Does Portfolio Optimization Beat Splitting Evenly?

Adrian Erlikhman Ryan Erlikhman
Los Angeles Center for Enriched Studies
erlikhman.adrian@gmail.com rerlikhman@gmail.com

2026-07-24

Abstract

Modern portfolio theory says an investor should hold assets in proportions that maximize the portfolio's reward-to-risk ratio—the Markowitz portfolio. A simpler approach holds equal amounts of each asset (the "$1/N$" portfolio). We evaluate both on a factor model calibrated to real stock market data. The $1/N$ portfolio returns more money. Across 250 simulations of ten assets with a ten-year estimation window, the $1/N$ portfolio earns a Sharpe ratio of 0.70 compared to the Markowitz portfolio's 0.51—with 385% annual volatility and 260 times the portfolio turnover—and wins 74% of the time. No strategy outperforms the $1/N$ portfolio; the best performers ignore the estimated expected returns for the assets. Given the true covariance matrix but estimated expected returns, the optimizer still trails $1/N$; given the true expected returns but an estimated covariance matrix, it is nearly optimal. Estimating expected returns is the critical difficulty. Covariance shrinkage and a no-shorting constraint each help in ways that match the theory, but shrinkage does nothing for the mean returns. With enough data—around two decades of returns that remain stationary over time—the shrinkage-based portfolio will overtake the $1/N$ portfolio, by a wider margin as the number of assets grows. This condition is rarely met in practice, though. Thus, optimization is not useless, but it is unlikely to help at the sample sizes markets provide. All code is provided for readers to test these strategies on live market data.

Introduction

In 1952, Harry Markowitz gave finance its first real theorem. Given the expected returns, variances, and correlations of a set of assets, he showed that there is a unique portfolio that would deliver the maximum expected return for a given level of risk (Markowitz, 1952). The theorem won a Nobel Prize, and it is the foundation of modern finance. From it emerges a seemingly logical corollary: if there is an optimal portfolio according to some rigorous set of mathematical rules, then any simpler strategy for building a portfolio must be missing out on some money.

The simplest possible portfolio to consider is the $1/N$ portfolio, which assigns an equal fraction of wealth to each asset. The $1/N$ portfolio requires no statistics, no mathematical equations to solve, and no forecasts of the future. Yet if the Markowitz model is any good at all, then the $1/N$ portfolio should be beaten by it. It is well known that it is not. DeMiguel et al. (2009) tested fourteen sophisticated portfolio optimization techniques against the $1/N$ portfolio on seven datasets, and found that none of them reliably beat the naive rule once the parameters had to be estimated from the data. In this paper, we reproduce those results in a more controlled setting, explain why they occur, and ask whether the fixes recommended for the optimization techniques actually fix the problem.

The answer is that optimization techniques tend to amplify estimation error. Markowitz's method assigns each asset a weight proportional to the inverse of the covariance matrix times the vector

of expected returns, $\Sigma^{-1}\mu$. Both of these quantities must be estimated from observed data, and the estimation of expected returns carries an inherent problem: small differences in the expected return of an individual asset are difficult to observe over short time periods. As Merton (1980) showed, observing the expected return of any asset with any precision requires an extremely long observation period, because the signal is tiny next to the noise. An optimizer that does not know of the noise in these estimates will end up adding money to assets that performed well by chance alone; it will become, as Michaud (1989) calls it, an error-maximizing machine. The $1/N$ portfolio, in contrast, does not make any estimates, and so it cannot be fooled in this way.

We make three contributions. First, we re-create the DeMiguel–Garlappi–Uppal phenomenon using linear algebra and Monte Carlo simulations. Second, using an *oracle decomposition*, we evaluate the impact of using estimated rather than true values of the inputs to the portfolio optimizer, separately for each of the four possible combinations of estimated versus true expected returns and covariance matrix. The underlying insight is not new—Chopra and Ziemba (1993) estimated that errors in expected returns are roughly an order of magnitude more costly than errors in variances—but we reproduce it in the same out-of-sample setting in which $1/N$ is evaluated, so that the mechanism and the horse race are measured on one common scale. Third, we evaluate two remedies for the estimation problem: shrinkage estimators for the covariance matrix and expected returns (Ledoit and Wolf, 2004; Jorion, 1986), and the no-short-sale constraint.

In designing the simulations, we were careful to observe the single most important habit in empirical finance: to test every strategy out of sample, and always to include a naive strategy as a baseline. Studies that skip this routinely report fantastical performance that evaporates on contact with new data. Every number reported in this paper was calculated on data that the strategy did not use to choose its portfolio weights, and $1/N$ serves as the yardstick throughout.

Background and related work

Mean–variance optimization.

Suppose there are N assets in a portfolio, with the weights represented by the vector $w \in R^N$. The constraint that the weights sum to one means that w must satisfy $w^\top 1 = 1$. Suppose further that the assets have expected excess returns μ and covariance matrix Σ, so that the portfolio has expected excess return $w^\top \mu$ and variance $w^\top \Sigma w$. Then the mean–variance investor with risk aversion γ aims to solve

$$\max_{w : w^\top 1 = 1} \; w^\top \mu - \frac{\gamma}{2} w^\top \Sigma w.$$

The solution to this optimization problem is the well-known tangency (or "maximum-Sharpe") portfolio

$$w_{MV} = \frac{\Sigma^{-1}\mu}{1^\top \Sigma^{-1}\mu}.$$

In the special case in which the investor ignores expected returns entirely and simply minimizes portfolio variance subject to being fully invested—equivalently, the limit of as risk aversion γ grows without bound—the resulting portfolio is the global minimum-variance (GMV) portfolio

$$w_{GMV} = \frac{\Sigma^{-1}1}{1^{\top}\Sigma^{-1}1}.$$

The comparison between these two portfolios is the main experimental lever of this paper. The GMV portfolio requires only the covariance matrix Σ, whereas the tangency portfolio additionally requires the vector μ of expected returns. If both quantities were known exactly, the tangency portfolio would have the higher Sharpe ratio by construction. Comparing them out of sample therefore isolates the cost of having to estimate μ: whenever the GMV portfolio outperforms the tangency portfolio in practice, that gap is attributable to estimation error in μ.

