Proceedings of ICFTBA 2026 Symposium: Strategic Management & Business Analytics: Data-Driven Performance Models
DOI: 10.54254/2754-1169/2026.LD39216

A Survey of Matrix Decomposition Methods in Financial Data Analysis

Ruihan Tang

College of Arts and Science, New York University, New York, USA
ruihantang0920@gmail.com

Abstract. This paper reviews the application of matrix decomposition techniques in financial data analysis, with particular emphasis on singular value decomposition, principal component analysis, non-negative matrix factorization, and robust principal component analysis. High-dimensional financial datasets, such as stock return matrices, bond yield curves, and high-frequency trading records, often contain substantial noise and hidden factor structures. Matrix decomposition techniques address these difficulties by transforming complex financial datasets into lower-dimensional representations. Drawing on classical studies of low-rank approximation, principal component analysis, factor models, covariance estimation, asset pricing, and portfolio theory, this survey shows that matrix decomposition methods serve several major functions in financial analysis: extracting latent risk factors, improving high-dimensional covariance matrix estimation, supporting asset pricing and portfolio optimization, and separating systematic risks from idiosyncratic shocks. In yield curve analysis, principal components can be economically interpreted as level, slope, and curvature factors. In risk management, low-rank structures help identify common risk drivers, while residual components reflect asset-specific shocks. However, the review also highlights several limitations, including market non-stationarity, weak economic interpretation of statistical factors, and limited capacity to model tail risk. Overall, matrix decomposition provides a useful bridge between mathematical dimension reduction and practical financial modeling, while future research should further develop dynamic, nonlinear, and economically interpretable decomposition frameworks.

Keywords: Matrix Decomposition, Financial Data Analysis, Principal Component Analysis, Factor Models, Covariance Estimation

1. Introduction

Financial markets produce increasingly large and complex datasets, including stock returns, bond yield curves, portfolio holdings, and high-frequency trading records. These datasets are often high-dimensional, noisy, and strongly correlated across assets and time. In many financial contexts, the observed movements of a large number of assets are not fully independent but are driven by a smaller number of common market, sectoral, or macroeconomic factors. Therefore, an important task in financial data analysis is to extract low-dimensional structures from high-dimensional observations while reducing the influence of noise.

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2754-1169/2026.LD39216

Matrix decomposition methods provide a mathematical framework for this task. Singular value decomposition and low-rank approximation compress data while preserving essential information, principal component analysis (PCA) identifies dominant variation directions, non-negative matrix factorization (NMF) improves interpretability for non-negative financial data, and robust PCA separates common structures from abnormal components [1-4]. In financial applications, these tools are closely connected to statistical factor models and factor number selection, as well as to large-dimensional covariance matrix estimation [5, 6].

Recent research has extended matrix decomposition in finance beyond classical dimension reduction toward nonlinear factor extraction, high-dimensional covariance estimation, and dynamic or sparse factor modeling [7-15]. These developments increasingly connect traditional decomposition methods with machine learning and modern asset-pricing applications [16].

This paper reviews matrix decomposition techniques used in financial data analysis, with a focus on singular value decomposition, principal component analysis, non-negative matrix factorization, and robust principal component analysis. It discusses the mathematical foundations of these techniques and explains their practical roles in factor extraction, covariance matrix estimation, asset pricing, portfolio optimization, yield curve analysis, and risk management. The purpose of this survey is to connect matrix decomposition theory with financial modeling and to provide a reference point for future research on dynamic, nonlinear, and interpretable decomposition methods.

2. Mathematical foundations

2.1. Singular value decomposition

Financial datasets are often represented as matrices, where rows correspond to observations and columns represent financial variables. Because these matrices are usually high-dimensional and noisy, matrix decomposition methods are used to extract lower-dimensional structures while preserving important information. Singular value decomposition (SVD) provides a fundamental tool for this purpose [1]. For any real matrix $A \in R^{(m \times n)}$, SVD is given by:

$$A = U \Sigma V^{\top} \tag{1}$$

Where U and V are orthogonal matrices, and Σ is a diagonal matrix. The diagonal entries of Σ are singular values, usually ordered as $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r \geq 0$. These values are usually arranged in descending order, with larger values representing components that explain more variation in the original data. In financial applications, truncated SVD is commonly applied by retaining only the first k largest singular values:

$$A_k = U_k \Sigma_k V_k^{\top} \tag{2}$$

Here, A_k is the rank-k approximation of the original matrix A. The theory behind this approximation is the Eckart-Young theorem, which states that truncated SVD provides the optimal low-rank approximation under the Frobenius norm [1].

