An Empirical Study of Robust Portfolios Based on Statistical Noise Reduction and Regularization Constraints

Zichun Fu

School of South China Agricultural University, Guangzhou 510642, China.

1956979929@qq.com

Abstract. This research paper discusses the issue of out-of-sample robustness of portfolio optimization within a complex market environment. Within the mean-variance framework, we compare four methods for covariance estimation and weighting: a baseline model based on sample covariances, the Ledoit-Wolf shrinkage estimator, Garber's robust covariance estimator, and a minimum-variance model incorporating an L2 regularization term. This study uses data from 60 American stocks from 2019 to 2021, with four rebalancing cycles (5, 20, 60, and 100 days) were established, and a moving-window backtesting method was applied to evaluate the performance of the various methods in terms of returns, risk, and out-of-sample stability. The empirical results indicate that the L2 normalization model generates higher cumulative returns and Sharpe ratios during most rebalancing periods, demonstrating effective stabilization of the weights; Garbner's robust covariance technique demonstrates consistent resilience to downturns in high-volatility market environments; And Reduit-Wolf's contraction estimates also show a relatively stable improvement in sample skewness compared to the benchmark model. These results suggest that including a covariance noise or normalization term in the traditional expected value and variance-based approach to portfolio management helps reduce the impact of estimation error and market volatility on portfolio performance. The period of analysis for this paper is specified. Furthermore, it covers the observation period, highlights limitations related to market comprehensiveness and parameter selection, and provides guidance for future research by incorporating financial data from diverse sources and more advanced robust optimization techniques.

Keywords: robust portfolio optimization, covariance estimation, L2 regularization, rolling backtesting, out-of-sample performance.

1. Introduction

1.1. The Portfolio Selection Problem and Its Research Significance

When creating an investment plan, the main question is how to allocate funds among different investment avenues in order to balance the returns and risks of a changing market. Unlike investing in a single asset class, managing an investment portfolio is not only about achieving high returns, but also about managing risks, It also involves managing risks and ensuring the stability of the entire investment portfolio. Therefore, the proper structuring of financial instruments is a major area of study within the disciplines of financial engineering and quantitative finance, encompassing areas such as fund management, asset allocation, quantitative investing, and risk management. Moreover, given the market's growing complexity and its constantly changing interrelationships, the question of how to structure a solid investment portfolio that can deliver good returns beyond benchmark performance is a crucial one, both in theory and in practice. is of great importance, both in theory and in practice.

1.2. The Classic Status and Modern Challenges of the Markowitz Mean-Variance Model

Modern portfolio theory is based on the mean-variance model proposed by Markowitz [1]. This model rests on two assumptions: first, investors seek to achieve higher returns for a given level of risk; second, once the target return is achieved, they seek to minimize portfolio risk. Mathematically, risk is largely defined by the variance of portfolio returns; therefore, the portfolio problem can be formulated as an optimization problem concerning asset weights. Furthermore, the lower the total variance compared to portfolios with similar characteristics, the lower its cost. The goal is to minimize portfolio variance, while the sum of all weightings must equal one, and fractional weightings are generally not permitted. These observations are of fundamental importance because they were the first to transform "diversification" from an empirical concept into a measurable and rational mathematical model.

However, in practice, the performance of the min-variance model largely depends on the quality of the estimates of the covariance matrix. The covariance matrix describes the relationships among individual assets; it not only determines the overall risk of the portfolio but also directly influences the optimal allocation of weighting coefficients. With real financial data, accurately estimating the covariance matrix is often difficult. First, asset return series are typically subject to noise, fat-tailed distributions, and abnormal volatility; a few extreme outliers can significantly distort the sample covariance. Second, if the amount of data is limited, then the sample covariance matrix itself suffers from a significant estimation error, and the optimizer further amplifies this error, resulting in extreme solutions that exhibit unstable weights or even excessive centralization [2]; Moreover, the market environment itself is not static, and there is no guarantee that the correlation structures observed in historical periods will persist stably in the future. Therefore, relying directly on sample cross-correlations when optimizing a portfolio often leads to results that seem plausible.

