International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(23s) 2026 pp. 1794-1801
https://doi.org/10.70917/ijcisim-2026-5801
Article

Tail Robustness Inversion in Portfolio Optimisation

Rakshay Pawar

Senior Quantitative Finance Associate, Codeflow Capital LLC, United States
ORCID 0009-0008-7367-4811 | Rakshay@codeflowcapital.com

Abstract: This study identifies and quantifies tail-robustness inversion: improving a simulated covariance-tail objective can increase realised portfolio risk. A prespecified walk-forward evaluation covers 49 value-weighted US industry portfolios, 648 monthly rebalances, and 13,614 daily out-of-sample returns from January 1972 through December 2025. Bootstrap covariance-tail expected-shortfall optimisation lowers its in-sample tail objective by 1.938% relative to Ledoit–Wolf minimum variance, with an objective no higher at any rebalance (647 improvements and one tie in the saved results). Yet average out-of-sample realised variance is 2.588% higher, with a paired block-bootstrap 95% confidence interval of [1.121%, 4.057%] for the increase. The inversion persists across all 12 reported specifications and all five chronological regimes in annualised volatility. Diversification constraints reduce the variance penalty, providing a practical direction for controlling the effect. The analysis also distinguishes the covariance fragility ratio's descriptive value from its ability to forecast subsequent risk. Together, these findings provide a reproducible empirical counterexample to equating simulated tail protection with effective portfolio protection, and a structured validation approach for assessing covariance-based allocation methods before deployment.
Keywords: Bootstrap covariance estimation; minimum-variance portfolio; expected shortfall optimisation; tail-robustness inversion; concentration-turnover channel; walk-forward backtesting; out-of-sample portfolio risk

1. Introduction

Covariance estimation is central to quantitative portfolio risk management. Markowitz [1] established the mean–variance portfolio framework; subsequent research documented how estimation error can destabilise its portfolio weights [9, 10, 11]. When an optimiser responds to noise in estimated covariances, an allocation that appears efficient in sample may deliver weaker risk control out of sample. This distinction motivates direct evaluation of realised risk alongside the criterion used to select portfolio weights.

Linear shrinkage offers a well-established response to covariance estimation error. Ledoit and Wolf [2, 7] combine the sample covariance matrix with a structured target using a data-dependent shrinkage intensity. Their analytical nonlinear-shrinkage estimator [17] extends this approach through eigenvalue-specific adjustments. The present study uses the scaled-identity linear estimator as a transparent and computationally tractable benchmark for testing the incremental value of bootstrap-tail optimisation.

Bootstrap resampling offers a different route to the same destination. Rather than regularising a single covariance estimate, it constructs a large ensemble of covariance matrices from resampled return histories and optimises portfolio weights against the ensemble distribution. Michaud [11] documented the extreme weight instability of classical mean-variance optimisation, motivating subsequent resampling-based attempts to stabilize portfolio allocations. The bootstrap covariance-tail variant sharpens the objective further: instead of the full bootstrap variance distribution, it targets only the worst 10% of bootstrap variance scenarios — a tail expected shortfall — to make the portfolio explicitly robust to the covariance realisations that most threaten out-of-sample stability.

This paper makes the relationship between simulated tail protection and realised risk the central object of evaluation. Across 648 monthly rebalances on 49 US industry portfolios from 1972 to 2025, the bootstrap-tail strategy matches or improves the benchmark objective at every rebalance but increases average realised variance relative to shrinkage minimum variance. We call this documented divergence tail-robustness inversion. The contribution combines a precisely measured risk penalty, a broad specification check, and diversification and turnover diagnostics that make the result useful for portfolio model validation.

The remainder of the paper is organised as follows. Section 2 reviews the portfolio estimation and bootstrap literatures that frame the investigation. Section 3 describes the data, portfolio construction methods, and statistical tests. Section 4 presents the empirical results. Section 5 examines the practical implications and limitations. Section 6 draws conclusions.

