

Research on Improving Markowitz Portfolio Optimization

Chenyan Luo

Ulink College of Shanghai, Shanghai, China
amyluochenyan@163.com

Abstract. This study is focused on how mean, covariance shrinkage and regularization methods works within Markowitz Portfolio Optimization based on a micro asset pool, in which return dispersion can be extreme and one or two winners can dominate the cross-section. The analysis proceeds in two stages. Firstly, this study is focused on how to make the estimation of mean and covariance matrix more accurately, including historical estimates, James-Stein mean shrinkage, Ledoit-Wolf covariance shrinkage, joint shrinkage, confidence-bound robust means, and Michaud resampling are compared over risk-aversion coefficients of 2, 3, and 5 and estimation windows of 1, 2, and 5 years. Secondly, based on estimation window of 1 year with risk-aversion coefficients of 2, this study introduced different regularization methods including higher moments, an L2 weight penalty, a 25% position cap, a turnover penalty, and short selling with gross exposure limited to 1.5. Every test includes an equal position sizing portfolio and S&P 500 ETF as baselines. Furthermore, to avoid the bias of asset choosing manually, this study made 500 resampled asset pool with 7 stocks from a 36-stock historical Dow Jones constituent union. The result shows that Ledoit-Wolf shrinkage increases little in performance in the chosen asset pool while 25% position cap attains the highest median Sharpe ratio, 0.724, and exceeds the 1/N Sharpe ratio in 70.0% of samples across random pools, which indicates that model complexity is not uniformly valuable in a micro-universe: a method improves performance only when its reduction in estimation error exceeds the cost of suppressing informative cross-sectional signals.

Keywords: Markowitz optimization, shrinkage estimation, portfolio regularization, repeated random subsampling

1. Theoretical foundation

Markowitz's contribution [1] was not a perfect return-prediction model. It was the first systematic algebraic formulation of security selection as a trade-off among expected returns, variances, and covariances. The same structure creates the model's central empirical weakness. Expected returns and covariances must be estimated from finite samples, and small input errors can be transformed into concentrated and unstable portfolio weights [2]. A useful evaluation of mean-variance optimization must therefore connect each source of estimation error to a specific remedy rather than treating model extensions as interchangeable improvements.

The literature offers three broad responses. James and Stein [3] established that shrinking multiple means toward a common target can reduce aggregate estimation loss; Jorion [4] adapted this logic to portfolio choice. Ledoit and Wolf [5] stabilized sample covariance matrices by shrinking them toward a structured target. Michaud [6] addressed optimizer sensitivity by repeatedly simulating plausible samples, solving each sample-specific portfolio, and averaging the resulting weights. Goldfarb and Iyengar [7] instead placed uncertain parameters in explicit uncertainty sets and selected portfolios for their worst admissible outcomes.

A second literature changes the decision problem rather than the estimated inputs. Jagannathan and Ma [8] showed that apparently incorrect position constraints can act as implicit covariance regularizers. DeMiguel et al. [9] controlled extreme allocations through portfolio-norm constraints, while Lobo et al. [10] incorporated transaction costs into the objective. DeMiguel et al. [11] demonstrated that complex optimizers must also be compared with the parameter-free 1/N portfolio. Without this benchmark, a sophisticated model can appear successful while failing to beat a feasible allocation exposed to the same asset opportunity set.

Most of this evidence is motivated by universes containing tens or hundreds of securities, where the asset dimension approaches or exceeds the time-series dimension and the sample covariance matrix becomes ill-conditioned. A micro-universe has a different statistical structure. With only seven assets, the covariance matrix can be full rank and reasonably conditioned, yet long-run returns may be sharply unequal. A concentrated solution can therefore represent either an estimation-error corner solution or a rational response to a persistent winner. Shrinkage and diversification constraints reduce variance, but they can also erase genuine information.