1.3. Existing Improvement Approaches and Research Trends

The optimization methods employed in this study can be divided into three main categories with respect to covariance estimation error and optimization instability. The first category includes statistical smoothing methods, an example of which is the Lasso-Woolf smoothing estimator [3]. This method, which does not rely entirely on the sample covariance, smooths it in the direction of the sought structural matrix so as to balance bias and variance, thereby increasing estimation stability. The second group comprises robust covariance methods, represented by the Garberian statistical estimator [5]. Through a threshold mechanism, this method focuses only on significant covariances among the return series, minimizes noise from low-variance elements, and reduces the influence of outliers on the covariance estimates. Methods in the third category are not limited to modifications of the covariance input data, but impose direct constraints on the structure of the solution during the optimization phase. For example, there is a method that includes an L2 regularization term in the objective function to limit excessive concentration of weights and improve the smoothness and feasibility of the portfolio distribution [6]. These three categories of methods correspond to two complementary technical approaches: "enhanced input data" and "stabilized output data" [3].

1.4. Research Questions, Design, and Contributions of This Paper

Based on the previous discussion, This research paper conducts a comparative empirical study focusing on four representative methods for covariance estimation or stabilization within the framework of a long-only model for a combined 'medium-reactive' portfolio. Specifically, this paper selects Sample Correlation as the reference model, the Ledoit-Wolf method as the shrinkage estimator, and Garber Statistics as the robust correlation estimator, and introduces a min-variance model with an L2 normalization term as a weight stabilization technique. To simulate as accurately as possible the dynamic rebalancing process encountered in practice, a sliding-window backtesting method is used in this study. Furthermore, four rebalancing horizons (5 days, 20 days, 60 days, and 100 days) are defined to investigate whether the change in the rebalancing cycle affects the performance ranking of the individual methods.

The main contributions of this paper are as follows. First, within a unified mean-variance portfolio framework, we conduct a systematic comparison of four typical methods: sample covariance, shrinkage estimation, robust covariance, and regularization optimization. Second, we analyze portfolio rebalancing frequency as an independent research dimension and compare differences in out-of-sample performance across different holding periods. Third, we report experimental results across three dimensions—cumulative returns, volatility control, and tail risk management—to provide an empirical comparison of the out-of-sample performance of different robust portfolio methods.

This document is organized as follows: Chapter II describes the basic principles and performance indicators of the four methods; Chapter III presents the experimental methodology and analysis of the results; finally, Chapter IV presents the conclusions of the study, identifying its limitations and suggesting directions for further research.

2. Methodology

2.1. A Unified Analytical Framework for Mean-Variance Optimization

The methodology presented in this paper is based on the Markowitz mean-variance optimization framework [1]. Its core idea is to consider both return and risk simultaneously when constructing a portfolio. In this paper, we adopt a penalty function as a unified analytical framework and formulate the portfolio optimization problem as follows:

$$\max_{w} w^T \mu - \lambda w^T \sum w$$

Here, w denotes the asset weight vector, where $1^T w = 1$ and $w \geq 0$; μ denotes the expected return vector of the assets; Σ denotes the covariance matrix of asset returns; and λ denotes the investor's risk aversion coefficient. The first term of this objective function captures "Return-seeking," while the second term captures "Risk aversion."

In the empirical design of this paper, the expected return vector μ is estimated from the historical average returns of each asset within a rolling training window, while the covariance matrix Σ is constructed using methods such as sample covariance, the Ledoit-Wolf contraction estimate, and the Gerber robust covariance. To ensure comparability across different models, this study uniformly sets the risk aversion coefficient $\lambda = 2$ and solves for the mean-variance portfolio subject to the constraints that the sum of weights equals 1 and short selling is not permitted. The core of this comparison lies in determining whether, under the same return estimation method, different covariance estimation methods and weight regularization constraints can improve the portfolio' s out-of-sample performance; traditional methods for estimating the covariance matrix are referred to as the benchmark methods.