2. Literature Review

The sensitivity of estimated portfolio weights is well established. Britten-Jones [10] analyses sampling error through the relationship between portfolio estimation and regression. DeMiguel, Garlappi and Uppal [9] compare 14 portfolio models across seven datasets and find that none consistently outperforms the 1/N benchmark across their performance measures. These results motivate careful out-of-sample validation; they do not imply that equal weighting must dominate in every universe or under every risk objective.

Ledoit and Wolf [2, 7] addressed the covariance component of this problem through linear shrinkage toward a structured target; the estimator implemented here uses a scaled-identity target. The analytical derivation of the shrinkage intensity distinguishes their estimator from ad hoc regularisation: the intensity is not a tuning parameter but a function of the data, derived under an asymptotic framework that allows both N and T to grow. Their 2020 analytical nonlinear-shrinkage extension [17] applies eigenvalue-specific shrinkage, recovering substantially more structure when N/T is large.

Bayesian and high-dimensional portfolio estimation address parameter uncertainty from complementary perspectives. Bauder et al. [14], Bodnar, Mazur and Okhrin [19], and Bodnar, Parolya and Schmid [20] develop methods for portfolio selection under estimation uncertainty. Kempf and Memmel [18] study estimation of the global minimum-variance portfolio. These contributions frame the present question: whether explicitly weighting adverse covariance draws improves the realised performance of a long-only risk-minimising allocation.

Michaud [11] describes the instability of classical optimised portfolios. Bootstrap methods offer a way to represent sampling uncertainty using repeated resamples [6]. For dependent returns, the circular block bootstrap of Politis and Romano [5] resamples contiguous observations with wraparound, preserving dependence within sampled blocks while approximating the serial structure of the original series.

Rockafellar and Uryasev [4] develop a tractable expected-shortfall optimisation framework, and Artzner et al. [8] establish the coherence framework for risk measures. In this study, expected shortfall is applied to the distribution of bootstrap portfolio variances, rather than directly to realised portfolio losses. Bongiorno and Challet [12] examine covariance filtering with bootstrapped hierarchies, while Oliveira, Guzman and Firoozye [13] propose non-parametric bootstrap robust optimisation. The present study focuses specifically on the empirical relationship between the upper-tail mean of bootstrap variances and subsequent realised variance.

Kolm, Tutuncu and Fabozzi [21] review practical challenges in portfolio optimisation, including estimation error and implementation. Roncalli [16] discusses asset-management risk methods, while Lopez de Prado [22] and Coqueret and Guida [15] provide broader context for financial model evaluation. The contribution here is a unified assessment of the simulated tail objective, realised variance, diversification, turnover, and covariance-fragility diagnostics within one prespecified walk-forward design.

3. Methodology

3.1 Data

Daily value-weighted returns for 49 US industry portfolios are obtained from the Kenneth R. French Data Library [23]. The frozen complete-case input sample contains 14,120 daily observations from January 1970 through December 2025. The out-of-sample evaluation begins in January 1972 after the first 504-trading-day estimation window. These are academic industry portfolios, not individual stocks or directly executable funds; no security-level delisting, merger, dividend, or corporate-action adjustments are made beyond the published Data Library construction. The provider changed its underlying CRSP file format beginning with the January 2025 data release, so a separate sensitivity analysis ends in 2024.

3.2 Benchmark: Ledoit-Wolf Minimum-Variance Portfolio

The benchmark at each rebalancing date is the long-only global minimum variance portfolio constructed using the Ledoit-Wolf linear shrinkage covariance estimator [2, 7], with shrinkage intensity determined analytically. The optimisation solves the problem of minimising portfolio variance subject to the long-only simplex constraint. This benchmark is chosen because Ledoit-Wolf shrinkage is a widely used, well-established covariance estimator and provides a disciplined, well-conditioned baseline for evaluating the incremental effect of bootstrap-tail optimisation.