DOI: 10.54254/2754-1169/2026.LD39216

$$A_k = \arg\min_{\text{rank}(B)\le k} \|A - B\|_F \tag{3}$$

In financial data analysis, this result supports the use of SVD for dimension reduction and noise filtering. The larger singular values may correspond to shared market movements, while the remaining components can be treated as less informative variation.

2.2. Principal component analysis

PCA is closely related to SVD but focuses on the covariance structure of financial variables. Since asset returns often exhibit strong correlations, PCA transforms original variables into a smaller number of principal components that capture the largest sources of variation [2]. Let $X \in R^{(T\times N)}$ be a centered data matrix, where T is the number of observations and N is the number of assets or financial variables. The sample covariance matrix is:

$$\hat{\Sigma} = \frac{1}{T} X^T X \tag{4}$$

Here, Σ measures the joint movement of the variables in X. PCA performs eigendecomposition on the covariance matrix:

$$\hat{\Sigma} = P \Lambda P^\top \tag{5}$$

In this case, the columns of P represent the principal component directions, and the diagonal elements of Λ are eigenvalues. Each eigenvalue shows how much variance can be attributed to the corresponding principal component. The proportion of variance explained by the first k principal components is calculated as:

$$EVR(k) = \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{N} \lambda_i} \tag{6}$$

In financial analysis, this measure helps determine the number of factors required to describe market movements. PCA is also mathematically connected with SVD, since the principal component directions can be obtained from the singular vectors of the data matrix.

2.3. Covariance matrix and high-dimensional estimation

The covariance matrix is a fundamental component in financial modeling because it describes the co-movement of asset returns and supports portfolio optimization, risk management, and factor modeling. For N assets, the covariance between asset i and asset j is denoted as

DOI: 10.54254/2754-1169/2026.LD39216

$$\Sigma_{ij}=\mathrm{Cov}(r_i,r_j) \tag{7}$$

For a sample of return vectors, the sample covariance matrix can be estimated by

$$\hat{\Sigma}=\frac{1}{T\text{-}1}\sum_{t=1}^{T}(r_t-\bar{r})(r_t-\bar{r})^{\top} \tag{8}$$

In the equation above, r_t is the vector of asset returns at time t, while $\bar{r}$ is the sample average return vector. However, in high-dimensional financial datasets, covariance estimation become challenging when the number of assets is close to or larger than the number of observations. In such cases, sample covariance matrices may be severely distorted.

This limitation motivates the use of regularized and factor-based covariance estimators in financial applications [6, 8, 12, 14]. The Marchenko-Pastur law gives a theoretical threshold for distinguishing noise-driven eigenvalues:

$$\lambda_{\pm}=\sigma^2\left(1\pm\sqrt{\frac{N}{T}}\right)^2 \tag{9}$$

In a pure noise environment, sample eigenvalues are expected to remain within this region. Eigenvalues that exceed the upper threshold may reveal actual common factors rather than noise. This provides a justification for applying matrix decomposition techniques to separate signal from noise.

2.4. Non-negative matrix factorization

Unlike SVD and PCA, which allows positive and negative components, NMF imposes non-negativity constraints and produces an addictive representation, This property is useful for financial data that are naturally non-negative, such as transaction volumes, numbers of trades, portfolio weights, and holdings data [3]. Given a non-negative matrix V, NMF seeks an approximation of the form

$$V \approx WH \tag{10}$$

Where $W \geq 0$, $H \geq 0$. Here, V is the original non-negative data matrix, whereas W and H are lower-dimensional non-negative factor matrices. The non-negativity constraint enables the extracted components to be interpreted as , additive patterns. In financial applications, NMF can identify interpretable structures in transaction activity, portfolio composition, and investor holdings. Compared with PCA, NMF may provide clearer economic interpretation when the underlying financial variables are non-negative.

DOI: 10.54254/2754-1169/2026.LD39216

3. Core methods

Matrix decomposition techniques are widely applied in finance through factor modeling, covariance estimation, and risk decomposition. This section introduces their main methodological implementations.