2.2. Method 1: Ledoit-Wolf Compressed Estimation

The first category of improvements involves statistical shrinkage methods; this paper selects Ledoit-Wolf compressed estimation as a representative example [3]. The core intuition behind this method is that while the sample covariance matrix S is unbiased, its estimation variance is large, especially when there are many assets and the sample size is relatively small; conversely, a highly structured target matrix F (such as a matrix of constant correlation coefficients) may have a large bias but has a very small estimation variance. The Ledoit-Wolf method uses a shrinkage parameter $\delta \in [0, 1]$ to shrink the sample covariance matrix toward the target matrix, yielding an improved estimator:

$\hat{\Sigma}_{LW} = \delta F + (1 - \delta) S$

In particular, the optimal shrinkage strength δ^* is derived theoretically by minimizing the mean squared error of the estimator. This method strikes a balance between bias and variance, aiming to improve the stability of out-of-sample predictions.

2.3. Method 2: Robust Covariance Estimation Based on the Gerber Statistic

The second category of improvements involves robust coproduction methods. In this paper, we use the Gerber statistic to construct robust correlation and covariance matrices [5]. The core idea of the Gerber method is to focus only on "Significant coproduction" while filtering out "Minor noise." The specific steps are as follows:

1. Set a threshold: For each asset, calculate a volatility threshold (e.g., 0.5 times the standard deviation) based on its historical return series.

2. State discretization: For each trading day, compare the return of each asset with this threshold and classify it into three states: significant rise (return greater than the positive threshold), significant fall (return less than the negative threshold), or neutral (return between the thresholds).

3. Calculate robust correlation: For any two assets i and j, count the number of days n_c during the sample period when they exhibit "significant co-movement" (both rising or both falling), and the number of days n_d when they exhibit "significant counter-movement" (one rising and one falling). Days where at least one asset is "neutral" are ignored. The Gerber correlation coefficient is then defined as:

$$g_{ij} = \frac{n_c - n_d}{n_c + n_d}$$

When $n_c + n_d = 0$, this paper sets $g_{ij} = 0$; at the same time, g_{ii} is set to 1, thereby constructing the Gerber correlation matrix $G = g_{ij}$.

4. Constructing the covariance matrix: Using the Gerber correlation matrix G and the sample standard deviation D of each asset's return, construct a robust covariance matrix: $\hat{\Sigma}_{Gerber} = DGD$.

This method mitigates the impact of minor noise and extreme outliers on correlation estimates and is suitable for thick-tailed, non-normal market environments.

2.4. Method 3: L2-Regularized Mean-Variance Model

While the Ledoit-Wolf and Gerber methods aim to "correct the input covariance matrix," the third approach seeks to "Directly stabilize the optimizer itself." To this end, this paper incorporates an L2 regularization term into the traditional minimum-variance objective function to penalize the Euclidean norm of the weight vector (i.e., the sum of the squares of the weights):

$$\min_{w} - w^T\mu + \lambda w^T \sum w + \lambda_{reg} \sum_{i=1}^{N} w_i^2 \ s.t. \sum_{i} w_i = 1, w_i \geq 0$$

Here, λ_{reg} is the regularization strength hyperparameter. The newly added regularization term $\Sigma_i w_i^2$ penalizes "concentrated holdings": if the weights are overly concentrated on a few assets, the value of the penalty term will be larger; if the weights are more diversified, the penalty will be relatively smaller. The underlying principle is that, in real-world settings where parameter estimates are not entirely reliable, forcing the model to maintain a certain degree of diversification can reduce the risk of corner solutions and enhance the robustness of the optimization results to input errors, thereby improving out-of-sample performance [6].