3.3 Treatment: Bootstrap Covariance-Tail ES90 Portfolio

For each monthly rebalancing date, the treatment portfolio is constructed using a circular block bootstrap with 100 covariance-draw replications and a block length of 21 trading days [5, 6]. Each replication yields a resampled return matrix from which a full covariance matrix is estimated. The portfolio variance under each replication is computed as the quadratic form of portfolio weights and the bootstrap covariance matrix. The objective is to minimise the ES90 of the bootstrap variance distribution — the mean portfolio variance over the worst 10% of bootstrap replications — subject to the long-only simplex constraint. The empirical average of the largest 10% of bootstrap quadratic-variance draws is minimized using sequential least-squares programming (SLSQP) with an analytic subgradient; Rockafellar and Uryasev [4] provide the related expected-shortfall framework rather than the implemented solver.

3.4 Walk-Forward Harness

Both portfolios are rebalanced monthly on identical dates using only the preceding 504 daily returns available as of each rebalancing date. The out-of-sample walk-forward period runs from January 1972 through December 2025, yielding 648 monthly rebalances and 13,614 daily portfolio-return observations. Monthly realized variance is computed as the sample variance of gross daily portfolio returns within each month, annualized by 252 trading days. The primary comparison uses gross realized variance; net strategy diagnostics apply a 10-basis-point one-way transaction cost, with alternative cost assumptions evaluated separately.

3.5 Primary Test Statistic and Inference

The primary statistic is proportional realised-variance reduction: one minus the ratio of the treatment's mean monthly variance to the benchmark's mean monthly variance. Inference uses a paired circular block bootstrap with 5,000 replications and 12-month blocks [5]. Each resample preserves the pairing of the two strategies and recomputes the ratio from their resampled monthly outcomes. The estimate is −2.588%, with a 95% interval of [−4.057%, −1.121%]. For direct interpretation of the inversion, results are also expressed as the equivalent variance increase of 2.588%, with interval [1.121%, 4.057%].

3.6 Concentration-Turnover Diagnostics

Concentration is measured by effective N, the reciprocal of the sum of squared weights. One-way turnover is half the absolute weight change from drifted holdings to new targets; initial establishment has turnover one. Common

10% and 20% weight caps test whether diversification attenuates the variance penalty, consistent with the role of constraints in controlling estimation risk [3].

3.7 Covariance Fragility Ratio

The covariance fragility ratio (CFR) is the mean variance of the worst 10% of bootstrap draws divided by the mean variance of all draws, minus one. It measures the relative simulated tail surcharge. Its association with next-month variance surprise is tested separately.

4. Results

4.1 Evidence of Tail Robustness Inversion

Bootstrap-tail ES90 increases average monthly realised variance by 2.588% relative to Ledoit–Wolf minimum variance, with a paired 95% block-bootstrap interval of [1.121%, 4.057%] for the increase. Annualised gross volatility is 11.90% versus 11.74%, approximately 15 basis points higher before rounding. Monthly variance is higher in 412 of 648 months, lower in 235, and equal in one. Annualised volatility is also higher in all five prespecified chronological regimes.

The simulated objective improves by 1.938% on average: it is lower at 647 rebalances and equal at one at the saved precision. This verifies improvement over the benchmark on the selection criterion, without certifying a global numerical optimum. Its divergence from realised risk establishes the inversion. Table 1 reports all six strategies.

Table 1. Out-of-sample performance across six portfolio strategies

Strategy	Volatility gross / net (%)	Net Sharpe	Net max. drawdown (%)	Annual turnover (%)	Effective industries	Mean CFR
Equal weight	16.73 / 16.73	0.477	−57.3	21.4	49.00	0.342
Sample GMV	11.74 / 11.74	0.564	−40.8	72.5	4.28	0.331
Ledoit–Wolf GMV	**11.74 / 11.74**	**0.556**	**−40.9**	**71.5**	**4.69**	**0.334**
Bootstrap mean GMV	11.75 / 11.75	0.562	−41.0	79.1	4.26	0.331
Bootstrap-tail ES90	**11.90 / 11.90**	**0.559**	**−42.7**	**117.4**	**4.01**	**0.303**
Hierarchical risk parity	14.77 / 14.77	0.517	−52.7	103.8	34.80	0.349

January 1972–December 2025; 648 rebalances. Volatility is annualised; gross and net values coincide at displayed precision. Net statistics deduct 10 basis points per unit of one-way turnover. GMV = global minimum variance; CFR = covariance fragility ratio. Bold identifies the primary comparison.