Why estimation error in the expected return is fatal.

Suppose an asset has a true expected excess return of $\mu = 8\%$ per year and volatility $\sigma = 20\%$ per year. The standard error of the mean estimated from M years of data is $\sigma/\sqrt{M}$, so estimating μ with a standard error of just 2% per year requires $M = (20/2)^2 = 100$ years of returns. Volatility, by contrast, can be estimated well from a short calendar span, because the precision of a variance estimate improves with the *number* of observations and so can be sharpened by sampling monthly or daily; the precision of a mean estimate depends only on the calendar length of the sample and cannot be improved by sampling more often within it (Merton, 1980). The major issue with mean–variance optimization is therefore that the parameter most important to the problem is the one most difficult to estimate. Consequently, weights proportional to $\Sigma^{-1}\mu$ will tend to assign very large positions to assets whose expected returns are estimated with noise, since small differences in the estimated returns of similar assets are greatly amplified by the inversion of the covariance matrix. Best and Grauer (1991) discovered that mean–variance weights are exquisitely sensitive to small changes in the estimated expected returns. Chopra and Ziemba (1993) studied the impact of this sensitivity on investor welfare, finding that estimation errors in expected returns are roughly an order of magnitude more detrimental than estimation errors in the volatilities of the assets. These results are explored and verified in Experiment C.

The naive benchmark.

DeMiguel et al. (2009) found, through analysis of multiple datasets of real-world assets, that it is difficult to find a mean–variance portfolio that outperforms the simple $1/N$ rule for allocating weights to each of N assets. They also derived, for a stylized setting, the length of history an investor must observe before the estimation error in the mean–variance problem shrinks enough for the optimizer to overtake $1/N$; for realistic numbers of assets that "critical length" runs to hundreds or thousands of years. Our simulation study reveals the same effects, and in addition lets us measure how they depend on the number of assets N and the length of the estimation window M.

Remedies.

If the inputs are noisy, the first thing to do is to shrink the noisy estimates toward something stable. Ledoit and Wolf (2004) shrink the sample covariance matrix toward a target matrix with a

particular structure (a constant-correlation matrix), introducing bias in the estimate but reducing its variance and ensuring that it will be invertible. Jorion (1986) applies the same idea to the estimation of expected returns, shrinking each asset's estimated mean toward a common grand mean. Another approach is to place constraints on the weights: requiring that the portfolio weights be non-negative ($w \geq 0$) prevents the portfolio from making short sales of any of the constituent assets. Jagannathan and Ma (2003) show that imposing this constraint has an effect similar to a shrinkage estimator. We test all three approaches.

Methods

Data: a calibrated simulation, and why

For each strategy, we evaluate its performance on returns drawn from a factor model whose parameters are calibrated to the moments of real equity returns. Specifically, we model each asset's excess monthly return as

$$r_t = \mu + B f_t + \varepsilon_t, \qquad f_t \sim N\left(0, \Sigma_f\right), \quad \varepsilon_t \sim N(0, D),$$

where the factor loadings B are centered at one and the idiosyncratic variances D are chosen to achieve target levels of total volatility and pairwise correlation among the assets in the simulated universe. The resulting return distribution has covariance $\Sigma = B\Sigma_f B^{\top} + D$ and mean μ, where each asset's mean is chosen so that its annual returns achieve a target Sharpe ratio. Overall, the simulated assets feature an annual volatility of 0.28, an average correlation of 0.27 between each pair of assets, and an average Sharpe ratio of 0.40 per asset—all within the range of the volatilities, correlations, and Sharpe ratios of actual equity assets.

One potentially undesirable feature of such a construction is that each asset's expected return is chosen independently of the other assets. As a result, the Sharpe ratio of the optimal portfolio increases quickly with the number of assets: across the universes drawn for Experiment B, the true optimal portfolios average 1.04 for $N = 10$, 1.94 for $N = 25$, and 3.73 for $N = 50$. Sharpe ratios of that size are obviously not attained by any investor in the real market; they are an artifact of the assumption that expected returns are drawn independently of one another. The direction of the distortion matters, however: it makes the prize for successful optimization *larger* than in reality, so the finding that the optimizer still loses to $1/N$ is a conservative one. Comparisons across different values of N should be read with this in mind, which is why our headline result fixes $N = 10$, where the true tangency Sharpe ratio of 1.03 is realistic.

Rather than using actual market return data, we use simulated data for several reasons. First, because we control the true parameters of the data (μ and Σ), we can calculate the optimal portfolio and measure the error that results from estimating those parameters rather than knowing them; this is impossible with real market data, where the truth is unknown. Second, rather than testing a strategy on a single set of historical prices, simulating returns over hundreds of universes allows for the estimation of the systematic behavior of each strategy, and averaging over many simulations allows for the calculation of genuine standard errors. Third, the parameters of the simulated data are actually quite generous to the optimizer: returns are drawn independently *over time* from a stationary distribution, so there is no regime change or parameter

drift for the optimizer to be fooled by. In real markets prices and returns are not stationary, so this aspect of the design works to make the paper's finding stronger—the failure of the optimizer in this simulation is a lower bound on its difficulty in the real market. Finally, the code that accompanies this paper includes a loader built on the yfinance Python library, which allows the identical pipeline to be applied directly to real market data.

