

High-Dimensional Covariance Denoising and Geometric Robustness: A Statistical Framework for Portfolio Stability

Yuxuan Fang

Department of Jiangxi University of Finance and Economics, Nanchang, China

fyx12345622@qq.com

Abstract. In high-dimensional financial settings, covariance matrix estimation suffers from the curse of dimensionality and parameter instability, which substantially limit the practical effectiveness of the Markowitz mean–variance framework. To address this issue, this paper constructs a "covariance estimation - portfolio strategy" collaborative framework, systematically examining the performance differences of four classic methods - sample covariance, shrinkage estimation, MCD robust estimation, and Tyler M-estimation in global minimum variance portfolios, mean-variance portfolios, and risk parity portfolios, to enhance the stability of investment portfolios in high-dimensional, noisy, and heavy-tailed environments. The empirical analysis covers three scenarios: ideal Gaussian distribution, contaminated data scenario, and resampling of the S&P 500 real market data, and uses estimation error, Sharpe ratio, maximum drawdown, Sortino ratio, and the time series stability of portfolio weights as evaluation metrics. Simulation experiments show that in the contaminated data environment, the estimation error of MCD and Tyler estimators is significantly lower than that of sample covariance and shrinkage estimation, demonstrating stronger robustness against outliers; real market backtesting further reveals that embedding robust covariance estimation into low-sensitivity optimization models such as risk parity and global minimum variance substantially reduces the time series fluctuations of portfolio weights and improves risk-adjusted returns, where the Risk Parity_Shrinkage and Risk Parity_Robust strategies outperform traditional benchmark portfolios in Sharpe ratio, maximum drawdown, and stability indicators. The research shows that the "estimator-strategy" adaptation framework is an effective paradigm for eliminating high-dimensional estimation errors and enhancing strategy resilience. This framework offers practical value for high-dimensional asset allocation, quantitative strategy development, and systemic risk prevention and mitigation.

Keywords: High-dimensional covariance estimation, Robust estimation, Portfolio optimization.

1. Introduction

Since Markowitz proposed the mean-variance model, modern portfolio theory has become the theoretical foundation for the allocation of financial assets ^[1]. However, with the explosive growth of the number of assets, the construction of investment portfolios is undergoing a paradigm shift towards "high dimensions", and in this environment, the reliable estimation of the covariance matrix is facing severe challenges ^[2]. Although methods such as shrinkage estimation and robust estimation have made positive progress in high-dimensional scenarios, the construction of portfolio stability still faces three core problems: the curse of dimensionality, error amplification, and poor adaptability to heavy-tailed distributions, which constitute the practical motivation for this study ^[3].

The statistical framework for portfolio stability has undergone a systematic evolution from the classical mean-variance framework to the modern robust optimization framework, and from low-dimensional covariance estimation to high-dimensional covariance denoising. Early studies mainly focused on the parameter estimation problem within the mean-variance framework, but it was found that estimation errors would make the theoretically optimal investment portfolio perform worse in practice than the equal-weight allocation strategy ^[4]. This discovery prompted researchers to seek breakthroughs in two directions: one is to develop more robust covariance estimation methods to reduce input noise, and the other is to directly embed robustness constraints in the optimization model to enhance the stability of the portfolio. Current research further focuses on covariance matrix denoising in high-dimensional scenarios and the collaborative design of "estimator-strategy", aiming

to systematically improve the robustness of the investment portfolio in complex market environments from both the estimation at the input end and the optimization at the output end.

The current research on the construction of investment portfolio robustness can be classified into three categories: shrinkage estimation, robust covariance estimation, and robust optimization models. In the field of shrinkage estimation, Ledoit and Wolf effectively balanced the estimation bias and variance in high-dimensional scenarios by linearly contracting a structured target matrix, but this method still has insufficient ability to handle noise in batch eigenvalues^[5]. In the field of robust covariance estimation, Tyler's M-estimator has distribution-free properties under elliptically symmetric distributions and is robust to heavy-tailed distributions, but its computational complexity is relatively high in high-dimensional conditions^[6]; Rousseeuw and Van Driessen developed the MCD algorithm, which raised the breakdown point to 50%, significantly enhancing the ability to resist outliers, but it faces high computational costs and the risk of getting stuck in local optima^[7]; Shi et al. combined Kendall tau with diagonal dominance structure and achieved the highest Sharpe ratio in an empirical study of 235 stocks, but the setting of diagonal dominance parameters was relatively sensitive^[8]; Zhu and Welsch introduced the regularized vine structure into robust covariance estimation, demonstrating good nonlinear relationship characterization ability in 50 US stocks, but the selection of the vine structure becomes difficult when the number of assets is large^[9]. In the field of robust optimization models, Choueifaty and Coignard constructed the maximum diversification portfolio, which effectively reduced the common risk exposure by maximizing the diversification ratio, but this method is sensitive to covariance estimation errors^[10]. Maillard et al. proposed the risk parity portfolio, where each asset contributes equally to the total risk, not relying on return predictions, thus having natural robustness against mean estimation errors, but the weight solution complexity is relatively high in high-dimensional conditions^[11].