3.1. Statistical factor model

Building on the low-rank structure of SVD and PCA, statistical factor models describe the common variation in financial returns through a small number of latent factors. Instead of modeling each asset independently, factor models assume that asset returns consist of systematic components and idiosyncratic errors [5]. Let $X \in R^{(T \times N)}$ be a return matrix, where T is the number of time observations and N is the number of assets. The statistical factor model can be written as

$$X = FB^{\top} + E \tag{11}$$

F represents the factor realization matrix, B represents the factor loading matrix, and E denotes the idiosyncratic residual matrix. This framework is closely connected with approximate factor models, where high-dimensional financial variables are explained by a limited number of latent factors [5]. Using truncated SVD, the return matrix can be approximated as:

$$X \approx U_k \Sigma_k V_k^{\top} \tag{12}$$

This gives a natural estimate of the factor model by setting

$$F = U_k \Sigma_k, \ B = V_k \tag{13}$$

The strength of this approach is that factors are extracted directly from data rather than predefined economic variables. Therefore, SVD- and PCA-based factor models are useful when underlying risk factors are unknown or difficult to observe [7, 13, 15].

3.2. Factor number selection

After extracting principal components or singular vectors, determining the appropriate number of factors is essential. Retaining too few factors may lose important information, while excessive factors may capture noise. A common approach is the scree plot, which identifies the "elbow point" where additional components provide limited explanatory power. A more formal method is the Bai-Ng information criterion for approximate factor models [5]:

$$IC(k) = \ln V(k, \hat{F}_k) + k \cdot g(N, T) \tag{14}$$

DOI: 10.54254/2754-1169/2026.LD39216

In this case, $V(k,\widehat{F}k)$ is the variance of the residuals after accounting for k factors, while $g(N,T)$ is a penalty function depending on the number of assets and observations. The chosen number of factors is the value of k that minimizes the information criterion. This technique is useful when working with high-dimensional financial data because it balances goodness of fit with model complexity.

3.3. Large-dimensional covariance estimation

The covariance matrix is fundamental in portfolio optimization and risk management. However, when the number of assets is large relative to observations, sample covariance estimates become unstable.

Shrinkage methods improve covariance estimation by combining the sample covariance matrix with a structured target matrix. Ledoit and Wolf propose a linear shrinkage estimator [6]:

$$\hat{\Sigma}_{LW}=(1-\alpha)\hat{\Sigma}+\alpha\mu I,\ \ \mu=\frac{1}{N}\,tr(\hat{\Sigma}) \tag{15}$$

where $\hat{\Sigma}$ is the sample covariance matrix, I is the identity matrix, and α is the shrinkage intensity. The term μI provides a stable target matrix.

Another important technique is factor-structured covariance estimation:

$$\Sigma=B\Sigma_F B^{\top}+\Sigma_\varepsilon \tag{16}$$

In this case, $B\Sigma_F B^T$ represents the covariance matrix generated by common factors, while $\Sigma\varepsilon$ represents idiosyncratic risk. By reducing the number of estimated parameters, factor-structured covariance models provide more stable solutions for large-scale portfolio construction and risk management.

3.4. Robust PCA

Although PCA effectively extracts common factors, it can be sensitive to outliers and extreme observations. Robust PCA extends traditional PCA by separating systematic structures from abnormal components [4]. The basic decomposition is

$$M=L+S \tag{17}$$

Here, M is the observed matrix, L is a low-rank matrix, and S is a sparse matrix. A common optimization formulation is

$$\min_{L,S}\|L\|_*+\lambda\|S\|_1\ \text{s.t. } L+S=M. \tag{18}$$

DOI: 10.54254/2754-1169/2026.LD39216

In this case, $\|L\|_*$ is the nuclear norm, which encourages L to be low-rank, whereas $\|S\|_1$ encourages S to be sparse. In finance, L can represent systematic market movements, while S captures abnormal returns, shocks, or asset-specific events. Therefore, robust PCA has applications in risk decomposition, anomaly detection, and stress testing analysis.

4. Application in finance

Matrix decomposition techniques have been widely applied in finance to extract latent structures, estimate risk factors, and improve financial modeling. This section discusses four major applications: asset pricing, portfolio optimization, yield curve analysis, and risk management.