Strategies

Table summarizes the seven rules that we compare. For each rule, we first form an estimate of the average return within the estimation window, denoted $\hat{\mu}$, and an estimate of the covariance matrix of asset returns within that window, denoted $\hat{\Sigma}$. We then plug these values into one of the weight formulas given above, optionally substituting shrunk estimates of the mean or covariance matrix.

The seven portfolio rules. "LW" denotes Ledoit–Wolf covariance shrinkage; "BS" denotes Bayes–Stein mean shrinkage. The min-variance rules ignore expected returns entirely.

Strategy	**Inputs used**	**Weight rule**
1/N	none	w=1/N
MV (sample)	μ,Σ	Eq.
MV (long-only)	μ,Σ	Eq. s.t. w≥0
MV (LW cov)	μ,ΣLW	Eq. with shrunk Σ
MV (LW cov + BS mean)	μBS,ΣLW	Eq. , both shrunk
Min-Var (sample)	Σ	Eq.
Min-Var (LW cov)	ΣLW	Eq. with shrunk Σ

The Ledoit–Wolf estimator replaces the sample covariance matrix with a convex combination of that matrix and a target in which every pairwise correlation is replaced by the *average* sample correlation across all pairs of assets in the estimation window. The intensity of this shrinkage is chosen to minimize the expected squared error of the covariance estimate (the exact formula is given in Appendix). The Bayes–Stein estimator shrinks the sample mean vector toward the mean return of the global minimum-variance portfolio, with an intensity likewise chosen from the data (Jorion, 1986).

Out-of-sample protocol

We follow the rolling-window procedure of DeMiguel et al. (2009) exactly. For each month t from M onward, we estimate the parameters from the M most recent months of returns (months $t - M$ through $t - 1$), form the weights w_t, and record the portfolio return $w_t^\top r_t$ in the next month. No information from month t or any later month is used to form w_t, so there is no look-ahead bias. For each strategy this yields a time series of out-of-sample returns.

Performance metrics

We report the annualized out-of-sample Sharpe ratio, $SR = \sqrt{12}\, r^{-}/s_r$, where r^{-} and s_r are the mean and standard deviation of the out-of-sample monthly returns. We also report the annualized volatility of those returns; the certainty-equivalent return $CEQ = 12\left(r^{-} - \frac{\gamma}{2}s_r^2\right)$ with $\gamma = 1$, which is the guaranteed return a mean–variance investor would accept in place of the risky strategy; and turnover, the average fraction of the portfolio traded each period, which is a proxy for the transaction costs the strategy would suffer in practice.

Monte Carlo design

In each simulation we create a new "universe" of assets with a new set of true parameters, and we generate a new return history for those assets. The averages across simulations therefore reflect the phenomenon in general, rather than within a single universe of assets. Experiment A is performed with $N = 10$ assets, a window of $M = 120$ months (ten years), 180 months of out-of-sample data, and 250 replications. Experiment B uses the same protocol with $N \in \{10, 25, 50\}$ and $M \in \{60, 120, 240\}$ and 200 replications. Experiment C uses $N = 10$, $M = 120$, and 300 replications. Every replication draws its own seed, and all seeds are fixed and recorded, so the pipeline is deterministic and exactly reproducible.

Results

The headline: naive beats optimal

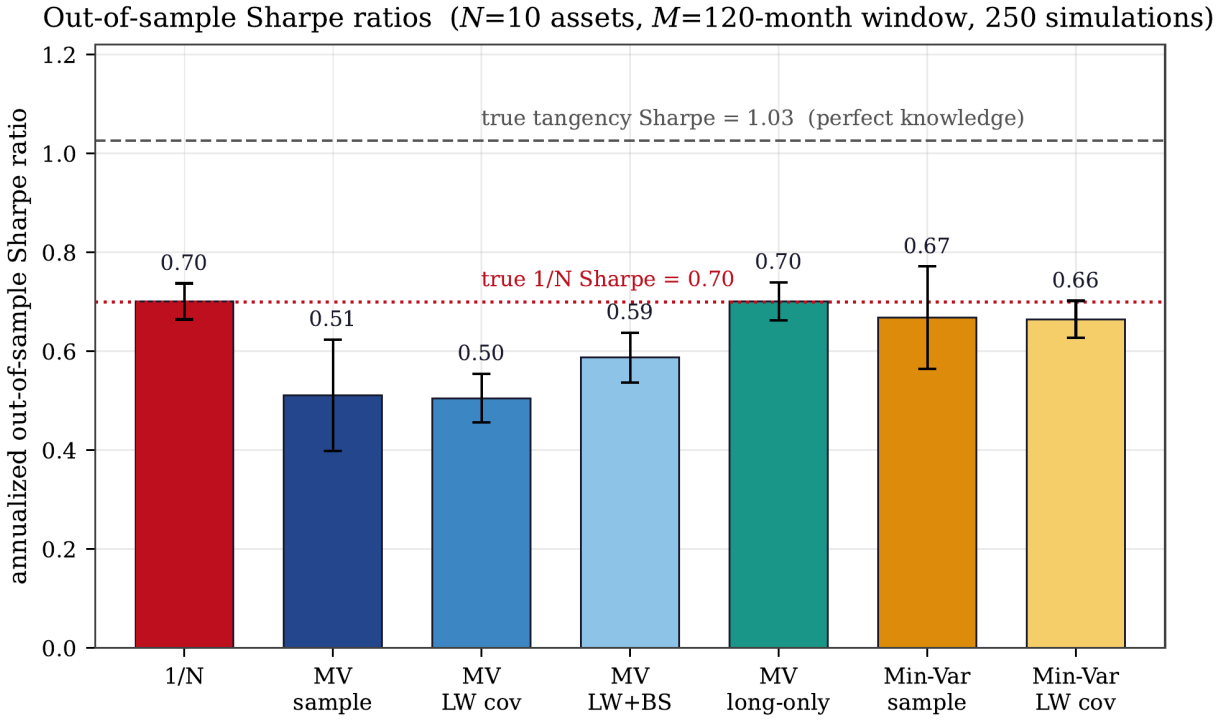
Table and Figure tell the story. For ten assets, over a ten-year estimation window, the naive $1/N$ portfolio achieves an out-of-sample Sharpe ratio of 0.70. The "optimal" portfolio constructed from the sample of realized returns within that ten-year window achieves only 0.51—while boasting a volatility of 385% and a certainty equivalent of -203, and trading its own value more than thirteen times per month (13.5). The theoretically optimal strategy is therefore beaten by a strategy that requires no statistical estimation at all, and takes on far more risk to lose. On average the $1/N$ portfolio outperforms the sample Markowitz portfolio by 0.19 in Sharpe ratio. That difference is highly significant (Wilcoxon $p \approx 3 \times 10^{-14}$, paired t-test $p \approx 4 \times 10^{-4}$), and $1/N$ wins in 74% of individual simulations.