Overall, methods such as shrinkage estimation, robust estimation, and RMT denoising have achieved remarkable results in their respective dimensions. However, most studies focus on improvements at a single level and there is relatively limited research on the systematic integration of robust covariance estimation and robust portfolio construction strategies. In particular, the applicable boundaries and performance differences of different "estimator-strategy" combinations are still lacking in systematic comparison and evaluation. Therefore, this paper aims to construct a collaborative framework that integrates multiple covariance estimation methods and portfolio strategies to evaluate the applicability and performance differences of different fusion paths in high-dimensional, noisy, and heavy-tailed environments. Specifically, this paper proposes a collaborative framework from the two levels of input estimation and portfolio strategies: at the input level, the estimation accuracy and robustness of four covariance estimation methods (sample covariance, shrinkage estimation, MCD robust estimation, and Tyler M-estimation) are systematically compared; in addition, the four estimation methods are respectively embedded in three types of investment portfolio strategies (global minimum variance portfolio, mean-variance portfolio, and risk parity portfolio) to form an "estimator-strategy" integration framework.

The main innovations and contributions of this work are as follows:

A "covariance estimation - portfolio strategy" collaborative adaptation framework was proposed, systematically exploring the integration paths of four estimators and three types of optimization models. It reveals the applicable boundaries of different matching patterns in high-dimensional environments from the dual perspectives of estimation bias and performance differentiation.

A three-step progressive empirical evaluation scheme was designed, integrating ideal distribution, contaminated data, and real market resampling. It was cross-validated from the dual dimensions of estimation accuracy and strategy robustness, and introduced the measurement of performance distribution dispersion to test the extrapolation stability.

The synergy mechanism between robust estimators and low-sensitivity optimization models was revealed, that is, the anti-deviation at the input end and the weight constraints at the optimization end jointly suppress error transmission. Empirical evidence shows that this coupling path can

simultaneously reduce estimation bias and weight fluctuations, providing a verifiable basis for estimator-strategy matching.

2. Related Work

2.1. Overall Structure of the Investment Portfolio Stability Framework

In high-dimensional financial data environments, the direct application of the traditional Markowitz mean-variance model faces severe challenges. When the ratio of asset dimensions to sample size is too large, the sample covariance matrix no longer exhibits consistency or invertibility. Minor estimation errors in the covariance matrix will be systematically amplified during the optimization process, resulting in extreme concentration or drastic fluctuations in weights. A single optimization model is unable to simultaneously address the dual problems of estimation bias and error amplification. Therefore, modern research has gradually formed a multi-module collaborative technical framework.

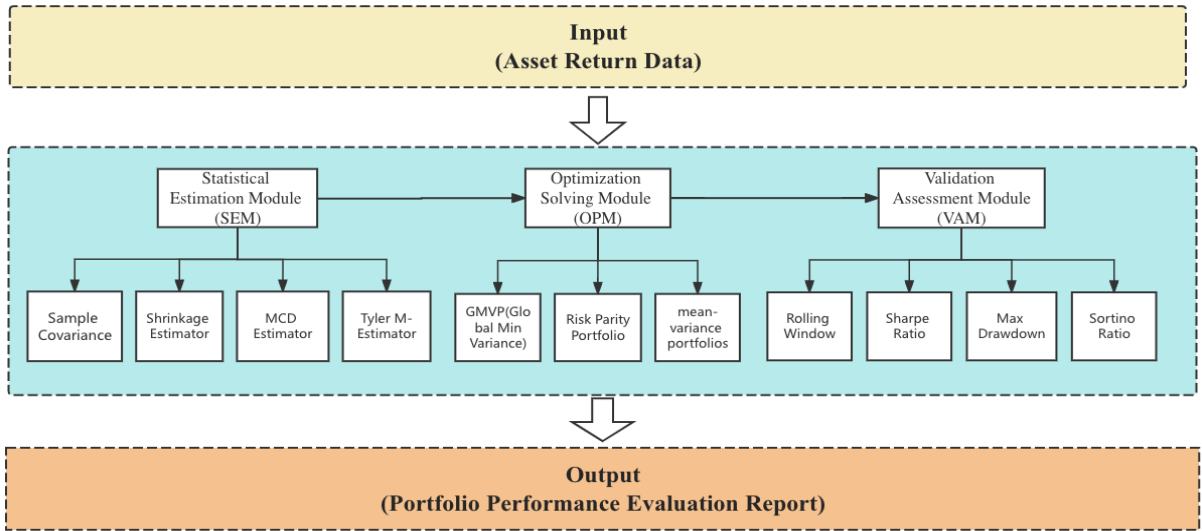


Figure 1. Multi-module Collaborative Technology Framework Diagram

As shown in Figure 1, a complete investment portfolio stability construction system typically consists of three core modules. The statistical estimation module is responsible for estimating the covariance matrix from asset return data, and the quality of its output directly affects all subsequent steps. This module needs to provide stable and reversible covariance estimation in high-dimensional, noisy, and heavy-tailed data environments. The optimization solution module, given the covariance matrix and expected returns, solves the optimal investment portfolio weights^[12]. The verification module uses methods such as rolling window backtesting to evaluate the out-of-sample performance of different strategies. There are clear data flows and collaborative relationships between these three modules.