This paper asks three questions. First, do classical stabilization methods fail when a micro-universe is dominated by a small number of high-return securities? Second, where is the boundary between correcting estimation error and suppressing an economically useful signal? Third, which method remains stable when the asset pool, risk preference, and estimation window change? The study's contribution is not a claim that seven securities represent a diversified market portfolio. The small universe is used intentionally as a limiting experiment in which the costs and benefits of regularization are unusually visible.

2. Model specification and empirical design

Let the simple return of asset i at date t be

$$r_{i,t} = P_{i,t}/P_{i,t-1} - 1. \quad (1)$$

Simple returns are used because a self-financing portfolio return is exactly linear in beginning-of-period weights, $r_{p,t} = w_t^T r_t$. This property matches both the optimizer and the monthly wealth recursion without requiring a log-return approximation. The standard portfolio solves

$$\begin{aligned} \max_w \quad & w^T \hat{\mu} - \frac{\lambda}{2} w^T \hat{\Sigma} w, \\ \text{s.t.} \quad & \mathbf{1}^T w = 1, \quad w_i \geq 0. \end{aligned} \quad (2)$$

where $\hat{\mu}$ and $\hat{\Sigma}$ are daily estimates and λ is risk aversion. The first empirical stage changes only these inputs. James-Stein shrinkage compresses cross-sectional mean differences; Ledoit-Wolf shrinkage moves the covariance matrix toward a structured target; joint shrinkage applies both transformations. The robust portfolio replaces each mean by the lower confidence bound $\hat{\mu}_i - 1.96\hat{\sigma}_i/\sqrt{n}$.

$\sqrt{T}$. Michaud resampling draws synthetic return histories from the estimated multivariate distribution, optimizes each history, and averages the resulting weights.

The estimation window $L \in \{1, 2, 5\}$ represents a bias-variance trade-off. A short window adapts rapidly to regime changes but has high sampling variance. A long window reduces variance but can combine observations from economically different states. Risk aversion $\lambda \in \{2, 3, 5\}$ is also an error-allocation parameter. When λ is low, expected return receives more weight relative to covariance risk, exposing the optimizer more directly to unstable mean forecasts.

The second stage uses a controlled one-factor-at-a-time design. Every specification begins with the historical mean, sample covariance, full investment, and long-only baseline. It then adds only one of the following: a reward for portfolio skewness and a penalty for kurtosis; an L2 penalty $-\gamma \|w\|_2^2$; a hard bound $w_i \leq 25\%$; a turnover penalty $-c\|w_t - w_{t-1}\|_1$; or short selling. The short-selling portfolio is constrained by

$$\sum_i |w_i| \leq 1.5. \quad (3)$$

This gross-exposure limit prevents economically implausible long-short amplification when highly correlated assets create estimated near-arbitrage directions. Because each treatment differs from the baseline in exactly one respect, its performance difference can be interpreted as a marginal effect rather than the result of a cumulative constraint bundle.

The main universe contains The Coca-Cola Company, Exxon Mobil Corporation, Dow Inc., NextEra Energy, Inc., Realty Income Corporation, NVIDIA Corporation, and Alphabet Inc. Class A. Prices cover 2015 through September 16, 2026. Dow Inc.'s shorter history makes the common out-of-sample period begin in 2019. At each month-end, the model uses only information available through that date and holds the resulting weights during the next period. The main Michaud experiment uses 20 simulations per rebalance. Benchmarks are the monthly rebalanced 1/N portfolio formed from the same seven stocks and the SPDR S&P 500 ETF. Main-sample series begin with Yahoo Finance adjusted closes [12] and are extended through September 16, 2026, with scaled Nasdaq closing prices [13]. The random universe and SPY use split-adjusted Nasdaq prices from the same source vintage.

The robustness experiment forms a union of historical Dow Jones Industrial Average constituents observed since 2015 [14]. After history-coverage screening, 36 securities remain. Five hundred distinct seven-stock pools are sampled without replacement within each pool. Nasdaq price availability runs from September 2016 through September 2026. Every pool uses $\lambda = 2$, a 1-year window, and monthly rebalancing; Michaud uses five simulations per rebalance to control computation. Repeated sampling substantially reduces dependence on the original winner-heavy selection, although incomplete delisting returns prevent a claim that survivorship bias has been eliminated. Random-universe securities and SPY use the same split-adjusted price-return convention and exclude cash dividends. Reported strategy returns are gross of commissions, bid-ask spreads, market impact, and taxes.