No strategy in the table *beats* the $1/N$ portfolio. Two of them—the no-short-sale portfolio and the sample minimum-variance portfolio—have Sharpe ratios statistically indistinguishable from $1/N$ ($p = 0.68$ and $p = 0.34$), so they tie with the naive rule rather than beat it. The three rules based on unconstrained estimates of expected returns all have Sharpe ratios between 0.50 and 0.59. The remaining rule, the minimum-variance portfolio built on the Ledoit–Wolf covariance estimate, scores 0.66 and so trails $1/N$ by only 0.04; with 250 simulations even that small gap is statistically detectable ($p = 2 \times 10^{-4}$), which is a useful reminder that statistical significance and economic importance are not the same thing. The ordering is governed by exposure to estimation error in expected returns: every rule that avoids or constrains that estimate

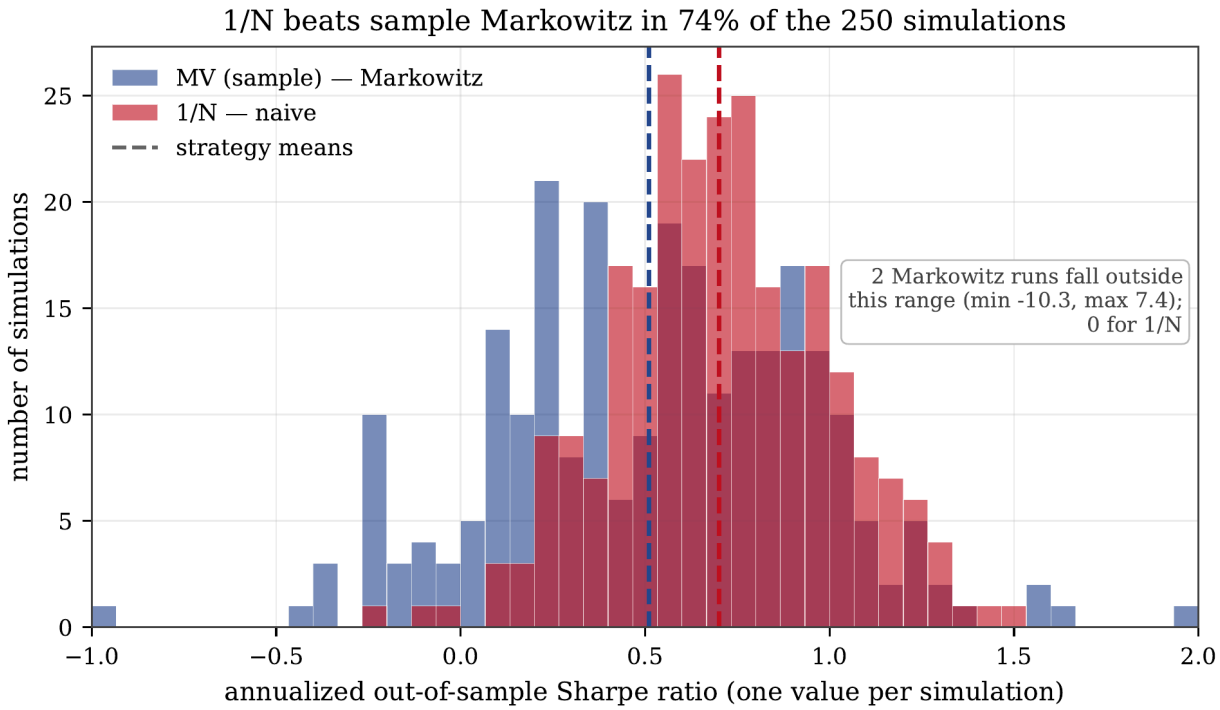
sits within 0.04 of $1/N$ in Sharpe ratio, while every rule that acts on it unconstrained falls at least 0.11 below.