2.2. Core Statistical Techniques for Portfolio Stability

2.2.1. Characteristics and Preprocessing of High-Dimensional Financial Data

The data of financial asset returns usually exhibit non-Gaussian characteristics such as heavy tails, skewness, and volatility clustering, and their distribution is time-varying. Traditional estimation methods based on the assumption of normality are difficult to effectively adapt to these situations^[13]. Moreover, in actual market data, outliers often occur due to extreme events or data recording errors, which can distort the estimation of the covariance matrix. To reduce the impact of these problems on estimation accuracy, common preprocessing methods include yield rate standardization and denoising, aiming to provide a relatively clean data basis for the robust estimation of the covariance matrix^[14].

2.2.2. Covariance estimators for portfolio stability

The estimation error of the covariance matrix is an important source of portfolio instability. To improve the estimation accuracy, researchers have proposed various robust covariance estimation methods. Among them, shrinkage estimation achieves a balance between bias and variance by linearly contracting a structured target matrix. Non-linear contraction further uses random matrix theory to differentiate and correct eigenvalues, thereby improving the estimation effect in high-dimensional environments. At the same time, robust estimation based on MCD enhances robustness to outliers by identifying and eliminating abnormal observations. Tyler M-estimation demonstrates strong adaptability to heavy-tailed financial data due to its distribution-free property.

2.2.3. The synergy between the input end estimation and the combined strategy

Relying solely on input end covariance denoising is not sufficient to ensure the stability of the investment portfolio. It also needs to be combined with a low-sensitivity portfolio optimization strategy. The global minimum variance portfolio only depends on the covariance matrix and is insensitive to mean estimation errors; the risk parity strategy reduces the reliance on return predictions by balancing the risk contributions of each asset, thus having strong robustness. When robust covariance estimators such as MCD and Tyler are embedded in the above portfolio strategies, they can not only reduce the impact of input data contamination but also suppress weight fluctuations, thereby improving the out-of-sample stability of the portfolio. In contrast, even with robust covariance estimation, the mean-variance model offers only limited stability improvements due to its heavy reliance on return predictions.

3. The core technical path and optimization strategies for high-dimensional covariance denoising and geometric robustness construction

This paper proposes a framework for high-dimensional covariance denoising and geometric robustness from three levels: statistical estimation, optimization solution, and structure construction. By coordinating multi-dimensional robustness methods, it enhances the stability of the investment portfolio.

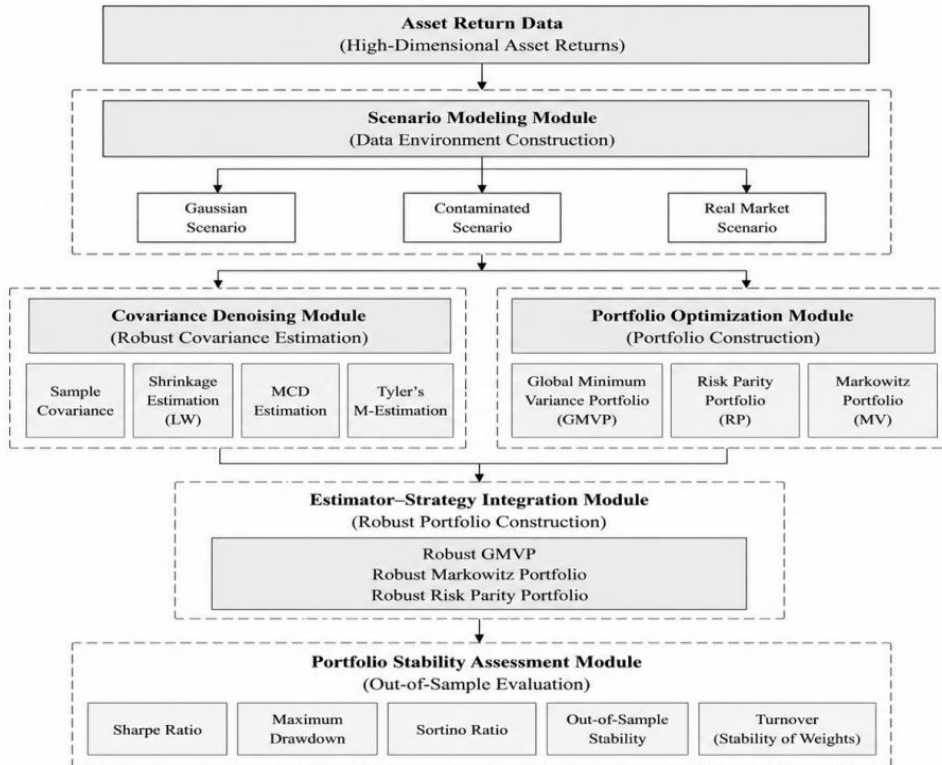


Figure 2. Technical roadmap for portfolio stability

3.1. Scenario Modeling of Portfolio Stability

To systematically evaluate the robustness of different covariance estimation methods and portfolio construction strategies, this study has constructed three types of financial market scenarios:

3.1.1. Idealized Gaussian Distribution Data

This scenario assumes that the asset return rate strictly follows a multivariate Gaussian distribution, which is used to examine the benchmark performance of each estimator when the data is relatively clean, and provides a reference benchmark for the subsequent scenario where outliers are added. The parameter settings are as follows: the number of assets $N=15$, the length of the time series $T=300$. This scale can reflect the complexity of multi-asset allocation and ensure that the sample size is sufficient to support the estimation of the covariance matrix. The true mean vector is composed of 15 positive numbers ranging from 0.0004 to 0.0017, simulating the heterogeneity of expected returns for different assets. The true covariance matrix Σ_{true} is generated using a multi-factor structure, including 3 common factors and asset-specific variances, and ensuring that the matrix is positive definite. Based on these parameters, an ideal dataset $R_{clean} \sim N(\mu_{true}, \Sigma_{true})$ without contamination is generated.