2.5. A Unified Comparative Framework for the Four Methods

Structurally, the four classes of methods compared in this paper can be divided into two groups. The first group modifies the covariance matrix itself, including the benchmark sample covariance, the Ledoit-Wolf compressed estimate, and the Gerber robust covariance; their common feature is that the expected return vector uses the sample mean estimate within the same rolling training window, with differences primarily reflected in the construction of the covariance matrix Σ in the risk term. The second group modifies the optimizer itself, specifically the L2-regularized minimum variance model; this approach adds a penalty for weight concentration on top of traditional risk minimization. Consequently, the method design in this paper actually corresponds to two distinct yet complementary technical approaches: one improves the stability of the risk measure by refining the covariance estimate, while the other imposes constraints on the weight structure at the optimization level to mitigate the risk of estimation errors being amplified by the optimizer.

3. Empirical Results and Discussion

3.1. Data Sources, Experimental Design, and Evaluation Metrics

This study utilized daily trading data for 60 U.S. stocks from 2019 to 2021, collected from cn.investing.com, as the sample for the experimental analysis. The period covered by the sample spans a time of extreme market volatility both before and after the impact of the COVID-19 pandemic, allowing us to analyze the performance of various portfolio strategies outside the sample in an unstable market environment. For each stock in this study, we first calculated simple returns based on daily price data:

$$r_{i,t} = \frac{P_{i,t}}{P_{i,t-1}} - 1,$$

In this context, $P_{i,t}$ denotes the price of the i-th stock on the t-th trading day, and $r_{i,t}$ denotes the corresponding daily return.

For this study, forward projection with a moving average was selected for the performance comparison. To be more specific, in the training period, the last two hundred trading days were used. For each backtest day, the covariance matrix of the asset's expected return vector with the return indicators from the training period is calculated, which in turn allows for the calculation of the portfolio weights. These weights are then applied to the daily returns of the subsequent period, which allows the calculation of the portfolio's actual, unobserved future return. This paper sets four holding periods—5 days, 20 days, 60 days, and 100 days—to compare the performance of each model under different rebalancing frequencies. Since the first 200 trading days are used for model training, the cumulative net value curve and performance evaluation are calculated starting from the 201st trading day.

Within each rolling training window, the expected return vector is estimated using the sample mean and annualized over 252 trading days:

$$\hat{\mu} = 252 \cdot \frac{1}{T} \sum_{t=1}^{T} r_t,$$

Here, T = 200 denotes the training window length, and r_t represents the vector of returns for all assets on the t-th trading day. The covariance matrix is estimated using sample covariance, Ledoit-Wolf reduced covariance, and Gerber robust covariance, respectively, depending on the model, and is similarly annualized. All models are solved within a unified long-only mean–variance framework:

$$\min_{w} - w^T \hat{\mu} + \lambda w^T \hat{\Sigma} w + \lambda_{reg} \sum_{i=1}^{N} w_i^2,$$

$$\text{s.t.} \quad \sum_{i=1}^{N} w_i = 1, 0 \le w_i \le 1,$$

Here, w denotes the portfolio weight vector, μ denotes the estimated annualized expected return vector, $\hat{\Sigma}$ denotes the annualized covariance matrix, and λ denotes the risk aversion coefficient. In this paper, λ is uniformly set to 2. For the Benchmark, Ledoit-Wolf, and Gerber models, λ_{reg} is set to 0; for the L_2 regularization model, λ_{reg} is set to 5.

After obtaining the portfolio weights, the out-of-sample return of the portfolio on the t-th trading day is calculated as:

$r_{p,t} = w^T r_t$.

The corresponding cumulative net value is:

$$V_t = \prod_{s=1}^{t} (1 + r_{p,s}).$$

To comprehensively evaluate the out-of-sample performance of different models, this paper selects annualized return, annualized volatility, Sharpe ratio, maximum drawdown, and 95% VaR as the primary evaluation metrics. Let {rp,t} denote the sequence of daily returns for the out-of-sample portfolio, where T0 is the number of out-of-sample trading days. Then, the average daily return is:

$$\bar{r}_p = \frac{1}{T_0}\sum_{t=1}^{T_0} r_{p,t}$$

This paper defines the annualized return as:

$$R_{ann} = (1 + \bar{r}_p)^{252} - 1$$

The annualized volatility is defined as:

$$\sigma_{ann} = \sqrt{252} \cdot sd(r_{p,t})$$

Without introducing a risk-free rate separately, the Sharpe ratio is calculated as:

$$Sharpe = \frac{R_{ann}}{\sigma_{ann}}$$

Maximum drawdown measures the largest decline in a portfolio's net value relative to its historical peak and is defined as:

$$MDD = \min_t\left(\frac{V_t}{max_{s\le t} V_s} - 1\right)$$

The 95% VaR is represented by the 5th percentile of out-of-sample daily returns:

$VaR_{95\%} = Q_{5\%}(r_{p,t})$

The above metrics characterize the out-of-sample performance of the portfolio from the perspectives of return levels, volatility risk, risk-adjusted returns, extreme drawdowns, and tail losses.

3.2. Comparison of Cumulative Net Value Curves

Figures 1 through 4 illustrate the changes in cumulative net value for the benchmark model, the Ledoit-Wolf contraction estimation model, the Gerber robust covariance model, and the L2 regularization model under four rebalancing cycles: 5 days, 20 days, 60 days, and 100 days. Overall, all four models were impacted by the severe market volatility at the beginning of 2020, resulting in a significant drawdown in cumulative net value at the start of the sample period. As market conditions subsequently improved, the net asset value of each fund gradually began to recover and has shown a more pronounced upward trend since 2021. This demonstrates that, within the framework of moving average and variance models, various techniques for estimating and adjusting covariance provide all investors with the ability to capitalize on profit opportunities during market recovery phases.

In terms of the models' relative performance, the Ledoit-Wolf contraction model delivered excellent cumulative returns in most reallocation cycles. This suggests that minimizing the estimation error by reducing the sample covariance matrix can, to some extent, improve the portfolio's extrapolation capacity. The performance curves of the benchmark model and the Gerber model generally show similarities, but there are some differences depending on the rebalancing cycle. In particular, the Gerber method proved to be particularly effective in scenarios where the frequency of rebalancing is low. In contrast, the cumulative growth in net assets of the L2 standardized model is relatively low, but the curve is smoother. This suggests that while standardizing weights may somewhat reduce return elasticity, it effectively minimizes portfolio risk and net asset volatility.

3.3. Comparative Analysis of Quantitative Metrics

To more accurately compare the results of the various models on out-of-sample data, Table 1 presents the annual return, annual volatility, Sharpe ratio, maximum drawdown, and 95% VaR for each model across four rebalancing cycles. In terms of returns and risk-adjusted returns, the Ledoit-Wolf model recorded the highest annual returns and Sharpe ratios for the 5-day, 20-day, and 60-day rebalancing cycles. In contrast, the Gerber model recorded the highest annual returns and Sharpe ratios for the 100-day rebalancing cycle, suggesting that reducing the rebalancing frequency can be beneficial when covariances are estimated reliably. The comparison modules generally demonstrated fairly stable performance but remained at a lower level overall than the Ledoita-Wolf module, suggesting that directly using the model's covariance remains sensitive to estimation errors.