Headline results (Experiment A: $N = 10$, $M = 120$, 250 simulations). Sharpe ratio, annualized volatility, certainty-equivalent return, and average monthly turnover are means across simulations. $Pr(1/N\ wins)$ is the fraction of simulations in which $1/N$ achieves a higher out-of-sample Sharpe ratio than the strategy in that row; p is the two-sided paired Wilcoxon test of the per-simulation Sharpe ratio difference between $1/N$ and that strategy. Note that MV (long-only) and Min-Var (sample) are statistically indistinguishable from $1/N$ ($p = 0.68$ and $p = 0.34$): they tie with the naive rule rather than beat it. For reference, the average true tangency Sharpe ratio is 1.03, and the true Sharpe ratio of $1/N$ is 0.70.

Strategy	Sharpe	Volatility	CEQ	Turnover	Pr(1/N wins)	p
1/N	0.70	0.16	0.098	0.05	—	—
MV (sample)	0.51	3.85	−203.2	13.51	0.74	3×10−14
MV (long-only)	0.70	0.19	0.112	0.10	0.50	0.68
MV (LW cov)	0.50	2.98	−96.0	14.40	0.78	2×10−19
MV (LW cov + BS mean)	0.59	1.46	−14.9	4.87	0.68	6×10−10
Min-Var (sample)	0.67	0.15	0.091	0.10	0.49	0.34
Min-Var (LW cov)	0.66	0.15	0.088	0.06	0.60	2×10−4

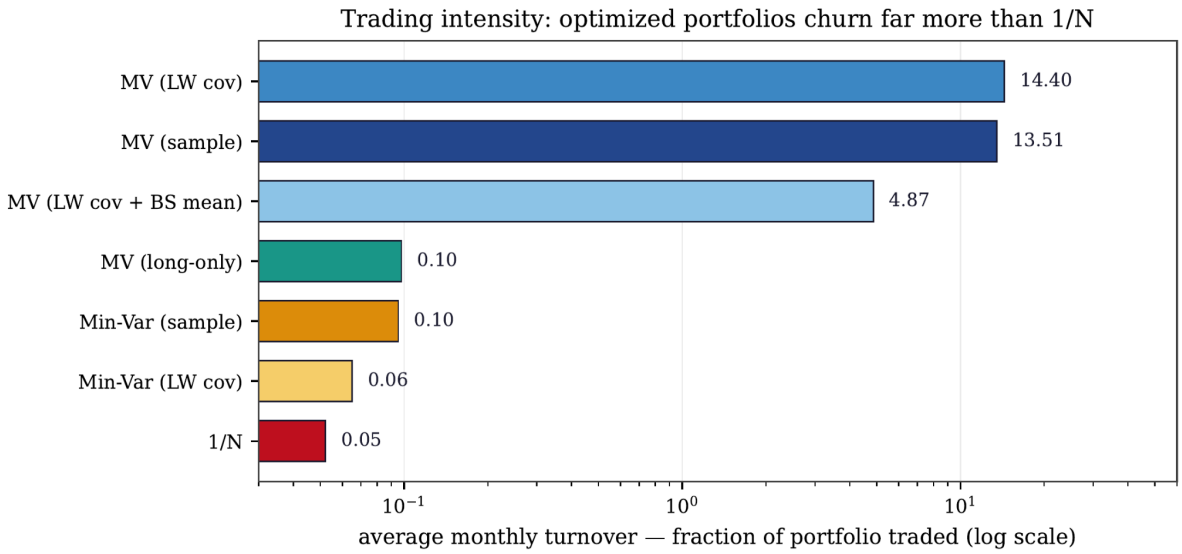


Out-of-sample Sharpe ratios by strategy (error bars are 95% *confidence intervals across simulations). The dashed line is the Sharpe ratio an investor could earn if the true parameters were known; the dotted line is the true* $1/N$ *Sharpe ratio. Every mean–variance variant that acts on unconstrained estimated expected returns falls short of* $1/N$*; only the rules that discard or constrain the mean estimate (Min-Var, long-only) catch up.*



Distribution of the out-of-sample Sharpe ratio across the 250 *simulations for* $1/N$ *(red) and the sample Markowitz rule (blue). The Markowitz distribution shows both a leftward shift and far greater dispersion, the signature of a rule whose weights are driven by noise. Vertical lines indicate the mean of each distribution. The horizontal axis is restricted to* $[-1, 2]$ *for clarity; two Markowitz simulations fall outside that range, at* -10.3 *and* 7.4*, while no* $1/N$ *simulation does. The wild fluctuations of the Markowitz rule run in both directions, and an investor cannot know in advance which tail they will land in.*