4.2 Diversification and Turnover Diagnostics

Effective N declines from 4.69 industries under Ledoit–Wolf to 4.01 under bootstrap-tail ES90, a difference of 0.69 industries before rounding. The tail objective is therefore associated with a more concentrated allocation. Common weight caps reduce the variance penalty from 2.588% to 1.412% at a 20% cap and 1.092% at a 10% cap. This sensitivity gives the result practical value: diversification constraints mitigate the inversion, although they do not eliminate it or establish concentration as its sole cause.

Annualised one-way turnover increases by 64.1%, from 0.715 to 1.174 times portfolio value. The resampled covariance distribution changes as the 504-day estimation window advances, and the resulting reallocation pattern is consistent with greater sensitivity to changing estimates. Turnover adds implementation cost; it cannot mechanically explain an increase in the gross-variance outcome, which excludes transaction costs. Concentration, turnover, and weight-cap sensitivity are therefore treated as related diagnostics rather than a quantified causal decomposition.

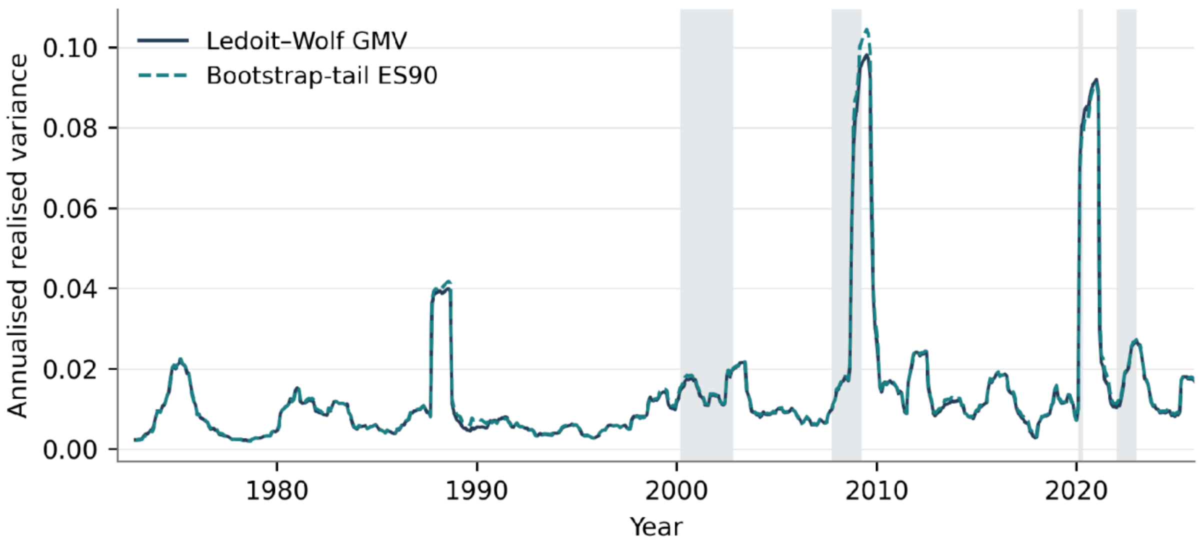

Figure 1. Trailing 12-month mean of annualised monthly realised variance. The underlying monthly series spans January 1972–December 2025; the first complete 12-month average is December 1972. Grey bands mark March 2000–October 2002, October 2007–March 2009, February–April 2020, and calendar 2022. The higher-variance count of 412 months refers to unsmoothed monthly observations.

4.3 Interpreting Covariance Fragility

The proposed strategy's mean covariance fragility ratio is 0.303 across 648 rebalances (Figure 2). This means that its upper-tail mean bootstrap variance is, on average, approximately 30.3% above its mean bootstrap variance. The ratio is an in-sample distributional diagnostic. Separately, the prespecified positive association with subsequent variance surprise is contradicted by the pooled six-strategy evidence: across 3,888 strategy-month observations, the Spearman correlation is −0.225, with a 95% block-bootstrap interval of [−0.367, −0.074]. This finding sharpens the interpretation of the diagnostic: a larger relative simulated surcharge is not a validated signal of larger subsequent variance surprise.