4.1. Asset pricing and factor models

Asset pricing is one of the most important applications of matrix decomposition techniques. Standard asset pricing models assume that asset returns are driven by several systematic risk factors. In this regard, PCA- and SVD-based factor models provide a statistical approach to extracting latent factors from large return matrices.

By decomposing return data into common components and residual risks, matrix decomposition methods help identify hidden sources of return variation across assets. Unlike predefined economic factors, statistical factors are extracted directly from observed data, which is useful when the underlying drivers of asset returns are unknown.

Recent studies extend this idea through nonlinear representation methods, such as autoencoder-based asset pricing models, which extract latent factors beyond linear PCA structures [7]. However, these statistically generated factors may lack clear economic interpretation. Therefore, an important challenge in factor-based asset pricing is balancing predictive performance with economic meaning [16].

4.2. Portfolio optimization

Portfolio optimization relies heavily on accurate covariance estimation because portfolio risk is determined by the covariance structure of asset returns. However, in high-dimensional markets, sample covariance matrices often become unstable due to limited observations.

Matrix decomposition techniques solve this problem by reducing covariance estimation complexity through factor structures. Instead of directly estimating all pairwise asset correlations, factor-based covariance models estimate common factor risks and idiosyncratic components separately.

Ding, Li, and Zheng show that statistical factor models improve high-dimensional minimum variance portfolio estimation by providing more stable covariance structures [8]. Similarly, nonlinear shrinkage and dynamic covariance estimation methods further improve covariance estimation in large-scale portfolio applications [10, 12]. Therefore, matrix decomposition contributes to portfolio optimization by improving covariance stability, reducing estimation errors, and enhancing portfolio robustness.

4.3. Yield curve analysis

Yield curve modeling is a typical and straightforward example of PCA applications in finance. A yield curve represents interest rates across different maturities, and movements across maturities are

DOI: 10.54254/2754-1169/2026.LD39216

usually highly correlated. PCA reduces the high-dimensional yield curve into several interpretable factors.

The three most commonly identified components are level, slope, and curvature. The level factor represents parallel shifts in interest rates across maturities, the slope factor captures changes between short-term and long-term rates, and the curvature factor describes movements in medium-term maturities.

Because these components have clear economic interpretations, PCA provides both dimensional reduction and financial meaning in fixed-income analysis. Recent research continues to apply PCA for yield curve scenario generation and relative value analysis in bond markets [9].

4.4. Risk management

Risk management requires separating common market movements from asset-specific shocks. This distinction is closely related to the low-rank structure underlying matrix decomposition. Robust PCA provides an effective framework by decomposing financial data into a low-rank component and a sparse component. The low-rank matrix captures systematic market movements, while the sparse matrix identifies abnormal shocks like corporate events, liquidity shocks, or jumps during the period of crises [4]. Recent developments further illustrate the importance of dynamic and high-dimensional risk models, where decomposition-based approaches help update dependence structures and improve risk estimation [12, 14].

Overall, matrix decompositions help in risk management by determining common risk factors and differentiating signal from noise.

5. Challenges and future directions

Even though matrix decomposition has become increasingly popular in finance applications, there are still several obstacles to its further implementation. First of all, these are issues associated with high dimensionality, non-stationarity, nonlinearity, poor interpretability, and heavy tails.

The problem of high dimensionality is especially important since financial data sets may have a large number of assets compared to the number of observations, which makes covariance matrices and eigenvectors obtained based on their decomposition extremely noisy. The issue of high-dimensional portfolios emphasizes the necessity to apply more robust techniques in such cases [8].

Also, the non-stationarity of financial data is evident. Different market structures and relations between returns may be observed during normal times, crises, or various regimes of the economy. Dynamic covariance estimates are an example of overcoming this issue [12].

Finally, classical techniques, such as PCA or SVD, are predominantly linear, however, nonlinear relations can be observed in finance. Nonlinear dimensionality reduction and autoencoder-based asset-pricing models prove that point [7].

6. Conclusion

This paper has presented an overview of matrix decomposition techniques and their uses in financial data analytics. The paper started with the presentation of the mathematical basics of singular value decomposition, principal component analysis, covariance matrices, and non-negative matrix factorization. Then it moved on to the discussion of ways to extend those techniques to statistical factor models, factor number determination, high-dimensional covariance matrix estimation, and robust principal component analysis.