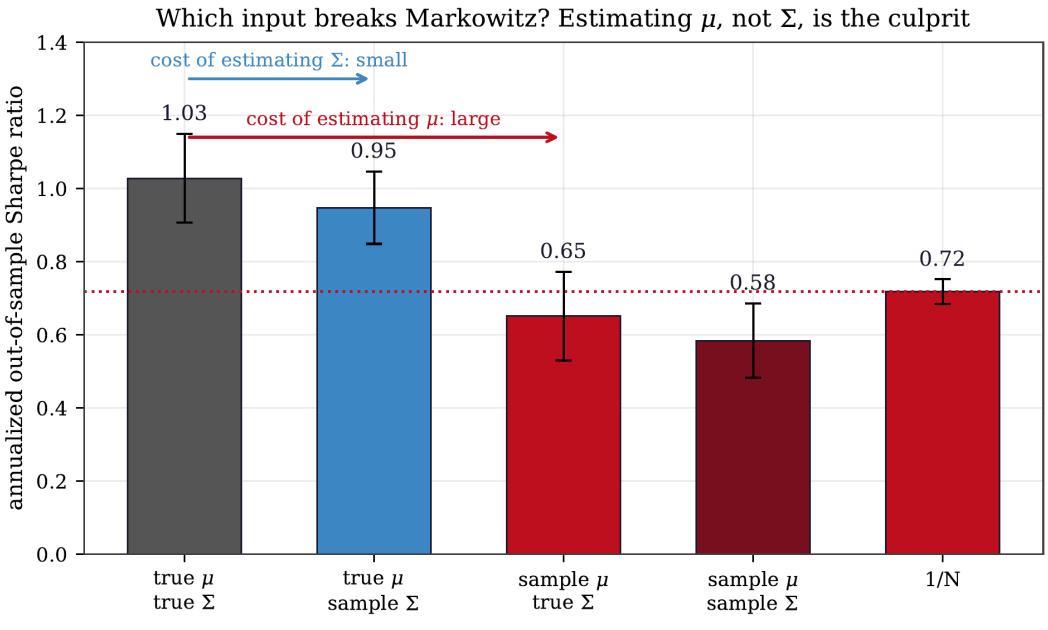
Two features of Table are worth noting. First, while the Sharpe ratio of the sample Markowitz rule (0.51) is only mediocre, its volatility (385%, some 24 times that of $1/N$) is enormous, and its certainty equivalent (-203) indicates that it is taking very large leveraged long and short positions; the turnover of 13.5 is roughly 260 times that of $1/N$, so the strategy would be wiped out by trading costs under any realistic cost per trade. Figure displays the turnover of each strategy on a log scale. Second, the two minimum-variance rules, which do not estimate expected returns at all, are close to $1/N$ in Sharpe ratio and nearly identical to it in volatility. That single contrast already points to the culprit.



Average monthly turnover (log scale). The optimized portfolios that use estimated means trade one to two orders of magnitude more than $1/N$ or the minimum-variance rules, because their weights chase estimation noise from month to month.

The mechanism: it is the means

Experiment C settles the diagnosis. Figure feeds the tangency formula the true and estimated moments in all four combinations and plots the resulting out-of-sample Sharpe ratio. Using the true expected returns and the true covariance matrix, the portfolio achieves a Sharpe ratio of 1.03—the attainable ceiling. Using the true expected returns and an *estimated* covariance matrix gives 0.95, very close to that ceiling. But using the true covariance matrix and *estimated* expected returns collapses performance to 0.65, below the 0.72 that $1/N$ achieves in the same experiment. (The universes and return paths in Experiment C were generated independently of those in Experiment A, which is why $1/N$ reads 0.72 here against 0.70 in Table ; the difference is Monte Carlo noise of the expected size, and within each experiment all strategies face the same set of universes and paths.) Estimating both inputs gives 0.58. The damage is overwhelmingly attributable to the estimation of expected returns; the covariance matrix is essentially a spectator.



Oracle decomposition (Experiment C). Each bar feeds the Markowitz formula a different combination of true and estimated inputs. Estimating the covariance matrix (second bar) costs almost nothing; estimating expected returns (third bar) is what breaks the optimizer, dragging it below the naive $1/N$ *benchmark (last bar).*

This finding explains the rest of the paper. The minimum-variance portfolio does well thanks to not estimating means. The long-only constraint helps because it limits how much a noisy mean estimate can move the weights. And, as we see next, shrinking the covariance does not rescue Markowitz on its own, because the covariance matrix was never the problem.

Do the fixes work? Ledoit–Wolf, constraints, and mean shrinkage

The fixes work exactly as the mechanism predicts.

Covariance shrinkage helps the covariance side of the problem but not the mean side. Table shows that using the Ledoit–Wolf estimator in place of the sample covariance matrix has only a marginal effect on the Sharpe ratio of the tangency portfolio, moving it from 0.51 to 0.50. In situations with many assets and short histories, however, the picture changes completely. At $N = 50$ assets estimated over $M = 60$ months—barely enough data to fit a 50×50 covariance matrix—the sample Markowitz portfolio collapses to a Sharpe ratio of 0.01, while the Ledoit–Wolf version reaches 0.33; over the same cells the minimum-variance portfolio rises from 0.33 to 0.63 (Table and Figure). Covariance shrinkage is therefore essential for high-dimensional data with short histories, and close to irrelevant when the sample covariance matrix is already well estimated.