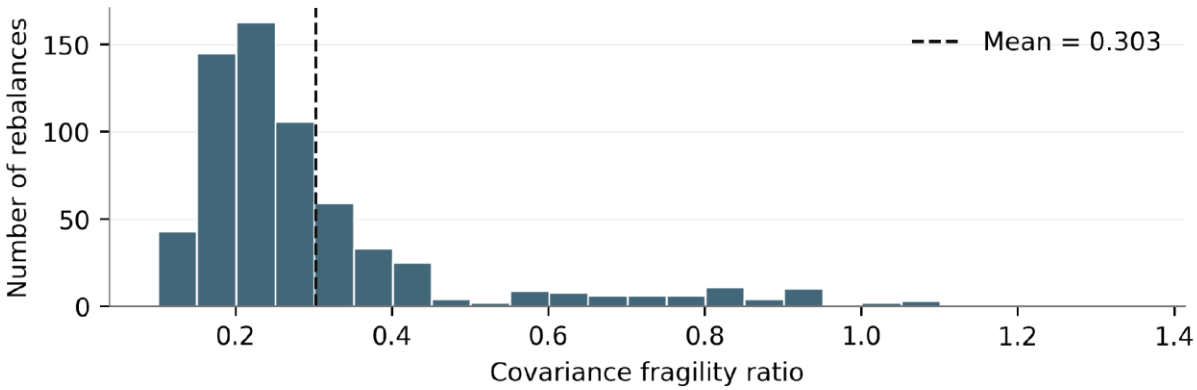

Figure 2. Distribution of the proposed strategy's covariance fragility ratio across 648 rebalances. The dashed line marks the mean of 0.303. All observed ratios are positive, consistent with an upper-tail mean exceeding the mean of the same bootstrap distribution.

4.4 Robustness

The inversion persists across all 12 reported specifications: the base case, ten prespecified parameter variants, and one sample-end sensitivity. Every 95% interval for the variance increase lies above zero (Table 2). Each specification uses its stated tail level and evaluation period. Transaction-cost scenarios of 0, 5, 10, and 25 basis points quantify implementation burden separately; they leave gross variance unchanged by construction. A post-protocol multistart check produces a 2.399% variance increase, with a 95% interval of [1.024%, 3.795%], supporting the same economic conclusion.

Table 2. Persistence of the inversion across the specification grid

Specification	OOS months	Tail objective reduction (%)	Realised variance increase (%)	95% CI for increase (%)
Base specification	**648**	**1.94**	**2.59**	**[1.12, 4.06]**
252-day window	660	2.67	3.20	[2.03, 4.49]
756-day window	636	1.70	2.59	[0.97, 4.45]
5-day blocks	648	1.65	2.10	[1.16, 3.11]
63-day blocks	648	2.18	2.24	[1.22, 3.47]
75% tail level	648	1.27	1.47	[0.48, 2.53]
95% tail level	648	2.29	2.80	[1.19, 4.41]
50 bootstrap draws	648	2.03	2.15	[0.85, 3.46]
200 bootstrap draws	648	1.86	2.35	[0.89, 3.77]
10% maximum weight	648	0.89	1.09	[0.54, 1.75]
20% maximum weight	648	1.13	1.41	[0.40, 2.49]
Sample ends in 2024	636	1.91	2.58	[0.99, 4.14]

Positive variance changes and intervals indicate higher realised risk than the corresponding Ledoit–Wolf benchmark. Tail levels are 90% unless otherwise stated. Window variants begin in 1971 and 1973, respectively; the final row ends in 2024. OOS = out of sample.

Table 3. Empirical contributions across four evaluation dimensions

Evaluation dimension	Measured evidence	Contribution
Objective versus realised risk	Tail objective improves by 1.938%; realised variance rises by 2.588%.	Identifies and quantifies the inversion.
Regime and stress behaviour	Higher volatility in all five regimes; worse drawdown in three of six stress episodes.	Shows persistence with clearly bounded stress evidence.
Fragility and subsequent risk	Pooled Spearman correlation −0.225; 95% CI [−0.367, −0.074].	Separates descriptive fragility from predictive risk.
Diversification and implementation	Turnover rises 64.1%; common weight caps reduce the variance penalty.	Identifies mitigation and implementation trade-offs.