3.1.2. Contaminated Data Scenario: Non-normal Data Set with Outliers

In the actual financial market, asset returns often suffer from "contamination" due to extreme events, data recording errors, or non-normal distribution characteristics, manifesting as outliers or abnormal observations. To simulate this reality, three types of anomalies were artificially injected into the ideal Gaussian data set R_{clean} : mean shift (to simulate global disturbances caused by systemic shocks), single asset extreme values (to simulate local disturbances caused by specific events of a company), and complete outlier observations (to simulate overall distortion caused by data source errors or extreme market scenarios), contaminating approximately 10% of the observations, and constructing the polluted data set $R_{contaminated}$ to test the tolerance and robustness of each estimator to outliers.

3.1.3. Typical Market Data Scenario: Resampled Data Set Based on Real Historical Data

To verify the performance of different methods in the real market, based on the historical data of the S&P 500 constituents from January 2000 to September 2020, a financial data resampling method was used to generate 50 sub-data sets, each containing 20 assets and 504 trading days, to simulate diverse market initial conditions and asset portfolios, avoiding the dependence of the results on a single historical path. This scenario does not rely on artificially generated assumptions and directly reflects the complex characteristics of the real market, used to evaluate the overall performance of each strategy in the real market environment.

3.2. Adaptive Optimization Strategy for Portfolio Stability

In high-dimensional financial data environments, the estimation error of the covariance matrix is the main cause of the instability of the investment portfolio. This section introduces four representative covariance estimation methods, aiming to denoise the covariance matrix and enhance robustness on the data at the input stage.

3.2.1. Classical Sample Covariance Estimation

The sample covariance matrix is the most fundamental estimator, defined as:

$$\hat{\Sigma}_{sample} = \frac{1}{T-1} \sum_{t=1}^T (R_t - \bar{R})(R_t - \bar{R})^T \quad (1)$$

Where R_t is the return vector at time t , and $\bar{R}$ is the sample mean vector. This estimator has excellent properties when the data is low-dimensional and follows a normal distribution. However, in high-dimensional cases or when there are outliers, the estimation accuracy sharply decreases, leading to unstable subsequent optimization results.

3.2.2. Shrinkage Estimator

To reduce the estimation variance of the sample covariance, the shrinkage estimator proposed by Ledoit and Wolf linearly shrinks the sample covariance towards a structured target matrix (such as the identity matrix or a constant correlation coefficient matrix):

$$\hat{\Sigma}_{shrink} = \delta F + (1 - \delta) \hat{\Sigma}_{sample} \quad (2)$$

Where F is the target matrix, $\delta \in [0, 1]$ is the shrinkage strength, and it is determined through an analytical method to minimize the mean square error. The shrinkage estimator can achieve a balance between bias and variance, and performs robustly in high-dimensional environments. However, its ability to resist outliers is still limited because outliers affect both the sample covariance and the target matrix simultaneously.

3.2.3. Robust Estimation with High Breakdown Point Based on MCD

The Minimum Covariance Determinant (MCD) estimator searches for a subset such that the determinant of the sample covariance matrix of this subset is minimized. Its core idea is to eliminate abnormal observations that distort the covariance structure. The breakdown point of the MCD estimator can reach $(n-h)/n$, and theoretically can withstand up to 50% of contaminated data. This study employs the FAST-MCD algorithm for efficient computation.

3.2.4. Tyler's M Estimator

Tyler's M estimator is a robust estimator with distribution-free properties under elliptical symmetric distributions. It solves the following fixed-point equation through iterative weighting:

$$\hat{\Sigma}_{Tyler} = \frac{p}{n} \sum_{t=1}^n \frac{(R_t - \mu)(R_t - \mu)^T}{(R_t - \mu)^T \hat{\Sigma}_{Tyler}^{-1} (R_t - \mu)} \quad (3)$$

where μ is usually replaced by a robust location estimator. This estimator is insensitive to the actual distribution shape of the data, and is particularly suitable for financial return sequences with heavy-tailed characteristics, having strong geometric robustness.

3.3. Framework for Integration of Covariance Estimation and Portfolio Strategy

A single input-end denoising is not sufficient to ensure the stability of the final investment portfolio. Robust mechanisms need to be introduced at the optimization-solving end and the portfolio structure end. This section constructs a "estimator-strategy" integration framework, embedding the above four estimators into various portfolio construction methods.