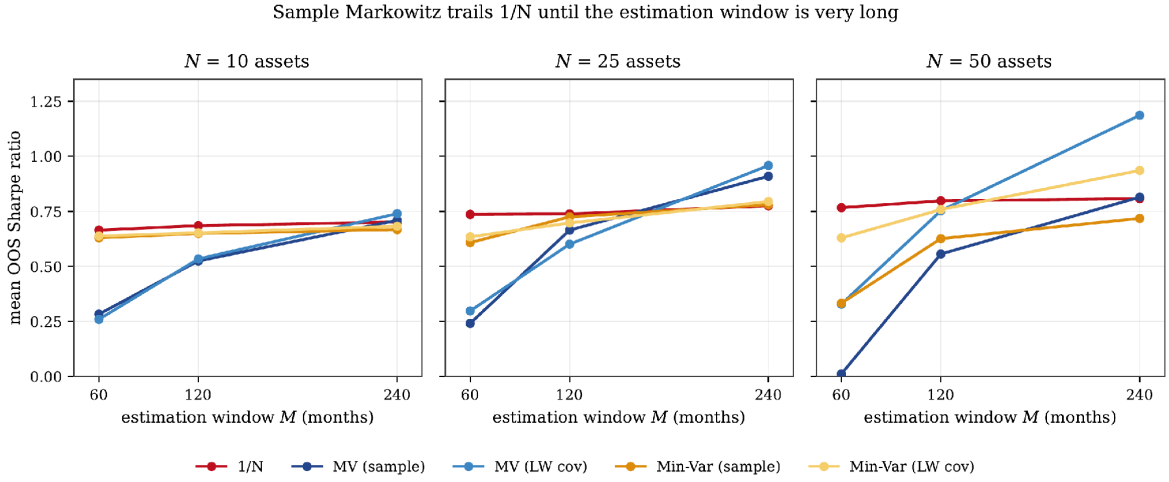
Shrinking the means helps where shrinking the covariance did not. In Table , the mean–variance portfolio that uses both the Ledoit–Wolf covariance matrix and Bayes–Stein mean shrinkage outperforms the one that shrinks only the covariance (0.59 against 0.50). Its volatility drops from 298% to 146%, its certainty equivalent improves from -96 to -15, and its turnover falls to roughly a third of its former value. So although the doubly shrunk portfolio still does not beat $1/N$, it is at least moving in the right direction, because it is finally attacking the right input.

Constraints are the most effective of the simple fixes. The long-only constraint, which prevents short sales, raises the Sharpe ratio of the Markowitz portfolio from 0.51 to 0.70 with far more reasonable volatility and turnover, leaving it statistically tied with $1/N$ ($p = 0.68$). The constraint works not by improving the estimates but by mechanically preventing the optimizer from acting on their worst errors, which is the mechanism identified by Jagannathan and Ma (2003).

Sensitivity grid (Experiment B): mean out-of-sample Sharpe ratios. "True tan." is the Sharpe ratio attainable with perfect knowledge. Sample Markowitz is worst when the window M is short and the number of assets N is large, and reaches $1/N$ *only at the longest windows. Ledoit–Wolf shrinkage helps most in the cells where N is large and M small.*

N	M	True tan.	1/N	MV (sample)	MV (LW)	Min-Var	Min-Var (LW)
10	60	1.04	0.66	0.28	0.26	0.63	0.64
10	120	1.04	0.68	0.52	0.53	0.65	0.65

N	M	True tan.	1/N	MV (sample)	MV (LW)	Min-Var	Min-Var (LW)
10	240	1.04	0.70	0.71	0.74	0.67	0.68
25	60	1.94	0.74	0.24	0.30	0.61	0.63
25	120	1.94	0.74	0.66	0.60	0.72	0.70
25	240	1.94	0.77	0.91	0.96	0.78	0.79
50	60	3.73	0.77	0.01	0.33	0.33	0.63
50	120	3.73	0.80	0.56	0.75	0.63	0.76
50	240	3.73	0.81	0.81	1.19	0.72	0.94



Out-of-sample Sharpe ratio versus estimation-window length M, for $N = 10, 25, 50$ *assets (all panels share a common vertical scale). Sample Markowitz (dark blue) starts far below* $1/N$ *(red) and climbs as more data become available, reaching* $1/N$ *only around a* 20*-year window. The collapse at short windows worsens sharply with N—at* $N = 50$, $M = 60$ *the sample optimizer is essentially worthless—which is precisely the regime where Ledoit–Wolf shrinkage (light blue) rescues it. Minimum-variance rules (orange) sit close to* $1/N$ *throughout because they never estimate expected returns.*

How much history would Markowitz need?

Figure also quantifies the "critical length" idea of DeMiguel et al. (2009). In our rather idealized simulation, it takes 240 months of history—twenty years—before the sample Markowitz portfolio reaches the $1/N$ portfolio.

What governs this outcome is the number of observations relative to the number of parameters to be estimated. The covariance matrix alone contains $N(N + 1)/2$ distinct entries, so the parameter count grows with the square of the number of assets while the data grow only with M. For short histories—an estimation window of 60 months—the performance of the optimized portfolio declines sharply as the universe grows, from 0.28 at $N = 10$ to 0.01 at $N = 50$, because there are barely more observations than parameters to fit.

At long estimation windows the ordering reverses. With $M = 240$ months, the margin by which the shrinkage-based mean–variance portfolio beats $1/N$ actually widens as the universe grows: 0.74 against 0.70 at $N = 10$, 0.96 against 0.77 at $N = 25$, and 1.19 against 0.81 at $N = 50$. A larger universe offers more opportunity to diversify, and an optimizer with enough data can exploit it. The honest summary is therefore conditional: given a long enough window, optimization wins, and it wins by more when there are more assets; below that threshold it loses badly, and loses worse when there are more assets.

The twenty-year figure applies only to stationary data, and markets are not stationary over twenty years. Expected returns drift, correlations spike in crises, and the parameters the optimizer is chasing are a moving target, so the effective sample is far shorter than the calendar sample. Only in the best case will an optimizer ever see twenty years of stationary data.

Discussion

Our results are a controlled reproduction of DeMiguel et al. (2009) and a direct measurement of its cause. Three lessons stand out.

First, *the failure is statistical, not mathematical*. Markowitz's theory is correct given the true parameters—our "true, true" oracle outperforms every other strategy—and the optimizer's sensitivity to μ is a virtue when μ is known. The problem arises only in the noisy estimation of the mean return vector from data.