The original directional hypotheses anticipated lower risk and a positive fragility–surprise relationship. Those predictions are not supported. The contributions above describe the observed counterintuitive findings; they do not redefine the original hypotheses.

5. Discussion

The central contribution is an empirical counterexample to a common modelling intuition: a portfolio with a better simulated covariance-tail score need not have lower realised risk. Circular block resampling and the tail objective perform their defined statistical and optimisation roles, yet the selected portfolio generalises differently from what the original directional hypotheses anticipated. The study makes that gap measurable and tests its persistence across alternative specifications.

This result complements the estimation-error concerns examined by DeMiguel, Garlappi and Uppal [9], but does not establish equal weighting as the best risk-control strategy in this dataset. Indeed, its volatility in Table 1 is higher than that of either minimum-variance strategy. The distinctive evidence here is the comparison with shrinkage: lower simulated tail variance coincides with fewer effective industries, higher turnover, and higher realised variance. That combination provides a specific model-validation finding rather than a general rejection of portfolio optimisation.

CFR describes the relative upper-tail surcharge in simulated variance. Its mean of 0.303 documents a material in-sample surcharge, while its negative association with subsequent variance surprise contradicts the prespecified positive prediction. This distinction makes CFR useful for describing the optimisation environment without treating it as a validated warning signal for future risk.

Bayesian portfolio approaches [14, 19] and bootstrap-based approaches represent uncertainty in different ways. The present results do not test Bayesian strategies or rank those frameworks. They instead demonstrate why uncertainty representation and objective selection should be evaluated separately: representing adverse scenarios does not establish that optimising a particular summary of those scenarios improves realised risk.

The findings are bounded by a single 49-industry research-portfolio universe, monthly rebalancing, long-only allocations, and the tested covariance and tail specifications. They do not establish a universal failure of robust optimisation or bootstrap inference. The academic portfolios also omit security-level execution frictions beyond the stated turnover-cost model. Multi-asset universes, alternative covariance estimators, nonlinear instruments, and other rebalancing frequencies require separate evaluation. These boundaries define where the reported counterexample has been demonstrated and where the validation approach can next be tested.

6. Conclusions

This study identifies tail-robustness inversion over 648 monthly rebalances on 49 US industry portfolios from 1972 to 2025. Bootstrap-tail optimisation matches or improves the benchmark's simulated objective at all 648 rebalances, yet raises average realised variance by 2.588% relative to Ledoit–Wolf minimum variance; the 95% block-bootstrap interval for the increase is [1.121%, 4.057%]. Annualised gross volatility is 11.90% versus 11.74%. The same inversion persists across all 12 reported specifications, with higher volatility in each of five chronological regimes.

The results also identify a practical mitigation: common diversification constraints reduce the variance penalty, although the remaining penalty stays statistically distinguishable from zero. Lower effective diversification and higher turnover help characterise the allocation changes accompanying the inversion. The covariance fragility ratio provides a complementary in-sample diagnostic, while its negative association with subsequent variance surprise establishes a clear boundary on its predictive interpretation.

The contribution is a reproducible demonstration that optimisation success and realised protection can diverge. Evaluating the simulated objective alongside realised risk, diversification, turnover, and specification sensitivity provides an actionable validation approach for covariance-based allocations before deployment.