3.3.1. Fusion of Global Minimum Variance Portfolio and Robust Estimation

The Global Minimum Variance Portfolio (GMVP) relies solely on the covariance matrix and is insensitive to mean estimation errors. By combining GMVP with a robust covariance estimator, a "Robust GMVP" that is highly immune to outliers can be obtained. Its weights are obtained by solving a quadratic programming problem:

$$\min_w w^T \hat{\Sigma} w \quad \text{s.t.} \quad w^T 1 = 1, w_i \geq 0 \quad (4)$$

The classical Markowitz mean-variance model uses both mean and covariance, and is more sensitive to estimation errors. This paper adopts the "Robust Mean-Variance Portfolio" strategy, substituting the robust estimated mean and covariance into the optimization problem and applying non-negative constraints.

3.3.2. Fusion of Risk Parity Strategy and Robust Estimation

The Risk Parity strategy aims to make the contribution of each asset to the total risk of the portfolio equal. Its weights satisfy:

$$w_i \cdot (\hat{\Sigma} w)_i = w_j \cdot (\hat{\Sigma} w)_j, \quad \forall i, j \quad (5)$$

This strategy has significant advantages over the traditional mean-variance framework: it does not rely on return predictions, avoids the largest source of estimation error, reducing the sensitivity of the strategy to input parameters; it is naturally diversified, avoiding excessive concentration of portfolio weights in a few assets at the risk contribution level, thereby reducing the risk of extreme losses. However, its calculation depends on the accurate estimation of the covariance matrix. By introducing the robust covariance estimator (such as MCD, Tyler) into the risk parity iterative solution process, the "Robust Risk Parity" strategy can be obtained. This strategy can retain structural robustness while resisting the contamination of input data.

4. Experimental Results and Comparative Analysis

4.1. Experimental Environment and Evaluation Indicators

This study conducted simulation experiments and portfolio backtesting analyses using R version 4.2 and above. The experiments involved modules such as covariance estimation, portfolio optimization, and performance evaluation, and were mainly implemented using relevant toolkits for robust statistics, portfolio optimization, and backtesting analysis.

This section aims to evaluate the estimation performance of each estimator in ideal and contaminated environments. The estimation accuracy is measured by the following two indicators:

Mean estimation error:

$$Err_{\mu} = \|\hat{\mu} - \mu_{true}\|^2 \quad (6)$$

Covariance estimation error:

$$Err_{\Sigma} = \|\hat{\Sigma} - \Sigma_{true}\|_F \quad (7)$$

4.2. Experimental Results

4.2.1. Objective Quantitative Analysis

Table 1. Comparison Table of Errors of Four Estimation Methods

Estimator	Clean Mu Error	Clean Sigma Error	Contaminated Mu Error	Contaminated Sigma Error	Mu Error Ratio	Sigma Error Ratio
Sample	0.000217	0.0065	0.000451	0.0121	2.08	1.86
Shrinkage	0.000217	0.0063	0.000451	0.0112	2.08	1.78
Robust	0.000228	0.0073	0.000383	0.0078	1.68	1.07
Tyler	0.000259	0.007	0.000429	0.0074	1.66	1.06

Table 1 presents the estimation error performance of four covariance estimation methods in the ideal data and contaminated data environments. In the ideal Gaussian distribution environment, the error differences among the estimators are relatively small, indicating that the traditional sample covariance estimation and shrinkage estimation still have good statistical efficiency. As outliers are introduced, the performance differences of different estimation methods gradually become apparent.

The mean error and covariance error of sample covariance estimation and shrinkage estimation both show significant growth. Among them, the covariance error increase reaches 1.86 times and 1.78 times respectively. In contrast, the error increases of MCD and Tyler estimation remain relatively low, with the covariance error increase being only 1.07 times and 1.06 times respectively. This indicates that robust estimation methods can effectively mitigate the influence of abnormal observations on the covariance structure and maintain high estimation accuracy in noisy environments.

4.2.2. Subjective Qualitative Analysis

Figure 3 shows the mean estimation errors of four estimators under different data environments. It can be observed that in the ideal data scenario, the error levels of each method are close; after the introduction of contaminated data, the sample estimation and shrinkage estimation errors increase significantly, while the MCD and Tyler estimations have a smaller variation. This phenomenon indicates that the robust estimation methods can effectively identify and reduce the impact of

abnormal observations on the mean estimation, thereby maintaining a more stable estimation result. Figure 4 further shows the changes in covariance estimation errors. Compared with the mean errors, the covariance estimation errors are more sensitive to outliers. The sample covariance estimation and shrinkage estimation have a significant increase in error under contaminated data, while the MCD and Tyler estimations have a smaller variation. From the graphical representation, the robust estimators maintain a relatively consistent error level before and after contamination, indicating that they can more accurately restore the true correlation structure between assets and provide more reliable input information for subsequent portfolio optimization.

