mapsto \langle \Phi \rangle^{avg}(w_i^\top X)$

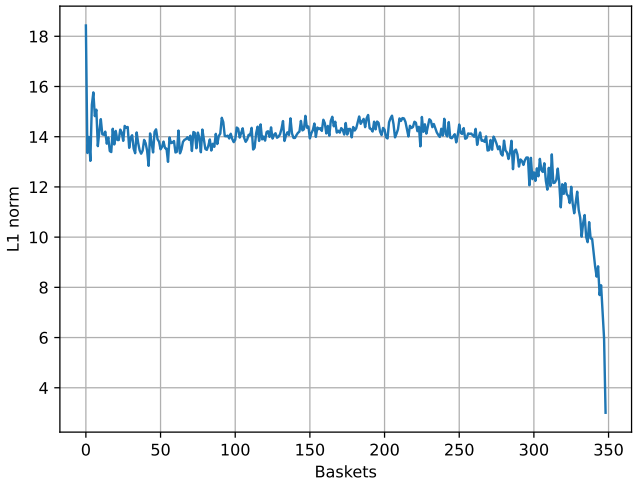


Figure 17: ℓ_1 norm of the PCA directions.

References

[1] Yacine Ait-Sahalia and Dacheng Xiu. Using principal component analysis to estimate a high dimensional factor model with high-frequency data. *Journal of Econometrics*, 2017.

[2] Cecilia Aubrun, Rudy Morel, Michael Benzaquen, and Jean-Philippe Bouchaud. Identifying new classes of financial price jumps with wavelets. *Proceedings of the National Academy of Sciences*, 122(6):e2409156121, 2025.

[3] M Benzaquen, I Mastromatteo, Z Eisler, and J-P Bouchaud. Dissecting cross-impact on stock markets: an empirical analysis. *Journal of Statistical Mechanics: Theory and Experiment*, 2017(2):023406, 2017.

[4] J-P Bouchaud, Laurent Laloux, M Augusta Miceli, and Marc Potters. Large dimension forecasting models and random singular value spectra. *The European Physical Journal B*, 55:201–207, 2007.

[5] Gary Chamberlain and Michael Rothschild. Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica*, 1983.

[6] Rémy Chicheportiche and Jean-Philippe Bouchaud. The fine-structure of volatility feedback i: Multi-scale self-reflexivity. *Physica A-statistical Mechanics and Its Applications*, 2014.

[7] Eugene F. Fama and Kenneth R. French. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 1993.

[8] Jianqing Fan, Yuan Liao, and Weichen Wang. Projected principal component analysis in factor models. *Annals of Statistics*, 2016.

[9] Jim Gatheral, Thibault Jaisson, and Mathieu Rosenbaum. Volatility is rough. *Quantitative Finance*, 18(6):933–949, 2018.

[10] Yong He, Lingxiao Li, Dong Liu, and Wen-Xin Zhou. Huber principal component analysis for large-dimensional factor models. *Journal of Econometrics*, 2025.

[11] I. Johnstone and D. Paul. Pca in high dimensions: An orientation. *Proceedings of the IEEE*, 2018.

[12] Benoit B. Mandelbrot. Fractals and scaling in finance: Discontinuity, concentration, risk. 1997.

[13] B. Mandlebrot. The variation of certain speculative prices. 1963.

[14] R. C. Merton. An intertemporal capital asset pricing model. 1973.

[15] Rudy Morel. *Compact models of multi-scale processes*. PhD thesis, École Normale Supérieure, 2023.

[16] Rudy Morel, Stéphane Mallat, and Jean-Philippe Bouchaud. Path shadowing monte carlo. *Quantitative Finance*, 24(9):1199–1225, 2024.

[17] Rudy Morel, Gaspar Rochette, Roberto Leonarduzzi, Jean-Philippe Bouchaud, and Stéphane Mallat. Scale dependencies and self-similar models with wavelet scattering spectra. *Applied and Computational Harmonic Analysis*, 75:101724, 2025.

[18] Stephen A Ross. The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13(3):341–360, 1976.

[19] Jonathon Shlens. A tutorial on principal component analysis. *arXiv.org*, 2014.

[20] Ali Siahkoohi, Rudy Morel, Randall Balestriero, Erwan Allys, Grégory Sainton, Taichi Kawamura, and Maarten V de Hoop. Martian time-series unraveled: A multi-scale nested approach with factorial variational autoencoders. *arXiv preprint arXiv:2305.16189*, 2023.

[21] Ali Siahkoohi, Rudy Morel, V Maarten, Erwan Allys, Grégory Sainton, and Taichi Kawamura. Unearthing insights into mars: Unsupervised source separation with limited data. In *International Conference on Machine Learning*, pages 31754–31772. PMLR, 2023.

[22] R. Tsay. Principal component analysis and factor models. 2005.

[23] Gilles Zumbach and Paul Lynch. Heterogeneous volatility cascade in financial markets. *Physica A-statistical Mechanics and Its Applications*, 2001.