REFERENCES

1. H. Markowitz, "Portfolio Selection," J. Finance, vol. 7, no. 1, pp. 77-91, 1952.
2. O. Ledoit and M. Wolf, "Honey, I Shrunk the Sample Covariance Matrix," J. Portfolio Mgmt., vol. 30, no. 4, pp. 110-119, 2004. DOI: 10.3905/jpm.2004.110.
3. R. Jagannathan and T. Ma, "Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps," J. Finance, vol. 58, no. 4, pp. 1651-1683, 2003.
4. R. T. Rockafellar and S. Uryasev, "Optimization of Conditional Value-at-Risk," J. Risk, vol. 2, no. 3, pp. 21-41, 2000.
5. D. N. Politis and J. P. Romano, "A Circular Block-Resampling Procedure for Stationary Data," in Exploring the Limits of Bootstrap, R. LePage and L. Billard, Eds. Wiley, 1992, pp. 263-270.
6. B. Efron and R. J. Tibshirani, An Introduction to the Bootstrap. Chapman and Hall/CRC, 1994. DOI: 10.1201/9780429246593.
7. O. Ledoit and M. Wolf, "A Well-Conditioned Estimator for Large-Dimensional Covariance Matrices," J. Multivariate Anal., vol. 88, no. 2, pp. 365-411, 2004.
8. P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath, "Coherent Measures of Risk," Math. Finance, vol. 9, no. 3, pp. 203-228, 1999.
9. V. DeMiguel, L. Garlappi, and R. Uppal, "Optimal Versus Naive Diversification: How Inefficient Is the 1/N Portfolio Strategy?" Rev. Financ. Stud., vol. 22, no. 5, pp. 1915-1953, 2009.
10. M. Britten-Jones, "The Sampling Error in Estimates of Mean-Variance Efficient Portfolio Weights," J. Finance, vol. 54, no. 2, pp. 655-671, 1999.
11. R. O. Michaud, "The Markowitz Optimization Enigma: Is Optimized Optimal?" Financ. Anal. J., vol. 45, no. 1, pp. 31-42, 1989.
12. C. Bongiorno and D. Challet, "Covariance Matrix Filtering with Bootstrapped Hierarchies," PLoS ONE, vol. 16, no. 1, article e0245092, 2021. DOI: 10.1371/journal.pone.0245092.
13. D. C. Oliveira, G. Guzman, and N. Firoozye, "(Non-Parametric) Bootstrap Robust Optimization for Portfolios and Trading Strategies," arXiv:2510.12725, 2025.
14. D. Bauder, T. Bodnar, N. Parolya, and W. Schmid, "Bayesian Mean-Variance Analysis: Optimal Portfolio Selection Under Parameter Uncertainty," Quant. Finance, vol. 21, no. 2, pp. 221-242, 2021. DOI: 10.1080/14697688.2020.1748214.
15. G. Coqueret and T. Guida, Machine Learning for Factor Investing. Chapman and Hall/CRC, 2020.
16. T. Roncalli, "Advanced Course in Asset Management (Presentation Slides)," SSRN 3773484, 2021. DOI: 10.2139/ssrn.3773484.
17. O. Ledoit and M. Wolf, "Analytical Nonlinear Shrinkage of Large-Dimensional Covariance Matrices," Ann. Statist., vol. 48, no. 5, pp. 3043-3065, 2020. DOI: 10.1214/19-AOS1921.
18. A. Kempf and C. Memmel, "Estimating the Global Minimum Variance Portfolio," Schmalenbach Business Rev., vol. 58, no. 4, pp. 332-348, 2006.
19. T. Bodnar, S. Mazur, and Y. Okhrin, "Bayesian Estimation of the Global Minimum Variance Portfolio," Eur. J. Oper. Res., vol. 256, no. 1, pp. 292-307, 2017. DOI: 10.1016/j.ejor.2016.05.044.
20. T. Bodnar, N. Parolya, and W. Schmid, "Estimation of the Global Minimum Variance Portfolio in High Dimensions," Eur. J. Oper. Res., vol. 266, no. 1, pp. 371-390, 2018. DOI: 10.1016/j.ejor.2017.09.028.

21. P. N. Kolm, R. Tutuncu, and F. J. Fabozzi, "60 Years of Portfolio Optimization: Practical Challenges and Current Trends," Eur. J. Oper. Res., vol. 234, no. 2, pp. 356-371, 2014.
22. M. Lopez de Prado, Advances in Financial Machine Learning. Wiley, 2018.
23. K. R. French, "49 Industry Portfolios," Kenneth R. French Data Library, Dartmouth College; daily value-weighted US industry portfolio return series.