Table 1. Research design and identification structure

Experiment	Methods	Common Benchmarks	Identified Effect
Stage 1	Six input estimators; $\lambda = 2, 3, 5$; 1-, 2-, and 5-year windows	Standard Markowitz, 1/N, SPY	Estimation method and parameter grid

Table 1. (continued)

Stage 2	Baseline plus, separately, higher moments, L2, a 25% cap, turnover, or bounded shorting	Standard Markowitz, 1/N, SPY	Marginal effect of one portfolio constraint
Random robustness	500 seven-stock pools from 36 historical constituents	1/N and price-return SPY	Asset-selection sensitivity and distributional stability

Note. Stage 1 changes only parameter estimation. Stage 2 changes exactly one portfolio feature at a time. The 1/N benchmark is rebalanced monthly within the same asset pool; SPY denotes the SPDR S&P 500 ETF.

3. Empirical results and robustness evidence

Table 2 reports the method with the highest Sharpe ratio within each risk-aversion/window cell. Ledoit-Wolf is marginally best for every risk preference under the 1-year window. Under the 2-year window, 1/N becomes the best method at $\lambda = 3$ and 5, showing that the optimized portfolios do not reliably dominate a simple benchmark. Across all nine parameter cells, the average Sharpe ratios of Michaud, Ledoit-Wolf, standard Markowitz, and 1/N are 1.231, 1.184, 1.179, and 1.043, respectively. Michaud's average rebalance turnover is 31.83%, compared with 0.28% for 1/N, so its gross-return advantage need not survive realistic implementation costs.

Table 2. Best strategy by risk aversion and estimation window

λ	Window	Highest-Sharpe Method	Method Return	Method Sharpe	1/N Return	1/N Sharpe	SPY Return
2	1 year	Ledoit-Wolf	58.88%	1.465	21.61%	1.134	15.20%
2	2 years	Michaud	36.86%	1.033	19.83%	1.030	14.72%
2	5 years	Standard Markowitz	61.68%	1.335	20.78%	0.964	14.86%
3	1 year	Ledoit-Wolf	55.80%	1.438	21.61%	1.134	15.20%
3	2 years	1/N	19.83%	1.030	19.83%	1.030	14.72%
3	5 years	Michaud	45.14%	1.302	20.78%	0.964	14.86%
5	1 year	Ledoit-Wolf	50.31%	1.457	21.61%	1.134	15.20%
5	2 years	1/N	19.83%	1.030	19.83%	1.030	14.72%
5	5 years	Michaud	34.18%	1.231	20.78%	0.964	14.86%

Note. The highest-Sharpe method is selected within each λ -window cell from the six input-estimation methods and 1/N. Returns are geometric annual returns. Sharpe ratios use a zero risk-free rate. Out-of-sample start dates differ by window, so horizontal comparisons are valid within rows.

Ledoit-Wolf's small gain has a direct mathematical explanation. In the 1-year windows, the median number of observations is 251 while $N = 7$. The median sample-covariance condition number is 22.89, so the matrix is full rank and does not display the severe pathology associated with high-dimensional problems. The median Ledoit-Wolf shrinkage intensity is only 7.06%, leaving

limited covariance noise to correct. In this fixed universe, expected-return uncertainty is more consequential than covariance instability. James-Stein and joint shrinkage reduce average returns to 12.73% and 12.98%, both below $1/N$. When cross-sectional returns are extremely dispersed, shrinking means toward a common center can misclassify a persistent winner as noise.