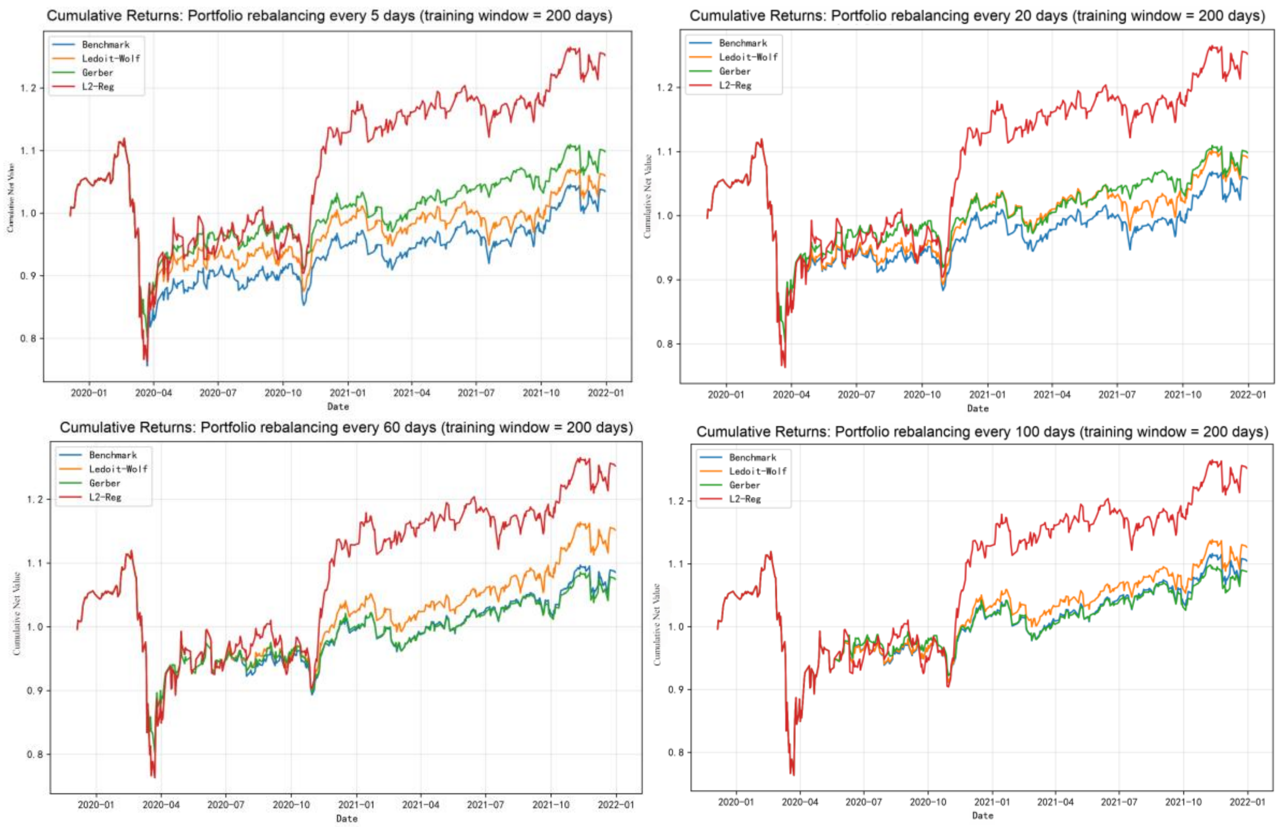

Figure 1. Cumulative net value curves for each model under different rebalancing cycles

From a risk management perspective, the model using L2 smoothing shows the lowest annualized volatility across all rebalancing cycles, while the maximum drawdown and the absolute value of the 95% -VaR is also relatively low. This demonstrates that the L2 smoothing factor can significantly reduce the risk of weight concentration and extreme losses. However, the annualized return of this model is significantly lower than that of the other three model types, leading to an uncompetitive Sharpe ratio during certain rebalancing cycles. Therefore, although L2 normalizations are suitable as a tool for risk management and weight stabilization, they do not necessarily guarantee optimal results in terms of return maximization.

Table 1. Out-of-sample performance comparison of four portfolio models across different rebalancing cycles

Portfolio Rebalancing Cycle	Model	Annualized Return	Annualized Volatility	Sharpe Ratio	Maximum Drawdown	95% VaR
5 Days	Benchmark	0.7630	0.4983	1.5310	-0.4719	-0.0460
	Ledoit-Wolf	0.8287	0.5224	1.5865	-0.4890	-0.0472
	Gerber	0.6858	0.4788	1.4323	-0.4750	-0.0453
	L2-Reg	0.4201	0.2855	1.4715	-0.3510	-0.0270
20 Days	Benchmark	1.0026	0.5059	1.9817	-0.4819	-0.0455
	Ledoit-Wolf	1.1686	0.5356	2.1817	-0.5012	-0.0467
	Gerber	0.9974	0.4887	2.0410	-0.4859	-0.0442
	L2-Reg	0.4393	0.2896	1.5169	-0.3410	-0.0260
60 Days	Benchmark	1.0574	0.4901	2.1573	-0.4717	-0.0445
	Ledoit-Wolf	1.2441	0.5234	2.3771	-0.4912	-0.0455
	Gerber	0.9807	0.4780	2.0516	-0.4723	-0.0422
	L2-Reg	0.3892	0.2802	1.3890	-0.3205	-0.0250
100 Days	Benchmark	0.9826	0.4194	2.3427	-0.3509	-0.0402
	Ledoit-Wolf	1.0517	0.4329	2.4297	-0.3535	-0.0406
	Gerber	1.0942	0.4252	2.5732	-0.3523	-0.0392
	L2-Reg	0.3364	0.2711	1.2410	-0.3200	-0.0244