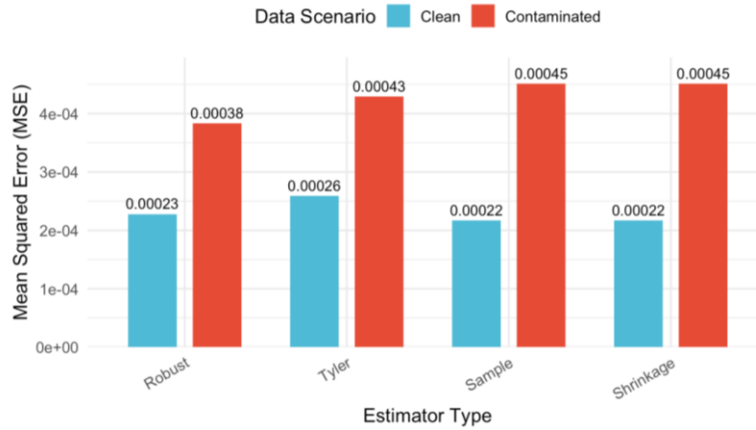


Figure 3. Comparison of Mean Estimation Error

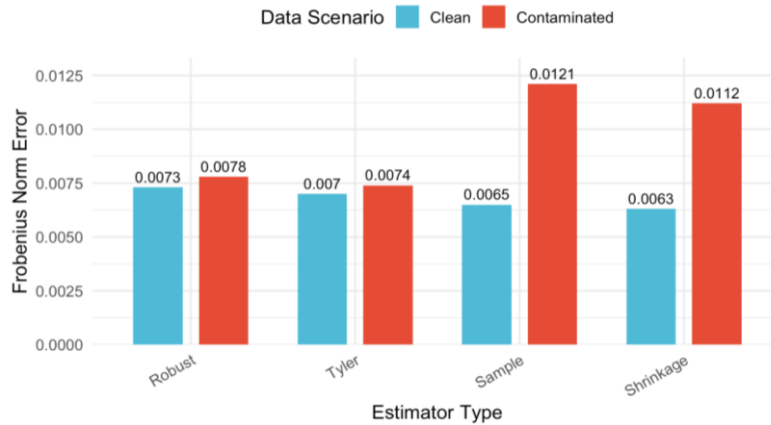


Figure 4. Comparison of Covariance Estimation Errors

4.3. Comprehensive Analysis

By integrating the quantitative and qualitative analysis results, it can be observed that the differences in performance among various covariance estimation methods are limited in an ideal environment. However, they exhibit significantly different robustness characteristics when there are outliers and heavy-tailed features. Both the MCD estimation and the Tyler estimation maintain a low error increase in the contaminated data environment, indicating that they can effectively mitigate the impact of abnormal observations on the covariance structure. In contrast, the sample covariance estimation and shrinkage estimation are more sensitive to data contamination, and the estimation errors significantly expand. These results indicate that outliers and deviations from the distribution in high-dimensional financial data have a significant impact on covariance estimation, while robust estimation methods can maintain a high estimation accuracy while improving the stability of the model. This provides a theoretical basis for embedding robust covariance estimation into the investment portfolio optimization model in the subsequent chapter. The fifth chapter will further examine the actual performance of the integration of different estimators and portfolio strategies in the real market data environment.

5. Comparative Analysis of Investment Portfolio Performance Based on Real Market Data

5.1. Dataset and Evaluation Indicators

To closely simulate the real investment scenario, this paper constructs a resampled dataset using historical data of the S&P 500 constituents, generating 50 independent sub-samples, each containing 20 assets and approximately two years of trading data. The backtesting strategy system is divided into three categories: basic strategies, including quintile stock selection strategies; robust fusion strategies, which incorporate samples, shrinkage, and MCD estimators into the global minimum variance portfolio (GMVP), mean-variance portfolio (Markowitz), and risk parity portfolio; enhanced strategies, including PCA factor risk parity, enhanced diversification strategies, and RSI momentum long-short strategies. At the same time, using the equal-weighted portfolio and the market index as benchmarks, the performance of different frameworks is comprehensively compared.

Backtesting indicators cover returns, risks, risk-adjusted returns, and tail risks, and the core indicators include: annualized return, annualized volatility, Sharpe ratio, maximum drawdown, downside risk indicators, etc.

5.2. Comprehensive Performance Comparison

Table 2. Summary of Portfolio Performance

	Sharpe ratio	max drawdown	annual return	annual volatility	Sortino ratio
RP_Shrinkage	1.42	10.00%	18.00%	15.00%	2.02
RP_Robust	1.37	11.0%	19.0%	15.0%	2.02
1/N	1.35	12.00%	19.00%	16.00%	2.06
RP_Sample	1.34	11.00%	18.00%	15.00%	1.98
RP_PCA	1.33	11.00%	17.00%	14.00%	1.99
GMVP_Robust	1.19	9.0%	15.0%	13.0%	1.80
GMVP_Shrinkage	1.16	9.00%	13.00%	13.00%	1.72
GMVP_Sample	1.14	9.00%	13.00%	13.00%	1.65
Quintile	0.97	16.00%	22.00%	24.00%	1.39
Enhanced_Div	0.94	9.00%	9.00%	10.00%	1.37
index	0.75	12.00%	11.00%	16.00%	1.08
Markowitz_Shrinkage	0.74	28.00%	21.00%	36.00%	1.08
Markowitz_Sample	0.73	28.00%	21.00%	36.00%	1.06
Markowitz_Robust	0.73	24.00%	24.00%	35.00%	1.11
RSI_LongShort	0.63	17.00%	11.00%	20.00%	0.94

Table 2 summarizes the average performance of each strategy on 50 resampled datasets. From the overall results, the risk parity-based strategies performed the best, among which the RP_Shrinkage strategy achieved the highest Sharpe ratio while maintaining a lower maximum drawdown, demonstrating a better risk-return balance capability. The GMVP series of strategies were slightly lower in terms of return level than the risk parity strategies, but had better volatility and drawdown control effects, showing a stronger defensive characteristic.