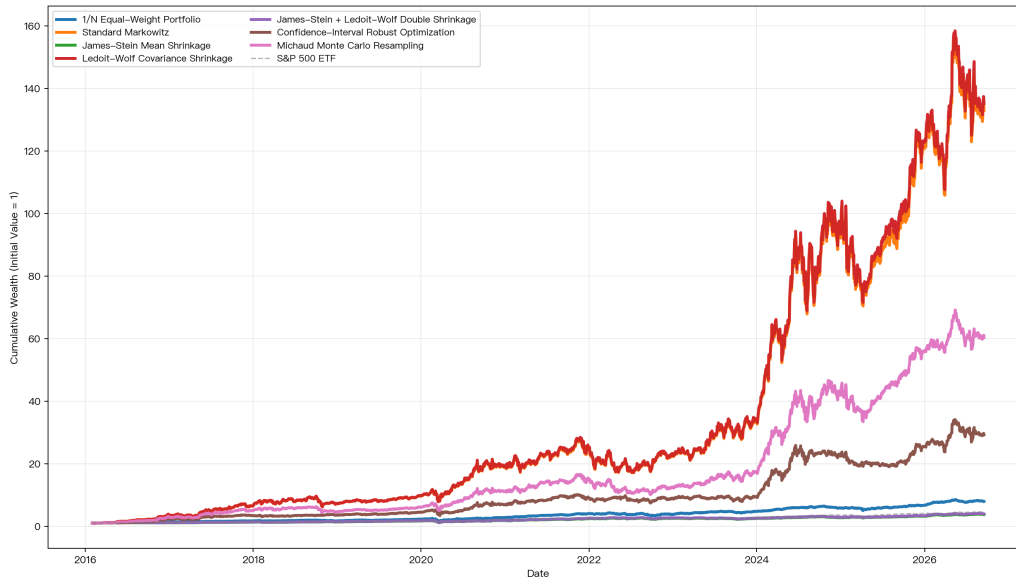


Figure 1. Out-of-sample cumulative wealth of input estimators, $1/N$, and the market benchmark

Note. Portfolios use $\lambda = 2$, a 1-year rolling window, and monthly rebalancing. Wealth is normalized to 1 on the common start date. Solid lines are portfolio strategies; the dashed line is the SPDR S&P 500 ETF. Results are gross of trading costs.

Stage 1 also identifies the 1-year window with $\lambda = 2$ as the experiment's maximum-error-exposure environment. It combines rapid signal updating with the largest relative weight on the estimated mean and therefore creates the strongest tendency to chase recent winners. Stage 2 fixes this setting and asks whether each constraint can stabilize the portfolio under deliberately adverse input noise. Table 3 shows that higher-moment terms leave the baseline nearly unchanged. L2 lowers the maximum weight from 91.09% to 79.85% and improves maximum drawdown from -39.01% to -33.67%, although its Sharpe ratio falls slightly. The 25% cap reduces volatility to 20.33% while retaining a return above both $1/N$ and SPY. The turnover penalty reduces average two-sided turnover from 23.90% to 7.75%, but maximum drawdown deteriorates to -58.34%; transaction smoothing is not a substitute for a risk constraint. Bounded shorting raises annual return to 65.84%, but volatility, drawdown, and turnover all increase, and the Sharpe ratio falls below the baseline. A leverage constraint prevents explosive long-short scaling without making the strategy stable.

Table 3. Marginal effects of individual portfolio constraints under maximum error exposure

Specification	Annual Return	Volatility	Sharpe	Maximum Drawdown	Monthly Turnover	Maximum Weight
Standard Markowitz	58.64%	40.08%	1.463	-39.01%	23.90%	91.09%
Higher moments only	58.55%	40.15%	1.458	-39.43%	24.04%	91.30%

Table 3. (continued)

L2 penalty only	55.15%	38.14%	1.446	-33.67%	27.65%	79.85%
25% cap only	26.62%	20.33%	1.310	-32.94%	22.97%	25.00%
Turnover penalty only	46.52%	42.95%	1.083	-58.34%	7.75%	88.13%
Bounded shorting only	65.84%	48.69%	1.352	-44.15%	47.74%	111.49%
1/N	21.61%	19.06%	1.134	-37.78%	0.22%	14.29%
SPY	15.20%	--	--	--	--	--

Note. Every optimized portfolio uses the historical mean, sample covariance, $\lambda = 2$, and a 1-year rolling window, with only the named feature changed. Turnover is the two-sided quantity $\sum_i |w_{i,t} - w_{i,t-1}|$. Bounded shorting satisfies $\sum_i |w_i| \leq 1.5$. Only concurrent annual return is reported for SPY.