4. Conclusions

This commentary compares the out-of-sample performance of a normative model, using the covariance model, the Ledoit-Wolf shrinkage estimation model, the Gerber robust covariance model, and the L2 regularization model, within the framework of a mean- and variance-based long-only strategy. This study uses daily data of sixty American stocks from 2019 to 2021 and employs a rolling backtest method with a fixed 200-day setup period and four rebalancing cycles of five, twenty, sixty, and one hundred days. Performance is evaluated using metrics such as cumulative net asset value, annualized return, annualized volatility, Sharpe ratio, maximum drawdown, and 95% VaR.

The empirical test results indicate that the equilibrium- and leverage-constraint-based decision-making methods have a significant impact on the out-of-sample performance of mean-variance-optimized portfolios. The Ledoit-Wolf reduced-form model produced higher annual returns and higher Sharpe ratios in most backtesting periods, indicating that reducing the bias helps to reduce the estimation noise under low sample size conditions. Gerber's robust balancing component performed very well under low cluster updating conditions, and demonstrated a level of noise resistance. Although the L2 adjustment component has a smaller return, it has advantages for risk management metrics such as annualized volatility, maximum loss, and VaR, indicating that weighting adjustments can reduce portfolio variance and the risk of large losses.

In general, the Ledoit-Wolf method is more suitable for improving returns and risk-adjusted returns, while the Gerber method offers certain advantages in specific situations where rebalancing is required only sporadically, and the L2 adjustment method is more effective for risk management and stabilization of the weights. The results of this article demonstrate that, within a mean-variance framework, reducing covariance, estimating robust correlations, or introducing adjustment constraints can improve a portfolio's out-of-sample stability in various ways.

This study has several limitations. First, the analysis period is limited to a relatively short timeframe, from 2019 to 2021. Second, the risk aversion coefficient, the Gaver threshold, and the L2 normalization level were set to fixed values, and no systematic sensitivity analysis was conducted for these parameters. Finally, this study did not account for constraints related to transaction costs or turnover rates. Future research could further increase the sample size, examine the model's performance under various parameter settings, and incorporate transaction costs, position concentration constraints, and financial information from multiple sources into the portfolio optimization process.

Reference

[1] Markowitz, H. (1952) . Portfolio selection. The Journal of Finance, 7(1) , 77–91.

[2] DeMiguel, V. , Garlappi, L. , & Uppal, R. (2009) . Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? The Review of Financial Studies, 22(5) , 1915–1953.

[3] Jagannathan, R. , & Ma, T. (2003) . Risk reduction in large portfolios: Why imposing the wrong constraints helps. The Journal of Finance, 58(4) , 1651–1683.

[4] Ledoit, O. , & Wolf, M. (2004) . A well-conditioned estimator for large-dimensional covariance matrices. Journal of Multivariate Analysis, 88(2) , 365–411.

[5] Gerber, S. , Markowitz, H. M. , Ernst, P. A. , Miao, Y. , Javid, B. , & Sargen, P. (2022) . The Gerber statistic: A robust co-movement measure for portfolio optimization. The Journal of Portfolio Management, 48(3) , 87–102.

[6] DeMiguel, V. , Garlappi, L. , Nogales, F. J. , & Uppal, R. (2009) . A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. Management Science, 55(5) , 798–812.