In contrast, the Markowitz series of strategies, although achieving a higher annualized return rate, had significantly higher volatility and maximum drawdown, resulting in relatively lower risk-adjusted returns. The results indicate that in a high-dimensional financial environment, combining robust covariance estimation with low-sensitivity portfolio strategies is more conducive to improving the overall stability of the investment portfolio.

The enhanced diversification strategy and the PCA risk parity strategy performed stably, lying between traditional strategies and robust fusion strategies, indicating that factor dimensionality reduction and diversification constraints can enhance robustness of the portfolio, but still fall short of the collaborative framework of "robust estimation + portfolio optimization". In terms of benchmarks, the Sharpe ratios of all robust portfolio strategies significantly exceeded the equal-weighted portfolio and the market index, indicating that the high-dimensional covariance denoising and portfolio framework proposed in this paper has certain advantages in risk-adjusted returns.

5.3. Comparison of Sharpe Ratio and Maximum Drawdown Dual Indicators

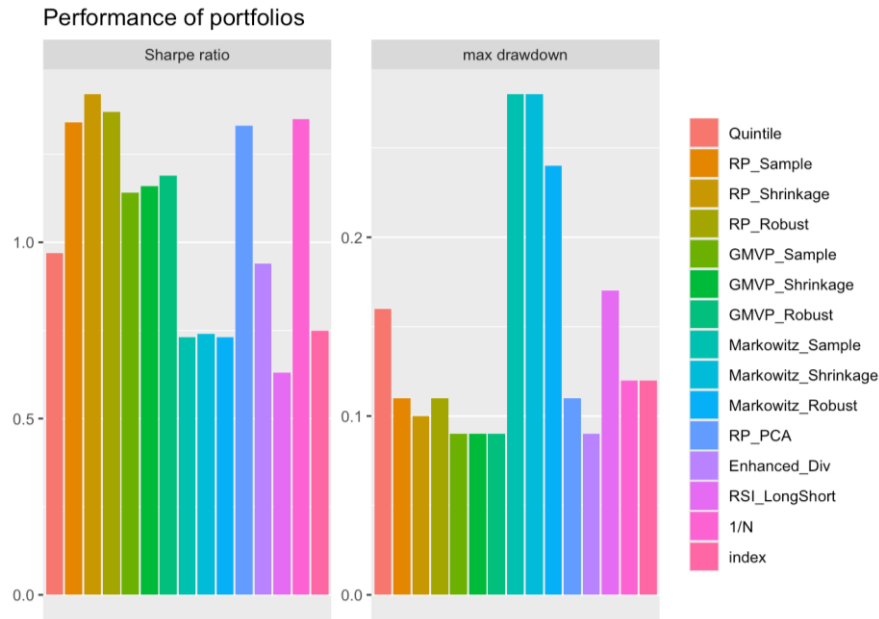


Figure 5. Comparison Bar Chart of Sharpe Ratio and Maximum Drawdown for Each Strategy

Figure 5 further compares the return-risk characteristics of each strategy. The risk parity and GMVP series of strategies generally have higher Sharpe ratios and lower maximum drawdowns. Among them, RP_Robust, RP_Shrinkage, and GMVP_Robust perform particularly well, indicating that robust covariance estimation can effectively improve the risk control ability of the portfolio. In contrast, the Markowitz series of strategies have higher returns but significantly higher drawdown levels, reflecting their sensitivity to parameter estimation errors. The results show that in high-dimensional financial data environments, reducing reliance on return predictions and combining with robust covariance estimation can help achieve more stable investment performance.