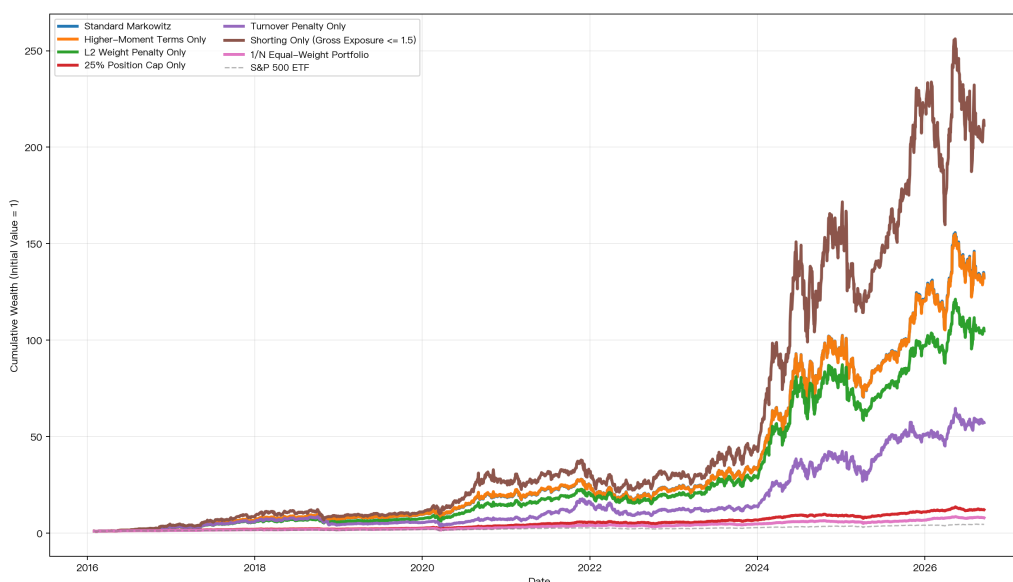


Figure 2. Out-of-sample cumulative wealth of individual constraints, 1/N, and the market benchmark

Note. Model parameters and the evaluation period match Table 3. Wealth is normalized to 1 at the common start. Each strategy differs from standard Markowitz by one portfolio feature, so curve differences identify that feature's marginal effect.

The fixed seven-stock result remains strongly affected by NVIDIA's trend. Table 4 and Figure 3 therefore report distributional evidence from 500 random asset pools. The 25% cap has a median annual return of 14.42%, below standard Markowitz's 15.64%, but its median Sharpe ratio of 0.724 is the highest among all methods. Its return and Sharpe ratio exceed same-pool 1/N in 78.2% and 70.0% of draws. The 1/N median Sharpe ratio is 0.632, ranking second and exceeding Michaud, L2, and standard Markowitz. Ledoit-Wolf and the standard model have nearly identical median Sharpe ratios, 0.534 and 0.538, again supporting the argument that covariance shrinkage has limited room to operate when N/T is small. Bounded shorting achieves the highest median return, 16.29%, but a median Sharpe ratio of only 0.491. Leverage increases return dispersion and tail exposure rather than generating a stable advantage.