DOI: 10.54254/2754-1169/2026.LD39216

From this literature survey, one can conclude that there are many ways to use matrix decomposition techniques in different domains in finance. Matrix decomposition is useful in asset pricing, where it allows one to find latent factors and is a complement to traditional economic factor models. Matrix decomposition improves the estimation of covariance matrix and helps decrease its instability due to high dimensionality in portfolio optimization. In yield curve studies, the PCA technique summarizes movements along the maturities using level, slope, and curvature factors. Finally, in risk management, decomposition techniques allow one to discriminate between systematic and idiosyncratic movements in the market.

However, there are still a number of shortcomings. First, both traditional SVD and PCA are linear models and do not necessarily cover the complex nonlinear relationships in the finance field. Second, factor structures can be regime-dependent, implying that any static decomposition technique would be less reliable. Third, statistical factors cannot always be interpreted from an economic perspective, while covariance models lack the ability to handle tail risks.

On the whole, matrix decomposition is a vital method in the field of financial data analysis due to its usefulness in reducing dimensions, uncovering hidden factors, estimating covariances, and managing risks.

References

[1] Eckart, C., & Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika, 1*(3), 211-218. doi:10.1007/BF02288367

[2] Jolliffe, I. T. (2002). *Principal component analysis* (2nd ed.). Springer, New York.

[3] Lee, D. D., & Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature, 401*, 788-791. https://doi.org/10.1038/44565

[4] Candes, E. J., Li, X., Ma, Y., & Wright, J. (2011). Robust principal component analysis? *Journal of the ACM, 58*(3), 1-37. https://doi.org/10.1145/1970392.1970395

[5] Bai, J., & Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica, 70*(1), 191-221. https://doi.org/10.1111/1468-0262.00273

[6] Ledoit, O., & Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis, 88*(2), 365-411. https://doi.org/10.1016/S0047-259X(03)00096-4

[7] Gu, S., Kelly, B., & Xiu, D. (2021). Autoencoder asset pricing models. *Journal of Econometrics, 222*(1), 429-450. DOI: 10.1016/j.jeconom.2020.07.009

[8] Ding, Y., Li, Y., & Zheng, X. (2021). High dimensional minimum variance portfolio estimation under statistical factor models. *Journal of Econometrics, 222*(1), 502-515. DOI: 10.1016/j.jeconom.2020.07.013

[9] Oprea, A. (2022). The use of principal component analysis (PCA) in building yield curve scenarios and identifying relative-value trading opportunities on the Romanian government bond market. *Journal of Risk and Financial Management, 15*(6), 1-37. https://ideas.repec.org/a/gam/jjrfmx/v15y2022i6p247-d829362.html

[10] Bongiorno, C., & Challet, D. (2023). Non-linear shrinkage of the price return covariance matrix is far from optimal for portfolio optimization. *Finance Research Letters, 52*, 103383. DOI: 10.1016/j.frl.2022.103383

[11] Fang, P., Gao, Z., & Tsay, R. S. (2023). Supervised kernel principal component analysis for forecasting. *Finance Research Letters, 58*, 104292. DOI: 10.1016/j.frl.2023.104292

[12] Jiang, B., Liu, C., & Tang, C. Y. (2024). Dynamic covariance matrix estimation and portfolio analysis with high-frequency data. *Journal of Financial Econometrics, 22*(2), 461-491. https://doi.org/10.1093/jjfinec/nbad003

[13] Bagnara, M. (2024). Asset pricing and machine learning: A critical review. *Journal of Economic Surveys, 38*(1), 27-56. https://doi.org/10.1111/joes.12532

[14] Daniele, M., Pohlmeier, W., & Zagidullina, A. (2025). A sparse approximate factor model for high-dimensional covariance matrix estimation and portfolio selection. *Journal of Financial Econometrics, 23*(1), nbae017. https://doi.org/10.1093/jjfinec/nbae017

[15] Nechvatalova, L. (2025). Autoencoder asset pricing models and economic restrictions: International evidence. *International Review of Financial Analysis, 107*, 104642. https://doi.org/10.1016/j.irfa.2025.104642

[16] Giglio, S., Kelly, B., & Xiu, D. (2022). Factor models, machine learning, and asset pricing. *Annual Review of Financial Economics, 14*, 337-368. https://doi.org/10.1146/annurev-financial-101521-104735