5.4. Test of Stability of Strategy Performance Distribution

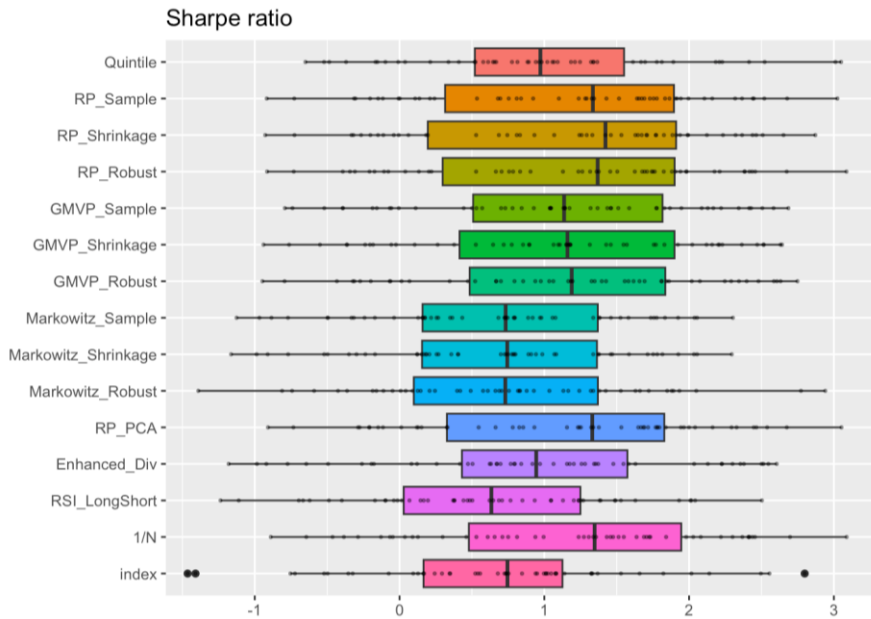


Figure 6. Box plot of Sharpe ratios for each strategy

Figure 6 presents the distribution of Sharpe ratios for each strategy across 50 resampled datasets. It can be observed that RP_Robust and RP_Shrinkage have higher median levels, and their box widths are narrower, indicating that they can maintain relatively stable performance in different sample

environments. In contrast, the distribution of traditional sample estimation strategies is wider and has more outliers, suggesting that they are more sensitive to changes in the market environment. The equal-weighted portfolio shows relatively stable performance but has a limited overall return level. Overall, the combination of robust covariance estimation and portfolio optimization models not only improves the average performance but also enhances the stability and consistency of the strategies in different market conditions.

6. Summary and Outlook

6.1. Summary

This paper conducts a systematic study on the issue of portfolio stability from the perspective of the synergy between high-dimensional covariance estimation and portfolio optimization. At the methodological level, a collaborative framework of "input-end covariance estimation + portfolio strategy construction" was established, and the performances of four methods - sample covariance, shrinkage estimation, MCD robust estimation, and Tyler M-estimation - in three portfolio strategies-global minimum variance portfolio, mean-variance portfolio, and risk parity portfolio - were compared. At the empirical level, the study was verified in three scenarios: ideal Gaussian distribution, contaminated data, and real market resampling. The main findings are as follows: At the estimation level, the errors of MCD and Tyler robust estimation under contaminated data are much lower than those of sample estimation and shrinkage estimation. At the portfolio level, embedding MCD or Tyler robust estimation into low-sensitivity optimization models such as risk parity and global minimum variance portfolio can simultaneously reduce covariance estimation errors and portfolio weight fluctuations. This indicates that robust estimators are not always the best in all strategies; their advantages are more likely to be manifested in low-sensitive portfolio models.

6.2. Outlook

This paper mainly studies the issue of portfolio stability based on the static covariance estimation framework. In the future, it can further combine dynamic covariance models and online learning methods to construct a robust estimation mechanism that can adapt to changes in market conditions. At the same time, the research can be extended to multi-market and multi-asset allocation scenarios to explore cross-market correlation structures and risk transmission characteristics. With the development of high-dimensional statistics and machine learning technologies, the collaborative design of covariance estimation and portfolio optimization still has broad research space.

References

- [1] Markowitz H. Portfolio selection. *The Journal of Finance*, 1952, 7(1): 77-91.
- [2] Li H, Liu L, Xia N, et al. High-dimensional expected utility portfolios under the spiked covariance model. *Finance Research Letters*, 2026, 93: 109637.
- [3] Bodnar T, Okhrin Y, Parolya N. Optimal shrinkage-based portfolio selection in high dimensions. *Journal of Business and Economic Statistics*, 2023, 41(1): 140-156.
- [4] Michaud R O. The Markowitz optimization enigma: Is 'optimized' optimal. *Financial Analysts Journal*, 1989, 45(1): 31-42.
- [5] Ledoit O, Wolf M. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 2004, 88(2): 365-411.
- [6] Tyler D E. A distribution-free M-estimator of multivariate scatter. *The Annals of Statistics*, 1987: 234-251.
- [7] Rousseeuw P J, Van Driessen K. A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, 1999, 41(3): 212-223.

- [8] Shi F, Shu L, Gu X. A robust latent factor model for high-dimensional portfolio selection. *Journal of Empirical Finance*, 2025, 83: 101623.
- [9] Zhu Z, Welsch R E. Robust dependence modeling for high-dimensional covariance matrices with financial applications. *The Annals of Applied Statistics*, 2018, 12(2): 1228-1249.
- [10] Choueifaty Y, Coignard Y. Toward maximum diversification. *The Journal of Portfolio Management*, 2008, 35(1): 40-51.
- [11] Maillard S, Roncalli T, Teiletche J. The properties of equally weighted risk contribution portfolios. *The Journal of Portfolio Management*, 2010, 36(4): 60-70.
- [12] Laloux L, Cizeau P, Potters M, et al. Random matrix theory and financial correlations. *International Journal of Theoretical and Applied Finance*, 2000, 3(3): 391-397.
- [13] Jay E, Soler T, Terreaux E, et al. Improving portfolios global performance using a cleaned and robust covariance matrix estimate. *Annals of Operations Research*, 2020: 1-24.
- [14] Vanni F, Hitaj A, Mastrogiamomo E. Enhancing portfolio allocation: A random matrix theory perspective. *Mathematics*, 2024, 12(9): 1389.