Table 4. Distribution of performance across 500 random seven-stock pools

Method	Median Return	Median Sharpe	Return Beats 1/N	Sharpe Beats 1/N	Return Beats SPY
25% position cap	14.42%	0.724	78.2%	70.0%	63.4%
1/N	12.44%	0.632	--	--	43.6%
Michaud	15.38%	0.618	74.0%	43.4%	66.0%
L2 penalty	15.11%	0.586	75.2%	38.0%	63.8%
Standard Markowitz	15.64%	0.538	70.0%	30.8%	61.8%
Ledoit-Wolf	15.53%	0.534	70.0%	31.4%	61.8%
Higher moments	15.38%	0.535	69.4%	31.0%	60.6%
Turnover penalty	13.46%	0.545	57.8%	24.2%	50.4%
Bounded shorting	16.29%	0.491	69.2%	25.4%	63.0%
Joint shrinkage	8.55%	0.490	4.2%	12.0%	5.4%
James-Stein	8.20%	0.473	4.0%	11.4%	5.4%
Robust mean	11.44%	0.446	43.0%	15.6%	41.0%

Note. Each row contains 500 distinct seven-stock pools. A win rate is the fraction of pools in which the method's metric is strictly greater than same-pool 1/N or concurrent SPY. All portfolios use $\lambda = 2$, a 1-year rolling window, and monthly rebalancing. SPY comparisons use price returns excluding cash dividends.

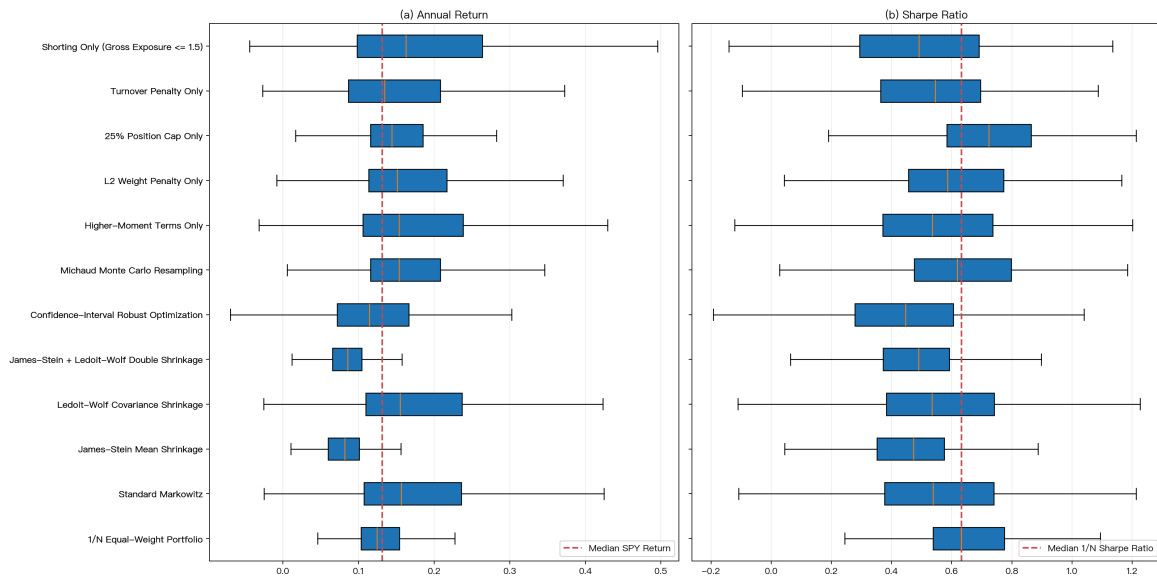


Figure 3. Annual return and Sharpe ratio distributions across 500 random seven-stock pools

Note. Boxes span the 25th to 75th percentiles, internal lines indicate medians, and whiskers extend to the most extreme observation within 1.5 interquartile ranges; outliers are not plotted. The dashed line in panel (a) is the median SPY annual return. The dashed line in panel (b) is the median 1/N Sharpe ratio.