Second, *the naive $1/N$ rule is a form of extreme shrinkage*. It can be read as the limit of shrinking every estimate toward a common value, until no asset looks different from any other. The excellent performance of this strategy implies that at realistic sample sizes the optimal amount of shrinkage is very large—close to the limit where the estimates are ignored entirely.

Third, *the fixes are partial and targeted*, and we should be careful not to overstate the case against optimization. The Ledoit–Wolf shrinkage estimator helps, though only for the covariance matrix and only where the covariance matrix is hard to estimate from the available data. Crucially, our own grid of values for N and M contains a regime in which the shrinkage-based mean–variance rule *outperforms* $1/N$: at $N = 50$ and $M = 240$, its Sharpe ratio is 1.19 against 0.81 for $1/N$. The appropriate conclusion is therefore not that optimization never works, but that it requires far more data than is ordinarily assumed in practice. With insufficient data it is worse than doing nothing clever at all.

Limitations.

Our simulations assume i.i.d. Gaussian returns generated from a stationary factor model. In reality returns have fat tails, time-varying volatility, and parameters that change over time; each of these characteristics makes estimation more difficult and would tend to increase the advantage of the naive $1/N$ rule, so our conclusions are conservative. We measure the turnover of each strategy but do not subtract transaction costs from returns; doing so would only further favor $1/N$. We do not include a risk-free asset or leverage constraints beyond the long-only case, and we consider only a single covariance-shrinkage target and a single mean-shrinkage estimator. Finally, because the study is based on simulation, the numbers reported are calibrated to—not

drawn from—market data; the accompanying code performs the same steps on real market data obtained through the yfinance library for readers who would like to see the results reproduced on market history.

Future work.

It would be of interest to perform the same study on real returns from market sectors or individual countries; to include transaction costs in the calculation of each strategy's returns; to study non-Gaussian and regime-switching data-generating processes that may better represent market returns; and to evaluate more recent estimators of the mean and covariance matrix of asset returns, including nonlinear shrinkage and factor-model covariances.

Conclusion

Does Markowitz mean–variance optimization beat the even split? At realistic sample sizes it does not: across ten assets and a decade of monthly returns, the even split matches or beats every variant of mean–variance optimization we tried, and it does so with a fraction of the risk and trading costs.

The reason is that the optimizer is broken by noise in the estimated expected returns—not by the covariance matrix—and the usual fixes help exactly to the extent that they prevent the optimizer from blindly trusting that noisy estimate.

This finding is a statement about data, not about mathematics. With enough stationary data—about twenty years of returns in our simulations—mean–variance optimization does overtake the even split, and with covariance shrinkage it overtakes it by a wide margin that grows with the number of assets. Until that threshold is met, though, the sophisticated allocation rule is not merely unhelpful; it is actively harmful relative to the simple rule that assumes little about the market and thus runs less risk of assuming too much.

Reproducibility.

All code, random seeds, and figures are available in the accompanying repository. The simulations take only a few minutes to run on a laptop using only the numpy, scipy, scikit-learn, and matplotlib packages. A yfinance-based data loader is provided so that the identical pipeline can be run on live market data.

The Ledoit–Wolf constant-correlation estimator

Let S be the sample covariance of M demeaned return vectors $x_t \in R^N$, with entries s_{ij} and sample variances s_{ii}. The constant-correlation target F replaces every pairwise correlation by the average sample correlation r^-:

$$f_{ii} = s_{ii}, \qquad f_{ij} = r^- \sqrt{s_{ii} s_{jj}} \ (i \neq j), \qquad r^- = \frac{2}{N(N-1)} \sum_{i<j} \frac{s_{ij}}{\sqrt{s_{ii} s_{jj}}}.$$

The shrinkage estimator is $\hat{\Sigma}_{LW} = \delta F + (1 - \delta)S$ with optimal intensity $\delta = max\{0, min\{1, \hat{\kappa}/M\}\}$ and $\hat{\kappa} = \left(\hat{\pi} - \hat{\rho}\right)/\hat{\gamma}$, where $\hat{\pi} = \sum_{ij} \hat{Asy}\hat{V}ar\left(s_{ij}\right)$ measures the noise in S, $\hat{\gamma} = \sum_{ij} \left(f_{ij} - s_{ij}\right)^2$ measures the misfit of the target, and $\hat{\rho}$ corrects for the covariance between S and F (Ledoit and Wolf, 2004). Intuitively, δ is large when the sample covariance is noisy (small M) and the target fits well, and small otherwise. In Experiment A the average estimated intensity was $\delta \approx 0.48$.

References

Best, M. J., and Grauer, R. R. (1991). On the sensitivity of mean–variance-efficient portfolios to changes in asset means. *Review of Financial Studies*, 4(2), 315–342.

Chopra, V. K., and Ziemba, W. T. (1993). The effect of errors in means, variances, and covariances on optimal portfolio choice. *Journal of Portfolio Management*, 19(2), 6–11.

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

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

Jorion, P. (1986). Bayes–Stein estimation for portfolio analysis. *Journal of Financial and Quantitative Analysis*, 21(3), 279–292.

Ledoit, O., and Wolf, M. (2004). Honey, I shrunk the sample covariance matrix. *Journal of Portfolio Management*, 30(4), 110–119.

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

Merton, R. C. (1980). On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics*, 8(4), 323–361.

Michaud, R. O. (1989). The Markowitz optimization enigma: Is ‘optimized’ optimal? *Financial Analysts Journal*, 45(1), 31–42.