4. Conclusion and boundary conditions

The two-stage design separates the role of input estimation from the role of portfolio constraints. In the fixed seven-stock universe, short-window Markowitz portfolios exploit strong trending assets and earn high gross returns. Ledoit-Wolf provides only a marginal adjustment because the sample covariance matrix is already well conditioned. Mean shrinkage can be harmful because it removes cross-sectional information that is unusually strong in a winner-dominated universe. The controlled experiment shows that L2 regularization and position caps reduce concentration and volatility. A turnover penalty solves a trading-demand problem rather than a risk problem, and leverage-bounded shorting does not consistently improve risk-adjusted performance.

Repeated random sampling changes the interpretation of the "best" method. Standard Markowitz performs exceptionally in the ex post winner sample, but the 25% cap supplies the highest and most stable risk-adjusted performance across random pools. The 1/N portfolio remains a difficult benchmark to defeat. The central result is therefore conditional: robust optimization in a micro-universe does not mean adding every available layer of complexity. Transparent diversification constraints should be considered first. More elaborate shrinkage and resampling methods are most likely to justify their cost when the asset dimension rises, the covariance matrix becomes materially ill-conditioned, or the investor's implementation objective changes.

Four limitations remain. The historical constituent union lacks complete delisting returns and therefore cannot eliminate survivorship bias. The random experiment uses price returns rather than dividend-inclusive total returns. Main results are gross of trading costs. Finally, the large number of method and parameter comparisons is not adjusted for multiple testing. Future work should use a survivorship-free database such as CRSP, incorporate net transaction costs, test performance differences relative to 1/N, and select hyperparameters in a nested out-of-sample procedure.

References

- [1] Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, 7(1), 77-91. <https://doi.org/10.1111/j.1540-6261.1952.tb01525.x>
- [2] Britten-Jones, M. (1999). The sampling error in estimates of mean-variance efficient portfolio weights. *The Journal of Finance*, 54(2), 655-671. <https://doi.org/10.1111/0022-1082.00120>
- [3] James, W., & Stein, C. (1961). Estimation with quadratic loss. In J. Neyman (Ed.), *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1, pp. 361-379). University of California Press. <https://projecteuclid.org/euclid.bsm/1200512173>
- [4] Jorion, P. (1986). Bayes-Stein estimation for portfolio analysis. *Journal of Financial and Quantitative Analysis*, 21(3), 279-292. <https://doi.org/10.2307/2331042>
- [5] Ledoit, O., & Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2), 365-411. [https://doi.org/10.1016/S0047-259X\(03\)00096-4](https://doi.org/10.1016/S0047-259X(03)00096-4)
- [6] Michaud, R. O. (1989). The Markowitz optimization enigma: Is 'optimized' optimal? *Financial Analysts Journal*, 45(1), 31-42. <https://doi.org/10.2469/faj.v45.n1.31>
- [7] Goldfarb, D., & Iyengar, G. (2003). Robust portfolio selection problems. *Mathematics of Operations Research*, 28(1), 1-38. <https://doi.org/10.1287/moor.28.1.1.14260>
- [8] Jagannathan, R., & Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The Journal of Finance*, 58(4), 1651-1683. <https://doi.org/10.1111/1540-6261.00580>
- [9] 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. <https://doi.org/10.1287/mnsc.1080.0986>
- [10] Lobo, M. S., Fazel, M., & Boyd, S. (2007). Portfolio optimization with linear and fixed transaction costs. *Annals of Operations Research*, 152(1), 341-365. <https://doi.org/10.1007/s10479-006-0145-1>
- [11] 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. <https://doi.org/10.1093/rfs/hhm075>

- [12] Yahoo Finance. (n.d.). *Historical data* [Data set]. Retrieved September 16, 2026, from <https://finance.yahoo.com/>
- [13] Nasdaq. (n.d.). *Historical stock data* [Data set]. Retrieved September 16, 2026, from <https://www.nasdaq.com/market-activity/stocks>
- [14] S&P Dow Jones Indices. (n.d.). *Dow Jones Industrial Average* [Data set]. Retrieved September 16, 2026, from <https://www.spglobal.com/spdji/en/indices/equity/dow-jones-industrial-average/>